主管单位：中华人民共和国工业和信息化部
主办单位：北京航空航天大学

ISSN 1001-5965
CN 11-2625/V

北京航空航天大學学报

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS



飞机地面结霜机理及分阶段预测模型

陈 斌　朱庆民　庄 淇　王立文

5 2026
Vol.52

北京航空航天大学学报

第 52 卷　第 5 期　（总第 399 期）　2026 年 5 月

目　　次

飞机地面结霜机理及分阶段预测模型 …… 陈斌，朱庆民，庄淇，王立文（1355）

双注意力引导的 U-Net++遥感图像语义分割模型 …… 刘春娟，辛钰强，吴小所，闫浩文（1366）

主动侧杆系统多余力机理及敏感性分析 …… 杨建忠，左金鹏，孙晓哲（1378）

基于多尺度特征融合的滚动轴承寿命预测 …… 火久元，李昕，常琛，李宇峰，张耀南（1391）

基于活动链分析的高校学生行为模式研究 …… 翟宜凯，蒋晓桐，田琼，黄海军（1406）

基于感知增强与多尺度特征融合的小目标车辆检测 ……
…… 沈瑜，李阳阳，李博昊，高宝渠，魏子易，白珊（1422）

基于自适应复合卷积的航拍小目标检测算法 …… 邓天民，余洋，陈月田，谢鹏飞（1433）

基于特征信息补充与增强的无人机视角图像目标检测算法 …… 邬开俊，蒲卓（1445）

基于深度学习的固体姿轨控发动机性能预示方法 …… 杨慧欣，王旭，李响（1456）

基于流体单位质量力的飞机燃油测量算法 …… 王晓阳，任恒英，饶波（1467）

加速栅小孔对变推力离子推力器性能影响的实验研究 ……
…… 胡竟，郭德洲，耿海，杨福全，李建鹏，陈娟娟（1476）

航天员舱外维修操作时间的量化评估 …… 刁常堃，王昊，孟令子，苏南，高鸿，邓忠民（1487）

考虑擦机尾动态限制角的起飞抬轮过程优化 …… 庄南剑，杨雪雅，谷润平（1496）

绿色低碳导向的航班推出策略与牵引作业计划联合优化 …… 寇伟彬，张世杰，王佳玉（1504）

模拟与模数混合芯片宽带传导抗扰度量化仿真建模方法 …… 陈曦，谢树果，魏梦圆，李圆圆（1513）

基于深度特征修正的目标跟踪算法 …… 陈凯，黄煜杰，赵晓冬，王鹏飞，陈凯，林岩泽（1523）

基于图像语义分割的目标跟踪尺度自适应算法 ……
…… 陈凯，赵晓冬，黄煜杰，王鹏飞，雷一辰，张彪（1536）

触爆载荷作用下的混凝土重力坝动态响应规律 …… 张克钒，彭永，李翔宇，卢芳云，王哲（1547）

航天 AFRP 的紫外飞秒激光加工特性与均质化去除方法 ……
…… 路明雨，张明，卫宇璇，李波，崔志刚，张开虎（1557）

一种高超飞行器用 CFD 网格不确定度分析方法 …… 郭文娟，李强，周岭（1567）

高隐身飞翼布局翼型敏感性分析及优化设计 …… 张伟，周琳，陈宪，舒博文，黄江涛，高正红（1578）

基于点云分割算法的战术导弹表面压力分布快速预测 …… 蔺佳哲，何磊，程明，周岭，杨春明（1587）

木卫四探测任务转移轨道状态序列搜索与优化 …… 井泉，李明涛，王有亮（1596）

基于风险概率分解的 RTK 完好性评估方法 ……
…… 宋远，黄智刚，李锐，王岳辰，沈军，王永超，聂欣（1605）

基于分布式安装 MSCSG 的挠性航天器姿态振动测控方法 …… 李磊，王卫杰，庞维坤，王丽芬，段乐强，王宸宇（1615）
基于空时谱熵的分布式阵列雷达同步误差估计方法 …… 耿雪胤，王俊，杨彬，孙进平（1627）
基于多种微结构的机翼结构/材料多尺度气动弹性优化 …… 黎珂宇，杨超，王晓喆，万志强，李畅（1635）
基于数字孪生的健康服役框架及其在永磁同步电机的应用 …… 郭浩宇，王少萍，石健，张育玮（1647）
轰炸机挂飞导弹设计突风载荷谱编制方法 …… 卫昆钰，李晨迪，李博文，袁愿，贺小帆（1657）
航空柱塞泵缸体倾覆行为仿真及试验研究 …… 谷建宁，张小波，夏天翔，王天照（1666）
一维阵列多液滴蒸发模型及实验 …… 韩琪炜，金捷，赖根鸿，黄俊琦，王方（1680）
基于平方马氏距离的空间目标碰撞风险分析算法 …… 黄梦蝶，王禄丰，黄旭星，李爽（1691）
基于三维格网误差建模的 GNSS/IMU 抗差自适应定位算法 …… 盛琪，孙蕊，何雨霖，张珩瑜（1701）
考虑多环节间隙非线性的电动舵机多学科耦合动力学建模 …… 杨石琳，乔凯俐，万晓忠，段绵超，郭宁，徐超（1712）
飞翼布局飞机颤振主动抑制的多输入/多输出控制律设计 …… 冯宇轩，霍应元，李俊杰（1720）
吊舱组合导航系统弹性杆臂误差补偿方法 …… 王海鹏，刘海娜，王振东，缪存孝，张贺，叶文（1728）
机载中空纤维膜分离特性的预测分析 …… 柴婷，付子祺，刘冠男，白文涛，胡万俊，冯诗愚（1739）
基于桥臂共享型多电平变换器的 SRM 瞬时转矩控制 …… 陈凡强，李存贺，杜钦君，柳健，焦提操（1746）
基于卷积注意力与特征融合的火灾检测算法 …… 田佳麒，秦国轩，张为（1756）
考虑动态重构机制的 IMA 可靠性建模和评估方法 …… 胡耕硕，焦健，胡朗霄，景永峰（1767）

期刊基本参数：CN 11-2625/V * 1956 * m * A4 * 422 * zh * P * ¥50.00 * 400 * 40 * 2026-05

（**编　辑**　张　嵘　孙　芳　万玉敏　卞欢欢　王　茜　李亚泰）

JOURNAL OF BEIJING UNIVERSITY OF
AERONAUTICS AND ASTRONAUTICS

Vol. 52 No. 5 （Sum 399） May 2026

CONTENTS

Mechanism of aircraft ground frosting and staged prediction model
…… CHEN Bin，ZHU Qingmin，ZHUANG Qi，WANG Liwen （1355）
Semantic segmentation model for remote sensing images based on U-Net++ guided by dual attention
…… LIU Chunjuan，XIN Yuqiang，WU Xiaosuo，YAN Haowen （1366）
Mechanism and sensitivity analysis of surplus force in active sidestick system
…… YANG Jianzhong，ZUO Jinpeng，SUN Xiaozhe （1378）
Rolling bearing life prediction based on multi-scale feature fusion
…… HUO Jiuyuan，LI Xin，CHANG Chen，LI Yufeng，ZHANG Yaonan （1391）
Study of university students' behavioral patterns via activity chain analysis
…… ZHAI Yikai，JIANG Xiaotong，TIAN Qiong，HUANG Haijun （1406）
Small target vehicle detection based on perceptual enhancement and multi scale feature fusion
…… SHEN Yu，LI Yangyang，LI Bohao，GAO Baoqu，WEI Ziyi，BAI Shan （1422）
Small object detection algorithm for aerial photography based on adaptive compound convolution
…… DENG Tianmin，YU Yang，CHEN Yuetian，XIE Pengfei （1433）
Object detection algorithm for UAV viewpoint images based on feature information complementation and enhancement
…… WU Kaijun，PU Zhuo （1445）
A prediction method for solid divert and attitude control motor performance based on deep learning
…… YANG Huixin，WANG Xu，LI Xiang （1456）
Aircraft fuel quantity measurement algorithm based on fluid unit mass force
…… WANG Xiaoyang，REN Hengying，RAO Bo （1467）
Experimental study on effect of aperture of accelerator grid on performance of variable thrust ion thruster
…… HU Jing，GUO Dezhou，GENG Hai，YANG Fuquan，LI Jianpeng，CHEN Juanjuan （1476）
Quantitative assessment of extravehicular maintenance operation time for astronauts
…… DIAO Changkun，WANG Hao，MENG Lingzi，SU Nan，GAO Hong，DENG Zhongmin （1487）
Optimization of take-off rotation process considering tail striking dynamic limit angle
…… ZHUANG Nanjian，YANG Xueya，GU Runping （1496）
Joint optimization of green low-carbon oriented flight launch strategy and towing operation plan
…… KOU Weibin，ZHANG Shijie，WANG Jiayu （1504）
Simulation modeling methodology for broadband conducted immunity quantization of analog and analog-digital hybrid chips
…… CHEN Xi，XIE Shuguo，WEI Mengyuan，LI Yuanyuan （1513）
Object tracking algorithm based on deep feature modification
…… CHEN Kai，HUANG Yujie，ZHAO Xiaodong，WANG Pengfei，CHEN Kai，LIN Yanze （1523）
Adaptive algorithm for target tracking scale based on image semantic segmentation
…… CHEN Kai，ZHAO Xiaodong，HUANG Yujie，WANG Pengfei，LEI Yichen，ZHANG Biao （1536）
Dynamic response law of concrete gravity dam under contact explosion load
…… ZHANG Kefan，PENG Yong，LI Xiangyu，LU Fangyun，WANG Zhe （1547）
Characteristics and homogenization removal methods of ultraviolet femtosecond laser processing of aerospace AFRP
…… LU Mingyu，ZHANG Ming，WEI Yuxuan，LI Bo，CUI Zhigang，ZHANG Kaihu （1557）
A CFD grid uncertainty analysis method for hypersonic aircraft
…… GUO Wenjuan，LI Qiang，ZHOU Ling （1567）
Sensitivity analysis and optimization design of high stealth flying wing layout airfoil
…… ZHANG Wei，ZHOU Lin，CHEN Xian，SHU Bowen，HUANG Jiangtao，GAO Zhenghong （1578）
Rapid prediction of surface pressure distribution of tactical missile based on point cloud segmentation algorithm
…… LIN Jiazhe，HE Lei，CHENG Ming，ZHOU Ling，YANG Chunming （1587）
Sequence search and optimization for transfer trajectory status in Callisto exploration mission
…… JING Quan，LI Mingtao，WANG Youliang （1596）
RTK integrity evaluation method based on risk probability decomposition
…… SONG Yuan，HUANG Zhigang，LI Rui，WANG Yuechen，SHEN Jun，WANG Yongchao，NIE Xin （1605）

Attitude measurement and control method of flexible spacecraft based on distributed installation MSCSG
…… LI Lei，WANG Weijie，PANG Weikun，WANG Lifen，DUAN Leqiang，WANG Chenyu （1615）
Space-time spectral entropy based synchronization error estimation method for distributed array radar
…… GENG Xueyin，WANG Jun，YANG Bin，SUN Jinping （1627）
Aeroelastic optimization of wing structure and material using multiple microstructures
…… LI Keyu，YANG Chao，WANG Xiaozhe，WAN Zhiqiang，LI Chang （1635）
Digital twin-based health service framework and its application in permanent magnet synchronous motors
…… GUO Haoyu，WANG Shaoping，SHI Jian，ZHANG Yuwei （1647）
Developing of tornado missile design gust load spectrum for bomber-mounted air-to-ground missiles
…… WEI Kunyu，LI Chendi，LI Bowen，YUAN Yuan，HE Xiaofan （1657）
Simulation and experimental research for overturning behavior of aviation plunger pump cylinder bodies
…… GU Jianning，ZHANG Xiaobo，XIA Tianxiang，WANG Tianzhao （1666）
Experimental study of multi-droplet evaporation model of one-dimensional array
…… HAN Qiwei，JIN Jie，LAI Genhong，HUANG Junqi，WANG Fang （1680）
Space target collision risk analysis algorithm based on the square Mahalanobis distance
…… HUANG Mengdie，WANG Lufeng，HUANG Xuxing，LI Shuang （1691）
A robust adaptive positioning algorithm for GNSS/IMU based on 3D grid error modeling
…… SHENG Qi，SUN Rui，HE Yulin，ZHANG Hengyu （1701）
Multidisciplinary coupled dynamics modeling of electro-mechanical actuator considering nonlinear clearances in multiple stages
…… YANG Shilin，QIAO Kaili，WAN Xiaozhong，DUAN Mianchao，GUO Ning，XU Chao （1712）
Design of multiple-input/multiple-output control law for active flutter suppression of flying-wing aircraft
…… FENG Yuxuan，HUO Yingyuan，LI Junjie （1720）
Pod integrated navigation system elastic lever arm error compensation method
…… WANG Haipeng，LIU Haina，WANG Zhendong，MIAO Cunxiao，ZHANG He，YE Wen （1728）
Prediction and analysis of separation performance of on-board hollow fiber membrane
…… CHAI Ting，FU Ziqi，LIU Guannan，BAI Wentao，HU Wanjun，FENG Shiyu （1739）
Instantaneous torque control of SRM based on bridge arm shared multilevel converter
…… CHEN Fanqiang，LI Cunhe，DU Qinjun，LIU Jian，JIAO Ticao （1746）
Fire-and-smoke detection algorithm based on convolutional attention and feature fusion
…… TIAN Jiaqi，QIN Guoxuan，ZHANG Wei （1756）
Reliability modeling and evaluation method of IMA under dynamic reconfiguration
…… HU Gengshuo，JIAO Jian，HU Langxiao，JING Yongfeng （1767）

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0109

飞机地面结霜机理及分阶段预测模型

陈斌[1,2,*]，朱庆民[1,2]，庄淇[1,2]，王立文[2,3]

(1. 中国民航大学 电子信息与自动化学院，天津 300300；2. 中国民航大学 航空地面特种设备研究基地，天津 300300；
3. 中国民航大学 科技创新研究院，天津 300300)

摘　　要：飞机地面结霜影响飞行安全，造成航班延误，精准的结霜预测是机场除冰雪参数生成的重要依据。飞机地面结霜受气象、机体表面结构和航班时刻等诸多因素影响，且机场对飞机结霜预报指标特殊。为此，将飞机地面结霜分为液核形成阶段和霜层增长阶段进行研究。在液核形成阶段，基于经典成核理论建立了结霜温度模型，将冻结液核覆盖率达到阈值时的温度定义为结霜温度，研究了飞机表面特性及外部环境对结霜温度的影响情况；在霜层增长阶段，基于欧拉多相流建立了霜层增长模型，从霜层的生成原理入手，加入了霜层生长的识别判据，研究了不同工况环境对飞机表面霜层特性的影响。多种工况仿真和现场实验结果表明：结霜温度模型的平均误差为0.65 ℃；霜层增长模型的平均误差为5.99%。研究结果表明所提的分阶段建模方法可为预测飞机地面结霜问题提供理论支持。

关 键 词：成核；结霜温度；霜层增长；数值模拟；预测
中图分类号：V219; TB61^{+}1
文献标志码：A　　**文章编号**：1001-5965(2026)05-1355-11

含湿量高的空气在接触到低于冰点温度的固体表面时会发生相变，造成固体表面结霜。冬季，停靠在机场的飞机表面饱受结霜现象的困扰，除北方机场外，合肥新桥国际机场和地处亚热带的南宁吴圩国际机场均发生过飞机表面结霜现象，可见霜层形成的环境与气象条件复杂多变。当结霜发生后，机场地勤人员仅能通过现场观察的方式来判断飞机表面是否结霜，缺乏对结霜的预报功能。因此，为减少机场人工和除冰资源成本的投入，保障航班的正常运行，对飞机地面结霜的准确预测显得尤为重要。

结霜现象分为液核形成阶段和霜层增长阶段，包括3个时期：霜晶形成期、霜层生长期、充分生长期[1-3]。多数情况下，霜层的形成和扩散通常始于表面上液核的形成。液核的形成不仅是结霜现象的初始阶段，同时也对随后霜层的发展产生着至关重要的影响。在液核形成阶段，空气中的水蒸气在冷表面附近发生相变凝结成核。Chaudhary等[4]对液滴成核的研究发现，过冷水滴的成核发生在某一点后，迅速扩展到整个水滴。Matsumoto等[5]利用理论公式对水分子中的氢键自发形成初始冰核进行了模拟。Huang等[6]通过对比实验研究了接触角大小对液滴冻结时间的影响。Bucher等[7]研究了不同粗糙度表面对液滴的接触影响。陆晨律等[8]通过实验手段研究了过冷水与表面的接触面积对积冰成核的影响。Na[9]和Piucco[10]等研究了平板表面上霜成核的过程，并探究了外界环境因素对成核的影响。

当冻结液核在冷表面上稳定形成后，空气中处于过饱和状态的水蒸气在冻结液核上相变结霜，纵向生长，此时结霜过程演变进入下一个阶段，即霜层增长阶段。针对这一阶段，众多专家学者建立了霜层生长数学模型来预测霜层的变化。Lee等[11]通

收稿日期：2024-02-28；录用日期：2024-10-25；网络出版时间：2025-02-26 16:51
网络出版地址：link.cnki.net/urlid/11.2625.V.20250226.1048.001
基金项目：国家自然科学基金面上项目(52472460)
*通信作者. E-mail：chenbindavid@163.com

引用格式：陈斌，朱庆民，庄淇，等. 飞机地面结霜机理及分阶段预测模型[J]. 北京航空航天大学学报，2026，52(5)：1355-1365.
CHEN B，ZHU Q M，ZHUANG Q，et al. Mechanism of aircraft ground frosting and staged prediction model[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1355-1365 (in Chinese).

过对湿空气中水蒸气的分子扩散进行研究，计算了霜层的增长速率。Xu等[12]通过对结霜过程中传质传热机理进行分析，建立了霜层生长预测模型。Hermes等[13-14]将霜层表示为湿空气和冰组成的多孔介质，并研究了建立的霜层生长模型随外界环境变化的规律。Lenic等[15-16]将计算区域划分为湿空气区域与霜层区域，并分别计算，模拟了翅片管换热器上的霜层生长。Kim等[17]将水蒸气的浓度梯度作为相变的驱动力，基于欧拉多相流建立了霜层生长数学模型。赵泽扬等[18]通过实验和数值模拟的手段研究了倾斜表面上霜层的生长过程。

综上，众多专家学者对结霜过程进行了深入的研究。在液核形成阶段，研究人员对液核的形成进行了细致的分析，但缺少对机翼表面结霜温度的关注，利用该参数可以有助于结霜预报结果，这是机场及航空公司最关注的问题；在霜层增长阶段，学者给出了关于霜层生长过程的建模思路，并将其应用于冷表面及翅片管换热器等平面区域，但对于飞机表面来说，存在不同的流体构型曲面，导致了不同曲率下霜层物性不同。基于众多研究中的不足之处，本文以地面停驻的飞机为研究背景，以结霜现象的演变过程为研究对象，分阶段建立结霜预测模型。针对目前民航领域缺乏的结霜预报功能，探究地面飞机结霜温度起始点及霜层的生长趋势，为飞机地面结霜预测提供理论支持。

1 飞机地面结霜问题描述

飞机表面的结霜与其他物体结霜现象不同，具有大面积和复杂性的特点。飞机地面结霜示意图如图1所示。过夜航班表面的结霜是机场内最普遍的结霜现象。同时，冷浸燃油霜冻也是机场内频发的现象，是指飞机在高空中飞行，被寒冷气流浸泡过的油箱在落地后仍保持极低的温度，因内外温差较大，机翼上表面极易发生结霜。飞机在停机坪停靠的位置也会导致结霜存在差异性。例如，在空旷地带停靠的飞机与停靠在航站楼附近的飞机受到的风场不同，导致2架飞机在开始结霜时间和结霜厚度上存在很大区别。

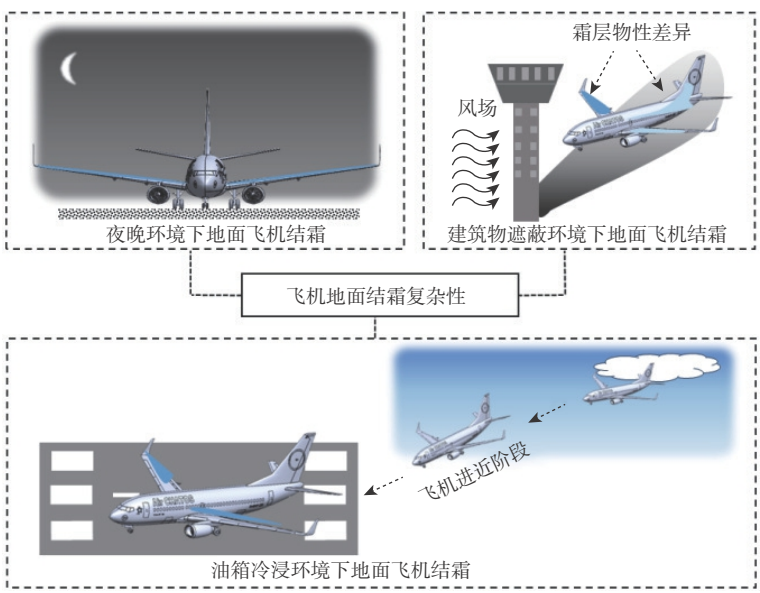


图1 飞机地面结霜复杂性

Fig. 1 Aircraft ground frost complexity

中国民用航空局发布的《民航冬季安全运行指导意见》要求机场监测和预报飞机结霜情况，以减少航班延误和安全风险[19]。在这一过程中，机场工作人员需要提前掌握结霜温度和结霜厚度等关键信息。精准预测结霜温度可以帮助机场及相关部门提高飞机地面结霜的预报效果，结霜厚度的精准预测有助于机场对结霜严重等级进行评估，为飞机离港前的除霜工作做好充分参数决策。为弥补机场及航空公司对结霜关键信息预报方面的不足，依据所建立的2个模型完成对整个结霜过程的预测，帮助机场及航空公司获取飞机地面结霜的关键信息。

2 模型建立

2.1 建模思路

模型建立思路如图2所示。

针对结霜演变第1阶段，从微观角度入手，基于经典成核理论求解水蒸气冷凝成核过程中的成核率和半径生长速率，并最终求解冻结液核在冷表面的覆盖率。将覆盖率达到覆盖阈值时的温度定义为结霜的起始温度，即为模型预测的结霜温度。

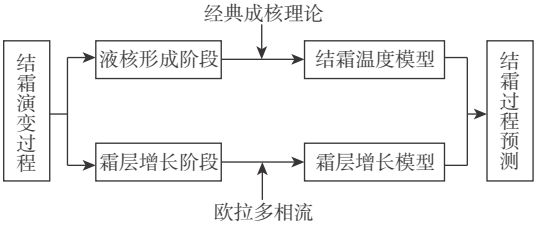


图 2　建模思路示意图

Fig. 2　Modeling idea diagram

针对结霜演变第 2 阶段, 从宏观角度入手, 基于欧拉多相流模型推导水蒸气相变成霜过程中的质量传递等控制方程, 并求解飞机表面霜层的物性参数, 最终得到霜层厚度的预测模型。

2.2　结霜温度模型

2.2.1　成核理论

空气中处于过饱和状态的水蒸气在飞机表面冷凝成核, 同时随着温度的降低, 水分子逐渐冻结并生长形成枝状或柱状的冰晶体, 这是霜层形成的主要基础, 并且是影响整个结霜过程的关键因素。通过对该过程中成核自由能、液核形成率、液核半径生长速率的计算, 推导出用于预测结霜温度的覆盖率。

1) 成核自由能。成核过程中, 体系的体积自由能减少和表面自由能增加之间存在着竞争关系, 这种竞争关系如图 3 所示。

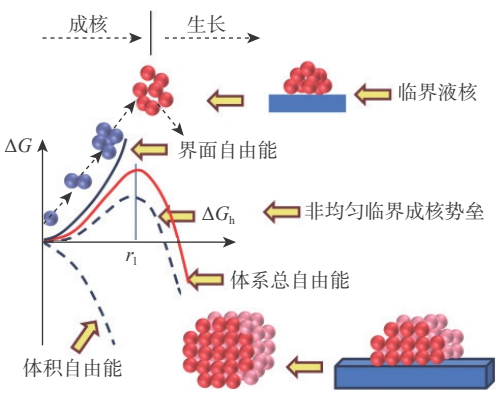


图 3　成核过程自由能变化

Fig. 3　Free energy changes during nucleation process

图 3 反映了成核过程中体系自由能的变化。从自由能角度分析, 液核的形成过程中总是存在着界面自由能的增加和体积自由能的减少, 只有当生成液核的体系总自由能大于临界成核功ΔG^*时, 液核才能生成并生长; 从成核半径角度分析, 只有当液核的半径大于临界半径r_l时, 液核才能生成并生长。临界成核功ΔG^*表征了相变过程中必须克服的能量势垒, 即形成一个能够稳定生长的临界尺寸液核所需的最小功。临界成核功和临界半径表示为

$$r_l = \frac{2\gamma_{qw} v_w}{\Delta\mu} \tag{1}$$

$$\Delta G^* = \frac{4\pi\gamma_{qw}}{3}\left(\frac{2\gamma_{qw} v_w}{\Delta\mu}\right)^2 \tag{2}$$

式中: r_l为液核的临界半径; γ_{qw}为液核的界面自由能; v_w为液核的摩尔体积; $\Delta\mu$为水蒸气凝结相变过程的自由能差值, 表示为[10]

$$\Delta\mu = \frac{h_c(T_l - T_w)}{T_l} + v_l(p_w - p_s(T_l)) - R_0 T_w \ln\left(\frac{p_s(T_s)}{p_s(T_l)}\right) - R_0 T_l \ln s_d \tag{3}$$

其中: h_c为单位质量内水蒸气相变换热产生的热量; T_l为露点温度; T_w为水的温度; v_l为液核的比体积; p_w为液核的压力; $p_s(T_l)$为露点温度下空气的饱和水蒸气分压; $p_s(T_s)$为当前表面温度下空气的饱和水蒸气分压; R_0为理想气体常数; s_d为湿度。

2) 成核率及半径生长速率。根据经典成核理论可知, 自然界中的成核方式分为均匀成核和非均匀成核。本文认为在飞机表面上的成核行为是非均匀成核的结果, 故其成核功可表示为

$$\Delta G_h = f(\theta)\Delta G^* \tag{4}$$

式中: ΔG_h为飞机表面成核的成核功; $f(\theta)$为关于接触角的函数, 可以表示为

$$f(\theta) = \left(2 + \cos^3\theta_{WA} - 3\cos\theta_{WA}\right)/4 \tag{5}$$

其中: θ_{WA}为冻结液核与飞机表面的接触角。

航班时间会影响地面飞机机翼表面温度, 加速结霜过程。飞机表面的液核成核率可以表示为

$$J_h = \frac{N_a k T_a}{h}\exp\left(-\frac{E_w}{N_A k T_s}\right)\exp\left(-\frac{\Delta G_h}{k T_s}\right) \tag{6}$$

式中: N_a为液核的分子数密度; k为玻耳兹曼常数; T_a为飞机表面温度; h为普朗克常数; E_w为液核的扩散激活能; N_A为阿伏伽德罗常数。

因此, 可得出飞机表面的冻结液核半径生长速率V_{Aiw}[11]为

$$V_{Aiw} = V_{iw}\sin\theta_{WA} \tag{7}$$

式中: V_{iw}为理想状态下的冻结液核半径生长速率。该速率是基于热平衡原理推导的, 即假定液核冻结过程中释放的相变潜热完全通过热传导和对流传热散发到周围环境中。考虑液核的球冠状几何特征及气液固三相界面的热阻影响, 可表示为

$$V_{iw} = \frac{(T_s - T_a)(r - r_l)}{2r\rho_w L_m}\left[\frac{r\theta_{WA}(1-\cos\theta_{WA})}{4k_w\sin\theta_{WA}} + \frac{1-\cos\theta_{WA}}{\sin^2\theta_{WA}} + \frac{1}{2h_a}\right] \tag{8}$$

式中: T_s为飞机周围空气温度; ρ_w为液相的密度; L_m为凝结相变的换热系数; h_a为相间传递的传热系数; r为液核的半径; k_w为冻结液核的热导率。

2.2.2 结霜温度求解

当液核在飞机表面稳定生成后，越来越多的液核随之冻结生长扩张，形成霜层的基本骨架，如图4所示。表面冻结液核扩张覆盖达到阈值时的温度为结霜开始时的温度。

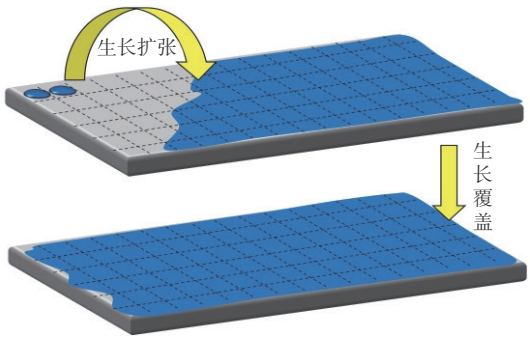


图4 液核覆盖示意图

Fig. 4 Liquid core covering diagram

定义飞机表面的面积为 S，冻结液核在 t=0 时开始形成并生长，则在 t_1 到 t_1+dt 这段时间内新生成的液核，生长至时刻t时所覆盖的面积增量dS_i可表示为“单核面积”与“新核数量”的乘积，即

$$\mathrm{d}S_i=\pi[r_0+V_{Aiw}(t-t_1)]^2J_hS_w(t_1)\mathrm{d}t \tag{9}$$

式中：S_i为经过冻结液核覆盖的面积；S_w为未经过冻结液核覆盖的面积；r_0为初始半径。

当空气温度在−30～20 ℃、风速在0.2～5 m/s范围内时，冻结液核的半径约为10^{-8}数量级，冻结液核的表面积可忽略不计[20]。则冻结液核的覆盖率可表示为

$$\varphi=1-\frac{S_w}{S}=\frac{S_i}{S} \tag{10}$$

将式(10)代入式(9)，忽略冻结液核的底面半径，同时对等式量变并进行积分，得到飞机表面冻结液核的覆盖率为

$$\varphi=1-\exp\left(-\frac{\pi J_hV_{Aiw}^2t^3}{3}\right) \tag{11}$$

式(6)和式(8)中求得的液核成核率和冻结液核半径生长速率中均含有飞机表面温度。将冻结液核覆盖率达到覆盖阈值时的温度定义为结霜起始温度。实际应用中，受环境因素和表面特性的影响，霜层覆盖率很难达到1，因此，定义覆盖阈值为0.999，则模型所预测的结霜温度可通过式(12)进行求解：

$$\varphi=1-\exp\left(-\frac{\pi J_h(T_i)V_{Aiw}^2(T_i)t^3}{3}\right)=0.999 \tag{12}$$

式中：T_i为模型预测的结霜温度，其存在于成核率和半径生长速率中。通过对模型中结霜温度的求解，可以得到结霜的起始点，方便机场人员对是否结霜进行判断。

2.3 霜层增长模型

当冻结的液核在飞机表面覆盖完成后，结霜过程进入霜层增长阶段。空气中的水蒸气通过2种方式附着于表面上，使霜层得到充分生长，如图5所示。一种方式是当湿空气中的水蒸气在飞机表面附近活动时，过饱和的一部分接触到表面转化为霜晶，具体表现为霜层厚度的增长；另一种方式是霜层中水蒸气的浓度低于外部水蒸气，导致外部水蒸气通过分子之间的扩散运动进入霜层内部，在接触低温环境后凝华成霜晶，具体表现为霜层密度的增加。

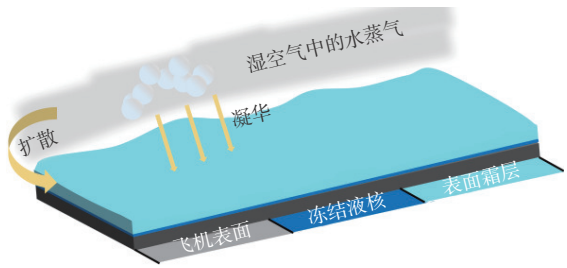


图5 霜层增长示意图

Fig. 5 Frost layer growth diagram

2.3.1 霜层增长控制方程

霜层增长作为宏观意义上的增长过程，涉及气态到固态的相变，选用以欧拉多相流为基础所建立的霜层增长模型来进行厚度预测。在飞机表面和湿空气附近两相混合，定义首相为湿空气，由干空气和水蒸气组成；副相为霜。该模型只针对霜层生长过程中厚度变化进行研究，因此，依次对两相之间通过组分输运而引起的体积分数变化进行求解。

1）体积分数守恒方程及源项。体积分数守恒方程对模拟多相流中的物质传递、相互作用和相变等过程至关重要。在欧拉多相流模型中，各相的体积分数是空间和时间的函数。各相都被视为相互联通的连续体[17]，每一项都用体积分数来表示。定义α为体积分数，α_d为首相体积分数，α_i为副相体积分数，则有

$$\alpha_d+\alpha_i=1 \tag{13}$$

当α_i=0时，表示网格内均为空气；当α_i=1时，表示网格内均为霜；当0<α_i<1时，表示网格内两相共存。霜层在生长过程中可能会产生因为气温波动导致的霜层升华，但霜层升华转移的质量很少。为简化模型计算，本文将霜层升华引起的质量转移定义为0，则副相的体积分数表示为

$$\frac{\partial\alpha_i}{\partial t}+\nabla\cdot(\alpha_i\boldsymbol{u}_i)=S_{\alpha i} \tag{14}$$

式中: ∇为向量微分算子; $\boldsymbol{u}_i$为副相的速度; $S_{\alpha i}$为副相体积源项, 其通过式(15)给出:

$$S_{\alpha i}=\frac{\dot{m}_{d\to i}}{\rho_i} \tag{15}$$

式中: $\dot{m}_{d\to i}$为首相中的水蒸气向副相转移的质量; ρ_i为副相的密度。

2) 质量守恒方程及源项。在欧拉多相流中, 质量守恒方程是描述多相流中各相质量分布随时间和空间变化的方程。在本文所建模型质量转移中, 仅考虑首相向副相的质量转移, 忽略因为霜层升华和融化损失的质量。

首相质量守恒方程为

$$\frac{\partial(\alpha_d\rho_d)}{\partial t}+\nabla\cdot(\alpha_d\rho_d\boldsymbol{u}_d)=S_{md} \tag{16}$$

副相质量守恒方程为

$$\frac{\partial(\alpha_i\rho_i)}{\partial t}+\nabla\cdot(\alpha_i\rho_i\boldsymbol{u}_i)=S_{mi} \tag{17}$$

式中: ρ_d和$\boldsymbol{u}_d$分别为首相的密度和速度; S_m为相变产生的质量源项, 由式(18)和式(19)给出:

$$S_{mi}=\dot{m}_{d\to i} \tag{18}$$

$$S_{md}=-\dot{m}_{d\to i} \tag{19}$$

其中: $\dot{m}_{i\to d}$为副相向首相的质量转移。

3) 组分输运守恒方程及源项。在欧拉多相流中, 组分输运守恒方程描述了多相流中各组分浓度的演化。首相到副相的质量传递实质上是湿空气中的水蒸气组分到霜层中的传递, 需要考虑湿空气中水蒸气的组分输运方程:

$$\frac{\partial(\alpha_d\rho_dF_{vd})}{\partial t}+\nabla\cdot(\alpha_d\rho_dF_{vd}\boldsymbol{u}_d)=\nabla\cdot(\rho_dD_H\nabla F_{vd})+S_{Fvd} \tag{20}$$

式中: F_{vd}为首相中水蒸气的质量分数; D_H为水蒸气扩散系数; S_{Fvd}为传质引起的组分源项, 表示为

$$S_{Fvd}=-\dot{m}_{d\to i} \tag{21}$$

2.3.2　霜层增长计算方法

1) 霜层增长阈值。当霜层增长控制方程给定后, 飞机附近的水蒸气在方程的控制下开始发生相变结霜。为准确计算出网格内产生的霜层, 根据结霜现象发生原理, 定义了合理的霜层增长阈值, 为霜层增长提供合理的识别判据。

首先, 定义了水蒸气向霜层转移的质量, 认为其与水蒸气分压和表面温度对应下的饱和水蒸气压力之差有关, 表示为

$$\dot{m}_{d\to i}=\tau_v\rho_e(F_{vd}-F_s(T)) \tag{22}$$

式中: ρ_e为湿空气的有效密度; τ_v为凝华时间的松弛系数; $F_s(T)$为单元网格温度T下对应的饱和水蒸气质量分数。

最后, 根据前文提出的假设, 忽略霜的升华现象, 对结霜过程中两相之间的质量传递, 将副相速度设置为 0。随着副相体积分数的增加, 黏附于飞机表面的霜层逐渐生成。当霜层增长到一定高度时, 就会影响飞机表面附近的流场$S_{mi}=\dot{m}_{d\to i}$, 阻碍空气的流动。当在流场空间中对质量传递不加以区分和限制时, 在计算过程中, 速度场迅速被改变, 偏离实际过程, 导致计算发散。因此, 对于霜层的生长计算, 需要加入合理的识别判据, 由式(23)给出[16]:

$$\dot{m}_{d\to i}=\begin{cases}\tau_v\rho_e(F_{vd}-F_s(T)) & T\leqslant\min(T_1,273.15K)\\ 0 & T>\min(T_1,273.15K)\end{cases} \tag{23}$$

式 (23) 中规定了 2 种情况: ①由水蒸气压力差造成质量转移; ②不存在质量转移。

图 6 为结霜计算区域划分示意图。根据本文所建立的霜层增长识别判据, A 区域为霜层内部区域, 霜层内部温度低于露点温度, 因此, 根据判据其符合情况①; B 区域为霜层表面区域, 霜层表面温度低于露点温度, 因此, 根据判据也符合情况①; C 区域为湿空气区域, 远离霜层的空气温度迅速增加, 因此, 根据判据其符合情况②, 即不产生质量转移。

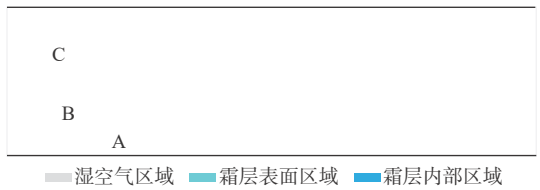


图 6　结霜计算区域划分

Fig. 6　Frost calculation area division

2) 霜层厚度计算。为对模型进行验证, 需求解模型计算后所产生霜层的厚度, 其能够最直接地体现出结霜载体表面的霜层变化。采用加权平均的方式计算出霜层厚度, 可表示为

$$\delta_f=\frac{\sum_{j=1}^{n}Y_j}{n} \tag{24}$$

式中: Y_j为霜层计算区域的纵坐标, 即某一点的霜层厚度; n为霜层计算区域内纵坐标的个数。可以通过测算结霜载体基准面法线方向上α_i>10^{-6}的网格数量得到霜层计算区域的纵坐标, 通过计算平均坐标值得到该计算区域的霜层厚度。

3　数值模拟

不同工况环境下, 飞机表面的结霜特性存在差

异，因此，通过改变模型参数来分析成核及结霜温度的变化，通过改变模拟环境来研究霜层增长变化。

3.1 结霜温度数值模拟

3.1.1 模型参数选取

选用2024铝板来模拟飞机表面，因为液核与飞机表面之间张力的存在，其多呈现半球形，所以，对于接触角的计算，可以先通过图像测量的方式确定液核与表面之间的表观接触角，再运用Thomas[21]提出的公式计算出冻结液核与表面之间的接触角。表观接触角的计算如图7所示。

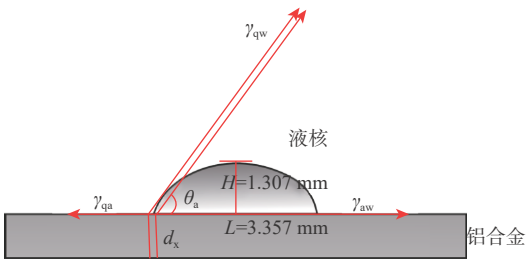


图7 2024铝板表面液核状态

Fig. 7 Liquid core status of 2024 aluminum plate surface

经图像测量手段得到液核高度H为1.307 mm，底部接触直径L为3.357 mm，据此得出液核与2024铝板表面的表观接触角为63.4°，将其代入Thomas[21]提出的公式，计算得到冻结液核与表面之间的接触角为46.3°。

除了对接触角进行确定，模型的准确性还与一些液核自身参数有关，如液核冻结过程中的潜热系数、液核的密度、冻结液核的密度、液核的扩散激活能、液核中水分子的面积数密度等。同时，模型中还涉及定理中常用的常数，如阿伏伽德罗常数、普朗克常数、玻耳兹曼常数等。模型中涉及的相关参数如表1所示。

表1 特性参数选取

Table 1 Characteristic parameter selection

参数	数值
液核分子数密度/cm^{-2}	$1.038\,3\times10^{15}$
玻耳兹曼常数/(J·K^{-1})	$1.380\,622\times10^{-23}$
普朗克常数/(J·s^{-1})	$6.626\,176\times10^{-34}$
阿伏伽德罗常数	$6.022\,045\times10^{23}$
液核密度/(g·cm^{-3})	1.000
凝固潜热/(J·kg^{-1})	334 388.9
液核扩散激活能/(J·mol^{-1})	13 000

3.1.2 结霜温度变化分析

当飞机在停机坪内停靠时，其表面受到多种因素影响，本节针对机场内可能存在的高湿、低温和飞机表面不同部位粗糙度差异来设定工况条件，设置了不同露点温度、不同表面温度、不同粗糙度下的对照组，并以此进行数值模拟，分析具体条件下各工况环境对地面飞机表面结霜温度的影响。

为探究在不同粗糙度下的结霜温度变化，给定0.5,1,1.5,2 μm这4个粗糙度值，当表面温度为−5℃时，露点温度在不同粗糙度下对结霜温度的影响情况如图8所示。当露点温度为5 ℃时，表面温度在不同粗糙度作用下的结霜温度变化情况如图9所示。

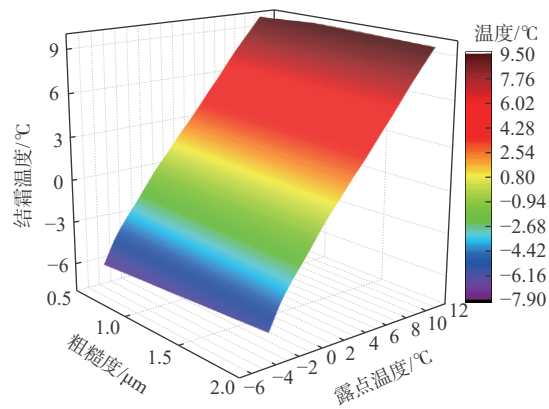


图8 露点温度对结霜温度的影响

Fig. 8 Effect of dew point temperature on frost formation temperature

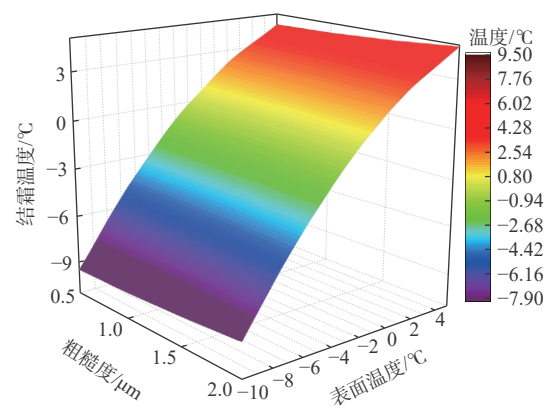


图9 表面温度对结霜温度的影响

Fig. 9 Effect of surface temperature on frosting formation temperature

通过观察图8可得，在同一粗糙度下，结霜温度与露点温度之间成正比，即结霜温度随着露点温度的升高而增大，当露点温度从−5 ℃提升到10 ℃时，结霜温度从−7.9 ℃提升到8.6 ℃；在同一露点温度下，结霜温度与粗糙度之间成正比，即结霜温度随着粗糙度的增大而增大，当粗糙度从0.5 μm提升到2 μm时，结霜温度从−7.9 ℃提升到−7 ℃。

由图9可得，当露点温度较高时，露点温度与表面温度相近。在同一表面温度下，低粗糙度表面

与高粗糙度表面结霜温度存在差异, 当粗糙度从 0.5 μm 提升到 2 μm 时, 结霜温度提升约 0.6 ℃。

3.2　霜层增长数值模拟

3.2.1　模型设置

选定长为 60 mm、宽为 20 mm 的 NACA0018 翼型为研究目标, 表面材料为铝。为简化模型计算, 考虑飞机在停机坪停靠的理想环境, 在其外侧建立一个足够大的流体域, 尺寸为: 长 500 mm、高 200 mm、宽 300 mm, 如图 10 所示。

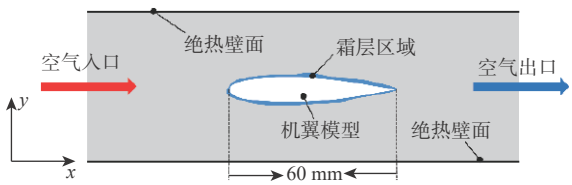


图 10　物理模型示意图

Fig. 10　Schematic diagram of physical model

使用 ANSYS Fluent 软件中的欧拉多相流模型来模拟空气中的水蒸气发生相变的过程, 为使模拟过程更加符合驻地飞机实际场景, 加入了不可压缩多相流的压力修正方程, 表示为

$$\frac{1}{\rho_{rd}}\left[\frac{\partial}{\partial t}\alpha_d\rho_d+\nabla\cdot\alpha_d\rho_d\boldsymbol{u}_d'+\nabla\cdot\alpha_d\rho_d\boldsymbol{u}_d^*-(\dot{m}_{i\to d}-\dot{m}_{d\to i})\right]+\frac{1}{\rho_{ri}}\left[\frac{\partial}{\partial t}\alpha_i\rho_i+\nabla\cdot\alpha_i\rho_i\boldsymbol{u}_i'+\nabla\cdot\alpha_i\rho_i\boldsymbol{u}_i^*-(\dot{m}_{d\to i}-\dot{m}_{i\to d})\right]=0 \tag{25}$$

式中: ρ_{rd}为首相的参考密度; ρ_{ri}为副相的参考密度; $\boldsymbol{u}_d'$为首相的速度修正值; $\boldsymbol{u}_i'$为副相的速度修正值; $\boldsymbol{u}_d^*$为当前首相的速度迭代值; $\boldsymbol{u}_i^*$为当前副相的速度迭代值。

将通过 UDF 编辑的结霜相变公式导入模型中, 对模拟中涉及的物性参数进行设定, 物性参数如表 2 所示。

表 2　物性参数选取

Table 2　Physical property parameter selection

物质	密度/(kg·m^{-3})	热导率/(W·(m·K)$^{-1}$)	黏性系数/(J·(g·K)$^{-1}$)
空气	1.225		1.79×10^{-5}
水蒸气	0.554		1.34×10^{-5}
霜	917	2.5	
铝	2 719	202.4	

3.2.2　霜层厚度变化

1) 机翼上表面霜层厚度分析。机翼作为存在曲率的结霜载体, 其表面不同部位处的霜层形貌必然存在差异。同时, 根据防除冰大纲要求, 在机翼的下方区域有霜的情况下可以起飞[22], 因此, 重点探究机翼上表面霜层生长情况。所建机翼模型全长 60 mm, 在模型上表面 x=5,15,25,45,55 mm 处选取了 5 个探测点, 以探究不同曲率对机翼霜层厚度的影响。在 T_a=−10, −8 ℃ 表面温度条件下, 计算所得不同位置的霜层厚度随时间的变化规律, 如图 11 所示。通过观察图 11 可得, 霜层厚度随时间增加而呈现增长趋势。沿着空气流动方向, 各时间段内霜层的厚度变化先从最小值增大达到峰值, 随后以极缓慢的速度逐渐下降; 然而在机翼的末段, 局部厚度再次略有提高。可以认为机翼表面的霜层生长存在端部效应, 机翼前沿和机翼末端的霜层生长行为较为明显, 前沿和末端区域的局部厚度大于主体区域。

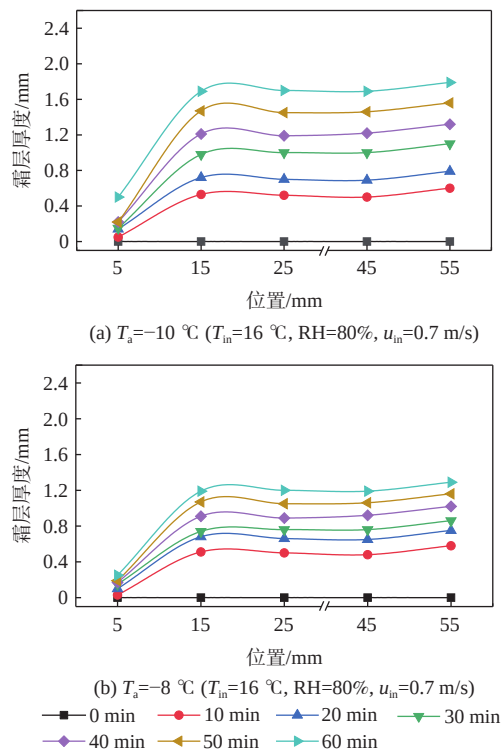
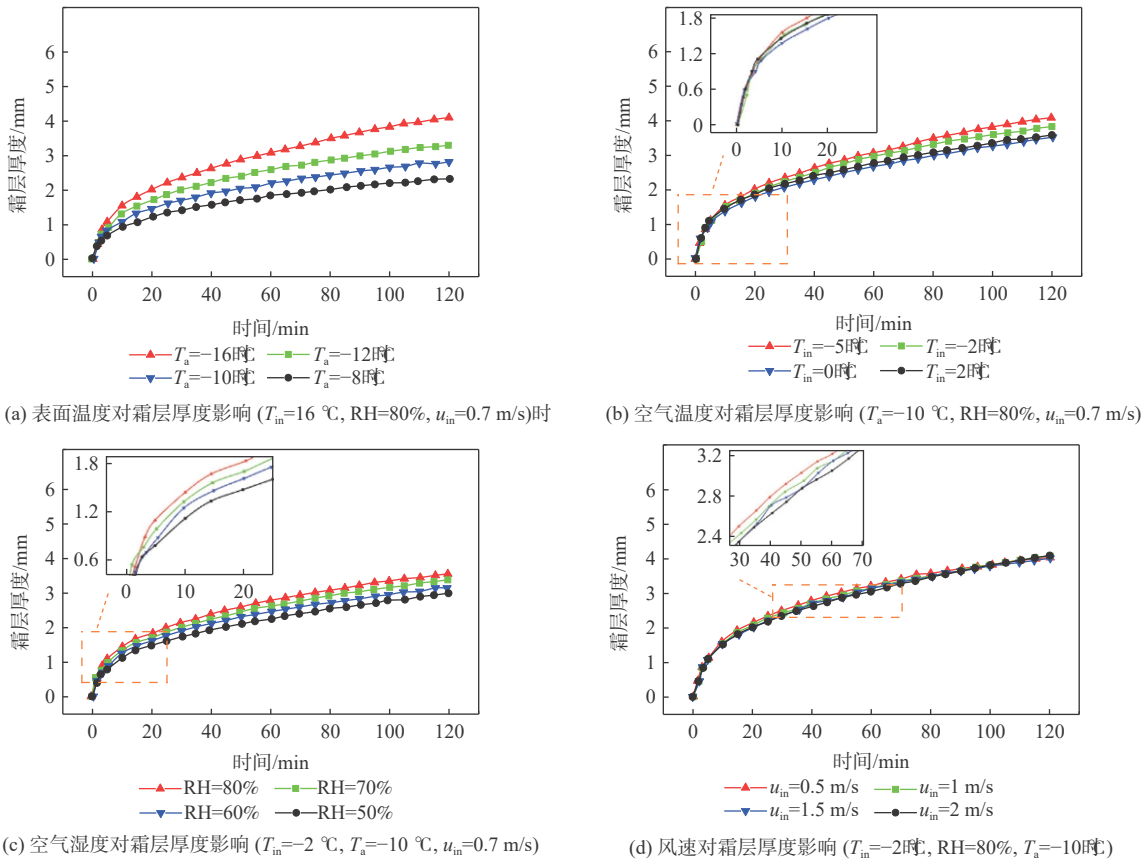


图 11　不同位置霜层厚度随时间变化

Fig. 11　Thickness of frost layer varies with time at different locations

2) 霜层厚度影响因素分析。在飞机表面霜层的生长过程中, 飞机表面温度、空气温度、空气湿度、风速等外部环境严重影响着飞机表面的霜层厚度变化。本文针对各参数影响下的飞机机翼上表面霜层的平均厚度变化进行数值模拟, 模拟结果如图 12 所示。可得, 在不同影响因素下, 飞机表面霜层厚度的变化趋势相类似。具体变化表现为: 霜层厚度随飞机表面温度的降低而升高; 随空气温度的降低而升高; 随空气湿度的增加而升高。但是, 霜层厚度对于空气流速的改变不敏感, 这是因为空气流速的增加会加快水蒸气的质量传递, 但流速越大对霜层



(a) 表面温度对霜层厚度影响 (T_{in}=16 ℃, RH=80%, u_{in}=0.7 m/s)时

(b) 空气温度对霜层厚度影响 (T_a=−10 ℃, RH=80%, u_{in}=0.7 m/s)

(c) 空气湿度对霜层厚度影响 (T_{in}=−2 ℃, T_a=−10 ℃, u_{in}=0.7 m/s)

(d) 风速对霜层厚度影响 (T_{in}=−2℃, RH=80%, T_a=−10℃)

图 12 环境因素对霜层厚度的影响

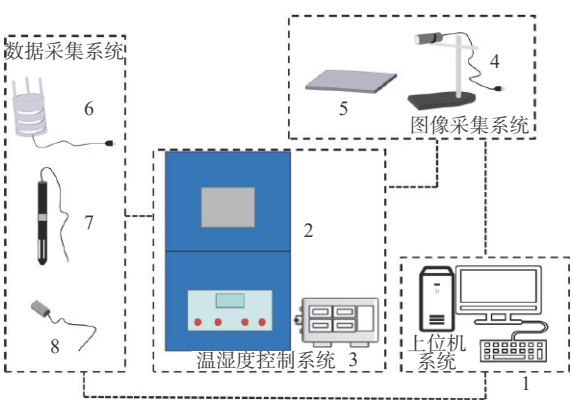
Fig. 12 Effect of environmental factors on frost layer thickness

的稳定生长影响越大，导致空气流速对霜层厚度影响较小。

4 实验验证

4.1 结霜温度实验验证

1）实验系统。实验系统主要针对结霜温度模型进行搭建，为准确把控变量，选择在室内环境下搭建实验系统，如图 13 所示。利用温湿度控制系统模拟适宜结霜的环境；采用数据采集系统中的 PT100 贴片式温度传感器来监测结霜温度；温湿度传感器测量低温试验箱中环境温湿度；利用图像采集系统中的电子显微镜来观测铝板表面结霜情况。



1. 上位机；2. 低温试验箱；3. 温湿度控制器；4. 相机；
5. 铝板；6. 气象传感器；7. 温湿度传感器；
8. PT100 贴片式表面温度传感器。

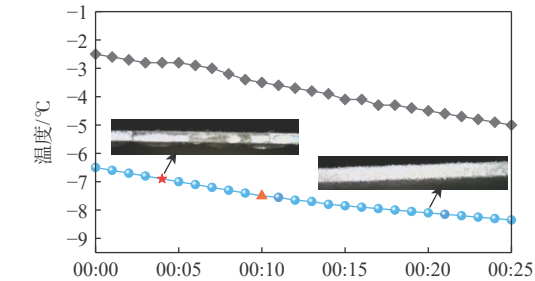
图 13 结霜温度实验系统

Fig. 13 Frost temperature test system

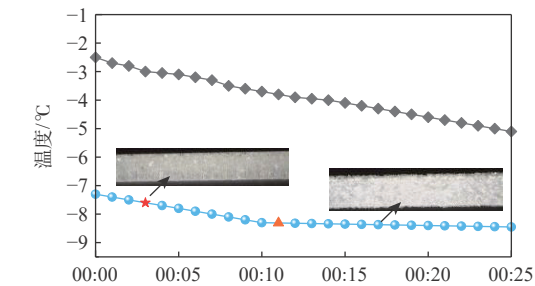
2）实验结果。为验证结霜温度模型的实际效果，以及结霜温度模型的预测精准度，选择 2024 铝板和 5052 铝板 2 种不同牌号铝板分别进行实验，实验与模型验证结果如图 14 所示。通过观察图 14 可得，在成核率方面，实验结果与数值模拟结果表现一致，即 5052 铝板表面成核优先于 2024 铝板。在结霜温度方面，2024 铝板实际结霜温度与预测结霜温度相差 0.6 ℃，5052 铝板实际结霜温度与预测结霜温度相差 0.7 ℃，两者的平均绝对误差为 0.65 ℃，模型所预测的结霜温度平均误差为 9.1%。

4.2 霜层增长实验验证

1）实验系统。实验系统主要针对霜层增长模型进行搭建，为更加真实地还原机场停驻飞机机翼表面的结霜情况，选择在机场附近的室外环境下搭建实验系统，如图 15 所示。模拟机翼整体材料为 2024 铝板，表面采用真实蒙皮工艺进行加工处理。运用添加机器视觉算法的电子显微镜作为图像采集系统，负责检测机翼表面的霜层生长情况。数据采集系统包括温湿度传感器和贴片式温度传感器，



(a) 2024铝板



(b) 5052铝板



图 14　结霜温度实验与模型结果对比

Fig. 14　Comparison of experimental and model results for frost temperature

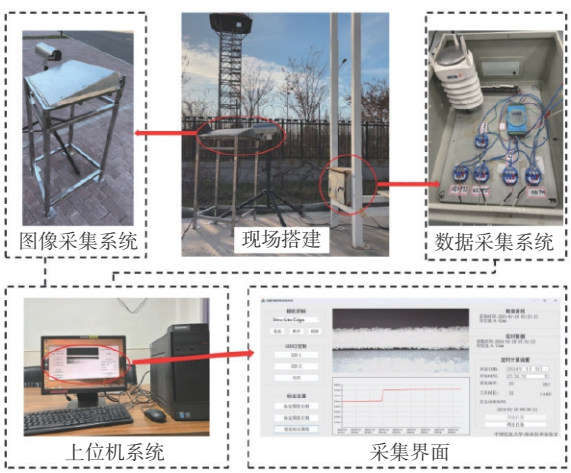


图 15　霜层增长实验系统

Fig. 15　Frost layer growth test system

负责采集机翼附近气象环境及机翼表面温度, 并将数据传递给上位机, 通过编写的采集界面实时显示。

2) 实验工况及参数。为准确验证模型的稳定性, 选取如图 16 所示 2 次不同外部环境下的结霜实验数据。2 次数据均取于晴朗少风的夜间, 工况 1 为冬季北方机场普遍的低温高湿环境, 工况 2 为低温但湿度较低情况下的机场环境。

3) 实验结果。2 种工况下的模型与实验对比结果如图 17 所示。可得, 2 种工况环境下模型所预测的霜层增长的走向与实验数据相似, 均呈现增长趋势。工况 1 因为湿度高, 霜层厚度高于工况 2。

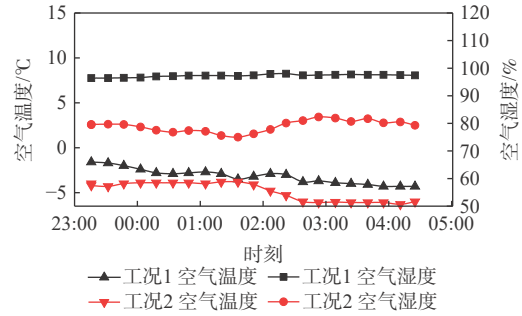
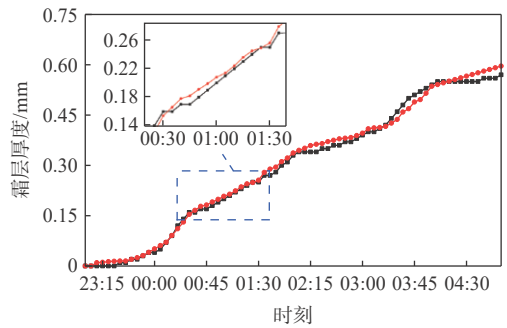
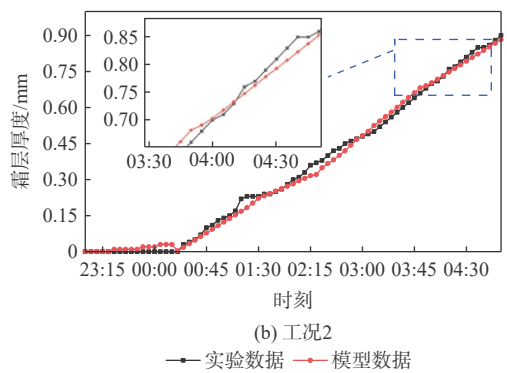


图 16　2 种实验工况

Fig. 16　Two experimental conditions



(a) 工况1



(b) 工况2



图 17　霜层厚度实验与模型结果对比

Fig. 17　Comparison of experimental and model results for frost thickness

在工况 1 中, 模型所预测的数据在霜层增长中间时期误差较大, 为 9.18%, 但整体来看结果良好, 平均误差不足 6.18%。在工况 2 中, 模型所预测的数据在霜层增长后段时期误差较大, 为 10.05%, 但模型数据与实验数据平均误差不足 5.81%。结合 2 组实验数据的对比验证, 霜层增长模型整体预测精度良好, 整个霜层增长过程的平均误差为 5.99%。

5　结　论

本文以地面停驻的飞机为研究背景, 以结霜现象的演变过程为研究对象, 分阶段研究飞机地面结霜多参数影响机制, 探究了不同环境条件对地面飞机结霜过程的影响, 预测了飞机表面结霜开始温度和霜层增长趋势。主要结论如下:

1）基于经典成核理论建立了结霜温度模型，求解了飞机表面冻结液核的覆盖率，并将冻结液核完全覆盖时的温度定义为结霜现象开始时的温度，完成了对结霜温度的预测。与实验结果进行对比，在2024铝板表面结霜温度预测误差为0.6 ℃，在5052铝板表面结霜温度预测误差为0.7 ℃，平均绝对误差为0.65 ℃，模型所预测的结霜温度平均误差为9.1%。

2）基于ANSYS Fluent中的欧拉多相流建立了霜层增长模型，给出了合理的结霜判据和空间判据，并结合地面飞机停靠的实际情况加入了不可压缩多相流的压力修正方程。与2组实验结果进行对比，工况1中平均误差为6.18%，工况2中平均误差为5.81%，整个霜层增长过程的平均误差为5.99%。

3）为解决机场面临的飞机地面结霜问题，提出分阶段建模的方法来预测结霜的整个过程，结合模型预测结果来看，本文所建立的模型可为机场、空管局等被飞机地面结霜现象困扰的单位提供一定的预测思路和理论支持。

参考文献（References）

[1] HAYASHI Y, AOKI A, ADACHI S, et al. Study of frost properties correlating with frost formation types[J]. Journal of Heat Transfer, 1977, 99(2): 239-245.

[2] WU X M, DAI W T, SHAN X F, et al. Visual and theoretical analyses of the early stage of frost formation on cold surfaces[J]. Journal of Enhanced Heat Transfer, 2007, 14(3): 257-268.

[3] CHENG C H, WU K H. Observations of early-stage frost formation on a cold plate in atmospheric air flow[J]. Journal of Heat Transfer, 2003, 125(1): 95-102.

[4] CHAUDHARY G, LI R. Freezing of water droplets on solid surfaces: an experimental and numerical study[J]. Experimental Thermal and Fluid Science, 2014, 57: 86-93.

[5] MATSUMOTO M, SAITO S, OHMINE I. Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing[J]. Nature, 2002, 416(6879): 409-413.

[6] HUANG L Y, LIU Z L, LIU Y M, et al. Effect of contact angle on water droplet freezing process on a cold flat surface[J]. Experimental Thermal and Fluid Science, 2012, 40: 74-80.

[7] BUCHER F, KNOTHE K, THEILER A. Normal and tangential contact problem of surfaces with measured roughness[J]. Wear, 2002, 253(1-2): 204-218.

[8] 陆晨律, 王利平, 孟航飞, 等. 过冷水-壁面接触面积对冰成核行为影响实验研究[J]. 上海交通大学学报, 2025, 59(8): 1145-1155.
LU C L, WANG L P, MENG H F, et al. Ice nucleation behavior in supercooled water with varying wall contact area[J]. Journal of Shanghai Jiao Tong University, 2025, 59(8): 1145-1155(in Chinese).

[9] NA B, WEBB R L. A fundamental understanding of factors affecting frost nucleation[J]. International Journal of Heat and Mass Transfer, 2003, 46(20): 3797-3808.

[10] PIUCCO R O, HERMES C J L, MELO C, et al. A study of frost nucleation on flat surfaces[J]. Experimental Thermal and Fluid Science, 2008, 32(8): 1710-1715.

[11] LEE K S, KIM W S, LEE T H. A one-dimensional model for frost formation on a cold flat surface[J]. International Journal of Heat and Mass Transfer, 1997, 40(18): 4359-4365.

[12] XU D D, WANG L W. Frost formation mechanism analysis and frost growth prediction on ground aircraft[J]. Heat and Mass Transfer, 2013, 49(10): 1385-1393.

[13] HERMES C J L, PIUCCO R, BARBOSA J, et al. A study of frost growth and densification on flat surfaces[J]. Experimental Thermal and Fluid Science, 2009, 33(2): 371-379.

[14] HERMES C J L. An analytical solution to the problem of frost growth and densification on flat surfaces[J]. International Journal of Heat and Mass Transfer, 2012, 55(23-24): 7346-7351.

[15] LENIC K, TRP A, FRANKOVIC B. Transient two-dimensional model of frost formation on a fin-and-tube heat exchanger[J]. International Journal of Heat and Mass Transfer, 2009, 52(1-2): 22-32.

[16] LENIC K, TRP A, FRANKOVIC B. Prediction of an effective cooling output of the fin-and-tube heat exchanger under frosting conditions[J]. Applied Thermal Engineering, 2009, 29(11-12): 2534-2543.

[17] KIM D, KIM C, LEE K S. Frosting model for predicting macroscopic and local frost behaviors on a cold plate[J]. International Journal of Heat and Mass Transfer, 2015, 82: 135-142.

[18] 赵泽扬, 夏鹏. 倾斜表面结霜过程实验及数值模拟研究[J]. 制冷技术, 2022, 42(5): 33-39.
ZHAO Z Y, XIA P. Experimental and numerical research on frosting process of inclined surfaces[J]. Chinese Journal of Refrigeration Technology, 2022, 42(5): 33-39(in Chinese).

[19] 中国民用航空局. 民航冬季安全运行指导意见: 局发明电〔2007〕4027号[Z]. 北京: 中国民用航空局, 2007.
Civil Aviation Administration of China. Guidance on winter safety operation of civil aviation: CAAC Dispatch [2007] No. 4027[Z]. Beijing: Civil Aviation Administration of China, 2007(in Chinese).

[20] 洪天顺, 梁彩华, 吕宁. 结霜工况下水蒸气在冷表面上冷凝成核的理论与实验研究[J]. 中南大学学报 (自然科学版), 2021, 52(6): 1865-1875.
HONG T S, LIANG C H, LV N. Theory and experimental study of water vapour nucleation process on a cold surface in frosting condition[J]. Journal of Central South University (Science and Technology), 2021, 52(6): 1865-1875(in Chinese).

[21] THOMAS Y. The Bakerian lecture. experiments and calculations relative to physical optics[J]. Philosophical Transactions of the Royal Society of London, 1804, 94: 1-16.

[22] 中国民用航空局. 地面结冰条件下的运行: AC-121-50R1[S]. 北京: 中国民用航空局, 2014.
Civil Aviation Administration of China. Operation under icy ground conditions: AC-121-50R1[S]. Beijing: Civil Aviation Administration of China, 2014(in Chinese).

Mechanism of aircraft ground frosting and staged prediction model

CHEN Bin[1, 2, *], ZHU Qingmin[1, 2], ZHUANG Qi[1, 2], WANG Liwen[2, 3]

(1. College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China;
2. Aviation Ground Special Equipment Research Base, Civil Aviation University of China, Tianjin 300300, China;
3. Research Institute of Science and Technology, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Aircraft ground frost has an impact on flight safety and causes flight delays. An accurate frost forecast is an essential basis for the generation of airport snow removal parameters. Frost formation on aircraft ground is influenced by many factors such as weather, body surface structure and flight time, and airports have special forecast indexes for aircraft frost formation. Therefore, frost formation on aircraft ground is divided into the nucleation stage and the frost layer growth stage. A frost formation temperature model based on classical nucleation theory is established during the nucleation stage, and the frost formation temperature is defined as the temperature at which the coverage rate of frozen liquid nuclei approaches the threshold. The surface characteristics of aircraft and the influence of the external environment on the frosting temperature were studied. During the growth stage, the Euler multiphase flow-based frost layer growth model is developed, the frost layer growth identification criterion is added from the frost layer formation principle, and the impact of various working conditions on the frost layer properties of the aircraft surface is investigated. The results of simulation and field experiments show that the average error of the frost formation temperature model is 0.65 ℃. The average error of the frost growth model is 5.99%. The results demonstrate that the proposed phased modeling method can provide theoretical support for predicting ground aircraft frosting.

Keywords: nucleation; frost temperature; frost layer growth; numerical simulation; predict

Received: 2024-02-28; **Accepted**: 2024-10-25; **Published Online**: 2025-02-26 16:51
URL: link.cnki.net/urlid/11.2625.V.20250226.1048.001
Foundation item: General Project of National Natural Science Foundation of China (52472460)
* **Corresponding author**. E-mail: chenbindavid@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0122

双注意力引导的U-Net++遥感图像语义分割模型

刘春娟[1]，辛钰强[1]，吴小所[1,*]，闫浩文[2]
(1. 兰州交通大学 电子与信息工程学院，兰州 730070；2. 兰州交通大学 测绘与地理信息学院，兰州 730070)

摘 要：利用语义分割算法为遥感图像中的像素赋予地物类别标签是遥感图像智能解译中的重要内容。针对高分辨率遥感图像中不同地物类别之间尺度差异大且场景复杂导致的物体边缘分割不完整、小尺度物体分割精度低的问题，提出双注意力引导的U-Net++语义分割模型。在网络的编码阶段构建双分支骨干网络提取特征，利用互注意力捕捉不同尺度特征图像素之间的依赖关系，自适应地融合相同网络深度的不同尺度特征，提升对小尺度物体的关注度；在网络的解码阶段引入空间与通道混合的注意力机制，缩小不同深度子解码器输出之间的语义差距，同时融合其中不同层次的语义信息和空间位置表征，解决复杂场景下精细分割的问题。实验结果表明：所提算法在Potsdam数据集与Vaihingen数据集上的平均交并比(mIoU)分别达到了86.77%与82.73%，F_1分数的均值分别达到了92.32%与90.79%，整体性能显著优于U-Net++、FarSeg、DMAU-Net、SAPNet等对比算法，且对小尺度物体的分割性能有明显提升。

关 键 词：遥感图像；语义分割；U-Net++；注意力机制；小尺度物体
中图分类号：TP751
文献标志码：A **文章编号**：1001-5965(2026)05-1366-12

语义分割作为计算机视觉技术中的主流任务之一，以其在场景理解中的独特优势成为遥感图像智能解译中的重要一环，在地质灾害监测、国土资源评估、边海安全预警等领域发挥着重要作用[1]。遥感图像语义分割是在像素层面对遥感图像中的地物类别进行密集预测、分配类别标签的一种图像处理技术。遥感图像中普遍存在地物种类繁多、光谱及纹理特征复杂等特点，如何精准预测出复杂场景下目标物体的语义信息、分配类别标签仍是长期的挑战[2]。

自Long等[3]提出全卷积神经网络(fully convolutional neural network, FCN)以来，这种在像素层面对图像进行端到端密集预测的学习范式成为目前语义分割算法的通用设计方案[4-5]。为提高深度卷积神经网络对目标物体的分割精度，各种算法相继引入语义分割模型，并在分割性能上取得了不错的提升。一种方法是通过融合多尺度特征增加空间细节信息，缓解目标物体边缘分割模糊的问题[6-9]。例如，Ronneberger等[10]提出的U-Net模型，通过跳跃连接将编解码过程中相同分辨率的特征图在通道维度进行拼接融合，弥补了编码器下采样损失的语义信息；Zhou等[11]提出在嵌套的U-Net结构中对同尺度特征进行密集的跳跃连接，以弥合编码器与解码器同尺度特征图之间的语义差距；Chen等[12]采用级联的方式将空洞空间金字塔池化模块嵌入编码器的深层网络，捕获了多尺度的特征信息；He等[13]将Swin Transformer编码器采用并行的方式嵌入U-Net，获取更加全局的上下文信息。另一种方法是通过注意力机制对语义信息加以筛选，提高特定语义的特征表示[14-16]。例如，SE-Net[17]可以根据任

收稿日期：2024-03-04；**录用日期**：2024-04-08；**网络出版时间**：2024-06-27 13:55
网络出版地址：link.cnki.net/urlid/11.2625.V.20240626.1534.002
基金项目：国家重点研发计划(2022YFB3903604)；甘肃省重点研发计划(20YF8GA035)；甘肃省自然科学基金(22JR5RA320)；兰州交通大学青年科学基金(2021029)
* **通信作者**. E-mail：wuxs_laser@lzjtu.edu.cn

引用格式：刘春娟，辛钰强，吴小所，等. 双注意力引导的U-Net++遥感图像语义分割模型[J]. 北京航空航天大学学报，2026，52(5)：1366-1377. LIU C J，XIN Y Q，WU X S，et al. Semantic segmentation model for remote sensing images based on U-Net++ guided by dual attention[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1366-1377 (in Chinese).

务目标的不同, 增强或抑制不同的通道信息, 提升网络的特征表达能力; 双重注意力网络(dual attention network, DANet)[18] 依据 2 个任意位置之间的空间相关性及不同通道维度特征映射之间的相互依赖关系, 自适应地整合局部特征信息和全局依赖关系; 语义分割的双向注意力网络(criss-cross attention network, CCNet)[19] 通过十字交叉注意力模块捕捉交叉路径及其周围像素的远距离依赖关系, 聚合密集的上下文信息。近年来, 将多尺度特征融合与注意力机制结合的算法在遥感图像语义分割任务中取得了不错的分割效果。例如, Oktay 等[20] 在 U-Net 的跳跃连接末端加入注意力门控单元, 通过优化不同层次语义特征间的融合方式提高了分割精度。双注意力交互式多尺度残差网络(dual-attention interactive multi-scale residual network, DA-IMRN)[21] 通过双重注意力引导空间信息分支和光谱信息分支间的交互式特征学习, 获取更加精细的分割结果; 双注意力多尺度 U 型网络(dual-attention multi-scale U-Net, DMAU-Net)[22] 通过基于密集连接的多尺度池化模块提高了网络的特征提取能力, 并引入高效的通道注意力机制增强有效特征图, 抑制冗余信息; 感知渐进的分割网络(segmentation-aware progressive network, SAPNet)[23] 引入协同注意力模块构建上下文亲和性特征图, 并通过协同注意力感知模块融合空间和通道中的细节信息。虽然上述方法通过注意力机制对编码器提取到的多尺度信息中的有效信息加以关注或冗余信息加以筛选的方法在计算机视觉图像分割任务中取得了极大的成果, 然而, 由于并未考虑遥感图像存在不同地物类别之间尺度差异较大且占比不平衡的特性, 导致多数算法在遥感图像语义分割任务中表现出对小尺度物体的分割精度较低。

为解决遥感图像语义分割任务中大尺度物体与小尺度物体间分割精度无法兼顾的问题, 本文提出一种由编码端注意力和解码端注意力共同引导的 U-Net++语义分割算法。在网络的编码端引入互注意力融合双编码结构提取到的多尺度信息, 并依据不同尺度特征图之间像素的相互依赖关系, 提升对小尺度物体的关注度; 在网络的解码端引入全局-局部注意力融合模块 (global-local attention fusion module, GLAM), 在通道域和空间域分别对不同子解码器的输出加以筛选, 增强输出特征图的语义辨析能力。

1　本文算法

1.1　网络结构

本文以 U-Net++为基础模型, 结合双分支骨干网络 (Resnet18) 的多尺度输入策略、侧面输出融合策略及注意力机制, 构建了一种双注意力引导的 U-Net++模型 (U-Net++ guided by dual attention, DAU-Net++) 用于处理场景愈发复杂的遥感图像语义分割问题, 其结构如图 1 所示。将输入分辨率为 512 像素×512 像素的遥感图像通过邻近插值法上采样得到尺寸为 1 024 像素×1 024 像素的图像, 通

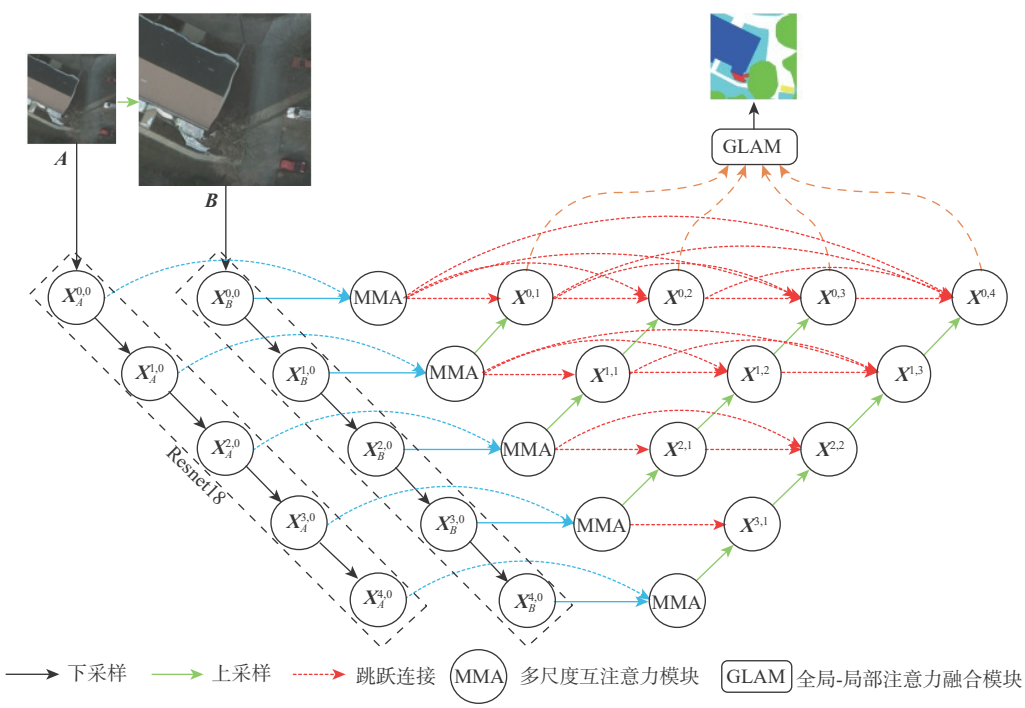


图 1　双注意力引导的 U-Net++网络

Fig. 1　U-Net++ guided by dual attention

过骨干网络对2幅不同尺度图像分别进行特征提取，得到不同尺度的特征图 $\boldsymbol{X}_A^{i,0}$、$\boldsymbol{X}_B^{i,0} \in \mathbf{R}^{2H\times 2W\times C}$ ($i\in\{0,1,2,3,4\}$)，其中，H、W、C 分别为特征图的高、宽、通道数。将获得的不同尺度遥感图像的特征图输入到多尺度互注意力 (multi-scale mutual attention, MMA) 模块中，通过计算各尺度特征图中像素间的依赖关系，在像素级层面上平衡不同尺度目标物体所占的权重，并自适应地融合不同尺度特征。为保持高分辨率细粒度特征的定位信息，在编码器和解码器之间使用密集跳跃连接机制，使得融合的高分辨率细粒度特征可以通过跳跃连接连续地传输到解码器，补偿解码器上采样导致的位置信息损失。解码器最终得到4组与原始图像分辨率大小相同的预测输出，分别为 $\boldsymbol{X}^{0,1}$、$\boldsymbol{X}^{0,2}$、$\boldsymbol{X}^{0,3}$、$\boldsymbol{X}^{0,4}$，其具有不同的语义层次和空间位置表征。浅层子解码器的输出富含细粒度的特征，更利于定位；而深层子解码器的输出具有更丰富的语义信息，有利于分类。利用GLAM模块中的通道注意力与空间注意力抑制不同深度解码器输出间的语义鸿沟和定位差异，解决复杂场景下的精细分割问题。

图2为图1所示卷积块的结构。设 $\boldsymbol{x}^{i,j}$ 为卷积块 $\boldsymbol{X}^{i,j}$ 的输出，则其可表示为

$$\boldsymbol{x}^{i,j}=\begin{cases}\lambda\left(\delta\left(\boldsymbol{x}^{i-1,j}\right)\right) & j=0\\ \delta\left(\varepsilon\left(\boldsymbol{x}_A^{i,0},\boldsymbol{x}_B^{i,0},\mu\left(\boldsymbol{x}^{i+1,j-1}\right)\right)\right) & j=1\\ \delta\left(\varepsilon\left(\boldsymbol{x}_A^{i,0},\boldsymbol{x}_B^{i,0},\varepsilon\left(\boldsymbol{x}^{i,k}\right)_{k=1}^{j-1}\mu\left(\boldsymbol{x}^{i+1,j-1}\right)\right)\right) & j>1\end{cases} \tag{1}$$

式中：$\delta(\cdot)$表示卷积运算；$\lambda(\cdot)$表示下采样的2×2最大池化操作；$\mu(\cdot)$表示使用转置卷积进行上采样；$\varepsilon(\cdot)$表示通道维度上的连接，目的是融合特征。

当 $j=0$ 时，编码器下采样并提取特征；当 $j>0$ 时，密集的跳跃连接机制开始工作，编码器中的细粒度特征依次传输到深层解码器。

1.2 多尺度互注意力模块

遥感图像的语义分割任务中存在着由于地物类别之间尺度差异较大导致的大尺度物体与小尺度物体间分割精度无法兼顾的问题。注意力机制是模拟人类对感兴趣物体的关注度提出的一种矩阵加权算法，其在网络训练过程中可以自适应地学习到最优的权重矩阵，为感兴趣的信息分配更多关注度，达到滤除冗余信息、关注有效信息的目的[24]。本文依据不同尺度特征图之间的相关性提出MMA模块，其结构如图3所示。

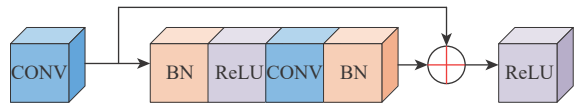


图2 卷积块结构

Fig. 2 Convolution block architecture

2幅不同尺度遥感图像经骨干网络 (Resnet18) 提取得到特征图，将同一网络层次不同尺度的2幅特征图输入MMA模块。该模块通过计算2幅输入的不同尺度特征图之间像素的依赖关系，在像素级的层面上平衡不同尺度目标物体所占的权重，缓解遥感图像中由于目标物体间尺度差异大、类别不平衡导致的分割精度下降的问题。

如图3所示，MMA模块需要2个输入特征图：$\boldsymbol{x}_A^{i,0} \in \mathbf{R}^{C\times H\times W}$ ($i\in\{0,1,2,3,4\}$)表示分辨率为512像素×512像素的遥感图像经骨干网络 (Resnet18) 提取得到的特征图；$\boldsymbol{x}_B^{i,0} \in \mathbf{R}^{C\times 2H\times 2W}$ ($i\in\{0,1,2,3,4\}$)表示对原图像经2倍上采样后由骨干网络 (Resnet18) 提取得到的特征图。特征图 $\boldsymbol{x}_A^{i,0}$ 经变形得到特征矩阵 $\boldsymbol{A}\in\mathbf{R}^{C\times N}$，矩阵 $\boldsymbol{A}$ 经转置得到矩阵 $\boldsymbol{A}'\in\mathbf{R}^{N\times C}$；特征图 $\boldsymbol{x}_B^{i,0}$ 经2倍下采样后通过1×1卷积得到新的特征图 $\boldsymbol{B}\in\mathbf{R}^{C\times H\times W}$，对特征图 $\boldsymbol{B}$ 变形得到矩阵 $\boldsymbol{B}'\in\mathbf{R}^{C\times N}$。其中，$N=H\times W$，代表输入特征向量的个数。将特征矩阵 $\boldsymbol{A}'$ 和 $\boldsymbol{B}'$ 相乘，经Softmax函数激活输出得到注意力权重图 $\boldsymbol{Q}\in\mathbf{R}^{N\times N}$。该过程可表示为

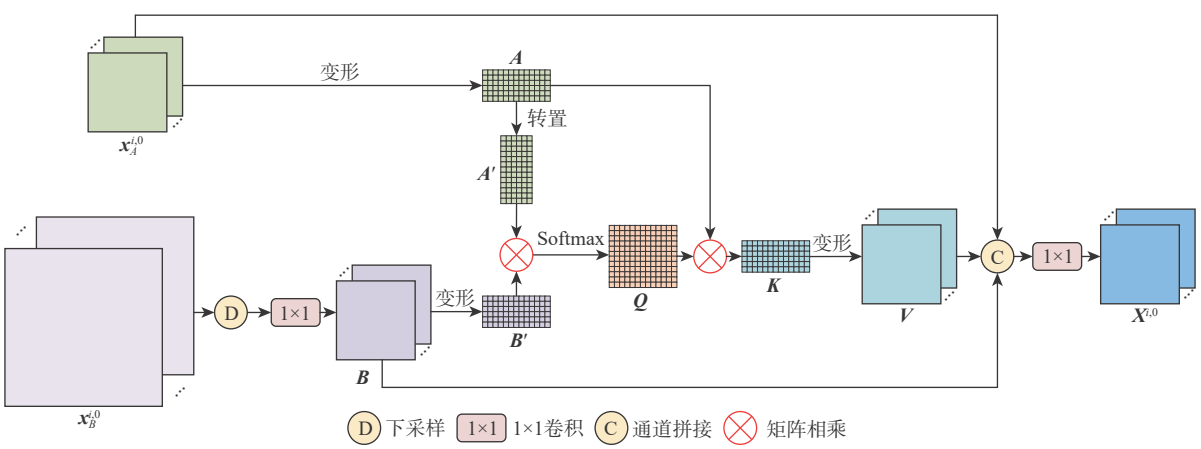


图3 多尺度互注意力模块

Fig. 3 Multi-scale mutual attention module

$$\boldsymbol{Q}=\boldsymbol{A}'\boldsymbol{B}'=\left[\text{transpose}(\boldsymbol{A})\right]\cdot\left[\text{reshape}(\boldsymbol{B})\right] \tag{2}$$

式中: transpose(·)表示对矩阵进行转置操作; reshape(·)表示对矩阵进行变形操作。

对矩阵$\boldsymbol{A}$和注意力权重图$\boldsymbol{Q}$矩阵相乘得到矩阵$\boldsymbol{K}\in\mathbf{R}^{C\times N}$, 将特征矩阵$\boldsymbol{K}$变形得到特征图$\boldsymbol{V}\in\mathbf{R}^{C\times H\times W}$。将特征图$\boldsymbol{V}$、特征图$\boldsymbol{x}_A^{i,0}$、特征图$\boldsymbol{B}$进行通道拼接, 并利用 1×1 卷积在通道维度进行压缩得到特征图$\boldsymbol{X}^{i,0}\in\mathbf{R}^{C\times H\times W}$。具体方法为

$$\boldsymbol{X}^{i,0}=\xi\left[\text{concat}\left(\boldsymbol{V},\boldsymbol{x}_A^{i,0},\boldsymbol{B}\right)\right]=\xi\left[\text{concat}\left(\text{reshape}(\boldsymbol{QA}),\boldsymbol{x}_A^{i,0},\boldsymbol{B}\right)\right] \tag{3}$$

式中: $\xi[\cdot]$表示将通道维度压缩 1/3 的降维操作; concat(·)表示通道拼接。

从式 (3) 可以看出, 特征图$\boldsymbol{X}^{i,0}$不仅包含了输入特征图$\boldsymbol{x}_A^{i,0}$、$\boldsymbol{x}_B^{i,0}$的特征信息, 还包含了不同尺度特征图中像素间的依赖关系。在遥感图像的语义分割任务中, 大尺度目标物体分割的精确度往往高于小尺度目标物体的分割精确度。通过将原始图像以临近插值法放大图像来增大小尺度目标物体的尺寸, 再以多尺度互注意力捕获不同尺度特征图中像素间的全局依赖关系, 生成注意力权重图。该权重图通过调整特征图中不同尺度目标物体所占权重, 在不影响大尺度目标物体分割精确度的前提下, 提高小尺度目标物体的分割精确度。

1.3　全局-局部注意力融合模块

U-Net++网络最终得到 4 组与输入图像尺度相同的预测图像, 即模块$\boldsymbol{X}^{0,j}(j\in\{1,2,3,4\})$的输出$\boldsymbol{x}^{0,j}(j\in\{1,2,3,4\})$。4 组输出的大小相同, 但具有不同的语义层次和空间位置表征。浅层子解码器的输出富含细粒度的特征, 更利于定位; 而深层子解码器的输出具有更丰富的语义信息, 有利于分类。通道注意力可以依据输出特征图中不同通道的特征信息对计算机视觉任务的贡献不同为不同通道给予不同的关注度, 以此节省计算资源, 提高算法的特征表达能力。空间注意力机制通过聚合所有通道的特征信息, 旨在获取对目标物体的空间位置信息的关注。为融合 4 组表征不同信息的特征, 同时抑制其中存在的语义鸿沟及定位差异, 自适应地选择和关注不同组特征之间更有意义的信息, 提出了 GLAM 模块, 其结构如图 4 所示。

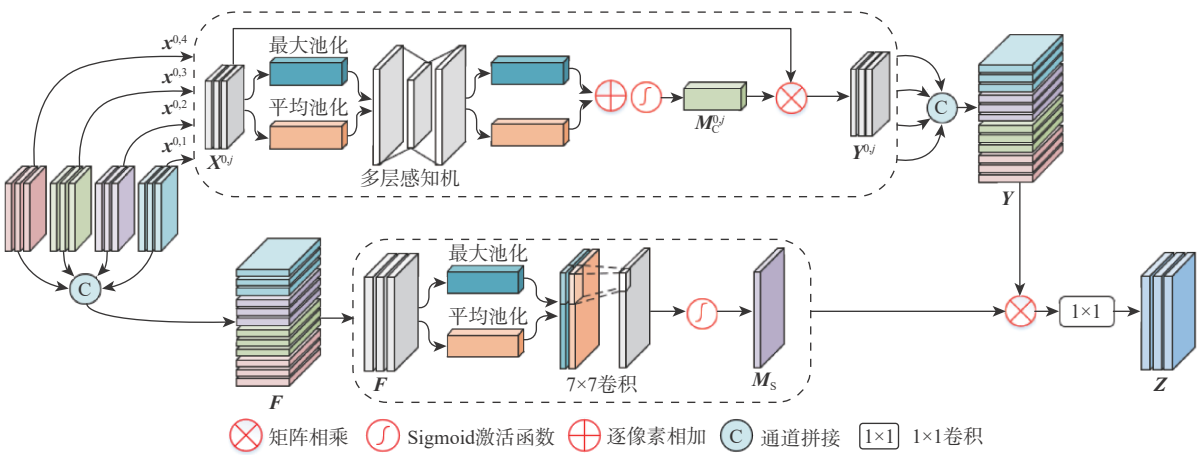


图 4　全局-局部注意力融合模块

Fig. 4　Global-local attention fusion module

首先, 将 4 组特征$\boldsymbol{x}^{0,j}(j\in\{1,2,3,4\})$分别通过通道注意力提取组内关系。具体过程为: 在通道注意力的压缩模块, 对输入特征图$\boldsymbol{x}^{0,j}\in\mathbf{R}^{C\times H\times W}$使用平均池化和最大池化聚合特征图的空间信息, 生成 2 个不同的空间上下文描述。将平均池化特征与最大池化特征输入通道注意力的激励模块, 通过多层感知机 (multilayer perceptron, MLP) 给不同的特征通道生成不同的权重值; 即通过 2 个全连接层在降维、升维的过程中构建特征通道之间的相关性, 在 2 个全连接层之间使用 ReLU 激活函数进行归一化处理。使用元素求和的方法合并 2 个特征向量, 并用 Sigmoid 函数激活输出得到通道注意力权重$\boldsymbol{M}_{\text{C}}^{0,j}\in\mathbf{R}^{C\times 1\times 1}$。利用通道注意力权重向量$\boldsymbol{M}_{\text{C}}^{0,j}$对输入特征$\boldsymbol{x}^{0,j}$矩阵加权, 得到特征图$\boldsymbol{Y}^{0,j}\in\mathbf{R}^{C\times H\times W}$。将 4 组特征图$\boldsymbol{Y}^{0,j}(j\in\{1,2,3,4\})$在通道维度拼接得到$\boldsymbol{Y}\in\mathbf{R}^{4C\times H\times W}$。该过程可表示为

$$\boldsymbol{M}_{\text{C}}^{0,j}=\partial\left(\text{MLP}\left(\text{MaxPool}\left(\boldsymbol{x}^{0,j}\right)\right)+\text{MLP}\left(\text{AvgPool}\left(\boldsymbol{x}^{0,j}\right)\right)\right) \tag{4}$$

$$\boldsymbol{Y}^{0,j}=\boldsymbol{x}^{0,j}\otimes\boldsymbol{M}_{\text{C}}^{0,j} \tag{5}$$

$$\boldsymbol{Y}=\text{concat}\left(\boldsymbol{Y}^{0,1},\boldsymbol{Y}^{0,2},\boldsymbol{Y}^{0,3},\boldsymbol{Y}^{0,4}\right) \tag{6}$$

式中: $\partial(\cdot)$表示 Sigmoid 激活函数; ⊗表示逐像素相乘。

其次, 同步地将 4 组特征$\boldsymbol{x}^{0,j}(j\in\{1,2,3,4\})$在通

道维度拼接，利用空间注意力提取组间关系。具体操作为：将4组特征$\boldsymbol{x}^{0,j}\in\mathbf{R}^{C\times H\times W}$在通道维度拼接，得到特征图$\boldsymbol{F}\in\mathbf{R}^{4C\times H\times W}$。对合并特征图$\boldsymbol{F}$在通道维度使用最大池化和平均池化进行压缩，得到2个不同的、聚合所有通道信息的特征向量。利用7×7卷积链接和压缩2个特征向量，并用Sigmoid函数激活输出，得到空间注意力权重$\boldsymbol{M}_{\mathrm{S}}\in\mathbf{R}^{1\times H\times W}$。其过程可以表示为

$$\boldsymbol{F}=\mathrm{concat}\left(\boldsymbol{x}^{0,1},\boldsymbol{x}^{0,2},\boldsymbol{x}^{0,3},\boldsymbol{x}^{0,4}\right) \tag{7}$$

$$\boldsymbol{M}_{\mathrm{S}}=\partial\left(f^{7\times7}\left(\left[\mathrm{AvgPool}(\boldsymbol{F});\mathrm{MaxPool}(\boldsymbol{F})\right]\right)\right) \tag{8}$$

式中：$f^{7\times7}(\cdot)$表示卷积核大小为7×7的卷积运算。

最后，利用$\boldsymbol{M}_{\mathrm{S}}$对特征矩阵$\boldsymbol{Y}\in\mathbf{R}^{4C\times H\times W}$矩阵加权，得到$\boldsymbol{Z}'\in\mathbf{R}^{4C\times H\times W}$。利用1×1卷积对特征图$\boldsymbol{Z}'$压缩通道数后得到融合特征$\boldsymbol{Z}\in\mathbf{R}^{C\times H\times W}$。方法过程可表示为

$$\boldsymbol{Z}=\psi(\boldsymbol{Y}\otimes\boldsymbol{M}_{\mathrm{S}}) \tag{9}$$

式中：$\psi(\cdot)$表示通道维度压缩1/4的降维处理。

2 数据集及评估指标

2.1 数据集

为验证本文模型的先进性及有效性，选择国际摄影测量和遥感学会发布的Potsdam数据集和Vaihingen数据集作为实验数据集。

Potsdam数据集是在德国勃兰登堡州的首府波茨坦上空拍摄的数字正射影像图，包含38张分辨率为6 000像素×6 000像素的遥感图像及对应标签；波茨坦市是一个典型的历史城市，有着众多的建筑群、狭窄的街道和密集的聚落结构。Vaihingen数据集包含33张分辨率大小不同、在韦兴根市上空拍摄的遥感图像及其标签，原始图像中包含众多的多层独立建筑。2个数据集标签被标注为背景及5种地物类别，分别为不透水表面、建筑物、低矮植被、树、汽车。

实验过程中，考虑到GPU的限制，将原始图像裁剪为尺寸大小为512像素×512像素的图像进行实验。对于Potsdam数据集，将其中8组图像裁剪后作为测试集(编号为2_10、2_11、3_10、3_11、4_10、4_11、5_10、5_11)，将其余30组图像裁剪后按3∶1的比例划分为训练集、验证集。Vaihingen数据集中，将其中7组图像裁剪后作为测试集(编号为1、4、7、11、14、17、22)，将其余26组图像裁剪后以3∶1的比例划分为训练集、验证集。

2.2 评估指标

为定量评估本文模型的各项性能，选择语义分割任务中常用的3个指标对其进行评估，分别为平均交并比(mean intersection over union, mIoU)、总体精度(overall accuracy, OA)和F_1分数。其中，平均交并比可以评估预测图像与标签图像的相似度；总体精度是被正确分类的像元数与总体像元数的比值；F_1分数是精准率和召回率之间的谐波均值，F_1值越接近1，模型性能越好。

3 实验设计与结果分析

实验过程在64位Windows10操作系统的服务器上开展，硬件配置为英特尔至强R处理器E5-2650 v4(2.20 GHz)，配备80 GB的内存(RAM)，搭载一张显存为11 GB的Nvidia GeForce GTX 1080 Ti显卡。模型基于Pytorch深度学习框架实现，环境配置为Python3.7+Pytorch1.11版本。网络训练过程中，使用Adam梯度优化算法和二元交叉熵损失函数更新模型参数，网络迭代次数的阈值设置为200，批样本数量(Batchsize)大小设置为4，初始学习率设置为3×10^{-4}，加载ImageNet数据的预训练权值文件对网络参数初始化。学习率衰减策略如下：

$$\alpha=\left(\frac{1}{2}\right)^{\frac{N_{\mathrm{epoch}}}{5}}\alpha_0 \tag{10}$$

式中：α为当前学习率；N_{epoch}为当前训练轮数；α_0为初始学习率。

3.1 消融实验

为验证DAU-Net++各模块的有效性及合理性，在2个数据集上对算法模型展开模块消融实验。对消融实验的策略描述如表1所示。

3.1.1 Potsdam数据集的消融实验

在Potsdam数据集上对DAU-Net++进行模块消融实验，各项评价指标及预测图像如表2、图5所示，图中，虚线框为预测图像的分割缺陷。

表2中，在Baseline基础上加入MMA模块后，模型的mIoU、OA、F_1分数的均值mF1分别增加了3.99%、2.21%、4.12%；不透水表面、建筑物、低矮植被、树、汽车等5种地物类别的IoU分别提升了1.67%、2.89%、3.95%、6.80%、7.16%，其中，低矮植被、树、汽车等小尺度目标物体的参数提升较大。对比图5的第2、3、4列可以看出，加入MMA模块后，模型对图像中占比较少或尺度较小的目标物体的分割精度有了较大提升。多尺度输入策略结合互注意力机制能够通过捕获不同尺度图像像素间的依赖关系，平衡不同尺度物体所占权重，提高占比少、尺度小的物体的分割精度，缓解遥感图像中类别不平衡导致的小尺度物体分割精度下降的问题。

在Baseline基础上加入GLAM模块后，模型的

表 1　消融实验策略描述

Table 1　Strategy description for ablation experiments

模型	策略描述
Baseline	骨干网络为Resnet18的U-Net++网络
Baseline + MMA	在Baseline基础上增加双分支骨干网络输入及MMA模块
Baseline + GLAM	在Baseline基础上增加GLAM模块
DAU-Net++(Baseline + MMA + GLAM)	在Baseline基础上增加双分支骨干网络输入、MMA模块、GLAM模块

表 2　Potsdam 数据集的消融实验定量评估结果

Table 2　Quantitative assessment results of ablation experiments on Potsdam dataset

网络模型	IoU/%					mIoU/%	OA/%	mF1/%
	不透水表面	建筑物	低矮植被	树	汽车			
Baseline	84.37	87.02	79.97	68.57	69.05	80.36	88.63	87.32
Baseline + MMA	86.04	89.91	83.92	75.37	76.21	84.35	90.84	91.44
Baseline + GLAM	85.87	90.71	84.64	76.19	75.39	84.66	91.21	91.73
DAU-Net++	**87.72**	**91.85**	**86.49**	**78.21**	**77.42**	**86.77**	**91.97**	**92.32**

注: 表中加粗字体表示各项指标中的最优数据, 下划线标出数据为各项指标的次优数据。

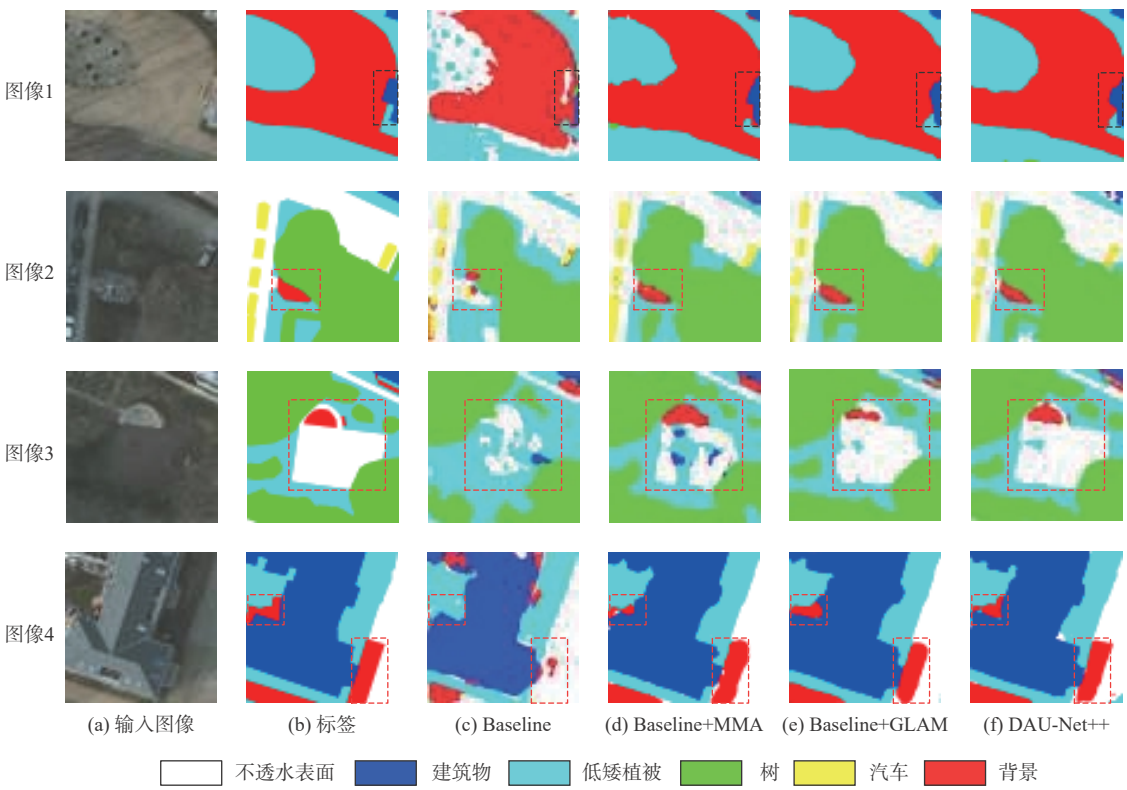


图 5　Potsdam 数据集的消融实验定性比较结果

Fig. 5　Qualitative comparison results of ablation experiments on Potsdam dataset

mIoU、OA、mF1 分别提升了 4.30%、2.58%、4.41%; 5 种地物类别的 IoU 分别提升了 1.50%、3.69%、4.67%、7.62%、6.34%, 其中, 以低矮植被、树、汽车等边界不规则、类间相似度高的目标物体其参数提升最为明显。从图 5 的第 2、3、5 列可以看出, 对比 Baseline, Baseline + GLAM 的预测图像能够更好地展现物体的细节信息, 对目标物体分割地更加完整。侧方输出融合策略结合注意力机制将不同网络深度子解码器的输出自适应融合, 弥合了浅层子解码器与深层子解码器间的语义鸿沟及定位差异, 加强了目标物体的边界信息, 提高了目标物体的分割精度。

在 Baseline 基础上加入 MMA 模块和 GLAM 模块后, 对比 Baseline + MMA, DAU-Net++的 mIoU、OA、mF1 分别增加了 2.42%、1.13%、0.88%; 相较于 Baseline + GLAM, DAU-Net++的 mIoU、OA、mF1 分

别增加了2.11%、0.76%、0.59%。对比图5各列图像可以看出，DAU-Net++语义分割模型集合了MMA模块与GLAM模块的优点，在保证占比少、尺度小的物体的分割精度的情况下，增加了空间细节信息，细化了物体的边界信息，提高了各种地物类别的分割精度。

3.1.2 Vaihingen数据集的消融实验

在Vaihingen数据集上对DAU-Net++进行模块消融实验，得到的各项指标及预测图像如表3、图6所示，图中，虚线框为预测图像的分割缺陷。

表3 Vaihingen数据集的消融实验定量评估结果

Table 3 Quantitative assessment results of ablation experiments on Vaihingen dataset

网络模型	IoU/%					mIoU/%	OA/%	mF1/%
	不透水表面	建筑物	低矮植被	树	汽车			
Baseline	88.06	89.64	69.82	78.17	66.02	78.49	89.66	86.58
Baseline + MMA	90.86	93.22	75.99	81.64	73.40	80.88	90.85	89.29
Baseline + GLAM	91.15	93.41	75.61	83.95	71.68	81.03	91.04	88.84
DAU-Net++	**92.26**	**94.51**	**77.44**	**85.53**	**75.31**	**82.73**	**91.68**	**90.79**

注：表中加粗字体表示各项指标中的最优数据，下划线标出数据为各项指标的次优数据。

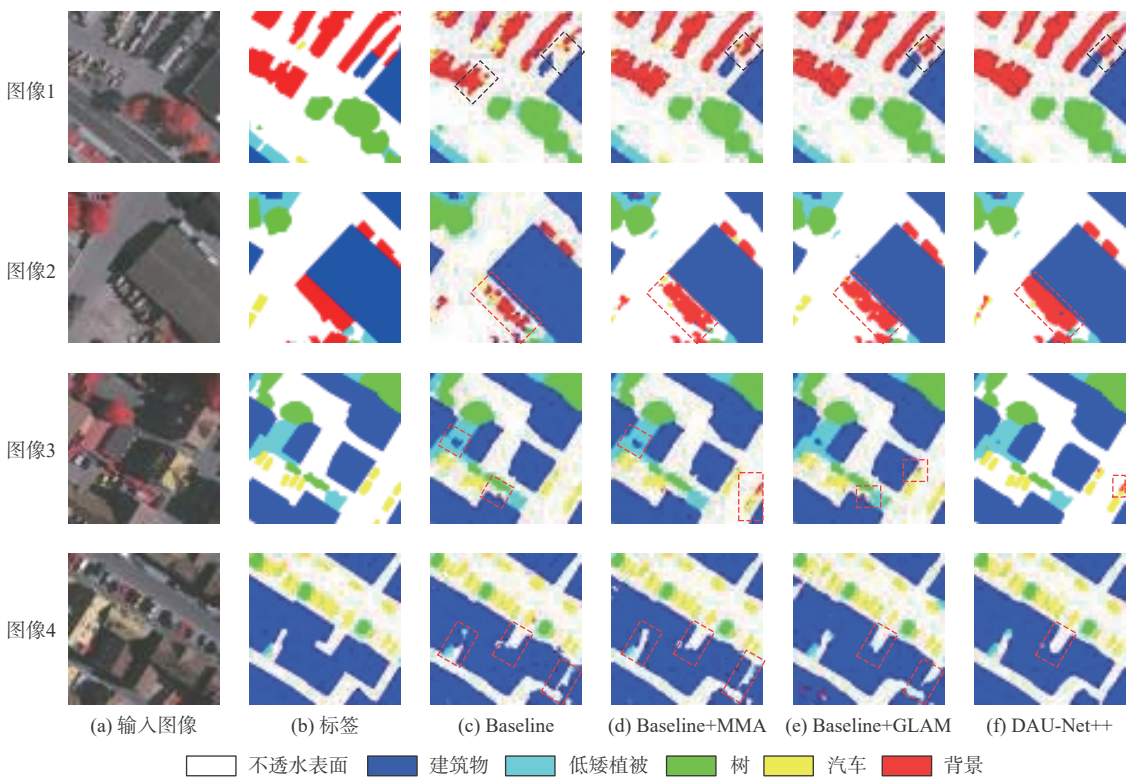


图6 Vaihingen数据集的消融实验定性比较结果

Fig. 6 Qualitative comparison results of ablation experiments on Vaihingen dataset

在Baseline基础上加入MMA模块后，模型的mIoU、OA、mF1分别增加了2.39%、1.19%、2.71%；不透水表面、建筑物、低矮植被、树、汽车等5种地物类别的IoU分别提升了2.80%、3.58%、6.17%、3.47%、7.38%，其中，低矮植被、树、汽车等小尺度目标物体的指标提升最为明显。对比图6的第2、3、4列可以看出，Baseline + MMA能够提高小尺度物体的分割精度，缓解遥感图像中类别不平衡导致的不同尺度目标物体间分割精度差异大的问题。

相较于Baseline，Baseline + GLAM的mIoU、OA、mF1分别增加了2.54%、1.38%、2.26%，5种地物类别的IoU分别提升了3.09%、3.77%、5.79%、5.78%、5.66%，其中，以低矮植被、树、汽车等与其他地物类别边界较难区分的目标物体的指标提升最为明显。对比图6的第4、5列可以发现，Baseline + MMA对小尺度物体如汽车、树等的分割更具有优势，而Baseline + GLAM分割的物体边界信息更加完整。

在Baseline基础上加入MMA模块和GLAM模块后，对比Baseline + MMA，DAU-Net++的mIoU、OA、mF1分别增加了1.85%、0.83%、1.50%；相较于

Baseline + GLAM, DAU-Net++的 mIoU、OA、mF1 分别增加了 1.70%、0.64%、1.95%。对比图 6 的各列图像可以看出, DAU-Net++的预测图像中目标物体的边界更加完整, 且对各类场景下小尺度物体的分割精度均有较大提升。DAU-Net++更加适合场景愈发复杂的遥感图像语义分割任务。

3.2　对比实验

为验证本文提出的 DAU-Net++语义分割模型的先进性, 在 2 个公开数据集上与 PSPNet[5]、U-Net[10]、U-Net++[11]、EMANet[25]、FarSeg[26]、MACU-Net[27]、DA-IMRN[21]、DMAU-Net[22]、SAPNet[23] 展开对比实验。

3.2.1　Potsdam 数据集的对比实验

各模型在 Potsdam 数据集上得到的评估指标及预测图像如表 4、图 7 所示, 图中, 虚线框为预测图像的分割缺陷。

表 4　Potsdam 数据集的对比实验定量评估结果

Table 4　Quantitative assessment results of comparative experiments on Potsdam dataset

网络模型	IoU/%					mIoU/%	OA/%	mF1/%	浮点运算速度/10^9 s^{-1}
	不透水表面	建筑物	低矮植被	树	汽车				
PSPNet[5]	80.78	85.32	78.55	56.21	65.84	76.17	88.32	85.79	67.95
U-Net[10]	81.63	84.51	76.57	60.38	67.57	76.75	87.69	85.26	**63.59**
U-Net++[11]	84.37	87.02	79.97	68.57	69.05	80.36	88.63	87.32	90.16
EMANet[25]	83.64	90.38	81.15	72.72	71.60	81.87	89.38	88.59	126.49
FarSeg[26]	84.39	89.63	80.95	74.62	70.46	82.55	89.13	87.89	193.23
DA-IMRN[21]	87.86	90.21	81.62	74.53	74.39	84.96	90.24	89.73	167.68
MACU-Net[27]	86.64	90.36	80.69	76.58	73.37	84.76	90.18	90.86	213.58
DMAU-Net[22]	**88.89**	92.03	84.94	76.39	75.46	85.68	91.40	91.15	186.74
SAPNet[23]	88.36	**92.14**	83.91	**78.52**	76.31	85.73	91.24	91.68	317.94
DAU-Net++	87.72	91.85	**86.49**	78.21	**77.42**	**86.77**	**91.97**	**92.32**	257.41

注: 表中加粗字体表示各项指标中的最优数据, 下划线标出数据为各项指标的次优数据。

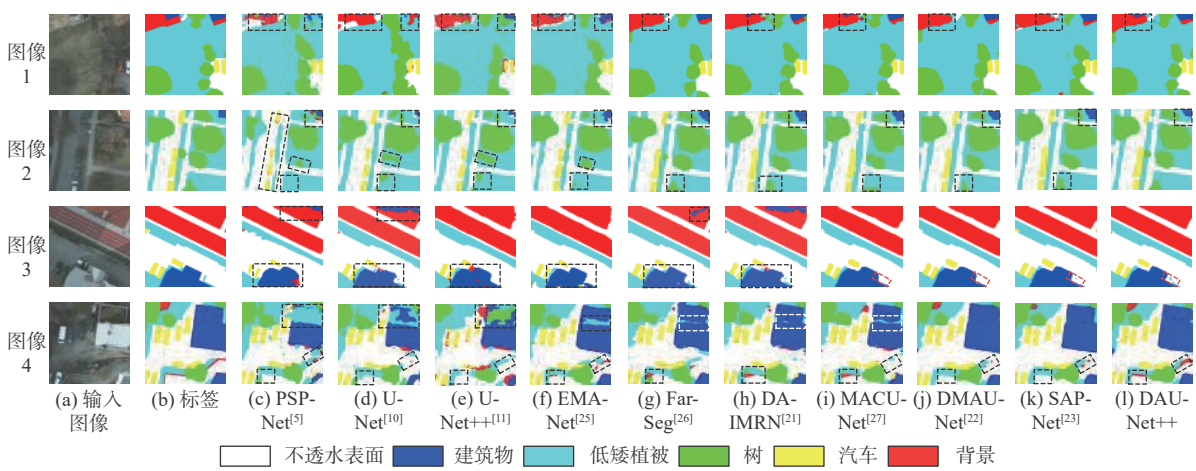


图 7　Potsdam 数据集的对比实验定性比较结果

Fig. 7　Qualitative comparison results of comparative experiments on Potsdam dataset

从表 4 可以看出, 各算法对大尺度目标物体(如不透水表面、建筑)都有较好的分割效果, 而对小尺度目标物体(如树、汽车)及边界信息复杂的地物类别(如低矮植被)的分割效果较差。受遥感图像场景复杂、地物类别丰富、类内差异较大及类间相似度较高等特点的影响, 传统的卷积神经网络虽然计算复杂度较低, 但在遥感图像语义分割任务中整体表现也相对较差, 如 PSPNet、U-Net 等; 而增加空间细节信息路径的优化算法, 计算复杂度较高, 同时在遥感图像语义分割任务中取得了相对较好的性能表现, 如 U-Net++、EMANet、FarSeg 等; 近年来提出的 DA-IMRN、MACU-Net、DMAU-Net、SAPNet 等通过注意力机制进行信息筛选的算法, 计算量较为庞大, 在遥感图像语义分割任务中占据一定的优势。

在表 4 中, DAU-Net++的 mIoU、OA、mF1 较次优网络分别提升了 1.04%、0.57%、0.64%, 模型的整体性能提升较大。DAU-Net++对低矮植被、树类别的 IoU 相较于 DA-IMRN 分别提升了 4.87%、

3.68%, 评估模型整体性能的 mIoU、OA、mF1 分别提升了 1.81%、1.73%、2.59%; 相较于 MACU-Net, DAU-Net++对低矮植被、汽车类别的 IoU 提升了 5.80%、4.05%; 相较于 DMAU-Net, DAU-Net++对树、汽车类别的 IoU 分别提升了 1.82%、1.96%; 与 SAPNet 相比, DAU-Net++对低矮植被的 IoU 提升了 2.58%。相较于其他算法, DAU-Net++在低矮植被、树、汽车类别的 IoU 上有明显提升。与建筑物相比, 树、汽车尺度相对较小, 而低矮植被具有边界不规则的特性, 多数神经网络对这 3 类物体分割效果较差。由此可见, DAU-Net++通过多尺度互注意力平衡不同尺度物体所占的权重, 提升小尺度物体分割精度的同时, 通过增加细粒度的空间信息路径, 提高了目标物体边界分割的完整性。

结合图 7 可以看出, PSPNet、U-Net 对遥感图像的分割效果相对较差, 如图像 2 中对汽车的边界部分错误分类为背景、图像 3 中将大面积背景错误分类为建筑物等; U-Net++、EMANet、FarSeg 对图像中边界信息复杂、色差较小的地物类别容易混淆, 产生错误分类现象, 如图像 2 中树误判为低矮植被、图像 4 中建筑物误判为低矮植被; DA-IMRN、MACU-Net、DMAU-Net、SAPNet 对遥感图像整体的分割效果相对较好, 只是对图像中占比较少、尺度较小的物体分割效果较差, 如图像 1 中汽车、图像 2 中树的边界分割不清晰; 相较于其他算法, DAU-Net++通过增加细粒度的空间信息路径, 融合浅层子解码器的空间细节信息和深层子解码器的抽象语义信息, 对目标物体的边界分割地更加完整, 对低矮植被、树、汽车等类别的分割效果提升最为明显, 整体表现出较好的性能。

3.2.2 Vaihingen 数据集的对比实验

各模型在 Vaihingen 数据集上得到的评估指标及预测图像如表 5、图 8 所示, 图中, 虚线框为预测图像的分割缺陷。

表 5 Vaihingen 数据集的对比实验定量评估结果

Table 5 Quantitative assessment results of comparative experiments on Vaihingen dataset

网络模型	IoU/%					mIoU/%	OA/%	mF1/%
	不透水表面	建筑物	低矮植被	树	汽车			
PSPNet[5]	85.30	86.44	62.53	76.44	57.39	73.78	86.55	83.18
U-Net[10]	84.76	87.49	63.28	74.81	59.36	72.18	85.79	83.24
U-Net++[11]	88.06	89.64	69.82	78.17	66.02	78.49	89.66	86.58
EMANet[25]	89.76	92.21	72.57	81.63	72.08	80.37	90.63	88.26
FarSeg[26]	90.58	91.99	73.08	82.79	71.02	79.62	89.98	88.67
DA-IMRN[21]	88.94	93.61	74.56	82.61	72.83	80.89	90.47	89.45
MACU-Net[27]	92.21	93.29	73.82	82.95	73.28	81.42	90.58	89.53
DMAU-Net[22]	**92.74**	**95.26**	76.51	83.79	73.64	81.38	90.75	89.96
SAPNet[23]	92.58	94.62	75.16	84.16	74.24	81.69	91.08	89.44
DAU-Net++	92.26	94.51	**77.44**	**85.53**	**75.31**	**82.73**	**91.68**	**90.79**

注: 表中加粗字体表示各项指标中的最优数据, 下划线标出数据为各项指标的次优数据。

从表 5 可以看出, 在 Vaihingen 数据集上, DAU-Net++的 mIoU、OA、mF1 较次优算法分别提升了 1.04%、0.60%、0.83%, 模型的整体性能显著优于其他对比算法。与 DA-IMRN 相比, DAU-Net++对低矮植被、树、汽车类别的 IoU 分别提升了 2.88%、2.92%、2.48%, mIoU、OA、mF1 分别提升了 1.84%、1.21%、1.34%; DAU-Net++对低矮植被、树、汽车类别的 IoU 相较于 MACU-Net 分别提升了 3.62%、2.58%、2.03%; 与 DMAU-Net 相比, DAU-Net++对树、汽车类别的 IoU 分别提升了 1.74%、1.67%; 相较于 SAPNet, DAU-Net++对低矮植被类别的 IoU 提升了 2.28%。由此可见, DAU-Net++对小尺度物体与边缘轮廓复杂的物体的分割具有一定优势, 更加适合处理背景信息复杂的遥感图像语义分割任务。

从图 8 可以看出, 在遥感图像场景信息趋于复杂的情况下, PSPNet、U-Net、U-Net++、EMANet、FarSeg 表现相对较差, 预测图像中均出现了大片的错误分类现象, 如将图像 3 中汽车类别与树类别的边界处错误分类为背景、将图像 4 中大片背景错误分类为不透水表面的现象; DA-IMRN、MACU-Net、DMAU-Net、SAPNet 的预测图像整体相对较好, 但对小尺度物体的分割效果相对较差, 如对图像 1 中汽车类别、图像 3 中树类别的分割效果仍不够理想; DAU-Net++的预测图像相较其他 8 种对比模型的预测图像取得了较好的分割效果, 特别是在背景、汽车等地物类别上具有明显的优势, 如图像 2 中的汽车类别、图像 3 中的树类别。DAU-Net++通过互注意力平衡不同尺度物体的权重, 使得小尺度物体

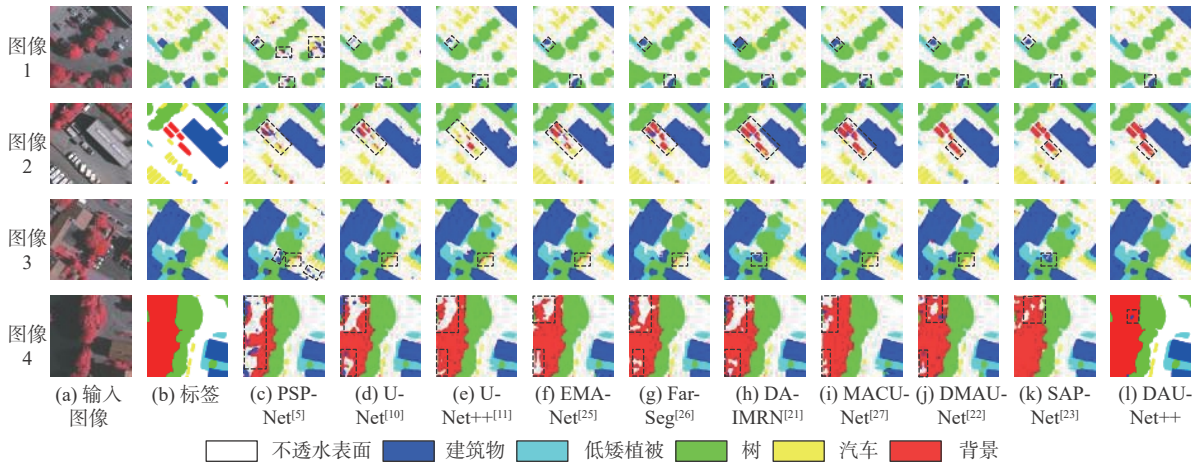


图 8　Vaihingen 数据集的对比实验定性比较结果

Fig. 8　Qualitative comparison results of comparative experiments on Vaihingen dataset

的分割效果得到明显提升。

4　结　论

基于现有算法对小尺度物体分割精度较低的问题, 本文结合注意力模块及 U-Net++在图像处理领域的突出优势, 提出了一种双注意力引导的 U-Net++语义分割模型, 通过在 2 个公开数据集上开展的模块消融实验与对比实验验证了本文算法的先进性, 具体结论如下:

1) 增加多尺度互注意力模块和全局-局部注意力融合模块后, 算法对低矮植被、汽车等地物类别的分割精度有明显提升, 预测图像中的边界信息也更为完整。

2) 相较于 DMAU-Net、SAPNet 等先进算法, 本文算法对小尺度物体的分割精度有较为明显的提升, 并提高了目标物体边界分割的完整性, 整体表现出较好的性能。

结合现有算法在图像分割精度方面遇到的瓶颈, 在未来的工作中尝试从损失函数或遥感数据预处理的角度着手, 以此解决尺度差异大、类别不平衡导致的小尺度、小比例物体分割精度较低的问题, 或许是不错的突破口。

参考文献（References）

[1] 杨军, 张金影. 嵌入自注意力机制的 U 型高分遥感影像语义分割网络[J]. 北京航空航天大学学报, 2025, 51(5): 1514-1527.
YANG J, ZHANG J Y. U-shaped semantic segmentation network of high-resolution remote sensing images embedded with self-attention mechanism[J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(5): 1514-1527(in Chinese).

[2] 吴云华, 张泽中, 华冰, 等. 应用卷积神经网络的遥感图像云层自主检测[J]. 哈尔滨工业大学学报, 2020, 52(12): 27-34.
WU Y H, ZHANG Z Z, HUA B, et al. Autonomous cloud detection for remote sensing images using convolutional neural network[J]. Journal of Harbin Institute of Technology, 2020, 52(12): 27-34(in Chinese).

[3] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 3431-3440.

[4] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-06-17)[2024-03-01]. https://arxiv.org/abs/1706.05587.

[5] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6230-6239.

[6] WANG X, LI Z S, HUANG Y P, et al. Multimodal medical image segmentation using multi-scale context-aware network[J]. Neurocomputing, 2022, 486: 135-146.

[7] DOU F R, ZHANG C F, HU D, et al. EASNet: a multiscale attention semantic segmentation network combined with asymmetric convolution[J]. Journal of Electronic Imaging, 2022, 31(4): 043034.

[8] LUO J, ZHAO L, ZHU L, et al. Multi-scale receptive field fusion network for lightweight image super-resolution[J]. Neurocomputing, 2022, 493: 314-326.

[9] LIN D, SHEN D G, SHEN S T, et al. ZigZagNet: fusing top-down and bottom-up context for object segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 7482-7491.

[10] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2015: 234-241.

[11] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHSH N, et al. U-Net++: a nested U-Net architecture for medical image segmentation[C]//Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Berlin: Springer, 2018: 3-11.

[12] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 833-851.

[13] HE X, ZHOU Y, ZHAO J Q, et al. Swin Transformer embedding U-Net for remote sensing image semantic segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 4408715.

[14] CUI W, WANG F, HE X, et al. Multi-scale semantic segmentation and spatial relationship recognition of remote sensing images based on an attention model[J]. Remote Sensing, 2019, 11(9): 1044.

[15] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 3-19.

[16] QI X Q, LI K Q, LIU P K, et al. Deep attention and multi-scale networks for accurate remote sensing image segmentation[J]. IEEE Access, 2020, 8: 146627-146639.

[17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.

[18] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 3141-3149.

[19] HUANG Z L, WANG X G, WEI Y C, et al. CCNet: criss-cross attention for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(6): 6896-6908.

[20] OKTAY O, SCHLEMPER J, LE FOLGOC L, et al. Attention U-Net: learning where to look for the pancreas[EB/OL]. (2018-05-20) [2024-03-01]. https://arxiv.org/abs/1804.03999.

[21] ZOU L, ZHANG Z F, DU H J, et al. DA-IMRN: dual-attention-guided interactive multi-scale residual network for hyperspectral image classification[J]. Remote Sensing, 2022, 14(3): 530.

[22] YANG Y, DONG J W, WANG Y H, et al. DMAU-Net: an attention-based multiscale max-pooling dense network for the semantic segmentation in VHR remote-sensing images[J]. Remote Sensing, 2023, 15(5): 1328.

[23] LI X, XU F, LIU F, et al. A synergistical attention model for semantic segmentation of remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5400916.

[24] 刘春娟, 乔泽, 闫浩文, 等. 基于双路径监督的遥感图像语义分割网络[J]. 北京航空航天大学学报, 2025, 51(3): 732-741.
LIU C J, QIAO Z, YAN H W, et al. Semantic segmentation network of remote sensing images based on dual path supervision[J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(3): 732-741(in Chinese).

[25] LI X, ZHONG Z S, WU J L, et al. Expectation-maximization attention networks for semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2020: 9166-9175.

[26] ZHENG Z, ZHONG Y F, WANG J J, et al. Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 4095-4104.

[27] LI R, DUAN C X, ZHENG S Y, et al. MACU-Net for semantic segmentation of fine-resolution remotely sensed images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 8007205.

Semantic segmentation model for remote sensing images based on U-Net++ guided by dual attention

LIU Chunjuan[1]，XIN Yuqiang[1]，WU Xiaosuo[1,*]，YAN Haowen[2]

(1. School of Electronic and Information Engineering，Lanzhou Jiaotong University，Lanzhou 730070，China；
2. College of Surveying and Geo-informatics，Lanzhou Jiaotong University，Lanzhou 730070，China)

Abstract: An essential component of the intelligent interpretation of remote sensing images is the use of semantic segmentation algorithms to assign feature class labels to individual pixels. Aiming at the problem of low segmentation accuracy of deep neural networks for small-scale objects caused by the large scale difference between different categories of objects in high-resolution remote sensing images, a U-Net++ guided by dual attention semantic segmentation model is proposed in this paper. In the encoding stage of the network, a dual parallel backbone network is constructed to extract features, and mutual attention is utilized to capture the dependencies between pixels of feature maps of different scales, adaptively fusing features of different scales with the same network depth to enhance the attention to small-scale objects. To address the issue of fine segmentation in complex scenes, a spatial and channel hybrid attention mechanism is introduced in the network's decoding stage to reduce the semantic gap between the outputs of various depth sub-decoders while fusing the semantic information and spatial location representations at various levels therein. The proposed algorithm achieves notable performance metrics, with the mean intersection over union (mIoU) values of 86.77% and 82.73% on the Potsdam dataset and Vaihingen dataset, respectively, accompanied by the mean F_1-score of 92.32% and 90.79%. These results underscore the algorithm's efficacy in delivering comprehensive segmentation of small scale objects, surpassing the performance of other state-of-the-art semantic segmentation algorithms such as U-Net++, FarSeg, DMAU-Net, and SAPNet.

Keywords: remote sensing image; semantic segmentation; U-Net++; attention mechanism; small scale objects

Received：2024-03-04；**Accepted**：2024-04-08；**Published Online**：2024-06-27 13：55

URL：link.cnki.net/urlid/11.2625.V.20240626.1534.002

Foundation items：National Key Research and Development Program of China (2022YFB3903604); Key R & D Projects in Gansu Province (20YF8GA035); Gansu Provincial Natural Science Foundation (22JR5RA320); Lanzhou Jiaotong University Youth Science Fund (2021029)

* **Corresponding author**. E-mail：wuxs_laser@lzjtu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0127

主动侧杆系统多余力机理及敏感性分析

杨建忠，左金鹏，孙晓哲*
(中国民航大学 安全科学与工程学院，天津 300300)

摘 要：主动侧杆系统存在的多余力问题造成了“抖杆打手”、态势感知差、特殊情况下易诱发飞行事故等问题。为更好地分析多余力的作用机理，确定系统参数对多余力的影响机理及规律，通过理论分析确定了与系统多余力相关的参数，基于蒙特卡罗仿真详细分析了各参数对系统多余力的影响，开展了多余力频谱分析，进行了各参数的多余力影响敏感性分析。通过定性定量分析确定了对系统多余力影响最大的因素，即驾驶员操纵频率、侧杆等效转动惯量、侧杆等效阻尼、转速传感器信号偏置、力矩传感器信号偏置，并分析了各参数对多余力的影响机理和程度，为主动侧杆系统多余力抑制方法设计和符合性验证研究提供了参考依据。

关 键 词：主动侧杆；多余力；蒙特卡罗仿真；敏感性分析；伺服系统
中图分类号：V249; TP29
文献标志码：A **文章编号**：1001-5965(2026)05-1378-13

民机驾驶员操纵装置中，侧杆具有占用空间小、质量轻、不遮挡驾驶员视线等优点，广泛应用于空客系列、C-919等飞机。但目前型号中采用的侧杆主要为被动侧杆，存在杆力与杆位移关系单调，无法向飞行员反馈真实飞机状态，紧急情况下易诱发飞行事故的问题[1]。为此，出现了主动侧杆技术。目前，美国有机型安装了主动侧杆，中国商飞也计划将主动侧杆应用于民机以提升驾驶员的操纵感觉。相较于被动侧杆，主动侧杆不再采用弹簧-阻尼器装置提供简单力感反馈，而是采用电机等驱动装置，根据飞控系统的指令提供给驾驶员与飞机飞行状态相关联的反馈力，同时通过电子控制单元实现主副驾驶员的侧杆联动，有效提升了驾驶员的态势感知能力[2]。

主动侧杆是一种精密的机电一体化设备，常见的工作模式有位置伺服模式和力伺服模式。在驾驶员操纵过程中，侧杆工作于力伺服模式，驾驶员操纵导致侧杆偏转一定角度，飞控计算机根据角度指令和飞机的运动状态实时更新杆力-位移曲线，由侧杆控制单元控制驱动电机产生反馈力。在接通自动驾驶仪或单驾驶员操纵情况下，侧杆工作于位置伺服模式，接收来自自动驾驶仪或另一侧侧杆的角度指令，跟随指令作动[3]。在操纵过程中，驾驶员对侧杆施加操纵力，使侧杆产生角加速度，进而产生偏转角度，导致角度滞后于驾驶员操纵力，根据杆力-位移曲线生成的反馈力滞后于操纵力导致了多余力发生，这是多余力产生机理中最核心的因素。在实际的主动侧杆系统中，由于驾驶员操纵力造成的强位置扰动，以及系统运行过程中存在的机械结构磨损、传感器精度下降、信号偏差、信号错误等因素，也进一步加大了多余力。多余力降低了主动侧杆的力加载精度，飞行员无法通过杆力信息准确感知飞行状态，在特殊飞行条件下易诱发飞行事故，影响飞行安全。因此，对民机采用主动侧杆系统的设计和适航符合性验证提出了更高的要求[4-5]。

现有对主动侧杆系统多余力的研究内容较少且普遍集中于多余力的抑制方法，对多余力产生机理的研究较少[6]。2020年，南京航空航天大学陈悦[7]分析了多余力的产生机理，将其归纳为主动侧杆杆力加载系统中存在的强位置扰动导致的扰动力，并

收稿日期：2024-03-05；录用日期：2024-08-06；网络出版时间：2024-09-04 11:00
网络出版地址：link.cnki.net/urlid/11.2625.V.20240903.1848.003
*通信作者. E-mail：xzsun@cauc.edu.cn

引用格式：杨建忠，左金鹏，孙晓哲. 主动侧杆系统多余力机理及敏感性分析[J]. 北京航空航天大学学报，2026，52(5)：1378-1390.
YANG J Z，ZUO J P，SUN X Z. Mechanism and sensitivity analysis of surplus force in active sidestick system[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1378-1390 (in Chinese).

总结了多余力的抑制方法。2021 年, 南京航空航天大学张旭[8]分析了多余力的产生原因, 将多余力视为“强位置扰动”, 认为其与主动侧杆的位置变化密切相关, 并且无法从源头消除, 只能采取一定的方法进行抑制和补偿。2022 年, 中国运载火箭技术研究院的研究人员将多余力视为扰动力矩, 并将其归纳为电机内部的摩擦力矩和电机为驾驶员施加的负载转矩, 指出了扰动力矩受驾驶员操纵的影响[9]。目前, 根据主动侧杆的结构特点和驱动电机的工作特点, 将多余力归纳为 2 种: 惯性多余力和电磁多余力。当前研究缺乏对多余力产生机理更细致的分析和归纳总结。主动侧杆系统中, 常存在参数变化、外部扰动、未建模动态、摩擦间隙等非线性不确定因素, 其对多余力影响程度的定量分析及参数敏感性分析还有不足, 有必要对多余力的影响机理做细致的分析研究。

由于主动侧杆在制造和安装使用过程中, 内部设备参数并非固定值, 且随着运行时间或内外受扰变化, 在物理实验平台通过改变参数开展多余力的研究相当困难, 通过仿真的方式改变相关参数则比较容易, 且能高精度地反映实际参数变化带来的影响, 可以考虑通过提取系统中多余力相关参数进行分析。本文建立了主动侧杆俯仰轴力伺服功能模型, 通过数值仿真得到多余力的产生机理, 分析多余力的来源、相关参数的敏感性和影响规律, 通过蒙特卡罗仿真对不确定性参数的影响进行分析[10], 再根据蒙特卡罗仿真结果进行对应参数的多余力频谱分析, 得到对多余力影响较大的因素, 明确各因素对多余力的定量影响, 为实际工程中主动侧杆的设计及系统多余力抑制方法和故障诊断、容错控制算法的开发提供参考依据。

1　主动侧杆系统动力学建模

1.1　主动侧杆工作原理

主动侧杆通常由电控制单元(electronic control unit, ECU)、电机驱动单元(power drive unit, PDU)、驱动电机、力矩传感器、减速齿轮、角度传感器、驾驶员推杆装置(二自由度转台)[11]、侧杆握把等组成, 其结构框图如图 1 所示。在飞行中, 驾驶员操纵导致侧杆偏转一定角度, 角度指令由侧杆 ECU 传输到飞控计算机(flight control computer, FCC), FCC 根据飞机的运动状态实时更新杆力-位移曲线传输给侧杆 ECU, 由 ECU 控制驱动电机产生反馈力[12]。

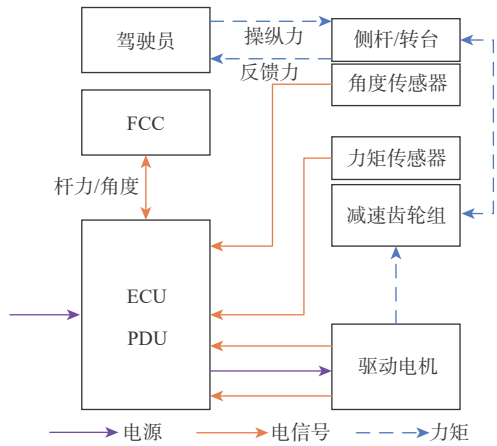


图 1　主动侧杆结构框图

Fig. 1　Structure block diagram of active sidestick

1.2　系统建模

主动侧杆驱动电机常采用永磁同步电机, 永磁同步电机三相电压平衡方程经过坐标变换可以得到d-q电压方程为[13]

$$\begin{cases} u_d = Ri_d + L_d\dfrac{\mathrm{d}i_d}{\mathrm{d}t} - \omega_{\mathrm{m}}p_{\mathrm{n}}L_q i_q \\ u_q = Ri_q + L_q\dfrac{\mathrm{d}i_q}{\mathrm{d}t} + \omega_{\mathrm{m}}p_{\mathrm{n}}(L_d i_d + \psi_{\mathrm{f}}) \end{cases} \tag{1}$$

永磁同步电机电磁转矩方程为

$$T_{\mathrm{e}} = 1.5p_{\mathrm{n}}[\psi_{\mathrm{f}}i_q + (L_d - L_q)i_d i_q] = k_{\mathrm{t}}i_q \tag{2}$$

电机运动方程为

$$J\frac{\mathrm{d}\omega_{\mathrm{m}}}{\mathrm{d}t} = T_{\mathrm{e}} - T_{\mathrm{L}} - B\omega_{\mathrm{m}} \tag{3}$$

式中: u_d、u_q、i_d、i_q、L_d、L_q分别为电压、电流、电感在d-q参考系中的量; R为相电阻; ψ_{f}为永磁体磁链; T_{e}和T_{L}分别为电磁转矩和负载转矩; B为电机和负载的阻尼系数; p_{n}为电机极对数; J为电机转子转动惯量; ω_{m}为电机转速; k_{t}为电机转矩常数。

主动侧杆的力伺服模式是利用电机来实时地为驾驶员提供所需要的反馈力, 从而使驾驶员在操纵侧杆时受到逼真的力感觉, 进而判断飞机的状态。以俯仰方向为例, 驾驶员作用于侧杆上的力矩与电机反馈力矩间的关系可表示为[14]

$$T = J_0\ddot{\theta}_{\mathrm{s}} + C\dot{\theta}_{\mathrm{s}} + K\theta_{\mathrm{s}} + T_0 + T_{\mathrm{d}} \tag{4}$$

式中: T为驾驶员作用于侧杆转轴上的力矩; $\ddot{\theta}_{\mathrm{s}}$、$\dot{\theta}_{\mathrm{s}}$和$\theta_{\mathrm{s}}$分别为驾驶员操纵侧杆转动的角加速度、角速度和角度; J_0为驾驶员手臂到电机的等效转动惯量; C为阻尼系数; K为 ASS 机械系统的等效弹性系数; T_0为侧杆绕转轴的摩擦力矩; T_{d}为电机根据飞机的飞行状态确定的杆力模型通过减速齿轮作用于侧杆上的力矩, 即反馈力。

当驾驶员操纵侧杆手柄运动时, 由于主动侧杆手柄的各部件间为刚性连接, 可以忽略等效弹性系数K, 则可以推导出在电机及驾驶员共同作用下合力矩到手柄转角的方程为[15]

$$T - T_{\mathrm{d}} - T_0 = J_0\ddot{\theta}_{\mathrm{s}} + C\dot{\theta}_{\mathrm{s}} \tag{5}$$

电机的负载转矩T_L传递到俯仰轴上的反力矩T_d可表示为

$$T_d = iT_L \tag{6}$$

式中：i为齿轮减速比；电机负载转矩T_L由光栅式力矩传感器测得，其计算公式为[9]：$T_L = g_c(\theta_m - i\theta_s)$，$\theta_m$为电机转角，即$\omega_m = \mathrm{d}\theta_m/\mathrm{d}t$，$g_c$为扭转刚度。

为更好地接近实际系统，加入摩擦[16-17]、外部负载扰动等不确定性环节，各参数标定值如表1所示[18]。

表1 参数标定值[18]

Table 1 Parameter calibration value[18]

参数	数值
电机相电阻R/Ω	1.3
电机相电感L/mH	1.175
永磁体磁链ψ_f/Wb	0.033 3
电机极对数p_n	20
电机转子转动惯量J/(kg·m²)	0.000 231
扭转刚度g_c/(N·m·rad^{-1})	550
齿轮减速比i	8
侧杆长度l/m	0.15

系统控制框图如图2所示。系统采用三闭环PI控制，从内到外依次为电流环、转速环和力闭环，杆力-位移曲线梯度固定为10。为验证本文所建立模型的合理性与正确性，将仿真结果(见图3、图4)与文献[15,19]中主动侧杆的结果做对比。在相同的驾驶员操纵力输入情况下，主动侧杆提供的反馈力效果(见图3)与文献[15,19]的结果一致。在输入指令均为正弦操纵力的情况下，本文的多余力曲线(见图3)与文献[19]的多余力曲线一致。此外，按照固定梯度为10的杆力-位移曲线加载，系统侧杆偏角输出(见图4)为10°，符合主动侧杆工作原理。由此，验证了主动侧杆系统模型的正确性与有效性。

由于模型仅采用了传统PI控制器，未能实现很好的多余力抑制效果，多余力的特征较为明显。随着仿真的进行，多余力呈现逐渐减小的趋势，期间最大多余力为8.99 N，为方便，下文称其为“标称多余力”[20]。

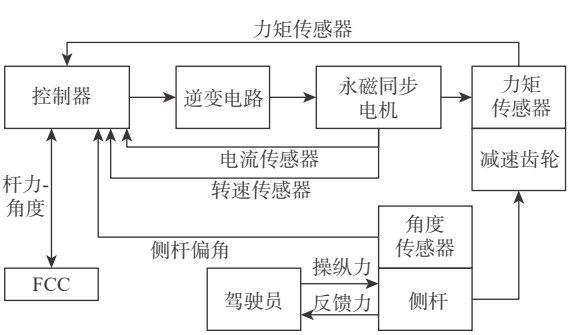


图2 主动侧杆系统控制框图

Fig. 2 System control block diagram of active sidestick

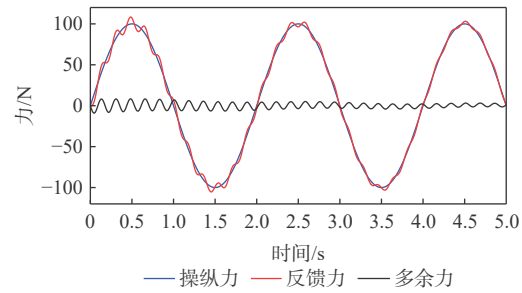


图3 操纵力-反馈力-多余力示意图

Fig. 3 Diagram of control force-feedback force-surplus force

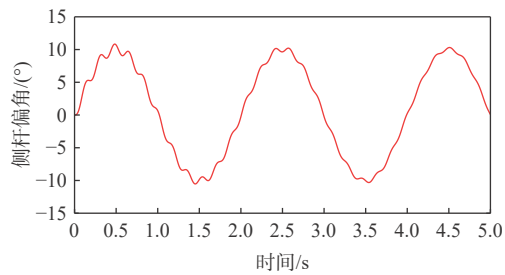


图4 侧杆偏角示意图

Fig. 4 Diagram of steering column offset angle

当驾驶员操纵侧杆运动时，力伺服模式实际上完成的是被动加载过程，这是一个典型的位置扰动型系统，即电机根据驾驶员操纵侧杆转动的角度输出相应的力矩，从而给侧杆提供相对应的反力，而驾驶员同时又在操纵侧杆运动，反过来又干扰了电机的输出。

2 多余力机理分析

基于图2所示系统简化后，可得系统的开环传递函数框图如图5所示。从图5可以得到系统输入指令力F_c到输出力F_d的传递函数为加载通道传递函数[21]，表示如下：

$$G_1(s) = (g_c k_t i J_0 s + k_t g_c i C)/A(s) \tag{7}$$

从驾驶员操纵力F到系统输出力F_d的传递函数为驾驶员操纵通道传递函数，表示如下：

$$G_2(s) = \left(-g_c i^2 JLs^2 - g_c i^2 JRs - g_c i^2 k_t k_e\right)/A(s) \tag{8}$$

式中：L为电机相电感；R为电机绕组电阻；k_e为反电动势常数。

$$\begin{cases} A(s) = G_a s^4 + G_b s^3 + G_c s^2 + G_d s + G_e \\ G_a = JJ_0Ll \\ G_b = JJ_0Rl + JCLl \\ G_c = JCRl + J_0 k_t k_e l + J_0 L g_c l - JLg_c i^2 \\ G_d = Clk_t k_e + J_0 R g_c l + CLg_c l - JRg_c i^2 \\ G_e = CRg_c l - k_t k_e g_c i^2 \end{cases} \tag{9}$$

式中：l为侧杆长度。

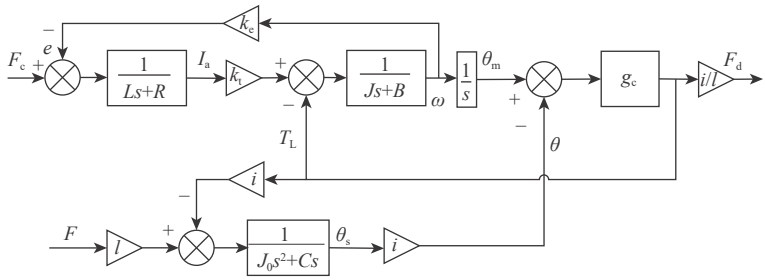


图 5　主动侧杆系统开环控制框图

Fig. 5　System open loop control block diagram of active sidestick

可得到系统输出力为

$$F_d = G_1(s)F_c(s) + G_2(s)F(s) \tag{10}$$

令主动侧杆的输入指令F_c为 0, 可以得到主动侧杆多余力的传递函数为[22-23]

$$G(s) = \left(-g_c i^2 JLs^2 - g_c i^2 JRs - g_c i^2 k_t k_e\right)/A(s) \tag{11}$$

由此可知, 多余力的影响因素众多, 包括驾驶员操纵指令F, 侧杆等效参数J_0、C, 扭转刚度g_c, 驱动电机的参数等。为明确影响多余力的关键因素及各因素的影响程度, 以下将对多余力影响因素进行定性分析。

根据系统实际运行中可能的参数变化, 将影响多余力的因素归纳为 4 类共 16 种, 详细的参数变化范围如表 2 所示[24-25]。

表 2　多余力的影响参数[24-25]

Table 2　Influence parameter of surplus force[24-25]

分类	影响多余力的因素	表现形式
驾驶员指令	驾驶员操纵频率	操纵频率在0～5 rad/s变化
传动装置	侧杆等效转动惯量	等效转动惯量有20%的偏差
	侧杆等效阻尼	等效阻尼有30%的偏差
	扭转刚度	扭转刚度有25%的偏差
电机参数	电机相电阻	相电阻有30%的误差
	电机相电感	相电感有10%的误差
	电机转子转动惯量	转子转动惯量有10%的误差
	电机转矩系数	转矩系数有6%的差异
传感器、控制器	力指令信号延迟	力指令信号延迟2%
	力矩传感器信号延迟	力矩传感器信号延迟0.5%
	转速传感器信号延迟	转速传感器信号延迟0.5%
	力矩传感器信号偏置	力矩传感器信号偏置20%
	转速传感器信号偏置	转速传感器信号偏置20%
	电流传感器信号偏置	电流传感器信号偏置20%
	转速传感器信号噪声	向转速传感器注入2%、频率为50 Hz的白噪声
	电流传感器信号噪声	向电流传感器注入2%、频率为50 Hz的白噪声

根据以上选定的参数, 对各因素影响做蒙特卡罗分析, 从而得到多余力影响因素的结果。

3　参数影响分析

3.1　参数对系统性能影响分析

由加载电机本身性能引起的多余力, 其输入为来自 FCC 的力加载指令, 输出为加载系统输出力, 系统性能越好, 加载输出力越能更好地跟踪指令力, 多余力越小。

1) 扭转刚度对前向通道的影响。取扭转刚度g_c分别为 300,550,800 N·m/rad, 在其他因素不变的情况下, 对系统前向通道开环频率特性进行仿真, 如图 6 所示。可知, 在低频段, 扭转刚度的不同取值对系统的性能影响差异不大。当频率增大时, 扭转刚度对多余力的影响增大, 引起机械谐振, 扭转刚度越大, 谐振峰值越大, 影响系统的稳定性。因此, 要在保证系统性能的前提下选取较小的扭转刚度。

2) 电机转子转动惯量对前向通道的影响。取电机转子转动惯量J分别为 0.000 1,0.000 3,0.000 5 kg·m^2,

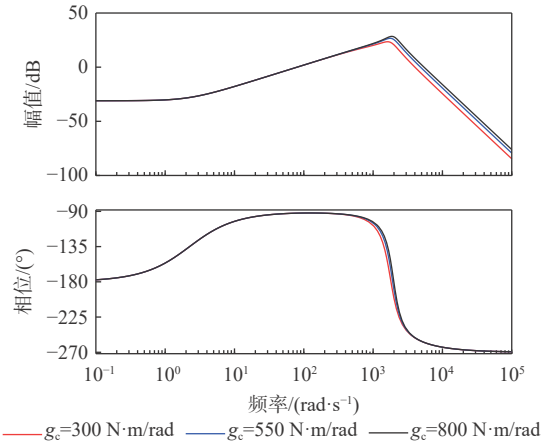


图6 扭转刚度对系统性能影响Bode图

Fig. 6 Bode diagram of effect of torsional stiffness on system performance

在其他因素不变的情况下，对系统前向通道开环频率特性进行仿真，如图7所示。可知，在低频段，不同转子转动惯量的系统幅频特性曲线大致一致；在高频段，随着转子转动惯量的增大，系统的谐振峰值略有减小，但谐振频率和带宽减小，响应速度变慢。为保证系统在高频率情况下的性能，应在保证系统正常工作的前提下尽量减小转子转动惯量，采用小惯量电机。

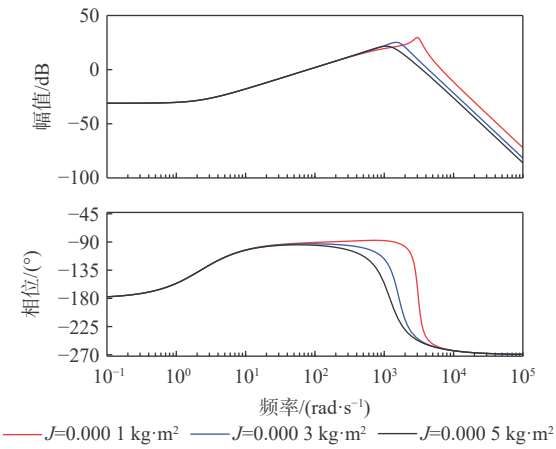


图7 电机转子转动惯量对系统性能影响Bode图

Fig. 7 Bode diagram of effect of motor rotor moment of inertia on system performance

3.2 参数对多余力的影响

3.2.1 多余力大小分析

针对表2中每个参数，在参数变化范围内平均采样，随机选取500组数值进行仿真，记录每次系统运行仿真过程中多余力的最大值，以标称多余力为基准，从多余力的变化分布范围来综合分析确定各参数对系统多余力的影响程度。仿真结果如图8所示。图中，超出标称多余力的比例越高，表明超出标称多余力值的次数越多。

由图8可知：

1）图8(a)中，在驾驶员不同的操纵频率情况下，有50%会超出标称多余力。随着驾驶员操纵频率增大，电机的加载频率增大，导致电机转速波动变大，系统输出多余力增大[26]。

2）图8(b)～(d)中，按照影响从大到小的顺序为：侧杆等效阻尼、侧杆等效转动惯量、扭转刚度。等效阻尼变化时，有85%会超出标称多余力，等效转动惯量变化只有45%超出标称多余力，扭转刚度变化时，有1%会超出标称多余力。阻尼和惯量是传动装置侧杆的固有特征，侧杆在驾驶员和加载电机的合力下绕枢轴旋转，大阻尼和大惯量阻碍了侧杆快速运动，有助于减小系统多余力。

3）图8(e)～(h)中，电机的转子转动惯量、转矩系数、相电阻、相电感变化导致了多余力小于1%的范围超出了标称多余力。电机的参数代表了电机的工作能力，参数变化导致电机工作能力下降，电机工作能力仍能满足系统要求的前提下，电机参数变化不会对系统多余力造成影响。

4）图8(i)～(o)中，可得出：

① 图8(i)、(j)中，力矩传感器信号延迟造成系统不能正常工作，力指令信号延迟导致80%超出标称多余力，转速传感器信号延迟造成了2%超出标称多余力。传感器信号是控制器的反馈回路，对于三闭环控制器，外环的信号延迟会造成多余力超出，甚至系统不能正常工作，内环反馈回路的信号延迟对系统性能和多余力的影响小于外环。

② 图8(k)～(m)中，力矩传感器的信号偏置造成50%超出了标称多余力，转速传感器信号偏置造成40%超出了标称多余力，电流传感器的信号偏置造成了2%超出了标称多余力。说明控制器反馈回路的信号偏置会造成系统多余力增大，且内环的影响小于外环。

③ 图8(n)、(o)中，力矩传感器信号的噪声导致系统不能正常工作，转速传感器信号噪声造成98%超出标称多余力，电流传感器信号噪声导致100%超出标称多余力。控制器反馈回路信号噪声导致反馈信号精度下降，因此，反馈外环的信号噪声导致系统不能正常工作，反馈内环信号噪声导致系统多余力增大。

3.2.2 多余力频谱分析

根据蒙特卡罗仿真结果，选择使多余力数值最大的对应参数进行多余力的频谱分析，分析系统参数变化对多余力变化频率的影响。由典型参数的频谱分布(见图9)和参数频谱分布(见表3)可知，系统参数变化导致多余力频谱分布变化，主要体现在

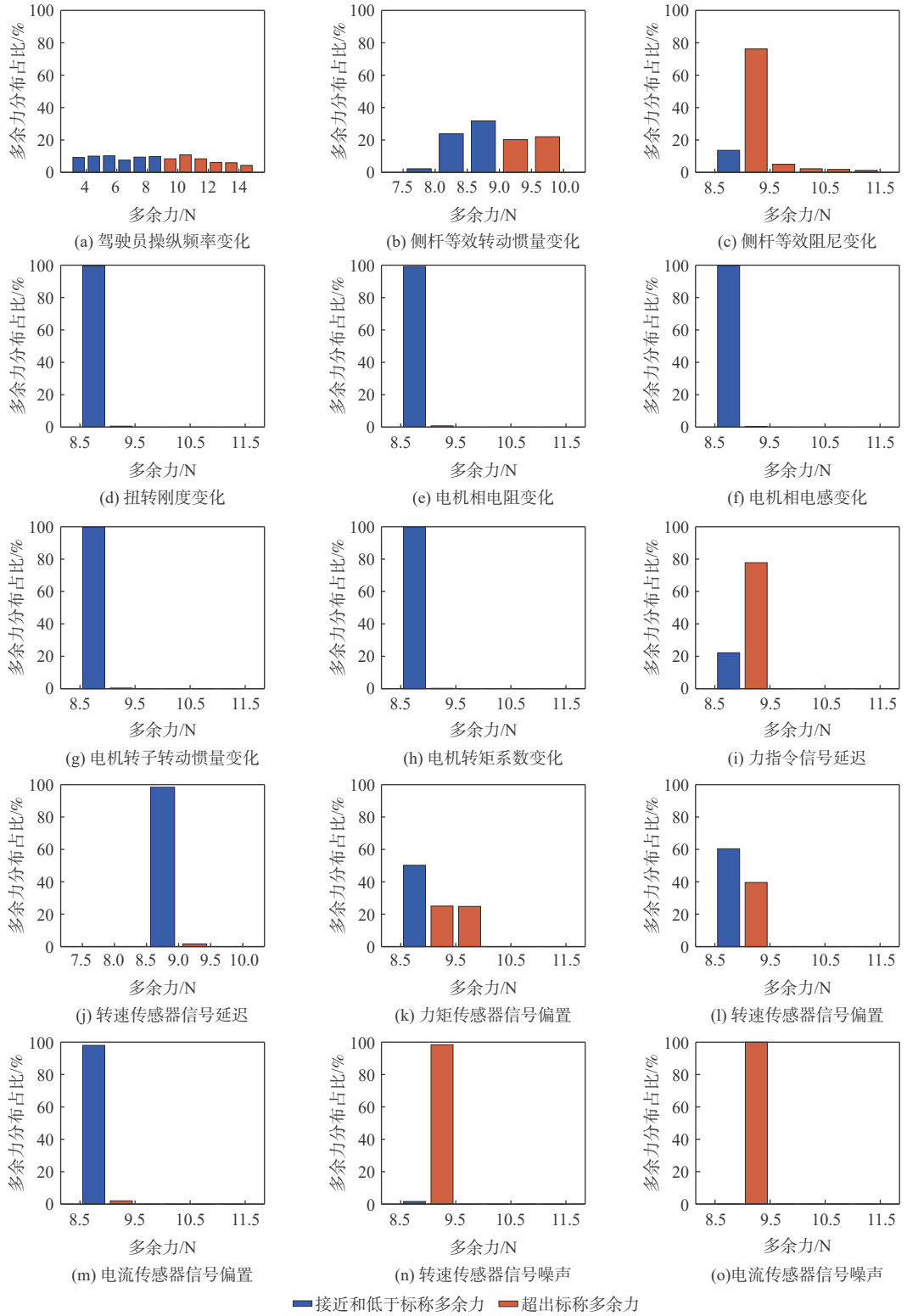


图 8　参数对多余力分布占比的影响

Fig. 8　Influence of parameters on distribution proportion of surplus force

主频率变化和总谐波失真 (total harmonic distortion, THD) 变化。主频率代表多余力各频率分布中幅值最大的频率, THD 代表多余力中主频率之外的谐波, 比值越大, 则谐波越丰富。对驾驶员操纵影响较大的因素为主频率及其幅值, 主频率越大、幅值越高, 对驾驶员操纵感觉影响越大。THD 值对驾驶员操纵感觉的影响较小, 但其越大, 则影响越明显。由表 3 可知, 侧杆等效转动惯量和力矩传感器信号偏置会导致多余力主频率减小至 5.18 Hz, 同时, 前者主频率幅值增大, 后者主频率幅值减小, THD 值都增大, 如图 9(c)、(e)所示。侧杆等效阻尼和驾驶员操纵频率变化, 则会导致多余力 5.7 Hz

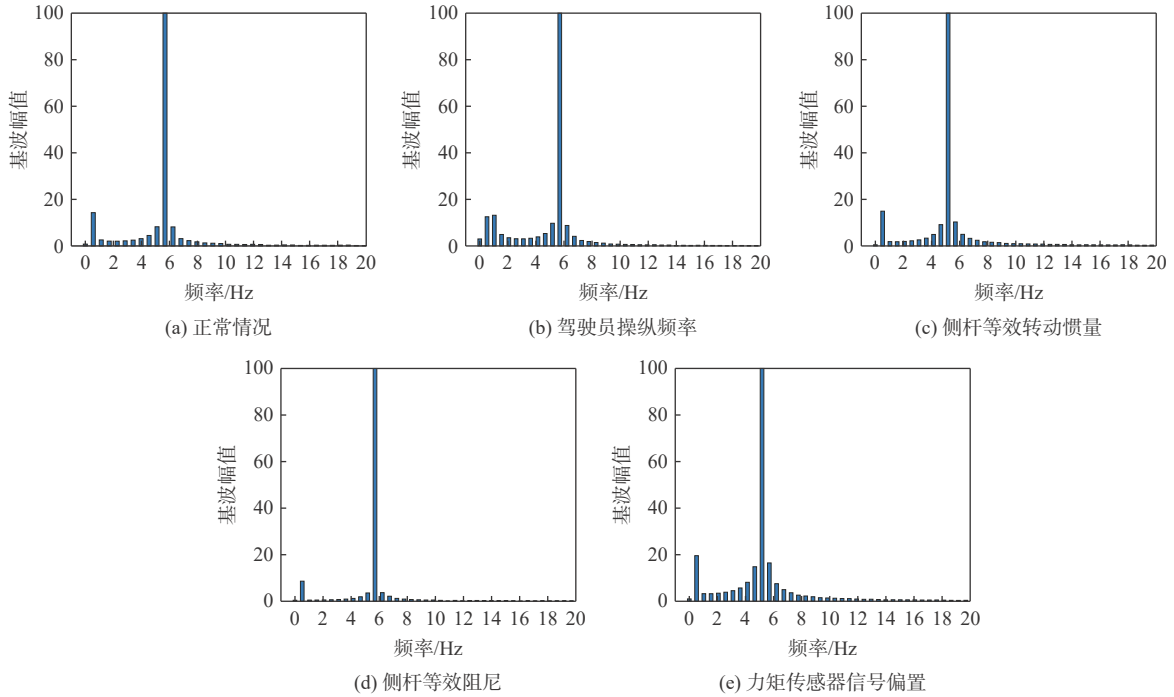


图9 单参数多余力频谱图

Fig. 9 Signal parameter surplus force spectrum diagram

表3 多余力频谱仿真结果

Table 3 Spectrum simulation results of surplus force

参数	主频率/Hz	主频率幅值	THD/%
电流传感器信号噪声	5.7	6.418	12.85
力矩传感器信号偏置	5.18	6.444	20.07
电机转矩系数	5.7	6.742	10.94
转速传感器信号噪声	5.7	6.747	11.09
扭转刚度	5.7	6.782	10.70
转速传感器信号延迟	5.7	6.784	10.67
电机相电阻	5.7	6.784	10.69
电机相电感	5.7	6.784	10.69
电机转子转动惯量	5.7	6.8	10.56
电流传感器信号偏置	5.7	6.924	9.88
力指令信号延迟	5.7	7.543	6.73
转速传感器信号偏置	5.7	7.722	5.93
侧杆等效转动惯量	5.18	7.724	12.97
侧杆等效阻尼	5.7	9.648	4.91
驾驶员操纵频率	5.7	10.68	10.55

注: 正常情况下, 主频率为5.7 Hz, 主频率幅值为6.933, THD为9.70%。

的主频率幅值增大, 如图9(d)、(b)所示。但是, 侧杆等效阻尼变化会导致高次谐波减少, 驾驶员操纵频率变化则会导致高次谐波增多。

由表3可知, 力指令信号延迟和转速传感器信号偏置均会导致主频率幅值增大和THD减小。除此之外, 电机参数变化和电流、转速内环的噪声、延迟、偏置等参数均会导致多余力主频幅值减小和THD增大。上述结果与3.2.1节多余力数值分析的结果一致。

3.2.3 结果分析

仿真结果表明, 系统参数变化不仅导致系统多余力数值增大, 而且导致多余力主频率及其幅值发生变化, 同时影响高次谐波的占比。主频率及其幅值增大导致驾驶员操纵感觉变差, 由于谐波幅值较小, 对驾驶员操纵感觉的影响较小。总体来说, 对系统多余力影响较大的因素主要为驾驶员操纵频率、侧杆的等效阻尼和等效转动惯量变化、力矩传感器信号偏置、力指令信号延迟、转速和电流传感器信号噪声、转速传感器信号偏置。对系统性能影响最大的因素为力矩传感器信号延迟和噪声, 小幅度的噪声和延迟便会造成多余力增大, 更大幅度的噪声和延迟则会导致系统无法正常工作。力矩传感器是控制器最外环, 因此, 有必要通过传感器冗余、滤波等手段提高力矩传感器信号的精度, 以确保系统的工作能力。

3.3 参数敏感性分析

针对表2选定的参数, 参数变化时多余力的变化量可以表征参数对多余力的影响程度, 即为参数敏感度。在此引入表征参数敏感度的指标函数$s_e(x)$, 指标函数表征了参数每次变化1%时所引起的多余力的变化程度, σ为修正因子。选取最大敏感度作为参数的敏感度指标, 并确定敏感区间。敏感区间是散点图中$G_r(x)$斜率最大的一段, 可以为系统控制性能验证和鲁棒性验证提供参考, 图10

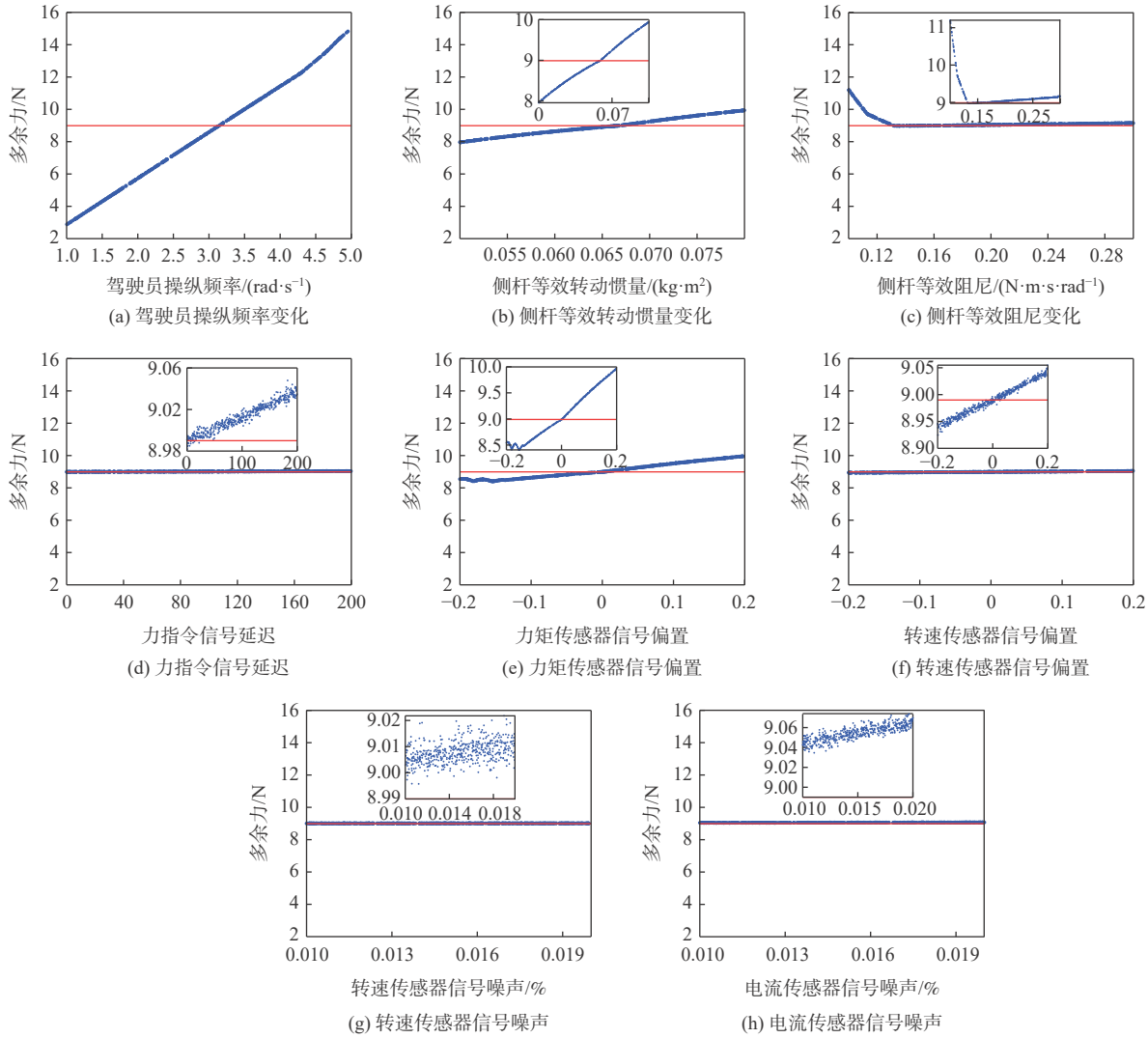


图 10　单参数变化时多余力分布散点图

Fig. 10　Surplus force distribution scatter plot under single parameter variation

中红色横线表示标称多余力。指标函数为: $s_e(x)=\max\left(\dfrac{G_r(x+\Delta x)-G_r(x)}{\Delta x\cdot\sigma}\right)$。通过公式计算多余力敏感度参数, 将各参数对应的敏感度及敏感度区间整理于表 4 中。

通过对多余力影响参数的敏感度分析得到对多余力敏感度影响的优先级为: 驾驶员操纵频率、侧杆等效转动惯量、侧杆等效阻尼、力矩传感器信号偏置, 与上文分析结果一致。而多余力对以下参数的敏感度较小: 转速传感器信号偏置和噪声、电流传感器信号噪声、力指令信号延迟。驾驶员操纵频率越高, 系统多余力越大, 随着操纵频率增大, 多余力从 2.88 N 增长至 14.82 N。这是因为电机加载能力有限, 驾驶员操纵频率的增大导致了多余力增大。

侧杆等效转动惯量增大, 系统的多余力增大, 这是因为随着转动惯量越大, 侧杆改变状态需要的驱动力越大, 导致系统多余力越大。侧杆等效阻尼在 [0.1, 0.14] N·m·s/rad 区间时, 多余力从 10.52 N 减小至 8.98 N, 等效阻尼较小时, 侧杆对操纵力的响应快, 而加载电机加载能力有限导致多余力较大, 随着等效阻尼增大, 系统的多余力逐渐减小。侧杆等效阻尼在 [0.14, 0.3] N·m·s/rad 区间时, 多余力逐渐增大, 从 8.98 N 增长到最大 9.15 N, 这说明主动侧杆系统中存在一个最优阻尼, 可以满足在其

表 4　参数敏感性

Table 4　Parameter sensitivity

参数	敏感度	敏感区间
驾驶员操纵频率	305.526	All
侧杆等效转动惯量	303.347	All
侧杆等效阻尼	241.265	0.101～0.133
力矩传感器信号偏置	12.576	−0.15～0.2
转速传感器信号偏置	0.275 9	All
电流传感器信号噪声	0.183 2	All
转速传感器信号噪声	0.107 5	All
力指令信号延迟	0.094 2	All

他参数不变情况下多余力水平最小，在本系统中最优阻尼为 0.133 8 N·m·s/rad。力矩传感器信号负偏置导致系统多余力小于标称多余力，最小为 8.414 N，力矩传感器信号正偏置导致系统多余力大于标称多余力，偏置程度越大，多余力水平越大，最大为 9.96 N。力矩传感器是三闭环系统最外环反馈通道传感器，传感器信号负偏置时，反馈值小于实际值导致转速环输入指令变大，电机加载能力提升，导致系统多余力减小。传感器信号正偏置时，反馈值大于实际值导致转速环输入指令变小，电机加载能力下降，导致系统多余力增大。说明可以通过转速前馈方式提升电机加载能力和响应速度，进而减小系统多余力。

敏感度的数值和敏感区间代表了参数对系统的影响程度，参数越高、区间越大，影响的程度越大，在开发和设计过程中，要考虑参数设置对系统的影响，在故障检测设计过程中，也可以参考参数的敏感度及敏感区间以提高故障检测的快速性和准确性。

3.4 多参数影响分析

3.4.1 多余力大小分析

在主动侧杆系统实际工作过程中，各参数的变化通常一起出现，为明确各参数同时变化时系统多余力的发生情况，选择表 2 中的参数，同时改变参数做 1 000 组仿真，分析系统多余力的发展变化情况。考虑到驾驶员操纵频率对多余力的影响最大，仿真探究了其他因素叠加时驾驶员操纵频率对多余力的影响，以及驾驶员操纵频率固定时其他各因素组合作用对多余力的影响，如图 11 所示，图中红色横线表示标称多余力。

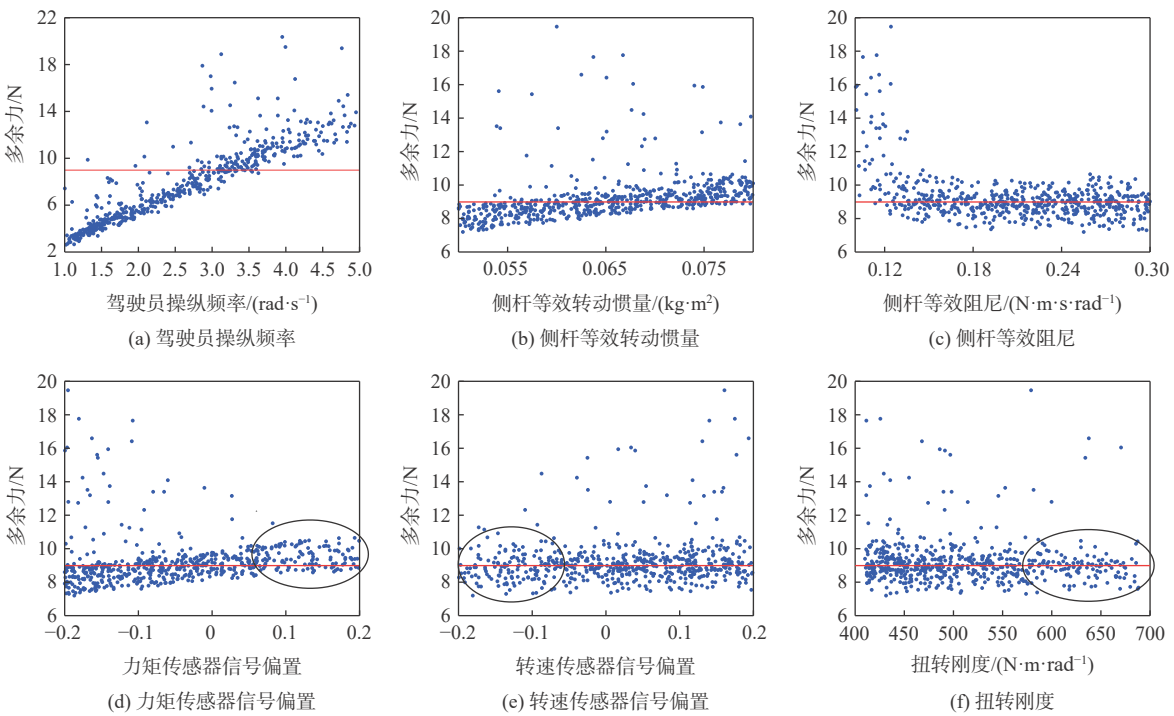


图 11 多参数变化时多余力分布散点图

Fig. 11 Surplus force distribution scatter plot under multi-parameter variation

仿真结果显示了不同参数变化对多余力均存在不同程度的影响，在参数变化的不同区间呈现不同程度的影响。驾驶员操纵频率对多余力的影响最为直接，操纵频率越大，则多余力越大。

侧杆等效阻尼、侧杆等效转动惯量、力矩传感器信号偏置、转速传感器信号偏置、扭转刚度变化等因素变化导致多余力在标称多余力附近波动。具体分析如下：

1）随着侧杆等效转动惯量参数变大，导致多余力呈现整体变大趋势，多余力数值从 7.20 N 增大至 10.63 N（图 11 中上部离散点与侧杆等效阻尼有关，在此不作考虑），说明转动惯量增大会使多余力变大。这是因为转动惯量越大，在侧杆速度变化时改变状态需要的驱动力越大，导致系统多余力越大。

2）侧杆等效阻尼变化对多余力的影响集中体现在：等效阻尼数值较小时，在等效阻尼小于 0.14 N·m·s/rad 时，系统多余力分布于 7.93～19.46 N 区间内。等效阻尼大于 0.14 N·m·s/rad 时，系统多余力均匀分布于 7.20～10.65 N 区间内。这是由于等效阻尼较小时，侧杆对操纵力的响应快，而加载电机加载能力有限导致多余力较大，随着等效阻尼增大，系统的多余力保持稳定。

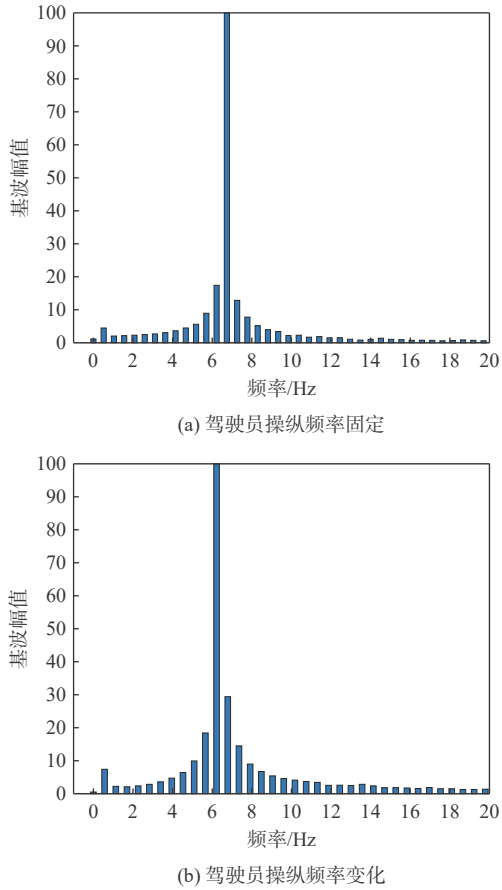
3) 力矩传感器信号偏置会造成多余力变化, 随着偏置参数由负到正的变化, 多余力呈现整体增大的趋势, 多余力数值从 7.20 N 增大至 10.65 N(图 11 中上部离散点与等效阻尼有关, 在此不作考虑)。力矩传感器是三闭环系统最外环反馈通道传感器, 传感器信号负偏置时, 反馈值小于实际值导致转速环输入指令变大, 电机加载能力提升, 导致系统多余力减小。传感器信号正偏置时, 反馈值大于实际值导致转速环输入指令变小, 电机加载能力下降, 导致系统多余力增大。

4) 转速传感器信号偏置对多余力的影响程度不明显, 随着传感器信号偏置数值变化, 多余力分布在 7.20～10.65 N 区间内。

5) 扭转刚度参数变化对系统多余力的影响程度不明显, 随着扭转刚度参数数值变化, 多余力分布在 7.20～10.65 N 区间内。

3.4.2　多余力频谱分析

根据多参数蒙特卡罗仿真结果, 选择使多余力最大的参数进行多余力频谱分析, 分析系统参数变化对多余力频率变化的影响。考虑到驾驶员操纵频率对多余力的影响, 仿真探究了其他因素叠加时驾驶员操纵频率对多余力的影响, 以及驾驶员操纵频率固定时其他各因素组合作用对多余力的影响, 仿真结果如图 12 所示。驾驶员操纵频率固定时, 多余力主频率增大至 6.74 Hz, 且主频率幅值和 THD 均增大。驾驶员操纵频率变化时, 多余力主频率增大至 6.22 Hz, 主频率幅值和 THD 均增大, 且大于驾驶员操纵频率固定的情况。这说明多参数变化会导致系统工作状态变化, 电机工作特性变化, 进而导致系统多余力主频率及其幅值变大, 同时, 高次谐波占比增大, 使驾驶员操纵感觉变差。



(a) 驾驶员操纵频率固定

(b) 驾驶员操纵频率变化

图 12　多参数多余力频谱图

Fig. 12　Multiparameter surplus force spectrum diagram

3.4.3　结果分析

由上述结果可知, 在多参数变化情况下, 对多余力的影响按从大到小的顺序为: 驾驶员操纵频率、侧杆等效阻尼、侧杆等效转动惯量、力矩传感器信号偏置。

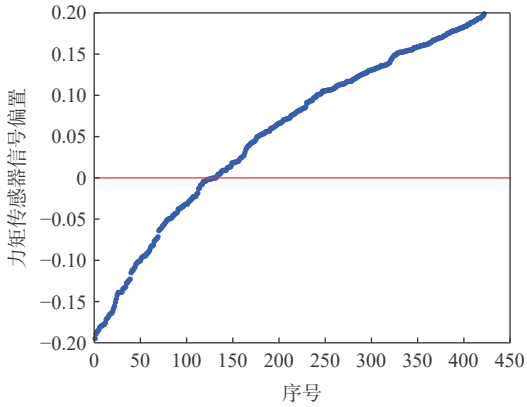
仿真中出现了多参数变化影响系统性能进而造成系统不能正常工作的情况, 经过梳理分析, 得到对系统性能影响较大的参数为力矩传感器信号正偏置、转速传感器信号负偏置、扭转刚度这 3 个参数导致系统不能正常工作的分布, 如图 13 所示。力矩传感器信号正偏置时, 力闭环反馈回路反馈值大于实际值, 导致系统加载力始终小于指令, 系统工作状态异常, 直到失去工作能力。转速传感器信号负偏置时, 转速环反馈回路反馈值小于实际值, 导致电机转速始终大于期望转速, 系统工作状态异常, 直到失去工作能力。力矩传感器是侧杆与加载电机的连接部件, 扭转刚度过大, 易引起谐振导致系统不能正常工作。因此, 在侧杆开发过程中, 可以考虑提高传感器冗余数量以保证反馈数据可靠性, 在控制算法开发过程中, 可考虑采用状态观测器类控制算法以补偿传感器信号异常对系统性能造成的影响, 选取扭转刚度小的力矩传感器。

综上所述, 为减小主动侧杆系统的多余力, 对主动侧杆系统的设计提出以下建议:

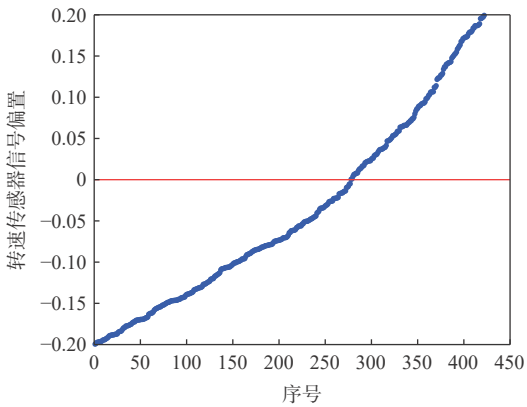
1) 尽可能减小驾驶员操纵装置的等效转动惯量(小于 0.066 7 kg·m^2), 增大驾驶员操纵装置的等效阻尼(大于 0.133 8 N·m·s/rad), 同时, 有助于防止在驾驶员较为激进的操纵行为下引发驾驶员诱发振荡(pilot-induced oscillation, PIO)。

2) 针对加载电机, 选取转动惯量小的电机有助于减小系统多余力。

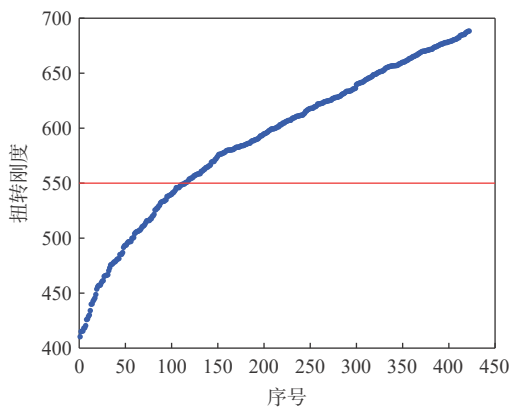
3) 针对力矩传感器, 选取扭转刚度小的传感器有助于减小系统多余力(小于 600 N·m/rad)。考虑到传感器的信号精度, 尤其是力矩传感器的信号精度会对系统多余力造成影响, 在主动侧杆系统设计时可以考虑增加力矩传感器冗余, 通过信号加权、滤波等方式弥补传感器信号精度退化问题。在控制算法开发过程中, 可考虑采用前馈等控制算法提



(a) 力矩传感器信号偏置分布散点图



(b) 转速传感器信号偏置分布散点图



(c) 扭转刚度分布散点图

图 13 故障情况影响参数分布散点图

Fig. 13 Fault condition affects parameter distribution scatter plot

高系统的响应速度，考虑采用状态观测器类控制算法以补偿传感器信号异常对系统性能造成的影响。

4 结 论

1）通过单参数蒙特卡罗仿真和多余力频谱分析确定了对多余力影响较大的因素为：驾驶员操纵频率、侧杆的等效转动惯量和等效阻尼、力矩传感器和转速传感器信号偏置、转速传感器和电流传感器信号噪声。

2）对各因素进行敏感性分析，确定了对多余力敏感度影响较大的因素为：驾驶员操纵频率、侧杆等效转动惯量和侧杆等效阻尼。

3）对多参数共同作用下的多余力影响分析验证了单参数影响和敏感性分析的结果，并进一步明确了参数的影响机理。

4）可以参考本文方法进行结构参数设计和设备优选，在算法开发过程中可关注传感器信号精度对多余力的影响。

后续可基于研究结果开发符合驾驶员操纵需求的多余力抑制算法，也可考虑驾驶员操纵品质问题进行多余力抑制算法调优。

参考文献（References）

[1] YAO Z Y, YANG X Z, QIU X Y. Research and design of the control law of passive side-stick system with human factor fault[J]. IOP Conference Series: Materials Science and Engineering, 2020, 751(1): 012014.

[2] FU W, VAN PAASSEN M M, MULDER M. Developing active manipulators in aircraft flight control[J]. Journal of Guidance, Control, and Dynamics, 2019, 42(8): 1755-1767.

[3] LIANG Q X L, LIU S Q, LU Z R, et al. Adaptive sliding mode based servo tracking control of aircrafts' active inceptor control systems[J]. Journal of Aeronautics, Astronautics and Aviation, Series A, 2017, 49: 385-396.

[4] 孟小宝, 邹泉, 何欢欢, 等. 安装主动侧杆的飞机的人机交互控制系统研究[J]. 航空科学技术, 2023, 34(8): 57-67.
MENG X B, ZOU Q, HE H H, et al. Study on the human-machine interaction control system of aircraft equipped with active side stick[J]. Aeronautical Science & Technology, 2023, 34(8): 57-67(in Chinese).

[5] SAE International. Aerospace active inceptor systems for aircraft flight and engine controls[R]. Warrendale: SAE International, 2018.

[6] CHEN H X, SHEN D K, YANG X. Effects of mechanical properties of electric loading system on surplus torque[C]//Proceedings of the IEEE Chinese Guidance, Navigation and Control Conference. Piscataway: IEEE Press, 2017: 913-917.

[7] 陈悦. 飞机主动侧杆高精度杆力控制研究[D]. 南京: 南京航空航天大学, 2020.
CHEN Y. Research on high accuracy force control of aircraft active side-stick[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2020(in Chinese).

[8] 张旭. 高性能主动侧杆控制方法研究与实现[D]. 南京: 南京航空航天大学, 2021.
ZHANG X. Research and implementation of high performance active side-stick control method[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2021(in Chinese).

[9] ZHOU Y J, LIU J, WANG Q Y, et al. Mixed sensitivity servo control of active control systems[J]. Machines, 2022, 10(10): 842.

[10] 孙晓哲, 侯东, 杨建忠. 双余度机电作动器力纷争机理及敏感性[J]. 航空学报, 2023, 44(增刊 1): 727661.
SUN X Z, HOU D, YANG J Z. Mechanism and sensitivity of force fight in dual redundant electromechanical actuators[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(Sup 1): 727661(in Chinese).

[11] FELLAH K, GUIATNI M, OURNID A K, et al. Fuzzy-PID side-stick force control for flight simulation[J]. The Aeronautical Journal, 2016, 120(1227): 845-872.

[12] ZHOU G R, LIU S Q, SANG Y J, et al. LPV robust servo control of aircraft active side-sticks[J]. Aircraft Engineering and Aerospace Technology, 2020, 92(4): 599-609.

[13] ZHOU K, AI M, SUN Y C, et al. PMSM vector control strategy based on active disturbance rejection controller[J]. Energies, 2019, 12(20): 3827.

[14] 王志刚, 李锋, 姚李阳, 等. 主动操纵装置电机伺服控制方法研究[J]. 飞机设计, 2020, 40(5): 4-7.
WANG Z G, LI F, YAO L Y, et al. Research on servo motor control method of active control device[J]. Aircraft Design, 2020, 40(5): 4-7(in Chinese).

[15] 贾会栋. 飞机主动杆结构设计及控制策略研究[D]. 大连: 大连理工大学, 2019.
JIA H D. Research on structural design and control strategy of aircraft active side-stick[D]. Dalian: Dalian University of Technology, 2019(in Chinese).

[16] XUE H J, WU H G, ZHANG X Y. Analysis of friction effect on dynamic characteristics of artificial feel system[J]. Applied Mechanics and Materials, 2013, 401-403: 460-464.

[17] 王灿, 杨忠, 陈旭, 等. 飞机主动侧杆非线性系统的建模与控制[J]. 应用科技, 2023, 50(5): 157-162.
WANG C, YANG Z, CHEN X, et al. Modeling and control of nonlinear system of aircraft active sidestick[J]. Applied Science and Technology, 2023, 50(5): 157-162(in Chinese).

[18] 陈致君, 张诚, 朱姝姝, 等. 飞机主动侧杆的弹簧-阻尼系统力矩控制策略[J]. 飞控与探测, 2023, 6(1): 32-40.
CHEN Z J, ZHANG C, ZHU S S, et al. Torque control for spring-damping system of aircraft active side-stick[J]. Flight Control & Detection, 2023, 6(1): 32-40(in Chinese).

[19] 吴昊, 孙永荣, 孙旭东, 等. 基于 MPC 算法的飞机主动侧杆杆力控制研究[J]. 机械与电子, 2023, 41(8): 33-38.
WU H, SUN Y R, SUN X D, et al. Research on stick force control of active side stick based on MPC algorithm[J]. Machinery & Electronics, 2023, 41(8): 33-38(in Chinese).

[20] ZILINU H, LU Z R. On active sidestick force control strategy technology[C]//Proceedings of the IEEE CSAA Guidance, Navigation and Control Conference. Piscataway: IEEE Press, 2020: 1-5.

[21] CHEN Y D, SHAN L, SUN J F, et al. Surplus torque research on electric motor loading simulator using active disturbance rejection controller[C]//Proceedings of the 29th Chinese Control and Decision Conference. Piscataway: IEEE Press, 2017: 5124-5130.

[22] 郑国. 基于改进 PSO-PID 的电动负载模拟器关键技术研究[D]. 太远: 中北大学, 2023.
ZHENG G. Research on key technologies of electric load simulator based on improved PSO-PID[D]. Taiyuan: North University of China, 2023(in Chinese).

[23] CHEN Y Q, MA J, FANG Y. Mechanism modeling and analysis of surplus torque in aeroload simulator[J]. International Journal of Modeling, Simulation, and Scientific Computing, 2015, 6(1): 1550005.

[24] 何晶晶, 高亚奎, 张家盛. 液压伺服作动系统力纷争敏感度及机理研究[J]. 测控技术, 2018, 37(10): 39-43.
HE J J, GAO Y K, ZHANG J S. Sensitivity and mechanism for force fighting of servo-hydraulic actuation system[J]. Measurement & Control Technology, 2018, 37(10): 39-43(in Chinese).

[25] 陈宇青. 电动负载力矩模拟器多余力矩产生机理及抑制方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2015.
CHEN Y Q. Research on the mechanism of excess torque in electric load torque simulator and its suppression method[D]. Harbin: Harbin Institute of Technology, 2015(in Chinese).

[26] ZAICHIK L E, YASHIN Y P, DESYATNIK P A, et al. Handling quality of aircraft equipped with sidesticks: AIAA 2014-2589[R]. Reston: AIAA, 2014.

Mechanism and sensitivity analysis of surplus force in active sidestick system

YANG Jianzhong，ZUO Jinpeng，SUN Xiaozhe*

(College of Safety Science and Engineering，Civil Aviation University of China，Tianjin 300300，China)

Abstract：In the active sidestick system, there is a problem of surplus force, which causes problems such as ‘jigging rod beater’, poor situational awareness, and easy to induce flight accidents under special circumstances. The parameters associated with the surplus force of the system were first identified through theoretical analysis. The influence of each parameter on the surplus force of the system was then thoroughly examined through Monte Carlo simulation. Based on the results of the Monte Carlo simulation, the spectrum analysis of the surplus force and the sensitivity analysis of the influence of the surplus force of each parameter were conducted in order to better understand the mechanism of the surplus force and determine the influence mechanism and law of the system parameters on the surplus force. The driver’s control frequency, the equivalent inertia of the sidestick, the equivalent damping of the sidestick, the signal bias of the speed sensor, and the signal bias of the torque sensor are among the factors that have the biggest effects on the surplus force of the system. The degree to which each parameter influences the surplus force is determined through qualitative and quantitative analysis, which serves as a guide for the design and compliance verification method of the surplus force suppression method of the active sidestick system.

Keywords：active sidestick；surplus force；Monte Carlo simulation；sensitivity analysis；servosystem

Received：2024-03-05；**Accepted**：2024-08-06；**Published Online**：2024-09-04 11：00

URL：link.cnki.net/urlid/11.2625.V.20240903.1848.003

* **Corresponding author**. E-mail：xzsun@cauc.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0161

基于多尺度特征融合的滚动轴承寿命预测

火久元[1,2,*]，李昕[1]，常琛[1]，李宇峰[1]，张耀南[2]

(1. 兰州交通大学 电子与信息工程学院，兰州 730070；2. 国家冰川冻土沙漠科学数据中心，兰州 730000)

摘 要：滚动轴承是机械设备中的常用部件，有效预测滚动轴承的剩余使用寿命 (RUL) 对于制定合理的维修计划和确保设备的安全性具有重要作用。传统的深度学习方法难以提取滚动轴承的多尺度退化特征，而非平稳信号噪声的存在也使 RUL 更难预测。因此，提出了一种 RUL 预测模型 EEMD-AFP-FSBLformer，该模型结合了集成经验模态分解 (EEMD)、离散小波变换 (DWT)、注意力特征金字塔 (AFP) 和 FSBLformer 网络。通过 EEMD 分解与 DWT 降噪处理，对低频模态函数与降噪处理后的高频模态函数进行滚动轴承时域退化特征处理，以产生更多具有代表性的退化特征；将退化特征输入到 AFP 网络中以提取多尺度的特征；将退化特征作为 FSBLformer 模型的输入，FSBLformer 模型的编码器引入了特征注意力机制和自注意力机制，解码器使用了双向长短期记忆 (BiLSTM) 网络，使模型在特征提取与时序预测方面更具优势。分别在 PHM2012 数据集与 XJTU-SY 数据集的不同工况下进行实验，结果表明：所提模型决定系数达 94% 以上，可有效提取滚动轴承的多尺度退化特征并预测其 RUL。

关 键 词：滚动轴承；寿命预测；集成经验模态分解；注意力特征金字塔；FSBLformer
中图分类号：TH133.3; TP183
文献标志码：A **文章编号**：1001-5965(2026)05-1391-15

随着现代工业的蓬勃发展，轴承已经是机械设备中不可缺少的部件。轴承的意外失效或故障可能导致巨大的经济损失和灾难性后果。开展滚动轴承寿命预测研究对提高机械设备运行效率、实现机械设备健康管理具有重要意义[1]。机械设备剩余使用寿命 (remaining useful life, RUL) 是在当前运行环境中从当前时刻到故障发生，即机械设备失效的时间间隔。由于机械设备的运行状态不断变化，RUL 是时刻变化的。因此，RUL 作为机械设备预测与健康管理系统的核心内容之一，其目的是能够发现早期故障，并实现对 RUL 的预测。

机械设备 RUL 预测大体可以概括为基于模型的方法、数据驱动方法和混合方法[2]。基于模型的方法是用数学或物理模型来确定退化程度。例如，Le Son[3] 等综合了来自噪声的退化数据，借助吉布斯采样技术，利用非齐次含噪声的伽马过程完成部件的退化过程建模，通过退化状态得出系统的 RUL; Carpinteri 等[4] 将裂纹的形状、大小与疲劳寿命一起分析，得出了它们之间的关系。但是，这些方法需要特定的模型，且很难准确地建立模型，在处理随机噪声方面的局限性还会导致其在 RUL 预测中发生较大的误差。与基于模型的方法完全不同，数据驱动方法通过映射监控信号与 RUL 值之间的关系来分析问题，且不需要特定的模型。因此，数据驱动方法逐渐成为 RUL 预测领域的主要方法。

在数据驱动的轴承 RUL 预测方面，主要包括人工神经网络 (artificial neural network, ANN)、隐马尔可夫模型 (hidden Markov model, HMM)、贝叶斯信念网络 (Bayesian belief network, BBN) 和支持向量机 (support vector machine, SVM) 等。例如，Loutas 等[5] 利用 SVM 结合时间序列技术，对机器设备的 RUL

收稿日期：2024-03-22；**录用日期**：2024-05-11；**网络出版时间**：2024-06-20 08:58
网络出版地址：link.cnki.net/urlid/11.2625.V.20240618.1120.002
基金项目：国家自然科学基金 (62262038)；甘肃省重点研发计划工业项目 (22YF7GA145)
* **通信作者**. E-mail：huojy@mail.lzjtu.cn

引用格式：火久元，李昕，常琛，等. 基于多尺度特征融合的滚动轴承寿命预测 [J]. 北京航空航天大学学报，2026，52（5）：1391-1405. HUO J Y，LI X，CHANG C，et al. Rolling bearing life prediction based on multi-scale feature fusion [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（5）：1391-1405 （in Chinese）.

进行预测；Rathore等[6]使用一种综合特征选择策略，包括粒子群优化(particle swarm optimization, PSO)技术和特征转换，将敏感预测特征集输入双向长短期记忆(bidirectional long short-term memory, BiLSTM)网络，进行RUL预测。然而，这些方法在处理非平稳的长序列数据时效率较低，预测精度也并不是很理想，因此，针对非平稳信号提出混合方法是非常重要的。

常见的混合方法主要通过数据预处理将原始序列分解为若干个连续序列，再对每个子序列分别建模，将所有连续序列的预测结果叠加，得到满足精度要求的组合预测结果[7]。例如，Gao等[8]提出一种使用集成经验模态分解(ensemble empirical mode decomposition, EEMD)、能量矩熵和增强长短期记忆(long short-term memory, LSTM)的方法进行轴承早期故障预测，通过选择敏感的固有模态函数(intrinsic mode function, IMF)，并利用增强的LSTM通过特征提取进行故障预测；Li等[9]提出一种基于风险评估和退化状态系数的RUL预测方法，根据振动信号在时域和时频域中的特性来计算马氏距离(Mahalanobis distance, MD)，通过使用变分模态分解-奇异值分解(variational mode decomposition-singular value decomposition, VMD-SVD)提取时频域中的特征，使用MD1-CUMSUM获得单调递增的健康指标(health indicator, HI)，提出风险评估来自适应地确定初始故障和故障阈值，从而进行测试轴承的RUL预测；Guo等[10]通过经验模态分解(empirical mode decomposition, EMD)，提出一种基于欧几里得距离和余弦相似度的相似度测量方法，以选择具有更多退化特征的IMF，提高了RUL预测准确度。序列预处理方法主要包括EMD和小波分解(wavelet decomposition, WD)等。EMD作为一种优秀的时频分析方法，适用于非线性、非平稳信号的分析[11]，其根据信号的内在特征，将动态信号分解为一系列的IMF函数，有效提取了信号的本质特征，但也存在边界效应和模态混叠方面的问题。为解决上述问题，对EMD进行改进，提出了EEMD[12]，通过多次添加不同的高斯白噪声来进行集成，其目的是通过平均多次分解结果来减小噪声对IMF提取的影响，从而提高分解的稳定性和可靠性。然而，分解后的IMF中的高频IMF也含有较多无关或冗余的噪声，这些噪声对RUL预测是无用的，甚至会干扰寿命预测的效果[13]。因此，为提高预测精度，减少计算负担，有必要对含有噪声较多的高频IMF进行降噪处理，将低频IMF和降噪后的高频IMF分别进行时域特征处理。

由于条件限制，很多设备无法布置多个传感器，往往只能利用单个传感器获得的数据进行分析。因此，如何从有限的原始数据中提取更全面的退化特征是提高RUL预测精度的关键。Zhao等[14]通过并行使用不同卷积核的结构实现了对航空发动机多尺度空间退化特征的提取，并通过对比实验证明了该结构的优越性。Zhu等[15]利用多尺度卷积神经网络(convolutional neural network, CNN)实现了轴承退化特征提取。上述基于深度学习的模型在特征提取任务中取得了一定的成效，但大多只能获得单维度、单尺度的性能退化特征。因此，有必要提出可同时从设备中提取的多尺度退化特征模块。多尺度特征可以提取不同尺度的信息，能够全面反映系统或信号的变化和趋势，具有更好的鲁棒性和预测能力。

注意力机制的变体，如自我注意力[16]、空间注意力[17]及特征注意力等[18]，已在深度学习方法中得到广泛研究，以提高方法性能。因此，本文针对滚动轴承RUL预测提出FSBLformer模型，其具有特征注意力机制与自注意力机制，可以提取特征表示，且能帮助模型选择和突出输入特征中最具信息量或最相关的部分。BiLSTM作为全连接层也可以帮助捕捉输入序列的上下文信息和时序关系。因此，可将多尺度特征作为FSBLformer网络的输入，对滚动轴承RUL进行预测。

鉴于此，本文研究的必要性体现在以下方面：①由于EEMD是一种解决机械运行过程中非平稳信号的有效方法，但EEMD在分解过程中也有其自身的缺点，即对滚动轴承振动信号进行分解得到的IMF中只有一部分能够反映退化特征，其余部分都是噪声或干扰成分。因此，有必要提出一种合理的方法来对IMF进行降噪处理，使降噪后分解得到的IMF更加纯净，且准确得到不同频率分量的IMF。这些IMF能够完整地表达信号的频域特征，包括滚动轴承性能退化所引起的频率变化。②对滚动轴承振动信号进行EEMD分解、降噪和时间序列特征处理后，得到的多个时间序列特征通过注意力特征金字塔(attention feature pyramid, AFP)网络中分割、捕获、合并、堆栈等操作，进行从粗到细不同尺度的特征提取。③针对众多特征信息难以被完全提取的问题，创建了FSBLformer网络，其使用自注意力和特征注意力作为编码器的主要组成部分，旨在提取关键信息。解码器中运用BiLSTM与全连接网络更好地进行时序预测。因此，本文提出EEMD-AFP-FSBLformer预测模型，可达到在多尺度下充分提取各个时域特征、捕获特征数据全局时空信息的效果，同时在有噪声干扰时完成较为准确的RUL预测。

1 理论基础

1.1 集成经验模态分解

EEMD是一种信号处理方法，用于将非平稳信

号分解为多个 IMF。EEMD 可以从不同时间尺度上实现对非平稳信号的分解, 并且对故障特征反应较为敏感, 广泛应用于信号处理和故障诊断中, 是针对 EMD 方法在处理一些随机和噪声信号时容易出现模式混合和模式伪影问题提出的改进。EEMD 是一种基于信号本身特征的自适应信号分解方法, 其将信号分解为一系列 IMF, 每个 IMF 都应满足 2 个条件: ①在局部区域内只有一个极值点; ②均值为 0。在 EEMD 中, 将原始信号加上随机高斯白噪声, 并进行多次重复实验, 通过在多次实验中对不同随机噪声进行分解, 取平均值, 可以有效减少噪声的影响, 获得更可靠和准确的分解结果[19]。

该方法实现的具体过程如下:

1) 在原始信号 $x(t)$ 中添加白噪声 $w(t)$, 将添加的白噪声次数记为 N 次, 如下:

$$x_i(t)=x(t)+w_i(t)\qquad i=1,2,\cdots,N \tag{1}$$

2) 将处理后的原始信号采用 EMD 方法分解得到 m 个 IMF 分量和 1 个趋势项 $r_i(t)$, 如下:

$$x_i(t)=\sum_{j=1}^{m}C_{ij}(t)+r_i(t) \tag{2}$$

3) 重复添加白噪声, 将每次得到的 IMF 分量进行集成平均处理, 作为最终的分解结果, 如下:

$$I_i=\frac{1}{N}\sum_{i=1}^{N}C_{ij}(t) \tag{3}$$

式中: $x_i(t)$为原始信号; $w_i(t)$为符合高斯正态分布的白噪声信号; N 为添加的白噪声次数; $C_{ij}(t)$为添加的第 j 个白噪声后的第 i 个 IMF 分量。

1.2　离散小波变换

对于非平稳信号, 本文采用离散小波变换 (discrete wavelet transform, DWT) 和离散小波逆变换对输入数据进行预处理。根据 Percival 和瓦尔登在 2000 年提出的小波分析理论, 原始时间序列信号 $x(t)$ 的 DWT 定义如下[20]:

$$\varphi_{j,k}(t)=a_0^{-j/2}\varphi\left(a_0^{-j}-b_0k\right) \tag{4}$$

$$\mathrm{DWT}(j,k)=\int_{-\infty}^{\infty}x(t)\varphi_{j,k}^{*}(t)\,\mathrm{d}t \tag{5}$$

式中: $\varphi(t)$为母小波; φ^*为母小波的复共轭; a_0、b_0为常数; 伸缩参数 j 和平移参数 $k(j,k\in\mathbf{R})$为 2 个标度参数, 分别决定了振荡频率和位移位置。

1.3　特征金字塔

特征金字塔用于在不同尺度上进行特征提取, 其思想是通过构建多层级的特征金字塔来解决问题。特征金字塔是一种多尺度的特征表示, 每个层级包含了原始数据的不同分辨率版本, 通过在每个层级上提取特征, 使得方法能够对不同尺度的特征进行提取[21]。一般来说, 金字塔的每个层级都对应不同的分辨率, 在每个层级上应用特征提取方法, 如使用 CNN 提取特征, 通过特征金字塔, 可以获得来自不同尺度的特征表示, 这些特征表示可以捕捉到不同大小的目标和细节信息[22]。

1.4　Transformer

当需要处理滚动轴承等复杂时序数据, 且这些数据具有长距离依赖关系和高维信息时, 使用 Transformer 模型能够更有效地捕捉数据之间的复杂关系, 并提供更好的预测性能。Transformer 能够在编码和解码阶段同时访问输入序列的所有位置, 因此, 能够更好地捕捉全局信息, 也可以同时处理输入序列中的所有位置, 使得 Transformer 在处理长序列时效率更高。Transformer 擅长提取上下文特征, 因为自注意力机制允许其根据序列中不同元素与其他元素的关系来衡量它们的重要性。

Transformer 模型由编码器和解码器组成。编码器用于将输入序列编码作为一系列的隐藏表示, 而解码器则将这些隐藏表示解码为输出序列。Transformer 模型使用多层自注意力层和前馈神经网络层来实现编码器和解码器, 通过在编码器和解码器之间进行自注意力计算、位置编码和前馈网络映射等操作, 实现对输入序列的编码和生成目标序列的解码过程。

2　EEMD-AFP-FSBLformer 预测模型

2.1　EEMD 数据预处理

在使用 EEMD 分解后, 产生了高频 IMF 和低频 IMF。高频 IMF 主要包含信号中较高频率和较小振幅的成分, 往往包括较多噪声和细节部分。低频 IMF 更接近原始信号的整体趋势与主要特征。为使 EEMD 分解后产生的 IMF 质量更高, 针对产生的高频 IMF, 提出一种基于 DWT 并重构的降噪方法。首先, 样本熵的值越高, 说明含有噪声占比越高, 经过样本熵计算后, 确认了高频 IMF 所具备的特点, 即包括较多噪声和细节部分。然后, 对降噪后的高频 IMF 与 EEMD 分解产生的低频 IMF 进行时序特征处理, 通过该方法进行的特征提取会得到更加明显的退化特征信息, 可以对滚动轴承寿命进行更精准的预测。整体流程如图 1 所示。

2.2　注意力特征金字塔网络

预处理后产生了众多时域退化特征, 如何将其充分提取是非常具有挑战性的。针对这一问题, 诸多多尺度特征提取方法被提出, 但使用本文提出的 AFP 能够最大程度地发挥在特征提取中比较有优势的特征金字塔网络和注意力机制的优势, 其引导网络从粗到细依次发现不同尺度的显著线索, 实现全面互补的感知。

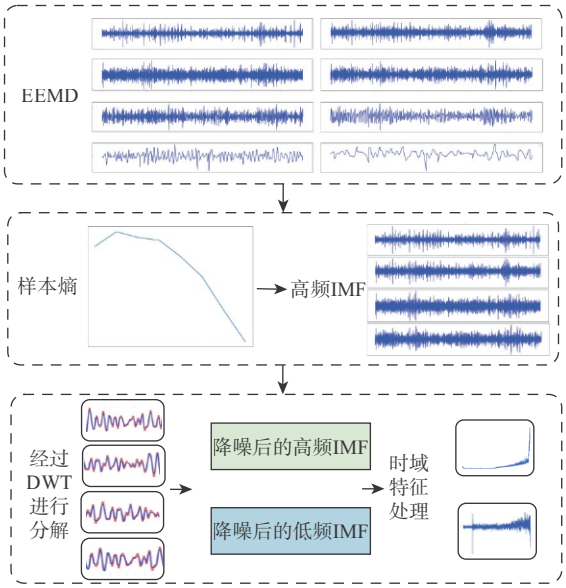


图 1 EEMD 特征提取流程

Fig. 1 EEMD feature extraction flowchart

首先，通过骨干网络块提取特征映射 $\boldsymbol{X}\in\mathbf{R}^{C\times H\times W}$，其中，$C$ 为通道维数，$H\times W$ 表示空间域。然后，学习注意力特征金字塔 P 以发现特征 $\boldsymbol{X}$ 上的多尺度显著部分。AFP 包含具有不同级别的注意力图，如 $P=\{P_i\}$，其中，P_i 为第 i 个金字塔层。在不同的金字塔层中，学习不同尺度的判别线索。粗略的注意力引导了细粒度注意力学习，其可以被公式化为一个连续的学习过程，如下：

$$\boldsymbol{X}_i=P_i(\boldsymbol{X}_{i-1}) \tag{6}$$

式中：$\boldsymbol{X}_i$ 为第 i 层中的特征图。

如图 2 所示，对于每个级别的注意力学习，先将其拆分为更细粒度，并将这些学习到的注意力合并为当前级别的注意力地图。

1）分割。将特征图 $\boldsymbol{X}_i$ 作为输入，并输出 n 个分割特征张量 $\boldsymbol{X}_{i,j}$，其中，$j=\{1,2,\cdots,n\}$ 为分割特征部分的索引。具体地，分割部分的数量是基于基数的指数增长，如当金字塔等级是 i 且基数是 2 时，$n=2^i$。随着金字塔层次 i 的增加，分割部分的数量相应增加，表明这些特征部分的粒度更细。分割操作应用于注意力模块的对应域。例如，倾向于将特征图 $\boldsymbol{X}_i$ 分割为多个部分 $\boldsymbol{X}_{i,j}\in\mathbf{R}^{C\times\frac{H}{n}\times W}$，并在通道域 C 上应用除法，以获得通道注意的 $\boldsymbol{X}_{i,j}\in\mathbf{R}^{\frac{C}{n}\times H\times W}$。

2）捕获与合并。给定分割特征张量 $\{\boldsymbol{X}_{i,j}\}$，学习一组子注意力模型 $\{A_{i,j}\}$ 来捕获每个特征张量的线索。将所有子注意力模型合并为分割操作的逆过程，以获得相同大小的整个注意力图 A_i，如下：

$$A_i=\left[A_{i,j}\left(\boldsymbol{X}_{i,j}\right)\right]_{j=1}^{n} \tag{7}$$

式中：$[\cdot]_{j=1}^{n}$ 表示 n 个注意力地图的集中度。该聚合注意力图 A_i 指示第 i 层中的显著区域。

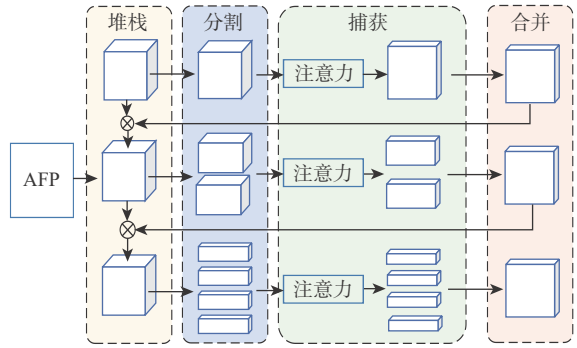


图 2 AFP 网络模型结构

Fig. 2 Architecture of AFP network model

3）堆栈。通过将注意力从粗到细堆叠成金字塔，引导网络逐步关注重要特征。有了学习到的注意力 A_i，网络发现了更多的区别性特征，如下：

$$\boldsymbol{X}_i=\sigma A_i*\boldsymbol{X}_{i-1} \tag{8}$$

式中："$*$" 表示元素级乘积；σ 表示 Sigmoid 激活函数。

这些特征被归一化的注意力图重新加权，并被馈送到下一个金字塔级别，以指导更细粒度的注意力学习。随着金字塔层数的增加，在原始粗特征的基础上发现了更多的细粒度线索。

2.3 FSBLformer 预测模型

本文基于 Transformer 模型提出了 FSBLformer 滚动轴承 RUL 预测模型，以实现对滚动轴承 RUL 的高精度预测。在编码阶段，引入 2 种不同类型的注意力，分别为特征注意力和自注意力。通过捕获特征和数据自身之间的依赖关系，增强了模型模拟时间序列数据的能力。BiLSTM 可以关注所有时间步骤的信息，通过将 Transformer 结构与 BiLSTM 相结合，更好地捕获序列中的长期依赖关系。FSBLformer 预测模型的总体架构如图 3 所示。

FSBLformer 预测模型采用编码器解码器架构。模型的每层由 3 个子层组成，分别包括多头特征注意力机制、多头自注意力机制和全连接前馈网络。每个子层之后是残差连接和层归一化。

在编码器中，采用两级特征-自多头注意力机制，由多头特征注意力机制和多头自注意机制组成，如图 4 所示。

来自给定集合的向量的加权组合可以通过注意力机制来计算。查询和键之间的相似性被缩放的点积注意力用来为值分配权重。注意力计算如下：

$$\mathrm{Att}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{9}$$

式中：$\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 分别为查询、键和值；$\sqrt{d_k}$ 为键的尺寸的平方根，被用作缩放因子以降低点积方差对特征尺寸的敏感性，确保了训练期间的稳定梯度。

在特征维度中，应用特征注意力，利用基于互

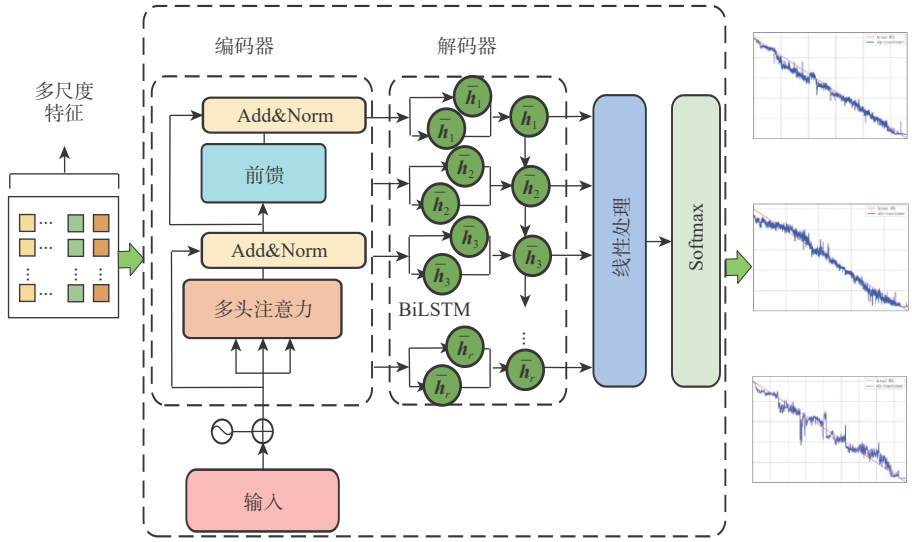


图 3　FSBLformer 预测模型总体架构

Fig. 3　The overall architecture of the prediction model

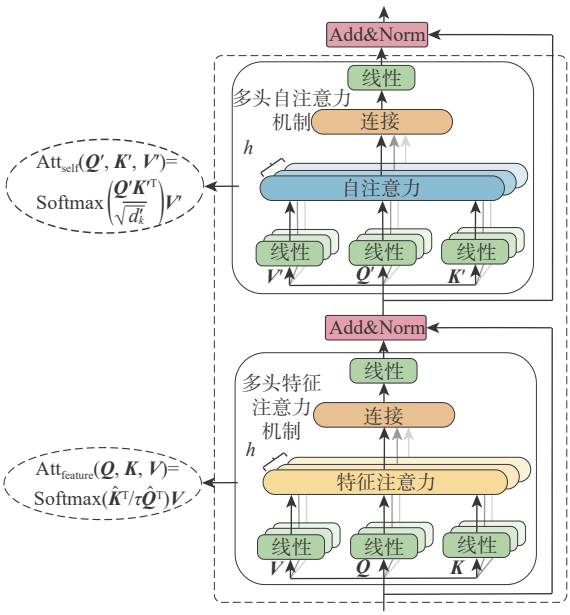


图 4　FSBLformer 预测模型编码器结构

Fig. 4　Encoder structure of FSBLformer prediction model

协方差的注意力函数[23]。特征注意力函数如下:

$$\mathrm{Att}_{\mathrm{feature}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Softmax}\left(\frac{\hat{\boldsymbol{K}}^{\mathrm{T}}}{\tau\hat{\boldsymbol{Q}}^{\mathrm{T}}}\right)\boldsymbol{V} \tag{10}$$

每个输出标记嵌入被定义为 $\boldsymbol{V}$ 中的对应响应标记嵌入特征 d_v 的凸组合。每列通过 L_2 归一化, 归一化的特征查询和关键矩阵为 $\hat{\boldsymbol{Q}}$ 和 $\hat{\boldsymbol{K}}$。互协方差矩阵的元素 $\hat{\boldsymbol{K}}^{\mathrm{T}}\hat{\boldsymbol{Q}}$ 在 [−1,1] 范围内。Softmax 中的可学习参数 τ 导致注意力权重更均匀分布。

$$\mathrm{MultiHead}_{\mathrm{feature}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Concat}(h_1,h_2,\cdots,h_h)\boldsymbol{W}^{\mathrm{O}} \tag{11}$$

式中: $\mathrm{MultiHead}_{\mathrm{feature}}$、Concat分别表示多头特征注意力、拼接; $\boldsymbol{W}^{\mathrm{O}}$ 为特征输出投影矩阵。$L_i=\mathrm{Att}_{\mathrm{feature}}\left(\boldsymbol{QW}_i^{\boldsymbol{Q}},\boldsymbol{KW}_i^{\boldsymbol{K}},\boldsymbol{VW}_i^{\boldsymbol{V}}\right)$ 表示将缩放点积特征注意力机制应用于第 i 个头部的结果, $\boldsymbol{W}_i^{\boldsymbol{Q}}$、$\boldsymbol{W}_i^{\boldsymbol{K}}$、$\boldsymbol{W}_i^{\boldsymbol{V}}$ 分别为 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 的加权矩阵。

自注意力函数如下:

$$\mathrm{Att}_{\mathrm{self}}(\boldsymbol{Q}',\boldsymbol{K}',\boldsymbol{V}')=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}'\boldsymbol{K}'^{\mathrm{T}}}{\sqrt{d_k'}}\right)\boldsymbol{V}' \tag{12}$$

式中: $\boldsymbol{Q}'$、$\boldsymbol{K}'$、$\boldsymbol{V}'$ 分别为时间查询矩阵、键矩阵、值矩阵; d_k' 为查询维度。

时间多头自注意力机制如下:

$$\mathrm{MultiHead}_{\mathrm{self}}(\boldsymbol{Q}',\boldsymbol{K}',\boldsymbol{V}')=\mathrm{Concat}\left(h_1',h_2',\cdots,h_h'\right)\boldsymbol{W}'^{\mathrm{O}} \tag{13}$$

式中: $\boldsymbol{W}'^{\mathrm{O}}$ 为时间输出投影矩阵。$\boldsymbol{H}'=\mathrm{Att}_{\mathrm{self}}\left(\boldsymbol{QW}_i^{\boldsymbol{Q}'},\boldsymbol{KW}_i^{\boldsymbol{K}'},\boldsymbol{VW}_i^{\boldsymbol{V}'}\right)$ 表示将缩放点积时间自注意力机制应用于第 i 个头部的结果, $\boldsymbol{W}_i^{\boldsymbol{Q}'}$、$\boldsymbol{W}_i^{\boldsymbol{K}'}$、$\boldsymbol{W}_i^{\boldsymbol{Q}'}$ 分别为 $\boldsymbol{Q}'$、$\boldsymbol{K}'$、$\boldsymbol{V}'$ 的加权矩阵。

解码器由多个解码器层组成, 每个解码器层都包括一个 BiLSTM 网络和一个全连接网络, 如图 5 所示。

BiLSTM 模型的前向传递公式如下:

$$\boldsymbol{i}_t=\sigma(\boldsymbol{W}_i*[\boldsymbol{h}_{t-1};\boldsymbol{x}_t]+\boldsymbol{b}_i) \tag{14}$$

$$\boldsymbol{f}_t=\sigma\left(\boldsymbol{W}_f*[\boldsymbol{h}_{t-1};\boldsymbol{x}_t]+\boldsymbol{b}_f\right) \tag{15}$$

$$\boldsymbol{o}_t=\sigma(\boldsymbol{W}_o*[\boldsymbol{h}_{t-1};\boldsymbol{x}_t]+\boldsymbol{b}_o) \tag{16}$$

$$\tilde{\boldsymbol{c}}_t=\tanh(\boldsymbol{W}_c*[\boldsymbol{h}_{t-1};\boldsymbol{x}_t]+\boldsymbol{b}_c) \tag{17}$$

$$\boldsymbol{c}_t=\boldsymbol{f}_t*\boldsymbol{c}_{t-1}+\boldsymbol{i}_t*\tilde{\boldsymbol{c}}_t \tag{18}$$

$$\boldsymbol{h}_t=\boldsymbol{o}_t*\tanh(\boldsymbol{c}_t) \tag{19}$$

式中: $\boldsymbol{i}_t$ 为时刻 t 时的输入门; 初始输出通过使用 S 形激活函数 σ 产生; $\boldsymbol{x}_t$ 为时刻 t 时的输入; $\boldsymbol{h}_{t-1}$ 和 $\boldsymbol{h}_t$ 分别为先前隐藏的信息和当前隐藏的信息; 由遗忘门在时刻 t 产生的输出表示为 $\boldsymbol{f}_t$, 其确定先前单元状态 $\boldsymbol{c}_{t-1}$ 被传输的程度, 部分传输或移除; 由 tanh 激活层生成的新信息值为候选细胞状态 $\tilde{\boldsymbol{c}}_t$,

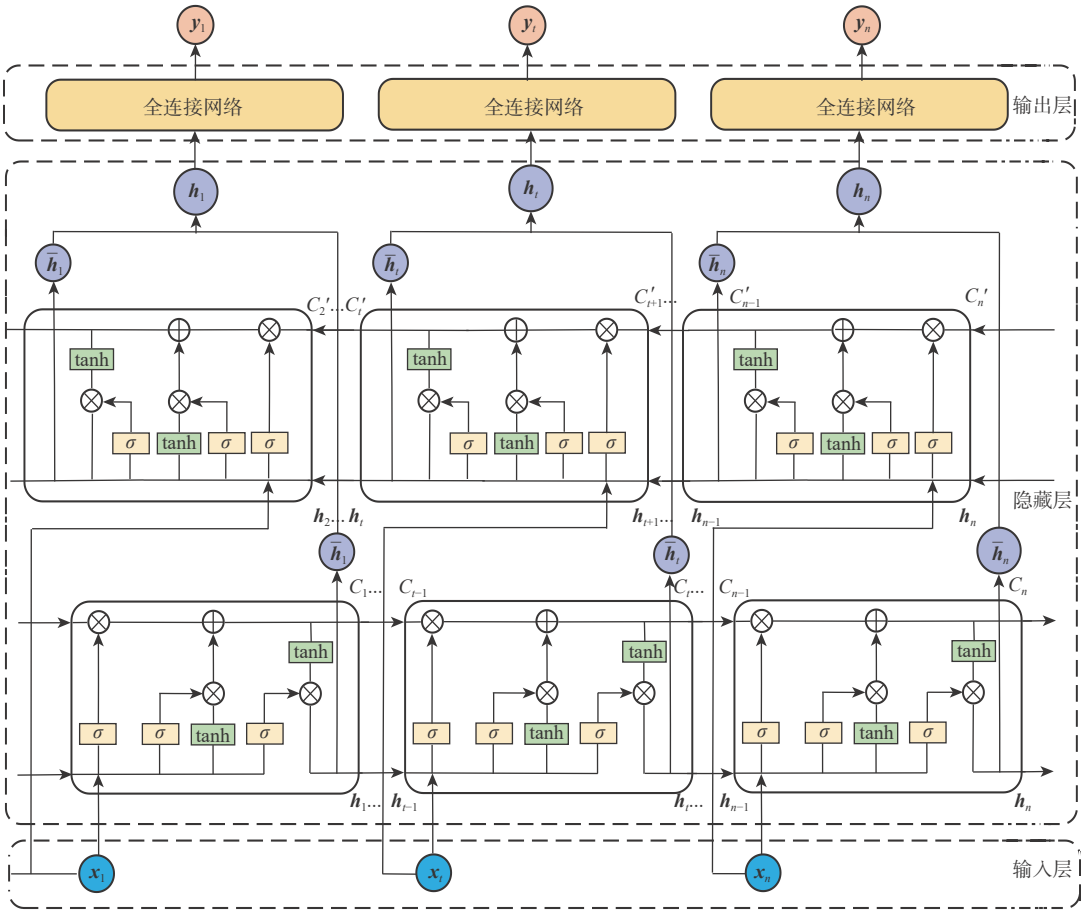


图 5 FSBLformer 预测模型解码器结构

Fig. 5 Decoder structure of FSBLformer prediction model

$\boldsymbol{c}_t$为当前细胞状态；$\boldsymbol{o}_t$为时刻 t 的输出；$\boldsymbol{W}_i$、$\boldsymbol{W}_f$、$\boldsymbol{W}_c$、$\boldsymbol{W}_o$为相应门的权重系数；$\boldsymbol{b}_i$、$\boldsymbol{b}_f$、$\boldsymbol{b}_c$、$\boldsymbol{b}_o$为相应门的偏置向量。

BiLSTM 网络模型在 2 个方向上处理预测。正向 LSTM 单元生成正向隐藏状态信息，反向 LSTM 单元生成反向隐藏状态信息。联合表示通过连接在前向和后向方向上的信息获得，如下：

$$\boldsymbol{h}_t = \left[\overrightarrow{\boldsymbol{h}_t}, \overleftarrow{\boldsymbol{h}_t}\right] \tag{20}$$

式中：LSTM 单元在时刻 t 处的正向方向上的隐藏状态由$\overrightarrow{\boldsymbol{h}_t}$表示，LSTM 单元在时刻 t 处的反向方向上的隐藏状态由$\overleftarrow{\boldsymbol{h}_t}$表示。

2.4 寿命预测框架

基于 EEMD-AFP-FSBLformer 方法的 RUL 预测具体过程如下：

步骤 1 采集滚动轴承振动信号，分别在不同工况中将轴承划分为训练集、测试集和验证集。

步骤 2 使用 EEMD 与 DWT 方法对数据进行预处理。通过对经 EEMD 分解后的 IMF 进行样本熵计算，得到含有较高噪声的 IMF，并由 DWT 来对含有较高噪声的 IMF 进行降噪处理，以减小网络规模，提高 IMF 质量。

步骤 3 分别将预处理后的训练集、测试集和验证集数据放入 AFP 网络中进行多尺度特征提取。

步骤 4 将得到的多尺度特征拼接后送入 FSBLformer 网络中进行训练，得到滚动轴承的 RUL 值，利用评价指标和可视化技术评价生成样本的质量。

3 实验验证与分析

为验证本文提出的 RUL 预测方法的性能，采用 PHM2012（PRONOSTIA 的加速老化平台）公开数据集进行实验[24]，旨在提供真实的信息来描述滚动轴承在 RUL 期间的自然退化。实验数据集是从 IEEE 2012 PHM 数据挑战中获得的振动加速度数据，采集的振动信号由 2 560 个样本组成，每 10 s 记录一次，采样频率为 25.6 kHz。数据集的具体划分信息如表 1 所示。采用的深度学习框架为 Pytorch。计算机的配置为：Inter(R) Xeon(R) Silver 4210R CPU @ 2.40 GHz 2.39 GHz, NVIDIA Tesla T4 显卡，64 GB 内存。为客观反映实验的有效性并排除偶然性，每组实验重复多次，取平均值作为最终结果。

3.1 评价指标

为评估经过 EEMD 分解后 IMF 的噪声含量，将样本熵作为评价指标使用。样本熵是一种表征序

表 1　IEEE 2012 PHM 预测挑战的数据集

Table 1　Datasets for the IEEE 2012 PHM prediction challenge

运行工况	转速/(r·min^{-1})	载荷/N	训练集	测试集
工况1	1 800	4 000	Bearing1_1, Bearing1_2	Bearing1_3,Bearing1_4, Bearing1_5,Bearing1_6, Bearing1_7
工况2	1 650	4 200	Bearing2_1, Bearing2_2	Bearing2_3,Bearing2_4, Bearing2_5, Bearing2_6,Bearing2_7

列复杂程度的测度方法, 不存在对自身数据的比较而产生计算偏差。序列的自我相似性越高, 则样本熵值越小, 时间序列越复杂, 则样本熵值越大。具体原理如下:

1) 序列$\boldsymbol{G}=[g(1),g(2),\cdots,g(N)]$构造 m 维矢量, 如下:

$$\boldsymbol{G}(i)=[g(i),g(i+1),\cdots,g(i+m-1)]\quad i=1,2,\cdots,N-m+1 \tag{21}$$

2) 定义$\boldsymbol{G}(i)$与$\boldsymbol{G}(j)$间对应元素的距离最大值为$d[\boldsymbol{G}(i),\boldsymbol{G}(j)]$, 如下:

$$d[\boldsymbol{G}(i),\boldsymbol{G}(j)]=\max_{k\in(0,m-1)}|g(i+k)-g(j+k)| \tag{22}$$

3) 给定阈值$r=0.2s$, s 为序列标准差; 统计$d[\boldsymbol{G}(i),\boldsymbol{G}(j)]<r$的数目及与总矢量个数 N–m 的比值, 如下:

$$B_i^m(r)=\frac{\text{num}\{d[\boldsymbol{G}(i),\boldsymbol{G}(j)]<r\}}{N-m+1} \tag{23}$$

4) 对式(23)结果求平均, 如下:

$$B^m(r)=\frac{1}{N-m+1}\sum_{r=1}^{N-1}B_i^m(r) \tag{24}$$

5) 将维数 m 加 1, 重复步骤 1)～步骤 4)。

6) 由于实际上 N 为有限值, 则样本熵估计值为

$$\text{SampEn}(m,r,N)=-\ln\left[\frac{B^{m+1}(r)}{B^m(r)}\right] \tag{25}$$

为评估本文方法的性能, 同时将平均绝对误差 (mean absolute error, MAE)、均方根误差 (root mean squared error, RMSE) 和决定系数R^2作为评价指标[25]。

MAE 是绝对误差的平均值, 能更好地反映预测值误差的实际情况, 用来衡量预测值与真实值之间的平均绝对误差, MAE 越小, 模型越好。

$$E_{\text{MAE}}=\frac{1}{n}\sum_{i=1}^{n}|o_i-o_j|\in[0,+\infty) \tag{26}$$

式中: o_i为实际值; o_j为预测值。

RMSE 是预测数据和原始数据对应点误差的平方和的均值, 用于 RUL 估计广泛使用的评估度量, 指示模型在预测中会产生多大的误差。

$$E_{\text{RMSE}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i-o_j)^2}\in[0,+\infty) \tag{27}$$

R^2是表征回归方程在多大程度上解释了因变量的变化, 或者说方程对观测值的拟合程度如何, 其用于 RUL 估计广泛使用的评估度量, 最大值为 1, 越大表示模型越好。

$$R^2=1-\sum_{i=1}^{n}(o_i-o_j)^2\Big/\sum_{i=1}^{n}(o_i-\bar{o})^2\in[0,1] \tag{28}$$

式中: $\bar{o}$为o_i的平均值。

3.2　数据处理结果分析

PHM2012 数据集包括 x 和 y 方向的振动信号。本文主要对 x 方向的振动信号进行研究, 在 PHM 2012 数据集中, 滚动轴承生命周期数据子集 Bearing1_1 和 Bearing1_2 被分配给训练集, 子集 Bearing1_3 被分配给测试集。分别对训练集和测试集进行 EEMD 分解, 得到若干个 IMF 和残差。如图 6 所示, Bearing1_1 在 EEMD 分解后的 IMF 局部放大图中可以观察到高频 IMF 的特点与低频 IMF 的特点有明显差别。如表 2 可知, 高频 IMF 的样本熵更高, 说明可能包含更多的噪声。因此, 本文选择前 4 个 IMF 作为高频 IMF 来进行降噪处理。

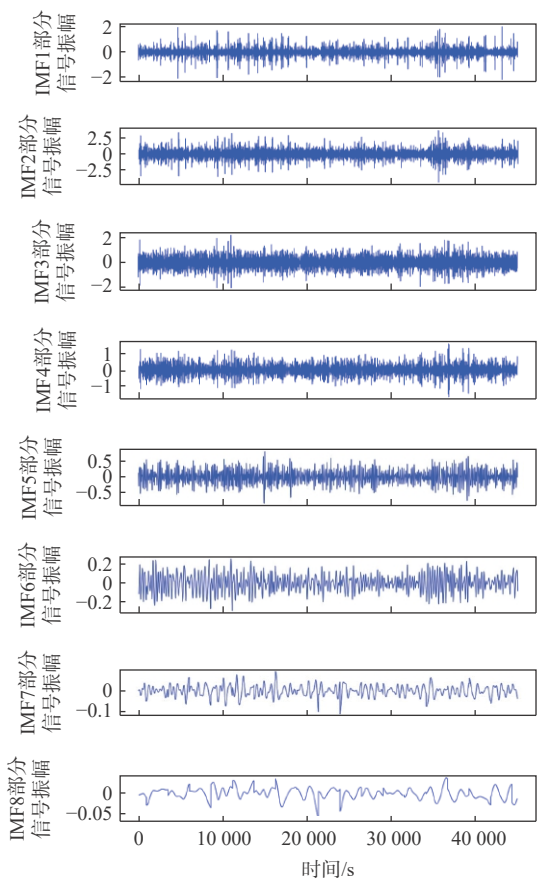


图 6　Bearing1_1 的 IMF 局部放大图

Fig. 6　Partial enlarged IMF image of Bearing1_1

表 2 Bearing1_1 的 IMF 样本熵

Table 2 Sample entropy of IMF of Bearing1_ 1

模态分量	样本熵	模态分量	样本熵
IMF1	9.40	IMF5	8.93
IMF2	10.06	IMF6	8.02
IMF3	9.80	IMF7	6.51
IMF4	9.67	IMF8	5.10

对 IMF1～IMF4 进行 DWT 与重构，可以完成高频 IMF 的降噪处理，通过多次计算降噪后，高频 IMF 重构信号与降噪前高频 IMF 重构信号的相关性系数 S 及与原始振动信号的相关性系数取平均值 O 来判断 DWT 层数，结果如表 3 所示。

本文使用斯皮尔曼秩相关系数评估降噪前后信号质量，其不是衡量 2 组数据的线性相关性，而是衡量 2 组数据秩序间的线性相关性[26]，如下：

$$S = 1 - \left(6\sum_{i=1}^{n} v_i^2\right) \Big/ z\left(z^2 - 1\right) \tag{29}$$

式中：z 为数据的长度；v_i 为第 i 个主观评分和客观评分的等级差。斯皮尔曼秩相关系数表明 x（独立变量）和 y（依赖变量）的相关方向。当 x 增加时，y 趋向于增加，斯皮尔曼秩相关系数则为正；当 x 增加时，y 趋向于减少，斯皮尔曼秩相关系数则为负；当 x 和 y 完全单调相关时，斯皮尔曼秩相关系数的绝对值为 1。

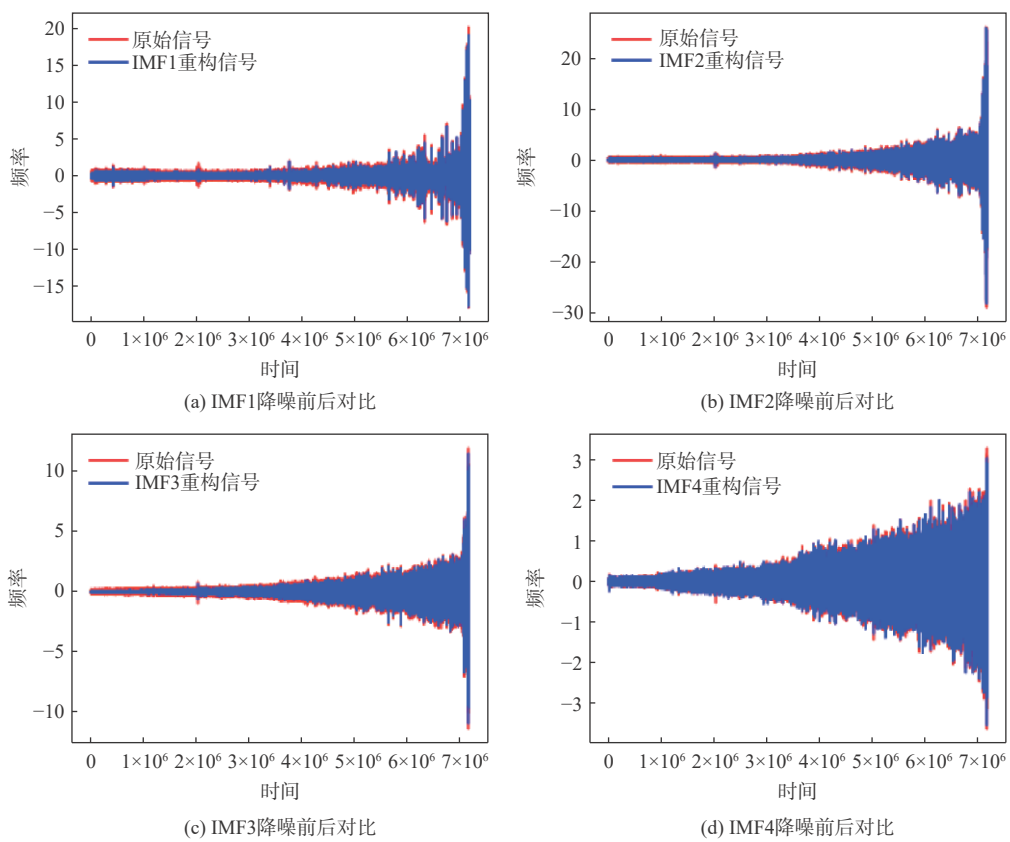
表 3 DWT 分解层数与相关性系数

Table 3 DWT decomposition layers and correlation coefficients

层数	相关性系数	平均值
1	0.001 001	0.996 612
2	0.000 104 7	0.959 676
3	0.001 326	0.850 582
4	0.002 175	0.717 504
5	0.002 692	0.751 011
6	0.002 298	0.689 021
7	0.001 241	0.537 907

已知降噪前高频 IMF 重构信号与原始振动信号的相关性系数的值为 0.000 96，通过 S 值和 O 值可判断出 DWT 层数为 5 时降噪信号质量最好。以工况 1 下的 Bearing1_1 为例，图 7 为其降噪前后对比，图 8 为 Bearing1_1 的 IMF1 部分信号降噪前后对比。可以发现，经过 DWT 降噪后的信号更加平滑。

时域退化特征是研究滚动轴承 RUL 时非常关键的部分，一般用方差、平均值、中位数、均方根、能量、峰度、熵和反双曲正弦函数等时域特征来对滚动轴承退化情况进行分析。在对高频 IMF 进行降噪处理后，将其与低频 IMF 中的每个 IMF 信号



(a) IMF1降噪前后对比　(b) IMF2降噪前后对比

(c) IMF3降噪前后对比　(d) IMF4降噪前后对比

图 7 Bearing1_1 的高频 IMF 降噪前后对比

Fig. 7 Comparison of Bearing 1_1 before and after high-frequency IMF denosing

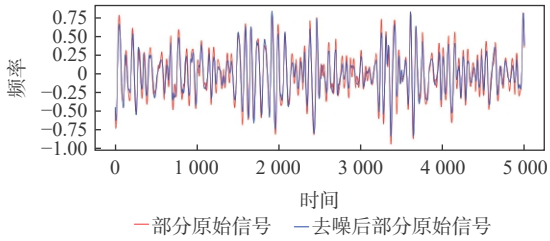


图 8　Bearing1_1 的 IMF1 部分信号降噪前后对比

Fig. 8　Comparison of IMF1 part signal of Bearing 1_1 before and after noise reduction

都进行时域特征处理, 如图 9 所示, 以工况 1 下的 Bearing1_1 为例, 给出其各个具有代表性的时域退化特征。

将得到的多尺度特征拼接融合后作为输入放入 FSBLformer 网络之前, 需要对 FSBLformer 网络的参数进行初始化, 设置 FSBLformer 网络的 nhead 参数为 2, 该参数定义了注意力机制中的多头注意力的数量。每个注意力头都可以关注不同的位置和语义信息, 从而增加模型对序列的表征能力。一般而言, 较大的 nhead 值可以提供更多的信息交互和表示能力。设置 num_layers 参数为 2, 该参数表示 FSBLformer 模型中堆叠的编码器或解码器层数。增加层数可以提高模型的表示能力, 使其能够更好地捕捉序列中的依赖关系。dropout 参数设置为 0.2, 通过随机地丢弃一定比例的输出, 可以减少模型的过拟合风险, 并提高模型的泛化能力。此外, 设置 SGD 优化算法来更新网络超参数, 将学习率初始化为 0.001, 根据验证集的损失函数值动态更新学习率, 并且设置权重衰减参数进一步减小过拟合的风险。

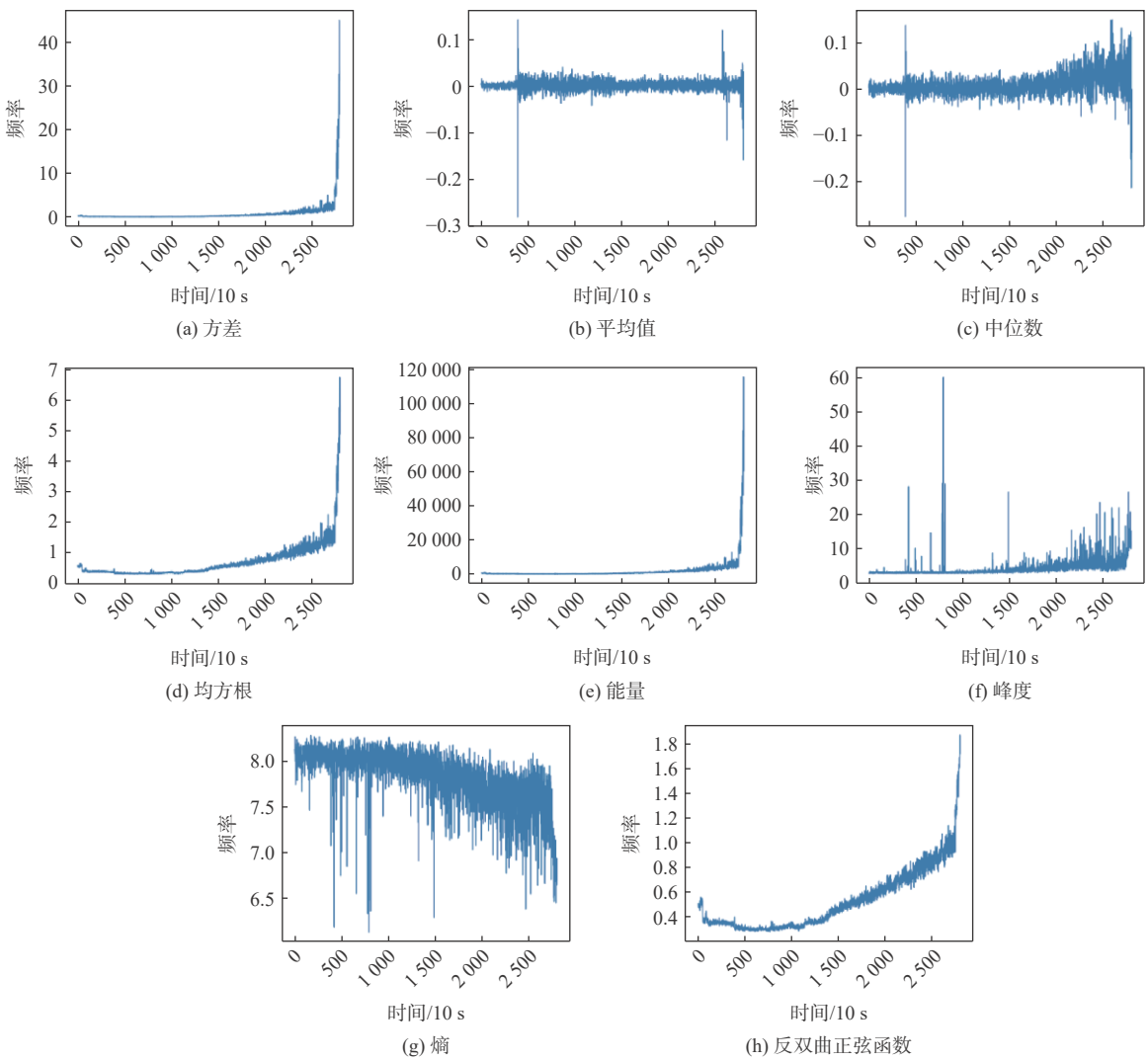


图 9　时域退化特征

Fig. 9　Time domain degradation characteristics

3.3　工况 1 的实验结果分析

模型训练后, 需要利用数据集来验证其性能。以工况 1 下的 Bearing1_1 和 Bearing1_2 作为训练集进行模型训练, 以 Bearing1_3、Bearing1_4、Bearing1_5、Bearing1_6 和 Bearing1_7 为测试集验证方法性能。图 10 显示了不同滚动轴承的 RUL 预测结果。实

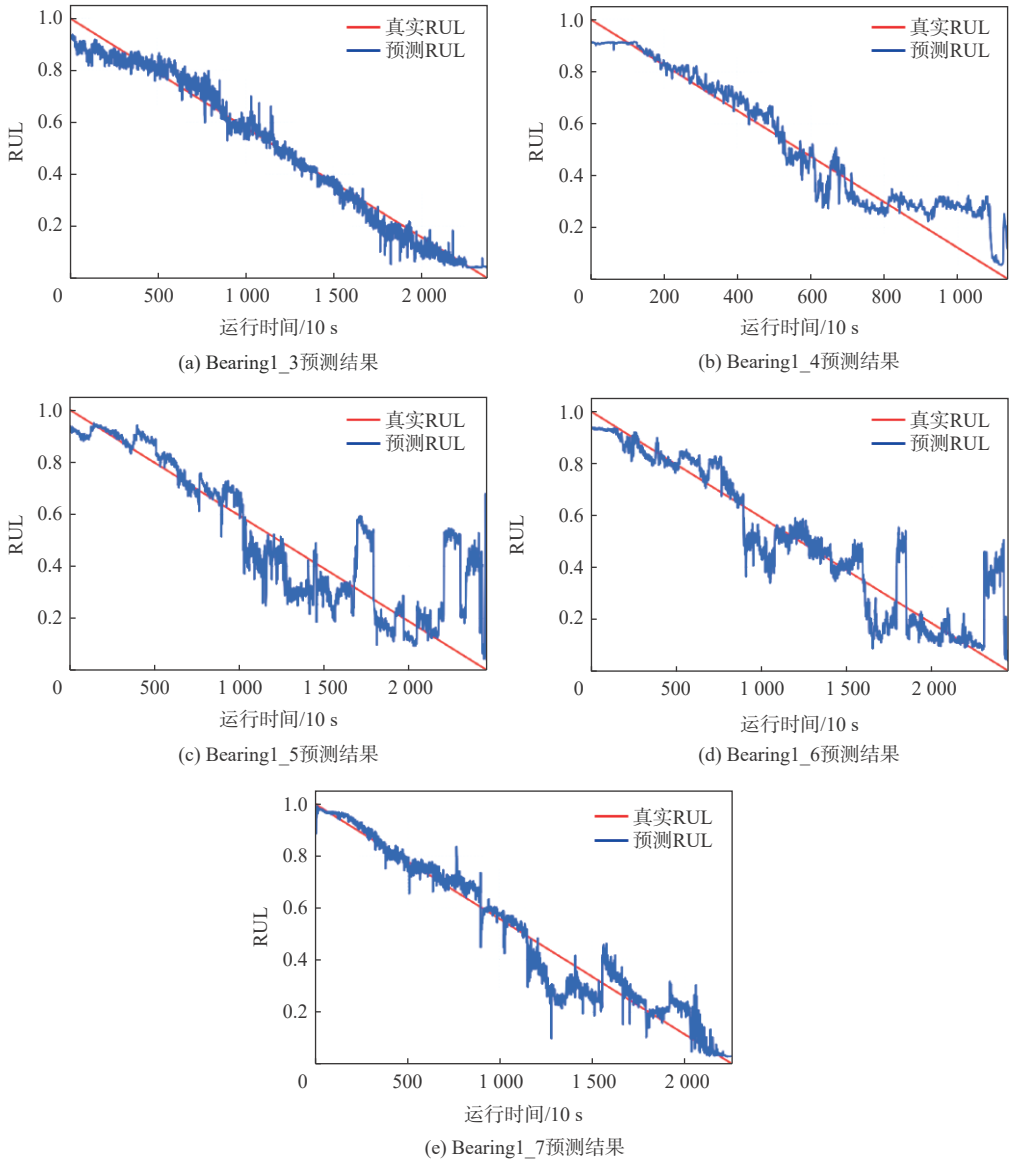


图 10 工况 1 下不同滚动轴承的 RUL 预测结果

Fig. 10 RUL prediction results of different roll bearings under the first operating condition

验结果表明，本文方法能有效拟合滚动轴承的退化过程，RUL 预测精度较高，能有效解决滚动轴承 RUL 预测问题。

在对滚动轴承退化状态进行建模时，由于训练过程中存在随机因素，方法的稳定性备受关注。本文以测试集的 Bearing1_3 为例，给出了相同实验参数设置下，Bearing1_3 重复运行 3 次的 RUL 预测结果，如图 11 所示。实验结果表明，多次运行的 RUL 预测结果差异较小，均能较好地反映滚动轴承退化过程，验证了 EEMD-AFP-FSBLformer 模型的稳定性。

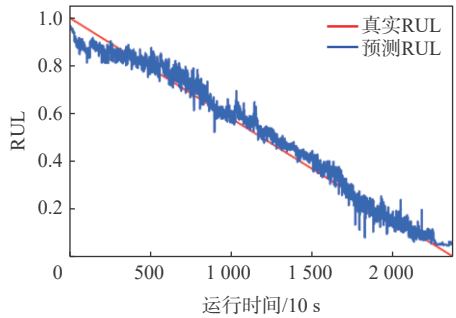
3.4 工况 2 的实验结果分析

为进一步验证本文方法的稳定性，分别采用工况 2 下的 Bearing2_1 和 Bearing2_2 作为训练集对模型进行训练，Bearing2_3、Bearing2_4、Bearing2_5、Bearing2_6 和 Bearing2_7 作为测试集验证方法性能。由图 12 可以看出，Bearing2_4 和 Bearing2_7 的预测曲线基本拟合真实值。Bearing2_7 的 RUL 预测实验在相同实验参数设置下重复 3 次，如图 13 所示。从对比实验结果可以看出，重复 3 次的 RUL 预测实验结果差异不大，均能较好地反映滚动轴承退化过程，进一步验证了本文模型在不同工况下的稳定性。

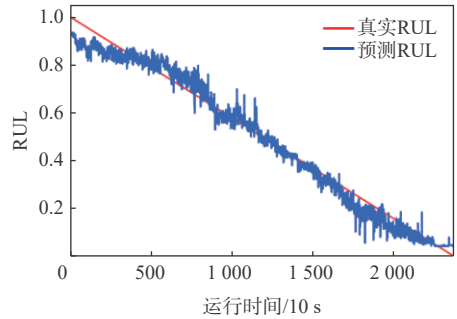
综上所述，本文方法能有效拟合滚动轴承的退化过程，甚至不同工况下的退化过程，并且 RUL 预测精度较高，能有效解决滚动轴承 RUL 预测问题。

3.5 XJTU 滚动轴承数据集下的实验结果分析

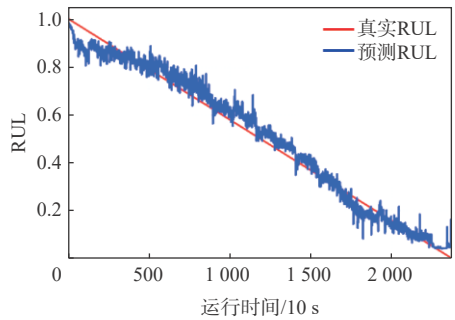
为体现本文方法的泛化能力，选用西安交通大学 XJTU-SY 滚动轴承数据集进行实验。具体而言，采用工况 1 下的 Bearing1_1 与 Bearing1_2、工况 2 下的 Bearing2_1 与 Bearing2_2 作为训练集；同时，选用工况 1 下的 Bearing1_3 与工况 2 下的 Bearing2_3



(a) Bearing1_3第1次预测结果



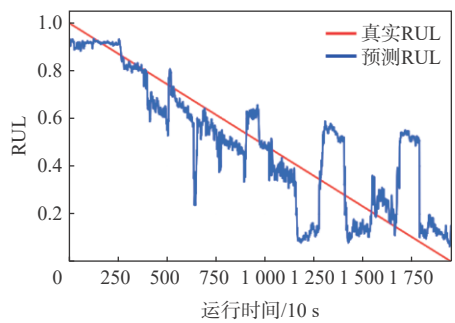
(b) Bearing1_3第2次预测结果



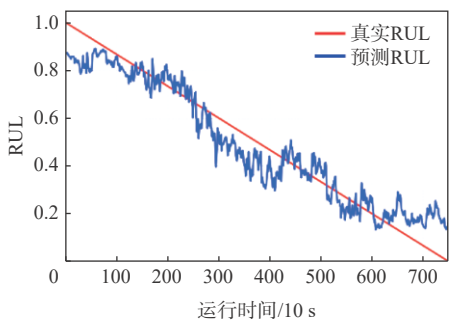
(c) Bearing1_3第3次预测结果

图 11　Bearing 1_3 重复运行 3 次的 RUL 预测结果

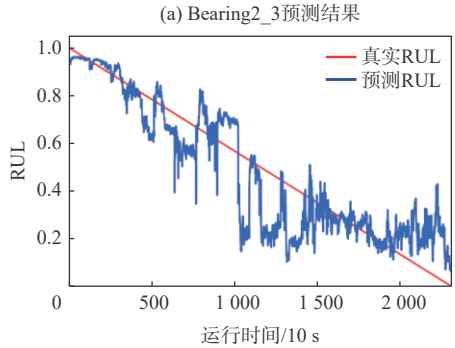
Fig. 11　RUL prediction results for three repeated runs of Bearing1_3



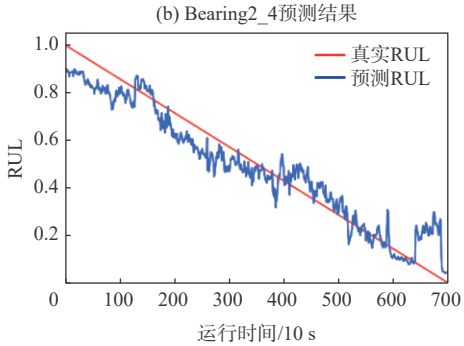
(a) Bearing2_3预测结果



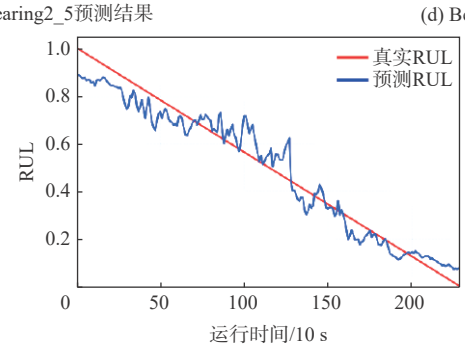
(b) Bearing2_4预测结果



(c) Bearing2_5预测结果



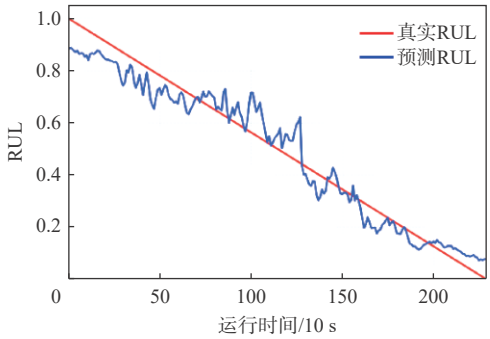
(d) Bearing2_6预测结果



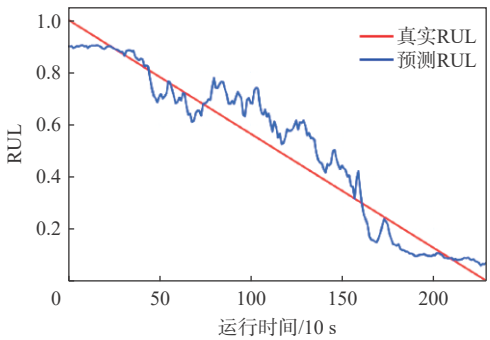
(e) Bearing2_7预测结果

图 12　工况 2 下不同滚动轴承的 RUL 预测结果

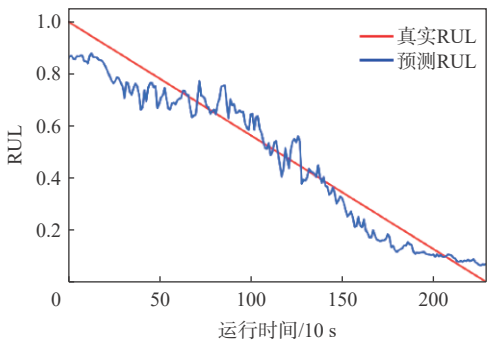
Fig. 12　RUL prediction results of different roll bearings under the second operating condition



(a) Bearing2_7第1次预测结果



(b) Bearing2_7第2次预测结果



(c) Bearing2_7第3次预测结果

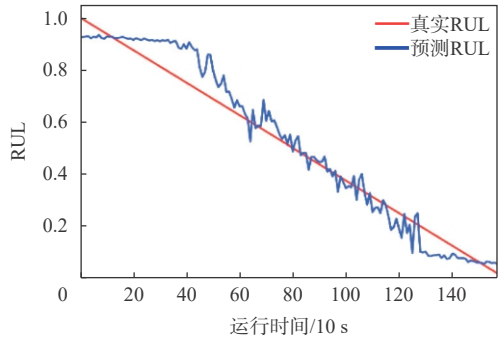
图 13 Bearing 2_7 重复运行 3 次的 RUL 预测结果

Fig. 13 RUL prediction results for three repeated runs of Bearing2_7

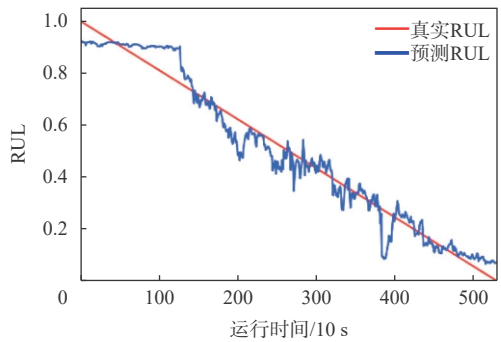
作为测试集，以验证模型在不同工况下的预测性能。由图 14 可以看出，在不同数据集上，RUL 预测实验结果依然较好，体现了本文模型较好的泛化能力。

3.6 预测模型的消融实验分析

为验证本文提出的数据处理的优越性，以 PHM2012 数据集中 Bearing1_3 和 Bearing2_7 为例，在 EEMD 分解后，将未进行去噪的高频 IMF 和对高频 IMF 进行去噪的 IMF 分别进行时域特征处理，为验证对高频 IMF 进行降噪处理有助于更有效地提取退化特征，本文将这 2 种高频时域特征处理后的信号特征与经低频 IMF 时域特征处理后的信号特征，分别输入 AFP-FSBLformer 网络进行 RUL 预测，比较结果如图 15 所示，并且对二者预测结果的评价指标进行比较，如表 4 所示。



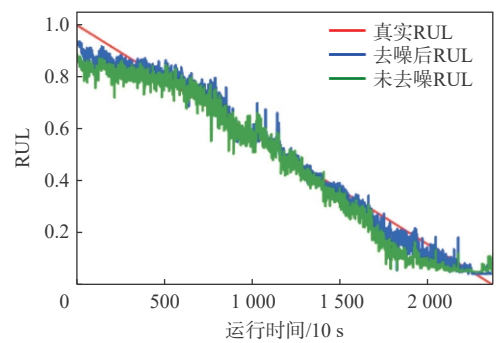
(a) Bearing1_3预测结果



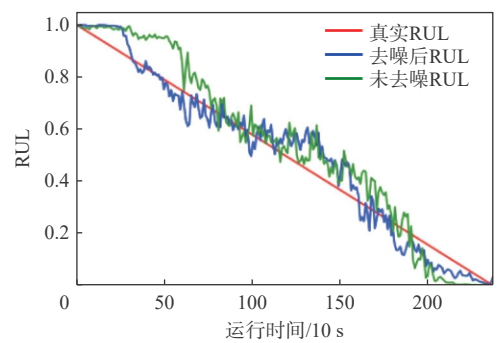
(b) Bearing2_3预测结果

图 14 XJTU-SY 数据集的 RUL 预测结果

Fig. 14 RUL prediction results for the XJTU-SY dataset



(a) Bearing1_3



(b) Bearing2_7

图 15 对高频 IMF 进行降噪前后的 RUL 预测结果

Fig. 15 RUL prediction results before and after denoising for high-frequency IMF

3.7 预测模型的对比实验分析

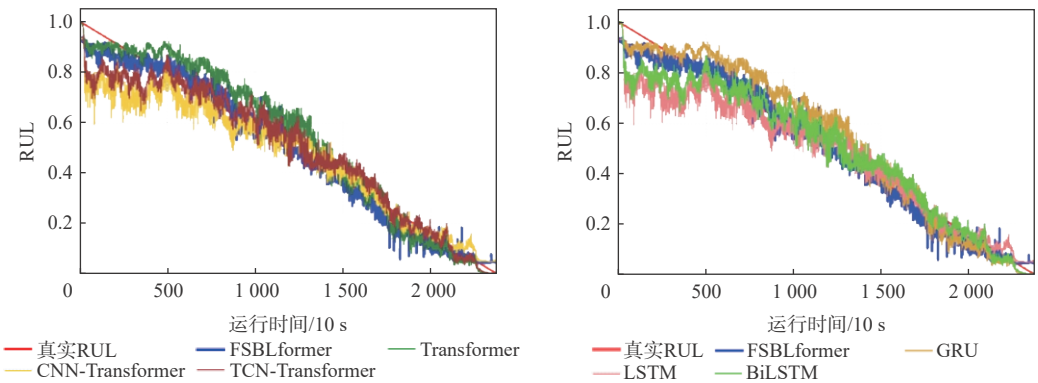
为验证本文提出的预测结构的优越性，基于 PHM2012 数据集中工况 1 的 Bearing1_3 和工况 2

表 4　Bearing1_3 和 Bearing2_7 降噪前后信号进行 RUL 预测后评价指标

Table 4　Evaluation metrics after RUL prediction of signals before and after denoising in Bearing1_3 and Bearing2_7

轴承	降噪前/后	MAE	RMSE	R^2
Bearing1_3	降噪前	0.058 1	0.076 6	0.929 1
	降噪后	0.046 6	0.070 4	0.888 3
Bearing2_7	降噪前	0.050 5	0.069 1	0.941 5
	降噪后	0.040 2	0.066 1	0.931 3

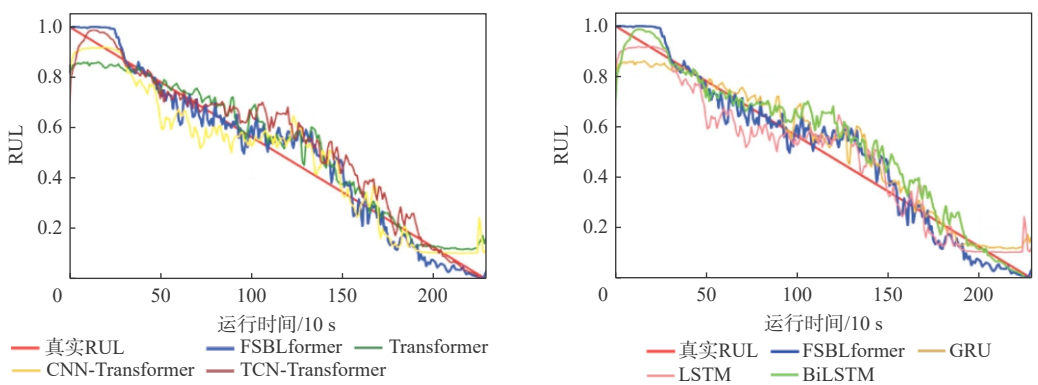
的 Bearing2_7 进行了一系列不同预测方法的对比实验。针对这些滚动轴承, 图 16 与图 17 中分别显示本文模型与 3 种不同方法的预测结果, 将本文模型分别与 Transformer、CNN-Transformer、TCN-Transformer、GRU、LSTM、BiLSTM 进行对比。将所有方法的评价指标汇总在表 5 中。通过比较图 16 与图 17 中的结果可以看出, 基于 EEMD-AFP-FSBLformer 的 RUL 预测结果与实际值更为接近。



(a) Bearing1_3对比实验1　　(b) Bearing1_3对比实验2

图 16　不同方法下 Bearing1_3 的 RUL 预测结果

Fig. 16　RUL prediction results of Bearing1_3 for different methods



(a) Bearing2_7对比实验1　　(b) Bearing2_7对比实验2

图 17　不同方法下 Bearing2_7 的 RUL 预测结果

Fig. 17　RUL prediction results of Bearing2_7 for different methods

表 5　RUL 预测后评价指标

Table 5　RUL Forecast Evaluation Indicators

方法	MAE		RMSE		R^2	
	Bearing1_3	Bearing2_7	Bearing1_3	Bearing2_7	Bearing1_3	Bearing2_7
Transformer	0.044 6	0.073 9	0.064 6	0.106 1	0.950 1	0.896 7
CNN-Transformer	0.074 3	0.072 1	0.095 1	0.104 3	0.896 9	0.903 8
TCN-Transformer	0.058 1	0.057 1	0.076 6	0.080 3	0.933 1	0.925 1
GRU	0.048 0	0.060 1	0.061 3	0.092 1	0.948 6	0.902 7
LSTM	0.079 6	0.058 5	0.087 7	0.080 3	0.881 7	0.910 1
BiLSTM	0.064 1	0.059 1	0.083 2	0.081 6	0.923 6	0.920 3
EEMD-AFP-FSBLformer	0.033 4	0.052 5	0.044 8	0.074 3	0.975 8	0.941 9

单个滚动轴承的误差不具有代表性，因此，需要计算多个滚动轴承在其工况下的RMSE、MAE和R^2的平均值，以对模型进行评估。如表5所示，将本文结构与其他方法预测结果的评价指标进行比较，可以看出，滚动轴承训练集和测试集的预测评价指标中，对比实验中绝大部分的MAE、RMSE高于EEMD-AFP-FSBLformer方法，R^2低于EEMD-AFP-FSBLformer方法。综上所述，本文模型对于不同工况下滚动轴承预测精度更高、更具有鲁棒性。

4 结 论

1）本文模型可实现较为优异的滚动轴承RUL预测，各项评价指标中，MAE至少提升33%，RMSE至少提升36%，R^2高达94%。

2）数据预处理时，进行降噪处理可以有效降低振动信号噪声干扰，降噪后的信号可以较好地体现振动信号的特征。

3）振动信号的时域特征可以显示其退化情况，而AFP可以更加充分地提取其退化特征。

4）本文模型可以在PHM2012与XJTU-SY滚动轴承数据集的不同工况下进行，并且在大多数滚动轴承的RUL预测中效果良好。

为使本文模型能更好地处理各种工况下的滚动轴承，仍需要优化数据预处理的降噪方法，在提高预测精度的同时，减少计算资源的消耗。

参考文献（References）

[1] 王焱, 丁华, 孙晓春, 等. 基于改进ECANet-TCN和迁移学习的轴承剩余寿命预测[J]. 振动与冲击, 2023, 42(21): 149-159.
WANG Y, DING H, SUN X C, et al. Bearing residual life prediction based on improved ECANet-TCN and transfer learning[J]. Journal of Vibration and Shock, 2023, 42(21): 149-159(in Chinese).

[2] HENG A, ZHANG S, TAN A C C, et al. Rotating machinery prognostics: state of the art, challenges and opportunities[J]. Mechanical Systems and Signal Processing, 2009, 23(3): 724-739.

[3] LE SON K, FOULADIRAD M, BARROS A. Remaining useful lifetime estimation and noisy gamma deterioration process[J]. Reliability Engineering & System Safety, 2016, 149: 76-87.

[4] CARPINTERI A, PAGGI M. The effect of crack size and specimen size on the relation between the Paris and Wöhler curves[J]. Meccanica, 2014, 49(4): 765-773.

[5] LOUTAS T H, ROULIAS D, GEORGOULAS G. Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic E-support vectors regression[J]. IEEE Transactions on Reliability, 2013, 62(4): 821-832.

[6] RATHORE M S, HARSHA S P. Prognostics analysis of rolling bearing based on bi-directional LSTM and attention mechanism[J]. Journal of Failure Analysis and Prevention, 2022, 22(2): 704-723.

[7] LI J X, WANG C, DING H B, et al. EMD and spectrum-centro-baric-correction-based analysis of vortex street characteristics in mist annular flow of wet gas[J]. IEEE Transactions on Instrumentation and Measurement, 2018, 67(5): 1150-1160.

[8] GAO Z H, LIU Y, WANG Q J, et al. Ensemble empirical mode decomposition energy moment entropy and enhanced long short-term memory for early fault prediction of bearing[J]. Measurement, 2022, 188: 110417.

[9] LI Q, YAN C F, CHEN G Y, et al. Remaining useful life prediction of rolling bearings based on risk assessment and degradation state coefficient[J]. ISA Transactions, 2022, 129: 413-428.

[10] GUO R X, WANG Y, ZHANG H C, et al. Remaining useful life prediction for rolling bearings using EMD-RISI-LSTM[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 3509812.

[11] JIN S Y, SU Y, GUO C J, et al. Offshore ship recognition based on center frequency projection of improved EMD and KNN algorithm [J]. Mechanical Systems and Signal Processing, 2023, 189: 110076.

[12] LI S S, YANG Y, LI C, et al. Research on signal processing technology of ultrasonic non-destructive testing based on EEMD combined with wavelet packet[J]. IEEJ Transactions on Electrical and Electronic Engineering, 2023, 18(5): 686-700.

[13] LUO J, LING C, ZHANG L. Fault diagnosis of CRDM based on EEMD-wavelet threshold denoising and WOA-KELM[C]//Proceedings of the 29th International Conference on Nuclear Engineering. New York: ASME, 2022: 1-8.

[14] ZHAO K, JIA Z, JIA F, et al. Multi-scale integrated deep self-attention network for predicting remaining useful life of aero-engine[J]. Engineering Applications of Artificial Intelligence, 2023, 120: 105860.

[15] ZHU Y C, YANG S, TONG J G, et al. Multi-scale detector optimized for small target[J]. Optoelectronics Letters, 2024, 20(4): 243-248.

[16] 赵志宏, 李晴, 杨绍普, 等. 基于BiLSTM与注意力机制的剩余使用寿命预测研究[J]. 振动与冲击, 2022, 41(6): 44-50.
ZHAO Z H, LI Q, YANG S P, et al. Remaining useful life prediction based on BiLSTM and attention mechanism[J]. Journal of Vibration and Shock, 2022, 41(6): 44-50(in Chinese).

[17] SHU Z Y, GAO L, YI S, et al. Context-aware 3D points of interest detection via spatial attention mechanism[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2023, 19(6): 1-19.

[18] CHEN M, ZANG S R, AI Z H, et al. RFA-Net: residual feature attention network for fine-grained image inpainting[J]. Engineering Applications of Artificial Intelligence, 2023, 119: 105814.

[19] ZHAO J W, NIE G Z, WEN Y H. Monthly precipitation prediction in Luoyang city based on EEMD-LSTM-ARIMA model[J]. Water Science and Technology, 2023, 87(1): 318-335.

[20] SANG Y F, SUN F B, SINGH V P, et al. A discrete wavelet spectrum approach for identifying non-monotonic trends in hydroclimate data[J]. Hydrology and Earth System Sciences, 2018, 22(1): 757-766.

[21] ZHANG W H, ZENG Y Q, WANG J P, et al. Multi-scale feature pyramid approach for melt track classification in laser powder bed fusion via coaxial high-speed imaging[J]. Computers in Industry, 2023, 151: 103975.

[22] LIU S T, HUANG D, WANG Y H. Learning spatial fusion for single-shot object detection[EB/OL]. (2019-11-25)[2024-03-01].

https://arxiv.org/abs/1911.09516.

[23] EL-NOUBY A, TOUVRON H, CARON M, et al. XCiT: cross-covariance image Transformers[EB/OL]. (2021-06-18)[2024-03-01]. https://arxiv.org/abs/2106.09681.

[24] NECTOUX P, GOURIVEAU R, MEDJAHER K, et al. PRONOSTIA: an experimental platform for bearings accelerated degradation tests[C]//Proceedings of the IEEE International Conference on Prognostics and Health Management. Piscataway: IEEEE Press, 2012: 1-8.

[25] 曾大懿, 蒋雨良, 邹益胜, 等. 一种新的轴承寿命预测特征评价指标构建与验证[J]. 振动与冲击, 2021, 40(22): 18-27.
ZENG D Y, JIANG Y L, ZOU Y S, et al. Construction and verification of a new evaluation index for bearing life prediction characteristics[J]. Journal of Vibration and Shock, 2021, 40(22): 18-27(in Chinese).

[26] KANCHANAMALA P, LAKSHMANAN R, KUMAR B M, et al. AACO: aquila anti-coronavirus optimization-based deep LSTM network for road accident and severity detection[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2023, 37(5): 2252030.

Rolling bearing life prediction based on multi-scale feature fusion

HUO Jiuyuan[1, 2, *], LI Xin[1], CHANG Chen[1], LI Yufeng[1], ZHANG Yaonan[2]

(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. National Cryosphere Desert Science Data Center, Lanzhou 730000, China)

Abstract: Rolling bearings are commonly used components in mechanical equipment, and the effective prediction of the remaining useful life (RUL) of bearings plays an important role in formulating a reasonable maintenance plan, avoiding sudden downtime of mechanical equipment, and ensuring the safety of equipment. Traditional deep learning methods are difficult to extract multi-dimensional and multi-scale degradation features, which reduces the accuracy of RUL prediction. Meanwhile, there are uncertainties such as noise and model parameters, which make it difficult to meet the maintenance requirements for point prediction of RUL. In this paper, we propose an RUL prediction model called EEMD-AFP-FSBLformer, which integrates the FSBLformer network, attention feature pyramid (AFP), discrete wavelet transform (DWT), and ensemble empirical mode decomposition (EEMD). The low-frequency modal functions are firstly processed by EEMD decomposition with DWT noise reduction and bearing time-domain degradation features with the noise reduction processed high-frequency modal functions in order to produce more representative degradation features; then the degradation features are inputted into the AFP network in order to extract the multi-scale features; and finally, these degradation features are used as inputs to the FSBLformer model. The FSBLformer model's encoder incorporates the self-attention and feature attention mechanisms, while the decoder employs the bidirectional long short-term memory (BiLSTM) network, which improves the model's performance in time prediction and feature extraction. The experiments are conducted in different working conditions of the PHM2012 dataset and XJTU-SY dataset, and the comparative experimental analysis shows that the model has a high coefficient of determination of more than 94%, which can effectively extract the multi-scale degradation features of bearings and predict their RUL.

Keywords: rolling bearing; life prediction; ensemble empirical mode decomposition; attention feature pyramid; FSBLformer

Received: 2024-03-22; **Accepted**: 2024-05-11; **Published Online**: 2024-06-20 08:58

URL: link.cnki.net/urlid/11.2625.V.20240618.1120.002

Foundation items: National Natural Science Foundation of China (62262038); Gansu Provincial Key R & D Program-Industrial Projects (22YF7GA145)

* **Corresponding author**. E-mail: huojy@mail.lzjtu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2025.0719

基于活动链分析的高校学生行为模式研究

翟宜凯[1]，蒋晓桐[1]，田琼[1,2,*]，黄海军[1]

(1. 北京航空航天大学 经济管理学院 复杂系统分析与管理决策教育部重点实验室，北京 100191；
2. 杭州市北京航空航天大学国际创新研究院，杭州 311100)

摘　　要：面对教育现代化的时代要求，校园数据的场景化应用为高校数字化转型提供了全新机遇。为此，基于校园3万名师生3个月的Wi-Fi日志和区域兴趣点(POI)数据，通过轨迹重构、语义映射与模式挖掘揭示校园行为的时空规律。将狄利克雷多项式回归(DMR)主题模型和基于手机数据的时空规律挖掘(STRMM)模型引入学生行为分析。DMR主题模型有效融合动态轨迹和静态地标数据，识别出10类校园区域功能；STRMM模型增强周期行为与不确定行为的捕捉能力，将本科生日常活动归纳为10类典型模式，包括标准教学型、专注科研型等，进一步识别出6类具有不同活动演变规律的本科生群体，呈现从低年级课程主导到高年级自主科研与弹性作息的转变，揭示了学生行为的年级动态性。研究证实，基于Wi-Fi数据的行为分析可有效识别区域功能与学生行为特征，为管理精准化、资源优化与“三全育人”实践提供数据支撑，对推动高校数字化转型具有重要参考价值。

关 键 词：活动链分析；模式挖掘；出行行为；交通管理；教育大数据
中图分类号：F572; TP391
文献标志码：A　　**文章编号**：1001-5965(2026)05-1406-16

随着高校信息化改革持续深入，教育数据挖掘与行为分析已成为优化教学管理和校园服务的重要研究方向[1-2]。传统研究多依托校园卡消费、图书馆借阅、在线学习记录等多源数据构建学生行为分析模型[3-5]，为学业预警、社交网络分析等特定场景提供了重要支撑。然而，作为校园信息化基础设施的核心组成部分，Wi-Fi网络产生的海量接入日志数据尚未得到充分挖掘。这类数据不仅记录了网络连接特征，更蕴含了丰富的时空行为信息，为构建全场景校园出行行为分析提供了新的研究视角。

现有基于Wi-Fi数据的学生行为研究主要聚焦于网络活动分析层面。刘科生等[6]对3 000名学生的网络行为数据进行聚类分析，发现网络高使用量群体的学业表现呈现显著下降趋势；胡祖辉等[7]基于决策树算法验证了“上网时长、上网次数”等指标与学业成绩的负相关关系；李有增等[4]在构建5维学生行为分析模型时，整合了课堂学习、校园生活等多维数据，但对网络数据的分析仍局限于上网活动特征。国际研究中，Hsu等[8]收集4所大学的WLAN数据，仅利用Wi-Fi接入点(access point, AP)连接频次识别轻度和重度网络依赖的学生群体。通过探索校园内学生活动，可以有效区分校园内轻度和重度网络使用用户分布，及时定位校园内的紧急情况[9]，推断校园内的社交互动情况[10]，估计校园路网的行程流量[11]，助力校园区域功能的识别与优化[12]等。

以上研究揭示了网络行为数据分析的应用潜力，但普遍存在2个局限：①分析维度单一，主要关

收稿日期：2025-10-10；**录用日期**：2025-11-21；**网络出版时间**：2025-12-05 09:43
网络出版地址：link.cnki.net/urlid/11.2625.V.20251204.1719.003
基金项目：国家重点研发计划(2023YFE0115600)；国家自然科学基金(72288101,72394374)
* **通信作者**. E-mail：tianqiong@buaa.edu.cn

引用格式：翟宜凯，蒋晓桐，田琼，等. 基于活动链分析的高校学生行为模式研究[J]. 北京航空航天大学学报，2026，52(5)：1406-1421.
ZHAI Y K，JIANG X T，TIAN Q，et al. Study of university students' behavioral patterns via activity chain analysis[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1406-1421 (in Chinese).

注网络连接的显性指标, 如在线时长、流量消耗等, 未能充分挖掘数据中蕴含的时空动态特征; ②场景关联薄弱, 缺乏对行为模式与校园场景语义的深度关联分析。例如, 仅仅依赖网络用量指标时, 无法区分长时间网络连接是“在实验室持续工作”或“在宿舍长时间观看视频”的行为差异, 尽管二者网络使用特征相似, 但反映的学习状态截然不同。

理论层面, 人类活动模式具有显著的规律性和可预测性。González 等[13] 通过百万手机用户的移动轨迹分析, 证实个体行为遵循简单可重复的模式; Song 等[14] 进一步发现个人流动模式的可预测性通常不低于 80%; Schneider 等[15] 证明仅需 17 个网络主题即可捕获 90% 人口的行为特征。人类典型活动模式被应用于日常生活中各类服务, 如景点处游客活动模式分析[16]、城市规划和交互服务[17]、交通规划、需求生成和出行预测[9]。以上研究和应用为基于轨迹数据的模式识别提供了理论基础。

技术层面, 相关研究经历了 3 个演进阶段: ①传统统计与聚类阶段。早期研究采用主成分分析结合 K-means 聚类方法, 将用户活动划分为多类时空模式。Hsu 与 Helmy[8] 通过统计指标量化移动性特征, 发现用户 AP 访问行为具有时空集中性, 群体行为具有低熵特征。这类方法虽能识别粗粒度模式, 但难以刻画时间动态性与多尺度特性。Calabrese 等[18] 使用 Wi-Fi 数据对 MIT 校园内的区域功能进行聚类, 并分析活动强度。②概率主题模型阶段。Farrahi 与 Gatica-Perez[19] 将隐含狄利克雷分布 (latent Dirichlet allocation, LDA) 模型引入移动数据分析, 将每日轨迹视为“文档”, 时空单元活动作为“词汇”, 成功识别出晚归、持续工作等主题行为; Yuan 等[20] 利用兴趣点 (point of interest, POI) 数据和人口流动模式识别北京市不同功能区; Qin 等[21] 提出的基于手机数据的时空规律挖掘 (spatio-temporal routine mining on mobile phone data, STRMM) 模型通过轨迹-日模式-常规模式 3 层隐变量捕捉多时间跨度行为不确定性, 在预测任务中表现优于传统方法。③活动链分析阶段。采用关系马尔可夫网络 (relational Markov network, RMN) 构建出行链, 结合土地利用数据聚类活动类型。Yin 等[22] 从手机数据中提取带目的地的活动链, 模式分布验证精度达 85%。这类方法虽然增强了语义解释性, 但对数据精度和连续性要求较高, 较难适应 Wi-Fi 日志固有的断连问题, 如设备休眠、信号盲区导致的 AP 切换缺失。相比之下, 使用概率主题模型和传统统计与聚类挖掘校园用户的活动模式具有更低的成本和更高的鲁棒性。因此, 选用基于 K-means、LDA 和 STRMM 这 3 类方法进行活动模式的挖掘。

校园 Wi-Fi 日志数据挖掘包含 3 层关键信息, 分别为连接时长和流量使用等网络活动指标、通过 AP 位置映射移动路径的时空轨迹、区域功能与活动类型对应关系揭示的场景语义, 现有研究对后 2 层的价值开发明显不足, 缺乏对时空行为模式与场景语义的深度关联分析, 直接影响了行为分析的粒度与质效。本文突破传统 Wi-Fi 数据的使用局限, 创新性地构建“轨迹重构-活动解析-模式识别-群体识别”4 层分析框架, 建立物理时空与活动语义的关联体系, 实现从 Wi-Fi 日志到时空行为模式再到场景语义的完整映射。主要贡献如下: ①创新性地将狄利克雷多项式回归 (Dirichlet multinomial regression, DMR) 主题模型和 STRMM 模型引入校园行为分析。DMR 主题模型解决了传统 K-means 聚类方法及 LDA 等自然语言模型利用地标等单维度数据评估的局限性, 有效融合动态轨迹和静态地标数据进行活动语义识别, 提升语义识别准确性。STRMM 模型综合考虑人类行为多时间尺度和在时间维度上的不确定性变化, 通过概率层次模型量化并完整表征全天行为模式, 克服传统周期性建模预设刚性模型并叠加残差噪声来表征不确定的不足, 提升周期行为与不确定行为特征的捕捉能力。②识别出 10 类校园区域功能, 并将学生日常活动归纳为 10 类典型模式, 包含标准教学型、教研结合型、专注科研型等, 构建学生在校园内的全场景行为模式。③识别出 6 类具有不同活动演变规律的群体。这种时空动态特征的解析和活动模式识别, 对增强个性化服务与智能化管理决策具有重要意义。该框架能够实现校园场景下高精度活动链构建、活动语义解析、活动模式识别、异质群体发现。

1　问题提出与数据预处理

1.1　问题提出

智慧校园如何基于可探测数据, 实现活动链与校园场景语义的深度关联, 构建高精度的学生行为模式识别模型, 揭示其活动规律及群体异质性, 已经成为高校人才培养改革的一个现实问题。

随着高校信息化基础设施的不断完善, Wi-Fi 网络已基本实现校园全覆盖, 由此产生的海量日志数据为研究校园移动行为提供了宝贵资源。这些数据蕴含了丰富的时空动态信息, 能够通过 AP 设备的位置分布、连接时长及切换规律, 重构学生在校园内的全场景行为轨迹, 更好地了解大学生的学习模式和校园行为特征。然而, 现有研究对 Wi-Fi 数据的利用仍停留于网络使用量统计层面, 缺乏对

时空行为模式与场景语义的深度关联分析。

本文创新性地利用Wi-Fi接入日志搭建“轨迹重构-活动解析-模式识别-群体识别”的活动模式和群体识别框架，如图1所示。轨迹重构模块根据Wi-Fi数据特征进行活动链提取，通过数据清洗、增量聚类实现移动轨迹重构。活动解析模块结合动态轨迹时空特征和静态校园POI类别数据，实现活动序列语义的有效识别。模式识别模块对比3类主流算法效果，创新引入STRMM模型识别典型的学生活动模式。群体识别模块在典型活动模式的基础上，利用STRMM模型识别出各类拥有不同活动规律的群体，构建学生行为画像。

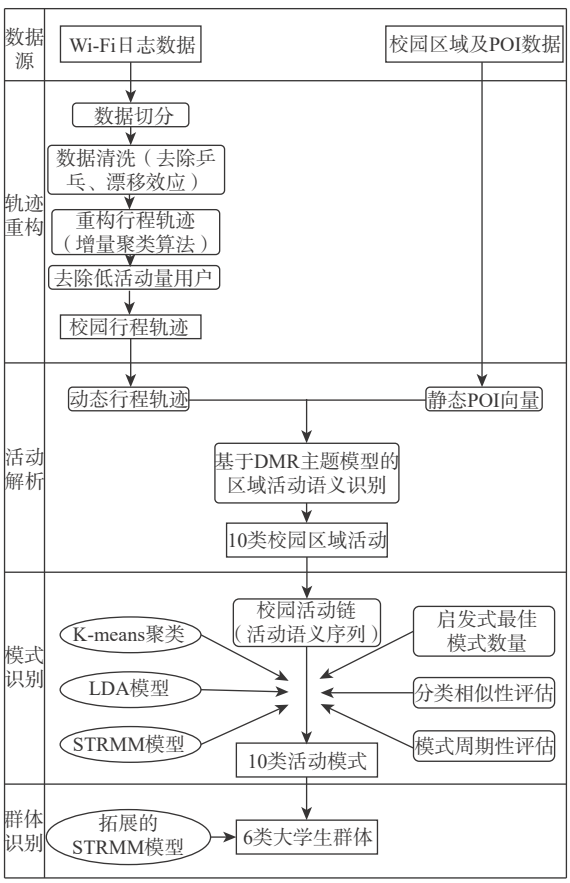


图1 活动模式和群体类别识别框架

Fig. 1 Recognition framework of activity pattern and group category

1.2 数据预处理

本节对数据集进行简单描述，根据数据特征完成数据匹配和切分、噪声数据清洗、行程轨迹重构、低活动量用户去除，得到出行轨迹数据，作为活动链生成的基础数据。

1.2.1 数据介绍

本文采用覆盖全校园的Wi-Fi数据进行研究。每个Wi-Fi数据连接记录包括用户匿名ID、用户设备ID、AP设备ID及位置、用户身份类别、学生用户年级及接入时间戳。

智慧校园内，AP设备遍布校园区域，同时，师生普遍开通了Wi-Fi账号且长期处于开放状态。师生设备登录过Wi-Fi网络的任意一个AP后，校园全部AP会在该设备靠近的一定范围内（室外大功率AP覆盖半径为40～100 m，室内吸顶式AP覆盖半径为10～12 m）自动记录师生用户设备连接和断开Wi-Fi网络的行为。因此，Wi-Fi网络接入记录能够真实地反映校园人员活动。观测AP日志记录，校内师生基本全部使用过校园Wi-Fi网络，每个AP设备都被连接过。

某高校近万个AP覆盖校园164个区域，通过梳理AP设备的位置分布、连接日志数据，整理了某年春季学期3—5月共92天的所有连接数据。为了确保匿名性，对全部接入师生的ID进行哈希处理，生成了新的随机标识符作为匿名ID。对重复记录进行清洗，3个月内共有3万名师生、10.1万台设备进行共9 982万次Wi-Fi网络连接。同时，在校园地图标注区域的原始POI类别，分为学生公寓、食堂、体育场、教学楼等。

1.2.2 基于Wi-Fi数据的轨迹提取与筛选

采集全校3个月匿名师生Wi-Fi数据进行预处理，从原始记录中提取出轨迹数据，形成完整、无噪声的出行轨迹，主要包含以下步骤：

步骤1 规范数据格式并切分完整出行轨迹。为准确描述校园人员活动轨迹，以日为单位梳理并规范数据格式。分别从3个Wi-Fi网络相关系统日志中提取关键字段，即Radius认证日志、DHCP日志、AP日志，如表1所示。通过其中时间戳和关联字段信息匹配出完整用户轨迹记录，记录用户类别和学生年级，用于后续用户群体类别分析。

表1 用于生成轨迹的WiFi数据关键字段

Table 1 WiFi data key fields for generating trajectories

设备系统	时间	关键字段	关键字段
Radius认证日志	时间戳1	用户匿名ID、用户类别及信息	IP
DHCP日志	时间戳2	用户MAC	IP
AP日志	时间戳3	用户MAC	区域位置

利用时间间隔切分记录，处理缺失数据，剔除错误信息，并清洗跳跃记录。当用户不在校内或用户设备关闭时，将不会产生相应的记录。若用户的2条相邻记录位于同一个区域内，较长的断连极可能是因为用户设备关闭而不是用户离校（如进行长时间的自习，或在宿舍休息等）；若用户的2条相邻记录位于不同区域且记录之间间隔时间较长，则极有可能是因为用户离开了校园。考虑校园生活语境，对应上午、下午、晚间课程科研活动，时长约为5 h。同时，考虑记录间隔超过5 h的活动，前后轨

迹关联性显著降低。因此, 选择 5 h 作为行程的切分时间阈值。具体地, 如果 2 个连续记录之间的时间间隔超过 5 h, 则将其切分成 2 段活动。根据步行合理速度及校内行车限速, 设定用户最大移动速度为 5 m/s, 若 2 条相邻记录的区域距离大于间隔时间内用户可以到达的最远距离, 则认为其为跳跃记录并清洗。

步骤 2　根据真实驻留点重构行程轨迹。借鉴 Widhalm 等[23] 在 2015 年提出的将低通滤波器特性与增量聚类算法相结合的方法, 可以稳健地检测停留并将原始 Wi-Fi 轨迹转换为访问地点序列, 为每个访问地点生成合适的停留时间。实验中, 考虑 Wi-Fi 数据的空间精度和校园建筑物尺度, 确保属于同一建筑同一次停留准确汇集, 选择聚类半径为 50 m。最终获得了 3 个月内共 2 081 万条行程轨迹。本文拓展考虑单个用户的多个设备记录。一方面, 在真实数据集中, 约 60% 的教职工和 83% 的学生拥有 2 台及以上设备; 另一方面, 3 个月内约有 48% 的每日轨迹是由 2 台及以上设备产生的。因此, 尽管大多数人都有一台突出的主要使用设备, 多设备的考虑可以使行程重建更加完整。

步骤 3　去除活动量低的用户数据。由于关注典型活动模式挖掘, 一些非常见的低活动量用户需要被删去, 否则这些“噪声”会对最终结果产生影响。筛选在 3 个月内至少出现了 30 d 的用户, 作为可识别模式的轨迹数据基础, 获得 1 332 万条校园行程轨迹数据。

2　基于主题模型的活动语义识别

2.1　出行活动的潜在语义信息

运动轨迹能够反映校园用户每日活动的物理位置和时间, 但无法阐明用户在出行中进行的实际活动和动机。因此, 需要将出行行为模型建立在语义空间而不是地理空间。

高校校园活动区域具有明确的核心功能类别, 同时存在功能重叠与演化现象。静态校园地标 POI 有效揭示了区域的核心功能特征, 但具有相似 POI 分布的区域因功能强度差异可能划归不同功能类别, 如教学楼有自助售货机, 但不会像校内超市被划归为购物类别。动态轨迹数据能够有效反映区域实际功能类别及强度, 综合分析出发地、目的地、停留时间及人员类别等信息, 能够有效揭示区域的实际功能。例如, 原本的教学楼宇因统筹规划调整为学院实验室, 虽然楼宇名称仍为教学楼, 但其访问轨迹呈现出同科研楼宇一致, 而不同于教学楼的特征。

通过有效结合静态地标 POI 与动态轨迹, 能够准确识别移动轨迹中驻留点区域的活动语义。Calabrese 等[18] 使用特征分解来描述每个 AP 随时间变化的连接设备数量, 并提取前 4 个特征向量作为聚类的输入。Poucin 等[12] 生成了更多表征连接动态和用户耗时的变量作为聚类输入。考虑到校园活动中工作日和周末的区分度较大, 本文在前人的特征选择基础上新增工作日和周末的特征, 聚类特征既包含了信号分解得到的连接幅度动态特征, 也包含了用户数量及时长等静态特征, 如表 2 所示。

表 2　各区域在工作日/周末的聚类特征

Table 2　Clustering characteristics of various construction and areas during weekdays/weekends

类型	特征变量	具体含义
静态特征	平均用户数量	每分钟内该区域用户平均数
	平均连接时长	每分钟内该区域用户平均停留时间
	连接时长中位数	每分钟内该区域用户停留时间中位数
	连接时长标准差	每分钟内该区域用户停留时间标准差
动态特征	连接幅度	以平均连接次数为信号计算的幅度值
	连接时长幅度	以平均连接时长为信号计算的幅度值

通过 K-means 聚类 POI 信息, 能够简单分析各类校园活动语义特征, 绘制教学、科研、服务 3 类活动在连接流量、停留时间上的动态特征(见图 2)和静态特征(见图 3), 展现活动区域的语义区分情况。

图 2 展示了一天中 3 类活动区域的动态特征、AP 连接数量分布(见图 2(a)、(b)、(c))和用户停留时间分布(见图 2(d)、(e)、(f))。教学活动所在的区域, AP 连接数量分布在工作日和周末的差异明显, 工作日存在 3 个显著的高峰且为 2 节课时长, 对应了早、中、晚 3 个常规的课程时段。科研型活动所在区域的连接数量也存在 3 个高峰, 但数量随时间的波动更平缓, 并且周末与工作日的差异较小, 突出了科研活动时间更自由, 没有明确时间间隔。服务类活动所在的区域, 如咖啡厅、服务大厅等, 工作日和周末的连接数量均较少, 停留时间较短, 且分布较为平缓。

图 3 展示了一天中 3 类活动区域的静态特征, 分别为平均连接数量(见图 3(a))和接入用户的平均连接时长(见图 3(b))。教学活动的平均连接数量在周中比周末要高很多, 但平均连接时长都接近 1 节课时长。科研活动的平均连接数量和平均连接时长在周中和周末差异小, 与教学活动相比, 连接数量低, 但平均连接时长高。服务活动的静态特征在周中和周末的差异不大, 连接数量标准差最小, 连接时长标准差最大。

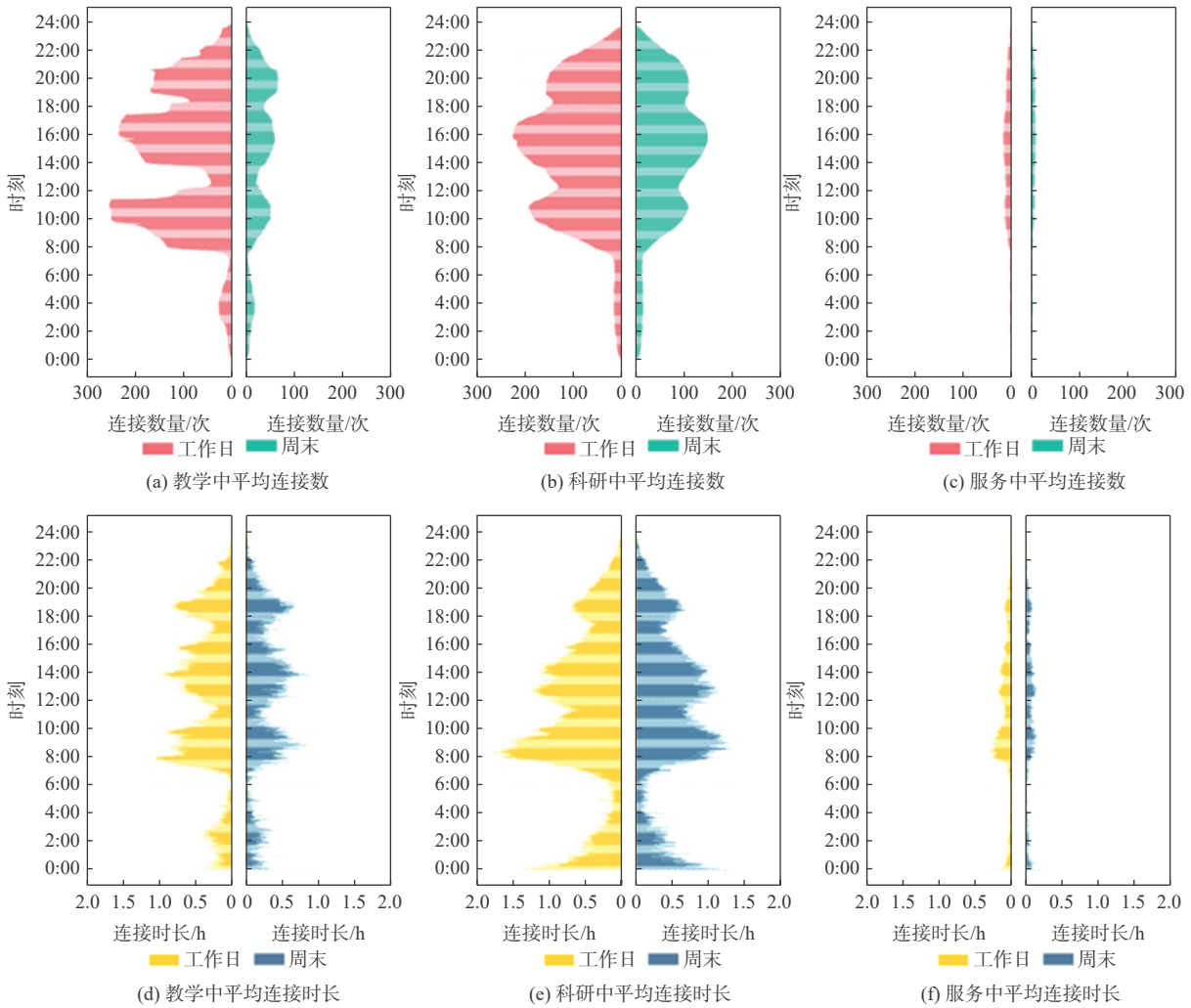


图2 教学、科研、服务活动的Wi-Fi数据动态特征

Fig. 2 Dynamic characteristics of Wi-Fi data in teaching, scientific research, and service activities

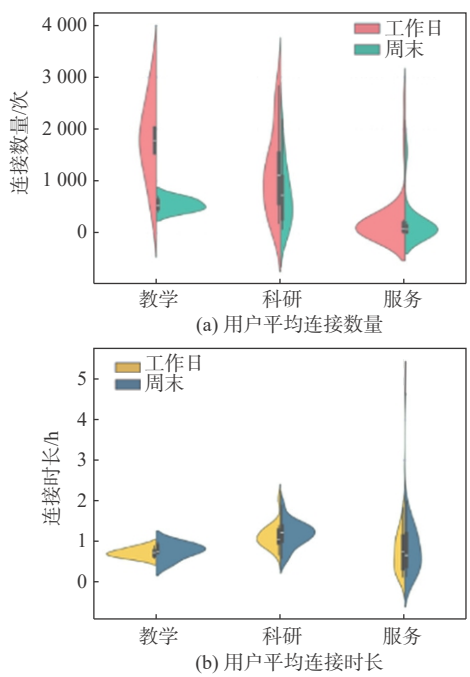


图3 教学、科研、服务活动的Wi-Fi数据静态特征

Fig. 3 Static characteristics of Wi-Fi data in teaching, scientific research, and service activities

2.2 基于DMR主题模型的活动语义识别模型

2.2.1 模型原理

主题模型利用单词空间的映射挖掘文档中的潜在语义信息和结构信息[20]。LDA是主题模型的经典算法之一，其是一种概率生成模型，通过分析每个文档的单词分布来推断文档背后的主题结构。LDA算法将语料库中每篇文档都看作由多个主题混合产生，而主题表示为词的概率分布，每个词都来自特定的主题，因此，每篇文档是单词通过概率生成的结果。引入Dirichlet先验分布，初始化文档-主题分布矩阵和主题-词分布矩阵，使用Gibbs Sampling迭代更新2个矩阵直至收敛，LDA模型可以根据文档中词的分布观测识别文档主题。

校园区域功能语义识别和文档主题识别具有相似性，在区域的功能发现和文档的主题识别之间进行类比，将一个区域类比为一篇文档，区域功能类比文档主题，拥有多个功能的区域就像一个包含多种主题的文档。然而，区域功能识别需要静态地标和动态轨迹2个观测变量共同刻画，DMR主题

模型引入先验数据, 符合本文中动静结合的双参数需求, 使得实验结果更具备实用性。区域功能语义可以通过静态地标 POI 分布和人员动态轨迹进行表征, 其中, 静态地标 POI 相当于文档的题目、作者、关键词等元数据, 而人员动态出行轨迹相当于文档中的单词。区域功能特征就体现在 POI 分布带来的活动聚集和人员流动, 借鉴主题模型通过关键词等元数据和词汇分布推断文档主题的方法, 实现区域功能准确识别。

将移动出行链表示为 DMR 主题模型的步骤如下: 首先, DMR 主题模型用于区域功能(文档主题)识别的生成过程原理如图 4 所示, 灰色节点表示可观测的出行轨迹(单词)分布 $m_{i,n}$ 和 POI 特征向量(文档元数据) $\boldsymbol{x}_i$。其次, DMR 中 Dirichlet 先验概率 α_i 将每个区域(文档)观测到的元数据纳入考虑, 对于元数据类别的不同组合, 所得 α 值不同。最后, 结合功能-轨迹分布(主题-单词)分布先验概率 β, 可以将静态元数据和动态分布数据两者观测值有效结合, 解决多源数据有效融合问题, 推断每个区域的活动功能。

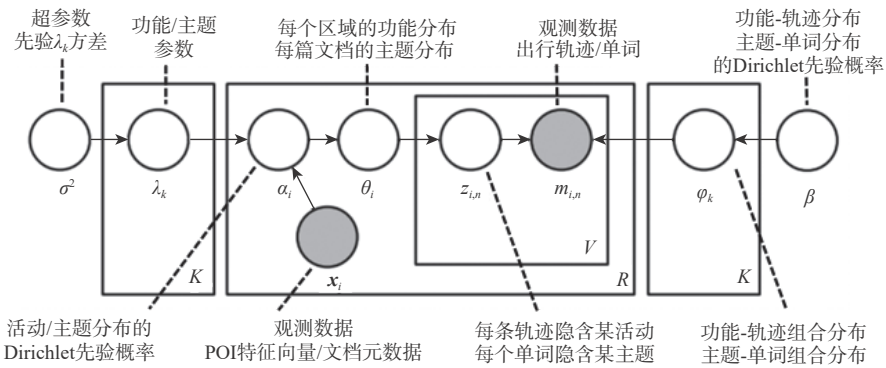


图 4　基于 DMR 主题模型的区域功能识别原理

Fig. 4　Principle of regional function recognition based on DMR topic model

2.2.2　实证分析

基于校园 POI 分布、Wi-Fi 设备位置及连接记录数据, 提取静态地标和动态轨迹信息, 通过 DMR 主题模型识别校园活动实际分布。为评估识别准确性, 利用校园平面规划图中区域功能类别作为参考, 同时调研 100 位长期生活、工作在校园的各类别师生标注校园各区域实际功能语义和功能强度排序, 验证本文模型的准确性。

动态轨迹数据利用第 1 节得到的出行轨迹, 以小时为单位将接入时间离散处理到各时间片中, 每天 24 h, 区分工作日或周末共计 48 个时间片, 统计每个活动区域在各时间片内的驻留次数, 即为出行轨迹中驻留点次数, 如表 3 所示。构造<来源区域, 区域位置, 时间片, 驻留次数>四元组, 表示校园区域间移动模式, 作为模型输入。

表 3　校园区域间移动模式

Table 3　Movement patterns between campus areas

来源区域	区域位置	时间片	驻留次数
10号公寓	1号教学楼	工作日8:00—9:00	16
6号公寓	1号教学楼	工作日9:00—10:00	53
操场	停车场	周末8:00—9:00	0
图书馆	2号食堂	周末11:00—12:00	112

借鉴自然语言处理中词频-逆文档频率 (term frequency-inverse document frequency, TF-IDF)衡量某一类 POI 在特定区域中的功能强度, 将每个区域视作一个文档, 将 POI 类别视为文档中的词。TF 值为区域 POI 类别频率, 计算区域 i 中第 j 类 POI 的数量 n_{ij} 和该区域内所有 POI 总数 V_i 的比值。IDF 值衡量该类 POI 在全校区域中的分布稀有性, 计算全校区域总数 R 与包含第 j 类 POI 的区域数 $|\{r|j\in P(r)\}|$ 比值的自然对数。TF、IDF 相乘得到第 j 类 POI 在区域 i 中的权重。

$$F_{\mathrm{tf},ij}=n_{ij}/V_i \tag{1}$$

$$v_{ij}=F_{\mathrm{tf},ij}F_{\mathrm{idf},j}=\frac{n_{ij}}{V_i}\ln\left(\frac{R}{|\{r|j\in P(r)\}|}\right) \tag{2}$$

对于给定区域 r_i, 定义其 POI 特征向量 $\boldsymbol{f}_i=(v_{i1},v_{i2},\cdots,v_{ic})$, 其中, v_{ij} 为第 j 个 POI 类别的 TF-IDF 值, 该值越高, 表明该 POI 类别功能在区域中重要性越突出。

DMR 主题模型的输入是 2 个观测数据: 区域间移动轨迹和静态地标 POI 特征向量。由于 DMR 主题模型属于无监督学习, 还需要通过试验确定主题数量, 即校园内区域功能语义总数, 其选择直接影响识别的可解释性, 在 F 高校区域类别合理数量范围 1～30 内进行试验, 利用主题一致性得分初步判断类别数量, 选取合适的随机种子, 一致性得分随着主题数量增加而增加, 到某个数值趋于平稳。经过试验选取语义类别数量 K 为 10, 作为模型输入。

DMR主题模型得到的输出结果，将校园164个区域分为10类（A0～A9），并标注相应功能强度。表4列出了10类识别结果中20个各功能类别代表性区域，区域功能频率强度 (frequency intensity, FI) 为模型计算结果，根据每列FI排序得到的各类别中功能的强度排名在IR列显示，每行表示区域的

表4 校园典型区域活动语义识别结果与功能强度排序

Table 4 Semantic recognition results and functional intensity ranking of typical campus areas

区域	FI									
	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9
3号教学楼	0.249	0	0.002	0	0	0	0.007	0	0.003	0.002
1号教学楼	0.102	0	0.002	0	0	0	0.034	0.047	0.006	0.004
16号公寓	0	0.150	0	0	0.022	0.001	0	0	0.010	0.015
13号公寓	0.008	0.021	0	0	0.003	0	0.000	**0.031**	0	0.001
9号公寓	0	0.135	0	0	0	0	0.002	0	0.008	0.005
综合食堂	0.029	0	0.536	0	0	0	0	0.009	0.004	0.002
2号食堂	0	0.002	0.071	0	0.002		0.001	0	0.002	0.004
办公楼	0	0	0.007	0.147	0.014	0.001	0.003	0	0	0.005
办公楼二区	0	0.001	0.002	0.049	0.002	0.001	0	0	0.003	0.003
办公楼三区	0	0	0	0.013	0	0	0.001	0	0	0
家属区	0	0.001	0.003	0	0.333	0	0.005	0.011	0	0.127
校医院	0	0	0	0	0.140	0.131	0.017	0	0.031	0.072
操场	0.044	0	0	0	0.098	0.344	0.034	0.099	0.100	0.099
会议中心	0	0.001	0.004	0	0	0.170	0.001	0	0.008	0.006
足球场	0	0	0	0	0	0.167	0	0	0	0.001
主楼1栋	0	0	0	0	0	0	0.093	0.047	0.050	0.032
主楼7栋	0.010	0	0	0.055	0	0	0.138	0.061	0.055	0.033
主楼6栋	**0.065**	0	0.016	0.010	0.003	0	**0**	0.145	0.032	0.020
生医楼	0.012	0	0	0.036	0.007	0.001	0.034	0.034	0.057	0.036
土木楼	0	0	0	0.022	0	0	0.007	0	0.058	0.041

区域	IR									
	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9
3号教学楼	1		17				27		45	55
1号教学楼	2		20				7	9	30	40
16号公寓		1			35	24			22	18
13号公寓	23	19			39		72	12		65
9号公寓		2					46		26	36
综合食堂	10		1					24	41	52
2号食堂		38	2		119		62		57	37
办公楼			11	1	4	23	41			32
办公楼二区		50	19	6	6	22			46	49
办公楼三区				18			67			
家属区		54	16		1	9	32	23		1
校医院					2	4	18		11	4
操场	7				92	1	9	2	2	3
会议中心		47	15			2	69		25	31
足球场						3				61
主楼1栋							3	8	6	8
主楼7栋	20			5			1	5	5	7
主楼6栋	5		8	26	26			1	10	14
生医楼	18			9	9	26	8	11	4	6
土木楼				12			25		3	5

功能排名, 取前 2 名按照颜色深度表示, 颜色深的功能排名更高。

模型将全校园 164 个区域识别归类到 10 个功能语义类别中, 并综合研判功能强度排名中代表性区域(功能-区域排名)、区域各类功能分布(区域-功能排名), 得到 A0～A9 区域功能识别结果为教学区、宿舍区、食堂区、办公区、家属区、活动区、硕博科研区、本科科研区、教师科研区、其他。结果同 100 位师生标注的校园功能区高度一致。

区域 A0 为课堂教学区, 该功能最典型的代表性区域是教学楼, 其功能强度和内部排名明显高于其他区域。此外, 具备教学功能的区域场所, 如配备教室的科研楼宇和体育课场所操场在该类别区域配置中比较靠前。

区域 A1～A5 分别为宿舍区、食堂区、办公区、家属区、活动区。宿舍区特色 POI 为学生宿舍, 夜间驻留次数最高。食堂区特色 POI 为餐饮, 且典型移动模式具有时长较短且非常集中的特点, 平均活动时长约 0.5 h, 集中在早、中、晚 3 个时段。办公区 POI 功能强度较为平均, 分散在校园各区域, 列出的典型办公区为 3 个主要的校级办公区, 功能强度排名并非前三位, 表明部分学院办公区功能强度大于校级办公区。

DMR 主题模型可以将原本笼统的“科研类”区域细粒度区分成 A6 硕博科研区、A7 本科科研区、A8 教师科研区, 有效区分了不同师生群体同类行为模式的区域差异, 使区域功能识别更为精细。科研类别主要集中在该高校各学院的科研楼宇 POI, 但功能-区域排名有所区别。教师科研区分散在各科研楼宇, 表示各学院都有老师科研行为; 硕博科研区集中在实验室较多的学院楼宇, 本科科研区分散在学院楼宇和具备实验条件的教学区域。从时间维度上, A6 特征为长时间待在科研区域, A7 特征为白天不仅仅出现在实验区域, 也出现在教学区域。值得指出的是, A7 类别中, 某宿舍区域具备较强的科研属性, 经确认, 该宿舍区地下一层均为临时性大规模物理实验室, 服务于本科生科研实验, 该“异常”进一步证明了 DMR 语义识别的有效性。

至此, 通过 DMR 主题模型进行活动语义识别, 得到高校校园教学区、宿舍区、食堂区、办公区、活动区、本科科研区、其他、硕博科研区、教师科研区、家属区等 10 类细粒度活动区域语义类别。

考虑到不同类型用户活动存在较大差异, 如学生在食堂的活动大概率为吃饭, 而职工人员在食堂的活动可能是工作, 因此, 本文后续分析选用了行为规律性最强、同质性最高、数据量最大的 9 755 名本科生进行细粒度识别和分析。剔除硕博、教师、家属 3 类群体的活动, 行为类别中同本科生相关的是前 7 类大学生活动, 加上表示无连接信号的“断联无信号”类, 共 8 类大学生典型活动。

3　活动模式与群体分布检测模型

利用第 1 节得到的出行轨迹数据和第 2 节得到的典型活动, 构建校园用户的活动序列即活动链, 从中捕获典型活动模式。将 STRMM 模型引入校园行为分析, 提升周期行为与不确定行为的捕捉能力, 对比 K-means 聚类、LDA 模型和 STRMM 模型 3 种模式识别方法, 验证校园场景下最优活动模式检测方法。

3.1　活动链构建

假设一天被等分为τ个时间切片, 每个时间切片长度为$24/\tau$。假设共有C类活动, 活动集合为$S=\{A_1,A_2,\cdots,A_C\}$。则对于特定的一个用户, 其在某日的活动序列可以定义为一个长度为τ的列表: $[a_1,a_2,\cdots,a_\tau]$, 其中, $a_t\in S(t=1,2,\cdots,\tau)$表示用户$U$在时间切片$t$内进行的活动。显然, τ越大, 时间被切分得越细, 用户的活动轨迹就描述得越精准。进一步, 将一天划分为T个时间块, 时间块之间可以不等长, 并且设第k个时间块内包含了N_k个时间切片, 满足$\sum_{k=1}^{T}N_k=\tau$。假设该用户有D天的记录数据, 图 5 展示了定义的每日活动序列, 其中, TS表示时间切片, $a_{i,j}^{(k)}(i=1,2,\cdots,D;j=1,2,\cdots,N_k;k=1,2,\cdots,T;a_{i,j}^{(k)}\in S)$表示该用户在第$i$天第$k$个时间块内, 第$j$个时间切片内进行的活动。

第 2 节获得了 8 类大学生活动, 因此, 设定活动数 C=8。考虑到部分活动停留时间普遍较短, 如食堂吃饭等, 设定每 15 min 进行一次切分, 即一天

	时间区间 1				…	时间区间 T			
	TS 1	TS 2	…	TS N_1	…	TS $\tau-N_\tau$	…	TS $\tau-1$	TS τ
第1天	$a_{1,1}^{(1)}$	$a_{1,2}^{(1)}$	…	$a_{1,N_1}^{(1)}$	…	$a_{1,1}^{(T)}$	…	$a_{1,N_T-1}^{(T)}$	$a_{1,N_T}^{(T)}$
第2天	$a_{2,1}^{(1)}$	$a_{2,2}^{(1)}$	…	$a_{2,N_1}^{(1)}$	…	$a_{2,1}^{(T)}$	…	$a_{2,N_T-1}^{(T)}$	$a_{2,N_T}^{(T)}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
第D天	$a_{D,1}^{(1)}$	$a_{D,2}^{(1)}$	…	$a_{D,N_1}^{(1)}$	…	$a_{D,1}^{(T)}$	…	$a_{D,N_T-1}^{(T)}$	$a_{D,N_T}^{(T)}$

图 5　用户多日活动链序列图

Fig. 5　Activity chain sequence of a user over multiple days

的时间切片数τ=96。遵循 Qin[21] 和 Farrahi[19] 等的设定，将一天切分为 T=8 个时间间隔，分别为 0:00—7:00、7:00—9:00、9:00—11:00、11:00—14:00、14:00—17:00、17:00—19:00、19:00—21:00、21:00—24:00。该设定规则与校园生活节律适配，上下午核心教学科研时间、早中晚食堂高峰时段均能够合理捕捉；同时，对活动较少的夜间与活动较多的白天区分处理精度，避免过多无意义的夜间活动产生。

3.2 活动模式识别

3.2.1 3 类识别方法

本节考虑 3 种不同的模式识别方法，并对比其所捕获的典型模式之间的异同，确定适合校园 Wi-Fi 数据场景的识别方法。

1）基于余弦距离的 K-means 聚类。K-means 是聚类方法中最常见的一种无监督学习方法。特征工程的选择、距离度量方式的选择、聚类个数的选择是决定聚类结果的关键因素。本文输入样本为所有本科生的每日活动序列，根据 3.1 节中构建的活动链序列，每个样本是一个长度为 96 的类别型特征向量，其中，每个元素都表示一个活动。针对类别型数据，常用的方法为 one-hot 编码处理，将每个元素编码为一个长度为活动类别总数C的序列，并为对应类别标记 1，其他类别标记 0。为避免维度倍增和数据稀疏影响算法性能，考虑到“活动模式”可以理解为不同轨迹在相似时间段内偏向于进行相似的活动，从活动概率角度重建特征。针对第i条轨迹中第k个时间块$k\in\{1,2,\cdots,T\}$，计算活动c在该时间块内出现的概率$p_{i,c}^{(k)}$。

$$p_{i,c}^{(k)}=\frac{\sum_{j=1}^{N_k} f_c\left(a_{i,j}^{(k)}\right)}{N_k} \tag{3}$$

式中：N_k为属于第 k 个时间块的数据点总数量；$f_c(\cdot)$为指示函数，表示为

$$f_c(x)=\begin{cases}1 & x=c\\ 0 & x\neq c\end{cases} \tag{4}$$

通过上述特征构建方法，将原本特征向量的长度从τ =96 转换为$C\times T$=88，表示该日活动序列内T个时间块内C类活动的出现概率，如图 6 所示。96 个时间切片向量聚焦于每个时段在做什么，数据高纬稀疏且忽略时间块内分布规律，而 88 维概率特征向量关注一天内不同时段倾向于做什么，更符合活动模式的定义，且有效降低特征维度。

	时间区间 1					时间区间 T			
	C1	C2	···	C8	···	C1	C2	···	C8
第1天	$p_{1,1}^{(1)}$	$p_{1,2}^{(1)}$	···	$p_{1,8}^{(1)}$	···	$p_{1,1}^{(T)}$	$p_{1,2}^{(T)}$	···	$p_{1,8}^{(T)}$
第2天	$p_{2,1}^{(1)}$	$p_{2,2}^{(1)}$	···	$p_{2,8}^{(1)}$	···	$p_{2,1}^{(T)}$	$p_{2,2}^{(T)}$	···	$p_{2,8}^{(T)}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
第D天	$p_{D,1}^{(1)}$	$p_{D,2}^{(1)}$	···	$p_{D,8}^{(1)}$	···	$p_{D,1}^{(T)}$	$p_{D,2}^{(T)}$	···	$p_{D,8}^{(T)}$

图 6 K-means 方法中用户的轨迹特征向量示意图

Fig. 6 User trajectory feature vector in K-means method

考虑到活动链语义特征向量重点考量特征方向而非距离，选择余弦距离而非欧氏距离。聚类个数的选择，通过计算聚类结果的平均平方误差和（sum of squared errors, SSE），绘制折线图，寻找折线转折点“手肘”，作为聚类个数的最佳取值。即达到某个聚类数后，进一步增加聚类数对聚类效果的改善不再明显，SSE 的下降趋势趋于平缓。

2）基于主题概率模型的 LDA 方法。主题模型用于分析文本数据中隐藏的主题结构和单词分布，通过 LDA 方法进行活动模式识别，将每日活动序列视为一篇文档，将其中的特征向量视为词汇。针对序列中每个时间切片的活动，利用其所在的时间块和活动的语义信息共同构成单词。例如，一个单词可以表示为 (0:00—7:00, 宿舍休息) 或者 (11:00—14:00, 食堂吃饭) 等。通过这种方式，为每个序列，即 LDA 方法中的“文档”，创建了τ = 96个“单词”，如图 7 所示。其中，每个单词都包含了活动在时间和语义上的特征，相同词汇出现次数越多，说明这篇文档越倾向于存在某种模式。最终通过计算主题一致性，选择一致性最高的主题数作为最佳主题个数。

	时间区间 1					时间区间 T			
	TS 1	TS 2	···	TS N_1	···	TS $\tau-N_\tau$	···	TS $\tau-1$	TS τ
第1天	$(1, a_{1,1}^{(1)})$	$(1, a_{1,2}^{(1)})$	···	$(1, a_{1,N_1}^{(1)})$	···	$(T, a_{1,1}^{(T)})$	···	$(T, a_{1,N_T-1}^{(T)})$	$(T, a_{1,N_T}^{(T)})$
第2天	$(1, a_{2,1}^{(1)})$	$(1, a_{2,2}^{(1)})$	···	$(1, a_{2,N_1}^{(1)})$	···	$(T, a_{2,1}^{(T)})$	···	$(T, a_{2,N_T-1}^{(T)})$	$(T, a_{2,N_T}^{(T)})$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
第D天	$(1, a_{D,1}^{(1)})$	$(1, a_{D,2}^{(1)})$	···	$(1, a_{D,N_1}^{(1)})$	···	$(T, a_{D,1}^{(T)})$	···	$(T, a_{D,N_T-1}^{(T)})$	$(T, a_{D,N_T}^{(T)})$

图 7 LDA 方法中单个用户的轨迹特征向量示意图

Fig. 7 Trajectory feature vector for an individual user in LDA method

3) STRMM 模型。在 LDA 模型的基础上, 借鉴 Qin 等[21] 提出的 STRMM 模型, 考虑一天中不同时间段的行为不同程度的不确定性, 产生更真实和直观的行为模式, 并且在多个数值实验中发现, 相较于传统的聚类和主题模型, 该模型在模式分类准确度和轨迹预测准确度上均获得更好的表现。

STRMM 模型是由 LDA 方法衍生出的一种模型, 其具体生成路径如图 8 所示。假设共存在 L类模式, 该模型认为第 i条活动轨迹的主题 $l_i(l_i \in \{1,2,\cdots,L\}, i=1,2,\cdots,M)$都服从 L维先验参数 π的多项分布, 即 $l_i \sim \text{Multinomial}(\pi)$。模型还假设每个主题中, 不同时间块 T内的模式是相互独立的。因此, 第 i条轨迹中时间块 k的模式 $\theta_i^{(k)}$都服从于先验参数为 $\boldsymbol{\alpha}_{r_i}^{(k)}$的 Dirichlet 分布, 每个 $\boldsymbol{\alpha}_{l_i}^{(k)}$都是一个 C维向量, C为活动类别总数, 即 $\theta_i^{(k)} \sim \text{Dirichlet}\left(\boldsymbol{\alpha}_{l_i}^{(k)}\right)$。第 i条轨迹位于时间块 k的第 j个时间切片内的真实活动 $a_{i,j}^{(k)}(i=1,2,\cdots,M; j=1,2,\cdots,N_k; k=1,2,\cdots,T; a_{i,j}^{(k)} \in S)$将服从参数为 $\theta_i^{(k)}$的多项式分布, 即 $a_{i,j}^{(k)} \sim \text{Multinomial}(\theta_i^{(k)})$。Qin 等[21] 详细阐释了参数估计算法, 主题分布 π及每个主题下不同时间块的模式 $\boldsymbol{\alpha}_l^{(k)}$, 可以通过算法迭代得到。

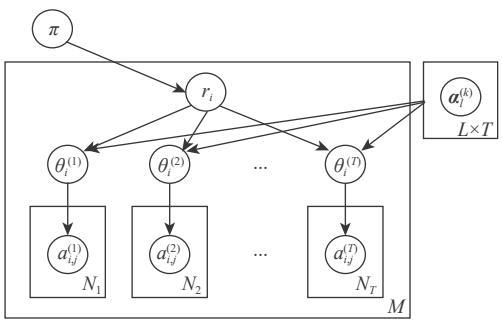


图 8　STRMM 模型的原理标注

Fig. 8　Plate notation of STRMM model

通过随机期望最大化算法, 得到稳定的 1 个 L维的 π和 $L\times T$个 C维的 $\boldsymbol{\alpha}_l^{(k)}$, 分别表示语义主题的概率分布和每个时间块的活动模式。通过计算模型的困惑度, 绘制手肘图, 确认困惑度降幅随模式类别增加显著减慢的拐点, 选择对应数量作为最佳模式个数。

3.2.2　识别效果对比

1) 模式识别类别数量选定。通过 K-means 方法绘制聚类数量-平方和误差手肘图, LDA 方法绘制主题一致性图, STRMM 方法绘制困惑度柱状图, 可以直观看出最优聚类数量。本文数据集下, K-means 聚类方法的最优类别个数为 $n=6$, LDA 模型和 STRMM 模型的最优类别个数为 $n=10$。为方便后续 3 种方法效果的横向对比, 选择各方法最优数量值。

2) 从识别结果相似度来看, 绘制 3 种方法分类结果的关联度分析如图 9 所示, 用“聚类方法. 类别序号”表示各类方法的识别结果。可以直观看到, LDA 模型与 K-means 方法得到的活动模式分类更为相似, LDA 方法倾向于将 K-means 方法中识别的类拆分成更细的子类, 而通过 STRMM 方法得到的模式与上述 2 种方法得到的模式具有更大的差异。分别计算 3 种方法两两间的调整兰德指数 (adjusted Rand index, ARI) 和归一化互信息 (normalized mutual information, NMI), 如表 5 所示。ARI 越接近 1, 说明聚类结果越相近, 越接近 0, 说明结果越随机, 越接近−1, 说明结果相反。NMI 越接近 1, 越相似, 越接近 0, 差异越大。STRMM 与 K-means、LDA 方法的识别结果差异显著(ARI、NMI 近 0), 因其通过多层隐变量刻画了行为在周、月尺度上的周期规律, 而传统方法仅关注单日静态分布, 凸显了 STRMM 方法捕捉长时间动态的独特优势。

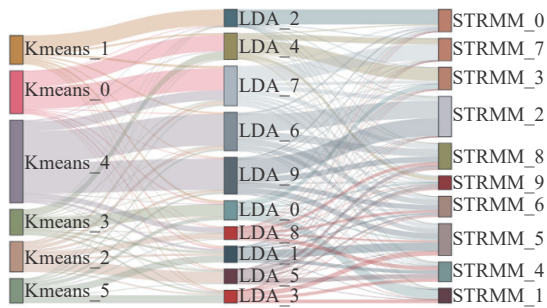


图 9　3 种方法识别结果的关联度分析

Fig. 9　Correlation analysis of recognition results from three methods

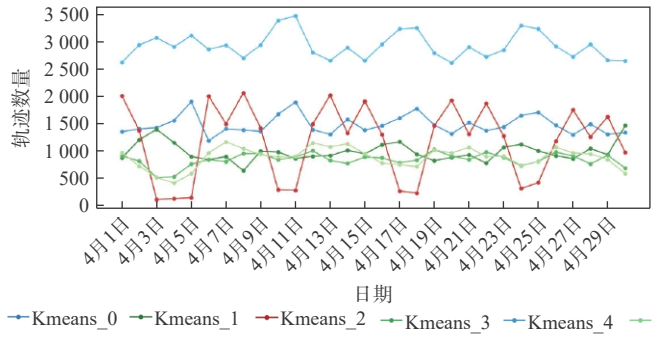
表 5　3 种方法识别结果的相似度

Table 5　Similarity of recognition results from three methods

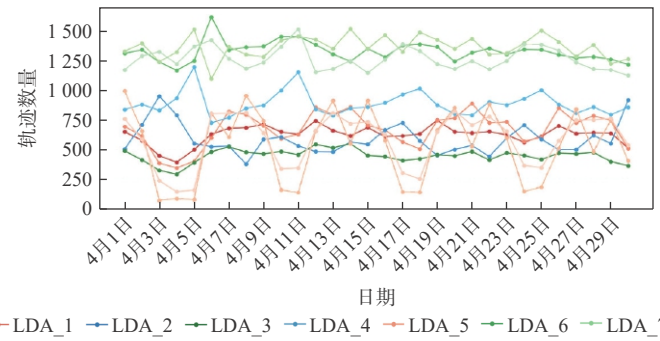
方法	ARI	NMI
K-means vs LDA	0.291	0.373
K-means vs STRMM	-1.665×10^{-5}	1.959×10^{-5}
LDA vs STRMM	2.393×10^{-5}	3.789×10^{-5}

3) 从模式识别效果来看, 相对于其他 2 个模型, STRMM 模型更能捕获周期性的典型模式。

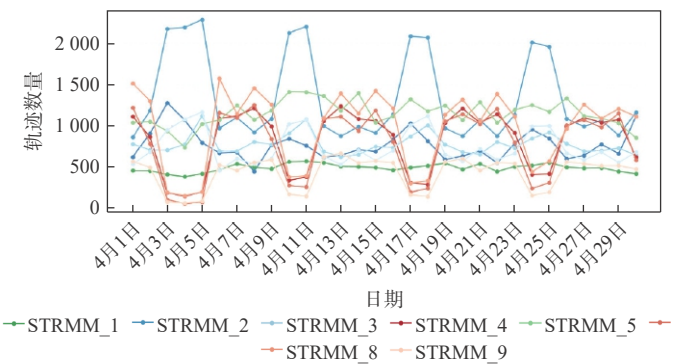
绘制 3 种方法所识别的各模式在 4 月每一天中包含的轨迹数量, 如图 10 所示, 分别用红色系、蓝色系、绿色系表示周中典型、周末典型、非典型自由活动轨迹。4 月 3—5 日是中国传统节日清明节的假期, 4 月 10 日、11 日、17 日、18 日、24 日、25 日为周末, 其余为常规的工作日。可以看到, 图 10(a)中 K-means 方法的类别 2 和图 10(b)中 LDA 方法的类别 0、1、5、8 为典型工作日模式, 而图 10(c)中 STRMM 方法的类别 4、6、8、9 为典型工作日模式。以图 10(c)为例, STRMM 方法识别出的类别 4、6、8、9 在全部工作日的轨迹数量均维持高位, 且在节假日骤降, 呈现出清晰的周期性; 而



(a) K_means识别结果



(b) LDA识别结果



(c) STRMM识别结果

图 10 3种方法识别结果在一个月内的轨迹数量

Fig. 10 Number of trajectories identified by three methods within one month

K-means与LDA方法的结果则存在更多噪声，表明STRMM方法的周期行为捕捉能力更强。同时，还识别出了图10(c)中的类别2等典型周末模式。尽管3种方法在模型输入时均未考虑轨迹是否处于工作日或是周末，但模式识别结果却可以看出工作日典型特征，说明了3种识别方法的有效性与实用性；STRMM模型进一步识别出周末特征，体现了本场景中STRMM模型的优势。

3.3 群体分布识别

在3.2节中，STRMM模型识别出了每日轨迹级别的10类活动模式。简单地说，给定特定用户在特定某日的全天活动轨迹，可以判断该用户于该日活动属于10类活动模式的哪一种，如常规学习型、外出社交型、长居宿舍型等。但由于最终目的是对用户群体行为进行分析，了解其长期活动模式演变情况，因此，需要进行用户群体识别。

使用拓展的STRMM模型完成这一聚类过程。具体操作步骤是：根据当前每个用户3个月内每日活动轨迹的所属模式，由于部分用户在某些天内存在记录缺失，在3.2节识别出的10类活动模式基础上，新增一个表示当日轨迹缺失的“空”类型，共计11类单日活动模式。令O表示总的单日活动模式数量($O=11$)，活动模式集合为$P=\{P_1,P_2,\cdots,P_O\}$。令U表示目前需要进行模式识别的本科生总数($U=9\,755$)，用u指代特定的一个用户($u=1,2,3,\cdots,U$)。令D表示目前数据覆盖的总天数($D=92$)，用d指代特定的一天($d=1,2,3,\cdots,D$)。类似地，与识别单日轨迹级别的活动主题时假定一天中的每个时间块的模式是相互独立的一样，将92 d以2周为间隔切分成7个周期(前6个周期各14 d，第7个周期为8 d)，并且认为每个周期的模式都是相互独立的。记周期数为$W(W=7)$，用w指代特定的一个

周期 ($w=1,2,3,\cdots,W$), 并且设第 w 个周期内包含 D_w 天, 满足 $\sum_{w=1}^{W} D_w = D$。则对于特定用户 u, 其长期活动模式序列可以定义为一个长度为 D 的向量 $[p_1, p_2, \cdots, p_D]$, 其中, $p_d \in P (d=1,2,3,\cdots,D)$ 表示用户 u 在第 d 天的活动模式情况。进一步加入周期信息, 特定用户 u 的活动演变序列如图 11 所示, 其中, TP 表示周期切片, $p_{u,l}^{(w)} (u=1,2,3,\cdots,U; l=1,2,\cdots,D_w; w=1,2,3,\cdots,W; P_{u,l}^{(w)} \in P)$ 表示第 u 个用户在第 d 个周期的第 l 天内的活动模式。

具体的模式识别方法与 3.2.1 节中模式识别方法类似, 不再赘述。通过计算不同类别数量下拓展的 STRMM 模型的困惑度, 最终获得了 6 类具有不同长期活动模式的用户群体。

	时间段 1				…	时间段 W			
	TP 1	TP 2	…	TP D_1	…	TP $D-D_W$	…	TP $D-1$	TP D
第1个用户	$p_{1,1}^{(1)}$	$p_{1,2}^{(1)}$	…	$p_{1,D_1}^{(1)}$	…	$p_{1,1}^{(W)}$	…	$p_{1,D_W-1}^{(W)}$	$a_{1,D_W}^{(W)}$
第2个用户	$p_{2,1}^{(1)}$	$p_{2,2}^{(1)}$	…	$p_{2,D_1}^{(1)}$	…	$p_{2,1}^{(W)}$	…	$p_{2,D_W-1}^{(W)}$	$p_{2,D_W}^{(W)}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
第U个用户	$p_{U,1}^{(1)}$	$p_{U,2}^{(1)}$	…	$p_{U,D_1}^{(1)}$	…	$p_{U,1}^{(W)}$	…	$p_{U,D_W-1}^{(W)}$	$a_{U,D_W}^{(W)}$

图 11　用户的长期活动模式序列图

Fig. 11　Long term activity pattern sequence of users

4　活动和群体模式识别结果分析

4.1　活动模式识别结果

对 3.2 节 STRMM 模型 10 类模式识别结果进行展示分析, 阐释每类模式含义。

以每个模式的单日轨迹为一行, 将时间切片作为横轴, 用不同的色块表示不同时间片的活动, 展示 10 类模式每日活动情况, 如图 12 所示。各类模式按照活动类别特征区分明显, 如各分类在 0:00—7:00 时段的表现能够有效区分睡觉时是否关闭信息设备, 以及起床时间差异。进一步绘制每类模式下每个时间块内不同活动出现的概率, 如图 13 所示。

结合每个模式在工作日或周末的轨迹数量分布(见图 10(c)), 根据轨迹数量在周末、周中更多或持平, 可以将这 10 类活动模式划分为第 0、2、3、7 类典型周末活动、第 4、6、8、9 类典型周中活动和第 1、5 类非典型自由活动轨迹。

分析图 10(c)和图 13 活动模式的概率分布, 周末活动对应类别为第 0、2、3、7 类, 占总活动轨迹的 45.8%, 其特征为含有较多的断联无记录活动(离校或休眠)和宿舍活动。其中, 第 2 类的周末“宿舍休息”活动在每个时间块的发生概率均为最高, 说明大多数学生在周末会选择整天在宿舍休息。第 7 类与第 2 类类似, 只是在早上和深夜出现更多的无记录活动, 说明学生此时可能关闭了电子设备进行休息。而第 0 类活动, 周六 14:00 前“宿舍休息”发生概率较高, 14:00 后“断联无记录”活动发生概率较高, 说明此类同学多选择在周末下午外出。第 3 类活动, 14:00 前“断联无记录”发生概率较高, 14:00 后“宿舍休息”和“教学区学习”活动发生概率上升, 说明此类同学多选择周末早起外出。综上所述, 将第 2、7 类定义为“长居宿舍型”, 第 0 类定义为“晚起外出型”, 第 3 类定义为“早起外出型”。

周中活动对应第 4、6、8、9 类, 占总活动轨迹的 34.2%, 核心特征表现为上下午工作时段(07:00—11:00、14:00—19:00)呈现高频的“教学区学习”活动。其中, 第 4 类除了有较多的“教学区学习”活动, 还会有较多的“食堂吃饭”活动和“文体区活动”; 第 8 类与第 9 类类似, 都是典型的规律上课活动, 只是第 9 类在早上和深夜会出现更多的无记录活动, 可能这部分轨迹的用户习惯于在睡眠时将电子设备关闭, 或者居于校外; 而第 6 类在工作时段除了较多的“教学区学习”活动以外, 还存在较多的“科研”活动, 体现了本科生在规律上课外的科研活动。综上所述, 将第 4 类定义为“学娱平衡型”, 第 8、9 类定义为“标准教学型”, 第 6 类定义为“教研复合型”。

自由活动对应类别为第 1、5 类, 占总活动轨迹的 19.9%, 活动以“科研活动”、“文体活动”和“宿舍休息”为主, 且在工作日和周末的发生概率无较大差异。其中, 第 1 类活动模式的“科研活动”活动发生概率远高于其他类活动模式, 该类的“其他”活动发生概率同样最高, 包括校医院就医、超市购物等生活活动。而第 5 类呈现出“科研活动、文体活动、宿舍休息”活动概率均较高, 体现上述 3 类活动相互交集的现象。综上所述, 将第 1 类定义为“专注科研型”, 第 5 类定义为“科研-休闲复合型”。

计算了 10 类活动模式在不同年级群体中的轨迹数量分布, 如图 14 所示。可以发现, 周末活动(第 0、2、3、7 类)均呈现年级越高分布概率越高的情况, 说明高年级的大学生由于课程量减少, 会出

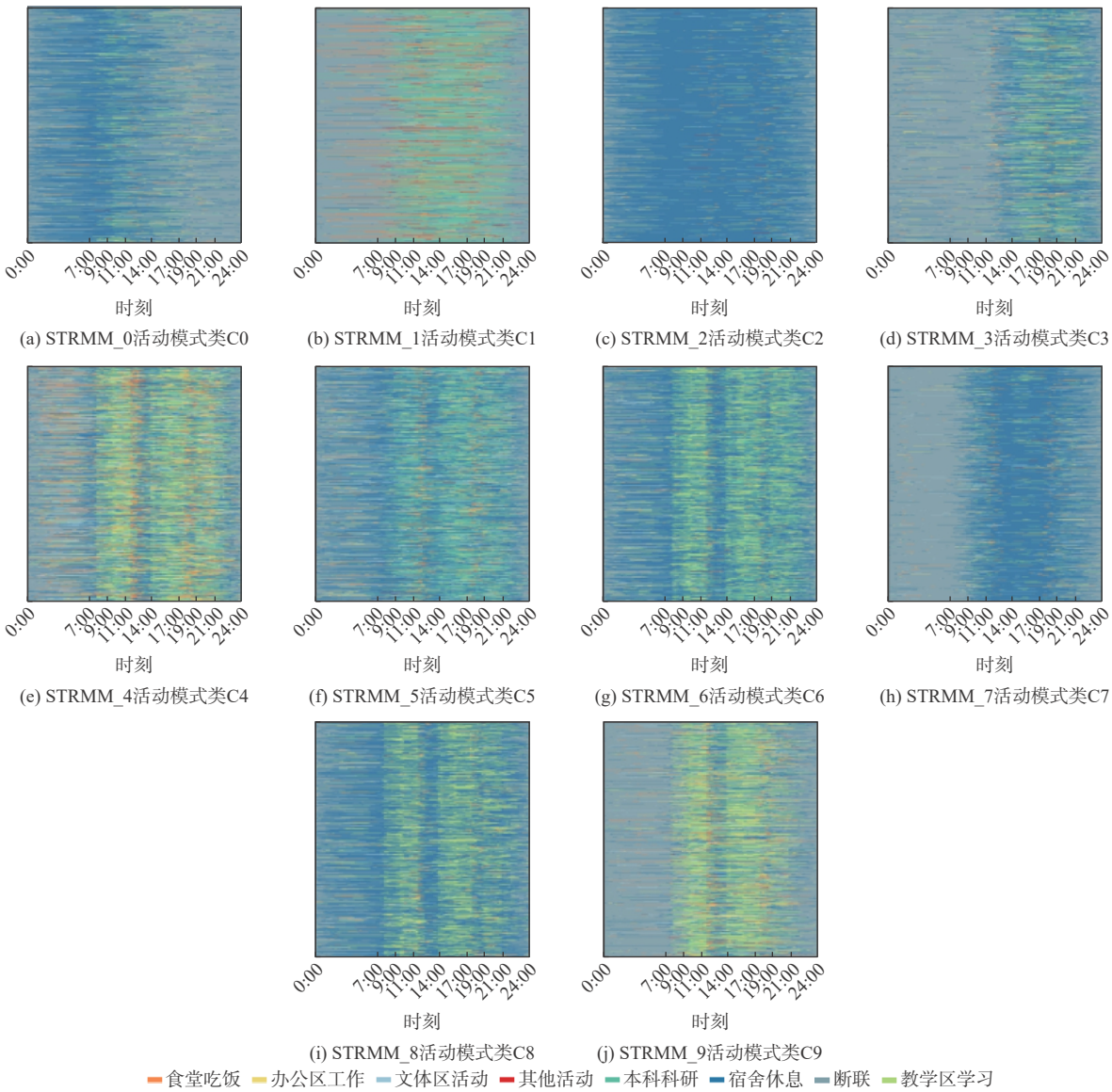


图 12 典型活动模式的时间序列分布图

Fig. 12 Time series distribution diagram of typical activity patterns

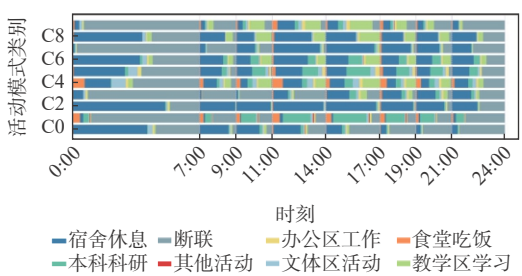


图 13 单日活动模式的概率分布

Fig. 13 Probability distribution of daily activity pattern

现更多的周末活动；而周中活动（第 4、6、8、9 类）均呈现年级越低分布概率越高的情况，说明大一、大二的新生会有更规律的周中活动；对于自由活动（第 1、5 类），其中，第 5 类“科研-休闲复合型”呈现年级越高分布概率越高的情况，说明高年级学生有较长自由时间进行休息、科研、文体活动，而第 1 类“专注科研型”总量较低，在大四、大一学生中出现概率较高，说明大学期间学习为主，投入科研的总量较低，毕业和入学年级对科研更加专注。

	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9
大一	0.069	0.061	0.095	0.092	0.144	0.079	0.129	0.081	0.164	0.086
大二	0.084	0.050	0.151	0.096	0.100	0.104	0.114	0.077	0.155	0.069
大三	0.096	0.056	0.172	0.094	0.076	0.156	0.093	0.091	0.106	0.060
大四	0.127	0.087	0.251	0.103	0.025	0.200	0.023	0.131	0.028	0.023

图 14 不同年级的活动模式分布

Fig. 14 Distribution of activity patterns in different grades

4.2　群体模式识别结果

3.3 节在单日活动模式基础上, 考虑每位本科生在 3 个月内的活动模式演变情况, 针对识别出的具有不同活动模式偏好的 6 类群体进行分析。

类似地, 以 3 个月 92 d 作为横轴, 以每位用户在每天的活动模式序列作为纵轴, 不同色块表示不同的活动模式, 展示这 6 类主要群体活动模式分布, 如图 15 所示。绘制 6 类群体在 3 个月内的活动模式概率分布, 如图 16 所示, 以及不同群体在各个年级的人数分布, 如图 17 所示。

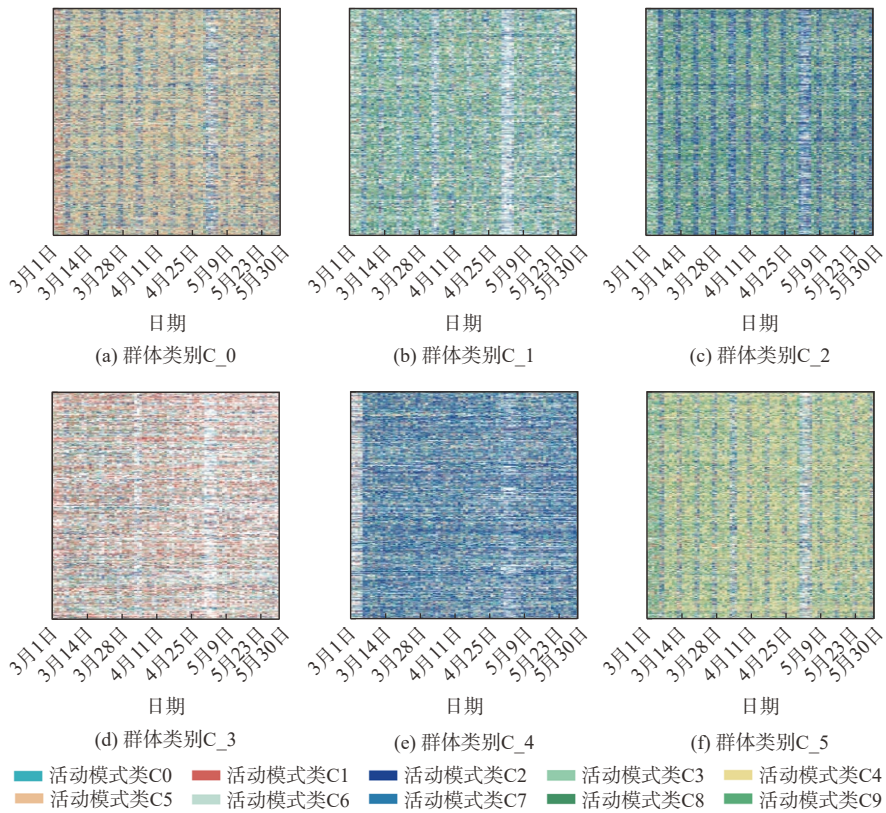


图 15　典型群体的活动模式序列分布

Fig. 15　Distribution of activity pattern sequence of typical groups

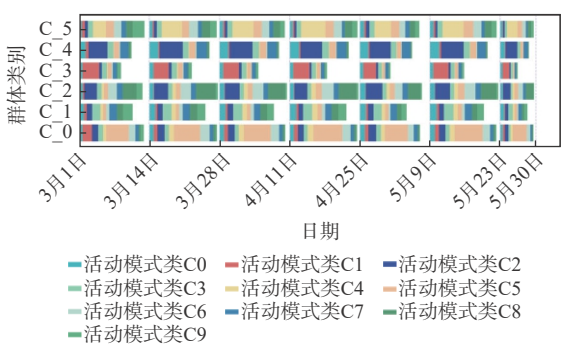


图 16　群体的活动模式概率分布

Fig. 16　Probability distribution of group activity patterns

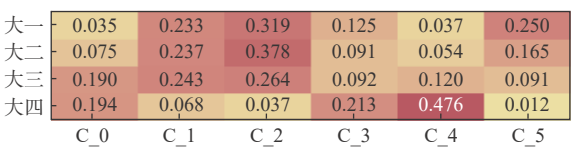


图 17　不同年级群体的活动模式分布

Fig. 17　Distribution of activity patterns among different grade groups

分析图 16 和图 17, 第 0 类群体“科研-休闲复合型”活动模式的出现概率最高, 且该群体随着年级越高占比越大, 大四占比达 19.4%, 较大一提升了 15.9%, 说明课程较少的高年级的大学生, 更倾向于较长时间宿舍休息、规律的科研活动, 同时坚持健身, 形成平衡的生活方式。

第 1 类群体与第 2 类相似, 含有较多的“标准教学型”模式及“教研复合型”模式, 且这 2 类群体随着年级越高占比越小, 说明 2 类群体大多为有较多课程的低年级大学生。第 1 类存在更多的记录缺失现象, 表征该群体休息时会关机或在校外居住。

第 3 类群体中“专注科研型”模式出现概率最高, 且在大四群体中占比最高。一方面, “专注科研型”模式拥有较多的科研活动, 符合大四学生科研论文写作需求强的实际规律, 同时说明高年级大学生就医等其他需求较强。

第 4 类群体的“长居宿舍型”模式出现概率最高, 且该群体随着年级越高占比越大, 大四占比达

47.6%, 较大一提升了 43.9%, 说明高年级同学由于没有较多规定课程, 更倾向于在宿舍完成大部分的日常活动。

第 5 类群体的“学娱平衡型”模式出现概率最高, 并且该群体随着年级越高占比越小, 侧面反映了低年级新生更愿意在学习之余, 积极参与健身和社交活动。

在对不同活动群体的分析中, 发现学生的日常行为模式与年级呈现出明显的相关性。随着年级的升高, 学生的行为模式逐渐从以学习和社交为主的规律性活动, 转向以休息、科研和自我管理为主的自主性活动。这种转变不仅反映了学生在学业压力和生活重心上的变化, 也揭示了他们在适应不同学习阶段时所表现出的灵活性和多样性。同时, 高年级群体中出现的高频科研活动与就医需求的增加, 进一步说明了年级较高学生在面对较大科研压力时, 更需要重视身体的健康状况。总体而言, 通过模式识别方法所获得的学生群体画像, 体现了学生群体在成长过程中的适应性与多样性, 以及他们在不同阶段对资源和时间分配的策略调整。

5 结 论

1) 充分挖掘校园 Wi-Fi 数据的时空特征, 建立了从物理空间到语义空间的全链条活动模式分析框架, 通过轨迹重构、活动解析、模式识别、群体识别, 系统性挖掘了高校学生活动规律与群体特征。

2) 引入 DMR 模型和 STRMM 模型, 准确识别出校园场景下 10 类功能区域、10 类典型活动模式与 6 类学生群体。结果表明, 低年级学生行为规律性强, 以课程与社团为主; 高年级则更多表现为科研、弹性作息等自主性活动, 体现出学生主体性发展的阶段性特征。

3) 创新点体现在 3 方面: ①突破传统 Wi-Fi 用于网络行为分析的局限, 实现低成本、高精度的校园区域功能与移动性刻画, 解决了校园场景下, 依赖单一出行方式或单一人群数据所带来的代表性不足与判断偏差问题; ②通过“区域-活动-模式-群体”分析链, 完成细粒度动态识别与语义映射, 有效结合动态轨迹和静态地标数据进行活动识别, 优于传统方法; ③揭示了学生行为随年级的动态演变规律, 验证了 STRMM 模型在周期行为捕捉上的优越性。为“三全育人”背景下的差异化教育策略的制定提供了定量依据。

本文仍存在一定局限性: 仅基于 Wi-Fi 数据进行轨迹描述与活动刻画, 未融合消费、门禁等多源信息; 考虑隐私保护未关联学业表现, 无法进行行为模式与学业发展之间深层关联的探索。未来将构建隐私保护下的多源数据融合框架, 推动学生管理从“事后治理”向“事前预防”的数字化转型。

参考文献 (References)

[1] 张文梅, 祁彬斌, 范文翔. 数据驱动的教学行为分析: 现状、逻辑与发展趋向[J]. 远程教育杂志, 2021, 39(1): 84-93.
ZHANG W M, QI B B, FAN W X. Analysis of data-driven teaching behavior: current situation, basic logic and development trend [J]. Journal of Distance Education, 2021, 39(1): 84-93(in Chinese).

[2] 谢承杰, 卢树华, 石洋宇, 等. 面向跨域人群计数的教师-学生复合优化方法[J/OL]. 北京航空航天大学学报, 2025[2025-10-01]. https://link.cnki.net/urlid/11.2625.V.20250106.1634.004.
XIE C J, LU S H, SHI Y Y, et al. A teacher-student hybrid optimization method for cross-domain crowd counting[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2025[2025-10-01]. https://link.cnki.net/urlid/11.2625.V.20250106.1634.004(in Chinese).

[3] 张涛, 张思. 教育大数据挖掘的学习者模型设计与计算研究[J]. 电化教育研究, 2020, 41(9): 61-67.
ZHANG T, ZHANG S. Research on design and computing of learner model of educational big data mining[J]. e-Education Research, 2020, 41(9): 61-67(in Chinese).

[4] 李有增, 曾浩. 基于学生行为分析模型的高校智慧校园教育大数据应用研究[J]. 中国电化教育, 2018(7): 33-38.
LI Y Z, ZENG H. Smart campus big data education application based on students' behavior analysis model[J]. China Educational Technology, 2018(7): 33-38(in Chinese).

[5] 陈辉, 白骏, 殷传涛, 等. 基于学习行为的 MOOC 用户持续学习预测框架[J]. 北京航空航天大学学报, 2023, 49(1): 74-82.
CHEN H, BAI J, YIN C T, et al. Behavior based MOOC user dropout predication framework[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(1): 74-82(in Chinese).

[6] 刘科生, 倪义坤. 高校学生工作数据挖掘的研究和实践[J]. 北京航空航天大学学报 (社会科学版), 2020, 33(5): 147-152.
LIU K S, NI Y K. Research and application of data mining on college student affairs management[J]. Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition), 2020, 33(5): 147-152(in Chinese).

[7] 胡祖辉, 施佺. 高校学生上网行为分析与数据挖掘研究[J]. 中国远程教育, 2017, 37(2): 26-32.
HU Z H, SHI Q. Analysis of college students' online behavior and research on data mining[J]. Chinese Journal of Distance Education, 2017, 37(2): 26-32(in Chinese).

[8] HSU W, HELMY A. Impact: investigation of mobile-user patterns across university campuses using WLAN trace analysis[EB/OL]. (2005-08-02)[2025-10-01]. https://arxiv.org/abs/cs/0508009.

[9] FARHAN A A, BAMIS A, WANG B. Locating emergencies in a campus using Wi-Fi access point association data[C]//Proceedings of the ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication. New York: ACM, 2013: 1123-1134.

[10] DAS SWAIN V, KWON H, SARGOLZAEI S, et al. Leveraging WiFi network logs to infer student collocation and its relationship with academic performance[J]. EPJ Data Science, 2023, 12(1): 22.

[11] JUNDEE T, PHITHAKKITNUKOON S, RATTI C. Inferring trips and origin-destination flows from Wi-Fi probe data: a case study of campus Wi-Fi network[J]. IEEE Access, 2023, 11: 63351-63364.
[12] POUCIN G, FAROOQ B, PATTERSON Z. Activity patterns mining in Wi-Fi access point logs[J]. Computers, Environment and Urban Systems, 2018, 67: 55-67.
[13] GONZÁLEZ M C, HIDALGO C A, BARABÁSI A L. Understanding individual human mobility patterns[J]. Nature, 2008, 453(7196): 779-782.
[14] SONG C M, QU Z H, BLUMM N, et al. Limits of predictability in human mobility[J]. Science, 2010, 327(5968): 1018-1021.
[15] SCHNEIDER C M, BELIK V, COURONNÉ T, et al. Unravelling daily human mobility motifs[J]. Journal of the Royal Society Interface, 2013, 10(84): 20130246.
[16] LIU W B, WANG B X, YANG Y, et al. Cluster analysis of microscopic spatio-temporal patterns of tourists' movement behaviors in mountainous scenic areas using open GPS-trajectory data[J]. Tourism Management, 2022, 93: 104614.
[17] JIANG S, FERREIRA J, GONZÁLEZ M C. Clustering daily patterns of human activities in the city[J]. Data Mining and Knowledge Discovery, 2012, 25(3): 478-510.
[18] CALABRESE F, READES J, RATTI C. Eigenplaces: segmenting space through digital signatures[J]. IEEE Pervasive Computing, 2010, 9(1): 78-84.
[19] FARRAHI K, GATICA-PEREZ D. Mining human location-routines using a multi-level approach to topic modeling[C]//Proceedings of the IEEE 2nd International Conference on Social Computing. Piscataway: IEEE Press, 2010: 446-451.
[20] YUAN J, ZHENG Y, XIE X. Discovering regions of different functions in a city using human mobility and POIs[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2012: 186-194.
[21] QIN T, SHANGGUAN W F, SONG G J, et al. Spatio-temporal routine mining on mobile phone data[J]. ACM Transactions on Knowledge Discovery from Data, 2018, 12(5): 1-24.
[22] YIN L, LIN N, ZHAO Z Y. Mining daily activity chains from large-scale mobile phone location data[J]. Cities, 2021, 109: 103013.
[23] WIDHALM P, YANG Y X, ULM M, et al. Discovering urban activity patterns in cell phone data[J]. Transportation, 2015, 42(4): 597-623.

Study of university students' behavioral patterns via activity chain analysis

ZHAI Yikai[1]，JIANG Xiaotong[1]，TIAN Qiong[1, 2, *]，HUANG Haijun[1]

(1. MoE Key Laboratory of Complex System Analysis and Management Decision，School of Economics and Management，Beihang University，Beijing 100191，China；
2. Hangzhou International Innovation Institute of Beihang University，Hangzhou 311100，China)

Abstract：In response to the demands of educational modernization, the scenario-based application of campus data provides new opportunities for the digital transformation of higher education. In order to identify spatiotemporal patterns of campus behavior, this study uses trajectory reconstruction, semantic mapping, and pattern mining on three months' worth of Wi-Fi logs and point of interest (POI) data from thirty thousand students and faculty. Innovatively introducing the Dirichlet multinomial regression (DMR) model and spatio-temporal routine mining on mobile phone data (STRMM) model for student behavior analysis, the DMR model effectively integrates dynamic trajectories and static landmark data to identify 10 categories of campus functional areas. The STRMM model enhances the ability to capture periodic and uncertain behaviors, categorizing undergraduate daily activities into 10 typical patterns, including standard teaching-oriented and focused research-oriented types. Additionally, grade-level dynamics in student behavior were shown by identifying six types of undergraduate groups with distinct evolutionary behavioral patterns. These groups demonstrated a shift from course-dominated activities in lower grades to self-directed research and flexible schedules in higher grades. The study confirms that Wi-Fi data-based behavioral analysis can effectively identify functional areas and student behavior characteristics, providing data support for precise management, resource optimization, and the practice of 'Three-Comprehensive Education' with important practical reference value for promoting the digital transformation of higher education.

Keywords：activity chain analysis; pattern mining; travel behavior; traffic management; educational big data

Received：2025-10-10；**Accepted**：2025-11-21；**Published Online**：2025-12-05 09：43

URL：link.cnki.net/urlid/11.2625.V.20251204.1719.003

Foundation items：National Key Research and Development Program of China (2023YFE0115600); National Natural Science Foundation of China (72288101,72394374)

*** Corresponding author**. E-mail：tianqiong@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0124

基于感知增强与多尺度特征融合的小目标车辆检测

沈瑜*，李阳阳，李博昊，高宝渠，魏子易，白珊
(兰州交通大学 电子与信息工程学院，兰州 730070)

摘 要：为解决车辆小目标携带信息匮乏，特征表达能力弱，导致现有算法检测精度低、漏检等问题，提出一种基于感知增强与多尺度特征融合的小目标车辆检测算法。设计空间局部特征感知增强主干网络SLFPB-ST，解决特征提取过程中小目标特征信息丢失严重的问题；提出一种多尺度特征融合网络(MSIFN)，通过分配权重，关注更多的细节信息，同时，在MSIFN中加入大目标抑制块(LRB)约束大目标特征，保留小目标特征表达；采用无锚框机制减少小目标漏检问题，提高检测精度。在UA-DETRAC数据集和Vehicle数据集上的实验结果表明：与Swin Transformer算法相比，所提算法在mAP、AP50和AP75指标上分别提升5.15%、9.35%和4.35%，参数量增加14 MB，检测速度降低0.4帧/s，验证了算法具有良好的鲁棒性和适用性。

关 键 词：小目标车辆检测；Swin Transformer；多尺度特征融合；空洞卷积；GIOU
中图分类号：V221+.3; TB553
文献标志码：A **文章编号**：1001-5965(2026)05-1422-11

随着经济的快速发展，车辆在人们的日常生活中已变得越来越普遍[1]，而与之相关的行车安全问题也变得日益严峻[2]。近年来，深度学习在工业、交通、建筑业维护[3]等领域取得了显著成果，为车辆检测提供了精准和高效的解决方案。

基于深度学习的目标检测算法大致分为3类：①以R-CNN[4]、Faster R-CNN[5]、Cascade R-CNN[6]等为代表的Two-stage算法，这类算法利用卷积神经网络(convolutional neural network, CNN)生成候选框，再进行分类。②以SSD[7]、RetinaNet[8]、CenterNet[9]、YOLO系列[10-11]算法为代表的One-stage算法，直接将图像通过CNN进行分类和定位，速度较快，但精度较低。③基于Transformer[12]框架的算法，如ViT[13]、DETR[14]和Swin Transformer[15]等，近年来，这类算法在目标检测领域取得了显著的进展。以上检测算法可以满足一般的检测需求，但对于目标尺寸小于32像素×32像素的小目标，检测效果并不理想。因此，研究小目标的检测算法和技术仍是当前研究的重要方向之一[16]。

目前，目标检测算法在处理小目标时，常面临大量漏检的情况。然而，随着目标检测算法的不断优化和改进，小目标的检测性能有所提升。一些学者提出了改进的小目标检测算法。例如，Fu等[17]提出了DSSD算法，通过反卷积操作和跳跃连接技术优化对小目标的检测性能，加强了对小目标细节信息的提取能力，但模型的复杂度较高；Zhang等[18]引入了M-CBAM注意力机制和自适应融合因子，有效减少了小目标的漏检问题，但模型计算量较高；Chen等[19]提出了一种动态拼接图像的方法，以减轻小目标检测中正负样本不均衡的问题，但检测精度慢，不能满足实时性的要求。还有学者专注于改进特征提取和融合方法。例如，Liu等[20]提出了一种新颖的特征融合方法，通过学习权重参数有效融合不同层的特征，提升小目标的检测性能。此外，Zoph等[21]将神经网络结构搜索方法用于目标检测领域，优化数据，增加小目标样本数量，提升检

收稿日期：2024-03-05；录用日期：2024-06-21；网络出版时间：2024-07-19 10:57
网络出版地址：link.cnki.net/urlid/11.2625.V.20240717.1624.002
基金项目：国家自然科学基金(61861025,62241106)
*通信作者. E-mail：18609311366@163.com

引用格式：沈瑜，李阳阳，李博昊，等. 基于感知增强与多尺度特征融合的小目标车辆检测[J]. 北京航空航天大学学报，2026，52(5)：1422-1432. SHEN Y，LI Y Y，LI B H，et al. Small target vehicle detection based on perceptual enhancement and multi scale feature fusion[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1422-1432 (in Chinese).

测性能, 但模型鲁棒性有待验证。面对小目标检测的挑战, 学术界正在积极探索各种新颖方法, 以提高目标检测算法在复杂场景中的性能表现。

尽管上述算法在一定程度上解决了小目标细节信息丢失和漏检的问题, 提高了小目标检测的准确性, 但仍不能完全满足任务精准度和鲁棒性的要求。基于此, 本文创新点如下:

1) 设计了一种空间局部特征感知块(spatial local feature perception block, SLFPB), 通过融合空洞卷积和残差连接来增强模型对图像局部特征信息的提取能力, 减少小目标细节信息的丢失。这种设计提高了模型的灵活性和自适应性。

2) 针对小目标样本携带信息匮乏、漏检率较高的问题, 在多尺度特征融合网络(multi-scale information fusion network, MSIFN)中设计了一个大目标抑制块(large objects restriction block, LRB), 其目的是利用深层特征图中的大目标特征信息, 保留对小目标更敏感的浅层特征图中的信息, 从而解决在特征融合过程中小目标特征信息容易丢失和漏检的问题, 使得小目标检测算法的鲁棒性及精准性得到有效提升。

3) 通过在 Vehicle 和 UA-DETRAC 数据集上进行实验, 算法检测精度有明显提升, 且具有更快的检测速度, 充分证明了本文算法的有效性和适用性。

1　本文模型

针对密集、遮挡的车辆数据集中, 小目标样本携带信息匮乏、特征表达能力弱、漏检率较高的问题, 本文提出了一种空间局部特征感知增强 Swin Transformer 的小目标车辆检测算法。整体网络模型由主干网络 SLFPB-ST、MSIFN 和无锚框(anchor free, AF)检测器组成, 算法结构如图 1 所示。

首先, 采用 Swin Transformer 作为特征提取的主干网络, 通过分层设计和移动窗口自注意力机制降低计算量, 增强深层特征的提取能力。为进一步加强模型对图像信息提取的能力, 提出了一种由空洞卷积和残差连接组成的 SLFPB, 将其放入 Swin Transformer 获取特征信息之前, 能够提取更多的空间局部上下文信息。其次, 在 MSIFN 中引入 LRB, 限制大目标特征的表达, 转移更多的小目标特征信息, 以便更容易地检测到小目标。然后, 采用软 RoI 选择机制, 对融合后的特征进行自适应学习, 以

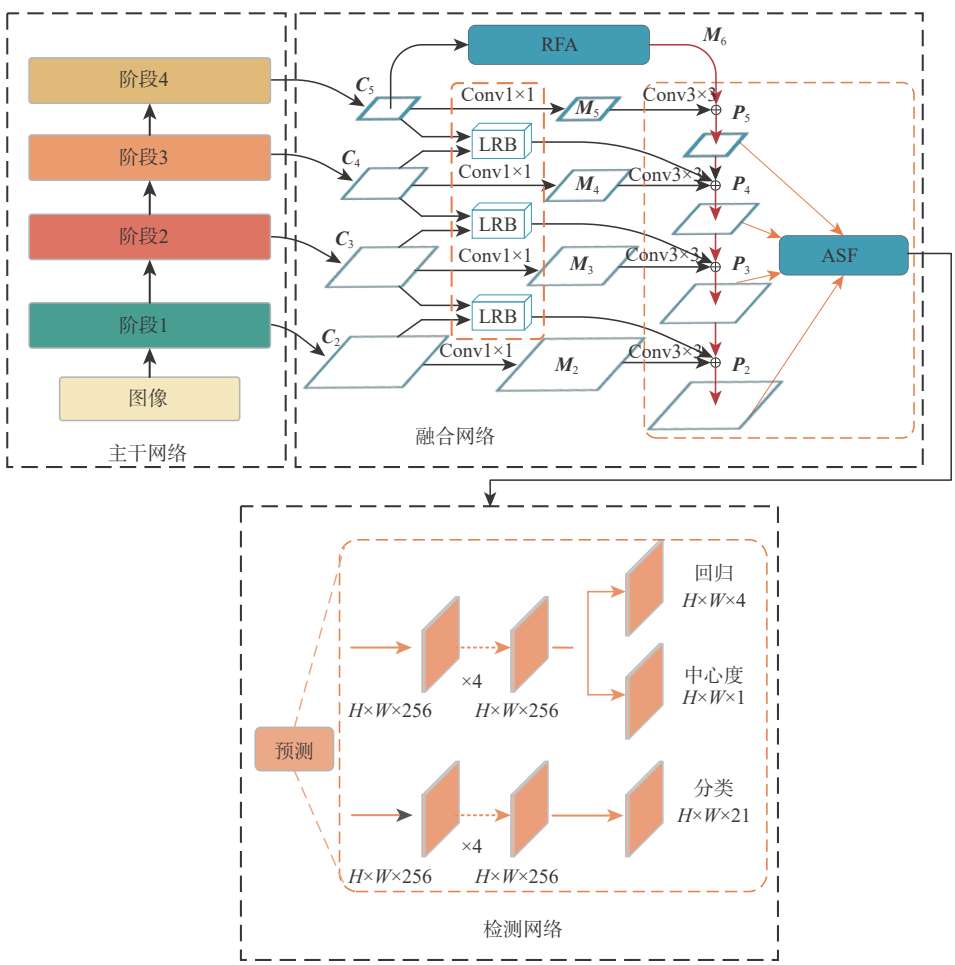


图 1　本文算法框架

Fig. 1　Framework of the proposed algorithm

输出固定大小的特征图。最后,使用无锚框检测机制来实现目标的准确分类和精准定位。设计旨在提高算法对小目标的检测性能和适用性,保持对大目标的有效识别能力。

1.1 Swin Transformer 网络

Swin Transformer 的网络结构如图 2 所示。将编码器分为 4 个阶段,均由补丁和若干 Swin Transformer Block 组成。首先,将图像进行分块操作;然后,通过线性嵌入层进行通道变换,利用 Swin Transformer Block 完成信息建模;最后,特征图分辨率变为原来的一半,通道数翻倍。这种结构设计不仅有效降低了计算成本,还提高了检测精度。

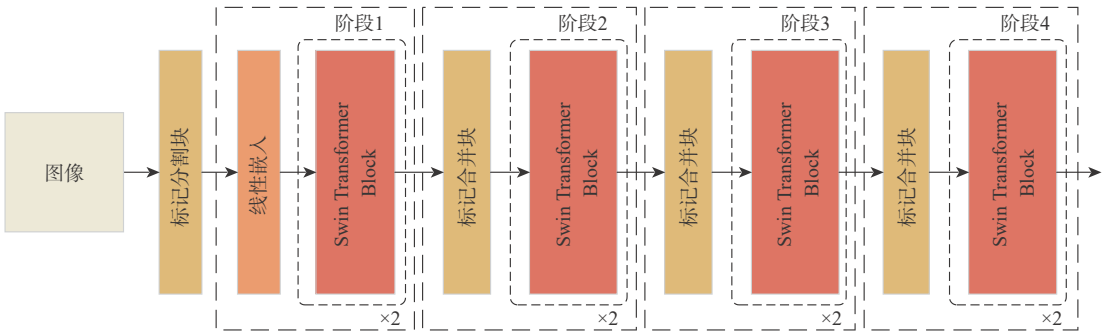


图 2 Swin Transformer 网络结构

Fig. 2 Network framework of Swin Transformer

1.2 改进的 Swin Transformer 主干网络

Swin Transformer Block 结构如图 3 所示。相较于传统的 Transformer Block, Swin Transformer Block 将原有的多头自注意力变为窗口多头自注意力 (window multi-head self attention, W-MSA) 和移动窗口多头自注意力 (shift window multi-head self attention, SW-MSA)。SW-MSA 利用窗口之间的移动来提取信息,通过循环移位和掩码操作来合并窗口并消除不关联部分,提高了检测性能。

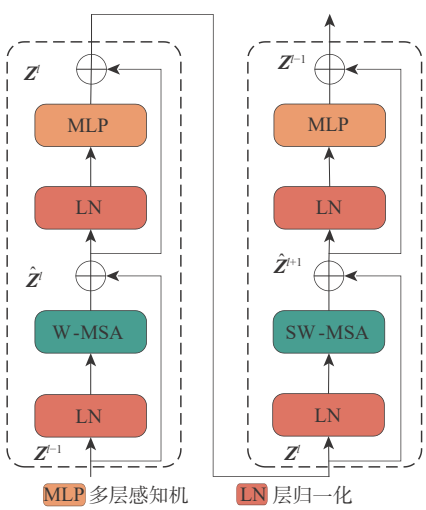


图 3 Swin Transformer Block 网络结构

Fig. 3 Network framework of Swin Transformer Block

对于车辆图像中较多小目标样本,Swin Transformer 还是无法针对大量空间上下文信息完成有效编码。针对该问题,提出了一种空间局部特征感知增强 Swin Transformer 主干网络 SLFPB-ST。网络结构如图 4 所示。SLFPB-ST 划分为 4 个阶段,任意阶段都由一个补丁、n个 SLFPB 及 n个 Swin Transformer Block 构成。其中,n表示对应阶段的模块数量,分别为 2、2、6、2。这种设计增加了复杂环境下对小目标细节信息的提取,减少了小目标的漏检。由于对网络参数量和检测性能高低的考虑,基准模型采用通用的 Swin-T 网络模型。

为提高模型对图像局部特征信息的获取能力,提出了一种由残差连接和空洞卷积构成的 SLFPB。SLFPB 对应放入到每个 Swin Transformer Block 前,来增强网络的性能。SLFPB 结构如图 5 所示。首先,调整输入的原始数据格式作为残差连接的第 1 个分支,并保持与 Swin Transformer 网络适应。其次,第 2 个分支由 3 层空洞卷积构成,卷积核大小为 3×3,空洞因子为 (1,2,5),使用不同的空洞因子提高模型的适应性。然后,使用 GELU 激活函数。最后,将 2 个输出分支相加,并将通道维度降至 256。设计无需进行池化操作,且计算量少,能够提供更大的感受野,获取更大范围的空间特征信息,避免了由常规下采样操作造成提取特征信息不充分且小目标信息丢失的问题。

1.3 多尺度特征融合网络

传统的特征金字塔只包含单向的特征融合路径,无法充分融合丰富的特征。为保留浅层特征图和深层特征图中小目标的特征信息,提出一种 MSIFN 结构,以实现不同尺度特征图的特征信息融合,网络结构如图 6 所示。图像经过主干阶段的提取,分辨率降低,影响图像中小目标信息的获取。因此,本文在 SLFPB-ST 主干网络后加入了 MSIFN 结构,融合浅层特征及深层语义信息,促使小目标检测性能不断提升。

图 6 中,SLFPB-ST 主干网络从下到上对图像

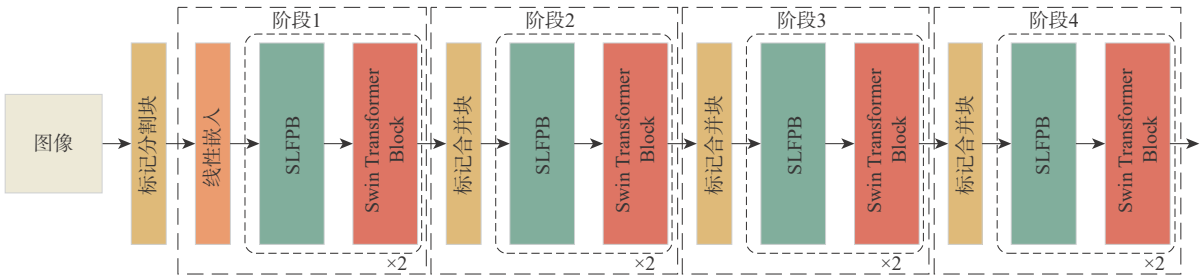


图 4　SLFPB-ST 主干网络

Fig. 4　SLFPB-ST backbone network

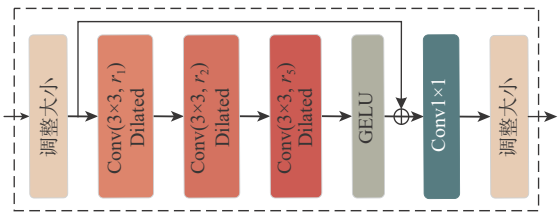


图 5　空间局部特征感知块

Fig. 5　Spatial local feature perception block

逐层进行特征提取, 生成$\boldsymbol{C}_2 \sim \boldsymbol{C}_5$这 4 种特征图。为消除不同层级融合的不兼容性, 需将$\boldsymbol{C}_2 \sim \boldsymbol{C}_5$的通道数统一降为 256 大小, 得到$\boldsymbol{M}_2 \sim \boldsymbol{M}_5$特征图。在自上而下的通道中, 从特征图$\boldsymbol{C}_5$开始融合, 利用高层次的语义信息增强低层特征图, 多样化的上下文信息就可直接提供出来。由于特征通道数的减少, 使$\boldsymbol{C}_5$层特征信息出现损失, 且每个通道只有单一尺度信息, 并不兼容其他层的特征。残差特征增强(residual feature augmentation, RFA)模块通过残差分支, 将不同的空间上下文信息融入到原始分支中, 减少了通道内信息丢失的问题, 提高了网络性能。这种方式可对各层次的特征进行充分利用, 确保层级之间的兼容性, 进而不断提升车辆检测的鲁棒性及准确性。

RFA 生成的$\boldsymbol{M}_6$经过 2 倍上采样后与$\boldsymbol{M}_5$经过 3×3卷积核进一步提取特征后的结果相加得到$\boldsymbol{P}_5$, $\boldsymbol{P}_5$经过 2 倍上采样后与$\boldsymbol{M}_4$经过3×3卷积核的结果相加得到$\boldsymbol{P}_4$, $\boldsymbol{P}_4$进行 2 倍上采样后与$\boldsymbol{M}_3$经过3×3卷积核的结果相加得到$\boldsymbol{P}_3$, $\boldsymbol{P}_3$进行 2 倍上采样后与$\boldsymbol{M}_2$经过3×3卷积核的结果相加得到$\boldsymbol{P}_2$。经融合后得到的特征图$\boldsymbol{P}_2 \sim \boldsymbol{P}_5$, 再经自适应空间融合(adaptive spatial fusion, ASF)模块将特征图自适应融合并输出。

图 7(a)为 RFA 模块实现不同上下文信息融合的过程。首先, 针对尺度为 S 的$\boldsymbol{C}_5$来说, 需要进行自适应池化操作, 以生成多种尺度$(\alpha_1 \times s, \alpha_2 \times s, \cdots, \alpha_n \times s)$的上下文特征($s$为特征图$\boldsymbol{C}_5$的基准尺度, 其高度和宽度分别为 H 和 W), 在自适应池化操作中, 以s为基准生成多尺度的上下文特征。α_n为一组缩放因子, 用于对基准尺度s进行缩放, 从而得到不同的特征图尺度)。其次, 每个特征单独经历卷积核为1×1的卷积层, 使通道数降低至 256。最后, 由双线性插值法得到上采样尺度 S, 并进行自适应融合。图 7(b)为 ASF 模块的实现细节过程。输入的上采样特征为每个特征生成相应的空间权重, 通过

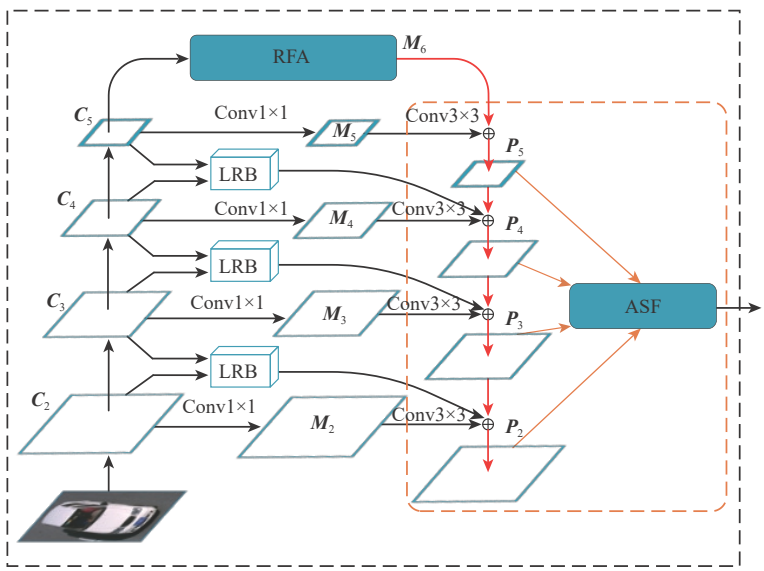


图 6　多尺度特征融合网络

Fig. 6　Multi-scale information fusion network

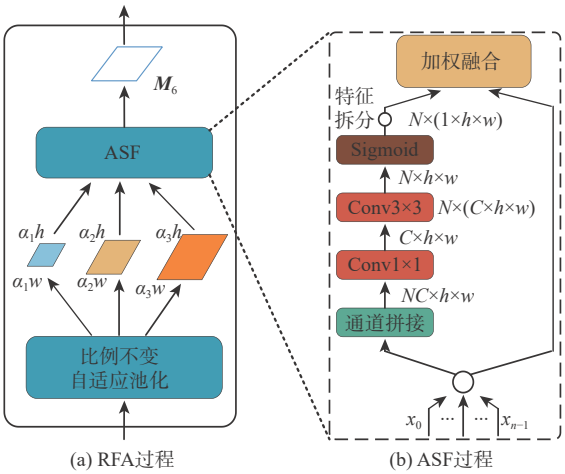


图7 RFA和ASF的过程

Fig. 7 Process of RFA and ASF

权重将上下文信息进行聚合，可直接赋予多尺度上下文信息，减少信息损失。

模型中，图像传递信息时，较深层特征图包含更少的小样本信息，相比之下，浅层特征图中更容易捕获小目标的特征信息。基于该特性，在$\boldsymbol{C}_2 \sim \boldsymbol{C}_5$特征图中每相邻2个特征图之间插入LRB，其结构如图8所示。LRB的主要功能在于利用深层特征图中的大目标信息，对浅层特征图中的大目标特征表达进行约束，确保小目标特征在浅层特征图中得以保留，并将输出连接至$\boldsymbol{P}_2 \sim \boldsymbol{P}_4$特征图。LRB具备$\boldsymbol{L}_i$、$\boldsymbol{L}_{i+1}$这2个输入端和$\boldsymbol{O}_i$这1个输出端，$\boldsymbol{L}_i$、$\boldsymbol{L}_{i+1}$分别接收浅层特征图和深层特征图。$\boldsymbol{L}_{i+1}$输入LRB后，先通过最近邻插值法提升深层特征图的分辨率，使其与浅层特征图相匹配；再借助包含卷积层、批量归一化BN和LeakReLU激活函数的CBL模块，将通道数调整一致；最终进行空间门控单元G_s处理，输出维度为$2H\times 2W\times C$的特征图（H为特征图高度，W为特征图宽度，C为特征图的通道数）。Sigmoid函数进行归一化操作，得到大目标的特征尺度函数$S(T)$，具体公式为

$$S(T)=\frac{1}{1+\mathrm{e}^{G_s(\mathrm{CBL}(\mathrm{up}(\boldsymbol{L}_{i+1})))}} \tag{1}$$

式中：up表示上采样操作；CBL表示1×1卷积+批量归一化+激活函数的统称，用于促进网络收敛。

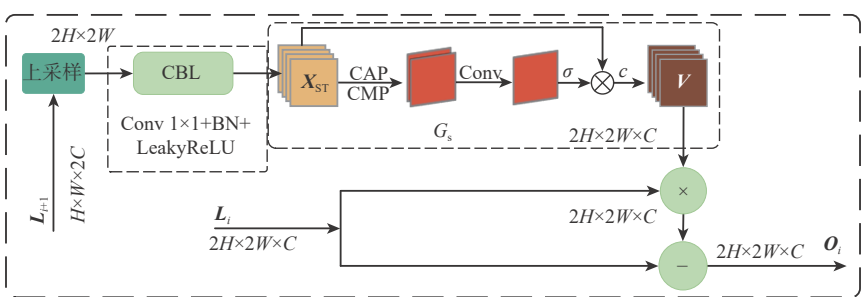
将大目标特征尺度函数与浅层特征图$\boldsymbol{L}_i$逐点相乘，增加原本具有大目标特征的部分，再与浅层特征图$\boldsymbol{L}_i$逐点相减，消除部分大目标特征信息，从而保留小目标特征信息，最终将得到的结果以逐点相加的方式融入到$\boldsymbol{O}_i$特征图中，生成新的特征图$\boldsymbol{O}_i^*$，目的是更好地保留小目标特征信息，提高小目标的检测性能。具体如下：

$$\boldsymbol{O}_i^*=(\boldsymbol{C}_i+\boldsymbol{C}_i\cdot S(T))+\boldsymbol{O}_i \tag{2}$$

空间门控单元G_s中通过卷积层生成二维权重矩阵，来更好地提取特征，再对矩阵归一化操作与$\boldsymbol{X}_{\mathrm{ST}}$矩阵相乘得到特征$\boldsymbol{V}$，具体公式如下：

$$\boldsymbol{V}=\boldsymbol{X}_{\mathrm{ST}}\cdot\sigma(C_{7\times7}(P_{\mathrm{a}}^{C}\boldsymbol{X}_{\mathrm{ST}}),(P_{\mathrm{m}}^{C}\boldsymbol{X}_{\mathrm{ST}})) \tag{3}$$

式中：P_{a}^{C}表示平均池化；P_{m}^{C}表示最大池化；$C_{7\times7}$表示卷积核大小为7×7的卷积层。



CAP：通道平均池化；σ：Sigmoid；CMP：通道最大池化；⊗：矩阵乘积。

图8 大目标抑制块

Fig. 8 Large objects restraint block

1.4 无锚框机制

FCOS算法的无锚框机制从像素点入手，不依赖预先定义的锚框，直接预测目标物体的中心点、宽度和高度，避免了与真实值匹配的问题。引入中心度损失有助于训练中区分正负样本，并抑制远离目标的低质量预测框，对中心度较低的小目标来说，极大减少了漏检率，提高了网络检测性能，并缓解了融合网络重合带来的影响。

1.5 模型损失函数

本文所提小目标车辆检测网络模型的损失函数定义如下：

$$L_{\text{center-ness}}=\sqrt{\frac{\min(l^*,r^*)}{\max(l^*,r^*)}\cdot\frac{\min(t^*,b^*)}{\max(t^*,b^*)}} \tag{4}$$

$$L=\frac{1}{N_{\mathrm{pos}}}\sum_{x,y}L_{\mathrm{cls}}(p_{x,y})+\frac{1}{N_{\mathrm{pos}}}\sum_{x,y}kL_{\mathrm{reg}}(t_{x,y})+L_{\text{center-ness}} \tag{5}$$

式中：l^*、r^*、t^*、b^*分别表示像素点到目标左、右、上、下边界的距离；L_{cls}为分类损失，采用 Focal Loss[8]焦点损失函数；L_{reg}为回归损失，采用 GIoU Loss[22]损失函数；$L_{center\text{-}ness}$为中心度损失函数；$p_{x,y}$为类别的预测概率；$t_{x,y}$为回归的预测坐标；N_{pos}为正类样本数；k为指数函数。

训练过程中，总损失函数的曲线如图 9 所示。当训练次数超过 65 时，损失函数数值趋于平稳，最终损失值稳定在 0.42 左右，训练结果较为理想。

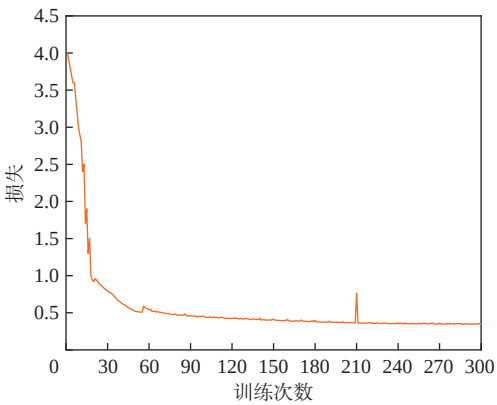


图 9　总损失曲线

Fig. 9　Total loss curve

2　实验结果及分析

2.1　数据集和评价指标

UA-DETRAC 是一个大规模多样化的车辆检测数据集，其涵盖的训练集和测试集图像分别为 52 085 张和 26 167 张，包含晴天、多云、夜间和雨天 4 类天气条件，是城市道路场景中数据量较大的评测数据集。为进一步在复杂场景下进行算法验证，证明本文算法的鲁棒性和实用性，使用了 Vehicle 数据集。Vehicle 数据集共标注了 21 种车辆类别，涵盖 3 000 多张图像，通过数据预处理将数据集扩充到 25 000 张，最终共计 103 252 张图像，按 8∶2 的比例划分为训练集和测试集，训练集将 10% 的比例划分为验证集。数据集的部分样本如图 10 所示。

为直观、全面地评价本文算法的性能，采用 COCO 数据集的指标（AP、AP(S)、AP(M)、AP(L)、AP50 和 AP75）作为本文的评价指标。常用的模型性评价指标公式如下：

$$P=\frac{T_P}{T_P+F_P} \tag{6}$$

$$R=\frac{T_P}{T_P+F_N} \tag{7}$$

$$F_1=\frac{2RP}{R+P} \tag{8}$$

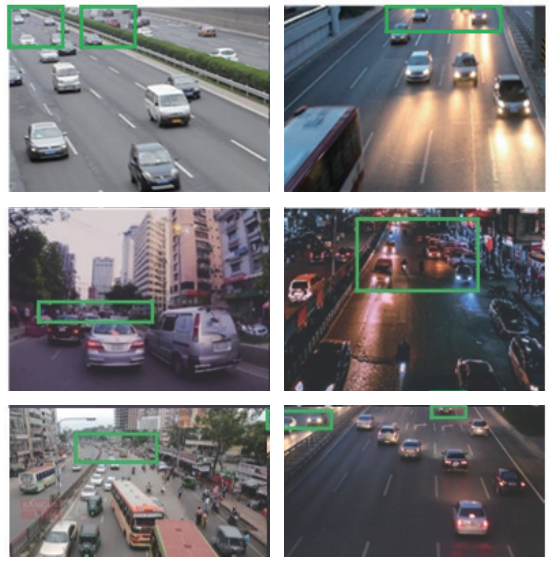


图 10　数据集部分样本示例

Fig. 10　Examples of partial samples of dataset

$$A_P=\int PR\mathrm{d}R \tag{9}$$

$$m_{AP}=\frac{1}{C}\sum_{K=1}^{C}A_{P,K} \tag{10}$$

式中：P 为精确率；R 为召回率；T_P 表示正确地将车辆分类的数量；F_P 表示错误地将车辆归类为其他类别的数量；F_N 表示未能识别出车辆的数量；A_P 为平均精度；m_{AP} 为均值平均精度。其中，将 mAP 作为目标检测性能的主要评价指标，AP(S)为 32 像素×32 像素以下目标的平均精度，AP(L)为 96 像素×96 像素以上目标的平均精度，AP(M)为介于 AP(S)和 AP(L)之间目标的平均精度。

2.2　实验平台及参数设置

本文使用的操作系统为 Windons 10，CPU 为 Intel Core i5-13600KF，GPU 为 NVIDIA RTX 3090，显存为 32 GB。软件环境为 Python 3.8.0、PyTorch 1.8.2、CUDA 11.1.0、OpenCV 4.8.0 和 mmcv-full 2.1.0。训练中，优化器选择 AdamW，动量因子和权重衰减因子分别为 0.9 和 0.002 5，初始学习率为 0.000 1，迭代次数为 300，训练批量大小设置为 8。

2.3　不同模型的对比实验

为验证本文算法的有效性，将其与主流目标检测算法在 Vehicle 数据集上进行训练和验证，结果如表 1 所示。本文算法的 mAP、AP50、AP75、AP(S)、AP(M)和 AP(L)分别达到了 58.60%、84.20%、60.80%、37.15%、57.60% 和 67.56%。与原始算法 Swin Transformer 相比，本文算法的 mAP、AP50、AP75、AP(S)、AP(M)和 AP(L)分别提升了 5.15%、9.35%、4.35%、2.51%、2.2% 和 2.91%。

在中目标检测精度方面，YOLOv4 算法检测精

表 1 本文算法与主流算法检测结果比较

Table 1 Comparison of detection results between the proposed algorithm and mainstream algorithms

算法	主干网络	mAP/%	AP50/%	AP75/%	AP(S)/%	AP(M)/%	AP(L)/%	参数量/MB	检测速度/(帧·s^{-1})
Faster R-CNN[5]	ResNet101	51.85	73.75	56.27	33.28	54.84	64.18	243.6	26.8
SSD[7]	VGG16	49.35	66.42	48.35	28.24	51.52	60.80	92.1	52.4
RetinaNet[8]	ResNet50	54.85	77.15	59.16	36.65	59.29	62.21	96.6	46.8
CenterNet[9]	DLA-34	55.65	79.04	59.24	36.12	58.14	66.10	88.8	45.5
YOLOv4[11]	Darknet53	56.85	81.24	62.12	36.55	60.10	66.90	198.5	30.6
DETR[14]	ResNet-50	48.74	68.22	47.20	27.56	49.50	58.85	142	35.7
原始算法[15]	Swin-T	53.45	74.85	56.45	34.64	55.40	64.65	114	42.8
本文	SLFPB-ST	58.60	84.20	60.80	37.15	57.60	67.56	128	42.4

度最高，达到了 60.10%，比本文算法高 2.5%，但 mAP 和 AP(S)相比本文算法低了 1.75% 和 0.6%。One-stage 算法的 RetinaNet 网络和 SSD 网络具备较快的检测速度，本文算法略低，但检测精度远超于其他算法，其中，mAP、AP(S)较 SSD 算法和 RetinaNet 算法高出 9.25%、8.91% 和 3.75%、0.5%。本文算法相比于其他算法，在 mAP、AP(S)和 AP(L)指标上都达到了最高，验证了算法对小目标检测的有效性和准确性。

目前，已有研究提出了较多基于 Transformer 的改进算法，表 2 展示了本文算法与较新算法的对比结果。其他算法多是将现有模块改进与 Swin-Transformer 相结合，而本文针对具体问题提出具体模块，不仅能够更好地捕捉图像中的关键信息，还能与 Swin-Transformer 无缝链接，充分发挥其强大的特征提取能力，同时，引入无锚框检测机制，进一步提升模型性能。从表 2 中可得出，本文算法检测精度最高，检测速度最快，可以满足实时检测的需要。

表 2 不同算法结果比较

Table 2 Comparison of results of different algorithms

算法	mAP/%	mAP提升/%	AP(S)/%	AP(S)提升/%	参数量/MB	参数量降低/MB	检测速度/(帧·s^{-1})	检测速度提升/(帧·s^{-1})
文献[16]	52.23	1.85	33.50	1.64	278	−36	24.6	1.2
文献[17]	56.40	0.70	26.64	0.65	124	14	42.0	4.5
文献[18]	51.65	3.97	29.28	1.75	198	−46	38.8	−9.1
本文	58.60	5.15	37.15	2.51	128	−14	42.4	−0.4

为更全面地评估本文提出的 MSIFN 的效果，图 11 展示了本文算法与原始算法各类别 mAP 值的对比，均有提升。在 IoU 阈值为 0.5 下，以 AP50、F_1 得分、精确率、召回率和检测速度作为评价指标，与原始算法进行比较，结果如表 3 所示。本文算法与原始算法相比，AP50 提升了 9.35%，F_1

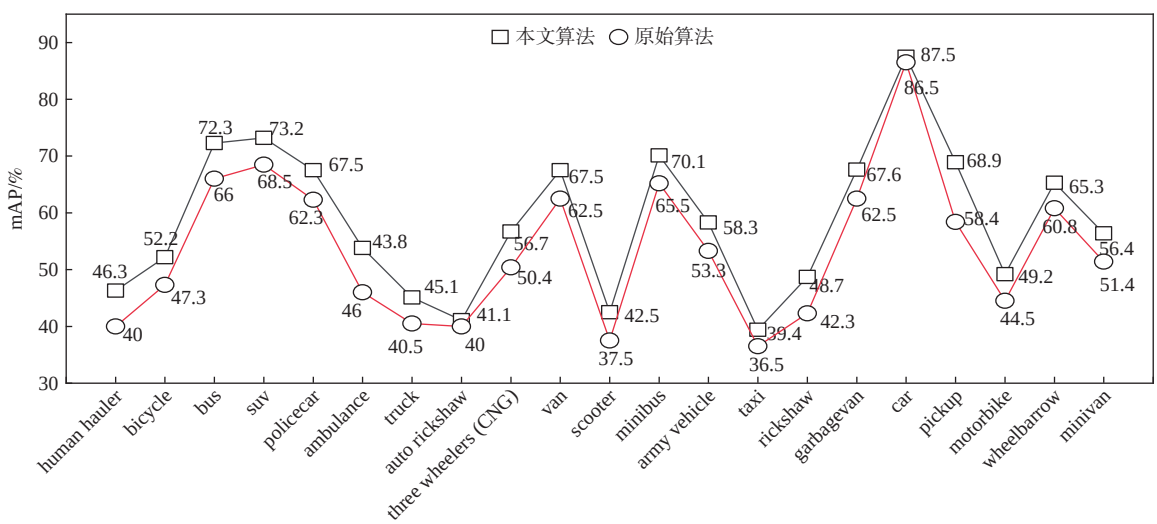


图 11 本文算法与原始算法各类别检测精度对比

Fig. 11 Comparison of detection accuracy of each category between the proposed algorithm and original algorithm

表 3　本文算法与原始算法的性能指标比较

Table 3　Performance comparison between the proposed algorithm and original algorithm

算法	AP50/%	F_1/%	精确率/%	召回率/%	检测速度/(帧·s^{-1})
原始算法	74.85	73.60	83.42	72.65	42.8
本文	84.20	85.16	91.30	79.48	42.4

得分提升了 11.56%, 精确率提升了 7.88%, 召回率提升了 6.83%, 检测速度下降了 0.4 帧/s。本文算法在除了检测速度以外的其他各项指标上均有提升。这是因为 SLFPB 有效扩大了每层特征图的感受野, 增强了信息的捕获能力。此外, 采用 MSIFN 来融合不同尺度的特征图, 并自适应关注贡献较大的特征图。最终融合得到的特征图具有更强大的辨识能力。

2.4　车辆目标预测结果

在 UA-DETRAC 数据集和 Vehicle 数据集上对 YOLOv4 算法、原始算法和本文算法的可视化预测结果进行展示, 如图 12 和图 13 所示。图 12(a) 中存在大量小目标漏检的情况。图 12(b) 中, 采用 Transformer 算法提升了小目标检测性能, 但仍存在漏检现象。图 12(c) 中, 本文算法对小目标定位和分类能力强, 可以捕获许多 YOLOv4 算法和原始算法无法检测到的小目标。图 13 对遮挡目标检测结果进行展示, 得出 YOLOv4 算法和 Swin Transformer 算法对遮挡部分目标检测效果较差, 而本文算法对遮挡部分目标检测效果较好, 可以完全准确定位和检测出遮挡目标。从视觉效果可以看出, 本文算法对小目标和遮挡目标具有强大的检测效果。

为验证本文算法在复杂环境下的检测效果, 将本文算法与原始算法在不同的光照情况下的预测结果进行对比, 结果如图 14 所示。在光照充足的情况下, 2 种算法都能较准确地检测到目标物体, 但本文算法对小目标的细节处理上更出色, 可以检测到更小目标物体。在光照较弱的情况下, 原始算法出现较多漏检情况, 而本文算法则表现出较强的鲁棒性, 依然可以准确地检测到小目标物体。这主要



(a) YOLOv4　(b) 原始算法　(c) 本文算法

图 12　Vehicle 数据集小目标检测结果对比

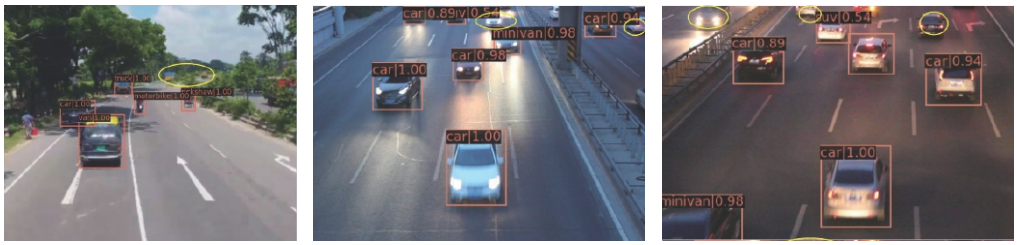
Fig. 12　Comparison of small target detection results in Vehicle dataset



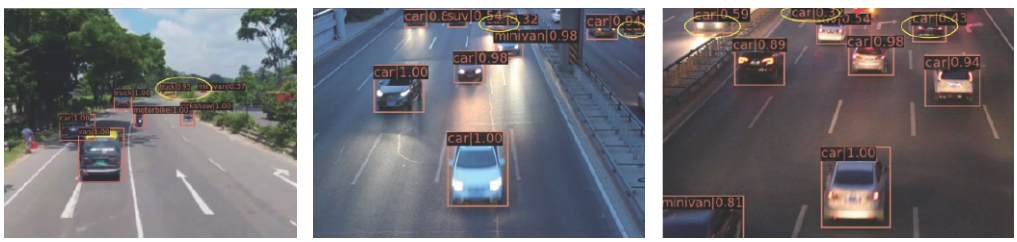
(a) YOLOv4　(b) 原始算法　(c) 本文算法

图 13　UA-DETRAC 数据集遮挡目标检测结果对比

Fig. 13　Comparison of occlusion target detection results for UA-DETRAC dataset



(a) 原始算法



(b) 本文算法

图 14 不同光照情况下检测结果对比

Fig. 14 Comparison of detection results under different light conditions

因为 MSIFN 中加入的 LRB 模块可以更有效地提取特征信息，更好地适应光照变化。该结果验证了本文算法的鲁棒性和有效性，为实际应用提供了有力支持。

2.5 消融实验

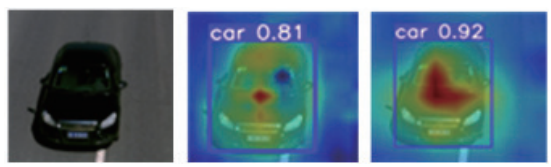
为对本文算法中各模块的有效性进行分析，逐个加入所提模块进行实验，结果如表 4 所示。可以看出，加入 SLFPB 后，mAP 值提高 0.5%，因为 SLFPB 采用分支方式，充分提取了小目标的细节信息，提高了小目标的检测性能。加入 MSIFN 后，mAP 值提高 1.67%，MSIFN 进行自适应多尺度特征融合，可以将深层语义信息和浅层特征进行有效整合，以增强不同层之间的联系，但网络输出部分仍存在大量开销，使检测速度降低。此外，将 SLFPB 和 MSIFN 同时使用，检测速度下降了 1.8 帧/s，但平均检测精度提升了 3.29%。在 MSIFN 网络中加入 LRB 模块，mAP 提升 1.86%，LRB 约束大目标，保留小目标，进一步促使小目标检测精度提升。再搭配无锚框检测机制，减少了小目标的漏检。

表 4 消融实验结果

Table 4 Results of ablation experiments

Swin-T	SLFPB	MSIFN	LRB	检测速度/(帧·s^{-1})	mAP/%
√				35.2	53.45
√	√			38.5	53.95
√		√		36.9	55.12
√	√	√		33.4	56.74
√	√	√	√	42.4	58.60

为更直观地验证加入 SLFPB 对模型特征提取能力的有效性，对输入的图像进行可视化热力图展示，如图 15 所示。可以看出，图 15(c) 的热力图更集中在目标中心处且置信度更高，这是因为 SLFPB 可以更好地捕捉目标的细节特征信息，从而准确定位目标位置。



(a) 原图 (b) 未加入SLFPB (b) 加入SLFPB

图 15 可视化热力图对比

Fig. 15 Comparison of visual thermal maps

3 结 论

本文提出一种感知增强与多尺度特征融合的小目标检测算法来解决图像中携带信息匮乏，特征表达能力弱，导致检测效果不理想的问题。

1）提出一种 SLFPB，扩大感受野，加强空间局部上下文信息的感知能力，并减少了计算复杂度。通过 MSIFN 实现不同特征图的特征信息融合，增强特征的辨识能力。对深层特征图中加入 LRB 来抑制大目标信息，进而提取更多小目标特征信息，有效缓解了小目标漏检问题。

2）考虑到图像中小目标的尺度分布特点，采用 FCOS 算法中无锚框检测头，进一步提升小目标的检测精度。

3）从 mAP、AP(S)、损失函数及场景检测结果等多方面，对本文算法和原始算法训练得到的模型进行对比分析。通过实验对比，mAP 值提升 5.15%，本文算法具有一定可行性。

与此同时，改进后的算法参数量和检测速度仍然不理想，进一步优化参数量和加快算法速度将是后续改进方向。

参考文献（References）

[1] XIAO Z Y, ZHANG G B. An attention-based odometry framework for multisensory unmanned ground vehicles (UGVs)[J]. Drones, 2023, 7(12): 699.

[2] 王坤, 项琦鑫. 改进 Yolov4 的车辆弱目标检测算法[J]. 中国惯性技术学报, 2023, 31(8): 797-805.
WANG K, XIANG Q X. Improved Yolov4 algorithm for vehicle weak object detection[J]. Journal of Chinese Inertial Technology, 2023, 31(8): 797-805(in Chinese).

[3] LIANG H, SEO S. UAV low-altitude remote sensing inspection system using a small target detection network for Helmet wear detection[J]. Remote Sensing, 2023, 15(1): 196.

[4] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2016: 1440-1448.

[5] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[6] CAI Z W, VASCONCELOS N. Cascade R-CNN: high quality object detection and instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5): 1483-1498.

[7] CHEN W P, QIAO Y T, LI Y J. Inception-SSD: an improved single shot detector for vehicle detection[J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(11): 5047-5053.

[8] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2999-3007.

[9] DUAN K W, BAI S, XIE L X, et al. CenterNet: keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2020: 6568-6577.

[10] DING X W, YANG R D. Vehicle and parking space detection based on improved YOLO network model[J]. Journal of Physics: Conference Series, 2019, 1325(1): 012084.

[11] SUMIT S S, WATADA J, ROY A, et al. In object detection deep learning methods, YOLO shows supremum to Mask R-CNN[J]. Journal of Physics: Conference Series, 2020, 1529(4): 042086.

[12] SU X Y, LIU H M, TAO L F, et al. An end-to-end framework for remaining useful life prediction of rolling bearing based on feature pre-extraction mechanism and deep adaptive Transformer model[J]. Computers & Industrial Engineering, 2021, 161: 107531.

[13] KOJIMA T, IWASAWA Y, MATSUO Y. Robustifying vision Transformer without retraining from scratch using attention-based test-time adaptation[J]. New Generation Computing, 2023, 41(1): 5-24.

[14] QI F, CHEN G M, LIU J Y, et al. End-to-end pest detection on an improved deformable DETR with multihead criss cross attention[J]. Ecological Informatics, 2022, 72: 101902.

[15] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: hierarchical vision Transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2022: 9992-10002.

[16] XU X. Research on a small target object detection algorithm for electric transmission lines based on convolutional neural network[J]. IAENG International Journal of Computer Science, 2023, 50(2): 375-380.

[17] FU C Y, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector [EB/OL]. (2017-01-23)[2024-03-01]. https://arxiv.org/abs/1701.06659.

[18] ZHANG Q, ZHANG H Y, LU X W. Adaptive feature fusion for small object detection[J]. Applied Sciences, 2022, 12(22): 11854.

[19] CHEN Y K, ZHANG P Z, KONG T, et al. Scale-aware automatic augmentations for object detection with dynamic training[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2367-2383.

[20] LIU S T, HUANG D, WANG Y H. Learning spatial fusion for single-shot object detection[EB/OL]. (2019-11-25)[2024-03-01]. https://arxiv.org/abs/1911.09516.

[21] ZOPH B, CUBUK E D, GHIASI G, et al. Learning data augmentation strategies for object detection[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2020: 566-583.

[22] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 658-666.

Small target vehicle detection based on perceptual enhancement and multi scale feature fusion

SHEN Yu*，LI Yangyang，LI Bohao，GAO Baoqu，WEI Ziyi，BAI Shan

(School of Electronic and Information Engineering，Lanzhou Jiaotong University，Lanzhou 730070，China)

Abstract：This paper suggests a perception-enhanced and multi-scale fusion-based algorithm for small target vehicle detection in order to address the problem of inadequate information and weak feature expression ability of small targets carried by vehicles, which leads to low detection accuracy and missed detections in current algorithms. Firstly, a spatial local feature enhancement backbone network SLFPB-ST is designed to solve the problem of severe loss of feature information for small targets during the feature extraction process. Secondly, a multi-scale information fusion network (MSIFN) is proposed to fuse features at multiple scales by allocating weights to focus on more detailed information. In order to restrict the characteristics of large objects while maintaining the feature representation of small targets, MSIFN also has a large objects restriction block (LRB). Finally, an anchor-free mechanism is adopted to reduce missed detections of small targets and improve detection accuracy. Experimental results on the UA-DETRAC dataset and Vehicle dataset demonstrate that compared with the Swin Transformer algorithm, our algorithm achieves an improvement of 5.15% in mAP, 9.35% in AP50, and 4.35% in AP75 with an increase in parameter size by 14 MB and a decrease in detection speed by 0.4 frames per second. This validates that our algorithm exhibits good robustness and applicability.

Keywords：small target vehicle detection; Swin Transformer; multi-scale feature fusion; dilated convolution; GIOU

Received：2024-03-05；**Accepted**：2024-06-21；**Published Online**：2024-07-19 10：57

URL：link.cnki.net/urlid/11.2625.V.20240717.1624.002

Foundation items：National Natural Science Foundation of China (61861025,62241106)

* **Corresponding author**. E-mail：18609311366@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0135

基于自适应复合卷积的航拍小目标检测算法

邓天民*，余洋，陈月田，谢鹏飞
(重庆交通大学 交通运输学院，重庆 400074)

摘 要：针对航拍图像中小目标的比例较高且特征提取效果差等问题，提出一种在精度和存储资源消耗上相对平衡的航拍小目标检测算法。提出一种轻量化自适应复合卷积(LACC)模块，加强对细粒度特征的提取能力，并摒弃背景信息达到自适应调节有效特征的输出；基于LACC设计一种多尺度特征融合网络，进一步降低小目标漏检率；使用空间上下文金字塔(SCP)的子分支替代快速空间金字塔池化(SPPF)模块，减少信息混淆与冗余的同时还能适应小目标检测场景；构建一种WiseIou-V3-NMS非极大值抑制算法，考虑检测框具有对象但被删除的情况，使其有效提高网络对遮挡重叠目标的检测定位能力；提出轻量化共享卷积GN检测头，保持对多尺度特征信息敏感的同时减少参数量及模型大小。在VisDrone2019公开数据集上，所提算法的平均精度平均值MAP0.5为0.466，相比于基线算法YOLOv8s提升0.077，网络参数量减少21.6%，模型大小减少18.7%。

关 键 词：航拍图像；小目标检测；YOLOv8；特征融合；细粒度特征
中图分类号：V221+.3; TB553
文献标志码：A **文章编号**：1001-5965(2026)05-1433-12

在近几十年中，计算机视觉领域的发展非常迅猛，其在民用和军用领域中的应用已得到广泛认可。随着计算机视觉算法的日益先进和图像分辨率的不断提高，该领域已经取得了显著成果。然而，当面对无人机航拍图像时，现有网络检测精度表现不足，核心原因是无人机航拍图像背景复杂，目标小且常被遮挡重叠。同时，传统大规模的目标检测算法无法满足无人机平台的计算和存储资源条件的要求。因此，如何兼顾检测精度与存储资源的消耗，是当前无人机航拍图像目标检测亟待解决的难题。

目标检测算法在精度与存储资源消耗上达到更好的平衡是计算机视觉领域的主要研究方向。在增加检测精度方面上，许多学者做出了贡献。例如，Luo等[1]优化YOLOv5网络，从不同尺度大规模获取信息之间的依赖关系，并加入改进的通道注意力机制，抑制复杂背景的干扰，指出空间金字塔池化(spatial pyramid pooling, SPP)[2]对小目标检测的低适用性，改进后增加了检测精度；Zhou等[3]基于欧几里得距离的测量方法对目标进行分类，并通过修改定位损失函数改进小目标定位的性能，使网络更适合小目标检测场景；何宇豪等[4]在网络中添加小目标检测层，并加入卷积块注意力模块(convolutional block attention module, CBAM)[5]，提取小目标丰富的空间信息，增强目标特征表达能力，针对性地使用Soft-NMS算法来替换非极大值抑制(non-maximum suppression, NMS)，实现对密集遮挡目标的精确识别；韩俊等[6]通过增大输入图像分辨率来适应小目标检测场景，但未考虑图像分辨率增大后对硬件要求更苛刻；刘一诺等[7]在特征提取过程中加入注意力模块，抑制背景信息的干扰，在特征融合过程中加入SPD-Conv模块[8]，加强了对细粒度信息的提取，以增强对小目标的检测精度；Lou等[9]提出一种新的下采样方法，更好地保留上下文特征信

收稿日期：2024-03-07；**录用日期**：2024-05-31；**网络出版时间**：2024-06-19 09:04
网络出版地址：link.cnki.net/urlid/11.2625.V.20240618.1106.001
基金项目：国家重点研发计划(2022YFC3800502)；重庆市技术创新与应用发展专项重点项目(2022TIAD-KPX0069)
*通信作者. E-mail：dtianmin@cqjtu.edu.cn

引用格式：邓天民，余洋，陈月田，等. 基于自适应复合卷积的航拍小目标检测算法[J]. 北京航空航天大学学报，2026，52(5)：1433-1444.
DENG T M，YU Y，CHEN Y T，et al. Small object detection algorithm for aerial photography based on adaptive compound convolution[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1433-1444 (in Chinese).

息，增强信息的流动率，通过对浅层及深层信息的有效结合，丰富网络中的目标信息。

为使目标检测模型更好地嵌入终端或移动设备，许多学者在减少存储资源消耗上做出了相应的研究。例如，Du 等[10]从轻量级的角度优化检测头，减少参数量与计算量，通过引入一种上下文增强稀疏卷积层和自适应多层掩蔽方案优化掩码比，实现丰富的前景信息覆盖，提高检测精度；Zhang 等[11]通过自上而下的方法将低分辨率与高分辨率重复融合，提取出全局的语义特征信息，并提出双曲池化模块，以减少特征信息的丢失，利用数据聚类的方法自适应生成目标建议框，保证具有良好的检测精度；薛珊等[12]通过在轻量化的 YOLOv7-tiny[13]模型上进行优化，设计一种新的注意力机制模块来加强对目标的识别精度，增加小目标检测层，丰富目标的空间信息，改进的损失函数提高了模型的收敛速度。

由于 YOLO 系列在检测精度与存储资源上的优异表现，很多研究者致力于对其进行改进。在 YOLOv5 算法被广泛应用于工业后，ULTRALYTICS 公司优化了检测算法，使其在速度、精度及使用便捷性上进一步提升，形成了 YOLOv8 算法。为应对航拍图像中目标小、检测难度大及在硬件条件差的终端部署难等问题，本文基于 YOLOv8 模型，针对小目标及遮挡重叠目标进行优化，提出一种在检测精度与存储资源 2 个方面达到相对平衡的检测算法。

1 基于自适应复合卷积的检测算法

以 YOLOv8 为基准模型，分析其在面临小目标及遮挡重叠目标时，考虑计算存储资源的消耗，从以下 4 个方面构建一种基于自适应复合卷积的航拍小目标检测算法。首先，提出一种轻量化自适应复合卷积 (lightweight adaptive compound convolution, LACC) 模块，利用分组卷积把输入特征以通道维度拆开，提高网络细粒度以适应小目标检测，并以级联的方式引入激活函数，赋予不同通道不同的特征权重，保证特征信息的多样性；在分组卷积的同时，利用平均池化和 1×1 卷积，以更少的参数成本交互目标特征信息，提高网络性能及鲁棒性。其次，基于 LACC 模块，提出了一种基于自适应复合卷积的多尺度特征融合网络，加强小目标特征信息的提取能力。再次，改进骨干结构，使用空间上下文金字塔 (spatial context pyramid, SCP)[14]模块中的子分支替代算力消耗大的快速空间金字塔池化（spatial pyramid pooling-fast, SPPF）模块，加强对局部信息的捕获能力，增加小目标检测精度。然后，构建 WiseIou-V3-NMS 非极大值抑制算法替代原本的 CIou-NMS 算法，采用权值衰减策略解决 NMS 算法带来的漏检问题，有效提高网络对遮挡重叠目标检测定位能力。最后，提出一种轻量化共享卷积 GN 检测头 (lightweight shared convolutional GN detection head, LSCGDH)，使用共享卷积减少参数量，增大感受野，并保持住语义及空间信息的提取能力，为应对不同尺度层级，使用 Scale 层对特征进行缩放，丰富多尺度目标语义信息。改进后的整体检测模型如图 1 所示。

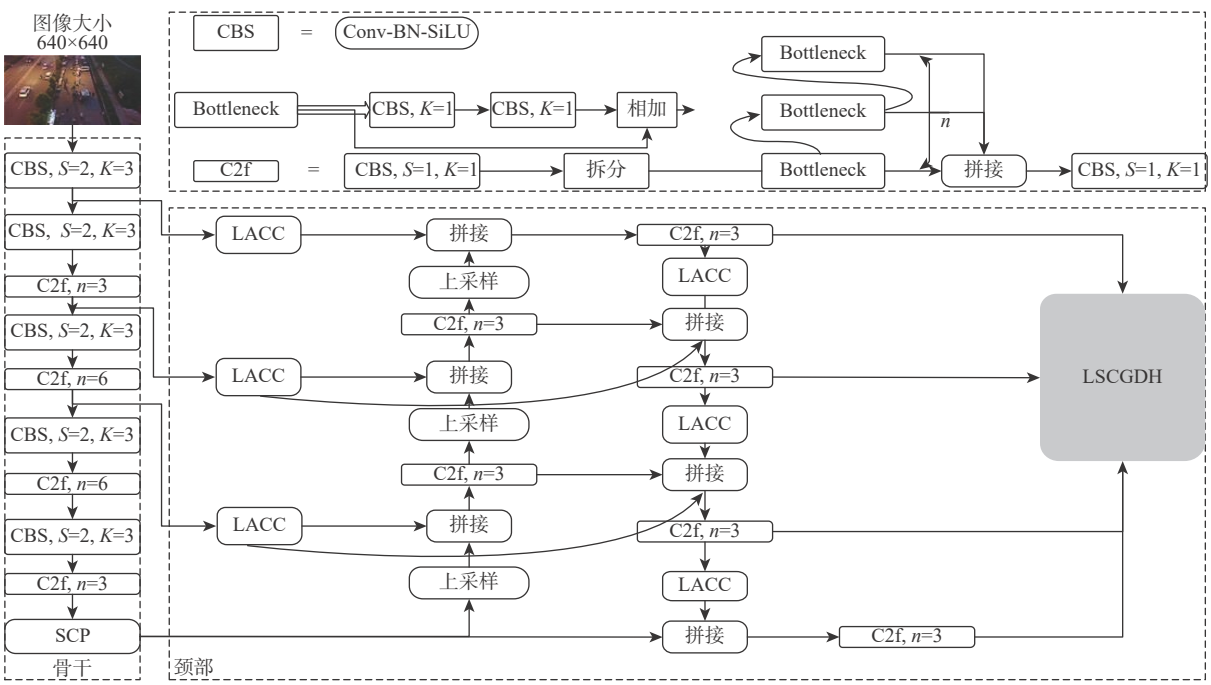


图 1 整体网络结构

Fig. 1 Overall network structure

1.1　多尺度特征融合网络

在 YOLOv8 中, 特征图按降序分为 5 类尺度特征, 分别在骨干、特征金字塔网络 (feature pyramid network, FPN)[15] 及路径聚合网络 (path aggregation network, PAN)[16] 结构中表示为 B1～B5、P3～P4 及 N4～N5。YOLOv8 使用的 PAN-FPN 结构是对传统 FPN 的补充, 又利用 P4-N4 与 P3-N3 的融合来加强目标定位特征的学习, 解决部分目标空间信息丢失的问题。然而, YOLOv8 应用于航拍图像目标检测时, 还存在改进的空间。一方面, 由于缺乏对大尺度特征图的关注, 检测模型可能会忽略一些航拍图像中小目标及遮挡目标的空间信息, 导致检测精度降低; 另一方面, 经过较多次数的卷积及上采样, 原始语义特征信息会丢失一部分, 导致对信息丰富度不足的小目标检测影响大。对于上述难题, 本文提出一种 LACC 模块(见图 2), 并基于该模块设计一种多尺度特征融合网络, 如图 3 所示。

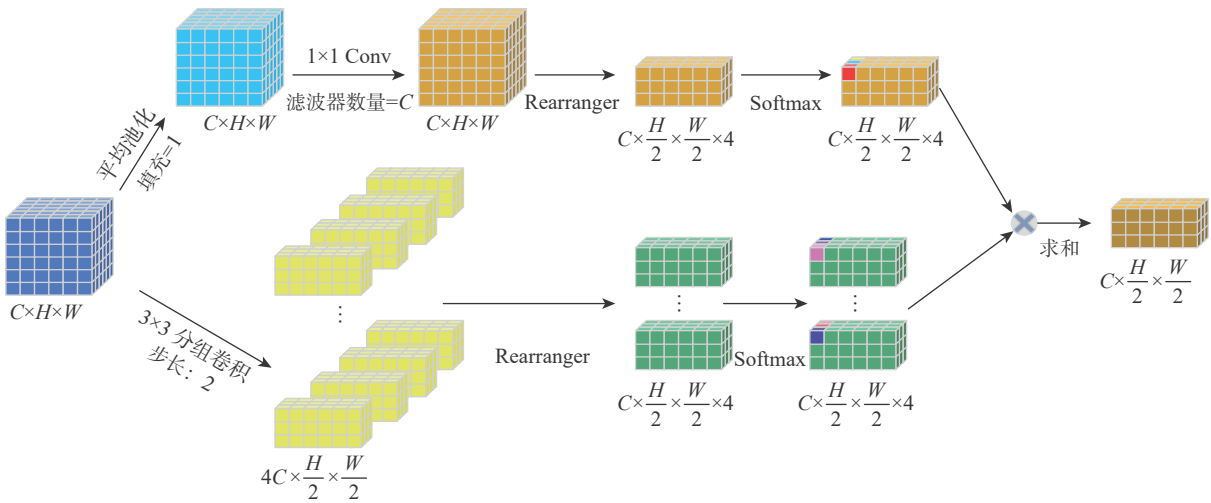


图 2　轻量化自适应复合卷积模块

Fig. 2　Lightweight adaptive compound convolution module

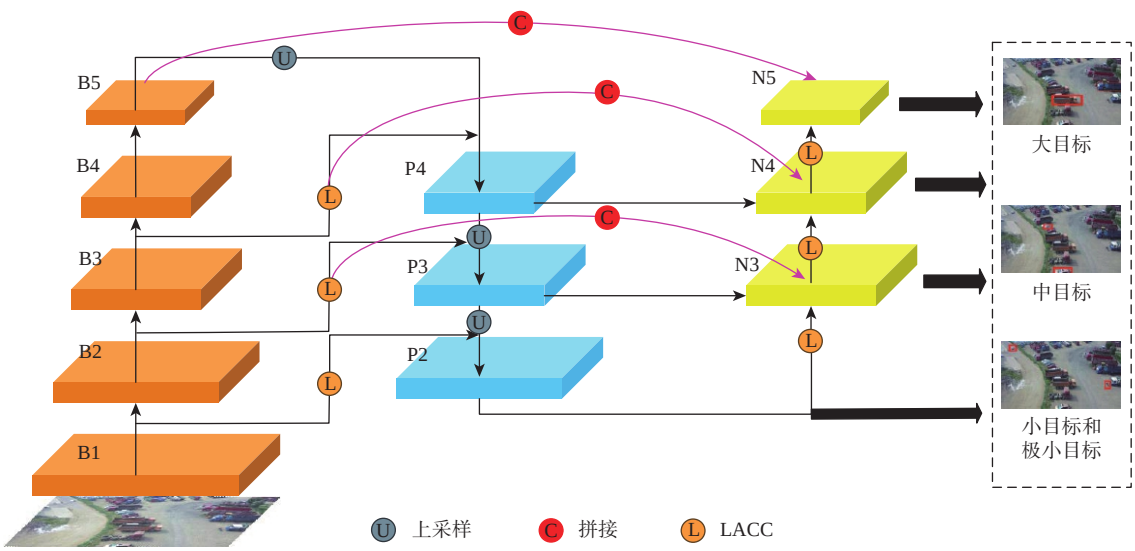


图 3　多尺度特征融合网络

Fig. 3　Multi-scale feature fusion network

在 YOLO 网络中, 对输入图像进行下采样提取特征, 一般采用步长为 2 的 3×3 卷积或步长为 2 的池化 (如最大池化、最小池化、平均池化)。采用标准卷积时, 未考虑图像中不同特征的不同重要性, 降低了提取特征的有效性, 最终对小目标众多的航拍图像检测精度有影响, 此外, 还会增加参数量与存储资源的消耗, 不利于在无人机等小型终端部署。采用池化时, 虽参数量明显减少, 但在很大程度上缺失小目标特征信息, 对航拍图像的检测而言是不利的。因此, 本文提出一种 LACC 模块, 可以解决标准卷积存在的卷积核参数共享问题, 并考虑标准卷积后会对像素信息少的小目标检测不友好[8],

实现对图像中细粒度信息的精确把握，相对于标准卷积降低网络参数量。LACC模块可以认为是一个即插即用模块，该模块可以替代标准卷积来减少参数成本，同时保持住图像细节特征信息，这对要求精度与资源消耗平衡的航拍小目标检测具有很大的优势。

LACC模块参考FOCUS切片操作，将图像语义信息传入通道维度，然而FOCUS结构时间耗费大，所需计算成本高，不适合嵌入无人机，因此，通过分组卷积及Rearrange函数来实现这一目的。利用相应大小的分组卷积动态生成基于接收域大小的特征，分组卷积的分组数取决于每个层级的通道数。计算公式为

$$G = C/N \tag{1}$$

式中：G为分组数；C为每层对应通道数；N取8, 16, 32，N因为通道数的限制最大取32。

使用分组卷积比FOCUS切片操作的速度快，且利用分组卷积提取感受野特征后，会在通道维度上将原始特征映射成新特征。根据最近的相关研究表明[17-19]，交互信息可以提高信息丰富性，对于本来像素信息少的小目标是有效的。然而，与每个感受野特征交互会导致更多的计算开销。因此，为降低交互信息时产生的计算开销与参数量，在另一条分支先使用平均池化来聚合每个接收域特征的全局信息，再使用1×1标准卷积操作提高网络非线性程度和鲁棒性，再次使用Rearrange函数将张量维度升维,并重新变换排序，使具有丰富的语义信息保存到最后一个维度，最终使用Softmax函数来动态强调图像中每个像素的重要性。LACC模块的计算可以表示为

$$\boldsymbol{F} = \sum_{i=1}^{r} \left[\mathrm{Softmax}\left(\mathrm{Conv}_{1\times1}(\mathrm{AvgPool}(\boldsymbol{X}_i))\right) \cdot \mathrm{Softmax}\left(g_{S=2}^{3\times3}(\boldsymbol{X}_i)\right)\right] \tag{2}$$

式中：$\boldsymbol{F}$为LACC模块的输出矩阵；$\mathrm{Conv}_{1\times1}$表示1×1标准卷积；$g_{S=2}^{3\times3}$表示卷积核大小为3×3、步长S为2的分组卷积；$\boldsymbol{X}_i$为输入图中的第i个信息，共有r个像素；Avgpool表示平均池化；Softmax表示激活函数。

根据提出的LACC模块，首先，在骨干的语义信息传入FPN之前增加3个分别与P2、P3、P4同层级的LACC模块，目的是降低网络总体参数的同时，使骨干丰富的原始语义特征尽快传入FPN。其次，为重点关注大尺度特征图，在FPN结构中添加一次上采样，并与骨干结构中的B1层进行特征融合，在此层级(P2)添加一个新的检测头，使得目标空间信息更加丰富，以提高小目标的检测精度。然后，在PAN结构中将原本的标准卷积换成LACC模块，减少网络整体参数量。最后，多尺度特征融合网络的中心思想是提高多尺度特征融合的次数和概率，增加空间信息与语义信息的丰富性，以提升对航拍图像中目标的检测精度。具体方法是：对于具有2个输入路径的特征图，从骨干部分引入第3条输入路径到PAN结构中，前提是第3条输入路径与将要融合的其他2条路径的特征图大小相同。这样的处理方法不会增加额外的参数成本，算法只添加了B3-N3与B4-N4的路径。例如，N4的输出可表达为

$$\boldsymbol{N}_4^{\mathrm{out}} = \mathrm{C2f}(\mathrm{Concat}(\mathrm{LACC}(\boldsymbol{N}_4^{\mathrm{in}}), \mathrm{LACC}_{\boldsymbol{B}_3}^{\mathrm{out}}, \boldsymbol{P}_4^{\mathrm{out}}), n) \tag{3}$$

式中：C2f表示对应的模块操作；$\mathrm{LACC}(\boldsymbol{N}_4^{\mathrm{in}})$表示N4的输入再经过LACC模块；$\mathrm{LACC}_{\boldsymbol{B}_3}^{\mathrm{out}}$表示B3经过LACC模块后的输出；$\boldsymbol{P}_4^{\mathrm{out}}$为P4层级的输出；$n$为C2f的使用次数；Concat表示对输入特征进行拼接。

1.2 优化的骨干网络

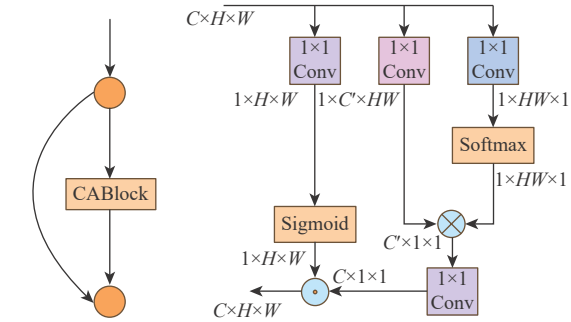
YOLOv8的骨干网络采用标准卷积和C2f模块后，通过SPPF模块使骨干网络最丰富的上层语义传入FPN结构，虽然SPPF模块比SPP模块参数量少，但由于SPPF模块频繁使用最大池化，对小目标的学习率不高，这对于航拍图像目标检测来说是不利的。因此，引用上下文聚合网络[14]中的SCP模块的子分支替换SPPF模块，减少参数量的同时，保持住网络对局部信息的关注度，加强细节信息的流动率。

图像传入骨干网络，通过聚合不同层级的特征图语义信息传入FPN时，仍会有局部细节信息的丢失，这对于目标仅覆盖一小部分区域的航拍图像是不利的。因此，采用SCP模块加强对局部信息的捕获能力。该模块学习从整个特征图中聚合特征，以此加强泛化性能，提高鲁棒性。使用自适应权重将特征组合到每个像素中，保证了只有有效的全局信息被融合到局部区域，同时减少信息混淆。

由于航拍图像中的目标仅覆盖一小部分区域，为不使太多无用的背景信息传入网络，SCP模块还通过在其结构上添加一条额外的路径来学习每个像素的信息，保证特征信息的多样性，作用是减少聚合来自其他空间位置特征信息的必要性，从而减少网络参数量与计算量。SCP模块的结构如图4(a)所示，其由带有一个残差连接的上下文聚合块(context aggregation block, CABlock)组成，CABlock如图4(b)所示。在SCP模块中，像素级空间上下文通过式(4)所示的方法聚合。在CABlock中，使用1×1卷积来执行映射，以交互通道信息，减少网络参数。对比广泛使用自注意力机制[20]的做法，SCP模块在很大程度上减少了参数量与计算量。

$$\begin{cases} \boldsymbol{Q}^j = \boldsymbol{P}^j + \boldsymbol{a}^j \sum_{j=1}^{L} \left[\dfrac{\exp\left(\boldsymbol{w}_k \boldsymbol{P}^j\right)}{\sum_{m=1}^{L} \exp\left(\boldsymbol{w}_k \boldsymbol{P}^m\right)} \boldsymbol{w}_v \boldsymbol{P}^j \right] \\ \boldsymbol{a}^j = [\boldsymbol{E} + \exp(-\boldsymbol{w}_a \boldsymbol{P}^j)]^{(-1)} \end{cases} \tag{4}$$

式中：$\boldsymbol{Q}^j$和$\boldsymbol{P}^j$分别为SCP模块中像素j输出和输



(a) SCP模块子分支结构　　(b) CABlock结构

图 4　SCP 模块子分支结构和 CABlock 结构

Fig. 4　SCP module sub-branch structure and CABlock structure

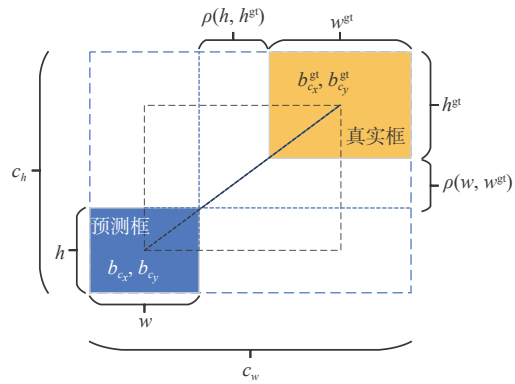


图 5　损失函数参数示意图

Fig. 5　Schematic diagram of loss function parameters

入的特征, 特征图包括 L 个像素; $j,m\in\{1,L\}$为每个像素的索引; $\boldsymbol{w}_k$、$\boldsymbol{w}_v$ 和 $\boldsymbol{w}_a$ 为用于投影特征图的线性变换矩阵; $\boldsymbol{a}^j$ 为与 $\boldsymbol{Q}^j$ 和 $\boldsymbol{P}^j$ 形状相同的加权矩阵, 达到平衡每个像素聚合上下文的程度; $\boldsymbol{P}^m$ 为像素 m 输入的特征; $\boldsymbol{E}$ 为全 1 矩阵; (-1)表示逐元素取倒数。

1.3　WiseIou-V3-NMS 非极大值抑制算法

YOLOv8 算法采用传统的 NMS 算法, 在待检测框与置信度最高的检测框之间的 IoU 大于设定的阈值时, 直接将待检测框的置信度设置为 0, 导致遗漏重叠区域较大的对象框, 使遮挡重叠目标的精度较低。为避免删除包含对象的检测框, 借鉴 Wise-IoU[21] 损失函数思想, 使用 WiseIou-V3-NMS 算法。该算法将 WiseIou-V3 作为 NMS 的准则, 使检测框置信度乘以一个加权函数, 与置信度最高的框的重叠越大, 权重衰减越快, 置信度衰减到设定的阈值时, 舍弃此框, 提高检测定位目标能力的同时, 考虑到检测框具有对象但被删除的情况, 以及 2 个可比较的框具有相同对象的情况, WiseIou-V3-NMS 算法的公式可定义为

$$s_i=\begin{cases}s_i & U-R_{\text{Wise-IoU}}(\boldsymbol{M},\boldsymbol{H}_i)<\varepsilon\\ 0 & U-R_{\text{Wise-IoU}}(\boldsymbol{M},\boldsymbol{H}_i)\geqslant\varepsilon\end{cases}\tag{5}$$

式中: $\boldsymbol{M}$ 为置信度最高的检测框; $\boldsymbol{H}_i$ 为第 i 个待检测框; ε为 NMS 阈值大小; $R_{\text{Wise-IoU}}(\boldsymbol{M},\boldsymbol{H}_i)$为 $R_{\text{Wise-IoU}}$ 关于 $\boldsymbol{M}$ 和 $\boldsymbol{H}_i$ 的值; s_i为不同检测类别的分类得分; U 为第 i 待检测框与 $\boldsymbol{M}$ 的交并比; i 为每个网格对应锚框的个数。

$$R_{\text{Wise-IoU}}=\exp\left[\frac{(b^{\text{gt}}_{c_x}-b_{c_x})^2+(b^{\text{gt}}_{c_y}-b_{c_y})^2}{(c_w^2+c_h^2)}\right]\tag{6}$$

式中: c_h 和 c_w 分别为预测框和真实框的最小闭合框的高度和宽度; b_{c_x}和b_{c_y}为预测框中心点的坐标; $b^{\text{gt}}_{c_x}$和$b^{\text{gt}}_{c_y}$为真实框中心点的坐标。

具体表示情况如图 5 所示。

WiseIou-V3-NMS 算法不仅考虑了方位、质心距离及重叠面积, 还使享有最高梯度增益的边界框的条件处于动态变化中, 且梯度增益随损失值的增加呈非单调变化, 使得对于小目标众多的航拍图像, 可以动态优化小目标的损失权重, 提高检测精度。图 6 为使用 WiseIou-V3-NMS 与 NMS 处理的结果对比。图 6 中存在 10 个遮挡重叠车辆, 且遮挡重叠程度高, 使用 NMS 算法仅检测出 7 个车辆目标, 而使用 WiseIou-V3-NMS 算法成功检测出所有遮挡重叠目标, 且检测精度比 NMS 算法更高。结果表明, WiseIou-V3-NMS 算法在遮挡重叠场景中具有更高的鲁棒性与准确率。



(a) NMS　　(b) WiseIou-V3-NMS

图 6　遮挡重叠目标检测结果对比

Fig. 6　Comparison of results of occlusion overlapping object detection

1.4　轻量化共享卷积 GN 检测头

YOLOv8 检测头较 YOLOv5 有两大改进。①将耦合头换成解耦头结构。解耦头结构含有 2 条并行分支, 分别对目标类别特征及空间特征进行提取, 为提高信息交换程度及网络鲁棒性, 2 条分支分别使用 1×1 卷积, 以提高目标检测精度。但解耦头结构使用较多次数的卷积交换通道信息, 使参数量较耦合头结构有大幅度增长。②将 Anchor-Based 换成 Anchor-Free。Anchor-Based 不需要多尺度遍历滑窗, 对小尺度的目标检测而言较为困难。Anchor-Free 不构建锚框, 直接学习目标位置信息, 通过回归目标的 4 个顶点识别目标, 兼顾大尺度与小尺度的目标信息, 有利于小目标检测。

本文模型为关注多尺度特征，含有 4 个检测头，因此，为降低模型参数量及大小，提出 LSCGDH。LSCGDH 保留了 YOLOv8 原本的解耦头结构，加强对小目标信息的提取，核心思想是：使用共享卷积在分类分支和回归分支共享参数，减少各分支使用标准卷积的次数，达到减少参数量的目的。Tian 等[22]证实使用 GN 归一化可提升检测头定位和分类的能力，因此，LSCGDH 在卷积结构中使用 GN 归一化。

LSCGDH 接收 P2～P4 层级输入的特征后，首先，在 4 个层级分别使用卷积核大小为 1×1 的 GN 归一化标准卷积（Conv-GN），增加通道维度的信息交换，丰富目标信息量，提升检测头分类及回归的性能。其次，使用 2 个卷积核大小为 3×3 的 GN 归一化共享卷积（ShareConv-GN），聚合来自 4 个层级丰富的空间及语义信息，相较于多次使用 1×1 卷积，可抑制冗余信息的流动，且 2 个卷积核大小为 3×3 的共享卷积可提高感受野，增加学习相邻特征信息的概率，同时减少参数量。最后，将共享卷积提取出的信息输入到分类（Cls）和回归头（Reg）中，同时为应对每个检测头所检测目标尺寸不一致的问题，预防小目标特征信息减少，使用 Scale 层对回归检测头中的特征进行缩放，加强对多尺度特征的保留能力。Scale 层含有一个可学习的缩放因子 Z，其与任意尺度的输入相乘，且随着训练的进行，Z 根据梯度下降算法更新，实现对输入特征的缩放。综上，LSCGDH 的目的是利用相对较少的参数去提取有效的目标信息，其总体结构如图 7 所示。

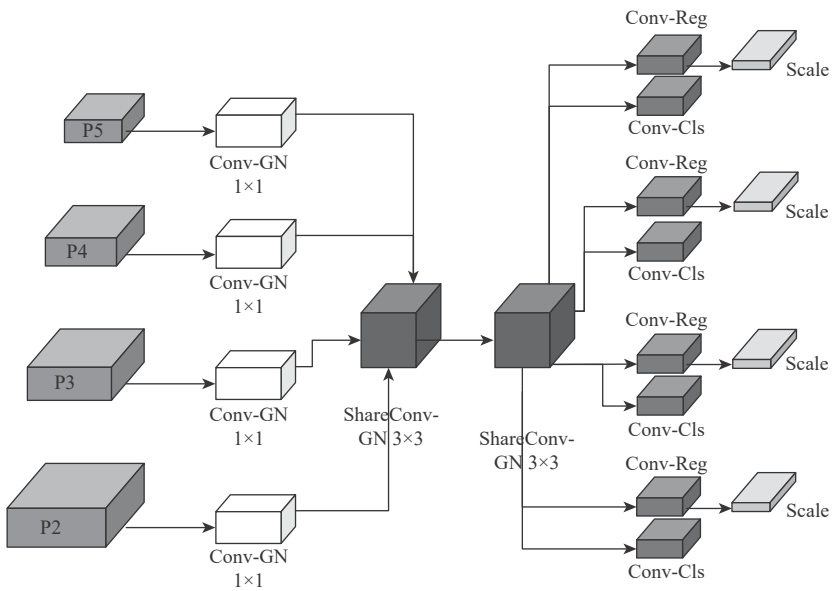


图 7 轻量化共享卷积 GN 检测头

Fig. 7 Lightweight shared convolutional GN detection head

2 实 验

2.1 数据集

本文选用国际上具有权威代表性的数据集 VisDrone2019 作为实验验证对象。VisDrone2019 数据集[17]由天津大学和数据挖掘团队 AISKYEYE 收集并开发。该数据集使用各种无人机进行多场景、多任务拍摄，涵盖 14 个不同地点、2 种不同环境、10 种物体类别、2 个不同环境密度条件、不同天气和照明条件。与传统的检测数据集不同，VisDrone2019 数据集的每张图像可能包含数百个待检测对象，共包含 260 万个边界框的手动注释。该数据集可用于图像、视频的目标检测及单、多目标跟踪，具有很大的挑战意义。为验证该数据集的小目标（小于 32 像素×32 像素）占比及遮挡重叠情况，用脚本遍历数据集中所有标注框，统计后得出数据集中含有 44.7% 的小目标，被遮挡目标比例有 49.97%，符合本文研究的目的。

按照 VisDrone2019 挑战赛的数据集划分方法，将整个数据集划分为训练集（6 471 个样本）、验证集（548 个样本）及测试集（1 610 个样本）。

2.2 实验环境及参数配置

实验环境如下：系统使用 Ubuntu 18.04，CPU 使用 12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50 GHz，GPU 使用 RTX3080（显存为 10 GB），深度学习框架采用 opencv-python 4.6.0.66、cuda 11.7、torch 2.0.1 及 torchvision 0.15.2。为保证模型效果的公平性和可比性，所有消融实验和对比实验涉及的每个模型都未使用任何预训练权重。模型训练过程中的关键参数设置如表 1 所示。

表 1　训练参数设置

Table 1　Training parameter setting

参数	设置
轮数	150
批次大小	4
优化器	SGD
预训练	False
冲量	0.937
权重衰减	5×10^{-4}
马赛克数据增强	1.0
初始学习率	1×10^{-2}
最终学习率	1×10^{-4}

选用更高分辨率的图像可以在一定程度上提高检测精度, 但对计算硬件要求高, 且训练时间过长, 所需算力消耗大, 不利于无人机等小型终端部署, 因此, 本文图像输入分辨率选用 640×640。表 2 为选用不同分辨率的结果对比。可以看出, 选用 640×640 输入图像分辨率在训练时间、运行状态、最高占用显存及批次大小方面都是最优的。

2.3　消融实验

为验证 LACC 模块中不同分组卷积分组数对模型造成的影响, 同时确定 LACC 模块最佳的分组数, 设置了一组消融实验, 以 YOLOv8s 只改进上文中提及的轻量化多尺度特征融合网络为基础, 将式 (1) 中的 N 设置为 8, 16, 32, 其他参数均保持一致进行实验。采用准确率 P、召回率 R、平均精度平均值 MAP0.5 和 MAP0.5:0.95、参数量及模型大小作为评价指标。表 3 展示了验证集不同分组数对应的实验结果。可以看出, N 取 8 时, 参数量与模型大小最小, 但减少的参数对于航拍图像中小目标的检测有效; N 取 32 时, 只有 MAP0.5 比 N 取 16 时高 0.001, MAP0.5:0.95 并未增加, 原因在于增加了冗余的分组数, 导致参数量与模型大小增加。综上, 为既能保证模型的检测精度, 又能兼顾模型参数量及大小, 将 N 确定为 16 进行后续实验。

根据表 3 得出的最佳结果 (N=16) 继续做总体消融实验。本文设计的航拍图像检测模型主要优化 YOLOv8s 的颈部、骨干、非极大值抑制及检测头。通过消融实验验证各个更改部分对网络性能及参数的影响, 即 A(基准模型)、B(轻量化多尺度特征融合网络)、C(骨干网络)、D(非极大值抑制)、E(LSCGDH)。模型在 VisDrone2019 数据集的验证集及测试集上的消融实验结果如表 4 所示。可得, 每种改进都分别起到了上文说明的作用。加入多尺度特征融合网络, 网络细粒度得到提高, 使验证集 MAP0.5 提高了 0.04, 且使模型参数量减少了 0.75×10^6, 模型大小减少了 1 MB。只优化骨干网络, 使局部细节信息传入颈部网络, 验证集 MAP0.5 提高了 0.007, 参数量减少了 0.6×10^6, 模型大小减少了 1 MB, 但当在加入多尺度特征融合网络后继续优化骨干网络, 验证集 MAP0.5 提高了 0.2%, 说明 2 种改进对小目标的语义信息有相似的学习程度。在前面 2 个改进基础上加入构建的非极大值抑制算法, 加强对遮挡重叠目标的识别, 使得验证集 MAP0.5 提高了 0.028, 测试集 MAP0.5 提升了 0.083。只加入 LSCGDH, MAP0.5 提高了 0.002, 参数量减少了 1.67×10^6, 但在前面 3 个改进的基础上加入 LSCGDH, MAP0.5 提高了 0.007, 原因在于 LSCGDH 聚合了 4 个检测头提供的更加丰富的目标空间信息。

2.4　对比实验

在航拍图像检测中, 因其硬件条件比其他常规

表 2　输入图像不同分辨率训练结果

Table 2　Training results of input images with different resolutions

分辨率	批次大小	运行状态	最高占用显存/GB	每轮训练时长/s
640×640	4	Normal	3.26	82～91
1 280×1 280	4	Crash		
1 280×1 280	2	Normal	7.18	234～243
1 536×1 536	4	Crash		
1 536×1 536	2	Normal	8.81	337～350

表 3　LACC 消融实验结果

Table 3　Results of LACC ablation experiment

数据集	N	P	R	MAP0.5	MAP0.5:0.95	参数量	模型大小/MB
验证集	8	0.501	0.419	0.425	0.251	**10.15×10^6**	**20.70**
	16	0.524	0.413	0.429	0.253	10.35×10^6	20.90
	32	0.526	0.417	**0.43**	**0.253**	10.73×10^6	21.90

注: 加粗字体为该列最优值。

表4 总体消融实验结果

Table 4 Results of overall ablation experiment

数据集	模块	*P*	*R*	MAP0.5	MAP0.5:0.95	参数量	模型大小/MB
验证集	A	0.505	0.375	0.389	0.234	11.10×10^6	21.90
	A+B	0.524	0.413	0.429	0.259	10.35×10^6	20.90
	A+C	0.511	0.380	0.396	0.237	10.50×10^6	20.90
	A+D	0.549	0.360	0.420	0.249	11.10×10^6	21.90
	A+E	0.510	0.382	0.391	0.234	9.43×10^6	19.30
	A+B+C	0.523	**0.414**	0.431	0.260	9.69×10^6	19.80
	A+B+C+D	0.561	0.384	0.459	0.301	9.69×10^6	19.80
	A+B+C+D+E	**0.564**	0.396	**0.466**	**0.308**	**8.70×10^6**	**17.80**
数据集	模块	*P*	*R*	MAP0.5	MAP0.5:0.95		
测试集	A	0.435	0.335	0.312	0.177		
	A+B	0.446	0.364	0.343	0.193		
	A+C	0.438	0.334	0.316	0.177		
	A+D	0.489	0.301	0.375	0.258		
	A+E	0.434	0.339	0.314	0.177		
	A+B+C	0.452	**0.365**	0.344	0.199		
	A+B+C+D	0.532	0.320	0.427	0.290		
	A+B+C+D+E	**0.532**	0.300	**0.430**	**0.290**		

注: 加粗字体为该列最优值。

检测更差, 不仅要考虑精度, 还要考虑模型本身的大小。为展示本文算法的优越性, 进行2组对比实验: 一组是与YOLO系列算法进行纵向对比实验, 另一组是与经典模型及近期的先进模型进行横向对比实验。采用平均精度平均值MAP0.5和MAP0.5:0.95、平均精度AP、参数量及模型大小作为评价指标。

本文算法与YOLOv4[23]、YOLOv5s、YOLOv5l、YOLOX-l[24]、YOLOv7-tiny[13]、YOLOv9-C[25]进行对比。对比实验的各项参数按照表1进行, YOLOv5是官方版本6.0, 其余模型均为官方版本。为确保对比实验的有效性, 所有实验都未用预训练权重进行训练。对比实验结果为训练集结果, 如表5所示。

表5 纵向对比实验结果

Table 5 Results of longitudinal comparison experiment

模型	AP										MAP0.5	MAP0.5:0.95	参数量	模型大小/MB
	行人	人	自行车	汽车	面包车	卡车	三轮车	遮阳篷三轮车	巴士	摩托车				
YOLOv4[23]	0.248	0.126	0.086	0.643	0.224	0.227	0.114	0.076	0.443	0.217	0.307	0.166	9.12×10^6	18.20
YOLOv5s	0.390	0.313	0.112	0.735	0.354	0.295	0.205	0.111	0.431	0.37	0.332	0.174	7.03×10^6	18.10
YOLOv5l	0.478	0.377	0.178	0.782	0.426	0.403	0.268	0.131	0.547	0.453	0.404	0.216	46.15×10^6	
YOLOX-l[24]	0.348	0.245	0.169	0.724	0.344	0.405	0.231	0.178	0.531	0.360	0.353	0.195	54.16×10^6	
YOLOv7-tiny[13]	0.379	0.346	0.094	0.761	0.363	0.298	0.201	0.106	0.432	0.418	0.340	0.181	**6.03×10^6**	**12.10**
YOLOv8s	0.415	0.302	0.119	**0.797**	0.458	0.353	0.291	0.140	0.584	0.439	0.389	0.231	11.10×10^6	21.90
YOLOv9-C[25]	0.340	0.184	0.154	0.775	0.452	**0.541**	0.248	0.241	**0.649**	0.383	0.397	0.240	50.90×10^6	98.30
本文	**0.482**	**0.432**	**0.35**	0.772	**0.505**	0.418	**0.372**	**0.242**	0.579	**0.455**	**0.466**	**0.308**	8.70×10^6	17.80

注: 加粗字体为该列最优值。

根据表5, YOLOv4、YOLOv5s、YOLOv7-tiny的参数量少, 利于在无人平台上部署, 但检测精度很低, 不能满足航拍图像的检测要求; YOLOX-l模型大, 且检测精度低; YOLOv5l及YOLOv9-C检测精度比其他YOLO系列高, 但参数量及模型大小很大, 很难满足在无人机等小型终端的部署; YOLOv8s没有达到精度与计算资源最优平衡模型。同时, 从表5中可以看出, 本文算法在行人、人、自行车、摩托车

及遮阳篷三轮车这样的小目标的检测效果最好, 表明网络细粒度对齐效果更好。

通过横向对比实验来进一步验证算法的先进性, 对比实验结果如表 6 所示。可以看出, Cascade R-CNN[26]、Faster R-CNN[27] 及 CDNet[28] 这些经典网络模型的精度不高; DC-YOLOv8[9] 基于 YOLOv8 改进后对目标检测精度不高; Yolov5_GBCS[4]、DA-YOLO[29] 及 FSD-YOLOv5[30] 都是基于 YOLOv5 改进, 但检测精度提升不高; TPH-YOLOv5[31] 精度提升高, 但参数量大, 不适合嵌入无人机等小型终端。综合 2 个对比实验结果, 本文算法在检测精度及存储资源消耗 2 个方面做到了良好的平衡。

表 6　横向对比实验结果

Table 6　Results of horizontal comparison experiment

方法	MAP0.5	MAP0.5:0.95	参数量
Cascade R-CNN[26]	0.319	0.161	
Faster R-CNN[27]	0.225	0.151	
CDNet[28]	0.342		
DC-YOLOv8[9]	0.415	0.247	
Yolov5_GBCS[4]	0.432		**5.97×10^6**
DA-YOLO[29]	0.380		
FSD-YOLOv5[30]	0.363		
TPH-YOLOv5[31]	0.449	0.294	60.42×10^6
本文	**0.466**	**0.308**	8.70×10^6

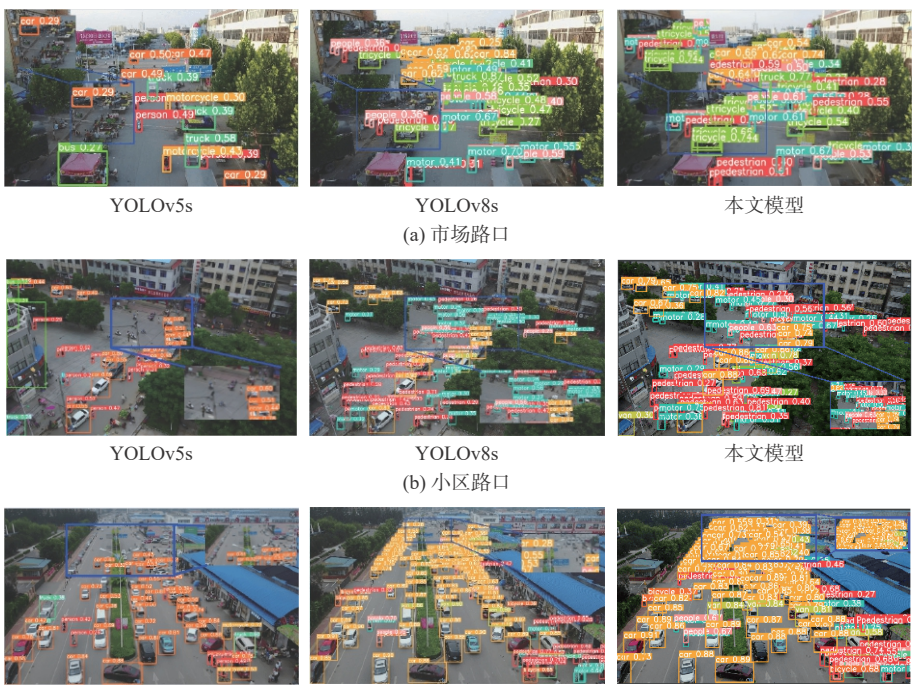
注: 加粗字体为该列最优值, 包含引用文献实验结果。

2.5　可视化分析

为更加直观、清楚地展示本文模型的检测效果, 进行可视化分析, 从模型推理结果和热力图 2 个不同角度来验证模型的检测性能。

使用本文模型、YOLOv8s 和 YOLOv5s 进行推理实验。选择小区路口、市场路口及城市主干路 3 种代表型场景作为分析依据, 这些场景目标小且众多, 相互之间遮挡严重, 适合进行推理实验。3 个模型的检测结果如图 8 所示。结果表明, 与 YOLOv8s 和 YOLOv5s 对比, 本文模型对小目标、遮挡重叠目标及视线远端目标具有最优检测精度, 且对漏检、误检、密集目标提高了检测率, 有效提高了检测性能。

利用 Grad-CAM[32] 为 YOLOv8s 和本文模型生成热力图。热力图能直观看出模型对图像具体区域的关注度。梯度值是通过 Grad-CAM 对模型输出类别的置信度进行反向传播而获得, 颜色越深, 代表关注度越高。实验结果如图 9 所示。从图 9(a)、(b) 中可以看出, YOLOv8s 对小目标与远方目标关注度及敏感度不够; 从图 9(c) 中可以看出, YOLOv8s 对背景噪声抑制作用不太理想。本文模型不仅比 YOLOv8s 对小目标与远方目标更加关注, 且对图像中背景有更好的去噪作用; 模型的关注度更加集中在物体的中心点, 使得预测的边界框更加准确, 从而提高模型的整体检测性能。以上结果都表明本文模型有更好的鲁棒性与有效性。



YOLOv5s　YOLOv8s　本文模型

(a) 市场路口

YOLOv5s　YOLOv8s　本文模型

(b) 小区路口

YOLOv5s　YOLOv8s　本文模型

(c) 城市主干路

图 8　3 种场景推理可视化结果对比

Fig. 8　Comparison of visualization results from three scenario inferences



原图 YOLOv8s 本文模型

(a) 密集场景



原图 YOLOv8s 本文模型

(b) 远景小目标



原图 YOLOv8s 本文模型

(c) 背景噪声

图 9 3 种场景可视热力图结果对比

Fig. 9 Comparison of visual heatmap result from three scenario

3 结 论

本文针对航拍图像中小目标及遮挡重叠目标检测困难的问题，提出一种在精度及存储资源相对平衡的航拍小目标检测算法。

1）提出一种 LACC 模块，关注输入特征中重要的语义特征信息及相邻区域之间的特征信息，再结合设计的多尺度特征融合网络，达到在维持参数成本的同时实现对图像中小目标的精确识别。

2）引用上下文聚合网络中的 SCP 模块子分支来替换 SPPF 模块，减少网络参数量及局部信息的丢失。

3）为针对航拍图像中存在密集遮挡目标的情况，构建 WiseIou-V3-NMS 算法筛选预测框，防止 2 个重叠程度高的检测框被直接删除，从而提高检测精度。

4）提出一种 LSCGDH，目的是降低参数的同时，保持对多尺度目标信息的提取能力。

5）实验结果表明，在检测精度方面，基于自适应复合卷积的航拍小目标检测算法的 MAP0.5 值由 0.389 提升至 0.466；在参数量方面，由 11.10×10^6 减少至 8.7×10^6。本文算法的 MAP0.5 高于 DC-YOLOv8、Yolov5_GBCS、DA-YOLO、FSD-YOLOv5 及 TPH-YOLOv5 算法。

参考文献（References）

[1] LUO X D, WU Y Q, WANG F Y. Target detection method of UAV aerial imagery based on improved YOLOv5[J]. Remote Sensing, 2022, 14(19): 5063.

[2] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[3] ZHOU H L, MA A T, NIU Y F, et al. Small-object detection for UAV-based images using a distance metric method[J]. Drones, 2022, 6(10): 308.

[4] 何宇豪, 易明发, 周先存, 等. 基于改进的 Yolov5 的无人机图像小目标检测[J]. 智能系统学报, 2024, 19(3): 635-645.
HE Y H, YI M F, ZHOU X C, et al. Small target detection in UAV image based on improved Yolov5[J]. CAAI Transactions on Intelligent Systems, 2024, 19(3): 635-645(in Chinese).

[5] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 3-19.

[6] 韩俊, 袁小平, 王准, 等. 基于 YOLOv5s 的无人机密集小目标检测算法[J]. 浙江大学学报 (工学版), 2023, 57(6): 1224-1233.
HAN J, YUAN X P, WANG Z, et al. UAV dense small target detection algorithm based on YOLOv5s[J]. Journal of Zhejiang University (Engineering Science), 2023, 57(6): 1224-1233(in Chinese).

[7] 刘一诺, 张琪, 王蓉, 等. 针对航拍小目标检测的 YOLOv7 改进方法[J]. 北京航空航天大学学报, 2025, 51(7): 2506-2512.
LIU Y N, ZHANG Q, WANG R, et al. Improved YOLOv7 method for aerial small target detection in aerial photography[J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(7): 2506-2512(in Chinese).

[8] SUNKARA R, LUO T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects[C]//Proceedings of the Joint European Conference on Ma-

chine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2023: 443-459.

[9] LOU H T, DUAN X H, GUO J M, et al. DC-YOLOv8: small-size object detection algorithm based on camera sensor[J]. Electronics, 2023, 12(10): 2323.

[10] DU B W, HUANG Y C, CHEN J X, et al. Adaptive sparse convolutional networks with global context enhancement for faster object detection on drone images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 13435-13444.

[11] ZHANG X X, WANG C Y, JIN J, et al. Object detection of VisDrone by stronger feature extraction Faster R-CNN[J]. Journal of Electronic Imaging, 2023, 32(1): 013018.

[12] 薛珊, 安宏宇, 吕琼莹, 等. 复杂背景下基于 YOLOv7-tiny 的图像目标检测算法[J]. 红外与激光工程, 2024, 53(1): 20230472.
XUE S, AN H Y, LV Q Y, et al. Image target detection algorithm based on YOLOv7-tiny in complex background[J]. Infrared and Laser Engineering, 2024, 53(1): 20230472(in Chinese).

[13] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 7464-7475.

[14] LIU Y, LI H F, HU C, et al. Learning to aggregate multi-scale context for instance segmentation in remote sensing images[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(1): 595-609.

[15] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 936-944.

[16] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8759-8768.

[17] ZHU P F, WEN L Y, DU D W, et al. Detection and tracking meet drones challenge[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7380-7399.

[18] BELLO I, ZOPH B, LE Q, et al. Attention augmented convolutional networks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2020: 3285-3294.

[19] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck Transformers for visual recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 16514-16524.

[20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[21] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[EB/OL]. (2023-04-08)[2024-03-01]. https://arxiv.org/abs/2301.10051.

[22] TIAN Z, SHEN C H, CHEN H, et al. FCOS: a simple and strong anchor-free object detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 1922-1933.

[23] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2024-03-01]. https://arxiv.org/abs/2004.10934.

[24] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-08-06)[2024-03-01]. https://arxiv.org/abs/2107.08430.

[25] WANG C Y, YEH I H, LIAO H M. YOLOv9: learning what you want to learn using programmable gradient information[EB/OL]. (2024-02-09)[2024-03-01]. https://arxiv.org/abs/2402.13616.

[26] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6154-6162.

[27] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[28] HUANG X B, LI Q F, SHEN L L, et al. CDNet: cross-frequency dual-branch network for face anti-spoofing[C]//Proceedings of the International Joint Conference on Neural Networks. Piscataway: IEEE Press, 2023: 1-9.

[29] 李校林, 刘大东, 刘鑫满, 等. 改进 YOLOv5 的无人机航拍图像目标检测算法[J]. 计算机工程与应用, 2024, 60(11): 204-214.
LI X L, LIU D D, LIU X M, et al. Improved target detection algorithm for UAV aerial image based on YOLOv5[J]. Computer Engineering and Applications, 2024, 60(11): 204-214(in Chinese).

[30] 郭业才, 孙京东, AMITAVE S. 基于 YOLOv5 改进的航拍图像目标检测算法[J]. 系统仿真学报, 2025, 37(2): 551-562.
GUO Y C, SUN J D, AMITAVE S. Improved target detection algorithm for aerial images based on YOLOv5[J]. Journal of System Simulation, 2025, 37(2): 551-562(in Chinese).

[31] ZHU X K, LYU S C, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on Transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Piscataway: IEEE Press, 2021: 2778-2788.

[32] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

Small object detection algorithm for aerial photography based on adaptive compound convolution

DENG Tianmin*, YU Yang, CHEN Yuetian, XIE Pengfei

(College of Traffic and Transportation, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: To remedy the problem of aerial imagery, including the high proportion of small objects and the poor feature extraction effect, a small object detection algorithm with relatively balanced accuracy and storage resource consumption is proposed. To improve the capacity to extract fine-grained features and produce adaptable features without background information, a lightweight adaptive compound convolutional (LACC) module is first suggested. Then, a multi-scale feature fusion network based on LACC is designed to further reduce the missing rate of small targets. Secondly, the sub-branches of the spatial context pyramid (SCP) are used to replace the spatial pyramid pooling-fast (SPPF) module, which can reduce information confusion and redundancy, and adapt to small target detection scenarios. In order for the network to successfully increase the detection and positioning capabilities of occlusion-overlapping targets, a WiseIou-V3-NMS non-maximum suppression method is then built, taking into account that the detection frame contains objects but is erased. Finally, a lightweight shared convolutional GN detection head is proposed to keep the sensitivity to multi-scale feature information while reducing the number of parameters and computation. On the VisDrone2019 public dataset, the proposed algorithm achieves a mean accuracy MAP0.5 of 0.466, which is 0.077 higher than the baseline algorithm YOLOv8s, the number of network parameters is reduced by 21.6%, and the model size is reduced by 18.7%.

Keywords: aerial images; small target detection; YOLOv8; feature fusion; fine-grained features

Received: 2024-03-07; **Accepted**: 2024-05-31; **Published Online**: 2024-06-19 09:04

URL: link.cnki.net/urlid/11.2625.V.20240618.1106.001

Foundation items: National Key Research and Development Program of China (2022YFC3800502); Special Key Project for Technological Innovation and Application Development in Chongqing (2022TIAD-KPX0069)

* **Corresponding author**. E-mail: dtianmin@cqjtu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0190

基于特征信息补充与增强的无人机视角图像目标检测算法

邬开俊，蒲卓*
(兰州交通大学 电子与信息工程学院，兰州 730000)

摘 要：针对无人机（UAV）视角图像中目标尺度变化大、小尺寸目标占比高且背景噪声干扰严重等问题，提出一种基于特征信息增强与补充的无人机视角图像目标检测算法。为利用高层语义信息捕捉更加丰富的多尺度信息，提出一种多元融合空间金字塔池化（MFSPPF）方法；设计多分支语义增强（MBSE）模块，可以通过多个分支提取丰富的多尺度特征并构建多尺度特征之间的联系，从而在特征融合传递信息时防止重要特征信息丢失；提出细节特征补充（DFC）模块，将低层特征信息提取细化后得到丰富细粒度特征信息，经过特征融合传递实现对高层特征中细节信息的补充。通过在VisDrone2021数据集上进行实验，结果表明：所提算法相较于基线算法YOLOv8m，平均精度（AP）、AP50、AP75、AP(s)、AP(m)、AP(l)分别提高3.7%、5.4%、3.9%、3.1%、4.0%、7.9%。并且所提方法在YOLOv8其他模型中同样适用。与其他算法相比，所提算法在不同交并比（IOU）阈值标准下和不同尺寸目标检测下都具有优异的检测效果，同时保持较快的检测速度，能够适用于无人机视角图像检测任务。

关 键 词：目标检测；无人机图像；YOLOv8；特征增强；细粒度特征
中图分类号：V279; TP391.41
文献标志码：A **文章编号**：1001-5965(2026)05-1445-11

目标检测是计算机视觉中的一项基础且重要的任务。近年来随着深度学习技术的进步，基于深度学习的目标检测取得了令人瞩目的突破。而无人机具有灵活性高和成本低廉等特点，现今在智慧交通、城市管理、智慧农林、灾害巡检等领域得到广泛的应用[1]。通过将目标检测技术应用到无人机视角图像，可以充分发挥无人机平台的优势，进一步拓展目标检测技术和无人机应用领域。但是通用检测器在检测MS COCO数据集[2]和PASCAL VOC数据集[3]等自然场景中的数据集时表现良好，应用到VisDrone2021数据集[4]等无人机视角图像数据集时检测效果不理想。这是由于相较于自然场景图像，无人机视角图像主要存在3个问题。首先，无人机作为空中平台高度变化较大，并且同时包含远近距离不同的物体，造成无人机视角中的物体尺度变化较大，小尺寸目标占比高，物体之间存在遮挡；其次，无人机应用场景的不同带来了光照影响、雨雾遮挡、目标角度变化大等问题；最后，无人机视角图像视野开阔，图像中覆盖了较多目标，但也覆盖了较多的背景物体。因此，如何提升无人机视角图像检测精度是一个值得研究的、具有挑战性的问题。

为提升无人机视角图像检测效果，加强无人机视角下目标检测的应用，本文充分利用了YOLOv8m[5]

收稿日期：2024-04-01；**录用日期**：2024-05-20；**网络出版时间**：2024-09-05 11:39
网络出版地址：link.cnki.net/urlid/11.2625.V.20240904.1623.002
基金项目：甘肃省自然科学基金(23JRRA913)；内蒙古自治区重点研发与成果转化计划项目(2023YFSH0043)；兰州交通大学重点研发项目资助(ZDYF2304)
* **通信作者**. E-mail：shiyuepz@163.com

引用格式：邬开俊，蒲卓. 基于特征信息补充与增强的无人机视角图像目标检测算法[J]. 北京航空航天大学学报，2026，52(5)：1445-1455.
WU K J，PU Z. Object detection algorithm for UAV viewpoint images based on feature information complementation and enhancement[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1445-1455 (in Chinese).

模型检测精度高、计算复杂度较低等特点，针对无人机视角下图像的特点和存在的问题，深入挖掘YOLOv8m模型的潜力，并在其基础上进行了改进和优化，补充和增强模型中目标特征信息，提高模型检测性能。首先，通过改进原网络中快速空间金字塔池化（spatial pyramid pooling-fast, SPPF）模块，对骨干网络提取特征进一步融合增强，提升了特征图的表征能力，使得网络能够更加准确捕捉目标的特征；其次，为加强特征融合部分特征传递能力，防止重要特征丢失，提供一个多分支结构并将其加入到特征融合部分。这一结构的引入使得网络能够充分利用各个分支的特征信息，增强其检测性能。对于高层特征图细节信息缺失的问题，使用浅层特征图获取鲁棒的细节信息，将其加入到特征融合自下而上传递部分，增强了高层特征图中的细节信息，以此增强网络对于目标的检测能力。本文算法在各种复杂的场景下都能够更准确地识别定位目标。概括起来，主要创新如下：

1）提出一种多元融合快速空间金字塔池化（multivariate fusion spatial pyramid pooling-fast, MFSPPF）模块替换YOLOv8m原始网络中SPPF模块。该模块使用最大池化操作和平均池化操作双分支取代原始网络SPPF模块中的最大池化操作，充分利用高层丰富语义信息生成高质量的多尺度空间特征，以有效提升无人机视角图像检测效果。

2）构建一种多分支语义增强（multi-branch semantic enhancement, MBSE）模块。该模块通过多个感受野不同的分支获取物体多尺度特征信息，并通过共享参数的通道注意力单元关联4个分支提取的特征信息，在特征融合时充分挖掘物体多尺度信息并防止特征传递时重要信息丢失，提高模型检测性能。

3）考虑到高层特征图中蕴含丰富的语义信息但是丢失了细粒度特征，提出一种细节特征补充（detailed feature complement, DFC）模块，将低层特征图中的细节信息经过提取后添加到特征融合自下而上传递路径中，补充高层特征图中的细粒度信息，以提高检测精度。

1 相关研究

1.1 自然场景图像目标检测

随着深度学习技术的发展，自然场景图像目标检测算法取得了令人瞩目的成果。这些算法主要可以分为两类：基于卷积神经网络（convolutional neural networks, CNN）的目标检测算法和基于Transformer的目标检测算法。基于CNN的目标检测算法使用CNN进行特征提取。其中，单阶段检测算法采用了回归分析的思想，通过神经网络模型直接预测输入图像中的目标位置坐标和类别概率，相关算法有RetinaNet算法[6]、单次多框目标检测（single shot multiBox detector, SSD）算法[7]、完全卷积的单阶段目标检测（fully convolutional one-stage object detection, FCOS）算法[8]、以及YOLO系列算法如YOLO[9]、YOLOX[10]、YOLOv7[11]等。这些算法简单、高效、实时性高，并且计算复杂度也较低。而两阶段检测算法首先通过区域建议网络等方法生成候选框，然后，对生成的候选框进行分类和位置回归。代表性的算法有快速区域卷积神经网络（faster region-based convolutional neural networks, Faster RCNN）算法[12]、Mask RCNN算法[13]、Cascade RCNN算法[14]、Double-Head RCNN算法[15]等。两阶段检测算法通常精度较高，特别是在目标尺寸较小或目标密集下具有优势。但是相比于单阶段检测器，速度较慢并且复杂度高。基于Transformer的目标检测算法在近些年取得很好的检测效果，其中，DETR（detection transformer）[16]通过将对象任务检测任务转化为直接集合预测问题来优化训练过程，利用匈牙利算法匹配预测框和真实框，不需要Anchor生成和非极大值抑制（non-maximum suppression, NMS）后处理。随后，其他基于DETR的改进算法不断提出，具体有Deformable DETR[17]、DN-DETR（denoising detection transformer）[18]、RT-DETR（real-time detection transformer）[19]等，但是这类算法对计算资源需求较高，并且小目标检测效果较差。

1.2 无人机视角图像目标检测

由于无人机视角图像高空拍摄带来的背景复杂等问题，研究人员根据无人机视角图像的特性进行了大量研究，当前研究重点围绕数据增强策略、骨干网络优化、多尺度特征融合、聚焦目标区域、轻量化等方法。Kiefer等[20]利用虚拟合成数据集探索合成数据在现实世界无人机目标检测应用潜力。Chen等[21]通过基于生成对抗网络（generative adversarial networks, GAN）的数据增强方法来检测无人机图像中的目标。利用数据增强方法可以丰富数据集的多样性，但是会使得模型训练成本增加，而且面对不同的任务场景需要重新设计。王殿伟等[22]在Double-Head RCNN骨干网络中加入Transformer及可变形卷积模块提高小目标特征提取能力，设计基于内容感知特征重组的特征金字塔网络结构减少背景噪声干扰，改善了无人机图像中小目标检测效果。冒国韬等[23]通过在骨干网络中加入多尺度分割注意力单元强化关键特征表达能力，同时设计自

适应加权特征融合方法在复杂背景情况下提高无人机图像检测能力。这类算法通过骨干网络优化增强目标特征信息提取, 利用多尺度特征融合充分挖掘不同尺寸目标特征, 以此提升检测精度。但是会增加计算复杂度和参数量, 影响模型的实际部署和应用。Li 等[24] 针对高分辨航拍图像中存在目标尺度变化、分别不均匀问题设计了一种密度图引导的目标检测网络(density-map guided object detection networks, DMNet), 在航拍数据集上取得优异的性能。Leng 等[25] 提出一个反向注意力探索模块聚焦具有检测困难区域, 并设计一个区域特定上下文学习模块增强对困难区域信息挖掘, 有效提高检测困难物体的检测性能。Deng 等[26] 提出一种全局-局部自适应网络, 实现图像局部区域更高分辨率检测, 提供更细粒度的特征, 以实现更精确的检测。聚焦图像局部区域的方法可以明显改善无人机视角图像检测性能, 然而模型较为复杂, 涉及较多的超参数设置, 阻碍了模型性能的进一步提升。Chen 等[27] 为满足资源受限设备的需求, 提出一种轻量级无人机视觉目标检测网络。Yu 等[28] 提出了一种新的实时对象检测器, 可以在移动设备的物体检测方面取得优异的性能表现。模型轻量化可以使算法部署到有限算力的设备上, 丰富使用场景, 但仅能用于简单场景, 在检测精度上还有较大差距。以上工作从不同的角度出发探索更好的无人机图像目标检测模型, 而本文从多尺度特征提取和细粒度特征补充方面出发, 基于 YOLOv8m 提出改进算法, 以有效检测无人机视角图像目标。

2　本文算法

YOLOv8 算法主要由输入端、骨干网络、特征融合网络、检测网络 4 部分组成。骨干网络对输入图像进行逐级特征提取, 生成从 $\boldsymbol{P}_1$～$\boldsymbol{P}_5$ 的多尺度特征表示。特征融合网络对不同层级的特征进行有效整合与增强。检测网络基于融合后的特征完成目标的分类与定位, 输出最终的检测结果。

本文算法基于 YOLOv8m 算法改进, 整体架构如图 1 所示。该算法使用 MFSPPF 模块替换原始 YOLOv8m 网络中的 SPPF 模块提取更丰富多尺度信息。在特征融合路径中加入 MBSE 模块, 充分挖掘多尺度特征信息并保留重要特征信息。DFC 模块将低层特征图的细粒度特征信息提取后经过特征融合自下而上传递, 以补充高层特征图中细粒度信息缺失的情况。

2.1　多元融合空间金字塔池化模块

高级特征图包含了丰富的语义信息, 有助于目标准确定位及分类。然而原网络中的 SPPF 模块采用多个最大池化操作合并的方式聚合高层特征图

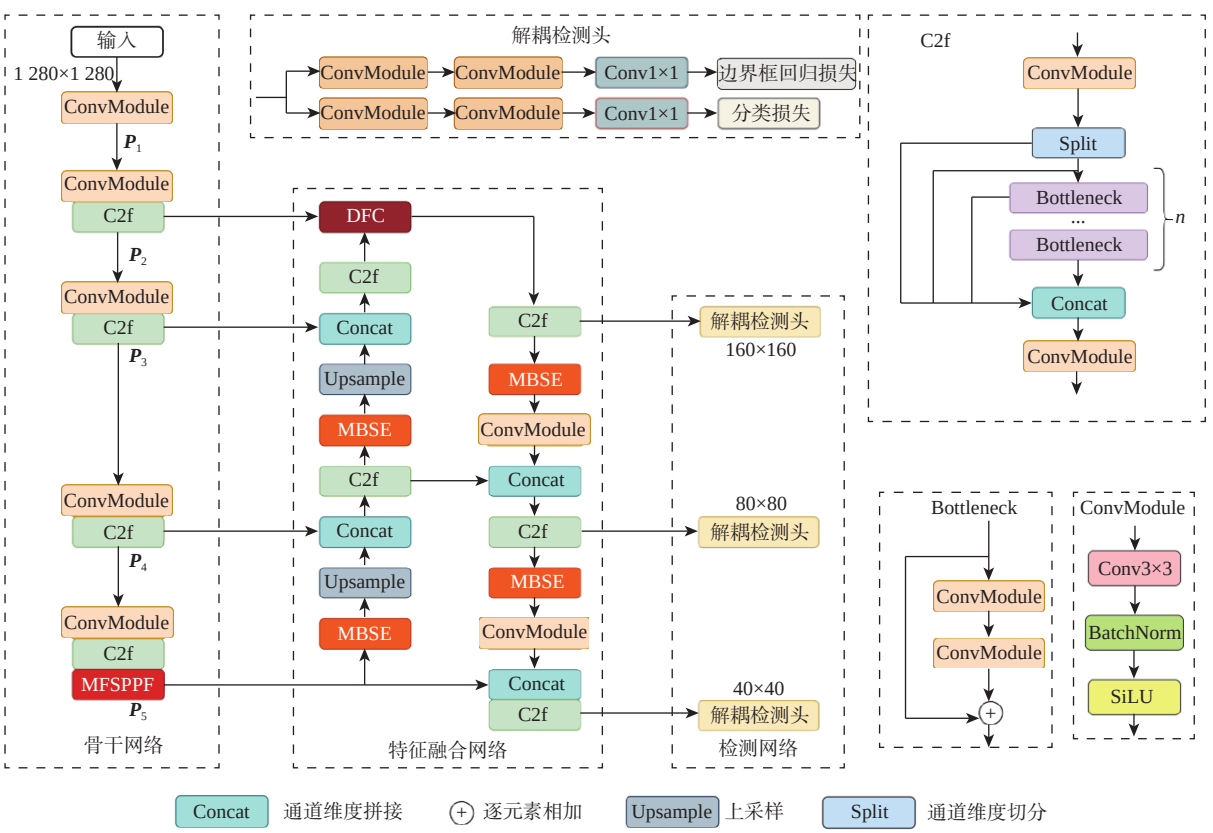


图 1　本文算法架构

Fig. 1　Architecture of the proposed algorithm

的空间信息，这种方式一定程度上聚合了多尺度特征，但实际上没有充分利用高层特征图中丰富的、深层次的特征信息，导致部分特征信息的丢失，并不能生成高质量的特征表示。而且特征融合部分需要将 SPPF 模块提取的特征信息向后续层传递进行特征融合，改进 SPPF 模块在高层特征图中捕捉并保留更加丰富的语义信息将有助于提升检测效果。

MFSPPF 模块在原有 SPPF 模块的基础上进行改进，整体结构如图 2 所示。最大池化操作可以有效提取目标显著特征，通过选取特征图中每个区域的最大值，高效的捕捉目标的边缘、纹理等重要细节信息。而平均池化操作通过计算特征图每个区域的平均值，能够获取更广泛的全局上下文信息。MFSPPF 模块同时使用最大池化操作和平均池化操作，构建双池化特征融合结构（dual-pooling feature fusion structure, DPFFS），以替代原始 SPPF 模块中单一最大池化操作。这种方式可以更加全面地挖掘目标区域显著特征，同时综合考虑全局上下文信息，有助于减少无人机视角图像复杂背景噪声信息对检测的影响，有效区分目标和背景，提高模型的鲁棒性。

具体来说，输入特征张量 $\boldsymbol{X}\in\mathbf{R}^{C\times H\times W}$，先对通道维度进行降维得到 $\boldsymbol{X}'\in\mathbf{R}^{C/2\times H\times W}$，即

$$\boldsymbol{X}'=\sigma(\theta(\mathrm{Conv}_{1\times1}(\boldsymbol{X}))) \tag{1}$$

式中：$\sigma(\cdot)$为 Sigmoid 激活函数；$\theta(\cdot)$表示 BatchNorm 操作；$\mathrm{Conv}_{1\times1}(\cdot)$表示1×1卷积操作。

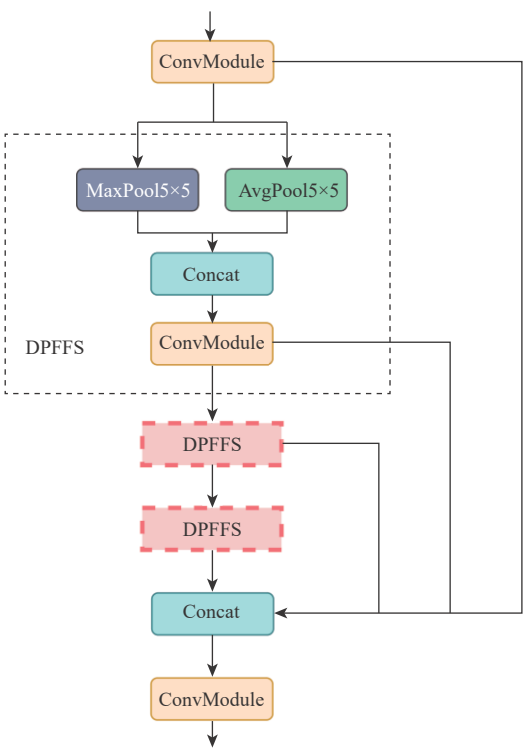


图 2 多元融合快速空间金字塔池化模块结构

Fig. 2 Multivariate fusion spatial pyramid pooling-fast module structure

原始 SPPF 模块使用 3 个连续核大小为5×5最大池化操作对转换后的特征张量处理，本文将其中的最大池化操作替换为双池化特征融合结构，即将特征张量沿着通道维度使用平均池化操作和最大池化操作来聚合特征图的信息得到 2 个不同的空间注意力图 $\boldsymbol{X}_{\mathrm{avg}}\in\mathbf{R}^{C/2\times H\times W}$ 和 $\boldsymbol{X}_{\mathrm{max}}\in\mathbf{R}^{C/2\times H\times W}$，并且在通道维度连接，特征张量形状为$C\times H\times W$，对通道进行降维形状变为$C/2\times H\times W$。

与 SPPF 模块类似，首先，将降维得到的 $\boldsymbol{X}'$通过双池化特征融合结构得到特征张量 $\boldsymbol{X}_1\in\mathbf{R}^{C/2\times H\times W}$，数学描述为

$$\boldsymbol{X}_1=\sigma(\theta(\mathrm{Conv}_{1\times1}([\mathrm{AvgPool}_{5\times5}(\boldsymbol{X}');\mathrm{MaxPool}_{5\times5}(\boldsymbol{X}')]))) \tag{2}$$

式中：$\mathrm{AvgPool}_{5\times5}(\cdot)$和$\mathrm{MaxPool}_{5\times5}(\cdot)$分别表示核大小为5×5的平均池化操作和最大池化操作；[⋯;⋯] 表示向量拼接。

然后，在$\boldsymbol{X}_1$的基础上利用双池化特征融合结构得到$\boldsymbol{X}_2$，即

$$\boldsymbol{X}_2=\sigma(\theta(\mathrm{Conv}_{1\times1}([\mathrm{AvgPool}_{5\times5}(\boldsymbol{X}_1);\mathrm{MaxPool}_{5\times5}(\boldsymbol{X}_1)]))) \tag{3}$$

再次，在$\boldsymbol{X}_2$基础上使用双池化特征融合结构得到$\boldsymbol{X}_3$，即

$$\boldsymbol{X}_3=\sigma(\theta(\mathrm{Conv}_{1\times1}([\mathrm{AvgPool}_{5\times5}(\boldsymbol{X}_2);\mathrm{MaxPool}_{5\times5}(\boldsymbol{X}_2)]))) \tag{4}$$

最后，将这些特征张量在通道维度进行合并，合并后通道维度为$2C$，再将通道维度进行降维得到输出特征张量 $\boldsymbol{X}_{\mathrm{out}}\in\mathbf{R}^{C\times H\times W}$，数学描述为

$$\boldsymbol{X}_{\mathrm{out}}=\sigma(\theta(\mathrm{Conv}_{1\times1}([\boldsymbol{X}';\boldsymbol{X}_1;\boldsymbol{X}_2;\boldsymbol{X}_3]))) \tag{5}$$

2.2 多分支语义增强模块

无人机图像中目标尺度变化大、背景复杂，需要网络能够充分挖掘目标不同尺度的空间特征信息。而原网络中特征融合部分在自上而下、自下而上层层传递特征信息时部分重要信息丢失，无法充分传播至下一层。为此，本文设计了 MBSE 模块，主要结构如图 3 所示。该模块通过使用多个并行分支提取多尺度特征信息，每个分支关注特定尺度特征。利用共享参数通道注意力单元构建不同尺度特征信息之间的联系，学习到通用特征表示。将所有分支特征信息在通道维度合并，通过卷积操作整合特征信息，以有效的利用多个分支提取的信息，产生具有丰富语义的多尺度特征表示。将该模块加入到特征融合网络部分中可以防止特征传递时重要信息丢失。

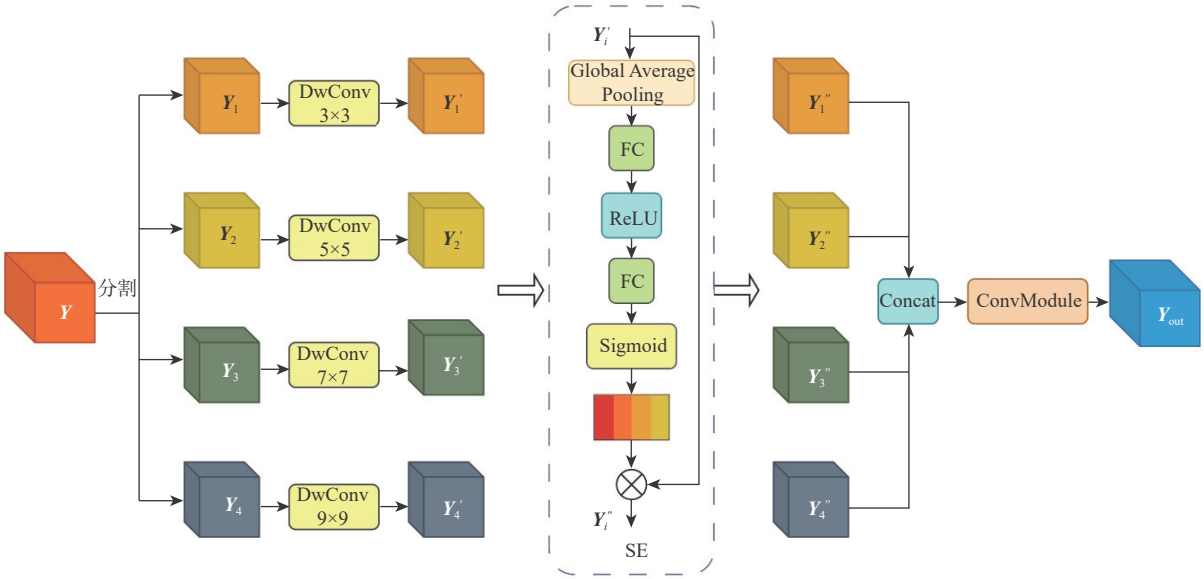


图 3　多分支语义增强模块结构

Fig. 3　Multi-branch semantic enhancement module structure

具体来说, 首先, 将输入的特征张量$\boldsymbol{Y}\in\mathbf{R}^{C\times H\times W}$在通道维度上切分为 4 部分: $\boldsymbol{Y}_1\in\mathbf{R}^{C/4\times H\times W}$、$\boldsymbol{Y}_2\in\mathbf{R}^{C/4\times H\times W}$、$\boldsymbol{Y}_3\in\mathbf{R}^{C/4\times H\times W}$、$\boldsymbol{Y}_4\in\mathbf{R}^{C/4\times H\times W}$。

将切分的特征张量分别通过卷积核大小 (3×3、5×5、7×7、9×9) 不同深度卷积的并行分支可以提取不同感受野下的目标多尺度特征信息。深度卷积相比于普通卷积, 其可以在降低计算成本的同时增强特征表示能力。该过程可以表示为

$$\boldsymbol{Y}_i'=\mathrm{DwConv}_{k_i\times k_i}(\boldsymbol{Y}_i)\quad k_i=3,5,7,9;\ i=1,2,3,4 \tag{6}$$

式中: $\mathrm{DwConv}_{k_i\times k_i}(\cdot)$表示不同卷积核大小$k_i$的深度卷积; i为不同分支。

对提取到的特征$\boldsymbol{Y}_i'$分别通过共享参数的通道注意力单元建立通道依赖关系。通道注意力单元采用的是挤压激励模块 (squeeze-and-excitation, SE)[29], 其可以分为 2 个步骤: 挤压和激励。

步骤 1　输入特征张量$\boldsymbol{Z}\in\mathbf{R}^{C\times H\times W}$, 通道挤压步骤可以收集全局上下文信息, 通常使用全局平均池化实现, 得到全局空间特征$\boldsymbol{Z}'\in\mathbf{R}^{C\times 1\times 1}$, 数学描述为

$$\boldsymbol{Z}'=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{Z}(i,j) \tag{7}$$

步骤 2　通过激励步骤捕获通道间的依赖关系。该步骤首先使用降维系数为r的全连接层(fully connected, FC)对特征张量$\boldsymbol{Z}'$进行降维操作得到维度为$C/r\times 1\times 1$的特征张量; 然后, 使用 ReLU 激活函数处理, 经过全连接层恢复原始特征张量的维度; 最后, 利用 Sigmoid 激活函数归一化得到输出特征张量$\boldsymbol{Z}_{\mathrm{out}}\in\mathbf{R}^{C\times 1\times 1}$。以上过程数学表示为

$$\boldsymbol{Z}_{\mathrm{out}}=\sigma(\mathrm{FC}_2(\mathrm{SiLU}(\mathrm{FC}_1(\boldsymbol{Z}')))) \tag{8}$$

式中: $\mathrm{FC}_1(\cdot)$和$\mathrm{FC}_2(\cdot)$分别表示系数r不同的全连接层; $\mathrm{SiLU}(\cdot)$表示 SiLU 激活函数。

使用$\mathrm{SE}(\cdot)$表示挤压激励模块操作, 将并行分支使用共享参数的通道注意力单元处理, 并将得到的全局空间特征与相应并行分支元素相乘, 该部分可以表示为

$$\boldsymbol{Y}_i''=\boldsymbol{Y}_i'\otimes\mathrm{SE}(\boldsymbol{Y}_i')\qquad i=1,2,3,4 \tag{9}$$

式中: $\otimes$表示元素相乘操作。

在通道维度合并并行分支, 并使用5×5卷积模块进行处理转换, 得到输出特征张量$\boldsymbol{Y}_{\mathrm{out}}\in\mathbf{R}^{C\times H\times W}$, 数学表示为

$$\boldsymbol{Y}_{\mathrm{out}}=\sigma(\theta(\mathrm{Conv}_{5\times 5}([\boldsymbol{Y}_1'';\boldsymbol{Y}_2'';\boldsymbol{Y}_3'';\boldsymbol{Y}_4'']))) \tag{10}$$

式中: $\mathrm{Conv}_{5\times 5}(\cdot)$表示卷积核大小为5×5的卷积操作。

2.3　细节特征补充模块

随着网络深度的增加, 卷积神经网络中高层特征图分辨率降低、感受野增大。高层特征图中蕴含了丰富的语义信息, 但是丢失了大量的细粒度特征。相比之下, 浅层的特征图则蕴含了丰富的细节信息。通过提取低层局部细节信息对高层特征图细粒度特征信息进行补充, 可以进一步增强高层特征图语义表征能力, 提高无人机视角下目标检测性能。因此, 本文提出细节特征补充模块, 具体结构如图 4 所示。该模块将特征$\boldsymbol{P}_2$通过膨胀率不同的膨胀卷积并行分支处理后合并, 使用卷积操作减小分辨率和通道维度。然后, 结合特征融合自上而下传播得到的$\boldsymbol{P}_3$特征, 生成富含细节信息的特征图。将该特征图与$\boldsymbol{P}_3$融合, 增强$\boldsymbol{P}_3$特征中的目标细节信息, 促使特征融合自下而上传递信息时将细

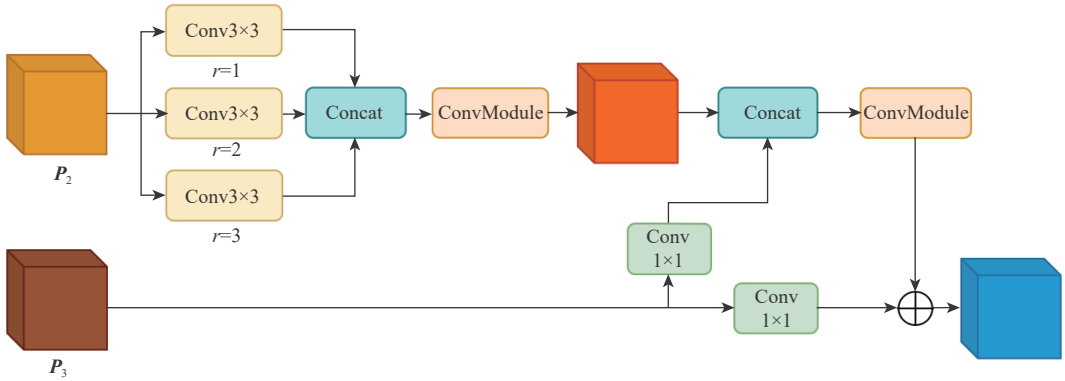


图4 细节特征补充模块结构

Fig. 4 Detailed features complement module structure

节信息向更高层特征图传递，以此增强无人机视角图像目标检测性能。

具体说来，输入高层特征图即网络中特征 $\boldsymbol{P}_3 \in \mathbf{R}^{C_h \times H \times W}$，低层特征图即特征 $\boldsymbol{P}_2 \in \mathbf{R}^{C_l \times 2H \times 2W}$。虽然低层特征包含丰富的细节信息，但是由于卷积次数不足，低层特征中也存在大量的背景噪声信息。因此，本文将特征张量 $\boldsymbol{P}_2$ 通过扩张率分别为1、2、3和核大小3×3的3个膨胀卷积并行分支，然后，将这些分支在通道维度合并通道维度为 $3C_l$，使用3×3卷积模块对通道维度进行转换并下采样，以此获得更加精确的特征信息 $\boldsymbol{P}'_2 \in \mathbf{R}^{C_h \times H \times W}$。以上过程数学表述为

$$\boldsymbol{P}'_2 = \sigma(\theta(\mathrm{Conv}_{3\times3}([\mathrm{Dilat}_{d_1}(\boldsymbol{P}_2); \mathrm{Dilat}_{d_2}(\boldsymbol{P}_2); \mathrm{Dilat}_{d_3}(\boldsymbol{P}2)]))) \quad d_i = 1,2,3; i = 1,2,3 \tag{11}$$

式中：$\mathrm{Dilat}_d(\cdot)$ 表示扩张率为 d 的膨胀卷积。

经过特征自上而下传播后，为降低低层特征图背景噪声的干扰，将细节特征信息 $\boldsymbol{P}'_2$ 通过3×3卷积转换，特征图 $\boldsymbol{P}_3$ 通过1×1卷积模块进行转换后在通道维度合并并卷积转换得到 $\boldsymbol{P}''_2 \in \mathbf{R}^{C_h \times H \times W}$，数学描述为

$$\boldsymbol{P}''_2 = \sigma(\theta(\mathrm{Conv}_{3\times3}([\mathrm{Conv}_{3\times3}(\boldsymbol{P}'_2); \mathrm{Conv}_{1\times1}(\boldsymbol{P}_3)]))) \tag{12}$$

将得到的细节特征 $\boldsymbol{P}''_2$ 与卷积后的 $\boldsymbol{P}_3$ 特征元素相加，以此得到经过细节信息增强的 $\boldsymbol{P}_3$ 特征，并在后续特征融合自下而上传递路径将细节信息传递到高层特征，增强目标检测性能。该部分数学表示为

$$\boldsymbol{P}'_3 = \mathrm{Conv}_{1\times1}(\boldsymbol{P}_3) \odot \boldsymbol{P}''_2 \tag{13}$$

式中：⊙表示逐元素相加操作。

3 实 验

3.1 数据集与评价指标

本文使用VisDrone2021数据集[4]作为模型训练及性能测试检测基准。VisDrone2021数据集是用于无人机图像目标检测的数据集，共包含10个类别，分别为“pedestrian”、“people”、“bicycle”、“car”、“van”、“truck”、“tricycle”、“awning-tricycle”、“bus”、“motor”。该数据集共有10 209张图像，并被分为3部分，分别是包含6 471张图像的训练集VisDrone2021-DEV-train、包含548张图像的验证集VisDrone2021-DEV-val和包含1 610张图像的测试集VisDrone2021-DEV-test。由于该数据集收集的图像拍摄角度变化大，目标尺寸变化大，目标类别之间语义差别较小，并且尺寸小于32像素的小目标众多，使得在无人机图像目标检测任务领域中具有挑战性。本文在训练集上进行模型训练，使用验证集进行消融实验，并使用测试集对不同模型之间的性能进行公平的比较。

为有效的评估本文算法，本文在实验中使用COCO数据集的检测评估指标：平均精度（average precision, AP）、交并比（intersection over union, IOU）阈值大于0.5时测试的平均精度AP50、IOU阈值大于0.75时的测试平均精度AP75、物体像素尺寸小于32像素×32像素时的测试平均精度AP(s)、物体像素尺寸大于32像素×32像素小于96像素×96像素之间时的测试平均精度AP(m)、物体像素尺寸大于96像素×96像素时的测试平均精度AP(l)。此外，通过单张图像推理时间评估算法的计算效率，所得出的推理时间由预处理时间、模型推理时间和后处理时间3部分组成。

3.2 实施细节

本文使用NVIDIA RTX3090 GPU进行训练和测试，在Pytorch 1.12.1框架构建模型，基线算法使用源码或者开源目标检测工具箱MMDetection中的源码。模型共训练500个epoch，其中，包括10个用于学习率预热的epoch。图像输入像素大小调整为1 280像素×1 280像素。模型训练到最后10个时期时关闭数据增强策略。训练批量大小设置为4，使用Adam优化器进行训练，初始学习率设

置为 0.001, 其他训练设置与基线模型设置保持一致。

3.3　消融实验

为证明本文提出的各个模块的有效性, 在 VisDrone2021-DET-val 数据集上进行了消融实验。表 1 为以 YOLOv8m[5] 为基线模型, 在其中加入不同模块时的检测结果, 最优结果加粗显示。

消融实验的结果可以表明, 当使用 MFSPPF 模块替换 SPPF 模块后, 算法的参数量增加 0.49×10^6, 整体检测性能得到显著提升, AP、AP50、AP75、AP(s)、AP(m)、AP(l) 分别提升了了 3.4%、5.0%、3.3%、2.8%、3.2%、4.8%。其中, AP50 提升的效果最显著, 而且对于不同尺度目标检测效果检测都有明显改善, 特别是对于大尺寸目标的检测效果。相较其他 2 个模块的作用, 该模块在小尺寸目标的检测效果提升最大。实验表明, MFSPPF 模块相较于 SPPF 模块能够只付出较少参数量的代价, 捕获更加精确的多尺度目标特征, 使得模型在检测不同尺寸目标时更加有效, 提升整体检测精度。

表 1　VisDrone 验证集上的消融实验结果

Table 1　Results of ablation experiments on VisDrone validation set

算法				AP/%	AP50/%	AP75/%	AP(s)/%	AP(m)/%	AP(l)/%	参数量
基线	MFSPPF	DFC	MBSE							
√				33.7	53.5	35.7	25.5	46.0	42.5	**25.85×10^6**
√	√			37.1	58.5	39.0	28.3	49.2	47.3	26.34×10^6
√		√		36.7	58.1	38.7	28.0	49.0	48.0	26.81×10^6
√			√	37.0	58.4	39.1	27.9	49.6	**52.3**	32.71×10^6
√	√	√	√	**37.4**	**58.9**	**39.6**	**28.6**	**50.0**	50.4	35.47×10^6

当在基线模型中加入 MBSE 模块时参数量增加 6.86×10^6, AP、AP50、AP75、AP(s)、AP(m)、AP(l) 与基线模型相比分别提升了 3.3%、4.9%、3.4%、2.4%、3.6%、9.8%, 整体检测性能提升, 特别是对于大目标的检测效果 AP(l) 达到了 9.8%, 提升最为明显。相较于其他模块, 对中大尺寸目标检测效果提升最好。实验表明, MBSE 模块可以有效提取具有丰富语义的多尺度特征信息, 促使特征融合传递特征信息时保留重要特征信息, 特别是大目标的特征信息。

在基线模型中加入 DFC 模块后, 与基线模型相比参数量增加 0.96×10^6, AP、AP50、AP75 分别提升了 3.0%、4.6%、3.0%, 在不同阈值要求下检测的准确性都得到了提升, DFC 模块对整体的检测有着积极的影响。AP(s)、AP(m)、AP(l) 分别提升了 2.5%、3.0%、5.5%, 说明细节特征补充模块对不同尺寸目标的检测效果均有改善, 特别是大尺寸目标的检测效果。

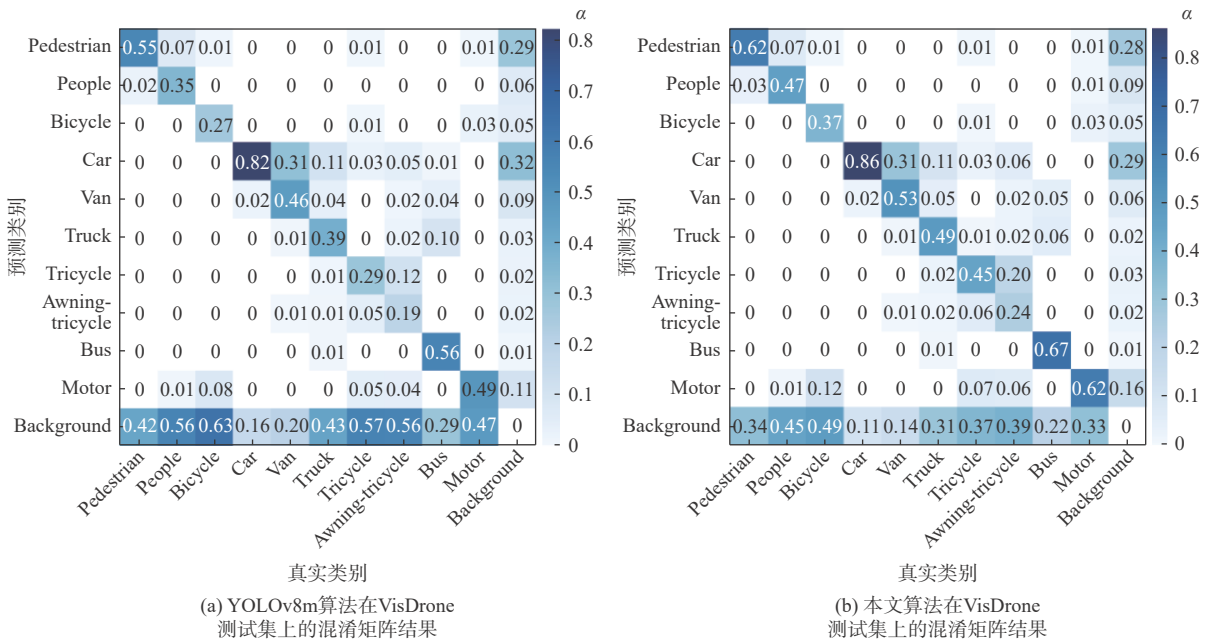
最终, 本文算法与基线模型相比参数量虽然增加 9.62×10^6, 但是 AP、AP50、AP75、AP(s)、AP(m)、AP(l) 分别提升了 3.7%、5.4%、3.9%、3.1%、4.0%、7.9%。整体的检测性能提升显著, 不同尺度目标检测效果也有明显提升, 大尺寸目标检测效果提升最多。相较于单个模块, 在 AP、AP50、AP75、AP(s)、AP(m) 这 5 个评价指标方面达到最优, 而 AP(l) 相较于 MBSE 模块的效果下降了 1.9%, 猜测是由于多个特征增强的方法使得特征信息之间相互影响, 降低了 MBSE 模块对于大尺寸目标检测的效果, 后续需要进一步研究多个特征增强模块的特征融合效果。

此外, 通过比较基线与本文算法的混淆矩阵如图 5 所示, 发现本文算法较基线模型具有更加出色的目标定位能力和分类能力, 能够更好地挖掘目标特征信息, 从复杂无人机图像中识别并分类目标, 真阳率大幅提高, 特别是对于“bicycle”、“tricycle”、“awning-tricycle”等无人机视角下较难识别的目标也有着明显改善, 其中, α 表示预测类别在真实类别中的比例。实验表明本文算法可以提取更加精确的目标多尺度特征, 减少误检漏检情况, 可以有效应用于无人机视角图像目标检测。

为验证本文算法的适用性和泛化能力, 在 YOLOv8n 和 YOLOv8s 中加入本文提出的模块, 结果如表 2 所示, 其中, 同类对比算法最优结果加粗显示。在加入本文提出模块后, YOLOv8n 相较于原模型 AP、AP50、AP75、AP(s)、AP(m)、AP(l) 分别提升 2.5%、3.3%、2.9%、3.0%、2.5%、1.8%, 模型参数量增加 1.93×10^6。改进后 YOLOv8s 相比于原模型 AP、AP50、AP75、AP(s)、AP(m)、AP(l) 分别提升 1.2%、1.7%、1.1%、1.4%、1.4%、2.3%, 模型参数量增加 5.49×10^6。2 种模型在改进后性能提升非常显著。因此, 这些结果表明本文算法在 YOLOv8 其他系列模型中具有良好的泛化性和可靠性。

3.4　对比实验

为验证本文算法的有效性及评估其检测性能, 将本文算法与 8 种目标检测算法 (RetinaNet[6]、Faster-RCNN[12]、YOLOv5[30]、YOLOX[10]、任务对齐的一阶段物体检测器 (task-aligned one-stage object



(a) YOLOv8m算法在VisDrone测试集上的混淆矩阵结果　(b) 本文算法在VisDrone测试集上的混淆矩阵结果

图 5　基线算法与本文算法在 VisDrone 验证集上的混淆矩阵结果

Fig. 5　Confusion matrix results for baseline algorithm and the proposed algorithm on VisDrone validation set

表 2　YOLOv8 其他算法改进后检测性能对比

Table 2　Comparison of detection performance of improved YOLOv8 other algorithms

算法	AP/%	AP50/%	AP75/%	AP(s)/%	AP(m)/%	AP(l)/%	参数量
Yolov8n[5]	28.6	46.8	29.5	19.7	40.1	45.1	**3.01×10^6**
Yolov8n+MFSPPF+DFC+MBSE	**31.1**	**50.1**	**32.4**	**22.7**	**42.6**	**46.9**	4.40×10^6
Yolov8s[5]	34.0	54.5	35.3	24.9	46.0	48.1	**11.14×10^6**
Yolov8s+MFSPPF+DFC+MBSE	**35.2**	**56.2**	**36.4**	**26.3**	**47.4**	**50.4**	16.63×10^6

detection, TOOD)[31]、VFNet[32]、YOLOv8[5]、Gold YOLOm[33])在 VisDrone2021-DET-test 数据集上进行比较，分别考察了 AP、AP50、AP75、AP(s)、AP(m)、AP(l) 和单张图像所需要的推理时间共 7 个指标，实验结果如表 3 所示，最优结果加粗显示。其中，AP 结果为 28.7%，较次优算法结果高 2.8%。在 IOU 大于等于 0.5 的标准下，AP50 结果为 47.4%，较次优算法结果提高 3.5%，相较于其他算法有着极大的优势。在 IOU 大于等于 0.75 标准下，AP75 达到 29.8%，相较于次优算法提高 2.7%，表明本文算法在更加严格的 IOU 阈值下能够取得较其他方法更加准确的检测结果。并且，实验结果还显示出对小尺寸和中等尺寸目标的检测效果，AP(s)、AP(m) 相较于次优结果分别高 1.6%、3.6%，避免了由于目标尺寸较小而导致的误检和漏检。AP(l) 相较于次优结果略有提升，增加了 0.3%，相较其他算法显著提升。在推理时间方面，本文算法为 28.6 ms，相较于最快的 YOLOv8m 算法还有较大的提升空间。综

表 3　VisDrone 测试集上的对比试验结果

Table 3　Comparative test results on the VisDrone test set

算法	骨干网络	AP/%	AP50/%	AP75/%	AP(s)/%	AP(m)/%	AP(l)/%	推理时间/ms
RetinaNet[6]	ResNet-50	8.0	15.5	7.6	2.1	13.2	23.7	22.7
Faster RCNN[12]	ResNet-50	12.8	23.9	12.6	5.2	21.1	29.7	33.4
Yolov5m[30]	CSPDarkNet	21.3	37.7	21.7	13.2	31.2	33.4	21.1
YoloXm[10]	CSPDarkNet	19.7	36.5	19.1	13.0	28.0	23.9	28.1
TOOD[31]	ResNet-50	22.9	38.2	23.8	13.9	34.3	43.1	43.8
VFNet[32]	ResNet-50	17.6	30.3	17.9	10.0	26.7	36.4	43.4
YOLOv8m[5]	CSPDarkNet	25.8	42.4	27.1	16.8	37.4	39.3	**18.4**
Gold YOLOm[33]	CSPDarkNet	25.9	43.9	26.5	15.6	37.1	48.7	28.2
本文算法	CSPDarkNet	**28.7**	**47.4**	**29.8**	**18.4**	**41.0**	**49.0**	28.6

上所述, 本文通过对比实验清楚的表明, 本文算法在无人机视角图像上具有显著优势, 不仅提供了较快的推理时间, 而且在不同 IOU 阈值标准和不同尺寸物体方面实现优异的检测精度。

为更加充分验证本文算法在无人机视角图像检测方面的优异性, 在图中提供了本文算法的可视化示例如图 6 所示, 图 6(a)～图 6(c)为可视化白天场景, 图 6(d)～图 6(f)为可视化夜晚场景。图 6(a)拍摄的是白天街边道路场景, 从检测结果中可以看到, 本文算法可以很好的检测出各类别、各尺寸目标。图 6(b)是在有太阳光干扰情况下的检测结果, 本文算法可以正确检测出各类别目标, 包括遮挡较严重的目标, 但也存在几处漏检和误检情况(黑色虚框)。图 6(c)是十字路口下的检测结果, 图中小目标较多, 造成类别间语义差别较小, 如“people”与“pedestrians”、“motor”与“bicycle”, 而且易被背景信息干扰, 识别难度显著增加。各类别目标还存在严重遮挡问题, 如“motor”和“people”。从图中可以看到, 虽然识别难度增加, 存在部分漏检误检情况(黑色虚框), 但是本文算法依然能够较准确的识别大部分目标。图 6(d)为夜晚街道场景检测结果, 由于夜晚可视信息变少, 城市街道存在各种灯光干扰, 而且车速较快会存在运动模糊等问题, 识别难度大幅增加, 检测结果中可以看到误检漏检情况(红色虚框), 但是整体能够很好的检测各类目标。图 6(e)相较于图 6(d)桥上车辆速度更快, 运动模糊情况进一步加剧, 而且由于大视角存在大量远距离小目标, 外加各种灯光干扰, 使得检测难度提升, 造成漏检误检情况增加(蓝色虚框), 但本文算法仍然可以发挥出色的检测性能, 特别是对于远距离小目标, 以及可视条件较好情况下各类目标的检测(红色框)。图 6(f)是在夜晚街道各种灯光干扰较严重情况下的检测结果, 本文算法可以准确检测, 但存在部分灯光干扰严重及灯光较暗情况下造成的漏检(黑色虚框)。从总的检测效果可以看出, 本文算法对无人机视角图像有强大的检测能力。



(a) 白天街边道路场景　(b) 强光照场景　(c) 交通路口场景

(d) 夜晚街道场景　(e) 夜晚高速桥面场景　(f) 夜晚灯光干扰场景

图 6　本文算法可视化结果

Fig. 6　Results of visualization of the proposed algorithm

4　结　论

本文针对无人机视角图像面临的物体尺度变化大、小尺寸目标占比高、背景干扰大等问题提出一种基于特征信息增强的无人机视角图像目标检测算法。

1) 本文算法利用 MFSPPF 模块替代原网络中 SPPF 模块, 能够更加精准的识别不同尺寸目标。另外, 提出 MBSE 模块有效防止特征融合部分重要信息丢失, 增强了不同尺寸特别是大尺寸目标的检测效果。DFC 模块将低层特征图中丰富的细粒度特征信息传递到高层, 有效提升整体检测性能。

2) 在 VisDrone2021 的实验结果中, 验证了本文模块的有效性。相较于 YOLOv8m, 本文算法 AP、AP50、AP75、AP(s)、AP(m)、AP(l) 分别提高了 3.7%、5.4%、3.9%、3.1%、4.0%、7.9%, 整体检测结果明显提升, 减少了误检和漏检情况。并且本文方法具有良好的适用性, 在 YOLOv8 其他规模不同的模型中也表现出优异的检测性能。

3) 在测试集的实验表明, 本文算法较其他比较算法可以获得更好的检测效果, 并且同时保持了较快的检测速度。从可视化结果可以看出, 本文算法可以在白天、夜晚及光照干扰等情况下很好的应对无人机视角图像目标检测任务。

从检测结果可知，本文算法在强光干扰、运动模糊、可视条件极差、目标之间遮挡、分辨率较低的目标存在少量误检和漏检情况，而且模块参数量较大，后续研究需要进一步改进。

参考文献（References）

[1] 江波, 屈若锟, 李彦冬, 等. 基于深度学习的无人机航拍目标检测研究综述[J]. 航空学报, 2021, 42(4): 524519.
JIANG B, QU R K, LI Y D, et al. Object detection in UAV imagery based on deep learning: review[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(4): 524519(in Chinese).

[2] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2014: 740-755.

[3] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[4] ZHU P, WEN L, DU D, et al. Detection and tracking meet drones challenge[J]. IEEE Trans Pattern Anal Mach Intell, 2022, 44(11): 7380-7399.

[5] JOCHER G. YOLOv8 by ultralytics[EB/OL]. (2023-09-27)[2024-02-10]. https://github.com/ultralytics/ultralytics.

[6] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2999-3007.

[7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multiBox detector[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2016: 21-37.

[8] TIAN Z, SHEN C H, CHEN H, et al. FCOS: a simple and strong anchor-free object detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 1922-1933.

[9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 779-788.

[10] GE Z, LIU S T, WANG F, et al. Yolox: exceeding yolo series in 2021[EB/OL].(2021-08-06)[2024-03-29].https://arxiv.org/abs/2107.08430.

[11] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 7464-7475.

[12] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[13] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2980-2988.

[14] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6154-6162.

[15] WU Y, CHEN Y P, YUAN L, et al. Rethinking classification and localization for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 10183-10192.

[16] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Proceedings of the Computer Vision-ECCV . Berlin: Springer, 2020: 213-229.

[17] ZHU X Z, SU W J, LU L W, et al. Deformable detr: deformable transformers for end-to-end object detection1[EB/OL]. (2021-03-18)[2024-03-29]. https://arxiv.org/abs/2010.04159.

[18] LI F, ZHANG H, LIU S L, et al. DN-DETR: accelerate DETR training by introducing query DeNoising[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 13609-13617.

[19] ZHAO Y, LV W Y, XU S L, et al. Detrs beat yolos on real-time object detection[EB/OL]. (2023-04-17)[2024-03-29]. https://arxiv.org/abs/2304.08069.

[20] KIEFER B, OTT D, ZELL A. Leveraging synthetic data in object detection on unmanned aerial vehicles[C]//Proceedings of the 26th International Conference on Pattern Recognition. Piscataway: IEEE Press, 2022: 3564-3571.

[21] CHEN Y T, LI J, NIU Y F, et al. Small object detection networks based on classification-oriented super-resolution GAN for UAV aerial imagery[C]//Proceedings of the Chinese Control and Decision Conference. Piscataway: IEEE Press, 2019: 4610-4615.

[22] 王殿伟, 胡里晨, 房杰, 等. 基于改进 Double-Head RCNN 的无人机航拍图像小目标检测算法[J]. 北京航空航天大学学报, 2024, 50(7): 2141-2149.
WANG D W, HU L C, FANG J, et al. Small target detection algorithm based on improved Double-Head RCNN for UAV aerial images[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(7): 2141-2149(in Chinese).

[23] 冒国韬, 邓天民, 于楠晶. 基于多尺度分割注意力的无人机航拍图像目标检测算法[J]. 航空学报, 2023, 44(5): 268-278.
MAO G T, DENG T M, YU N J. Object detection in UAV images based on multi-scale split attention[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(5): 268-278(in Chinese).

[24] LI C L, YANG T, ZHU S J, et al. Density map guided object detection in aerial images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2020737-746.

[25] LENG J X, MO M, ZHOU Y H, et al. Pareto refocusing for drone-view object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(3): 1320-1334.

[26] DENG S T, LI S, XIE K, et al. A global-local self-adaptive network for drone-view object detection[J]. IEEE Transactions on Image Processing, 2020, 30: 1556-1569.

[27] CHEN N Y, LI Y, YANG Z M, et al. LODNU: lightweight object detection network in UAV vision[J]. The Journal of Supercomputing, 2023, 79(9): 10117-10138.

[28] YU G H, CHANG Q Y, LV W Y, et al. PP-PicoDet: a better real-time object detector on mobile devices[EB/OL]. (2021-11-01)[2024-03-30]. https://arxiv.org/abs/2111.00902.

[29] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.

[30] JOCHER G. YOLOv5 by ultralytics (version 7.0) [EB/OL]. (2022-11-22)[2024-02-10]. https://doi.org/10.5281/zenodo.3908559.

[31] FENG C J, ZHONG Y J, GAO Y, et al. TOOD: task-aligned one-stage object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 3490-3499.

[32] ZHANG H Y, WANG Y, DAYOUB F, et al. VarifocalNet: an IoU-aware dense object detector[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 8510-8519.

[33] WANG C C, HE W, NIE Y, et al. Gold-YOLO: efficient object detector via gather-and-distribute mechanism[EB/OL]. (2023-09-20)[2024-03-30]. https://arxiv.org/abs/2309.11331.

Object detection algorithm for UAV viewpoint images based on feature information complementation and enhancement

WU Kaijun，PU Zhuo*

(School of Electronic and Information Engineering，Lanzhou Jiaotong University，Lanzhou 730000，China)

Abstract：To address the challenges of significant object scale variation, a high proportion of small-sized objects, and severe background noise interference in images captured from unmanned aerial vehicle (UAV) viewpoint images, a novel object detection algorithm based on feature information enhancement and complementation is proposed. First, to leverage high-level semantic information for capturing richer multi-scale features, a multivariate fusion spatial pyramid pooling-fast (MFSPPF) method is introduced. Second, in order to avoid losing important information during feature fusion, a multi-branch semantic enhancement (MBSE) module is created to extract rich multi-scale features over several branches and create links between these features. In addition, the detailed feature complementation (DFC) module is proposed to extract and refine the low-level feature information to obtain rich, fine-grained feature information and achieve the complementation of the detailed information in the high-level features after feature fusion. Experiments conducted on the VisDrone2021 dataset demonstrate that the proposed algorithm outperforms the baseline YOLOv8m method, with average precision (AP), AP50, AP75, AP(s), AP(m), and AP(l) improving by 3.7%, 5.4%, 3.9%, 3.1%, 4.0%, and 7.9%, respectively. Furthermore, the proposed method is applicable to other YOLOv8 models. The proposed algorithm is appropriate for UAV perspective image detection jobs because it maintains a fast detection speed while achieving better detection performance across a range of intersection over union (IOU) thresholds and object sizes.

Keywords：object detection; UAV images; YOLOv8; feature enhancement; fine-grained feature

Received：2024-04-01；**Accepted**：2024-05-20；**Published Online**：2024-09-05 11：39

URL：link.cnki.net/urlid/11.2625.V.20240904.1623.002

Foundation items：Natural Science Foundation of Gansu Province (23JRRA913); Inner Mongolia Autonomous Region Key Research and Development and Achievement Transformation Program Project (2023YFSH0043); Supported by Key Research and Development Project of Lanzhou Jiaotong University (ZDYF2304)

* **Corresponding author**. E-mail：shiyuepz@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0182

基于深度学习的固体姿轨控发动机性能预示方法

杨慧欣[1,*]，王旭[1]，李响[2]
(1. 沈阳航空航天大学 航空宇航学院，沈阳 110136；2. 西安交通大学 现代设计及转子轴承系统教育部重点实验室，西安 710049)

摘 要：为实现对固体姿轨控发动机性能快速预示，以及评估不同参数对发动机性能影响，提出一种基于特征可视化卷积神经网络仿真代理模型预示方法。通过对固体姿轨控发动机进行数值仿真建模，得到多组仿真数据进行训练，建立一种具有快速推理特性的基于特征可视化卷积神经网络仿真代理模型，可实现发动机性能快速预示。结果表明：针对发动机压强和推力，所建模型预测误差分别为0.2%和6%；针对固体姿轨控发动机多参数输入问题，所提方法可实现鉴定参数对发动机性能的影响程度；针对燃气阀门工作过程的压强，燃气阀门入口压强对压强变化影响更显著；针对燃气阀门推力，阀门开度为3.5 mm以下，阀栓行程对推力预示结果影响更显著，而阀门开度在3.5 mm以上时，影响推力预示结果更多的是入口压强。所提方法有良好的泛用性，可在算法层面直接评估关键参数，免去了多组仿真实验对比的繁琐，大大提高研究效率。

关 键 词：固体姿轨控发动机；代理模型；性能预示；深度学习；特征可视化
中图分类号：V231.1
文献标志码：A **文章编号**：1001-5965(2026)05-1456-11

固体转向与姿态控制系统简称固体姿轨控系统，广泛应用于飞行器飞行的姿态控制、轨道控制及末端制导等。固体姿轨控发动机推力精准调控是飞行器机动性和灵活性的根本保障。固体姿轨控发动机的核心是燃气发生器，负责给每个推力喷管提供高温高压燃气，通过驱动装置调节燃气阀的开度，从而实现推力矢量的调节[1]。固体姿轨控发动机在工作过程中的内部参数，如压强、推力等变化复杂，直接影响发动机性能。针对固体姿轨控发动机基本理论研究的传统方法是通过对固体姿轨控发动机工作过程简化，在设置假设条件基础上进行建模仿真，结合实验数据对发动机开展设计与优化研究。然而，实验数据的获取成本高，周期长，且固体姿轨控系统工作过程物理参数变化较为复杂，涉及多门学科，需要考虑高度耦合动态效应。传统动力学仿真建模研究难度高，难以保证计算精度。虽然仿真建模在一定程度上降低了实验成本，但其运算过程复杂，耗时且效率低。

固体姿轨控发动机阀门在调节过程涉及参数众多，机理复杂，使发动机性能的精确预示变得越发困难。对此，研究者们进行了诸多研究。李娟等[2]建立喉栓二维参数化模型，并基于搜索法确定了一种等效喉部面积与喉栓位置的计算方法，构建了阀栓位置与发动机推力之间的关系；并通过二维仿真模型，研究不同燃气阀型面及喉栓位置对发动机性能的影响。武婷文[3]通过建立稳态仿真模型来研究喉栓式变推力固体火箭发动机，分析了喉栓处于不同位置和喉栓采用不同型面时的发动机流场及发动机性能的变化规律，发现喉栓型面的选取对发动机性能参数的影响较小，而压强恒定与压强变化2种情况对发动机性能的影响十分显著。苗禾状[4]建立稳态仿真模型进行数值模拟，研究了喉栓位置和喉栓速度等属性对燃气阀推力调节的影响规律。吴超[5]建立了基于二维流场仿真对固体姿轨

收稿日期：2024-03-29；录用日期：2024-07-26；网络出版时间：2024-08-13 17:50
网络出版地址：link.cnki.net/urlid/11.2625.V.20240812.1637.001
*通信作者. E-mail：yanghuixin2014@163.com

引用格式：杨慧欣，王旭，李响. 基于深度学习的固体姿轨控发动机性能预示方法[J]. 北京航空航天大学学报，2026，52(5)：1456-1466.
YANG H X，WANG X，LI X. A prediction method for solid divert and attitude control motor performance based on deep learning[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1456-1466 (in Chinese).

控发动机推力快速预示方法, 一定程度解决理论推力计算公式精度低、三维流场仿真计算成本高等问题, 但相较于代理模型方法依旧需要更多时间成本。张杰等[6]基于采用递归演化的拉丁超立方试验方法, 设计包括 4 种维度下喉栓位移、喉栓沉积量、喷管沉积量、压强条件均匀分布的仿真实验, 并根据仿真实验数据, 提出改进径向基的代理模型方法, 能够对考虑沉积效应的发动机推力进行快速预示。尽管最终实验结果达到较高的精度, 但该方法对仿真数据数量要求较高, 需要耗费大量时间成本, 且难以评估各参数对发动机性能影响。

为简化模型复杂度, 提升工作效率, 有研究者提出了代理模型的方法, 如径向基函数 (radial basis functions, RBF)[7]、高斯过程 (Gaussian process, GP)[8]、支持向量基 (support vector regression, SVM) 回归模型[9]等, 这些都是在仿真数据的基础上, 构建“数据拟合模型”, 直接建立输入输出的对应关系, 以实现快速仿真。代理模型是一类近似数学模型, 此模型可替代在分析和优化设计过程中比较复杂和费时的数值分析, 也称为近似模型或元模型[10]。然而这些传统代理模型对于高复杂度、高非线性、高维度数据处理能力存在欠缺。

深度学习复杂系统建模在图像识别、语义分割、自然语言处理、机械故障诊断、寿命评估等多种工程领域验证了其可行性。深度学习的核心思想是模拟神经元间的相互连接, 以实现复杂的信号传递。深度神经网络模型在处理高度非线性变量时具有独特优势, 可进行快速推理, 对专家知识需求较低, 效率更高。在近些年来, 以深度学习为基础的代理模型成为了一种新的流行趋势。Tang 等[11]构建一种基于深度卷积和循环神经网络架构的“循环 R-U 网络”作为代理模型, 用于准确预测渠道化地质模型中的动态地下流的动态压力变化。何磊等[12]建立基于长短期记忆(long short-term memory, LSTM)神经网络的非定常气动力模型, 实现了对三角翼大迎角失速状态下, 升力系数、阻力系数和俯仰力矩系数的预测。Sun 等[13]设计了一种物理约束的深度学习方法的流体流动的代理模型, 通过深度神经网络架构来加强初始和边界条件, 将控制的雷诺平均 Navier-Stokes 方程纳入神经网络的损失进行训练, 并成功通血液动力学流体力学仿真结果验证。

深度学习模型都是黑箱模型, 在训练过程缺乏可解释性。在深度学习领域, 显著图(saliency map, SM)最早应用于图像领域, 其通过可视化分类结果相对于输入的梯度分布, 来阐明模型内部输入与输出之间的对应关系[14]。Yang 等[15]研究滚动轴承的振动信号, 通过分析输出结果相对于输入的梯度变化, 阐释了每个输入如何影响结果。Zeiler 等[16]对卷积神经网络 (convolutional neural networks, CNN) 中的不同层使用反卷积操作, 将特征重新映射到像素空间。同时, 引入遮蔽实验, 通过掩盖图像中的不同部分来研究这些部分对预测结果的影响, 以增强卷积神经网络的可解释性。

无论是基于数值仿真的方法还是基于代理模型的方法都存在局限性。构建基于深度神经网络的代理模型, 在充足的仿真数据支持下, 不仅能快速准确地预测固体姿轨控发动机性能, 还能显著评估不同输入参数对输出参数的影响, 从而有效识别关键参数, 为固体姿轨控发动机设计提供参考。本文提出一种特征可视化卷积神经网络的代理模型方法, 实现了发动机性能的快速预测。同时, 通过显著图进行特征可视化, 阐明了不同输入参数对输出参数的具体影响, 为固体姿轨控发动机的智能化建模研究提供参考。

1　模型原理

基于深度学习的固体姿轨控发动机性能预示方法采用卷积神经网络和特征可视化等技术, 除用于固体姿轨控发动机性能快速预示的同时, 可以通过特征可视化鉴别关键参数。代理模型通过卷积神经网络以足量仿真数据训练而成。特征可视化通过计算输出结果对输入参数的梯度来实现, 以显著图方式展现。根据梯度大小说明不同输入参数对输出结果的影响程度。

1.1　基于特征可视化的卷积神经网络

1.1.1　卷积神经网络

卷积神经网络最初用于图像分类, 通过多个卷积核来提取目标的不同特征信息, 不断堆叠卷积层和池化层, 提取目标的更深层次、更抽象的特征信息。通过展平层将这些特征整合, 以获取全局的特征信息。在提取特征后, 网络使用全连接层, 激活函数在层与层之间实现非线性映射。与全连接神经网络相比, 卷积神经网络通过权重共享特性显著减少了对计算资源的需求, 从而提高计算效率。

在第 i 层卷积神经网络的表达式为

$$y_i^{l+1}(j)=f(X^l(j)W_i^l+b_i^l) \tag{1}$$

式中: f 为激活函数; l 为该网络中的卷积层; i 为在第 l 层的第 i 个卷积核; W_i^l 为层 l 中的共享权重; b_i^l 为与之相应的偏置项; $X^l(j)$ 为对应于 l 层的第 j 个感受野, 将经过卷积后结果输入到非线性激活函数中, 可以获得该层神经元的输出; $y_i^{l+1}(j)$ 为经过激活函数后提取的特征。

1.1.2 显著图

显著图是一种用于可视化深度学习内部特征的方法，不同的输入参数会影响神经网络的预测结果，而输入参数越复杂，解释这些参数对设计的影响就越困难。固体姿轨控发动机的动力学物理仿真模型涉及许多物理参数，这使得其设计具有较大的复杂性。因此，这种结合了显著图的深度卷积神经网络可以作为一种新型智能化方法，用于定性解释物理参数对输出结果的影响。对于由 n 个参数组成的输入 $X=\{x_0,x_1,\cdots,x_n\}$，第 m 个输入参数梯度的表达式为

$$W_m=\partial y/\partial x_m \qquad m\in[0,n] \tag{2}$$

式中：y 为神经网络的输出，处在神经网络最后一层；W_m 为输出 y 对第 m 个输入参数传播的梯度。每个梯度结果都对应着一个输入特征，在模型训练完成后，通过显著图可以监测输出相对于输入的梯度变化，从而清晰反映出输入参数的特征分布。这种特征可视化卷积神经网络经过充分训练后，不仅能够实现快速预测，还辅助识别关键输入参数，为固体姿轨控发动机设计提供有利参考。图 1 为特征可视化的卷积神经网络模型示意图，对包括 m 组样本训练集训练完成后，计算不同样本对应输出对输入的梯度结果，通过显著图的方式进行可视化，来说明不同输入参数对于输出的影响。梯度反馈越大，则输入参数对于输出的影响越大。

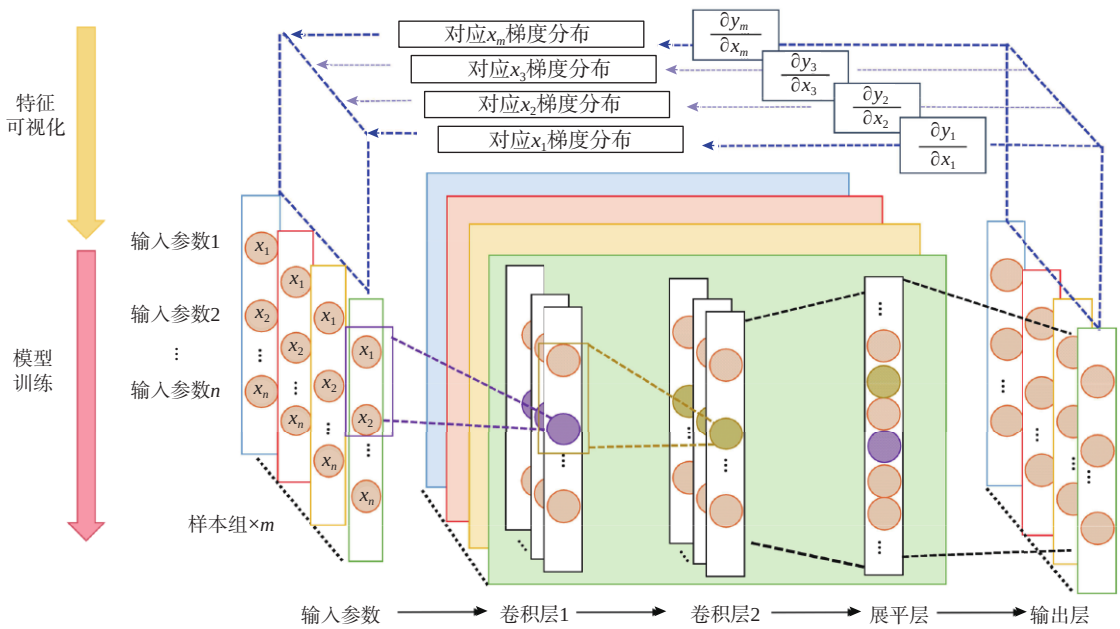


图 1 特征可视化的卷积神经网络模型示意图

Fig. 1 Schematic diagram of a convolution neural networks model for feature visualization

1.2 模型训练流程

神经网络在训练过程中会进行多次迭代，每轮迭代网络输出一个预测值，作为这一轮的标签。特征可视化的卷积神经网络主要基于 Adam 优化算法和小批量算法，通过反向传播算法，根据预测值与实际标签之间的误差，误差值通过损失函数进行计算，自适应地调整网络中每层的权值。经过不断训练迭代，使预测值与实际标签误差变小，直到损失误差达到收敛状态。训练完成后，通过监测输出对输入的梯度变化，揭示不同输入参数对输出结果的影响程度。特征可视化的卷积神经网络算法流程如图 2 所示，其具体步骤如下。

步骤 1 将输入物理参数进行归一化处理，以适应不同量级的物理量输入，使神经网络能够准确监测梯度训练过程中传播大小。归一化采用最大最小归一化方式，使输入大小保持在 [0,1] 之间。其数学表达式为

$$x'_j=(x_j-x_{\min})/(x_{\max}-x_{\min}) \tag{3}$$

式中：x'_j 为单组样本中的第 j 个归一化后的节点；x_j 为单组样本中第 j 个归一化前的节点；$x_{\min}$ 为单组样本中数值最小的节点；$x_{\max}$ 为单组样本中数值最大的节点。

步骤 2 搭建神经网络，设置神经网络层数，隐藏层内部节点数量、学习率、正则化方法、激活函数、确定迭代次数。

步骤 3 划分训练集和测试集，并用小批量数据集对网络进行训练，每遍历一次全部训练集即完成一次迭代。根据精度指标或预设的训练轮次判断是否停止训练，若是，则停止训练，否则继续下一步。

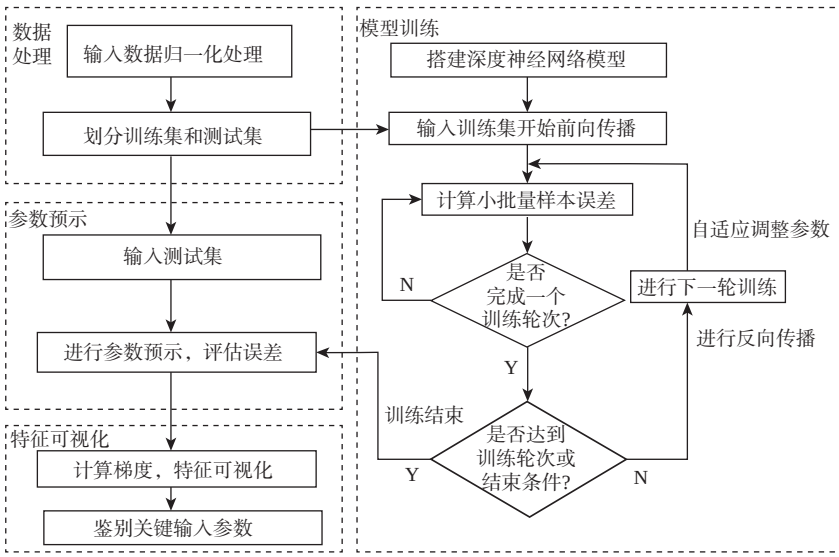


图 2　特征可视化的卷积神经网络算法流程

Fig. 2　Flow chart of convolutional neural network algorithm based on feature visualization

步骤 4　神经网络模型根据模型损失函数值进行反向更新，自适应进行优化调整深度神经网络中每层的权值。

步骤 5　基于上述训练完成的模型，对固体姿轨控发动机性能进行快速预示。

2　固体姿轨控发动机燃气调节阀性能快速预示

压强和推力是燃气调节阀实现精确调节的重点关注对象。燃气主要通过阀栓移动改变等效喉部面积及燃气流量，影响燃气阀内部压强，进行推力调节，从而影响燃气阀性能。在推力调节过程中涉及众多参数，无论是根据理论公式进行计算还是采用传统的动力学仿真方法进行研究，都存在实现困难、效率低下等问题。本节基于本文方法，分别对固体姿轨控发动机燃气调节阀压强、推力进行预示，并通过特征可视化来分析不同输入参数对发动机性能的影响。本文方法具有良好的泛用性，能在算法层面直接寻找关键参数，显著提高研究效率。

2.1　数据集

本算例数据集来源于固体姿轨控发动机燃气调节阀仿真数据，分析了燃气调节阀在不同工作压强下的性能。初始条件设置如下：入口总压分别设置为 6 MPa、8.3 MPa、10 MPa、12 MPa，出口压强为 101 325 Pa，温度均为 300 K。数值仿真开始时，喉栓从完全闭合位置启动，并假设在整个移动过程中保持匀速。在每个入口压强条件下，阀芯的运动速度 v 分别设置为 5 mm/s、10 mm/s、25 mm/s 和 2 mm/s。将阀芯运动自完全闭合，到完全打开，再到完全闭合定义一个周期的工作流程。阀芯运动的最大开度为 4.5 mm。图 3 为燃气阀工作压强为 10 MPa 下，阀栓以 5 mm/s 速度运动的仿真结果。

可以看出，随着燃气阀开度逐渐增大，燃气阀喉部压力及其分布范围逐渐增大，且喉栓头部压力也逐渐增大。这是因为随着燃气阀开度的增大，喉栓头部与喷管收缩段间的距离的增加，使喉栓头部附近存在高压气低速区域面积不断增大，最终导致燃气阀喉部压强及其分布范围的增大。

为验证本文仿真模型数据的有效性，对仿真结果与实验结果进行对比。实验主要在阀腔内设置了 2 个监测点，实时采集喉栓实际运动过程中的压强数据。2 个监测点位于燃气阀内壁同一平面上，用于反映阀腔内压力的变化，其具体位置如图 4 所示。三维仿真模型如图 5 所示。

实验数据为燃气阀门入口压强在约为 8.2～8.3 MPa 范围下，移动速度分别为 2 mm/s、5 mm/s、10 mm/s 及 25 mm/s 时燃气阀门监测点的压强数据。燃气阀动态实验应用步进电机按照预定指令控制喉栓从完全关闭位置以一定速度运动到预定位置，经过一定时间后步进电机控制喉栓从开启状态至闭合状态。采样频率为 10 kHz，相关实验描述如下：固体姿轨控发动机燃气阀调节实验系统主要由燃气阀、供气系统、伺服系统和测试系统等组成，燃气阀调节实验装置如图 6 所示。

实验过程如下：供气系统阀门开启，实验系统管路充气，直到管路压强稳定；伺服控制器发出指令，控制驱动器发出脉冲电流，控制喉栓按指令运行；喉栓运动指令结束，供气系统结束供气，实验管路降压，结束数据采集。

为验证喉栓动密封可靠性，燃气阀门入口压强

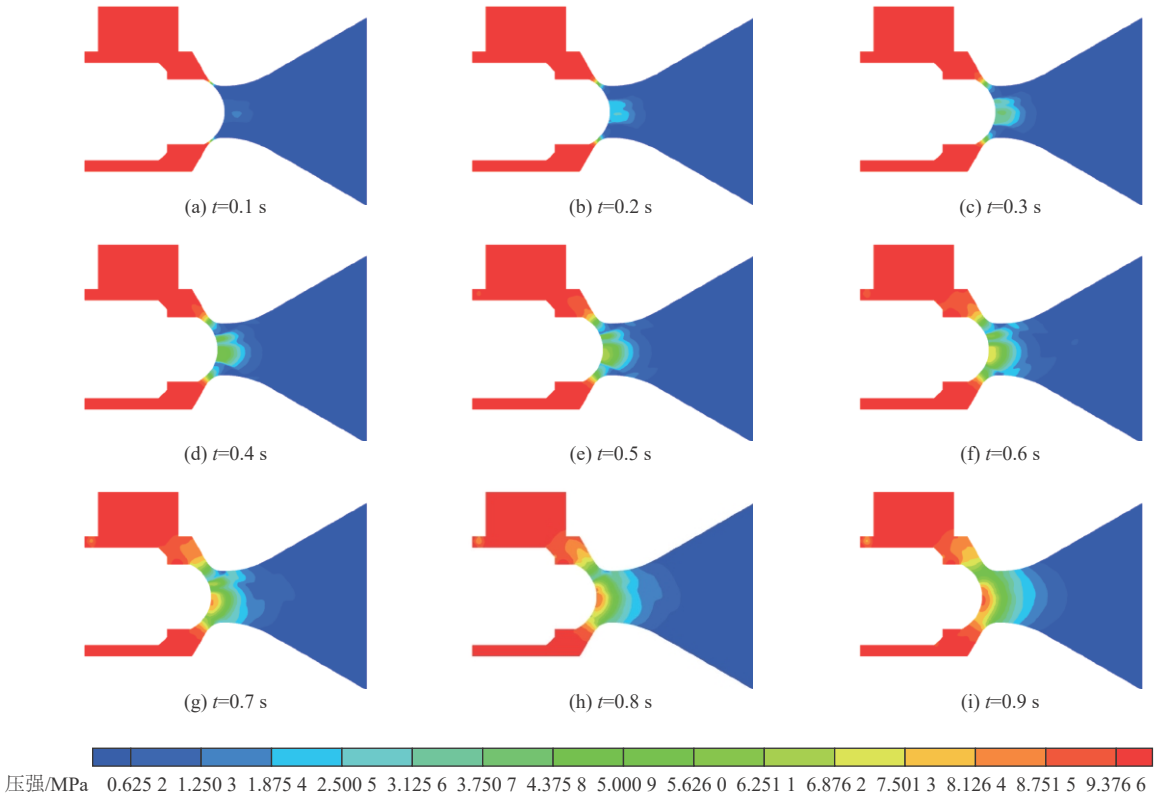


图 3 10 MPa 入口压强下燃气阀随时间变化的压强云图

Fig. 3 Pressure cloud of gas valve over time with 10 MPa inlet pressure

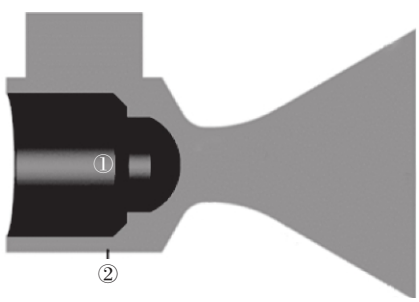


图 4 燃气阀监测点对应位置

Fig. 4 Corresponding location of gas valve monitoring points

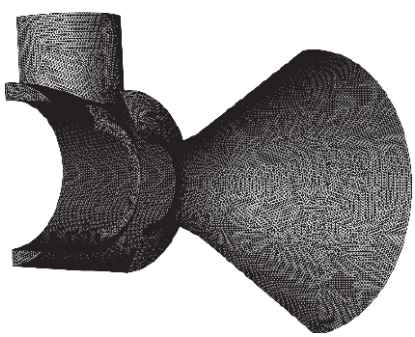


图 5 燃气阀三维模型

Fig. 5 Gas valve 3D model

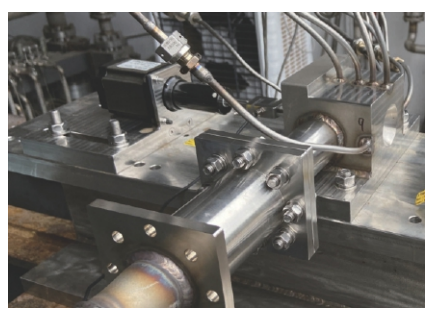


图 6 燃气阀调节实验装置

Fig. 6 Gas valve adjustment experiment device

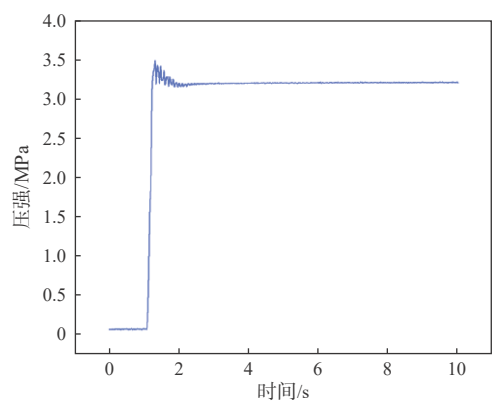


图 7 燃气阀密封性能测试

Fig. 7 Gas valve sealing performance test

为 3.2 MPa 时对燃气阀进行压力测试，通过步进电机使喉栓处于完全闭合状态，对燃气阀管路加压，监测燃气阀内部压强，测试结果如图 7 所示。根据测试结果可以看出，加压后燃气阀内部压强迅速增加，最大压强峰值为 3.49 MPa，压强峰值与稳定压强比值为 1.09；约 1 s 后管路内部压强稳定在 3.2 MPa，

说明使用喉栓动密封满足实验需求。

之后进行燃气阀动态调节实验, 喉栓运动速度为 5 mm/s, 燃气阀进气总压为 8.2 Mpa, 对燃气阀喉栓部分燃气流量进行实时监控, 实验结果如图 8 所示。

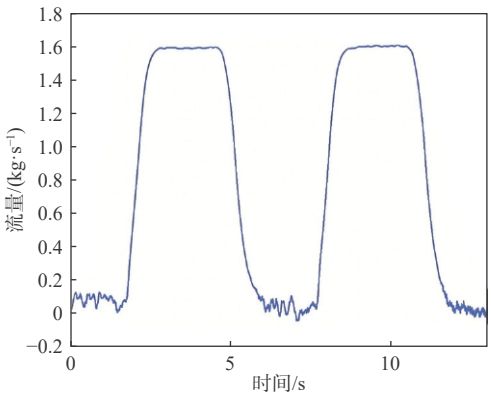


图 8　燃气阀喉栓部位流量曲线

Fig. 8　Flow curve of gas valve throat plug section of gas valve

由图 8 可知, 燃气流量上升过程随时间呈线性变化, 能够较好地实现燃气流量线性调节。

研究中采用衡量指标以平均绝对误差(mean absolute error, MAE) 和平均绝对百分比误差 (mean absolute percentage error, MAPE) 为主, 计算式分别为

$$E_{\text{mae}}=\frac{1}{n}\sum_{i=1}^{n}\left|Y_i-\hat{Y}_i\right| \tag{4}$$

$$E_{\text{mape}}=\frac{1}{n}\sum_{n=1}^{n}\left[\left(\left|Y_i-\hat{Y}_i\right|/Y_i\right)\times 100\%\right] \tag{5}$$

式中: Y_i为样本数据中第 i 个样本点压强真实值; $\hat{Y}_i$为第 i 个样本点对应的模型预测值。

4 组测试的实验结果与仿真计算结果平均百分比误差如表 1 所示。

表 1　监测点压强平均绝对百分比误差对比

Table 1　Comparison of mean absolute percentage error of pressure at monitoring points

监测点	MAPE/%			
	v=2 mm/s	v=5 mm/s	v=10 mm/s	v=25 mm/s
1	6.17	4.24	6.09	7.80
2	5.89	4.31	5.82	7.11

2.2　结果与分析

建立压强快速预示模型时, 模型的输入参数包括: 阀栓的入口压强和移动速度, 输出参数是 2 个监测点的压强。鉴于仿真数据集中入口压强与阀栓移动速度参数的均匀分布, 选取 14 组样本作为训练集, 随机选取 2 组具有不同入口压强和阀栓移动速度的样本作为测试集。对于不同阀栓移速分别取不同时间步长, 输出 180 个样本点。在训练完成后, 再计算预测数据相对于输入参数的梯度, 并进行特征可视化, 分析不同输入对结果的影响, 以寻找关键参数。

针对本文模型的建立, 模型的输入参数包括入口压强和阀栓行程, 输出参数为对应的出口推力。考虑入口压强和阀栓行程都是影响推力的重要参数, 其中, 阀栓行程直接影响喉部面积。根据推力调节公式, 等效喉部面积的减小导致燃气流量减低, 从而影响推力。而入口压强直接影响燃气阀门工作过程内部流场, 在训练完成后, 利用特征可视化评估这 2 个因素的重要性。推力快速预示模型仍为 2.1 节中仿真模型产生的仿真数据集, 但仅包括在 4 个不同入口压强 p 下, 分别为 6, 8.3, 10, 12 MPa, 阀栓移速为 5 mm/s 进行周期性运动的推力变化, 阀栓行程每改变 0.5 mm 计算一次, 共产生 40 组仿真数据。样本按 4:1 的比例划分为训练集和测试集, 即 32 组作为训练集, 8 组样本作为测试集, 后者仅用于评估模型性能, 不参与训练过程。图 9 为燃气阀以 5 mm/s 速度运动下推力变化。

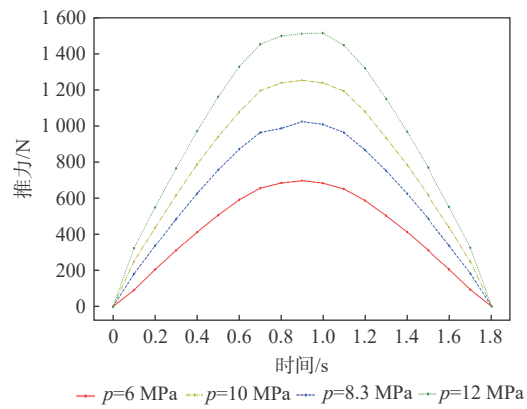


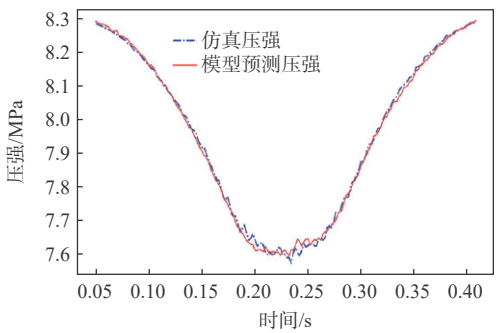
图 9　燃气阀速度运动 5 mm/s 时的出口推力

Fig. 9　Gas valve outlet thrust with 5 mm/s speed

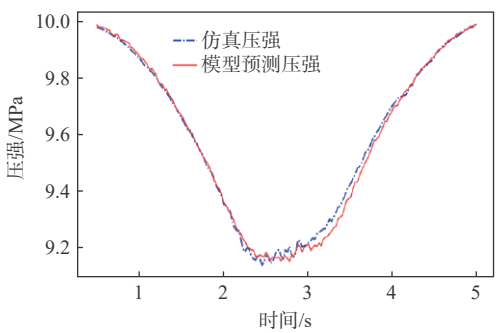
在同一个周期内, 阀门在开启和关闭过程中的推力变化曲线基本对称, 且推力随进气口压强的增大而增大。随着阀门逐渐打开, 燃气阀的质量流率和压强逐渐增大, 当阀门完全打开时, 其出口推力达到最大。在阀门关闭过程中, 燃气阀的质量流率和压强逐渐减小, 直至阀门完全关闭时, 推力恢复至初始值。因此, 研究重点放在阀门从完全闭合到达到最大行程的半个周期过程。

燃气阀门压强快速预示模型的预测结果如图 10 和图 11 所示, 预测误差分别如表 2 和表 3 所示。仿真数据集中, 燃气阀门内壁监测点压强随着阀门逐渐开启, 呈现减小趋势; 随着阀门逐渐闭合呈现增大趋势。由于阀门打开, 阀腔内部气体由于

内外压差，受到挤压，气体向喷管溢出，此时阀栓头部在阀门开启过程中，在阀栓头部气体流速降低，逐渐形成高压区域，燃气阀门内壁监测点压强会随着阀门开度增大而降低。而随着阀门闭合，气体在腔体内部难以溢出，燃气阀门内壁监测点压强会逐渐增大，压强变化过程与阀门开启过程相反。阀门内部测试集的拟合误差均小于 0.2%；平均绝对误差最大为 14 300 Pa，相较于以 MPa 为单位的仿真结果，模型预测达到相当高的精度。另外，训练好的模型预测结果仅需 1 s 以内，相较于仿真模型耗费的时间，大大提高了模型的计算效率。



(a) p=8.3 MPa, v=25 mm/s



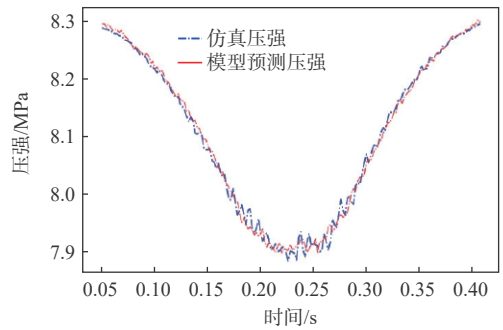
(b) p=10 MPa, v=2 mm/s

图 10　2 组工况下燃气阀监测点 1 的预测结果

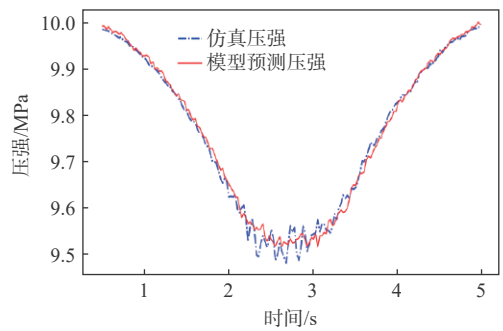
Fig. 10　Predicted results for gas valve monitoring point 1 for two operating conditions groups

本文模型在训练完成后，计算所有训练样本中模型输出结果对输入的梯度，并以显著图方式进行特征可视化，以此评价不同参数对结果的影响程度。输入参数样本梯度分布如图 12 所示。

根据图 12 显示，入口压强的梯度量级远远大于阀栓移速的梯度反馈，入口压强对于输出结果影响更显著。为验证结果对比了多组仿真结果。图 13 为入口压强为 8.3 MPa 时不同阀栓移速下，2 个监测点压强变化结果。横坐标阀栓为阀栓位置，令处于闭合状态下阀门的位置为零点。图中展示阀门自闭和状态逐渐开启到 4.5 mm 行程，之后逐渐闭合到初始状态的过程。当仿真模型保持同一入口压强时，随着阀栓移速的改变，监测点变化趋势基本不受影响。图 14 为在阀栓移动速度为 5 mm/s 时，



(a) p=8.3 MPa, v=25 mm/s工况



(b) p=10 MPa, v=2 mm/s工况

图 11　2 组工况下燃气阀监测点 2 的预测结果

Fig. 11　Predicted results for gas valve monitoring point 2 for two operating conditions groups

表 2　燃气阀监测点压强预测平均绝对百分比误差

Table 2　Comparison of mean average percent error of pressure at gas valve monitoring points

模型输入	MAPE/%	
	监测点1	监测点2
p=8.3 MPa, v=25 mm/s	0.12	0.11
p=10 MPa, v=2 mm/s	0.15	0.12

表 3　燃气阀监测点压强预测平均绝对误差

Table 3　Comparison of mean absolute error of pressure at gas valve monitoring points

模型输入	MAE/Pa	
	监测点1	监测点2
p=8.3 MPa, v=25 mm/s	9 500	8 700
p=10 MPa, v=2 mm/s	14 300	11 700

不同入口压强的监测点压强变化结果。在相同阀栓移速下，入口压强的改变对仿真结果的影响显著。

构建燃气阀门推力快速预示模型过程中，为确保模型有效性并减少偶然性，本文在固定模型参数的基础上，将样本随机分为 2 组不同的训练集和测试集。燃气阀门出口推力快速预示模型预测结果如图 15 所示，测试集预测绝对误差及平均百分比误差如表 4 和表 5 所示，其中，L 为阀栓行程。根据 2 组测试集结果反馈，推力预测误差都在 6% 以下。验证了该代理模型的有效性。

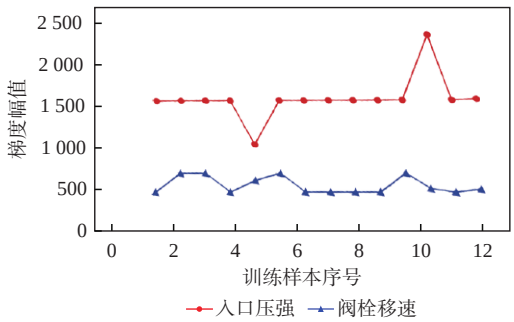
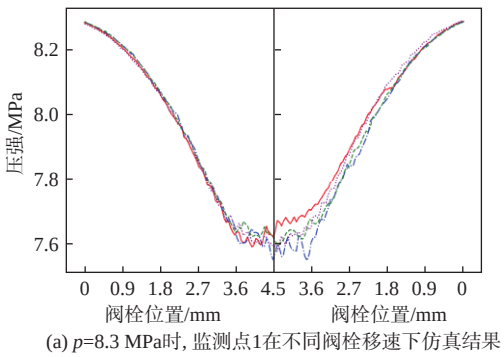
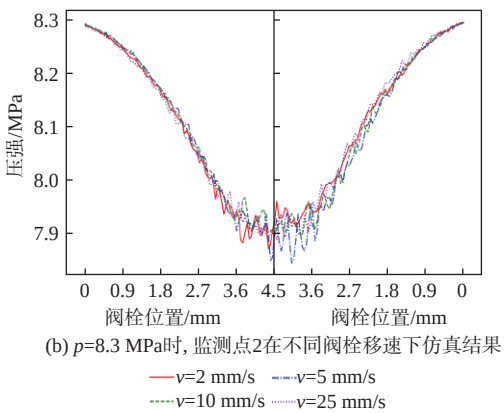


图 12　入口压强和阀栓移速梯度分布

Fig. 12　Gradient distribution of inlet pressure and pintle velocity



(a) p=8.3 MPa时, 监测点1在不同阀栓移速下仿真结果

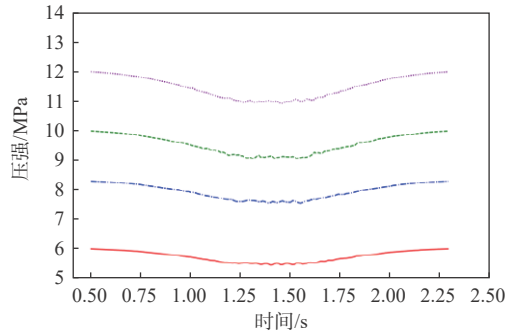


(b) p=8.3 MPa时, 监测点2在不同阀栓移速下仿真结果

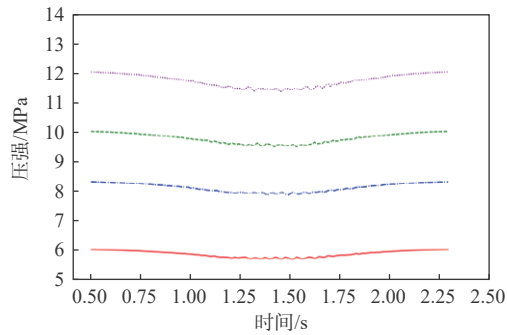
图 13　燃气阀监测点在不同阀栓移速下仿真结果

Fig. 13　Simulation results of gas valve monitoring points with different pintle velocity

同样的, 在训练完成后, 通过计算所有训练样本中模型输出对输入参数的梯度, 利用显著图实现特征可视化, 识别关键参数。输入参数样本梯度分布如图 16 和图 17 所示。图 16 和图 17 显示了输入参数样本的梯度分布, 其中, 每组训练集包含 32 个样本。在多数样本中, 入口压强对预测结果的梯度反馈较小。相比之下, 当阀栓行程大于 3.5 mm 时, 入口压强的梯度反馈超过阀栓行程。这是因为阀门在工作初期对燃气流量的影响更大, 导致在不同入口压强下推力差异不明显。而随着阀栓开度的增大, 阀门对于燃气流量影响逐渐减小, 而阀门推力随着阀栓行程增加, 在不同入口压强下的差异越来越显著。此外, 测试集中推力预测误差较大的样本主要出现在阀门低开度时, 表明在低开度条件下阀栓行程对推力的影响更为显著。因此, 对于在阀栓开度为 3.5 mm 以下, 阀栓行程对推力预示结果



(a) v=5 mm/s时, 监测点1在不同入口压强下仿真结果

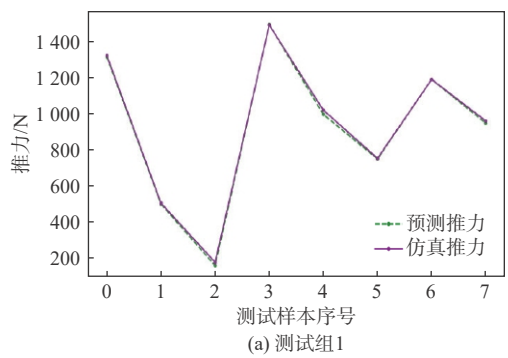


(b) v=5 mm/s时, 监测点2在不同入口压强下仿真结果

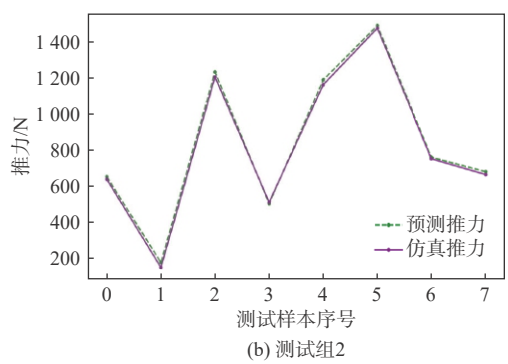
p=6 MPa　p=10 MPa　p=8.3 MPa　p=12 MPa

图 14　燃气阀监测点在不同入口压强下仿真结果

Fig. 14　Simulation results of gas valve monitoring points with inlet pressure



(a) 测试组1



(b) 测试组2

图 15　不同测试组的推力预测结果

Fig. 15　Thrust prediction results for different test groups

表 4 测试集 1 误差对比

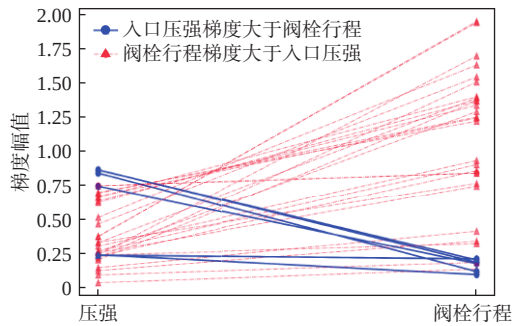
Table 4 Error comparison for test set 1

模型输入	平均百分比误差/%	平均绝对误差/N
p=6 MPa, L=0.5 mm	16.54	14.16
p=6 MPa, L=1.5 mm	9.33	29.18
p=6 MPa, L=2 mm	1.19	4.95
p=6 MPa, L=2.5 mm	0.52	2.62
p=8.3 MPa, L=4.5 mm	3.82	39.07
p=10 MPa, L=4.0 mm	0.05	0.56
p=12 MPa, L=0.5 mm	10.45	9.85
p=12 MPa, L=2.5 mm	0.94	10.94

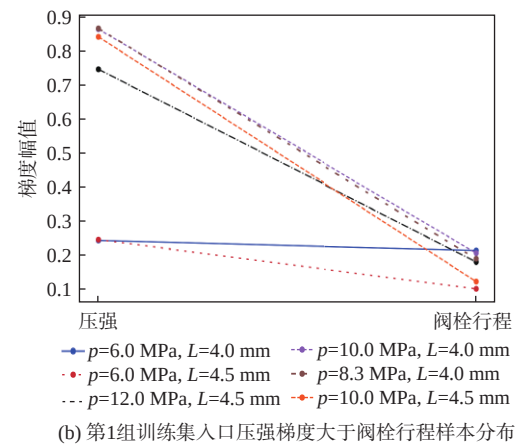
表 5 测试集 2 误差对比

Table 5 Error comparison for test set 2

模型输入	平均百分比误差/%	平均绝对误差/N
p=6 MPa, L=2.5 mm	1.42	7.21
p=6 MPa, L=3.5 mm	3.63	23.87
p=6 MPa, L=4.0 mm	2.90	19.92
p=8.3 MPa, L=1.5 mm	5.22	9.45
p=10 MPa, L=3.5 mm	1.20	14.30
p=10 MPa, L=4.0 mm	2.33	28.85
p=12 MPa, L=1.5 mm	1.14	8.71
p=12 MPa, L=4.0 mm	0.21	3.03



(a) 第1组训练集入口压强和阀栓行程梯度分布

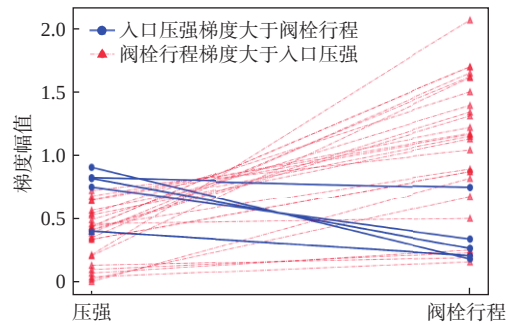


(b) 第1组训练集入口压强梯度大于阀栓行程样本分布

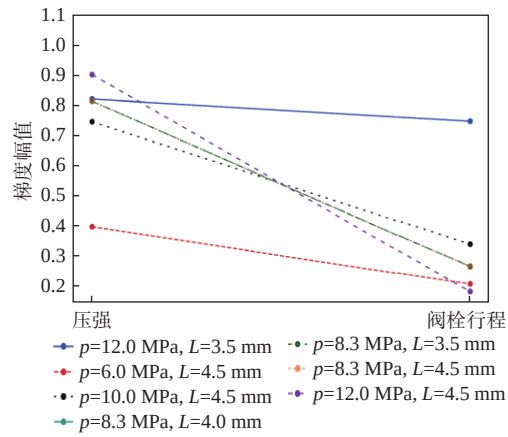
图 16 第 1 组测试集梯度结果对比

Fig. 16 Comparison of gradient results for the first test set

影响更显著，而在阀栓开度在 3.5 mm 以上时，影响推力预示结果的更多是入口压强。



(a) 第2组训练集入口压强和阀栓行程梯度分布



(b) 第2组训练集入口压强梯度大于阀栓行程样本分布

图 17 第 2 组测试集梯度结果对比

Fig. 17 Comparison of gradient results for the second test set

3 结 论

1）本文提出一种基于特征可视化的卷积神经网络代理模型构建方法，该方法可实现固体姿轨控发动机性能的快速预示，并通过对固姿燃气阀调节过程仿真数据的训练，实现阀门压强和推力快速预示，预示误差分别达到 0.2% 和 6%；预示仅需 1 s 以内，相较于仿真方法，本文方法可以大大提高效率。

2）通过引入显著图方法实现特征可视化，针对训练完成的燃气阀压强及推力预示模型，通过分别展示燃气阀压强、推力对各个输入参数的梯度，在算法上直接评估不同输入参数对燃气阀性能的影响程度，无需通过多组仿真结果对比，显著提高了研究效率。为固体姿轨控发动机智能化发展提供参考。

参考文献（References）

[1] 万东, 何国强, 王占利, 等. 针栓喷管技术在固体姿轨控系统中的应用研究[J]. 现代防御技术, 2011, 39(3): 48-54.
WAN D, HE G Q, WANG Z L, et al. An application study of pintle-nozzle technology for solid divert and attitude control system[J]. Modern Defence Technology, 2011, 39(3): 48-54(in Chinese).

[2] 李娟, 李江, 王毅林, 等. 喉栓式变推力发动机性能研究[J]. 固体火箭技术, 2007, 30(6): 505-509.
LI J, LI J, WANG Y L, et al. Study on performance of pintle controlled thrust solid rocket motor[J]. Journal of Solid Rocket Techno-

logy, 2007, 30(6): 505-509(in Chinese).

[3] 武婷文. 喉栓式变推力固体火箭发动机稳态与瞬态工作特性研究[D]. 西安: 航天动力技术研究院, 2023.
WU T W. Study on steady and transient operating characteristics of throat plug variable thrust solid rocket motor[D]. Xi'an: Academy of Aerospace Solid Propulsion Technology, 2023(in Chinese).

[4] 苗禾状. 喉栓式可控固体火箭发动机推力调节研究[D]. 哈尔滨: 哈尔滨工程大学, 2009.
MIAO H Z. Study on thrust adjustment of throat-plug controllable solid rocket motor[D]. Harbin: Harbin Engineering University, 2009(in Chinese).

[5] 吴超. 针栓式固体姿轨控发动机性能分析与设计优化[D]. 长沙: 国防科技大学, 2022.
WU C. Performance analysis and design optimization of pin-bolt solid attitude and orbit control engine[D]. Changsha: National University of Defense Technology, 2022(in Chinese).

[6] 张杰, 李国盛, 文谦, 等. 基于改进增广径向基的固体姿轨控发动机推力快速预示[J]. 推进技术, 2023, 44(8): 201-209.
ZHANG J, LI G S, WEN Q, et al. Fast thrust prediction method for solid divert and attitude control system based on improved augmented radial basis functions[J]. Journal of Propulsion Technology, 2023, 44(8): 201-209(in Chinese).

[7] REGIS R G, SHOEMAKER C A. A stochastic radial basis function method for the global optimization of expensive functions[J]. INFORMS Journal on Computing, 2007, 19(4): 497-509.

[8] ATKINSON S, ZABARAS N. Structured Bayesian Gaussian process latent variable model: applications to data-driven dimensionality reduction and high-dimensional inversion[J]. Journal of Computational Physics, 2019, 383: 166-195.

[9] YAN C, YIN Z Y, SHEN X L, et al. Surrogate-based optimization with improved support vector regression for non-circular vent hole on aero-engine turbine disk[J]. Aerospace Science and Technology, 2020, 96: 105332.

[10] 刘轲. 基于代理模型的运载火箭多学科优化及可视化研究[D]. 绵阳: 西南科技大学, 2022.
LIU K. Research on multidisciplinary optimization and visualization of launch vehicle based on proxy model[D]. Mianyang: Southwest University of Science and Technology, 2022(in Chinese).

[11] TANG M, LIU Y M, DURLOFSKY L J. A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems[J]. Journal of Computational Physics, 2020, 413: 109456.

[12] 何磊, 张显才, 钱炜祺, 等. 基于长短时记忆神经网络的非定常气动力建模方法[J]. 飞行力学, 2021, 39(5): 8-12.
HE L, ZHANG X C, QIAN W Q, et al. Unsteady aerodynamics modeling method based on long short-term memory neural network[J]. Flight Dynamics, 2021, 39(5): 8-12(in Chinese).

[13] SUN L N, GAO H, PAN S W, et al. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data[J]. Computer Methods in Applied Mechanics and Engineering, 2020, 361: 112732.

[14] SIMONYAN K, VEDALDI A, ZISSERMAN A. Deep inside convolutional networks: visualising image classification models and saliency maps[EB/OL]. (2014-04-19)[2024-03-20]. https://arxiv.org/abs/1312.6034.

[15] YANG H X, LI X, ZHANG W. Interpretability of deep convolutional neural networks on rolling bearing fault diagnosis[J]. Measurement Science and Technology, 2022, 33(5): 055005.

[16] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//Proceedings of the Computer Vision. Berlin: Springer, 2014: 818-833.

A prediction method for solid divert and attitude control motor performance based on deep learning

YANG Huixin[1,*], WANG Xu[1], LI Xiang[2]

(1. School of Aerospace Engineering, Shenyang Aerospace University, Shenyang 110136, China;
2. Key laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A feature visualization convolutional neural network-based surrogate model is suggested to achieve the quick prediction of solid divert and attitude control motor performance and assess the influence of various parameters. After obtaining a number of simulation datasets for training from numerical simulation modeling of the solid divert and attitude control motor, a feature visualization convolutional neural network simulation surrogate model with fast inference characteristics is established, which can realize fast performance prediction. The results show that for pressure and thrust of the solid divert and attitude control motor, the model prediction error is 0.2% and 6%, respectively. For multi-parameters input problem of the solid divert and attitude control motor, the proposed method can identify the influence of parameters on performance. Pressure changes have a significant impact on the prediction results for solid divert and attitude control motor gas valve pressure during operation; for gas valve thrust, the pintle stroke has a greater impact on the thrust prediction result when the valve opening is less than 3.5 mm, while the inlet pressure has a greater impact when the valve opening is greater than 3.5 mm. The method has good generalizability and can directly evaluate the key parameters at the algorithmic level, while eliminating the cumbersome comparison of multiple sets of simulation experiments, which can greatly improve the research efficiency.

Keywords: solid divert and attitude control motor; surrogate model; performance prediction; deep learning; feature visualization

Received: 2024-03-29; **Accepted**: 2024-07-26; **Published Online**: 2024-08-13 17:50
URL: link.cnki.net/urlid/11.2625.V.20240812.1637.001
* **Corresponding author**. E-mail: yanghuixin2014@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0108

基于流体单位质量力的飞机燃油测量算法

王晓阳，任恒英*，饶波

(航空工业第一飞机设计研究院 机电系统设计研究所，西安 710089)

摘　　要：为准确测量飞机油箱内的燃油体积，提出基于流体单位质量力的飞机燃油测量算法（AFUMF）。在机体坐标系下基于三轴加速度计信号等效流体单位质量力直接确定油面单位法向量，将燃油箱内油量测量传感器测量的油面高度转换为燃油面上的点坐标，得到油面距离机体原点的有向距离用于确定油面高度，建立燃油测量数据库并实现燃油测量传感器高度至燃油体积解算。基于CATIA二次开发，采用蒙特卡罗方法对最小二乘（LMS）拟合和AFUMF开展误差评估，结果表明：在传感器存在误差的条件下，AFUMF对测量传感器误差有减小作用且仅需1根传感器有效即可实现油量解算，LMS对测量传感器误差有放大作用且需至少3根传感器有效以实现油量解算。通过动态燃油箱台架试验对AFUMF进行验证，试验结果表明：在测量范围内姿态稳定条件下，燃油测量最大误差为±0.5%油箱容积。

关 键 词：飞机燃油测量；燃油测量数据库建立；燃油体积解算；燃油测量误差评估；燃油测量系统试验

中图分类号：V228.1; TP391.9

文献标志码：A　　**文章编号**：1001-5965(2026)05-1467-09

飞机燃油测量指示系统应能在经批准的飞机使用包线范围内始终提供准确和可靠的剩余可用燃油指示[1]。如果飞机的燃油测量系统误差较大，则会影响飞机的经济性[2]。

飞机的燃油测量系统由安装在燃油箱中的多个传感器组成，均采用传感器测量得到燃油高度，使用高度解算燃油体积，将燃油体积与密度相乘得到燃油质量[1]。

在实际使用中，飞机燃油箱内的燃油高度与燃油体积成非线性关系[3]。工程上通常采用多维查表插值法，根据传感器实时测量高度、飞行姿态、机体加速度共同确定的当前燃油平面，在预先确定的燃油质量特性数据库中插值求得剩余燃油量[4-6]。因此，飞机燃油测量工作主要包括:建立各油箱的燃油质量特性数据库；根据传感器和姿态等信息确定燃油平面，通过多维查表插值法实现燃油体积计算[7]。

初期燃油质量特性数据库按照实际传感器的位置建立，需要对油箱内每根传感器都建立相应表格，存在数据库规模庞大的问题[8-9]。文献[10]在多维查表插值法基础上提出等效传感器的方法，将3根传感器的输出融合到1根特定位置的虚拟传感器上，实现了数据量减少2/3，但该方法局限于3根传感器，且要求在最大油面俯仰和横滚姿态下所有传感器都具有有效的输出，同时数据库与传感器的位置相关，当传感器的布局变化时，数据库需要重新建立。文献[11]提出在沿着油箱高度各截面几何形心拟合直线位置处，假定安装虚拟传感器进行数据库建立和燃油解算，而虚拟传感器位置的选取成为影响燃油体积解算精度的主要因素，并针对不同油箱均需进行截面几何形心分析和计算。文献[12]提出一种基于等效传感器的自适应步长切割法建立燃油质量特性数据库，等效传感器位置与实际传

收稿日期：2024-02-28；录用日期：2024-04-19；网络出版时间：2024-05-14 12:11
网络出版地址：link.cnki.net/urlid/11.2625.V.20240513.1217.001
*通信作者. E-mail：renhyool@avic.com

引用格式：王晓阳，任恒英，饶波. 基于流体单位质量力的飞机燃油测量算法[J]. 北京航空航天大学学报，2026，52(5)：1467-1475.
WANG X Y，REN H Y，RAO B. Aircraft fuel quantity measurement algorithm based on fluid unit mass force[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1467-1475 (in Chinese).

感器位置独立，但各等效传感器基准点不统一，当单个测量单元传感器全浸没无效时无法有效实现系统重构；同时该文献提出利用最小二乘 (least mean squares, LMS) 拟合 3 根以上油量传感器输出值，或通过飞机姿态和加速度解算得到等效燃油平面姿态角的方法确定燃油平面（以下简称加速度航姿法），但 LMS 要求有效传感器数量不少于 3 根，系统代价较大，加速度航姿法不必要环节较多增加了误差源。

本文提出一种基于流体单位质量力的飞机燃油测量算法（aircraft fuel quantity measurement algorithm based on fluid unit mass force, AFUMF），在机体坐标系下采用飞机三轴线加速度计的测量值确定油箱内的油面单位法向量在 x 轴和 y 轴的分量以确定有效油面姿态，结合油量测量传感器与油面的交点坐标得到油面距离原点的有向距离以确定油面高度，通过油面姿态和油面高度检索燃油测量数据库进行燃油体积解算，减少了油面解算环节，降低了燃油体积解算误差。基于 CATIA 二次开发建立了基于蒙特卡罗方法的油面算法评估程序，对 LMS 和 AFUMF 开展误差评估，结果表明，AFUMF 在有效油量传感器数量和测量误差方面更有优势。建立飞机燃油箱试验件并在六自由度摇摆台环境下验证了本文算法的正确性。

1 测量算法推导

1.1 坐标系介绍

本文涉及的坐标系共 4 个，其中：铅垂地面固定坐标系 $(S'_gO'_gx'_gy'_gz'_g)$、飞机牵连铅垂地面坐标系 $(S_gO_gx_gy_gz_g)$ 和机体坐标系 $(S_bO_bx_by_bz_b)$[13] 主要用于说明三轴线加速度计输出值、机体加速度和飞机姿态之间的关系。民用飞机机体坐标系 $(S_cO_cx_cy_cz_c)$ 主要用于燃油测量数据库建立和测量解算以减少坐标系转换环节，其原点 O_c 位于机头前水平基准面与飞机对称面的交线上且距机头某一确定位置处，纵轴 x_c 为飞机水平基准面与对称面的交线并指向航向后方，横轴 y_c 位于飞机水平基准面内垂直于纵轴并指向航向右方，竖轴 z_c 位于飞机对称面内垂直于横轴并指向上方[14]。

1.2 假设及前提条件

本文算法基于以下假设或前提条件：

1）燃油为不可压缩的连续流体。

2）飞机稳定状态下油箱内的燃油处于平衡流体状态，不考虑燃油因飞机状态改变导致的反复晃动、表面波动及燃油在结构边界反射导致的喷溅等因素影响；不考虑燃油表面张力对液面的影响。

3）假定飞机为刚体，在飞机静置、匀速直线飞行、稳定加速或稳定盘旋条件下，全机所有位置加速度相同。

4）加速度计壳体刚性固定在飞机对称面上，且纵向位置邻近飞机重心。

5）在选定作为测量单元的燃油箱区域内，燃油流动不受任何限制，即飞机状态变化之后燃油可立即或在一个测量周期内稳定至最新的平衡状态。

1.3 平衡流体

当飞机处于静置、匀速直线飞行、稳定加速或稳定盘旋条件下，燃油箱内的燃油相对机体静止，处于平衡流体状态。

流体的平衡状态定义为：流体的每个质点都处于静止状态，或者相互间都没有相对运动；前者整个流体系统都不存在加速度，后者则整个流体系统都可以有加速度[15]。

平衡流体的平衡微分方程（欧拉平衡方程）为

$$\begin{cases} F_x-\dfrac{1}{\rho}\cdot\dfrac{\partial p}{\partial F_x}=0 \\ F_y-\dfrac{1}{\rho}\cdot\dfrac{\partial p}{\partial F_x}=0 \\ F_z-\dfrac{1}{\rho}\cdot\dfrac{\partial p}{\partial F_x}=0 \end{cases} \tag{1}$$

式中：F_x、F_y、F_z 分别为流体所受到的单位质量力（加速度）在 x、y、z 轴的分量；p 为流体压强；ρ 为流体密度。

由式（1）可得到等压面方程[16]：

$$F_x\mathrm{d}x+F_y\mathrm{d}y+F_z\mathrm{d}z=0 \tag{2}$$

根据流体的等压面方程可知，在平衡的流体中，通过每个质点的等压面必与该点所受到的单位质量力垂直，燃油箱内燃油的等压面即为与油平面相平行的一系列平面，因此，燃油质点所受到的单位质量力的方向即为油面的法向量方向。

1.4 加速度计

三轴线加速度计的敏感轴处于机体的 3 个轴向，可测量飞机 3 个轴向的线加速度[17]。加速度计通过弹簧力或电磁力使其内部的质量块相对壳体保持静止，从而根据其输出信号得到作用在质量块上的外力 $\boldsymbol{R}$，进一步得到质量块所受到的单位质量力 $\boldsymbol{b}$，其值为飞机机体或加速度壳体相对铅垂地面固定坐标系的加速度 $\boldsymbol{a}$ 与重力加速度 $\boldsymbol{g}$ 在测量轴上的差值[18]。得到如下关系式：

$$\begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix}=\begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix}-\begin{bmatrix} -\sin\theta \\ \cos\theta\sin\Phi \\ \cos\theta\cos\Phi \end{bmatrix}\boldsymbol{g} \tag{3}$$

式中：b_x、b_y、b_z 为三轴线加速度计直接测量的单位

质量力分量; a_x、a_y、a_z 为机体加速度 $\boldsymbol{a}$ 在机体坐标系各轴的分量; θ、Φ 分别为 S_b 相对于 S_g 的俯仰角和横滚角。

文献 [12, 19-20] 均提出通过坐标旋转矩阵将飞机机体加速度 $\boldsymbol{a}$ 与重力加速度 $\boldsymbol{g}$ 在机体坐标系下进行合并解算得到流体的单位质量力 $\boldsymbol{b}$, 以确定油面相对机体的俯仰和横滚姿态。由式(3)可知, 实际情况为飞机通过三轴线加速度计直接测量得到单位质量力 $\boldsymbol{b}$, 之后根据陀螺仪测量得到的飞机姿态角解算得到机体加速度 $\boldsymbol{a}$, 直接使用三轴线加速度计的测量值即可确定燃油质点所受到的单位质量力, 进而确定油面姿态, 无需进行额外的坐标系转换。

1.5　油面方程推导

为便于测量解算和燃油测量数据库建立, 在民用飞机机体坐标系 S_c 下开展燃油测量算法推导, 加速度计各轴方向应与坐标系各轴方向一致, 若相反则将对应加速度数值取相反数。

燃油平面的点法式方程为

$$A(x-x_0)+B(y-y_0)+C(z-z_0)=0 \tag{4}$$

式中: (A, B, C) 为燃油平面的法向量; (x_0, y_0, z_0) 为平面上的点。

飞机机体在稳定状态时为刚体, 燃油箱刚性固定在机体上, 其内部的燃油流体相对机体坐标系静止, 则全机所有流体质点所受到的单位质量力方向均相同(大小随流体深度变化)。加速度计壳体刚性固定在机体上, 加速度计内部的质量块相对机体坐标系静止, 则加速度计的质量块所受到的单位质量力与全机流体所受到的单位质量力方向相同, 因此, 可将三轴加速度计的加速度信号数值 $\boldsymbol{a}_m=(a_{mx}, a_{my}, a_{mz})$ 直接代入式 (4) 并转换为一般式方程:

$$a_{mx}x+a_{my}y+a_{mz}z-(a_{mx}x_0+a_{my}y_0+a_{mz}z_0)=0 \tag{5}$$

通常一个测量单元内会有多个有效传感器(设为 n), 已知各有效传感器顶点坐标 (x_{Hi}, y_{Hi}, z_{Hi}) 和底点坐标 (x_{Li}, y_{Li}, z_{Li}), 可通过式 (6) 将各有效传感器的高度 H_i 转换为坐标点 (x_i, y_i, z_i):

$$\begin{cases} t_i=H_i/\sqrt{(x_{Hi}-x_{Li})^2+(y_{Hi}-y_{Li})^2+(z_{Hi}-z_{Li})^2} \\ x_i=x_{Li}+t_i(x_{Hi}-x_{Li}) \\ y_i=y_{Li}+t_i(y_{Hi}-y_{Li}) \\ z_i=z_{Li}+t_i(z_{Hi}-z_{Li}) \end{cases} \tag{6}$$

式中: $i=1, 2, \cdots, n$。

理论上所有传感器与油面的交点均应位于同一平面, 但由于传感器误差、安装误差、数据采样不同步、燃油波动等因素, 各传感器的高度读数解算得到的点处于相互平行的多个平面上, 通过式 (7) 对多根传感器确定的式 (6) 取均值得到“居中”平面作为油平面:

$$a_{mx}x+a_{my}y+a_{mz}z+\frac{\sum_{i=1}^{n}-(a_{mx}x_i+a_{my}y_i+a_{mz}z_i)}{n}=0 \tag{7}$$

根据油平面方程可知, 为实现测量单元的燃油体积解算, 需保证加速度计数据和至少 1 根传感器有效。将平面方程的向量单位化, 从而实现对油量数据库的检索。

令

$$|\boldsymbol{a}_m|=\sqrt{a_{mx}^2+a_{my}^2+a_{mz}^2} \tag{8}$$

将油面方程简写为

$$a_{mx0}x+a_{my0}y+a_{mz0}z+D_0=0 \tag{9}$$

式中: a_{mx0}、a_{my0}、a_{mz0} 分别为油面单位法向量 $\boldsymbol{a}_{m0}$ 在纵轴 x_c、横轴 y_c、竖轴 z_c 的分量; D_0 为油平面距原点 O_c 的有向距离。由于平面有 2 个相反的单位法向量, 为防止测量数据库软件与机上软件出现油面方向相反的情况并简化编程, 本文将 a_{mz0} 的数值强制定义为正值, 如下:

$$a_{mx0}=\begin{cases} a_{mx}/|\boldsymbol{a}_m| & a_{mz}\geqslant 0 \\ -a_{mx}/|\boldsymbol{a}_m| & a_{mz}<0 \end{cases} \tag{10}$$

$$a_{my0}=\begin{cases} a_{my}/|\boldsymbol{a}_m| & a_{mz}\geqslant 0 \\ -a_{my}/|\boldsymbol{a}_m| & a_{mz}<0 \end{cases} \tag{11}$$

$$a_{mz0}=\begin{cases} a_{mz}/|\boldsymbol{a}_m| & a_{mz}\geqslant 0 \\ -a_{mz}/|\boldsymbol{a}_m| & a_{mz}<0 \end{cases} \tag{12}$$

$$D_0=\sum_{i=1}^{n}-(a_{mx0}x_i+a_{my0}y_i+a_{mz0}z_i)/n \tag{13}$$

图 1 为油面单位法向量与姿态角关系示意。a_{mx0} 为单位法向量 $\boldsymbol{a}_{m0}$ 与 $O_cy_cz_c$ 平面之间夹角 θ_f 的正弦值, 与油面俯仰姿态相关。a_{my0} 为单位法向量

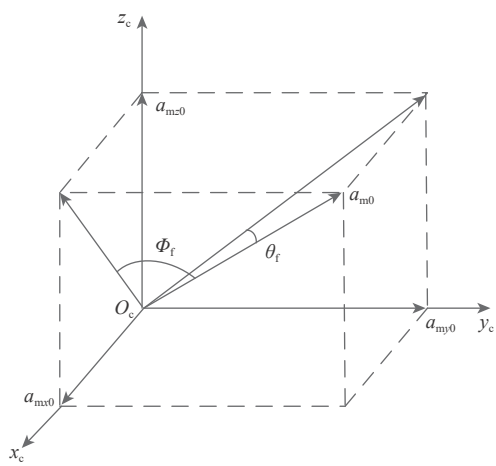


图 1　油面单位法向量与姿态角关系

Fig. 1　Relationship between unit normal vector and attitude angles of fuel surface

$\boldsymbol{a}_{m0}$ 与 $O_c x_c z_c$ 平面之间夹角 Φ_f 的正弦值，与油面横滚姿态相关。本文定义的 θ_f、Φ_f 仅用于根据以上对应关系得到测量范围内 a_{mx0} 和 a_{my0} 的最小值与最大值。

D_0 可视为垂直于燃油面的等效传感器，其最小和最大值不固定，而是取决于各油面姿态下测量单元轮廓与原点的最小和最大有向距离。最小值 D_{0min} 对应最大燃油体积，最大值 D_{0max} 对应最小燃油体积。

2 测量数据库建立

为减少解算环节和误差源，直接采用油面单位法向量建立燃油测量数据库。根据飞机的姿态包线和机体加速度范围确定燃油测量系统需要覆盖的有效油面姿态角范围，包括油面相对机体的俯仰角 θ_f 和横滚角 Φ_f，求正弦值即可分别得到 a_{mx0} 和 a_{my0} 的最小值和最大值。

燃油测量数据库的建立过程如下：

1）对油箱测量单元的CATIA模型预处理获得油箱载油空腔模型[20-21]。

2）a_{mx0} 和 a_{my0} 分别按均步长或变步长遍历油面姿态范围，两两组合确定各油面姿态。

3）对于每个 a_{mx0} 和 a_{my0} 组合，根据式(14)可得到对应的 a_{mz0} 值，该值在测量数据库建立和燃油体积解算时恒为正值。

$$a_{mz0}=\sqrt{1-a_{mx0}^2-a_{my0}^2} \tag{14}$$

4）将 a_{mx0}、a_{my0}、a_{mz0}、和0分别赋值给CATIA的方程平面的 A、B、C、D 参数建立基准面，CATIA的方程平面参数与一般式的 D 值相反，如式(15)所示：

$$Ax+By+Cz=D \tag{15}$$

5）基于基准面，通过平面平移建立切割面对油箱载油空腔模型按所需距离步长（均步长或根据横截面积变化率动态调整步长）进行遍历切割（SPLIT功能）操作，利用CATIA的测量功能同步测量得到各切割平面下方的燃油体积、重心坐标和对应的 D_0 值，以及该姿态下的 D_{0min} 和 D_{0max}，并存储至数据库，数据库可以是数据表或多项式系数。

燃油测量数据库与测量传感器的布局相互独立，更改传感器布局无需重新建立数据库，解决了文献[10-11]存在的问题。

3 燃油测量解算

燃油测量解算需要在机上存储测量单元内各油量传感器的顶点、底点坐标及燃油测量数据库。

图2为测量传感器有效性示意，测量传感器只有被燃油面部分浸没时才是有效的，未浸没或全部浸没均无法解算得到正确油量。

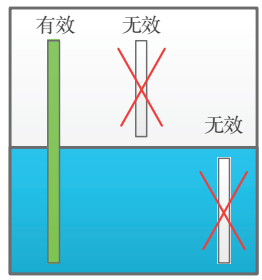


图2 测量传感器有效性示意图

Fig. 2 Schematic diagram of validity of fuel probe

图3为燃油测量解算流程示意，先采集加速度计的输出信号，根据式(10)～式(12)解算得到油面单位法向量 $(a_{mx0}, a_{my0}, a_{mz0})$，结合各有效油量传感器的顶点、底点坐标和高度值根据式(6)解算得到油面上至少1个点坐标，根据式(13)解算得到 D_0。使用 a_{mx0} 和 a_{my0} 检索燃油测量数据库，确定燃油体积 V 与 D_0 的对应关系数据，采用数据表插值或求解多项式解算得到燃油体积。

由于燃油测量数据库的姿态为按照姿态步长

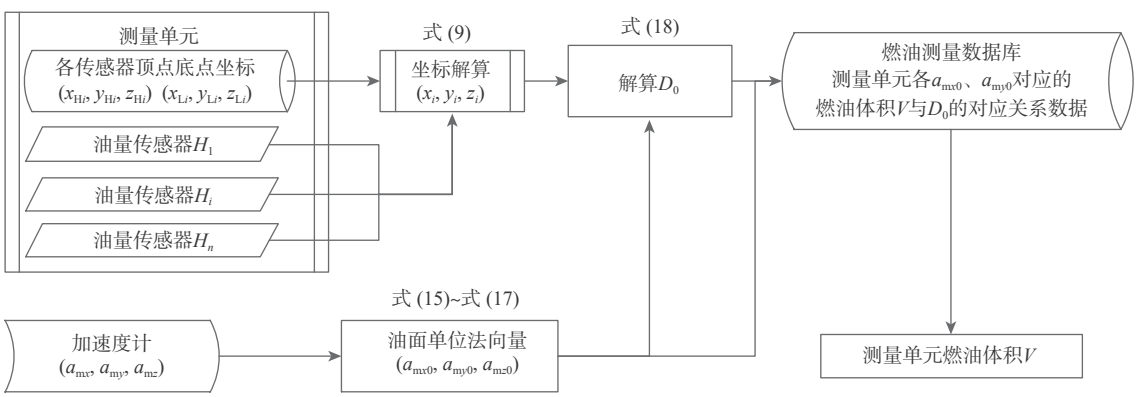


图3 燃油测量解算流程示意图

Fig. 3 Schematic diagram of fuel gauging process

存储的离散点, 实际解算的油面姿态正好与数据库存储姿态重合的概率为 0, 理论上如果存储足够多的姿态数据, 可通过将姿态四舍五入至最近的测量数据库姿态, 但该方法将占用大量的存储空间。通过解算实际油面姿态邻近的 4 个姿态的燃油体积进行双线性插值能够得到较好的解算结果。

4　燃油测量算法对比

4.1　解算环节对比

图 4 为 AFUMF 与加速度航姿法的解算环节对比。可知三轴线加速度计与陀螺仪的输出信号解算得到机体加速度, 之后加速度航姿法根据机体加速度和飞机姿态再次解算得到三轴线加速度的测量值, 进一步解算得到油面相对机体的俯仰和横滚角, 检索燃油测量数据库解算得到燃油体积。该方法解算环节较多, 且飞机姿态必须按照先俯仰再滚转的顺序进行坐标系旋转, 油面姿态角需确保与燃油测量数据库的坐标系旋转顺序一致, 否则将会导致解算油面姿态错误。

AFUMF 直接采用三轴线加速度计的输出信号确定油面单位法向量, 通过向量检索燃油测量数据库解算得到燃油体积, 省略了陀螺仪 2 次参与油面姿态解算, 以及油面向量转换为角度的环节, 算法简单且减少了误差源。

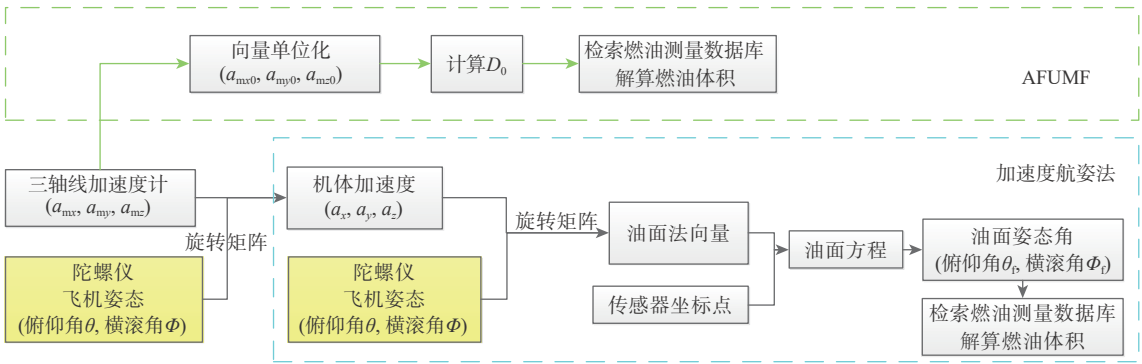


图 4　AFUMF 与加速度航姿法解算环节对比

Fig. 4　Comparison between AFUMF and algorithm of attitude combined acceleration

4.2　等效传感器对比

由于所有测量单元的 D_0 值均相对机体坐标系原点, 当单个测量单元的所有传感器均被燃油浸没无法解算燃油体积时, 可直接使用相邻存在自主串油的测量单元的 D_0 值进行燃油体积解算。已有的虚拟/等效传感器方法由于每个测量单元单独设置基准, 无法有效达到该效果。

图 5 为典型油面姿态下测量解算示意, 图中共 3 个测量单元, 每个测量单元各安装 1 根油量测量传感器, 3 个测量单元之间存在底部单向活门和顶部串油通气孔进行自主串油。

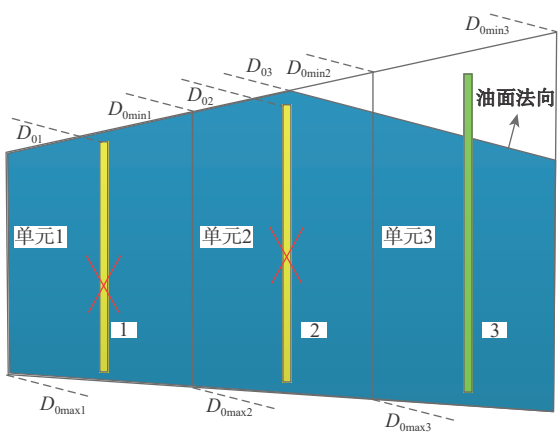


图 5　典型油面姿态测量解算示意图

Fig. 5　Schematic diagram of fuel quantity solution for typical fuel surface attitude

在当前油面状态下, 单元 1 处于满容积状态, 传感器 1 全部浸没无效, 按照传感器顶端的 D_{01} 进行解算则得到偏低值, 若按满容积保护则会得到正确值。单元 2 处于高油面但未处于满容积, 传感器 2 全部浸没无效, 按照传感器顶端的 D_{02} 进行解算则得到偏低值, 若按满容积保护则会得到偏高值。单元 3 处于高油面但未满容积, 传感器 3 部分浸没有效, 按传感器确定的 D_{03} 进行解算则得到正确值。将单元 1 和单元 2 均采用 D_{03} 进行解算, 单元 1 将根据边界值保护逻辑按 D_{0min1} 进行解算得到正确值, 单元 2 将按 D_{03} 进行解算则得到正确值。该解算逻辑经试验室验证有效, 解决了文献 [12] 存在的无法有效实现系统重构的问题。

4.3　算法解算速度评估

目前常用的燃油面算法包括: 三点确定油面法[10]、LMS[12]、加速度航姿法[12] 及 AFUMF。当 LMS 的有效油量传感器数量为 3 时, 其与三点确定油面法等效。加速度航姿法与 AFUMF 主要是在解算环节数量上存在差异。因此, 对 LMS 和 AFUMF 进行评估可覆盖当前的燃油面算法。

当有效油量传感器数量不少于3时，可使用LMS平面拟合的方法来计算燃油平面方程，一般情况下，燃油平面不会与z轴平行，可设平面方程为[12]

$$Ax+By+z+D=0 \tag{16}$$

求解三元一次方程组即可确定油面方程[12]：

$$\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} y_i^2 & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i & n \end{bmatrix} \begin{bmatrix} A \\ B \\ D \end{bmatrix} = \begin{bmatrix} -\sum_{i=1}^{n} x_i z_i \\ -\sum_{i=1}^{n} y_i z_i \\ -\sum_{i=1}^{n} z_i \end{bmatrix} \tag{17}$$

式中：n为有效传感器的数量；x_i、y_i、z_i分别为各有效传感器与油面的交点坐标，其中，i=1, 2, …, n。

为解算燃油平面方程，LMS和AFUMF均需根据式(6)先解算各有效传感器与油面的交点坐标，该步骤的计算量相同。

LMS需要先计算式(17)方程组的各项数值，有效传感器数量越多则计算量越大，再求解方程组得到A、B、D数值，之后，将平面方程单位向量化，最后，计算反三角函数得到油面姿态角。

AFUMF通过式(10)～式(12)将加速度信号向量单位化，之后，通过式(13)求均值确定D_0值。

忽略计算量相同的步骤，二者之间的差异为：LMS需要求解式(17)的方程，AFUMF需要按式(13)求均值，AFUMF所需的计算量小于LMS。

4.4 算法误差评估

对LMS和AFUMF的解算误差进行评估，选择典型油箱内形模型作为评估算例，如图6所示，并在油面姿态范围内按照3根有效布置15根测量传感器(实线)用于LMS评估，按照1根有效布置5根测量传感器(虚线)用于AFUMF评估。

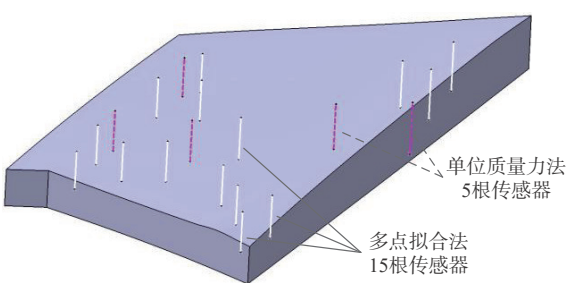


图6 算法评估算例油箱模型示意图

Fig. 6 Schematic diagram of case study of fuel tank model

在VB.NET环境下对CATIA二次开发，基于蒙特卡罗方法建立燃油面算法误差评估程序，包括基准数据建立模块和误差评估模块。

图7为油面算法误差评估流程，首先，通过基准数据建立模块在油面俯仰角为−30°～30°、横滚角为−15°～15°范围内按照1°步长确定油面姿态，按照10 mm等距步长对该姿态的燃油箱内形数模进行切割，利用CAITA的测量功能得到每次切割的燃油体积及对应的切割面与各传感器的交点坐标作为基准数据，切割工况共计133 375个，存储各工况的油面姿态角度、燃油体积、20根传感器交点坐标数据。

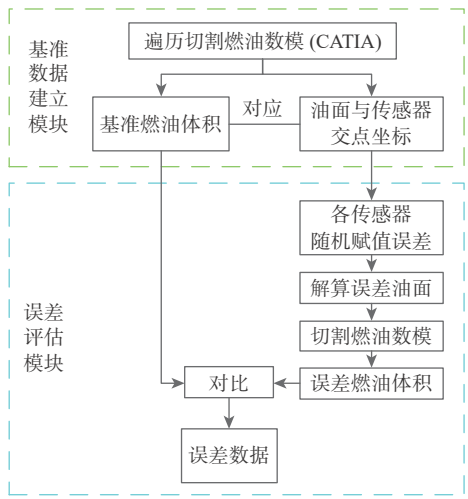


图7 油面算法误差评估流程

Fig. 7 Error evaluation process of fuel surface algorithm

通过误差评估模块对LMS和AFUMF分别分析，LMS与AFUMF分别忽略少于3根和1根有效传感器的工况，为每个工况的传感器按高斯分布随机赋值10次误差，解算得到误差油面，通过误差油面对燃油模型进行切割，测量误差燃油体积并减去基准燃油体积之后除以总燃油体积，得到每个工况随机10次的误差百分比并存储最大和最小值。按传感器规范赋值最大误差，油量传感器误差按0.5%(3σ)传感器长度赋值，加速度计误差按0.2%(3σ)量程赋值。

图8为误差分析结果，由图8(a)可知，LMS各工况的最大误差主要分布于0.5%满油量以内，但存在一些工况的最大误差接近1%满油量(仅4根传感器有效)，个别工况的最大误差接近2%满油量(仅3根传感器有效，等效三点确定油面法)。由于油量传感器最大误差为0.5%满油量，LMS对油量传感器的误差有放大作用。

由图8(b)可知，AFUMF各工况的最大误差主要分布于0.2%满油量以内，存在一些工况的最大误差接近但均未超出0.5%满油量。当多根油量传感器有效时会对各传感器确定的平行油面取均值，其最大误差不会超出单根传感器的最大误差，主要误差分布范围取决于加速度计的精度。

根据以上误差评估可得出以下结论：

1) LMS对油量传感器误差有放大作用，且需

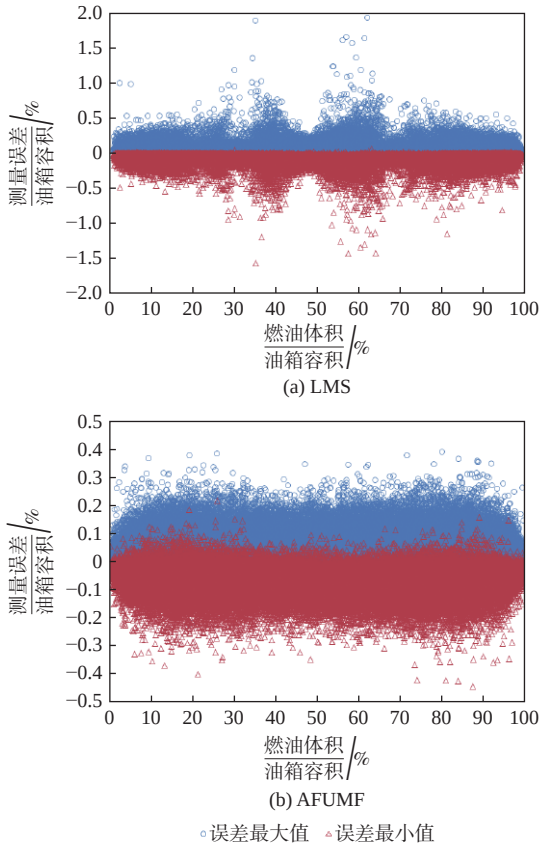


图 8　误差分析结果

Fig. 8　Error evaluation results

要至少 3 根传感器有效才可正常解算, 系统代价较大。

2) AFUMF 对油量传感器误差有抑制作用, 且仅需保证 1 根传感器有效即可正常解算, 系统代价较小, 此外, 加速度计为飞机必装设备, 未增加系统代价。

5　试验验证

为验证燃油测量算法的正确性, 建立飞机燃油箱试验件并在六自由度摇摆台环境下对测量算法进行验证。图 9 为燃油测量试验台示意, 台架处于负俯仰姿态。试验台架主要包括作动筒、试验台及试验油箱 3 部分。作动筒共 12 个, 分 3 组呈 120°与试验台对接。试验台与试验油箱之间通过安装支撑固连, 并在对接处安装三轴线加速度计, 其三轴与民用飞机机体坐标系重合, 最大误差为 0.2% 量程。试验油箱包括 4 个存在自主串油的测量单元, 分别安装 1, 2, 4, 4 根电容式油量测量传感器, 传感器最大误差为 0.5% 传感器长度, 同时在最外侧的测量单元内安装补偿传感器和密度传感器。在油箱加油管路上安装质量流量计对加注进油箱的燃油量进行测量并作为基准油量。

在 VB.NET 环境下对 CATIA 二次开发, 基于燃油箱试验件的制造模型建立燃油测量数据库, 按照

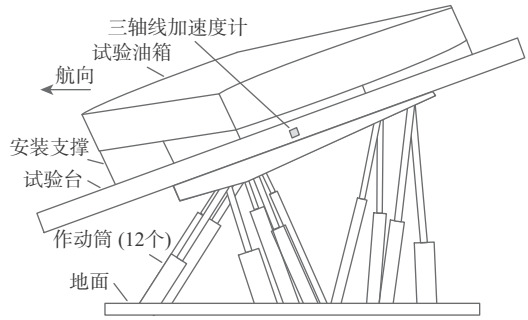


图 9　燃油测量试验台示意图

Fig. 9　Schematic diagram of fuel quantity measurement test rig

a_{mx0} 取值范围 [−0.5,0.5], a_{my0} 取值范围 [−0.26,0.26] 为 4 个测量单元分别建立燃油测量数据库, 当$|a_{mx0}|$和$|a_{my0}|$均不大于 0.035 时姿态步长为 0.001, 均不大于 0.11 时姿态步长为 0.005, 其余姿态下的姿态步长为 0.01, 每个测量单元的油面姿态数量为 11 665, 每个油面姿态的燃油体积 V 和 D_0 对应关系以 15 次多项式系数的形式进行存储, 多项式最大拟合误差为 0.07% 满油量。

采用 AFUMF 解算燃油量, 按以下流程开展试验:

1) 向油箱加注燃油并记录基准油量。

2) 六自由度摇摆台作动至所需的姿态并保持稳定, 记录燃油量解算结果。

3) 继续作动至下一姿态, 重复步骤 2 和步骤 3 直至试验结束。

部分试验结果如表 1 所示, 基准值为基准油量除以总油量, 指示值为稳定状态下与基准值偏差最大的测量数值除以总油量, 误差为指示值减去基准值。

表 1　燃油测量试验误差

Table 1　Error of fuel quantity measurement test

a_{mx0}	a_{my0}	基准值/%	指示值/%	误差/%
0	0	4.59	4.58	−0.01
−0.12	0	4.59	4.24	−0.35
−0.26	0.09	4.59	4.17	−0.42
0	0.26	4.59	4.47	−0.12
0	0	45.51	45.46	−0.05
−0.12	0	45.51	45.87	0.36
−0.26	−0.09	45.51	45.91	0.40
0.17	−0.09	45.51	45.03	−0.47
0	0	90.96	91.05	0.10
−0.12	0	90.96	90.73	−0.23
−0.17	0.09	90.96	91.09	0.14
0.17	0.09	90.96	91.28	0.32
0	−0.26	90.96	91.47	0.51
0	0.26	90.96	90.62	−0.34
0	0	100.00	100.03	0.03
−0.26	0.09	100.00	99.78	−0.22
0.17	0.09	100.00	100.09	0.09
0	−0.26	100.00	100.39	0.39
0	0.26	100.00	99.89	−0.11

燃油测量的动态台架试验的误差源包括燃油箱制造误差、传感器误差、传感器安装误差和燃油测量数据库解算误差，未包括油箱变形误差。试验结果表明在稳定状态下，AFUMF可以在油面姿态测量范围内将测量误差控制在±0.5%满油量范围内，小于文献[2]提出的最大误差要求：地面和巡航1%满油量，稳态爬升和下降2%满油量，其余持续状态3%满油量。

6 结 论

1）采用三轴加速度计信号直接确定油面单位法向量，并使用油面单位法向量代替油面姿态角建立燃油测量数据库，减少了油量解算环节和误差源。

2）各测量单元的燃油测量数据库与传感器布局相互独立，更改传感器布局无需重新建立数据库。

3）全机测量单元的基准为坐标系原点，便于测量单元内传感器全部浸没无效时采用存在相互自主串油的测量单元信息对系统进行重构。

4）基于CATIA二次开发建立油面算法评估程序，结果表明LMS对油量传感器误差有放大作用，AFUMF对油量传感器误差有抑制作用。

5）开展动态燃油箱台架试验对AFUMF进行验证，在各稳定姿态下大部分测量误差不大于±0.5%满油量，测量精度高于标准要求。

参考文献（References）

[1] SAE International. Guidance for the design and installation of fuel quantity indicating systems: AIR5691[R]. Warrendale: SAE International, 2017: 11-33.

[2] Airlines Electronic Engineering Committee. Guidance for the design and installation of fuel quantity systems: ARINC REPORT 611-1[R]. Annapolis: Aeronautical radio, Inc., 1999: 1-8.

[3] LANGTON R, CLARK C, HEWITT M. 飞机燃油系统[M]. 颜万亿, 译. 上海: 上海交通大学出版社, 2010: 137-176.
LANGTON R, CLARK C, HEWITT M. Aircraft fuel systems [M]. YAN W Y, translated. Shanghai: Shanghai Jiaotong University Press, 2010: 24-27(in Chinese).

[4] 袁梅, 林柯, 崔德刚. 飞机燃油油量测量及姿态误差修正方法[J]. 航空计测技术, 2001, 21(1): 24-26.
YUAN M, LIN K, CUI D G. Attitude error correction method of measurement of aircraft fuel[J]. Aviation Metrology & Measurement Technology, 2001, 21(1): 24-26(in Chinese).

[5] 金宇林, 庄达民, 杨钦. 民航客机油箱建模与油量传感器姿态误差修正[J]. 北京航空航天大学学报, 2005, 31(2): 218-222.
JIN Y L, ZHUANG D M, YANG Q. Modeling of airliner gasoline tank and modification of attitude error of sensor[J]. Journal of Beijing University of Aeronautics and Astronautics, 2005, 31(2): 218-222(in Chinese).

[6] 王向杨, 庄达民, 刘建民. 飞机燃油测量方法研究[J]. 飞机设计, 2004(1): 47-51.
WANG X Y, ZHUANG D M, LIU J M. A study of aircraft fuel quantity gauging method[J]. Aircraft Design, 2004(1): 47-51(in Chinese).

[7] 单宝峰, 张广涛, 李景春, 等. 航空油量测量技术研究及其发展现状[J]. 自动化仪表, 2013, 34(4): 32-33.
SHAN B F, ZHANG G T, LI J C, et al. Research on the measurement technology for aviation fuel volume and its current developing status[J]. Process Automation Instrumentation, 2013, 34(4): 32-33(in Chinese).

[8] GAO N, QU Z H. Modified particle swarm optimization based algorithm for BP neural network for measuring aircraft remaining fuel volume[C]//Proceedings of the 31st Chinese Control Conference. Piscataway: IEEE Press, 2012: 3398-3401.

[9] 段福宽, 揭裕文, 许应虎, 等. 飞机燃油系统油量计算与误差分析[J]. 南京航空航天大学学报, 2005, 37(6): 811-815.
DUAN F K, JIE Y W, XU Y H, et al. Aircraft fuel calculation and attitude error analysis[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2005, 37(6): 811-815(in Chinese).

[10] 欧阳平超, 刘红梅, 焦宗夏. 基于等效传感器和插值法的燃油量测量算法[J]. 北京航空航天大学学报, 2006, 32(8): 950-953.
OUYANG P C, LIU H M, JIAO Z X. Measurement arithmetic based on linear interpolation and equivalent transducer in aircraft fuel quantity gaging[J]. Journal of Beijing University of Aeronautics and Astronautics, 2006, 32(8): 950-953(in Chinese).

[11] 常伟, 苏三买, 王卉, 等. 基于虚拟传感器理论的飞机油量测量方法[J]. 计算机仿真, 2011, 28(12): 53-57.
CHANG W, SU S M, WANG H, et al. Measurement method of aircraft fuel quantity based on dummy sensor theory[J]. Computer Simulation, 2011, 28(12): 53-57(in Chinese).

[12] 袁梅, 何一强, 牛奔, 等. 变速变姿态下飞机燃油体积解算技术[J]. 北京航空航天大学学报, 2016, 42(9): 1827-1835.
YUAN M, HE Y Q, NIU B, et al. Aircraft fuel volume calculation technique in state of variable velocity and attitude[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(9): 1827-1835(in Chinese).

[13] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 飞行力学 概念、量和符号 第1部分: 坐标轴系和运动状态变量: GB/T 14410.1—2008[S]. 北京: 中国标准出版社, 2008: 299-300.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Flight mechanics-concepts quantities and symbols Part 1: axis systems and motion state variables: GB/T 14410.1—2008[S]. Beijing: Standards Press of China, 2008: 299-300(in Chinese) .

[14] 中华人民共和国国防科学技术工业委员会. 基于CATIA建模要求 第2部分: 坐标系: HB7756.2-2014[S]. 北京: 中国航空综合技术研究所, 2014: 1-2.
Commission of Science, Technology and Industry for National Defense of China. CATIA modeling requirements Part 2: coordinate axis system: HB7756.2-2014[S]. Beijing: AVIC China Aero-Polytechnology Establishment, 2014: 1-2(in Chinese).

[15] HUGHES W F, BRIGHTON J A. 流体动力学[M]. 徐燕侯, 过明道, 徐立功, 等, 译. 北京: 科学出版社, 2002: 12-14.
HUGHES W F, BRIGHTON J A. Schaum's outline of theory and problems of fluid dynamics[M]. XU Y H, GUO M D, XU L G, et

al, translated. Beijing: China Science Publishing & Media Ltd, 2002: 12-14(in Chinese).

[16] 蒋新生. 工程流体力学[M]. 重庆: 重庆大学出版社, 2017: 23-27.
JIANG X S. Engineering fluid mechanics[M]. Chongqing: Chongqing University Press, 2017: 23-27(in Chinese).

[17] 吴森堂, 费玉华. 飞行控制系统[M]. 北京: 北京航空航天大学出版社, 2005: 122-125.
WU S T, FEI Y H. Flight control[M]. Beijing: Beijing University of Aeronautics & Astronautics Press, 2005: 122-125(in Chinese).

[18] RUDOLF B. 飞行控制[M]. 金长江, 肖业伦, 译. 北京: 国防工业出版社, 1999: 35-227.
RUDOLF B. Flight control[M]. JIN C J, XIAO Y L, translated. Beijing: National Defense Industry Press, 1999: 35-227(in Chinese).

[19] 苏三买, 常伟, 王卉. 飞机油箱燃油量体积特性计算与油量测量仿真[J]. 计算机测量与控制, 2011, 19(9): 2091-2094.
SU S M, CHANG W, WANG H. Aircraft fuel tank volume characteristic database establishing and fuel quantity measurement simulation[J]. Computer Measurement & Control, 2011, 19(9): 2091-2094 (in Chinese).

[20] 宋伟, 钟若瑛. 飞机燃油质量特性计算的截面自适应分割法[J]. 北京航空航天大学学报, 2020, 46(3): 505-514.
SONG W, ZHONG R Y. Adaptive cross-section segmentation method for aircraft fuel mass properties calculation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(3): 505-514(in Chinese).

[21] 苏三买, 陈金平, 常伟. 采用 CATIA 的飞机燃油质量特性计算方法[J]. 计算机仿真, 2015, 32(10): 22-26.
SU S M, CHEN J P, CHANG W. Calculate algorithm of aircraft fuel mass property based on CATIA[J]. Computer Simulation, 2015, 32(10): 22-26(in Chinese).

Aircraft fuel quantity measurement algorithm based on fluid unit mass force

WANG Xiaoyang，REN Hengying*，RAO Bo

(Institute of Mechanical and Electrical System Design，AVIC The First Aircraft Institute，Xi'an 710089，China)

Abstract: In order to accurately measure the remaining fuel in the aircraft fuel tank, the aircraft fuel quantity measurement algorithm based on fluid unit mass force (AFUMF) is proposed. The unit normal vector of the fuel surface in the aircraft-body coordinate frame can be directly determined using the three-axis accelerometer's output signals as the equivalent of fluid unit mass force. The fuel quantity database can then be created and utilized for the fuel volume calculation by combing the point coordinates obtained from the fuel probes' height signal to determine the directed distance from origin as the fuel surface height. A CATIA secondary development program is developed to evaluate the error distribution of least mean square (LMS) and AFUMF by using the Monte-Carlo analysis method. The result shows AFUMF will alleviate the fuel probes' error and only 1 valid probe is needed, on the contrary, LMS will amplify the fuel probes' error and at least 3 valid probes are required. Using a six-degree-of-freedom platform, a fuel tank test item was constructed to validate AFUMF. The test findings indicate that, under steady conditions, the maximum error is ±0.5% full fuel tank quantity.

Keywords: aircraft fuel quantity measurement; fuel quantity measurement database establishing; fuel volume calculation; error evaluation of fuel quantity measurement; fuel quantity measurement system test

Received：2024-02-28；**Accepted**：2024-04-19；**Published Online**：2024-05-14 12：11
URL：link.cnki.net/urlid/11.2625.V.20240513.1217.001

* **Corresponding author**. E-mail：renhyool@avic.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0180

加速栅小孔对变推力离子推力器性能影响的实验研究

胡竟*，郭德洲，耿海，杨福全，李建鹏，陈娟娟
(兰州空间技术物理研究所 真空技术与物理重点实验室，兰州 730000)

摘 要：离子推力器以其比冲高、寿命长的突出优势，在长寿命航天器的姿轨控任务中得到广泛应用。以实现大范围、高精度推力调节为目标，针对加速栅小孔对连续变推力模式离子推力器的性能影响开展了综合实验研究，分析加速栅小孔补偿设计的机理与影响因素，基于10 cm口径连续变推力离子推力器，设计2种加速栅小孔中心距补偿系数的栅极系统，在不同推力模式下研究加速栅小孔中心距补偿系数对推力器放电室内等离子体宽范围放电特性、电子反流极限电压与加速栅电流、离子束流分布特性及离子溅射刻蚀效应的影响作用关系。以此为基础，开展多时间尺度下变推力模式对加速栅溅射刻蚀的影响评价。结果表明：在总加速电压恒定的情况下，改变加速栅小孔中心距补偿系数可调节放电室的基本性能和束流离子的密度分布，且推力越小，补偿系数变化的影响权重越明显；加速栅小孔中心距补偿系数的变化不会明显改变栅极系统的电子返流极限电压，但会显著影响宽范围变推力工作点下束流离子对栅极系统的离子溅射刻蚀效应，对此可通过调整总加速电压予以改善。研究结果为变推力离子推力器栅极系统匹配性设计及高性能变推力控制策略优化提供了技术支撑。

关 键 词：加速栅小孔中心距补偿系数；变推力；离子推力器；溅射刻蚀；放电损耗
中图分类号：V439+.4
文献标志码：A **文章编号**：1001-5965(2026)05-1476-11

离子推力器以其累计工作时间长[1]、比冲高[2-3]、推力连续可调[4-5]的突出优势在航天器姿态控制、轨道转移、无拖曳飞行及深空探测主推进任务中得到广泛应用，是国际上实现商业化批产应用的主要电推进产品之一。

为满足近地轨道高分辨率观测卫星精确维轨任务及重力梯度测量卫星无拖曳飞行任务的应用需求，离子推力器必须具备宽范围、连续变推力的工作能力，宽范围的功能模式要求离子推力器具有较宽的推力变比范围，卫星在轨可实现推力大跨步的输出调节；连续变推力的功能模式要求离子推力器能够不间断地实现推力精细调节(分辨率一般优于10 μN)，以保障卫星载荷的正常运行。基于以上应用需求，兰州空间技术物理研究所研制了具有宽范围连续变推力调节能力的10 cm口径Kaufman型离子推力器[6-8]，产品可在推力为1～20 mN范围内实现优于50 μN分辨率的推力调节响应[9]。

栅极系统是离子推力器的关键结构，其束流引出特性是决定离子推力器工作性能和服役寿命的关键因素之一，而束流引出特性又取决于栅极系统自身的结构设计及其与上游放电室内等离子体之间的匹配性，良好的栅极结构是离子推力器实现宽范围连续变推力的基础。为避免在力学环境中发生栅片的剧烈形变、破损及工作过程中出现不规则

收稿日期：2024-03-28；录用日期：2024-05-15；网络出版时间：2024-06-04 15:14
网络出版地址：link.cnki.net/urlid/11.2625.V.20240604.1120.004
基金项目：国家重点研发计划(2021YFC2202704)；国家自然科学基金(62371213)；甘肃省自然科学基金(22JR5RA787,22JR5RA789,22JR5RA784)
*通信作者. E-mail：hjing37615486@163.com

引用格式：胡竟，郭德洲，耿海，等. 加速栅小孔对变推力离子推力器性能影响的实验研究[J]. 北京航空航天大学学报，2026，52(5)：1476-1486. HU J，GUO D Z，GENG H，et al. Experimental study on effect of aperture of accelerator grid on performance of variable thrust ion thruster[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1476-1486 (in Chinese).

热形变和栅间短路的问题, 10 cm 及以上口径的离子推力器栅极系统都采用球面结构。相比凸球面栅极系统, 凹面栅极系统以其栅间距变化规律性佳、与工作环境温度匹配性好的显著优势成为宽范围连续变推力离子推力器的最优选择。实践应用中发现, 受栅片固有构型制约, 凹球面栅极系统所引离子束流容易出现过聚集现象, 在造成推力损失的同时会出现离子直接轰击并刻蚀加速栅小孔侧壁的情况, 从而加快栅极系统的磨损, 最终降低离子推力器的服役寿命。因此, 解决凹球面栅极系统的束流离子过聚集现象是宽范围连续变推力离子推力器实现在轨工程化应用必须解决的关键技术之一。

国内外学者就栅极系统的结构设计与工作特性等方面开展了大量研究。文献 [10-11] 建立了 IBEX-T 全粒子网格(particle-in-cell, PIC)法仿真模型, 可用于分析栅极系统束流引出的瞬间状态和交换电荷离子的分布。文献 [12] 采用 PIC-蒙特卡罗碰撞 (Monte Carlo collision, MCC) 方法建立了栅极系统的仿真模型, 可研究不同放电条件下加速栅电流与束电流的变化关系。文献 [13-14] 利用 iGun 程序代码构建了双栅和三栅组件的磨损模型。文献 [15] 以加速栅中心凹槽腐蚀深度数据为基础, 采用累积损伤理论建立了离子推力器多功率段工作模式下的加速栅寿命预测模型, 并对美国 NEXT-40 cm 离子推力器的加速栅寿命开展了预测。文献 [16] 采用三维粒子模拟方法分析了工作模式变化对栅极系统电场分布和束流状态的影响及离子推力器的多模式设计需求。文献 [17-19] 基于小天体探测任务对多模式离子推力器的应用需求, 开展了凸球面三栅变孔径栅极系统设计, 并确定了栅极系统在各工作点的工作参数。文献 [20] 提出并验证了利用减小屏栅边缘孔径消除束流离子对减速栅边缘孔溅射刻蚀的方法。文献 [21] 采用 PIC-MCC 方法构建了栅孔二维寿命预估模型, 分析了加速栅电压、屏栅孔径及加速栅与减速栅冷态间距等栅设计参数对凸球面三栅极系统寿命的影响。文献 [22] 采用实验方法对比研究了凹凸球曲面构型的栅极系统对推力器放电损耗、束流密度分布和低温启动特性的影响, 研究指出凹球面构型的栅极系统所引离子束流更容易聚焦, 且工作状态下的放电损耗相对较大。

上述研究工作主要针对单模式或多模式离子推力器栅极系统的分析, 且多聚焦于栅极系统宏观的结构尺寸、工作参数对推力器性能与寿命的影响作用关系。对于连续变推力模式的栅极系统, 特别是凹球面栅极系统仅开展了工作性能的对比分析, 并未据此进行更深入的机理分析与研究, 而涉及凹球面栅极系统的关键参数设计与分析更是鲜有相关报道。

本文以实现大范围、高精度推力调节为目标, 针对加速栅小孔对连续变推力模式离子推力器性能的影响开展了综合实验研究, 分析了加速栅小孔补偿设计的机理与影响因素, 基于 10 cm 口径连续变推力离子推力器, 设计 2 种加速栅小孔中心距补偿系数的栅极系统, 在不同推力模式下研究加速栅小孔中心距补偿系数对推力器放电室内等离子体宽范围放电特性、电子反流极限电压与加速栅电流、离子束流分布特性及离子溅射刻蚀效应的影响作用关系。以此为基础, 开展了多时间尺度下变推力模式对加速栅溅射刻蚀的影响评价。

1　加速栅小孔补偿原理及影响

凹球面构型下推力器离子束流的聚集现象会造成明显的推力损失[23]。凹球面栅极系统出现束流过聚集的原因主要体现在以下 2 个方面: ①凹面结构自身形状所产生的束流聚焦; ②凹面栅片加工过程中屏栅、加速栅小孔中心线的相对位移造成束流聚焦。图 1 为凹面栅孔束流引出时的聚焦示意图。

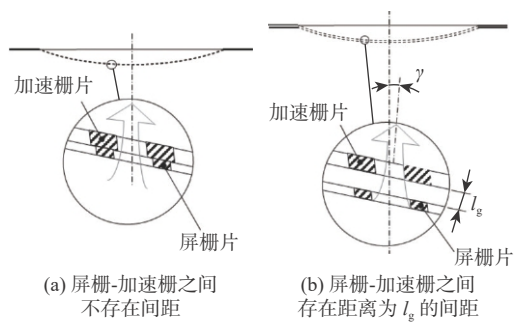


图 1　凹面栅孔束流引出时的聚焦示意图

Fig. 1　Schematic diagram of beam induced by concave grid hole

由图 1 可知, 当屏栅、加速栅之间不存在间距时, 栅片上单个小孔所引束流的流线与推力器的轴向相互平行; 而当屏栅、加速栅之间存在距离为 l_g 的间距时, 小孔所引束流相对推力器的轴向存在大小为 γ 的夹角, 整体已出现内敛的趋势。为引导栅片小孔束流的流线近似轴向, 在保证栅片间距的情况下需要对加速栅小孔中心距进行调整, 这种调整称为“补偿”。对于凹面构型的栅极系统, 补偿措施是将加速栅的小孔中心距相对屏栅适当缩小, 以此让每个小孔的束流沿平行于推力器轴线的方向发生偏转, 进而弥补凹面所产生的束流聚焦作用。

定义凹球面栅极系统的小孔中心距补偿系数 δ 表征加速栅小孔中心距的相对变化量, 减小后的加速栅小孔中心距是小孔中心距原始设计值的 $1-\delta$ 倍。一般存在以下经验关系[24]:

$$\delta=\frac{4kh^2l_g\Phi_h+2kh^2\Phi_h(D_b^2+h^2)(t_s+t_a)/(D_b^2-h^2)+2hd_a\left(D_b^2+h^2\right)\arcsin\left[2h\Phi_h/\left(D_b^2+h^2\right)\right]}{k\left(D_b^2+h^2\right)^2\arcsin\left[2h\Phi_h/\left(D_b^2+h^2\right)\right]} \tag{1}$$

式中：δ 为小孔中心距补偿系数；k 为经验系数；h 为凹面拱高；Φ_h 为凹球面直径；D_b 为凹球面上束口径直径；t_s 和 t_a 分别为屏栅片和加速栅片的厚度；d_a 为加速栅片小孔直径。可以看出，小孔中心距补偿系数 δ 与栅极系统的凹型拱高、栅极间距、屏栅与加速栅厚度及加速栅小孔的孔径等栅极系统的结构参数密切相关，同时也受经验系数取值的影响。在其他结构参数相同的情况下，加速栅小孔中心距会因经验系数取值的不同而产生差异。

2 实验系统和方法

采用兰州空间技术物理研究所研制的 10 cm 口径连续变推力离子推力器开展了凹球面栅极系统加速栅小孔中心距补偿系数对推力器工作性能影响的实验研究。在既定的推力器状态和栅片开孔口径方案下，设计形成了 2 种加速栅小孔中心距补偿系数的凹面栅极系统，除栅极系统外的离子推力器放电室、空心阴极等其他状态完全一致。对于 2 种栅极系统，加速栅片除小孔中心距因补偿系数不同而存在差异外其他结构参数完全一致，屏栅片的所有结构参数均完全一致；2 种栅极系统的屏栅片与加速栅片间距利用薄厚规控制在 0.7±0.02 mm 的范围内、屏栅片小孔与加速栅片小孔的对中同轴度利用高精度三维轮廓仪控制在 0.1 mm 范围内。其中，第 1 种栅极系统的加速栅小孔中心距补偿系数为 0.407%，第 2 种栅极系统的加速栅小孔中心距补偿系数为 0.204%。

推力器实验利用兰州空间技术物理研究所 TS-6S 离子推力器性能测试设备，设备为卧式结构，真空舱直径 φ=1.5 m，长度为 3 m，配备了 2 台对氮气抽速为 1×10^4 L/s 的外置式低温泵、2 台对氙气抽速为 1.49×10^4 L/s 的内置式低温泵作为抽气系统的主泵组，采用罗茨泵、螺杆泵及分子泵机组构成系统的粗抽泵组。系统空载抽真空 24 h 后的本底真空度优于 5.0×10^{-5} Pa，带载 1 mg/s 氙气时的工作真空度优于 1.0×10^{-3} Pa。设备氙气供给系统由高压氙气瓶、减压、稳压及质量流量控制器等部件组成，其中，质量流量控制器采用 ALICAT 公司的 M 系列高精度质量流量计，推力器阳极的流量控制精度为±(0.2% 满量程+0.4% 读数)、中和器与主阴极的流量控制精度为±(0.2% 满量程+0.8% 读数)。设备供电系统由 MAGNA 公司的 LS 系统电源组成，屏栅电源、阳极电源及励磁电源的电信号稳定度均为±0.01%。

表 1 所列为 10 cm 口径连续变推力离子推力器的实验工作条件。

表 1 推力器实验工作条件

Table 1 Experimental working conditions of thrusters

工作参数	参数配置	参数模式
屏栅电压	1 150 V	恒压输出
加速栅电压	−150 V	恒压输出
阳极电流	1.1 A(5 mN)、2.2 A(10 mN)、3.1 A(20 mN)、3.9 A(25 mN)	恒流输出
励磁电流	0.19 A(5 mN)、0.30 A(10 mN)、0.45 A(20 mN)、0.54 A(25 mN)	电流宽范围可调
中和器触持电流	1.4 A	恒流输出
主阴极触持电流	0.8 A	恒流输出
主阴极供气	0.08 mg/s	恒定流率输出
中和器供气	0.08 mg/s	恒定流率输出
阳极供气	0.15 mg/s(5 mN)、0.25 mg/s(10 mN)、0.50 mg/s(20 mN)、0.54 mg/s(25 mN)	不同推力点下恒定流率输出

注：括号内数值表示推力大小。

基于表 1 所列工作参数，在满足不同推力工作点供气条件的情况下监测恒压输出电源的电流值、恒流输出电源的电压值，并以此为基准核算推力器的功率、放电损耗等项目，以分析不同加速栅小孔中心距补偿系数下推力器工作性能的变化情况。放电损耗计算式为

$$\zeta=(I_dV_d+I_kV_k)/I_s \tag{2}$$

式中：ζ 为放电损耗；I_d 和 V_d 分别为放电电流和放电电压；I_k 和 V_k 分别为主阴极触持电流和主阴极触持电压；I_s 为屏栅电流。

为对比 2 种补偿系数下 10 cm 口径连续变推力离子推力器的束流密度分布变化情况，采用等离子体诊断装置对推力器径向不同位置的束流密度进行测试。等离子体诊断装置以收集盘直径为 1.1 cm 的单法拉第探针为基础，探针的束流密度测量范围为 0～50 mA/cm^2，测量精度优于±1%，用 500 Ω 的电阻进行采样，其安装在距离推力器加速栅下游表面 50 mm 处，并以 2 mm 的步进长度沿垂直于推力器轴线的方向水平扫描，每个步进点测试 100 个数值，通过去除最值求平均的方式获得该步进点位的测试结果。图 2 为等离子体诊断装置测试原理图。

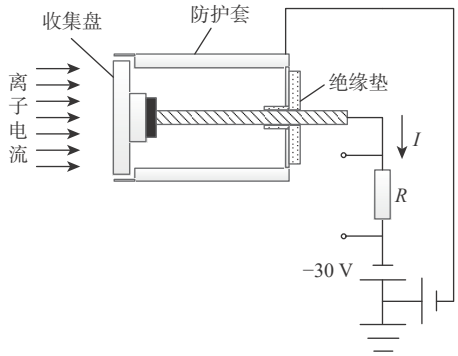


图 2　等离子体诊断装置测试原理图

Fig. 2　Test schematic diagram of plasma diagnostic device

为抑制电子和杂散离子对离子电流密度的干扰, 图 2 中收集盘、防护套均加载了−30 V 的偏置电压。结合 10 cm 口径连续变推力离子推力器的离子束流特性, 测试前对收集盘与防护套之间的间隙开展了优化, 确保间隙小于等离子体的鞘层厚度, 并对探针电流和偏置电压的稳定度与准确度进行了校核以消除不必要的测试误差。

3　结果和讨论

3.1　等离子体宽范围放电影响

离子推力器的放电损耗是评价推力器放电室内等离子放电性能的关键参数指标之一。放电损耗用以表征放电室产生单位离子束流所需消耗的放电功率, 数值越小表明放电损耗越低, 放电室的放电效率越高。对于连续变推力离子推力器而言, 各个推力工作点的束流功耗(束电流与束电压的乘积)是相对固定的, 由此推力器的放电损耗便决定了其总效率, 放电损耗越低推力器效率越高。图 3 为 2 种加速栅小孔中心距补偿系数的凹球面栅极系统在 10 cm 口径连续变推力离子推力器不同工作点的放电损耗及放电电压变化情况。

由图 3 可知, 10 cm 口径连续变推力离子推力器分别装配加速栅小孔中心距补偿系数 δ=0.204%、δ=0.407% 的凹球面栅极系统时放电室内等离子体的放电呈现出相类似的变化规律, 但性能状态存在明显差异。推力器在推力为 1～25 mN 范围内变推力调节过程中, 随着推力器输出推力的增大, 放电室的放电损耗呈现出快速下降并逐渐趋于稳定的变化特性, 而放电电压则以缓慢上升的形式趋于稳定。对于推力为 12 mN 以下的小推力工作点, 放电室的放电损耗在 2 种补偿系数状态下均处在较高水平, 分析认为该现象主要是由于为实现不同浓度工质气体的稳定电离放电, 放电室中电子在阳极壁面的损失比例在宽范围推力工作点下均保持在较高的水平, 并不会因推力小而出现本质性下降, 对

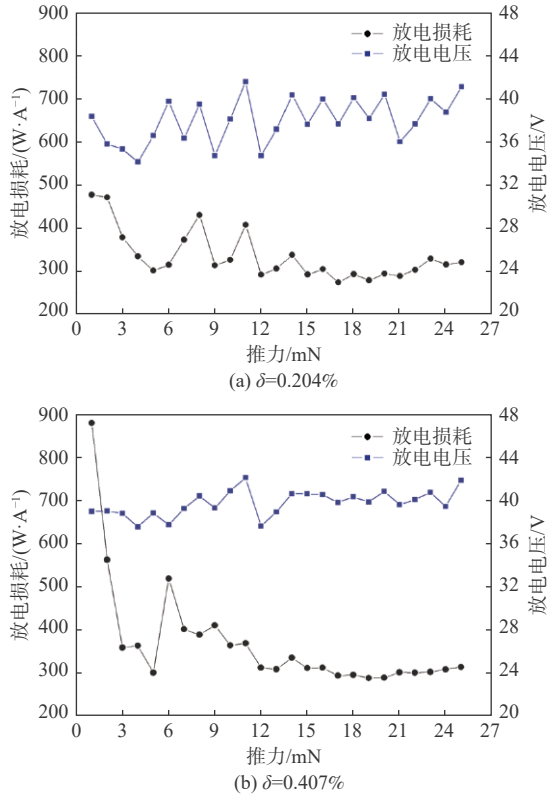


图 3　2 种加速栅小孔中心距补偿系数不同工作点放电损耗及放电电压变化情况

Fig. 3　Discharge loss and discharge voltage variation at different operating points under two accelerator grid compensation coefficients

于小推力工作点而言, 放电室内部工质气体利用率不高, 等离子体密度相对较低, 放电消耗的能量主要是维持原初电子的能级, 此时必然会出现较高的放电损耗。

加速栅小孔中心距补偿系数由 0.204% 增加到 0.407% 时, 推力器在推力为 1 mN 工作点的离子透明度下降约 7.14%。事实上, 由栅极系统小孔离子导流性能计算理论可知, 加速栅小孔中心距补偿系数的变化会改变栅极系统的离子透明度, 进而影响到离子引出效率, 并最终改变放电室的放电损耗。因此, 当补偿系数由 0.204% 增加到 0.407% 时, 推力器在推力为 1 mN 工作点的放电损耗由 477.2 W/A 上升至 868.7 W/A, 增大近 82%。此后, 随着放电室等离子密度的上升(推力器输出推力逐渐增大), 栅片小孔所引离子束流逐渐接近其最大束流引出能力, 加速栅小孔中心距补偿系数差异对栅极系统的离子透明度的影响权重逐渐降低, 使得推力器在推力为 12 mN 之后的推力工作点下的放电损耗与放电电压均趋于稳定。

3.2　电子反流极限电压及加速栅电流影响

电子反流极限电压和加速栅电流是衡量离子推力器工作可靠性与寿命的重要参量。加速栅电

压主要是为推力器提供足够强度的电势阱以防止中和器发射的电子反流到放电室内部，电子反流极限电压就是指额定工作条件下，电子通过栅极系统进入放电室的最小加速栅电压。

在连续变推力离子推力器的诸多失效模式中，加速栅孔壁磨损出现的电子反流失效和交换电荷离子对加速栅下游表面溅射磨损导致的加速栅结构失效已成为决定推力器在轨服役寿命的最重要失效模式[25]。因此，控制电子反流极限电压抑制电子返流失效、降低交换电荷离子碰撞加速栅形成的加速栅电流杜绝加速栅结构失效成为加速栅小孔中心距补偿系数影响研究的关键。

栅极系统小孔的束流强度可表示为

$$I_b = \frac{\pi\varepsilon_0}{9}\sqrt{\frac{2e}{M_i}}\frac{V_t^{3/2}d_s^2}{l_e^2} \tag{3}$$

式中：I_b 为单孔的束流强度；ε_0 为介电常数；e 为电子电荷；M_i 为离子质量；V_t 为总加速电压；d_s 为屏栅孔径；l_e 为有效加速长度，与栅极间距、屏栅厚度及屏栅孔径密切相关，其表达式为

$$l_e = \sqrt{(l_g + t_s)^2 + (d_s/2)^2} \tag{4}$$

可以看出，在保持其他条件不变的情况下，栅极系统的鞘发射离子束流与总加速电压之间遵守二分之三次方定律。理想情况下栅极系统鞘面发射的离子会全部通过加速场，不会被加速栅孔边缘所截获，而且通过加速栅小孔后的离子束流发散角最小。但当加速栅电压不足时，会使得鞘面形状与位置发生变化，最终改变离子束流的聚焦特性，导致加速栅电流的异常变化。图 4 为 2 种补偿系数下 10 cm 口径连续变推力离子推力器在推力分别为 5 mN、10 mN、20 mN 及 25 mN 工作点的离子束流、加速栅电流随加速栅电压的变化关系。

由图 4 可知，在 2 种补偿系数下加速栅电压调节过程中 10 cm 口径连续变推力离子推力器在不同工作点的离子束流、加速栅电流变化规律基本相同。总体而言，随着加速栅电压的下降，离子束流因总加速电压的降低而相应下降，但随着加速栅电压的进一步下降，加速栅电压在加速栅小孔区域构建的电势阱低于电子动能，造成电子反流进入放电室，从而产生离子束流上升的现象。2 种补偿系数的栅极系统在不同工作点的电子反流极限电压基本保持在 70～80 V 之间，并无明显差异。此外，加速栅小孔中心距补偿系数 δ=0.204% 下的栅极系统离子引出能力要优于补偿系数 δ=0.407% 的状态，其离子束流对加速栅电压的变化更为敏感，究其原因主要是相同的栅间距状态下，补偿系数 δ=0.204%

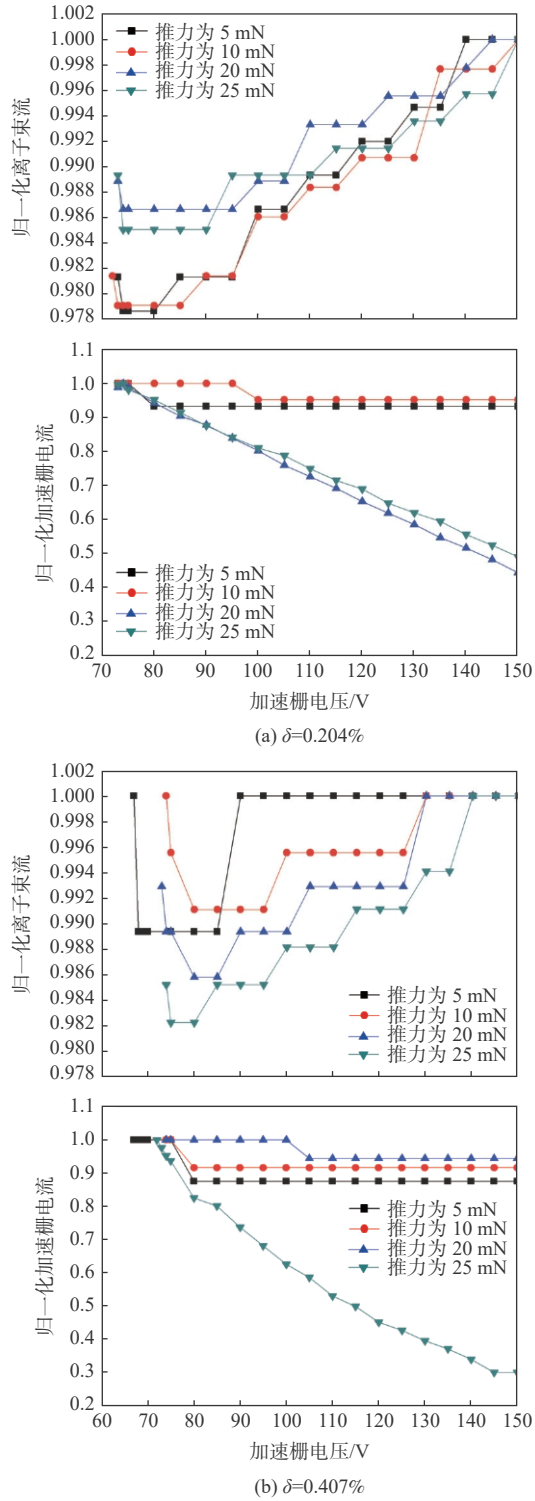


图 4 不同工作点下离子束流、加速栅电流随加速栅电压变化关系

Fig. 4 Relationship between ion beam current and acceleration current at different operating points and acceleration voltage

下的加速栅小孔与屏栅小孔的对中性优于 δ=0.407% 的状态，由此可以建立更优的离子加速电场构型，加速栅电压的变化可以更为灵敏的体现在电场构型的差异上，从而对离子束流产生更精细、微小的影响。

分析不同工作点下加速栅电压调节过程中加

速栅电流变化情况可以看出, 与加速栅小孔中心距补偿系数 δ=0.204% 的状态不同, 当补偿系数 δ=0.407% 时, 推力器在推力为 5～20 mN 较宽的推力范围内加速栅电流均保持在较高量级, 显然其对加速栅电压调节的敏感性不足, 由此降低了卫星通过调节加速栅电压以改善栅极系统工作寿命措施的有效性。事实上, 当加速栅小孔中心距补偿系数过大时, 其对离子的聚焦、引出特性已发生变化, 离子会直接轰击到加速栅的结构本体形成冲刷, 在形成较大量级加速栅电流的同时也加快了加速栅的溅射刻蚀速率, 最终降低了栅极系统的工作寿命。显然, 加速栅小孔中心距补偿系数 δ=0.204% 的栅极系统离子束流引出性能、工作寿命要优于补偿系数 δ=0.407% 时的栅极系统状态。

3.3　离子束流分布特性影响

离子束流分布特性反映的是离子推力器栅极系统整个开孔区域内离子束流分布的均匀程度, 常采用束流平直度来表示。束流平直度的计算式为[26]

$$f_b = j_{b.avg}/j_{b.max} \tag{5}$$

式中: f_b 为束流平直度; $j_{b.avg}$ 和 $j_{b.max}$ 分别为平均束流密度和束流密度最大值。显然, 束流平直度数值越大, 其均匀性越好, 束流密度在栅极系统径向的分布越倾向于扁平化。在束流大小相同的情况下, 束流平直度较低的推力器中心轴线区域就会存在较高的束流密度, 而边缘区域的束流密度将对较低。中心区域过高的离子束流密度在加快屏栅上游表面溅射磨损的同时对加速栅下游表面交换电荷离子的刻蚀效应同样起到助推作用, 严重制约了栅极系统的工作寿命。此外, 较低的束流平直度也会影响栅极系统的离子聚焦引出能力, 降低栅极小孔的最大可引出束流, 最终会压缩推力器的变推力调节范围。

图 5 给出了 2 种补偿系数状态下 10 cm 口径连续变推力离子推力器在推力分别为 5 mN、10 mN、20 mN 及 25 mN 工作点的束流密度分布归一化曲线。

分析图 5 可以看出, 10 cm 口径连续变推力离子推力器的束流密度沿径向呈现出近似平顶的分布形式, 束流分布较为均匀。既定的栅片开孔口径下, 加速栅小孔中心距补偿系数由 0.204% 增大到 0.407% 后, 束流密度分布曲线略微变窄变高, 束流均匀性稍有下降, 平直度随之降低。对于上述 4 个不同推力工作点, 2 种补偿系数状态下大推力状态的束流平直度下降幅度要小于小推力状态, 整体平均下降约 5%。结果表明, 加速栅小孔中心距补偿系数对变推力离子推力器的离子束流分布存在影响, 束流密度越小影响越大。对此主要是因为补偿系数由 0.204% 增大到 0.407% 后, 栅极系统的离子透明度降低, 离子引出效率变差。小推力工作点的离子束流密度远低于栅片小孔的最大束流引出能力, 在此情况下引出小孔的变化便对束流密度的分布产生相对较为明显的影响; 区别于此, 大推力工作点的离子束流密度已接近栅片小孔的最大引出能力, 离子束流引出效率趋于最大化, 加速栅小孔中心距补偿系数变化所产生的影响基本可以忽略不计, 使得离子束流密度分布的变化幅度明显弱于小推力器工作点。至此也进一步确认了 3.1 节中大推力工作点下放电室内放电损耗与放电电压趋于稳定的现象。

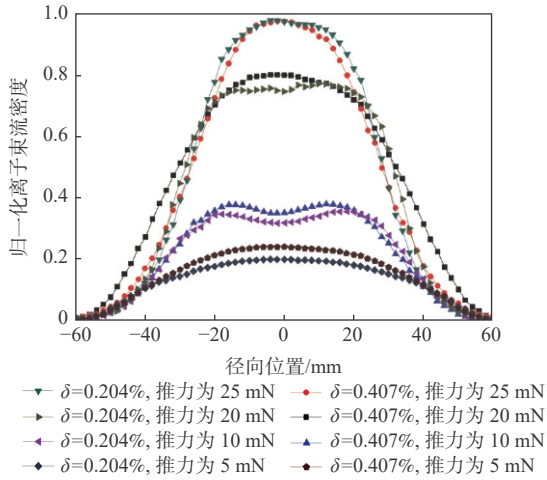


图 5　2 种补偿系数状态下推力器在不同工作点的束流密度分布归一化曲线

Fig. 5　Normalized curves of beam density distribution of thrusters at different operating points under two compensation coefficients

就离子推力器而言, 交换电荷离子密度与离子束流密度基本呈线性关系[27]。在既定的栅片开孔口径下, 加速栅小孔中心距补偿系数由 0.204% 增大到 0.407% 后, 束流密度平顶区域峰值的增加将导致加速栅下游交换电荷离子密度峰值相应增大。由此, 加速栅小孔中心距补偿系数增加后, 推力器在宽范围工作点下交换电荷离子的溅射腐蚀寿命均会下降, 平均降低约 5%。

3.4　工作效能与离子刻蚀效应影响

离子推力器的工作原理决定了其等离子体放电过程与束流离子聚焦引出过程之间的弱耦合性和相对分离性, 推力器工作过程中栅极系统会不可避免的受到高密度离子束流的快速冲刷而产生溅射刻蚀。为此, 本文研究了加速栅小孔中心距补偿系数对持续性离子溅射刻蚀效应的影响, 主要通过 200 h 短期磨损实验对 2 种加速栅小孔中心距补偿

系数下的加速栅下游表面及小孔直径的离子刻蚀影响进行了分析，考虑到20 mN工作点的要求工作时长占推力为1～25 mN宽范围工作点的90%以上，短期磨损实验主要在推力为20 mN工作点下开展，实验过程中监测了放电损耗、放电电压、加速栅电流及整机功耗等关键指标项目的变化情况。图6给出了2种补偿系数状态下10 cm口径连续变推力离子推力器在推力为20 mN工作点工作200 h时的放电损耗、放电电压、加速栅电流及整机功耗变化情况。

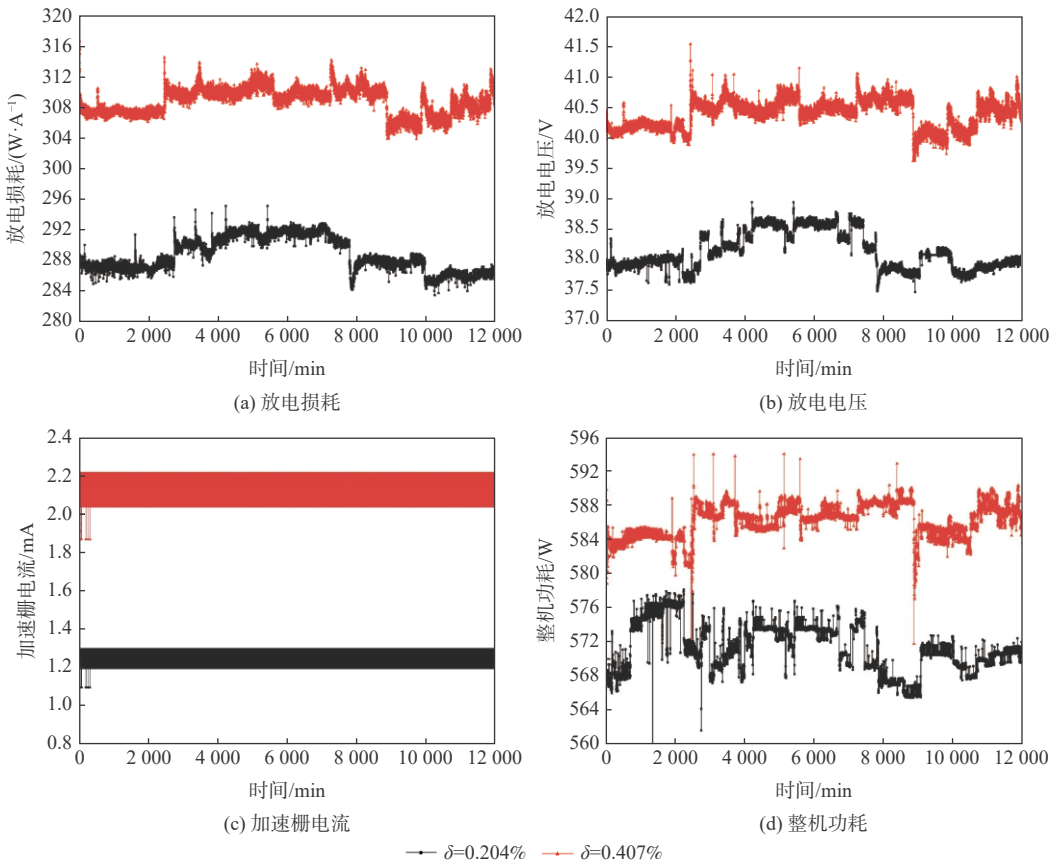


图6 2种补偿系数状态下20 mN工作点工作200 h时的关键参数变化情况

Fig. 6 Changes of key parameters at 20 mN for 200 h under two compensation coefficients

可以看出，10 cm口径连续变推力离子推力器的栅极系统采用2种加速栅小孔中心距补偿系数时整机工作效能存在显著差异，加速栅小孔中心距补偿系数δ=0.204%下的工作效能明显优于补偿系数δ=0.407%的状态，其200 h内的平均放电损耗和整机功耗分别约为289 W/A和571 W，较补偿系数δ=0.407%状态下的放电损耗、整机功耗分别低20 W/A和15 W。与此同时，平均放电电压和加速栅电流分别仅为38.1 V和1.25 mA，相对于δ=0.407%状态下的参数水平均出现大幅下降。放电电压的下降可有效降低离子的能量，进而减小离子对栅极系统的溅射刻蚀速率，最终提升整机工作寿命；加速栅电流的下降意味着直接轰击到加速栅结构上的束流离子和交换电荷离子数量的降低，同样可以提高工作寿命。

为对加速栅下游表面及小孔直径的离子刻蚀情况进行量化描述，采用激光共聚焦显微镜对工作200 h后束流密度最大的加速栅下游表面中心区域的刻蚀深度、小孔直径进行了测量。考虑到离子束流的轴对称性，主要对加速栅表面ϕ=10 mm区域内2个垂直方向上的中心点位(1#)和边缘点位(2#)的刻蚀深度与小孔直径进行了测量，测试结果如表2所示。

分析表2数据可以看出，10 cm口径连续变推力离子推力器栅极系统采用2种加速栅小孔中心距补偿系数时的加速栅下游表面刻蚀与小孔直径变化情况存在明显差异，但同一种补偿系数下加速栅水平方向与垂直方向的不同位置处的刻蚀深度、小孔直径变化基本一致。加速栅小孔中心距补偿系数δ=0.204%下的加速栅下游表面刻蚀深度、小孔直径较补偿系数δ=0.407%的状态分别减小约61%和24%。综合图6和表2可以发现，加速栅小孔中心距补偿系数δ=0.204%下的整机工作效能与离子溅射刻蚀情况均优于补偿系数δ=0.407%的状

表 2　2 种补偿系数状态下加速栅表面刻蚀深度与小孔直径测量结果

Table 2　Measurement results of etching depth and aperture diameter of the accelerator grid surface under two compensation coefficients

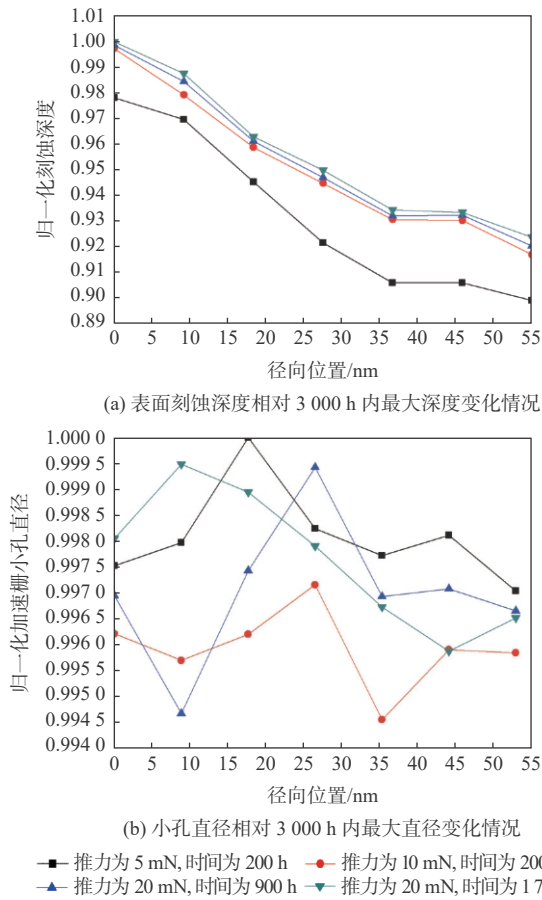
测试方向	刻蚀深度与最大刻蚀深度之比				小孔直径与最大轮廓直径之比			
	δ=0.204%		δ=0.407%		δ=0.204%		δ=0.407%	
	1#	2#	1#	2#	1#	2#	1#	2#
水平	0.378 0	0.380 1	1.000 0	0.991 2	0.759 2	0.753 3	0.993 4	0.995 7
垂直	0.378 5	0.380 6	0.967 6	0.989 3	0.760 5	0.757 5	0.993 7	1.000 0

态。分析其原因主要是加速栅小孔中心距补偿系数由 0.204% 增大到 0.407% 时, 栅极系统中屏栅与加速栅的小孔中心线发生较大位移, 栅极系统对离子的聚焦引出性能变差, 离子束流出现过聚焦现象, 加速栅对离子束流形成直接截获, 表征为加速栅电流的增加、真实离子束流的下降。与此同时, 过大的加速栅小孔补偿导致屏栅与加速栅共同构建的离子束流输运通道总面积减小, 离子透明度降低, 为保障实际所引离子束流指标的符合性, 唯有增加放电功耗以提升离子束流输运通道单位面积内的离子密度, 由此将导致放电损耗、放电电压的大幅增加。

3.5　多时间尺度下变推力模式的溅射刻蚀评价

基于上述影响研究可以发现, 10 cm 口径连续变推力离子推力器采用加速栅小孔中心距补偿系数 δ=0.204% 时的等离子体放电特性、离子束流分布趋势及加速栅短期耐离子溅射能力均优于 δ=0.407% 的状态。鉴于此, 基于加速栅小孔中心距补偿系数为 0.204% 的凹球面栅极系统对 10 cm 口径连续变推力器离子推力器进行了累计 3 000 h 不同推力模式的溅射刻蚀评价实验。结合变推力工作特性, 推力器先后在推力为 5 mN、10 mN 工作点各累计工作 200 h, 最后在推力为 20 mN 工作点累计工作 2 600 h, 各工作点下每工作 20 h 后推力器关机冷却 0.5 h。

实验期间采用激光共聚焦显微镜分别在推力器推力为 5 mN 工作 200 h 后、推力为 10 mN 工作 200 h 后、推力为 20 mN 工作 900 h 后及推力为 20 mN 工作 1 700 h 后对加速栅下游表面刻蚀深度、小孔直径进行测量。表面刻蚀深度测量时以加速栅结构中心轴为起点, 在半径方向上等间距选择 7 个位置点开展了测量; 小孔直径测量时以加速栅中心小孔为起点, 在半径方向上以 2 个小孔为间隔选择了 7 个小孔开展了测试。图 7 给出了加速栅小孔中心距补偿系数为 0.204% 下 10 cm 口径连续变推力离子推力器不同阶段时加速栅不同半径处下游表面刻蚀深度与小孔直径的相对变化情况。



(a) 表面刻蚀深度相对 3 000 h 内最大深度变化情况

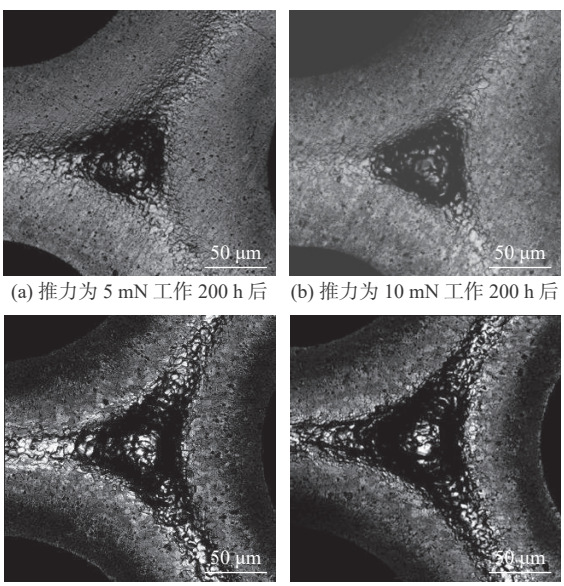
(b) 小孔直径相对 3 000 h 内最大直径变化情况

图 7　加速栅不同半径处表面刻蚀深度与小孔直径相对变化情况

Fig. 7　Relative change of etching depth and aperture diameter at different radii on downstream surface of accelerator grid

由图 7(a) 可见, 随着推力器工作累计时长的增加, 加速栅下游表面已出现规律性的溅射刻蚀。可以看出, 加速栅中心区域表面溅射刻蚀程度要远高于边缘区域, 对此主要是因为发散场构型下的变推力离子推力器放电室下游中心区域的离子密度要高于边缘区域, 而栅极系统所引离子束流轮廓与屏栅附近的等离子体密度分布具有对称相似性, 由此使得中心区域的离子密度要远高于边缘区域, 从而导致其溅射刻蚀速率要高于边缘区域。此外, 相对于 10 mN 工作点及 20 mN 工作点而言, 5 mN 工作

点下的放电室等离子体密度相对较低，在总加速电压恒定的情况下束流离子被栅极系统引出过程中处于过聚焦的状态，加速栅的离子轰击电流较大；而且此时放电室内工质利用率较低，大部分推进剂透过栅极系统逃逸并在加速栅下游形成交换电荷离子，2方面因素导致5 mN工作点下加速栅电流偏大且下游表面溅射刻蚀速度快。此后，随着放电室等离子体密度的增大，束流离子的引出逐渐由过聚焦状态过渡到正常聚焦状态，离子不再直接轰击到加速栅结本体上，加速栅电流中的离子轰击电流占比趋于0，而且工质利用率也伴随放电室等离子体密度的增大而上升，逃逸的推进剂减少，由此产生的交换电荷离子密度下降，使得加速栅电流降低的同时下游表面溅射刻蚀趋于平缓。显然，为降低栅极系统在小推力工作模式下因束流过聚焦所出现的快速刻蚀效应，应配合调整总加速电压，以提高离子束流密度与总加速电压之间的匹配性。由此，整体表征为推力为5 mN工作点下的溅射刻蚀程度高于推力为10 mN和推力为20 mN工作点，且推力为10 mN和推力为20 mN工作点下的溅射刻蚀速率低，变化缓慢。区别于图7(a)中表面刻蚀深度的规律性变化，图7(b)中不同阶段、不同位置的小孔直径变化并无明显规律，且相对变化量较小，显然累计3 000 h的溅射刻蚀评价实验中小孔直径并未出现明显变化，其最大直径与公称直径相比基本一致。图8为10 cm口径连续变推力离子推力器不同阶段时加速栅不同半径处下游表面刻蚀深度显微测试结果。



(a) 推力为5 mN工作200 h后 (b) 推力为10 mN工作200 h后

(c) 推力为20 mN工作900 h后 (d) 推力为20 mN工作1 700 h后

图8 加速栅下游表面刻蚀中心区域坑槽腐蚀深度显微测试结果

Fig. 8 Microscopic test results of pit corrosion depth in center area of downstream surface of accelerator grid

4 结 论

1）以实现大范围、高精度推力调节为目标，开展了凹球面栅极系统加速栅小孔中心距补偿系数对推力器工作性能影响的综合实验研究。结果表明，在推力器口径确定的情况下，改变加速栅小孔中心距补偿系数将对推力器的变推力工作性能产生重要影响。

2）在总加速电压恒定的情况下，改变加速栅小孔中心距补偿系数可调节放电室的基本性能和束流离子的密度分布，且推力越小，补偿系数变化的影响权重越明显，当补偿系数由0.204%增加到0.407%时，推力器在推力为1 mN工作点的放电损耗由477.2 W/A上升至868.7 W/A，增大近82%。

3）加速栅小孔中心距补偿系数的变化不会明显改变栅极系统的电子返流极限电压，但会显著影响宽范围变推力工作点下束流离子对栅极系统的离子溅射刻蚀效应，当补偿系数由0.204%增加到0.407%时，推力器在宽范围工作点下交换电荷离子的溅射腐蚀寿命平均降低约5%，对此可通过调整总加速电压予以改善。本文研究为变推力离子推力器栅极系统匹配性设计及高性能变推力控制策略优化提供技术支撑。

参考文献（References）

[1] WILLIAMS L T, WALKER M L R. Initial performance evaluation of a gridded radio frequency ion thruster[J]. Journal of Propulsion and Power, 2014, 30(3): 645-655.

[2] KORKUT B, LEVIN D A. Three-dimensional simulations of back-flows from ion thruster plumes using unstructured grid refinement[J]. Journal of Propulsion and Power, 2017, 33(1): 264-275.

[3] 龙建飞, 张天平, 李娟, 等. 离子推力器栅极透过率径向分布特性研究[J]. 物理学报, 2017, 66(16): 63-71.
LONG J F, ZHANG T P, LI J, et al. Optical transparency radial distribution of ion thruster[J]. Acta Physica Sinica, 2017, 66(16): 63-71 (in Chinese).

[4] CANUTO E, MASSOTTI L. All-propulsion design of the drag-free and attitude control of the European satellite GOCE[J]. Acta Astronautica, 2009, 64(2-3): 325-344.

[5] 胡竟, 王亮, 张天平, 等. LIPS-300离子推力器环形会切磁场等效磁路分析研究[J]. 推进技术, 2018, 39(3): 715-720.
HU J, WANG L, ZHANG T P, et al. Research on equivalent magnetic circuit of ring-cusp magnet field for LIPS-300 ion thruster[J]. Journal of Propulsion Technology, 2018, 39(3): 715-720(in Chinese).

[6] 杨福全, 王蒙, 郑茂繁, 等. 10cm离子推力器放电室性能优化研究[J]. 推进技术, 2017, 38(1): 235-241.
YANG F Q, WANG M, ZHENG M F, et al. Optimization of performance of discharge chamber of a 10cm diameter ion thruster[J]. Journal of Propulsion Technology, 2017, 38(1): 235-241(in Chinese).

[7] 胡竟, 杨福全, 郭德洲, 等. 基于CFD的10cm氙离子推力器阳极

推进剂供给方式优化[J]. 北京航空航天大学学报, 2020, 46(8): 1476-1484.
HU J, YANG F Q, GUO D Z, et al. Optimization of anode propellant allocation manner of 10cm xenon ion thruster based on CFD[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(8): 1476-1484(in Chinese).

[8] 胡竟, 耿海, 王东升, 等. 连续变推力离子推力器双荷离子特性分析与诊断[J]. 北京航空航天大学学报, 2023, 49(12): 3303-3310.
HU J, GENG H, WANG D S, et al. Characteristic analysis and diagnosis of double charged ions of continuous variable-thrust ion thruster[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(12): 3303-3310(in Chinese).

[9] 胡竟, 杨福全, 郭德洲, 等. 10cm 氙离子推力器变推力特性研究[J]. 推进技术, 2020, 41(10): 2382-2389.
HU J, YANG F Q, GUO D Z, et al. Analysis on variable-thrust characteristic of 10cm xenon ion thruster[J]. Journal of Propulsion Technology, 2020, 41(10): 2382-2389(in Chinese).

[10] OKAWA Y, TAKEGAHARA H. Numerical study of beam extraction phenomena in an ion thruster[J]. Japanese Journal of Applied Physics, 2001, 40(1R): 314.

[11] OKAWA Y, SATOW H, TAKEGAHARA H. Investigation of ion beam extraction phenomena using particle simulation code[C]//Proceedings of the 22nd International Symposium On Space Technology and Science. Tokyo: ISTS Press, 2000: 1432-1437.

[12] NAKANO M, NAKAMURA K, NAKAGAWA Y, et al. Numerical simulation of full-aperture-pair ion optics in a miniature ion thruster[J]. Physics of Plasmas, 2018, 25(1): 013524.

[13] TARTZ M, HARTMANN E, NEUMANN H. Extraction grid erosion in long-time ion thruster operation: comparison with simulation[C]//Proceedings of the 39th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit. Reston: AIAA, 2003.

[14] TARTZ M, HARTMANN E, DELTSCHEW R, et al. Study of extraction grid erosion in a long-time ion thruster operational test: IEPC-2003-176[R]. Washington, D. C. : IEPC, 2003: 1-9.

[15] 李军星, 张勇波, 王治华, 等. 基于加速栅溅射腐蚀失效的离子推力器寿命预测[J]. 航空动力学报, 2016, 31(5): 1047-1052.
LI J X, ZHANG Y B, WANG Z H, et al. Life prediction of ion thruster based on sputtering erosion failure of accelerator grid[J]. Journal of Aerospace Power, 2016, 31(5): 1047-1052(in Chinese).

[16] 陈茂林, 夏广庆, 毛根旺. 多模式离子推力器栅极系统三维粒子模拟仿真[J]. 物理学报, 2014, 63(18): 189-197.
CHEN M L, XIA G Q, MAO G W. Three-dimensional particle in cell simulation of multi-mode ion thruster optics system[J]. Acta Physica Sinica, 2014, 63(18): 189-197(in Chinese).

[17] 李建鹏, 靳伍银, 赵以德. 多模式离子推力器输入参数设计及工作特性研究[J]. 物理学报, 2022, 71(7): 259-267.
LI J P, JIN W Y, ZHAO Y D. Design of input parameters and operating characteristics for multi-mode ion thruster[J]. Acta Physica Sinica, 2022, 71(7): 259-267(in Chinese).

[18] 李建鹏, 靳伍银, 赵以德. 加速电压和阳极流率对离子推力器性能的影响[J]. 物理学报, 2022, 71(1): 166-173.
LI J P, JIN W Y, ZHAO Y D. Influence of acceleration grid voltage and anode flow rate on performance of ion thruster[J]. Acta Physica Sinica, 2022, 71(1): 166-173(in Chinese).

[19] 李建鹏, 赵以德, 靳伍银, 等. 多模式离子推力器放电室和栅极设计及其性能实验研究[J]. 物理学报, 2022, 71(19): 247-257.
LI J P, ZHAO Y D, JIN W Y, et al. Design and performance test of discharge chamber and grid for multi-mode ion thrusters[J]. Acta Physica Sinica, 2022, 71(19): 247-257(in Chinese).

[20] 赵以德, 李娟, 吴宗海, 等. 屏栅边缘小孔孔径对离子推力器性能的影响[J]. 物理学报, 2020, 69(11): 234-241.
ZHAO Y D, LI J, WU Z H, et al. Influence of screen gird aperture diameter in outer region on performance of dual-mode ion thruster[J]. Acta Physica Sinica, 2020, 69(11): 234-241(in Chinese).

[21] 孙明明, 张天平, 高俊, 等. 30cm 离子推力器三栅极组件设计参数对寿命的影响研究[J]. 推进技术, 2019, 40(2): 472-480.
SUN M M, ZHANG T P, GAO J, et al. Effects of three-grids assembly design parameters on estimated lifetime for 30cm diameter ion thruster[J]. Journal of Propulsion Technology, 2019, 40(2): 472-480(in Chinese).

[22] 郭德洲, 胡竟, 杨福全, 等. 曲面栅极朝向对离子推力器影响的试验研究[J]. 真空与低温, 2022, 28(1): 115-121.
GUO D Z, HU J, YANG F Q, et al. Experimental of the influence of grid optical system spherical orientation on the operating characteristics of ion thruster[J]. Vacuum and Cryogenics, 2022, 28(1): 115-121(in Chinese).

[23] GOEBEL D M, KATZ I. Fundamentals of electric propulsion: ion and hall thrusters[M]. Hoboken: Wiley, 2008.

[24] BROPHY J R, GARNER C E, GOODFELLOW K D. Electric propulsion system technology: NASA CR-194100[R]. Washington, D. C. : NASA, 1990: 93-103.

[25] 王二蒙. 离子推进器加速栅极腐蚀机理的研究[D]. 哈尔滨: 哈尔滨工业大学, 2013: 1-18.
WANG E M. Study on accelerated grid corrosion mechanism of ion thruster[D]. Harbin: Harbin Institute of Technology, 2013: 1-18(in Chinese).

[26] 李得天, 张天平, 张伟文, 等. 空间电推进测试与评价技术[M]. 北京: 北京理工大学出版社, 2017: 30-73.
LI D T, ZHANG T P, ZHANG W W, et al. Testing and evaluation technologies for space electric propulsion[M]. Beijing: Beijing Insititute of Technology Press, 2017: 30-73(in Chinese).

[27] 赵以德, 江豪成, 张天平, 等. 离子推力器束流均匀性改善方法研究[J]. 真空与低温, 2015, 21(3): 157-160.
ZHAO Y D, JIANG H C, ZHANG T P, et al. Method of improving beam current density uniformity for ion thruster[J]. Vacuum and Cryogenics, 2015, 21(3): 157-160(in Chinese).

Experimental study on effect of aperture of accelerator grid on performance of variable thrust ion thruster

HU Jing*，GUO Dezhou，GENG Hai，YANG Fuquan，LI Jianpeng，CHEN Juanjuan

(Science and Technology on Vacuum Technology and Physics Laboratory，Lanzhou Institute of Physics，Lanzhou 730000，China)

Abstract：Ion thrusters have been widely used in attitude and orbit control missions of long-duration spacecraft because of their outstanding advantages of high specific impulse and long life. A comprehensive experimental study was carried out to investigate the effect of the accelerator grid compensation coefficient on the performance of the continuous variable thrust ion thruster to achieve a large range and high precision thrust regulation. The mechanism and influencing factors of accelerator grid compensation design were analyzed. Based on 10cm diameter continuous variable thrust ion thrusters, two types of grid systems with different accelerator grid compensation coefficients were designed. The effects of accelerator grid compensation coefficients on a wide range of discharge characteristics of the ion thruster plasma in the discharge chamber, electron backflow limited voltage and acceleration current, ion beam distribution and ion sputtering etching effect were studied under different thrust modes. Two types of grid systems with varying accelerator grid compensation coefficients were devised based on continuous variable thrust ion thrusters with a diameter of 10 cm. The result shows that under constant total acceleration voltage, changing the accelerator grid compensation coefficient can adjust the basic performance of the discharge chamber and the density distribution of the ion beam. And the smaller the thrust, the more pronounced the effect of the compensation coefficient variation. At a wide range of variable-thrust operating points, the accelerator grid compensation coefficient variation has a significant effect on the ion sputtering erosion effect of the ion beam on the grid system, but it has little effect on the electron backflow limited voltage of the grid system. It can be improved by adjusting the total acceleration voltage. The above research will provide technical support for the matching design of the grid system of a prototype ion thruster and optimization of high-performance variable-thrust control strategies, facilitating the illustration and implementation of such systems.

Keywords：acceleration grid aperture pitch compensation coefficients；variable-thrust；ion thruster；sputtering erosion；discharge loss

Received：2024-03-28；**Accepted**：2024-05-15；**Published Online**：2024-06-04 15：14

URL：link.cnki.net/urlid/11.2625.V.20240604.1120.004

Foundation items：National Key Research and Development Program of China (2021YFC2202704); National Natural Science Foundation of China (62371213); Natural Science Foundation of Gansu Province (22JR5RA787,22JR5RA789,22JR5RA784)

* **Corresponding author**. E-mail：hjing37615486@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0185

航天员舱外维修操作时间的量化评估

刁常堃[1,*]，王昊[1]，孟令子[1]，苏南[1]，高鸿[2]，邓忠民[2]

(1. 北京空间飞行器总体设计部，北京 100094；2. 北京航空航天大学 宇航学院，北京 100191)

摘 要：航天员舱外维修操作是维持空间站在太空环境中安全运行的必要措施，量化评估维修操作时间是确保舱外活动任务成功和航天员安全的基础。基于此，提出一种航天员舱外维修操作时间量化评估方法，用于设计和分析航天员舱外活动操作时间，以满足航天员出舱时间的要求。该方法考虑微重力环境的影响，运用模特（MOD）法对已有数据中的关节基准动作时间进行修正，采用最小二乘法对已有数据进行拟合得到关节基准动作时间的不确定性范围。针对不同航天员、不同舱外任务的工况，与公开的舱外活动视频进行对比分析，验证了所提方法的准确性，为航天员舱外维修操作所需时间提供一种可行的依据。

关 键 词：舱外维修；时间量化评估；模特法；最小二乘法；关节基准动作

中图分类号：V423

文献标志码：A **文章编号**：1001-5965(2026)05-1487-09

合理的载荷照料和设备维护维修是实现载人航天器长寿命和维持系统高可靠性的必要手段，可以解决航天器出现的各种故障并延长其寿命，从而确保飞行安全和任务完成[1-2]。然而，准确预计维修任务所需时间对于空间站的维修性设计具有重大意义。这样的预估可以为空间站的维修性设计和维修时间分配提供支持[3-4]。因此，开发出准确的维修任务时间评估方法对于保障空间站的正常运行和任务完成至关重要。

目前，在评估太空环境中的舱外维修时间方面，通常采用基于工程经验的在轨维修时间度量方法[5]。此方法主要依赖地面上的维修时间方法进行估计，并通过将地面上的维修时间乘一定的倍数来近似在轨维修时间。然而，这种粗略的估算方法只能提供大致的出舱时间，缺乏设计性和适应性。另外还有基于任务分配的在轨维修时间方法[6]、基于动作时间度量的方法[7]及基于虚拟现实的维修时间度量方法[8]。在微重力环境中，舱外维修动作与地面动作完全不同，维修任务也各有不同，使得维修动作更加复杂。因此，目前舱外维修操作时间的量化方法的准确性仍有待提高。为更精确地评估舱外维修时间，需要开发出适应微重力环境的新方法，并结合实际任务需求进行精确的时间估算。

虽然在轨维修时间评估方法的文献报道相对较少，但在其他方面的维修时间评估已经进行了大量的研究工作。张力匀等[9]基于虚拟现实技术，开发了一种维修时间评估方法，该方法结合了动作实验和预定动作时间法。Chen 等[10]提出了一种维修时间评估方法，该方法基于虚拟仿真技术，并考虑了维修动作要素（maintenance action elements, MAEs）和模特（modular arrangement of predetermined time standards, MOD）法。Dong 等[11]设计的集成平台用于维修性验证和维修时间预测。Su 等[12]通过研究维修程度与维修时间之间的关系，发展了一种计算可用时间内维修设备数量的算法。吴振亚等[13]综述了系统平均修复时间（mean time to repair, MTTR）验证

收稿日期：2024-04-01；**录用日期**：2024-08-23；**网络出版时间**：2024-10-11 13:44
网络出版地址：link.cnki.net/urlid/11.2625.V.20241010.1844.003
基金项目：国家重点研发计划 (2021YFC2202801)
* **通信作者**. E-mail：diaochangkun@qq.com

引用格式：刁常堃，王昊，孟令子，等. 航天员舱外维修操作时间的量化评估 [J]. 北京航空航天大学学报，2026，52（5）：1487-1495.
DIAO C K，WANG H，MENG L Z，et al. Quantitative assessment of extravehicular maintenance operation time for astronauts [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（5）：1487-1495（in Chinese）.

方法的相关内容。这些方法为舱外维修操作时间的评估方法的开发提供了一定的借鉴。

本文考虑了微重力环境，提出一种航天员舱外维修操作时间量化评估方法。与公开的舱外视频进行了对比分析，验证了该方法在评估舱外维修操作时间方面的准确性和可靠性，为航天员舱外维修任务的规划和安排提供有利的参考。

1 舱外维修操作时间评估方法框架

由于微重力环境[14]的影响，航天员舱外维修操作与地面维修操作存在显著差异。因此，地面上维修操作的时间评估方法已经不适用于舱外维修操作的时间评估。为解决这个问题，本文考虑微重力环境的影响，基于MOD法与最小二乘法对文献[5]数据进行修正与拟合，建立一套适用于舱外维修操作时间评估的方法。

1.1 模特法原理

MOD法是预定动作时间标准(predetermined time standards, PTS)法中最具代表性的方法之一。其将生产操作动作归纳为21种基本动作，如图1所示[15]，并通过分解实际生产中的操作，将其归纳为这21种基本动作中的一种或多种。通过重新组合和计算这些分解后的动作，可以得到基准动作MOD值，并由此计算出时间值(1MOD = 0.129 s)[15]。

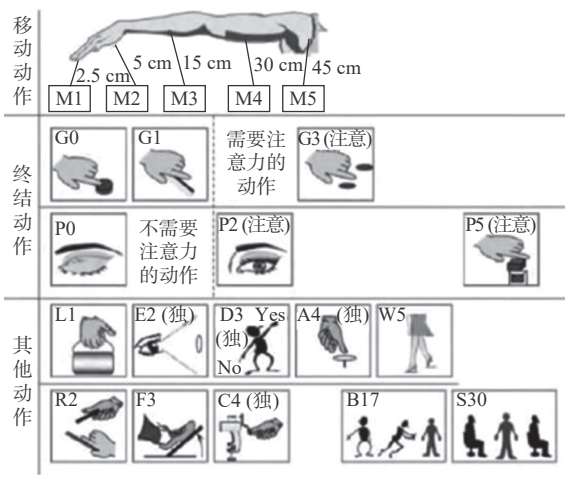


图1 模特法基本动作[15]

Fig. 1 Basic movements of MOD method[15]

在MOD法中，相同条件下，不同人执行同一动作所需的时间值基本相等。此外，MOD法还考虑了身体其他部位动作相对于手指动作的时间关系。具体而言，身体其他部位的动作所需时间通常是手指动作一次MOD数所需时间的整数倍。例如，腕关节动作所需时间是手指动作所需时间的2倍，复杂放置动作所需时间是手指动作所需时间的5倍，具体如表1所示[16]。基于MOD值的时间计算方法可以帮助人们更准确地评估和预测不同动作所需的时间[16]。

表1 模特法基本动作及对应的MOD值[16]

Table 1 Basic movements of MOD method and corresponding MOD values[16]

基本动作	动作描述	MOD值
移动动作	手指动作M1	1
	腕关节动作M2	2
	肘关节动作M3	3
	肩关节动作M4	4
	伸直手臂动作M5	5
终结动作	触摸动作G0	0
	简单抓取动作G1	1
	复杂抓取动作G3	3
	简单放置动作P0	0
	较复杂放置动作P2	2
	复杂放置动作P5	5
其他动作	走步动作W5	5
	起身坐下动作S30	30
	目视动作E2	2
	校正动作R2	2
	判断动作D3	3
	加压动作A4	4
	旋转动作C4	4

1.2 微重力下维修动作分析

在空间站中，一段完整的维修任务主要包括出舱准备、出舱过闸、舱外活动和返回过闸，如图2所示。主要对航天员舱外维修操作的时间进行评估，基于模特法，将航天员舱外维修操作分解成指关节、腕关节、肘关节和肩关节的动作单元，最后，通过时间的累加来确定舱外维修操作的总时间。

评估一段舱外维修时间，采用累计法进行串行作业和并行作业的计算[17]：

串行作业维修时间：

$$T = k_1 t_1 + k_2 t_2 + \cdots + k_m t_m = \sum_{i=1}^{m} k_i t_i \tag{1}$$

式中：t_i为该维修任务中第i项维修活动的时间；k_i为第i项维修活动对整个维修任务的影响权重；m为维修任务中维修活动的总数。

并行作业维修时间：

$$T = \mathrm{Max}(k_1 t_1, k_2 t_2, \cdots, k_m t_m) \tag{2}$$

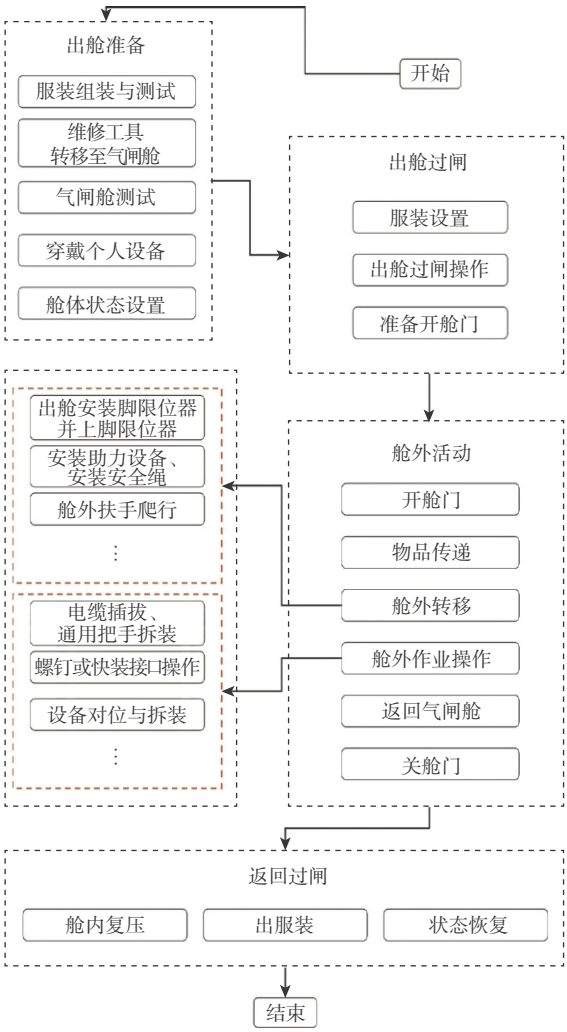


图 2　舱外维修操作时间流程

Fig. 2　Process of extravehicular maintenance operation time

1.3　微重力下维修动作的判定标准

为确定舱外维修操作所需的关节基准动作时间, 需要明确微重力下维修动作的判定标准。微重力下的移动类动作时间计算式为[5]

$$t = s/v \tag{3}$$

式中: s 为维修人员到维修地点的距离; v 为微重力环境下漂浮的平均速度。

本文主要研究微重力环境下维修类动作的时间判定标准, 根据维修时指关节、腕关节、肘关节和肩关节进行划分。通过模特法判定维修操作动作确定时间。

为确定基准动作时间, 引入节段自由度来度量代表相应关节的动作。节段自由度是用来定量确定关节段在三维空间中形态特征的量, 可以理解为人体各个关节的各种夹角。为建立节段自由度和旋转角度的关系, 引入笛卡儿坐标系来建立节段自由度和旋转角度之间的关系, 将关节固定在坐标原点, 并通过节段自由度的变化来分别求得最终位姿与初始位姿之间的夹角。

由于微重力环境下不会影响关节的旋转角度的大小, 并且各个关节的旋转角度是一定的。在维修操作过程中, 4 个关节会一起协调控制整个操作活动。人体关节在动作过程中各个关节的活动类型和活动角度如表 2 所示[5]。

表 2　关节的活动类型和活动角度[5]

Table 2　Types and angles of joint movement[5]

部位	活动角度	活动角度范围/(°)
指关节	屈曲θ_{KQ1}	0～ 90
	伸展θ_{KQ2}	0～15
腕关节	掌曲θ_{WQ1}	0～80
	背伸θ_{WQ2}	0～30
	桡偏θ_{WP1}	0～20
	尺偏θ_{WP2}	0～30
肘关节	屈曲θ_{EQ1}	0～150
	伸展θ_{EQ2}	0～10
	旋前θ_{EX1}	0～80
	旋后θ_{EX2}	0～80
肩关节	屈曲θ_{SQ1}	0～180
	伸展θ_{SQ2}	0～120
	外展θ_{SZ1}	0～90
	内收θ_{SZ2}	0～30
	旋内θ_{SX1}	0～70
	旋外θ_{SX2}	0～90

各关节活动角度和旋转角度间的关系为[9]

$$\theta = \arcsin\left(\sin^2(\theta_{SQ1}+\theta_{SQ2})+\cos^2\theta_{SQ1}\cos^2\theta_{SQ2}\sin^2(\theta_{SZ1}-\theta_{SZ2})-\frac{1}{2}\sin 2\theta_{SX1}\sin 2\theta_{SX2}(\cos(\theta_{SQ1}-\theta_{SQ2})+1)\right)^{\frac{1}{2}} \tag{4}$$

式中: θ 为以关节为中心点的旋转角度。

指尖到肩关节的长度平均值作为指尖活动的最大值 r_{max}, 肘关节到肩关节的长度作为指尖活动的最小半径 r_{min}。

根据弧长公式[5]:

$$l = \alpha r \tag{5}$$

式中: r 为半径; l 为圆心角弧长; α 为弧度。

可得指尖活动距离和肘关节的移动距离为[5]

$$\begin{cases}\theta_a = (\theta_{SQ}+\theta_{SZ})/2\\ \theta_a r_{min} < l < \theta_a r_{max}\end{cases} \tag{6}$$

式中: θ_a 为肩关节屈曲/伸展和外展/内收角度的平均值。

计算得到肩关节活动时指尖的移动距离范围(单位: cm) 为[5]

$$17.1 < l < 42.7 \tag{7}$$

将式(4)～式(7)求得的肩关节活动角度和肩关节活动时指尖的移动距离作为判定肩关节的动作标准为[5]

$$\begin{cases} 17.1 < l < 42.7 \\ \theta > 35.8° \text{或} |\theta_{SX1} - \theta_{SX2}| \geqslant 42.0° \end{cases} \tag{8}$$

根据上述步骤，分别可以求得肘关节、腕关节和手指关节的活动角度范围和各关节动作时指尖活动的距离，各个关节的判定标准如下。

肘关节的动作标准为[5]

$$\begin{cases} 8.2 < l < 18.7 \\ \theta < 11.8° \\ 42° < |\theta_{EX1} - \theta_{EX2}| < 48.7° \end{cases} \tag{9}$$

腕关节的动作标准为[5]

$$\begin{cases} 2.8 < l < 8.2 \\ |\theta_{WQ1} - \theta_{WQ2}| < 20.6° \end{cases} \tag{10}$$

指关节的动作标准为[5]

$$\begin{cases} 0 < l < 2.8 \\ |\theta_{KQ1} - \theta_{KQ2}| < 24° \end{cases} \tag{11}$$

1.4 舱外维修时间的修正

通过对文献数据的修正与拟合得到1MOD的关节基准动作时间区间值，文献[5]数据如表3所示。

考虑到舱外维修操作时间可能受到各种因素的影响而产生波动，因此，本文采用了修正值和不确定性范围的方法来评估舱外维修操作时间。

表3 文献[5]数据

Table 3 Literature[5] data

动作	基准动作时间平均值/s	基准动作时间方差/s^2	基准动作时间中位数/s	基准动作时间最大值/s	基准动作时间最小值/s	基准动作时间极差/s
指关节动作M1	0.72	0.03	0.71	0.96	0.52	0.44
腕关节动作M2	1.39	0.09	1.30	2.16	0.98	1.18
肘关节动作M3	1.95	0.19	2.01	2.82	1.19	1.63
肩关节动作M4	3.02	0.26	2.95	3.81	2.06	1.75

首先，通过使用MOD法修正基准动作时间，从而可以更准确地反映在特定环境下的操作时间；其次，应用最小二乘法对文献[5]数据进行拟合，得到关节基准动作时间的不确定性范围。此范围可以评估操作时间的可能波动范围；最后，将修正值和不确定性范围结合起来，以得到对舱外维修操作时间的评估。本文方法可以更全面地考虑舱外维修操作时间的不确定性，并提供一个更准确的操作时间范围。

对文献[5]中的基准动作时间进行修正，由模特法可知，各部分动作时间为1MOD的整数倍，因此，可得1MOD的关节基准动作时间修正值为

$$T_{1MOD} = \left(\frac{0.72}{1} + \frac{1.39}{2} + \frac{1.95}{3} + \frac{3.02}{4}\right) \Big/ 4 = 0.705 \tag{12}$$

对表3中数据采用最小二乘法进行拟合，得到关节基准动作时间修正值的不确定性范围。设各关节动作时间的MOD数为自变量x，各关节动作时间的平均值为因变量y。因此，表3中列出的关节动作的自变量分别为(0, 0)、(1, 0.72)、(2, 1.39)、(3, 1.95)、(4, 3.02)。通过对以上数据进行拟合可得：

$$\bar{x} = \sum_{i=1}^{N} x_i \Big/ N = \frac{0+1+2+3+4}{5} = 2 \tag{13}$$

$$\bar{y} = \sum_{i=1}^{N} y_i \Big/ N = 1.416 \tag{14}$$

$$r_t = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}} = 0.994 \tag{15}$$

式中：$\bar{x}$为各组数据的自变量平均值；x_i为第i组数据的自变量值；N为数据组数；$\bar{y}$为各组数据的因变量平均值，s；y_i为第i组数据的因变量值；r_t为该组数据在拟合过程中的相关系数。

由于相关系数较高，可以认为该拟合过程能够准确表达各组数据间的关系，因此，结果较为可靠。

因此，关节基准动作时间修正值的不确定性范围为

$$\sigma = \sqrt{(1/r_t^2 - 1) \Big/ (N-2)} = 0.064 \tag{16}$$

式中：σ为关节基准动作时间修正值的不确定性范围，s。

由式(12)和式(16)分析结果可知，1MOD的关节基准动作时间区间(单位：s)为

$$t_{1MOD} = 0.705 \pm 0.064 \tag{17}$$

航天员舱外维修操作时间的量化评估流程图如图3所示。

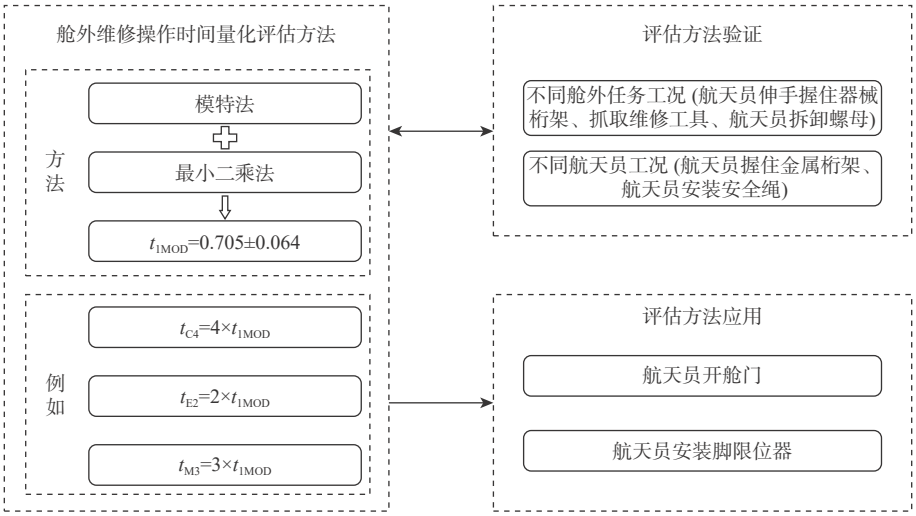


图 3　航天员舱外维修操作时间的量化评估流程图

Fig. 3　Framework for quantitative assessment of extravehicular maintenance operation time for astronauts

2　本文方法验证分析

针对不同航天员、不同舱外任务的工况, 对本文方法的准确性进行验证。使用公开的舱外活动视频进行对比分析, 重点观察了视频中的舱外维修操作与实际时间。在观察过程中, 记录下了每个操作的详细步骤和所需时间。接着, 使用本文方法对这些操作进了时间量化评估。通过对比分析和评估结果, 发现本文方法在不同航天员和不同舱外任务的工况下都表现出较高的准确性。评估结果与实际观察到的操作时间相吻合, 证明了本文方法的可行性。

2.1　不同舱外任务工况

2.1.1　案例 1—航天员伸手握住器械桁架、抓取维修工具

图 4 为公开舱外视频片段截图。在航天员执行伸手握住器械桁架的动作中(见图 4(a)), 肘关节和腕关节是控制手靠近桁架的主要关节。通过评估标准旋转角度和手指尖的移动范围, 将这个动作分解为肘关节动作 M3 和简单抓取动作 G1。分析结果如表 4 所示, 这个动作总共需要 4 个基准时间单元。类似地, 在航天员执行伸手抓取维修工具的动作中(见图 4(b)), 与伸手握住器械桁架相似, 同样由肘关节和腕关节控制手接近维修工具, 这个动作可以分解为肘关节动作 M3 和简单抓取动作 G1。并且同样需要 4 个基准时间单元。



(a) 伸手握住器械桁架　　(b) 伸手抓取维修工具

图 4　案例 1 公开舱外视频片段截图

Fig. 4　Public extravehicular video segments screenshots of case 1

表 4　案例 1 结果分析

Table 4　Case 1 result analysis

动作描述	动作分解	t_{1MOD}值	本文方法时间/s	文献[5]方法时间/s	实际时间/s
伸手握住器械桁架	M3-G1	4	[5.13,6.15]	5.34	5.72
伸手抓取维修工具	M3-G1	4			

2.1.2　案例 2—航天员拆卸螺母

航天员接近待拆卸螺母的动作中(见图 5(a)), 肩关节和肘关节一起动作来控制手接近螺母。通过判断标准旋转角度和手指尖的移动范围, 确定肩关节为主动作关节, 这个动作可以分解为肩关节动作 M4。在旋转动作中(见图 5(b)), 航天员将维修工具套上螺母后, 以肘关节为中心进行旋转动作,



(a) 接近待拆卸螺母　　(b) 旋转动作

(c) 拧下螺母

图 5　案例 2 公开舱外视频片段截图

Fig. 5　Public extravehicular video segments screenshots of case 2

其他关节基本保持不变。每转一周对应 2 个肘关节动作 M3，总共转了 5 周。在拧下螺母的动作中（见图 5(c)），待螺母松懈后，航天员直接用食指和大拇指拧下螺母。以指关节为中心进行动作，其他关节基本保持不变，这个动作可以分解为 5 个手指动作 M1，结果分析如表 5 所示。

表 5　案例 2 结果分析

Table 5　Case 2 result analysis

动作描述	动作分解	$t_{1\mathrm{MOD}}$值	本文方法时间/s	文献[5]方法时间/s	实际时间/s
接近待拆卸螺母	M4	4			
旋转动作	M3	30	[25.00,29.99]	26.12	28.00
拧下螺母	M1	5			

2.1.3　结果分析

为定量评估本文方法相较于文献 [5] 方法在时间计算精度上的优势，本文引入相对误差作为无量纲的评价指标，用于衡量计算值与真实值之间的偏离程度，其计算式为

$$R_{\mathrm{e}}=\left|\frac{\overline{T}_{\mathrm{proposed}}/T_{\mathrm{reference}}-T_{\mathrm{true}}}{T_{\mathrm{true}}}\right|\times 100\% \tag{18}$$

式中：R_{e} 为相对误差；$\overline{T}_{\mathrm{proposed}}$ 为本文方法预测时间均值；$T_{\mathrm{reference}}$ 为文献 [5] 方法预测时间；T_{true} 为实际时间。

如图 6 和图 7 所示，案例 3 的动作为航天员握住金属桁架，本文方法计算得到的结果与实际时间更加接近，相对误差也更小，本文方法计算的平均时间与实际时间的平均相对误差为 2.03%，而文献 [5] 方法计算的时间与实际时间的平均相对误差为 6.23%。说明本文方法更加准确地反映了实际工况，能够更准确地评估舱外维修操作时间。

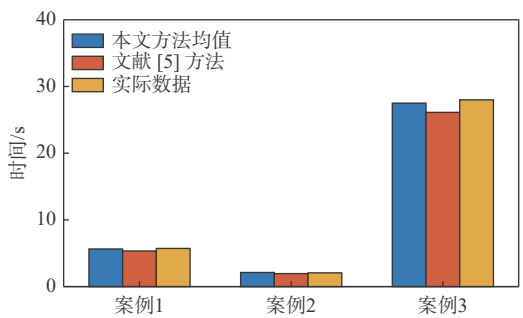


图 6　2 种方法时间数据对比

Fig. 6　Comparison of time data between two methods

2.2　不同航天员工况

本节中对出舱活动中常见动作所需时间进行了统计，总共统计了 3 个乘组的 4 次公开出舱活动视频，分别为神舟 7 号出舱、神舟 12 号首次出舱、神舟 12 号第 2 次出舱及神舟 13 号首次出舱。在神舟 13 号首次出舱活动中，统计了女性航天员王亚平的动作用时。

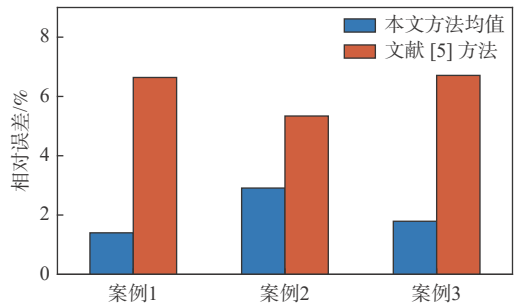


图 7　本文方法与文献 [5] 方法的相对误差对比

Fig. 7　Comparison of relative errors between the proposed method and literature[5] method

2.2.1　案例 3—航天员握住金属桁架

不同航天员（男、女）握住金属桁架的工况如图 8 所示，本文研究不同航天员对本文方法的准确性影响。在出舱活动过程中，航天员需要握住舱门外的金属桁架以稳定自身姿态，并提供向上的力帮助身体移出舱门。将航天员握住金属桁架的动作分解为肘关节动作 M3 和简单抓取动作 G1。肘关节动作 M3 主要通过肘关节控制小臂旋转，使航天员的手腕移动至桁架处。而简单抓取动作 G1 则是航天员手指的移动，抓住金属桁架。针对这 2 部分运动，使用本文方法对航天员握住金属桁架的工况进行准确性分析。观察不同航天员进行这个动作的视频，并记录下了每个动作的详细步骤和所需时间，结果分析如图 9 和表 6 所示。



(a) 男航天员握住桁架　(b) 女航天员握住桁架

图 8　案例 3 公开舱外视频片段截图

Fig. 8　Public extravehicular video segments screenshots of case 3

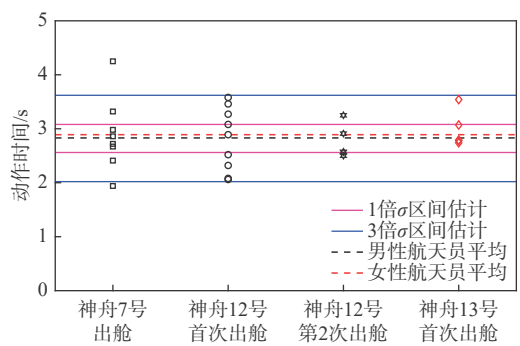


图 9　案例 3 动作时间统计

Fig. 9　Time statistics of case 3

2.2.2　案例 4—航天员安装安全绳

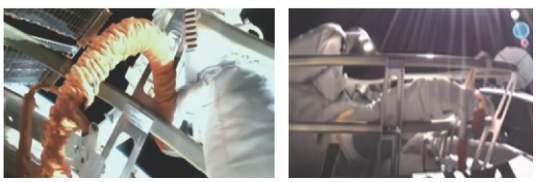
航天员在执行任务时，为确保其不会从空间站

表 6　案例 3 结果分析

Table 6　Case 3 result analysis

动作描述	动作分解	t_{1MOD}值	本文方法时间/s	男航天员平均实际时间/s	女航天员平均实际时间/s
握住金属桁架	M3-G1	4	[2.56,3.08]	2.83	2.89

表面脱落, 需要将连接在舱外航天服上的安全绳悬挂于空间站把手上, 如图 10 所示。在安装安全绳的过程中, 航天员需要通过肩关节动作 M4 带动小臂和大臂, 使安全绳卡扣移动至合理的位置。接着, 航天员需要进行较为复杂的动作 P2, 集中注意力将卡扣放置于把手上。最后, 航天员需要进行旋转动作 C4, 将卡扣旋转至正确方位。综合这 3 部分运动, 可将安装安全绳的动作分解为肩关节动作 M4、较复杂动作 P2 和旋转动作 C4, 结果分析如图 11 和表 7 所示。



(a) 男航天员安装安全绳　　(b) 女航天员安装安全绳

图 10　案例 4 公开舱外视频片段截图

Fig. 10　Public extravehicular video segments screenshots of case 4

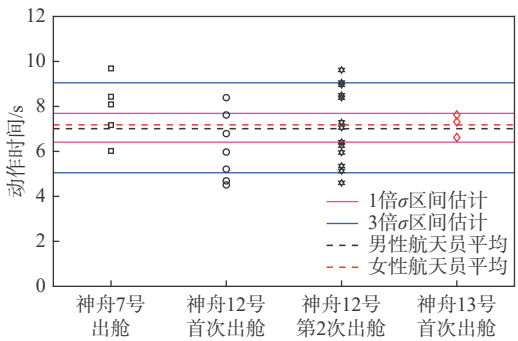


图 11　案例 4 动作时间统计

Fig. 11　Time statistics of case 4

表 7　案例 4 结果分析

Table 7　Case 4 result analysis

动作描述	动作分解	t_{1MOD}值	本文方法时间/s	男航天员平均实际时间/s	女航天员平均实际时间/s
安装安全绳	M4-P2-C4	10	[6.41,7.69]	7.01	7.18

在案例 3 和案例 4 中, 对 2 名不同性别的航天员在执行相同动作时所需的时间进行了分析。结果显示, 尽管 2 名航天员的动作时间存在差异, 但仍在本文方法计算的时间范围内。这表明本文方法具有一定的适用性, 可以适应不同航天员的需求, 并预测他们的工作时间。因此, 本文方法可以有效地评估不同航天员在执行特定动作时所需的时间。

3　本文方法应用

采用发展的舱外维修操作时间评估方法, 对航天员开舱门、航天员安装脚限位器的完整过程进行了出舱操作时间评估。

3.1　航天员开舱门

舱门是航天器与外部太空环境之间的关键通道, 开舱门使得航天员能够安全地进出舱, 执行太空任务或维修工作。

航天员开启舱门环节可以划分为 4 个关键动作: 伸手取工具、旋转工具、放回工具和伸手拉住舱门。首先, 航天员利用肩关节控制手来抓取工具, 这一步骤包括肩关节动作 M4 和复杂的抓取动作 G3, 总共需要 7 个 t_{1MOD} 的时间; 接着, 航天员使用腕关节控制手进行旋转动作, 对工具进行旋转以解锁舱门, 这一步可被视为腕关节动作 M2 和旋转动作 C4 的组合, 由于在整个开启舱门的过程中需执行约 18 次, 因此, 总共需要 108 个 t_{1MOD} 的时间; 然后, 航天员利用肩关节控制手放回工具, 这一步骤包括肩关节动作 M4 和复杂放置动作 P5, 需要 9 个 t_{1MOD} 的时间; 最后, 航天员伸手拉住舱门将其打开, 本文的动作由肩关节动作 M4 和复杂抓取动作 G3 组成, 大约需要执行 3 次, 总共需要 21 个 t_{1MOD} 的时间。

由以上结果分析可知, 航天员开舱门的动作共需要消耗 145 个 t_{1MOD} 值, 整个航天员开舱门的动作流程如表 8 所示。

表 8　航天员开舱门的时间评估

Table 8　Estimated time for astronauts to open the door of the spacecraft

动作描述	动作分解	次数	t_{1MOD}值	评估时间/s
伸手取工具	M4-G3	1	7	[4.49,5.38]
旋转工具	M2-C4	18	108	[69.23,83.05]
放回工具	M4-P5	1	9	[5.77,6.92]
伸手拉住舱门	M4-G3	3	21	[13.46,16.15]

3.2　航天员安装脚限位器

航天员安装脚限位器是一个关键的步骤, 其需要航天员进行一系列动作来完成。首先, 航天员需要准确地拿取脚限位器, 这个动作需要肩关节 M4 进行控制, 航天员必须准确地定位手部并抓取脚限位器。这个动作可以分为 2 个部分: 大臂的运动 M4 和复杂的抓取动作 G3。大臂的运动需要航天员调整肩关节的角度, 以便将手部移动到正确的位置。复杂的抓取动作包括手指的灵活操作, 航天员需要通过控制手指的力度和位置来抓取脚限位器。大臂的运动 M4 和复杂的抓取 G3 需要一定的

时间来完成，大约需要7个t_{1MOD}值。接下来，航天员需要将脚限位器按对位标识插入机械臂对应的基座上。此步骤需要航天员调节肘关节M3的角度，将脚限位器按对位标识移动到机械臂对应的基座位置。接着，航天员需要进行复杂放置动作P5，集中注意力E2将脚限位器插入基座。航天员需要将脚限位器按对位标识插入机械臂对应的基座上的动作需要多次重复进行，大约进行了4次，共计需要40个t_{1MOD}值。最后，航天员拨动锁定脚限位器末端的锁紧把手完成锁紧，这个动作同样由肘关节M3进行控制。航天员需要将手部旋转，以便锁定脚限位器末端的锁紧把手。锁定脚限位器末端的锁紧把手动作进行分解动作可以分为2个部分：小臂的运动M3和旋转动作C4。小臂的运动需要航天员调整肘关节的角度，以便将手部旋转到正确的位置。旋转动作是为确保锁紧把手的稳固性，航天员需要旋转手部以锁定脚限位器末端。锁定末端的锁紧把手的动作需要多次重复进行，大约进行了2次，共需要14个t_{1MOD}值。具体完整的动作流程如表9所示。

表9 航天员安装脚限位器的时间评估

Table 9 Estimated time for astronauts to install the foot restraint device

动作描述	动作分解	次数	t_{1MOD}值	评估时间/s
拿取脚限位器	M4-G3	1	7	[4.49,5.38]
将脚限位器末端插入基座	M3-P5-E2	4	40	[25.64,30.76]
锁定末端的锁紧把手	M3-C4	2	14	[8.97,10.77]

由表9可知，整个过程共需要61个t_{1MOD}值，其中，包括了手部移动、抓取、插入、放置和旋转等动作。因此，可以评估航天员安装脚限位器的过程需要大约39.10～46.91 s的时间。这个时间范围考虑了航天员进行各种动作和操作的时间，包括拿取脚限位器、将脚限位器末端插入基座、锁定末端的锁紧把手等步骤。

4 结 论

1）本文采用模特法修正已有数据中的关节基准动作时间，利用最小二乘法拟合关节基准动作时间的不确定性范围，提出了一种新的方法来评估舱外维修操作时间，研究成果为航天员舱外维修操作时间的设计和分析提供了理论基础。

2）相对于已有方法，本文方法更全面地考虑了操作时间的不确定性，并能够提供更准确的操作时间评估结果。本文方法计算的平均时间与实际时间的平均相对误差为2.03%，而已有方法计算的时间与实际时间的平均相对误差为6.23%。

由于可调研资料的限制，本文方法仍存在一些不足之处，需要进一步改进和完善。例如，在航天员出舱操作动作流程的分析方面，可以进行更详细的细化。今后的研究可以通过收集更多的数据样本来提高评估的准确性，并进一步探索航天员舱外维修操作时间的影响因素和误差来源。

参考文献（References）

[1] 李兴乾, 魏传锋, 张伟, 等. 载人航天器在轨维修性设计体系[J]. 系统工程与电子技术, 2016, 38(1): 84-89.
LI X Q, WEI C F, ZHANG W, et al. On-orbit maintainability design system for manned spacecraft[J]. Systems Engineering and Electronics, 2016, 38(1): 84-89(in Chinese).

[2] 张杰, 朱欣, 尹玉梅, 等. 国外载人航天器维修性标准综述[J]. 载人航天, 2023, 29(6): 824-831.
ZHANG J, ZHU X, YIN Y M, et al. Review of foreign standards for manned spacecraft maintainability[J]. Manned Spaceflight, 2023, 29(6): 824-831(in Chinese).

[3] 张伟, 侯永青, 夏侨丽, 等. 空间站维修设计与实践[J]. 中国科学: 技术科学, 2022, 52(9): 1332-1344.
ZHANG W, HOU Y Q, XIA Q L, et al. Maintenance design and practice for the China space station[J]. Scientia Sinica (Technologica), 2022, 52(9): 1332-1344(in Chinese).

[4] GE X Y, ZHOU Q X, LIU Z Q. Assessment of space station on-orbit maintenance task complexity[J]. Reliability Engineering & System Safety, 2020, 193: 106661.

[5] 薛龙. 面向在轨的维修时间度量方法研究[D]. 北京: 北京航空航天大学, 2018.
XUE L. Research on on-orbit maintenance time measurement methods[D]. Beijing: Beihang University, 2018(in Chinese).

[6] REMBALA R, OWER C. Robotic assembly and maintenance of future space stations based on the ISS mission operations experience[J]. Acta Astronautica, 2009, 65(7-8): 912-920.

[7] FENG Z H, DING X R, ZHANG H, et al. An energy consumption estimation method for the tool setting process in CNC milling based on the modular arrangement of predetermined time standards[J]. Energies, 2023, 16(20): 7064.

[8] 康乐. 基于虚拟现实的维修时间度量方法研究[D]. 北京: 北京航空航天大学, 2012.
KANG L. Research of maintenance time measurement based on virtual reality[D]. Beijing: Beihang University, 2012(in Chinese).

[9] 张力匀, 刘毅, 刘佳. 基于虚拟现实的维修时间预计方法[J]. 计算机辅助设计与图形学学报, 2016, 28(8): 1383-1392.
ZHANG L Y, LIU Y, LIU J. Maintenance time prediction using virtual reality[J]. Journal of Computer-Aided Design & Computer

Graphics, 2016, 28(8): 1383-1392(in Chinese).

[10] CHEN J Y, ZHOU D, KANG L, et al. A maintenance time estimation method based on virtual simulation and improved modular arrangement of predetermined time standards[J]. International Journal of Industrial Ergonomics, 2020, 80: 103042.

[11] DONG Z, LE K, CHUAN L. A virtual reality-based maintenance time measurement methodology for complex products[J]. Assembly Automation, 2013, 33(3): 221-230.

[12] SU X B, GAO Q, MA W N, et al. Algorithm of maintenance time and maintenance amount based on maintenance degree[C]//Proceedings of the IEEE International Conference on Power, Intelligent Computing and Systems. Piscataway: IEEE Press, 2020: 230-234.

[13] 吴振亚, 郝建平, 陈汉青, 等. 装备平均修复时间验证方法研究综述[J]. 系统工程与电子技术, 2020, 42(12): 2931-2943.
WU Z Y, HAO J P, CHEN H Q, et al. Review on research of equipment mean time to repair demonstration methods[J]. Systems Engineering and Electronics, 2020, 42(12): 2931-2943(in Chinese).

[14] 陈善广. 航天员出舱活动技术[M]. 北京: 中国宇航出版社, 2007.
CHEN S G. Astronaut extravehicular activity technology[M]. Beijing: China Astronautic Publishing House, 2007(in Chinese).

[15] 聂喜荣, 李鹏. 预定时间标准方法对比分析[J]. 价值工程, 2019, 38(19): 252-254.
NIE X R, LI P. Comparative analysis of predetermined time standard methods[J]. Value Engineering, 2019, 38(19): 252-254(in Chinese).

[16] 易树平, 郭伏. 基础工业工程[M]. 3 版. 北京: 机械工业出版社, 2022.
YI S P, GUO F. Fundament of industrial engineering[M]. 3rd ed. Beijing: China Machine Press, 2022(in Chinese).

[17] 李涛, 魏传锋, 李伟, 等. 载人航天器在轨维修地面仿真验证技术[J]. 航天器环境工程, 2016, 33(5): 510-515.
LI T, WEI C F, LI W, et al. Simulation and validation technology of manned spacecraft on-orbit maintenance[J]. Spacecraft Environment Engineering, 2016, 33(5): 510-515(in Chinese).

Quantitative assessment of extravehicular maintenance operation time for astronauts

DIAO Changkun[1,*], WANG Hao[1], MENG Lingzi[1], SU Nan[1], GAO Hong[2], DENG Zhongmin[2]

(1. Beijing Institute of Spacecraft System Engineering, Beijing 100094, China;
2. School of Astronautics, Beihang University, Beijing 100191, China)

Abstract: For space stations to operate safely in the space environment, astronauts' extravehicular maintenance tasks are essential. Therefore, it is essential to quantitatively assess the time required for these operations to ensure the success of extravehicular activities and the safety of astronauts. In this study, we present a quantitative technique for evaluating astronaut operational time during extravehicular maintenance activities, with the goal of designing and analyzing the duration of astronaut extravehicular operations to fulfill extravehicular time requirements. The proposed method takes into consideration the influence of the microgravity environment and the reference motion time of the joint in the literature data, which was modified by the modular arrangement of predetermined time standard (MOD) method, and the uncertainty range of the reference motion time of the joint was obtained by fitting the literature data with the least square method. For the working conditions of different astronauts and different extravehicular tasks, a comparative analysis with publicly available videos of extravehicular activities was conducted to verify the accuracy of the assessment method and provide a feasible basis for the time required for astronauts' extravehicular maintenance operations.

Keywords: extravehicular maintenance; quantitative assessment of time; modular arrangement of predetermined time standard method; least square method; joint reference action

Received: 2024-04-01; **Accepted**: 2024-08-23; **Published Online**: 2024-10-11 13:44

URL: link.cnki.net/urlid/11.2625.V.20241010.1844.003

Foundation item: National Key Research and Development Program of China (2021YFC2202801)

* **Corresponding author**. E-mail: diaochangkun@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0189

考虑擦机尾动态限制角的起飞抬轮过程优化

庄南剑*，杨雪雅，谷润平
(中国民航大学 空中交通管理学院，天津 300300)

摘 要：为防止擦机尾的发生，同时优化起飞抬轮过程，将抬轮过程中擦机尾角的变化考虑在内，建立计算飞机俯仰角、机尾离地高度等性能参数的抬轮阶段动力学模型。以波音737-800为例，探究抬轮速率变化对抬轮段距离的影响，对比擦机尾动态限制角和擦机尾固定限制角这2种抬轮方式的差异，分析动态限制角的优势。结果表明：大多数情况下，该机型最优抬轮速率范围为2.5～3.0 (°)/s。总体而言，相比于固定限制角，使用动态限制角的抬轮段距离减小约10%，从抬轮到离地的时间缩短约10%。对有关后机身长度和起落架高度的飞机设计提出建议，即后机身长度的选取与起落架高度应是相匹配的，选取比值约为5.0，若采用擦机尾动态限制角抬轮方式，比值应略微增大。研究结果可为飞行员提供抬轮操作参考，改善飞机的起飞性能，在防止擦机尾的同时，提高飞机载重。

关 键 词：航空运输；起飞性能；擦机尾；抬轮速率；动态限制角
中图分类号：X949; V212.2
文献标志码：A **文章编号**：1001-5965(2026)05-1496-08

随着民航安全技术的发展,飞行事故率有所降低,但总的事故征候数量却呈现明显的上升趋势,其中,擦机尾是近年来事故征候增长较快的一类事件[1-2]。起飞擦机尾事件取决于起飞抬轮过程是否妥当。起飞的抬轮过程为到达抬轮速度时,机组施加向后带杆,用大约3(°)/s的抬轮速率,使飞机的俯仰姿态连续增加到起飞目标姿态的过程。起飞抬轮的响应机理复杂,影响因素众多,包括气象风切变、机组操作、飞机性能等,对于机身较长的机型,飞机的起飞性能更易受擦机尾角限制。

针对起飞擦机尾方面的研究,Chen等[3]分析大量历史样本数据,基于逐步回归方法,利用Python建立了最优多元回归模型,找到了擦机尾密切相关的重要变量。Li等[4]根据各机型的几何特性确定了飞机擦机尾建模的关键参数,利用历史航班数据,筛选出风险最大的航班。孙瑞山等[5]基于快速存取记录器(quick access recorder, QAR)数据,结合数理统计分析,模拟出飞机起飞离地仰角值的可能分布。杨星月等[6]建立擦机尾人因操作分析模型,探究飞行员模型参数变化对俯仰角的影响。这些研究大多是基于历史数据的原因分析和风险评估,对起飞抬轮过程进行建模,分析起飞离地擦机尾角的影响,但未对起飞过程中各时刻的起飞抬轮参数进行计算分析。

起飞过程的仿真需要结合起飞抬轮阶段的动力学模型,从而更加清晰的分析各个因素是如何影响导致擦机尾事件。蔡良才等[7]将起飞抬轮段简化成始终以离地俯仰角滑跑至离地,认为飞机抬前轮是在瞬间完成的。刘静等[8]将飞机滑跑俯仰角表示为速度的函数,提高了建模的准确性。孟祥光等[9]建立了起飞过程中飞机动力学模型、起落架运动模型和驾驶员操纵模型。Chan[10]以香港国际机场为例,

收稿日期：2024-04-01；**录用日期**：2024-05-15；**网络出版时间**：2024-05-24 17:30
网络出版地址：link.cnki.net/urlid/11.2625.V.20240524.0944.002
基金项目：国家重点研发计划(2023YFB4302903)；中央高校基本科研业务费自然科学一般项目(KJZ53420210102)
* **通信作者**. E-mail：njzhuang@cauc.edu.cn

引用格式：庄南剑，杨雪雅，谷润平. 考虑擦机尾动态限制角的起飞抬轮过程优化[J]. 北京航空航天大学学报，2026，52(5)：1496-1503.
ZHUANG N J，YANG X Y，GU R P. Optimization of take-off rotation process considering tail striking dynamic limit angle[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1496-1503 (in Chinese).

研究了风切变对擦机尾的影响。Cuerno-Rejado 等[11]通过研究起飞关键参数之间的线性或非线性关系, 对后机身和水平尾翼的设计进行了优化。Balachandran 等[12]由俯仰角和机尾离地高度确定了飞机的擦机尾限制角。Jonathan[13]设计了一种用于防止飞机起降时擦机尾的动态控制器, 控制器通过控制俯仰角减少擦机尾的风险。Oudin 等[14]根据机尾接地点的位置, 给定一个安全裕度, 计算俯仰角极限值。上述研究中评判擦机尾风险的一个关键参数是俯仰角, 俯仰角的最大限值是擦机尾限制角, 研究中将擦机尾角记为固定值。但是实际上起落架支柱为联接机体和机轮的弹性装置[15], 因此, 擦机尾角是会变化的; 其次, 上述研究中也并未考虑机组抬轮速率对起飞性能的影响。

针对上述问题, 本文将抬轮过程中擦机尾角的变化考虑在内, 同时考虑了抬轮速率对抬轮过程的影响, 建立计算飞机俯仰角、机尾离地高度等重要性能参数的抬轮阶段动力学模型; 分析不同机场标高和温度下, 抬轮速率变化对抬轮段距离的影响, 对比采用擦机尾动态限制角和擦机尾固定限制角时, 飞机姿态变化过程和抬轮段距离的差异; 最后, 计算改变后机身长度的变化和起落架高度对起飞性能的影响。

1　动力学建模及分析

1.1　抬轮过程的距离计算模型

飞机顺利起飞的过程为松刹车加速, 达到抬轮速度 V_R 后进行抬轮操作, 在飞机离地后收起落架, 直至起飞阶段结束, 可以将起飞过程分为加速滑跑段、抬轮段和爬升段。由于擦机尾风险仅取决于抬轮段, 所以重点分析抬轮段的运动过程, 抬轮段飞机受力如图 1 所示。

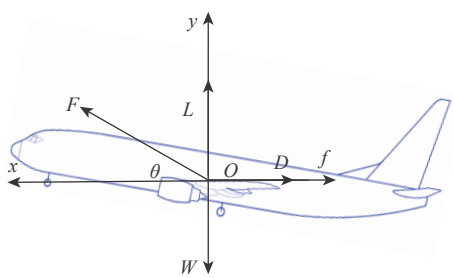


图 1　地面抬轮段受力分析

Fig. 1　Ground-level wheel-lift section stress analysis

如图 1 所示, 飞机运动坐标系以重心 O 为坐标原点, 沿运动方向为 x 轴, 垂直运动方向为 y 轴。飞机受力包括重力 W, N、推力 F, N、升力 L, N、阻力 D, N 和摩擦力 f, N, 升力和阻力的计算式分别为

$$\begin{cases} L = qS_W C_L = \dfrac{1}{2}\rho V^2 S_W C_L \\ D = qS_W C_D = \dfrac{1}{2}\rho V^2 S_W C_D \end{cases} \tag{1}$$

式中: q 为动压, Pa; S_W 为机翼面积, m^2; C_L 为升力系数; ρ 为空气密度, kg/m^3; V 为飞机真空速, m/s; C_D 为阻力系数。

根据 LEAP-1A 航空发动机性能手册[16], 由高度、温度和马赫数可以查询到 F/δ, 其中, δ 表示压强比, 计算式为

$$\delta = P/P_0 \tag{2}$$

式中: P 为大气压强, Pa; P_0 为标准海平面大气压强, Pa。

推力可表示为

$$F = F(H, T, V) \tag{3}$$

式中: H 为机场标高, m; T 为温度, ℃。

飞机受力沿运动坐标系分解得

$$\begin{cases} F_x = F\cos\theta - D - \mu(W - L) \\ F_y = W - L - F_v \end{cases} \tag{4}$$

式中: F_x 为水平合外力, N; θ 为飞机的俯仰角, (°); μ 为摩擦系数; F_y 为机身对起落架的压力, N; F_v 为推力在垂直方向的分量, N, 计算式为

$$F_v = F\sin\theta \tag{5}$$

飞机在抬轮后, 俯仰角不断增加, 假定飞机的抬轮速率为固定值 w_γ, (°)/s。则飞机的俯仰角 θ 为

$$\theta = w_\gamma t \tag{6}$$

式中: t 为距离抬轮开始的时间, s。

已知飞机在运动坐标系中的合力, 飞机的加速度 a, m/s^2 可表示为

$$a = \frac{1}{m}\left[F\cos\theta - \mu mg - (C_D - \mu C_L) qS_W\right] \tag{7}$$

式中: g 为重力加速度, m/s^2; m 为飞机的质量, kg。

当机身对起落架的压力为 0, 即 $W = L + F_v$ 时, 飞机刚好离地, 则抬轮段距离 S, m 可表示为

$$S = \int_{V_R}^{V_{LOF}} \frac{mV\mathrm{d}V}{F\cos\theta - \mu mg - (C_D - \mu C_L) qS_W} \tag{8}$$

式中: V_R 为抬轮速度, m/s; V_{LOF} 为飞机后轮离地时刻的速度, m/s。

1.2　擦机尾固定限制角

当飞机机尾触地时, 地面、机身和起落架构成一个三角形, 同种机型机身长度为固定值, 因此, 主起落架的高度决定擦机尾角, 如图 2 所示。

擦机尾限制角在擦机尾角的基础上设定机尾离地高度安全余量, 是抬轮法则中的一个擦机尾保

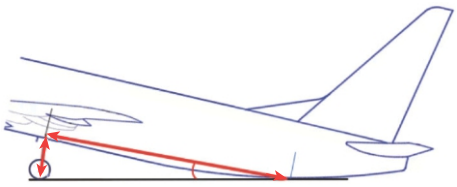


图2 擦机尾角示意图

Fig. 2 Schematic diagram of wiping tail corner

护功能，用于防止飞机起飞和着陆过程中俯仰角过大产生擦机尾。限制角大小与起落架高度有关，当起落架不受压力时，起落架减震支柱处于全伸出状态，对应计算的限制角最大，当起落架受压力时，减震支柱处于压缩状态，计算的限制角减小。因为飞机在起飞过程中，起落架减震支柱必然会受一定程度的压缩，一般采用较为保守数值，根据完全压缩后的最小起落架高度$H_{\min}$，cm计算得到固定限制角。

擦机尾固定限制角θ_{F}，(°) 计算式为

$$\theta_{\mathrm{F}}=\arcsin((H_{\min}-h_{\mathrm{sF}})/L_{\mathrm{B}}) \tag{9}$$

式中：h_{sF}为采用固定限制角时的机尾离地高度安全余量，cm；L_{B}为主起落架到擦机尾点的后机身长度，cm。

擦机尾固定限制角抬轮段距离计算流程如图3所示。

1.3 擦机尾动态限制角

擦机尾固定限制角是根据完全压缩后的起落架高度计算得到的，但是在起飞过程中，由于机身对起落架的压力逐渐减小，起落架高度会增加，限制角也会逐渐增大，采用固定限制角的方式过于保守。因此，可以根据飞机起飞过程中压力的变化计算不同时刻的动态变化的限制角，这种抬轮方式称之为擦机尾动态限制角。动态限制角与压缩率、起飞过程中机身对起落架的压力、机尾离地安全余量和后机身长度有关。

当主起落架机轮受力时，主起落架会被压缩，主起落架高度H_{m}，cm可表示为

$$H_{\mathrm{m}}=H_0-kF_y \tag{10}$$

式中：H_0为未受压力时主起落架的初始高度，cm；k为压缩率，cm/N。

开始抬轮到主起落架离地过程中机尾离地高度h，cm 可表示为

$$h=H_{\mathrm{m}}-L_{\mathrm{B}}\sin\theta \tag{11}$$

擦机尾动态限制角θ_{D}，(°) 的计算式为

$$\theta_{\mathrm{D}}=\arcsin((H_{\mathrm{m}}-h_{\mathrm{sD}})/L_{\mathrm{B}}) \tag{12}$$

式中：h_{sD}为擦机尾动态限制角下的机尾离地高度安全余量，cm。

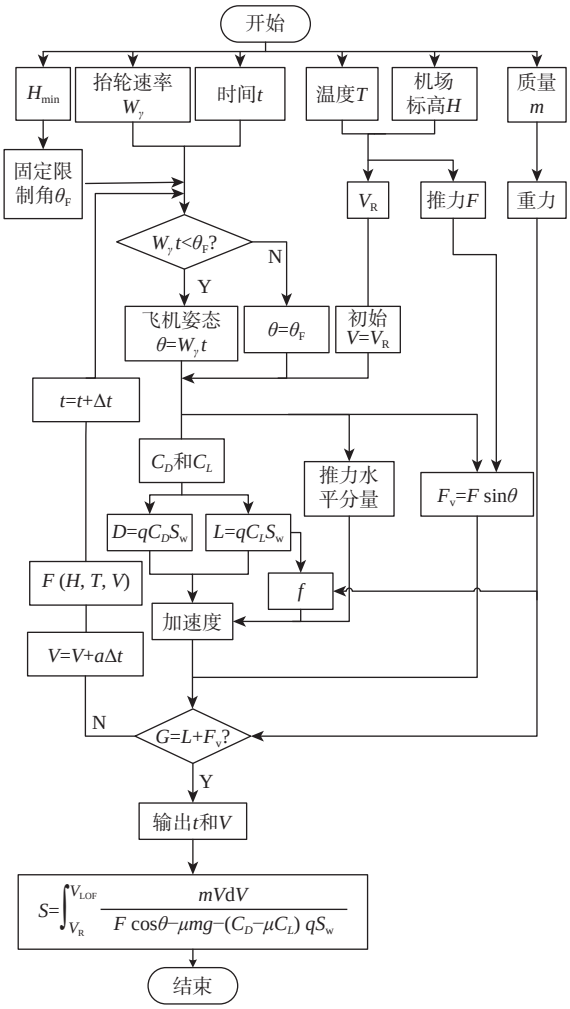


图3 抬轮段距离计算流程图

Fig. 3 Flow chart for calculating rotation distance

应用擦机尾动态限制角时，抬轮段距离的计算方法与如图3所示的流程图相似，但是需要计算抬轮段每个时刻的起落架高度，根据变化的起落架高度求出每个时刻的限制角，然后，计算出抬轮段距离。

2 典型机型算例分析

2.1 模型准确性分析

以典型机型波音 737-800 为例进行验证，机型具体参数如表1所示。

模型的对比验证数据是通过 BCOP 软件计算的起飞距离结果。BCOP 是波音公司开发并提供给航司用于计算飞机飞行航迹、爬升梯度、噪声的性能软件。依据如图3所示的流程图，时间步长Δt取值 0.1 s，计算不同的机场标高和温度下的抬轮段距离，然后与 BCOP 对比，计算结果如图4所示。

由图4可知，本文模型计算的抬轮段距离与BCOP结果接近，算例误差平均为 1.3 m，随温度、

表 1　机型参数

Table 1　Aircraft parameters

飞机质量/kg	机翼面积/m^2	标准抬轮速率/((°)·s^{-1})	后机身长度/m	起落架初始高度/cm	完全压缩后起落架高度/cm	压缩率/10^{-5}(cm·N^{-1})
67 000	124.5	3	700	136	112	3

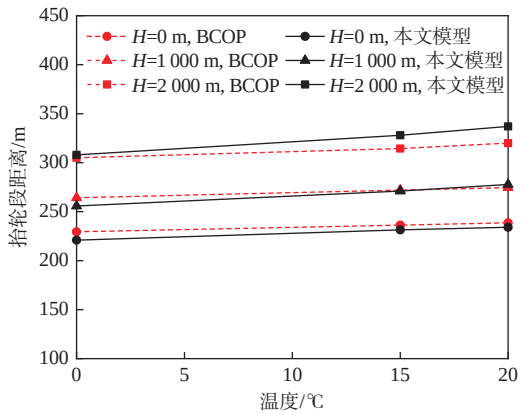


图 4　本文模型与 BCOP 计算结果对比

Fig. 4　Comparison between results of the proposed model and BCOP

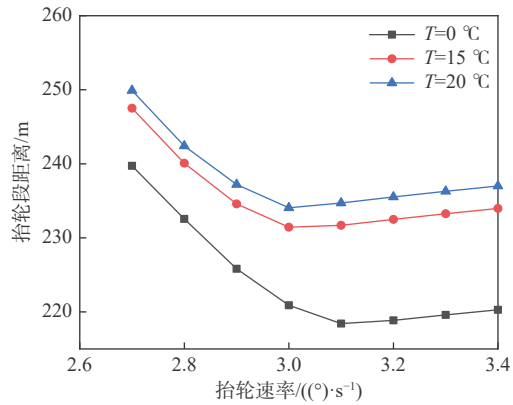


图 5　H=0 m, 抬轮段距离随抬轮速率的变化

Fig. 5　H=0 m, rotation distance varies with rotation rate

标高的变化趋势与 BCOP 吻合良好。

2.2　抬轮速率对起飞性能的影响分析

计算机场标高为 0 m 时, 不同温度情况下抬轮速率的变化对抬轮段距离的影响, 结果如图 5 所示。

由图 5 可知, 抬轮速率一定时, 温度增加, 抬轮段距离随之增加。距离增加的主要原因是温度增加, 空气密度减小, 飞机抬轮速度对应的真空速增大, 因此, 在抬轮段的滑跑过程中距离增大。还可以观察到, 随抬轮速率增大, 抬轮段距离是先减小后增加。在各温度条件下, 抬轮速率为 3.0～3.1(°)/s 时抬轮段距离最短, 起飞性能最优。

以 T=0 ℃ 为例, 飞机以不同抬轮速率抬轮时, 离地参数如表 2 所示。抬轮速率在 2.7～3.1(°)/s 范围内时, 每增加 0.1(°)/s, 距离平均减少 7 m 左右, 原因是随着抬轮速率增加, 飞机的姿态变化快, 升力系数急剧增加, 飞机离地姿态大, 推力在垂直方向的分力也提供了一定作用, 因此, 离地速度较小, 抬轮段距离短。

表 2　H=0 m, T=0 ℃, 不同抬轮速率下起飞性能参数变化

Table 2　H=0 m, T=0 ℃, variation of relevant parameters at different rotation speeds

距离抬轮开始的时间/s	θ/(°)			C_L			D/N			F_V/N		
	V_R= 2.7 (°)/s	V_R= 3.1 (°)/s	V_R= 3.4 (°)/s	V_R= 2.7 (°)/s	V_R= 3.1 (°)/s	V_R= 3.4 (°)/s	V_R= 2.7 (°)/s	V_R= 3.1 (°)/s	V_R= 3.4 (°)/s	V_R= 2.7 (°)/s	V_R= 3.1 (°)/s	V_R= 3.4 (°)/s
0	0	0	0	0.350	0.350	0.350	28 371	28 371	28 371	0	0	0
0.5	1.35	1.55	1.70	0.519	0.544	0.563	29 958	30 259	30 485	4 551	5 225	5 731
1	2.70	3.10	3.40	0.677	0.721	0.754	32 923	33 644	34 328	9 099	10 446	11 456
1.5	4.05	4.65	5.10	0.825	0.888	0.936	36 964	38 309	39 620	13 643	15 660	17 171
2	5.40	6.20	6.80	0.967	1.051	1.114	41 837	44 646	46 980	18 179	20 862	22 872
2.5	6.75	7.75	8.50	1.109	1.214	1.290	48 174	52 972	56 743	22 704	26 049	28 552
3	8.10	9.04	9.04	1.250	1.344	1.344	56 379	61 535	61 448	27 217	30 351	30 351

抬轮速率在 3.1～3.4(°)/s 范围内, 每增加 0.1(°)/s, 距离平均增加 0.6 m 左右。原因是抬轮太快, 飞机姿态过早到达限制姿态, 大姿态下的气动阻力也会变大, 加速度减小, 所以加速距离会稍微增加。

分析了 H=0 m 的情况后, 分析标高 H=1 000 m 和 H=2 000 m 时, 抬轮段距离随抬轮速率的变化如图 6 和图 7 所示。

对比图 5～图 7 可知, 机场标高增加, 抬轮段各参数随抬轮速率的变化规律与标高为 0 m 的情况基本相同。随着机场标高的增加, 抬轮段距离变大, 其原因为标高增加, 密度减小, 离地真空速增大, 另一方面, 标高增加, 飞机推力减小, 所以其影响比温度影响更大。

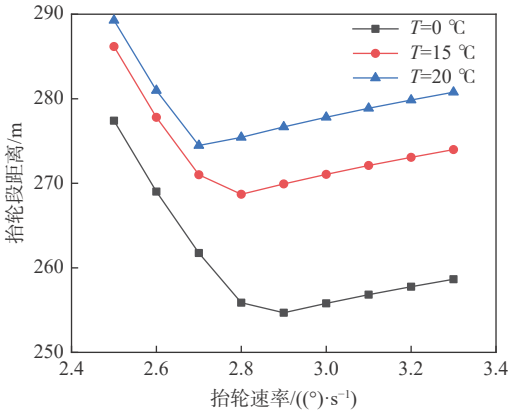


图6 H=1 000 m,抬轮段距离随抬轮速率变化

Fig. 6 H=1 000 m, rotation distance varies with rotation rate

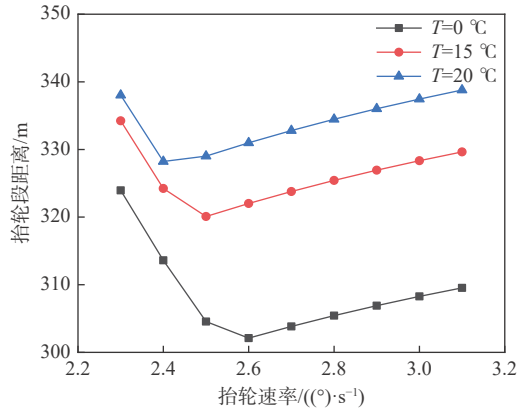


图7 H=2 000 m, 抬轮段距离随抬轮速率的变化

Fig. 7 H=2 000 m, rotation distance varies with rotation rate

此外，随标高和温度增加，最佳抬轮速率是减小的，H=0 m 时，最佳抬轮速率为 3.0～3.1(°)/s，H=1 000 m 时，最佳抬轮速率为 2.7～2.9(°)/s，H=2 000 m 时，最佳抬轮速率为 2.4～2.6(°)/s。抬轮速率小于最佳抬轮速率时，每减小 0.1(°)/s，起飞距离平均增加 7 m，在大于最佳抬轮速率时，每增加 0.1(°)/s，距离平均增加 0.8 m。不同标高、不同温度的离地参数如表 3 所示，其中，加粗部分为最佳结果。

表3 不同标高、不同温度的离地参数

Table 3 Ground clearance parameters at different elevations and temperature

抬轮速率/((°)·s⁻¹)	从抬轮到离地时间/s			速率×时间		
	H=0 m,T=0 ℃	H=1 000 m, T=15 ℃	H=2 000 m, T=20℃	H=0 m, T=0 ℃	H=1 000 m, T=15 ℃	H=2 000 m, T=20 ℃
2.3			3.91			8.98
2.4			**3.80**			**9.11**
2.5		3.56	3.81		8.91	9.51
2.6		3.46	3.83		9.01	9.95
2.7	3.27	3.38	3.85	8.83	9.13	10.39
2.8	3.18	**3.35**	3.87	8.90	**9.38**	10.82
2.9	3.09	3.37	3.88	8.96	9.77	11.26
3	3.03	3.38	3.90	9.08	10.15	11.70
3.1	**2.99**	3.40	3.91	**9.28**	10.53	12.12
3.2	3.00	3.41		9.60	10.91	
3.3	3.01	3.42		9.93	11.28	
3.4	3.02			10.26		

由表 3 可以发现，不同标高、不同温度情况下，最佳抬轮速率对应的时间都是最短的，还可以发现最佳抬轮速率与抬轮时间的乘积范围在 9.1～9.4，略大于离地限制角。随着标高和温度增加，最优抬轮速率从 3.0(°)/s 减小至 2.5(°)/s，同时，最优抬轮速率对应的离地时间增加。抬轮速率小于最佳抬轮速率时，起飞距离显著增加，在大于最佳抬轮速率时，距离增加较少。因此，起飞过程中，建议飞机以最佳抬轮速率或稍大于最佳抬轮速率抬轮。

2.3 擦机尾动态限制角的优势分析

H=2 000 m, T=20 ℃ 时，在起飞抬轮阶段，起落架的高度和机尾最低点离地高度变化曲线如图 8 所示，横坐标表示距离抬轮开始的时间。

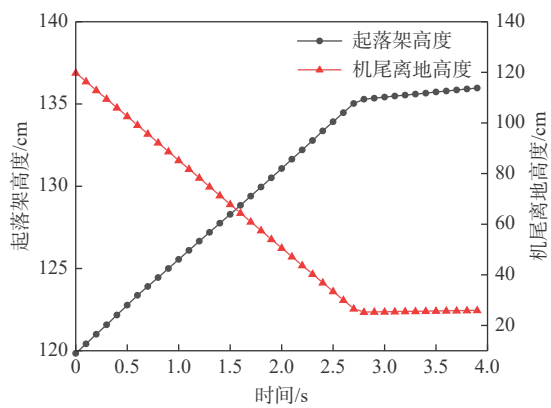


图8 起落架高度和机尾离地高度随时间变化

Fig. 8 Landing gear height and tail height above ground change over time

由图 8 可知, 开始抬轮后, 起落架的高度逐渐增加, 因为起落架是弹性装置, 刚开始是处于压缩状态, 随着速度的增加, 升力会增加, 同时由于飞机俯仰角的增加, 推力在垂直方向的分力增加, 所以机身对起落架的压力逐渐减小, 起落架会由完全压缩状态逐渐伸展变长。

如图 8 所示, 时间范围为 2.7～3.9 s 时, 在标高、温度较高时, 飞机起飞离地受擦机尾角限制, 飞机需要更长的时间和距离来达到离地速度, 因此, 起飞距离较长。

由图 8 变化曲线可知, 机尾离地高度逐渐减小后保持不变, 因为当飞机姿态逐渐增大到限制角后, 会保持俯仰角不变, 同时机尾离地高度也几乎不变。但是可以发现, 其实飞机离地时存在大约 25 cm 安全余量, 因为起落架高度在抬轮过程中会逐渐增大, 相比于刚开始起落架完全压缩状态, 机尾离地安全余量会增加, 所以采用固定限制角的方式过于保守。因此, 在保证离地最低安全余量的同时, 限制角是可以进一步增大的。

设定 h_{sD}=10 cm, 选取 H=0 m 和 2 000 m 时, 擦机尾动态限制角和擦机尾固定限制角下的抬轮段距离对比如图 9 所示。

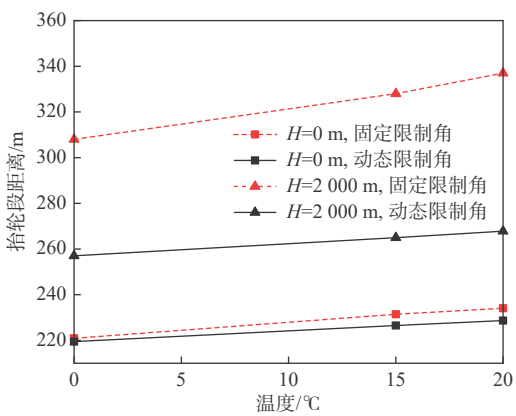


图 9　2 种限制角下的抬轮距离对比

Fig. 9　Comparison of rotation distance of two restricted angles

可以观察到, 使用动态限制角的抬轮段距离小于固定限制角, 并且高度越高, 使用动态限制角的起飞性能优势越明显。其中, 在 H=2 000 m 时, 相比于固定限制角, 使用动态限制角的抬轮段距离减小 17%以上, 从抬轮到离地的时间也缩短 20%, 总体而言, 使用动态限制角的抬轮段距离减小约 10%, 从抬轮到离地的时间缩短约 10%。

为探究擦机尾动态限制角能有效减小起飞距离的原因, 以 H=2 000 m, T=20 ℃ 时为例, 分析应用动态限制角和固定限制角时飞机俯仰角的变化规律, 如图 10 所示。

由图 10 可知, 固定限制角为 9.0°, 动态限制角

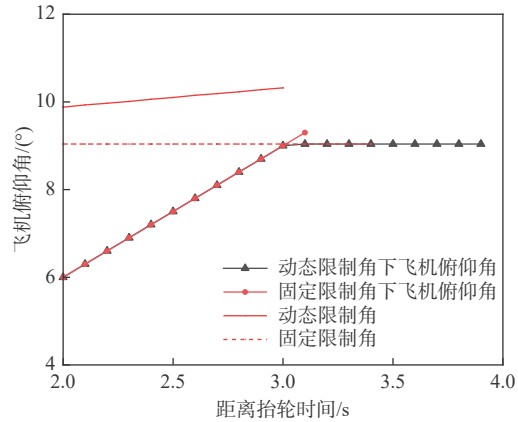


图 10　飞机俯仰角随时间变化

Fig. 10　Pitch angle of aircraft varies with time

在刚抬轮时与固定限制角相同, 随着抬轮姿态的增大, 动态限制角逐渐增加, 在 3.0 s 时增大至 10.3°。在固定限制角下, 飞机俯仰角增加到 9°后, 会保持姿态加速到飞机离地, 并且当俯仰角受限制时, 飞机速度增加较为困难, 所以需要较长的时间 3.9 s 才能离地; 而动态限制角下, 飞机俯仰角超过 9°, 最终可以达到 9.3°, 虽然俯仰角增加的幅度不大, 但俯仰角增加所引起的升力增加是显著的, 所以只需 3.1 s 就能离地, 因此, 起飞距离被大幅缩减。

2.4　应　用

在研究了擦机尾动态限制角的优势后, 可针对后机身长度和起落架高度优化设计进行研究。

在 H=1 000 m, T=15 ℃ 时, 采用固定限制角和动态限制角计算后机身长度的变化对抬轮段距离的影响, 如图 11 所示。

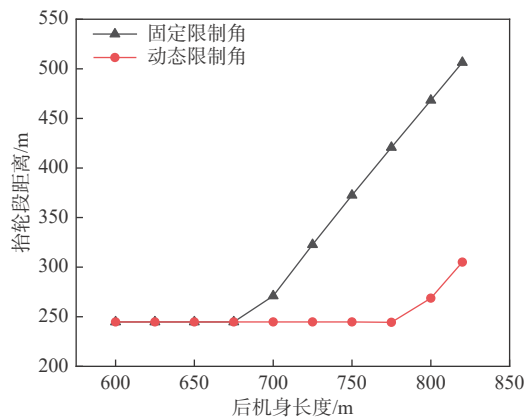


图 11　后机身长度的变化对抬轮段距离的影响

Fig. 11　Effect of change in length of rear fuselage on rotation distance

由图 11 可知, 随着后机身长度增加, 抬轮段距离先不变后增加, 拐点记为最优后机身长度。采用固定限制角时, 最优后机身长度为 680 m, 采用动态限制角时, 优后机身长度为 780 m。因此, 在设计飞机阶段, 相较于固定限制角, 采用动态限制角可适

当增加飞机的机身长度。

若改变起落架初始高度，最优后机身长度的变化如图 12 所示。

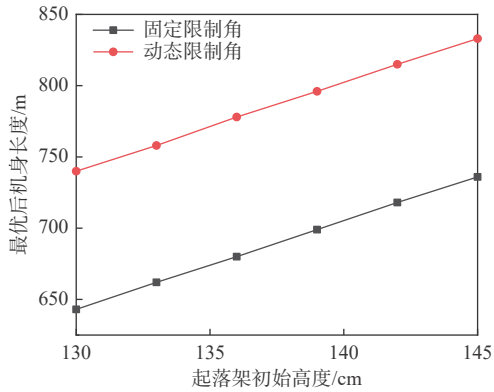


图 12 最优后机身长度与起落架初始高度的关系

Fig. 12 Relationship between length of optimal rear fuselage and height of landing gear

由图 12 可知，相比于固定限制角，采用动态限制角的最优后机身长度能增加将近 100 m。同时，最优后机身长度与起落架的初始高度几乎是线性相关的，采用擦机尾固定限制角时，最优机身长度与起落架高度的比值约为 5.0，采用动态限制角时，比值约为 5.6。因此，在设计飞机时，后机身长度 L_B 与起落架高度 H_0 的选取应是相匹配的。

3 结 论

本文以波音 737-800 为例，研究了该机型抬轮速率的变化对抬轮过程的影响，并将擦机尾角的变化考虑在内，得到的具体结论如下：

1）大多数情况下，最优抬轮速率范围为 2.5～3.0(°)/s，随着标高和温度的增加，最优抬轮速率变小。

2）抬轮速率小于最佳抬轮速率时，每减小 0.1(°)/s，起飞距离平均增加 7 m，在大于最佳抬轮速率时，每增加 0.1(°)/s，距离平均增加 0.8 m。因此，起飞过程中，建议飞机以最佳抬轮速率或稍大于最佳抬轮速率抬轮。

3）相比于固定限制角，使用动态限制角的抬轮段距离减小 10% 左右，从抬轮到离地的时间缩短 10% 左右。

4）在设计飞机时，后机身长度的选取与起落架初始高度应是相匹配的，两者的比值 L_B/H_0 约为 5.0，若采用擦机尾动态限制角，比值应略微增大。

参考文献（References）

[1] Boeing. Statistical summary of commercial jet airplane accidents worldwide operations: 1959-2014[R]. Washton, D. C.: Boeing, 2017.

[2] 中国民用航空局. 中国民航不安全事件统计分析报告[EB/OL]. (2018-07-18)[2024-03-10]. https://safety.caac.gov.cn.
Civil Aviation Administration of China. Statistical analysis report of Chinese civil aviation unsafe incidents[EB/OL]. (2018-07-18)[2024-03-10]. https://safety.caac.gov.cn(in Chinese).

[3] CHEN Y D, HAN Z, GU J. Analyzing the causes of tail strike event during takeoff with stepwise regression[C]//Proceedings of the IEEE 1st International Conference on Civil Aviation Safety and Information Technology. Piscataway: IEEE Press, 2019: 418-424.

[4] LI C X, ZHONG M Z, ZHOU X T. An algorithm for aircraft tail-strike screening based on geometric constraint model[C]//Proceedings of the IEEE 4th International Conference on Civil Aviation Safety and Information Technology. Piscataway: IEEE Press, 2022: 346-351.

[5] 孙瑞山, 杨绎煊. 飞机起飞擦机尾事件的风险预测研究[J]. 安全与环境工程, 2016, 23(2): 153-156.
SUN R S, YANG Y X. Risk prediction of aircraft tail strike events during take-off phase[J]. Safety and Environmental Engineering, 2016, 23(2): 153-156(in Chinese).

[6] 杨星月, 汪磊. 基于飞行员模型的民机擦机尾事件分析[J]. 飞行力学, 2022, 40(5): 89-94.
YANG X Y, WANG L. Analysis of civil aircraft tail striking event based on pilot model[J]. Flight Dynamics, 2022, 40(5): 89-94(in Chinese).

[7] 蔡良才, 王海服, 朱占卿, 等. 高原机场飞机起飞滑跑距离计算方法[J]. 交通运输工程学报, 2013, 13(2): 66-72.
CAI L C, WANG H F, ZHU Z Q, et al. Calculation method of running distance for aircraft takeoff on plateau airport[J]. Journal of Traffic and Transportation Engineering, 2013, 13(2): 66-72(in Chinese).

[8] 刘静, 成婷婷, 屈飞舟. 运输类飞机污染跑道起飞性能建模与分析[J]. 飞行力学, 2021, 39(2): 22-26.
LIU J, CHENG T T, QU F Z. Modeling and analysis of takeoff performance on contaminated runway for transport aircraft[J]. Flight Dynamics, 2021, 39(2): 22-26(in Chinese).

[9] 孟祥光, 王立新, 刘海良. 民机起飞爬升梯度适航符合性数学仿真评估[J]. 北京航空航天大学学报, 2016, 42(10): 2222-2230.
MENG X G, WANG L X, LIU H L. Mathematical simulation and assessment of airworthiness compliance of climb gradient during takeoff of civil aircraft[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(10): 2222-2230(in Chinese).

[10] CHAN P W. A tail strike event of an aircraft due to terrain-induced wind shear at the Hong Kong International Airport[J]. Meteorological Applications, 2014, 21(3): 504-511.

[11] CUERNO-REJADO C, SANCHEZ-CARMONA A. Preliminary sizing correlations for the rear-end of transport aircraft[J]. Aircraft Engineering and Aerospace Technology, 2016, 88(1): 24-32.

[12] BALACHANDRAN S, OZAY N, ATKINS E M. Verification guided refinement of flight safety assessment and management system for takeoff[J]. Journal of Aerospace Information Systems, 2016, 13(9): 357-369.

[13] JONATHAN A. Dynamic tail strike avoidance controller fortakeoff and landing[D]. Gothenburg: Chalmers University of Technology, 2023.

[14] OUDIN S, FROT C. Optimizing take-off performances using a protected rotation law[C]//Proceedings of the AIAA SCITECH Forum.

Reston: AIAA Press, 2022.

[15] 刘海良, 王立新. 基于数字虚拟飞行的民用飞机纵向地面操稳特性评估[J]. 航空学报, 2015, 36(5): 1432-1441.
LIU H L, WANG L X. Assessment of longitudinal ground stability and control for civil aircraft based on digital virtual flight testing method[J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(5): 1432-1441(in Chinese).

[16] CFM 国际公司. LEAP-1A 航空发动机性能手册: P/N 312-001-100-042)[M/OL]. (2022-09-15)[2025-03-10]. https://www.cfmaero-engines.com/manuals/leap-1a-performan-ce-v42.
CFM International. LEAP-1A aircraft engine performance manual: P/N 312-001-100-042)[M/OL]. (2022-09-15)[2025-03-10]. https://www.cfmaeroengines.com/manuals/leap-1a-performan-ce-v42(in Chinese).

Optimization of take-off rotation process considering tail striking dynamic limit angle

ZHUANG Nanjian*, YANG Xueya, GU Runping

(College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In order to prevent the occurrence of tail striking and optimize the take-off rotation process, this paper established a dynamic model of rotation to calculate the performance parameters such as pitch angle of the aircraft and tail ground clearance, taking into account the change of the tail striking angle during the rotation process. Using the Boeing 737-800 as an example, this study examined how the rotation rate affected the rotation distance, contrasted the tail striking dynamic limit angle with the tail striking fixed limit angle, and examined the benefits of the latter. The results show that in most cases, the optimal rotation rate ranges from 2.5 to 3.0 (°)/s. In comparison to the tail striking fixed limit angle, the rotation distance and lift-to-lift-off time are reduced by approximately 10% and 10%, respectively, with the tail striking dynamic limit angle. Finally, it is suggested that the ratio of the length of the rear fuselage to the length of the landing gear should be about 5.0, and that the ratio should be increased slightly if the tail striking dynamic limit angle is used. This study can provide pilots with a reference, improve the take-off performance of the aircraft, and increase the load of the aircraft while preventing tail striking.

Keywords: air transportation; take-off performance; tail striking; rotation rate; dynamic limit angle

Received: 2024-04-01; **Accepted**: 2024-05-15; **Published Online**: 2024-05-24 17:30

URL: link.cnki.net/urlid/11.2625.V.20240524.0944.002

Foundation items: National Key Research and Development Program of China (2023YFB4302903); General Program of Nature Science of the Fundamental Research Funds for the Central Universities (KJZ53420210102)

* **Corresponding author**. E-mail: njzhuang@cauc.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0192

绿色低碳导向的航班推出策略与牵引作业计划联合优化

寇伟彬[1,*]，张世杰[1]，王佳玉[2]
(1. 中国民航大学 交通科学与工程学院，天津 300300；2. 陆军军事交通学院，天津 300161)

摘 要：为减少离港航空器在推出滑行过程中产生的碳排放，提出一类基于绿色低碳导向的航空器牵引推出作业调度优化模型。根据航空器场面滑行状态及延误情况，在考虑航空器机坪运行规则的基础上，建立航班推出时刻动态调整机制，定义柔性推出缓冲时间，量化动态牵引推出条件，建立考虑碳排放的航空器牵引调度混合整数规划模型，结合牵引推出机-车协同调度优化计划降低场面滑行碳排放水平。考虑到整数规划模型为复杂非线性结构，求解难度较大，提出基于线性迭代的两阶段算法求解。采用天津滨海国际机场数据开展案例分析，标定调度优化模型中关键参数并验证该模型的可行性。结果表明：与现情况相比，通过对牵引推出计划的优化，航空器场面运行排放降低7.98%。研究结果对机场飞行区低碳运行及机-车协同调度提供决策支持。

关 键 词：航空运输；机场；牵引推出；线性化算法；绿色低碳
中图分类号：V355.2; U8
文献标志码：A **文章编号**：1001-5965(2026)05-1504-09

由于高峰期机场流量上升,航空器场面滑行与排队延误现象普遍,控制航空器场面运行碳排放已经成为民航运输业碳减排工作的关键[1]。航空器在滑行道滑行及排队等待时发动机持续运行会产生大量碳排放,相关研究表明,航空器在起降着陆循环过程中滑行的污染物排放占比最高,约为72%[2]。因此,降低航空器地面滑行碳排放水平是机场碳减排的核心。

当前民航业碳排放的研究主要集中在航空器碳排放预测、区域差异、演变特征和减排措施上[3]。针对机场场面运行碳减排而言,多以航空器滑行、停机位的分配及航空器的推出为研究对象,通过建立数学优化模型的方法来缓解拥堵以降低碳排放水平。杨培颖等[4]建立了针对接送机场服务中以最小碳排放量为目标的车次分配与调度模型,采用了改进的基于时间和地域划分的极线扫描算法对现实接送机场服务运作中车辆配置和设置服务水平提供了有价值的建议。王超等[5]通过引入航空器空中走向约束与航班接续约束,以减少过长滑行距离的停机位分配方案。机场跑道滑行系统容量有限,高峰时段同时推出大量航空器会导致拥堵,而航空器的推出主要依靠牵引车牵引[6],因此,在一定的时间窗范围内动态调整航空器推出时刻,合理调整航班推出控制决策,优化机-车协同调度计划,降低航空器场面滑行延误整体水平,进而控制整个地面碳排放水平是一项可行的方案。

航空器推出过程需要考虑牵引车与航空器的协同运行与资源调配,通常指牵引车根据航班的起降时刻表、停机位分布特征及相关客流属性,确定服务车次、接续关系、保障目标及相关关键时间参

收稿日期：2024-04-03；**录用日期**：2024-06-07；**网络出版时间**：2024-06-28 16:33
网络出版地址：link.cnki.net/urlid/11.2625.V.20240627.1628.002
基金项目：天津市教委科研计划 (2023KJ224)
*通信作者. E-mail：wbkou@cauc.edu.cn

引用格式：寇伟彬，张世杰，王佳玉. 绿色低碳导向的航班推出策略与牵引作业计划联合优化[J]. 北京航空航天大学学报，2026，52(5)：1504-1512.
KOU W B，ZHANG S J，WANG J Y. Joint optimization of green low-carbon oriented flight launch strategy and towing operation plan[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1504-1512 (in Chinese).

数等, 为车次分配牵引任务, 将航空器牵引到滑行道的过程[7]。Du 等[8] 面向机-车协同牵引推出调度优化问题建立一套整数规划模型, 其中, 该模型的目标为推出作业时间及牵引运行延误惩罚最小。Liu 等[9] 提出一种以总额外时间成本最小和多种服务车辆数量最少为目标的双目标机-车协同调度优化模型, 利用国内某机场一天高峰时段的航班数据进行了研究。van Oosterom 等[10] 为确保机-车协同牵引推出调度行程的最短距离, 提出基于端到端优化框架的两阶段混合整数线性规划模型。在计划推出时间的前提下, 航空器的推出时刻可根据现场情况进行动态调整, 然而当推出时刻过早, 滑行时间预留过多, 航空器在滑行道等待排队时间会增加; 当推出时刻过晚, 可能会造成航空器无法按时起飞及延误。目前, 对于航班推出时刻研究方面主要采用仿真和分析建模 2 种方法[11]。仿真方面主要通过对机场跑滑系统进行拓扑建模, 模拟机场航空器离场过程[12]; 分析建模方面主要以航空器延误时间最小、航空器排队时间最短或航空器推出成本最少等方式调整推出策略[13]。由于航空器推出过程中涉及动态时间窗、最小尾流间隔和翼展等约束, 因此, 不少学者在算法方面开发了对应的解析式及启发式求解算法(如自适应大邻域搜索)进行求解[14]。

目前关于航空器推出的研究多以考虑经济成本、推出效率与最短路径为研究目标, 较少研究航班推出与航空器滑行碳排放的耦合关系。然而, 高峰期产生的航空器在滑行道大量的排队延误与额外的碳排放, 核心原因之一是大量航空器需要在同一时段内集中推出离港, 因此, 在满足现行规定的前提下合理的调整各航班推出调度计划会对降低机场场面运行碳排放水平有着显著的影响。本文将航空器场面滑行与牵引推出调度耦合关联, 以绿色低碳为导向, 定义“柔性推出缓冲时间”的概念, 建立牵引推出作业调度优化模型, 在考虑航空器滑行碳排放的基础上, 通过柔性推出时间的动态调整, 降低场面航空器拥堵与延误, 进而达到场面运行碳减排水平的目标。

1　问题描述与建模

1.1　问题描述

航空器牵引调度作业指牵引车从车库出发前往停机位进行航空器牵引推出工作, 在完成一次航空器服务任务后, 牵引车辆根据调度计划决定是否进行下一次航空器的牵引服务或返回车库。与传统的车辆路径规划相比, 该问题属于“多对多”类型, 并允许牵引车辆进行多次作业。航空器牵引调度优化模型流程示意图如图 1 所示。

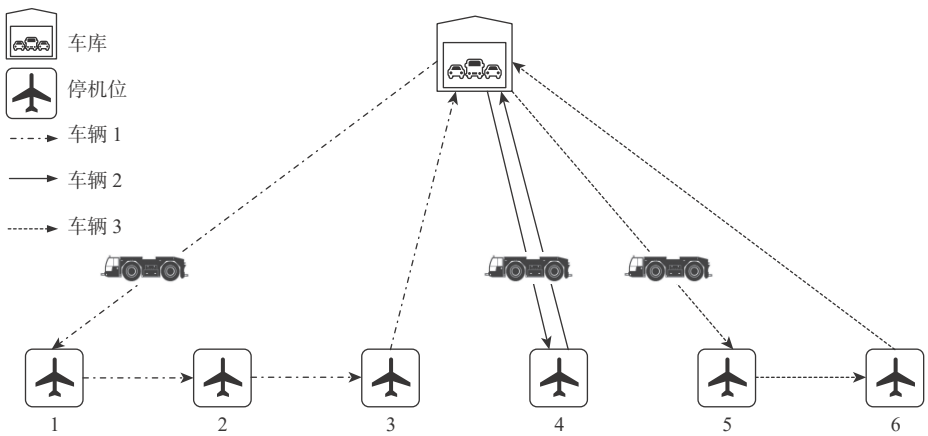


图 1　航空器牵引调度优化模型流程示意图

Fig. 1　Schematic diagram of aircraft towing scheduling optimization model process

在牵引作业过程中, 航空器牵引流程包含一系列关键时间点[15], 包括最早开始服务时刻、最晚开始服务时刻、航空器在滑行道等待时间及起飞时刻等, 关键时间点间的时间序列如图 2 所示, 其中, 航空器柔性推出缓冲时间表示在服务时间窗内, 牵引服务可根据实际需求灵活调整实际开始服务时刻, 该时间也是优化模型中重点优化的决策变量。值得注意的是, 随着机场流量的增加, 如果在同一时间段内有 2 架或更多航空器准备同时牵引推出, 就可能导致航空器需要在滑行道上产生排队延误。此时, 由于航空器发动机仍在运行, 将产生大量燃油消耗和碳排放。因此, 本文调度优化模型中航空器柔性推出缓冲时间的量化[16] 是降低航空器场面运行碳排放水平的关键因素。

1.2　本文调度优化模型构建

本文调度优化模型主要假设如下: ①每架航空

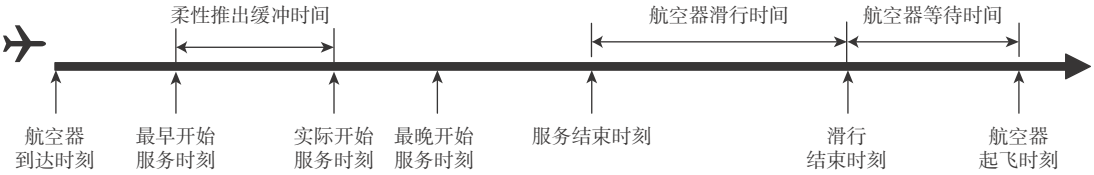


图2 航空器牵引流程时间轴

Fig. 2 Aircraft towing process timeline

器有且仅有一辆牵引车对其进行牵引推出服务；②牵引车在前往下一个节点之前需要完成上一个节点的牵引任务；③牵引车需要在规定时间窗之内完成牵引任务；④牵引车需从车库出发并最终返回车库。

基于上述假设，建立如下混合整数规划模型：

$$\begin{cases}\min Z=\omega_1 Z_1+\omega_2 Z_2\\ \text{s.t. } Z_1=\sum\limits_{k\in K}\sum\limits_{j,h\in N}x_{jhk}(T_{jhk}+S_{jk})C_k+\\ \qquad vC_v+\sum\limits_{j\in N,j\neq n_0}\left(D_j+T_j^{\text{taxiing}}\right)C_j+\\ \qquad\sum\limits_{j\in N,j\neq n_0}T^{\text{gate}}C_{\text{d}}\\ Z_2=\sum\limits_{j\in N,j\neq n_0}D_jC_j^{\text{oil}}IP\end{cases}\tag{1}$$

式中：Z_1为车辆运行成本与航空器运行成本；Z_2为航空器碳排放成本；ω_1与ω_2为加权系数；k为牵引车编号；K为牵引车辆集合；N为所有节点集合，其中，n_0节点为车库，其余节点为所需服务航空器节点；x_{jhk}为0-1变量，如果车辆k从节点j到达节点h，为1，其他情况，为0；T_{jhk}为牵引车k从节点j前往节点h所需的路程时间；S_{jk}为牵引车k服务航空器航班j所需服务时间；C_k为车辆k单位时间的行驶成本；v为车辆使用数目；C_v为单车固定启用成本；D_j为航空器j的滑行道等待时间，$D_j\geqslant 0$；T_j^{taxiing}为航空器j的滑行时间；C_j为航空器j单位时间的运行成本；d_j为航空器柔性推出缓冲时间；T^{gate}为停机位超时占用时间，当航空器超过机位最长占用时间时，为柔性推出缓冲时间减去机位最长占用时间，其他情况，为0；C_{d}为超时占用惩罚成本；C_j^{oil}为航空器j单位时间的耗油量；I为燃油碳排放系数，kg/kg；P为单位碳排放量的排放成本。

本文混合整数规划模型确保车辆从车库出发的约束如下：

$$\sum_{0,h\in N}x_{0hk}=1\qquad\forall k\in K\tag{2}$$

流量平衡约束指使车辆k进入节点j完成任务后离开节点j，具体如下：

$$\sum_{i,j\in N}x_{ijk}=\sum_{j,h\in N}x_{jhk}\qquad\forall k\in K\tag{3}$$

确保车辆最终回到车库的约束如下：

$$\sum_{j\in N}x_{j(n+1)k}=1\qquad\forall k\in K\tag{4}$$

确保一个航班只能由一辆牵引车进行服务的约束如下：

$$\sum_{k\in K}\sum_{j,h\in N}x_{jhk}=1\qquad\forall j\in N\tag{5}$$

确保车辆k先完成节j所需服务才可以前往下一个节点h，具体约束如下：

$$b_{jk}+S_{jk}+T_{jhk}-M(1-x_{jhk})\leqslant a_{hk}\qquad\forall j,h\in N\tag{6}$$

式中：M为特定参数；a_{jk}为车辆k到达节点j的时间；b_{jk}为车辆k服务节点j的开始时间。

决定了航空器离场推出柔性推出缓冲时间约束如下：

$$\sum_{k\in K}b_{jk}-L_j\leqslant d_j\qquad\forall j\in N,j\neq n_0\tag{7}$$

式中：d_j为对于航空器j的离场柔性推出缓冲时间（与L_j开始服务航空器节点j的最晚时间相关联）；L_j为开始服务航空器j的最晚时间。由于d_j是非负的，如果开始服务时间b_{jk}早于最晚服务时间L_j，则柔性推出缓冲时间为0，否则为开始服务时间b_{jk}与最晚服务时间L_j之差。

开始服务时间约束如下：

$$b_{jk}\geqslant a_{jk}\qquad\forall j\in N,j\neq n_0,k\in K\tag{8}$$

$$b_{jk}=x_{jhk}\max\left\{a_{jk},E_j\right\}\qquad\forall j\in N,j\neq n_0,k\in K\tag{9}$$

式中：E_j为开始服务离站航空器j的最早时间。如果车辆k提前到达航班j，则需要等待至最早开始服务时间E_j开始服务。

第2架航空器h的起飞时间与前一架航空器j的起飞时间间隔不小于$I_{\min}$，否则，航空器h需要在滑行道等待D_h才可起飞。具体约束如下：

$$D_h\geqslant z_{jh}\left[I_{\min}-(b_{hk}+S_{hk}+T_h^{\text{taxiing}}-T_j^{\text{takeoff}})\right]\qquad\forall j,h\in N,\ k\in K,\ j\neq h\neq n_0,\ j\neq n_1\tag{10}$$

式中：z_{jh}为0-1参数，表示航空器h是否紧接在航空器j后起飞，如果航空器h紧接在航空器j后起飞，为1，其他情况，为0；$I_{\min}$为2架航空器之间允许起飞的最小时间间隔；T_j^{takeoff}为航空器j的起飞时间。

约束(10)考虑航空器起飞间隔,减少或避免航空器滑行道等待时间,将其转移到机坪等待时间,降低航空器在滑行道等待时发动机燃油消耗从而促进机场场面低碳运行。

决策变量取值约束如下:

$$x_{jhk} \in \{0,1\} \qquad \forall j,h \in N, k \in K \tag{11}$$

$$a_{jk}, b_{jk} \geqslant 0 \qquad \forall j \in N, j \neq n_0 \tag{12}$$

$$0 \leqslant d_j \leqslant d_{\max} \qquad \forall j \in N, j \neq n_0 \tag{13}$$

$$z_{jh} \in \{0,1\} \qquad \forall j,h \in N, j \neq h \neq n_0 \tag{14}$$

式中: $d_{\max}$为航空器柔性推出缓冲的最大时间,可根据机场需求灵活调整。

2　求解算法

由于约束 (9) 与约束 (10) 为非线性结构,因此,1.2 节的混合整数规划模型是非线性规划模型,其求解难度相对较大。本文开发基于线性迭代的两阶段算法,通过 2 次线性迭代将原模型转化为线性规划模型进行求解。

1) 第 1 阶段线性化

约束 (9) 中为非线性项为 0-1 变量和一个最大值变量相乘,其中,最大值变量仍为非线性结构。先对最大值变量的部分进行处理,引入辅助决策变量y_1、y_2和β构造辅助约束变量,令$\beta = \max\{a_{jk}, E_j\}$,建立约束条件 (15)～约束条件(19):

$$\beta \geqslant a_{jk} \tag{15}$$

$$\beta \leqslant a_{jk} + M(1-y_1) \tag{16}$$

$$\beta \geqslant E_j \tag{17}$$

$$\beta \leqslant E_j + M(1-y_2) \tag{18}$$

$$y_1 + y_2 \geqslant 1 \tag{19}$$

式中: y_1为 0-1 变量,当$a_{jk} \geqslant E_j$时, y_1取 1,其他情况,取 0; y_2为 0-1 变量,当$a_{jk} \leqslant E_j$时, y_2取 1,其他情况,取 0。

当牵引车到达节点时间晚于最早开始服务时间时,即$a_{jk} > E_j$, y_1取 1, y_2取 0,由约束 (15) 和约束(16)可知, $a_{jk} \leqslant \beta \leqslant a_{jk}$,即$\beta = a_{jk}$;当牵引车到达节点时间早于最早开始服务时间时, $a_{jk} < E_j$, y_1取 0, y_2取 1,由约束 (17) 和约束 (18) 可知, $E_j \leqslant \beta \leqslant E_j$,即$\beta = E_j$;当牵引车到达节点时间等于最早开始服务时间时,即$a_{jk} = E_j$, y_1取 1, y_2取 1,由约束 (15)～约束(18) 可知$a_{jk} \leqslant \beta \leqslant a_{jk}, E_j \leqslant \beta \leqslant E_j$,即$\beta = a_{jk} = E_j$。

约束 (10) 中, 0-1 变量z_{jh}与连续变量b_{hk}相乘的二次项为非线性约束,通过引入辅助决策变量$y_3 (y_3 \geqslant 0)$,令$y_3 = z_{jh} b_{hk}$,构造辅助约束条件如下:

$$y_3 \leqslant M z_{jh} \tag{20}$$

$$y_3 \leqslant b_{hk} \tag{21}$$

$$y_3 \geqslant b_{hk} - M(1 - z_{jh}) \tag{22}$$

式中: 0-1 变量z_{jh}表示航空器h是否紧接在航空器j后起飞;当z_{jh}取 1,由约束 (21) 和约束(22)可知, $b_{hk} \leqslant y_3 \leqslant b_{hk}$,即$y_3 = b_{hk}$,此时$y_3 = z_{jh} b_{hk}$;当$z_{jh}$取 0,由约束 (20) 可知, $y_3 \leqslant 0$,即$y_3 = 0$,此时$y_3 = z_{jh} b_{hk}$。

2) 第 2 阶段线性化

经过第 1 阶段线性化,约束 (9) 变为$b_{jk} = x_{jhk}\beta$,其中, 0-1 变量x_{jhk}与连续变量β相乘的二次项为非线性约束,同理引入辅助决策变量$y_4 (y_4 \geqslant 0)$,令$y_4 = x_{jhk} b$构造辅助约束条件如下:

$$y_4 \leqslant M x_{jhk} \tag{23}$$

$$y_4 \leqslant \beta \tag{24}$$

$$y_4 \geqslant \beta - M(1 - x_{jhk}) \tag{25}$$

式中: 0-1 变量x_{jhk}表示车辆k是否服务节点j后前往节点h,当x_{jhk}取 1,由约束 (23) 和约束(24)可知, $\beta \leqslant y_4 \leqslant \beta$,即$y_4 = \beta$,此时$y_4 = x_{jhk} b$;当$x_{jhk}$取 0,由约束 (25) 可知$y_4 \leqslant 0$,即$y_4 = 0$,此时$y_4 = x_{jhk} b$。

通过式 (15)～式 (25) 的线性迭代,使得模型线性化,大大降低模型复杂程度,最后,通过商业求解器 Gurobi 在 Python 中编码进行求解。

3　实例分析

3.1　案例数据

以天津滨海国际机场为案例,选取某天 9:00—11:00 时段近、远机位共 35 个航班放行数据及相应的停机位情况,表 1 包括航空器序号、停机位序号、最早开始服务时间、计划推出时间窗、柔性推出缓冲时间上限等信息。对于柔性推出缓冲时间的约束,下限按照计划推出时间窗最小值确定,上限则根据《民航航班正常统计办法》确定为 10 min[17]。牵引车在机位之间行驶的速度及航班接受服务的时间要求按实际机场规定确定。机场的平面布局如图 3 和图 4 所示,其中,案例机坪区域共包括 35 个机位用于停放航空器,牵引车库位置位于 3 号停机位附近。此外,任意节点之间距离和走向由机场规定的飞行区内特种车辆路径所确定。

案例具体参数如表 2 所示,根据《民用航空空中交通管理规则》(CCAR-93TM-R4)相关规定,将航空器最小安全起飞间隔设定为 3 min[18]。单车固定启用成本取值为 1 000 元,车辆单位时间运行成本取值

表 1 航班基础数据

Table 1 Basic flight data

航空器序号	停机位序号	最早开始服务时刻	计划推出时刻	柔性推出缓冲时刻上限
1	1	9:06	9:11	9:16
2	2	9:20	9:25	9:30
3	3	9:00	9:05	9:10
4	4	9:08	9:16	9:18
5	5	9:33	9:38	9:43
6	6	10:07	10:12	10:17
7	7	10:20	10:25	10:30
8	8	9:04	9:09	9:14
9	9	9:35	9:40	9:45
10	10	9:23	9:28	9:33
11	11	9:53	9:58	10:03
12	12	9:02	9:07	9:12
13	13	10:17	10:22	10:27
14	14	10:04	10:09	10:14
15	15	10:59	11:04	11:09
16	16	9:26	9:31	9:36
17	17	9:10	9:15	9:20
18	18	9:14	9:19	9:24
19	19	9:17	9:22	9:27
20	20	9:25	9:30	9:35
21	21	9:28	9:33	9:38
22	22	9:54	9:59	10:04
23	23	9:38	9:43	9:48
24	24	9:57	10:02	10:07
25	25	9:40	9:45	9:50
26	26	10:01	10:06	10:11
27	27	10:10	10:16	10:20
28	28	10:14	10:19	10:24
29	29	10:33	10:38	10:43
30	30	10:23	10:28	10:33
31	31	10:40	10:45	10:50
32	32	10:57	11:02	11:07
33	33	10:46	10:51	10:56
34	34	10:48	10:53	10:58
35	35	10:53	10:58	11:03

为 1 元/min。航空器单位滑行成本取 133.76 元/min[19]，航空器停机位最长占用时间取 10 min[20]。参考已有研究成果[21]，燃油碳排放系数取 3.115 kg/kg；碳排放量的排放成本取 0.28 元/kg。



图 3 天津滨海国际机场平面布局

Fig. 3 Layout of Tianjin Binhai International Airport

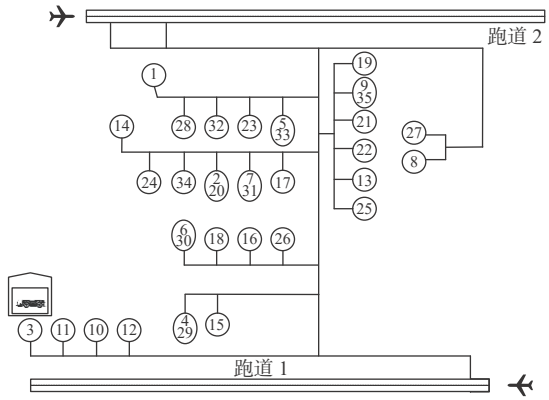


图 4 机场场面网络无向图

Fig. 4 Undirected graph of airport surface networks

3.2 案例结果分析

表 3 为牵引调度方案服务次序，括号内为航空器柔性推出缓冲时间（单位：min），其中，共派出 8 辆牵引车服务 35 架次航空器。图 5 为优化方案对应的调度计划甘特图，由图可知，全部牵引车在 150.08 min 的时间内完成牵引作业，其中，第 8 辆牵引车完成服务回库时间最晚，第 6 辆牵引车完成服务回库时间最早。此外，除 4 号、6 号牵引车外，其余牵引车所牵引航空器均存在柔性推出缓冲时间，以保证航空器滑行碳排放的优化。为更清晰的反映场面运行碳排放情况，做出最优调度方案下各航空器滑行碳排放如图 6 所示。其中，第 5 架航空器碳排放量最高，第 10 架航空器碳排放量最低。

优化结果如表 4 所示，与现状调度方案相比，目标函数值由 88 568.16 元降低到 81 820.80 元，降低 7.62%；牵引作业总时间由 392.61 min 降低到 367.24 min，降低 6.46%；航空器滑行碳排放量由 27 427.58 kg 降低到 25 238.70 kg，降低了 7.98%。优化后的调度计划成功降低了碳排放，提高了场面运行效率。

表 2 案例参数

Table 2 Case parameters

单车固定启用成本/元	车辆运行成本/(元·min^{-1})	燃油碳排放系数/(kg·kg^{-1})	碳排放量的排放成本/(元·kg^{-1})	最小安全起飞间隔/min	最晚推出时刻/min	航空器单位运行成本/(元·min^{-1})	航空燃油价格/(元·kg^{-1})
1 000	1	3.115	0.28	3	20	133.76	4.5

表 3　优化后航空器牵引调度方案

Table 3　Aircraft traction scheduling scheme

牵引车辆编号	牵引航空器序号
1	18→24→14(0.18)→34
2	17→2→20(0.9)→31→7(0.9)
3	16→6→30→26(4.2)
4	8→25→22→13
5	12→10→11→3(5.82)→15(2.58)
6	4→29
7	19→9→21(0.84)→27→35
8	5→23→28→1(5.22)→32(0.96)→33(3.3)

为进一步探究柔性推出缓冲时间对航空器滑行碳排放的降低效果, 在案例中增加了当柔性推出缓冲时间取上限时的航空器滑行碳排放优化结果, 并与最优模型下的方案进行对比, 定义为“最大柔性推出时间方案”。所得结果如表 4 所示, 该方案中航空器滑行碳排放量为 26 415.20 kg, 高于本文调度优化模型 1 176.5 kg, 柔性推出缓冲时间取得上限时, 航空器滑行碳排放未继续降低, 可以看出, 在一定区间内最合理的柔性推出缓冲时间可以最大程度的降低航空器滑行过程中的碳排放量, 过度延长柔性推出缓冲时间可能会带来相反的效果。

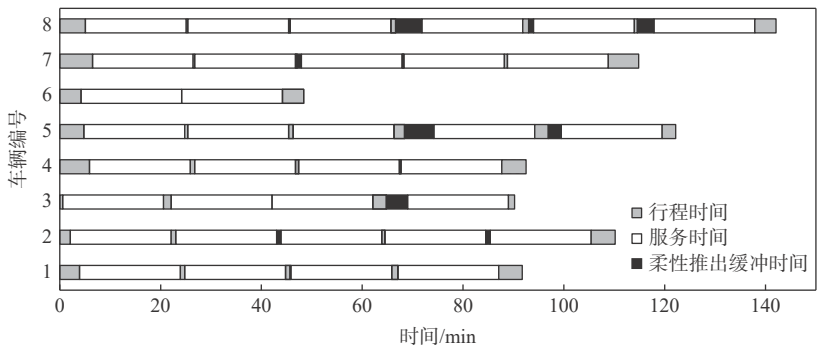


图 5　牵引调度时间轴

Fig. 5　Traction scheduling timeline

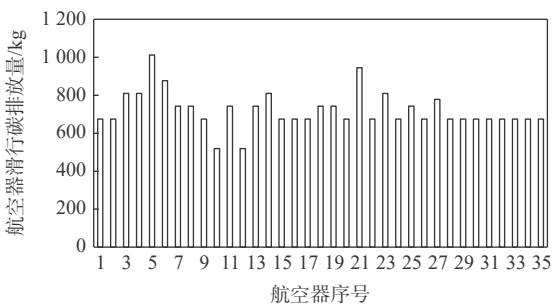


图 6　航空器滑行碳排放量

Fig. 6　Aircraft taxiing carbon emissions

表 4　优化结果

Table 4　Optimization results

牵引调度模型	牵引作业总时间/min	航空器滑行碳排放量/ kg	目标函数值/元
现状	392.61	27 427.58	88 568.16
本文调度优化模型	367.24	25 238.70	81 820.80
最大柔性推出时间方案	389.44	26 415.20	83 239.80

3.3　灵敏度分析

3.3.1　车辆数目灵敏度分析

实际案例中牵引车数量为 8 辆, 本文进一步研究牵引车数量对于场面碳排放量化影响, 取牵引车车辆数分别为 5～12 辆, 优化结果如表 5 所示, 可以看出, 随着车辆数目的增加, 车辆运行成本不断增加, 而航空器滑行碳排放量与航空器成本均先减小后增加, 其中, 当车辆数目为 9 时, 降低至最小值。分析其中原因, 开始阶段随着车辆数目增加, 牵引推出效率增加, 航空器排队等待时间降低, 所以航空器碳排放及航空器运行成本均会降低。当车辆数目超过 9 时, 牵引车数目满足已航空器推出需求, 在并不会显著增加航空器推出效率的前提下, 不断增加的车辆运行成本会导致碳排放优化结果反而变差, 因此, 可以得出, 本案例中当牵引车数目为 9 时为最合适的牵引推出方案。

3.3.2　航空器数目灵敏度分析

实际案例中航空器数目为 35 架, 为研究对于不同数目航空器对于场面碳排放量化影响, 取航空器数目分别为 20, 25, 30, 35, 40 架, 运行牵引调度模型, 计算航空器整体碳排放。如表 6 所示, 可以看出, 随着航空器数目的上升, 航空器滑行碳排放量和目标函数值优化幅度不断增大, 当航空器数目为 40 时, 航空器滑行碳排放量和目标函数值优化百分比最高, 分别为 15.01% 与 12.38%。可以看出, 随着待推出航空器数目上升, 滑行道排队现象愈发严重, 通过确定合理的柔性推出缓冲时间对降低航空器场面滑行的碳排放优化效果也越来越明显。

3.3.3　节假日高峰期场景案例分析

为验证多场景下本文调度优化模型对降低场面

表5 不同车辆数的优化结果

Table 5 Optimization results for different numbers of vehicles

车辆数目	航空器滑行碳排放量/kg	车辆运行成本/元	航空器运行成本/元
5	27 770.22	5 259.97	42 827.50
6	25 898.56	6 263.62	41 730.00
7	25 840.19	7 266.59	41 535.00
8	25 238.70	8 268.32	37 147.50
9	23 529.20	9 273.36	36 660.00
10	25 723.45	10 274.32	37 160.68
11	25 752.63	11 283.87	37 202.84
12	25 956.93	12 291.36	37 497.97

表6 不同航空器数目的优化结果

Table 6 Optimization results for different numbers of aircrafts

航空器数目	优化前航空器滑行碳排放量/kg	优化后航空器滑行碳排放量/kg	碳排放量优化百分比/%	优化前目标函数值/元	优化后目标函数值/元	目标函数值优化百分比/%
20	15 658.06	15 050.64	3.88	49 074.00	47 319.00	3.58
25	19 842.55	18 627.70	6.12	62 005.20	58 495.20	5.66
30	24 263.25	22 339.74	7.93	74 974.25	70 426.20	6.07
35	27 427.58	25 238.70	7.98	88 568.16	81 820.80	7.62
40	33 727.66	28 665.79	15.01	111 834.90	97 989.90	12.38

碳排放量的有效性，案例选取同年五一节假日前一天7:00—9:00高峰时段的43个航班场景进行分析。

本文调度优化模型优化结果如表7所示，与现状调度方案相比，目标函数值降低了14.24%，牵引作业总时间降低了10.49%，航空器滑行碳排放量降低了15.35%。节假日高峰时段43架推出场景经过优化之后，对比平峰期航空器碳排放量，优化幅度由7.98%增加到了15.35%，节假日高峰期得到更好的碳排放优化效果。因此，后续高峰期在制定联合作业策略时，可通过优化牵引车数目、扩大柔性推出缓冲时间窗等策略来进一步加强对航空器滑行碳排放量的优化效果。

表7 高峰期航空器优化结果

Table 7 Peak time optimization results for aircraft

牵引调度模型	牵引作业总时间/min	航空器滑行碳排放量/kg	目标函数值/元
现状	539.81	43 415.31	145 257.45
本文调度优化模型	483.17	36 749.21	124 575.90

3.4 算法对比分析

为验证两阶段线性迭代算法的可行性，本文引入遗传算法进行求解并开展效果对比分析。图7为遗传算法收敛图，设定种群大小为100，终止进化代数设定为300，交叉概率设定为0.8，变异概率设定为0.1，遗传算法计算结果如表8所示。可以看出，线性迭代算法较遗传算法目标函数值减少0.64%，求解时间减少3 min，因此，本文提出的两阶段线性迭代算法相较于传统启发式算法在求解性能与求解效率上均具备优势。

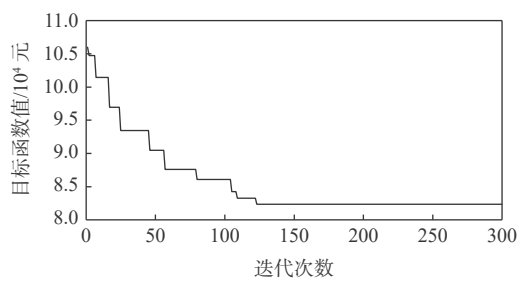


图7 遗传算法收敛

Fig. 7 Genetic algorithm convergence graph

表8 算法对比结果

Table 8 Comparison of algorithms

算法类型	求解时间/ min	目标函数值/元
遗传算法	6	82 347.20
线性迭代算法	3	81 820.80

4 结 论

本文基于绿色低碳导向，通过优化牵引车调度计划与推出时间将航空器在滑行道的等待时间减少，构建了机场航空器牵引调度问题的混合整数规划模型，并以天津滨海国际机场进行了案例分析，所得主要结论如下：

1）定义了航空器柔性推出缓冲时间，并通过案例验证证明，结合柔性推出策略的绿色低碳导向的

机场牵引调度计划优化能够有效的降低机场场面碳排放, 为机场场面低碳运行及调度策略提供决策支持。

2) 建立考虑碳排放的航空器牵引推出作业调度优化模型, 考虑动态时间窗, 对柔性推出缓冲时间进行量化。针对双重复杂非线性模型, 开发了基于线性迭代的两阶段算法。

3) 以天津滨海国际机场为实例进行分析。本文调度优化模型优化下的航空器滑行碳排放量较现状方案降低 7.98%。对于航空器牵引调度计划的车辆数目进行灵敏度分析, 存在最优的牵引车数目使得目标函数值最小。

本文以机场牵引车调度计划及航空器场面碳排放量降低为研究对象, 模型局限在航空器牵引推出作业, 未来的研究可考虑航空器牵引推出作业与其他机场运行环节的协同优化。

参考文献（References）

[1] 李耀辉, 于敬磊, 谷娟, 等. 民航飞行 CO_2 排放及其对气候变化影响研究综述[J]. 地球科学进展, 2023, 38(1): 9-16.
LI Y H, YU J L, GU J, et al. Review of research on CO_2 emissions from civil aviation flights and its impact on climate change[J]. Advances in Earth Science, 2023, 38(1): 9-16(in Chinese).

[2] BAAREN E V, ROLING P C. Design of a zero emission aircraft towing system[C]//Proceedings of the AIAA Aviation Forum. Reston: AIAA, 2019.

[3] 胡荣, 王德芸, 冯慧琳, 等. 碳达峰视角下的机场航空器碳排放预测[J]. 交通运输系统工程与信息, 2021, 21(6): 257-263.
HU R, WANG D Y, FENG H L, et al. Prediction of aircraft CO_2 emission from perspective of CO_2 emission peak[J]. Journal of Transportation Systems Engineering and Information Technology, 2021, 21(6): 257-263(in Chinese).

[4] 杨培颖, 唐加福, 于洋, 等. 面向最小碳排放量的接送机场服务的车辆路径与调度[J]. 自动化学报, 2013, 39(4): 424-432.
YANG P Y, TANG J F, YU Y, et al. Minimizing carbon emissions for vehicle routing and scheduling in picking up and delivering customers to airport service[J]. Acta Automatica Sinica, 2013, 39(4): 424-432(in Chinese).

[5] 王超, 任云鸿. 面向节油减排的平行多跑道混合运行机场停机位分配模型[J]. 交通信息与安全, 2021, 39(5): 144-152.
WANG C, REN Y H. A model of gate allocation for parallel multi-runway hybrid operation from the perspective of fuel-saving and carbon emission reduction[J]. Journal of Transport Information and Safety, 2021, 39(5): 144-152(in Chinese).

[6] 姜雨, 王雅莎, 薛清文, 等. 基于碳排放的航空器时空滑行优化研究[J]. 华南理工大学学报 (自然科学版), 2023, 51(10): 152-159.
JIANG Y, WANG Y S, XUE Q W, et al. Research on space-time taxiing optimization of aircraft based on carbon emission[J]. Journal of South China University of Technology (Natural Science Edition), 2023, 51(10): 152-159(in Chinese).

[7] BAO D W, ZHOU J Y, ZHANG Z Q, et al. Mixed fleet scheduling method for airport ground service vehicles under the trend of electrification[J]. Journal of Air Transport Management, 2023, 108: 102379.

[8] DU J Y, BRUNNER J O, KOLISCH R. Planning towing processes at airports more efficiently[J]. Transportation Research Part E: Logistics and Transportation Review, 2014, 70: 293-304.

[9] LIU Y H, WU J J, TANG J, et al. Scheduling optimisation of multi-type special vehicles in an airport[J]. Transportmetrica B: Transport Dynamics, 2022, 10(1): 954-970.

[10] VAN OOSTEROM S, MITICI M, HOEKSTRA J. Dispatching a fleet of electric towing vehicles for aircraft taxiing with conflict avoidance and efficient battery charging[J]. Transportation Research Part C: Emerging Technologies, 2023, 147: 103995.

[11] 姜雨, 刘振宇, 胡志韬, 等. 大型机场进场航空器联合调度模型[J]. 交通运输工程学报, 2022, 22(1): 205-215.
JIANG Y, LIU Z Y, HU Z T, et al. Coordinated scheduling model of arriving aircraft at large airport[J]. Journal of Traffic and Transportation Engineering, 2022, 22(1): 205-215(in Chinese).

[12] 赵文涛, 万莉莉, 彭秋萍. 基于航班离场成本的离场航空器滑行策略优化[J]. 科学技术与工程, 2021, 21(16): 6930-6936.
ZHAO W T, WAN L L, PENG Q P. Optimization of taxiing strategy of departing aircraft based on flight departure cost[J]. Science Technology and Engineering, 2021, 21(16): 6930-6936(in Chinese).

[13] 王博, 王剑辉, 彭笑非, 等. 基于机场协同决策系统的机坪牵引车调度方法[J]. 科学技术与工程, 2021, 21(4): 1667-1673.
WANG B, WANG J H, PENG X F, et al. The method for scheduling towing tractors based on A-CDM system[J]. Science Technology and Engineering, 2021, 21(4): 1667-1673(in Chinese).

[14] 张文义, 唐雨拉尔, 王旭兰, 等. 考虑双时间窗特性的机场多车型摆渡车调度优化[J]. 北京航空航天大学学报, 2025, 51(10): 3345-3353.
ZHANG W Y, TANG Y L E, WANG X L, et al. Optimization of multi-type airport ferry vehicles schedules with double service time windows characteristic[J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(10): 3345-3353(in Chinese).

[15] 张颖, 黎晓英, 尹嘉男, 等. 基于概率滑行时间的航空器离场推出柔性控制[J]. 航空学报, 2023, 44(10): 207-220.
ZHANG Y, LI X Y, YIN J N, et al. Flexible control of aircraft departure pushback time based on probabilistic taxiing time[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(10): 207-220(in Chinese).

[16] 赵俊, 陈旭梅, 刘志硕, 等. 航权资源双边谈判的序贯互惠博弈建模与分析[J]. 哈尔滨工业大学学报, 2022, 54(3): 1-11.
ZHAO J, CHEN X M, LIU Z S, et al. Modeling and analysis of sequential reciprocal game problem of bilateral negotiation for air traffic rights resources[J]. Journal of Harbin Institute of Technology, 2022, 54(3): 1-11(in Chinese).

[17] 中国民航局. 民航航班正常统计办法[EB/OL]. (2013-10-21) [2024-03-30]. http://www.caac.gov.cn/XXGK/XXGK/ZFGW/202406/P020240611543423562244.pdf.
Civil Aviation Administration of China. Statistical method for air transport flight normalcy[EB/OL]. (2013-10-21) [2024-03-30]. http://www.caac.gov.cn/XXGK/XXGK/ZFGW/202406/P020240611543423562244.pdf(in Chinese).

[18] 中华人民共和国交通运输部. 民用航空空中交通管理规则: CCAR-93TM-R4[S]. 北京: 中华人民共和国交通运输部, 2015: 83-88.
Ministry of Transport of the People's Republic of China. Civil aviation air traffic management regulations: CCAR-93TM-R4[S]. Beijing: Ministry of Transport of the People's Republic of China, 2015: 83-88(in Chinese).

[19] 郑丽君, 胡荣, 张军峰, 等. 高峰时段下离港航空器绿色滑行策略设计与评价[J]. 北京航空航天大学学报, 2019, 45(11): 2320-2326.
ZHENG L J, HU R, ZHANG J F, et al. Design and evaluation of green taxiing strategy for departure aircraft during peak hours[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(11): 2320-2326(in Chinese).

[20] 张亚平, 廉冠, 邢志伟, 等. 飞机推出控制停机位等待惩罚策略[J]. 哈尔滨工业大学学报, 2018, 50(3): 39-45.
ZHANG Y P, LIAN G, XING Z W, et al. Aircraft departure push-back control strategy based on gate-hold penalty[J]. Journal of Harbin Institute of Technology, 2018, 50(3): 39-45(in Chinese).

[21] 朱佳琳, 胡荣, 张军峰, 等. 中国航空器碳排放测算与演化特征研究[J]. 武汉理工大学学报 (交通科学与工程版), 2020, 44(3): 558-563.
ZHU J L, HU R, ZHANG J F, et al. Research on the measurement and evolution characteristics of aircraft carbon emissions in China[J]. Journal of Wuhan University of Technology(Transportation Science & Engineering), 2020, 44(3): 558-563(in Chinese).

Joint optimization of green low-carbon oriented flight launch strategy and towing operation plan

KOU Weibin[1,*], ZHANG Shijie[1], WANG Jiayu[2]

(1. School of Traffic Science and Engineering, Civil Aviation University of China, Tianjin 300300, China;
2. Army Military Transportation University, Tianjin 300161, China)

Abstract: To reduce carbon emissions during the departure taxiing process of outbound aircraft, this paper proposes a type of optimization model for aircraft towing operations based on a green and low-carbon approach. The flexible launch buffer time is defined, the dynamic traction launch conditions are quantified, the aircraft traction scheduling mixed integer programming model taking carbon emissions into account, and the dynamic adjustment mechanism of flight launch time is established based on the operation rules of the aircraft apron based on the taxiing state and delay of the aircraft surface. The carbon emission level of surface taxiing is reduced by combining the optimization of the traction roll-out motor-vehicle collaborative scheduling plan. The model integrates the optimization of aircraft-towing dispatch plans to reduce carbon emissions during apron taxiing. Considering the complexity and non-linearity of the model, a two-stage algorithm based on linear iteration is proposed for solving. Tianjin Binhai International Airport data is used for case studies, and the model's viability is confirmed and important model parameters are calibrated. The results of the case study indicate that, compared to the current situation, optimizing the towing departure plan reduces emissions during apron operations by 7.98%. The research findings provide decision support for achieving low-carbon operations in the airport apron area and optimizing aircraft-towing dispatch plans.

Keywords: air transportation; airport; towing; linearization algorithm; green and low-carbon

Received: 2024-04-03; **Accepted**: 2024-06-07; **Published Online**: 2024-06-28 16:33
URL: link.cnki.net/urlid/11.2625.V.20240627.1628.002
Foundation item: Scientific Research Program of Tianjin Municipal Education Commission (2023KJ224)
* **Corresponding author**. E-mail: wbkou@cauc.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0193

模拟与模数混合芯片宽带传导抗扰度量化仿真建模方法

陈曦[1]，谢树果[1,*]，魏梦圆[1]，李圆圆[2]

(1. 北京航空航天大学 电子信息工程学院，北京 100191；2. 北京航空航天大学 网络空间安全学院，北京 100191)

摘 要：针对目前模拟与模数混合芯片在射频（RF）传导敏感性建模方法集成电路传导抗扰度模型（ICIM-CI）中存在模型非线性特性不足、预测与实际电路在特定干扰条件下的表现存在较大偏差及无法开展级联量化敏感性仿真等问题，提出一种基于多谐波失真理论的模拟和模数混合芯片传导敏感度ICIM-CI改进建模方法，并集成电路传导抗扰度模型-非线性敏感行为模块（ICIM-CI-NIB）。该方法通过在敏感度行为（IB）模块中引入多谐波非线性参数，显著提升了传导敏感模型的非线性特性，将ICIM-CI模型从二值判定结果模型改进为具有非线性量化输出响应的级联仿真模型。通过采用ICIM-CI-NIB方法，能够快速生成芯片的传导敏感性频域宽带模型，具备在宽频大功率干扰注入下的高精度传导敏感性仿真模型，并支持级联仿真。通过典型模拟与模数混合芯片的实测结果显示：相较于ICIM-CI方法，所提方法归一化均方根误差（NMSE）提高了18.5 dB，同时建模时间减少了约98%。

关 键 词：模拟与模数混合芯片；传导敏感性；直接功率注入；电磁干扰；建模；传导抗扰度

中图分类号：TN401; TN453; TN601; V234

文献标志码：A **文章编号**：1001-5965(2026)05-1513-10

在电子设备的设计过程中，确保其通过敏感性测试以保障电磁安全是一项重要任务。仿真验证和评估是一种高效且快捷的方法。特别是对于电子设备上放置的集成电路(integrated circuit, IC)的抗扰度评估尤为重要，因为这会直接影响电子设备的电磁安全性和整体性能。因此，准确预测和评估集成电路在电磁干扰下的响应变得至关重要[1]。Loeckx等[2]的研究表明，即便是非常微小的电路拓扑结构变化，也能在整体的电磁抗扰性射频功率直接注入法(direct radio frequency power injection, DPI)[3]测试中显著提升电路的抗扰性，这强调了在设计阶段考虑电磁兼容性的重要性。

但使用芯片制造商提供芯片的功能性模型，通常无法满足电磁兼容性仿真的需求，这是因为功能性模型通常只关注芯片的功能和逻辑，而不直接关注芯片在特定电磁环境下的响应和性能。而如何评估芯片在电路上受到外部电磁干扰(electromagnetic interference, EMI)时的响应行为和稳定性，往往无法通过功能性模型实现。但芯片制造商往往不会提供芯片在特定干扰下的抗扰性仿真模型，这使得电路设计人员在进行敏感性仿真时面临挑战，特别是需要对芯片在受到宽频段大功率干扰信号时响应行为的仿真。因此，研究集成电路的传导敏感性行为模型[4]，特别是所谓的“黑盒模型”，成为电路设计领域的一项重要工作。“黑盒模型”在提取过程无需了解器件内部物理细节，而是通过输入输出信号的映射关系，获得抽象的数学表达式来描述记录到的输入信号 $x(t)$ 与输出信号 $y(t)$ 之间的关系。

收稿日期：2024-04-07；**录用日期**：2024-05-17；**网络出版时间**：2024-06-07 13:36

网络出版地址：link.cnki.net/urlid/11.2625.V.20240607.1032.002

基金项目：国家自然科学基金 (62202030)

* **通信作者**. E-mail：xieshuguo@buaa.edu.cn

引用格式：陈曦，谢树果，魏梦圆，等. 模拟与模数混合芯片宽带传导抗扰度量化仿真建模方法 [J]. 北京航空航天大学学报，2026，52（5）：1513-1522. CHEN X，XIE S G，WEI M Y，et al. Simulation modeling methodology for broadband conducted immunity quantization of analog and analog-digital hybrid chips [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（5）：1513-1522（in Chinese）.

当黑盒模型和被测件受到相同的输入信号激励时，黑盒模型的输出应尽可能接近被测件的实际响应。

在集成电路传导敏感性“黑盒模型”建模领域，已有部分学者在时域领域内开展了深入的研究。如人工神经网络(artificial neural networks, ANN)[5-6]、Volterra级数[7]、包络域模型[8]、双路径记忆模型[9]等。这些建模方法可以构建芯片在时域下的敏感性行为模型。这些模型在处理单一频点受扰后的电磁干扰问题时能够提供较为准确的输出，尤其在模拟芯片故障(如芯片抖动、输出波动)等方面表现出色。但在宽带电磁兼容性评估中，时域建模方法通常需要大量的测试和表征[10]，这表明时域黑盒模型在提供宽带仿真信息方面存在不足，这促使研究者探索新的解决方案。其中，集成电路传导抗扰度模型(integrated circuit immunity model-conducted immunity, ICIM-CI)[11]便是这类尝试之一，其在频域[12]分析方面表现出一定的优势，依赖于敏感度行为(immunity behavior, IB)模块和电源分布网络(power distribution network, PDN)的方法[13]，能够给出芯片宽带抗扰性仿真模型。然而，该模型在获取IB网表时采用线性假设[14]，根据注入功率输出通过/失败准则，这与模拟与模数混合器件在遭受电磁干扰后产生的非线性效应相矛盾，将导致模型与实际测量结果存在差异，这限制了ICIM-CI在精确预测和细致分析集成电路在复杂电磁干扰环境中表现的能力。此外，由于ICIM-CI建模方法是基于二值判定的敏感判据，模型通常只关注一个单一的输出指标(如输出最大电压值)，这对于单纯判断芯片的抗扰性似乎足够，但其无法进行参数化级联仿真。对于电路设计人员而言，他们更多关注的是芯片在放入电路中后电路的整体抗扰性指标，而非单一芯片的抗扰性指标，这直接阻碍了电路设计人员将芯片整合到整个电路板的级联仿真中，从而无法全面评估电路的电磁抗扰特性。尽管ICIM-CI提供了频域分析的框架，但其在效率和非线性特性处理方面的局限性[15]，以及无法支持电路板级的整体量化仿真，都是该模型需要克服的重要挑战。

本文提出基于多谐波失真理论(polyharmonic distortion, PHD)下的IB模块的改进建模方法，旨在解决现有ICIM-CI对于模拟与模数混合器件建模精度不足、无法开展量化仿真等方面的局限性。模型同时包含准确的时域和频域信息，能够在时域和频域中快速和准确的表征芯片在受到干扰后的非线性输出特性，并支持芯片在电路中的参数化传导敏感性仿真，并适应于电路设计人员进行量化级联抗扰性冗余设计。相较于传统ICIM-CI建模方法的线性假设，本文方法能够提供准确的非线性输出响应。除此之外，本文方法宽带建模时间相较于传统ICIM-CI方法缩短了98%，提高了建模效率。

1 非线性IB建模理论

传统ICIM-CI由PDN模块和IB模块组成，模型结构如图1所示，图中，P_{abs}为吸收功率，f为频率。

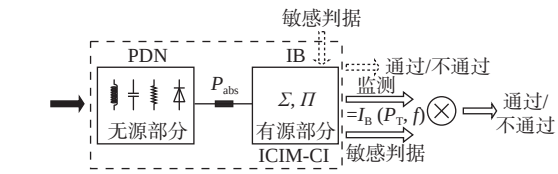


图1 ICIM-CI结构

Fig. 1 ICIM-CI structure

PDN表示集成电路的输入网络，包括干扰注入端口(disturbance input, DI)、干扰输出端口(disturbance output, DO)和内部端口。PDN主要描述线性行为。

IB是描述包括带内和带外集成电路频率响应行为的模块，IB为通过监测输出端口(observable output, OO)输出的信息描述集成电路对应用于DI的干扰信号的响应。IB包括频率、传输功率及OO的变化等参数。根据OO的测试方式，IB可以有2种建模方法：①可以使用通过/不通过敏感标准直接获得IB二进制数据(见图1虚线箭头)；②则使用非通过/失败测试提取故障检测模拟数据再利用敏感判据预测器件的通过/不通过结果(见图1实线箭头)。相较于第1种，第2种方法需要的测试次数往往更少。

在第2种方法时，在获得故障检测模拟数据时，常规需要建立线性拟合方程，即对每个频点记录干扰注入功率/电压、频率及敏感性能指标等参数。

拟合函数形式如下：

$$J_{Sen} = a(f_i)P_{abs} + b(f_i) \tag{1}$$

式中：J_{Sen}为敏感性能指标，用于量化IC在参数变化下的性能敏感程度；$a(f_i)=\sum_{i=0}^{n} x_i f_i^{p_i}$，$x_i$、$p_i$、$n$为拟合阶数；$b(f_i)=\sum_{i=0}^{n} y_i f_i^{q_i}$，$q_i$为待求的常数，$y_i$为拟合阶数；$P_{abs}$为吸收功率，$P_{abs}=P_T(1-|S_{11}|^2)$，$P_T$为传输功率；$P_T=P_F-P_R$，$P_F$为前向功率，$P_R$为反射功率。

对每个频点注入干扰功率和敏感性能指标进行拟合，基于最小二乘法对曲线和提出的函数进行拟合，找到最优的x_i、p_i、y_i、q_i，整个IB函数模型流程如图2所示，$P_{abs}(h,g)$表示第h个频率下的第g个功率值点。通过求解线性方程组(方程)建立敏感

性能指标和P_{abs}之间的关系, 芯片最终模型结构如图 3 所示。

但芯片在面临传导干扰的注入发生敏感时, 通常芯片会表现出非线性, 这也是为什么线性假设往往会导致最终预测结果存在偏差的原因。

因此, 引入一组多变量复变函数$\boldsymbol{F}_{pm}(\cdot)$, 该函数将所有相关输入频谱分量$\boldsymbol{A}_{qn}$与输出频谱分量$\boldsymbol{B}_{pm}$(其中, q 和 p 分别为输入端口号和输出端口号, q 和 p 的范围从 1 到信号端口数, m 和 n 分别为输出谐波指数和输入谐波指数, m 和 n 的范围从 0 到最高谐波指数)相关联。$\boldsymbol{B}_{pm}$ 数学表达式为

$$\boldsymbol{B}_{pm}=F_{pm}\left(\boldsymbol{D}_{\text{C}},\boldsymbol{A}_{11},\boldsymbol{A}_{12},\cdots,\boldsymbol{A}_{21},\boldsymbol{A}_{22},\cdots,\boldsymbol{A}_{qn}\right) \tag{2}$$

式中: $\boldsymbol{A}_{11}$为输入的干扰基频, $\boldsymbol{A}_{12},\cdots,\boldsymbol{A}_{21},\boldsymbol{A}_{22},\cdots\boldsymbol{A}_{qn}$依次表示端口 1 的 2 次谐波、端口 2 的 1 次频率分量、端口 2 的 2 次谐波, 以及通用形式下第 q 个端口的第 n 次频率分量。

芯片在受到干扰的情况下, 是由一个大信号干扰输入分量$\boldsymbol{A}_{11}$、正常驱动信号输入分量及其他输入分量(谐波频率分量)组成, 如图 4 所示[16-18]。此时, 相对较小的输入分量满足叠加原理。

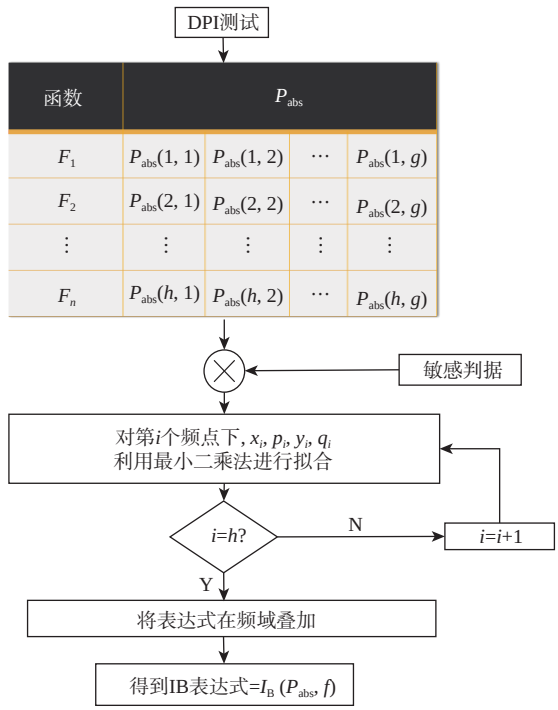


图 2　IB 原始方法函数建立流程

Fig. 2　Flowchart of function establishment for the traditional IB method

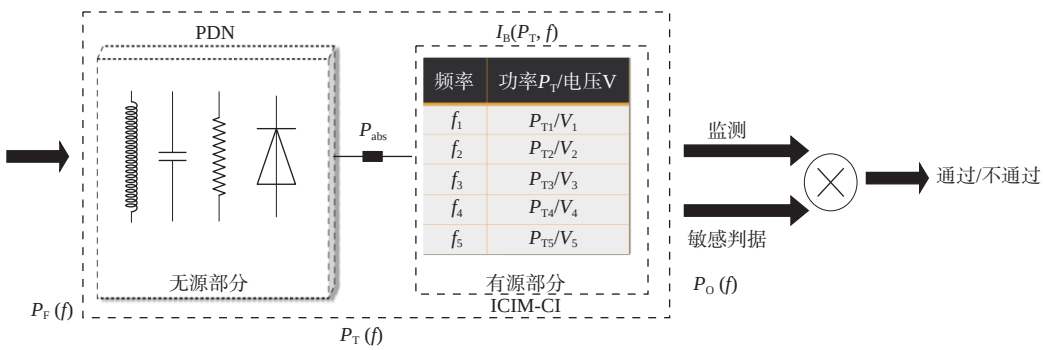


图 3　ICIM-CI 结构(方法 2)

Fig. 3　ICIM-CI structure (methodology 2)

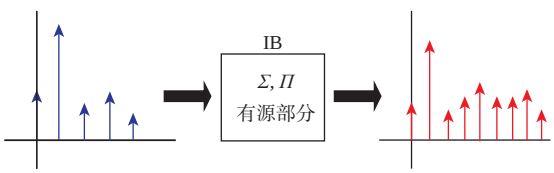


图 4　干扰信号和驱动信号在模拟与模数混合 IC 的非线性映射[16-18]

Fig. 4　Nonlinear mapping of interference and driving signals in analog-digital hybrid IC[16-18]

因此, 可以将输入信号到输出信号的映射定义为多变量非线性映射由大信号非线性映射和描述小信号入射引起的同频与跨频扰动的线性非解析映射组成, 如图 5 所示, 在 k 端口的 j 次谐波时, 表达式为

$$\boldsymbol{B}_k(\boldsymbol{D}_{\text{C}},\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\cdots)=\boldsymbol{X}_k^{\text{F}}(\boldsymbol{D}_{\text{C}},\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,\cdots)+\sum[\boldsymbol{X}_{kj}^{\text{S}}(\boldsymbol{D}_{\text{C}},\boldsymbol{A}_1)\boldsymbol{A}_j+\boldsymbol{X}_{kj}^{\text{T}}(\boldsymbol{D}_{\text{C}},\boldsymbol{A}_1)\boldsymbol{A}_j^*] \tag{3}$$

式中: $\boldsymbol{B}_k$为 k 端口的总输出散射波; $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{A}_3$, $\cdots$ 表示入射的散射波; $\boldsymbol{D}_{\text{C}}$ 表示直流偏置; $\boldsymbol{A}_j^*$为 $\boldsymbol{A}_j$ 的共轭。

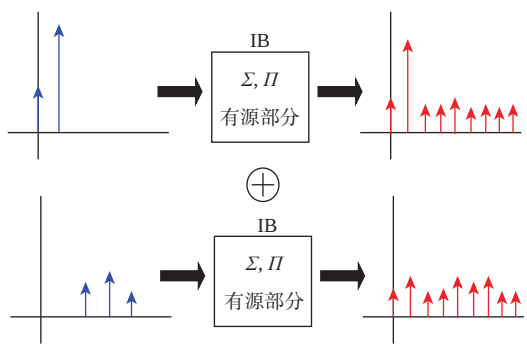


图 5　分为简单的非线性映射和线性非解析映射

Fig. 5　Divided into simple nonlinear mapping and linear non-analytic mapping

为方便计算，引入一个向量$\boldsymbol{P}$，定义为：$\boldsymbol{P}=\mathrm{e}^{+\mathrm{j}\phi(\boldsymbol{A}_{11})}$。

可将干扰基频$\boldsymbol{A}_{11}$进行正实数化，因为每个谐波都有2种贡献，此时，式(3)可以变为

$$\boldsymbol{B}_{pm}=F_{pm}(|\boldsymbol{A}_{11}|,\boldsymbol{A}_{12}\boldsymbol{P}^{-2},\boldsymbol{A}_{13}\boldsymbol{P}^{-3},\cdots,\boldsymbol{A}_{21}\boldsymbol{P}^{-1},\boldsymbol{A}_{22}\boldsymbol{P}^{-2},\cdots)\boldsymbol{P}^{+m} \tag{4}$$

进而式(4)变为

$$\boldsymbol{B}_{pm}=\boldsymbol{X}^{\mathrm{F}}_{pm}(|\boldsymbol{A}_{11}|)\boldsymbol{P}^{m}+\sum_{(q,n)\neq(1,1)}\boldsymbol{X}^{\mathrm{S}}_{pm,qn}(|\boldsymbol{A}_{11}|)\boldsymbol{P}^{m-n}\boldsymbol{A}_{qn}+\sum_{(q,n)\neq(1,1)}\boldsymbol{X}^{\mathrm{T}}_{pm,qn}(|\boldsymbol{A}_{11}|)\boldsymbol{P}^{m+n}\boldsymbol{A}^{*}_{qn} \tag{5}$$

式中：$\boldsymbol{A}^{*}_{qn}$为$\boldsymbol{A}_{qn}$的共轭项；系数$\boldsymbol{X}^{\mathrm{F}}_{pm}$、$\boldsymbol{X}^{\mathrm{S}}_{pm,qn}$、$\boldsymbol{X}^{\mathrm{T}}_{pm,qn}$为待求矩阵参数，$\boldsymbol{X}^{\mathrm{F}}_{pm}$描述了大信号工作点的影响。

$\boldsymbol{X}^{\mathrm{S}}_{pm,qn}$、$\boldsymbol{X}^{\mathrm{T}}_{pm,qn}$描述由于小信号入射在被测器件的端口上引起的同频与跨频扰动，其决定了被测器件对于端口上次谐波失配的敏感程度。

可以看出，相较传统ICIM-CI提供的方法2(见式(1))，这种根据单一输出响应监测建立线性化敏感预测模型而言，使用式(5)建立非线性IB(nonlinear IB, NIB)模型时，模型除包含大信号干扰输入分量和其他输入分量(谐波频率分量)能够准确模拟受扰状态下的输出响应外，还包含了正常驱动信号输入分量，因此，模型还可以准确的模拟芯片在正常工作下的输出响应，这对于量化端口输出响应、级联电路仿真提供了新思路和新方法。

2 ICIM-CI-NIB模型的仿真应用

本节介绍了使用芯片仿真电路来提取NIB的方法，并验证了所搭建的芯片在不同干扰功率下敏感性模型的准确性。使用基于白盒理论搭建的一款运算放大器等效电路模型。放大器为正向放大电路，设置正常工作时参数为：正相输入端V^{+}_{in}：f = 10 kHz，V = 100 mV的正弦波信号，直流偏置电压设置V^{+} = +15 V，V^{-} = −10 V。当芯片处于正常工作时，电路输出应如图6所示。依据IEC 62132-4[3]中DPI测试方法，对放大器输入端口和电源端口施加不同频率和功率的干扰信号，并设置敏感判据后，观察监测端口的输出响应敏感性特征主要可以分为以下4种现象，可根据IEC 62132-1[19]中规定的IC性能等级将故障描述为4种情况。具体描述如表1所示，其输出波形示意图如图6～图9所示。

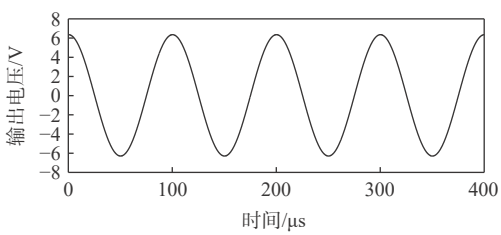


图6 芯片正常工作时(无干扰注入时)

Fig. 6 During chip normal operation (without interference)

表1 干扰后的运算放大器的输出类型和描述

Table 1 Output types and descriptions of operational amplifier after interference

案例	是否敏感	IC性能等级[19]	描述[19]
无干扰注入	否	Class A_{IC}	输出正常
干扰注入功率为−25 dBm	是	Class A_{IC}	输出波形发生畸变，但仍在允差要求内
干扰注入功率为0 dBm	是	Class C_{IC}	输出波形发生畸变并抖动，超出允差要求
干扰注入功率为10 dBm	是	Class C_{IC}	输出信号发生严重畸变，超出允差要求

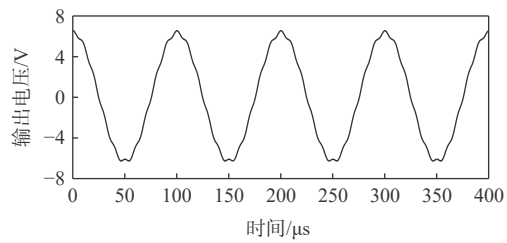


图7 芯片输出受到轻微扰动时(干扰注入−25 dBm时)

Fig. 7 Slight jitter in output waveform when subjected to −25 dBm interference, not exceeding tolerance requirements

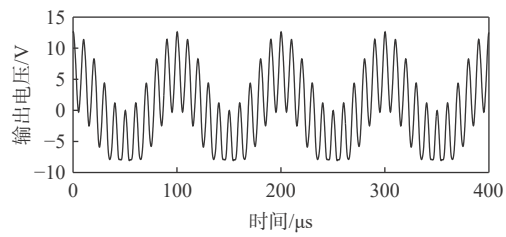


图8 芯片输出受到较大扰动时(干扰注入0 dBm时)

Fig. 8 Jittery output waveforms exceeding tolerance requirements when subjected to 0 dBm interference

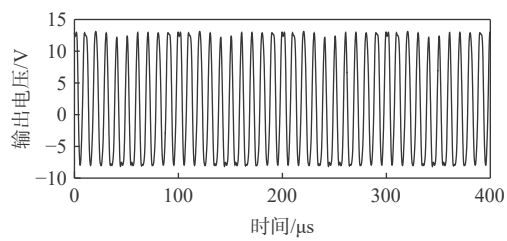


图9 芯片输出严重受扰时(干扰注入10 dBm时)

Fig. 9 Severe jitter in output waveform exceeding tolerance requirement when subjected to 10 dBm interference

通过使用仿真方式提取NIB模块并搭建ICIM-CI-NIB仿真模型，并验证模型的准确性及模型是否能够准确反映芯片在不同受扰情况下的输出响应。将本文ICIM-CI-NIB模型与传统ICIM-CI模型在频域抗扰性预测曲线上做对比。

2.1 模型提取

搭建模型并给芯片提供正常信号激励，使芯片

处于正常工作, 此时, 将干扰注入信号分别从放大器的负电源端口 V^- 和正相输入端 V_{in}^+ 注入, 如图 10 和图 11 所示, 放大器按照正向放大功能搭建电路。R1 和 R2 为实现正向放大的匹配电路, L1 和 C1 实现射频与直流的隔离。干扰信号的频段设定为 100～3 000 kHz, 干扰信号功率设定为−40～20 dBm 范围。按照第 2 节提出的方法求解式(5)中各个参数, 并完成放大器抗扰性模型。

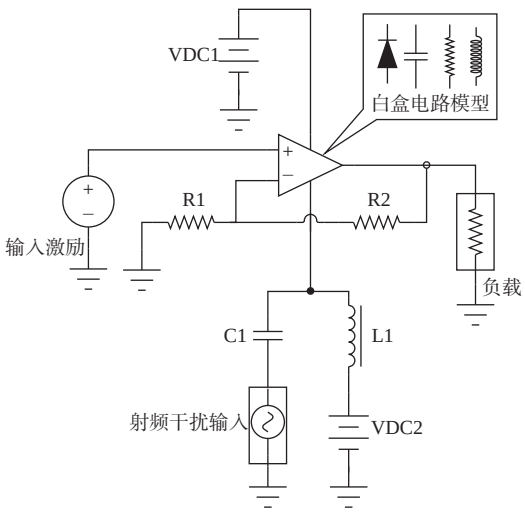


图 10　对电源端口 V^- 提取注入抗干扰模型

Fig. 10　Extraction of injection immunity models for power terminal V^-

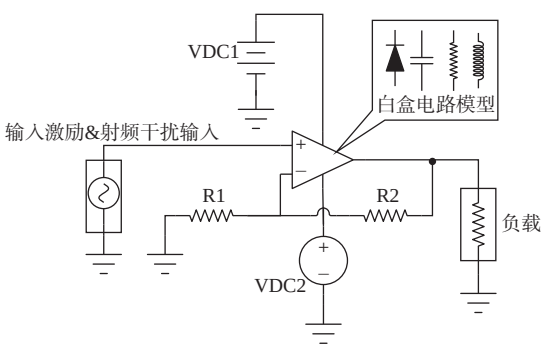


图 11　对输入端 V_{in}^+ 提取注入抗干扰模型

Fig. 11　Extraction of injected immunity model for input terminal V_{in}^+

2.2　模型验证

对本文抗扰性模型的准确性进行检验, 验证模型能否输出准确的时域响应波形, 以对输入端干扰注入频率为 100 kHz 为例, 验证模型的准确性。可以看出, 模型能够精确模拟器件在正常工作(见图 12)、输出波形发生畸变(见图 13)、输出波形发生畸变并抖动(见图 14)及输出信号发生严重畸变(见图 15)等各种受扰/非受扰状态下的响应波形, 表明该模型能够准确预测受扰行为并提供量化波形。

因为本文 ICIM-CI-NIB 模型可以直接输出芯片在受扰后的响应波形, 因此, 只需要在一次测试后,

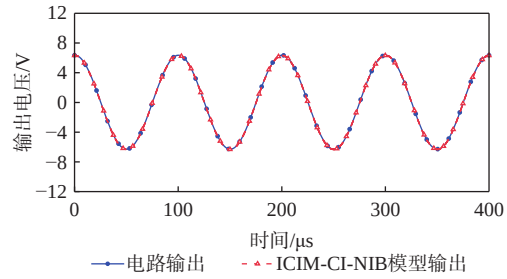


图 12　芯片正常工作时仿真模型输出与实际输出结果对比

Fig. 12　Comparison of simulation model output and actual output results during normal operation of chip

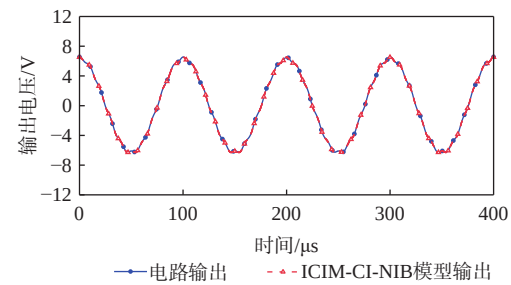


图 13　芯片输出受到轻微扰动时(干扰注入−25 dBm 时)仿真模型输出与实际输出结果对比

Fig. 13　Comparison of simulation model output and actual output when slight jitter occurs at the chip output (interference injection −25 dBm)

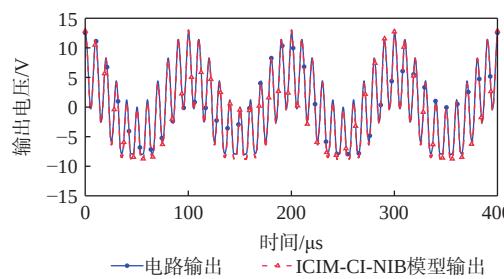


图 14　芯片输出受到较大扰动时(干扰注入 0 dBm 时)仿真模型输出与实际输出结果对比

Fig. 14　Comparison of simulation model output and actual output results when severe jitter occurs at chip output (when interference is injected at 0 dBm)

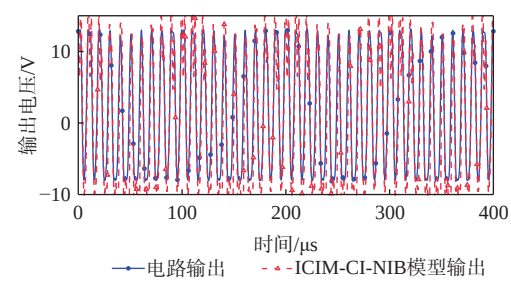


图 15　芯片输出严重受扰时(干扰注入 10 dBm 时)仿真模型输出与实际输出结果对比

Fig. 15　Comparison of simulation model output and actual output results when chip output is severely disturbed (interference injection at 10 dBm)

便可建立不同敏感判据的宽带抗扰预测曲线，相对于传统 ICIM-CI 需要先确定敏感判据，在建立抗扰预测曲线而言，ICIM-CI-NIB 模型不会由于敏感判据的变化而需要重新建模，这将大大节约整体建模时间。

为方便对比，设置敏感判据为输出电压峰峰值允许变化 $\Delta V_{\text{outp-p}}\leqslant 5\%$ 的条件，图 16 为当干扰从电源端注入时，使用本文方法和传统 ICIM-CI 建模方法结果的对比。可以看出，在此敏感判据下，由于系统整体仍有较高的线性度，因此，2 种建模方法对敏感性的预测均较好。

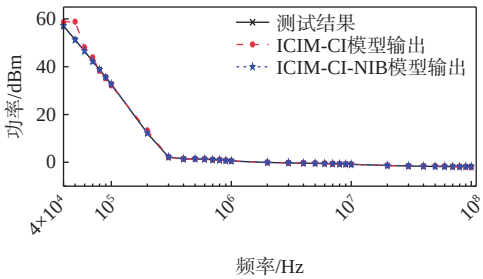


图 16 2 种建模方法的敏感性预测与实际输出结果的比较

Fig. 16 Comparison of sensitivity predictions of two modeling methods with actual output results

在对 V^{+}注入干扰时，设置 $\Delta V_{\text{outp-p}}\leqslant 10\%$ 时，如图 17 所示，当系统由于干扰出现非线性响应后，ICIM-CI 的预测精度变差，但 ICIM-CI-NIB 预测精度仍比较好。

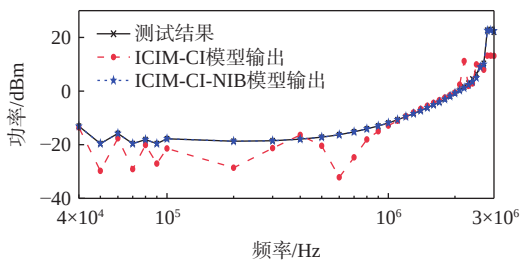


图 17 2 种建模方法的敏感性预测与实际输出结果的比较（V^{+}端口，敏感判据为 $\Delta V_{\text{outp-p}}\leqslant 10\%$）

Fig. 17 Comparison of the sensitivity prediction of two modeling methods with actual output results (at V^{+}port, immunity criterion at $\Delta V_{\text{outp-p}}\leqslant 10\%$)

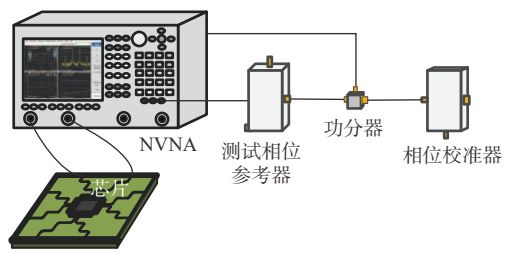
由图 17 可知，ICIM-CI-NIB 模型可以提供在时域和频域都非常准确的输出响应，相较于 ICIM-CI 模型而言，即使当敏感判据发生改变时，ICIM-CI-NIB 模型仍可以准确地预测被测器件的敏感现象，而不需要重新测量和建模，这将大大节省建模和测试的时间。

3 ICIM-CI-NIB 模型的测试应用

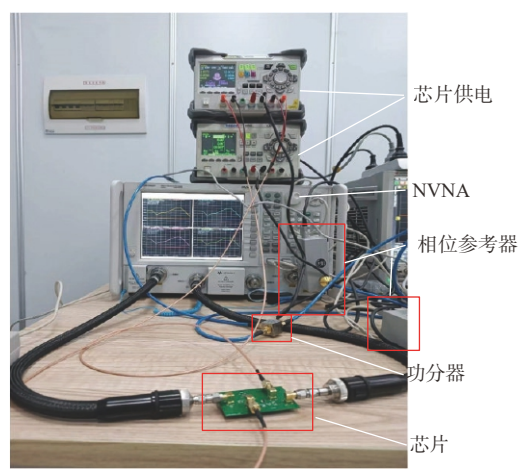
本节通过测试获得器件的传导敏感性模型并验证模型的准确性。选择 LMV651 运算放大器作为测试及建模对象，放大器设置为电压跟随器配置[20]，这种配置下，输入差分对受到的干扰可能导致放大器输出发生频率偏移或抖动[21]，致使放大器无法正常工作。此外，在所有可能的干扰信号中，叠加在运放输入引脚上的信号最难预防[22]。运算放大器由±2.5 V 的电源电压供电，V^{+}引脚设置为 0 V。干扰信号通过该引脚上的偏置三通注入。

3.1 NIB 模型提取方法

测量 NIB 的基本原理是使用方向性耦合器将入射波和反射波分离。利用非线性矢量网络分析仪（nonlinear vector network analyzer, NVNA）、信号源（可选）、直流源、外部相位参考发生器及适当的仪器控制和处理软件可以提取。直流源提供芯片基本工作环境。非线性矢量网络分析仪和信号源提供器件所面临的射频干扰环境，设置配置[23]如图 18 所示。



(a) 测试连接示意图



(b) 测试连接现场图

图 18 模型提取测试配置

Fig. 18 Model extraction test setup

使用上述测试配置，提取该运算放大器在 10～100 MHz 频率范围内。施加干扰频率在−10～10 dBm 范围内的 NIB 参数，并建立芯片的传导敏感性仿真模型。

3.2 NIB 模型测试验证

将从 DPI 测量获得的结果与所获得的模型进行比较，验证 NIB 模型的准确性，DPI 测试配置如图 19 所示。

DPI 测量是通过偏置三通向被测器件（在 V^{+}引脚或输入上）注入干扰信号来进行的，运算放大器用作跟随电路，并由±3 V 的电源电压供电。V^{+}引脚设置为 0 V。干扰信号通过该引脚上的偏置三通注入。

将测试结果和仿真模型进行对比，当注入

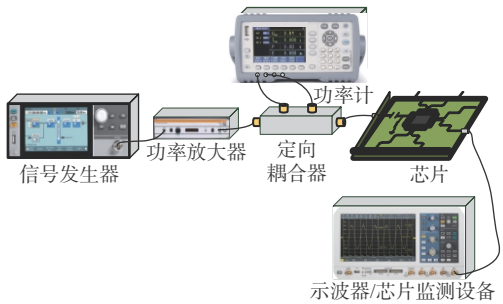


图 19　DPI 测试配置

Fig. 19　DPI test setup

f = 10 MHz、f = 40 MHz、f = 50 MHz、f = 100 MHz 等不同频率、不同干扰等级的干扰时, 时域波形的归一化均方根误差(normalized mean square error, MNSE)的统计情况如表 2 所示, 时域波形对比如图 20～图 27 所示。

表 2　不同干扰频率和功率下模型仿真结果和测试数据的 NMSE

Table 2　NMSE of model simulation results and test data at different interference frequencies and powers

干扰注入频率/MHz	干扰注入功率P_F/dBm	NMSE/dB
10	0	−27.487 6
40	−5	−36.849 0
50	−5	−28.672 5
50	0	−34.242 2
50	5	−30.998 5
50	10	−19.889 4
60	5	−28.981 4
100	10	−25.829 1

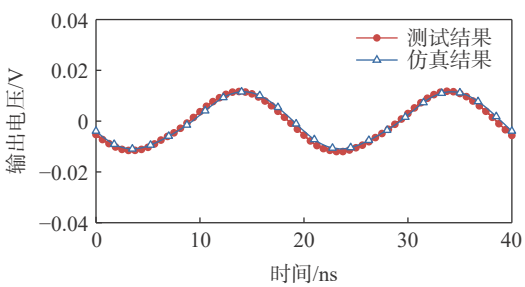


图 20　f=50 MHz, P_F=−5 dBm 时, 测试结果和仿真结果的对比

Fig. 20　Comparison of test and simulation results at f=50MHz, P_F=−5 dBm

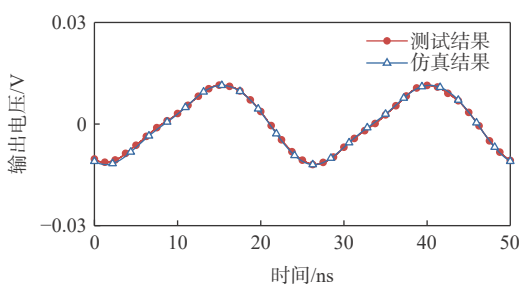


图 21　f=40 MHz, P_F=−5 dBm 时, 测试结果和仿真结果的对比

Fig. 21　Comparison of test and simulation results at f=40 MHz, P_F=−5 dBm

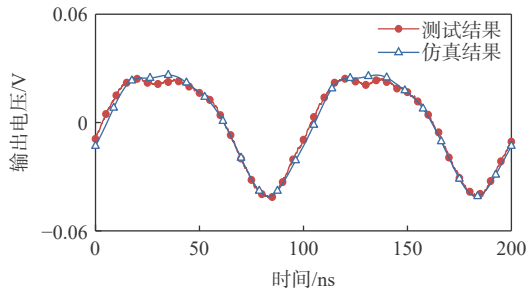


图 22　f=10 MHz, P_F=0 dBm 时, 测试结果和仿真结果的对比

Fig. 22　Comparison of test and simulation results at f=10 MHz, P_F=0 dBm

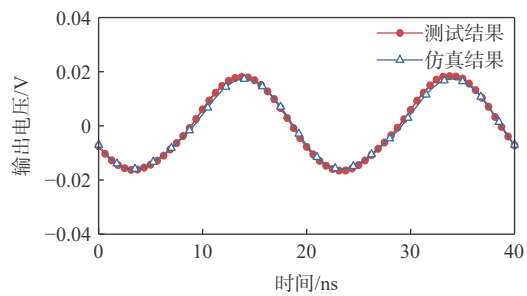


图 23　f=50 MHz, P_F=0 dBm 时, 测试结果和仿真结果的对比

Fig. 23　Comparison of test and simulation results at f=50 MHz, P_F=0 dBm

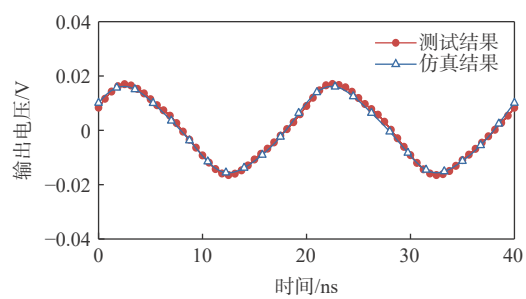


图 24　f=50 MHz, P_F=5 dBm 时, 测试结果和仿真结果的对比

Fig. 24　Comparison of test and simulation results at f=50 MHz, P_F=5 dBm

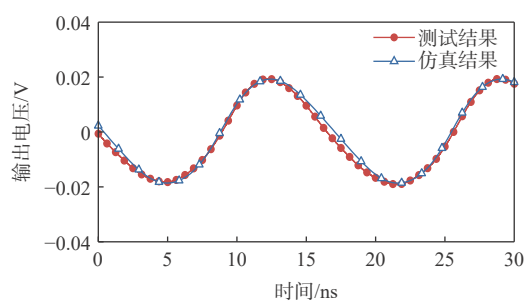


图 25　f=60 MHz, P_F=5 dBm 时, 测试结果和仿真结果的对比

Fig. 25　Comparison of test and simulation results at f=60 MHz, P_F=5 dBm

由表 2 可知, 在不同干扰注入频率及注入功率下, ICIM-CI-NIB 能够保持较高的准确性, 证明了其在模拟芯片正常运行、轻微和严重抖动条件

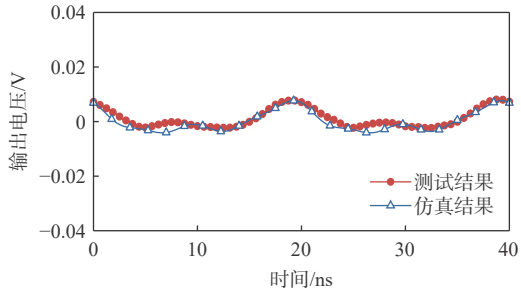


图 26 f=50 MHz, P_F=10 dBm 时测试结果和仿真结果的对比

Fig. 26 Comparison of test and simulation results at f=50 MHz, P_F=10 dBm

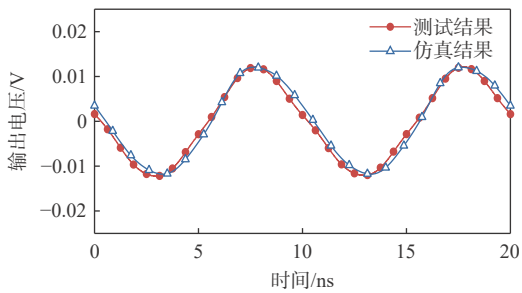


图 27 f=100 MHz, P_F=10 dBm 时, 测试结果和仿真结果的对比

Fig. 27 Comparison of test and simulation results at f=100 MHz, P_F=10 dBm

下的量化仿真波形方面的高精度性。当然, 可以看出, 随着系统抖动的加剧, 尤其是在严重抖动或失真的情况下, ICIM-CI-NIB 的归一化均方根误差有所下降。这种性能下降主要归因于系统在严重抖动情况下产生的波形是随机且不稳定的, 这对于任何模型来说都是一个挑战, 因为预测随机性较高的动态行为需要更高级的模拟技术和算法。尽管如此, ICIM-CI-NIB 仍然能够提供对系统敏感性的有效仿真和分析。这表明该模型不仅能够准确地模拟正常和轻度抖动状态下的系统行为, 而且即使在遇到高度不稳定和不可预测的抖动时, 也能够捕捉到足够的信息, 以便进行有意义的分析。

当引入敏感判据 $\Delta V_{\text{outp-p}} \leqslant 30$ mV 的条件时, 图 28 为使用本文方法和传统 ICIM-CI 预测结果对比。表 3 综合分析了 2 种方法在时域和频域下的建模精度、建模时间及是否可以开展级联量化仿真等方面的差异, 可以看出, 本文方法在建模精度、建模方法及可量化输出后开展级联仿真的方面具有明显优势。本文方法能够在时域输出可量化的仿真结果, 可支持电路板的级联量化仿真需求, 同时, 相比传统 ICIM-CI 建模方法, 本文方法在频域的建模缩短了约 98%, NMSE 降低了 18.5 dB。

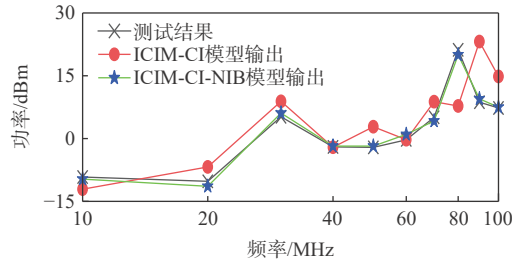


图 28 2 种建模方法的敏感性预测与 DPI 测量结果的比较对比

Fig. 28 Comparative comparison of immunity prediction and DPI measurements for the two modeling approaches

表 3 2 种模型归一化均方根误差和建模时间等信息对比

Table 3 Comparison of information such as normalized root mean square error and modeling time for two models

信号类型	建模方法	NMSE	建模时间/s	是否可以开展级联量化仿真
时域	ICIM-CI-NIB	−30.920 0	21.0	是
	ICIM-CI			否
频域	ICIM-CI-NIB	−31.335 2	34.8	是
	ICIM-CI	−12.798 2	498.0	否

4 结 论

1) 本文提出 ICIM-CI-NIB 为集成电路的传导敏感度建模提供了一种精准的可级联的建模方法。

2) 通过数值模拟和实验验证了 ICIM-CI-NIB 方法的准确性。通过引入 PHD, 可以有效地对 IB 模块进行非线性量化处理。实验结果表明, ICIM-CI-NIB 方法足以描述和预测集成电路传导敏感特性。与 ICIM-CI 方法相比, ICIM-CI-NIB 能够有效缩短建模时间, 且频域下的 NMSE 降低了 18.5 dB, 同时, ICIM-CI-NIB 能够在时域输出可量化的仿真结果, 可支持电路板的级联量化仿真需求。

尽管本文模型在运算放大器上显示出有效性, 但对于其他类型的芯片, 未来需要进一步评估模型的通用性和可移植性, 为其在实际应用中的广泛应用奠定基础。此外, 模型按照 IEC62132-4 中 DPI 测试方法对芯片在正弦波干扰下的有效性已得到验证, 但在实际应用中, 集成电路可能面临多种类型的电磁干扰, 如随机噪声、调制信号干扰等。每种干扰波形都有其独特的频谱特性和时域表现, 对集成电路的传导敏感度产生不同的影响。因此, 为全面评估模型的性能, 未来需要在更多类型的干扰波形下进行验证。

参考文献 (References)

[1] MASETTI G, GRAFFI S, GOLZIO D, et al. Failures induced on

analog integrated circuits by conveyed electromagnetic interferences: a review[J]. Microelectronics Reliability, 1996, 36(7-8): 955-972.

[2] LOECKX J, GIELEN G. Assessment of the DPI standard for immunity simulation of integrated circuits[C]//Proceedings of the IEEE International Symposium on Electromagnetic Compatibility. Piscataway: IEEE Press, 2007: 1-5.

[3] International Electrotechnical Commission. Integrated circuits-measurement of electromagnetic immunity 150 kHz to 1 GHz-Part 4: direct RF power injection method: IEC 62132-4: 2006[S]. Geneva : International Electrotechnical Commission, 2006.

[4] GAZDA C, GINSTE D V, ROGIER H, et al. An immunity modeling technique to predict the influence of continuous wave and amplitude modulated noise on nonlinear analog circuits[C]//Proceedings of the International Symposium on Electromagnetic Compatibility. Piscataway: IEEE Press, 2013.

[5] CHAHINE I, KADI M, GABORIAUD E, et al. Characterization and modeling of the susceptibility of integrated circuits to conducted electromagnetic disturbances up to 1 GHz[J].IEEE Transactions on Electromagnetic Compatibility, 2008, 50(2): 285-293.

[6] CEPERIC V, BARIC A. Modelling of electromagnetic immunity of integrated circuits by artificial neural networks[C]//Proceedings of the 20th International Zurich Symposium on Electromagnetic Compatibility. Piscataway: IEEE Press, 2009: 373-376.

[7] SCHROTER M, PEHLKE D R, LEE T Y. Compact modeling of high-frequency distortion in silicon integrated bipolar transistors[J]. IEEE Transactions on Electron Devices, 2000, 47(7): 1529-1535.

[8] WANG H D, JNGFU B, WU Z D. Multislice behavioral modeling based on envelope domain for power amplifiers[J]. Journal of Systems Engineering and Electronics, 2009, 20(2): 274-277.

[9] NGOYA E, QUINDROIT C, NEBUS J M. On the continuous-time model for nonlinear-memory modeling of RF power amplifiers[J]. IEEE Transactions on Microwave Theory and Techniques, 2009, 57(12): 3278-3292.

[10] LAFON F, RAMDANI M, PERDRIAU R, et al. An industry-compliant immunity modeling technique for integrated circuits[C]//Proceedings of the International Symposium on Electromagnetic Compatibility. Piscataway: IEEE Press, 2009.

[11] International Electrotechnical Commission. Integrated circuit EMC IC modeling-Part 4: ICIM-CI, integrated circuit immunity model, conducted immunity: IEC 62433-4: 2016[S]. Geneva: International Electrotechnical Commission, 2016.

[12] KHAN Q M, KOOHESTANI M, LEVANT J L, et al. Validation of IC conducted emission and immunity models including aging and thermal stress[J]. IEEE Transactions on Electromagnetic Compatibility, 2023, 65(3): 780-793.

[13] BOYER A, SICARD E. A case study to apprehend RF susceptibility of operational amplifiers[C]//Proceedings of the 12th International Workshop on the Electromagnetic Compatibility of Integrated Circuits. Piscataway: IEEE Press, 2019: 13-15.

[14] LAFON F, DE DARAN F, RAMDANI M, et al. Immunity modeling of integrated circuits: an industrial case[J]. IEICE Transactions on Communications, 2010, E93. B(7): 1723-1730.

[15] WANG Z Y, ZHOU C L, LIU T, et al. Nonlinear behavior immunity modeling of an LDO voltage regulator under conducted EMI[J].IEEE Transactions on Electromagnetic Compatibility, 2016, 58(4): 1016-1024.

[16] VERSPECHT J, ROOT D E, WOOD J, et al. Broad-band, multiharmonic frequency domain behavioral models from automated large-signal vectorial network measurements[C]//Proceedings of the IEEE MTT-S International Microwave Symposium Digest. Piscataway: IEEE Press, 2005: 1975-1978.

[17] ROOT D E, VERSPECHT J, SHARRIT D, et al. Broad-band polyharmonic distortion (PHD) behavioral models from fast automated simulations and large-signal vectorial network measurements[J]. IEEE Transactions on Microwave Theory and Techniques, 2005, 53(11): 3656-3664.

[18] VERSPECHT, ROOT D E. Polyharmonic distortion modeling[J]. IEEE Microwave Magazine, 2006, 7(3): 44-57.

[19] International Electrotechnical Commission. Integrated circuits—Measurement of electromagnetic immunity 150 kHz to 1 GHz-Part 1: General conditions and definitions: IEC 62132-1: 2015[S]. Geneva: International Electrotechnical Commission, 2015.

[20] RICHELLI A. EMI susceptibility issue in analog front-end for sensor applications[J]. Journal of Sensors, 2016, 2016: 1082454.

[21] POULTON A S. Effect of conducted EMI on the DC performance of operational amplifiers[J]. Electronics Letters, 1994, 30(4): 282-284.

[22] SETTI G, SPECIALE N. Design of a low EMI susceptibility CMOS transimpedance operational amplifier[J].Microelectronics Reliability, 1998, 38(6-8): 1143-1148.

[23] Keysight Technologies, Inc. NVNA help-general hardware configuration: E5080A-90001[R]. Santa Rosa：Keysight Technologies, 2023: 1-8.

Simulation modeling methodology for broadband conducted immunity quantization of analog and analog-digital hybrid chips

CHEN Xi[1]，XIE Shuguo[1,*]，WEI Mengyuan[1]，LI Yuanyuan[2]

(1. School of Electronic Information Engineering，Beihang University，Beijing 100191，China；
2. School of Cyber Scicence and Technology，Beihang University，Beijing 100191，China)

Abstract：There are issues with the current analog and analog-digital hybrid chip in radio frequency (RF) conduction sensitivity modeling method models of integrated circuits for RF integrated circuit immunity model-conducted immunity modeling (ICIM-CI). These issues include the model's lack of nonlinear characteristics, significant discrepancies between the prediction and the actual circuits' performance under particular interference conditions, and the inability to perform cascade quantization sensitivity simulation. The study proposes an improved model, integrated circuit immunity model-conducted immunity modeling-nonlinear immunity behavior (ICIM-CI-NIB), an improved modeling method for ICIM-CI with analog and analog-digital hybrid chip conduction sensitivity based on the theory of multi-harmonic distortion, which significantly improves the nonlinear characteristics of the conduction sensitivity model by introducing the multi-harmonic nonlinear parameter into the immunity behavior（IB）module, which improves the ICIM-CI model from the binary judgment result model to the cascaded quantized sensitivity simulation with nonlinear quantized output response of the cascade simulation model. By adopting the ICIM-CI-NIB method, it can quickly generate the frequency domain broadband model of the chip's conduction sensitivity, have a high-precision conduction sensitivity simulation model under the wide-band large-power interference injection, and support the cascade simulation. The observed results on a common analog and analog-digital hybrid chip demonstrate that this method reduces the modeling time by approximately 98% while improving the normalized mean square error (NMSE) by 18.5 dB when compared to the ICIM-CI method.

Keywords：analog and analog-digital hybrid chips；conducted sensitivity；direct power injection；electromagnetic interference；modeling；conducted immunity

Received：2024-04-07；**Accepted**：2024-05-17；**Published Online**：2024-06-07 13：36

URL：link.cnki.net/urlid/11.2625.V.20240607.1032.002

Foundation item：National Natural Science Foundation of China (62202030)

* **Corresponding author**. E-mail：xieshuguo@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0196

基于深度特征修正的目标跟踪算法

陈凯[1]，黄煜杰[1,*]，赵晓冬[1]，王鹏飞[2]，陈凯[1]，林岩泽[1]

(1. 南京航空航天大学 机电学院，南京 210016；2. 中国电子科技集团第二十八研究所, 南京 210007)

摘 要：在复杂的跟踪环境中，现有的跟踪器主要面临深度卷积特征冗余和目标跟踪过程正样本缺乏的问题。为去除深度卷积网络中的冗余特征，提出一种基于空间域和通道域融合的注意力机制模型，包括空间自注意力、通道注意力和空间注意力3个连续的子模块；以此为基础，构建一种适用于各种视觉算法的深度卷积网络；设计基于特征修正的提取目标深度卷积特征及特征增强策略，解决冗余特征负反馈及正样本缺乏等问题。实验结果表明：在主流残差网络（ResNet）目标分类任务中，加入特征修正模块可使Top-1和Top-5错误率明显降低，且不会造成额外的计算量或网络结构调整，实现轻量化插入；在多个相关跟踪算法中加入特征修正模块后实现了更加优秀的跟踪性能，并解决了判别器的过拟合问题。

关 键 词：目标跟踪；注意力机制；特征修正；轻量化插入；残差网络

中图分类号：TP391.4

文献标志码：A **文章编号**：1001-5965(2026)05-1523-13

随着智能交通、机器人技术的发展和需求，目标跟踪成为计算机视觉领域的基础研究问题之一。深度学习也为各类目标跟踪模型的性能提供了强大的推动力。在复杂环境中，跟踪目标的运动过程面临很多挑战：目标视觉特征和背景噪声不断变化、目标被不定时遮挡、跟踪过程伴随场景光照变化和目标形变[1-3]等，都会对视觉特征的应用产生影响[4]。需要从多方面研究目标跟踪相关算法以得到一个具有鲁棒性的高性能跟踪器。

视觉对象跟踪是仅在给定第1帧中的位置信息的情况下连续定位和跟踪视频中目标的任务。跟踪器所能依赖的仅仅是目标在第1帧中的目标特征。一类算法主要采用卷积神经网络(convolutional neural networks, CNN)的深层特征以端到端的方式来训练跟踪器[5-7]，通过在大规模数据集[8-9]上离线训练CNN模型来获得强大的识别能力；另一类是基于相关滤波器(correlation filter, CF)的算法，主要使用循环矩阵为相关滤波参数，在线学习生成密集采样。基于相关滤波的跟踪器[10-13]依靠具有大量参数和频繁在线更新的强大功能表示。

然而，大多数跟踪器只是将各种特征粗略地合并在一起，可以将其定义为多特征融合，其在网络中的每个层次都引入了很多冗余特征并掩盖了分层特征。这些冗余特征一方面增加了判别器的在线训练时间，另一方面加大了判别器的学习难度。如何从多层次的深度卷积特征提取真正有效的特征并对这些特征增强是提高跟踪器性能的一个关键技术难题。此外，文献[3,13]已经证实大量的可训练参数提高了严重过度拟合的风险，并且由于CNN的参数较大且特征稀疏，更深的残差网络(residual networks, ResNet)[14]等显示出了对视觉跟踪任务更有效的性能。如何有效利用丰富的深层CNN功能来构建高性能跟踪器仍然是一项具有挑战性的研究。

收稿日期：2024-04-07；**录用日期**：2024-05-09；**网络出版时间**：2024-06-11 08:40

网络出版地址：link.cnki.net/urlid/11.2625.V.20240607.0959.001

基金项目：国家自然科学基金(52202417)；中国博士后科学基金(2022TQ0155，2022M721605)；虚拟现实技术与系统全国重点实验室（北京航空航天大学）开放课题基金(VRLAB2023A02)

* **通信作者**. E-mail：yj_huang@nuaa.edu.cn

引用格式：陈凯，黄煜杰，赵晓冬，等. 基于深度特征修正的目标跟踪算法[J]. 北京航空航天大学学报，2026，52(5)：1523-1535.

CHEN K，HUANG Y J，ZHAO X D，et al. Object tracking algorithm based on deep feature modification[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1523-1535 (in Chinese).

1 相关工作

近年来，目标跟踪模型获得了广泛的发展。Li等[15]利用CNN端到端的训练方法将跟踪问题转化为分类问题。多域卷积神经网络(multi-domain convolutional neural networks, MDNet)[16]进一步结合离线多领域训练和在线更新分类器来识别特定目标。遵循端到端的想法，一些作品进一步使用孪生匹配结构来学习相似性度量，该度量将深度相关滤波器视为网络的一部分。孪生全卷积网络(siamese fully convolutional networks, SiamFC)[6]使用ILSVRC数据集[9]进行离线训练，并且不在线更新参数。判别相关滤波网络(discriminative correlation filter networks, DCF-Net)[7]提出一种端到端的网络体系结构，以学习卷积特征并同时执行相关跟踪过程。孪生区域建议网络(siamese region proposal networks, SiamRPN)[17]引入特征提取和区域建议子网，包括分类和回归分支。基于视觉几何组(visual geometry group, VGG)[18]骨干的孪生全卷积网络(VGG-based siamese network, SiamVGG)[19]在SiamFC的基础上用VGG替换了基本网络深度卷积神经网络(ImageNet classification with deep convolutional neural networks, AlexNet)[20]，以提高跟踪性能。这种类型的算法通常以第1帧的基础为模板或采用简单的移动平均策略来更新模板。

对于相关滤波器算法，由于傅里叶域中的高计算效率，在视觉跟踪中受到了广泛的关注。Bolme等[21]通过学习目标外观的最小平方误差平方和，提出一种相关滤波器跟踪器，使其可以高速运行。核循环结构跟踪器(circulant structure with kernels tracking, CSK)[11]使用圆形矩阵进行密集采样，以产生大量的计算负荷低的样本。核化相关滤波(kernelized correlation filter, KCF)[10]采用岭回归和多通道特征来解决相关滤波器参数。空间正则化判别相关滤波(spatially regularized discriminative correlation filters, SRDCF)[22]在滤波器参数上使用负高斯惩罚权重来克服边界效应。CSK将CNN的深度卷积特征引入SRDCF[23]，并取得了良好的效果。连续卷积算子跟踪器(learning continuous convolution operators for visual tracking, C-COT)[12]进一步将不同分辨率的特征图转换为连续的空间域，以实现更高的精度。随后的基于高效卷积算子跟踪器(efficient convolution operators for tracking, ECO)[13]在性能和效率方面改进了C-COT。ECO加权卷积响应相关滤波(correlation filters with weighted convolution responses, CFWCR)[24]规范化了从不同层提取的每个单独特征，以获得更可靠的结果。尽管上述方法在目标跟踪任务中取得了一定进展，但仍存在特征冗余、判别能力不足等问题。为此，本文提出一种基于深度特征修正的目标跟踪算法。

2 特征修正卷积神经网络

有效的目标特征能够大幅提高跟踪器的性能。深度网络随着层数的增加，特征中不可避免地存在很多冗余信息。同时，输入图像的背景也会对有效特征信息产生干扰[25-26]。本文抛弃已有的深度神经网络作为特征提取器，在ResNet的基础上利用通道域和空间域融合的注意力机制构建权重激励机制，选取对目标判别有用的特征，构建一个即插即用、适用于各种现有跟踪模型的深度目标分类网络。

2.1 特征修正注意力机制模型

软注意力关注区域或通道，是一种确定的、可微分的注意力机制。本文着重从注意力域的方向来构建相应的注意力机制，主要包括：空间域和通道域。空间域注意力将原始图片中的空间信息变换到另一个空间中利用网络的学习来保留；通道域注意力则将每个通道上的信号都增加一个权重，表示每个通道与关键信息的相关度。

基于2种域的结合，本文设计一种改进的注意力机制模块，即特征修正模块。如图1所示，每个

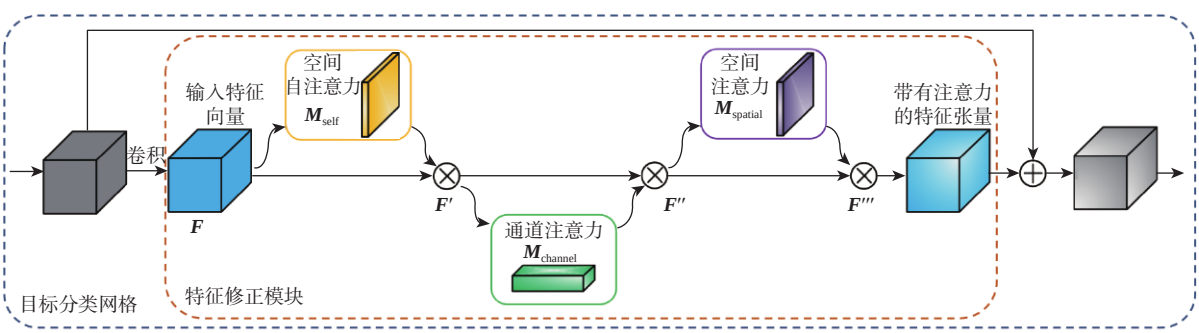


图1 基于通道域和空间域融合的注意力机制模块整体结构

Fig. 1 Overall structure of attention mechanism module based on fusion of channel and spatial domain

分支结构都可以有效地学习到在通道和空间域上对目标判别有意义的关键信息。在该模块中, 空间自注意力可以看作是特征张量的自激励响应, 其不使用最大池化和均值池化, 目的是初步筛选出特征张量中某一通道的异常较大值。而空间注意力的作用对象是经过通道注意力进行元素筛选之后的张量, 所以在空间位置上对特征进行细粒度筛选。因此, 采用初步特征筛选 (空间自注意力)-粗粒度特征筛选 (通道注意力)-细粒度特征筛选 (空间注意力) 的顺序来构建特征修正模块。

给定卷积网络的一个中间特征图, 通过特征修正模块深层卷积的每个卷积块可以进行自适应调整, 最终得到对目标分类判别有积极意义的修正卷积特征, 即, 强调有用的信息同时抑制作用小甚至起负作用的信息。给定一个中间特征图 $\boldsymbol{F} \in \mathbf{R}^{C\times H\times W}$ 作为输入, 修正特征模块依次推导二维自注意力权重 $\boldsymbol{M}_{\text{self}} \in \mathbf{R}^{1\times H\times W}$、一维通道注意力权重 $\boldsymbol{M}_{\text{channel}} \in \mathbf{R}^{C\times 1\times 1}$ 和二维空间注意力权重 $\boldsymbol{M}_{\text{spatial}} \in \mathbf{R}^{1\times H\times W}$。总体注意力卷积特征输出 $\boldsymbol{F}'''$ 的过程可以概括为

$$\begin{cases} \boldsymbol{F}' = \boldsymbol{M}_{\text{self}}(\boldsymbol{F}) \otimes \boldsymbol{F} \\ \boldsymbol{F}'' = \boldsymbol{M}_{\text{channel}}(\boldsymbol{F}') \otimes \boldsymbol{F}' \\ \boldsymbol{F}''' = \boldsymbol{M}_{\text{spatial}}(\boldsymbol{F}'') \otimes \boldsymbol{F}'' \end{cases} \tag{1}$$

式中: $\otimes$ 表示按元素相乘; $\boldsymbol{F}'$、$\boldsymbol{F}''$、$\boldsymbol{F}'''$ 为依次经过 3 个连续子模块的结果, $\boldsymbol{F}'''$ 为最终的修正特征。在乘法过程中, 通道注意力权重是沿空间维度传播的, 反之亦然。

2.1.1　空间自注意力模块

卷积网络的中间特征都是三维的张量, 其通道轴上可以理解为图像的“内容”, 空间轴上为图像的“位置”。通道注意力是对图像内容的初步粗粒度筛选, 空间注意力是对图像二维特征的细粒度筛选。如果卷积特征某一通道包含较强的噪声信息, 单纯依靠池化操作很容易让网络学习到噪声信息。因此, 在通道注意力模块之前构建空间注意力模块对卷积特征每个通道二维空间信息进行初步细粒度筛选。

空间自注意力模块结构如图 2 所示, 目标特征经过 1×1 卷积操作降维得到通道数为 1 的特征张量, 之后经过 Sigmoid 激励函数将特征值转换到 [0,1] 区间, 紧接着利用 SoftMax 函数获得空间自注意力权重 $\boldsymbol{M}_{\text{self}}$, 注意力权重和输入特征逐元素相乘获得自注意力修正特征, 该操作可以理解为输入特征的自激励响应输出。$\boldsymbol{M}_{\text{self}}$ 的计算过程为

$$\boldsymbol{M}_{\text{self}}(\boldsymbol{F}) = \text{SoftMax}\left(\int^{1@1\times 1}(\boldsymbol{F})\right) \tag{2}$$

式中: $\int^{1@1\times 1}$ 表示对 $\boldsymbol{F}$ 进行的二维卷积操作, @前、后分别表示卷积核个数为 1、卷积核的大小为 1×1。

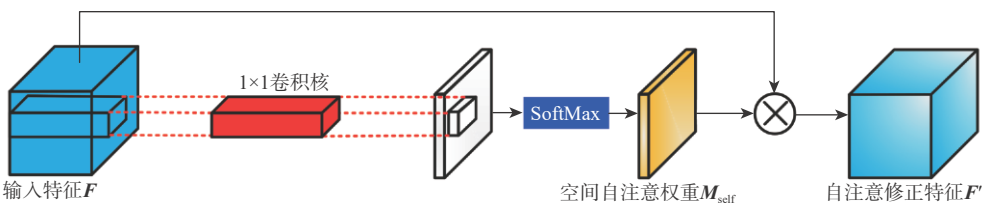


图 2　空间自注意力模块结构

Fig. 2　Structure of spatial self-attention module

2.1.2　通道注意力模块

通道注意力图是经过特征通道之间的关系来生成的, 定义特征图的每个通道为特征检测器[27], 即可强调对图像判别起正作用的内容同时忽略起负作用的内容。为获得通道域的注意力权重参数, 要将特征的空间域的尺寸进行压缩。文献 [28] 表明通过平均池化来计算空间统计量, 最大池化收集对象的独特特征, 可以大大提高网络的表示能力, 以推断出更精细的通道注意力。这也印证了特征修正模块设计的有效性。

如图 3 所示, 首先, 对输入的自注意修正特征 $\boldsymbol{F}'$ 采用最大池化和均值池化, 将特征在空间域上的信息进行聚合, 产生相应的空间特征描述上下文信息: 均值池化特征 $\boldsymbol{F}_{\text{avg}}^{\text{channel}}$ 和最大池化特征 $\boldsymbol{F}_{\text{max}}^{\text{channel}}$。

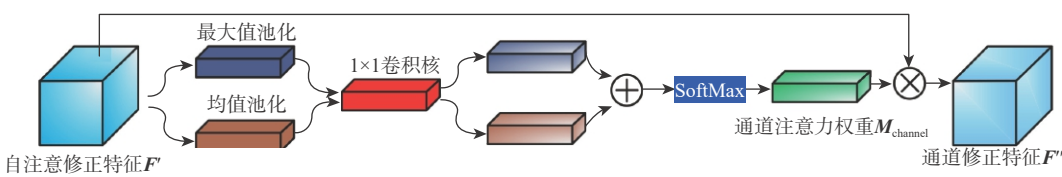


图 3　单个通道注意力模块整体流程图

Fig. 3　Overall flowchart of a single channel attention module

之后，利用共享卷积网络和SoftMax将$\boldsymbol{F}_{\text{avg}}^{\text{channel}}$和$\boldsymbol{F}_{\text{max}}^{\text{channel}}$生成通道注意力权重$\boldsymbol{M}_{\text{channel}} \in \mathbf{R}^{C\times1\times1}$。共享卷积网络由具有$C$个$1\times1$的卷积核构成。将该共享卷积网络应用于描述特征$\boldsymbol{F}_{\text{avg}}^{\text{channel}}$和$\boldsymbol{F}_{\text{max}}^{\text{channel}}$后，对描述特征使用逐元素求和合并输出特征向量。合并后的特征向量经过SoftMax之后获得最终的通道注意力权重$\boldsymbol{M}_{\text{channel}}$。通道注意力权重的计算式为

$$\boldsymbol{M}_{\text{channel}}(\boldsymbol{F}) = \text{SoftMax}\left(\int^{C@1\times1}(\text{AvgPool}(\boldsymbol{F}))+\int^{C@1\times1}(\text{MaxPool}(\boldsymbol{F}))\right) \tag{3}$$

式中：$\int^{C@1\times1}$表示二维卷积操作，@前、后分别表示卷积核个数为C、卷积核的大小为1×1。

2.1.3 空间注意力模块

如图4所示，空间注意力特征图主要通过每个通道特征元素之间的空间关系来生成，其关注的重点是“哪里”，其为信息，是对通道注意力的补充。首先，沿通道轴应用平均池化和最大池化操作，并将其在通道维度上连接起来，合并成在最大响应和均值响应两方面均有效的描述特征。文献[29]提出在突出显示信息区域方面沿通道轴合并操作是有效的。在级联特征描述特征上，应用卷积层以生成空间注意力图$\boldsymbol{M}_{\text{channel}} \in \mathbf{R}^{C\times1\times1}$，对该值编码即对信息进行强调或抑制。

空间注意力模块使用最大池化和均值池化操作来聚合特征图的通道信息，生成2个二维图：$\boldsymbol{F}_{\text{avg}}^{\text{spatial}} \in \mathbf{R}^{C\times1\times1}$和$\boldsymbol{F}_{\text{max}}^{\text{spatial}} \in \mathbf{R}^{C\times1\times1}$。其中，$\boldsymbol{F}_{\text{max}}^{\text{spatial}}$为通道中的平均合并特征，$\boldsymbol{F}_{\text{max}}^{\text{spatial}}$为最大合并特征。紧接着采用标准卷积层将其连接起来并进行卷积，从而生成相应的二维空间注意力权重：

$$\boldsymbol{M}_{\text{spatial}}(\boldsymbol{F}) = \text{SoftMax}\left(\int^{1@1\times1}([\text{AvgPool}(\boldsymbol{F});\text{MaxPool}(\boldsymbol{F})])\right) \tag{4}$$

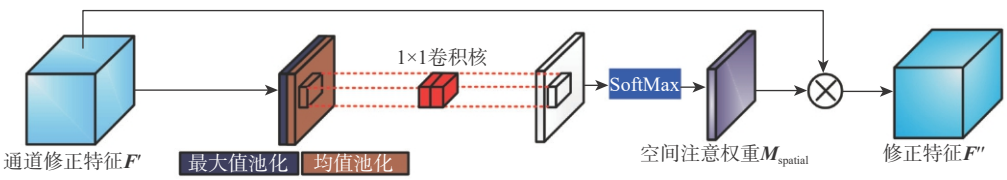


图4 单个空间注意力模块整体流程图

Fig. 4 Overall flowchart of a single spatial attention module

2.1.4 特征修正模块优化深度残差模块

由于特征修正模块不改变输入输出大小的同时具有很少的参数量，因此，本文提出的特征修正模块可以实现在现有的主流卷积网络中无缝插入。最新的研究表明，基于ResNet的卷积网络在目标分类和目标跟踪的特征提取过程中都取得了较好的成绩，因此，本文在ResNet的残差模块中加入特征修正模块构建带有注意力的目标分类网络，结构如图1所示。不仅可以提高目标分类网络的准确性，还能够根据修正模块中通道注意力和空间注意力的权重参数选择对目标判别起正相关作用的深度卷积特征。

2.2 基于特征修正网络提取深度卷积特征

在目标跟踪中，通常使用迁移学习将训练好的分类模型卷积层用于提取特征。如果这些特征全部用来训练判别器，一方面是对计算资源的浪费，另一方面卷积层中本身也会存在冗余特征。本文基于2.1节中训练的基于通道域和空间域融合注意力机制的分类卷积神经网络(ResNet-50+特征修正)，选择性提取卷积网络中每个层次的有效特征。为使卷积网络中不同层次的特征各尽其责，对卷积网络层中不同层次的特征信息对目标跟踪所贡献的特定性能进行了剖析，同时采用特征增强策略以增加目标跟踪过程中的正样本数量。

2.2.1 同一层次中卷积特征选取

在深层卷积网络构建中，网络层次逐渐加深时由于冗余层的存在会产生梯度消失、梯度爆炸和网络退化等问题。ResNet在深层网络中通过恒等映射屏蔽冗余层，设计一个残差块结构保证数据的输入和输出不变。然而，在实际的网络中冗余信息不仅存在于冗余层，在非冗余层中的通道和空间2个域上也同样存在。现有方法直接将一个完整的特征张量用于训练相关滤波器，不仅造成了资源浪费，也会对判别器的学习产生消极影响。因此，深层卷积网络中只有一个简单的残差块是不够的，需要更加智能的网络结构，在同一层次特征张量中自行学习记忆有用的通道和空间分布信息，并抑制负作用通道和冗余空间信息。

本文通过通道注意力权重$\boldsymbol{M}_{\text{channel}}$和空间注意力权重$\boldsymbol{M}_{\text{spatial}}$，分别从特征张量的通道和空间2个轴上选取有效的特征信息。如图5所示，首先，粗粒度选取ResNet残差块中$\boldsymbol{F}''$的通道注意力权重$\boldsymbol{M}_{\text{channel}}$高于阈值$T_{\text{channel}}$的元素所对应的通道特征得到$\boldsymbol{F}'_{\text{effective}}$，然后，将$\boldsymbol{F}'_{\text{effective}}$对应与空间注意力权重

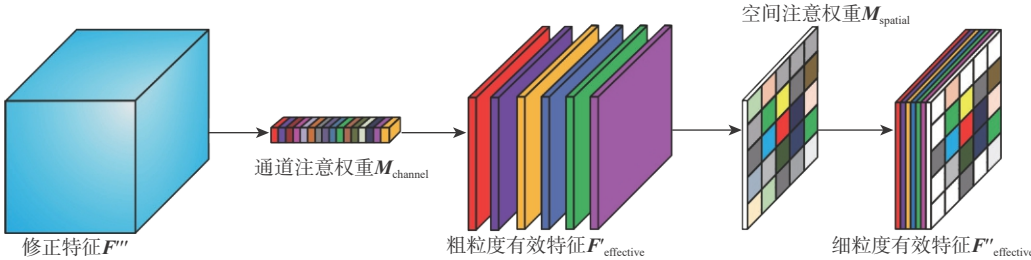


图 5　特征张量中有效特征的提取过程

Fig. 5　Extraction process of effective features in feature tensor

$M_{spatial}$中低于阈值$T_{spatial}$的元素值设置为 0, 使最终选择出的有效特征$F''_{effective}$将会在通道上包含更少的冗余信息, 在空间上包含更少的噪声和背景信息。

通道域上根据通道注意力权重筛选有效通道特征:

输入: $F'''_{N@h\times w}$, $M^{channel}_{N@1\times 1}$。

输出: $F'_{effective}$。

1. $i = 1 : N$ do
2. if $M^{channel}_{i@1\times 1} > T_{channel}$ then
3. $F'_{effective}$ append $(F'''_{i@h\times w})$
4. End for

$T_{channel}$和$T_{spatial}$的定义如下:

$$T_{channel} = \frac{1}{N} \tag{5}$$

$$T_{spatial} = \frac{1}{w\times h} \tag{6}$$

空间域上根据空间注意力权重筛选有效空间特征:

输入: $F'_{effective_{n@h\times w}}$, $M^{spatial}_{1@h\times w}$。

输出: $F''_{effective}$。

1. for $i = 1 : h$ do
2. for $j = 1 : w$ do
3. if $M^{spatial}_{1@(i,j)} < T_{spatial}$ then
4. $M^{spatial}_{1@(i,j)} = 0$
5. End for
6. End for
7. for $k = 1 : n$ do
8. $F''_{effective}$append $(F'_{effective k@h\times w} \otimes M^{spatial}_{1@h\times w})$
9. End for

2.2.2　卷积网络中卷积层次位置选取

传统的特征金字塔网络[30]使用自上而下和横向连接的方式来捕获不同尺度的物体。在线跟踪期间, 目标的大小经常变化。因此, 在跟踪框架中引入了分层功能。深层特征主要包含目标的高层语义信息, 但随着网络加深, 卷积特征的分辨率越来越低, 精度变差; 浅层网络主要表示目标的纹理和颜色信息, 但只要目标稍微产生形变, 浅层特征就会发生很大的变化, 鲁棒性很差。因此, 本文要做的核心改进是 2 种特征的区别对待, 利用深度特征训练跟踪器鲁棒性的同时, 利用浅层特征提高跟踪器的准确性。

卷积网络层从浅到深的语义信息对跟踪问题具有特定的性能。图 6 为 3 种跟踪情况 (遮挡、光照和简单场景) 下 ECO 跟踪器使用 ResNet50(“50”指的是 50 层残差网络)及其改进版 ResNet50+特征修正网络中不同层次卷积特征的性能。ResNet50 使用修正线性单元 (rectified linear unit, ReLU) 激活之前的网络层作为输出。由于相邻层的特征差异是相似的, 选择相邻层将引入更多冗余和干扰。选择图中每个阶段更好的性能水平作为特征提取层, 包括 1×1 卷积层 (1×1 convolution, Conv1x)、三维残差网络 (3D residual network, Res3d) 和 Res4f 残差块, 可以在增加特征的特异性的同时很容易地适应突变性的尺度变化。

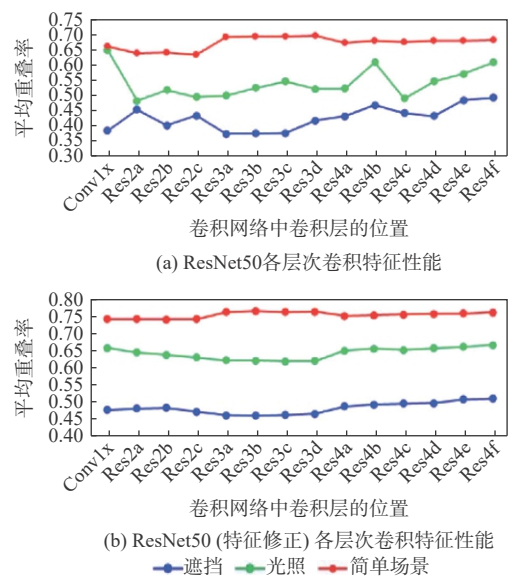


图 6　ECO 跟踪器在 3 种场景下不同层次卷积特征性能

Fig. 6　Different levels of convolutional feature performance of ECO tracker in three scenarios

对于 ResNet50+特征修正, 参考 2.2.1 节中基于特征修正网络提取目标的深度卷积特征, 通过

图 6(a)和图 6(b)对比，不难发现相比于 ResNet50，基于特征修正选取的卷积特征在 3 种场景下具有稳定的性能，同时也表现出不同层次卷积特征之间的差异性。选择图中每个阶段的更好的性能水平作为特征提取层，不仅可以增加特征的特异性，而且很容易地适应目标随机性的尺度变化。与 ResNet50 不同的是，经过特征修正网络的筛选之后的特征在数量上更少，在性能上更有效，能够在精度和实时性 2 个方面同时提高跟踪器性能。

2.2.3 目标视觉特征增强

通用目标跟踪任务中，仅第 1 帧是真实目标，其他训练样本都是跟踪器在跟踪过程中依据跟踪结果产生的，这就导致训练过程中匮乏真实的目标样本，而且样本的产生完全依赖于循环矩阵，这样的样本只有水平和纵向的移动信息，而缺乏目标旋转、形变和遮挡等信息。因此，本文采用以下增强方法：①镜像，水平翻转；②旋转，从−60°～60°的固定 12 个角度旋转；③平移，水平和竖直平移 n 像素再提取特征，等价于特征图平移 n/s 像素，s 为卷积网络中卷积核的步进；④模糊，高斯滤波模糊，模拟跟踪场景中常见的运动模型和尺度变化；⑤丢弃，通道级特征丢弃，随机 20% 的特征通道特征设置为 0，其余通道放大以保持样本能量。图 7 利用曲线下的面积(area under the curve, AUC)分数[10]表征数据增强对跟踪性能的影响。可以看出，深层特征始终受益于数据增强。除“平移”外，所有增强功能都会使跟踪效果提高 1% 以上。使用“模糊”增强功能可最大程度地提高性能，比不使用数据增强的基线 ECO 提高 4%。同时，浅层特征不会从数据扩充中受益。深度特征反映的语义信息，对这些特征增强都有不变性，能从增加的正样本获益。而浅层特征经过特征增强后其所反映的纹理信息会发生剧变，会扰乱跟踪器并损害其性能，该情况下得到的样本可以视为负样本。

由图 7 可知，ResNet50 的特征在经过特征修正网络的通道域、空间域修正和选取之后，在深层特征方面对跟踪器保持稳定的高增益，同时在浅层特征保持稳定的负增益。这也从另一方面验证了经过特征修正后的 ResNet50 内部包含更少的冗余信息，特征与目标保持很好的正相关性。

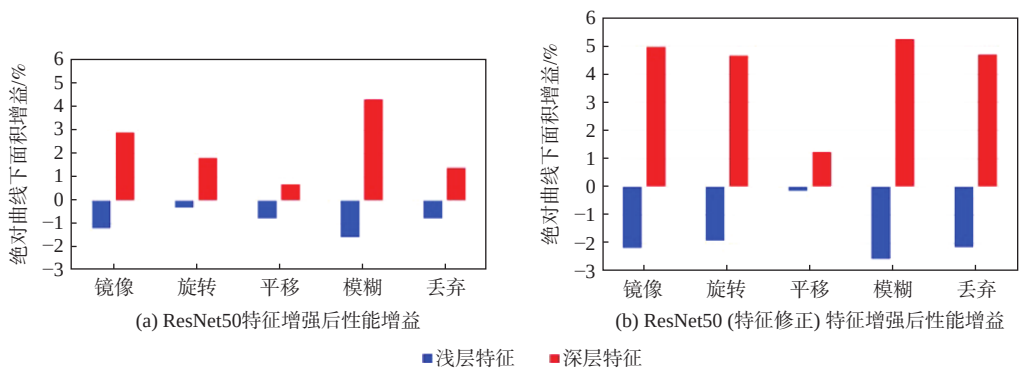


图 7 基于 ECO 跟踪器各层次特征增强后性能增益对比

Fig. 7 Performance gain comparison based on feature enhancement at each level of ECO tracker

3 实验与分析

3.1 特征修正模块分类效果实验

对于本文提出的特征修正模块，在 ImageNet-1K 上进行分类实验以严格评估，同时将其与现有的基于通道注意力的网络模型 SE-ResNet[31]和基于通道域和空间域的卷积块注意力模块(convolutional block attention module, CBAM)[32]在各项实验指标上进行对比。

3.1.1 数据集与评价指标

在 ImageNet 中，总共划分为 21 841 个类别，每个概念图像都是质量控制和人为标注的。主要包括：两栖动物、覆盖物、装置、车辆、人等。

本实验基于模型的参数量、浮点运算速度、帧率、Top-1、Top-5 错误率对特征修正模块、SE-ResNet 和 CBAM 进行评估。浮点运算速度常用于衡量模型计算复杂度。Top-1 和 Top-5 分别表示模型预测的前 1 和前 5 个类别中命中真实类别的比例(或错误率)。

3.1.2 目标分类网络实验结果及分析

表 1 总结了对比实验结果。具有特征修正模块的 ResNet 网络明显优于所有其他算法，这表明本文提出的特征修正模块可以很好地在大规模数据集中的各种模型上推广。此外，Top-1、Top-5 错误率显著降低，验证了使用特征修正模块可以在 SE-ResNet 和 CBAM 的基础上提高准确性。在不同层次的 ResNet 网络中，加入特征修正模块与加入压缩激励(squeeze-and-excitation, SE)模块、CBAM 对比所产生的参数量、浮点运算速度均近似或降低。

表 1　基于 ResNet 的不同网络结构在 ImageNet-1K 上的分类结果

Table 1　Classification results of different network architectures based on ResNet on ImageNet-1K

网络结构	参数量/MB	浮点运算速度/10^9 s^{-1}	Top-1 错误率/%	Top-5 错误率/%
ResNet18	11.69	1.814	29.60	10.55
ResNet18+SE	11.78	1.814	29.41	10.22
ResNet18+CBAM	11.78	1.815	29.27	10.09
ResNet18+空间自注意力	11.74	1.814	29.46	10.34
ResNet18+空间自注意力+通道注意力	11.76	1.814	29.23	10.07
ResNet18+特征修正	11.76	1.814	29.12	9.95
ResNet34	21.08	3.664	26.69	8.60
ResNet34+SE	21.96	3.664	26.13	8.35
ResNet34+CBAM	21.96	3.665	25.99	8.24
ResNet34+空间自注意力	21.95	3.664	26.02	8.50
ResNet34+空间自注意力+通道注意力	21.96	3.664	25.64	8.26
ResNet34+特征修正	21.96	3.664	25.45	8.11
ResNet50	25.56	3.858	24.56	7.50
ResNet50+SE	28.09	3.860	23.14	6.70
ResNet50+CBAM	28.09	3.864	22.66	6.31
ResNet50+空间自注意力	28.08	3.861	23.08	6.75
ResNet50+空间自注意力+通道注意力	28.09	3.862	21.69	6.20
ResNet50+特征修正	28.09	3.862	21.01	6.01
ResNet101	44.55	7.570	23.38	6.88
ResNet101+SE	49.33	7.575	23.35	6.19
ResNet101+CBAM	49.33	7.581	21.51	5.69
ResNet101+空间自注意力	49.32	7.579	22.48	6.24
ResNet101+空间自注意力+通道注意力	49.33	7.579	21.09	5.33
ResNet101+特征修正	49.33	7.579	20.69	5.07

如 ResNet50 中, 加入对比模块与特征修正模块产生了相同的参数量, 仅比纯 ResNet50 网络多出 2.53 MB, 而浮点运算速度持平。这显示出本文模型在不增大实际计算资源的基础上提高了模型的准确性, 并且特征修正模块取得优秀的目标分类性能, 在 ResNet50+特征修正结构中, Top-1 和 Top-5 错误率分别为 21.01% 和 6.01%, 相比纯 ResNet50 结构分别提升 3.55% 和 1.49%。因此, 就参数和计算而言, 特征修正模块的总体开销很小, 其可以实现在各种卷积网络中的轻量级即插即用。

3.1.3　特征修正模块消融试验

在表 1 中, 同时对本文特征修正模块中的 3 个子模块 (空间自注意力、通道注意力和空间注意力) 进行消融试验, 分别在 ResNet 网络中加入不同子模块, 分析其对最终性能的贡献。具体而言, 在每个 ResNet 网络中, 分别加入空间自注意力、空间自注意力+通道注意力及完整的特征修正模块 3 组实验, 观察在 ImageNet 中的分类结果。不难发现, 分别加入子模块后, 模型的错误率持续降低, 且参数量和浮点运算速度指标均相近, 可以认为各个子模块均对分类任务产生了正向效果。这验证了本文各子模块的有效性及其合并算法的功效, 生成更丰富的描述特征可以有效地补充通道注意力和空间注意力。

3.1.4　特征修正模块对注意力的改进可视化

在定性分析中, 研究使用来自 ImageNet 验证集的图像将类激活热力图(gradient-weighted class activation heatmap, Grad-CAM)[33] 应用于不同的网络。类激活热力图使用渐变来计算卷积层中空间位置的重要性, 其结果可清晰显示对目标判别起正影响的区域。通过观察网络认为对预测类别重要的区域可以查看该网络如何充分利用注意力机制。

将特征修正集成网络 (ResNet50+特征修正) 与 ResNet50 和 SE 集成网络 (ResNet50+SE) 及 CBAM 集成网络 (ResNet50+CBAM) 的可视化结果进行比较, 如图 8 所示。可以发现本文的特征修正集成网络的类激活热力图遮罩更好地覆盖了目标对象区域, 能够有效利用目标对象区域中的信息并从中聚集特征, 目标类别分数也相应增加。

空间自注意力权重参数是通过1×1的卷积生成的, 该网络可以在整体模型的优化过程中不断学习目标本身的空间位置分布, 减少背景噪声的影响, 这是通道注意力-空间注意力所不具备的能力; 同时, 本文将 CBAM 中的全连接和3×3卷积全部替

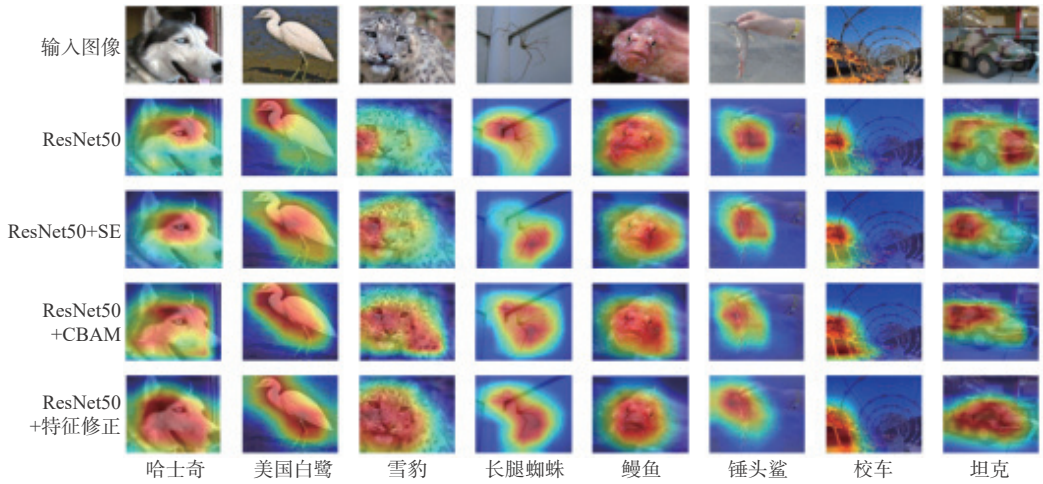


图8 3种分类网络对目标识别的类激活热力图可视化比较

Fig. 8 Visual comparison of class activation heatmaps of three classification networks for object recognition

换为1×1卷积，使得在增加空间自注意力后模型参数量不增加的同时，提高了计算速度和模型预测精度。因此，该轻量级注意力模块在保持较小开销的同时，实现了可观的性能改进，并且相对独立，可以实现对现有的大部分卷积网络 (AlexNet, VGG, ResNet) 的即插即用。

3.2 有效卷积特征跟踪验证实验

对于2.2节提出的对深度卷积特征的构建、提取和增强之后的有效卷积特征应用于现有的相关滤波器优秀跟踪算法在不同的数据集中进行特征优化，对比分析以验证基于空间域和通道域融合的注意力机制提取目标卷积特征的有效性。

3.2.1 数据集、对比模型与评价指标

目标跟踪领域已经形成的几个重要的基准数据集：OTB-100[34]、VOT[35-38]、UAV123[39]等。本文对最为广泛使用的OTB100和VOT系列数据集进行实验。本节对现有的基于相关滤波器的优秀算法ECO、C-COT、DeepSRDCF、互补学习器实时跟踪器(sum of template and pixel-wise learners, Staple)[30]、判别尺度空间跟踪器(discriminative scale space tracker, DSST)[40]统一使用ResNet50网络提取深度特征，同时加入修正卷积特征或利用修正后的特征替换其本身的深度特征得到优化算法：ECO-M、DeepSRDCF-M、Staple-M、DSST-M(“-M”表示在基础跟踪器上加入本文提出的特征修正模块后的改进版本)。

在OTB数据集中，使用目标中心位置误差绘制跟踪过程的精度图，计算一个视频序列中所有跟踪结果的平均中心位置误差，用于汇总跟踪器的整体性能。作为每个跟踪器的代表精度得分，OTB100使用阈值为20像素的得分。此外，通过成功率指标评估跟踪结果与人工标注的真实边界框的重叠率。成功率图显示了阈值从0～1变化的成功帧比率。而使用特定阈值的成功率值进行评估可能不具有代表性。因此，为避免单一阈值带来的偏差，本文采用成功率AUC作为综合评价指标，并据此对各跟踪算法进行评估。

VOT数据集主要使用3个评价指标：精确率A、鲁棒性R、预期平均重叠(expected average overlap, EAO)。精确率衡量跟踪器预测的边界框与地面真值的重叠程度。在时间步t处的跟踪精确率定义为跟踪器预测边界框A_t^{P}与人工标注真实边界框A_t^{G}之间的重叠：

$$\Phi_t = \frac{A_t^{\mathrm{G}} \cap A_t^{\mathrm{P}}}{A_t^{\mathrm{G}} \cup A_t^{\mathrm{P}}} \tag{7}$$

鲁棒性衡量的是跟踪器在跟踪过程中失去目标的次数，其通过故障率来衡量。故障率是跟踪器偏离目标并必须重新初始化的次数，一旦重叠度低于阈值，就定义为跟踪器丢失目标。可以预期，如果跟踪器在特定帧中发生故障，在下一帧中对其进行初始化，则很有可能会再次失败。为降低这种直接相关性，在跟踪失败5帧后初始化跟踪器。在给定的帧序列上对故障率进行多次测量，其平均值即平均鲁棒性。EAO是希望一个好的跟踪器同时拥有好的A和R，但直接用A和R的2个数加权和有失公允。因此，把理想EAO定义为把N_{s}帧的视频序列从1到一个期望的极大值对应的$\Phi_{N_{\mathrm{s}}}$求平均值，即期望平均覆盖率：

$$\Phi_{N_{\mathrm{s}}} = \frac{1}{N_{\mathrm{s}}} \sum_{i=1}^{N_{\mathrm{s}}} \Phi_i \tag{8}$$

3.2.2 OTB100实验分析对比

图9展示了现有优秀相关滤波器算法和在其基础上加入修正特征之后的相应优化算法在OTB数据集100个视频序列上的成功率和精确率，图例

中数值为成功率均值。

不难发现, 在目标定位的成功率上, 修正后的特征对相应的算法基本保持 2% 的增益; 在目标定位精确率方面本文提出的修正特征稳定保持 1% 的增益。

图 10 显示了在 3 个示例序列上, 加入修正特征之后的算法与基础算法相比较的部分可视化案例。在这 3 种情况下, 基础算法都在目标特定区域严重过拟合。在比例尺变化、变形和面外旋转的情况下使目标估计不佳。而加入修正特征的跟踪器能更好解决判别器的过拟合现象, 从而更好地预测目标外观。

3.2.3　VOT2018 实验分析对比

对于 VOT 数据集中的每个序列, 使用第 1 帧中目标的矩形框进行初始化来评估跟踪器。只要目标丢失, 该工具包就会重新启动跟踪器。表 2 显示了在 VOT2018 数据集上的 A、R、EAO 和帧率。本文提出的修正特征对基础相关滤波器算法在 EAO 得分方面获得普遍增益。值得注意的是在 VOT 数据集上, 加入修正特征的 C-COT-M 和 Stalple-M 的 EAO 得分要高于 ECO 算法。

对比不同跟踪器在 VOT2018 数据集上进行跟踪任务的帧率。不加特征修正的原模型采用 ResNet 提取深度特征, 加入特征修正模块后, 由于对特征进行了不同维度的筛选, 从通道和空间维度剔除了大量冗余特征并保留了真正有效的特征, 帧率均有所提高, 且保持了较高的跟踪性能。图 11 为修正后跟踪算法及其基础算法在 VOT2018 基准上的性能。由图可知, 修正特征不仅可以提高目标跟踪的精确率, 同时可以提高鲁棒性。尤其值得注意的是, Staple 算法在加入修正特征之后其准确性得到了较高的提升, 因其本身是不带有深度特征的, 这也证实了深度特征在目标特征表示方面的有效性。图 12 为在 VOT2018 数据集 3 个示例序列上, 加入修正特征

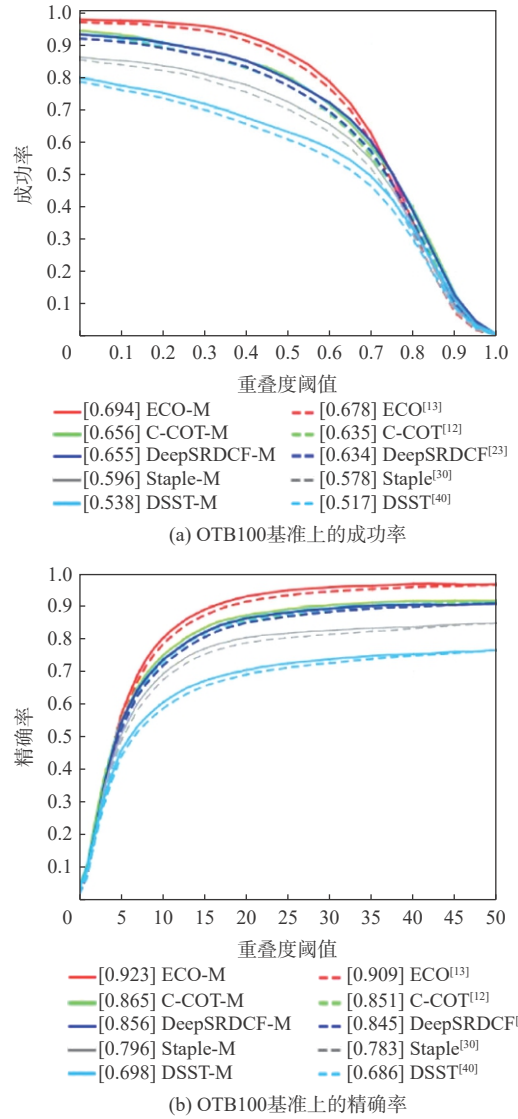


图 9　OTB100 基准上的成功率和精确率

Fig. 9　Success rate plot and precision rate plot on the OTB100 benchmark

之后的算法与基础算法相比较的部分可视化结果。在比例尺变化的情况下, 基础算法均导致目标

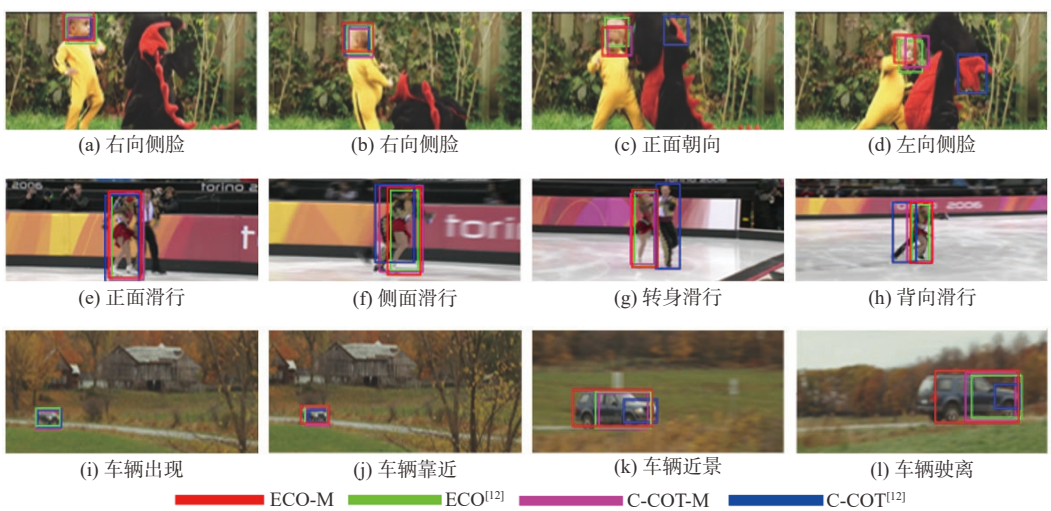


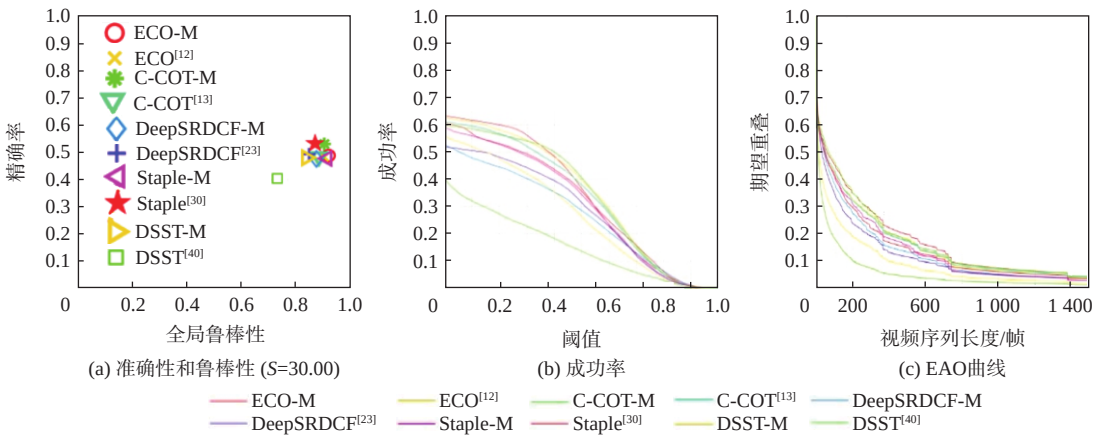
图 10　4 种跟踪算法在 OTB100 数据集 3 个视频序列上的定性结果

Fig. 10　Qualitative results of four tracking algorithms on three video sequences of OTB100

表 2 VOT2018 视频序列基准上的性能比较

Table 2 Performance comparison on VOT2018 video sequence benchmark

跟踪器	精确率	鲁棒性/次	EAO	帧率/(帧·s^{-1})
ECO-M	0.508 0	13.900 9	0.329 4	32
ECO[13]	0.497 8	13.511 2	0.307 7	25
C-COT-M	0.537 8	14.422 8	0.329 2	27
C-COT[12]	0.505 7	15.228 8	0.303 9	18
DeepSRDCF-M	0.480 4	20.921 2	0.254 1	13
DeepSRDCF[23]	0.501 6	23.964 4	0.228 2	6
Staple-M	0.498 5	14.629 6	0.289 0	34
Staple[30]	0.540 5	19.883 6	0.273 3	28
DSST-M	0.472 1	30.122 5	0.178 0	25
DSST[40]	0.400 5	63.072 3	0.097 6	20



(a) 准确性和鲁棒性 (S=30.00) (b) 成功率 (c) EAO曲线

图 11 在 VOT2018 基准上修正后的跟踪算法及其基础算法的性能

Fig. 11 Performance of corrected tracking algorithms and their baseline counterparts on VOT2018 benchmark



(a) 进出展示 (b) 全展展示 (c) 侧身展式 (d) 遮挡展示

(e) 空中翻滚 (f) 空中俯冲 (g) 车辆落地 (h) 正常行驶

(i) 远距助跑 (j) 即将起跳 (k) 腾空翻转 (l) 落地结束

图 12 4 种算法在 VOT2018 数据集 3 个视频序列上的定性结果

Fig. 12 Qualitative results of four algorithms on three video sequences of VOT2018

估计不佳。而本文提出的跟踪器能够有效解决过拟合问题。

4　结　论

1）经过本文特征修正网络的筛选之后的特征在数量上更少，在性能上更有效，能够在精度和实时性两方面同时提高跟踪器性能。

2）本文特征修正模块取得优秀的目标分类性能，在 ResNet50+特征修正结构中，Top-1 和 Top-5 错误率分别为 21.01% 和 6.01%，相比纯 ResNet50 结构分别提升 3.55% 和 1.49%。

3）就参数和计算而言，本文特征修正模块的总体开销很小，其可以实现在各种卷积网络中的轻量级即插即用。

4）本文特征修正模块应用在目标跟踪中，基础算法在增加特征修正模块后均取得了更优秀的跟踪性能，包括精确率、成功率、鲁棒性等，有效解决判别器在不同场景下的过拟合情况。

参考文献（References）

[1] ZHANG Y F, SUN P Z, JIANG Y, et al. ByteTrack: multi-object tracking by associating every detection box[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2022: 1-21.

[2] 黄煜杰, 陈凯, 王子源, 等. 多目视觉下基于融合特征的密集行人跟踪方法[J]. 北京航空航天大学学报, 2025, 51(7): 2513-2525.
HUANG Y J, CHEN K, WANG Z Y, et al. A dense pedestrian tracking method based on fusion features under multi-vision[J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(7): 2513-2525(in Chinese).

[3] STADLER D, BEYERER J. Improving multiple pedestrian tracking by track management and occlusion handling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 10953-10962.

[4] CHEN K, ZHU H H, TANG D B, et al. Future pedestrian location prediction in first-person videos for autonomous vehicles and social robots[J]. Image and Vision Computing, 2023, 134: 104671.

[5] CHEN K, SONG X, YUAN H T, et al. Fully convolutional encoder-decoder with an attention mechanism for practical pedestrian trajectory prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11): 20046-20060.

[6] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional Siamese networks for object tracking[C]//Proceedings of the Computer Vision-ECCV Workshops. Berlin: Springer, 2016: 850-865.

[7] WANG Q, GAO J, XING J L, et al. DCFNet: discriminant correlation filters network for visual tracking[EB/OL]. (2017-04-13)[2024-02-01]. https://arxiv.org/abs/1704.04057.

[8] REAL E, SHLENS J, MAZZOCCHI S, et al. YouTube-boundingboxes: a large high-precision human-annotated data set for object detection in video[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 7464-7473.

[9] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[10] ZHANG L, WANG Y J, SUN H H, et al. Robust visual correlation tracking[J]. Mathematical Problems in Engineering, 2015, 2015: 238971.

[11] LUQUE-BAENA R M, ORTIZ-DE-LAZCANO-LOBATO J M, LÓPEZ-RUBIO E, et al. A competitive neural network for multiple object tracking in video sequence analysis[J]. Neural Processing Letters, 2013, 37(1): 47-67.

[12] DANELLJAN M, ROBINSON A, SHAHBAZ KHAN F, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2016: 472-488.

[13] DANELLJAN M, BHAT G, KHAN F S, et al. ECO: efficient convolution operators for tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6931-6939.

[14] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

[15] LI H X, LI Y, PORIKLI F. DeepTrack: learning discriminative feature representations online for robust visual tracking[J]. IEEE Transactions on Image Processing, 2016, 25(4): 1834-1848.

[16] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 4293-4302.

[17] LI B, YAN J J, WU W, et al. High performance visual tracking with Siamese Region proposal network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8971-8980.

[18] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2024-02-01]. https://arxiv.org/abs/1409.1556.

[19] LI Y H, ZHANG X F, CHEN D M. SiamVGG: visual tracking using deeper Siamese networks[EB/OL]. (2022-07-04)[2024-02-01]. https://arxiv.org/abs/1902.02804.

[20] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. communications of the ACM, 2017, 60(6): 84-90.

[21] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2010: 2544-2550.

[22] DANELLJAN M, HÄGER G, KHAN F S, et al. Learning spatially regularized correlation filters for visual tracking[C]// Proceedings of the IEEE International Conference on Computer Vision.Piscataway: IEEE Press, 2016: 4310-4318.

[23] DANELLJAN M, HÄGER G, KHAN F S, et al. Convolutional features for correlation filter based visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision Workshop. Piscataway: IEEE Press, 2016: 621-629.

[24] HE Z Q, FAN Y R, ZHUANG J F, et al. Correlation filters with weighted convolution responses[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE Press, 2018: 1992-2000.

[25] YU E, LI Z L, HAN S D, et al. RelationTrack: relation-aware multiple object tracking with decoupled representation[J]. IEEE Transactions on Multimedia, 2023, 25: 2686-2697.

[26] ZENG F G, DONG B, ZHANG Y A, et al. MOTR: end-to-end multiple-object tracking withTransformer[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2022: 659-675.

[27] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2014: 818-833.

[28] ZHOU B L, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 2921-2929.

[29] LI H, LIN Z C. Accelerated proximal gradient methods for nonconvex programming[C]//Proceedings of the 29th International Conference on Neural Information Processing Systems-Volume 1. New York: ACM, 2015: 379-387.

[30] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: complementary learners for real-time tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1401-1409.

[31] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.

[32] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the Computer Vision-ECCV 2018. Berlin: Springer, 2018: 3-19.

[33] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks *via* gradient-based localization[J].International Journal of Computer Vision, 2020, 128(2): 336-359.

[34] WU Y, LIM J, YANG M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.

[35] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking VOT2016 challenge results[C]//Proceedings of the Computer Vision-ECCV Workshops. Berlin: Springer, 2016: 777-823.

[36] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking VOT2017 challenge results[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops.Piscataway: IEEE Press, 2018: 1949-1972.

[37] KRISTAN M, LEONARDIS A, MATAS J, et al. The sixth visual object tracking VOT2018 challenge results[C]// Proceedings of the Computer Vision-ECCV Workshops. Berlin: Springer, 2019: 3-53.

[38] KRISTAN M, MATAS J, LEONARDIS A, et al. The seventh visual object tracking VOT2019 challenge results[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop. Piscataway: IEEE Press, 2020: 2206-2241.

[39] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for UAV tracking[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2016: 445-461.

[40] CHEN Y T, WANG J, XIA R L, et al. The visual object tracking algorithm research based on adaptive combination kernel[J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10(12): 4855-4867.

Object tracking algorithm based on deep feature modification

CHEN Kai[1]，HUANG Yujie[1,*]，ZHAO Xiaodong[1]，WANG Pengfei[2]，CHEN Kai[1]，LIN Yanze[1]

(1. College of Mechanical and Electrical Engineering，Nanjing University of Aeronautics and Astronautics，Nanjing 210016，China；
2. The 28th Research Institute of China Electronics Technology Group，Nanjing 210007，China)

Abstract：In a complex tracking environment, existing trackers mainly face the problems of deep convolutional feature redundancy and a lack of positive samples in the target tracking process. In order to remove redundant features in deep convolutional networks, an attention mechanism model based on the fusion of the spatial domain and channel domain, which includes three continuous sub-modules: spatial self-attention, channel attention and spatial attention. On this basis, a deep convolutional network suitable for various vision algorithms is constructed. A feature enhancement technique based on feature modification is meant to extract target deep convolutional features in order to address the issues of negative feedback of redundant features and lack of positive samples. The experimental results show that in the mainstream residual networks (ResNet) object classification tasks, the addition of the feature modification module can significantly reduce the Top-1 and Top-5 error rates, and will not cause additional calculation or network structure adjustment, achieving lightweight insertion. Several tracking methods are combined with the feature modification module to improve tracking performance and address the discriminator's over-fitting issue.

Keywords：object tracking; attention mechanism; feature modification; lightweight insertion; residual network

Received：2024-04-07；**Accepted**：2024-05-09；**Published Online**：2024-06-11 08：40

URL：link.cnki.net/urlid/11.2625.V.20240607.0959.001

Foundation items：National Natural Science Foundation of China (52202417); China Postdoctoral Science Foundation (2022TQ0155,2022M721605); Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (VRLAB2023A02)

* **Corresponding author**. E-mail：yj_huang@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0197

基于图像语义分割的目标跟踪尺度自适应算法

陈凯[1]，赵晓冬[1,*]，黄煜杰[1]，王鹏飞[2]，雷一辰[1]，张彪[1]

(1. 南京航空航天大学 机电学院，南京 210016；2. 中国电子科技集团第二十八研究所，南京 210007)

摘 要：围绕如何在复杂的场景中充分利用图像语义分割得到目标的语义信息，并对目标跟踪器的输出进行尺度优化，设计了基于注意力机制优化的图像语义分割网络。针对目标跟踪器的输出和特征输入2个方面进行优化，该网络可以实现对各类算法的即插即用。利用图像语义分割掩码获得目标的旋转框边界，根据目标的旋转框和非旋转框边界对目标输入阶段的特征进行去噪优化，减弱背景噪声对跟踪器的判别影响。分别从所设计网络的结构、训练、目标旋转框标定及对跟踪器的输入特征进行去噪等方面进行讨论。在公开数据集OTB100、VOT2016和VOT2018上进行实验，对比验证了目标运动模型在解决目标跟踪过程中，目标尺度优化的准确率和鲁棒性。

关 键 词：目标跟踪；语义分割；注意力机制；尺度优化；背景去噪

中图分类号：TP391.4

文献标志码：A **文章编号**：1001-5965(2026)05-1536-11

目标跟踪的尺度自适应问题一直是相关滤波器跟踪算法的一大难点，也是目标跟踪的重点研究内容之一。因为能够应用到实际生产中的跟踪器，应该在提供精确目标尺度的同时保证较高的计算效率。但是，在目标跟踪算法中添加更多的尺度估计需要高昂的计算代价，因此，需要探索设计一种高效且准确的尺度估计算法对跟踪的输出结果进行优化。

相关滤波器尺寸估计经典算法——尺度自适应多特征融合(scale adaptive multiple feature, SAMF)[1]是一种无参数、不更新的算法，其尺度估计是否准确比较依赖于定位器的性能是否鲁棒。尺度估计和跟踪定位器的学习可以分开单独进行，但是多尺度在一定程度上可以对定位器定位起促进作用。现有的研究大多将两者融合互补进行应用，这样便有效避免了尺度估计偏移过大情况的发生。目前广泛应用于空间正则化判别相关滤波器(spatially regularized discriminative correlation filters, SRDCF)[2]系列目标跟踪算法。判别式尺度空间跟踪器(discriminative scale space tracker, DSST)[3]的流程是训练->定位->尺度估计->更新(尺度估计器更新，定位器更新)->定位->尺度估计->更新不断循环迭代。缺点是目标变化大则估计不准且定位精度需求高。bboxRR[4]在深度学习跟踪器中被普遍使用，对目标变化大的情况比DSST更加鲁棒，对定位器的准确度要求没有DSST高，但是缺乏较好的更新策略。

最新的基于深度孪生网络结构的跟踪算法SiamMask[5]给出了目标跟踪尺度自适应的优化方向，在网络结构中加入图像语义分割模块获得目标的分割掩码，根据该掩码信息可以获得高精度的目标旋转矩形框标定结果。但是基于孪生网络结构算法最大的缺点为过度依赖于网络在大量数据集中的离线训练结果，在线跟踪时并不会更新网络的参数以适应新的目标，从而导致目标在急剧旋转的情况下容易出现跟踪漂移的问题。因此，本文基于目标的图像语义分割信息对相关滤波器算法的输

收稿日期：2024-04-07；**录用日期**：2024-05-24；**网络出版时间**：2024-06-08 10:11

网络出版地址：link.cnki.net/urlid/11.2625.V.20240607.1715.003

基金项目：国家自然科学基金(52202417)；中国博士后科学基金(2022TQ0155,2022M721605)；虚拟现实技术与系统全国重点实验室(北京航空航天大学)开放课题基金(VRLAB2023A02)；中国科协青年科技人才托举工程(2023QNRC001)；江苏省科协青年科技人才托举工程(JSTJ-2023-XH032)

* **通信作者**. E-mail：xdzhao@nuaa.edu.cn

引用格式：*陈凯，赵晓冬，黄煜杰，等. 基于图像语义分割的目标跟踪尺度自适应算法[J]. 北京航空航天大学学报，2026，52(5)：1536-1546. CHEN K，ZHAO X D，HUANG Y J，et al. Adaptive algorithm for target tracking scale based on image semantic segmentation[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1536-1546 (in Chinese).*

入和输出进行优化, 一方面可以提高目标跟踪器输出的鲁棒性, 还可以提高输出尺度的精确性。

1　相关工作

最近的高性能跟踪算法主要分为 2 大类别: 预测分值和边界框回归。前者主要通过模型输出的响应图定位目标的位置, 因为其用到多尺度的目标特征信息, 故可以通过不同尺度的特征获得目标的尺度估计, 缺点为难以获得目标的长宽比信息, 其中, 典型的代表有相关滤波器跟踪算法和全卷积孪生网络 (fully-convolutional siamese networks, SiamFC)[6]。而对于边界框回归算法则大多是基于目标检测算法的 anchor 机制, 能够得到相比于预测分值算法更精确的目标尺度边界框, 其中, 代表性的成果有基于回归网络的通用目标跟踪 (generic object tracking using regression networks, GOTURN)[7] 和孪生区域建议网络 (siamese region proposal networks, AD-SiamRPN)[8]。

最新基于孪生网络结构的 SiamMask 创新性地将图像语义分割网络融入 AD-SiamRPN 中, 通过目标的掩码进一步提升目标边界框标定的准确性, 在公开数据集的测试中表现优异。证明了高精度的图像语义分割算法是解决目标跟踪边界框尺度优化的一个有效途径。

现有的许多研究都对使用卷积神经网络 (convolutional neural networks, CNN)[9] 进行对象实例分割[10-13] 的算法进行了相关探索。标准前馈 CNN[14-17] 中交替堆叠的卷积层和池化层可以合并控制模型的能力并增加接受域, 从而导致粗糙的、高度语义化的特征表示。虽然这对于提取对象级信息是必要的, 但这种体系结构会呈现与像素级变化无关的低分辨率现象。这对于分类和识别对象实例是有益的, 但对像素标记任务提出了挑战。因此, 仅使用上层网络进行对象实例分割的 CNN[10-12] 可以有效地生成粗糙的对象掩码, 但是很难实现像素精确的分割。

对于诸如语义分割和边缘检测之类的像素标记任务, “跳过” 连接[13,18-20] 获得了广泛的应用。在实践中, 常见的跳过体系结构等效于对每个网络层进行独立的预测, 并对结果进行上采样和平均池化。这对于语义分割是有效的, 因为早期层中的本地接受域可以为像素标记提供足够的数据。但是, 对于目标分割任务中的目标实例而言, 局部接收域是不充分的, 因此, 有必要区分目标实例之间的区别 (如可以标记羊皮的局部斑块, 但是如果没有目标层次的信息, 则很难确定它们是否属于同一动物)。

2　本文算法

本文提出基于图像语义分割的目标跟踪尺度自适应算法, 该算法基于注意力机制, 结合 ResNet50[21] 及自主设计的特征修正网络构建了一个端到端的图像语义分割网络, 并利用该网络对相关滤波器跟踪算法的目标尺度进行优化。其主要流程如图 1 所示, 在跟踪过程中, 考虑到不同旋转角度标定框的外接矩形要大于目标掩码本身的尺寸, 所以本文对上一帧的跟踪结果扩大 1.5 倍范围之后经过图像语义分割网络获得目标的掩码。根据目标的掩码可以获得目标的旋转矩形和非旋转矩形 2 个标定框, 并根据这 2 个标定框缩减目标在图像中的范围。当新一帧图像输入到特征提取器之后, 利用新的目标掩码对包含目标特征的矩阵进行背景降噪, 经过降噪后的目标特征可以提高判别器的置信度以获得更准确的新的目标跟踪结果, 如此循环下去最终实现对整个视频的目标跟踪。

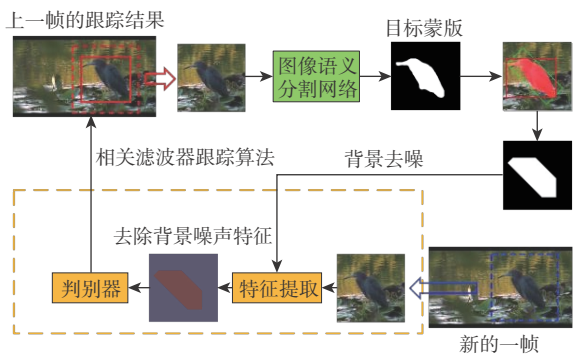


图 1　本文算法整体流程图

Fig. 1　Overall flow chart of the proposed algorithm

2.1　基于注意力机制的图像语义分割模型

本文目的是有效地合并来自底层功能空间丰富的信息和在高层网络中编码的语义信息。本文的图像语义分割算法遵循 3 个原则: ①分割对象通常需要对象级别的信息; ②给定对象级别的信息, 分割应该以自后向前的方式进行, 依次整合来自较早层的信息; ③本文算法应将合并带来的分辨率损失 (最终输出与输入的分辨率匹配) 相抵消。

为满足这些原则, 本文通过自后而前的优化过程增强了标准前馈网络, 具体算法如图 2 所示。本文引入了 “上采样优化模块” U_{p}, 其负责上采样并使输入掩码编码的分辨率加倍。每个上采样优化模块 $U_{\mathrm{p}i}$ 将自后而前的上采样中生成的掩码编码 $\boldsymbol{M}^{(\mathrm{Res}\text{-}i)}$ 及自前而后的下采样中生成的修正特征 $\boldsymbol{F}^{(\mathrm{Res}\text{-}i)}$ 作为输入, 并学习合并信息以生成新的上采样对象编码 $\boldsymbol{M}^{(\mathrm{Res}\text{-}i)}_{(\mathrm{M}\text{-up})}$。可以得到: $\boldsymbol{M}^{(\mathrm{Res}\text{-}i)}_{(\mathrm{M}\text{-up})}=U_{\mathrm{p}i}(\boldsymbol{M}^{(\mathrm{Res}\text{-}i)},\boldsymbol{F}^{(\mathrm{Res}\text{-}i)})$, 此处

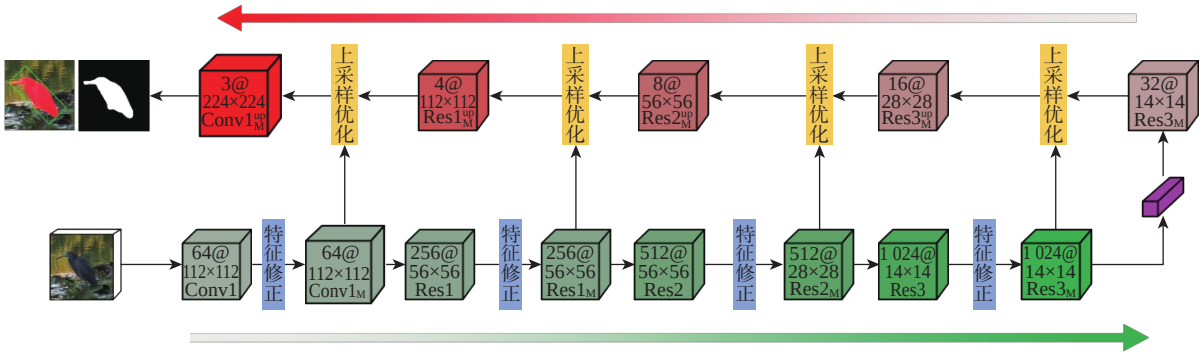


图2 本文图像语义分割网络的结构

Fig. 2 Structure of the image semantic segmentation network

可参考图3所示，整个前馈过程中堆叠了多个这样的模块。整个网络自前而后的过程中，本文自主设计了ResNet50+特征修正网络模块提取带有注意力参数的下采样特征。网络的最终输出是与输入图像具有相同分辨率的像素标签。

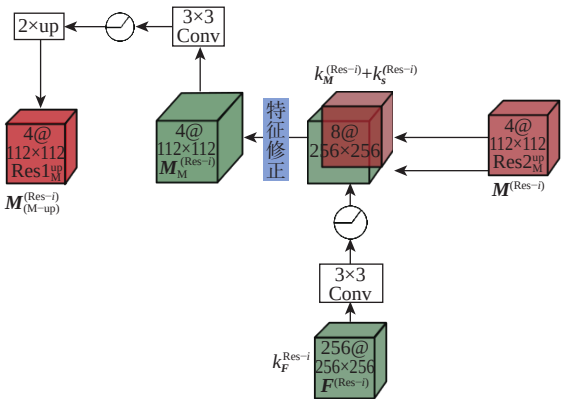


图3 上采样优化模块的结构

Fig. 3 Structure of the oversampling optimization module

2.1.1 上采样优化模块

上采样优化模块主要用来逆转下采样过程中的池化作用，输入来自上层的掩码编码和自前而后传递的匹配特征图，并且融合2个过程的信息来生成一个有2倍空间分辨率的掩码编码。这个过程一直持续到掩码的全部分辨率被恢复，最终的输出为精细的目标掩码。

如图3所示，“上采样优化模块”U_{p}的输入为来自上一层的$\boldsymbol{M}^{(\mathrm{Res}\text{-}i)}$和下一层的修正特征$\boldsymbol{F}^{(\mathrm{Res}\text{-}i)}$(池化前的卷积层的输出)，两者的空间分辨率相同，通道数分别为$k_{\boldsymbol{M}}^{(\mathrm{Res}\text{-}i)}$和$k_{\boldsymbol{F}}^{(\mathrm{Res}\text{-}i)}$，上采样优化要完成$\boldsymbol{M}_{(\mathrm{M\text{-}up})}^{(\mathrm{Res}\text{-}i)}=U_{pi}(\boldsymbol{M}^{(\mathrm{Res}\text{-}i)},\boldsymbol{F}^{(\mathrm{Res}\text{-}i)})$。通常，$k_{\boldsymbol{F}}^{(\mathrm{Res}\text{-}i)}$远大于$k_{\boldsymbol{M}}^{(\mathrm{Res}\text{-}i)}$，如果直接将2个通道级联，会产生巨大的计算开销，同时淹没$\boldsymbol{M}^{(\mathrm{Res}\text{-}i)}$中的信号。

本文通过3×3的卷积层作用在$\boldsymbol{F}^{(\mathrm{Res}\text{-}i)}$上以减少其通道数$k_{\boldsymbol{F}}^{(\mathrm{Res}\text{-}i)}$，但是保持空间分辨率不变则产生跳跃特征$\boldsymbol{S}^{(\mathrm{Res}\text{-}i)}$，$k_{\boldsymbol{S}}^{(\mathrm{Res}\text{-}i)}$为$\boldsymbol{S}^{(\mathrm{Res}\text{-}i)}$的通道数且远远小于$k_{\boldsymbol{F}}^{(\mathrm{Res}\text{-}i)}$。将$\boldsymbol{S}^{(\mathrm{Res}\text{-}i)}$和$\boldsymbol{M}^{(\mathrm{Res}\text{-}i)}$直接在通道维度级联起来，生成$k_{\boldsymbol{S}}^{(\mathrm{Res}\text{-}i)}+k_{\boldsymbol{M}}^{(\mathrm{Res}\text{-}i)}$通道的特征图。然后，通过2.1.2节中空间域和通道域融合的注意力机制的特征修正模块对$(\boldsymbol{S}^{(\mathrm{Res}\text{-}i)},\boldsymbol{M}^{(\mathrm{Res}\text{-}i)})$修正得到$\boldsymbol{M}_{\mathrm{M}}^{(\mathrm{Res}\text{-}i)}$。通过一个3×3的卷积层，再经过双线性上采样得到上采样优化模块的最终输出$\boldsymbol{M}_{(\mathrm{M\text{-}up})}^{(\mathrm{Res}\text{-}i)}$，其通道数由卷积层中卷积核的个数决定。

上采样优化模块的操作只有卷积、非线性激活函数ReLU、双线性上采样和级联，因此，完全是可以反向传播的。其中，比较重要的是$k_{\boldsymbol{S}}^{(\mathrm{Res}\text{-}i)}$和$k_{\boldsymbol{M}}^{(\mathrm{Res}\text{-}i)}$既不能过大使得计算量繁重，也不能过小(会导致信息丢失)，所以在初始化时可以采用一个较大的值，随着空间分辨率越来越大，逐渐减小通道数目。显然，上采样优化模块的数目应该等于前向网络中池化层的数目。

2.1.2 特征修正网络模块

为在卷积的运算过程中实现将通道和空间信息融合在一起计算以获取信息特征。因此，本文改进设计了通道和空间注意力模块，目的是为使每个分支结构都可以有效地学习到在通道和空间域上对目标判别有意义的关键信息。基于注意力的特征修正模块通过学习之后可以有效地强调或抑制网络中间特征层来帮助信息在网络中传播。在特征修正模块中，空间自注意力可以看作是特征张量的自激励响应，其不使用最大值和均值池化，目的是初步筛选出特征张量中某一通道的异常较大值。而空间注意力的作用对象是经过通道注意力在通道维度上进行元素筛选之后的张量，所以可以使用最大值和均值池化在空间位置上对特征进行细粒度筛选。因此，在本文的特征修正模块中采用初步特征筛选(空间自注意力)→粗粒度特征筛选(通道注意力)→细粒度特征筛选(空间注意力)的顺序来构建特征修正模块。

如图4所示，特征修正模块具有3个连续的子模块：空间自注意力、通道注意力和空间注意力。给定卷积网络的一个中间特征图，通过特征修正模块可以进行自适应调整，最终得到对目标分类

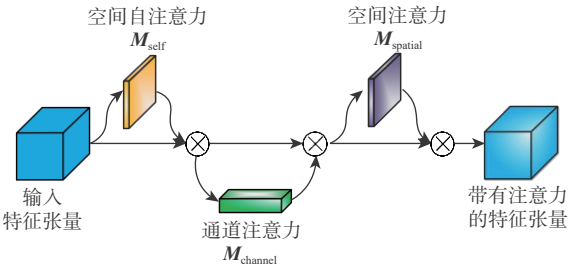


图 4　深度残差模块中加入特征修正模块的结构

Fig. 4　Structure of adding feature correction module to depth residual module

判别有积极意义的修正卷积特征。给定一个中间特征图$\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$作为输入，修正特征模块依次推导二维自注意图$\boldsymbol{M}_{\text{self}}\in\mathbf{R}^{1\times H\times W}$、一维通道注意图$\boldsymbol{M}_{\text{channel}}\in\mathbf{R}^{C\times 1\times 1}$和二维空间注意图$\boldsymbol{M}_{\text{spatial}}\in\mathbf{R}^{1\times H\times W}$。总体注意力卷积特征输出的$\boldsymbol{F}'''$获取过程可以概括为

$$\boldsymbol{F}'=\boldsymbol{M}_{\text{self}}(\boldsymbol{F})\otimes\boldsymbol{F} \tag{1}$$

$$\boldsymbol{F}''=\boldsymbol{M}_{\text{channel}}(\boldsymbol{F}')\otimes\boldsymbol{F}' \tag{2}$$

$$\boldsymbol{F}'''=\boldsymbol{M}_{\text{spatial}}(\boldsymbol{F}'')\otimes\boldsymbol{F}'' \tag{3}$$

式中：⊗表示按元素相乘。在乘法过程中，注意力权重会相应地传播：通道注意力权重是沿空间维度传播的，反之亦然。

因为特征修正模块不改变输入输出大小的同时具有很少的参数量，因此，本文的特征修正模块可以在现有的主流卷积网络中无缝插入。最新的研究表明，基于 ResNet 的卷积网络无论是在目标分类还是在目标跟踪的提取特征过程中都取得了较好的成绩，因此，本文在 ResNet 的残差模块中加入本文自主改进设计的特征修正模块构建带有注意力的目标分类网络，特征修正模块一方面可以提高目标分类网络的准确性，另一方面可以根据修正模块中的通道注意力权重参数和空间注意力权重参数为目标跟踪选择对目标判别具有正相关作用的深度卷积特征。

2.1.3　计算复杂度分析

本文提出的模型包括卷积、激活函数、上采样、级联及注意力机制。下面将对计算复杂度逐项分析。

对于初始卷积层，使用 3×3 卷积核，输入通道数$k_{\boldsymbol{F}}^{(\text{Res-}i)}$，输出通道数$k_{\boldsymbol{S}}^{(\text{Res-}i)}$。空间分辨率保持不变，则该卷积层的复杂度$C_1$为

$$C_1=O\left(k_{\boldsymbol{F}}^{(\text{Res-}i)}\times k_{\boldsymbol{S}}^{(\text{Res-}i)}\times 3\times 3\times H\times W\right) \tag{4}$$

式中：H为图像高度；W为图像宽度。

级联主要涉及通道数的叠加，计算复杂度可以忽略不计，主要表现为内存方面的操作。

注意力机制模块通过对$(\boldsymbol{S}^{(\text{Res-}i)},\boldsymbol{M}^{(\text{Res-}i)})$修正得到$\boldsymbol{M}_{\text{M}}^{(\text{Res-}i)}$，在空间自注意力，通道注意力和空间注意力的处理中对特征的尺寸大小并没有改变，只是通道维度的运算，故复杂度C_2为

$$C_2=O\left(\left(k_{\boldsymbol{F}}^{(\text{Res-}i)}\times k_{\boldsymbol{S}}^{(\text{Res-}i)}\right)\times k\times H\times W\right) \tag{5}$$

式中：k为注意力模块内部通道数。

之后再次使用 3×3 卷积核，激活函数和双线性上采样操作。又因为双线性上采样复杂度可以忽略不计，因此，该过程复杂度C_3为

$$C_3=O\left(k_{\boldsymbol{S}}^{(\text{Res-}i)}\times k'\times 3\times 3\times H\times W\right) \tag{6}$$

式中：k'为卷积后的通道数。

综上分析得到总体的计算复杂度C_{all}为

$$C_{\text{all}}=C_1+C_2+C_3 \tag{7}$$

与其他模型的对比试验分析，本文将在 3.3 节展开详细阐述。

2.2　基于图像语义分割优化目标尺度

目标的尺度信息是目标表示的关键信息之一，在目标中心点定位准确的基础上，如果目标的尺度计算存在偏差会导致目标特征信息提取不准确，如果尺度过大会带入更多背景噪声，如果尺度过小则会导致目标关键信息丢失。从 VOT2015[22] 开始首创性地提出抛弃传统地使用与坐标轴对齐的矩形框转而使用旋转的矩形框标定目标。VOT2016[23] 则提出使用目标的图像语义分割信息生成目标的旋转标定框，其指出包含目标的最小旋转矩形框实际是目标语义信息的一种最大近似，因此，可以对跟踪器的输出利用图像意义分割算法进行优化。本文可以通过一个像素极端到端的语义分割网络把目标的语义信息完整地从视野中提取出来，然后，根据目标的语义信息来确定目标的旋转矩形尺度和非旋转矩形尺度。同时，根据目标掩码生成矩形框和旋转矩形框对目标的深度特征进行背景降噪处理，提高目标跟踪判别器的判断能力。

2.2.1　目标矩形尺度的标定

目标矩形尺度的生成：典型的跟踪基准测试需要边界框作为目标对象的最终表示形式。本文考虑了 3 种（轴对齐的边界矩形 (minimum bounding box, Min-max)、旋转的最小边界矩形 (MBR)、VOT2016 中自动边界框生成优化策略（Opt））不同的策略从二进制掩码生成边界框，如图 5 所示。

现有的跟踪算法通常以固定或可变长宽比来预测与轴对齐的边界框。为测试产生的每帧二进制掩码可以改善跟踪效果并专注于表示精度。对于本实验，随机采样视频帧。实验中描述的算法在 VOT2016 序列中的随机裁剪搜索补丁 (随机偏移在±16 像素以内，缩放变形最大为 21±0.25 像素) 上进行了测试。



图 5 3 种从二进制掩码生成边界框的算法对比

Fig. 5 Comparison of three algorithms for generating bounding boxes from binary masks

如表 1 所示，本文比较了图 1 显示的基于高效卷积算子（efficient convolutional operators, ECO）和图像语义分割网络使用 Min-max、MBR 和 Opt 这 3 种矩形框标定算法的平均准确率。表中，mIOU 表示平均交并比，mAP@0.5 表示检测阈值为 0.5 的情况下检测精度均值，检测结果对比方式采用交并比，mAP@0.7 表示检测阈值为 0.7 的情况下检测精度均值，检测结果对比方式采用交并比。作为参考，对比了 SiamFC 和 SiamRPN 的结果，分别代表了固定长宽比算法和可变长宽比算法。SiamMask 作为利用图像语义分割算法也参与对比。

表 1 中无论使用哪种矩形框标定策略，本文算法都能实现最佳 mIOU。尽管 ECO-Opt 提供了最高 mIOU 和 mAP，但由于其缓慢的优化过程而需要大量的计算资源[24-26]。ECO-MBR 的 mAP@0.5 为 88.31%，优于现有的基于孪生网络结构的深度学习跟踪算法，本文使用 MBR 根据目标语义分割掩码获得目标边界矩形框。对于目标与轴对齐的边界框本文使用 Min-max 策略。

表 1 VOT-2016 上不同包围矩形框标定策略的性能

Table 1 Performance of different enclosing rectangular box calibration strategies on VOT-2016

算法	mIOU/%	mAP@0.5/%	mAP@0.7/%
SiamFC[6]	50.48	56.42	9.28
SiamRPN[8]	60.02	76.20	32.47
SiamMask-Min-max	65.05	82.99	43.09
SiamMask-MBR	67.15	85.42	50.86
SiamMask-Opt	71.68	90.77	60.47
ECO-Min-max	69.33	85.14	50.23
ECO-MBR	72.57	88.31	56.74
ECO-Opt	74.97	91.85	68.56

2.2.2 基于图像语义信息对目标进行背景去噪

基于相关滤波器的跟踪算法在提取目标的特征之后往往需要对图像中的背景进行去噪。现有的相关滤波器算法大都采用余弦窗对目标进行背景去噪，这种算法存在以下 2 个缺点：

1）过度依赖跟踪结果的精度，余弦窗是以目标中心为最高点生成二维高斯矩阵，如果跟踪结果中心与真实目标中心存在偏差，则生成的余弦窗将会保留背景中的噪声并忽略真正目标特征；

2）余弦窗本身是一个服从二维正态分布的高斯矩阵，其不可避免地过度强调目标中心的特征而忽略了目标距离中心较远的特征。如图 6(a) 所示，对于一只鸟而言，最具代表性的特征应该在头部而不是其身体中心部分。

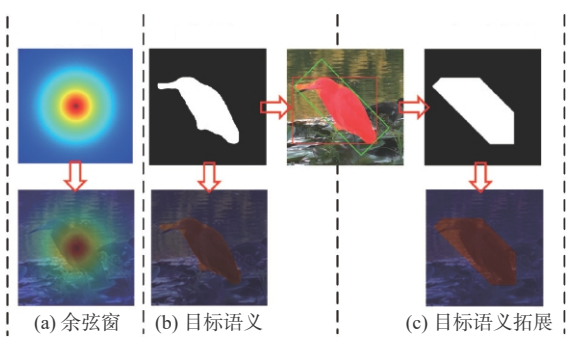


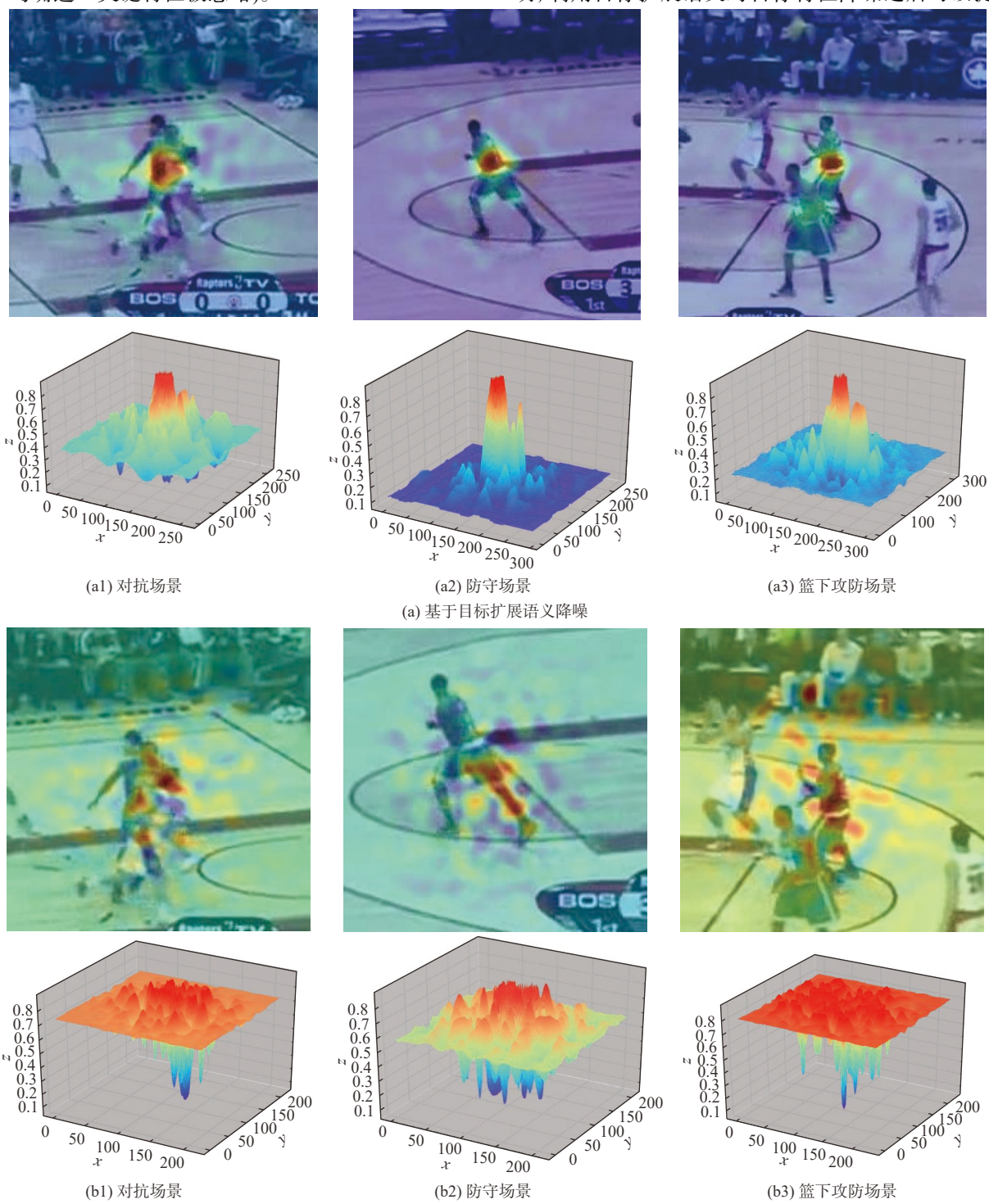
图 6 3 种掩码对目标背景去噪的示意图

Fig. 6 Schematic diagram of three masks for denoising target background in target images

在目标跟踪过程中, 相邻图像之间目标的形变几乎可以忽略不计, 因此, 可以利用上一帧中目标跟踪结果的图像语义信息对新的图像帧中的目标特征进行背景降噪。

如图 6(b) 所示, 由于图像语义分割网络本身存在一定的误差, 如果直接使用图 6(b) 中目标语义, 由于其自身的误差会导致目标的部分信息丢失 (图 6(b) 中鸟嘴这一关键特征被忽略)。

为避免由于语义分割误差导致的目标信息丢失问题, 本文基于目标的 MBR 旋转框和 Min-max 与坐标轴对齐的矩形框的交集作为目标的扩展语义信息对目标的特征进行降噪处理。如图 7 所示, 图中, x、y 分别为循环矩阵横纵坐标, z 为目标响应值。图 7(a) 显示了利用本文目标扩展语义替换 ECO[27] 中的余弦窗 (图 7 (d)) 对 ECO 的优化结果。实验表明, 利用目标扩展语义对目标特征降噪之后可以提



(a) 基于目标扩展语义降噪

(b) 基于余弦窗降噪

图 7　语义扩展和利用余弦窗对目标特征降噪后相关滤波跟踪算法 ECO 判别器的输出响应

Fig. 7　Semantic expansion and output response of ECO discriminator of correlation filter tracking algorithm after noise reduction of target features using a cosine window

高判别器的判别能力，ECO 的输出响应更尖锐。

3 实验结果与分析

本文主要基于 2.2 节中算法将 2.1 节中的目标语义分割模型对现有的相关滤波器算法进行优化，并且在具有代表性的公开数据集 OTB100[28]，VOT2016 和 VOT2018[29] 上进行实验对比分析，以验证本文中目标运动模型在解决目标跟踪过程中目标的尺度标定的优良表现，从而提高目标跟踪的准确性和鲁棒性。尤其在无人驾驶领域，良好的行人跟踪[30] 为行进的安全性与准确性提供了巨大保障，同时也为轨迹预测[31-32] 奠定了基础。

本文使用图像语义分割算法对基于相关滤波器的算法 ECO、DeepSRDCF[33] 进行优化，得到 ECO-S、DeepSRDCF-S。利用本文提出的图像语义分割算法对基于孪生网络结构的深度学习算法 SiamRPN 的输出进行优化得到 SiamRPN-S。并将这些算法与最新的深度学习算法 SiamMask 在公开数据集 OTB 和 VOT 上进行实验对比。

3.1 OTB100 数据集上的实验对比

图 8 中显示了现有基于相关滤波器的高精度跟踪算法 ECO 和 DeepSRDCF，以及在其基础上加入本文目标语义分割模型之后，相应的优化算法 (ECO-S、DeepSRDCF-S) 和最新的跟踪算法 SiamMask 在 OTB 数据集 100 个视频序列上的成功率和准确率，图中，图例数值为成功率均值。图像语义分割算法使得相关滤波器的性能得到较大提升，尤其是 ECO 算法无论是在成功率还是在准确率方面都优于现有的最新 SiamMask 算法。在对 SiamRPN 跟踪器的输出进行优化之后，SiamRPN-S 的性能也略有提升，这也证明了目标的尺度优化对于目标的跟踪结果的优化至关重要。

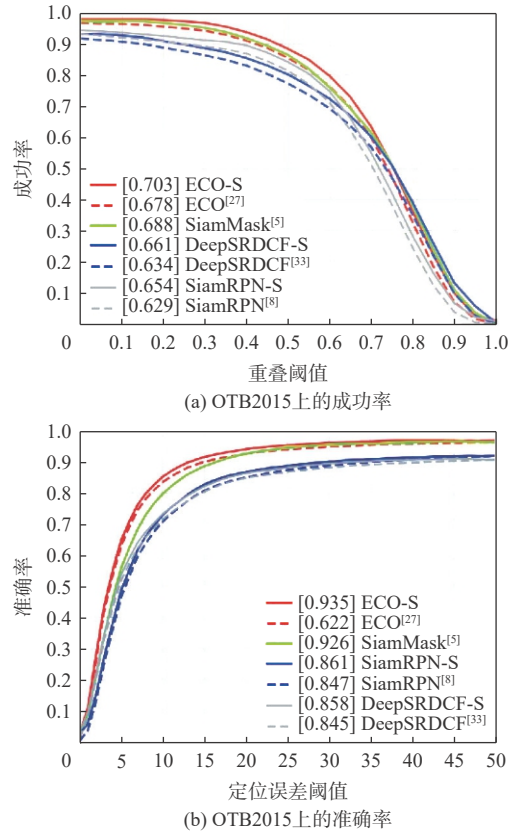


图 8 OTB2015 基准上的成功率和准确率

Fig. 8 Success rate and accuracy on OTB2015 benchmarks

图 9 为在 2 个视频序列上，ECO-S 和 SiamMask 定性跟踪结果比较。2 个视频序列有一个共同的特



图 9 ECO-S 和 SiamMask 在 OTB100 数据集 2 个视频序列上的定性结果

Fig. 9 Characterization results of ECO-S and SiamMask on two video sequences of OTB100

点: 场景中存在与目标具有相似视觉特征的同类物体, 在这类场景中相关滤波器算法 ECO 在目标被相似物体遮挡或消失之后跟踪器很容易将相似物体定位新的目标或丢失目标。经过本文的图像语义算法优化之后, ECO-S 增加了目标尺度标注的准确性, 同时经过目标扩展语义信息的背景降噪有效的聚焦目标的关键特征, 所以在图 9 中的鸟消失和人在被遮挡之后能稳定地定位到正确的目标。基于孪生网络的深度学习算法 SiamMask 虽然有目标语义信息的输出, 但是仍然存在孪生网络结构算法目标漂移问题。

3.2　VOT2016 和 VOT2018 数据集上的实验对比

视觉对象跟踪数据集 VOT 是包含 60 个序列的短期视觉跟踪算法之间的竞争。对于数据集中的每个序列, 通过使用第 1 帧中目标的矩形进行初始化来评估跟踪器。只要目标丢失, 该工具包就会重新启动跟踪器, 并计算平均故障数与统计平均覆盖率来描述跟踪器的鲁棒性和准确性。表 2 为 VOT2016 和 VOT2018 数据集的准确率, 鲁棒性及平均期望重叠 (expected average overlap, EAO)。可以发现, 目标语义分割模型可以同时提高目标跟踪的准确性和鲁棒性, 并且在预期的平均重叠方面也对目标跟踪器保持稳定的增益, 以 VOT2018 数据集为例, ECO-S 的 EAO 为 0.408 3, 而 ECO 的 EAO 为 0.307 7。图 10 和图 11 分别为加入目标运动模型之后的跟踪算法及其基础算法在 VOT2016 和 VOT2018 基准上的准确性、鲁棒性、EAO、成功率、EAO 曲线。由图 10 和图 11 可知, 目标运动模型在提高跟踪器的鲁棒性的同时对准确率也有提高, 对 EAO 的增强也非常明显。尤其值得注意的是基于相关滤波器的跟踪算法 ECO 在经过目标语义分割模型的优化后在 VOT2016 和 VOT2018 数据集上都优于 SiamMask。同时, 基于孪生网络的深度学习算法 SiamRPN 在加入目标运动模型之后, 其天然的低鲁棒性缺陷得到解决, 减少了目标跟踪过程中的漂移现象。

图 12 为在 VOT 数据集 2 个视频序列上, 加入目标语义分割模型之后的 ECO-S 模型与最新的深度学习算法 SiamMask 相比较的可视化跟踪结果。在前一个序列中, 由于跟踪目标 (橘色的鱼) 与背景具有十分相似的颜色特征, 这为跟踪带来了巨大挑战, 但是 ECO-S 可以一直准确地定位目标, 相反 SiamMask 则会因为鱼的形变丢失目标, 可见, 目标语义分割模型对目标特征背景去噪这一表现在跟

表 2　VOT2016 和 VOT2018 视频序列基准上的性能比较

Table 2　Performance comparison on VOT2016 and VOT2018 video sequence benchmarks

算法	准确率		鲁棒性		EAO	
	VOT2016	VOT2018	VOT2016	VOT2018	VOT2016	VOT2018
ECO-S	0.529 5	0.540 5	16.581 7	7.239 8	0.329 3	0.408 3
ECO[27]	0.484 7	0.497 8	15.043 7	13.511 2	0.308 9	0.307 7
SiamMask[5]	0.547 0	0.533 7	17.939 3	7.441 0	0.325 1	0.401 6
SiamRPN-S	0.570 2	0.551 5	19.272 0	10.255 1	0.320 6	0.372 7
SiamRPN[8]	0.538 6	0.539 9	21.081 7	14.204 0	0.257 9	0.317 7
DeepSRDCF-S	0.507 2	0.526 3	19.543 8	17.469 5	0.277 3	0.279 9
DeepSRDCF[33]	0.522 0	0.501 6	20.346 2	23.964 4	0.275 6	0.228 2

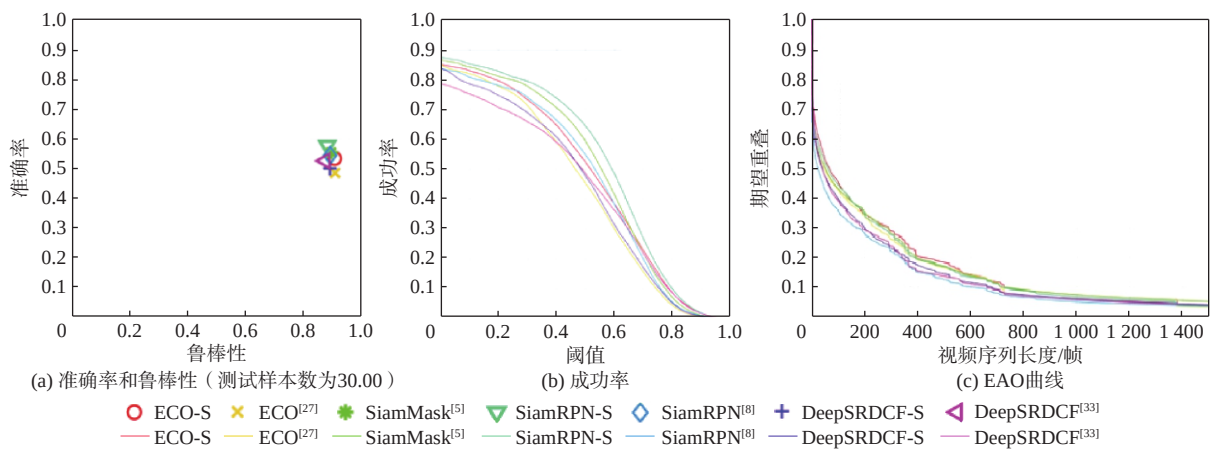


图 10　VOT2016 基准上的实验对比

Fig. 10　Comparison of experiments on VOT2016 benchmark

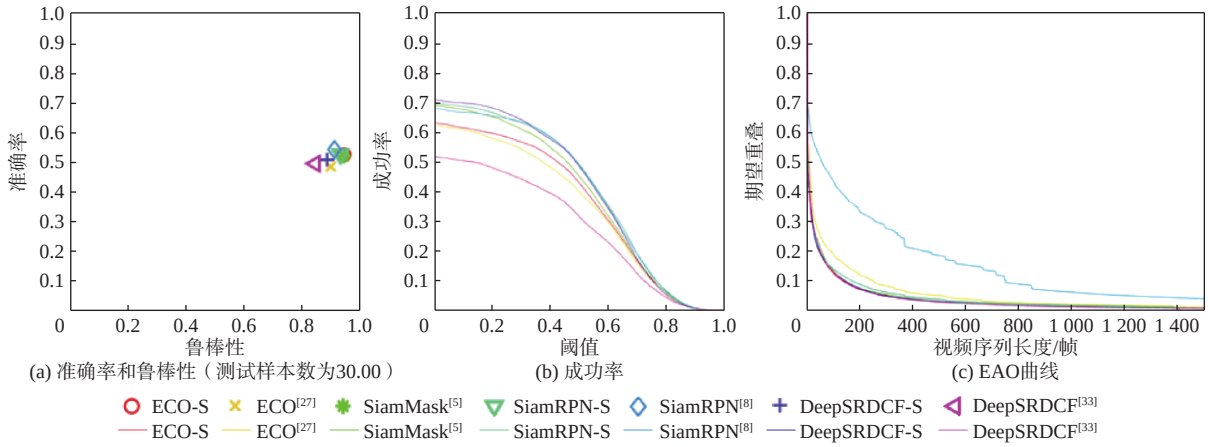


图 11 VOT2018 基准上的实验对比

Fig. 11 Comparison of experiments on VOT2016 benchmark

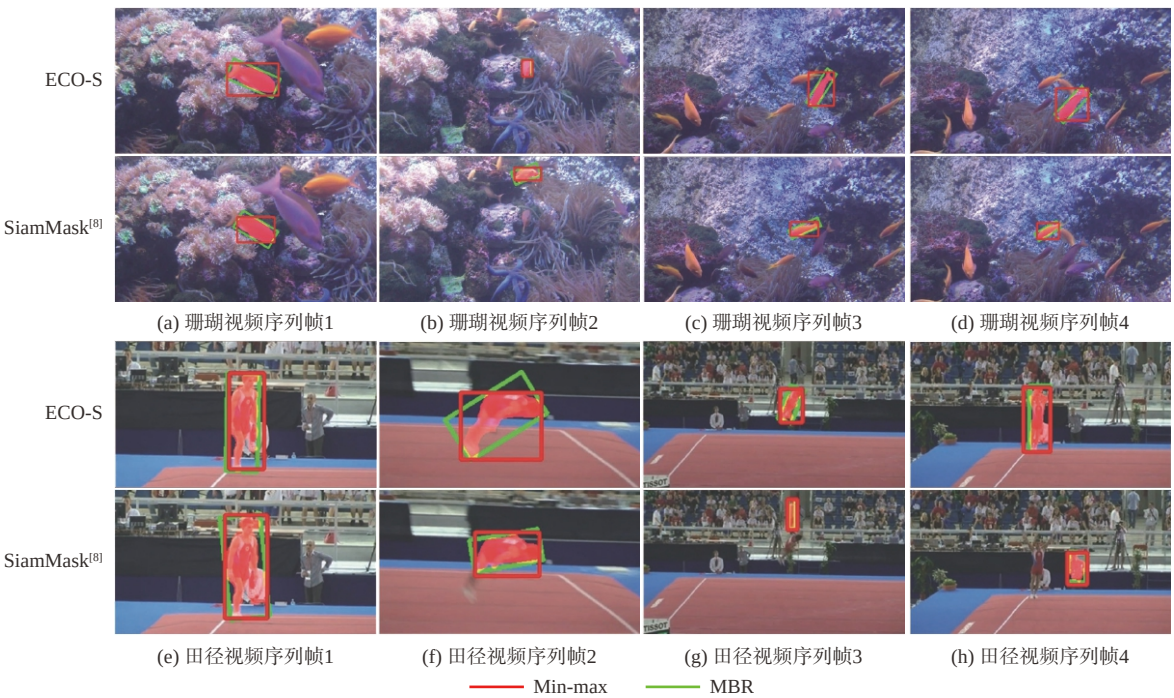


图 12 ECO-S 和 SiamMask 在 VOT 数据集 2 个视频序列上的定性结果

Fig. 12 Qualitative results of ECO-S and SiamMask on two video sequences of VOT

踪器性能的提升上具有关键作用。在第 2 个序列中面对运动员的剧烈运动，ECO-S 不仅可以一直定位目标的位置，同时还能保持目标的最优尺度输出，而 SiamMask 则将目标漂移到别的行人身上。由此可见，目标语义分割模型对目标尺寸表示更加贴合实际，更加符合人类对于目标跟踪的感官表现。

3.3 计算复杂度对比分析

3.1 节和 3.2 节实验呈现了目标语义分割模型对现有模型优化的良好效果。但是理想实验结果的代价是否建立在高昂的计算开销基础之上是本文进一步关注和验证的。本文通过参数量和浮点运算速度这 2 个指标进行对比实验。其中，浮点运算速度为每秒所执行的浮点运算次数。其常被用来估算电脑的执行效能，尤其是在使用到大量浮点运算的科学计算领域中。

由表 3 可知，ECO-S 的参数量相对于 ECO 仅仅增长 0.58×10^6，SiamRPN-S 和 DeepSRDCF-S 分别增长 0.52×10^6 和 0.66×10^6。而在浮点运算速度指标中，改进模型和原模型的差别在 10^9 的数量级下依旧在 0.1 以内。可见，本文设计的目标语义分割模型在增强准确率，鲁棒性的同时并没有以大幅度增长的计算开销作为代价。这意味着本文算法的功能强大，显示出该方法可以生成更丰富的描述特征，同时可以实现在各种卷积网络中的轻量级即插即用。

表 3　OBT100、VOT-2016、VOT-2018 上的计算开销比较

Table 3　Comparison of computational overhead on OBT100, VOT-2016, VOT-2018

算法	参数量	浮点运算速度/10^9 s^{-1}		
		OBT-100	VOT-2016	VOT-2018
ECO-S	5.92×10^6	0.39	0.46	0.51
ECO[27]	5.34×10^6	0.40	0.46	0.50
SiamMask[5]	12.96×10^6	0.86	0.97	1.07
SiamRPN-S	11.13×10^6	0.81	0.89	0.91
SiamRPN[8]	10.61×10^6	0.75	0.81	0.85
DeepSRDCF-S	12.68×10^6	0.86	0.96	1.07
DeepSRDCF[33]	12.02×10^6	0.85	0.96	1.06

4　结　论

1）基于图像语义分割的尺度优化策略能够提升目标跟踪的整体性能。在 OTB100、VOT2016 和 VOT2018 数据集上，本文算法在准确率、鲁棒性和 EAO 等指标上均优于对应的基础算法。以 VOT2018 数据集为例，ECO-S 的 EAO 为 0.408 3，而 ECO 的 EAO 为 0.307 7。

2）基于目标语义掩码生成的旋转 MBR 用于尺度标定，在多项评价指标上优于传统轴对齐边界框算法。在 VOT2016 数据集上，采用 MBR 的算法在 mIOU、mAP@0.5 和 mAP@0.7 指标上均取得更高数值。

3）利用目标语义信息对跟踪器输入特征进行背景去噪，有助于提高判别器输出的稳定性。定性实验结果表明，在目标遮挡及相似目标干扰场景中，本文算法能够保持连续、稳定的跟踪结果。

4）在计算复杂度方面，本文算法在参数量和计算开销小幅增加的情况下实现了性能提升。

参考文献（References）

[1] DING W Z, XU Q J, LIU S Y, et al. SAMF: a self-adaptive protein modeling framework[J]. Bioinformatics, 2021, 37(22): 4075-4082.

[2] TANG L N, XU X C, WANG X, et al. An automatic object attention likelihood map correlation filter for visual tracking[C]//Proceedings of the IEEE 11th Asia-Pacific Conference on Antennas and Propagation. Piscataway: IEEE Press, 2024: 1-2.

[3] LI B L, WANG Y, XU Y M, et al. DSST: a dual student model guided student–teacher framework for semi-supervised medical image segmentation[J]. Biomedical Signal Processing and Control, 2024, 90: 105890.

[4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 580-587.

[5] HU W M, WANG Q, ZHANG L, et al. SiamMask: a framework for fast online object tracking and segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(3): 3072-3089.

[6] LEE S, KIM S. Comparative analysis of IR-IR image matching applying the deep learning-based template matching techniques[C]//Proceedings of the 23rd International Conference on Control, Automation and Systems. Piscataway: IEEE Press, 2023: 441-444.

[7] HELD D, THRUN S, SAVARESE S. Learning to track at 100 FPS with deep regression networks[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2016: 749-765.

[8] WANG S Q, QIAN K, SHEN J L, et al. AD-SiamRPN: anti-deformation object tracking *via* an improved siamese region proposal network on hyperspectral videos[J]. Remote Sensing, 2023, 15(7): 1731.

[9] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[10] HARIHARAN B, ARBELÁEZ P, GIRSHICK R, et al. Simultaneous detection and segmentation[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2014: 297-312.

[11] PINHEIRO P O, COLLOBERT R, DOLLÁR P. Learning to segment object candidates[C]//Proceedings of the 29th International Conference on Neural Information Processing Systems. New York: ACM, 2015(2): 1990-1998.

[12] DAI J F, HE K M, SUN J. Instance-aware semantic segmentation *via* multi-task network cascades[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 3150-3158.

[13] HARIHARAN B, ARBELÁEZ P, GIRSHICK R, et al. Hypercolumns for object segmentation and fine-grained localization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 447-456.

[14] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2024-02-01]. https://arxiv.org/abs/1409.1556.

[16] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 1-9.

[17] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

[18] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE Press, 2016: 640-651.

[19] XIE S N, TU Z W. Holistically-nested edge detection[J]. International Journal of Computer Vision, 2017, 125(1): 3-18.

[20] SERMANET P, KAVUKCUOGLU K, CHINTALA S, et al. Pedestrian detection with unsupervised multi-stage feature learning [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2013: 3626-3633.

[21] KOONCE B. ResNet 50[C]//Proceedings of the Convolutional

Neural Networks with Swift for Tensorflow. Berkeley: Apress, 2021: 63-72.

[22] HAN F, JIANG S K, WU J M, et al. Real-time object tracking in the wild with siamese network[J]. Multimedia Tools and Applications, 2023, 82(16): 24327-24343.

[23] CAO D, DAI R H, WANG J, et al. Fast visual tracking with squeeze and excitation region proposal network[J]. Human-centric Computing and Information Sciences, 2023, 13(7): 20.

[24] WANG Q, GAO J, XING J L, et al. Dcfnet: discriminant correlation filters network for visual tracking[EB/OL]. (2017-04-13)[2024-02-01]. https://arxiv.org/abs/1704.04057.

[25] ZHANG H Y, LIU G X, ZHANG Y, et al. Robust multi-model visual tracking with distractor-aware template-coupled correlation filters joint learning[J]. IEEE Transactions on Multimedia, 2024, 26: 1813-1828.

[26] FANG H, LIAO G S, LIU Y J, et al. Shadow-assisted moving target tracking based on multidiscriminant correlation filters network in video SAR[J]. IEEE Geoscience and Remote Sensing Letters, 2023, 20: 4006205.

[27] DANELLJAN M, BHAT G, SHAHBAZ KHAN F, et al. ECO: efficient convolution operators for tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6638-6646.

[28] RAHMAN M M. Target focused shallow transformer framework for efficient visual tracking[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(21): 23409-23410.

[29] FU J Y, LIANG Q F, XIE Q S, et al. Object tracking based on foreground adaptive bounding box and motion state redetection[C]//Proceedings of the 3rd International Conference on Artificial Intelligence and Computer Engineering. Bellingham: SPIE, 2023: 155.

[30] 黄煜杰, 陈凯, 王子源, 等. 多目视觉下基于融合特征的密集行人跟踪方法[J]. 北京航空航天大学学报, 2025, 51(7): 2513-2525.
HUANG Y J, CHEN K, WANG Z Y, et al. A dense pedestrian tracking method based on fusion features under multi-vision[J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(7): 2513-2525(in Chinese).

[31] CHEN K, SONG X, YUAN H T, et al. Fully convolutional encoder-decoder with an attention mechanism for practical pedestrian trajectory prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11): 20046-20060.

[32] CHEN K, ZHU H H, TANG D B, et al. Future pedestrian location prediction in first-person videos for autonomous vehicles and social robots[J]. Image and Vision Computing, 2023, 134: 104671.

[33] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Convolutional features for correlation filter based visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE Press, 2015: 58-66.

Adaptive algorithm for target tracking scale based on image semantic segmentation

CHEN Kai[1], ZHAO Xiaodong[1,*], HUANG Yujie[1], WANG Pengfei[2], LEI Yichen[1], ZHANG Biao[1]

(1. College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;
2. China Electronics Technology Group 28th Research Institute, Nanjing 210007, China)

Abstract: In order to optimize the scale of the tracker's output, this paper focuses on how to fully utilize the semantic information of the target obtained by image semantic segmentation in complex scenes. It also designs an image semantic segmentation network based on the optimization of the attention mechanism to optimize the target tracker's output and the input of the features, which can realize plug-and-play for various algorithms. The image semantic segmentation mask is used to obtain the rotating frame boundary of the target, and the denoising optimization of the features in the input phase of the target is carried out according to the rotating and non-rotating frame boundaries of the target to attenuate the influence of the background noise on the discriminator of the tracker. The structure of the designed network, training, calibration of the target's rotating frame, and denoising of the tracker's input features are discussed, respectively. The correctness of the target motion model in resolving the scale calibration of the target during target tracking is verified through experimental comparison analysis on public datasets OTB100, VOT2016 and VOT2018. This enhances the accuracy and resilience of target tracking.

Keywords: target tracking; semantic segmentation; attention mechanisms; scale optimization; background denoising

Received: 2024-04-07; **Accepted**: 2024-05-24; **Published Online**: 2024-06-08 10:11

URL: link.cnki.net/urlid/11.2625.V.20240607.1715.003

Foundation items: National Natural Science Foundation of China (52202417); China Postdoctoral Science Foundation (2022TQ0155,2022M721605); Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (VRLAB2023A02); Young Elite Scientists Sponsorship Program by China Association for Science and Technology (2023QNRC001); Young Elite Scientists Sponsorship Program by Jiangsu Association for Science and Technology (JSTJ-2023-XH032)

* **Corresponding author**. E-mail: xdzhao@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0105

触爆载荷作用下的混凝土重力坝动态响应规律

张克钒[1,*]，彭永[2]，李翔宇[2]，卢芳云[2]，王哲[1]

(1. 国防科技大学 试验训练基地，西安 710000；2. 国防科技大学 理学院，长沙 410072)

摘 要：针对触爆载荷作用下混凝土重力坝的动态响应研究缺乏试验验证和规律性较弱、应用难度较大的问题，利用缩比试验验证后的仿真模型开展了不同爆点位置、不同当量大小共计69个工况的仿真模拟研究，考察了触爆载荷作用下重力坝的坝底帷幕振动速度、坝体加速度、受损坝体占比3个参量的变化时程曲线，初步归纳了坝体动态响应的规律。在此基础上，进一步提出能够将爆炸载荷参量和动态响应相关联的无量纲数 B 和 C，据此整合了所有仿真计算结果的参量变化信息，得到基于分析爆炸条件下重力坝动态响应的变化规律和趋势。结果表明：无量纲形式的坝体动态响应规律可有效地关联坝体物理毁伤和功能毁伤，以支撑重力坝易损性分析工作，同时可为其他工程类目标的毁伤机理研究提供参考。

关 键 词：混凝土重力坝；触爆载荷；动态响应；毁伤规律；无量纲数

中图分类号：V414.3^{+}2; O389

文献标志码：A **文章编号**：1001-5965(2026)05-1547-10

混凝土重力坝是大型的水工建筑物，在航运、灌溉、发电等行业和领域发挥着重要的政治、经济作用[1]，一旦失事将对下游人民生命财产造成不可估量的严重后果。恐怖袭击猖獗、空中弹药威胁日益增大的今天，重力坝逐渐成为不可忽略的重要战略目标。因此，开展重力坝在爆炸载荷下的相关研究对探究重力坝结构破坏模式的机理、重力坝目标的要地防护和毁伤评估、目标易损性分析方法等研究领域都具有较强的理论与工程实践意义。

目前，重力坝在爆炸载荷作用下的相关研究主要集中于3个方面：水下爆炸、侵彻爆炸和接触爆炸。由于重力坝自身尺寸及实验条件等因素的限制，数值模拟是大多数学者采用的主要方法和途径。

水下爆炸载荷下重力坝的相关研究工作开展时间最久，成果也相对丰富，涉及重力坝模态响应、应力应变变化、位移加速度变化、破坏模式分析等方面。Huang等[2]利用离心机和数值模拟相结合的方法研究了重力坝在近场水下爆炸载荷下的破坏，分析了坝体内部裂纹生成、扩展背后的潜在机制，气泡脉动对重力坝的破坏，不同起爆位置处的破坏模式对比。Wang等[3]开展了重力坝在水下爆炸的仿真计算，提出一种利用峰值速度和平均频率(peak velocity summation-mean frequency, PVS-MF)评估重力坝易损性的方法，其结果可用于水利工程的抗爆设计。此外，还开展了对地基岩性的研究，提出利用裂缝穿透率量化爆炸引起的裂缝扩展问题。在考虑了初始应力场的影响后[4]，给出更准确的重力坝水下爆炸受损的仿真计算结果。Li等[5]分别考虑了静水压力、爆点位置等因素对坝体受损结果的影响，并总结归纳出潜在威胁最大的破坏模式。Chen等[6]分析了不同间距、不同爆点位置下重力坝的不同破坏模式，并考虑了溢流隧道对重力坝整体受损结果的影响。Xu等[7]分析了重力坝在水下爆炸载荷下的破坏机理，并提出了用泡沫铝作

收稿日期：2024-02-28；**录用日期**：2024-05-09；**网络出版时间**：2024-05-31 08:05
网络出版地址：link.cnki.net/urlid/11.2625.V.20240530.1606.001
基金项目：国家自然科学基金 (12302450)
* **通信作者**. E-mail：kfzhang_v@163.com

引用格式：张克钒，彭永，李翔宇，等. 触爆载荷作用下的混凝土重力坝动态响应规律[J]. 北京航空航天大学学报，2026，52(5)：1547-1556. ZHANG K F，PENG Y，LI X Y，et al. Dynamic response law of concrete gravity dam under contact explosion load[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1547-1556 (in Chinese).

为抗爆措施的方法，其结果显著提升了重力坝的抗爆性能。Liu 等[8] 利用粒子群优化算法对水下爆炸载荷影响下的重力坝动力响应参数进行了反演，结果表明与实测值基本一致。值得注意的是，这些水中爆炸的相关研究大多以 2011 年 Vanadit-Ellis 等[9] 基于离心机进行的重力坝的缩比试验研究结果验证仿真计算结果，这从侧面反映了重力坝水下爆炸试验（或空中爆炸试验）的相关研究存在一定开展难度。

混凝土重力坝在接触爆炸和侵彻爆炸载荷下的相关研究主要侧重于整体毁伤结果的分析和归纳。Xu 等[10] 分别对重力坝上下游面的接触爆炸进行了仿真计算，将各类结构动力响应进行对比后归纳得到损伤程度最大的爆点位置。Zhang 等[11] 对侵彻爆炸和空爆情况下的大坝毁伤模式进行了数值仿真计算，分别总结分析了不同情况的爆炸载荷作用特点、坝体破坏效应、整体毁伤模式。李本平等对多种情况下侵彻弹打击重力坝目标的毁伤结果进行了仿真计算，对坝体的振速变化情况、应力云图分布情况、破坏范围情况进行了定量分析，总结了单弹打击[12]、连续打击条件下[13-14] 坝体的破坏效应分布和损伤模式种类。同样地，对于混凝土重力坝这类大型水利工程目标，基本没有开展过触爆载荷下的原尺寸坝体响应试验研究。

上述研究多数着眼于爆炸载荷作用下坝体特定参量的变化情况或特定条件的爆炸载荷对坝体完整性、稳定性的影响，尚未能够形成可支撑坝体毁伤结果判定的综合性、规律性结果，缺乏提升研究结果普适性的方式。

因此，本文考虑利用基于缩比模型试验验证的重力坝仿真计算模型[15]，开展不同当量、不同爆点、不同介质的触爆载荷作用下混凝土重力坝的动态响应规律研究。依据坝体失稳理论，考察坝体底部振动速度、坝体加速度、坝体受损体积占比 3 个参量的变化规律，在归纳和分析坝体动态响应规律的同时，尝试提出动态响应规律的无量纲形式，给出准确性良好、实用性较强的重力坝动态响应规律。

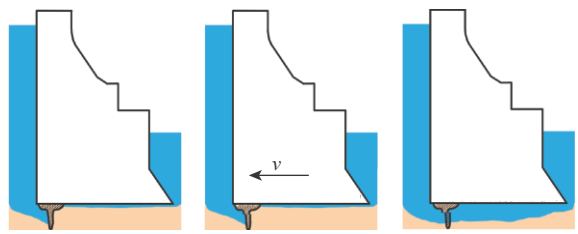
本文所得结果可用于量化、确定混凝土重力坝在接触爆炸载荷下的动态响应和破坏模式，同时可以为重力坝目标的易损性分析研究和毁伤评估工作提供参考。

1 重力坝动态响应的表征参量

根据重力坝的抗滑稳定分析理论[1]，坝体正常工作条件下，岩基面上所有法向合力带来的静摩擦力应大于使坝体具有翻转趋势的所有切向合力，此时坝体处于抗滑稳定状态。当高强度爆炸载荷以切向外力作用于坝体上时，会引发坝底帷幕振动速度、坝体加速度、坝体受损体积占比 3 个参量的变化，从而影响其抗滑稳定状态。

1.1 坝底帷幕振动速度

爆炸载荷作用于坝体后，使坝体发生了一定量的位移或振动，造成坝底部的防水帷幕断裂，进而引起坝体所受扬压力急剧增大，所有法向合力减小，最终影响坝体抗滑稳定状态，过程如图 1 所示。



(a) 正常状态　(b) 受到爆炸载荷后发生振动　(c) 帷幕断裂致使渗流量增大

图 1 帷幕断裂导致坝体失稳的原理示意图

Fig. 1 Diagram of principle of dam destabilization due to curtain fracture

坝底帷幕振动速度主要是指坝踵处帷幕位置的振动速度，拟通过仿真计算得到相应位置处节点的 y 方向振动速度作为帷幕振动速度 v（单位：cm/s）。

1.2 坝体加速度

爆炸载荷所产生的压力作用于坝体一侧，导致所有切向合力大于坝体在岩基面上的静摩擦力，导致坝体失稳。此时，爆炸载荷下坝体整体受到的切向力是坝体失稳的主要原因，可利用坝体加速度度量坝体动态响应（此处加速度为坝体等效整体加速度）。由于坝体上游侧通常存在静水压力，因此，只需考虑上游侧爆炸载荷对坝体加速度的影响。

坝体加速度是上游水中爆炸载荷作用到坝面后，坝体受到的向下游方向作用力的表征。而在仿真计算中，外力作用下的坝体整体加速度会由于坝体发生了弹性变形、塑性变形等力学响应而受到影响，通常来说会偏小，可认为外部载荷提供的能量一部分用于产生了变形能或转化成了弹性势能，故而得到的坝体加速度无法完全反映外部载荷使坝段整体受到的向下游方向的作用力。

因此，在计算坝体加速度时，本文采用将坝体视作刚体的方式减小其他能量耗散的影响，并对比了相同工况下（爆点位于坝体上游侧高程 170 m 处）刚体坝段和非刚体坝段的加速度峰值变化范围，结果如图 2 所示，其中，TNT 表示三硝基甲苯（trinitrotoluene）。

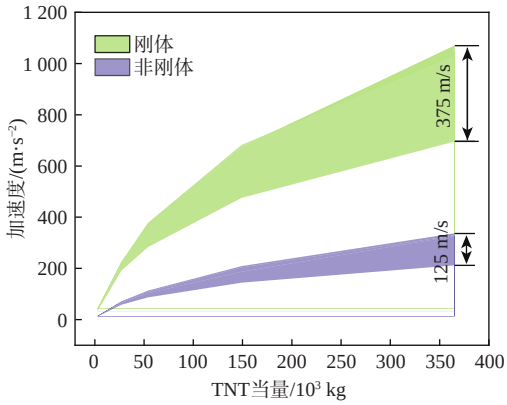


图 2　坝体刚体加速度和非刚体加速度的对比

Fig. 2　Comparison of rigid body acceleration and non-rigid body acceleration in dam structures

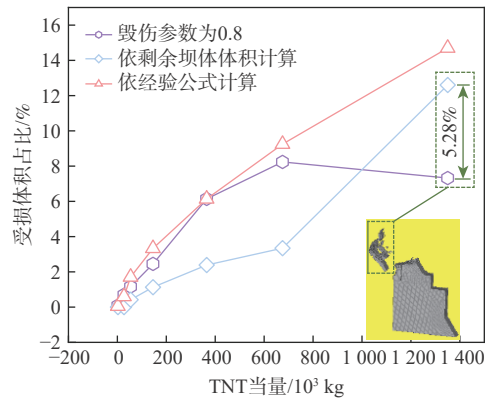


图 3　不同方法计算坝体受损体积的差异

Fig. 3　Difference of dam damage fraction calculated by different methods

由图 2 可知, 将坝体视为刚体后, 坝体加速度峰值普遍增大了约 2～4 倍, 不同工况之间的坝体加速度峰值之差从 125 m/s^2 增大到了 375 m/s^2。该方法有效地减小了其他形式能量损耗对坝体所受向下游方向作用力的影响, 使仿真计算得到的坝体加速度更贴近其实际意义。

1.3　坝体受损体积

爆炸载荷作用下坝体发生部分断裂、震塌后, 其自重和结构稳定性的变化也将影响坝体的稳定性。此时, 可利用坝体的受损体积占比作为毁伤准则[16-18]。

受损体积占比的具体数值无法直接从仿真计算中提取, 一般可通过分析损伤网格占比得出。损伤网格占比的计算方式利用了*MAT_RHT 材料模型中的 Damage Value 对网格进行损伤结果分类, 以超过某阈值的网格数与总网格数之比作为受损体积占比的具体数值。Damage Value(D_F)的阈值一般为 0.6～0.8, 其计算式为

$$D_F=\int_{\varepsilon_p^i}^{\varepsilon_p^f}\frac{\mathrm{d}\varepsilon_p}{\varepsilon_p^f} \tag{1}$$

式中: D_F 为坝体受损体积占比; ε_p 为塑性应变; i 和 f 分别表示硬化和失效。

此外, 在坝体受到大当量爆炸载荷作用下, 坝体顶部会产生较大范围的断裂和崩落, 这部分混凝土结构由于已经与剩余坝体发生整体脱离, 并未呈现明显应变, 因此, Damage Value 值较低。实际上, 这部分混凝土结构应纳入受损坝体的体积的计算范围。此时可通过计算剩余坝体体积的方式(即去除已脱落的坝体)对大当量工况下的受损体积占比进行修正, 以弥补 Damage Value 计算方法的局限性。不同方法计算坝体受损体积的区别如图 3 所示, 以坝顶位置处不同当量的工况为例, 不考虑崩落坝体时, 坝体受损体积占比会出现最大 5.28% 的误差。

2　坝体动态响应的仿真计算

本文针对爆炸载荷下混凝土重力坝的数值仿真缺乏试验验证的问题, 利用基于缩比试验验证的仿真模型开展重力坝动态响应的计算。文献 [15] 中, 通过开展静爆载荷作用下混凝土重力坝的缩比试验, 分析了 4 种不同缩放比尺寸的重力坝动态响应, 结合仿真计算结果建立了重力坝动态响应的相似律模型, 然后, 利用相似律规律分析、早期试验数据对比归纳出了原尺寸重力坝仿真模型的材料参数、网格尺寸和边界条件, 得到能够反映原尺寸重力坝动态响应的仿真模型。

2.1　仿真模型及工况

采用任意流固耦合(arbitrary lagrangian-eulerian, ALE)的方法计算混凝土重力坝在爆炸载荷下的动态响应, 整体仿真模型由空气、TNT、水、重力坝和地基组成, 仿真计算依托 LS-DYNA 完成, 各部分位置关系如图 4 所示, 图中, 1#～11#为 TNT 爆点。

空气、水使用*MAT_NULL 材料模型和*EOS_LINEAR_POLYNOMIAL 状态方程描述, 划分为欧拉网格(网格尺寸为 0.01 m), 并将网格边界全部设置为无反射边界。使用*MAT_HIGH_EXPLOSIVE_BURN 材料模型和*EOS_JWL 物态方程描述 TNT 炸药, 并划分欧拉网格(网格尺寸为 0.01 m)。混凝土重力坝材料模型为*MAT_RHT, 划分拉格朗日网格(网格尺寸为 0.01 m), 坝底设置位移约束。地基使用刚体材料模型*MAT_RIGID, 用于模拟地面, 划分拉格朗日网格。整体上利用*CONSTRAINED_LAGRANGE_IN_SOLID 和 *ALE_MULTI-MATERIAL_GROUP 关键字完成流固耦合的设置。具体模型参数如表 1～表 3 所示, 表中, A 为高压项压力系数, B_{EOS} 为中压项压力系, R_1 为高压项体积衰减指数, R_2 为中压项体积衰减指数, ω 为等熵指数, E_0 为爆轰释放的能量密度, A_1、A_2、A_3 为 Hugoniot 多项

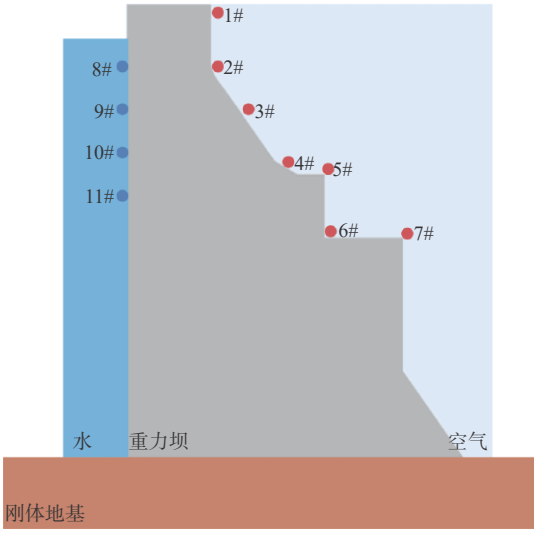


图4 仿真模型及爆点位置示意图

Fig. 4 Simulation model and location of blast point diagram

表1 仿真计算*MAT_HIGH_EXPLOSIVE_BURN模型参数

Table 1 Parameters in simulation computational *MAT_HIGH_EXPLOSIVE_BURN model

密度 ρ/(kg·m^{-3})	爆速 D/(m·s^{-1})	C-J压力 P_{cj}/GPa
1.67×10^3	7.1×10^3	21

表2 仿真计算*EOS_JWL模型参数

Table 2 Parameters in simulation computational *EOS_JWL model

A/GPa	B_{EOS}/GPa	R_1	R_2	w	E_0/(GPa·m^{-3})
370	3.55	4.15	0.95	0.35	7

表3 仿真计算*MAT_RHT模型参数

Table 3 Parameters in simulation computational *MAT_RHT model

密度ρ/(kg·m^{-3})	弹性剪切模量/Pa	侵蚀塑性应变	A_1/Pa
2.315×10^3	1.67×10^{13}	2.0	3.527×10^{10}
A_2/Pa	A_3/Pa	B_0	B_1
3.958×10^{10}	9.04×10^{10}	1.2	1.2
T_1/Pa	T_0/Pa	损伤参数D_1	损伤参数D_2
3.527×10^{10}	0	0.01	1.0
抗压强度 f_c/Pa	抗拉强度 f_t/f_c	抗剪强度f_s/f_c	弹性强度 f_t
4×10^7	0.1	0.18	0.7

式系数，B_0为多项式EOS中内能项的常数系数，B_1为多项式EOS中内能项的线性系数，T_1为压力体积关系的一次项系数，T_0为压力体积关系的二次项系数。

该模型中，坝顶高程为235 m，设计水位高程为225 m，最大坝高为180 m，坝顶宽为33 m，上游面为垂直面，下游面坡度为1∶0.7，坝段宽为22 m。坝体上游工作水位为170 m高程，下游为60 m高程。7个空中爆点位于坝体下游侧，由坝顶分布至下游水位上方的直角折坡处，基本覆盖了坝体下游侧所有的潜在打击位置。4个水中爆点等距分布于坝体上游侧的竖直面，基本覆盖了坝体上游侧潜在的触爆打击范围。

每个爆点分别设置了不同TNT当量爆炸载荷的工况，具体如表4所示。需要注意的是，虽然TNT当量递增的梯度较为平滑，但具体数值并不规整，这主要由于仿真计算模型的欧拉网格尺寸较大，无法通过网格选取或体积填充的方式得到整数的具体TNT当量数值。

表4 爆点对应具体工况明细

Table 4 Detailed specifications for critical points corresponding to specific operating conditions

爆点序号	爆点区域	爆点高程/m	触爆介质	TNT当量/10^3 kg
1#	下游侧	237	空气	2.7～1 350
2#	下游侧	224	空气	2.7～1 350
3#	下游侧	210	空气	2.7～1 350
4#	下游侧	190	空气	2.7～1 350
5#	下游侧	188	空气	2.7～1 350
6#	下游侧	166	空气	2.7～1 350
7#	下游侧	166	空气	2.7～1 350
8#	上游侧	230	水	2.7～365
9#	上游侧	220	水	2.7～365
10#	上游侧	210	水	2.7～365
11#	上游侧	200	水	2.7～365

触爆载荷工况与空爆载荷工况的关注点不同，空爆工况侧重于分析坝体的动力学响应(压力、加速度等)，因此，工况中所需TNT当量不需要很大；而触爆载荷工况更关注坝体的局部破坏情况，所以工况的设计以“可使得坝体产生明显破坏”为原则，在条件允许的范围内尽可能增大TNT当量。此外，考虑到水中爆炸载荷的衰减弱于空中爆炸，水中爆炸所致的坝体损伤程度明显高于空中爆炸，因此，上游侧各爆点的最大TNT当量均小于下游侧各爆点。

2.2 仿真计算结果

本节给出各个爆点位置处不同TNT当量造成的坝体毁伤结果和3个参量的变化规律曲面，坝体毁伤结果采用Damage Value值的云图显示。

为避免赘述，仅以特定爆点为例说明坝体局部破坏规律和参量变化规律。

2.2.1 坝体下游侧爆点

以1#爆点为例，不同TNT当量的坝体毁伤结果如图5所示。由图5可知，随TNT当量不断增大，坝体下游侧各折坡处产生明显的高应变区域，逐渐向上游延伸；在较大TNT当量的爆炸载荷作用下，坝顶部分在下游折坡约212 m高程处发生断裂，脱离剩余坝体并向上游侧飞散。

综合考虑坝体下游侧7个不同爆点(1#～7#)、7种不同TNT当量共计49个触爆工况的仿真计算

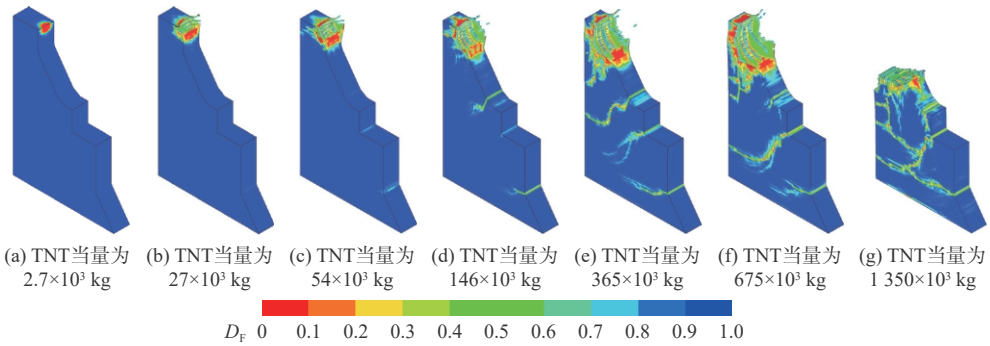
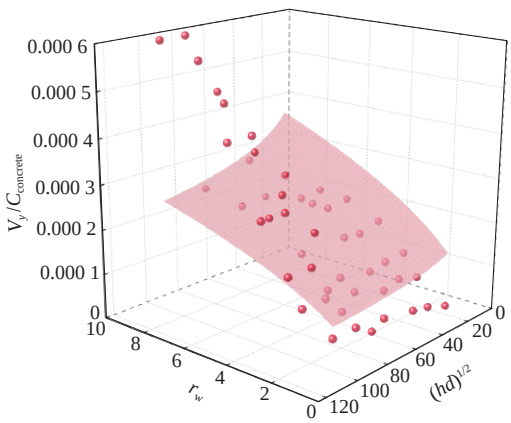


图 5　1#爆点的坝体毁伤结果

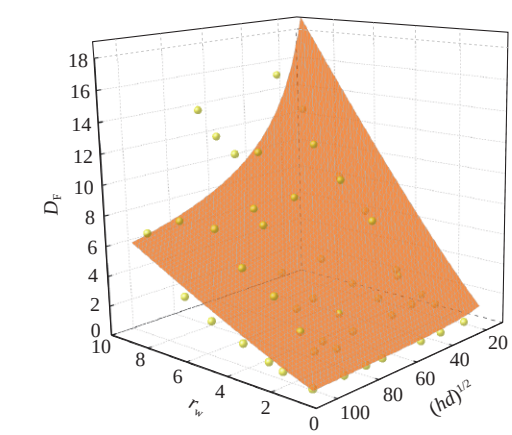
Fig. 5　Dam damage results at blast point 1#

结果, 2 个参量随爆炸当量和爆点位置的变化规律如图 6 所示, 爆炸 TNT 当量的自变量为球型装药半径 r_w, 爆点位置的自变量为爆点距上游侧水平距离 d 与爆点距坝顶的垂直距离 h 之积的平方根, $V_y/C_{concrete}$ 为帷幕振动速度/混凝土声速。图 6 中各拟合曲面公式和拟合系数 R^2 如表 5 所示。

定量分析坝体受损结果和坝体底部帷幕振动速度等坝体动态响应参量, 归纳出下游触爆载荷下的物理毁伤规律如下:



(a) 坝底帷幕振动速度变化规律



(b) 坝体受损体积变化规律

图 6　坝体在下游侧爆点作用下的动态响应规律

Fig. 6　Dynamic response law of dam under action of downstream side explosion points

1) 爆点附近出现较明显的局部破坏区域, 破坏范围随 TNT 当量增大而增大; 当爆点高程大于 210 m 时, 较大 TNT 当量的爆炸载荷使得坝顶部分发生断裂, 从而脱离剩余坝体。

2) 坝底帷幕振动速度随爆点位置不断下移, 呈逐渐增大的趋势; 随爆炸 TNT 当量增大而增大, 峰值约为 0.000 6 倍的混凝土声速。可见, 坝底帷幕振动速度的变化趋势与作用于剩余坝体(考虑坝体断裂情况)的爆炸能量密切相关。随爆点位置下移, 该能量增大; 随 TNT 当量增加, 坝体出现局部毁伤和断裂, 但作用于剩余坝体的爆炸能量也在增大。

3)坝体受损体积占比的变化规律起伏程度较大, 随 TNT 当量增大而增大, 受爆点位置影响较明显。在坝顶侧、折坡处的爆点均造成较大范围的坝体断裂和脱落, 因此, 坝体受损程度较高, 体积占比峰值达到 15% 左右。其余爆点由于未造成坝体断裂, 因此, 坝体受损体积占比均在 10% 以内。

表 5　下游侧爆点计算结果的拟合结果

Table 5　Fitting results of calculation results of downstream side explosion points

动态响应	拟合公式	拟合系数R^2
坝底帷幕振动速度	$V_y/C_{concrete}=0.000\,26\left(r_w/\sqrt{hd}\right)^{0.48}$	0.26
坝体受损体积占比	$D_F=6.86\left(r_w/\sqrt{hd}\right)^{1.13}$	0.7

2.2.2　坝体上游侧爆点

以 8#爆点为例, 不同 TNT 当量的坝体毁伤结果如图 7 所示。由图 7 可知, 爆点附近出现较明显的局部破坏区域, 破坏范围随 TNT 当量增大而增大, 且明显大于下游侧触爆载荷所致的破坏范围; 在大 TNT 当量爆炸载荷的情况下, 坝体损伤仍表现为局部破坏, 未出现明显的断裂现象或趋势。

综合考虑坝体上游侧 4 个不同爆点(8#～11#)、5 种不同 TNT 当量共计 20 个触爆工况的仿真计算结果, 3 个参量随爆炸 TNT 当量和爆点位置的变化规律如图 8 所示, 图中, 各拟合曲面公式和拟合系数 R^2, 如表 6 所示, a_{dam}/g 为坝体加速度/重力加速度。

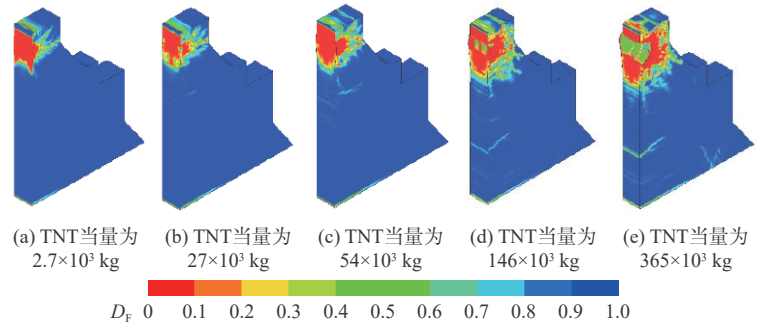


图 7 8#爆点的坝体毁伤结果

Fig. 7 Dam damage results at blast point 8

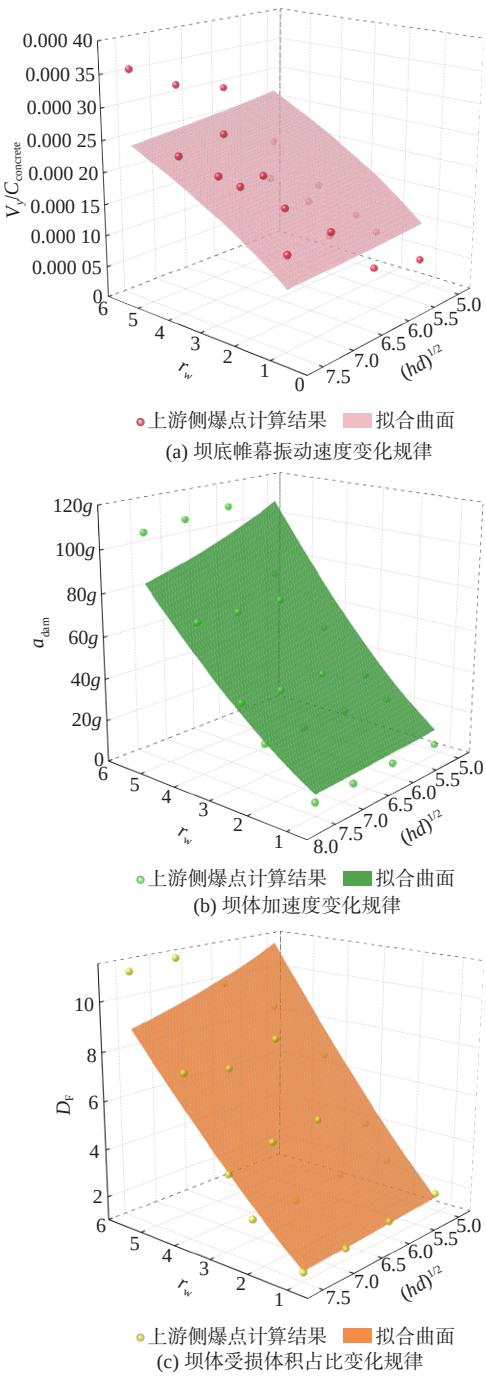


图 8 坝体在上游侧爆点作用下的动态响应规律

Fig. 8 Dynamic response law of dam body under action of upstream side explosion points

上游侧爆点所造成的坝体动态响应规律与下游侧存在以下 2 点区别：

1）由于水中爆炸的原因，相较于下游侧爆点（见图 5），上游侧爆点（见图 7）造成坝体受损范围更大，但未出现坝体大范围断裂的毁伤模式。

2）与下游侧的复杂坝面结构相比，上游侧坝面结构较为简单，因此，坝体动态响应规律同一性相对明显。坝底帷幕振动速度随 TNT 当量和爆点水深的增加而增大，峰值达到 0.003 5 倍的混凝土声速，较深爆点处的振动速度呈现出更明显的增大趋势。坝体加速度和坝体受损体积占比呈现出类似的变化趋势，具体数值随 TNT 当量增加而平滑上升，随爆点深度增加而缓慢增大，峰值分别达到 105g 和 9.1%。

表 6 上游侧爆点计算结果的拟合结果

Table 6 Fitting results of calculation results of upstream side explosion points

动态响应	拟合公式	拟合系数R^2
坝底帷幕振动速度V_y	$V_y/C_{concrete}=0.000\,15\left(r_w/\sqrt{hd}\right)^{0.6}$	0.34
坝体加速度a_{dam}	$a_{dam}/g=28.2\left(r_w/\sqrt{hd}\right)^{1.4}$	0.81
坝体受损体积占比D_F	$D_F=3.56\left(r_w/\sqrt{hd}\right)^{1.2}$	0.86

3 考虑无量纲形式的坝体毁伤规律

3.1 引入无量纲数

谈庆明在《量纲分析》[19]一书中提到：“特别是在研究新现象或新问题的时候，需对现象和问题中蕴含的物理环节、关系和过程进行初步地中肯分析，运用物理学中的基本规律，明确哪些参数对现象或问题起到控制作用。”

混凝土重力坝在触爆载荷下的动态响应是极其复杂的物理过程，仅简单地归纳爆炸 TNT 当量和各类动态响应参量之间的对应关系具有一定的局限性，不能够充分反映实际的物理背景，而且很难拓展分析结果的应用范围。

因此, 为更好地反映坝体动态响应与不同爆炸载荷之间函数关系的物理内涵, 引入无量纲数 B 和 C 分别度量坝体下游侧和上游侧不同工况的爆炸载荷参数, 表达式为

$$B = \frac{r_w}{l} \cdot \frac{h'}{b} = \frac{r_w}{l} \tan\theta \tag{2}$$

$$C = r_w / h' \tag{3}$$

式中: r_w 为触爆载荷的装药半径(假设触爆载荷为球形装药); l 为单个坝段的长度; h'为爆点位置距坝底的高度; b 为爆点到坝体另一侧的水平距离(局部宽度); $\tan\theta$ 为 h'与 b 的比值。图 9 为各参量的示意图。

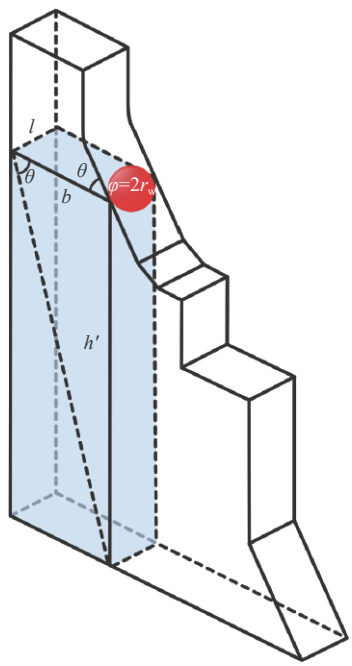


图 9　无量纲数中的参量示意图

Fig. 9　Parameter diagram of dimensionless numbers

无量纲数 B 由 r_w/l 和 h'/b 组成。r_w/l 的物理意义为坝体单位长度上所受到的爆炸载荷, 由于仿真计算中所有坝段长度 l 均相同, 因此, 该项仅与爆炸载荷的实际 TNT 当量大小相关。h'/b 的物理意义直观上为爆点具体位置的度量, 随着爆点位置的不断下降, $\tan\theta=h'/b$ 也在不断减小, 而重力坝下游侧形状并不规则, 所以事实上 θ 或 h'/b 定量地体现了下游侧坝面由竖直逐渐变化为水平面的过程。由图 9 可知, 在下游侧坝顶处, 爆炸载荷作用于坝顶的竖直面, 随着爆点位置的下降, 沿水平方向作用于坝面的爆炸载荷比例不断减小, 上游侧爆点同理。因此, h'/b 的物理意义为爆炸载荷对坝体的水平方向作用比例的大小, 对于特定坝段几何结构, 该项仅与爆点位置相关。

无量纲数 C 的物理意义为坝体受上游侧爆炸载荷的影响情况, 即随 TNT 当量增大深度减小, 对应坝体上游侧的竖直坝面。

基于无量纲数 B 和 C, 考虑无量纲形式处理方法, 可分别归纳出重力坝在触爆载荷作用下各动态响应参数(具体化表述)的无量纲变化规律:

$$V_y / c_{\text{concrete}} = f_1(B) \tag{4}$$

$$D_F = f_2(B) \tag{5}$$

$$V_y / c_{\text{concrete}} = f_3(C) \tag{6}$$

$$a_{\text{dam}} / g = f_4(C) \tag{7}$$

$$D_F = f_5(C) \tag{8}$$

式中: c_{concrete} 为混凝土中的声速, 取 4 800 m/s; g 为重力加速度, 取 9.8 m/s^2。

除爆炸载荷有关参量外, 重力坝在触爆载荷下的动态响应与混凝土材料参数(抗压、抗拉强度等)、炸药材料参数(密度、爆速等)等很多参量同样相关[20], 但由于本文仿真计算工况设置的限制及无量纲表达式的简化需求, 这些参量暂未纳入考虑。

3.2　坝底帷幕振动速度

重力坝在下游侧爆点作用下的坝底帷幕振动速度随无量纲数 B 变化规律如图 10(a)所示。重力坝的下游侧坝面形状并不规则, 因此, 坝体在下游侧触爆载荷作用下的动态响应不仅与爆炸 TNT 当量相关, 也明显受到爆点位置变化的影响, 无量纲数 B 的引入较好地将这 2 个影响因素结合起来了。

然而, 在对下游侧爆点作用下坝体受损占比体积变化规律分析时发现, 将所有数据点统一考虑时规律性不明显, 但 1#～3#爆点的结果与 4#～7#爆点存在明显的不同趋势。因此, 拟建立分段函数, 并利用爆点位置数据区分 2 组数据, 以提高规律曲线和数据点的一致性。由于无量纲数 B 不仅包含爆点位置信息还包含爆炸 TNT 当量信息, 不宜直接作为分段依据。参考无量纲数 B 的表达式, 可取 3#爆点和 4#爆点在坝面上的中点位置信息, 即 h'/b=2.2 作为函数分段点。图 10(a)中黑色三角形散点为 4#～7#爆点的计算结果, 对应 $h'/b \leqslant 2.2$, 拟合曲线为大红色, R^2=0.51; 灰色菱形散点为 1#～3#爆点的计算结果, 对应 $h'/b>2.2$, 拟合曲线为浅红色, R^2=0.69。因此, 坝体在下游侧触爆载荷作用下的坝底帷幕振动速度可表达为

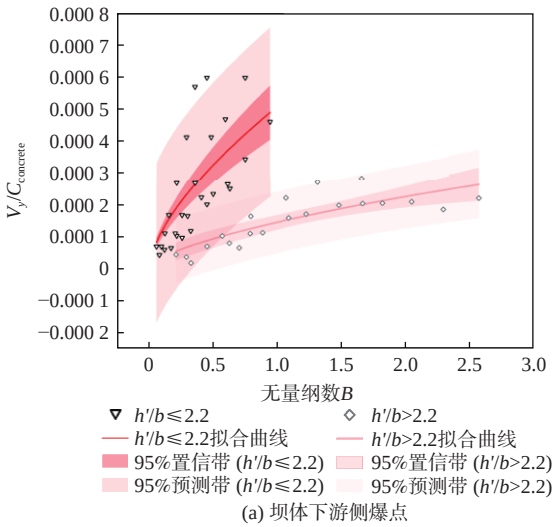
$$V_y / c_{\text{concrete}} = \begin{cases} 1.4\times10^{-4} B^{0.64} & h'/b \leqslant 2.2 \\ 5.1\times10^{-4} B^{0.64} & h'/b > 2.2 \end{cases} \tag{9}$$

重力坝在上游侧爆点作用下的坝底帷幕振动

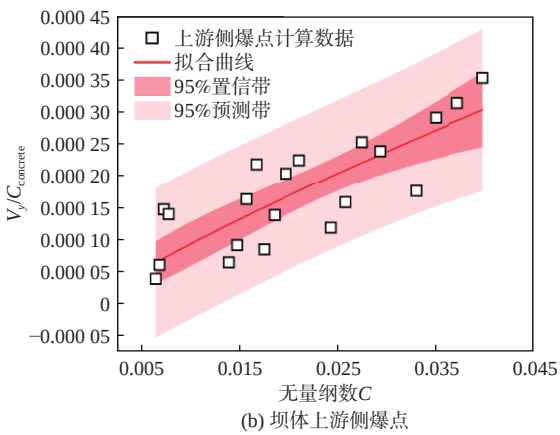
速度随无量纲数 C 变化规律如图 10(b)所示，图中，黑色方形散点为 8#～11#爆点的计算结果，红色曲线为拟合曲线，R^2=0.65。

坝体在上游侧触爆载荷作用下的坝底帷幕振动速度计算式为

$$V_y/c_{\text{concrete}} = 0.004\,7C^{0.85} \tag{10}$$



(a) 坝体下游侧爆点



(b) 坝体上游侧爆点

图 10 坝底帷幕振动速度规律曲线

Fig. 10 Law curves of curtain vibration velocity at dam base

3.3 坝体受损体积占比

重力坝在下游侧爆点作用下的坝体受损体积占比随无量纲数 B 变化规律如图 11 所示。

图 11 中浅蓝色三角形散点为 4#～7#爆点的计算结果，对应 h'/d≤2.2，拟合曲线为橙红色，R^2=0.81；深蓝色菱形散点为 1#～3#爆点的计算结果，对应 h'/d >2.2，拟合曲线为黄色，R^2=0.87。因此，坝体在下游侧触爆载荷作用下的受损体积占比可表达为

$$D_{\text{F}}=\begin{cases}3.8B^{1.4} & h'/b>2.2\\ 15B^{1.4} & h'/b\leqslant 2.2\end{cases} \tag{11}$$

图 11 和式 (11) 所表现出的规律性与无量纲数 B 的物理意义相契合。相同 TNT 当量的情况下，坝

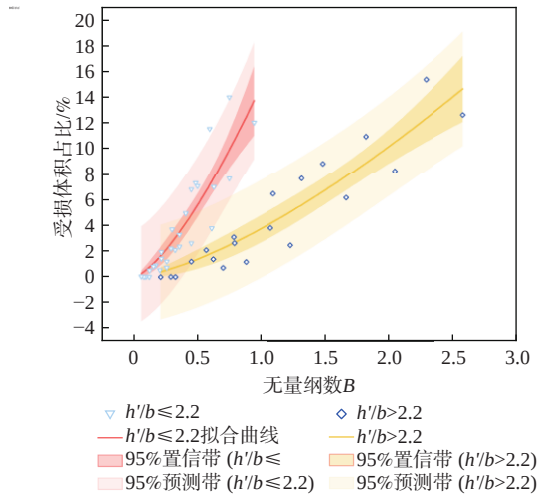


图 11 受损体积占比规律曲线(无量纲数 B)

Fig. 11 Law curves of damage fraction for blast points on downstream side (dimensionless number B)

体下游侧上部分爆点所造成的受损体积占比大于下部分爆点，此时无量纲数 B 的物理意义可以简化地等同于 h'/d，即爆炸载荷对坝体的水平方向作用比例大小。h'/b >2.2 时，爆炸载荷更多地作用于水平方向，更易造成坝体顶部断裂；h'/b ≤2.2 时爆炸载荷作用于竖直方向的比例更多，坝体受损体积占比也相应下降。

上游侧爆点作用下的坝体受损体积占比随无量纲数 C 变化规律如图 12 所示。

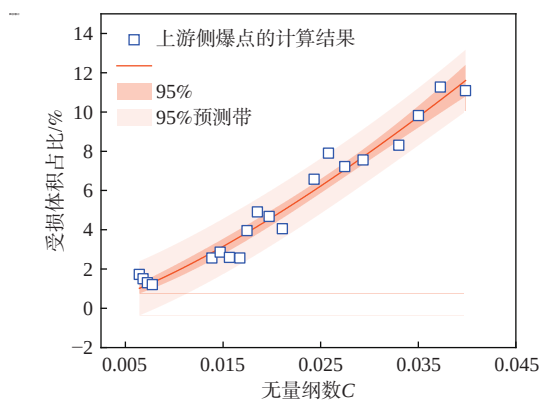


图 12 受损体积占比规律曲线(无量纲数 C)

Fig. 12 Law curve of damage fraction for blast points on upstream side (dimensionless number C)

由 12 图可知，所有上游侧爆点的数据点呈较好的一致性，拟合曲线 R^2 达到 0.96。坝体在上游侧触爆载荷作用下的受损体积占比计算式为

$$D_{\text{F}} = 854.1C^{1.3} \tag{12}$$

3.4 坝体加速度

重力坝在上游侧爆点作用下的坝体整体加速度随无量纲数 C 变化规律如图 13 所示，图中，紫色方形散点为 8#～11#爆点的计算结果，拟合曲线为

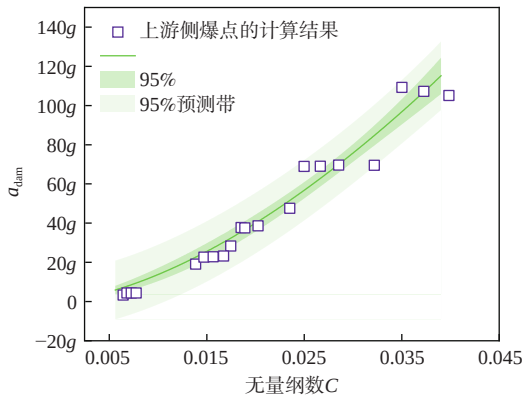


图 13　坝体加速度规律曲线

Fig. 13　Law curve of dam acceleration

绿色, R^2=0.96。

坝体在上游侧爆点作用下的坝体整体加速度计算式为

$$a_{dam}/g = 2.2\times10^4 C^{1.62} \qquad (13)$$

4　结　论

1) 与坝体加速度和坝体受损体积占比 2 个坝体动态响应参量相比, 坝底帷幕振动速度的变化规律受爆点位置影响更明显, 结构特殊性较强的下游侧可采用分段表征的方式提升规律性。

2) 对坝体动态响应规律数据的归纳和拟合过程中, 发现以爆炸 TNT 当量和爆点位置为自变量时, 坝体动态响应规律可通过幂次曲面函数进行拟合, 可为建立相关工程算法提供理论支撑。

3)通过建立具有实际物理意义的无量纲数, 可将坝体动态响应规律的表达式由多自变量转换为单自变量, 相较于曲面形式进一步提升了其规律性和简洁性, 能够有效支撑重力坝的易损性分析和毁伤评估相关工作。

参考文献（References）

[1] 陈叶青, 吕林梅, 王高辉, 等. 大坝抗爆性能评估研究进展[J]. 土木工程学报, 2021, 54(10): 9-19.
CHEN Y Q, LYU L M, WANG G H, et al. Review on the blast-resistance performance evaluation of dams[J]. China Civil Engineering Journal, 2021, 54(10): 9-19(in Chinese).

[2] HUANG X P, HU J, ZHANG X D, et al. Effect of bubble pulse on concrete gravity dam subjected to underwater explosion: Centrifuge test and numerical simulation[J]. Ocean Engineering, 2022, 243: 110291.

[3] WANG X H, ZHANG S R, WANG C, et al. Blast-induced damage and evaluation method of concrete gravity dam subjected to near-field underwater explosion[J]. Engineering Structures, 2020, 209: 109996.

[4] WANG X H, ZHANG S R, DAI J G, et al. Evaluation of base damage and stability of concrete gravity dam subjected to underwater explosion[J]. Structures, 2022, 38: 1502-1514.

[5] LI Q, WANG G H, LU W B, et al. Failure modes and effect analysis of concrete gravity dams subjected to underwater contact explosion considering the hydrostatic pressure[J]. Engineering Failure Analysis, 2018, 85: 62-76.

[6] CHEN J Y, LIU X P, XU Q. Numerical simulation analysis of damage mode of concrete gravity dam under close-in explosion[J]. KSCE Journal of Civil Engineering, 2017, 21(1): 397-407.

[7] XU Q, CHEN J Y, LI J, et al. Numerical study on antiknock measures of concrete gravity dam bearing underwater contact blast loading[J]. Journal of Renewable and Sustainable Energy, 2018, 10: 014101.

[8] LIU Y J, LU W B, WANG G H, et al. Structural parameter inversion of a gravity dam based on the dynamic response induced by an underwater explosion[J]. Journal of Vibration and Control, 2023, 29(7-8): 1577-1589.

[9] VANADIT-ELLIS W, DAVIS L K. Physical modeling of concrete gravity dam vulnerability to explosions[C]//Proceedings of the International WaterSide Security Conference. Piscataway: IEEE Press, 2011: 1-11.

[10] XU Q, CAO Y, CHEN J Y, et al. Study on the response and damage characteristic of concrete gravity dam subjected to contact explosion[J]. Journal of Hunan University, 2016(43): 62-74.

[11] ZHANG S R, KONG Y, WANG G H. Dynamic responses of a concrete gravity dam subjected to underwater and air explosions[J]. Journal of Vibration Shock, 2014, 33(17): 47-54.

[12] 李本平, 王永, 卢文波. 制导炸弹在坝前水面爆炸破坏效应研究[J]. 爆破, 2007, 24(4): 7-10.
LI B P, WANG Y, LU W B. Study on the damage effect for water surface explosion of precision guided bomb[J]. Blasting, 2007, 24(4): 7-10(in Chinese).

[13] 李本平, 卢文波. 制导炸弹水平侵彻爆炸作用下混凝土重力坝毁伤效应数值仿真[J]. 爆破, 2007, 24(1): 1-5.
LI B P, LU W B. Numerical simulation for the demage of concrete gravity dam under horizontal penetration and explosion of GBU-28[J]. Blasting, 2007, 24(1): 1-5(in Chinese).

[14] 李本平. 制导炸弹连续打击下混凝土重力坝的破坏效应[J]. 爆炸与冲击, 2010, 30(2): 220-224.
LI B P. Damage effect of a concrete gravity dam under continuous attacks of guided bombs[J]. Explosion and Shock Waves, 2010, 30(2): 220-224(in Chinese).

[15] ZHANG K F, LU F Y, PENG Y, et al. Study on dynamic response of gravity dam under air blast load based on similarity law[J]. Engineering Failure Analysis, 2022, 138: 106225.

[16] 王高辉, 张社荣, 卢文波, 等. 水下爆炸冲击荷载下混凝土重力坝的破坏效应[J]. 水利学报, 2015, 46(6): 723-731.
WANG G H, ZHANG S R, LU W B, et al. Damage effects of concrete gravity dams subjected to underwater explosion[J]. Journal of Hydraulic Engineering, 2015, 46(6): 723-731(in Chinese).

[17] 黄谢平, 孔祥振, 陈祖煜, 等. 近水面、库中、库底水下爆炸荷载作用下混凝土重力坝的破坏模式对比[J]. 土木工程学报, 2023, 56(3): 116-128.
HUANG X P, KONG X Z, CHEN Z Y, et al. Comparison of failure modes of concrete gravity dams induced by underwater explosion

loads near the water free surface, the middle, and the bottom of the reservoir[J]. China Civil Engineering Journal, 2023, 56(3): 116-128 (in Chinese).

[18] SHU Y Z, WANG G H, LU W B, et al. Damage characteristics and failure modes of concrete gravity dams subjected to penetration and explosion[J]. Engineering Failure Analysis, 2022, 134: 106030.

[19] 谈庆明. 量纲分析[M]. 合肥: 中国科学技术大学出版社, 2005.
TAN Q M. Dimensional analysis[M]. Hefei: University of Science and Technology of China Press, 2005(in Chinese).

[20] 林继镛, 张社荣. 水工建筑物[M]. 6 版. 北京: 中国水利水电出版社, 2019.
LIN J Y, ZHANG S R. Hydraulic structure[M]. 6th ed. Beijing: China Water & Power Press, 2019(in Chinese).

Dynamic response law of concrete gravity dam under contact explosion load

ZHANG Kefan[1,*], PENG Yong[2], LI Xiangyu[2], LU Fangyun[2], WANG Zhe[1]

(1. Test Center, National University of Defense Technology, Xi'an 710000, China;
2. School of Science, National University of Defense Technology, Changsha 410072, China)

Abstract: This paper uses the simulation model verified by scale test to conduct the simulation research of 69 working conditions with different explosion point positions and different equivalent sizes, given the lack of experimental verification, weak regularity, and challenging application of the dynamic response research of concrete gravity dam under the action of contact explosion load. The time history curves of three parameters of dam bottom curtain vibration velocity, dam acceleration and damaged dam ratio of gravity dam under contact explosion load are investigated, and the law of dynamic response of dam body is preliminarily summarized. On this basis, the dimensionless numbers B and C that can correlate the explosion load parameters with the dynamic response are further proposed. The variation law and trend of the dynamic reaction of the gravity dam based on the analysis of explosion conditions are derived by integrating the parameter values change information of all simulation results. The dimensionless dynamic response law of the dam body can effectively correlate the physical damage and functional damage of the dam body, so as to support the vulnerability analysis of gravity dams, and provide a reference for the study of the damage mechanism of other engineering targets.

Keywords: concrete gravity dam; contact explosion load; dynamic response; damage law; dimensionless number

Received: 2024-02-28; **Accepted**: 2024-05-09; **Published Online**: 2024-05-31 08:05
URL: link.cnki.net/urlid/11.2625.V.20240530.1606.001
Foundation item: National Natural Science Foundation of China (12302450)
* **Corresponding author**. E-mail: kfzhang_v@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0096

航天 AFRP 的紫外飞秒激光加工特性与均质化去除方法

路明雨[1]，张明[2]，卫宇璇[3]，李波[3]，崔志刚[3]，张开虎[3,*]

(1. 北京卫星环境工程研究所，北京 100094；2. 中国空间技术研究院，北京 100094；3. 北京卫星制造厂有限公司，北京 100094)

摘 要：针对 Kevlar 49/4211 芳纶纤维/环氧树脂复合材料的紫外飞秒激光（343 nm，260 fs）加工阈值和形貌特性进行研究。通过面积外延法测算了芳纶纤维增强树脂基复合材料 (AFRP) 中树脂基体、增强纤维及复合材料整体的激光去除阈值及阈值孵化效应，并预测三者的单脉冲去除阈值；分析入射能量通量与等效入射激光脉冲数量对材料加工边缘质量的影响规律。结果表明：Kevlar 49 芳纶纤维的紫外飞秒激光去除阈值和阈值孵化系数在不同脉冲数作用下均低于 4211 环氧树脂，AFRP 整体的等效去除阈值介于树脂和芳纶纤维之间。选择 4211 环氧树脂阈值若干倍的入射通量与 100 个以上等效脉冲数的紫外飞秒激光，采用原位叩击打孔或循环扫描切割的方法，能够获得入口宽度均匀、边缘整齐、材料准均质化去除的加工形貌，从而有效提高航天 AFRP 产品的加工精度与质量，满足其高性能制造需求。

关 键 词：航天复合材料；芳纶纤维；环氧树脂；紫外飞秒激光；去除阈值；孵化效应

中图分类号：V461; TB332

文献标志码：A **文章编号**：1001-5965(2026)05-1557-10

随着航天器朝着极端尺寸、高稳定性、高承载能力的趋势发展，其对结构机构及载荷的强度、刚度和轻量化需求日益突出。芳纶纤维增强树脂基复合材料 (aramid fiber reinforced polymer, AFRP) 作为一种具有高比强度、高比刚度、高韧性、高绝缘性等一系列优点的复合材料，已经成为空间飞行器制造过程中不可或缺的关键材料[1-2]，目前广泛应用于卫星固面天线副反射器与各类管路产品中。

AFRP 中的树脂基体与增强纤维的物理性质差异较大，且整体为各向异性非均质的层叠材料，这些特性使其成为一种典型的难加工材料。常见的机加工方法易使 AFRP 产生毛边、分层及纤维撕裂等缺陷，导致产品的制造精度和质量下降[3]。

近年来，针对传统方法加工 AFRP 存在的上述问题，国内外科研人员对其新型特种加工方法进行了一系列探索，研究方向包括激光加工、电加工、磨粒水射流加工、超声振动辅助切削加工等。其中，飞秒激光技术以其独特的物理机制和加工优势[4]展现出了满足航天 AFRP 产品高性能制造需求的潜力。

飞秒激光通常是指脉冲宽度小于 1 000 fs 量级的激光，其脉冲宽度极窄、能量密度极高、与材料作用时间极短，能够提供极端的制造精度，实现亚微米乃至纳米级的超高精度制造目标。其是一种具有非接触式、超强去除能力、超高加工精度、低热损伤等优势的新型加工技术，有望克服上述机械接触式加工存在的问题，成为一种实现航天 AFRP 产品高性能加工、满足其航天制造需求的理想方法。

然而，AFRP 多相层合的非均质特性也伴随着

收稿日期：2024-02-26；**录用日期**：2024-04-03；**网络出版时间**：2024-04-30 18:31

网络出版地址：link.cnki.net/urlid/11.2625.V.20240430.1406.002

基金项目：国家重点研发计划 (2022YFB4601303)；国家自然科学联合基金重点项目 (U2037205)

* **通信作者**. E-mail：zhangkh86@163.com

引用格式：路明雨，张明，卫宇璇，等. 航天 AFRP 的紫外飞秒激光加工特性与均质化去除方法 [J]. 北京航空航天大学学报，2026，52（5）：1557-1566. LU M Y，ZHANG M，WEI Y X，et al. Characteristics and homogenization removal methods of ultraviolet femtosecond laser processing of aerospace AFRP [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（5）：1557-1566（in Chinese）.

各组分之间激光吸收特性的差别，这使得激光与 AFRP 之间相互作用的物理过程十分复杂，影响实际加工效果的因素较多。因此，为实现航天 AFRP 的飞秒激光高性能加工，需要对相应的理论、技术和工艺方法进行研究。

针对上述问题，国内外科研人员已经开展了一些研究工作。在国外，Al-Sulaiman 等[5]研究了利用 CO_2 激光在 Kevlar 纤维树脂基 AFRP 层合板上切割通孔的效果及辅助气体射流对加工质量的影响规律。结果表明，辅助气体的气压对 AFRP 的切割断面质量有较大影响，当辅助气体的气压较高时，切缝的宽度会有所增加，壁面粗糙度也有明显的改善。Yilbas 等[6]通过 CO_2 激光对不同厚度的 AFRP 层合板进行切割实验。结果表明，随着激光切割速度的改变，AFRP 表面的平均切缝宽度和热影响区尺寸均会发生显著地变化。Rao 等[7]研究了激光脉宽、重复频率、气压和切割速度等参数对 AFRP 切割效果的影响，并利用实验数据建立了二阶非线性回归模型，预测了材料切口顶部和底部的切口偏差。结果表明，激光切割复合材料时，切割壁面的几何特征和加工质量很大程度上取决于加工参数的选择。Mishra 等[8]采用 Nd:YAG 激光切割 AFRP 层压板（纤维体积分数约为 55.2%，材料厚度约为 1.92 mm），测量了切缝顶部和底部的宽度，并讨论了各参数对切口质量的影响。Hirogaki[9]也对激光加工 AFRP 开展了研究。在国内，王含妮等[10]利用 COMSOL 软件建立了描述 AFRP 在激光作用下温度场变化规律的理论模型提出纤维增强树脂基复合材料的激光精密加工技术，分析了激光脉宽和波长，并采用不同参数的纳秒激光对 AFRP 进行了加工实验，结果表明，实验结果与数值模型仿真结果比较吻合，最大相对误差约为 12.49%。Zhang 等[11]采用超短脉冲激光切割航天复合材料的蜂窝叠层结构，得到了较高精度和质量的切割效果。杨剑等[12]开展了 AFRP 材料的皮秒激光切割工艺研究。采用紫外皮秒激光（343 nm 波长）、绿皮秒激光（532 nm 波长）和近红外皮秒激光（1 064 nm 波长）分别对 1 mm 厚的 AFRP 层合板进行单次扫描切割实验。结果表明，紫外光切割 AFRP 的质量更好，效率也更高。AFRP 的切缝宽度、深度和表面热影响区尺寸与入射的激光功率呈正相关，与激光扫描速度呈负相关。而切缝的锥度则与入射激光功率呈负相关，与激光扫描速度呈正相关。路明雨等[13]综述了激光加工非金属复合材料的研究与应用进展，其中，针对航天领域应用较为广泛的芳纶纤维复合材料，分析了当前国内外激光加工 AFRP 的研究与应用进展。并且对激光加工非金属复合材料的重点研究方向和航天应用前景进行了总结与展望。

从国内外研究现状可以看出，目前大多数研究采用了 CO_2 激光、毫秒、纳秒等传统激光加工 AFRP，并且工艺与方法研究主要集中于加工热影响区和锥角的控制。由于传统激光的热烧蚀现象比较明显，加工 AFRP 的热损伤尺寸通常为 50～1 000 μm，通常难以满足航天产品的精密高性能加工需求。相比之下，采用飞秒/皮秒等超短脉冲激光能够显著提高 AFRP 的加工精度和质量，但有关的研究内容比较有限，相应的机理、方法、技术与工艺研究尚处于起步阶段。

本文围绕航天 AFRP 复合材料的高性能加工需求，采用波长较短的紫外飞秒激光（波长 λ=355 nm，脉冲宽度 f=260 fs）对航天 AFRP（Kevlar 49/4211）进行原位打孔和扫描切割试验，利用光学显微镜和扫描电镜观察了激光加工后材料表面的烧蚀坑形貌与尺寸。基于烧蚀坑尺寸通过面积外延法拟合测算了 AFRP 中环氧树脂、Kevlar 纤维及复合材料整体的飞秒激光去除阈值，并研究了三者阈值随入射激光脉冲数的变化规律。此外，基于各组分的阈值特性与实际加工工况，提出实现 AFRP 材料飞秒激光准均质化去除的方法。定性观察 AFRP 在不同参数下的加工形貌差异，得到了加工精度和质量均较好的工艺参数。

1 实验设备与方法

1.1 实验材料

在航天 AFRP 复合材料中，环氧树脂及其改性树脂是目前应用最广泛、牌号最多、用量最大的树脂基体，也是未来制造高性能 AFRP 的主要基体材料。环氧树脂分子中含有 2 个或 2 个以上的环氧基，在固化剂作用下形成三向交联结构化合物，经过高温固化后具有出色的工艺性能、综合力学性能（高强度、高韧性）、耐空间环境性能和耐高低温性能[14]。环氧树脂主要成分的分子结构式如图 1 所示。

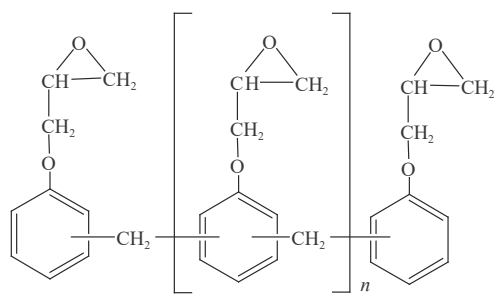


图 1 环氧树脂材料的基本分子式

Fig. 1 Basic molecular formula of epoxy resin materials

首先，选择 4211 环氧树脂作为研究对象，该树脂由 648 酚醛环氧树脂与三氟化硼单乙胺按 100∶3 的配比组成，是航天领域中最典型的环氧树脂基

体。此处采用湿法或热熔工艺将其制备为预浸料，再经高温加压流平固化后粘贴于铝合金垫片表面制成树脂样件，作为紫外飞秒激光加工的试验材料，4211 环氧树脂试片如图 2 所示。

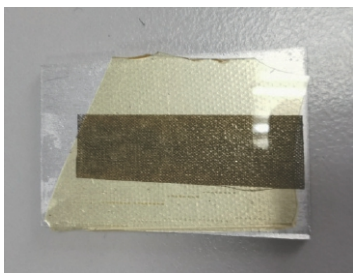


图 2　4211 环氧树脂实验试片

Fig. 2　4211 epoxy resin experimental specimen

AFRP 中使用的芳纶纤维主要分为聚对苯二甲酰对苯二胺纤维（如美国杜邦公司的 Kevlar 49、荷兰恩卡公司的 TwaronHM、中国的芳纶 1414）和聚对苯甲酰胺纤维（如 Kevlar 29、芳纶 14）2 类。其中，Kevlar 49 是一种低密度、高强度、高模量、耐高温的有机纤维，也是目前用于卫星结构的 AFRP 产品中使用最多的增强纤维之一。

因此，以 Kevlar 49 芳纶纤维为研究对象。将纤维束粘贴于铝合金垫片上制成试片作为紫外飞秒激光加工的试验材料如图 3 所示。通过实验测算 Kevlar49 芳纶纤维的超快激光加工阈值特性。

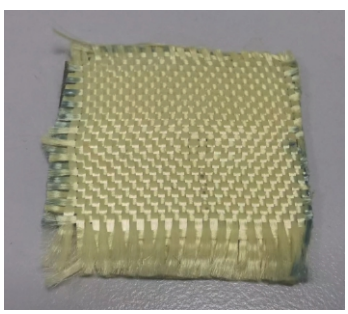


图 3　Kevlar 49 芳纶纤维束实验试片

Fig. 3　Kevlar 49 aramid fiber bundle experimental specimen

最后，以 4211 环氧树脂与 Kevlar 49 芳纶纤维作为基体材料与增强材料，把相应的 Kevlar 纤维网格布与树脂胶黏固化为预浸料，再经过多层铺放压叠裁剪等工艺制成 0.8 mm 厚的 AFRP 复合材料层合板试片如图 4 所示，将固化成型后的 Kevlar 49/4211 AFRP 整体作为加工实验对象。

1.2　实验设备

实验采用的紫外飞秒激光加工系统和加工过程如图 5 和图 6 所示。

采用紫外飞秒激光作为实验光源，通过扫描振镜对材料表面进行聚焦加工。所用的线偏振脉冲高斯激光束参数和性质如表 1 所示。

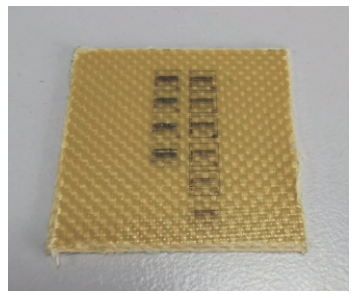


图 4　Kevlar 49/4211 复合材料实验试片

Fig. 4　Kevlar 49/4211 composite material experimental specimen

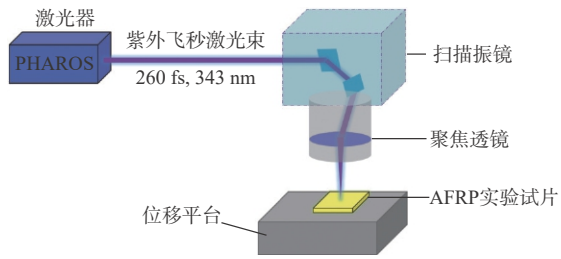


图 5　紫外飞秒激光加工系统示意图

Fig. 5　Schematic diagram of ultraviolet femtosecond laser processing system

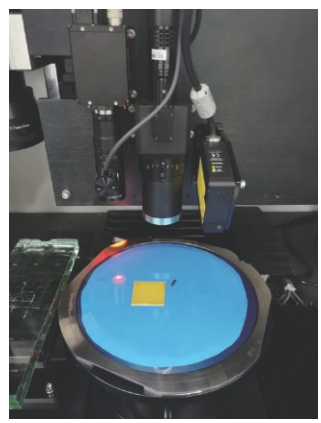


图 6　AFRP 试片加工试验过程

Fig. 6　AFRP specimen processing and testing process

表 1　激光器基本参数

Table 1　Basic parameters of laser

平均功率P/W	脉冲重复频率f_r/kHz	脉冲宽度(半高全宽)/fs	波长λ/nm
0.01～20	1～1 000	260	343

1.3　实验方法

将 4211 环氧树脂试片、Kevlar 49 纤维试片及 AFRP 试片分别置于激光的焦平面上，利用 1.2 节紫外飞秒激光对其进行原位叩击。再利用 OLYMPUSGX51 型光学金相显微镜对激光叩击实验的烧蚀坑形貌及孔径进行观测与数据采集。根据面积外延法[15]理论公式：

$$D^2 = 2\omega_0^2 \ln(E/E_{\text{th}}) \tag{1}$$

式中：D 为激光烧蚀坑直径；ω_0 为作用在材料表面的实际光斑半径 ($1/e^2$ 型光斑半径定义)；E 为每个激光脉冲的能量，可由激光平均功率 P 与脉冲重复频率 f_c 的比值得到；E_{th} 为材料的去除阈值能量。由式(1)可以得到 D^2 与 $\ln E$ 之间的线性关系，通过最小二乘法线性拟合若干对($\ln E$, D^2)坐标点，即可根据拟合直线的斜率和截距求出 ω_0 和 E_{th}。在得到 ω_0 后，再利用峰值通量 F(高斯光斑中心通量)与其高斯脉冲能量 E 之间的关系：

$$F = 2E/\pi\omega_0^2 \tag{2}$$

将激光脉冲的能量值转化为通量值，即可确定与 E_{th} 对应的去除阈值通量 F_{th}。

2 结果与分析

2.1 4211 环氧树脂的紫外飞秒激光去除阈值与孵化效应

将 4211 环氧树脂样件置于激光焦平面上，用上述紫外飞秒激光对其进行原位叩击，并对激光叩击实验后树脂表面的烧蚀孔径进行观测与数据采集。

在脉冲宽度为 260 fs、波长为 343 nm、脉冲重复频率为 50 kHz 的紫外飞秒激光叩击作用下，不同激光能量 E 和作用脉冲数 N 对应的环氧树脂烧蚀孔径形貌如图 7 所示。

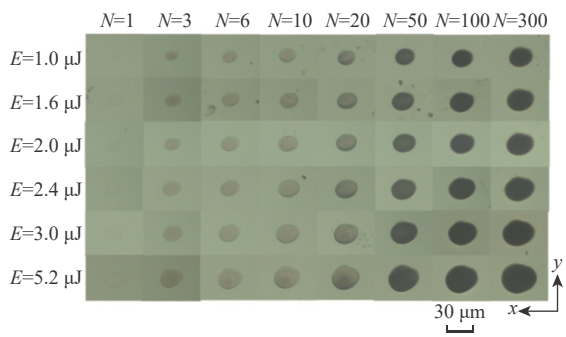


图 7 4211 环氧树脂在不同激光入射能量和入射脉冲数作用下的烧蚀坑

Fig. 7 Erosion pits of 4211 epoxy resin under different laser incident energies and incident pulse numbers

从图 7 中测量出 4211 环氧树脂在不同激光通量和不同脉冲数作用下的烧蚀坑直径(实验过程中 x 方向存在多次叩击产生的轻微位移误差，将 y 方向的尺寸作为实际烧蚀直径)。并根据面积外延法式(1)和式(2)得到 4211 环氧树脂在不同脉冲数作用下烧蚀坑直径平方随激光通量的变化关系，如图 8 所示。

根据图 8 中数据拟合计算得到紫外飞秒激光加工 4211 环氧树脂时的实际光斑直径 d_s 和去除阈值通量 F_{th}，如表 2 所示。

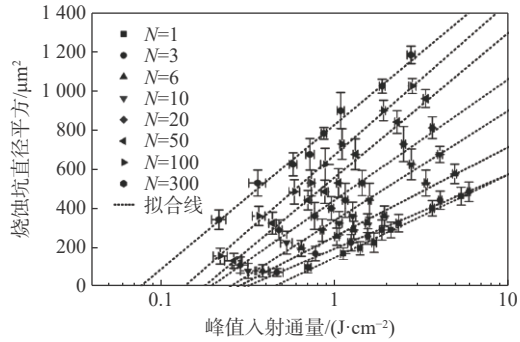


图 8 4211 环氧树脂在不同脉冲数作用下的烧蚀坑直径平方 D^2 与入射通量 F_0 关系

Fig. 8 Relationship between square of ablation pit diameter (D^2) and incident flux (F_0) of 4211 epoxy resin under different pulse numbers

表 2 4211 环氧树脂在不同激光脉冲数下的实际光斑直径与去除阈值通量

Table 2 Actual spot diameter and removal threshold flux of 4211 epoxy resin under different laser pulse numbers

脉冲数 N	实际光斑直径 d_s/μm	阈值通量 F_{th}/(J·cm^{-2})
1	19.16±0.36	0.44±0.04
3	18.17±0.46	0.31±0.02
6	19.95±0.41	0.28±0.02
10	22.10±0.54	0.25±0.02
20	23.18±0.77	0.19±0.02
50	24.22±0.72	0.15±0.01
100	26.50±0.86	0.14±0.01
300	26.69±0.98	0.09±0.01

由表 2 可知，环氧树脂的去除阈值会随着入射激光脉冲数的变化而改变，这一现象被称为阈值“孵化”效应[16]。这种效应是材料被激光脉冲多次照射后诱导的缺陷积累所导致的，这些缺陷会促进材料对激光能量的吸收，降低材料的有效去除阈值并增加材料的去除率。因此，需要讨论 4211 环氧树脂的去除阈值与入射激光脉冲数之间的定量关系。根据表 2 中 4211 环氧树脂在不同脉冲数作用下的紫外飞秒激光去除阈值，可以得到其孵化效应曲线，如图 9 所示。

由图 9 可知，4211 环氧树脂的紫外飞秒激光去除阈值随激光入射脉冲数的增加而下降，并且去除阈值的下降速度呈逐渐减小的趋势。其脉冲阈值约为 0.42 J/cm^2，当脉冲数从 1 增加到 50 时，去除阈值快速减小到 0.15 J/cm^2，而当脉冲数从 50 增加到 300 时，去除阈值仅减小到 0.09 J/cm^2。结果表明，当脉冲数目增加到一定程度后树脂材料的阈值会达到饱和。

得到激光脉冲积累效应的孵化曲线后，还需对材料阈值的“孵化”程度进行表征。材料的 N 脉冲

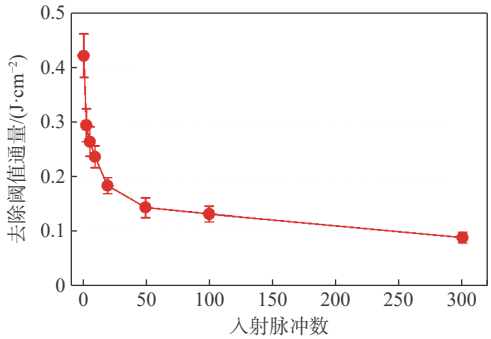


图 9　4211 环氧树脂去除阈值随激光入射脉冲数变化规律

Fig. 9　Variation law of 4211 epoxy resin removal threshold with number of laser incident pulses

去除阈值通量 $F_{th}(N)$ 与单脉冲去除阈值通量 $F_{th}(1)$ 之间关系的模型为[17]

$$F_{th}(N) = F_{th}(1)N^{S-1} \tag{3}$$

式中: S 为孵化系数, 能够体现材料中产生孵化效应的程度, 当 S=1 时, N 脉冲阈值与单脉冲阈值相等, 说明材料的去除阈值不随入射激光脉冲数而改变, 即没有产生孵化效应。对式(3)进行变换(两端同时乘以 N 并取 lg), 可得

$$\lg(NF_{th}(N)) = S\lg(N) + \lg(F_{th}(1)) \tag{4}$$

根据式(4)对多脉冲去除阈值 $F_{th}(N)$ 进行拟合, 得到 4211 环氧树脂在紫外飞秒激光作用下的孵化系数 S 和单脉冲去除阈值 $F_{th}(1)$, 结果如图 10 所示。

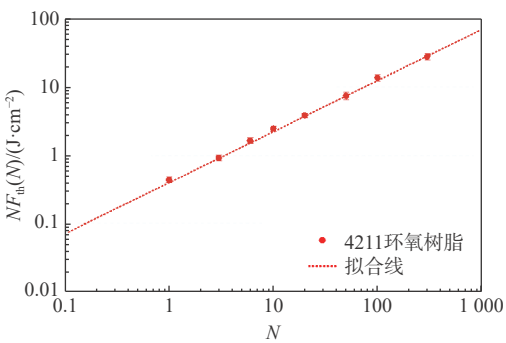


图 10　$NF_{th}(N)$ 随入射脉冲数目 N 的变化关系

Fig. 10　Relationship between $NF_{th}(N)$ and number of incident pulses N

由式(4)可知, 拟合直线的斜率代表孵化系数 S, 截距为单脉冲阈值 $F_{th}(1)$。根据图 10 中数据可得, 4211 环氧树脂的孵化系数约为 0.73, 单脉冲阈值约为 0.42 J/cm^2, 该结果也与实验直接测出的单脉冲去除阈值十分接近, 误差约为 4%, 验证了模型式(3)和式(4)拟合树脂孵化效应的合理性。

2.2　Kevlar 49 芳纶纤维的紫外飞秒激光去除阈值与孵化效应

采用紫外飞秒激光对芳纶纤维试片进行实验并采集加工后的数据, 同样通过面积外延法拟合计算得到 Kevlar 49 芳纶纤维材料的去除阈值, 并通过孵化效应模型计算芳纶纤维的单脉冲去除阈值及阈值孵化系数, 结果如图 11～图 13 所示。



图 11　Kevlar 49 芳纶纤维在紫外飞秒激光原位叩击时的烧蚀形貌

Fig. 11　Ablation morphology of Kevlar 49 aramid fibers during in-situ impact of ultraviolet femtosecond laser

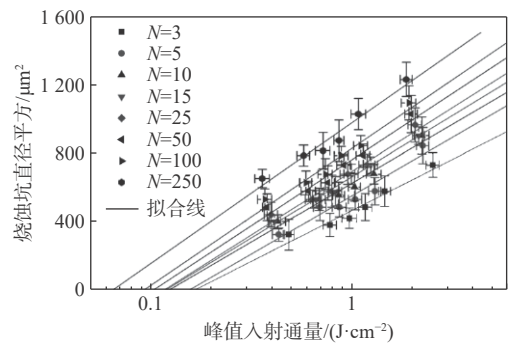


图 12　紫外飞秒脉冲作用下 Kevlar 49 芳纶纤维材料的烧蚀坑直径平方与入射通量关系

Fig. 12　Relationship between square of ablation pit diameter and incident flux of Kevlar 49 aramid fiber material under action of ultraviolet femtosecond pulses

结合图 12 中的实验数据与式(1)能够计算出, kevlar 49 芳纶纤维在不同入射脉冲数作用下的紫外飞秒激光去除阈值如表 3 所示。

由图 13 可知, Kevlar 49 芳纶纤维的紫外飞秒激光去除阈值也随激光入射脉冲数的增加而降低, 且阈值的降低速度同样呈现逐渐减小的趋势, 这表明, 当脉冲数增加到一定程度后芳纶纤维的阈值也会达到饱和。由图 13 中数据拟合计算可以得到, kevlar 49 纤维的紫外飞秒激光单脉冲去除阈值约为 0.21 J/cm^2, 孵化系数约为 0.80。

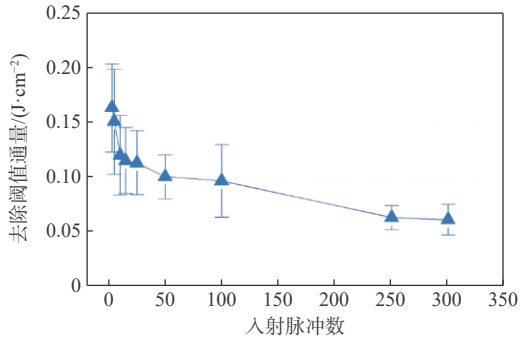
2.3　AFRP 整体的等效去除阈值与孵化效应

环氧树脂与芳纶纤维层合成型后的 Kevlar 49/4211 AFRP 基本参数如表 4 所示。

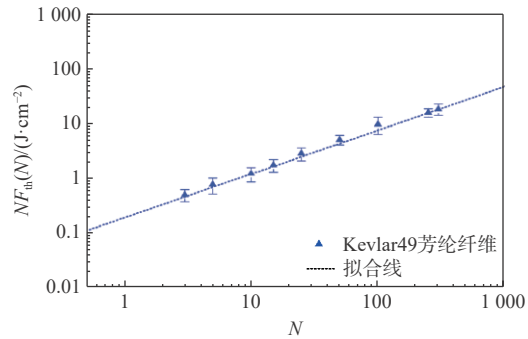
采用相同的紫外飞秒激光源和实验方法拟合计算 AFRP 作为整体时的等效去除阈值、单脉冲去除阈值和阈值孵化系数, AFRP 试片经激光原位叩击加工后的烧蚀坑形貌如图 14 所示。

测量 AFRP 的烧蚀坑孔径并采集数据后, 即可通过面积外延法拟合得到如图 15 和图 16 所示结果[18]。

根据图 15 中的实验数据及式(1)可知, Kevlar



(a) 去除阈值随入射激光脉冲数变化规律



(b) $NF_{th}(N)$随入射脉冲数目N变化关系

图 13 紫外飞秒激光作用下 Kevlar 49 芳纶纤维的去除阈值孵化效应

Fig. 13 Incubation effect of removal threshold on Kevlar 49 aramid fiber under UV femtosecond laser irradiation

表 3 Kevlar 49 芳纶纤维在不同激光脉冲数下的实际光斑直径与去除阈值通量

Table 3 Actual spot diameter and removal threshold flux of Kevlar 49 aramid fiber under different laser pulse numbers

N	d_s/μm	F_{th}/(J·cm^{-2})
3	22.89±0.81	0.17±0.04
5	24.23±1.06	0.15±0.05
10	24.34±1.22	0.13±0.04
15	25.43±1.14	0.12±0.03
25	26.41±1.05	0.12±0.02
50	25.92±1.30	0.11±0.03
100	27.32±1.22	0.10±0.02
250	27.57±1.47	0.06±0.01
300	27.68±1.52	0.06±0.01

表 4 AFRP 复合材料参数

Table 4 AFRP composite material parameters

厚度/mm	纤维质量分数/%	密度/(kg·m^{-3})	热导率/(W·(m·K)$^{-1}$)	拉伸强度/MPa	拉伸模量/GPa
1	65	1.3	<1	1 800	100

49/4211 复合材料在不同入射脉冲数作用下的紫外飞秒激光等效去除阈值如表 5 所示。

由图 16 可知，在紫外飞秒激光作用下，Kevlar 49/4211 的单脉冲等效去除阈值约为 0.38 J/cm^2，孵化系数约为 0.72。

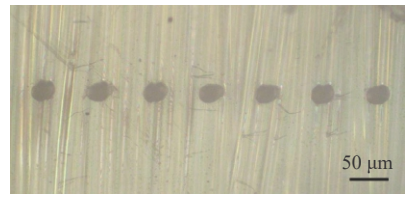


图 14 Kevlar 49/4211 AFRP 在紫外飞秒激光原位叩击时的烧蚀形貌

Fig. 14 Ablation morphology of Kevlar 49/4211 AFRP during in situ impact of ultraviolet femtosecond laser

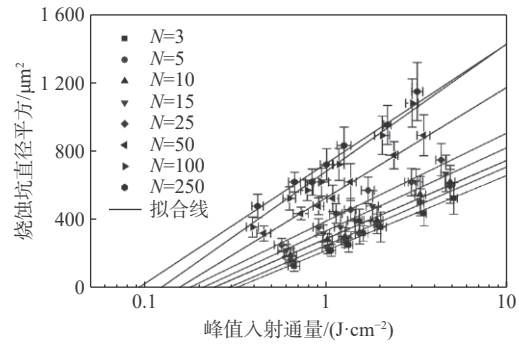
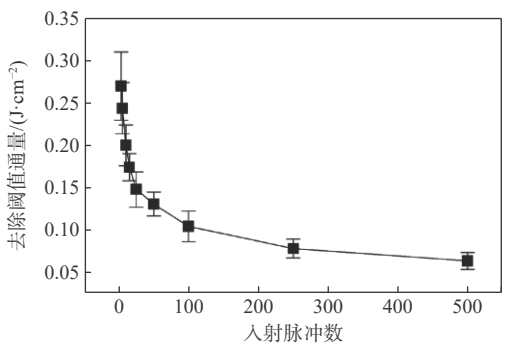
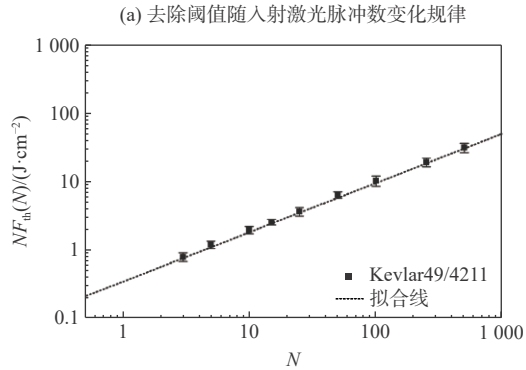


图 15 紫外飞秒激光作用下 Kevlar 49/4211 的烧蚀坑直径平方与入射通量关系[18]

Fig. 15 Relationship between square of ablation pit diameter and incident flux of Kevlar 49/4211 under action of ultraviolet femtosecond laser[18]



(a) 去除阈值随入射激光脉冲数变化规律



(b) $NF_{th}(N)$随入射脉冲数目N变化关系

图 16 紫外飞秒激光作用下 AFRP 等效去除阈值孵化效应[18]

Fig. 16 Incubation effect of AFRP equivalent removal threshold under the action of ultraviolet femtosecond laser[18]

2.4 AFRP 整体与各组分的加工特性关系

得到环氧树脂基体、Kevlar 纤维和 AFRP 整体

表 5　Kevlar49/4211 芳纶碳纤维在不同激光脉冲数下的实际光斑直径与去除阈值通量

Table 5　Actual spot diameter and removal threshold flux of Kevlar49/4211 aramid carbon fiber under different laser pulse numbers

N	d_s/μm	F_{th}/(J·cm^{-2})
3	19.56±1.75	0.29±0.04
5	17.63±1.60	0.27±0.04
10	20.01±2.41	0.22±0.03
15	20.62±1.77	0.19±0.03
25	21.20±2.42	0.16±0.02
50	23.75±4.63	0.14±0.02
100	25.55±6.01	0.12±0.02
250	24.73±5.29	0.09±0.02
500	25.61±6.67	0.07±0.01

的超快激光加工阈值特性后, 对比不同激光源作用下三者的加工去除阈值, 得到各组分的烧蚀阈值差异及其对 AFRP 整体等效阈值的影响规律。基体相、增强相与芳纶复合材料整体的去除阈值孵化曲线如图 17 所示。

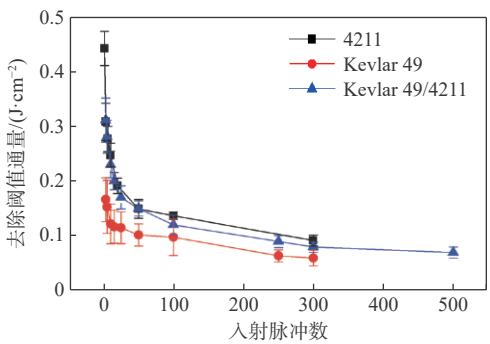


图 17　Kevlar 49/4211 AFRP 复合材料在紫外飞秒激光作用下各个组分的阈值对比

Fig. 17　Comparison of thresholds for various components of Kevlar 49/4211 AFRP composite materials under UV femtosecond laser irradiation

采用紫外飞秒激光加工 AFRP 时, Kevlar 芳纶纤维的去除阈值在不同脉冲数下始终低于 4211 环氧树脂, 并且树脂的阈值孵化系数同样小于芳纶纤维, 树脂的孵化效应更加明显, 从单脉冲到多脉冲的阈值下降幅度更大。AFRP 的等效去除阈值介于环氧树脂和 Kevlar 49 纤维之间。该结果体现了超快激光光场整形对 AFRP 加工特性的调控作用。

2.5　AFRP 的超快激光均质化去除方法

由于在实际工程中, 航天 AFRP 产品的加工需求以微孔制备和精密切割为主, 主要目标是实现 Kevlar 纤维与树脂基体的协同去除, 加工出壁面光滑边缘齐整均匀的微孔或切缝。因此, 本节基于 Kevlar 49/4211 芳纶纤维树脂基复合材料的去除阈值特性, 开展了相应的紫外飞秒激光准均质化去除方法研究。

从图 17 中可以看出, 虽然 Kevlar 49/4211 AFRP 中 4211 环氧树脂的紫外飞秒激光去除阈值始终高于 Kevlar 49 芳纶纤维, 但在多脉冲作用下二者之间的阈值差距相对较小。当采用紫外飞秒激光对 AFRP 进行原位钻孔和扫描切割加工时, 可以根据不同组分的阈值孵化效应曲线, 通过改变入射激光脉冲数调控芳纶纤维与环氧树脂材料之间的阈值差异。选取二者去除阈值差距相对较小的入射脉冲数和适当的激光通量与加工工艺参数, 便有望实现 AFRP 的准均质化去除。

2 种组分在紫外飞秒激光作用下的去除阈值如图 18 所示。

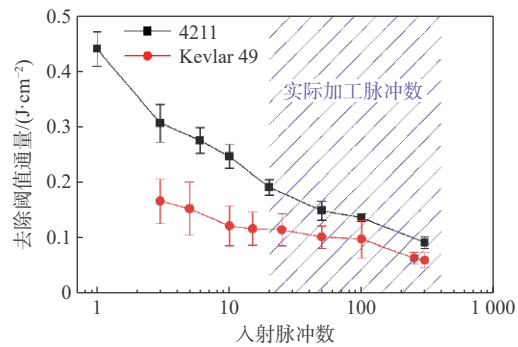


图 18　4211 环氧树脂与 Kevlar 49 芳纶纤维在紫外飞秒激光作用下的阈值孵化曲线

Fig. 18　Threshold incubation curves of 4211 epoxy resin and Kevlar 49 aramid fibers under UV femtosecond laser irradiation

由图 18 可知, 当入射脉冲数 N>100 时, Kevlar 49 纤维和 4211 环氧树脂的阈值相对接近。因此, 对于 AFRP 的紫外飞秒激光原位打孔和扫描切割加工, 将入射脉冲数控制在 100 左右时能够最大程度的令芳纶纤维与环氧树脂协同去除, 有效降低材料切口边缘的粗糙度从而提升加工精度和质量。因此, 选择 4211 环氧树脂阈值若干倍的激光通量, 入射通量 F 为 3.3 J/cm^2, 将入射脉冲数控制在 100 个左右进行 AFRP 的紫外飞秒激光原位打孔实验, 加工效果如图 19 所示。

对于脉冲激光的扫描切割加工, 通常利用激光的扫描次数、实际光斑半径、激光重频、扫描速度等参数推导作用于材料的等效入射脉冲数。计算等效入射脉冲数的经验公式为[19-21]

$$N_{\text{eff}} = \sqrt{\pi/2} n r_s f / v_s \tag{5}$$

式中: N_{eff} 为等效入射脉冲数; n 为扫描次数; r_s 为实际光斑半径; f 为激光重复频率; v_s 为扫描速度。将计算后的等效入射脉冲数控制在 100 个左右进行 Kevlar 49/4211 AFRP 的紫外飞秒激光扫描刻蚀实

验，入射通亮 F 为 3.3 J/cm^2，加工效果如图 20 和图 21 所示。

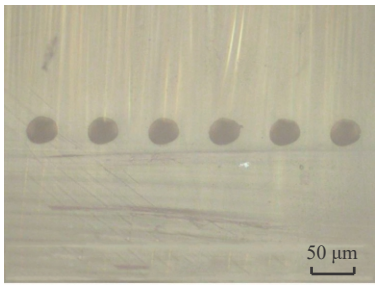


图 19 Kevlar 49/4211 在 100 个脉冲原位叩击下的去除效果

Fig. 19 Removal effect of Kevlar 49/4211 under 100 pulse in situ percussion

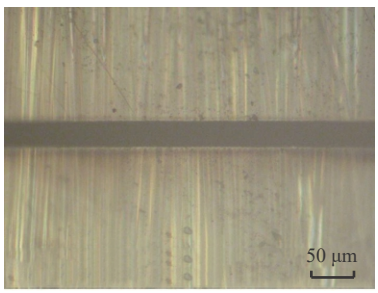


图 20 AFRP 在 100 等效脉冲作用下的扫描刻蚀效果

Fig. 20 Scanning etching effect of AFRP under 100 equivalent pulses

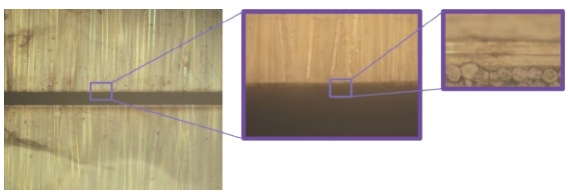


图 21 100 个等效紫外飞秒激光脉冲切割 AFRP 的切口剖面

Fig. 21 100 equivalent ultraviolet femtosecond laser pulses for cutting kerf profiles of AFRP

由图 20 和图 21 可知，采用紫外飞秒激光叩击或扫描 Kevlar 49/4211，能够在一定程度上将树脂基体与芳纶纤维接近协同去除。利用该参数切割的 AFRP 样件如图 22 所示。

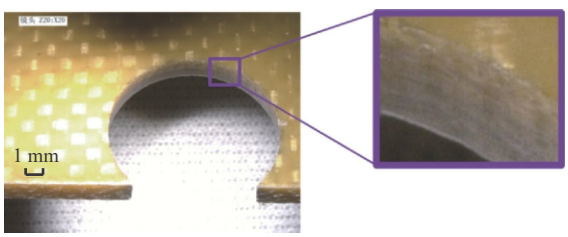


图 22 紫外飞秒激光在 AFRP 层合板上切割的通孔

Fig. 22 Through holes cut by ultraviolet femtosecond laser on AFRP laminates

由图 21 和图 22 可见，紫外飞秒激光加工 AFRP 样件的加工壁面均匀齐整，切割壁面光滑洁白，几乎不产生热损伤，且在约 1 mm 厚的 AFRP 层合板上加工单个 10 mm 直径的通孔仅需 10 s 左右。以 3.3 J/cm^2 入射通量和 100 个等效脉冲数的紫外飞秒激光加工 Kevlar 49/4211 AFRP 时，切口质量较好，材料表面齐整度较高，热影响区极小，且加工用时较短。证明了采用参数适当的紫外飞秒激光能够实现航天 AFRP 的精密减材加工，有效提高其加工质量和效率。

3 结 论

1）利用面积外延法定量计算了高分子环氧树脂、芳纶纤维和 AFRP 整体的紫外飞秒激光去除阈值通量。4211 环氧树脂的紫外飞秒激光去除阈值高于 Kevlar 49 芳纶纤维，AFRP 的等效加工阈值介于环氧树脂与芳纶纤维之间，表明 AFRP 的等效阈值受到树脂基体和增强纤维的共同影响。

2）通过孵化效应模型测算了树脂基体、增强纤维及 AFRP 整体的阈值孵化系数与单脉冲阈值。其中，4211 环氧树脂的孵化系数约为 0.73，单脉冲阈值约为 0.42 J/cm^2（与实验直接测出的单脉冲去除阈值 0.44 J/cm^2 相近）。Kevlar 49 芳纶纤维的单脉冲去除阈值约为 0.21 J/cm^2，孵化系数约为 0.80。Kevlar 49/4211 复合材料的单脉冲等效去除阈值约为 0.38 J/cm^2，孵化系数约为 0.72。

3）以材料的去除阈值为依据，选择各组分加工阈值相近的加工参数调控复合材料中树脂基体和芳纶纤维的去除同步性，能够实现 AFRP 的准均质化去除加工。对于 Kevlar 49/4211，选择通量为 4211 环氧树脂阈值若干倍的紫外飞秒激光，并采用 100 个等效脉冲以上原位叩击或循环扫描的方法，能够有效提高 AFRP 的加工边缘齐整度（提高尺寸精度）和加工效率，满足其均匀切割与精密制孔的航天制造需求。

参考文献（References）

[1] 石文静, 高峰, 柴洪友. 复合材料在航天器结构中的应用与展望[J]. 宇航材料工艺, 2019, 49(4): 1-6.
SHI W J, GAO F, CHAI H Y. Application and expectation of composite in spacecraft structure[J].Aerospace Materials & Technology, 2019, 49(4): 1-6(in Chinese).

[2] 陈昊宇, 殷莎, 胡建星, 等. 航天器冲击防护用热塑性仿生复合材料的弯曲性能研究[J]. 航天器环境工程, 2019, 36(2): 151-155.
CHEN H Y, YIN S, HU J X, et al. Bending properties of thermoplastic bioinspired helicoidal laminated composites used in impact protection of spacecraft[J].Spacecraft Environment Engineering, 2019, 36(2): 151-155(in Chinese).

[3] 石文天, 刘玉德, 张永安, 等. 芳纶纤维复合材料切削加工研究进展[J]. 表面技术, 2016, 45(1): 28-35.

SHI W T, LIU Y D, ZHANG Y A, et al. Research progress on the cutting process of aramid fiber composites[J]. Surface Technology, 2016, 45(1): 28-35(in Chinese).

[4] PADMANABHAM G, BATHE R. Laser materials processing for industrial applications[J]. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, 2018, 88(3): 359-374.

[5] AL-SULAIMAN F, YILBAS B S, KARAKAS F C, et al. Laser hole cutting in Kevlar: modeling and quality assessment[J]. The International Journal of Advanced Manufacturing Technology, 2008, 38(11): 1125-1136.

[6] YILBAS B S, AL-SULAIMAN F, KARAKAS C, et al. Laser cutting of multilayered kevlar plates[J]. Journal of Materials Engineering and Performance, 2007, 16(6): 663-671.

[7] RAO R V, SAVSANI V J, BALIC J. Teaching–learning-based optimization algorithm for unconstrained and constrained real-parameter optimization problems[J]. Engineering Optimization, 2012, 44(12): 1447-1462.

[8] MISHRA D R, BAJAJ A, BISHT R. Optimization of multiple kerf quality characteristics for cutting operation on carbon–basalt–Kevlar29 hybrid composite material using pulsed Nd: YAG laser using GRA[J]. CIRP Journal of Manufacturing Science and Technology, 2020, 30: 174-183.

[9] HIROGAKI K. Structural coloration of textiles: controlling structure of colloidal crystal array on fiber and textile[J]. Journal of Fiber Science and Technology, 2018, 74(12): 597-601.

[10] 王含妮, 谭勇, 张喜和. 芳纶纤维复合材料在脉冲激光作用下的热损伤研究[J]. 光电技术应用, 2015, 30(1): 26-31.
WANG H N, TAN Y, ZHANG X H. Research on thermal damage of aramid fiber composites acted by pulse laser[J]. Electro-Optic Technology Application, 2015, 30(1): 26-31(in Chinese).

[11] ZHANG K H, ZHANG X H, JIANG G, et al. Laser cutting of fiber-reinforced plastic laminate and its honeycomb sandwich structure[C]//Proceedings of the Advanced Laser Technology and Application. Bellingham: SPIE, 2020.

[12] 杨剑, 张瑞, 赵煜, 等. 皮秒激光切割 AFRP 复合材料实验研究[J]. 复合材料学报, 2022, 39(1): 147-157.
YANG J, ZHANG R, ZHAO Y, et al. Experimental study on picosecond laser cutting AFRP composites[J]. Acta Materiae Compositae Sinica, 2022, 39(1): 147-157(in Chinese).

[13] 路明雨, 张明, 张加波, 等. 激光加工非金属复合材料的研究与应用进展[J]. 宇航材料工艺, 2022, 52(1): 30-38.
LU M Y, ZHANG M, ZHANG J B, et al. Advances in research and application of laser processing of nonmetallic composites[J]. Aerospace Materials & Technology, 2022, 52(1): 30-38(in Chinese).

[14] WEN P G, ZHANG L W, ZHONG X Y, et al. Study on properties of 4211 epoxy resin system[J]. Journal of Aeronautical Materials, 2016, 36(2): 51-55.

[15] LIU J M. Simple technique for measurements of pulsed Gaussian-beam spot sizes[J]. Optics Letters, 1982, 7(5): 196-198.

[16] ROSENFELD A, LORENZ M, STOIAN R, et al. Ultrashort-laser-pulse damage threshold of transparent materials and the role of incubation[J]. Applied Physics A, 1999, 69(1): S373-S376.

[17] MANNION P T, MAGEE J, COYNE E, et al. The effect of damage accumulation behaviour on ablation thresholds and damage morphology in ultrafast laser micro-machining of common metals in air[J]. Applied Surface Science, 2004, 233(1-4): 275-287.

[18] 路明雨, 张明, 张开虎, 等. 高模量碳纤维增强树脂基复合材料的皮秒激光加工阈值特性[J]. 复合材料学报, 2021, 38(11): 3601-3609.
LU M Y, ZHANG M, ZHANG K H, et al. Threshold properties of high modulus carbon fiber reinforced plastic composite with picosecond laser processing[J]. Acta Materiae Compositae Sinica, 2021, 38(11): 3601-3609(in Chinese).

[19] CRAWFORD T H R, BOROWIEC A, HAUGEN H K. Femtosecond laser micromachining of grooves in silicon with 800nm pulses[J]. Applied Physics A, 2005, 80(8): 1717-1724.

[20] GUAY J M, CALÀ LESINA A, CÔTÉ G, et al. Laser-induced plasmonic colours on metals[J]. Nature Communications, 2017, 8: 16095.

[21] 李天舒, 王绍凯, 武清, 等. 宇航级 T800 碳纤维复合材料界面调控[J]. 北京航空航天大学学报, 2023, 49(8): 2011-2020.
LI T S, WANG S K, WU Q, et al. Interface adjustment of aerospace-grade T800 carbon fiber composite material[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(8): 2011-2020(in Chinese).

Characteristics and homogenization removal methods of ultraviolet femtosecond laser processing of aerospace AFRP

LU Mingyu[1]，ZHANG Ming[2]，WEI Yuxuan[3]，LI Bo[3]，CUI Zhigang[3]，ZHANG Kaihu[3,*]

(1. Beijing Institute of Spacecraft Environment Engineering，Beijing 100094，China；
2. China Academy of Space Technology，Beijing 100094，China；
3. Beijing Spacecraft，Beijing 100094，China)

Abstract：The UV femtosecond laser (343 nm, 260 fs) processing threshold and morphology characteristics of Kevlar 49/4211 aramid fiber/epoxy resin composite materials were studied. The area epitaxy method was used to calculate the laser removal threshold and threshold incubation effect of resin matrix, reinforced fibers, and the overall composite material in aramid fiber reinforced polymer (AFRP). The single pulse removal thresholds of the three were predicted; the impact of incident energy flux and equivalent number of input laser pulses on the quality of material processing edges was examined. The results showed that the UV femtosecond laser removal threshold and threshold incubation coefficient of Kevlar 49 aramid fiber were lower than those of 4211 epoxy resin under different pulse numbers, and the overall equivalent removal threshold of AFRP was between resin and aramid fiber. A processing morphology with uniform inlet width, neat edges, and quasi-homogeneous material removal can be achieved by choosing a UV femtosecond laser with an incident flux several times the threshold of 4211 epoxy resin and more than 100 equivalent pulse numbers and employing in-situ tapping or cyclic scanning cutting techniques. This effectively improves the processing accuracy and quality of aerospace AFRP products and satisfies their high-performance manufacturing needs.

Keywords：aerospace composite materials；aramid fiber；epoxy resin；ultraviolet femtosecond laser；remove threshold；incubation effect

Received：2024-02-26；**Accepted**：2024-04-03；**Published Online**：2024-04-30 18：31

URL：link.cnki.net/urlid/11.2625.V.20240430.1406.002

Foundation items：National Key Research and Development Program of China (2022YFB4601303); Key Program of the Joint Funds of the National Natural Science Foundation of China (U2037205)

* **Corresponding author**. E-mail：zhangkh86@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0099

一种高超飞行器用CFD网格不确定度分析方法

郭文娟，李强*，周岭
(中国空气动力研究与发展中心计算空气动力研究所，绵阳 621000)

摘　　要：针对高超飞行器计算流体力学(CFD)仿真中网格不确定度评估缺乏工程实用方法的难点问题，提出一种便于工程实际应用的CFD数据不确定度评估方法。所提方法采用Richardson插值实现高超飞行器用网格不确定度量化评估，同时建立高超飞行器仿真网格绘制规范，进而对4套网格计算得到的气动力结果进行了网格不确定度评估和分析。分析结果表明：所提方法在马赫数为4.95和7.95时，小迎角（低于8°）范围内，对乘波体飞行器气动力系数的网格不确定度评估性能良好。同时，湍流模型对所提方法影响不明显，故该高超飞行器所用CFD网格不确定度评估技术具有良好的工程适用性，对其他飞行器网格不确定度评估有一定的借鉴和指导价值。

关 键 词：乘波体；不确定度评估；Richardson模型；高超飞行器；计算流体力学
中图分类号：V221.3
文献标志码：A　　**文章编号**：1001-5965(2026)05-1567-11

近年来，随着计算机硬件的迅猛发展和仿真理论/算法的日臻成熟，计算流体力学(computational fluid dynamics, CFD)仿真已与风洞试验、理论分析并列成为飞行器改型和创新研制的主要手段。此外，随着CFD技术在高超飞行器研制中的应用占比日益增大，CFD数据可信度的重要性也日益凸显。CFD数据可信度研究已成为CFD仿真研究必不可少的部分。然而，CFD仿真引入了连续介质假设和理想气体方程等以简化实际物理现象，导致CFD仿真结果与真实物理数据间存在误差，故需系统评估CFD仿真结果的不确定度。

不确定度，即误差估计[1]，常用于评价真实值不确定或不可知时，计算值或试验值的误差情况。在宇航领域，CFD可信度评价一直受到国外政府和研究部门高度重视。文献[2]中明确指出不确定度的确定和控制是将来CFD计算所必备的基本能力。

CFD不确定度和误差研究也称作CFD验证与确认研究工作。Roy[3]全面综述了CFD验证与确认工作研究进展，深入分析了各种数值误差的区别和研究方法，并对数值离散误差进行了详细总结。陈鑫等[4]总结了CFD不确定度量化方法研究进展，详细探讨了CFD不确定性来源和表现形式及量化方法。

国内学者也广泛开展了CFD可信度评价研究。张涵信等[5]对CFD不确定度概念进行了研究。陈坚强等[6]采用Richardson插值法，研究了多套连续变化网格下T2-97模型表面压力数值解的空间离散误差和收敛性。陈江涛等[7]将数值离散和数值模型不确定度融合研究。上述研究丰富了学术界对CFD不确定性的认识。

CFD计算的误差产生于4个方面[8]：①物理建模误差：物理模型(控制方程)不能精确地描述物理现象，包括湍流模型和转捩模型带来的误差、简化气体状态方程、简化边界层和简化黏性项带来的误差；②数值离散误差：对控制方程离散化产生的误差，主要包括计算格式误差和网格离散误差；③舍入误差；④程序设计误差。③和④两个方面误差目前无法避免，而物理建模误差中简化气体方程和简

收稿日期：2024-02-28；录用日期：2024-07-02；网络出版时间：2024-07-05 13:34
网络出版地址：link.cnki.net/urlid/11.2625.V.20240704.1643.001
*通信作者. E-mail：409195707@qq.com

引用格式：郭文娟，李强，周岭. 一种高超飞行器用CFD网格不确定度分析方法[J]. 北京航空航天大学学报，2026，52(5)：1567-1577.
GUO W J，LI Q，ZHOU L. A CFD grid uncertainty analysis method for hypersonic aircraft[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1567-1577 (in Chinese).

化边界层的误差是目前对物理世界进行数学刻画所不能避免的，因此，在研究数值模拟的不确定度时，主要研究物理建模误差中湍流模型、转捩模型等带来的不确定度和数值离散误差中方程离散、计算网格带来的不确定度。 数值模拟的不确定度按研究方法分为2大类：①可参数化求解的不确定度，即数值模拟过程中一些参数无法精确得到，依赖经验，这类可用概率密度函数方法来求解，如模型参数不确定度[9-10]、几何外形的不确定度[11-12]、飞行来流条件引起的不确定度[13]；②无法参数化求解的不确定度，主要包括网格的不确定度[14-17]和数值模型的不确定度[18-19]。对于数值模拟过程中可参数化求解的不确定度国内外研究较多，但对无法参数化求解的不确定度研究相对匮乏，因此，对网格不确定度的研究十分必要。

计算网格的质量对数值模拟的结果影响是不可忽视的，目前，对计算网格不确定度的研究大致分为2类：通过获得较高精度的数值解从而求得误差和基于残差的方法。前者具有代表性的有网格细化方法、精度细化方法等。其中，网格细化方法是众多专家认为实际可行及推荐的。网格细化方法是利用在不同网格尺度下的计算结果，基于Richardson插值法获得插值网格解，从而给出数值计算误差的评估。国外一些学者，如Roache[20]和Roy等[21-22]在早期对此展开了一系列的研究，取得了不少研究成果。

目前，针对马赫数低于5飞行条件下CFD网格不确定度研究一般多将二维翼型或简单翼身融合体作为研究对象[7,23]，而对于马赫数高于5的超高声速飞行条件下CFD网格不确定度研究多采用简单典型高超声速问题模型如T2-97模型[6]开展不确定度规律研究。以上研究都基于简单模型且要求目标函数随网格加密严格收敛(即单调)，而在超高声速条件下高超飞行器CFD气动力数据往往不是严格平滑收敛，而是振荡收敛，马赫数越高，振荡收敛越明显，尤其是阻力由多成分组成，所以阻力系数随网格加密严格收敛是极难保证的，此时想达到目标函数随网格加密严格收敛(即单调)这个条件对网格要求比较严苛，不适合工程推广。而且当前对影响计算网格的不确定度的关键因素研究也不充分，如网格拓扑、湍流模型等因素对网格不确定度评估的影响。因此，本文提出一种基于Richardson插值方法的网格不确定度评估新方法，该方法与已有的研究方法相比放宽了对网格绘制和求解器精度的要求，不要求目标函数随网格加密严格收敛，这使得该方法会更适用于工程应用中网格不确定度评估，尤其是高超飞行器的网格不确定度评估。同时，鉴于网格质量对气动力计算结果影响明显，故本文制定了一套网格加密的规范，按照该规范绘制了指定飞行器的4套网格用于计算本文需要的气动力系数。本文在马赫数为4.95和7.95条件下，对乘波体飞行器的气动力系数的网格不确定度进行了深入的评估分析，并深入探讨了湍流模型等对网格不确定度的影响，为工程上超高马赫数条件下其他飞行器气动力系数的网格不确定度评估提供了参考。

1 高超飞行器用CFD网格不确定度分析方法建立

目前，已有的基于Richardson插值方法的网格离散不确定度评估方法均要求目标函数随网格加密严格收敛[6-7,23]。由于超高马赫数时气动力系数容易振荡收敛，满足目标函数随网格加密严格收敛这一条件对网格质量和求解器精度就极为严苛，极大地增加了网格不确定度评估工作量。因此，现有基于Richardson插值方法的CFD网格不确定度研究多采用简单模型，仅能在范围比较小的情况下分析网格不确定度评估规律，难以应用于较高马赫数复杂外形飞行器。本文结合Richardson插值分析原理，并考虑了目标函数随网格不严格收敛的情况，发展了一种高超飞行器用CFD网格不确定度分析方法，可有效降低对CFD网格划分的要求，进而极大地降低评估工作难度和时间。

Richardson插值原理就是将目标函数进行Taylor展开，假定目标函数的离散解满足k阶展开：

$$f_i = f_0 + \alpha h_i^k \tag{1}$$

式中：f_i为目标函数在第i套网格上的解；f_0为理想中目标函数的真值；$h_i \approx N_i^{-1/3}$为第i套网格的特征尺度，N_i为第i套网格量，网格细化比$R_i = h_i + 1/h_i$，为方便工程应用，不对R_i做过多要求；α为未知系数；k为目标函数的收敛精度。

式(1)中有3个未知数f_0、α、k，理论上只需要3套网格上目标函数的数值结果就能计算得到所有未知量的值。本文设置k的取值范围为$-30 \leqslant k \leqslant 30$(30为本文采用值，可根据具体拟合数据的特征进行调整)，若计算得到实际收敛精度k在1附近时，实际收敛精度与理论收敛精度差异不大，认为用式(1)拟合得到的误差是合理的；当k值偏离1较多时，实际收敛精度与理论收敛精度差异可采用其他更恰当的拟合式。本文使用CFD软件NNW-Flowstar的二阶空间离散格式进行仿真。当实际收

敛速度 $0.5\leqslant k\leqslant 2.1$ 时, 通过式(1)进行离散误差估计是合理的, 应用于工程中该取值范围也可适当放宽。当 $-30<k<0.5$ 时, 数值误差估计过大, 估计过于保守。当 $2.1<k<30$ 时, 数值误差估计过小。这 2 种情况下应使用其他拟合式。之前关于 Rechardson 插值方法计算网格不确定度的研究只应用于目标函数随网格加密严格收敛的情况, 但是在高马赫数条件下, 气动力系数往往存在振荡收敛现象, 此时需满足目标函数随网格加密严格收敛这一条件对网格质量和求解器精度十分严苛, 不利于工程应用。本文研究发现当目标函数随网格加密严格收敛时, 收敛精度 k 值不会偏离 0 很多, 本文研究结果中采用 $|k|<30$, 而目标函数不随网格加密单调收敛时, $|k|$取极值 30, 所以在本文中将$|k|$取极值情况(即不要求目标函数随网格加密严格收敛)也考虑进来, 将 $k=-30$ 条件下拟合式的选取等同于$-30<k<0.5$ 条件, 将 $k=30$ 条件下拟合式的选取等同于 $2.1<k<30$ 条件。

因为 NNW-Flowstar 软件的空间离散格式为二阶, 因此, 可采用一阶、二阶和一阶/二阶混合方程进行目标函数拟合, 采用以下数学形式。

一阶拟合为

$$f_i = f_0 + \alpha h_i \tag{2}$$

二阶拟合为

$$f_i = f_0 + \beta h_i^2 \tag{3}$$

式中: β 为未知系数。

混合一阶/二阶拟合为

$$f_i = f_0 + \alpha h_i + \beta h_i^2 \tag{4}$$

若计算软件精度更高, 可采用更高阶的拟合式, 式(1)～式(4)给出了 4 种目标函数拟合式, 对应的误差(即目标函数与其真值的标准差)如下:

$$\begin{cases} \sigma_k = \sqrt{\dfrac{\sum\limits_{i=1}^{n_g}\left[f_i-\left(f_0+\alpha h_i^k\right)\right]^2}{n_g-3}} \\ \sigma_1 = \sqrt{\dfrac{\sum\limits_{i=1}^{n_g}\left[f_i-\left(f_0+\alpha h_i\right)\right]^2}{n_g-2}} \\ \sigma_2 = \sqrt{\dfrac{\sum\limits_{i=1}^{n_g}\left[f_i-\left(f_0+\beta h_i^2\right)\right]^2}{n_g-2}} \\ \sigma_{12} = \sqrt{\dfrac{\sum\limits_{i=1}^{n_g}\left[f_i-\left(f_0+\alpha h_i+\beta h_i^2\right)\right]^2}{n_g-3}} \end{cases} \tag{5}$$

式中: n_g 为网格套数, 当只有 3 套网格时, σ_k 和 σ_{12} 值为 0。

本文采用的方法为: 根据式(1)拟合得到实际收敛精度 k, 当 $0.5\leqslant k\leqslant 2.1$ 时, 目标函数按式(1)展开拟合; 当 $2.1<k\leqslant 30$ 时, 将目标函数按一阶和二阶拟合式中标准差较小的拟合; 当 $-30\leqslant k<0.5$ 时, 将目标函数按一阶、二阶和混合一阶/二阶拟合式中标准差最小的拟合。根据选定的目标函数拟合式进行回归求解, 得到目标函数的真值 f_0、拟合值 F_i 和拟合标准差。

当实际收敛精度 $0.5\leqslant k\leqslant 2.1$, 并且拟合标准差 σ 小于目标函数随网格变化平均差量时, 误差估计 Σ 是比较可靠的, 如式(6)所示, 对误差结果引入安全因子 h_s 取为 1.5(由于计算得到的标准差可能太小, 可采用加安全因子的方法加以补偿, 即用标准差乘安全因子[24], 本文重点考察 4 套网格的情况, 故安全因子取值保守情况可取 1.5～2)。当拟合标准差 σ 不小于目标函数随网格变化平均差量时, 误差估计可靠性较低一些, 安全因子 h_s 此时应取更大的值来保证置信度(本文取为 3), 在工程应用中为保证置信度还可以根据拟合数据的实际特征将安全因子取更大的值。

$$\sigma < \Sigma = \frac{f_{i\max}-f_{i\min}}{n_g-1} \tag{6}$$

目标函数 95% 置信区间如下:

$$\left[f_i-\Sigma_c, f_i+\Sigma_c\right] \tag{7}$$

当 $\sigma<\Sigma$ 时, 对置信区间进行修正, 可得

$$\Sigma_c = h_s|f_i-f_0|+\sigma+|f_i-F_i| \tag{8}$$

当$\sigma\geqslant\Sigma$时, 对置信区间进行修正, 可得

$$\Sigma_c = \frac{h_s\sigma}{\Sigma}(|f_i-f_0|+\sigma+|f_i-F_i|) \tag{9}$$

最终目标函数的置信区间显示在密网格上:

$$\left[f_1-\Sigma_c, f_1+\Sigma_c\right] \tag{10}$$

Σ_c 即为目标函数的网格不确定度。现有研究在进行上述拟合过程中会进行加权最小二乘拟合, 即对式(5)进行加权, 文献 [25] 发现在没有明显异常值的拟合问题中, 只用等权值拟合方法即可。本文在评估网格不确定度之前对气动力系数与试验值(真值)进行了校验, 都符合工程计算要求, 没有异常值存在, 所以采取等权值最小二乘拟合即可。

本文方法相对传统的网格不确定评估方法扩宽了适用范围, 降低了网格质量的要求, 更具有工程适用性。

本节方法可归纳为如图 1 所示的流程。

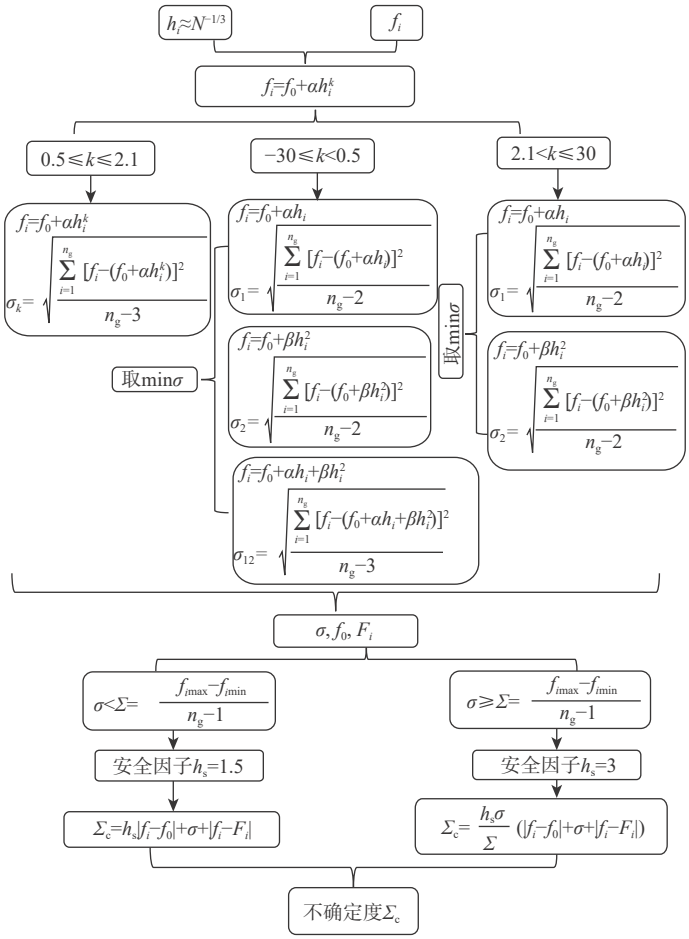


图 1 高超飞行器用 CFD 网格不确定度评估流程

Fig. 1 Flow of CFD grid uncertainty evaluation for hypersonic aircraft

2 案例设计与建模策略

2.1 高超飞行器案例设计

本文以具有水平和垂直尾舵的典型乘波体模型为案例验证本文方法的有效性。试验模型主要由机头/机身尾段组成，机身和机翼融合为一体，机翼尾段有左右水平舵，机身尾部上表面正中有立尾，如图 2 所示。

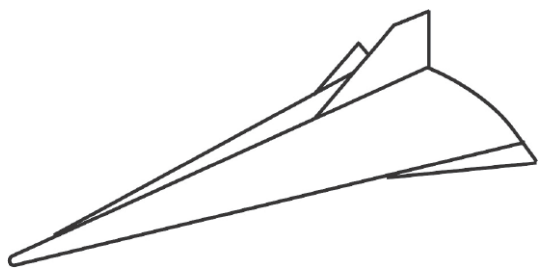


图 2 乘波体结构示意图

Fig. 2 Schematic diagram of wave rider structure

本文选取的计算条件与状态参数如表 1 所示。其中，纵向气动导数在迎角为 0°～8°范围内采用最小二乘法求取。

2.2 高超飞行器用 CFD 网格生成规范

为验证本文方法的正确性，需排除个人绘制网格习惯对网格质量的影响，使网格数量成为影响气动力计算结果的唯一因素。本文建立了如下的高超飞行器网格离散规范和策略。

步骤 1 先绘制一套中等密度的网格，对中等密度网格的要求如下：①第 1 层网格高度使 Y^+ 约等于 1；②每个方向的网格增长率不超过 1.25；③翼前缘和翼后缘的网格流向长度不超过当地弦长的 0.1%；④翼根处和翼梢处网格展向长度不超过半展长的 0.1%；⑤接近飞行器头部和机身后部的网格长度约为参考长度的 2%；⑥中等密度网格对飞行器阻力的预测精度必须满足工程要求。

步骤 2 在中等密度网格基础上绘制粗网格，Y^+

表 1 飞行器的计算条件

Table 1 Calculation conditions for aircraft

计算条件	来流马赫数	单位雷诺数$Re/10^7$ m^{-1}	力矩系数参考长度L /m	气动系数参考面积S_{ref} /m^2	参考点X_G	迎角/(°)
条件1	4.95	6.416	0.335 29	0.002 491 4	0	0、2、6、8
条件2	7.95	3.016	0.335 29	0.002 491 4	0	0、2、6、8

约等于中等密度网格 Y^+的 1.5 倍(此时 $Y^+=1.5$), 每条网格线的网格点数量为原本的 2/3。

步骤 3　在中等密度网格基础上绘制细网格, Y^+约等于中等密度网格 Y^+的 2/3(此时 $Y^+=2/3$), 每条网格线的网格点数量为原本的 1.5 倍。

步骤 4　若需要一套更粗的网格, 就在已绘制的粗网格基础上绘制, 绘制原理与步骤 2 相同, 只是将基础网格由中等密度的换为粗网格; 同理若需要更细的一套网格, 就在已绘制的细网格基础上按照步骤 3 原理进行加密。以此类推, 可获得需要的网格套数。

采用上述策略, 本文绘制了一组(4 套)乘波体飞行器网格, 具体信息如表 2 所示。该绘制规范要求并不严格, 在保证中等密度网格质量符合基本工程要求的同时, 也方便快速按要求绘制, 便于应用于工程计算。

表 2　4 套网格规模

Table 2　Scale of four grid sets

网格	网格量N_i	网格的特征尺度h_i
H1	1 196 800	0.009 418 700
H2	2 196 736	0.007 692 616
H3	4 409 728	0.006 098 121
H4	16 825 344	0.003 902 522

2.3　仿真工具选择和验证试验数据来源

本文的数值模拟计算工作基于 2.2 节所述的网格绘制策略, 使用了 4 套结构网格。由于求解的是高超声速状态, 风洞试验结果中的洞壁和支撑干扰可以忽略。计算设置的远场边界条件、气动力和流场计算采用中国空气动力研究与发展中心自主研发并免费开放的软件 NNW-Flowstar 程序, 在气动中心的工作站集群上进行。NNW-Flowstar 软件在空间上采用有限体积法离散, 在时间步上采用 LU-SGS 隐式方法。该软件在各类型号的气动任务中得到了广泛应用, 具有良好的工程应用背景。

为验证气动力计算数据的合理性, 本文采用了乘波体模型的气动力试验数据, 该数据来源于中国空气研究与发展中心的 FL-31 风洞。

3　不确定度评估方法正确性校验与分析

本节首先验证了网格不确定度分析的输入合理性, 然后, 使用本文方法对上述指定飞行器的 4 套网格计算所得的气动力结果进行了详细的网格不确定度分析。

3.1　输入校验

本节对网格不确定度分析的输入进行了研究, 输入为指定飞行器的气动力系数。为输入的合理性, 本文将马赫数为 4.95 和 7.95 条件下, 基于第 2 套粗网 H2 的气动力系数数值计算结果与高超声速风洞 FL-31 的试验数据进行对比。表 3 和表 4 分别为典型数据升力系数斜率和不同迎角的纵向压心系数 X_p 的对比。由表 3 可知, 计算和试验结果的升力系数斜率差值小于 1%, 纵向压心系数差值也都小于 3%。由表 4 可知, 计算和试验结果的升力系数斜率差值小于 2%, 纵向压心系数差值也都小于 1%。这些结果表明: 绘制的网格拓扑和质量符合计算的基本要求, 为下一步分析提供了保障。

表 3　*Ma* 为 4.95 时 H2 计算结果与试验结果对比

Table 3　Comparison between H2 calculation results and experimental results at *Ma*=4.95

结果对比	升力系数斜率C_L'	纵向压心系数X_p		
		迎角为2°	迎角为6°	迎角为8°
试验结果	0.180 05	0.618 11	0.620 77	0.622 09
计算结果	0.179 45	0.601 10	0.609 30	0.610 80

表 4　*Ma* 为 7.95 时 H2 计算结果与试验结果对比

Table 4　Comparison between H2 calculation results and experimental results at *Ma*=7.95

结果对比	升力系数斜率C_L'	纵向压心系数X_p		
		迎角为2°	迎角为6°	迎角为8°
试验结果	0.144 30	0.591 29	0.598 71	0.601 18
计算结果	0.141 53	0.586 80	0.596 70	0.599 20

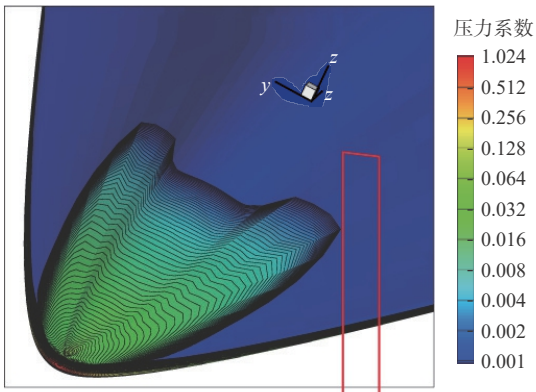
3.2　马赫数为 4.95 条件下不确定度评估分析

图 3 为马赫数为 4.95、迎角为 6°条件下, 网格 H1～H4 计算得到的乘波体头部流场压力图。图中, 红色框线是辅助观察的标记。由图可知, 随着网格加密, 激波强度逐渐增强, 激波变窄且边缘更加清晰。流场图随网格密度的变化明显改变, 因此, 在马赫数为 4.95 条件下, 乘波体外形对网格密度是敏感的。

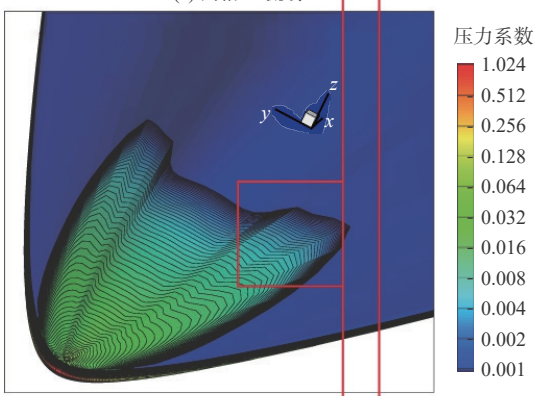
图 4～图 8 为气动力系数随网格量增加的变化。图中, SST 表示采用 *k-ω* SST 湍流模型计算的结果, SA 表示采用 SA 湍流模型计算的结果。

由图 4 和图 5 可知, 在所有条件下, 升力系数和俯仰力矩系数均随网格加密严格收敛; 由图 7 和图 8 可知, 在绝大部分条件下, 激波阻力系数和摩擦阻力系数均随网格加密严格收敛。然而, 激波阻力系数随网格量增加单调减小, 而摩擦阻力系数随网格量增加整体呈增加趋势(在较高马赫数条件下, 摩擦阻力系数本身很小, 收敛较慢且易于振荡收敛, 因此, 摩擦阻力系数要保持单调增加较难)。

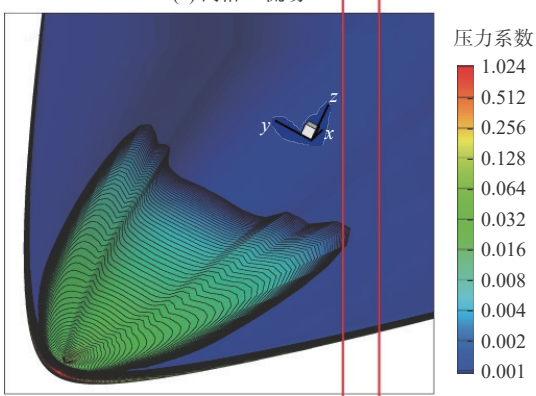
总阻力系数是激波阻力系数和摩擦阻力系数之和, 由于二者的单调变化规律不一致, 总阻力系数随网格量变化很难保持严格单调。实际结果也



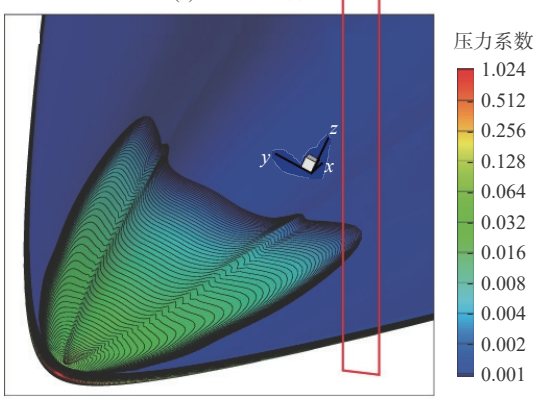
(a) 网格H1流场



(b) 网格H2流场



(c) 网格H3流场



(d) 网格H4流场

图 3 *Ma*=4.95 时乘波体头部压力流场随网格量变化

Fig. 3 Pressure flow field at head of wave rider at *Ma*=4.95 as function of grid size

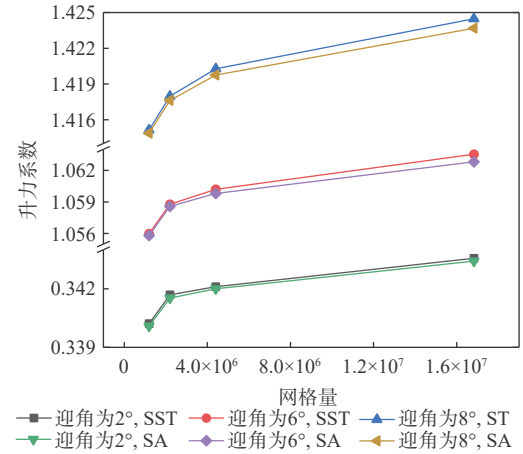


图 4 *Ma*=4.95 时升力系数随网格量变化

Fig. 4 Lift coefficient variation with grid quantity at *Ma*=4.95

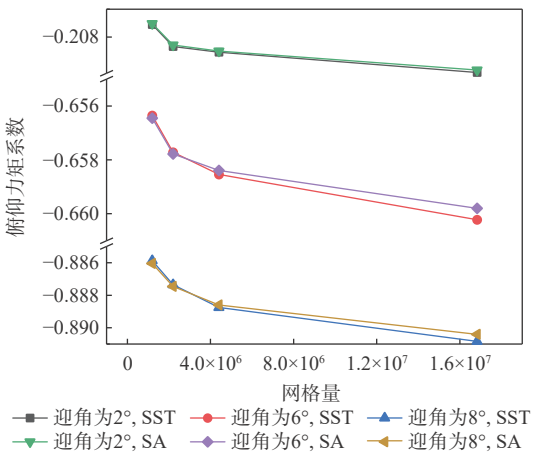


图 5 *Ma*=4.95 时俯仰力矩系数随网格量变化

Fig. 5 Pitch moment coefficient variation with grid quantity at *Ma*=4.95

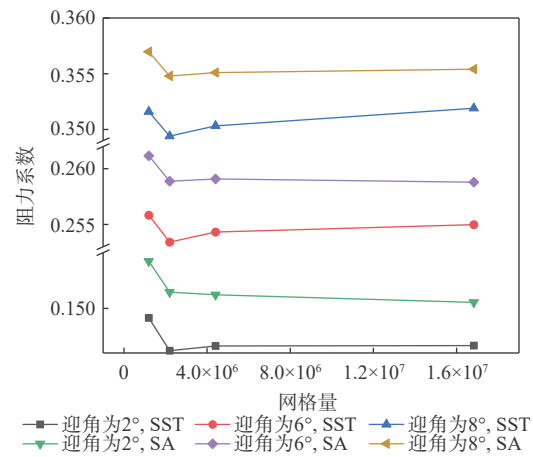


图 6 *Ma*=4.95 时阻力系数随网格量变化

Fig. 6 Resistance coefficient variation with grid quantity at *Ma*=4.95

确实如此，如图 6 所示，除在 SA 湍流模型迎角为 2°的条件下，总阻力系数随网格单调变化外，其他条件下的总阻力系数均不单调。

图 9～图 11 为马赫数为 4.95 条件下，不同迎角的气动力系数不确定度。图中，SST 表示使用 *k*-*ω* SST 湍流模型的计算结果，SA 表示使用 SA 湍流模

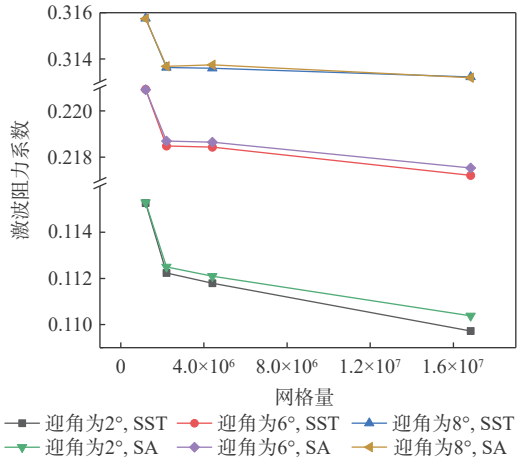


图 7　*Ma*=4.95 时激波阻力系数随网格量变化

Fig. 7　Shock drag coefficient variation with grid quantity at *Ma*=4.95

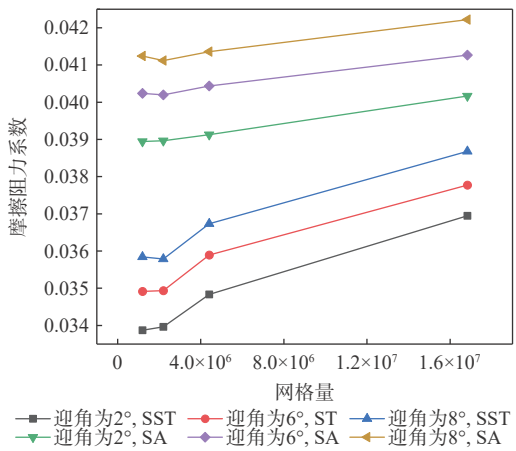


图 8　*Ma*=4.95 时摩擦阻力系数随网格量变化

Fig. 8　Variation of friction resistance coefficient with grid quantity at *Ma*=4.95

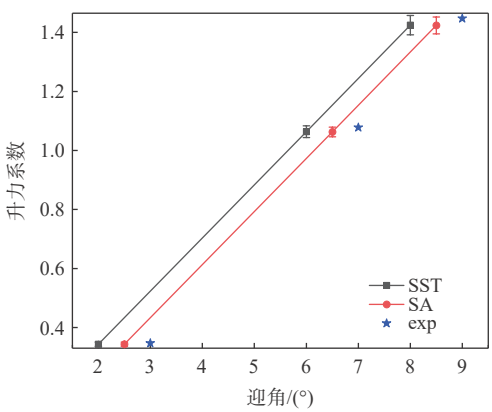


图 9　*Ma*=4.95 时升力系数不确定度

Fig. 9　Lift coefficient uncertainty at *Ma*=4.95

型的计算结果, exp 表示试验值。误差杆表示不确定度。为避免线条交叠影响观察, 绘图时将 SA 模型的横坐标向右平移 0.5°, 试验值向右平移 1°。

由图 9 和图 10 可知, 在马赫数为 4.95 条件下, SST 和 SA 湍流模型计算的升力系数和俯仰力矩系

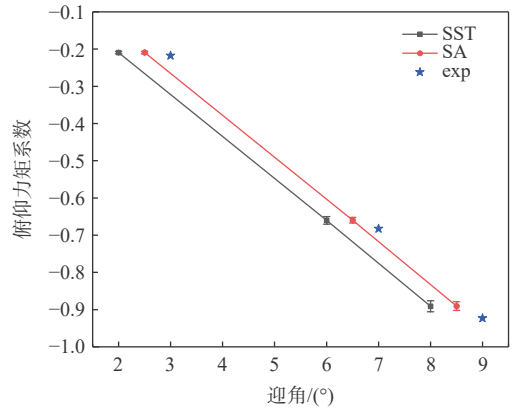


图 10　*Ma*=4.95 时俯仰力矩系数不确定度

Fig. 10　Uncertainty in the pitch moment coefficient at *Ma*=4.95

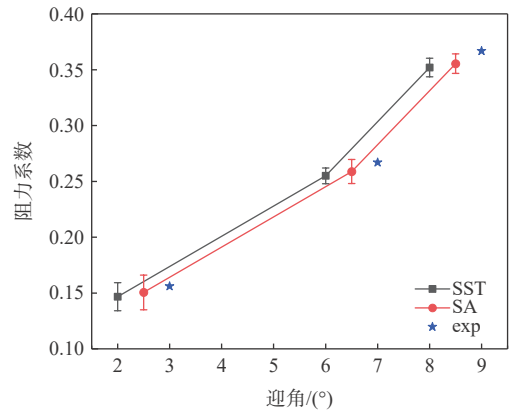


图 11　*Ma*=4.95 时阻力系数不确定度

Fig. 11　Uncertainty in the drag coefficient at *Ma*=4.95

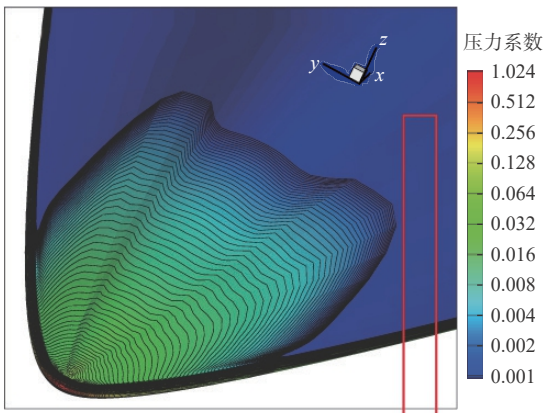
数与试验值吻合较好。SST 和 SA 湍流模型计算的升力系数和俯仰力矩系数在所有迎角处的不确定度十分相近, 说明在马赫数 4.95 条件下, 湍流模型对升力系数和俯仰力矩系数的不确定度影响很小。

由图 11 可知, 在马赫数为 4.95 条件下, SST 和 SA 湍流模型计算的阻力系数与试验值吻合较好, 误差在可接受范围内。SST 和 SA 湍流模型计算的阻力系数在 6°和 8°迎角处的不确定度十分相近, 说明在马赫数为 4.95 条件下, 湍流模型对阻力系数的不确定度影响有限。而在 2°迎角处, SST 模型计算的阻力系数随网格变化非单调, 而 SA 模型的结果是单调的, 但两者的不确定度没有显著差异。因此, 在马赫数为 4.95 条件下, 阻力系数随网格量变化的单调性不影响不确定度评估。

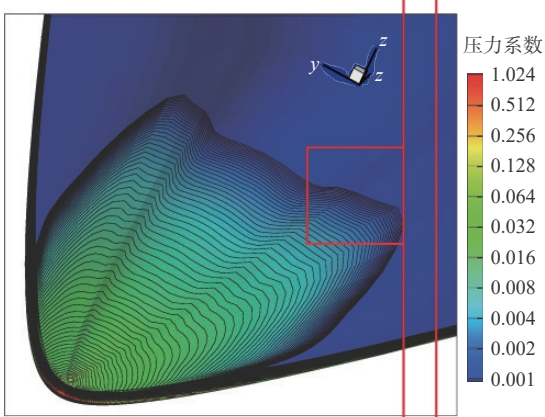
综上所述, 在马赫数为 4.95 条件下, 湍流模型对各气动力系数的网格不确定度评估影响不显著, 阻力系数随网格量变化的单调性对不确定度评估影响也不明显。本文方法在马赫数为 4.95 条件下具有良好的工程适用性。

3.3　马赫数为 7.95 条件下不确定度评估分析

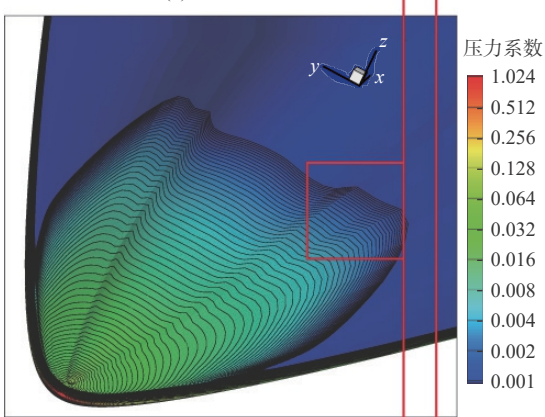
图 12 为马赫数为 7.95、迎角为 6°条件下, 网格



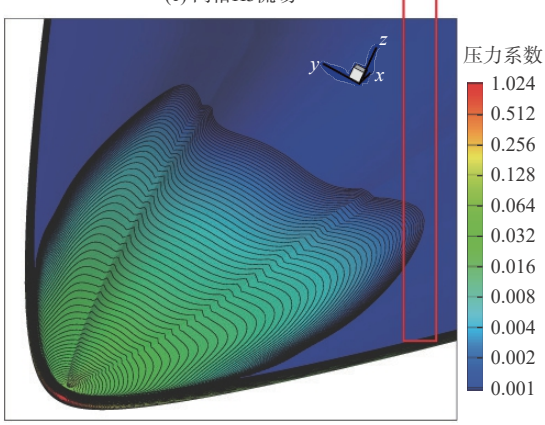
(a) 网格H1流场



(b) 网格H2流场



(c) 网格H3流场



(d) 网格H4流场

图 12 *Ma*=7.95 时乘波体头部压力流场随网格量变化

Fig. 12 Pressure flow field at head of wave rider at *Ma*=7.95 as function of grid size

H1～H4 的乘波体头部流场压力分布。在图中辅助参考线的帮助下，可以清晰地看出，随着网格的加密，激波的强度逐渐增强，激波变得更窄，边缘更加清晰。流场图随网格密度的变化有显著改变，因此，在马赫数为 7.95 的条件下，乘波体的外形对网格密度是敏感的。

图 13～图 17 为气动力系数随网格数量增加的变化情况，所有数据均在马赫数为 7.95 的条件下进行，Lim 表示层流模型。由图 13 和图 14 可知，升力系数和俯仰力矩系数在所有条件下均随着网格的增加严格单调收敛；由图 16 和图 17 可知，激波阻力系数在大多数条件下也随着网格的增加严格收敛，且单调减少。然而，在马赫数为 7.95 的条件下，摩擦阻力系数的收敛比低马赫数条件更困难，由于

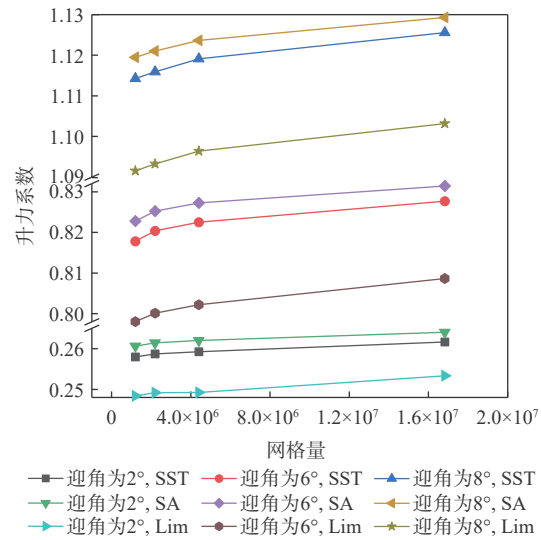


图 13 *Ma*=7.95 时升力系数随网格量变化

Fig. 13 Lift coefficient variation with grid quantity at *Ma*=7.95

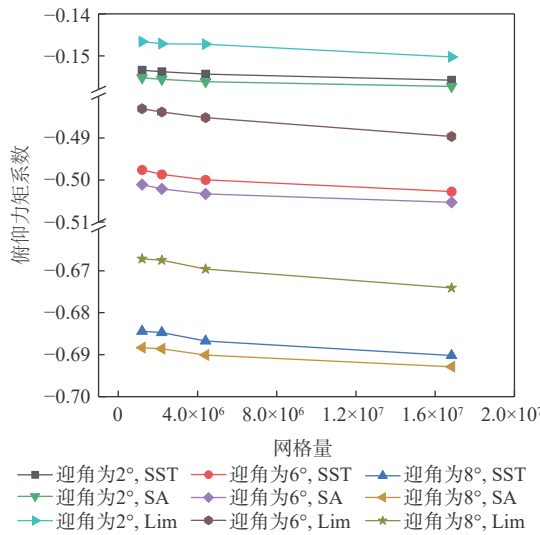


图 14 *Ma*=为 7.95 时俯仰力矩系数随网格量变化

Fig. 14 Pitch moment coefficient variation with grid quantity at *Ma*=7.95

其数值本身很小, 容易在小范围内波动, 但整体随网格增加呈现增加趋势(在高马赫数条件下, 摩擦阻力系数本身就很小, 收敛较慢且容易振荡, 因此, 摩擦阻力系数难以随网格数量单调增加)。总阻力为激波阻力系数和摩擦阻力系数之和, 由于激波阻力系数和摩擦阻力系数随网格量变化规律不一致, 且摩擦阻力系数易波动, 总阻力系数随网格数量变化难以保持严格单调。如图 15 所示, 除 SA 湍流模型在迎角为 2°条件下的阻力系数随网格数量严格单调变化外, 其他条件下的阻力系数均不单调。

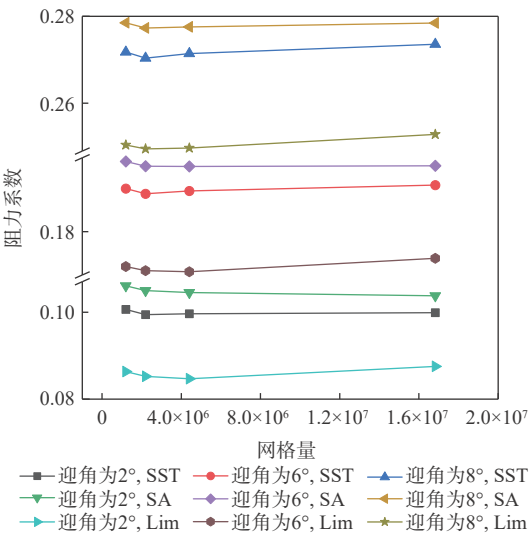


图 15　*Ma*=7.95 时阻力系数随网格量变化

Fig. 15　Resistance coefficient variation with grid quantity at *Ma*=7.95

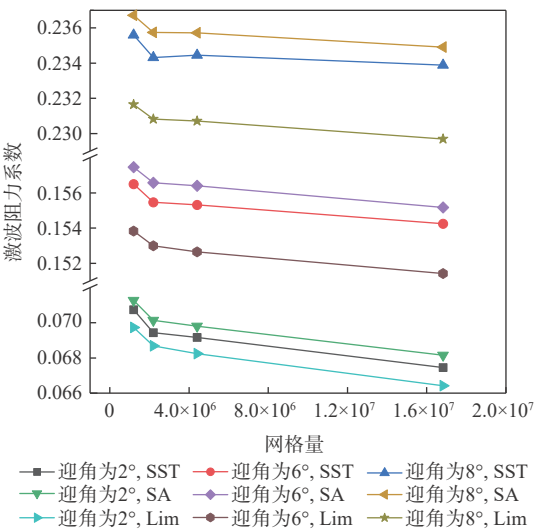


图 16　*Ma*=7.95 时激波阻力系数随网格量变化

Fig. 16　Shock drag coefficient variation with grid quantity at *Ma*=7.95

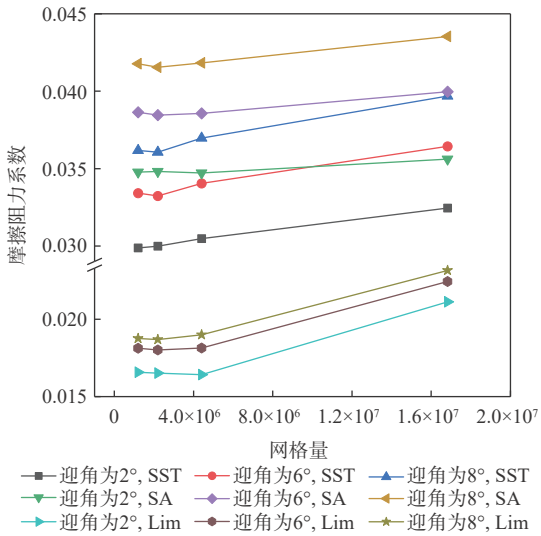


图 17　*Ma*=7.95 时摩擦阻力系数随网格量变化

Fig. 17　Variation of friction resistance coefficient with grid quantity at *Ma*=7.95

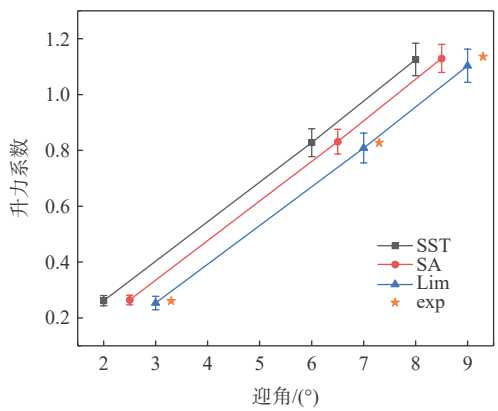


图 18　*Ma*=7.95 时升力系数不确定度

Fig. 18　Lift coefficient uncertainty at *Ma*=7.95

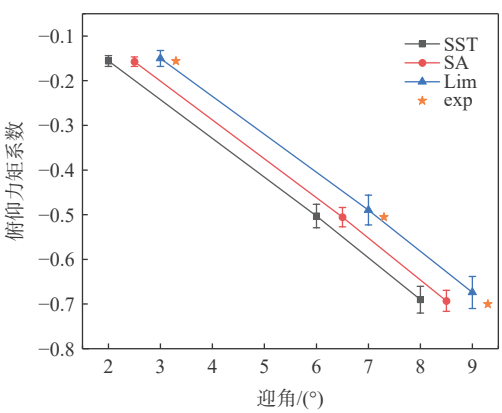


图 19　*Ma*=7.95 时俯仰力矩系数不确定度

Fig. 19　Uncertainty in the pitch moment coefficient at *Ma*=7.95

图 18～图 20 为马赫数为 7.95 条件下, 不同迎角的气动力系数的不确定度。图中, Lim 表示层流模型, 其他符号含义与前文描述一致。为避免线条重叠影响观察, 作图时对 SA 模型的横坐标平移 0.5°迎角, 对层流模型平移 1°迎角, 对试验值平移 1.3°迎角。

由图 18 和图 19 可知, 在马赫数为 7.95 条件下, SST、SA 湍流模型和层流模型计算得到的升力

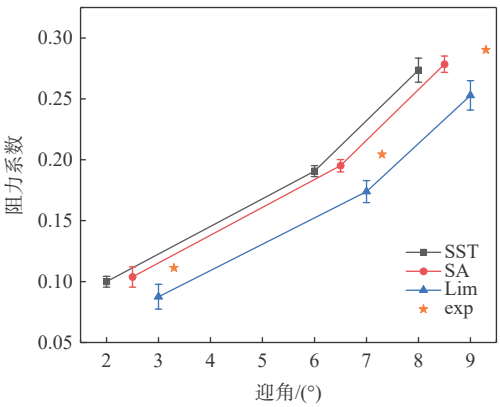


图 20 *Ma*=7.95 时阻力系数不确定度

Fig. 20 Uncertainty in the drag coefficient at *Ma*=7.95

系数和俯仰力矩系数与试验值吻合较好。由于此时雷诺数较大为 $1.011\,23\times10^7$，因此，湍流模型的计算结果相比层流模型与试验值更吻合。3 个模型计算得到的升力系数和俯仰力矩系数在所有迎角处的不确定度都非常相近，说明在马赫数为 7.95 条件下，湍流和层流模型对升力系数和俯仰力矩系数的不确定度影响很小。

由图 20 可知，在马赫数为 7.95 条件下，3 个模型计算得到的阻力系数与试验值的误差在可接受范围内。SST 和 SA 湍流模型及层流模型计算的阻力系数在 6°和 8°迎角处的不确定度差异不明显，表明在此条件下，湍流和层流模型对阻力系数的不确定度影响有限。在 2°迎角处，SST 模型和层流模型计算的阻力系数随网格量变化是非单调的，SA 模型的结果是单调的，但 SA 的不确定度介于 SST 模型和层流模型之间。在 2°迎角处，SST 模型、SA 模型和层流模型的阻力系数不确定度分别占自身阻力系数的 5%、8% 和 10%，三者区别不大，因此，在马赫数为 7.95 条件下，阻力系数随网格量变化的单调性不影响不确定度评估。

综合以上分析可知，在马赫数为 7.95 条件下，湍流模型对各气动力系数网格不确定度评估的影响并不显著，阻力系数随网格量变化的单调性对不确定度评估的影响也不明显。本文方法同样适用于马赫数为 7.95 条件下的工程问题。

4 结 论

本文提出一种基于 Richardson 插值方法的网格不确定度评估方法，并使用该方法评估了乘波体模型在马赫数为 4.95 和 7.95 及不同湍流模型条件下的网格不确定度，得到了如下结论：

1）由于阻力系数由激波阻力系数和摩擦阻力系数组成，而激波阻力系数与摩擦阻力系数随网格量变化的趋势不同，所以乘波体阻力系数随网格量变化难以单调收敛。然而，乘波体阻力系数随网格量变化是否严格收敛对网格不确定度评估的影响不明显。这说明本文使用的网格绘制策略和网格不确定度评估方法在高马赫数条件下具有良好的适用性，并且该方法放宽了对网格绘制和求解器精度的要求，适合工程推广使用。

2）在马赫数为 4.95 和 7.95 的条件下，湍流模型对气动数据的网格不确定度评估影响不明显。因此，本文方法的鲁棒性良好，在高马赫数条件下不受湍流模型等因素的影响。

接下来，将进一步探讨本文方法对其他类型飞行器的适用性及其关键影响因素，以发展更具工程应用性的 CFD 不确定度评估方法。

参考文献（References）

[1] ROACHE P J. Verification of codes and calculations[J]. AIAA Journal, 1998, 36(5): 696-702.

[2] SLOTNICK J P, KHODADOUST A, ALONSO J J, et al. CFD vision 2030 study: a path to revolutionary computational aerosciences: NF1676L-18332 [R]. Washington, D. C. : NASA, 2014.

[3] ROY C J. Review of code and solution verification procedures for computational simulation[J]. Journal of Computational Physics, 2005, 205(1): 131-156.

[4] 陈鑫, 王刚, 叶正寅, 等. CFD 不确定度量化方法研究综述[J]. 空气动力学学报, 2021, 39(4): 1-13.
CHEN X, WANG G, YE Z Y, et al. A review of uncertainty quantification methods for computational fluid dynamics[J]. Acta Aerodynamica Sinica, 2021, 39(4): 1-13(in Chinese).

[5] 张涵信, 查俊. 关于 CFD 验证确认中的不确定度和真值估算[J]. 空气动力学学报, 2010, 28(1): 39-45.
ZHANG H X, ZHA J. The uncertainty and truth-value assessment in the verification and validation of CFD[J]. Acta Aerodynamica Sinica, 2010, 28(1): 39-45(in Chinese).

[6] 陈坚强, 张益荣. 基于 Richardson 插值法的 CFD 验证和确认方法的研究[J]. 空气动力学学报, 2012, 30(2): 176-183.
CHEN J Q, ZHANG Y R. Verification and validation in CFD based on the Richardson extrapolation method[J]. Acta Aerodynamica Sinica, 2012, 30(2): 176-183(in Chinese).

[7] 陈江涛, 章超, 吴晓军, 等. 考虑数值离散误差的湍流模型选择引入的不确定度量化[J]. 航空学报, 2021, 42(9): 226-237.
CHEN J T, ZHANG C, WU X J, et al. Quantification of turbulence model-selection uncertainties considering discretization errors[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(9): 226-237(in Chinese).

[8] OBERKAMPF W L, BLOTTNER F G. Issues in computational fluid dynamics code verification and validation[J]. AIAA Journal, 1998, 36: 687-695.

[9] 向星皓, 张毅锋, 陈坚强, 等. 横流转捩模型参数不确定度量化分析与应用研究[J]. 宇航学报, 2020, 41(9): 1141-1150.
XIANG X H, ZHANG Y F, CHEN J Q, et al. Uncertainty quantification analysis and application research on cross-flow transition model parameters[J]. Journal of Astronautics, 2020, 41(9): 1141-1150 (in Chinese).

[10] 章超, 刘骁, 陈江涛, 等. 烧蚀热响应计算中的不确定性传播分析

方法研究[J]. 宇航学报, 2020, 41(11): 1401-1409.
ZHANG C, LIU X, CHEN J T, et al. Study on uncertainty propagation analysis method in ablative thermal response calculation[J]. Journal of Astronautics, 2020, 41(11): 1401-1409(in Chinese).

[11] LOEVEN A, BIJL H. Airfoil analysis with uncertain geometry using the probabilistic collocation method[C]//Proceedings of the 49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston : AIAA, 2008.

[12] PARUSSINI L, PEDIRODA V, POLONI C. Prediction of geometric uncertainty effects on fluid dynamics by polynomial chaos and fictitious domain method[J]. Computers & Fluids, 2010, 39(1): 137-151.

[13] MARIOTTI A, SALVETTI M V, SHOEIBI OMRANI P, et al. Stochastic analysis of the impact of freestream conditions on the aerodynamics of a rectangular 5: 1 cylinder[J]. Computers & Fluids, 2016, 136: 170-192.

[14] RICHARDSON L F. The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam[J]. Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character, 1911, 210(459-470): 307-357.

[15] CELIK I, KARATEKIN O. Numerical experiments on application of Richardson extrapolation with nonuniform grids[J]. Journal of Fluids Engineering, 1997, 119(3): 584-590.

[16] CELIK I B, GHIA U, ROACHE P J, et al. Procedure for estimation and reporting of uncertainty due to discretization in CFD applications[J]. Journal of fluids Engineering-Transactions of the ASME, 2008, 130(7): 078001.

[17] SCHAEFER J A, HOSDER S, MANI M, et al. The effect of grid topology and flow solver on turbulence model closure coefficient uncertainties for a transonic airfoil[C]//Proceedings of the 46th AIAA Fluid Dynamics Conference. Reston : AIAA, 2016.

[18] XIAO H, CINNELLA P. Quantification of model uncertainty in RANS simulations: a review[J]. Progress in Aerospace Sciences, 2019, 108: 1-31.

[19] DURAISAMY K, IACCARINO G, XIAO H. Turbulence modeling in the age of data[J]. Annual Review of Fluid Mechanics, 2019, 51: 357-377.

[20] ROACHE P J. Perspective: a method for uniform reporting of grid refinement studies[J]. Journal of Fluids Engineering, 1994, 116(3): 405-413.

[21] ROY C, OBERKAMPF W. A complete framework for verification, validation, and uncertainty quantification in scientific computing [C]//Proceedings of the 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition. Reston : AIAA, 2010.

[22] ROY C J. Grid convergence error analysis for mixed-order numerical schemes[J]. AIAA Journal, 2003, 41(4): 595-604.

[23] 左玲玉, 司海青, 李耀, 等. 基于 Richardson 外推法的 CFD 离散误差分析[J]. 指挥控制与仿真, 2022, 44(1): 58-62.
ZUO L Y, SI H Q, LI Y, et al. Analysis of CFD discrete error based on Richardson extrapolation method[J]. Command Control & Simulation, 2022, 44(1): 58-62(in Chinese).

[24] 罗刚. 不确定度 A 类评定及不确定度 B 类评定的探讨[J]. 计量与测试技术, 2007, 34(12): 42-43.
LUO G. Serches of the uncertainty a and the uncertainty B[J]. Metrology & Measurement Technique, 2007, 34(12): 42-43(in Chinese).

[25] TILO S. 数据拟合与不确定度: 加权最小二乘拟合及其推广[M]. 王鼎, 唐涛, 尹洁昕, 等, 译. 北京: 国防工业出版社, 2019.
TILO S. Data fitting and uncertainty: weighted least squares fitting and its generalizations[M]. WANG D, TANG T, YIN J X, et al, translated. Beijing: National Defense Industry Press, 2019(in Chinese).

A CFD grid uncertainty analysis method for hypersonic aircraft

GUO Wenjuan，LI Qiang*，ZHOU Ling

(Institute of Computational Aerodynamics，China Aerodynamics Research and Development Center，Mianyang 621000，China)

Abstract：Addressing the challenge of the lack of practical engineering methods for grid uncertainty assessment in computational fluid dynamics (CFD) simulations of hypersonic vehicles, this paper proposes a CFD data uncertainty assessment method that is convenient for practical engineering applications. The method utilizes Richardson interpolation to quantitatively assess grid uncertainty for hypersonic vehicles, and establishes a grid drawing specification for hypersonic vehicle simulations. Subsequently, grid uncertainty assessment and analysis are conducted on aerodynamic results obtained from four sets of grid calculations. The analysis results indicate that the proposed method performs well in assessing grid uncertainty for aerodynamic coefficients of waverider vehicles within a small angle of attack (below 8°) at Mach numbers of 4.95 and 7.95. Additionally, the turbulence model has an insignificant impact on this grid uncertainty assessment method. Therefore, the CFD grid uncertainty assessment technology used for hypersonic vehicles exhibits good engineering applicability and provides valuable reference and guidance for grid uncertainty assessment in other aircraft.

Keywords：wave rider; uncertainty assessment; Richardson model; advanced aircraft; computational fluid dynamics

Received：2024-02-28；**Accepted**：2024-07-02；**Published Online**：2024-07-05 13：34

URL：link.cnki.net/urlid/11.2625.V.20240704.1643.001

* **Corresponding author**. E-mail：409195707@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0112

高隐身飞翼布局翼型敏感性分析及优化设计

张伟[1]，周琳[1]，陈宪[1]，舒博文[2]，黄江涛[1,*]，高正红[2]
(1. 中国空气动力研究与发展中心，绵阳 621000；2. 西北工业大学 航空学院，西安 710072)

摘 要：高隐身飞翼布局必须满足宽频段隐身设计要求，传统气动外形隐身设计主要面向高频光学区，针对谐振区翼型气动外形隐身设计的研究较少，无法为飞翼布局气动外形宽频隐身设计提供有效的指导。开展飞翼布局翼型宽频隐身设计研究，利用基本效应法开展了不同频段下翼型外形参数敏感性研究，发现高、低频段隐身设计与气动设计要求对翼型气动外形敏感性区域各不相同，不同频段隐身设计存在复杂的矛盾关系。在此基础上，提出飞翼布局宽频隐身设计模型，以NACA65,3-018翼型为基础开展考虑宽频隐身的气动优化设计。结果表明：所提模型可以提高飞翼布局翼型气动性能，同时实现宽频段高隐身特性。

关 键 词：高隐身飞翼布局；宽频隐身；敏感性分析；基本效应法；优化设计
中图分类号：V211.3
文献标志码：A **文章编号**：1001-5965(2026)05-1578-09

相较于常规布局，飞翼布局[1-5]由于翼-身高度融合的布局特征，其翼型的作用对于布局影响更加突出。由于尾翼的缺失，除传统翼型设计的升阻特性要求以外，飞翼布局翼型设计还需要考虑俯仰力矩要求[6]。此外，翼型的隐身设计是飞翼布局雷达散射截面(radar cross section, RCS)减缩的重要手段之一[7-9]。对于先进飞机而言，隐身特性关系到飞机的生存能力和任务效能，因此，考虑隐身特性的高隐身飞翼布局翼型设计受到广泛关注。

针对高隐身飞翼布局翼型设计，Vinh等[10]研究了翼型几何参数对气动隐身特性的影响，并在此基础上开展了翼型的气动隐身一体化设计。Bondeson等[11]采用基于矩量法连续伴随方程的梯度计算方法对翼型进行了隐身及气动、隐身一体化优化。张彬乾等[7]针对某飞翼布局飞机分别进行了内翼段和外翼段翼型的气动隐身一体化设计，并分析了不同站位处翼型气动与隐身矛盾关系。陈曦等[12]利用雷诺平均Navier-Stokes方程及物理光学法开展了气动、隐身综合优化设计研究。

目前对飞行器外形RCS减缩机理的研究集中在光学区，随着未来隐身向全频、全向的目标发展，飞行器谐振区的散射特性愈发重要。目前针对飞行器宽频隐身设计方面的研究较少，周琳[13]针对飞翼布局翼型在入射频率500 MHz和9 GHz下的气动隐身综合设计问题展开研究，并指出了相位相消是谐振区RCS减缩的主要原因，但是由于其并未考虑谐振区和光学区内不同频率对气动外形要求的差异，因此，难以对飞行器宽频隐身设计提供有力的支撑。本文针对飞行器在隐身光学区和谐振区的气动外形设计问题开展研究，利用基本效应法[14]分析了光学区和谐振区不同频率对气动外形敏感性，以及对气动外形的要求，总结了不同频率隐身设计的矛盾和协同关系，进而建立了考虑宽频隐身的多学科设计模型，以NACA 65,3-018翼型[5]为基础开展了考虑宽频隐身的气动外形设计。结果表明，优化的高隐身飞翼布局翼型在满足纵向力矩配平要求和高升阻比要求的同时，实现了谐振区和光学区前向RCS特性的提升，为飞行器宽频隐身设计提供了参考。

收稿日期：2024-02-29；录用日期：2024-06-28；网络出版时间：2024-09-12 08:46
网络出版地址：link.cnki.net/urlid/11.2625.V.20240911.1330.001
*通信作者. E-mail：hjtcfd@163.com

引用格式：张伟，周琳，陈宪，等. 高隐身飞翼布局翼型敏感性分析及优化设计[J]. 北京航空航天大学学报，2026，52(5)：1578-1586. ZHANG W，ZHOU L，CHEN X，et al. Sensitivity analysis and optimization design of high stealth flying wing layout airfoil[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1578-1586 (in Chinese).

1　基本方法

1.1　基本效应法

本文采用全局敏感性分析方法-基本效应法(effect of element, EE)分析翼型的气动、隐身特性对其外形的敏感性[14-15]。该方法通过式(1)计算任一维设计变量X_i对响应值Y的影响[16]:

$$E_{Ei}=(f(X_1,X_2,\cdots,X_i+\Delta_i,\cdots,X_n)-f(X_1,X_2,\cdots,X_i,\cdots,X_n))/\Delta_i \tag{1}$$

式中: Δ_i为任一维设计变量的单位尺度变化量。

设计变量全局敏感度信息S_{Ai}计算式为[14]

$$S_{Ai}=\left(\sum_{i=1}^{r}|E_{Ei}|\right)\Big/m \tag{2}$$

式中: m为随机取样次数。

1.2　二维矩量法算例验证

本文采用二维矩量法[17]计算翼型的 RCS 特性, 其雷达散射宽度σ'计算式为[15]

$$\sigma'=\lim_{r\to\infty}2\pi r\frac{|\boldsymbol{E}^{s}|^2}{|\boldsymbol{E}^{i}|^2}=\lim_{\rho\to\infty}2\pi r\frac{|\boldsymbol{H}^{s}|^2}{|\boldsymbol{H}^{i}|^2} \tag{3}$$

式中: r为柱体到散射观察点的距离; $\boldsymbol{E}^{s}$为散射电场; $\boldsymbol{H}^{s}$为散射磁场; $\boldsymbol{E}^{i}$为入射电场; $\boldsymbol{H}^{i}$为入射磁场。

三维情况下, 雷达散射截面σ与σ'之间关系如下[17]:

$$\sigma=2l^2\sigma'/\lambda \qquad l\gg\lambda \tag{4}$$

式中: l为目标长度; λ为入射波长。

为验证本文应用的二维矩量法的可靠性, 下面以 NACA 65, 3-018 翼型为例, 并与商用软件 FEKO 进行计算对比, 其计算模型如图 1 所示。计算入射频率分别为 3 GHZ、5 GHZ 和 7 GHZ, 入射角度为 −90°～90°。FEKO 采用三维矩量法, 本文二维矩量法的计算结果经式 (4) 转化为三维结果。

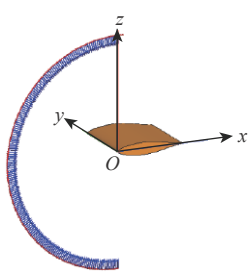


图 1　FEKO 三维计算模型

Fig. 1　3D analysis model of FEKO

由图 2 可知, 二维矩量法程序在不同频率下与 FEKO 软件的计算结果吻合良好, 能够较为精确捕捉波峰特性, 因此, 计算精度满足要求。

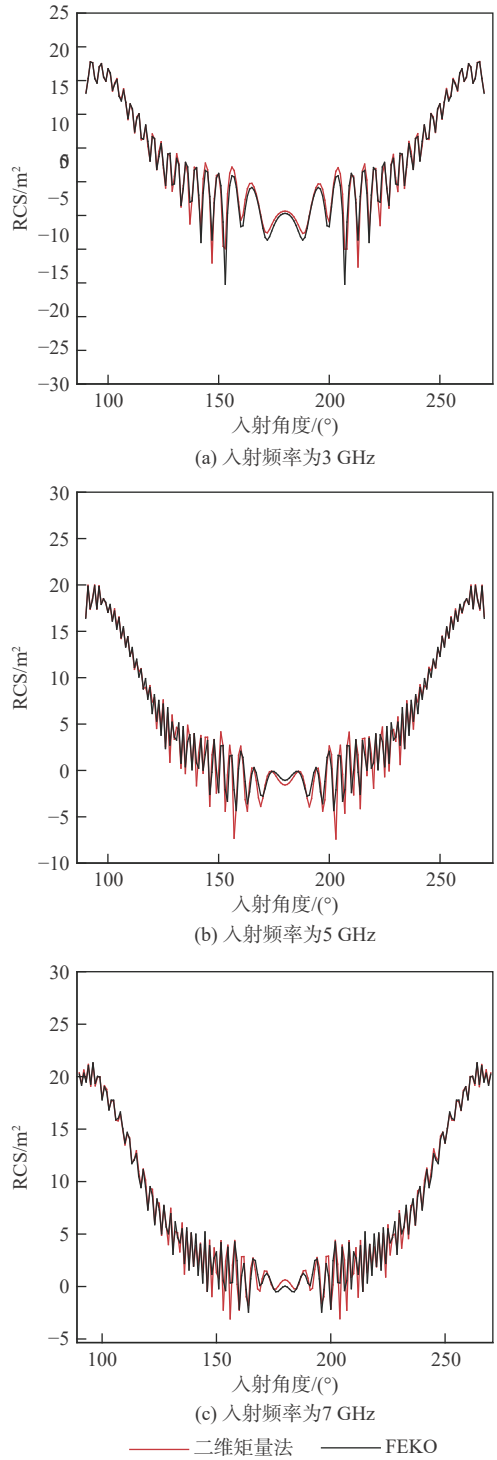


图 2　FEKO 与二维矩量法计算结果对比

Fig. 2　Results comparison of FEKO and 2D moment method calculations

2　设计变量敏感性

2.1　设计变量敏感性对比

根据目标尺寸与入射波长的相对关系, 目标的雷达散射可分为瑞利区、谐振区和光学区。图 3 为以金属球的电磁散射为例, 说明这 3 种电磁散射的特征, 其中, k 为入射波的波数(k=2π/λ), a 为球半径, σ_{spere}为球体的雷达散射截面, $\sigma_{spere}=\pi a^2$。

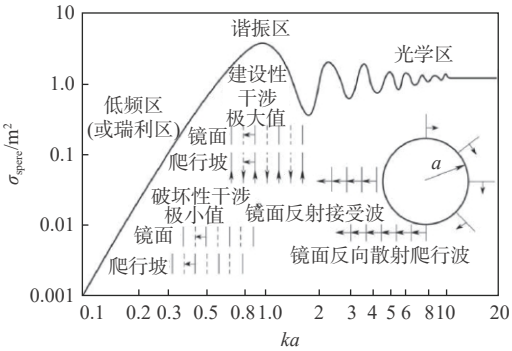


图3 金属球单站RCS随频率变化[18]

Fig. 3 Metal ball single station RCS variation with frequency plot[18]

对于飞行器而言，瑞利区的入射波长通常远大于目标尺寸(ka<1)，此时目标RCS与入射频率的4次方成正比，因此，飞行器外形隐身设计一般不考虑瑞利区散射。在谐振区，入射波长与目标尺寸往往在同一个量级(1<ka<10)，导致入射场沿目标长度上的相位变化很显著，因此，其散射特性比较重要。在光学区的入射波长一般远小于目标尺寸(ka>10)，此时，目标的散射场与局部几何形状密切相关，在计算时须考虑目标细节的几何结构[13]。

本文以相对厚度为18%的对称翼型NACA 65,3-018为基础[19]，开展设计变量关于气动、隐身目标特性敏感性研究，气动设计状态为$Ma=0.70$，升力系数$C_l=0.25$，$Re=2\times10^7$[20]，隐身特性入射频率分别为300 MHz、600 MHz、900 MHz、1 GHz、3 GHz、6 GHz、9 GHz，入射角度为前向±30°，入射角θ示意如图4所示。

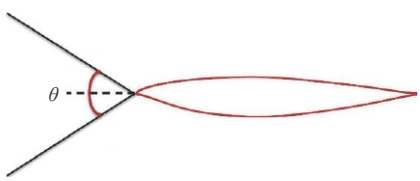


图4 入射角度示意图

Fig. 4 Schematic diagram of incident angle

本文采用CFL3D程序[21]计算翼型的气动特性，隐身特性计算采用二维矩量法[15,17]。采用Bspline方法[22]对翼型进行参数化，设计变量数为上下表面各6个，共12个设计变量，每个维度取值范围为[−0.02,0.02]，设计空间如图5所示，图中，横坐标x/c为x方向单位尺度，纵坐标y/c为y方向单位尺度。

首先，利用基本效应法开展翼型的阻力系数C_d、前向RCS均值关于其气动外形的敏感性研究，结果如图6和图7所示。图中，横轴D1～D6为上表面设计变量，后D7～D12为下表面设计变量。由图6可知，对于阻力特性而言，翼型上表面外形对

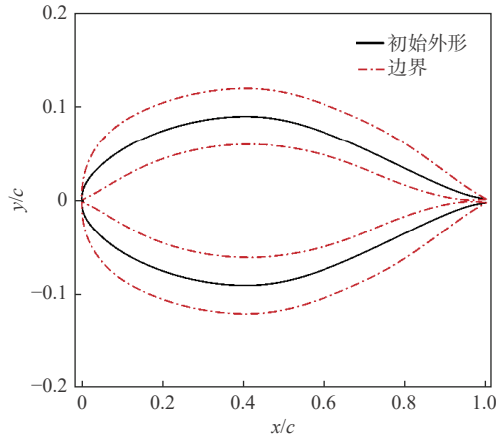


图5 初始设计空间

Fig. 5 Initial design space

阻力特性的敏感性更强；而对于隐身特性，翼型光学区(1～9 GHz)和谐振区(300～900 MHz)隐身特性对外形的要求截然不同，图7(a)～图7(c)为谐振区设计变量敏感性结果，图7(d)～图7(f)为光学区敏感性对比。可以看出，翼型光学区隐身特性主要由翼型前缘外形决定，因此，翼型上、下表面第1个设计变量的影响远大于其他变量；相较于光学区，翼型谐振区的隐身特性对外形的要求更为复杂，当入射频率f=300 MHz时，翼型整体外形对RCS特性影响都很大，其中，前缘和中部区域的外形较为突出；当f=600 MHz时，翼型的中部和后缘对RCS特性的影响更为显著；当f=900 MHz时，翼型的RCS特性主要由前缘和后缘的外形特征决定。综上所述，翼型谐振区RCS特性对气动外形要求不但与光学区大为不同，并且谐振区内不同频率之间对气动外形要求也有明显差异。

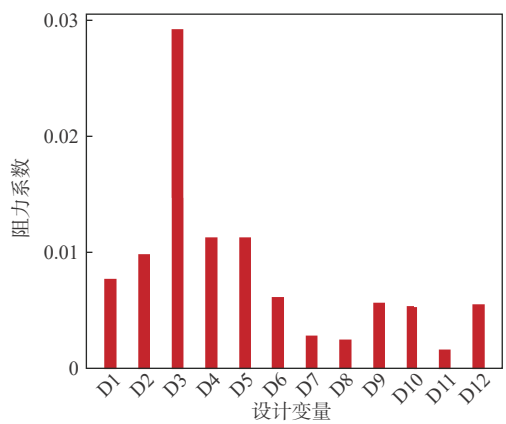


图6 设计变量关于C_d敏感性

Fig. 6 Sensitivity to C_d of design variables

2.2 不同频率设计结果对比

本节继续以NACA 65,3-018翼型为基础，分别开展不同频率下翼型隐身外形优化设计，隐身特性入射频率分别为300 MHz、600 MHz、900 MHz、

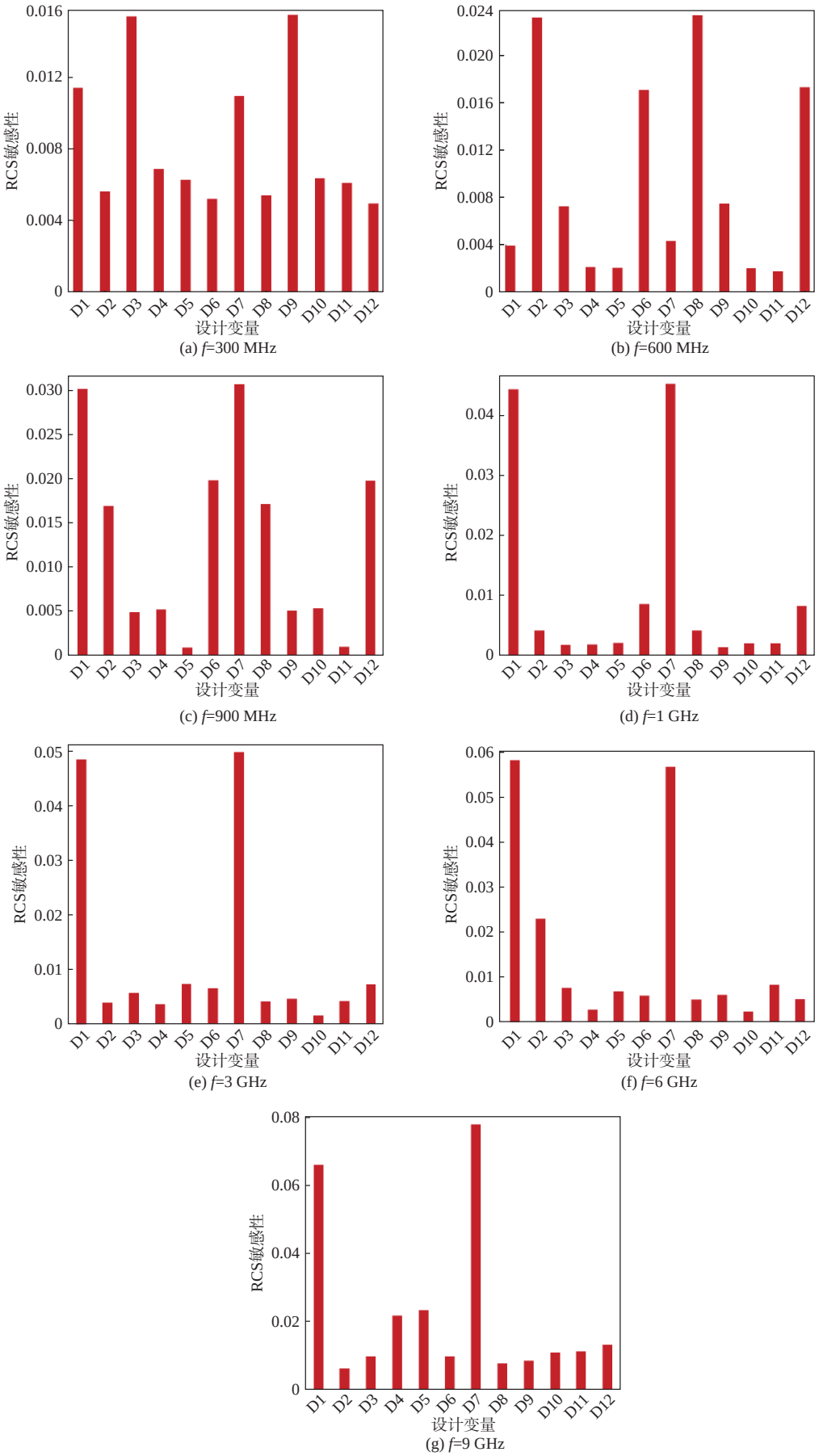


图 7　设计变量关于 RCS 敏感性

Fig. 7　Sensitivity to RCS of design variables

1 GHz、3 GHz、6 GHz、9 GHz, 采用基于代理模型的基于小生境的自适应空间重构 (niching-based adaptive space reconstruction, NASR) 算法[6] 开展优化设计, 该方法利用基于 Niching 的并行采样方法, 实现复杂设计空间高效搜索, 尤其适用于气动隐身多学科设计问题。

不同频率翼型设计结果如图 8 所示, 其中, 相较于初始翼型 NACA 65,3-018, 优化翼型 Opt_300 MHz 外形变化较大, 前缘半径增加, 呈明显的“钝头体”, Opt_300 MHz 的 10%～50% 弦长处的厚度减小, 50% 弦长后的翼型厚度明显增加。随着翼型设计频率提升至 900 MHz, 各优化翼型的前缘半径逐渐减小, 但当设计频率达到 1 GHz 时, 翼型前缘半径增加, 甚至大于 Opt_600 MHz, 随后又逐渐减小。此外, 当翼型设计频率在 300～1 000 MHz 范围内时, 翼型外形的厚度和弯度变化较大, 而当隐身设计频率在 3～9 GHz 范围内时, 优化翼型外形变化趋势较为接近, 主要区别在于前缘半径大小的不同。

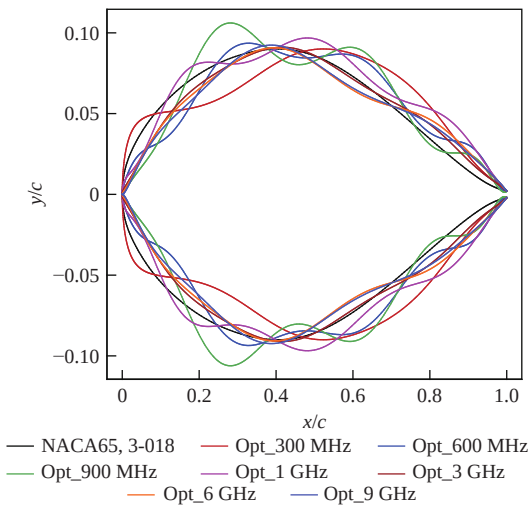


图 8 翼型设计外形对比

Fig. 8 Shape comparison of optimized airfoil

图 9 为各优化翼型在不同频率下隐身特性评估, 由图可知, 相较于初始翼型, 优化翼型 Opt_300 MHz 在 f=100～1 000 MHz 范围内前向 RCS 均值降低, 但当 f>600 MHz 时, Opt_300 MHz 在各频率的 RCS 均值明显提升; 优化翼型 Opt_600 MHz 和 Opt_900 MHz 在 f=200～400 MHz 范围内 RCS 均值提升, 在 f>400 MHz 时, RCS 均值明显降低; 优化翼型 Opt_1 GHz 在 f=100～600 MHz 的频率范围内 RCS 均值提升, 在 f>600 MHz 时, RCS 均值明显降低; 随着设计频率的进一步提升, 优化翼型 Opt_3 GHz、Opt_6 GHz 和 Opt_9 GHz 在 f>300 MHz 时 RCS 均值皆明显降低。

从上述结果可知, 谐振区和光学区内不同频率优化翼型的外形和特性差异很大, 谐振区 300 MHz

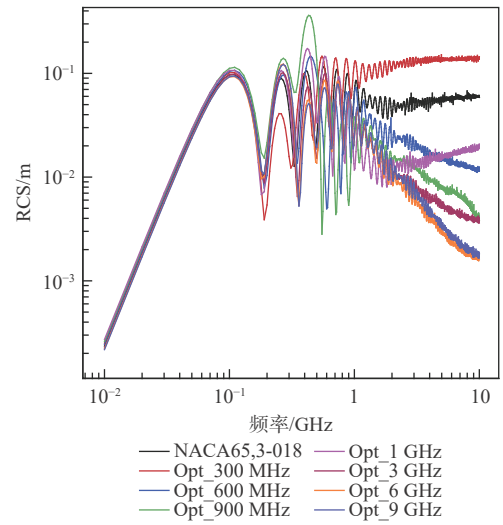


图 9 单点优化翼型 RCS 特性对比

Fig. 9 Comparison of RCS performance of airfoils subjected to single-point optimization

优化翼型的 RCS 特性在设计点改善明显, 但是在谐振区其他频率和光学区的 RCS 特性显著变差; 谐振区其他频率的优化翼型在设计点和光学区的 RCS 特性可以大幅改善, 但是在谐振区其他频率上 RCS 特性会急剧变差; 在隐身入射频率 f>3 GHz 时, 优化翼型的 RCS 特性在 f<600 MHz 范围内会变差, 但是在 f>600 MHz 范围内 RCS 特性显著提升, 并呈现相近的变化趋势。这是因为谐振区的 RCS 特性改善主要依赖相位相消[13], 不同频率对相位要求不同, 因此, 优化翼型在谐振区特性呈现明显的矛盾, 而当 f>600 MHz 时, 优化翼型前缘半径减小, 同时有利于减少光学区的镜面反射, 进一步提升高频的 RCS 特性。

3 考虑宽频 RCS 特性的翼型设计

由 2.2 节可知, 当隐身设计频率为 300 MHz 时, 优化翼型在 f=100～1 000 MHz 范围内 RCS 特性提升, 但是在光学区 RCS 特性严重受损; 当隐身设计频率大于 3 GHz 时, 优化翼型的隐身特性在 f=300～10 000 MHz 范围内均有所提升。

根据以上分析, 本节以 NACA 65,3-018 翼型为基础, 选取 f=(300 MHz, 9 GHz), 开展翼型宽频隐身设计研究。设计目标为提高 f=(300 MHz, 9 GHz) 时翼型的前向 RCS 均值, 入射角度如图 4 所示, 优化前后翼型最大厚度不减小。本文将 2 个设计目标进行正则化, 并采用线性加权的方式转换为一个目标进行优化搜索, 权重比例 1:1, 设计模型如下所示:

$$\begin{cases} \text{obj} \dfrac{R_{\text{CS}300\text{ MHz}}}{R_{\text{CS}300\text{ MHz_Baseline}}} + \dfrac{R_{\text{CS}9\text{ GHz}}}{R_{\text{CS}9\text{ GHz_Baseline}}} \\ \text{s.t.} \quad t/c_{\max} \geqslant 0.18 \end{cases} \tag{5}$$

式中: $R_{\text{CS}300\text{ MHz}}$ 和 $R_{\text{CS}9\text{ GHz}}$ 为设计外形的 RCS 特性;

$R_{\mathrm{CS300\ MHz_Baseline}}$和$R_{\mathrm{CS9\ GHz_Baseline}}$为初始外形的 RCS 特性; $t/c_{\max}$为设计外形最大相对厚度。

如表 1 所示, 在 f=300 MHz 时, 相较于初始外形 NACA65,3-018, 优化翼型 Opt_300 MHz9 GHz 的前向 RCS 均值降低了 0.004 7 m, 而当 f=9 GHz 时, 优化翼型的前向 RCS 均值降低了 0.055 6 m, RCS 特性大幅提升。图 10 为翼型设计外形对比, 相较于初始外形 NACA65,3-018, 优化翼型 Opt_300 MHz 9 GHz 前缘半径大幅减小, 前缘厚度变薄, 优化前后翼型最大厚度位置基本保持, 并且除了 20% 弦长处位置的弯度有所增加以外, 前缘其余位置弯度均有降低的趋势, 翼型后缘外形变化较小。

如图 11 所示, 相较于初始翼型 NACA65,3-018, 优化翼型 Opt_300 MHz9 GHz 实现了谐振区(f>300 MHz)和光学区等宽频范围内 RCS 特性的提升, 其中, 在 f=300～500 MHz 和 f>800 MHz 范围内, RCS 特性提升尤为明显。

表 1　优化翼型特性

Table 1　Performance of optimized airfoil

外形	RCS/m	
	f=300 MHz	f=9 GHz
NACA65,3-018	0.044 9	0.059 6
Opt_300 MHz 9 GHz	0.040 2	0.004 0

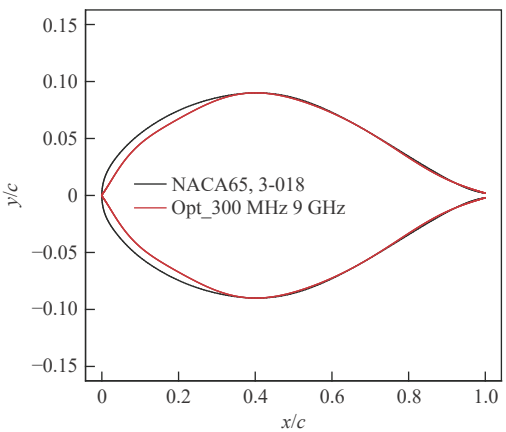


图 10　优化翼型外形对比

Fig. 10　Shape comparison of optimized airfoil

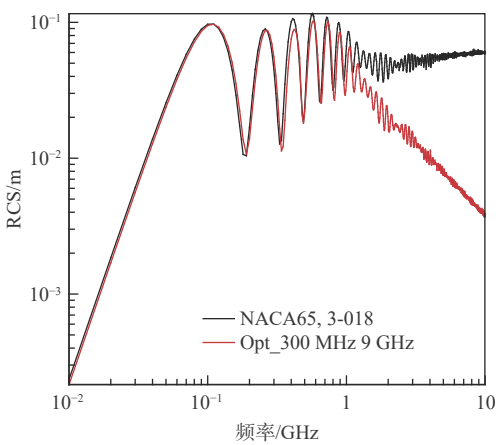


图 11　多点优化翼型 RCS 特性对比

Fig. 11　Comparison of RCS performance of airfoils subjected to multi-point optimization

4　考虑宽频 RCS 特性的翼型气动隐身设计

在第 3 节构造的宽频隐身设计模型基础上, 以 NACA 65,3-018 翼型为基础翼型, 继续开展考虑宽频 RCS 特性的翼型气动隐身综合设计, 其中, 气动设计目标为减小阻力系数, 其设计状态为 $Ma=0.70, C_l=0.25$[20], 隐身设计目标为降低 f=(300 MHz, 9 GHz) 时翼型的前向 RCS 均值, 隐身入射角度如图 4 所示。气动约束为力矩系数大于 0.03, 几何约束为翼型最大厚度不减小, 同样采用线性加权法, 将其转化为单目标设计问题, 设计模型如下所示:

$$\begin{cases} \mathrm{obj}\ \dfrac{R_{\mathrm{CS300\ MHz}}}{R_{\mathrm{CS300\ MHz_Baseline}}}+\dfrac{R_{\mathrm{CS9\ GHz}}}{R_{\mathrm{CS9\ GHz_Baseline}}}+\dfrac{C_d}{C_{d_\mathrm{Baseline}}} \\ \mathrm{s.t.}\quad C_l=0.25 \\ \qquad C_m\geqslant 0.03 \\ \qquad t/c_{\max}\geqslant 0.18 \end{cases} \tag{6}$$

式中: C_d为设计外形的阻力系数; C_m为力矩系数; C_{d_Baseline}为初始外形的阻力系数。

表 2 为优化翼型特性对比, 相较于初始翼型 NACA65,3-018, 优化翼型 Opt_Aero/Stealth 气动和隐身特性均有明显提升, 阻力系数降低了 2.07×10^{-3}, 力矩系数达到了 0.030 1, 满足约束要求, 在隐身特性方面, 当 f=300 MHz 时, 优化翼型 Opt_Aero/Stealth 前向 RCS 均值降低了 0.004 8 m, 当 f=9 GHz 时, 优化翼型 Opt_Aero/Stealth 前向 RCS 均值降低了 85.6%。

表 2　优化翼型特性

Table 2　Performance of optimized airfoil

外形	C_d	C_m	RCS/m	
			f=300 MHz	f=9 GHz
NACA65,3-018	0.011 41	−0.010	0.044 9	0.059 6
Opt_Aero/Stealth	0.009 34	0.030 1	0.040 1	0.008 6

图 12 为翼型设计外形对比, 相较于初始翼型, 优化翼型前缘半径减小, 下表面前缘 20% 弦长位置处正弯度增加, 使其出现明显的正加载特征, 配合后缘 60% 弦长处的反加载, 有利于提高翼型的抬头力矩, 如图 13 所示。图 14 为优化翼型 RCS 特性对比, 相较于初始翼型 NACA65,3-018, 优化翼型 Opt_Aero/Stealth 前向 RCS 特性明显提升, 在 200～

300 MHz范围内前向RCS均值降低了0.004 m左右，300～600 MHz范围内RCS特性基本保持，当f>600 MHz时，优化翼型前向RCS均值显著降低，并且随着入射频率的提升，改善量逐渐增加。

相较于初始外形NACA65,3-018，隐身设计结果Opt_300 MHz9 GHz和气动隐身综合设计结果Opt_Aero/Stealth在谐振区隐身特性改善不大，在光学区改善明显，而相较于Opt_300 MHz 9 GHz、Opt_Aero/Stealth的光学区RCS特性略有损失，但二者在谐振区的RCS特性基本持平，说明气动特性与宽频隐身特性的主要矛盾集中于光学区。

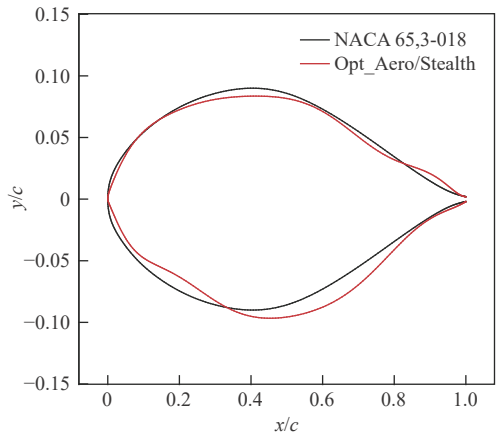


图12 气动隐身优化翼型外形对比

Fig. 12 Comparison of aerodynamic and stealth optimized airfoil shapes

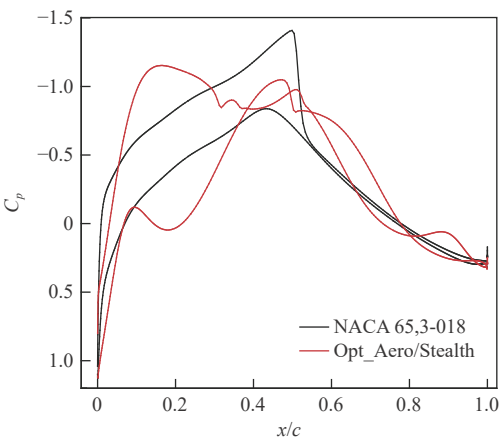


图13 优化翼型压力分布对比

Fig. 13 Pressure coefficient comparison of optimized airfoil

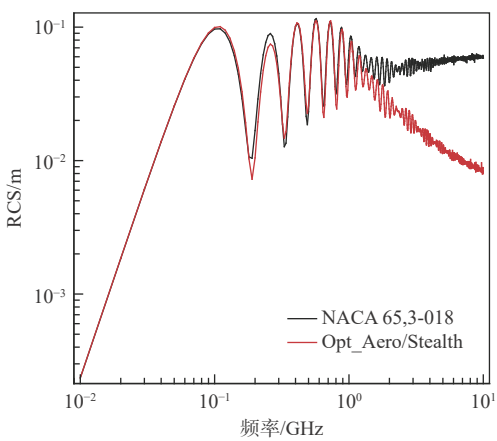


图14 气动隐身优化翼型RCS特性对比

Fig. 14 Comparison of RCS performance of airfoils from aerodynamio-stealth optimization

5 结 论

1）翼型外形参数对气动、隐身谐振区和光学区不同入射频率RCS特性敏感性不同，其中，翼型阻力特性主要受上表面外形影响，隐身谐振区RCS特性与翼型整体外形相关，并且不同频率RCS特性对翼型外形要求不同，存在强烈的矛盾关系。光学区RCS特性主要与翼型前缘外形密切相关，各频率下RCS特性对翼型外形要求一致。此外，光学区RCS特性提升有利于谐振区f>300 MHz范围内RCS特性的提升。

2）当f=300 MHz时，翼型RCS特性在设计点改善明显，但是在谐振区其他频率和光学区的前向RCS特性显著变差，而谐振区其他频率（600 MHz～1 GHz）的优化翼型在设计点和光学区的前向RCS特性改善明显，但在谐振区其他频率上RCS特性会损失，在光学区f>3 GHz时，翼型前向RCS特性在f<600 MHz范围内会变差，但是在f>600 MHz范围内RCS特性显著提升，并呈现相近的变化趋势。

3）基于以上结论，本文开展考虑宽频隐身特性的翼型气动隐身综合设计，结果表明，气动特性方面，优化翼型满足抬头力矩约束要求，提高了升阻特性，隐身特性方面，优化翼型在f=200～300 MHz和f>600 MHz范围内，前向RCS均值明显降低，300～600 MHz范围内RCS特性基本保持。

参考文献（References）

[1] ROMAN D, ALLEN J, LIEBECK R. Aerodynamic design challenges of the Blended-Wing-Body subsonic transport[C]//Proceedings of the 18th Applied Aerodynamics Conference. Reston: AIAA, 2000.

[2] ROMAN D, GILMORE R, WAKAYAMA S. Aerodynamics of high-subsonic blended-wing-body configurations[C]//Proceedings of the 41st Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2003.

[3] LIEBECK R H. Design of the blended wing body subsonic transport[J]. Journal of Aircraft, 2004, 41(1): 10-25.

[4] KROO I. Innovations in aeronautics[C]//Proceedings of the 42nd AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2004.

[5] 张伟, 赵轲, 夏露, 等. 飞翼布局翼型系列设计进展[J]. 空气动力学学报, 2021, 39(6): 37-52.
ZHANG W, ZHAO K, XIA L, et al. A multi-disciplinary

global/local optimization method for flying-wing airfoils design[J]. Acta Aerodynamica Sinica, 2021, 39(6): 37-52(in Chinese).

[6] ZHANG W, GAO Z H, WANG C, et al. The niching-based adaptive space reconstruction method for airfoil aerodynamic/stealth design[J]. Structural and Multidisciplinary Optimization, 2023, 66(7): 159.

[7] 张彬乾, 罗烈, 陈真利, 等. 飞翼布局隐身翼型优化设计[J]. 航空学报, 2014, 35(4): 957-967.
ZHANG B Q, LUO L, CHEN Z L, et al. On stealth airfoil optimization design for flying wing configuration[J]. Acta Aeronautica et Astronautica Sinica, 2014, 35(4): 957-967(in Chinese).

[8] TAJ Z U D, BILAL A, AWAIS M, et al. Design exploration and optimization of aerodynamics and radar cross section for a fighter aircraft[J]. Aerospace Science and Technology, 2023, 133: 108114.

[9] ZHANG W, GAO Z H, ZHOU L, et al. Research on multimodality in aerodynamic/stealth airfoil design optimization[C]//Proceedings of the 12th International Conference on Mechanical and Aerospace Engineering. Piscataway: IEEE Press, 2021: 384-393.

[10] VINH H, DWYER H, VAN DAM C. Finite-difference algorithms for the time-domain Maxwell's equations-A numerical approach to RCS analysis[C]//Proceedings of the 23rd Plasmadynamics and Lasers Conference. Reston: AIAA, 1992.

[11] BONDESON A, YANG Y, WEINERFELT P. Shape optimization for radar cross sections by a gradient method[J]. International Journal For Numerical Methods In Engineering, 2004, 61(5): 687-715.

[12] 陈曦, 白俊强, 李权. 某飞翼布局隐身飞行器的翼型优化[J]. 航空计算技术, 2013, 43(6): 46-49.
CHEN X, BAI J Q, LI Q. Airfoil optimization of a stealth flying wind aerial vehicle[J]. Aeronautical Computing Technique, 2013, 43(6): 46-49(in Chinese).

[13] 周琳. 基于全局/梯度耦合的低可探测布局气动隐身设计研究[D]. 西安: 西北工业大学, 2022.
ZHOU L. Research on global/local based aerodynamic/stealth optimization of low observable aircrafts[D]. Xi'an: Northwestern Polytechnical University, 2022(in Chinese).

[14] SALTELLI A, ANNONI P, AZZINI I, et al. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index[J]. Computer Physics Communications, 2010, 181(2): 259-270.

[15] 张伟, 高正红, 周琳, 等. 基于代理模型全局优化的自适应参数化方法[J]. 航空学报, 2020, 41(10): 160-171.
ZHANG W, GAO Z H, ZHOU L, et al. Adaptive parameterization method for surrogate-based global optimization[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(10): 160-171(in Chinese).

[16] MORRIS M D. Factorial sampling plans for preliminary computational experiments[J]. Technometrics, 1991, 33(2): 161-174.

[17] KNOTT E F. A progression of high-frequency RCS prediction techniques[J]. Proceedings of the IEEE, 1985, 73(2): 252-264.

[18] 桑建华. 飞行器隐身技术[M]. 北京: 航空工业出版社, 2013.
SANG J H. Low-observable technologies of aircraft[M]. Beijing: Aviation Industry Press, 2013(in Chinese).

[19] SEARS W. Flying-wing airplanes——The XB-35/YB-49 program [C]//Proceedings of the Symposium. Reston: AIAA, 1980.

[20] LIEBECK R. Design of the blended-wing-body subsonic transport[C]//Proceedings of the 40th AIAA Aerospace Sciences Meeting & Exhibit. Reston: AIAA, 2002.

[21] BARTELS R E, RUMSEY C L, BIEDRON R T. CFL3D version 6.4-general usage and aeroelastic analysis[EB/OL]. (2022-08-12) [2024-01-15]. https://nasa.github.io/CFL3D/.

[22] PRAUTZSCH H, BOEHM W, PALUSZNY M. Bézier and B-spline techniques[M]. Berlin: Springer, 2002.

Sensitivity analysis and optimization design of high stealth flying wing layout airfoil

ZHANG Wei[1]，ZHOU Lin[1]，CHEN Xian[1]，SHU Bowen[2]，HUANG Jiangtao[1,*]，GAO Zhenghong[2]

(1. China Aerodynamics Research and Development Center，Mianyang 621000，China；
2. School of Aeronautics，Northwestern Polytechnical University，Xi'an 710072，China)

Abstract：High-stealth flying wing layouts must meet the requirements of wideband stealth design. Traditional aerodynamic shape stealth design mainly targets the high-frequency optical range, and there is relatively little research on aerodynamic shape stealth design for wing profiles in the resonance range, making it incapable of providing effective guidance for wideband stealth design of flying wing aerodynamic shapes. To address these issues, research on wideband stealth design of wing layouts was conducted. Firstly, the basic effects method was used to study the sensitivity of wing profile parameters in different frequency bands, and it was found that the requirements for low-frequency stealth design and aerodynamic design have different sensitivity areas to the aerodynamic shape of the airfoil, and there are complex contradictions in stealth design in different frequency bands. This is the basis for a wideband stealth design model for wing layout, and NACA65,3-018 airfoils are used for aerodynamic optimization design that takes wideband stealth into account. The results show that the established wideband stealth design model can improve the aerodynamic performance of wing layout airfoils, simultaneously achieving wideband high stealth characteristics.

Keywords：high stealth flying wing layout；broadband stealth；sensitivity analysis；basic effects method；optimization design

Received：2024-02-29；**Accepted**：2024-06-28；**Published Online**：2024-09-12 08：46
URL：link.cnki.net/urlid/11.2625.V.20240911.1330.001
* **Corresponding author**. E-mail：hjtcfd@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0172

基于点云分割算法的战术导弹表面压力分布快速预测

蔺佳哲[1]，何磊[1]，程明[1]，周岭[1,*]，杨春明[2]
(1. 中国空气动力研究与发展中心 计算空气动力研究所，绵阳 621000；2. 西南科技大学 计算机科学与技术学院，绵阳 621000)

摘 要：如何缩短飞行器的设计周期是我国航空航天领域亟待解决的难点问题，其中，气动设计是飞行器设计的关键环节，快速精确获取飞行器气动特性可有力加速概念设计方案的迭代改进。因此，基于中国空气动力研究与发展中心现有气动数据库，利用“1+*N*”模式的深度神经网络建模算法，构建来流参数、气动外形到表面压力分布数据的映射关系。为提升预测精度，结合压力分布数据原有的物面网格划分信息，改进点云分割PointNet++算法，准确识别导弹不同部件，自动增加不同部件的标签特征。测试案例表明，采用改进的点云分割算法和“1+*N*”模式的深度神经网络建模算法，战术导弹全弹表面压力分布预测平均相对误差(MRE)基本控制在10%以内。所提算法建模效率较高，适用于各类复杂外形飞行器的压力分布预测，具有较好的工程应用前景。

关 键 词：战术导弹；压力分布预测；点云分割；深度神经网络；PointNet++算法
中图分类号：V211.24; V211.3
文献标志码：A **文章编号**：1001-5965(2026)05-1587-09

飞行器是涉及多学科的复杂系统，气动性能作为其他子系统开展工作的基础和出发点，具有全局性和先导性，气动性能的优劣在很大程度上决定了飞行器的总体性能。对于某型战术导弹，研究表面压力分布对于改进战术导弹气动性能有重要的意义：①通过研究导弹在不同飞行条件下的压力变化情况，识别表面高/低压区域，改善飞行器的气动性能；②压力分布是评估导弹气动性能的重要指标之一，通过测量和分析导弹表面压力分布，可以得到导弹的升力、阻力和力矩系数等相关参数，评估导弹的性能，并指导后续改进工作。

德国航空航天中心正在不断开发基于Python的航空数据代理建模工具箱SMARTy，目的是为飞行器设计工程师提供先进的数据驱动预测技术[1-2]。SMARTy按照应用程序编程接口方法设计，可将不同模块组合成更大、更复杂的应用程序，包括代理模型、多保真度模型、数据融合、降阶建模及高度集成的任务。由此可以看出，基于深度学习的气动建模已成为未来气动特性预测领域的主要发展方向[3]，现已应用于翼型流场预测[4-6]、真空羽流预测[7]、结冰翼型气动系数预测[8-9]、流场特征识别[10-12]、机翼气动系数预测[13]、智能赋能流场计算[14]、转捩建模[15]及飞行器压力分布预测[16]等领域。

由于全弹表面压力分布规律较为复杂，且预测数据量较大，直接通过神经网络模型建立来流条件、气动外形与表面压力分布的映射关系，建模精度通常难以满足需求。为解决上述问题，部分学者采用数据降阶的方式，该方法主要包括线性模态分析和非线性特征提取2类[17-18]。基于数据驱动的非线性特征提取可以通过神经网络直接学习，即构建自编码器(auto-encoder, AE)实现流场快照的自映射。这种方法的降维通过设置神经网络隐藏层神经元个数小于输入层来实现[19]，在高超声速流场预测[20-22]、复杂外形飞行器热环境预测[23]方面取得

收稿日期：2024-03-26；录用日期：2024-05-24；网络出版时间：2024-08-29 10:43
网络出版地址：link.cnki.net/urlid/11.2625.V.20240828.1409.003
*通信作者. E-mail：zlkk72@sina.com

引用格式：蔺佳哲，何磊，程明，等. 基于点云分割算法的战术导弹表面压力分布快速预测[J]. 北京航空航天大学学报，2026，52(5)：1587-1595. LIN J Z，HE L，CHENG M，et al. Rapid prediction of surface pressure distribution of tactical missile based on point cloud segmentation algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1587-1595 (in Chinese).

较好效果。近年来，以深度学习为代表的机器学习技术取得了令人瞩目的发展和成就，基于深度神经网络的特征提取和降维方法再次得到关注，并应用于流体力学预测领域。Wang等[24]发展了一种基于深度学习的变分AE来提取和描述超临界翼型周围流场的特征，并预测其稳态流场。统计结果表明，该方法在超临界翼型周围流场的重建和预测中具有较好的适用性，在激波区同样可以得到准确的预测结果。Wu等[25]在卷积自编码器(conventional auto-encoder, CAE)中引入自注意机制，进一步提高了网络特征提取能力，以圆柱绕流为例，验证了引入自注意机制的CAE流场预测误差比一般CAE流场预测误差降低了42.9%。类似的基于卷积神经神经网络/CAE的流场预测方法见文献[26-28]。需要注意的是，上述文献研究的气动外形一般较为光滑。对于具体的工程实践，如某型战术导弹表面压力分布的预测任务，由于导弹外形复杂等原因，相应的表面压力分布数据很难进行整体降阶处理。

为快速获取高超声速飞行器表面热流数据，张智超等[29]提出一种基于径向基神经网络的气动热快速预测代理模型方法，即在飞行器表面每个离散化的网格节点单独构造一种正则化的径向基神经网络，通过训练集对所有网络同时进行训练，获得各自网络的连接权值，所有网格节点的神经网络协同预测飞行器表面不同位置的热流，但飞行器表面热流分布不是独立的，而是相互关联的，上游流场势必对下游的热流分布带来影响。

综上，无论是降阶建模还是多点预测方法，对于战术导弹表面压力分布预测任务均难以胜任。为提升战术导弹表面压力分布预测精度，不仅需要关注输入(气动外形、来流参数)与输出(表面压力分布数据)之间映射模型的建模算法，而且输入输出数据本身的物理特征还需要进一步挖掘。战术导弹表面压力分布数据的输入包含导弹表面的一组点，由三维坐标组成，可以利用点云数据成熟的处理算法进行数据挖掘[30]。全弹表面压力分布建模过程中，导弹不同部件的表面压力分布规律相差较大，需要对不同区域的数据打上相应标签特征，这一流程可借助点云分割算法自动化实现。主流点云分割算法主要包括基于深度学习的PointNet及其改进算法，由于PointNet无法准确获取局部特征，很难对复杂场景进行分析。PointNet++采用层级结构，将点云数据逐层细化处理，通过聚合操作(如最大池化)来获取局部特征，并通过共享和非共享的多层感知器(multi-layer perceptron, MLP)来提取全局特征。这种层级结构能够有效处理不同尺度的物体和细节，并捕获点云中的空间关系[31]。

本文基于中国空气动力研究与发展中心气动数据库中的部分战术导弹气动数据集，构建导弹来流条件、气动外形与表面压力分布之间的快速预测模型，通过点云分割算法提升预测精度，有力支撑飞行器概念设计方案的高效迭代。

1 建模方案

中国空气动力研究与发展中心现有气动数据库中包含某型战术导弹共计3 000余种气动外形(包含不同弹翼/尾翼外形、不同弹头位置及不同弹翼/尾翼位置，如图1所示)数据集(气动力系数数据较全，表面压力分布数据较少)，其中，挑选出5种典型气动外形的战术导弹，每种外形现有200个工况条件下的表面压力分布数据，以此开展飞行器表面压力分布建模研究。战术导弹气动外形为常见的“*x*-*x*”轴对称外形，表面压力分布数据工况范围为：战术导弹马赫数0.6～3.0、攻角0°～30°、滚转角0°～45°。

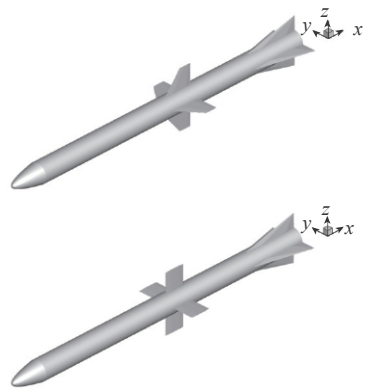


图1 不同类型的战术导弹气动外形(举例)

Fig. 1 Aerodynamic shapes of different types of tactical missiles (examples)

采用CAE进行压力数据降阶预测，在低维空间构建基于神经网络的表面压力分布预测模型，最终将预测结果还原到原始维度。上述方法中，表面压力分布的预测误差由CAE降阶重构误差和模型预测误差2部分累积而成。测试结果表明，CAE降阶重构误差较大，最终表面压力分布的预测平均相对误差(mean relative error, MRE)难以控制在20%以内。综合考虑建模效率和精度，本文设计2种建模算法。

1.1 建模算法1（原始算法）

对于全弹表面压力分布的预测，仅构造一个预测模型可充分考虑各区域之间的交互关系，实现来流条件、气动外形到表面压力分布的快速预测。通过对比深度神经网络和梯度提升决策树(gradient

boosting decision tree, GBDT)等传统机器学习算法，由于全弹表面压力分布预测的映射关系较为复杂，只有采用深度神经网络模型(如 MLP)，经过多轮超参数调试，才能最大限度保证表面压力分布的预测精度。

1.2　建模算法 2（改进算法）

通过分析基于数据驱动的建模误差来源，充分发掘气动数据的潜在物理信息以提升预测精度。一般而言，战术导弹的部件可以简单划分为 8 类：弹头、前弹身、弹翼、弹身与弹翼融合段、后弹身、尾翼、弹身与尾翼融合段及弹尾喷口。不同部件的表面压力分布规律相差较大，通过对不同部件输入数据增加标签特征，理论上可以提升输入数据的表征能力，进而提升表面压力分布的建模精度。

将表面压力分布数据的坐标点转换为点云数据，再通过点云分割算法 PointNet++进行部件识别和分割。实现局部特征提取需要解决 2 个关键问题：①如何将点集划分为不同区域；②如何利用特征提取器获取不同区域的局部特征。实际上，2 个问题是相关的，如果通过特征提取器对不同区域进行特征提取，需要每个分区具有相同结构。可以类比卷积神经网络(convolutional neural networks, CNN)来理解，在 CNN 中，卷积块作为基本的特征提取器，对应的区域都是(n,n)的像素区域。而在三维点云中，同样需要找到结构相同的子区域和对应的区域特征提取器。PointNet++的网络总体是 Encoder-Decoder 结构。Encoder 为降采样过程，通过多个集合抽象化结构实现多层次的降采样，得到不同规模的逐点特征，输出值可以认为是全局特征。其中，集合抽象化由采样、分组和点云网络 3 个模块构成。Decoder 根据分类和分割应用，又有所不同，分割任务 Decoder 为上采样过程，通过反向插值和跳跃链接实现在上采样的同时，能够获得局部加全局的逐点特征。具体原理如图 2 所示。

点云数据的分割任务实际是为原始点云中的每个点分配一个语义标签(即弹头/弹身/弹翼等)。比较简单的方式是：在采样过程中，采样所有的点为圆心进行局部特征提取。然而，该方式比较耗时。本文利用基于距离插值的分层特征传播策略，可以有效解决上述问题。

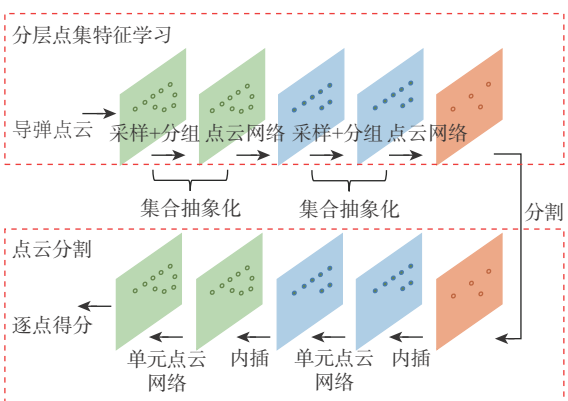


图 2　PointNet++点云分割基本原理

Fig. 2　Basic principle of PointNet++ point cloud segmentation

1) 基于k近邻的反向距离加权平均的插值方式，实现了丢失点(待插值点)特征的求解。假设丢失点(待插值点)x的待求解特征为$f(x)$，并假设其k个近邻点(特征已知)的特征为$f_i(x), i=1,2,\cdots,k$，则$f(x)$的求解如下：

$$\begin{cases} f(x)=\dfrac{\sum\limits_{i=1}^{k} w_i(x) f_i(x)}{\sum\limits_{i=1}^{k} w_i(x)} \\ w_i(x)=\dfrac{1}{d(x,x_i)} \end{cases} \tag{1}$$

式中：d为衡量距离的函数；w_i为d的倒数，距离d越小，权重越大。

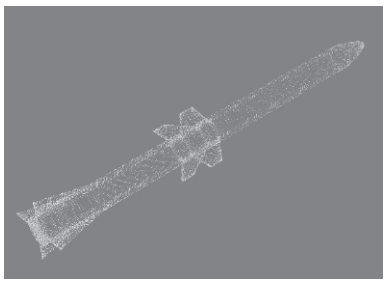
2) 将插值特征与集合抽象化阶段的特征(两者具有相同数量的特征点)通过结构连接后进行特征堆叠。

3) 堆叠的特征输入单元点云网络(类似于 CNN)中，实现特征的进一步提取。

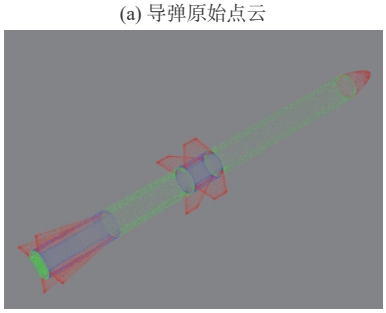
4) 执行步骤 1)～步骤 3)若干次，输出分割预测结果。

基于 PointNet++算法，观察到点云分割后，部件边缘部分点存在分类偏差。统计发现，部件分割的精度约为 80%～90%，需要借助相关物理信息帮助消除分割误差。表面压力分布数据集本身包括飞行器物面网格划分信息，对于某型战术导弹，表面划分为约 150 个物面网格区域(不同战术导弹气动外形的划分区域个数有所差别)。此时，对于不同部件交界处的点云分割，设定 60% 的点所在的物面网格属于该部件的网格区域，按照该规则，对每个部件点云分割后的结果进行补全或裁剪。最终，采用改进的 PointNet++算法可以高效处理气动数据库中的大量表面压力分布数据，其中，点云部件分割的准确性接近 100%，如图 3 所示。

点云分割完成后，需要构建表面压力分布的映射模型。建模算法 1 仅采用 1 个 MLP 模型，可以充分考虑全弹的表面压力分布规律，且各部件之间表面压力分布的预测结果更平顺，但全弹结构的复杂性使 1 个 MLP 模型难以实现表面每个位置的精确预测。通过构建“1+N”模式的 MLP 模型，即 1 个 MLP 模型预测全弹的表面压力分布趋势，N 个 MLP 模型分别预测导弹不同位置的表面压力分布数据。利用改进的 PointNet++算法，可以将战术导弹外形



(a) 导弹原始点云



(b) 导弹部件分割

图3 战术导弹点云部件分割示意图

Fig. 3 Schematic diagram of tactical missile point cloud component segmentation

点云快速分割为弹头、弹身、弹翼和尾翼4类。将弹身细分为前弹身、弹身与弹翼融合段、后弹身、弹身与尾翼融合段及弹尾喷口5小类，因此，“1+N”模式的N取值为8。

以弹头表面压力分布预测为例，全弹MLP模型在弹头部位的预测结果为$C_{\text{p-all}}$，弹头MLP模型预测结果为$C_{\text{p-head}}$，$\alpha C_{\text{p-all}}+(1-\alpha)C_{\text{p-head}}$为最终弹头表面压力分布的预测结果，权重$\alpha$的值可以通过多次模型训练，同时结合寻优算法得到。

2 建模流程设计

以建模算法2为例，本文设计的战术导弹表面压力分布建模算法流程如图4所示。

1）战术导弹气动数据集是采用国家数值风洞工程（National Numerical Wind, NNW）系列软件NNW-FlowStar计算的某型战术导弹不同工况下的流场数据，借助NNW-TopViz等程序后处理得到飞行器表面压力分布数据集。

2）利用改进的PointNet++算法，根据物面网格划分信息，采用“对于不同部件交界处的点云分割，设定60%的点所在的物面网格属于该部件的网格区域”这一判定原则，实现表面压力分布空间点云数据的精准分割处理，对导弹不同部件的输入数据增加标签值。

3）将表面压力分布数据按照一定比例，根据工况条件划分为训练集、验证集和测试集。

4）利用“1+N”模式的深度神经网络MLP模型，建立输入数据（马赫数、攻角、滚转角、x、y、z、标签值）到输出数据（压力系数C_p）的映射关系。

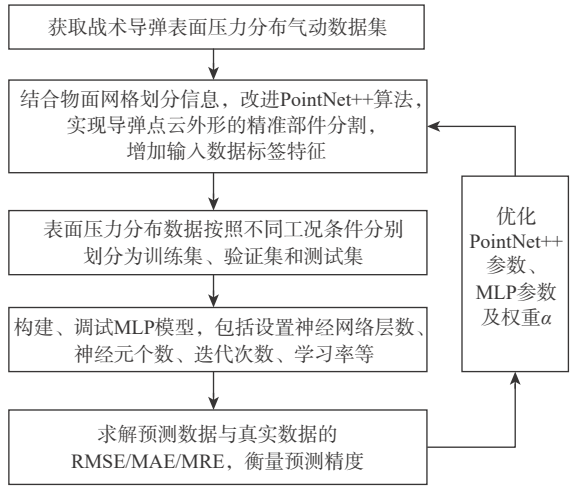


图4 战术导弹表面压力分布预测流程

Fig. 4 Flow chart of surface pressure distribution prediction for tactical missiles

5）测试集由不同工况条件下飞行器表面压力分布数据组成，计算预测值和真实值之间的均方根误差（root-mean-square error, RMSE）、平均绝对误差（mean absolute error, MAE）及MRE，对预测结果的优劣进行衡量。

6）迭代改进PointNet++算法和“1+N”模式的MLP模型，确定参数α，使表面压力分布预测模型达到最佳预测精度。

与现有文献相比，上述算法具有如下有益效果：

1）构建预测表面压力分布的“1+N”模式的MLP模型，建模精度满足导弹总体概念设计需求，解决了飞行器表面压力分布数据复杂导致建模精度低的难题。

2）现有气动数据库中的战术导弹气动数据集结合已有的物面网格划分信息，通过改进的PointNet++算法，精确分割导弹不同部件，对表面压力分布数据中的空间坐标添加标签值，可有效提升输入数据中气动外形的表征能力，提高全弹MLP模型的预测精度及部件MLP模型的建模效率。

3）本文建模算法旨在充分且高效利用现有气动数据库进行某型战术导弹表面压力分布预测。通过建立点云分割、预测模型构建、参数寻优等主要环节的自动化流程，减少气动数据清洗规整的工作量，大幅提升建模效率和气动数据库利用率。

3 测试案例分析

战术导弹5种气动外形的每组外形中，195组表面压力分布数据作为训练集，随机抽取5组作为测试集，通过计算全部测试集的RMSE、MAE及MRE数值，衡量表面压力分布预测精度。训练集数据样本数为195×5=975组，5种导弹外形相近（差异主要为弹翼尾翼相对弹身的位置、弹翼和尾翼的

形状), 一般对于基于数据驱动的气动建模, 训练集数据量接近 1 000 才能确保较好的预测精度。

3.1　测试结果

对战术导弹 5 种气动外形下, 共计 25 组测试集进行预测。图 5(a)展示了选取导弹模型 3, 马赫数 1.2、攻角 2°、滚转角 22.5°条件下弹头表面压力分布的预测结果。可以看出, 建模算法 1 预测结果中, 弹头表面压力分布等压线波折较多, 不符合战术导弹表面压力分布的实际规律。建模算法 2 预测结果中, 表面压力分布的变化规律与计算流体动力学(computational fluid dynamics, CFD)计算结果基本相同, 该算法的预测精度较建模算法 1 高。为进一步比较 2 种建模算法的预测效果, 如图 5(b)所示, 对弹头进行切片处理, 比较切片压力曲线的细节特征。不难发现, 建模算法 2 预测的切片压力曲线更接近 CFD 计算的表面压力分布曲线, 建模算法 1 预测的切片压力曲线与 CFD 计算的表面压力分布曲线的变化规律存在明显偏差。

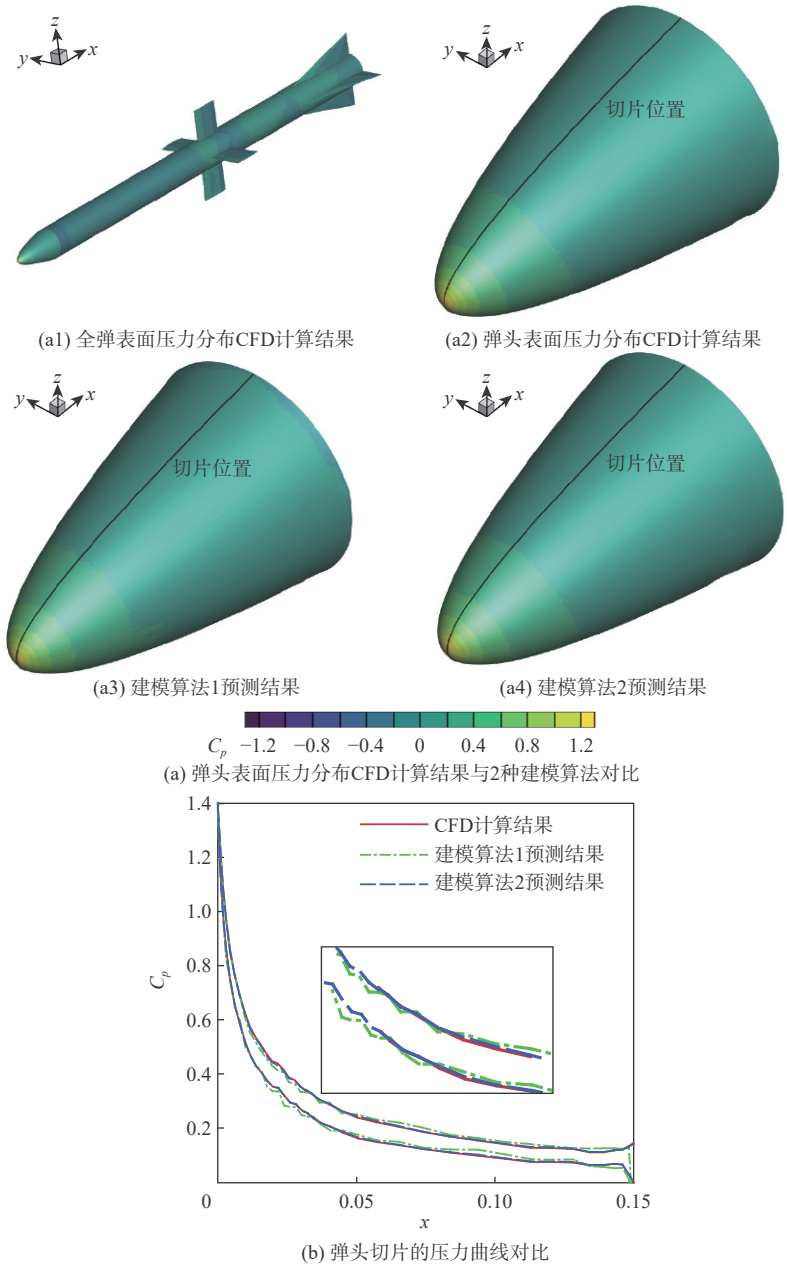
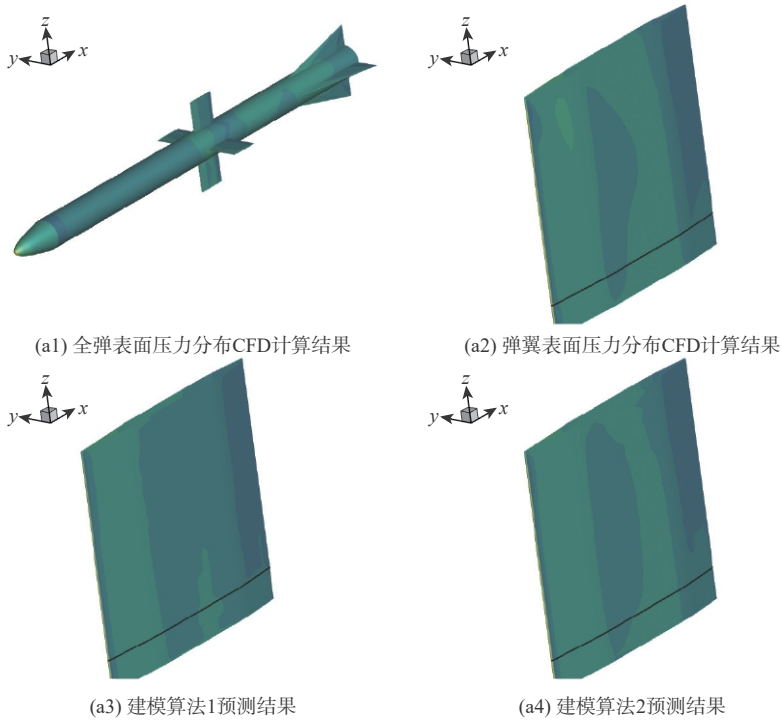


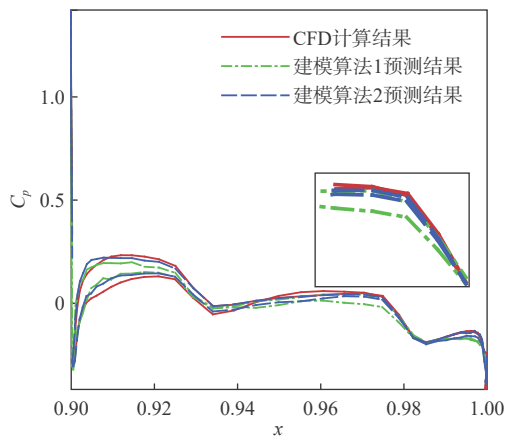
图 5　导弹模型 3 在马赫数 1.2、攻角 2°、滚转角 22.5°条件下弹头表面压力分布 CFD 计算结果及 2 种建模算法预测结果

Fig. 5　CFD calculation results and two modeling algorithms' prediction results of warhead surface pressure distribution for missile model 3 under Mach number 1.2, angle of attack 2° and roll angle 22.5°

图 6(a)所示的弹翼表面压力分布预测结果中, 建模算法 2 预测结果的细节规律更接近 CFD 计算结果, 且表面压力分布变化过渡更自然, 即建模算法 2 均优于建模算法 1。为进一步比较 2 种建模算法的预测效果, 如图 6(b)所示, 对弹翼进行切片处理, 比较切片压力曲线的细节特征。不难发现, 建



(a) 弹翼表面压力分布CFD计算结果与2种建模算法对比



(b) 弹翼切片的压力曲线对比

图 6 导弹模型 3 在马赫数 1.2、攻角 2°、滚转角 22.5°条件下弹翼表面压力分布 CFD 计算结果及 2 种建模算法预测结果

Fig. 6 CFD calculation results and two modeling algorithms' prediction results of missile wing surface pressure distribution for missile model 3 under Mach number 1.2, angle of attack 2° and roll angle 22.5°

模算法 2 预测的切片压力曲线更接近 CFD 计算的表面压力分布曲线。

图 7(a)所示的尾翼表面压力分布预测结果中，建模算法 1 的等压线变化情况与 CFD 计算结果相差较大，且整体云图形状相差较大，而建模算法 2 预测结果的细节规律更接近 CFD 计算结果。为进一步比较 2 种建模算法的预测效果，如图 7(b)所示，对尾翼进行切片处理，建模算法 2 预测的切片压力曲线基本贴合 CFD 计算的表面压力分布曲线。

对于其余 4 个战术导弹气动外形表面压力分布预测结果，建模算法 2 在预测精度上均优于建模算法 1。

3.2 误差分析

计算 5 个导弹模型共计 25 组测试集的平均误差指标，包含 RMSE、MAE 和 MRE，以此衡量建模算法 2 的可行性和有效性。RMSE 是模型预测结果与真实值之差的平方和均值的平方根；MAE 是模型预测结果与真实值之差的绝对值的平均数；MRE 是模型预测结果与真实值之差的绝对值与真实值之比的平均值。MRE 误差指标比较形象，可以直观了解模型的建模精度。表 1 为 25 个不同外形测试集的预测误差。

如图 8 所示，与未进行部件点云分割的建模算法 1 相比，建模算法 2 的全弹表面压力分布预测 MRE

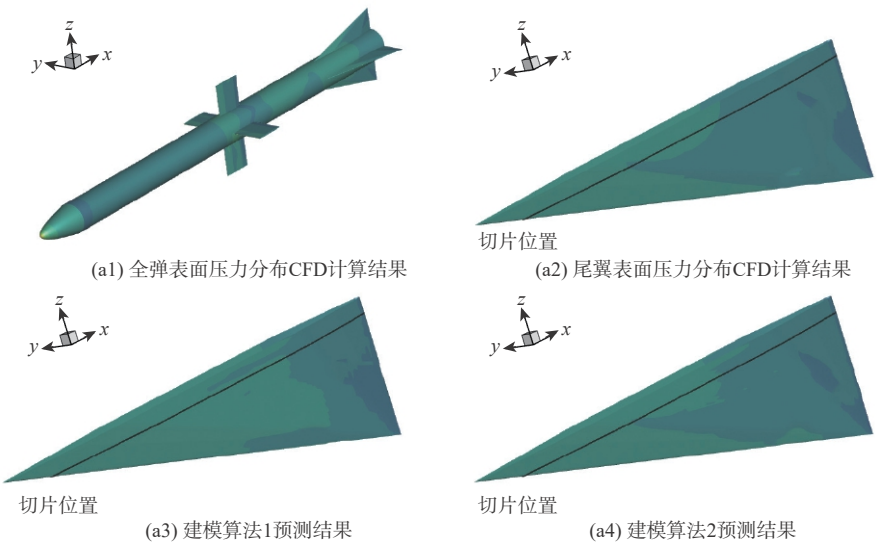


图 7　导弹模型 3 在马赫数 1.2、攻角 2°、滚转角 22.5°条件下尾翼表面压力分布 CFD 计算结果及 2 种建模算法预测结果

Fig. 7　CFD calculation results and two modeling algorithms' prediction results of tail wing surface pressure distribution for missile model 3 under Mach number 1.2, angle of attack 2° and roll angle 22.5°

表 1　战术导弹表面压力分布预测误差

Table 1　Prediction error of tactical missile surface pressure distribution

对象	建模算法	RMSE	MAE	MRE/%
全弹	算法1	0.058 6	0.026 9	26.56
	算法2	0.017 6	0.008 96	8.59
弹头	算法1	0.037 7	0.019 6	17.52
	算法2	0.005 0	0.003 7	3.33
弹尾喷口	算法1	0.029 2	0.015 1	7.61
	算法2	0.004 7	0.003 0	1.56
弹翼	算法1	0.111 4	0.054 8	39.46
	算法2	0.033 8	0.020 0	14.67
弹身与弹翼融合段	算法1	0.042 9	0.031 0	35.03
	算法2	0.018 1	0.012 9	13.38
前弹身	算法1	0.035 7	0.018 9	21.02
	算法2	0.009 0	0.005 2	5.97
后弹身	算法1	0.029 3	0.020 6	30.66
	算法2	0.011 6	0.008 0	11.04
尾翼	算法1	0.045 1	0.024 0	26.93
	算法2	0.012 3	0.007 4	7.79
弹身与尾翼融合段	算法1	0.023 8	0.014 5	19.98
	算法2	0.005 8	0.004 3	5.37

为 8.59%。建模算法 2 的弹头表面压力分布预测 MRE 为 3.33%, 而弹翼表面压力分布预测 MRE 高达 14.67%,

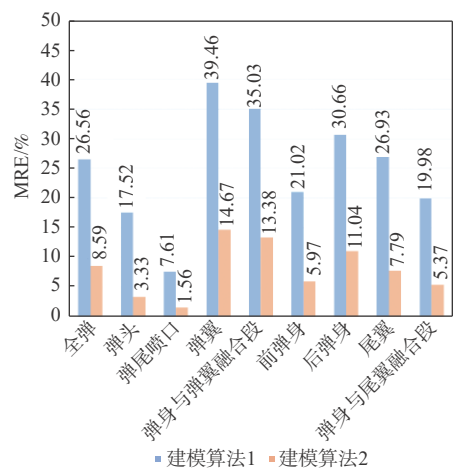


图 8　2 种建模算法的预测 MRE 对比

Fig. 8　Comparison of predicted MRE between two modeling algorithms

究其原因为：弹翼部分表面压力系数幅值接近0.1，虽然MRE指标较为直观，但接近0.1的数据计算相对误差时，误差容易被放大。

4 结 论

1）本文构建了基于部件特征增强的高精度表面压力分布预测模型。针对战术导弹全弹气动特性预测难题，提出结合改进PointNet++算法与“1+*N*”深度神经网络模式的映射算法。通过自动识别导弹部件，并增加标签特征，有效解决了复杂外形下流场特征提取困难的问题，实现了从来流参数与气动外形到表面压力分布的快速映射。

2）验证了分部件协同预测策略的有效性。测试结果表明，该策略能够准确捕捉全弹表面压力分布趋势及各部件局部细节。在典型测试案例中，战术导弹全弹表面压力分布预测的MRE控制在10%以内，显著提升了概念设计阶段气动数据的获取效率，满足了快速迭代的设计需求。

针对当前模型在复杂来流及新外形下的泛化局限性，下一步将借鉴数据驱动的当地化建模思想（如逆向半拉格朗日方法（backward semi-Lagrangian method, BSLM）），计划引入Euler方程计算的边界层外缘信息作为辅助特征，融合多保真度数据（如雷诺平均纳维-斯托克斯（Reynolds-averaged Navier-Stokes, RANS）方程计算结果），构建适应不同来流状态和壁面温度的气动热/力高效精准预测模型，以进一步拓展模型的适用范围和预测精度。

参考文献（References）

[1] BEKEMEYER P, BERTRAM A, HINES-CHAVES D A, et al. Data-driven aerodynamic modeling using the DLR SMARTy toolbox[C]//AIAA AVIATION 2022 Forum. Reston: AIAA, 2022: 3899.

[2] SABATER C, STÜRMER P, BEKEMEYER P. Fast Predictions of aircraft aerodynamics using deep learning techniques[C]//AIAA AVIATION 2021 Forum. Reston: AIAA, 2021: 2549.

[3] 张伟伟, 寇家庆, 刘溢浪. 智能赋能流体力学展望[J]. 航空学报, 2021, 42(4): 20-65.
ZHANG W W, KOU J Q, LIU Y L. Prospect of artificial intelligence empowered fluid mechanics[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(4): 20-65(in Chinese).

[4] 曹晓峰, 李鸿岩, 郭承鹏, 等. 基于深度学习的二维翼型流场重构技术研究[J]. 航空科学技术, 2022, 33(7): 106-112.
CAO X F, LI H Y, GUO C P, et al. Research on two-dimensional airfoil flow field reconstruction based on deep learning[J]. Aeronautical Science & Technology, 2022, 33(7): 106-112(in Chinese).

[5] HU J W, DOU Z H, ZHANG W W. Fast fluid-structure interaction simulation method based on deep learning flow field modeling[J]. Physics of Fluids, 2024, 36(4): 045106.

[6] HU J W, ZHANG W W. Flow field modeling of airfoil based on convolutional neural networks from transform domain perspective [J]. Aerospace Science and Technology, 2023, 136: 108198.

[7] 蔡国飙, 张百一, 贺碧蛟, 等. 真空羽流智能化计算[J]. 航空学报, 2022, 43(10): 102-114.
CAI G B, ZHANG B Y, HE B J, et al. Intelligent computation of vacuum plume[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(10): 102-114(in Chinese).

[8] 何磊, 钱炜祺, 董康生, 等. 基于卷积神经网络的结冰翼型气动特性建模[J]. 航空学报, 2023, 44(5): 59-72.
HE L, QIAN W Q, DONG K S, et al. Aerodynamic characteristics modeling of iced airfoil based on convolution neural networks[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(5): 59-72(in Chinese).

[9] 柴聪聪, 王强, 易贤, 等. 基于卷积神经网络的结冰翼型气动参数预测[J]. 飞行力学, 2021, 39(5): 13-18.
CHAI C C, WANG Q, YI X, et al. Aerodynamic parameters prediction of airfoil ice accretion based on convolutional neural network[J]. Flight Dynamics, 2021, 39(5): 13-18(in Chinese).

[10] 叶舒然, 张珍, 王一伟, 等. 基于卷积神经网络的深度学习流场特征识别及应用进展[J]. 航空学报, 2021, 42(4): 179-193.
YE S R, ZHANG Z, WANG Y W, et al. Progress in deep convolutional neural network based flow field recognition and its applications[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(4): 179-193(in Chinese).

[11] OBAYASHI W, AONO H, TATSUKAWA T, et al. Feature extraction of fields of fluid dynamics data using sparse convolutional autoencoder[J]. AIP Advances, 2021, 11(10): 105211.

[12] MURATA T, FUKAMI K, FUKAGATA K. Nonlinear mode decomposition with convolutional neural networks for fluid dynamics [J]. Journal of Fluid Mechanics, 2020, 882: A13.

[13] 吕召阳, 聂雪媛, 赵奥博. 基于CNN机翼气动系数预测[J]. 北京航空航天大学学报, 2023, 49(3): 674-680.
LYU Z Y, NIE X Y, ZHAO A B. Prediction of wing aerodynamic coefficient based on CNN[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(3): 674-680(in Chinese).

[14] CAO W B, SONG J H, ZHANG W W. A solver for subsonic flow around airfoils based on physics-informed neural networks and mesh transformation[J]. Physics of Fluids, 2024, 36(2): 027134.

[15] WU L, CUI B, WANG R, et al. Artificial neural network-substituted transition model for crossflow instability: modeling strategy and application prospect[J]. Physics of Fluids, 2024, 36(4): 045110.

[16] SHEN Y, HUANG W, WANG Z G, et al. A deep learning framework for aerodynamic pressure prediction on general three-dimensional configurations[J]. Physics of Fluids, 2023, 35(10): 107111.

[17] LE CLAINCHE S, LI J I, THEOFILIS V, et al. Flow around a hemisphere-cylinder at high angle of attack and low Reynolds number. part I: experimental and numerical investigation[J]. Aerospace Science and Technology, 2015, 44: 77-87.

[18] LE CLAINCHE S, RODRÍGUEZ D, THEOFILIS V, et al. Flow around a hemisphere-cylinder at high angle of attack and low Reynolds number. part II: POD and DMD applied to reduced domains[J]. Aerospace Science and Technology, 2015, 44: 88-100.

[19] AGOSTINI L. Exploration and prediction of fluid dynamical systems using auto-encoder technology[J]. Physics of Fluids, 2020, 32(6): 067103.

[20] CAO C Q, NIE C S, PAN S C, et al. A constrained reduced-order method for fast prediction of steady hypersonic flows[J]. Aerospace Science and Technology, 2019, 91: 679-690.

[21] 完颜振海, 徐嘉, 梁磊, 等. 基于特征代理模型的复杂流场预测方

法研究[J]. 弹箭与制导学报, 2017, 37(6): 105-108.
WANYAN Z H, XU J, LIANG L, et al. Research on method of prediction for complicated flow fields based on characteristic surrogate model[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2017, 37(6): 105-108(in Chinese).

[22] 侯强, 苏纬仪, 孙斐, 等. 高超声速乘波体扩容设计及流场快速预测[J]. 航空动力学报, 2021, 36(3): 564-574.
HOU Q, SU W Y, SUN F, et al. Design of improving volumetric efficiency of hypersonic waverider and rapid prediction of flow field[J]. Journal of Aerospace Power, 2021, 36(3): 564-574(in Chinese).

[23] 聂春生, 黄建栋, 王迅, 等. 基于 POD 方法的复杂外形飞行器热环境快速预测方法[J]. 空气动力学学报, 2017, 35(6): 760-765.
NIE C S, HUANG J D, WANG X, et al. Fast aeroheating prediction method for complex shape vehicles based on proper orthogonal decomposition[J]. Acta Aerodynamica Sinica, 2017, 35(6): 760-765 (in Chinese).

[24] WANG J, HE C, LI R Z, et al. Flow field prediction of supercritical airfoils via variational autoencoder based deep learning framework[J]. Physics of Fluids, 2021, 33(8): 086108.

[25] WU P, GONG S Q, PAN K K, et al. Reduced order model using convolutional auto-encoder with self-attention[J]. Physics of Fluids, 2021, 33(7): 077107.

[26] HASEGAWA K, FUKAMI K, MURATA T, et al. Machine-learning-based reduced-order modeling for unsteady flows around bluff bodies of various shapes[J]. Theoretical and Computational Fluid Dynamics, 2020, 34(4): 367-383.

[27] HASEGAWA K, FUKAMI K, MURATA T, et al. CNN-LSTM based reduced order modeling of two-dimensional unsteady flows around a circular cylinder at different Reynolds numbers[J]. Fluid Dynamics Research, 2020, 52(6): 065501.

[28] EIVAZI H, VEISI H, NADERI M H, et al. Deep neural networks for nonlinear model order reduction of unsteady flows[J]. Physics of Fluids, 2020, 32(10): 105104.

[29] 张智超, 高太元, 张磊, 等. 基于径向基神经网络的气动热预测代理模型[J]. 航空学报, 2021, 42(4): 297-306.
ZHANG Z C, GAO T Y, ZHANG L, et al. Aeroheating agent model based on radial basis function neural network[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(4): 297-306(in Chinese).

[30] 朱安迪, 达飞鹏, 盖绍彦. 对融合特征敏感的三维点云识别与分割[J]. 西安交通大学学报, 2024, 58(5): 52-63.
ZHU A D, DA F P, GAI S Y. Recognition and segmentation of point clouds sensitive to fusion features[J]. Journal of Xi'an Jiaotong University, 2024, 58(5): 52-63(in Chinese).

[31] 孙端正, 高飞, 叶周润, 等. 改进 PointNet++模型在道路杆状物提取中的应用[J]. 测绘通报, 2023(11): 95-99.
SUN D Z, GAO F, YE Z R, et al. Application of improved PointNet++ model in extracting road rods[J]. Bulletin of Surveying and Mapping, 2023(11): 95-99(in Chinese).

Rapid prediction of surface pressure distribution of tactical missile based on point cloud segmentation algorithm

LIN Jiazhe[1], HE Lei[1], CHENG Ming[1], ZHOU Ling[1,*], YANG Chunming[2]

(1. Computational Aerodynamic Institute, China Aerodynamics Research and Development Center, Mianyang 621000, China;
2. School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621000, China)

Abstract: How to shorten the design period of aircraft is a difficult problem to be solved urgently in the field of aerospace. Aerodynamic design is the key link of aircraft design. Rapid and accurate acquisition of aircraft aerodynamic characteristics can effectively accelerate the iterative improvement of conceptual design schemes. The deep neural network modeling algorithm of "1+*N*" mode is used to build the mapping relationship between flow parameters, aerodynamic profile, and surface pressure distribution data based on the China Aerodynamics Research and Development Center's current aerodynamic database. In order to improve the prediction accuracy, the point cloud segmentation PointNet++ algorithm was improved by combining the original surface grid division information of the pressure distribution data to accurately identify different components of the missile and automatically add label features of different components. The test case shows that by using the improved point cloud segmentation algorithm and the "1+*N*" mode deep neural network algorithm, the mean relative error (MRE) of the pressure distribution prediction of the tactical missile is basically controlled within 10%. The aforementioned approach has a high modeling efficiency, is appropriate for predicting the pressure distribution of a variety of intricate aircraft forms, and has a promising engineering application.

Keywords: tactical missiles; pressure distribution prediction; point cloud segmentation; deep neural network; PointNet++ algorithm

Received: 2024-03-26; **Accepted**: 2024-05-24; **Published Online**: 2024-08-29 10:43
URL: link.cnki.net/urlid/11.2625.V.20240828.1409.003

* **Corresponding author**. E-mail: zlkk72@sina.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0158

木卫四探测任务转移轨道状态序列搜索与优化

井泉[1,2]，李明涛[1,2,*]，王有亮[1]

(1. 中国科学院国家空间科学中心，北京 100190；2. 中国科学院大学，北京 100049)

摘 要：针对使用图形化分析方法设计木卫四环绕任务转移轨道时人工参与度高的问题，提出基于图形化分析的转移轨道状态序列搜索算法。通过对图形性质及共振引力辅助的分析，归纳出轨道状态的变化模式，进而提出迭代求解状态转移序列的搜索算法。对木卫四环绕任务转移轨道进行仿真，结果表明：所提算法可以快速高效获得多个转移序列；对得到的序列求解对应的转移轨道，探测器可以在2.13年内消耗2.108 km/s的速度增量，实现从初始环木大椭圆轨道到木卫四环绕轨道的转移，比对比算法得到的转移序列节省约100～200 m/s的速度增量及约1年的转移时间。

关 键 词：木星系统；木卫四探测；轨道优化设计；引力辅助；共振引力辅助

中图分类号：V249.122^{+}.5

文献标志码：A **文章编号**：1001-5965(2026)05-1596-09

木星是太阳系中最大的行星，与类地行星不同，木星是气体行星，人们对其了解较少。木星系统包含多颗卫星，4颗最大的被称为伽利略卫星，根据之前的研究，伽利略卫星上可能存在水和有机物[1]。木卫四(Callisto)是伽利略卫星中距木星最远的卫星，周围环境辐射较弱，适合长期探测，且木卫四具有古老的地质结构，保存有形成之初的信息。因此，对木星系统、木卫四的探索对于研究生命起源及行星演化具有重要意义。

1989年发射的“伽利略号”是最早的木星探测器，在轨进行了多次伽利略卫星的飞越，并对木星大气进行了探测[2]。2011年发射的“朱诺号”进入木星大倾角轨道，主要对木星极区进行观测[3]，对木星成分进行分析[4]。2023年发射的“冰月探测者”预计将多次飞越木卫二、木卫三、木卫四，对该3颗冰卫星进行环境、宜居性等评估[5]。也有不少计划中或取消了计划的任务，如：Europa Clipper重点观测木卫二[6]，为后续Europa Lander做先行研究[7]；Io Volcano Observer重点对木卫一的火山进行研究[8]；Laplace-P计划着陆在木卫三[9]；Europa Orbiter曾计划多次飞越伽利略卫星后被木卫二捕获[10]。多项计划表明伽利略卫星是将来的研究重点。

当前的任务集中关注于木卫二和木卫三，少有对木卫四还没有专门的任务。我国在2030年左右计划进行木星探测相关任务[11]和太阳系边际探测任务[12]，包括对木卫四的探测。国内也有学者对木卫四环绕任务转移轨道进行研究，文献[13]中的任务方案为先构造4次木卫四共振引力辅助来减小轨道半长轴，再利用3次木卫四共振借力配合机动来调节轨道偏心率，该方案可以在木星系内以2.335 km/s的速度增量完成任务，最终被木卫四捕获到圆轨道；文献[14]提出的多种方案可以多次飞越伽利略卫星，分别对木卫二、木卫三进行有计划地飞越，并最终被木卫四捕获到椭圆轨道；文献[15]提出一种3层优化策略，用来求解木星系内多次引力辅助的轨迹。

共振引力辅助是深空任务中常用的技术手段[16]。文献[17]研究了杠杆机动技术，通过在近星点或远星点施加机动来调整轨道周期与引力辅助天体周期成一定的倍数。文献[18]将该技术应用

收稿日期：2024-03-20；录用日期：2024-04-26；网络出版时间：2024-05-13 09:57
网络出版地址：link.cnki.net/urlid/11.2625.V.20240511.1535.001
基金项目：中国科学院青年创新促进会(2022146)
*通信作者. E-mail：limingtao@nssc.ac.cn

引用格式：井泉，李明涛，王有亮. 木卫四探测任务转移轨道状态序列搜索与优化[J]. 北京航空航天大学学报，2026，52(5)：1596-1604. JING Q，LI M T，WANG Y L. Sequence search and optimization for transfer trajectory status in Callisto exploration mission[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1596-1604 (in Chinese).

到非圆形轨道的多个天体的情形。文献 [19] 推导了共振引力辅助成立的充要条件，但过程中没有注意到天体速度在参考坐标系中的 z 轴分量为 0，导致推导出的条件略显复杂。木星具有较强的引力场，如果直接施加机动调节轨道，则需要很大的速度增量，共振引力辅助可以利用伽利略卫星的引力改变航天器能量，从而以较低的代价多次飞越伽利略卫星，甚至被木卫捕获。T-r_p 图本质是将航天器可能的状态描绘在周期 T、近木点半径 r_p 关系图中，文献 [14] 利用 T-r_p 图进行任务初步设计，决定了探测序列。在不考虑倾角的情况下，2 个参数即可确定轨道的形状，因此，针对不同的重点，也可选择不同的参数，如文献 [20] 选择近木点半径-远木点半径图来设计可以清楚了解轨道状态的变化。

在利用 T-r_p 图时，一般存在人工参与度高的问题。本文在文献 [19] 的基础上，对共振引力辅助的充要条件进行一定的简化，通过分析轨道状态点的转变模式，提出转移轨道状态序列搜索算法，用于求解木卫四环绕任务转移轨道，解决了人工参与度高的问题，得到速度增量与转移时间更优的转移轨道。

1　轨迹优化相关模型

由于当前火箭运载能力有限，从地球出发直接前往木星需要巨大能量，一般木星任务行星际段均采用行星引力辅助来节省燃料，木星系内航天器也常利用木卫来引力辅助改变轨迹。本节对轨迹优化时采用的引力辅助模型进行介绍。

1.1　等效脉冲引力辅助模型

将引力辅助效果等效为瞬时施加速度脉冲，简化后的引力辅助模型如图 1 所示。图中，$\boldsymbol{v}_{\mathrm{m}}$ 为引力辅助天体的速度矢量，引力辅助后的速度为 $\boldsymbol{v}_2$。引力辅助后的双曲超速 $\boldsymbol{v}_{\infty,2}$ 可由引力辅助前的双曲超速 $\boldsymbol{v}_{\infty,1}$、引力辅助高度 h、$\boldsymbol{v}_{\infty,2}$ 在 Ojk 平面的投影与 k 轴夹角参数 φ 来确定：

$$\boldsymbol{v}_{\infty,2}=\boldsymbol{v}_{\infty,1}\left[(\cos\delta)\boldsymbol{i}+(\sin\delta\cos\varphi)\boldsymbol{j}+(\sin\delta\sin\varphi)\boldsymbol{k}\right] \tag{1}$$

式中：$\boldsymbol{i}\parallel\boldsymbol{v}_{\infty,1}$、$\boldsymbol{k}\parallel\boldsymbol{v}_{\mathrm{m}}\times\boldsymbol{v}_{\infty,1}$、$\boldsymbol{j}=\boldsymbol{k}\times\boldsymbol{i}$ 均为单位矢量，组成引力辅助坐标系，其中，“$\parallel$” 表示前后 2 个向量平行。

双曲超速偏转角 δ 与引力辅助高度 h 的关系为

$$\sin\left(\frac{\delta}{2}\right)=\left[1+\frac{(h+R_p)\left|\boldsymbol{v}_{\infty,1}\right|^2}{\mu_{\mathrm{b}}}\right]^{-1} \tag{2}$$

式中：R_p 为引力辅助天体半径；μ_{b} 为引力辅助天体引力常数。

二体模型假设下，由于引力辅助轨道对于引力辅助天体是双曲线，引力辅助前后的双曲超速模长不发生变化，即 $\left|\boldsymbol{v}_{\infty,1}\right|=\left|\boldsymbol{v}_{\infty,2}\right|$。引力辅助的效果是改变了双曲超速矢量方向，即改变了 δ 角度。

于是，引力辅助模型可写为

$$\boldsymbol{\sigma}_2=\mathrm{GA}(\boldsymbol{\sigma}_1,h,\varphi,\boldsymbol{v}_{\mathrm{m}},\mu_{\mathrm{b}}) \tag{3}$$

式中：GA(·)为一般引力辅助模型；$\boldsymbol{\sigma}_1$ 为引力辅助前的轨道根数；$\boldsymbol{\sigma}_2$ 为引力辅助后的轨道根数。

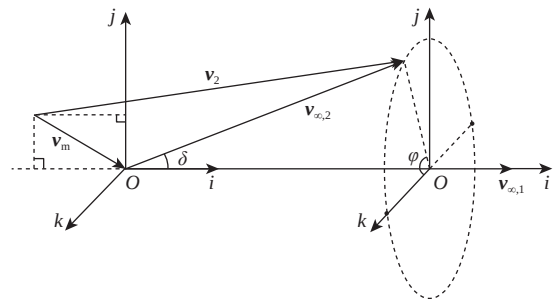


图 1　等效脉冲引力辅助模型示意图

Fig. 1　Illustration of the gravity-assist impulsive model

1.2　共振引力辅助模型

理论上可以构造任意共振比的轨道，一般共振比 n 为整数。以在行星系统中构造 n:1共振轨道为例，即航天器轨道周期 T_{s} 与行星卫星的轨道周期 T_{m} 之比为 n:1：

$$\frac{T_{\mathrm{s}}}{T_{\mathrm{m}}}=n \tag{4}$$

根据椭圆轨道周期与半长轴的关系，得到引力辅助后的轨道半长轴 a_{s} 与卫星的半长轴 a_{m} 满足：

$$a_{\mathrm{s}}=a_{\mathrm{m}}\sqrt[3]{n^2} \tag{5}$$

航天器和卫星关于中心天体的能量方程分别为

$$\frac{1}{2}\left|\boldsymbol{v}_{\mathrm{s}}\right|^2-\frac{\mu}{r}=-\frac{\mu}{2a_{\mathrm{s}}} \tag{6}$$

$$\frac{1}{2}\left|\boldsymbol{v}_{\mathrm{m}}\right|^2-\frac{\mu}{r}=-\frac{\mu}{2a_{\mathrm{m}}} \tag{7}$$

式中：μ 为行星的引力常数；r 为引力辅助时卫星的位矢模长；$\boldsymbol{v}_{\mathrm{s}}$ 为航天器速度。

式 (6) 和式 (7) 相减，并将 a_{s} 用 a_{m} 表示，可得

$$\left|\boldsymbol{v}_{\mathrm{s}}\right|^2=\left|\boldsymbol{v}_{\mathrm{m}}\right|^2+\frac{\mu}{a_{\mathrm{m}}}\left(1-\sqrt[3]{n^{-2}}\right) \tag{8}$$

引力辅助结束后的速度为 $\boldsymbol{v}_{\mathrm{s}}=\boldsymbol{v}_{\mathrm{m}}+\boldsymbol{v}_{\infty}$，引力辅助天体速度垂直于引力辅助坐标系 $\boldsymbol{k}$ 轴，设 $v_{\mathrm{m}i}$、$v_{\mathrm{m}j}$、$v_{\mathrm{m}k}$ 分别为引力辅助天体速度在辅助坐标系中的分量，则引力辅助天体的速度在引力辅助坐标系下可表示为

$$\boldsymbol{v}_{\mathrm{m}}=v_{\mathrm{m}i}\boldsymbol{i}+v_{\mathrm{m}j}\boldsymbol{j} \tag{9}$$

则引力辅助后航天器的速度在引力辅助坐标系下可表示为

$$\boldsymbol{v}_{\mathrm{s}}=(v_{\mathrm{m}i}+|\boldsymbol{v}_{\infty}|\cos\delta)\boldsymbol{i}+\left(v_{\mathrm{m}j}+|\boldsymbol{v}_{\infty}|\sin\delta\sin\varphi\right)\boldsymbol{j}+(|\boldsymbol{v}_{\infty}|\sin\delta\cos\varphi)\boldsymbol{k} \tag{10}$$

将式(9)和式(10)代入式(8),得到

$$\sin\varphi=Q \tag{11}$$

$$Q=\frac{\dfrac{\mu}{a_{\mathrm{m}}}\left(1-\sqrt[3]{n^{-2}}\right)-2v_{\mathrm{m}i}\left|\boldsymbol{v}_{\infty}\right|\cos\delta-\left|\boldsymbol{v}_{\infty}\right|^{2}}{2v_{\mathrm{m}j}\left|\boldsymbol{v}_{\infty}\right|\sin\delta} \tag{12}$$

由式(11)可知,共振引力辅助成立时,应有

$$-1\leqslant Q\leqslant 1 \tag{13}$$

式(11)左端含φ,右端含偏转角δ,该角度由引力辅助高度h决定。φ和h给定一个,可推导出另一个,即可求得引力辅助后的速度矢量。

综上,Q满足式(13)时,引力辅助后成为共振轨道的充要条件为

$$\varphi=\begin{cases}\arcsin Q\\ \pi-\arcsin Q\end{cases} \tag{14}$$

因此,共振引力辅助模型可写为

$$\left[\boldsymbol{v}_{\mathrm{s}}^{\mathrm{res}},\varphi_{\mathrm{s}}^{\mathrm{res}},Q\right]=\mathrm{ResGA}\left(\boldsymbol{v},\boldsymbol{v}_{\mathrm{m}},r_{\mathrm{ga}},n,a_{\mathrm{m}},\mu_{\mathrm{m}},\mu\right) \tag{15}$$

式中:$\boldsymbol{v}$为共振引力辅助前航天器的速度;r_{ga}为引力辅助半径,为式(2)中$h+R_p$项;μ_{m}为引力辅助天体引力常数;$\varphi_{\mathrm{s}}^{\mathrm{res}}$为求出的2个解对应的引力辅助角度参数;$\boldsymbol{v}_{\mathrm{s}}^{\mathrm{res}}$为引力辅助结束后的速度;ResGA(·)为共振引力辅助模型。

需要指出的是,在文献[19]中,由于未能注意到二体模型下$v_{\mathrm{m}k}$=0,导致推导出

$$\sin(\varphi+\xi)=Q' \tag{16}$$

$$\xi=\arctan\left(\frac{v_{\mathrm{m}k}}{v_{\mathrm{m}j}}\right) \tag{17}$$

从而需要对Q'与0的关系进行分情况讨论。进行简化后,本文给出的式(14)更简洁。

2 转移轨道状态序列搜索算法

木星系内对伽利略卫星的探测一般依靠多次引力辅助及多个天体的交替引力辅助,T-r_p图可以帮助设计转移序列。本节对木卫四任务转移轨道设计和优化时用到的T-r_p图进行介绍,并提出转移轨道状态序列搜索算法。

2.1 T-r_p图及其分析

2.1.1 T-r_p图

假设航天器轨道与引力辅助天体轨道共面,且引力辅助天体轨道为圆轨道,引力辅助时,两者轨道的一般关系如图2所示,其中,$\boldsymbol{r}_{\mathrm{m}}$为引力辅助天体的位矢。定义航天器双曲超速$\boldsymbol{v}_{\infty}$与引力辅助天体速度$\boldsymbol{v}_{\mathrm{m}}$的夹角为速度角$\alpha$,根据几何关系可以得出航天器的速度$\boldsymbol{v}_{\mathrm{s}}$为

$$\boldsymbol{v}_{\mathrm{s}}=(|\boldsymbol{v}_{\mathrm{m}}|+|\boldsymbol{v}_{\infty}|\cos\alpha)\boldsymbol{i}+\left[(1-2S_{\mathrm{OI}})|\boldsymbol{v}_{\infty}|\sin\alpha\right]\boldsymbol{j} \tag{18}$$

式中:$\boldsymbol{i}\parallel\boldsymbol{v}_{\mathrm{m}}$、$\boldsymbol{j}\parallel\boldsymbol{r}_{\mathrm{m}}$均为单位矢量,且可定义$\boldsymbol{k}=\boldsymbol{i}\times\boldsymbol{j}$;$S_{\mathrm{OI}}$等于0或1,表示飞越方向。

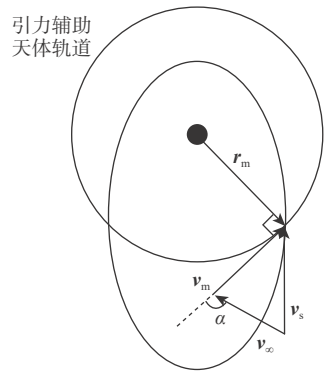


图2 引力辅助时速度角定义

Fig. 2 Definition of velocity angle when gravity assist

已知航天器位矢$\boldsymbol{r}_{\mathrm{s}}=\boldsymbol{r}_{\mathrm{m}}$、速度$\boldsymbol{v}_{\mathrm{s}}$,可以求出轨道能量$E$和动量矩$\boldsymbol{h}_{\mathrm{s}}$为

$$E=\frac{\left|\boldsymbol{v}_{\mathrm{s}}\right|^{2}}{2}-\frac{\mu}{|\boldsymbol{r}_{\mathrm{s}}|} \tag{19}$$

$$\boldsymbol{h}_{\mathrm{s}}=\boldsymbol{r}_{\mathrm{s}}\boldsymbol{v}_{\mathrm{s}} \tag{20}$$

进而求出轨道半长轴a和偏心率e:

$$a=-\frac{\mu}{2E} \tag{21}$$

$$e=\sqrt{1+\frac{2E\left|\boldsymbol{h}_{\mathrm{s}}\right|^{2}}{\mu^{2}}} \tag{22}$$

轨道周期T和近木点半径r_p为

$$T=2\pi\sqrt{\frac{a^{3}}{\mu}} \tag{23}$$

$$r_p=a(1-e) \tag{24}$$

给定引力辅助天体和引力辅助时双曲超速$\boldsymbol{v}_{\infty}$,则对不同的速度角α,对应一系列(T,r_p)点,将一系列点在图中描成曲线,形成了T-r_p图。图3给出了对于木卫二、木卫三、木卫四,双曲超速$\boldsymbol{v}_{\infty}$为特定值的T-r_p曲线,其中,横坐标为对数坐标,R_{J}为木星半径(R_{J}=71 492 km)。图中在各曲线旁标记了对应双曲超速的数值,单位为km/s,并标记了木卫轨道对应的点。一条线上所有点对应α取0°~180°,每隔20°标记一个点做参考。

2.1.2 状态点转移模式

根据引力辅助性质、木卫分布及T-r_p图,可以得知以下信息:

1)木卫轨道半径越大,其改变航天器近木点半径的范围就越大。鉴于木卫三和木卫四是质量最

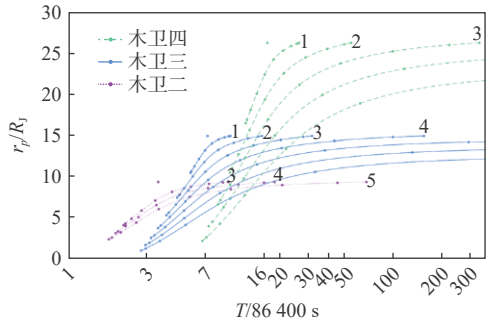


图 3　木卫二、木卫三、木卫四 T-r_p 图

Fig. 3　T-r_p graph of Europa, Ganymede and Callisto

大的木卫, 任务一般选择木卫三引力辅助来减小轨道周期和轨道半长轴, 选择木卫四引力辅助改变近木点半径, 进而改变轨道偏心率。

2) 由于引力辅助过程改变了α, 并不改变相对速度$\boldsymbol{v}_\infty$, 体现在图中, 状态点沿曲线移动到引力辅助后对应的点, 移动的“距离”由式 (2) 决定。

3) 任一木卫曲线上的任意一点表示存在特定周期、近木点的轨道, 满足与该木卫交会时的相对速度为特定值。则不同木卫曲线的交点表示: 在合适的条件下, 存在一个轨道可以同时以特定的相对速度飞越 2 个木卫。

根据以上分析, 可以得到轨道状态点在图上的移动模式: ①共振引力辅助导致轨道状态点沿曲线移动, 移动的终点是给定共振比对应周期的状态点; ②不同木卫曲线交点表示可以在此状态切换引力辅助天体。

对于木卫四环绕任务转移轨道, 由于涉及木卫三、木卫四交替共振引力辅助, 则状态点移动方式为: 沿某曲线移动, 在不同木卫曲线交点处切换引力辅助天体, 沿切换后曲线继续移动。沿曲线移动的总方向为减小与木卫四的双曲超速方向, 过程中一般伴随着周期减小。

2.2　序列搜索算法

根据 2.1.2 节对状态点转移模式的总结, 对于木卫四环绕任务转移轨道, 问题可以转换为状态点根据模式转移至目标状态点的问题。为了搜索转移序列, 需要存储 T-r_p 图中部分曲线信息, 由于共振引力辅助终点为特定周期的轨道, 因而需要另外存储常见共振比对应的轨道状态。

2.2.1　存储 T-r_p 图状态信息

为标记轨道状态点, 使用以下方式来记录航天器轨道状态:

$$S=\{m,\boldsymbol{v}_\infty,T,r_p\} \tag{25}$$

式中: m为飞越木卫索引。

S.C$\{m\}$为木卫m特定双曲超速曲线上点的总集合; S.C.r$\{m,\boldsymbol{v}_\infty\}$为木卫$m$、双曲超速$\boldsymbol{v}_\infty$曲线上特定共振比状态点的集合。

由 2.1.1 节可知, 不同的α对应轨道不同的状态, 存储曲线信息时, 离散点应尽量密集, 即$\Delta\alpha$尽量小, 文中$\Delta\alpha$取为 0.01°。

选择双曲超速值, 应照顾到木卫三、木卫四曲线交会密集区, 进而保证蕴含较多的转移序列。木卫三双曲超速$|\boldsymbol{v}_{\infty,3}|$取值范围为 2～9 km/s, 间隔 1 km/s; 木卫四双曲超速$|\boldsymbol{v}_{\infty,4}|$取值范围为 1～6.5 km/s, 间隔 0.5 km/s, 从而得到S.C{3}、S.C{4}。

涉及对特定共振比状态点的存储, 取共振比范围为 0.5～6, 间隔为 0.5。注意到木卫三、木卫四曲线相交处也包含木卫四共振比小于 1 的区域, 故对于木卫四, 另外保存共振比为 0.4、0.6、0.8 的状态点, 从而得到S.C.r{3}、S.C.r{4}。

考虑一般木星系探测任务的初始轨道周期为 200～300 d, 在考虑的轨道状态点中, 保留周期小于 300 d 的状态点, 存储结果如图 4 所示。图中, T_C 和 T_G 分别为木卫四和木卫三的周期, 圈状离散点标记了特定共振比的状态点。

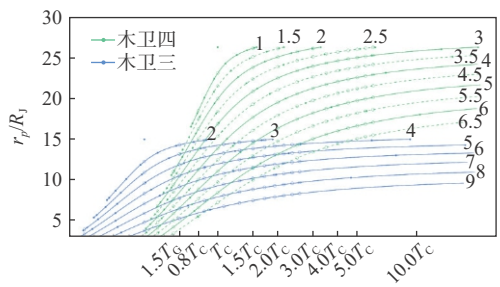


图 4　部分 T-r_p 图

Fig. 4　Part of T-r_p graph

2.2.2　算法流程

考虑初始轨道状态S_0周期为 200 d, 近木点半径为 $10R_J$, 即

$$S_0=\{m,\boldsymbol{v}_\infty,200\text{ d},10R_J\} \tag{26}$$

目标状态为环绕木卫四, 为以较小的速度增量为代价进入环绕轨道, 需要航天器与木卫四双曲超速尽可能小。由 T-r_p 图可知, 木卫四双曲超速 1 km/s 的曲线与木卫三曲线无交点, 即不能通过木卫三共振引力辅助抵达木卫四; 而木卫四双曲超速 1.5 km/s 的曲线与其他曲线有交点, 可以考虑以该状态为目标状态S_t来搜索转移序列, 即

$$S_t=\{4,1.5\text{ km/s},T,r_p\} \tag{27}$$

初始时航天器处在环木大椭圆轨道, 在远木点施加抬高近木点机动(perijove raise maneuver, PJR)后, 确定共振引力辅助天体, 从而完全确定S_0。假设引力辅助天体为m_1, 双曲超速为$\boldsymbol{v}_{\infty,1}$, 则计算此时

曲线上各共振点S.C.r$\{m_1,\boldsymbol{v}_{\infty,1}\}$与另一木卫曲线S.C$\{m_2\}$的交点，由于总的状态点移动方向是减小与木卫的相对速度，故只需计算小于当前状态对应双曲超速曲线的交点。

交点意味着可以在此处通过调相等轨道机动，从木卫m_1的共振轨道上转移到将与木卫m_2引力辅助的轨道上，即状态点由$S_i=\{m_1,\boldsymbol{v}_{\infty,1}\}$转移到$S_{i+1}=\{m_2,\boldsymbol{v}_{\infty,2}\}$，理想情况下，轨道周期与近木点不变。

状态点转移到另一木卫的曲线上，继续上述过程，即求当前曲线上周期为特定共振比状态点与另一木卫曲线的交点，如此迭代下去。当迭代中轨道状态等于目标状态$S_1=\{4,1.5\ \text{km/s}\}$时，算法停止，得到转移轨道状态序列。

2.2.3 估算速度增量

根据以上算法，可以得到多个转移轨道状态序列，通过估计序列所需速度增量对得到的序列进行保留。

速度增量主要有3部分：PJR、中途切换共振引力辅助天体所需速度增量Δv_{ph}和木卫四轨道插入机动（Callisto orbit insertion, COI）。

根据轨道周期与近木点半径，可以求出远木点速度，则根据前后大椭圆轨道周期、近木点半径，即可求出所需PJR。在本节的估算中，假设前后轨道周期不变。

实际上，施加PJR后，轨道在空间中的指向未必可以立即与对应木卫进行引力辅助，还需要额外的机动来改变轨道在空间中的位置。

对于切换引力辅助天体时施加的机动，由于前后轨道形状改变不大，可以视做只需较小的速度增量，而如果状态转变较大，则估计一个较大的速度增量。具体地：

$$\Delta v_{\text{ph}}=\begin{cases}0.05\ \text{km/s} & d<1\\ 0.2\ \text{km/s} & d\geqslant 1\end{cases} \tag{28}$$

式中：d为前后状态点在图上的距离，表示为

$$d=\left|\left(T_{i+1},r_{p,i+1}\right)-\left(T_i,r_{p,i}\right)\right| \tag{29}$$

仿真中，环绕木卫四轨道设置为距木卫四500 km圆轨道。COI的大小可根据到达双曲超速，利用能量守恒求得。

综上，对于序列的速度增量Δv_{seq}估计为

$$\Delta v_{\text{seq}}=\Delta v_{\text{PJR}}+\sum_i \Delta v_{\text{ph},i}+\Delta v_{\text{COI}} \tag{30}$$

式中：Δv_{PJR}为PJR的大小；Δv_{COI}为被木卫四捕获时所需的速度增量。

综上，算法可描述为：①根据初始状态S_0决定引力辅助天体m和相对速度，求出Δv_{PJR}；②求解当前曲线上各共振点S.C.r与另一木卫曲线S.C$\{m'\}$的交点，求解所需速度增量Δv_{ph}；③重复步骤②，直至状态为等于目标状态S_1；④求解得到序列对应的估计速度增量Δv_{seq}。

3 仿真结果与分析

3.1 序列搜索结果

仿真中增加了近木点半径约束，避开木星附近强烈辐射：

$$r_p>9R_{\text{J}} \tag{31}$$

保留速度增量小于1.6 km/s的解：

$$\Delta v_{\text{seq}}<1.6\ \text{km/s} \tag{32}$$

仿真使用的计算机CPU型号为Intel(R) Core (TM) i5-13500H，主频为2.6 GHz，单核，优化用时约为10 s。仿真结果如表1所示。

表1 S_0序列搜索结果

Table 1 Result of sequence search

m	$\lvert\boldsymbol{v}_\infty\rvert$/(km·s^{-1})	T/d	r_p/R_{J}	Δv_{seq}/(km·s^{-1})
3	5	200	10	1.398
3	6	200	10	1.465
3	7	200	10	1.529
4	5	200	10	1.548
4	6	200	10	1.568

可见，算法给出5个转移序列，其中包括3个先进行木卫三引力辅助的序列和2个先进行木卫四引力辅助的序列。给出以上2种情形下各自速度增量最小的状态序列转移图，对应的初始状态分别记为S_1、S_2，如图5和图6所示。

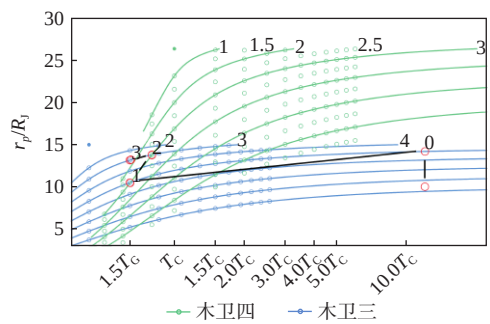


图5 S_1转移序列($S_0=(3,5\ \text{km/s},200\ \text{d},10R_{\text{J}})$)

Fig. 5 Transfer sequence of S_1 ($S_0=(3,5\ \text{km/s},200\ \text{d},10R_{\text{J}})$)

3.2 转移轨道优化结果

在求出的序列中，序列S_2只涉及2次木卫间转移，考虑木卫间转移通常需要较长的转移时间，综合速度增量与转移时间后，S_2为最优转移序列。建立对该序列的转移轨道分步求解模型，并将结果与之前类似序列的转移做对比。

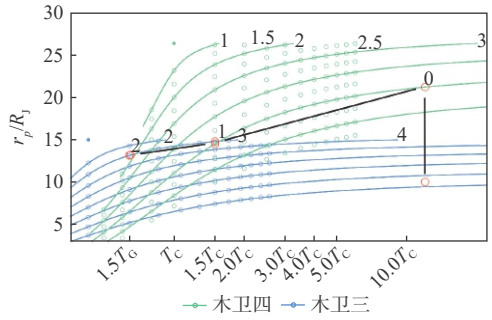


图 6　S_2转移序列($S_0=(4,5\ \text{km/s},200\ \text{d},10R_J)$)

Fig. 6　Transfer sequence of S_2 ($S_0=(4,5\ \text{km/s},200\ \text{d},10R_J)$)

序列 S_2 对应的转移策略为：首先，航天器从初始轨道抬高近木点半径，到可以与木卫四双曲超速 5 km/s 的状态上进行引力辅助，经过多次木卫四共振引力辅助后抵达共振比为 1.5 的轨道；然后，进行木卫间转移，转移后能以与木卫三双曲超速 3 km/s 的状态进行引力辅助，再经过多次木卫三共振引力辅助后抵达共振比为 1.5 的轨道；最后，进行木卫间转移，转移后状态与木卫四双曲超速为 1.5 km/s，进而进入环绕木卫四的轨道。

3.2.1　转移轨道优化模型

行星际段及木卫引力辅助捕获段优化采用文献 [21] 中的模型，通过优化可以得到航天器在近木点、施加木星轨道插入机动(Jupiter orbit maneuver, JOI)前的状态。随后，分段优化木卫四任务段。先优化 JOI 和 PJR，优化变量为

$$X_1=\left[\Delta v_{\text{JOI}},t_{\text{apo2m}}\right] \tag{33}$$

式中：Δv_{JOI}为 JOI 的大小；t_{apo2m}为远木点到第 1 次木卫四引力辅助的转移时间，通过求解对应的 Lambert 问题，即可求出 PJR。JOI 规定为沿速度切向，优化范围由施加切向机动后椭圆轨道的周期 T_{cap}限制，仿真中，限定 200 d<T_{cap}<350 d。

由于初始轨道与木星赤道面有一定的夹角，导致轨道状态点并不严格按照曲线移动，故需要适当放宽约束。第 1 次引力辅助的约束条件为交会木卫四时的相对速度$\boldsymbol{v}_{\infty,1}$，由图 6 可知：

$$\left|\boldsymbol{v}_{\infty,1}\right|<4.8\ \text{km/s} \tag{34}$$

目标函数为

$$J_1=\Delta v_{\text{JOI}}+\Delta v_{\text{PJR}} \tag{35}$$

为通过共振引力辅助尽快减小轨道周期，在每次共振引力辅助前，先确定共振比n，本文采用较为直接的方式来确定共振比。Q 的表达式(式 (12))可以视作共振比n与其他量的函数，简写为$Q=f(n)$。从小到大遍历可能的n，对每个n求出对应的 Q，判断 Q 是否满足式 (13)，若满足，共振比则确定。该方式的好处是通过解析法可以快速确定每次的共振比，由小到大的遍历也可以保证以最快的速度减小轨道周期。

需要注意的是，共振引力辅助程序可以计算出 2 个符合共振比的轨道，选择两者中倾角较小的解。这样选择可以保证，即使初始轨道相对木星赤道面有夹角，也能逐渐使倾角降低。由于木卫真实运动不是两体运动及航天器与木卫轨道面之间存在一个小的偏差，导致经过根据共振比计算出转移时间后，木卫位置会与上次交会位置有偏差，于是增加远木点机动来修正这次共振引力辅助的位置，通过小范围优化远木点到理论交会处的转移时间来确定机动大小。

2 次共振引力辅助终止条件为：对于木卫四共振引力辅助，目标共振比为 1.5，即图 6 中序号 1 处；对于木卫三共振引力辅助，目标共振比为 1.5，即图 6 中序号 2 处。

过程中还涉及木卫间转移，通过求解多圈 Lambert 问题来确定转移轨道。对于木卫三转移到木卫四，优化变量为

$$X_2=[t_0,t_{\text{f}}] \tag{36}$$

式中：t_0为在原轨道停泊时长；t_{f}为转移时间。

根据图 6，对飞越目标木卫时的相对速度有要求，故约束为

$$\left|\boldsymbol{v}_{\infty,2}\right|<3\ \text{km/s} \tag{37}$$

目标函数为转移所需速度增量Δv_{G2C}：

$$J_2=\Delta v_{\text{G2C}} \tag{38}$$

式中：Δv_{G2C}为木卫三转移到木卫四所需速度增量。

类似的，对于木卫三转移到木卫四并被捕获，优化变量为停泊时长及转移时间，目标函数为

$$J_3=\Delta v_{\text{C2G}}+\Delta v_{\text{COI}} \tag{39}$$

式中：Δv_{C2G}为木卫四转移到木卫三所需速度增量。

需要注意的是，在一定的转移时间内，可能不存在经济的转移轨道。可采取以下 2 种策略进行求解：①在任务时间允许的情况下，可适当放宽转移时间；②可以在优化转移轨迹时将最后一次共振引力辅助的参数也纳入到优化变量中，进而得到所需速度增量小的转移轨道。

3.2.2　转移轨道仿真结果

仿真起点为 JOI 前的状态，根据文献 [21]，具体状态如表 2 所示。

捕获后的目标轨道考虑 2 种，分别为文献 [13] 中高度 500 km 的圆轨道(C)和文献 [14] 中近地点高度 100 km、远地点高度为 3 倍木卫四半径的椭圆

轨道(E)。

仿真使用的计算机CPU型号为Intel(R) Core (TM) i5-13500H, 主频为2.6 GHz, 单核, 优化用时约为2 min。优化后的轨道数据如表3所示, 轨道在木星平赤道系中*Oxy*平面投影如图7～图9所示。图中, 红色线半径由小到大分别表示木卫一、木卫二、木卫三、木卫四的轨道, 蓝色线表示木卫间转移轨道, 黑色线表示共振轨道; “*”表示木卫间转移时施加机动的位置; 红点分别标记了木卫间转移段中, 航天器与木卫三、木卫四的交会点。转移轨道为顺行轨道, 即图中逆时针方向。

表2 木心J2000系下初始轨道状态

Table 2 The initial orbit state referenced in Jupiter-centered J2000 system

星历时间JD	半长轴/km	偏心率	倾角/(°)	升交点赤经/(°)	近地点幅角/(°)	真近点角/(°)
2 464 824.508	−7 593 520.642	1.094	22.535	350.752	295.361	0

表3 木星系内航天器轨道主要数据

Table 3 Main orbital data of spacecraft in Jovian system

事件	UTC时间	速度/(km·s^{-1})	共振比
JOI	2036-05-11	0.679	
PJR	2036-10-28	0.232	
C1	2037-04-22		5
C2	2037-07-14		3
C3	2037-09-02		2
C4	2037-10-06		1.5
Δv_{C2G}	2037-11-26	0.041	
G1	2038-01-18		2.5
G2	2038-02-22		1.5
Δv_{G2C}	2038-06-21	0.043	
COI	2038-06-28	1.086(C)/0.670(E)	

注: C1、C2、C3、C4为木卫四共振引力辅助; G1、G2为木卫三共振引力辅助。

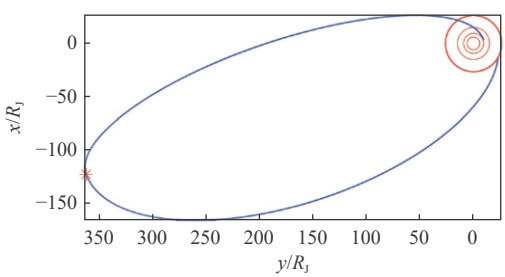


图7 被木星捕获轨迹

Fig. 7 Trajectory of being captured by Jupiter

表3表明, 航天器在远木点施加PJR后, 依次进行了4次木卫四共振引力辅助(C1～C4), 共振比分别为5、3、2、1.5, 随后施加1次机动交会木卫三; 再进行2次木卫三共振引力辅助(G1、G2), 共振比分别为2.5、1.5, 再施加1次机动交会木卫四并被捕获。木星系内, 施加JOI到被环绕木卫四, 任务时长为2.13年, 加上其他机动0.027 km/s, 总速度增量为2.108 km/s和1.691 km/s, 分别对应圆形环绕轨道和椭圆环绕轨道。

表4对比了与文献[13-14]中方案的速度增量。可见, 如果只考虑木星系内变轨, 本文方案比对比方案节省约100～200 m/s。

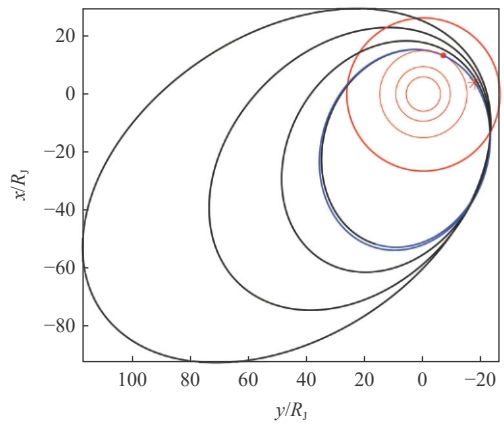


图8 木卫四共振引力辅助轨道及木卫四-木卫三转移轨道

Fig. 8 Trajectories with Callisto resonant gravity assist and Callisto-Ganymede transfer trajectory

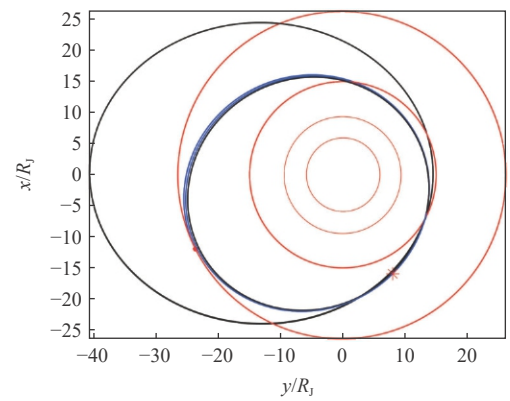


图9 木卫三共振引力辅助轨道及木卫三-木卫四转移轨道

Fig. 9 Trajectories with Ganymede resonant gravity assist and Ganymede-Callisto transfer trajectory

表4 各方案结果对比

Table 4 Comparison of outcomes across different strategies

方案	出发卫星逃逸地球时的能量/(km^2·s^{-2})	行星际速度增量/(km·s^{-1})	木星系内速度增量/(km·s^{-1})
文献[13]	17.771	0.708	2.335(C)
文献[14]	13.26	～0.4	～1.820(E)
本文	14.305	0.010	2.108(C)/1.691(E)

需要注意的是, 任务规划上, 文献[14]先进行木卫二探测, 再进行木卫三探测, 最后被木卫四捕获, 由于任务中还有木卫二、木卫三探测阶段, 可能需要相比仅木卫四捕获任务更多的速度增量。而

本文和文献 [13] 均为被木星捕获后直接前往木卫四捕获, 故重点分析与文献 [13] 的不同。

如果考虑行星际段的消耗, 文献 [13] 中行星际段速度增量为 0.708 km/s, 而本文的行星际段参考文献 [21], 速度增量仅为 0.010 km/s。因而从整个方案来看, 本文方案比文献 [13] 节省约 0.7 km/s。

图 10 给出了航天器全程状态的变化, 红色点标记了每次引力辅助及机动前后的状态。

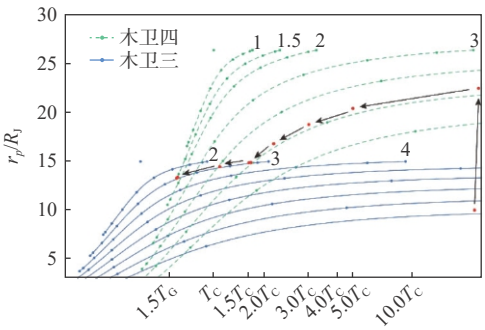


图 10　航天器轨道状态变化

Fig. 10　Changes in the spacecraft orbital status

本文木星系内任务时长为 2.13 年。文献 [13] 中采用 7 次木卫四共振引力辅助减小轨道半长轴和倾角, 联合优化木卫四共振引力辅助和深空机动最终被木卫四捕获的方案, 7 次共振引力辅助后的共振比分别为 6、4、3、2、2、2、2, 导致木星系内任务时长较长, 约为 3.49 年。本文任务时长节省约 1 年。

速度增量方面, 本文在优化木卫间转移轨道时求解多圈 Lambert 问题, 相比直接转移, 减小了速度增量消耗。从文献 [13] 方案中 COI 的值可以反推出捕获时相对木卫四的速度约为 1 km/s。7 次木卫四共振引力辅助后到达共振比为 2:1 的轨道, 后续需要从木卫四 2:1 共振轨道转移到与木卫四相对速度为 1 km/s 的状态, 然而根据 T-r_p 图来看, 无法仅通过木卫四引力辅助实现, 该阶段可以考虑通过木卫三引力辅助实现小代价转移。文献 [13] 为了将状态改变至与木卫四相对速度为 1 km/s 的状态, 没有进行木卫三引力辅助, 而是进行了 3 次较大的深空机动, 导致总速度增量较大。

4　结　论

1) 木卫四转移轨道状态序列搜索算法可以高效求解出多个转移序列, 解决了人工参与度高的问题。

2) 转移轨道优化算法可以高效求出序列搜索算法得到的序列对应的转移轨道。相比对比算法, 本文木星系内速度增量节省约 100～200 m/s, 任务时间节省约 1 年。

参考文献（References）

[1] VIDMACHENKO A P, STEKLOV A F. Life in Europa and on the Earth[C]//Proceedings of the 10th International Scientific and Practical Conference. Manchester: Cognum Publishing House, 2022: 264-273.

[2] LOEB A A, LAUKIEN F H. Overview of the Galileo project[J]. Journal of Astronomical Instrumentation, 2023, 12: 2340003.

[3] BOLTON S J, LUNINE J, STEVENSON D, et al. The Juno mission[J]. Space Science Reviews, 2017, 213(1-4): 5-37.

[4] STEVENSON D J. Jupiter's interior as revealed by Juno[J]. Annual Review of Earth and Planetary Sciences, 2020, 48: 465-489.

[5] 王建昭, 卜良霄. 木星系探测与 JUICE 任务概述[J]. 宇航学报, 2023, 44(6): 957-965.
WANG J Z, BU L X. Overview of Jupiter system exploration and JUICE mission[J]. Journal of Astronautics, 2023, 44(6): 957-965(in Chinese).

[6] BAYER T, BITTNER M, BUFFINGTON B, et al. Europa Clipper mission: preliminary design report[C]//Proceedings of the 2019 IEEE Aerospace Conference. Piscataway: IEEE Press, 2019: 1-24.

[7] PAPPALARDO R T, VANCE S, BAGENAL F, et al. Science potential from a Europa Lander[J]. Astrobiology, 2013, 13(8): 740-773.

[8] MCEWEN A, TURTLE E, HIBBARD K, et al. Io Volcano Observer (IVO): budget travel to the outer solar system[J]. Acta Astronautica, 2014, 93: 539-544.

[9] MARTYNOV M B, MERKULOV P V, LOMAKIN I V, et al. Advanced Russian mission Laplace-P to study the planetary system of Jupiter: scientific goals, objectives, special features and mission profile[J]. Solar System Research, 2017, 51(7): 555-562.

[10] CLARK K, BOLDT J, GREELEY R, et al. Return to Europa: overview of the Jupiter Europa Orbiter mission[J]. Advances in Space Research, 2011, 48(4): 629-650.

[11] 中华人民共和国国务院新闻办公室. 2021 中国的航天(2022 年 1 月)[M]. 北京: 人民出版社, 2022.
The State Council Information Office of the People's Republic of China. China's space program: a 2021 perspective (January 2022)[M]. Beijing: People's Publishing House, 2022(in Chinese).

[12] 张佳文. 太阳系边际探测任务轨道优化设计[D]. 北京: 中国科学院大学 (中国科学院国家空间科学中心), 2020.
ZHANG J W. Trajectory design and optimization for the solar system boundary exploration mission[D]. Beijing: National Space Science Center, University of Chinese Academy of Sciences, 2020(in Chinese).

[13] 陈诗雨, 杨洪伟, 宝音贺西. 木星系探测及行星穿越任务轨迹初步设计[J]. 深空探测学报, 2019, 6(2): 189-194.
CHEN S Y, YANG H W, BAOYIN H X. Preliminary design for the trajectories of Jovian and planetary mission[J]. Journal of Deep Space Exploration, 2019, 6(2): 189-194(in Chinese).

[14] 高博宇. 木星探测任务分析与轨道设计研究[D]. 北京: 中国空间技术研究院, 2022.
GAO B Y. Analysis and orbital design research for Jupiter exploration mission[D]. Beijing: China Academy of Space Technology, 2022(in Chinese).

[15] 田百义, 张熇, 冯昊, 等. 面向木星卫星交会任务的探测器飞行路径规划[J]. 宇航学报, 2022, 43(12): 1587-1596.
TIAN B Y, ZHANG H, FENG H, et al. Flight path planning for rendezvous mission of Jovian probe with the Jupiter's moons[J]. Journal of Astronautics, 2022, 43(12): 1587-1596(in Chinese).

[16] EISMONT N A, ZUBKO V A, BELYAEV A A, et al. Gravity assists maneuver in the problem of extension accessible landing areas on the Venus surface[J]. Open Astronomy, 2021, 30(1): 103-109.

[17] SIMS J A, LONGUSKI J M, STAUGLER A J. V8 leveraging for interplanetary missions: multiple-revolution orbit techniques[J]. Journal of Guidance, Control, and Dynamics, 1997, 20(3): 409-415.

[18] LANTUKH D V, RUSSELL R P, CAMPAGNOLA S. V-infinity leveraging boundary-value problem and application in spacecraft trajectory design[J]. Journal of Spacecraft and Rockets, 2015, 52(3): 697-710.

[19] 杨洪伟, 陈杨, 宝音贺西, 等. 共振引力辅助轨道的设计方法及应用[C]//中国宇航学会深空探测技术专业委员会第九届学术年会. 北京: 中国宇航学会深空探测技术专业委员会, 2012.
YANG H W, CHEN Y, BAOYIN H X, et al. Design methods and applications of resonant gravitational assisted orbits[C]//Proceedings of the 9th Annual Conference of the Deep Space Exploration Technology Committee of the Chinese Society of Astronautics. Beijing: Deep Space Exploration Technology Committee of the Chinese Society of Astronautics, 2012(in Chinese).

[20] CAMPAGNOLA S, BUFFINGTON B B, PETROPOULOS A E. Jovian tour design for orbiter and lander missions to Europa[J]. Acta Astronautica, 2014, 100: 68-81.

[21] JING Q, LI M T. Integrated trajectory optimization for Jupiter missions with single-satellite-aided capture[J]. Celestial Mechanics and Dynamical Astronomy, 2023, 135(6): 57.

Sequence search and optimization for transfer trajectory status in Callisto exploration mission

JING Quan[1, 2], LI Mingtao[1, 2, *], WANG Youliang[1]

(1. National Space Science Center, Chinese Academy of Science, Beijing 100190, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: A graphical analysis-based search algorithm for the transfer orbit state sequence is presented to address the problem of high manual involvement when designing the transfer orbit for the Callisto orbiting mission using graphical analysis approaches. Through the analysis of graphical properties and resonance gravity assists, the variation pattern of orbit states is summarized, and an iterative search algorithm for solving the state transition sequence is proposed. The algorithm can obtain multiple transfer sequences rapidly and efficiently, as demonstrated by the simulation of the transfer orbit for the Callisto orbiting mission. The solution of the corresponding transfer orbits for the obtained sequences shows that the spacecraft can achieve the transfer from a Jovian highly elliptical orbit to the Callisto orbiting orbit within 2.13 years by consuming a velocity increment of 2.108 km/s, saving approximately 100-200 m/s of velocity increment and about 1 year of transfer time compared to other methods. The method proposed in this paper solves the problem of high manual involvement when using graphical analysis methods.

Keywords: Jupiter system; Callisto exploration; trajectory optimization design; gravity assist; resonance gravity assist

Received: 2024-03-20; **Accepted**: 2024-04-26; **Published Online**: 2024-05-13 09:57
URL: link.cnki.net/urlid/11.2625.V.20240511.1535.001
Foundation item: Youth Innovation Promotion Association, CAS(2022146)
* **Corresponding author**. E-mail: limingtao@nssc.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0134

基于风险概率分解的RTK完好性评估方法

宋远[1]，黄智刚[1]，李锐[1,*]，王岳辰[2]，沈军[2]，王永超[3]，聂欣[4]

(1. 北京航空航天大学 电子信息工程学院，北京 100191；2. 北京合众思壮科技股份有限公司，北京 100176；
3. 民航数据通信有限责任公司，北京 100191；4. 中国空间技术研究院，北京 100094)

摘 要：实时动态定位(RTK)技术具有高精度和实时动态特性，在自动驾驶领域受到广泛关注，如何定量评估RTK定位完好性是否满足安全性能需求成为亟待解决的关键问题。当前针对RTK完好性的研究大多集中在用户端的算法改进与优化上，少有系统层面的风险分析与定量评估。基于此，提出一种基于风险概率分解的RTK完好性评估方法，并选取中高纬和低纬典型地区的多个观测站，使用不同类型接收机组合及不同年份的观测数据进行RTK完好性定量评估。结果显示，在中高纬电离层平静区域，定位完好性随接收机数据质量不同而存在明显差异，而在低纬度电离层活跃地区及活跃年份，电离层异常成为主要因素，完好性风险明显加大。根据评估结果，当前基于单频单星座及中等长度基线进行RTK定位时，其整体完好性风险概率位于10^{-4}～10^{-2}之间。所得结论与相关RTK数据处理的普遍结论一致，验证了所提方法的合理性，完好性的定量评估结果可作为后续各项监测器设计的重要依据。

关 键 词：完好性；完好性风险评估；RTK完好性；自动驾驶；接收机数据质量
中图分类号：V249.3; U666.134
文献标志码：A **文章编号**：1001-5965(2026)05-1605-10

实时动态定位(real-time kinematic, RTK)是一种广泛使用的基于全球卫星导航系统(global navigation satellite system, GNSS)的差分定位技术，在理想情况下能够快速收敛，达到动态厘米级和静态毫米级的定位精度。随着多模多频接收机成本的下降，RTK在自动驾驶领域受到越来越多的关注[1]。然而，在城市或市郊环境中，GNSS系统容易受到多径、非视距误差及射频干扰等影响，在低纬度地区容易受到电离层异常的影响，导致RTK性能下降，难以求得可靠的双差模糊度而产生错误的定位解，影响导航系统完好性，引起较大的安全风险。因此，如何对RTK完好性进行定量评估与针对性的提升是推进其在自动驾驶领域获得广泛应用的关键。

目前，对GNSS完好性的研究主要集中在航空应用中。在民航领域，对GNSS完好性的提升主要通过增强系统的建设来实现，包括星基增强系统(satellite-based augmentation system, SBAS)、地基增强系统(ground-based augmentation system, GBAS)和空基增强系统(airborne-based augmentation system, ABAS)[2]。国际民航组织(International Civil Aviation Organization, ICAO)根据航空飞行中对导航服务性能的要求定义了不同飞行阶段，并给出了对各个阶段的性能指标要求[3]。其中，SBAS和GBAS系统目前已能够支持一类垂直引导进近(approach with vertical guidance-I level, APV-I)和一类精密进近(category I, CAT-I)阶段的进近服务，达到2×10^{-7}/150 s的完好性指标要求。但由于民航用户均采用平滑后的伪距观测值进行定位，定位精度无法达到亚米级。此外，需要建立专用监测站、使用多台高质量

收稿日期：2024-03-07；**录用日期**：2024-04-19；**网络出版时间**：2024-05-08 14:01
网络出版地址：link.cnki.net/urlid/11.2625.V.20240507.1611.003
基金项目：国家重点研发计划(2022YFB3904302)；CAST创新基金(F-W-YY-2023-019)
*通信作者. E-mail：lee_ruin@263.net

引用格式：宋远，黄智刚，李锐，等. 基于风险概率分解的RTK完好性评估方法[J]. 北京航空航天大学学报，2026，52(5)：1605-1614.
SONG Y，HUANG Z G，LI R，et al. RTK integrity evaluation method based on risk probability decomposition[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1605-1614 (in Chinese).

接收机、多冗余备份的通信链路等一系列设施来保障监测网络的高可靠性。

在军用航空领域，美国联合精密进近与着陆系统(joint precision approach and landing system, JPALS)已完成初始性能认证[4]，使用双频载波观测值进行差分定位，能够实现全天候自主着舰，具备较强的抗干扰性能，能够达到10^{-6}/150 s完好性风险和1.5 m垂直告警门限(alert limit, AL)的指标要求。JPALS需要冗余的高质量舰载监测设备、机载抗干扰接收机、专用通信链路、干扰监测和抑制模块、双频P码及高质量L1/L2载波来保障定位完好性[5-6]，且各类完好性增强系统都建立在对风险源分析和完好性指标分解的基础上。

自动驾驶领域对定位精度和安全性均有较高的要求，其中，对精度的要求可分为道路级(<5 m)、车道级(<1.5 m)和车道内(<1 m)[7]。在安全性方面，美国汽车工程师学会制定的ISO 26262标准定义了4类汽车安全完整性等级(automotive safety integrity level, ASIL)，分别为ASIL-A/B/C/D [8]。Reid等基于ASIL-D，给出了针对整体车载定位系统10^{-8}/h定位故障概率和分米级AL的风险要求[9-10]。基于ASIL-B，文献[11]提出了针对单GNSS系统10^{-7}/h的完好性风险指标，对应3 m AL和开阔天空条件下99%的可用性，这也是本文后续进行风险分析与评估时所依照的指标。此外，根据中国公路工程技术标准中规定的车道宽度和车辆宽度，若要实现可靠的车道级导引，95%侧向定位误差应优于0.3～0.5 m[12]。

若要将RTK应用于自动驾驶，对精度的要求易于实现，而为了实现对安全性的要求，需要提升RTK定位的完好性，这依赖于对RTK故障源及误差特性的细致分析及对风险的定量评估。尽管RTK能够消除大部分公共误差，但部分与接收机和环境相关的误差项难以完全消除，反而因差分对其有放大作用[13-14]。部分学者将GNSS系统与惯导、转速计、雷达或3D城市模型等组合，进行粗差探测[15-19]，并提出了基于接收机自主完好性监测(receiver autonomous integrity monitoring, RAIM)或先进接收机自主完好性监测(advanced receiver autonomous integrity monitoring, ARAIM)的RTK完好性监测方法[1,20]，以保障定位误差超出AL的概率小于所给定的完好性风险。

然而，当前对RTK完好性的研究主要集中在用户端算法的提升与改进上，包括误差模型的改进、完好性监测方法的性能提升及保护级(protection level, PL)的优化方法等[20-21]，该类做法存在以下问题：①研究使用的风险指标(如整体完好性风险概率和AL的数值)大多是直接给定，没有给出具体的指标来源与合理性分析；②将保障RTK定位完好性的职责完全施加于用户端，用户端故障监测和PL计算方法复杂，不具有普适性，且未考虑接收机性能对定位完好性的影响；③所建立的误差模型参数对地域的依赖性很强，且大多只选取短时间的数据来验证PL的改进效果，无法证明基于所提方法的整体定位完好性能够达到所设定的指标。

根据国际标准中对完好性的定义[3,22]，完好性是导航系统提供信息正确性的可信程度，用以衡量导航服务的可靠性，体现为系统为用户提供及时有效告警的能力。当前对RTK完好性的研究更多关注PL的计算，然而PL只是系统内在完好性的一种外在体现形式，是误差模型和卫星几何构型等变量的函数，其目的是向用户提供超限告警的能力，应建立在对置信度的评估与系统初始风险概率的计算之上。不对RTK系统的原始风险和可信性进行评估和度量，而只关注PL算法的改进与提升，是对完好性本质的忽略，也是上述各种问题出现的原因。

为此，本文从完好性的基础定义出发，对RTK的通用架构进行分析，研究RTK定位中各环节所面临的风险来源及各自对整体风险的影响。为保障结果的保守性，选取单频单星座及中等长度基线定位场景，对当前RTK所能达到的完好性等级进行初步评估。在此基础上，研究通用架构的短板，为RTK完好性的提升奠定理论基础。

1 RTK完好性评估方法

1.1 RTK相对定位技术

RTK是一种实时动态精密相对定位技术，通过站星双差来消除公共误差项，定位过程中构建载波和伪距双差观测方程如下：

$$\begin{cases}\phi_{\mathrm{br},i}^{jk}=\dfrac{1}{\lambda_i}\left(r_{\mathrm{br}}^{jk}+T_{\mathrm{br}}^{jk}-I_{\mathrm{br},i}^{jk}+M_{\phi,\mathrm{br},i}^{jk}\right)+N_{\mathrm{br},i}^{jk}+\varepsilon_{\phi,\mathrm{br},i}^{jk}\\ P_{\mathrm{br},i}^{jk}=r_{\mathrm{br}}^{jk}+T_{\mathrm{br}}^{jk}-I_{\mathrm{br},i}^{jk}+M_{P,\mathrm{br},i}^{jk}+\varepsilon_{P,\mathrm{br},i}^{jk}\\ r_{\mathrm{br}}^{jk}=\left(|\boldsymbol{r}^k-\boldsymbol{r}_{\mathrm{r}}|-|\boldsymbol{r}^k-\boldsymbol{r}_{\mathrm{b}}|\right)-\left(|\boldsymbol{r}^j-\boldsymbol{r}_{\mathrm{r}}|-|\boldsymbol{r}^j-\boldsymbol{r}_{\mathrm{b}}|\right)\end{cases}\tag{1}$$

式中：$\phi_{\mathrm{br},i}^{jk}$和$P_{\mathrm{br},i}^{jk}$分别为第$i$个频点的载波和伪距双差观测值；$\lambda_i$为第$i$个频点波长；$\boldsymbol{r}$为三维位置坐标；下标b和r分别表示基准站和移动站；上标j和k为卫星标识，其中，j表示参考星；r_{br}^{jk}为双差距离真值，其中包含未知的移动站位置信息$\boldsymbol{r}_{\mathrm{r}}$；$N_{\mathrm{br},i}^{jk}$为未知的双差模糊度；$T_{\mathrm{br}}^{jk}$和$I_{\mathrm{br},i}^{jk}$分别为双差对流层误差和第$i$个频点双差电流层误差，在短基线条件下其影响可以忽略；$M_{\phi,\mathrm{br},i}^{jk}$和$M_{P,\mathrm{br},i}^{jk}$分别为第$i$个频点双差载波和伪距多径误差；$\varepsilon_{\phi,\mathrm{br},i}^{jk}$和$\varepsilon_{P,\mathrm{br},i}^{jk}$分别为第$i$个频点接收机

端载波相位和伪距双差的观测噪声。

观测方程 (1) 中的未知数包括移动站位置及整周模糊度。构建出观测方程后, 使用卡尔曼滤波器进行参数估计, 再利用 Bootstrap、最小二乘模糊度降相关平差 (least-squares ambiguity decorrelation adjustment, LAMBDA) 算法[23-25] 等求解模糊度整数解, 进而达到厘米级定位精度。

1.2　RTK 完好性风险来源分析

在 RTK 定位过程中, 一部分公共误差项(如星历星钟误差)可以通过双差消除, 但一些与传播路径、观测站环境和接收机相关的误差项仍然会保留, 甚至放大, 且由于整体数据传输过程中没有监测和校验环节的存在, 所有环节对定位完好性的风险项会相互叠加, 导致整体风险增加。对各风险项的分析评估是进行完好性指标分解与各项监测器设计的基础, 也是当前研究缺失的部分, 因此, 本文首先分析风险来源, 在此基础上对各风险项进行定量评估, 得到 RTK 完好性的初始评估结果。图 1 给出了 RTK 定位中面临的卫星段、传播段、基准站、通信链路和用户段的具体风险来源。

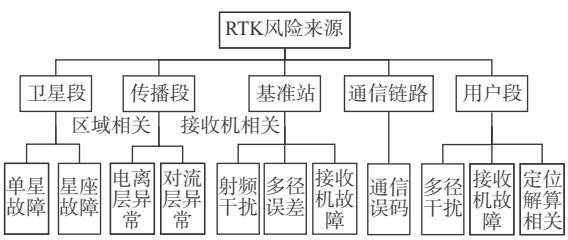


图 1　RTK 风险来源

Fig. 1　RTK risk sources

其中, 卫星段风险主要来源于卫星信号故障和电文故障, 根据各大 GNSS 在性能服务标准中的承诺[3], 全球定位系统(global positioning system, GPS)、北斗卫星导航系统(BeiDou navigation satellite system, BDS)、GLONASS 卫星导航系统及 Galileo 卫星导航系统的单星故障概率分别为 10^{-5}、10^{-5}、10^{-4} 及 3×10^{-5}, 星座故障概率分别为 10^{-8}、6×10^{-5}、10^{-4} 及 2×10^{-4}。本文使用单 GPS 系统进行 RTK 定位, 其星座故障概率远低于单星故障概率, 因此, 忽略星座故障对定位完好性的影响。假设平均可见卫星数为 10, 则有一颗卫星出现故障的概率为 10^{-4}。此外, 考虑利用外部辅助信息(如 SBAS 增强信息)来选择健康卫星, 并利用历史信息判断电文正确性, 因而卫星段风险有望得到显著降低。

传播段的风险主要来源于电离层和对流层异常。一般情况下, 当用户与基准站距离较近时, 通过站间单差可基本消除电离层和对流层误差。然而, 当出现电离层暴或对流层暴等异常情况时, 电离层和对流层具有较大的空间梯度, 无法通过差分完全消除。根据 Gallon 等的研究[26], 对流层异常对定位误差的影响主要集中在垂直方向, 因而暂时不考虑对水平定位要求更高的自动驾驶应用场景。而电离层异常与季节和区域具有较强的相关性, 且存在 11 年周期变化特性, 因而无法给出通用和明确的风险概率指标。本文通过对比典型低纬度地区在电离层活跃期和平静期的 RTK 完好性评估结果, 分析电离层异常对定位完好性的影响。

基准站和用户段的风险主要来源于接收机附近的射频干扰、多径误差及接收机故障, 与接收机数据质量密切相关。对于用户段, 需额外考虑定位算法引入的风险项; 对于基准站, 可以在选址阶段充分考虑对射频干扰和多径的规避, 且一般采用性能较好的接收机, 因此, 其对定位完好性的影响小于用户段。本文通过对几种典型的接收机组合进行 RTK 定位, 评估不同数据质量的接收机组合对应的完好性风险。

通信链路的风险主要来源于通信误码, 按照目前 4G 通信误码率来考虑, 其风险概率位于 10^{-5}～10^{-4} 之间, 如果能够通过重复播发或使用 2 种链路同时播发, 能够将该项风险进一步降低。

综合上述分析可知, 对于卫星段和通信链路的风险, 有较为明确的成因及数值, 通过改进措施可以降低至很小的范围, 而电离层异常和接收机数据质量是影响 RTK 完好性的主要因素, 后续将针对其展开更细致的分析。

1.3　RTK 完好性定量评估方法

本节研究基于风险概率分解的 RTK 完好性定量评估方法。根据定义, 完好性风险可以描述为定位误差超出给定阈值, 但系统未及时给出不可用告警的概率, 即危害误导信息(hazardous misleading information, HMI)的发生概率:

$$P_{\mathrm{HMI}} = P(\mathrm{HMI}) = \{\hat{\boldsymbol{x}} - \boldsymbol{x} \notin \Omega_{\mathrm{AL}} \cap D\} \tag{2}$$

式中: $\hat{\boldsymbol{x}}$为定位系统解算位置; $\boldsymbol{x}$为真实位置; Ω_{AL}为由 AL 所确定的定位误差边界; D表示无告警。

本文只关注水平定位解, 并选取水平告警门限(horizontal AL, HAL)为 3 m。根据 1.2 节中的分析, RTK 定位中各环节的风险项会相互叠加, 即不同来源的故障都可能导致完好性事件的发生, 因此, 可以将完好性风险分解为不同故障场景下的完好性风险之和:

$$P_{\mathrm{HMI}} = P_{\mathrm{HMI},F_0} + \sum_{l=1}^{5} P_{\mathrm{HMI},F_l} = P(\mathrm{HMI}|F_0)P(F_0) + \sum_{l=1}^{5} P(M_l)P(F_l) \tag{3}$$

式中：F_0表示无故障情况，反映定位解算相关风险项；$F_l(l=1,2,\cdots,5)$分别表示来自卫星段、传播段、通信链路、基准站及用户段的故障；M_l为漏检事件。

为了评估RTK的初始完好性，假设基准站未向用户提供观测数据或改正数的完好性保障，也未针对通信链路故障和传播段异常发出告警，且用户端自身不存在相应的故障监测和排除流程。因此，可以认为各类故障的漏检率均为1，则式(3)可以保守表示为各类故障概率与无故障定位误差超出告警限概率之和：

$$P_{\mathrm{HMI}}=P(\hat{\boldsymbol{x}}-\boldsymbol{x}\notin\Omega_{\mathrm{AL}},F_0)+\sum_{l=1}^{5}P(F_l) \tag{4}$$

根据1.2节中的分析，卫星段故障概率为10^{-4}，通信链路故障概率在$10^{-5}\sim10^{-4}$之间，假定在中高纬电离层平静期传播段风险为0，有$P(F_1)=10^{-4}$，$10^{-5}\leqslant P(F_2)\leqslant10^{-4}$，$P(F_3)=0$。对于基准站和用户段，主要的故障来源为多径误差、射频干扰及接收机故障，其对观测数据的影响主要体现在2个方面：①引起观测数据中断；②产生较大的观测误差，包括载波周跳和伪距粗差。两者对定位的影响主要体现在模糊度固定上，可能引起模糊度固定失败和模糊度固定错误。基于此分析定义相关事件：F_{c}表示数据中断或探测出周跳；F_{g}表示伪距粗差的出现；事件A表示定位解为固定解，其中，A_1表示固定解正确，A_2表示固定解错误；B表示定位解为浮点解，其中，B_1表示浮点解正确，B_2表示浮点解错误。

根据上述定义，基准站和用户段故障概率可以表示为

$$P(F_l)=P\left(F_{l,\mathrm{c}}\right)+P\left(F_{l,\mathrm{g}}\right)\qquad l=4,5 \tag{5}$$

式中：$P\left(F_{l,\mathrm{c}}\right)$为基准站($l=4$)或用户段($l=5$)出现数据中断或周跳的概率；$P\left(F_{l,\mathrm{g}}\right)$为基准站或用户段出现伪距粗差的概率。

定义事件$D\equiv\hat{\boldsymbol{x}}-\boldsymbol{x}\notin\Omega_{\mathrm{AL}}$，在未出现告警的情况下，可以保守认为整体定位解均为正常解，则有

$$P(D,F_0)=P(D|A)P(A)+P(D|B)P(B) \tag{6}$$

对于模糊度为固定解的情况，有

$$P(D|A)=P(D|A_1)P(A_1|A)+P(D|A_2)P(A_2|A) \tag{7}$$

式中：$P(A_1|A)$和$P(A_2|A)$分别为固定解中正确和错误解的比例，本文按照水平定位误差是否超过引言中所提95%定位精度(0.5 m)来判别定位解的正确性；$P(D|A_1)$为正确固定解所引起定位误差超过HAL的概率，在未进行解的正确性检验时，可以将满足定位精度要求的固定解视为正确固定解，因此，该项值必然为0，而在进行实时检验的情况下，认为通过检验的是固定解，则该项反映漏检率；$P(D|A_2)$为错误固定解所引起的定位误差超限概率，与模糊度解算方法及定位误差的正态性相关，可以通过统计得到。

同理，当模糊度为浮点解时，可以得到

$$P(D|B)=P(D|B_1)P(B_1|B)+P(D|B_2)P(B_2|B) \tag{8}$$

式中：各项含义与固定解的情况类似。

由此，式(6)可以化简为

$$P(D,F_0)=P(D|A_2)P(A_2|A)P(A)+P(D|B_2)P(B_2|B)P(B) \tag{9}$$

综上，根据基准站和移动站实际观测数据统计出$P\left(F_{l,\mathrm{c}}\right)$和$P\left(F_{l,\mathrm{g}}\right)$，$l=4,5$，再由RTK定位结果统计出式(9)中的各项概率大小，与其余风险项概率值共同代入式(4)，即可求出整体完好性风险概率。

2 接收机数据质量对RTK完好性的影响

2.1 不同数据质量接收机组合RTK定位结果

选取不同类型的接收机组合进行RTK定位和风险分析，所使用的观测数据来源、观测站名称、接收机类型编号及数据时段如表1所示。

表1 观测数据信息

Table 1 Observation data sources

数据来源	观测站名称	接收机类型编号	数据时段
美国连续运行参考站	ALCU ALDS	Model 1	2016-01-01—2016-01-31
陆态网	BJFS BJSH	Model 2	2020-12-01—2020-12-31
北京地区商用观测站	BJYZ BJWG	Model 3	2020-12-01—2020-12-31/ 2022-01-01—2022-01-20

Ren等[27]采用多径误差、周跳比、数据完整率、数据连续性4个指标对不同类型接收机的数据质量进行评估，得到的结论为Model 1类型接收机数据质量最优，Model 2次之，Model 3较差。基于该结论，将接收机按照数据质量分为高、中、低3档，针对高-高、中-中、低-中、低-低4种组合，使用GPS单频L1观测数据进行RTK定位解算，统计其在东、北、天3个方向95%、99.9%、99.999%定位误差、水平定位误差、最大粗差，以及99.9%定位误差与95%定位误差的比值(记为比值1，在正态分布假设下的理论值为1.68)、99.999%定位误差与95%定位误差的比值(记为比值2，在正态分布假设下的理论值为2.26)。本文使用的定位解算软件在实时动态定位库(real-time kinematic library, RTKLIB)[28]基础上开发，且算法具有通用性，其中，

对流层和电离层误差近似认为通过双差消除, 不用额外模型进行补偿, 模糊度固定算法为 LAMBDA。不同组合基线长度、模糊度固定率、所统计的历元总数及定位误差统计结果如表 2 所示。

表 2　不同接收机组合 RTK 定位结果

Table 2　RTK positioning results using different receiver combinations

观测站组合（移动站/基准站）	基线长度/km	模糊度固定率/%	历元总数	95%水平定位误差/m	99.999%水平定位误差/m	水平最大粗差/m	方向	95%定位误差/m	99.9%定位误差/m	99.999%定位误差/m	最大粗差/m	比值1	比值2
ALCU/ALDS	53	72	2 675 838	0.40	1.42	1.43	东	0.36	1.35	1.40	1.40	3.75	3.89
							北	0.19	0.63	1.20	1.21	3.32	6.32
							天	0.39	1.40	1.80	2.79	3.59	4.62
BJFS/BJSH	77	65	2 661 914	0.52	2.02	2.03	东	0.46	1.27	1.50	1.50	2.76	3.26
							北	0.26	0.92	1.36	1.44	3.54	5.23
							天	0.69	1.82	2.71	2.71	2.64	3.93
BJYZ/BJFS	61	77	2 643 782	0.43	1.53	1.53	东	0.32	1.45	1.51	1.51	4.53	4.72
							北	0.18	0.73	0.85	0.93	4.06	4.72
							天	0.39	1.58	2.08	2.09	4.05	5.33
BJYZ/BJWG	26	69	1 596 289	0.84	16.68	46.64	东	0.63	2.13	13.28	17.92	3.38	21.08
							北	0.54	2.46	13.67	45.74	4.56	25.31
							天	1.10	3.99	29.00	99.45	3.63	26.36

从表 2 可以看出, ALCU/ALDS 组合与 BJYZ/BJFS 组合的 95% 水平定位误差小于 0.5 m, 99.999% 水平定位误差及水平最大粗差小于 3 m。BJFS/BJSH 组合的 95% 水平定位误差略大于 0.5 m, 但由于其基线长度较长, 该结果具有合理性, 且其 99.999% 水平定位误差也小于 3 m。而 BJYZ/BJWG 组合定位误差较大, 几乎可达上百米, 该现象是由于 BJWG 站频繁出现数据中断导致, 这说明数据质量差的接收机会严重影响定位精度和完好性。此外, 比值 1 和比值 2 远大于理论值, 说明 RTK 定位误差整体正态性较差, 存在严重的“厚尾现象”, ALCU/ALDS 组合三维定位误差 QQ 图及分布直方图(图 2)也可以印证该结论。图 2(a)～2(c) 蓝色曲线反映了东北天 3 个方向实际定位误差与标准正态分布的分位数关系。可知, 其在尾部严重偏离反映正态分布的红色参考线。图 2(d)～2(f) 绘制了 3 个方向定位

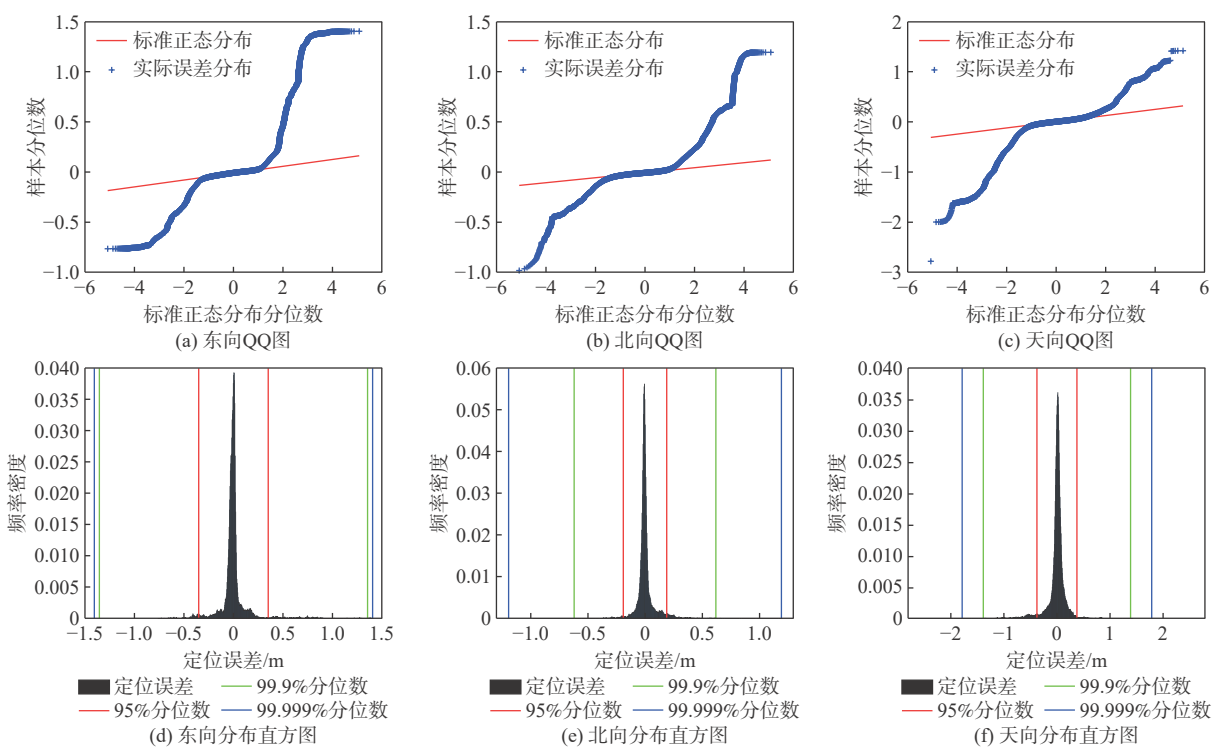


图 2　ALCU/ALDS 组合三维定位误差 QQ 图及分布直方图

Fig. 2　QQ plot and histogram of 3D positioning errors for ALCU/ALDS combination

误差分布直方图及表示 95%、99.9% 及 99.999% 定位误差分位数的参考线。可知，3 个方向定位误差均具有“尖峰厚尾”的特征。图 2 和表 2 均反映了 RTK 定位误差具有较差的正态性，说明 RTK 定位过程中，从观测数据到定位解算的各个环节容易出现较大的粗差或不确定性，对应较差的完好性。

需要说明的是，本文选取的观测站间距离均较远，进行单站 RTK 定位时，公共误差项难以完全消除，且只使用 GPS 单频观测数据进行定位解算，因此，固定率和定位精度均较差。若在该较为严苛的定位场景下，所需定位精度和完好性要求能够得以满足，则可以认为在更近的基线距离下，该指标也能得以满足。

2.2 不同数据质量接收机组合 RTK 完好性评估

根据 1.3 节方法，为了评估 RTK 定位完好性，需要统计不同数据质量接收机 $P(F_c)$ 和 $P\left(F_g\right)$ 的大小及与定位解相关的各项概率。$P(F_c)$ 为出现数据中断或探测出周跳的概率，根据文献 [26] 中对数据连续性的定义，可以保守认为 $P(F_c)$ 是滑动窗口时间为 1 的数据连续性指标，统计得到不同接收机类型对应的该项指标如表 3 所示。

表 3 不同类型接收机观测数据连续性指标统计

Table 3 Continuity statistics indicators of observation data for different receivers

接收机类型	观测站名称	周跳概率	数据中断概率	$P(F_c)$统计值
Model 1	ALCU	1.37×10^{-4}	3.86×10^{-7}	1.37×10^{-4}
	ALDS	9.65×10^{-5}	3.86×10^{-7}	9.69×10^{-5}
Model 2	BJFS	4.29×10^{-4}	2.06×10^{-5}	4.50×10^{-4}
	BJSH	2.08×10^{-4}	1.65×10^{-4}	3.73×10^{-4}
Model 3	BJYZ	2.18×10^{-4}	5.02×10^{-4}	7.20×10^{-4}
	BJWG	1.25×10^{-4}	3.13×10^{-3}	3.25×10^{-3}

$P\left(F_g\right)$ 为出现伪距粗差的概率，可以通过统计多径误差超出给定门限的概率来评估。由于本文中 HAL 取为 3 m，假设平均水平精度因子（horizontal dilution of precision, HDOP）为 1.5，因此，选取粗差门限为 2 m。对于不同接收机类型，粗差概率统计结果如表 4 所示。

根据几组 RTK 定位结果，统计固定解与浮点解比例、错误率及错误风险概率，结果如表 5 所示。可以看出，除较差数据质量的接收机组合 BJYZ/BJWG 外，RTK 定位固定解错误率在 10^{-2}～10^{-1} 之间，浮点解错误率在 10^{-1}～1 之间，后者约为前者的 10 倍；此外，由于前 3 组定位解中不存在定位误差超出 3 m 告警限的情况，为保守起见，将风险事件个数取为 1。结果中，错误浮点解风险概率的统计低于固定解风险概率，这是因为浮点解中错误解较多，基数大，并非是其实际的风险概率值小。为此，后续计算将浮点解风险概率保守取为固定解风险概率的 20 倍。

综合所有概率项，可以得到不同接收机组合整体完好性风险概率，如表 6 所示。可以看出，接收机数据质量的确对定位完好性有较大影响，使用高性能接收机有利于提升 RTK 的完好性，但即使是性能最优的接收机组合，其风险概率也仅能达到 10^{-4}～10^{-3} 量级，远远无法满足自动驾驶中 10^{-7} 的完好性风险要求。

表 4 不同类型接收机粗差概率统计结果

Table 4 Statistical results of gross error probabilities for different receivers

接收机类型	观测站名称	$P\left(F_g\right)$统计值
Model 1	ALCU	3.23×10^{-6}
	ALDS	1.33×10^{-6}
Model 2	BJFS	7.42×10^{-5}
	BJSH	1.62×10^{-4}
Model 3	BJYZ	1.57×10^{-5}
	BJWG	1.10×10^{-3}

表 5 不同接收机组合定位解风险概率项

Table 5 Probability term relevant to positioning solution risk for different receiver combinations

观测站组合（移动站/基准站）	$P(A)$	$P(A_2\|A)$	$P(B_2\|B)$	$P(D\|A_2)$	$P(D\|B_2)$	$P(D,F_0)$
ALCU/ALDS	7.22×10^{-1}	6.99×10^{-3}	9.57×10^{-3}	7.41×10^{-5}	1.40×10^{-5}	4.10×10^{-7}
BJFS/BJSH	6.48×10^{-1}	1.96×10^{-2}	1.27×10^{-1}	2.96×10^{-5}	8.40×10^{-6}	7.51×10^{-7}
BJYZ/BJFS	7.67×10^{-1}	1.62×10^{-2}	1.10×10^{-1}	3.05×10^{-5}	1.47×10^{-5}	7.56×10^{-7}
BJYZ/BJWG	6.90×10^{-1}	5.95×10^{-2}	2.77×10^{-1}	1.41×10^{-2}	8.60×10^{-3}	1.33×10^{-3}

表 6　不同接收机组合 RTK 整体完好性风险概率范围

Table 6　RTK integrity risk probability range for different receiver combinations

接收机组合	观测站组合(移动站/基准站)	整体完好性风险概率
Model 1/Model 1	ALCU/ALDS	3.49×10^{-4}～4.39×10^{-4}
Model 2/Model 2	BJFS/BJSH	1.17×10^{-3}～1.26×10^{-3}
Model 3/Model 2	BJYZ/BJFS	1.37×10^{-3}～1.46×10^{-3}
Model 3/Model 3	BJYZ/BJWG	6.53×10^{-3}～6.62×10^{-3}

3　电离层异常对 RTK 完好性的影响

第 2 节所使用的观测站均位于中高纬地区, 受电离层异常的影响较小, 因此, 选用纬度较低的香港地区连续运行参考站(continuous operational reference system, CORS)2015 年和 2020 年的观测数据来研究电离层异常对 RTK 完好性的影响。所选取的观测站信息如表 7 所示。

根据电离层 11 年周期变化特性, 2015 年是第 24 个太阳活动周期中地磁活动的高峰年, 伴随着较高的太阳活动强度, 电离层异常现象频发, 尤其在春秋季(3、4、9、10 月)频率最高[29], 而 2020 年是地磁活动和太阳活动的低谷年。因此, 选取 2015 年和 2020 年 3 月份的数据进行对比分析, 且同样使用 GPS 单频 L1 观测数据进行 RTK 定位。

表 7　低纬度地区观测数据信息

Table 7　Observation data sources in the low-latitude area

观测数据来源	观测站名称	接收机类型编号	数据时段
香港CORS站	HKCL, HKQT	Model 2	2015-03-01—2015-03-30
			2020-03-01—2020-03-30
	HKMW		2020-03-01—2020-03-30

从表 8 的定位统计结果可以看出, 2015 年 RTK 定位误差均远大于 2020 年, 说明电离层异常会对定位精度产生较大影响, 其中, 更为严重的是会引入大量异常定位解, 大大降低定位完好性。通过对观测数据的分析可以发现, 电离层异常会导致多颗卫星载波观测值瞬时缺失, 导致可用卫星数骤减, 引起高达数十米的定位误差, 这是由于电离层异常会引起信号幅值和相位的快速变化, 而载波跟踪环路相较于码跟踪环路更加脆弱, 因此, 更容易失锁, 导致无法输出正常载波观测值。从表 9 对不同年份 HKCL 和 HKQT 站观测数据 $P(F_c)$和$P(F_g)$的对比可以看出, 数据中断和周跳的概率在不同年份间的差异远大于伪距粗差概率, 印证了前述分析。

表 8　香港 CORS 站 RTK 定位结果

Table 8　RTK positioning results for HK CORS stations

观测站组合(移动站/基准站)	时间	基线长度/km	模糊度固定率/%	历元总数	95%水平定位误差/m	99.999%水平定位误差/m	水平最大粗差/m	方向	95%定位误差/m	99.9%定位误差/m	99.999%定位误差/m	最大粗差/m	比值1	比值2
								东	0.59	7.58	17.25	17.29	12.85	29.24
HKQT/HKCL	2015-03	31	40	2 584 500	0.8	17.82	17.84	北	0.51	7.66	11.10	11.21	15.02	21.76
								天	1.13	13.35	39.15	39.21	11.81	34.65
								东	0.24	0.67	0.72	2.41	2.79	3.00
HKQT/HKCL	2020-03	31	72	2 591 948	0.29	1.61	6.22	北	0.13	0.55	1.46	5.73	4.23	11.23
								天	0.38	1.67	4.51	6.53	4.39	11.87
								东	0.03	0.21	0.52	2.83	7.00	17.33
HKMW/HKCL	2020-03	11	91	2 591 981	0.05	1.01	7.26	北	0.04	0.20	0.93	6.68	5.00	23.25
								天	0.11	0.45	1.39	7.56	4.09	12.64

表 9　香港 CORS 站不同年份连续性概率及伪距粗差概率统计结果

Table 9　Statistical results of continuity probability and gross error probability of HK CORS stations across different years

时间	观测站名称	$P(F_c)$	$P(F_g)$
2015-03	HKCL	5.02×10^{-3}	1.79×10^{-5}
	HKQT	3.19×10^{-3}	9.86×10^{-4}
2020-03	HKCL	3.75×10^{-4}	4.95×10^{-5}
	HKQT	2.19×10^{-4}	5.09×10^{-4}
	HKMW	2.22×10^{-4}	7.76×10^{-4}

此外,通过对2020 年HKQT/HKCL 组合与HKMW/HKCL 组合定位结果的对比可以看出, 尽管短基线能够将 95% 定位精度提升至厘米级, 但 99.999% 定位精度并没有得到大幅提升, 且最大水平定位误差依然超出 3 m HAL, 这说明定位误差依然存在较强的“厚尾特性”, 定位完好性较差。

依据 RTK 定位结果统计式 (9) 中固定解、浮点解比例、错误率及错误风险, 如表 10 所示, 最终得到表 11 中整体完好性风险概率统计结果。

从表 10 可以看出, 电离层异常会大大降低模糊度固定率, 增加模糊度错误固定率, 使整体完好性风险增大。此外, 通过表 6 与表 11 中的数据对比可以发现, 即使在电离层平静期, 使用同类型的接收机和更短的基线, 香港地区 RTK 定位完好性风险也大于北京地区, 说明在电离层更为活跃的低

表10 香港CORS站定位解风险概率项

Table 10 Probability term relevant to positioning solution risk of HK CORS stations

观测站组合(移动站/基准站)	时间	$P(A)$	$P(A_2\|A)$	$P(B_2\|B)$	$P(D\|A_2)$	$P(D\|B_2)$	$P(D,F_0)$
HKQT/HKCL	2015-03	4.00×10^{-1}	1.38×10^{-1}	2.20×10^{-1}	1.05×10^{-2}	3.37×10^{-2}	5.01×10^{-3}
HKQT/HKCL	2020-03	7.19×10^{-1}	1.09×10^{-3}	3.44×10^{-2}	4.75×10^{-4}	3.91×10^{-4}	7.47×10^{-7}
HKMW/HKCL	2020-03	9.10×10^{-1}	3.19×10^{-5}	3.43×10^{-3}	1.56×10^{-2}	1.44×10^{-3}	8.93×10^{-7}

表11 香港CORS站不同年份RTK整体完好性风险概率

Table 11 RTK integrity risk probability of HK CORS stations in different years

时间	观测站组合(移动站/基准站)	整体完好性风险概率
2015-03	HKQT/HLCL	1.43×10^{-2}～1.44×10^{-2}
2020-03	HKQT/HKCL	1.26×10^{-3}～1.35×10^{-3}
2020-03	HKMW/HKCL	1.53×10^{-3}～1.62×10^{-3}

纬度地区,电离层是影响定位完好性的关键因素,这就对实际应用中接收机数据质量和基准站的密集度有更高的要求,也需要根据区域特性建立更为精确的电离层风险模型。

综合第2节与第3节的分析结果可以看出,在中高纬电离层平静地区,接收机数据质量是影响RTK完好性的主要因素,不同性能接收机组合对应的整体完好性风险概率位于10^{-4}～10^{-2}之间。而在低纬度电离层活跃区,电离层异常成为影响RTK完好性的主要因素,对于典型Model 2类型接收机组合,RTK整体完好性风险概率位于10^{-3}～10^{-2}之间。

考虑到完好性的本质是服务可靠性,若使用理想无故障接收机进行单点定位,其定位完好性只需考虑卫星端故障,风险概率约为10^{-4},而RTK通过增加额外环节提升了定位精度,但并未像GBAS和JPALS系统对这些环节进行冗余备份和监测,导致更多不可靠因素的引入,因此,整体完好性必然差于单点定位,这与目前对RTK定位的普遍认识相吻合,也从侧面验证了本文方法具有一定的合理性,能够对RTK完好性进行有效的定量评估。

4 结 论

1)本文提出一种基于风险概率分解的RTK完好性评估方法,通过对典型RTK架构的研究,分析各环节中潜在故障源,量化评估其对整体定位风险的影响,从系统设计和风险分解的角度提出完好性定量评估的理论公式,弥补了当前RTK完好性研究中只关注用户端算法改进而非系统性解决方案的不足。

2)使用本文方法,针对接收机数据质量和电离层异常2项影响定位完好性的关键因素,评估不同观测站组合和时空点处的风险大小,给出当前RTK定位完好性风险的保守范围。实验结果显示,基于自动驾驶3 m AL需求,使用单频单星座进行中等基线RTK定位,整体完好性风险概率位于10^{-4}～10^{-2}之间。

本文研究结果为后续RTK完好性的提升指出了可行的方向,其关键在于:从系统设计的角度对风险源进行更为细致的评估,充分挖掘并使用冗余观测信息提升故障监测能力,以保障和提升RTK服务的可靠性。本文对RTK完好性风险的定量评估结果可作为探讨RTK在自动驾驶中应承担的安全责任及后续各项监测器指标设计的重要依据。

致谢 感谢中国大陆构造环境监测网络(陆态网)、北京合众思壮科技股份有限公司、香港CORS站对论文提供数据支持。

参考文献(References)

[1] EL-MOWAFY A, KUBO N. Integrity monitoring for positioning of intelligent transport systems using integrated RTK-GNSS, IMU and vehicle odometer[J]. IET Intelligent Transport Systems, 2018, 12(8): 901-908.

[2] LI R, ZHENG S Y, WANG E S, et al. Advances in BeiDou navigation satellite system (BDS) and satellite navigation augmentation technologies[J]. Satellite Navigation, 2020, 1(1): 12.

[3] ICAO. International standards and recommended practices, annex 10 to the convention on civil aviation, aeronautical telecommunications: ANN-00010-008-01[S]. Montreal: ICAO, 2023: 3-88.

[4] U. S. Navy. Navy declares initial operational capability for joint precision approach and landing system[EB/OL]. [2021-05-18]. https://www.navy.mil/Press-Office/News-Stories/Article/2621944/navy-declares-initial-operational-capability-for-joint-precision-approach-and-l/.

[5] Department of the Navy. Draft system requirements document for joint precision approach and landing system (JPALS) increment 1: PMA2135-0100/R-5[R]. Washington, D. C. : U. S. Navy, 2007.

[6] RIFE J, KHANAFSEH S, PULLEN S, et al. Navigation, interference suppression, and fault monitoring in the sea-based joint precision approach and landing system[J]. Proceedings of the IEEE, 2008, 96(12): 1958-1975.

[7] BASNAYAKE C, WILLIAMS T, ALVES P, et al. Can GNSS drive V2X[EB/OL]. (2010-10-01)[2024-03-07]. https://www.gpsworld.com/transportationroadcan-gnss-drive-v2x-10611/.

[8] KAFKA P. The automotive standard ISO 26262, the innovative driver for enhanced safety assessment & technology for motor

cars[J]. Procedia Engineering, 2012, 45: 2-10.

[9] REID T G R, HOUTS S E, CAMMARATA R, et al. Localization requirements for autonomous vehicles[J]. SAE International Journal of Connected and Automated Vehicles, 2019, 2(3): 173-190.

[10] REID T G R, NEISH A, MANNING B. Localization & mapping requirements for level 2+ autonomous vehicles[C]//Proceedings of the 2023 International Technical Meeting of the Institute of Navigation. Manassas: ION, 2023: 107-123.

[11] BRYANT R, JULIEN O, HIDE C, et al. Mass market lane accurate positioning with high integrity[C]//Proceedings of the 32nd International Technical Meeting of the Satellite Division of the Institute of Navigation. Manassas: ION, 2019: 573-593.

[12] 中华人民共和国交通运输部. 公路工程技术标准: JTG B01—2014[S]. 北京: 人民交通出版社, 2015.
Ministry of Transport of the People's Republic of China. Technical standard of highway engineering: JTG B01—2014[S]. Beijing: China Communications Press, 2015(in Chinese).

[13] IMPARATO D, EL-MOWAFY A, RIZOS C. Integrity monitoring: from airborne to land applications[EB/OL]. (2018-11-05)[2024-03-07]. http://dx.doi.org/10.5772/intechopen.75777.

[14] IMPARATO D, EL-MOWAFY A, RIZOS C, et al. Vulnerabilities in SBAS and RTK positioning in intelligent transport systems: an overview[C]//International Global Navigation Satellite Systems Association Symposium 2018. Sydney: UNSW Australia, 2018: 1-19.

[15] EL-MOWAFY A, XU B, HSU L T. Integrity monitoring using multi-GNSS pseudorange observations in the urban environment combining ARAIM and 3D city models[J]. Journal of Spatial Science, 2022, 67(1): 91-110.

[16] RUWISCH F, SCHÖN S. Performance assessment of GNSS RTK positioning in urban environments: outlier detection versus 3DMA-FDE[C]//Proceedings of the 35th International Technical Meeting of the Satellite Division of the Institute of Navigation. Manassas: ION, 2022: 2649-2663.

[17] BARFORD T, LACAMBRE J B, GREER R. Optimizing GNSS time-differenced carrier phase measurements for high-integrity inertial navigation plus GNSS sensor fusion without ambiguity resolution[C]//Proceedings of the 2023 IEEE/ION Position, Location and Navigation Symposium. Piscataway: IEEE Press, 2023: 793-804.

[18] LI X X, QIN Z Y, SHEN Z H, et al. A high-precision vehicle navigation system based on tightly coupled PPP-RTK/INS/odometer integration[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(2): 1855-1866.

[19] SAIDANI M, MAYA D G, GUINAMARD A. High integrity multi-sensor navigation system for safety critical applications[C]//Proceedings of the 34th International Technical Meeting of the Satellite Division of the Institute of Navigation. Manassas: ION, 2021: 1767-1781.

[20] WANG K, EL-MOWAFY A. Effect of biases in integrity monitoring for RTK positioning[J]. Advances in Space Research, 2021, 67(12): 4025-4042.

[21] WANG K, EL-MOWAFY A, RIZOS C, et al. Integrity monitoring for horizontal RTK positioning: new weighting model and overbounding CDF in open-sky and suburban scenarios[J]. Remote Sensing, 2020, 12(7): 1173.

[22] Department of Defense, Department of Transportation, Department of Homeland Security. 2021 federal radionavigation plan: DOT-VNTSC-OST-R-15-01[R]. Springfield: National Technical Information Service, 2021: A-4.

[23] TEUNISSEN P J G. GNSS ambiguity bootstrapping: theory and application[C]//International Symposiumon Kinematic Systems in Geodesy, Geomatics and Navigation. Banff: University of Calgary, 2001: 246-254.

[24] TEUNISSEN P J G. The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation[J]. Journal of Geodesy, 1995, 70: 65-82.

[25] CHANG X W, YANG X, ZHOU T. MLAMBDA: a modified LAMBDA method for integer least-squares estimation[J]. Journal of Geodesy, 2005, 79: 552-565.

[26] GALLON E, JOERGER M, PERVAN B. Robust modeling of GNSS tropospheric delay dynamics[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(5): 2992-3003.

[27] REN H Y, HUANG Z G, LI R, et al. Monitoring station data quality analysis method[C]//China Satellite Navigation Conference Proceedings. Berlin: Springer, 2021: 22-32.

[28] TAKASU T, YASUDA A. Development of the low-cost RTK-GPS receiver with an open source program package RTKLIB[C]//International Symposium on GPS/GNSS. Jeju: ETRI, 2009: 4-6.

[29] LIU Y X, LI R, BAO J J, et al. A statistical study of the ionospheric anomalies affecting SBAS safety detected over China area in 2015[C]//China Satellite Navigation Conference 2020 Proceedings: Volume II. Berlin: Springer, 2020: 739-750.

RTK integrity evaluation method based on risk probability decomposition

SONG Yuan[1], HUANG Zhigang[1], LI Rui[1,*], WANG Yuechen[2], SHEN Jun[2], WANG Yongchao[3], NIE Xin[4]

(1. School of Electronic and Information Engineering, Beihang University, Beijing 100191, China;
2. Beijing UniStrong Science and Technology Corporation Limited, Beijing 100176, China;
3. Aviation Data Communication Corporation, Beijing 100191, China;
4. China Academy of Space Technology, Beijing 100094, China)

Abstract: The real-time kinematic (RTK) technique has attracted extensive attention in the field of autonomous driving for its high precision and high dynamic characteristics. As a result, the quantitative assessment of RTK integrity becomes a pressing issue that needs to be resolved in order to ascertain whether it can satisfy the safety standards in this industry. However, most current RTK integrity studies focus on the improvement and optimization of user-end algorithms rather than the analysis and evaluation of overall risks from system perspectives. Therefore, in this paper, an RTK integrity evaluation method based on risk probability decomposition is proposed, offering a theoretical calculation formula of integrity risks. To verify the rationality of this conclusion, observation data from multiple observation stations with different receiver types in typical regions of mid and low latitudes are used for RTK integrity evaluation. The findings demonstrate that, in middle-latitude regions, positioning integrity changes considerably with receiver data quality; nevertheless, in low-latitude regions, ionosphere anomalies become the primary factor influencing integrity. According to the evaluation results in the paper, the integrity risk probability of RTK positioning is between 10^{-4} and 10^{-2} based on single frequency, single constellation and medium-length baseline. The results of this paper are consistent with the general conclusion of RTK data processing, which proves that the proposed method is reasonable and can evaluate the quantitative RTK integrity effectively. Furthermore, the results of the quantitative evaluation of RTK integrity can be used as an important basis for the design of an integrity monitoring scheme.

Keywords: integrity; integrity risk evaluation; RTK integrity; autonomous driving; receiver data quality

Received: 2024-03-07; **Accepted**: 2024-04-19; **Published Online**: 2024-05-08 14:01

URL: link.cnki.net/urlid/11.2625.V.20240507.1611.003

Foundation items: National Key Research and Development Program of China (2022YFB3904302); Innovation Foundation of CAST (F-W-YY-2023-019)

* **Corresponding author**. E-mail: lee_ruin@263.net

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0139

基于分布式安装MSCSG的挠性航天器姿态振动测控方法

李磊[1]，王卫杰[2]，庞维坤[1]，王丽芬[2,*]，段乐强[1]，王宸宇[1]
(1. 航天工程大学 研究生院，北京 101416；2. 航天工程大学 宇航科学与技术系，北京 101416)

摘 要：为实现挠性航天器姿态振动的高精度测控一体化，提出一种基于分布式安装磁悬浮控制敏感陀螺(MSCSG)和径向基函数(RBF)神经网络的挠性航天器姿态振动测控一体化方法。建立挠性附件上分布式安装MSCSG的航天器动力学模型，分析MSCSG对挠性附件振动测量和控制的机理；在此基础上，设计神经网络鲁棒自适应状态反馈控制器，通过神经网络对分布式安装MSCSG带来的系统非线性项及外部扰动进行逼近，实现MSCSG输出控制力矩抑制挠性附件低频振动，同时，对挠性附件形变进行测量。刚体上安装的MSCSG群同时实现航天器姿态的测量与控制。仿真结果表明：所提方法相较于传统基于模态观测器的自适应控制方法，姿态控制精度及振动抑制效果均得到大幅提升。

关 键 词：磁悬浮控制敏感陀螺；挠性航天器；径向基函数神经网络；测量控制一体化；模态分析

中图分类号：V221+.3; TB553

文献标志码：A **文章编号**：1001-5965(2026)05-1615-12

随着航天器技术的不断发展，以及日益复杂的航天任务要求，现代航天器结构不仅趋于复杂化，也增加了更多桁架、大型帆板、大型天线、机械臂等挠性附件，其与航天器刚体部分动力学耦合，在轨道机动、姿态机动和外部扰动(如空间碎片、热辐射等)作用下，耦合引起的振动难以消除，会对航天器控制造成不利影响，甚至导致航天器失稳。现有方法主要从2个方面进行振动抑制。

在控制器设计层面，比较成熟的方法有输入成型法、滑模控制、鲁棒自适应控制等。输入成型法利用系统的线性可叠加性控制指令设置，达到消除系统振动的目的。宋晓娟等[1]提出一种用于刚-挠-液耦合航天器的振动抑制控制器，通过结合模糊控制和输入成型技术，在姿态机动过程中有效抑制了挠性附件振动。文献[2]提出一种结合输入成型器和自抗扰控制器的新型控制器，成功应用于多体刚柔耦合系统。文献[3]结合控制力矩陀螺(control moment gyroscope, CMG)和输入成型技术，设计了新型模态观测器，实现了陀螺柔性体的振动抑制。输入成型法虽然能在不增加系统资源的情况下抑制挠性附件引起的振动，但对系统建模精度有较高要求，在参数摄动条件下，控制器的鲁棒性受到制约。在此基础上，一些控制手段通过引入反馈控制的思想增加系统阻尼，从而达到振动抑制的目的。文献[4]设计了用于挠性航天器姿态机动的分数阶非奇异终端滑模控制器，保证了系统的有限时间收敛。文献[5]提出一种基于扩张状态观测器的鲁棒滑模控制方法，在姿态机动控制的同时，抑制了挠性振动。管萍等[6]提出一种基于径向基函数(radial basis function, RBF)神经网络的分数阶滑模控制器。姜野等[7]提出一种结合自适应输出反馈控制和神经网络的挠性航天器姿态控制方法，解决了执行机构

收稿日期：2024-03-11；录用日期：2024-04-05；网络出版时间：2024-05-29 14:33
网络出版地址：link.cnki.net/urlid/11.2625.V.20240528.1618.001
基金项目：国家自然科学基金(52075545)
*通信作者. E-mail：wanglifen_2009@139.com

引用格式：李磊，王卫杰，庞维坤，等. 基于分布式安装MSCSG的挠性航天器姿态振动测控方法[J]. 北京航空航天大学学报，2026，52(5)：1615-1626. LI L，WANG W J，PANG W K，et al. Attitude measurement and control method of flexible spacecraft based on distributed installation MSCSG[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1615-1626 (in Chinese).

饱和条件下的控制问题。这些方法本质上都是对滑模控制器的改进，通过将滑模控制技术和新型控制技术结合，达到提升控制器鲁棒性的目的。此外，余臻[8]提出一种自适应输出反馈控制器，在模态不可测且存在外部扰动的条件下完成了姿态机动和振动抑制，并证明了控制器的稳定性。以上方法中，控制力矩不是直接施加在挠性附件上，因而对挠性附件的振动抑制效果受到制约。

在器件层面，已有研究集中于采用压电陶瓷或CMG对挠性附件直接施加控制力矩，以抑制部分阶次的模态振动。文献[9-11]对采用压电陶瓷作为作动器的控制策略进行了深入研究，提出变结构输出反馈控制器、自适应非奇异终端滑模控制器、自适应反步控制器等，仿真和实验表明，将压电陶瓷直接黏附于挠性附件能够有效增加结构阻尼，从而抑制振动。Shahravi等[12]研究了压电陶瓷布置位置对振动抑制效果的影响，并设计了挠性卫星模型，验证所提理论。Gaudenzi等[13]研究了在欧洲ARTEMIS卫星上的半导体激光卫星间链路试验设备光学有效载荷安装压电陶瓷以抑制振动的控制方法，对卫星进行了有限元建模，并设计了线性二次型调节器（linear quadratic regulator, LQR）补偿器。这些方法利用压电智能材料的电致动特性对挠性附件输出控制力矩，然而由于测量需要，压电材料往往需要成对安装、测量和控制异位。此外，压电材料输出控制力矩有限，对较大幅振动无能为力，且存在电热效应，可能对载荷正常工作产生不利影响。

在挠性结构上分布安装力矩交换装置同样能实现振动抑制。Damaren等[14-15]研究了具有分布式角动量装置的挠性飞行器的可控性、可观测性及最优控制问题。然而，该方法假设系统的姿态运动很小，且包括模态变量在内的所有系统状态均可测量，限制了控制策略的应用。Hu等在2013年使用离散化逐点配置CMG来抑制回转弹性体的挠性振动[16]，然而，该控制方法不能直接应用于自由飞行的回转弹性体，其在2016年和2017年又构建了基于分布式CMG的自由飞行挠性航天器的动力学方程，并设计了自适应控制法及基于模态观测器的挠性抑振法[17-18]。Guo等[19]在2018年提出利用CMG产生的模态力控制航天器机动带来的挠性振动，进一步分析了模态力的奇异性，并基于动态逆设计了规避奇异的操纵律。以上方法将角动量装置嵌入挠性结构，使结构自身特性发生变化，在此基础上设计的控制器不仅能实现振动抑制，由于角动量装置输出力矩足够大，其控制力矩也能用于实现姿态机动。然而，这些方法仍未解决柔性结构形变测量和控制异位的问题。

2017年，航天工程大学任元团队成功研制磁悬浮控制敏感陀螺（magnetically suspended control sensitive gyroscope, MSCSG），作为具有测控一体化优势的新型惯性执行机构和敏感器，在解决挠性附件模态振动抑制和柔性结构形变测量上具有天然优势[19]，而神经网络控制器的强鲁棒性和精确逼近性能也为提升控制系统鲁棒性和参数摄动条件下的稳定性提供了一条可行的技术路径。基于此，本文设计了一种分布式安装MSCSG的挠性附件振动测控方法，通过将MSCSG嵌入挠性附件，不仅实现了振动的主动控制，而且实现了同位测量。

本文首先设计了一种分布式安装MSCSG的挠性航天器模型，并推导相应的动力学方程；然后，根据特定安装方式下的模型，提出一种基于柔性弹性转角测量方法，实现了测控同位；最后，提出一种神经网络鲁棒自适应状态反馈控制律，实现航天器姿态机动的同时，对挠性附件的振动进行抑制。

1 分布式安装MSCSG的挠性航天器建模方法

1.1 MSCSG及挠性航天器模型建模

MSCSG能够输出最大力矩为1.6 N·m、最大频率为140 Hz的微框架力矩，其结构如图1所示[20]，$Ox_ty_tz_t$为转子坐标系，$Ox_gy_gz_g$为磁轴承坐标系，转子分别绕Ox_g、Oy_g轴以α、β偏转。

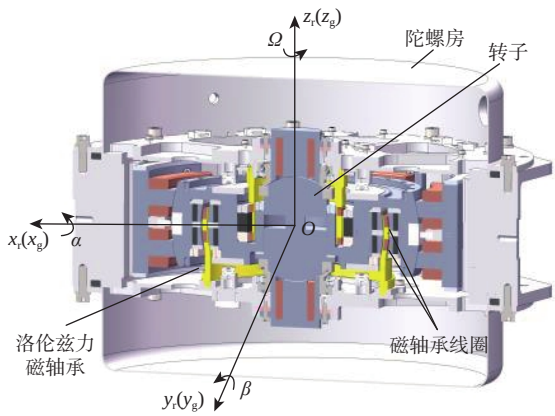


图1 MSCSG模型[20]

Fig. 1 MSCSG model[20]

动力学模型可以表示为

$$\begin{cases} J_x\ddot{\alpha}+J_z\Omega\dot{\beta}=4NBLl_{\mathrm{m}}i_\alpha \\ J_y\ddot{\beta}-J_z\Omega\dot{\alpha}=4NBLl_{\mathrm{m}}i_\beta \end{cases} \tag{1}$$

式中：$\boldsymbol{J}_{\mathrm{g}}=\left[J_x,\ J_y,\ J_z\right]^{\mathrm{T}}$为MSCSG转子转动惯量；$\Omega$为转子转速；$N$为洛伦兹力磁轴承（Lorentz force magnetic bearing, LFMB）中线圈匝数；B为磁感应强度；L为LFMB线圈长度；l_{m}为磁轴承半径；i_α、i_β分别为Ox_{g}、Oy_{g}轴方向上的线圈驱动电流。

考虑安装挠性附件的航天器，其附件上不安装

执行机构及传感器时, 动力学模型可以表示为

$$\begin{cases} \boldsymbol{J}\dot{\boldsymbol{\omega}}+\boldsymbol{M}^{\mathrm{T}}\ddot{\boldsymbol{\eta}}+\boldsymbol{\omega}\left(\boldsymbol{J}\boldsymbol{\omega}+\boldsymbol{M}^{\mathrm{T}}\dot{\boldsymbol{\eta}}\right)=\boldsymbol{u}+\boldsymbol{T}_{\mathrm{d}} \\ \ddot{\boldsymbol{\eta}}+\boldsymbol{C}\dot{\boldsymbol{\eta}}+\boldsymbol{K}\boldsymbol{\eta}+\boldsymbol{M}\dot{\boldsymbol{\omega}}=\boldsymbol{0} \end{cases} \tag{2}$$

式中: $\boldsymbol{J}$ 为航天器三轴转动惯量; $\boldsymbol{\omega}=[\omega_x, \omega_y, \omega_z]^{\mathrm{T}}$ 为航天器三轴角速度; $\boldsymbol{\eta}$ 为挠性附件的模态坐标; $\boldsymbol{M}$ 为挠性附件和航天器中心刚体的耦合矩阵; $\boldsymbol{T}_{\mathrm{d}}$ 为外部扰动; $\boldsymbol{u}=[u_x, u_y, u_z]^{\mathrm{T}}$ 为三轴姿态控制力矩; $\boldsymbol{C}$ 为系统挠性的模态阻尼矩阵; $\boldsymbol{K}$ 为系统挠性的模态频率矩阵。

$$\boldsymbol{C}=\mathrm{diag}\left(2\kappa_i\omega_i\right) \qquad i=1,2,\cdots,n \tag{3}$$

$$\boldsymbol{K}=\mathrm{diag}\left(\omega_i^2\right) \qquad i=1,2,\cdots,n \tag{4}$$

式中: κ_i 和 ω_i 分别为第 i 阶振动的阻尼比和固有频率。

可以看出, 由于挠性附件上没有控制机构, 航天器姿态机动引起的附件振动只能凭借自身结构阻尼缓慢降低, 且对于模态坐标缺乏相应的测量手段。

1.2　分布式安装 MSCSG 的挠性航天器模型

为解决上述问题, 在挠性附件上安装 MSCSG, 以中心固定有刚体的自由边界柔性板为例, 如图 2 所示。

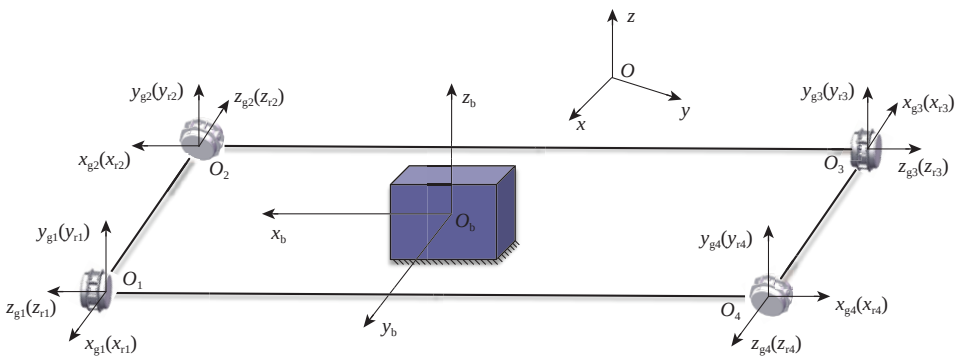


图 2　4 角安装 MSCSG 的挠性航天器模型

Fig. 2　Flexible spacecraft model with MSCSG mounted at the four corners

4 个 MSCSG 安装于柔性板的特定位置, 其中, $Oxyz$ 为惯性坐标系, $O_{\mathrm{b}}x_{\mathrm{b}}y_{\mathrm{b}}z_{\mathrm{b}}$ 为航天器本体坐标系, 与航天器中心刚体固连, 相对于惯性坐标系有 6 个自由度; $O_kx_{gk}y_{gk}z_{gk}$ 为第 k 个 MSCSG 定子安装坐标系, 且初始时刻(挠性附件不发生形变)时, 其三轴平行于航天器本体坐标系(非一一对应); $O_kx_{rk}y_{rk}z_{rk}$ 为第 k 个 MSCSG 转子坐标系, 与转子固连但不随转子转动, 当转子不偏转时, 转子坐标系与定子安装坐标系重合。

推导柔性形变条件下各坐标系之间的转换矩阵。当转子系绕 O_kx_{gk} 偏转 α、绕 O_ky_{gk} 偏转 β 时 ($\alpha,\beta\in[-2°, 2°]$), 第 k 个转子坐标系到第 k 个定子坐标系的转换矩阵可表示为

$$\boldsymbol{C}_{rk}^{\mathrm{gr}}=\begin{bmatrix}1&0&0\\0&\cos\alpha&\sin\alpha\\0&-\sin\alpha&\cos\alpha\end{bmatrix}\begin{bmatrix}\cos\beta&0&-\sin\beta\\0&1&0\\\sin\beta&0&\cos\beta\end{bmatrix}=\begin{bmatrix}\cos\beta&0&-\sin\beta\\\sin\alpha\sin\beta&\cos\alpha&\sin\alpha\cos\beta\\\cos\alpha\sin\beta&-\sin\alpha&\cos\alpha\cos\beta\end{bmatrix} \tag{5}$$

当初始时刻柔性板不发生变形时, 柔性板上 4 个 MSCSG 定子坐标系到航天系本体坐标系的转换矩阵为

$$\boldsymbol{C}_{\mathrm{g1}}^{\mathrm{b}}=\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix} \tag{6}$$

$$\boldsymbol{C}_{\mathrm{g2}}^{\mathrm{b}}=\begin{bmatrix}1&0&0\\0&0&-1\\0&1&0\end{bmatrix} \tag{7}$$

$$\boldsymbol{C}_{\mathrm{g3}}^{\mathrm{b}}=\begin{bmatrix}0&0&-1\\-1&0&0\\0&1&0\end{bmatrix} \tag{8}$$

$$\boldsymbol{C}_{\mathrm{g4}}^{\mathrm{b}}=\begin{bmatrix}-1&0&0\\0&0&1\\0&1&0\end{bmatrix} \tag{9}$$

取 MSCSG 安装处的挠性附件微元 dm 为研究对象, 在小变形假设下, 可以认为附件上每一点只在 Z 方向有弹性变形, 而在 x、y 方向有弹性转角, 则根据模态展开法, 第 k 个陀螺安装坐标 (x, y) 处, dm 相对于本体坐标系的弹性转角可表示为

$$\boldsymbol{\gamma}=\begin{bmatrix}\vartheta\\\xi\\0\end{bmatrix}=\boldsymbol{R}_k\boldsymbol{\eta}=\begin{bmatrix}\boldsymbol{R}_{xk}\boldsymbol{\eta}\\\boldsymbol{R}_{yk}\boldsymbol{\eta}\\\boldsymbol{R}_{zk}\boldsymbol{\eta}\end{bmatrix} \tag{10}$$

式中: ϑ、ξ 分别为 dm 相对于 $O_{\mathrm{b}}x_{\mathrm{b}}$、$O_{\mathrm{b}}y_{\mathrm{b}}$ 轴方向的弹性转角; $\boldsymbol{R}_k$ 为 (x, y) 处的转动模态矩阵; $\boldsymbol{R}_{xk}$、$\boldsymbol{R}_{yk}$、

$\boldsymbol{R}_{zk}$分别为$\boldsymbol{R}_k$的3个行向量。

当MSCSG安装处发生形变时，第k个定子坐标系到原定子坐标系的转换矩阵可表示为

$$\boldsymbol{C}_{\mathrm{gr}}^{gk}=\begin{bmatrix}1&0&0\\0&\cos\vartheta&-\sin\vartheta\\0&\sin\vartheta&\cos\vartheta\end{bmatrix}\begin{bmatrix}\cos\xi&0&\sin\xi\\0&1&0\\-\sin\xi&0&\cos\xi\end{bmatrix}=\begin{bmatrix}\cos\xi&0&\sin\xi\\\sin\vartheta\sin\xi&\cos\vartheta&-\sin\vartheta\cos\xi\\-\cos\vartheta\sin\xi&\sin\vartheta&\cos\vartheta\cos\xi\end{bmatrix}\tag{11}$$

综合式(5)可以得出，当转子偏转同时挠性附件发生形变时，第k个转子坐标系到本体坐标系的转换矩阵为

$$\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}=\boldsymbol{C}_{gk}^{\mathrm{b}}\boldsymbol{C}_{\mathrm{gr}}^{gk}\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{gr}}\tag{12}$$

由于转子偏转角很小，可认为

$$\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{gr}}\approx\begin{bmatrix}1&0&-\beta\\0&1&\alpha\\\beta&-\alpha&1\end{bmatrix}\tag{13}$$

考虑到模态振动一般情况下较小，同时，基于简化计算的目的，规定转子坐标系只绕O_kx_{gk}轴偏转，即当$\beta=0$时，柔性板4角的4个陀螺的转换矩阵分别为

$$\boldsymbol{C}_{\mathrm{r1}}^{\mathrm{b}}=\begin{bmatrix}\cos\vartheta(\beta\cos\xi-\sin\xi)&\cos\vartheta\cos\xi&\cos\vartheta(\cos\xi+\beta\sin\xi)+\alpha\sin\vartheta\\\cos\xi+\beta\sin\xi&-\alpha\sin\xi&-\beta\cos\xi+\sin\xi\\\sin\vartheta(\sin\xi-\beta\cos\xi)&\cos\vartheta+\alpha\sin\vartheta\cos\xi&-\sin\vartheta(\beta\sin\xi+\cos\xi)+\alpha\cos\vartheta\end{bmatrix}\tag{14}$$

$$\boldsymbol{C}_{\mathrm{r2}}^{\mathrm{b}}=\begin{bmatrix}\cos\xi+\beta\sin\xi&-\alpha\sin\xi&-\beta\cos\xi+\sin\xi\\\cos\vartheta(\sin\xi-\beta\cos\xi)&-\sin\vartheta+\alpha\cos\vartheta\cos\xi&-\cos\vartheta(\beta\sin\xi+\cos\xi)-\alpha\sin\vartheta\\\sin\vartheta(\sin\xi-\beta\cos\xi)&\cos\vartheta+\alpha\sin\vartheta\cos\xi&-\sin\vartheta(\beta\sin\xi+\cos\xi)+\alpha\cos\vartheta\end{bmatrix}\tag{15}$$

$$\boldsymbol{C}_{\mathrm{r3}}^{\mathrm{b}}=\begin{bmatrix}\cos\vartheta(\sin\xi-\beta\cos\xi)&-\sin\vartheta+\alpha\cos\vartheta\cos\xi&-\cos\vartheta(\beta\sin\xi+\cos\xi)-\alpha\sin\vartheta\\-\cos\xi-\beta\sin\xi&\alpha\sin\xi&\beta\cos\xi-\sin\xi\\\sin\vartheta(\sin\xi-\beta\cos\xi)&\cos\vartheta+\alpha\sin\vartheta\cos\xi&-\sin\vartheta(\beta\sin\xi+\cos\xi)+\alpha\cos\vartheta\end{bmatrix}\tag{16}$$

$$\boldsymbol{C}_{\mathrm{r4}}^{\mathrm{b}}=\begin{bmatrix}-\cos\xi-\beta\sin\xi&\alpha\sin\xi&\beta\cos\xi-\sin\xi\\\cos\vartheta(\beta\cos\vartheta\cos\xi-\sin\xi)&\sin\vartheta-\alpha\cos\vartheta\cos\xi&\cos\vartheta(\beta\sin\xi+\cos\xi)+\alpha\sin\vartheta\\\sin\vartheta(\sin\xi-\beta\cos\xi)&\cos\vartheta+\alpha\sin\vartheta\cos\xi&-\sin\vartheta(\beta\sin\xi+\cos\xi)+\alpha\cos\vartheta\end{bmatrix}\tag{17}$$

因此，根据欧拉动力学方程可知，对于安装在柔性结构上的MSCSG，由转子偏转和柔性振动引起的角动量方向变化与航天器姿态耦合产生的总力矩为

$$\boldsymbol{u}_{\mathrm{mscsg}}^{\mathrm{a}}=\sum_{k=1}^{4}\left[\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\dot{\boldsymbol{h}}_k+\boldsymbol{\omega}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)+\dot{\boldsymbol{\gamma}}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)\right]\tag{18}$$

式中：$\sum_{k=1}^{4}\left(\boldsymbol{\omega}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)+\dot{\boldsymbol{\gamma}}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)\right)$为MSCSG转子角动量与航天器姿态及柔性振动产生的耦合力矩；$\sum_{k=1}^{4}\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\dot{\boldsymbol{h}}_k$为挠性附件上安装MSCSG所产生的姿态控制力矩；$\boldsymbol{h}_k$为第$k$个MSCSG角动量矢量在转子坐标系的投影，表示为

$$\boldsymbol{h}_k=\begin{bmatrix}J_x\dot{\alpha}_k\\J_y\dot{\beta}_k\\J_z\Omega_k\end{bmatrix}\tag{19}$$

式中：Ω_k为第k个MSCSG的转速。

相应地，陀螺产生的模态力矩可表示为

$$\boldsymbol{u}_{\mathrm{mscsg}}^{\mathrm{m}}=\sum_{k=1}^{4}\boldsymbol{R}_k^{\mathrm{T}}\left[\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\dot{\boldsymbol{h}}_k+\boldsymbol{\omega}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)+\dot{\boldsymbol{\gamma}}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)\right]\tag{20}$$

式中：$\sum_{k=1}^{4}\boldsymbol{R}_k^{\mathrm{T}}\left[\boldsymbol{\omega}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)+\dot{\boldsymbol{\gamma}}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)\right]$为MSCSG转子角动量与航天器姿态及柔性振动产生的模态力矩；$\sum_{k=1}^{4}\boldsymbol{R}_k^{\mathrm{T}}\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\dot{\boldsymbol{h}}_k$为挠性附件上安装MSCSG所产生的模态控制力矩。

结合式(2)，安装MSCSG的挠性航天器动力学方程可表示为

$$\begin{cases}\boldsymbol{J}\dot{\boldsymbol{\omega}}+\boldsymbol{M}^{\mathrm{T}}\ddot{\boldsymbol{\eta}}+\boldsymbol{\omega}\times\left(\boldsymbol{J}\boldsymbol{\omega}+\boldsymbol{M}^{\mathrm{T}}\dot{\boldsymbol{\eta}}\right)+\sum_{k=1}^{4}\left[\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\dot{\boldsymbol{h}}_k+\boldsymbol{\omega}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)+\dot{\boldsymbol{\gamma}}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)\right]=\boldsymbol{u}_{\mathrm{b}}+\boldsymbol{T}_{\mathrm{d}}\\\ddot{\boldsymbol{\eta}}+\boldsymbol{C}\dot{\boldsymbol{\eta}}+\boldsymbol{K}\boldsymbol{\eta}+\boldsymbol{M}\dot{\boldsymbol{\omega}}+\sum_{k=1}^{4}\boldsymbol{R}_k^{\mathrm{T}}\left[\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\dot{\boldsymbol{h}}_k+\boldsymbol{\omega}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)+\dot{\boldsymbol{\gamma}}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)\right]=\boldsymbol{0}\end{cases}\tag{21}$$

式中：$\boldsymbol{u}_{\mathrm{b}}$为航天器刚体部分安装MSCSG所产生的姿态控制力矩。

针对刚体上的MSCSG，其产生的控制力矩直接施加于航天器刚体部分，用以控制航天器姿态，采用文献[21]提出的测控一体构型，出于简化计算的目的，规定转子只有1个偏转自由度，且构型倾角为45°，将包含转子偏转角加速度的量视为扰动，其构型总角动量可表示为

$$\boldsymbol{H}_{\mathrm{b}}=\begin{bmatrix} -\frac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}1}\sin\alpha_{\mathrm{b}1}-J_z\Omega_{\mathrm{b}2}\cos\alpha_{\mathrm{b}2}+\frac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}3}\sin\alpha_{\mathrm{b}3}+J_z\Omega_{\mathrm{b}4}\cos\alpha_{\mathrm{b}4}+\frac{\sqrt{2}}{2}J_x(\dot{\alpha}_{\mathrm{b}1}-\dot{\alpha}_{\mathrm{b}3}) \\ J_z\Omega_{\mathrm{b}1}\cos\alpha_{\mathrm{b}1}-\frac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}2}\sin\alpha_{\mathrm{b}2}-J_z\Omega_{\mathrm{b}3}\cos\alpha_{\mathrm{b}3}+\frac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}4}\sin\alpha_{\mathrm{b}4}+\frac{\sqrt{2}}{2}J_x(\dot{\alpha}_{\mathrm{b}2}-\dot{\alpha}_{\mathrm{b}4}) \\ \frac{\sqrt{2}}{2}J_z(\Omega_{\mathrm{b}1}\sin\alpha_{\mathrm{b}1}+\Omega_{\mathrm{b}2}\sin\alpha_{\mathrm{b}2}+\Omega_{\mathrm{b}3}\sin\alpha_{\mathrm{b}3}+\Omega_{\mathrm{b}4}\sin\alpha_{\mathrm{b}4})+\frac{\sqrt{2}}{2}J_x(\dot{\alpha}_{\mathrm{b}1}+\dot{\alpha}_{\mathrm{b}2}+\dot{\alpha}_{\mathrm{b}3}+\dot{\alpha}_{\mathrm{b}4})+J_z\Omega_{\mathrm{b}5} \end{bmatrix} \tag{22}$$

式中: $\Omega_{\mathrm{b}k}$为刚体上第 k 个陀螺的转速; $\alpha_{\mathrm{b}k}$为刚体上第 k 个陀螺的转子偏转角。

根据欧拉方程, 刚体上陀螺群产生的控制力矩为

$$\boldsymbol{u}_{\mathrm{b}}=-(\dot{\boldsymbol{H}}_{\mathrm{b}}+\boldsymbol{\omega}\times\boldsymbol{H}_{\mathrm{b}}) \tag{23}$$

针对挠性附件上直接安装的 MSCSG, 为方便控制器设计, 将其产生的力矩分为 4 部分: 姿态耦合力矩、姿态控制力矩、模态耦合力矩、模态控制力矩, 分别表示为

$$\boldsymbol{f}_{\mathrm{a}}=\sum_{k=1}^{4}\left[\boldsymbol{\omega}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)+\dot{\boldsymbol{\gamma}}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)\right] \tag{24}$$

$$\boldsymbol{u}_{\mathrm{a}}=\sum_{k=1}^{4}\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\dot{\boldsymbol{h}}_k \tag{25}$$

$$\boldsymbol{f}_{\mathrm{m}}=\sum_{k=1}^{4}\boldsymbol{R}_k^{\mathrm{T}}\left[\boldsymbol{\omega}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)+\dot{\boldsymbol{\gamma}}\times\left(\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\boldsymbol{h}_k\right)\right] \tag{26}$$

$$\boldsymbol{u}_{\mathrm{m}}=\sum_{k=1}^{4}\boldsymbol{R}_k^{\mathrm{T}}\boldsymbol{C}_{\mathrm{r}k}^{\mathrm{b}}\dot{\boldsymbol{h}}_k \tag{27}$$

航天器姿态动力学方程可以改写为

$$\begin{cases}\boldsymbol{J}\dot{\boldsymbol{\omega}}+\boldsymbol{M}^{\mathrm{T}}\ddot{\boldsymbol{\eta}}+\boldsymbol{\omega}\times\left(\boldsymbol{J}\boldsymbol{\omega}+\boldsymbol{M}^{\mathrm{T}}\dot{\boldsymbol{\eta}}\right)+\boldsymbol{f}_{\mathrm{a}}=-\boldsymbol{u}_{\mathrm{a}}+\boldsymbol{u}_{\mathrm{b}}+\boldsymbol{T}_{\mathrm{d}}\\ \ddot{\boldsymbol{\eta}}+\boldsymbol{C}\dot{\boldsymbol{\eta}}+\boldsymbol{K}\boldsymbol{\eta}+\boldsymbol{M}\dot{\boldsymbol{\omega}}+\boldsymbol{f}_{\mathrm{m}}=-\boldsymbol{u}_{\mathrm{m}}\end{cases} \tag{28}$$

2　MSCSG 对挠性附件振动的测控原理分析

2.1　基于分布式安装 MSCSG 的模态测量方法

一般情况下, 系统的模态坐标很难直接测量, 通过测量 MSCSG 安装处的弹性转角, 并合理设计控制器, 可以实现振动抑制。根据 MSCSG 动力学模型, 定子安装坐标系下, 姿态测量技术方程可表示为[22]

$$\begin{cases}J_x\dot{\psi}_x+J_z\Omega\psi_y=4NBLl_{\mathrm{m}}i_\alpha\\ J_y\dot{\psi}_y-J_z\Omega\psi_x=4NBLl_{\mathrm{m}}i_\beta\end{cases} \tag{29}$$

式中: ψ_x、ψ_y分别为转子在惯性坐标系 x、y 轴的角速度在初始定子安装坐标系的投影。

对于如图 2 安装在挠性附件的 MSCSG 而言, 方程可写为

$$\begin{bmatrix}\psi_{1x}\\ \psi_{1y}\end{bmatrix}=\begin{bmatrix}\omega_y+\dot{\xi}_1+\dot{\alpha}_1\\ \omega_z+\dot{\beta}_1\end{bmatrix} \tag{30}$$

$$\begin{bmatrix}\psi_{2x}\\ \psi_{2y}\end{bmatrix}=\begin{bmatrix}\omega_x+\dot{\vartheta}_2+\dot{\alpha}_2\\ \omega_z+\dot{\beta}_2\end{bmatrix} \tag{31}$$

$$\begin{bmatrix}\psi_{3x}\\ \psi_{3y}\end{bmatrix}=\begin{bmatrix}-\omega_y-\dot{\xi}_3+\dot{\alpha}_3\\ \omega_z+\dot{\beta}_3\end{bmatrix} \tag{32}$$

$$\begin{bmatrix}\psi_{4x}\\ \psi_{4y}\end{bmatrix}=\begin{bmatrix}-\omega_x-\dot{\vartheta}_4+\dot{\alpha}_4\\ \omega_z+\dot{\beta}_4\end{bmatrix} \tag{33}$$

在 4 个陀螺转速相同的条件下, 将式(30)～式(33)代入式(29)可得

$$\begin{cases}J_x\left(\dot{\omega}_y+\ddot{\xi}_1+\ddot{\alpha}_1\right)+J_z\left(\Omega+\dot{\vartheta}_1\right)\left(\omega_z+\dot{\beta}_1\right)=4NBLl_{\mathrm{m}}i_{1\alpha}\\ J_y\left(\dot{\omega}_z+\ddot{\beta}_1\right)-J_z\left(\Omega+\dot{\vartheta}_1\right)\left(\omega_y+\dot{\xi}_1+\dot{\alpha}_1\right)=4NBLl_{\mathrm{m}}i_{1\beta}\\ J_x\left(\dot{\omega}_x+\ddot{\vartheta}_2+\ddot{\alpha}_2\right)+J_z\left(\Omega-\dot{\xi}_2\right)\left(\omega_z+\dot{\beta}_2\right)=4NBLl_{\mathrm{m}}i_{2\alpha}\\ J_y\left(\dot{\omega}_z+\ddot{\beta}_2\right)-J_z\left(\Omega-\dot{\xi}_2\right)\left(\omega_x+\dot{\vartheta}_2+\dot{\alpha}_2\right)=4NBLl_{\mathrm{m}}i_{2\beta}\\ J_x\left(-\dot{\omega}_y-\ddot{\xi}_3+\ddot{\alpha}_3\right)+J_z\left(\Omega-\dot{\vartheta}_3\right)\left(\omega_z+\dot{\beta}_3\right)=4NBLl_{\mathrm{m}}i_{3\alpha}\\ J_y\left(\dot{\omega}_z+\ddot{\beta}_3\right)-J_z\left(\Omega-\dot{\vartheta}_3\right)\left(-\omega_y-\dot{\xi}_3+\dot{\alpha}_3\right)=4NBLl_{\mathrm{m}}i_{3\beta}\\ J_x\left(-\dot{\omega}_x-\ddot{\vartheta}_4+\ddot{\alpha}_4\right)+J_z\left(\Omega+\dot{\xi}_4\right)\left(\omega_z+\dot{\beta}_4\right)=4NBLl_{\mathrm{m}}i_{4\alpha}\\ J_y\left(\dot{\omega}_z+\ddot{\beta}_4\right)-J_z\left(\Omega+\dot{\xi}_4\right)\left(-\omega_x-\dot{\vartheta}_4+\dot{\alpha}_4\right)=4NBLl_{\mathrm{m}}i_{4\beta}\end{cases} \tag{34}$$

考虑到模态截断, 弹性转角的频率较低, 忽略式中弹性转角的二阶导数, 式 (34) 可转化为

$$\begin{cases}\boldsymbol{R}_{x1}\dot{\boldsymbol{\eta}}=\dot{\vartheta}_1\approx\dfrac{4NBLl_{\mathrm{m}}i_{1\alpha}-J_x\left(\dot{\omega}_y+\ddot{\alpha}_1\right)}{J_z\left(\omega_z+\dot{\beta}_1\right)}-\Omega\\ \boldsymbol{R}_{y1}\dot{\boldsymbol{\eta}}=\dot{\xi}_1\approx\dfrac{J_y\left(\dot{\omega}_z+\ddot{\beta}_1\right)-4NBLl_{\mathrm{m}}i_{1\beta}}{J_z\left(\Omega+\dot{\vartheta}_1\right)}-\omega_y-\dot{\alpha}_1\\ \boldsymbol{R}_{x2}\dot{\boldsymbol{\eta}}=\dot{\xi}_2\approx\Omega-\dfrac{4NBLl_{\mathrm{m}}i_{2\alpha}-J_x\left(\dot{\omega}_x+\ddot{\alpha}_2\right)}{J_z\left(\omega_z+\dot{\beta}_2\right)}\\ \boldsymbol{R}_{y2}\dot{\boldsymbol{\eta}}=\dot{\vartheta}_2\approx\dfrac{J_y\left(\dot{\omega}_z+\ddot{\beta}_2\right)-4NBLl_{\mathrm{m}}i_{1\beta}}{J_z\left(\Omega-\dot{\xi}_2\right)}-\omega_x-\dot{\alpha}_2\\ \boldsymbol{R}_{x3}\dot{\boldsymbol{\eta}}=\dot{\vartheta}_3\approx\dfrac{J_x\left(-\dot{\omega}_y+\ddot{\alpha}_3\right)-4NBLl_{\mathrm{m}}i_{3\alpha}}{J_z\left(\omega_z+\dot{\beta}_3\right)}+\Omega\\ \boldsymbol{R}_{y3}\dot{\boldsymbol{\eta}}=\dot{\xi}_3\approx\dfrac{4NBLl_{\mathrm{m}}i_{3\beta}-J_y\left(\dot{\omega}_z+\ddot{\beta}_3\right)}{J_z\left(\Omega-\dot{\vartheta}_3\right)}+\dot{\alpha}_3-\omega_y\\ \boldsymbol{R}_{x4}\dot{\boldsymbol{\eta}}=\dot{\xi}_4\approx\dfrac{4NBLl_{\mathrm{m}}i_{4\alpha}-J_x\left(-\dot{\omega}_x+\ddot{\alpha}_4\right)}{J_z\left(\omega_z+\dot{\beta}_4\right)}-\Omega\\ \boldsymbol{R}_{y4}\dot{\boldsymbol{\eta}}=\dot{\vartheta}_4\approx\dfrac{4NBLl_{\mathrm{m}}i_{4\beta}-J_y\left(\dot{\omega}_z+\ddot{\beta}_4\right)}{J_z\left(\Omega+\dot{\xi}_4\right)}+\dot{\alpha}_4-\omega_x\end{cases} \tag{35}$$

式中: 角速度分量 ω_x、ω_y、ω_z可由文献 [21] 给出的姿态测量方程通过刚体上安装的 MSCSG 构型信息解算; 分母的非负性可以通过操纵律设计保证。对于实模态, 当模态矩阵 $\boldsymbol{R}_{\mathrm{m}}=\left[\boldsymbol{R}_{x1},\boldsymbol{R}_{y1},\boldsymbol{R}_{y2},\boldsymbol{R}_{x2},\boldsymbol{R}_{x3},\boldsymbol{R}_{y3},\boldsymbol{R}_{y4},\boldsymbol{R}_{x4}\right]^{\mathrm{T}}$满秩时, 可解算出 8 阶模态坐标。

2.2　模态与可观测性分析

通过复模态分析来探讨 MSCSG 测量弹性变形机理, 无角速度激励条件下, 系统的无阻尼振动方

程可表示为

$$\ddot{\boldsymbol{\eta}}+\boldsymbol{K}\boldsymbol{\eta}+\sum_{k=1}^{4}\boldsymbol{R}_k^{\mathrm{T}}\left[\boldsymbol{C}_{rk}^{\mathrm{b}}\dot{\boldsymbol{h}}_k+\dot{\boldsymbol{\gamma}}\times\left(\boldsymbol{C}_{rk}^{\mathrm{b}}\boldsymbol{h}_k\right)\right]=\boldsymbol{0} \tag{36}$$

将式(36)改写为如下形式:

$$\boldsymbol{A}\dot{\boldsymbol{X}}+\boldsymbol{B}\boldsymbol{X}=\boldsymbol{F} \tag{37}$$

式中: $\boldsymbol{A}=\begin{bmatrix}\boldsymbol{0} & -\boldsymbol{E}\\ \boldsymbol{E} & \mathfrak{R}\end{bmatrix}$, $\boldsymbol{E}$ 为单位矩阵; $\boldsymbol{X}=\begin{bmatrix}\dot{\boldsymbol{\eta}}\\ \boldsymbol{\eta}\end{bmatrix}$; $\boldsymbol{B}=\begin{bmatrix}\boldsymbol{E} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{K}\end{bmatrix}$; $\boldsymbol{F}=\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{F}_{\mathrm{a}}\end{bmatrix}\dot{\boldsymbol{h}}_k$。矩阵$\mathfrak{R}$和$\boldsymbol{F}_{\mathrm{a}}$可表示为

$$\mathfrak{R}=\sum_{k=1}^{4}\boldsymbol{R}_k^{\mathrm{T}}\boldsymbol{R}_k\times\left(\boldsymbol{C}_{rk}^{\mathrm{b}}\boldsymbol{h}_k\right) \tag{38}$$

$$\boldsymbol{F}_{\mathrm{a}}=-\sum_{k=1}^{4}\boldsymbol{R}_k^{\mathrm{T}}\boldsymbol{C}_{rk}^{\mathrm{b}} \tag{39}$$

由于初始时刻MSCSG定子安装坐标系与航天器本体坐标系三轴平行,根据文献[23],$\mathfrak{R}$为反对称矩阵,将方程的$2n$个特征值设为$\lambda_g, g=1,2,\cdots,2n$。$2n$个特征向量为$\boldsymbol{\phi}_g, g=1,2,\cdots,2n$,该特征向量与物理空间特征向量一致,即系统有$2n$个复模态。有$\boldsymbol{X}=\tilde{\boldsymbol{\phi}}\mathrm{e}^{\lambda t}$,其中,$\tilde{\boldsymbol{\phi}}=\begin{bmatrix}\lambda\bar{\boldsymbol{\phi}} & \bar{\boldsymbol{\phi}}\end{bmatrix}^{\mathrm{T}}$为特征向量矩阵,$\lambda$和$\bar{\boldsymbol{\phi}}$分别为特征值和特征向量组成的矩阵,代入式(37)可得

$$(\lambda\boldsymbol{A}+\boldsymbol{B})\tilde{\boldsymbol{\phi}}=\boldsymbol{0} \tag{40}$$

由于$\mathfrak{R}$为反对称矩阵,$\boldsymbol{A}$也为反对称矩阵,即反Hermite矩阵[24],其具有$2n$个特征值,为共轭纯虚数。根据反Hermite矩阵的性质,$\boldsymbol{A}$为正规矩阵,即$\boldsymbol{A}$酉相似于对角阵:

$$\boldsymbol{A}=\boldsymbol{U}\begin{bmatrix}\lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_n & & & \\ & & & \lambda_{n+1} & & \\ & & & & \ddots & \\ & & & & & \lambda_{2n}\end{bmatrix}\boldsymbol{U}^{\mathrm{H}} \tag{41}$$

式中: $\boldsymbol{U}$为酉矩阵; $\lambda_g=\mathrm{j}\bar{\omega}_g$和$\lambda_{2n+1-g}$为共轭虚数,$\bar{\omega}_g$为正实数。

设相应的特征向量为

$$\tilde{\boldsymbol{\phi}}_g=\boldsymbol{P}_g+\mathrm{j}\boldsymbol{Q}_g \tag{42}$$

式中: $\boldsymbol{P}_g$、$\boldsymbol{Q}_g$为实矩阵, $\boldsymbol{P}_g=[-\bar{\omega}_g \quad \boldsymbol{q}_g \quad \boldsymbol{p}_g]^{\mathrm{T}}$, $\boldsymbol{Q}_g=[-\bar{\omega}_g \quad \boldsymbol{p}_g \quad \boldsymbol{q}_g]^{\mathrm{T}}$。

应用文献[23]的结论,由于特征向量为虚数对,可以直接用实部表示一对振型的叠加,有如下结论成立:

$$\mathrm{Re}\left(\boldsymbol{X}_g\right)=\mathrm{Re}\left(\boldsymbol{\phi}_g\mathrm{e}^{\lambda_g t}\right)=\boldsymbol{P}_g\cos\left(\bar{\omega}_g t\right)-\boldsymbol{Q}_g\sin\left(\bar{\omega}_g t\right) \tag{43}$$

取式(43)的后半部分,有

$$\mathrm{Re}\left(\boldsymbol{x}_g\right)=\mathrm{Re}\left(\boldsymbol{\phi}_g\mathrm{e}^{\lambda_g t}\right)=\boldsymbol{p}_g\cos\left(\bar{\omega}_g t\right)-\boldsymbol{q}_g\sin\left(\bar{\omega}_g t\right) \tag{44}$$

注意到

$$\mathrm{Re}\left(\dot{\boldsymbol{x}}_g\right)=-\lambda_g\boldsymbol{p}_g\sin\left(\bar{\omega}_g t\right)-\lambda_g\boldsymbol{q}_g\cos\left(\bar{\omega}_g t\right) \tag{45}$$

因此,设$\boldsymbol{P}_g=\begin{bmatrix}-\bar{\omega}_g\boldsymbol{q}_g & \boldsymbol{p}_g\end{bmatrix}^{\mathrm{T}}$, $\boldsymbol{Q}_g=\begin{bmatrix}\bar{\omega}_g\boldsymbol{p}_g & \boldsymbol{q}_g\end{bmatrix}^{\mathrm{T}}$。令$\boldsymbol{p}_g=\boldsymbol{q}_{2n+1-g}$, $\boldsymbol{p}_{2n+1-g}=-\boldsymbol{q}_g$,选取正交条件为[23]

$$\boldsymbol{P}_g^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_g=-2\bar{\omega}_g^3 \tag{46}$$

$$\boldsymbol{P}_g^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P}_g=\boldsymbol{Q}_g^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Q}_g=2\bar{\omega}_g^2 \tag{47}$$

令$\tilde{\boldsymbol{\Phi}}=\left[\tilde{\boldsymbol{\phi}}_1,\tilde{\boldsymbol{\phi}}_2,\cdots,\tilde{\boldsymbol{\phi}}_{2n}\right]_{2n\times 2n}$,有

$$\boldsymbol{X}=\tilde{\boldsymbol{\Phi}}\boldsymbol{\eta}=\sum_{g=1}^{2n}\boldsymbol{P}_g\eta_g \tag{48}$$

系统方程可写作:

$$\boldsymbol{P}_g^{\mathrm{T}}\boldsymbol{A}\sum_{g=1}^{2n}\boldsymbol{P}_g\dot{\eta}_g+\boldsymbol{P}_g^{\mathrm{T}}\boldsymbol{B}\sum_{g=1}^{2n}\boldsymbol{Q}_{2n+1-g}\eta_g=\boldsymbol{P}_g^{\mathrm{T}}\boldsymbol{F} \tag{49}$$

综上,第g个方程可表示为

$$\dot{\eta}_g-\bar{\omega}_g\eta_{2n+1-g}=\frac{1}{2\bar{\omega}_g^2}\boldsymbol{p}_g\boldsymbol{F}_{\mathrm{a}}\dot{\boldsymbol{h}} \tag{50}$$

令$\tilde{\boldsymbol{X}}=\left[\eta_1,\eta_2,\cdots,\eta_{2n}\right]^{\mathrm{T}}$,对于分布式安装4个MSCSG的挠性附件,有如下测量方程:

$$\boldsymbol{Y}=\begin{bmatrix}\boldsymbol{R}_{\mathrm{m}} & \boldsymbol{0}\end{bmatrix}\boldsymbol{X}=\begin{bmatrix}\boldsymbol{R}_{\mathrm{m}} & \boldsymbol{0}\end{bmatrix}\sum_{g=1}^{2n}\boldsymbol{P}_g\eta_g \tag{51}$$

则测量矩阵可以表示为

$$\boldsymbol{C}_{\mathrm{M}}=\boldsymbol{R}_{\mathrm{m}}\left[-\bar{\omega}_1\boldsymbol{q}_1,-\bar{\omega}_2\boldsymbol{q}_2,\cdots,-\bar{\omega}_{2n}\boldsymbol{q}_{2n}\right]^{\mathrm{T}} \tag{52}$$

据此,可以通过适当设计MSCSG的安装位置来满足部分截断模态的测量。若考虑4阶模态和挠性附件的对称性,假设MSCSG安装于挠性附件的某部位时满足: $|\boldsymbol{R}_{x1}(k)|=|\boldsymbol{R}_{x3}(k)|$、$|\boldsymbol{R}_{x2}(k)|=|\boldsymbol{R}_{x4}(k)|$、$|\boldsymbol{R}_{y1}(k)|=|\boldsymbol{R}_{y3}(k)|$、$|\boldsymbol{R}_{y2}(k)|=|\boldsymbol{R}_{y4}(k)|$,则测量方程可表示为

$$\begin{cases}\boldsymbol{R}_{x1}\dot{\boldsymbol{\eta}}=\dot{\vartheta}_1\approx\dfrac{4NBLl_{\mathrm{m}}i_{1\alpha}-J_x\left(\dot{\omega}_y+\ddot{\alpha}_1\right)}{J_z\left(\omega_z+\dot{\beta}_1\right)}-\Omega\\ \boldsymbol{R}_{y1}\dot{\boldsymbol{\eta}}=\dot{\xi}_1\approx\dfrac{J_y\left(\dot{\omega}_z+\dot{\beta}_1\right)-4NBLl_{\mathrm{m}}i_{1\beta}}{J_z\left(\Omega+\dot{\vartheta}_1\right)}-\omega_y-\dot{\alpha}_1\\ \boldsymbol{R}_{x2}\dot{\boldsymbol{\eta}}=\dot{\xi}_2\approx\Omega-\dfrac{4NBLl_{\mathrm{m}}i_{2\alpha}-J_x\left(\dot{\omega}_x+\ddot{\alpha}_2\right)}{J_z\left(\omega_z+\dot{\beta}_2\right)}\\ \boldsymbol{R}_{y2}\dot{\boldsymbol{\eta}}=\dot{\vartheta}_2\approx\dfrac{J_y\left(\dot{\omega}_z+\ddot{\beta}_2\right)-4NBLl_{\mathrm{m}}i_{1\beta}}{J_z\left(\Omega-\dot{\xi}_2\right)}-\omega_x-\dot{\alpha}_2\end{cases} \tag{53}$$

假设挠性附件上各点的模态向量已知,进一步将式(53)改写为

$$\dot{\boldsymbol{\eta}}=\left(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\right)^{-1}\boldsymbol{R}^{\mathrm{T}}\dot{\boldsymbol{\Phi}} \tag{54}$$

式中: $\boldsymbol{R}=\left[\boldsymbol{R}_{x1},\boldsymbol{R}_{y1},\boldsymbol{R}_{x2},\boldsymbol{R}_{y2}\right]^{\mathrm{T}}$; $\dot{\boldsymbol{\Phi}}=\left[\dot{\vartheta}_1,\dot{\xi}_1,\ \dot{\xi}_2,\dot{\vartheta}_2\right]^{\mathrm{T}}$。

3 控制器和操纵律设计

3.1 神经网络鲁棒自适应控制器设计

控制目标可以表述为:已知$\boldsymbol{\omega}=[\omega_x,\omega_y,\omega_z]^{\mathrm{T}}$、$\dot{\boldsymbol{\Phi}}=\left[\dot{\vartheta}_1,\dot{\xi}_1,\dot{\xi}_2,\dot{\vartheta}_2\right]^{\mathrm{T}}$及航天器三轴姿态角$\varphi$、$\theta$、$\psi$,使

航天器姿态与目标姿态φ_0、θ_0、ψ_0的误差$\boldsymbol{e}\to\boldsymbol{0}$, 且模态坐标$\boldsymbol{\eta}\to\boldsymbol{0}$。当截断前 4 阶模态时, 将航天器动力学方程改写为状态空间形式:

$$\boldsymbol{H}\ddot{\boldsymbol{X}}+\boldsymbol{P}\dot{\boldsymbol{X}}+\boldsymbol{\Theta}=\bar{\boldsymbol{U}} \tag{55}$$

式中: $\boldsymbol{H}=\begin{bmatrix}\boldsymbol{J} & \boldsymbol{M}^{\mathrm{T}}\boldsymbol{R}^{-1}\\ \boldsymbol{M} & \boldsymbol{R}^{-1}\end{bmatrix}$; $\boldsymbol{P}=\begin{bmatrix}\boldsymbol{\omega}\times\boldsymbol{J} & \boldsymbol{\omega}\times(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{R}^{-1})\\ \boldsymbol{0} & \boldsymbol{C}\boldsymbol{R}^{-1}\end{bmatrix}$; $\boldsymbol{X}=\begin{bmatrix}\boldsymbol{\omega}\\ \boldsymbol{\Phi}\end{bmatrix}$; $\boldsymbol{\Theta}=\begin{bmatrix}\boldsymbol{f}_{\mathrm{a}}-\boldsymbol{T}_{\mathrm{d}}\\ \boldsymbol{K}\boldsymbol{R}^{-1}\boldsymbol{\Phi}+\boldsymbol{f}_{\mathrm{m}}\end{bmatrix}$; $\bar{\boldsymbol{U}}=\begin{bmatrix}\boldsymbol{u}_1\\ \boldsymbol{u}_2\end{bmatrix}$。

设计神经网络对$\boldsymbol{\Theta}$进行估计, 设神经网络输出可以表示为$\tilde{\boldsymbol{\Theta}}=\begin{bmatrix}\tilde{\boldsymbol{\Theta}}_1\\ \tilde{\boldsymbol{\Theta}}_2\end{bmatrix}$, 其中:

$$\tilde{\boldsymbol{\Theta}}_1=\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{\Lambda}_1(\boldsymbol{X})+\boldsymbol{\varepsilon}_1 \tag{56}$$

$$\tilde{\boldsymbol{\Theta}}_2=\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{\Lambda}_2(\boldsymbol{X})+\boldsymbol{\varepsilon}_2 \tag{57}$$

式中: $\tilde{\boldsymbol{\Theta}}_1$、$\tilde{\boldsymbol{\Theta}}_2$分别为神经网络对$\boldsymbol{\Theta}$上、下 2 部分元素的估计值; $\boldsymbol{W}^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{W}_1^{\mathrm{T}} & \boldsymbol{W}_2^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$为理想权值, $\boldsymbol{W}_1^{\mathrm{T}}=\begin{bmatrix}w_{11}, w_{12}, \cdots, w_{1q}\end{bmatrix}^{\mathrm{T}}$, $\boldsymbol{W}_2^{\mathrm{T}}=\begin{bmatrix}w_{21}, w_{22}, \cdots, w_{2q}\end{bmatrix}^{\mathrm{T}}$, q为神经元个数; $\boldsymbol{\varepsilon}_1$、$\boldsymbol{\varepsilon}_2$为 2 个网络的逼近误差; $\boldsymbol{\Lambda}=[\boldsymbol{\Lambda}_1(\boldsymbol{X})\quad\boldsymbol{\Lambda}_2(\boldsymbol{X})]^{\mathrm{T}}$为网络核函数, 可以表示为

$$\begin{cases}\boldsymbol{\Lambda}_1\boldsymbol{X}=[\Lambda_{11}(x_1),\Lambda_{12}(x_2),\cdots,\Lambda_{17}(x_7)]^{\mathrm{T}}\\ \boldsymbol{\Lambda}_2\boldsymbol{X}=[\Lambda_{21}(x_1),\Lambda_{22}(x_2),\cdots,\Lambda_{27}(x_7)]^{\mathrm{T}}\end{cases} \tag{58}$$

式中: $\Lambda_{1z}x_z=\exp\left(\dfrac{\|x_z-c_{1z}\|^2}{2b_{1z}^2}\right)$ $\Lambda_{2z}x_z=\exp\left(\dfrac{\|x_z-c_{2z}\|^2}{2b_{2z}^2}\right)$, $z=1,2,\cdots,7$, c_{1z}、c_{2z}、b_{1z}、b_{2z}为网络参数。

定义目标姿态$\boldsymbol{X}_0=[\boldsymbol{\omega}_0\quad\boldsymbol{0}]^{\mathrm{T}}$, 跟踪误差$\boldsymbol{e}=[\boldsymbol{\omega}-\boldsymbol{\omega}_0\quad\boldsymbol{\Phi}]^{\mathrm{T}}$, 将神经网络输出代入式(55), 则有

$$\boldsymbol{H}\ddot{\boldsymbol{e}}+\boldsymbol{P}\dot{\boldsymbol{e}}+\tilde{\boldsymbol{\Theta}}+\boldsymbol{\varepsilon}=\bar{\boldsymbol{U}} \tag{59}$$

令神经网络权值$\tilde{\boldsymbol{W}}^{\mathrm{T}}\boldsymbol{\Lambda}=\begin{bmatrix}\tilde{\boldsymbol{W}}_1^{\mathrm{T}}\boldsymbol{\Lambda}_1(\boldsymbol{X}) & \tilde{\boldsymbol{W}}_2^{\mathrm{T}}\boldsymbol{\Lambda}_2(\boldsymbol{X})\end{bmatrix}^{\mathrm{T}}$, 其中, $\tilde{\boldsymbol{W}}^{\mathrm{T}}$为估计权值。设计鲁棒自适应状态反馈控制律为

$$\bar{\boldsymbol{U}}=-\rho(\boldsymbol{H}\ddot{\boldsymbol{e}}+\boldsymbol{P}\dot{\boldsymbol{e}})-\frac{1+\zeta^2}{2\zeta^2}(\dot{\boldsymbol{e}}+\rho\boldsymbol{e})+\mu\tilde{\boldsymbol{W}}^{\mathrm{T}}\boldsymbol{\Lambda}+\boldsymbol{P}(\dot{\boldsymbol{e}}+\rho\boldsymbol{e}) \tag{60}$$

式中: $\mu>0$; $\rho>0$; ζ为比例系数。

3.2　稳定性分析

令$\hat{\boldsymbol{e}}=\dot{\boldsymbol{e}}+\rho\boldsymbol{e}$, 定义 Lyapunov 函数为

$$\boldsymbol{L}=\frac{1}{2}\hat{\boldsymbol{e}}^{\mathrm{T}}\boldsymbol{H}\hat{\boldsymbol{e}}+\frac{1}{2}(\tilde{\boldsymbol{W}}-\boldsymbol{W})^{\mathrm{T}}(\tilde{\boldsymbol{W}}-\boldsymbol{W}) \tag{61}$$

则有

$$\dot{L}=\hat{\boldsymbol{e}}^{\mathrm{T}}\boldsymbol{H}\dot{\hat{\boldsymbol{e}}}+\frac{1}{2}\hat{\boldsymbol{e}}^{\mathrm{T}}\dot{\boldsymbol{H}}\hat{\boldsymbol{e}}+\dot{\tilde{\boldsymbol{W}}}^{\mathrm{T}}(\tilde{\boldsymbol{W}}-\boldsymbol{W})\cdots=\hat{\boldsymbol{e}}^{\mathrm{T}}[-\boldsymbol{P}\hat{\boldsymbol{e}}+\rho(\boldsymbol{H}\ddot{\boldsymbol{e}}+\boldsymbol{P}\dot{\boldsymbol{e}})-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Lambda}-\boldsymbol{\varepsilon}+\bar{\boldsymbol{U}}]+\frac{1}{2}\hat{\boldsymbol{e}}^{\mathrm{T}}\dot{\boldsymbol{H}}\hat{\boldsymbol{e}}+\dot{\tilde{\boldsymbol{W}}}^{\mathrm{T}}(\tilde{\boldsymbol{W}}-\boldsymbol{W}) \tag{62}$$

由于$\dot{\boldsymbol{H}}=\boldsymbol{0}$, 将控制律代入式 (62), 可得

$$\dot{L}=\hat{\boldsymbol{e}}^{\mathrm{T}}\left[-\boldsymbol{\varepsilon}-\frac{1+\zeta^2}{2\zeta^2}\hat{\boldsymbol{e}}+\mu(\tilde{\boldsymbol{W}}-\boldsymbol{W})\boldsymbol{\Lambda}\right]+\dot{\tilde{\boldsymbol{W}}}^{\mathrm{T}}(\tilde{\boldsymbol{W}}-\boldsymbol{W})=-\hat{\boldsymbol{e}}^{\mathrm{T}}\boldsymbol{\varepsilon}-\frac{1+\zeta^2}{2\zeta^2}\hat{\boldsymbol{e}}^{\mathrm{T}}\hat{\boldsymbol{e}}+\mu\hat{\boldsymbol{e}}^{\mathrm{T}}(\tilde{\boldsymbol{W}}-\boldsymbol{W})\boldsymbol{\Lambda}+\dot{\tilde{\boldsymbol{W}}}^{\mathrm{T}}(\tilde{\boldsymbol{W}}-\boldsymbol{W}) \tag{63}$$

进一步, 有

$$\dot{L}-\frac{\zeta^2}{2}\|\boldsymbol{\varepsilon}\|^2+\frac{1}{2}\|\hat{\boldsymbol{e}}\|^2=-\hat{\boldsymbol{e}}^{\mathrm{T}}\boldsymbol{\varepsilon}-\frac{1}{2\gamma^2}\hat{\boldsymbol{e}}^{\mathrm{T}}\hat{\boldsymbol{e}}-\frac{\zeta^2}{2}\|\boldsymbol{\varepsilon}\|^2+\mu\hat{\boldsymbol{e}}^{\mathrm{T}}(\tilde{\boldsymbol{W}}-\boldsymbol{W})\boldsymbol{\Lambda}+\dot{\tilde{\boldsymbol{W}}}^{\mathrm{T}}(\tilde{\boldsymbol{W}}-\boldsymbol{W})-\frac{1}{2}\hat{\boldsymbol{e}}^{\mathrm{T}}\hat{\boldsymbol{e}}+\frac{1}{2}\|\hat{\boldsymbol{e}}\|^2=-\frac{1}{2}\left(\zeta\|\boldsymbol{\varepsilon}\|+\frac{1}{\zeta}\hat{\boldsymbol{e}}\right)+\mu\hat{\boldsymbol{e}}^{\mathrm{T}}(\tilde{\boldsymbol{W}}-\boldsymbol{W})\boldsymbol{\Lambda}+\dot{\tilde{\boldsymbol{W}}}^{\mathrm{T}}(\tilde{\boldsymbol{W}}-\boldsymbol{W}) \tag{64}$$

选取自适应律为

$$\dot{\tilde{\boldsymbol{W}}}=-\mu\hat{\boldsymbol{e}}\boldsymbol{\Lambda}^{\mathrm{T}} \tag{65}$$

则有

$$\dot{L}-\frac{\zeta^2}{2}\|\boldsymbol{\varepsilon}\|^2+\frac{1}{2}\|\hat{\boldsymbol{e}}\|^2=-\frac{1}{2}\left(\zeta\|\boldsymbol{\varepsilon}\|+\frac{1}{\zeta}\hat{\boldsymbol{e}}\right)\leqslant 0 \tag{66}$$

根据 HJI 定理[25], 有

$$\frac{\|\hat{\boldsymbol{e}}\|_2}{\|\boldsymbol{\varepsilon}\|_2}\leqslant\zeta\qquad\|\boldsymbol{\varepsilon}\|\neq 0 \tag{67}$$

因此, 系统是鲁棒可镇定的。

结合式 (68)~式 (70):

$$\boldsymbol{u}_{\mathrm{m}}=J_z\Omega\begin{bmatrix}\boldsymbol{R}_1^{\mathrm{T}}\boldsymbol{C}_{\mathrm{r1}}^{\mathrm{b}}\\ \boldsymbol{R}_2^{\mathrm{T}}\boldsymbol{C}_{\mathrm{r2}}^{\mathrm{b}}\\ \boldsymbol{R}_3^{\mathrm{T}}\boldsymbol{C}_{\mathrm{r3}}^{\mathrm{b}}\\ \boldsymbol{R}_4^{\mathrm{T}}\boldsymbol{C}_{\mathrm{r4}}^{\mathrm{b}}\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\\ \alpha_4\end{bmatrix}=\boldsymbol{A}\dot{\boldsymbol{\Xi}} \tag{68}$$

$$\boldsymbol{u}_{\mathrm{a}}=J_z\Omega\begin{bmatrix}\boldsymbol{C}_{\mathrm{r1}}^{\mathrm{b}}\\ \boldsymbol{C}_{\mathrm{r2}}^{\mathrm{b}}\\ \boldsymbol{C}_{\mathrm{r3}}^{\mathrm{b}}\\ \boldsymbol{C}_{\mathrm{r4}}^{\mathrm{b}}\end{bmatrix}\begin{bmatrix}\dot{\alpha}_1\\ \dot{\alpha}_2\\ \dot{\alpha}_3\\ \dot{\alpha}_4\end{bmatrix}=\boldsymbol{B}\dot{\boldsymbol{\Xi}} \tag{69}$$

$$\boldsymbol{u}_{\mathrm{b}}=\begin{bmatrix}-\dfrac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}}\cos\alpha_{\mathrm{b1}} & J_z\Omega_{\mathrm{b}}\sin\alpha_{\mathrm{b2}} & \dfrac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}}\cos\alpha_{\mathrm{b3}} & -J_z\Omega_{\mathrm{b}}\sin\alpha_{\mathrm{b4}}\\ -J_z\Omega_{\mathrm{b}}\sin\alpha_{\mathrm{b1}} & -\dfrac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}}\cos\alpha_{\mathrm{b2}} & J_z\Omega_{\mathrm{b}}\sin\alpha_{\mathrm{b3}} & \dfrac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}}\cos\alpha_{\mathrm{b4}}\\ \dfrac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}}\cos\alpha_{\mathrm{b1}} & \dfrac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}}\cos\alpha_{\mathrm{b2}} & \dfrac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}}\cos\alpha_{\mathrm{b3}} & \dfrac{\sqrt{2}}{2}J_z\Omega_{\mathrm{b}}\cos\alpha_{\mathrm{b4}}\end{bmatrix}\begin{bmatrix}\dot{\alpha}_{\mathrm{b1}}\\ \dot{\alpha}_{\mathrm{b2}}\\ \dot{\alpha}_{\mathrm{b3}}\\ \dot{\alpha}_{\mathrm{b4}}\end{bmatrix}=\boldsymbol{G}\dot{\boldsymbol{\Xi}}_{\mathrm{b}} \tag{70}$$

由于 MSCSG 偏转角较小, 没有奇异现象, 设计伪逆操纵律如下:

$$\dot{\boldsymbol{\Xi}}=-(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_{\mathrm{m}} \tag{71}$$

$$\dot{\boldsymbol{\Xi}}_{\mathrm{b}}=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}(\boldsymbol{u}_1-\boldsymbol{B}\dot{\boldsymbol{\Xi}}) \tag{72}$$

4 仿真验证

考虑6 m×2 m×0.1 m的柔性板(如图3所示),质量为162 kg,泊松比为0.3。将板分为75个单元,共计96个节点。板中心40、41、56、57这4个节点固定作为刚体部分,其上安装金字塔构型MSCSG。在2、32、65、95节点安装MSCSG。

仿真系统参数如表1所示。

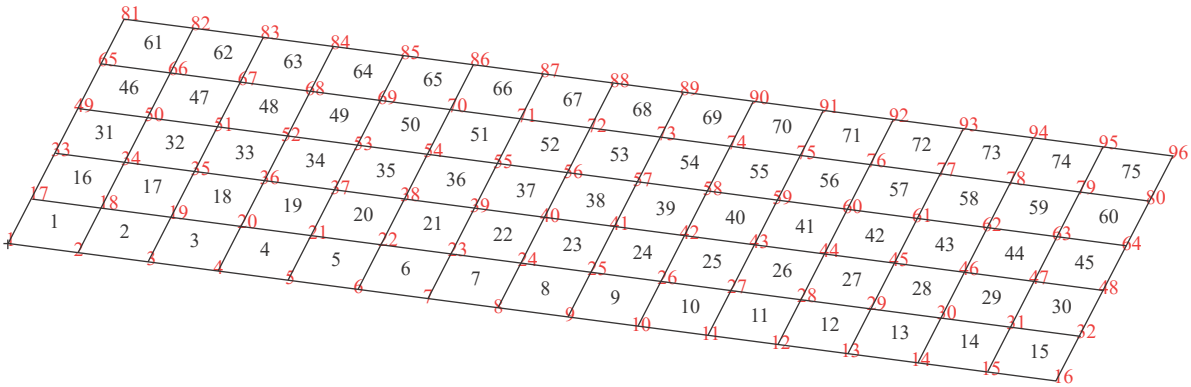


图3 安装MSCSG的挠性航天器模型

Fig. 3 Flexible spacecraft model with MSCSG installed

表1 仿真系统参数

Table 1 Simulation system parameters

转子X轴转动惯量/(kg·m^2)	转子Y轴转动惯量/(kg·m^2)	转子Z轴转动惯量/(kg·m^2)	挠性附件上转子转速/(r·min^{-1})	刚体上转子转速/(r·min^{-1})	转子质量/kg	ζ	μ	ρ
0.009 7	0.009 7	0.016 6	1 000	8 000	8.95	0.5	0.2	0.35

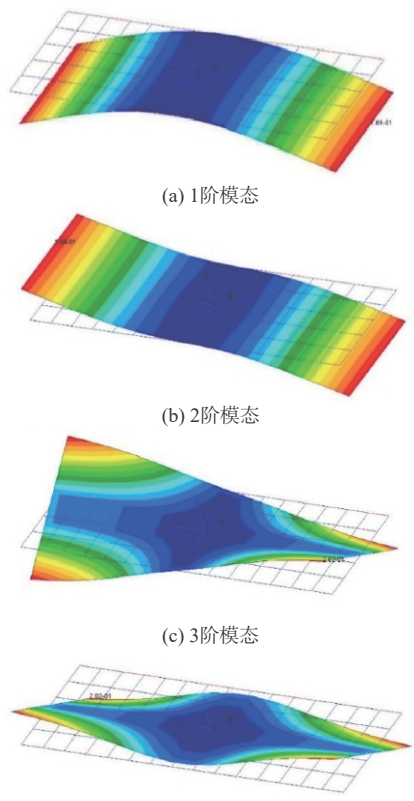
航天器的转动惯量为

$$\boldsymbol{J}=\begin{bmatrix}58.32 & 0.44 & 0\\ 0.44 & 490.33 & 0\\ 0 & 0 & 548.65\end{bmatrix}\ \text{kg}\cdot\text{m}^2$$

仿真主要验证航天器的小角度姿态机动和镇定性能,初始姿态为[−3°, 4°, 5°],目标姿态和模态坐标均为0,为尽可能降低姿态机动激发的挠性附件振动,采用余弦型机动路径。对比方法为文献[11]中提出的基于模态观测器的自适应控制律(传统方法)及本文提出的分布式鲁棒控制律(新方法)。自由边界下板的前4阶模态有限元仿真如图4所示。

在挠性附件不加控制的条件下,采用新方法提出的姿态控制律对航天器三轴姿态和模态进行仿真,其结果如图5和图6所示。可以看出,当仅有刚体上的MSCSG输出姿态控制力矩时,航天器三轴姿态能够按照设计的机动路径完成姿态机动,航天器三轴姿态角在20 s时达到目标姿态角[0°, 0°, 0°]。然而,由于未对挠性附件施加控制,其振动衰减较慢,对姿态控制精度造成不利影响。在50 s时,航天器姿态角仍有最大幅值为$\pm5\times10^{-5}$(°)的波动,并缓慢衰减。而航天器挠性附件的前4阶模态坐标最大幅值为$\pm1.2\times10^{-2}$,并在自身结构阻尼的影响下缓慢缩减,在第50 s时仍有最大幅值为$\pm3\times10^{-5}$的残余振动。值得注意的是,由于姿态机动路径的影响,模态坐标的最大值出现在2阶模态。

使挠性附件上分布式安装的MSCSG按照控制律输出控制力矩,加入模态控制力后,传统方法控制的航天器三轴姿态角及前4阶模态坐标如图7和图8所示。可以看出,当刚体上和挠性附件上的MSCSG同时输出姿态控制力矩及模态控制力矩时,航天器三轴姿态角在20 s时达到目标姿态角。[0°, 0°, 0°]传统方



(a) 1阶模态

(b) 2阶模态

(c) 3阶模态

(d) 4阶模态

图4 自由边界板前4阶模态

Fig. 4 The first four modes of free boundary plate

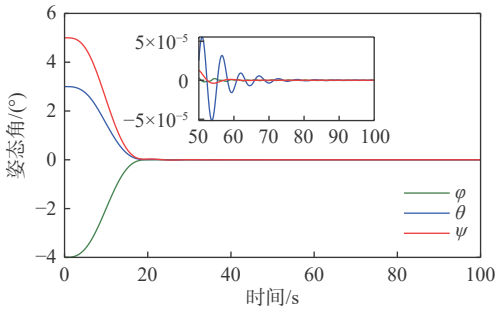


图 5　不加模态控制的航天器姿态

Fig. 5　Spacecraft attitude without modal control

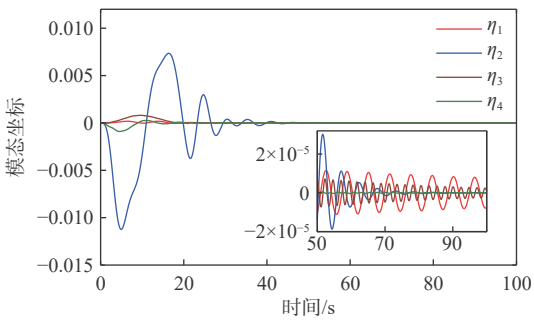


图 6　不加模态控制的挠性附件前 4 阶模态

Fig. 6　The first four orders of modes without modal control

法控制的三轴姿态角在 50 s 时, 航天器姿态角仍有最大幅值为$\pm 2\times 10^{-3}$(°)的波动, 并缓慢衰减。航天器挠性附件的前 4 阶模态坐标最大幅值为$\pm 2.5\times 10^{-5}$, 相较于不施加主动控制提升了 2 个数量级以上, 在第 50 s 时, 残余振动的最大幅值为$\pm 2.5\times 10^{-7}$。

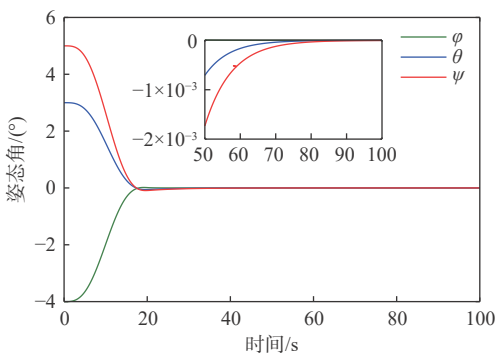


图 7　加入模态控制的航天器姿态

Fig. 7　Spacecraft attitude with modal control

在刚体上和挠性附件上的 MSCSG 同时输出姿态控制力矩及模态控制力矩的条件下, 新方法控制的航天器三轴姿态角及前 4 阶模态坐标如图 9 和图 10 所示。可以看出, 新方法控制的航天器三轴姿态角在 20 s 时达到目标姿态角 [0°, 0°, 0°]。在 50 s时, 航天器姿态角有最大幅值为$\pm 1.2\times 10^{-5}$(°)的波动, 相较于传统方法减少了 2 个数量级以上, 相较于新方法, 在不添加模态控制条件下的姿态角波动幅值减少了 3 倍。航天器挠性附件的前 4 阶模态坐标最大幅值为$\pm 6.1\times 10^{-6}$, 相较于不施加主动控制降低了 2 个数量级以上, 在第 50 s 时, 残余振动的最大幅值为$\pm 8\times 10^{-10}$。可以看

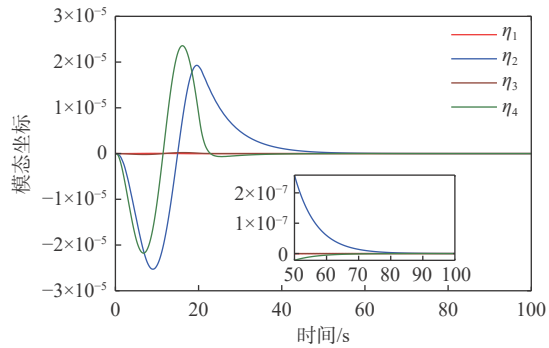


图 8　加入模态控制的挠性附件前 4 阶模态

Fig. 8　The first four modes of the flexible attachment with modal control

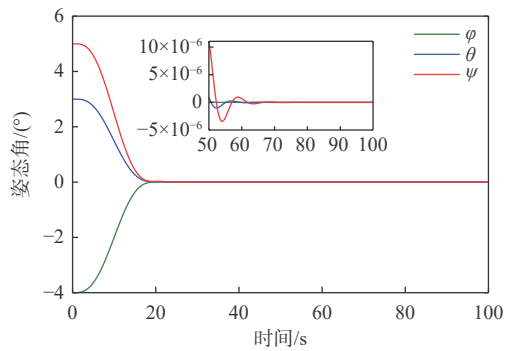


图 9　新方法加入模态控制的航天器姿态

Fig. 9　Spacecraft attitude controlled by new method with modal control

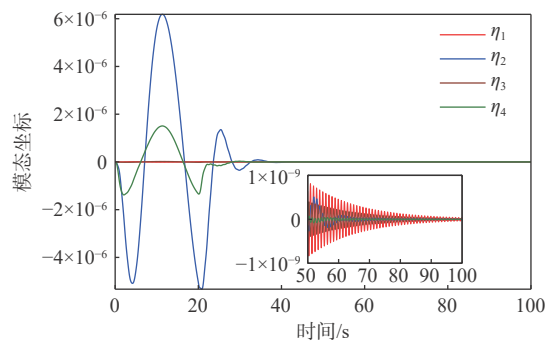


图 10　新方法加入模态控制的挠性附件前 4 阶模态

Fig. 10　The first four modes of the flexible attachment controlled by new method with modal control

出, 施加主动控制的挠性附件振动得到了极大削弱, 且衰减更快, 从而间接提升了姿态控制精度。图 11 和图 12 分别为新方法控制下航天器刚体上及挠性附件上安装的 MSCSG 的转子偏转角。可见, 刚体上 MSCSG 转子的最大偏转角为±5°, 挠性附件上 MSCSG 转子最大偏转角为±12°, 表明本文提出的安装构型可在转子小角度偏转的条件下同时实现航天器姿态镇定和振动抑制。图 13 为挠性附件上 4 个 MSCSG 安装位置处沿 Z 轴方向的弹性转角及其测量值。可见, MSCSG 能够在输出控制力矩的同时测量柔性弹性转角。图 14 为测量误差。可见, 最大测量误差保持在$\pm 2\times 10^{-4}$(°), 证明本文测控一体方案的可行性。

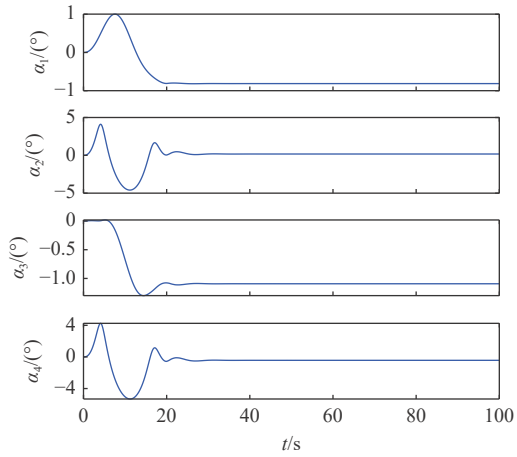


图 11 刚体上 4 个 MSCSG 转子偏转角

Fig. 11 Four MSCSG rotor deflection angles on rigid body

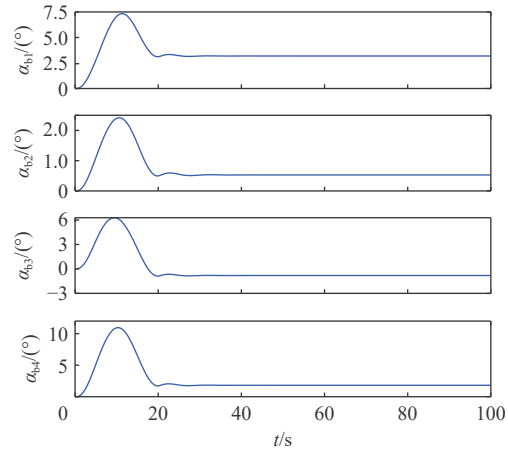


图 12 挠性附件上 4 个 MSCSG 转子偏转角

Fig. 12 Four MSCSG rotor deflection angles on flexible attachment

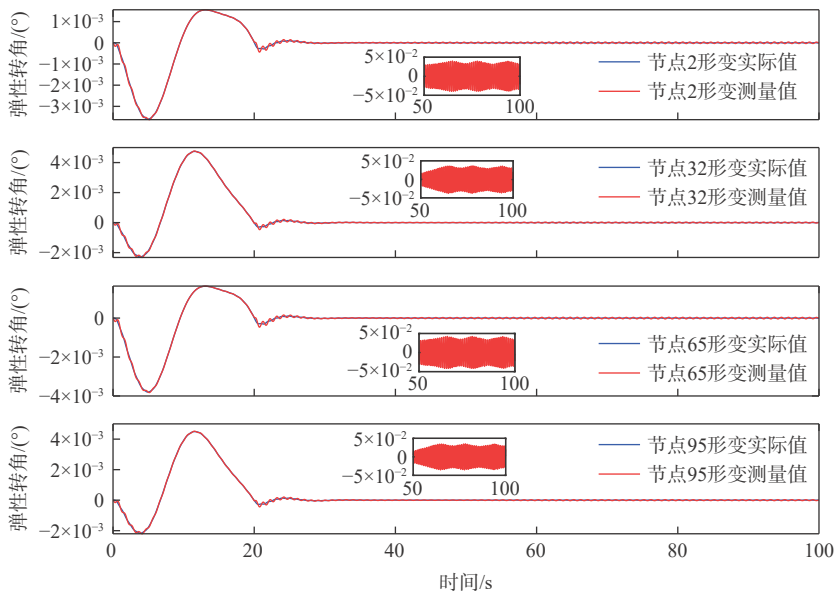


图 13 挠性附件上 4 个 MSCSG 处 Z 轴弹性转角及其测量值

Fig. 13 Z-axis elastic angles of rotation at 4 MSCSGs on the flexible attachment and their measured values

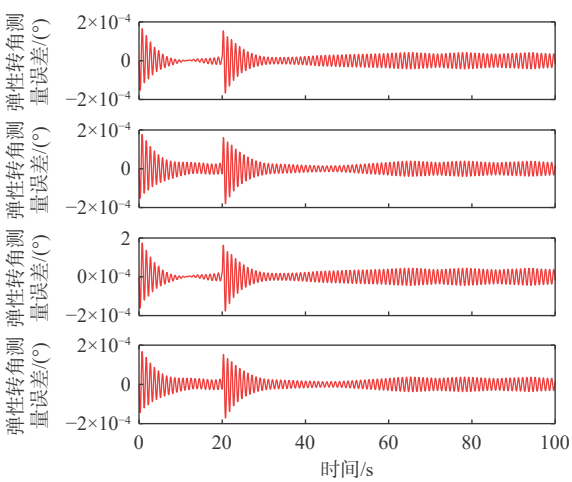


图 14 挠性附件上 4 个 MSCSG 处 Z 轴弹性转角测量误差

Fig. 14 Measuring errors of Z-axis elastic angles of rotation at 4 MSCSGs on the flexible attachment

5 结 论

本文提出一种基于分布式安装 MSCSG 和 RBF 神经网络的挠性航天器姿态振动测控一体化方法。建立分布式安装 MSCSG 的挠性航天器动力学模型，分析了 MSCSG 对挠性附件振动测量与控制机理，设计了神经网络鲁棒自适应状态反馈控制器，仿真结果表明：

1）本文方法能够实现挠性附件的振动测控一体化，最大弹性转角测量误差保持在$\pm 2\times10^{-4}$(°)，残余振动相较于传统方法降低了 2 个数量级以上。

2）本文方法能够实现航天器的高精度姿态控制，姿态角控制精度为$\pm 1.2\times10^{-5}$ (°)，相较于传统方法提升了 2 个数量级。

3）结果证明了本文方法的有效性和优越性。

参考文献（References）

[1] 宋晓娟, 李雪松, 何晓东, 等. 挠性充液航天器复合自适应姿态机动控制[J]. 空间控制技术与应用, 2023, 49(1): 11-20.
SONG X J, LI X S, HE X D, et al. Hybrid adaptive attitude maneuver control for flexible liquid-filled spacecraft[J]. Aerospace Control and Application, 2023, 49(1): 11-20(in Chinese).

[2] LI W P, LUO B, HUANG H. Active vibration control of flexible joint manipulator using input shaping and adaptive parameter auto disturbance rejection controller[J]. Journal of Sound and Vibration, 2016, 363: 97-125.

[3] JIA S Y, SHAN J J. Vibration control of gyroelastic spacecraft using input shaping and angular momentum devices[J]. Acta Astronautica, 2019, 159: 397-409.

[4] GAO J S, DENG L W, SONG S M. Fractional order nonsingular terminal sliding mode control for flexible spacecraft attitude tracking[J]. Instrumentation, 2016(1): 21-29.

[5] ZHOU C B, ZHOU D. Robust dynamic surface sliding mode control for attitude tracking of flexible spacecraft with an extended state observer[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2017, 231(3): 533-547.

[6] 管萍, 罗宗樾, 戈新生. 基于 RBF 神经网络的挠性航天器分数阶滑模姿态控制[J]. 飞行力学, 2021, 39(6): 68-74.
GUAN P, LUO Z Y, GE X S. Fractional order sliding mode attitude control of flexible spacecraft based on RBF neural network[J]. Flight Dynamics, 2021, 39(6): 68-74(in Chinese).

[7] 姜野, 胡庆雷, 马广富. 控制输入饱和的挠性航天器姿态机动智能鲁棒控制[J]. 宇航学报, 2009, 30(1): 188-192.
JIANG Y, HU Q L, MA G F. Intelligent adaptive variable structure attitude maneuvering control for flexible spacecraft with actuator saturation[J]. Journal of Astronautics, 2009, 30(1): 188-192(in Chinese).

[8] 余臻. 挠性航天器姿态机动与主动抑振控制研究[D]. 南京: 南京理工大学, 2017.
YU Z. On attitude maneuver control and active vibration suppression of flexible spacecraft[D]. Nanjing: Nanjing University of Science and Technology, 2017(in Chinese).

[9] 胡庆雷, 刘亚秋, 马广富. 挠性航天器姿态机动的变结构主动振动抑制[J]. 控制理论与应用, 2007, 24(3): 329-336.
HU Q L, LIU Y Q, MA G F. Active vibration suppression in flexible spacecraft with mismatched uncertainty via variable structure control[J]. Control Theory & Applications, 2007, 24(3): 329-336(in Chinese).

[10] HU Q L, MA G F. Vibration control of flexible spacecraft actuated by piezoceramics via variable structure strategy[J]. Journal of Harbin Institute of Technology (New Series), 2007, 14(5): 604-608.

[11] 马广富, 胡庆雷. 挠性航天器姿态机动的主动振动控制[J]. 哈尔滨工程大学学报, 2005, 26(5): 603-607.
MA G F, HU Q L. Active vibration control of flexible spacecraft during attitude maneuver[J]. Journal of Harbin Engineering University, 2005, 26(5): 603-607(in Chinese).

[12] SHAHRAVI M, AZIMI M. A comparative study for collocated and non-collocated sensor/actuator placement in vibration control of a maneuvering flexible satellite[J]. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2015, 229(8): 1415-1424.

[13] GAUDENZI P, GIARDA D, MORGANTI F, et al. Active microvibration control of an optical payload installed on the ARTEMIS spacecraft[J]. Journal of Intelligent Material Systems and Structures, 1998, 9(9): 740-748.

[14] DAMAREN C J, D'ELEUTERIO G M T. Optimal control of large space structures using distributed gyricity[J]. Journal of Guidance, Control, and Dynamics, 1989, 12(5): 723-731.

[15] DAMAREN C J, D'ELEUTERIO G M T. Controllability and observability of gyroelastic vehicles[J]. Journal of Guidance, Control, and Dynamics, 1991, 14(5): 886-894.

[16] HU Q, JIA Y H, XU S J. Simple adaptive control for vibration suppression of space structures using control moment gyroscopes as actuators[C]//Proceedings of the AIAA Guidance, Navigation, and Control Conference. Reston: AIAA, 2013: 4871.

[17] HU Q, ZHANG J R. Attitude control and vibration suppression for flexible spacecraft using control moment gyroscopes[J]. Journal of Aerospace Engineering, 2016, 29: 04015027.

[18] HU Q, GUO C D, ZHANG J. Singularity and steering logic for control moment gyros on flexible space structures[J]. Acta Astronautica, 2017, 137: 261-273.

[19] GUO J T, GENG Y H, WU B L, et al. Vibration suppression of flexible spacecraft during attitude maneuver using CMGs[J]. Aerospace Science and Technology, 2018, 72: 183-192.

[20] 李磊, 任元, 陈晓岑, 等. 基于 ADRC 和 RBF 神经网络的 MSCSG 控制系统设计[J]. 北京航空航天大学学报, 2020, 46(10): 1966-1972.
LI L, REN Y, CHEN X C, et al. Design of MSCSG control system based on ADRC and RBF neural network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(10): 1966-1972(in Chinese).

[21] REN Y, LI L, WANG W J, et al. Magnetically suspended control sensitive gyroscope rotor high-precision deflection decoupling method using quantum neural network and fractional-order terminal sliding mode control[J]. Fractal and Fractional, 2024, 8(2): 120.

[22] LI L, REN Y, WANG W J, et al. Spacecraft attitude measurement and control integration using a novel configuration of variable speed magnetically suspended control and sensing gyroscope[J]. IEEE Sensors Journal, 2023, 23(9): 9359-9369.

[23] D'ELEUTERIO G M T, HUGHES P C. Dynamics of gyroelastic continua[J]. Journal of Applied Mechanics, 1984, 51(2): 415-422.

[24] 陈佳晶, 刘东. 反 Hermite 矩阵的若干性质[J]. 湖州师范学院学报, 2009, 31(2): 7-10.
CHEN J J, LIU D. Some properties of skew Hermite matrices[J]. Journal of Huzhou Teachers College, 2009, 31(2): 7-10(in Chinese).

[25] 沙成满, 李岩, 杨冬梅. 基于 HJI 微分不等式的非线性不确定奇异系统鲁棒镇定[J]. 东北大学学报 (自然科学版), 2011, 32(12): 1680-1683.
SHA C M, LI Y, YANG D M. Robust stabilization for a class of uncertain nonlinear singular systems based on HJI differential inequality[J]. Journal of Northeastern University (Natural Science), 2011, 32(12): 1680-1683(in Chinese).

Attitude measurement and control method of flexible spacecraft based on distributed installation MSCSG

LI Lei[1], WANG Weijie[2], PANG Weikun[1], WANG Lifen[2,*], DUAN Leqiang[1], WANG Chenyu[1]

(1. Graduate School, Space Engineering University, Beijing 101416, China;
2. Department of Aerospace Science and Technology, Space Engineering University, Beijing 101416, China)

Abstract: In order to realize the integration of high-precision measurement and control of attitude vibration of a flexure spacecraft, an integrated method of attitude vibration measurement and control of a flexible spacecraft based on the distributed mounting of a magnetically suspended control sensitive gyroscope (MSCSG) and a radial basis function (RBF) neural network is proposed. Firstly, a spacecraft dynamics model with distributed mounted MSCSG on the flexible attachment is established, and the mechanism of vibration measurement and control of the flexible attachment by MSCSG is analyzed. Based on this, a neural network robust adaptive state feedback controller is designed, and the control torque output from MSCSG is realized suppression of low-frequency vibration of the flexible attachment while also measuring the deformation of the flexible attachment. The neural network approximates the nonlinear terms of the system and external perturbations brought about by distributed mounted MSCSG. The MSCSG cluster installed on the rigid body simultaneously realizes the measurement and control of spacecraft attitude. Simulation results show that the attitude control accuracy and vibration suppression effect of the proposed method are greatly improved compared with the traditional adaptive control method based on a modal observer.

Keywords: magnetically suspended control sensitive gyroscope; flexible spacecraft; radial basis function neural network; measurement and control integration; modal analysis

Received: 2024-03-11; **Accepted**: 2024-04-05; **Published Online**: 2024-05-29 14:33
URL: link.cnki.net/urlid/11.2625.V.20240528.1618.001
Foundation item: National Natural Science Foundation of China(52075545)
* **Corresponding author**. E-mail: wanglifen_2009@139.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0177

基于空时谱熵的分布式阵列雷达同步误差估计方法

耿雪胤，王俊，杨彬*，孙进平
(北京航空航天大学 电子信息工程学院，北京 100191)

摘 要：分布式阵列雷达(DAR)具有空域高分辨率、部署机动灵活等优点，但各单元雷达配置独立的时钟和振荡器及触发信号在馈线链路传输的不稳定性引入时间和相位同步误差，导致DAR相参合成精度降低。基于此，提出一种基于空时谱熵的时间、相位同步误差估计方法。建立含有时间和相位同步误差的空时协方差矩阵，以此构造距离-角度空时二维谱。依据信息熵理论建立同步误差与空时谱形状不确定度的对应关系，通过优化空时谱熵值，使其达到最小来估计时间和相位同步误差。仿真实验验证了所提方法的准确性，尤其在低信噪比时具有良好的估计性能。

关 键 词：分布式阵列雷达；时间相位同步误差；空时谱熵；空时二维谱；同步误差估计
中图分类号：TN958
文献标志码：A **文章编号**：1001-5965(2026)05-1627-08

阵列雷达引入多个天线对电磁波信号进行空域采样，能有效提高目标检测性能，具备强大的自适应抗干扰能力，已广泛应用在遥感、空间和空域监测、自动驾驶等多个领域[1]。阵列雷达角度维分辨率由波长与阵列孔径比值决定，高分辨的目标探测需求驱动其朝着大孔径、宽频带方向发展。然而，更大的阵列孔径导致雷达系统复杂度和工艺要求极大提升、制造成本昂贵、难以维护，限制了其实际应用范围[2]。一种有效的解决方案是将大孔径阵列分解为多个分布式的小孔径单元阵列雷达，通过相参合成以获得与大孔径相当的信噪比增益和分辨率[3-4]，这种分布式联合探测架构兼具机动性和高容错率，已成为近几年的研究热点。

早在1983年，Heimiller等系统介绍了分布式阵列雷达(distributed array radar, DAR)概念[5]，分析了均匀和随机排布的小孔径天线的合成方向图和分辨率特性及DAR接收的信噪比。2003年，美国“宙斯盾”舰载雷达[6]使用分布式阵列结构，并放置在舰船不同位置，经相参处理后，等效合成大孔径相控阵雷达，能够检测和跟踪具有小雷达散射截面的目标。近年来，北京理工大学在分布式雷达的时间、相位同步和全相参处理等方面开展广泛研究，2022年，联合多家单位建设的“中国复眼”，由多部25～30 m分布式雷达组成的超大孔径可实现高分辨率深空探测[1,7]。在自动驾驶等民用雷达应用中，分布式阵列体制也广受关注。文献[8]设计了一种大孔径调频连续波(frequency modulated continuous wave, FMCW)成像雷达，由15个分布式多发多收的节点组成，通过40 MHz参考时钟连接，具有0.1°的角度分辨率。文献[9]详细介绍了下一代汽车雷达网络中分布式阵列的结构、孔径设计和成像应用。

在DAR合成大孔径阵列时，时间和相位的高精度同步是相参合成的关键和前提。由于单元雷达配置独立的时钟和本振系统，触发时钟的参差和馈线链路不稳定性引入时间同步误差，振荡器的准确度和稳定性引入随机初始相位误差[1,10]。文献[11]分析了时间相位同步误差下分布式雷达同步误差估计的克拉美罗界(Cramer-Rao bound, CRB)，提出

收稿日期：2024-03-27；**录用日期**：2024-05-09；**网络出版时间**：2024-06-12 11:49
网络出版地址：link.cnki.net/urlid/11.2625.V.20240612.0919.001
基金项目：国家自然科学基金(62131001,62171029)；浙江省“领雁”研发攻关计划(2023C01148)
*通信作者. E-mail：young_being@buaa.edu.cn

引用格式：耿雪胤，王俊，杨彬，等. 基于空时谱熵的分布式阵列雷达同步误差估计方法[J]. 北京航空航天大学学报，2026，52(5)：1627-1634. GENG X Y，WANG J，YANG B，et al. Space-time spectral entropy based synchronization error estimation method for distributed array radar[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1627-1634 (in Chinese).

一种通用的信号处理方法，假设时间同步误差精确补偿，定量分析相位同步误差造成的增益损失。文献[12]分析了FMCW频率生成器中非理想锁相环引入的同步误差对距离估计的影响，表明单基地信号具有更高的信噪比和更低的相位噪声。DAR在相参合成时，按照是否需要合成目的地处的外部信号输入，可以分为闭环式架构和开环式架构[3]。闭环式架构根据回波信号估计相参合成所需的时间和相位同步误差，由于闭环架构无需精确已知坐标位置，在同步误差估计时备受关注[13]，常见的估计方法包括峰值法[5]、互相关法[14]和参数化方法[15]。文献[16]提出联合同步误差估计频率偏移，建立双基响应频率和相位偏的超定方程，在残差项中引入正弦函数最小二乘估计频率和相位同步误差。文献[17]依据最小输出能量(minimum output energy, MOE)准则，优化子阵权值向量，使目标函数最小来估计相位同步误差。文献[18]结合分布式阵列单元雷达与整个孔径的联合稀疏和低秩特征，设计了一种具有核范数正则化的$\ell_{1,2}$范数，求解子节点间的相位同步误差。上述估计方法都需要回波具有较高的信噪比，但实际中多种同步误差存在，回波信号难以直接通过多通道或多脉冲积累提高信噪比。

本文针对DAR合成大孔径阵列角度估计时，由于时间和相位同步误差导致距离-角度空时二维谱形状混乱的问题，提出一种基于空时谱熵的DAR同步误差估计方法。依据信息熵理论建立同步误差与空时谱形状不确定度的对应关系，通过最小化空时谱熵值来估计DAR的时间和相位同步误差。结果表明，本文方法能准确估计时间和相位同步误差，尤其在低信噪比时具有良好的估计性能。

1 分布式阵列雷达信号模型

考虑DAR由N个配有独立射频模块的单元阵列雷达节点组成，在雷达探测过程中，每个节点发射相互正交的FMCW波形形成宽波束，在接收端经匹配滤波形成N^2个虚拟节点来提高系统自由度。每个虚拟节点含有N_{t}个发射天线和N_{r}个接收天线，在虚拟节点接收信号时，需先分离出单基地信号和多基地信号[19]，其中，第p个节点($1\leqslant p\leqslant N$)发射的正交波形被第q个节点接收($1\leqslant q\leqslant N$)，当$p=q$时，信号为单基地信号，当$p\neq q$时，信号为多基地信号。DAR工作模式如图1所示。

以信号触发时刻为参考，信号抵达节点p时在馈线电路上传输时间为Δt_p，独立的振荡器具有初始相位ϕ_p。由于传输时间不一致及随机的初相，导致各节点发射的正交信号存在时间和相位同步误差，图2为FMCW波形中的同步误差示意图。在每个节点中以第1个发射阵元为参考，第p个节点中参考阵元发射信号$s_p(t)$的频率表示为

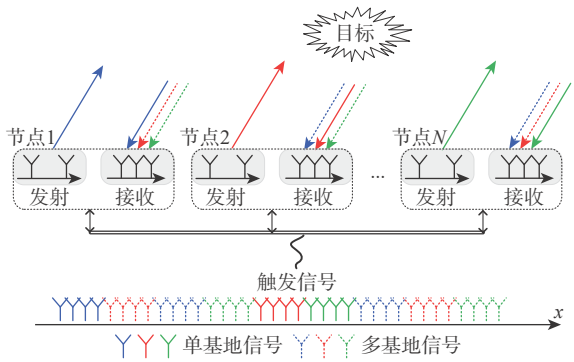


图1 分布式阵列雷达工作示意图

Fig. 1 Work diagram of distributed array radar

$$f_{\mathrm{T},p}(t)=f_{\mathrm{c}}+\frac{B}{T}(t-\Delta t_p) \tag{1}$$

式中：f_{c}为载频；B为发射信号带宽；T为脉冲重复周期。

调频斜率$\mu=B/T$，第p个节点相位表示为

$$\varphi_p(t)=2\pi\int_0^t f_{\mathrm{T},p}(t)\mathrm{d}t+\phi_p=2\pi\left(f_{\mathrm{c}}-\mu\Delta t_p\right)t+\pi\mu t^2+\phi_p \tag{2}$$

不失一般性，假设发射L个chirp幅度为1，有

$$s_p(t)=u_p(t)\mathrm{e}^{\mathrm{j}\varphi_p(t)} \tag{3}$$

式中：基带信号$u_p(t)=\sum_{l=0}^{L-1}h_p(t-lT)$，$h_p(t)$为发射信号的复包络。

假设分布式阵列相对探测场景是集中式布置，位于远场范围的目标k距离为R_k，速度为v_k，角度为θ_k。信号传播时延$\tau_k=2(R_k+v_kt)/c$，c为光速。接收端对正交波形分离，常见的波形正交方式有时分多

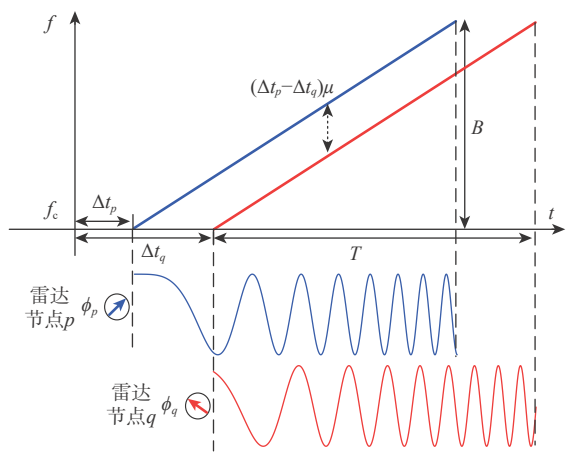


图2 p节点与q节点发射的FMCW信号含有时间和相位同步误差示意图

Fig. 2 Diagram of the FMCW signal transmitted by the pth and qth node with time and phase synchronization errors

址、频分多址和码分多址等, 对应的分离方法可参考文献 [20-22], 本文假设接收节点已分离出对应的发射波形。第p个节点发射波形$s_p(t)$经目标k反射的回波被第q个节点接收, 表示为

$$s_{pq,k}(t)=\alpha_k h_p(t-\tau_k)\mathrm{e}^{\mathrm{j}\varphi_p(t-\tau_k)} \tag{4}$$

式中: α_k为目标散射幅度。

接收信号$s_{pq,k}(t)$与本振信号$s_q(t)$混频后得到差频信号:

$$\tilde{s}_{pq,k}(t)=\begin{cases}\alpha_k h_p(t-\tau_k)\mathrm{e}^{\mathrm{j}2\pi\left[\left(f_k^{\mathrm{R}}+f_k^{\mathrm{D}}\right)t+\Phi_k^{\mathrm{R}}\right]}\cdot\mathrm{e}^{\mathrm{j}2\pi\left(\mu\Delta t_p-\mu\Delta t_q\right)t+\mathrm{j}\left(\phi_q-\phi_p\right)} & p\neq q\\ \alpha_k h_p(t-\tau_k)\mathrm{e}^{\mathrm{j}2\pi\left[\left(f_k^{\mathrm{R}}+f_k^{\mathrm{D}}\right)t+\Phi_k^{\mathrm{R}}\right]} & p=q\end{cases} \tag{5}$$

$$\begin{cases}f_k^{\mathrm{R}}=2R_k\mu/c\\ f_k^{\mathrm{D}}=2v_k(f_{\mathrm{c}}-\mu\Delta t_p)/c\\ \Phi_k^{\mathrm{R}}=2R_k(f_{\mathrm{c}}-\mu\Delta t_p)/c\end{cases} \tag{6}$$

由式(5)可知, 单基地信号中不含时间和相位同步误差, 但多基地信号引入由时间同步误差导致的频率偏置$\mu\Delta t_p-\mu\Delta t_q$及由初始相位误差导致的相位偏置$\phi_q-\phi_p$。

考虑波形分离后, 节点p发射信号被节点q接收形成的虚拟节点上, 子阵列的信号:

$$\boldsymbol{y}^{(pq)}(n_{\mathrm{s}})=\boldsymbol{A}^{(pq)}(\boldsymbol{\theta})\tilde{\boldsymbol{s}}_{pq}(n_{\mathrm{s}})+\boldsymbol{\eta}^{(pq)}(n_{\mathrm{s}}) \tag{7}$$

式中: $\tilde{\boldsymbol{s}}_{pq}(n_{\mathrm{s}})$为所有$K$个目标的回波差频信号经过模数转换器(analog-to-digital converter, ADC)采样得到的离散化序列信号, $\tilde{\boldsymbol{s}}_{pq}(n_{\mathrm{s}})=[\tilde{s}_{pq,1}(n_{\mathrm{s}}),\tilde{s}_{pq,2}(n_{\mathrm{s}}),\cdots,\tilde{s}_{pq,K}(n_{\mathrm{s}})]^{\mathrm{T}}$, $n_{\mathrm{s}}=1,2,\cdots,N_{\mathrm{S}}$为时间采样快拍; $\boldsymbol{\eta}^{(pq)}(n_{\mathrm{s}})$为满足空时独立同分布的零均值加性白噪声; $\boldsymbol{A}^{(pq)}(\boldsymbol{\theta})=[\boldsymbol{a}^{(pq)}(\theta_1),\boldsymbol{a}^{(pq)}(\theta_2),\cdots,\boldsymbol{a}^{(pq)}(\theta_K)]$为子阵导向矢量, 维度为$N_{\mathrm{t}}N_{\mathrm{r}}\times K$。

$$\boldsymbol{a}^{(pq)}(\theta_k)=\boldsymbol{a}_{\mathrm{T}}^{(pq)}(\theta_k)\otimes\boldsymbol{a}_{\mathrm{R}}^{(pq)}(\theta_k) \tag{8}$$

式中: "⊗"表示克罗内克积。

假设$d_{\mathrm{T},n_{\mathrm{t}}}^{(pq)}$和$d_{\mathrm{R},n_{\mathrm{r}}}^{(pq)}$($n_{\mathrm{t}}$=1,2, …, N_{t}; n_{r}=1,2, …, N_{r})是发射和接收阵元相对于坐标原点位置, 有

$$\boldsymbol{a}_{\mathrm{T}}^{(pq)}(\theta_k)=\left[\mathrm{e}^{-\mathrm{j}2\pi d_{\mathrm{T},1}^{(pq)}\sin\theta_k/\lambda},\mathrm{e}^{-\mathrm{j}2\pi d_{\mathrm{T},2}^{(pq)}\sin\theta_k/\lambda},\cdots,\mathrm{e}^{-\mathrm{j}2\pi d_{\mathrm{T},N_{\mathrm{t}}}^{(pq)}\sin\theta_k/\lambda}\right]^{\mathrm{T}} \tag{9}$$

$$\boldsymbol{a}_{\mathrm{R}}^{(pq)}(\theta_k)=\left[\mathrm{e}^{-\mathrm{j}2\pi d_{\mathrm{R},1}^{(pq)}\sin\theta_k/\lambda},\mathrm{e}^{-\mathrm{j}2\pi d_{\mathrm{R},2}^{(pq)}\sin\theta_k/\lambda},\cdots,\mathrm{e}^{-\mathrm{j}2\pi d_{\mathrm{R},N_{\mathrm{r}}}^{(pq)}\sin\theta_k/\lambda}\right]^{\mathrm{T}} \tag{10}$$

式中: 信号波长$\lambda=c/(f_{\mathrm{c}}-\mu\Delta t_p)\approx c/f_{\mathrm{c}}$。

将所有分布式虚拟节点采样信号排列为$N^2N_{\mathrm{t}}N_{\mathrm{r}}\times K$维接收矩阵$\boldsymbol{Y}=[\boldsymbol{y}(1),\boldsymbol{y}(2),\cdots,\boldsymbol{y}(N_{\mathrm{S}})]$, 其中, 在采样时刻$n_{\mathrm{s}}$的所有阵元接收信号$\boldsymbol{y}(n_{\mathrm{s}})$为

$$\boldsymbol{y}(n_{\mathrm{s}})=[\boldsymbol{y}^{(11)}(n_{\mathrm{s}})^{\mathrm{T}},\cdots,\boldsymbol{y}^{(1N)}(n_{\mathrm{s}})^{\mathrm{T}},\cdots,\boldsymbol{y}^{(NN)}(n_{\mathrm{s}})^{\mathrm{T}}]^{\mathrm{T}} \tag{11}$$

式中: $\boldsymbol{y}^{(NN)}(n_{\mathrm{s}})^{\mathrm{T}}$表示根据时间维度和空间维度确定的空时矩阵。

假设一个 chirp 内目标速度不变, 对每个虚拟天线的时域离散信号做傅里叶变换得到一维距离像, 从检测的距离谱峰相位中提取角度谱。图 3 显示时间误差导致一维距离谱峰偏移, 相位误差导致角度谱旁瓣抬升、主瓣展宽甚至散焦[23]。

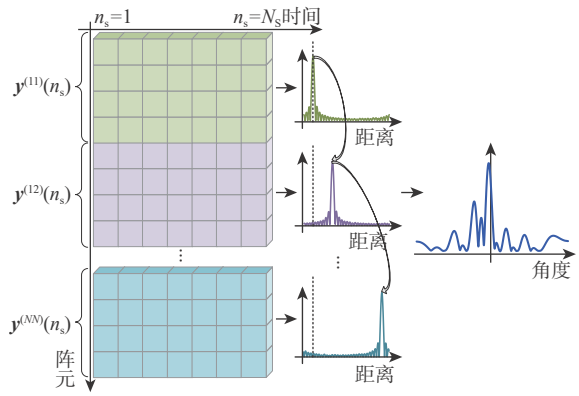


图 3　接收数据矩阵及具有时间与相位同步误差的距离谱与角度谱

Fig. 3　Received data matrix and range and angle spectrum with time and phase synchronization errors

2　基于空时谱熵的同步误差估计

首先, 构造含有时间和相位同步误差的距离-角度空时二维谱; 然后, 从谱形状不确定性度量的角度出发, 依据信息熵理论建立同步误差与空时二维谱图不确定性的对应关系; 最后, 通过最小化空时谱熵值来估计 DAR 的时偏和相偏误差。

2.1　空时二维谱构造

空时二维谱估计在空间谱的基础上增加时域信息, 可提高参数估计精度和分辨率[24]。在宽频段、广空域中, 空时二维谱具有同时测角和测频率的能力[25]。构造角度的谱搜索点$\bar{\boldsymbol{\theta}}=[\bar{\theta}_1,\bar{\theta}_2,\cdots,\bar{\theta}_{N_\theta}]$和距离的谱搜索点$\bar{\boldsymbol{r}}=[\bar{r}_1,\bar{r}_2,\cdots,\bar{r}_{N_{\mathrm{d}}}]$。角度和距离维搜索导向矢量:

$$\bar{\boldsymbol{a}}\left(\bar{\theta}_{n_\theta}\right)=\left[\boldsymbol{a}^{(11)}(\bar{\theta}_{n_\theta})^{\mathrm{T}},\cdots,\boldsymbol{a}^{(1N)}(\bar{\theta}_{n_\theta})^{\mathrm{T}},\cdots,\boldsymbol{a}^{(NN)}(\bar{\theta}_{n_\theta})^{\mathrm{T}}\right]^{\mathrm{T}} \tag{12}$$

$$\bar{\boldsymbol{a}}\left(\bar{r}_{n_{\mathrm{d}}}\right)=\left[\mathrm{e}^{-\mathrm{j}2\pi\mu\frac{2\bar{r}_{n_{\mathrm{d}}}}{c}t(1)},\mathrm{e}^{-\mathrm{j}2\pi\mu\frac{2\bar{r}_{n_{\mathrm{d}}}}{c}t(2)},\cdots,\mathrm{e}^{-\mathrm{j}2\pi\mu\frac{2\bar{r}_{n_{\mathrm{d}}}}{c}t(N_{\mathrm{S}})}\right]^{\mathrm{T}} \tag{13}$$

式中: $t(1),t(2),\cdots,t(N_{\mathrm{S}})$表示时间维$N_{\mathrm{S}}$个物理采样时刻; n_θ=1,2, …, N_θ; n_{d}=1,2, …, N_{d}。

由式(12)和式(13)可确定空时导向矢量:

$$\begin{aligned}\bar{\boldsymbol{A}}\left(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{r}}\right)&=\bar{\boldsymbol{A}}\left(\bar{\boldsymbol{\theta}}\right)\odot\bar{\boldsymbol{A}}\left(\bar{\boldsymbol{r}}\right)=\\ &\left[\bar{\boldsymbol{A}}\left(\bar{\theta}_1,\bar{r}_1\right),\bar{\boldsymbol{A}}\left(\bar{\theta}_2,\bar{r}_2\right),\cdots,\bar{\boldsymbol{A}}\left(\bar{\theta}_{N_\theta},\bar{r}_{N_{\mathrm{d}}}\right)\right]=\\ &\left[\bar{\boldsymbol{a}}\left(\bar{\theta}_1\right)\otimes\bar{\boldsymbol{a}}\left(\bar{r}_1\right),\bar{\boldsymbol{a}}\left(\bar{\theta}_2\right)\otimes\bar{\boldsymbol{a}}\left(\bar{r}_2\right),\cdots,\bar{\boldsymbol{a}}\left(\bar{\theta}_{N_\theta}\right)\otimes\bar{\boldsymbol{a}}\left(\bar{r}_{N_{\mathrm{d}}}\right)\right]\end{aligned} \tag{14}$$

式中：“⊙”表示Khatri-Rao积，是一种逐列求克罗内克积的运算。

为提高算法适应性，需对接收数据做平滑处理，将时间和相位同步误差表示在平滑矩阵中，通过求解各子协方差矩阵的均值来恢复满秩空时协方差矩阵。平滑矩阵$\breve{\boldsymbol{X}}$的构造方式如图4所示，其中，$l_1\times l_2$为空时滑窗维度，$\mathcal{P}_1=N^2N_tN_r-l_1+1$为接收数据空间维平滑次数，$\mathcal{P}_2=N_S-l_2+1$为时间维平滑次数。每个滑窗的数据重新排列形成$l_1l_2\times 1$维的列矢量数据，将所有滑窗数据重排列形成$l_1l_2\times\mathcal{P}_1\mathcal{P}_2$维的平滑矩阵$\breve{\boldsymbol{X}}$，此时，空时平滑协方差矩阵$\hat{\boldsymbol{R}}$表示为

$$\hat{\boldsymbol{R}}=\frac{1}{2\mathcal{P}_1\mathcal{P}_2}\left[\breve{\boldsymbol{X}}\breve{\boldsymbol{X}}^{\mathrm{H}}+\boldsymbol{J}\left(\breve{\boldsymbol{X}}\breve{\boldsymbol{X}}^{\mathrm{H}}\right)^{*}\boldsymbol{J}\right] \tag{15}$$

式中：置换矩阵为$l_1l_2\times l_1l_2$维的反对角矩阵：

$$\boldsymbol{J}=\begin{bmatrix}0&0&\cdots&0&1\\0&0&\cdots&1&0\\\vdots&\vdots&&\vdots&\vdots\\0&1&\cdots&0&0\\1&0&\cdots&0&0\end{bmatrix} \tag{16}$$

平滑后的协方差矩阵$\hat{\boldsymbol{R}}$的秩与平滑次数相关，表示为$\mathrm{rank}(\hat{\boldsymbol{R}})=\min(l_1l_2,\mathcal{P}_1\mathcal{P}_2)$。

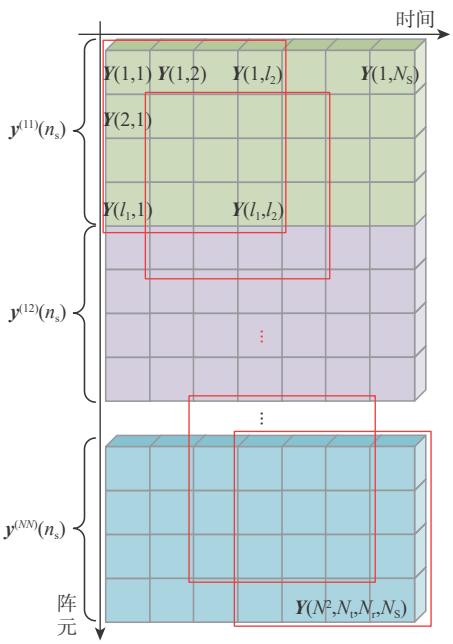


图4 空间平滑数据矩阵与滑窗

Fig. 4 Spatial smoothing data matrix and sliding window

Capon最小方差无畸变(minimum variance distortionless response, MVDR)波束形成器无需约束信号模型，在增强有用信号和抑制干扰方面具有优势[26]。本文采用Capon方法生成空时二维谱，通过谱峰搜索来获得功率谱函数，表示为

$$\boldsymbol{P}(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{r}})=\frac{1}{\overline{\boldsymbol{A}}(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{r}})^{\mathrm{H}}\hat{\boldsymbol{R}}^{-1}\overline{\boldsymbol{A}}(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{r}})} \tag{17}$$

当各节点没有同步误差时，目标会在距离-角度谱中聚焦。同步误差的存在导致期望信号导向矢量失配，而Capon波束形成器对导向矢量失配非常敏感，会引起空时二维谱主瓣展宽、积分旁瓣比下降等。

2.2 时间和相位同步误差估计方法

信息论中，熵用来定量描述样本中包含的信息平均量。文献[27]利用谱熵区分规则和不规则运动，规则运动是由一系列主频率组成，在不规则运动中频率分量扩散，导致谱熵增加。文献[28]在宽幅合成孔径雷达(synthetic aperture radar, SAR)中利用图像熵最小的准则校正多通道的相位差。受此启发，本节设计空时谱熵估计时间和相位同步误差。

从校准的思路出发，根据式(5)构造每个节点时间和相位待补偿的变量矩阵：

$$\widehat{\boldsymbol{\Phi}}=\mathrm{blkdiag}\left(\left[\widehat{\varphi}^{(11)}\boldsymbol{I}_{N_tN_r},\widehat{\varphi}^{(12)}\boldsymbol{I}_{N_tN_r},\cdots,\widehat{\varphi}^{(NN)}\boldsymbol{I}_{N_tN_r}\right]\right) \tag{18}$$

$$\widehat{\boldsymbol{\Gamma}}=\left[\underbrace{\widehat{\gamma}^{(11)},\cdots,\widehat{\gamma}^{(11)}}_{N_tN_r},\underbrace{\widehat{\gamma}^{(12)},\cdots,\widehat{\gamma}^{(12)}}_{N_tN_r},\cdots\underbrace{\widehat{\gamma}^{(NN)},\cdots,\widehat{\gamma}^{(NN)}}_{N_tN_r}\right] \tag{19}$$

式中：blkdiag(·)为块对角矩阵；$\boldsymbol{I}_{N_tN_r}$为N_tN_r维单位矩阵；$\widehat{\boldsymbol{\Phi}}$和$\widehat{\boldsymbol{\Gamma}}$的维度分别为$N^2N_tN_r\times N^2N_tN_r$和$N^2N_tN_r\times N_S$；$\widehat{\varphi}^{(pq)}$和$\widehat{\gamma}^{(pq)}$为每个虚拟节点应该补偿的相位因子，单基地信号的同步误差在匹配滤波时被对消掉，因此，只需补偿多基地信号的时间和相位同步误差。

$$\widehat{\varphi}^{(pq)}=\begin{cases}1 & p=q\\ \mathrm{e}^{\mathrm{j}\phi_{pq}} & p\neq q\end{cases} \tag{20}$$

$$\widehat{\gamma}^{(pq)}=\begin{cases}1 & p=q\\ \mathrm{e}^{\mathrm{j}2\pi\mu(\Delta t_{pq})t} & p\neq q\end{cases} \tag{21}$$

利用式(18)和式(19)的校正相位来重构空时二维谱：

$$\widehat{\boldsymbol{P}}\left(\widehat{\boldsymbol{\Phi}},\widehat{\boldsymbol{\Gamma}},\bar{\boldsymbol{\theta}},\bar{\boldsymbol{r}}\right)=\frac{1}{\overline{\boldsymbol{A}}(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{r}})^{\mathrm{H}}\widehat{\boldsymbol{R}}^{-1}\overline{\boldsymbol{A}}(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{r}})}\widehat{\boldsymbol{Y}}=\left(\widehat{\boldsymbol{\Phi}}^{*}\widehat{\boldsymbol{\Gamma}}^{*}\right)\circ\boldsymbol{Y} \tag{22}$$

式中：“*”表示共轭复数；“∘”表示哈达玛积；重构协方差矩阵$\widehat{\boldsymbol{R}}$是$\widehat{\boldsymbol{Y}}$由图4的方式构造平滑矩阵，依据式(15)得到。

不同的校正系数对应不同的空时谱形状，引入空时谱熵来评价重构距离角度谱的能量聚集情况：

$$\boldsymbol{E}\left(\widehat{\boldsymbol{\Phi}},\widehat{\boldsymbol{\Gamma}}\right)=\sum_{\bar{\theta}}\sum_{\bar{r}}\frac{\widehat{\boldsymbol{P}}\left(\widehat{\boldsymbol{\Phi}},\widehat{\boldsymbol{\Gamma}},\bar{\boldsymbol{\theta}},\bar{\boldsymbol{r}}\right)}{E_{\mathrm{sum}}}\ln\left(\frac{E_{\mathrm{sum}}}{\widehat{\boldsymbol{P}}\left(\widehat{\boldsymbol{\Phi}},\widehat{\boldsymbol{\Gamma}},\bar{\boldsymbol{\theta}},\bar{\boldsymbol{r}}\right)}\right) \tag{23}$$

$$E_{\text{sum}}=\sum_{\bar{\theta}}\sum_{\bar{r}}\widehat{\boldsymbol{P}}\left(\widehat{\boldsymbol{\Phi}},\widehat{\boldsymbol{\Gamma}},\bar{\theta}_{n_\theta},\bar{r}_{n_\mathrm{d}}\right)\tag{24}$$

式中: $\widehat{\boldsymbol{P}}\left(\widehat{\boldsymbol{\Phi}},\widehat{\boldsymbol{\Gamma}},\bar{\theta},\bar{r}\right)/E_{\text{sum}}$ 为每个空时谱搜索点中的能量占整个空时域能量的比例, 也可视为谱能量在各完备字典上的后验概率分布。由熵的性质可知, 当空时谱中所有能量都集中在一个分辨单元时, 熵值最小。当能量分布越随机, 表现为越“无序”, 包含的信息量越大, 即熵值越大; 当能量完全随机分布在各字典上时, 熵值达最大。由于同步误差引起的主瓣展宽和旁瓣抬升等问题最终都体现在熵值增加, 可通过优化补偿变量, 使二维谱熵最小:

$$\hat{\hat{\boldsymbol{\Phi}}},\hat{\hat{\boldsymbol{\Gamma}}}=\min_{\phi,\Delta t}\left(\boldsymbol{E}\left(\widehat{\boldsymbol{\Phi}},\widehat{\boldsymbol{\Gamma}}\right)\right)\tag{25}$$

式(25)表示在 DAR 中的时间和相位同步误差估计问题转化为空时谱熵最小问题求解。群体智能任务分配算法具有较高的鲁棒性和自适应能力, 其中, 粒子群优化算法[29]有全局寻优强、收敛速度快等优势。求解二维谱熵最小化问题(式(25))时, 采用粒子群优化算法求解得到时间和相位同步参数。

总结本文提出的基于空时谱熵的同步误差估计方法的设计步骤为: ①依据式 (18) 和式 (19) 构造时间和相位补偿矩阵, 依据式 (22) 获得重构的子阵列信号 $\widehat{\boldsymbol{Y}}$; ②依据式 (15) 和空时滑窗维度获得平滑协方差矩阵 $\widehat{\boldsymbol{R}}$; ③依据式 (24) 得到整个角度和距离搜索空间的空时谱熵值 E_{sum}; ④采用粒子群优化算法求解式 (25), 使其最小来估计时间和相位同步误差。

3　仿真分析

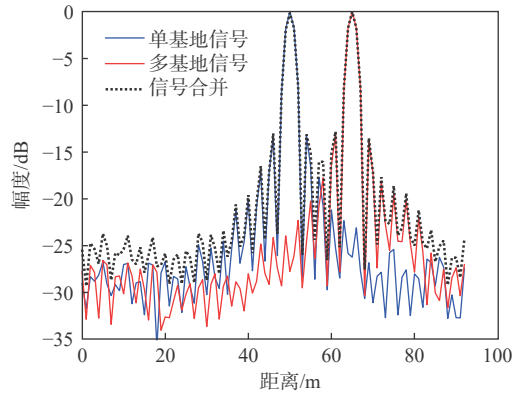
本节通过仿真数据验证本文提出的 DAR 时间和相位同步误差估计方法的有效性。接收的信号由单基地回波和多基地回波组成, 每个虚拟节点经匹配滤波后可得到 6 个均匀排布的虚拟阵元。雷达发射 FMCW 信号, 每个阵元接收 32 个采样快拍数据, 具体波形参数如表 1 所示。

表 1　雷达仿真参数

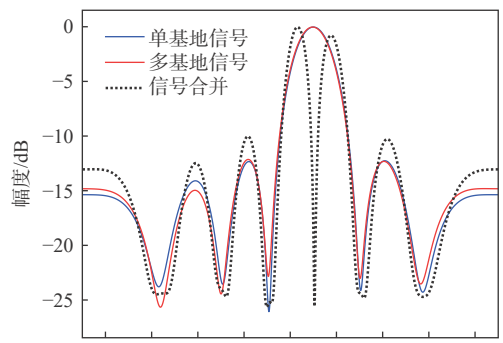
Table 1　Radar simulation parameters

载频/GHz	带宽/MHz	脉冲重复周期/μs	调频斜率/(MHz·μs^{-1})
77	500	7.3	68.2

1) 实验 1。考虑 DAR 多基地信号间有时间和相位同步误差, 分析同步误差对合成的距离、角度谱和空时二维谱的影响。在该仿真中, 设置单个目标位于角度 10°和距离 50 m 处, 多基地节点信号时间和相位同步误差分别为 100 ns 和 3 rad, 设置脉冲压缩后的信噪比为 10 dB。图 5(a) 和图 5(b) 给出了单基地、多基地基和合并后的节点信号的一维距离像和角度谱。时间同步误差导致的频率偏置在一维距离像上表现为虚假的谱峰, 该峰值位于距离为 64.85 m 处。节点间的相位误差不影响独立子阵测角性能, 当 2 个节点合并为大阵列相参测角时, 相位同步误差会导致角度谱出现伪峰及旁瓣抬高, 最终都体现为空时谱熵增加。



(a) 时间同步误差下各节点距离谱



(b) 相位同步误差下各节点角度谱

图 5　时间和相位同步误差下各节点距离和角度谱

Fig. 5　Range and angle spectrum with time and phase synchronization errors for each node

图 6 给出了单基地、多基地信号合并后的距离-角度空时谱熵值随同步误差大小的变化。在确保运算精度的同时, 保持较高的运算效率, 设置距离和角

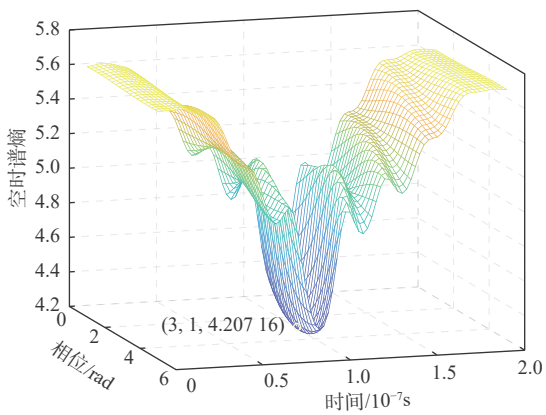
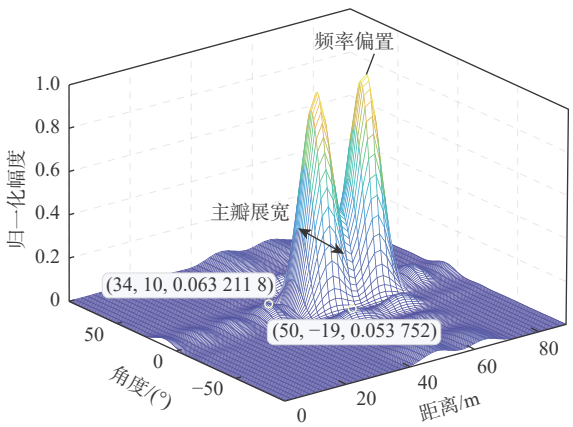


图 6　时间/相位同步误差与空时谱熵的关系

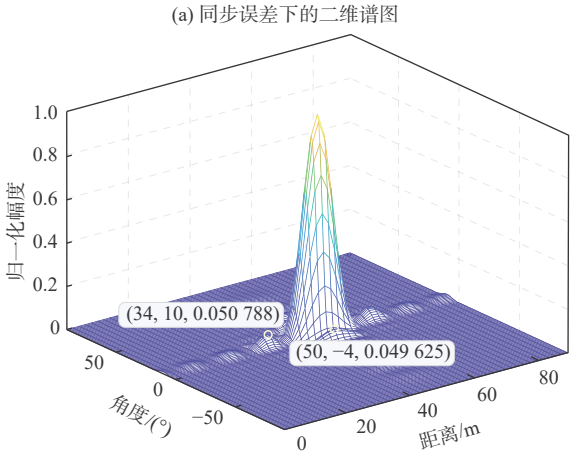
Fig. 6　Relationship between time/phase synchronization errors and space-time spectral entropy

度维滑窗长度l_1=6, l_2=8。由图 6 可以看出，基于二维空时谱熵的同步误差估计方法具有全局收敛特性，时间同步误差主要影响熵值大小，当熵值最小时，对应的时间和相位估计值为 100 ns 和 3 rad。

图 7 给出了优化校正同步误差前后的空时二维谱。粒子群求解时间和相位同步参数时，为了获得收敛的全局最优值，设置初始种群数量为 100，最大迭代次数为 30，时间和相位搜索范围分别是 [0,200] ns 和 [0,2π] rad，迭代一次的计算时间数量级为秒级。对比图 7(a)和图 7(b)可知，校准的角度谱主瓣宽度从 17°降低为 8°，由频率偏置导致的相位偏差与单基信号频率对齐。此外，距离维的归一化第 1 旁瓣幅度从 0.063 211 8 降低为 0.050 788，角度维的归一化旁瓣幅度从 0.053 752 降低为 0.049 625。



(a) 同步误差下的二维谱图

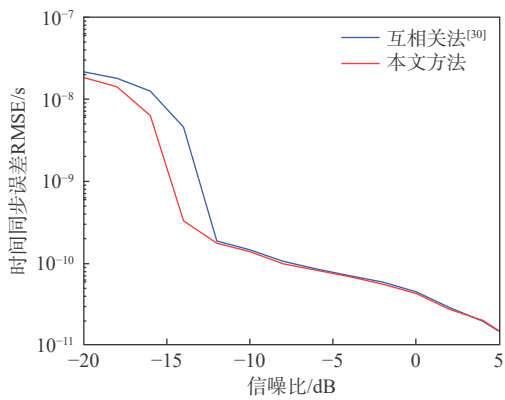


(b) 时间、相位误差同步后二维谱图

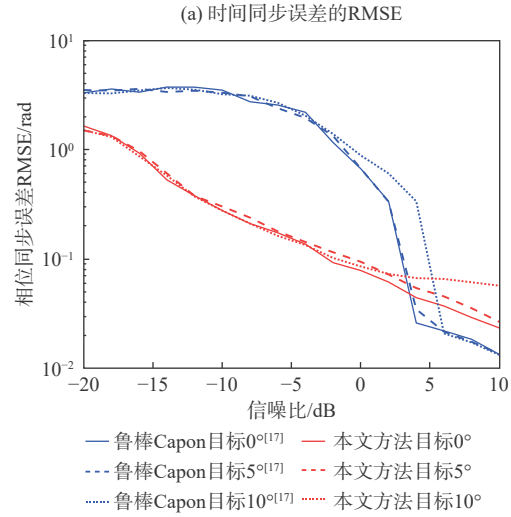
图 7 同步误差校正前后二维谱校正对比

Fig. 7 Comparison of two-dimensional spectrum before and after synchronization error compensation

2）实验 2。考虑 DAR 在不同信噪比下的同步误差估计精度，仿真设置的信噪比区间为 [−20,10] dB，步长为 2 dB，每组仿真进行 300 次蒙特卡罗实验，以均方根误差(root mean squared error, RMSE)作为参数估计性能指标。图 8(a) 给出了时间同步误差估计 RMSE 与信噪比关系，对比文献 [30] 中的互相关法求多基地信号间的时延参数，所提基于信息熵时延估计的方法总体都优于互相关法，且在低于 −12 dB 的低信噪比下，估计精度有明显优势。图 8(b) 给出了相位同步误差估计 RMSE 与信噪比的关系，对比文献 [17] 中的鲁棒 Capon 方法，本文方法通过最大化最小方差波束形成器的输出功率求解多基地信号间的相位误差，在输入信噪比低于 3 dB 的条件下，相位估计精度高。此外，相比鲁棒 Capon 方法[17]，本文方法只利用空间谱的谱形状混乱程度，无需联合估计目标到达角，计算复杂度较低。图 8(b) 比较了不同目标角度下的相位偏差估计性能，当目标位于 0°时，估计精度最高，随着目标角度偏离雷达法线方向，估计精度开始下降。



(a) 时间同步误差的RMSE



(b) 相位同步误差的RMSE

图 8 不同信噪比下时间与相位同步误差估计均方根误差

Fig. 8 Root mean squared error of time and phase synchronization error estimation under different signal-to-noise ratios

4 结 论

本文研究了 DAR 相参合成时同步误差的影响，提出一种基于空时谱熵的时间和相位同步误差估计方法，设计仿真实验对本文方法进行了验证。结论如下：

1）在 FMCW 信号体制下，时间同步误差导致

的频率偏置引起虚假距离谱峰, 相参测角时, 相位同步误差导致角度谱出现伪峰、旁瓣抬高, 这些现象都体现为空时谱熵值增加。

2) 建立含有时间和相位误差的单基地及多基地阵列回波信号模型, 基于 Capon 波束形成器构造了空时二维谱, 引入空时谱熵建立谱聚焦评价准则, 通过最小化空时谱熵值估计时间和相位同步误差。

3) 通过数值仿真验证了本文方法有效性, 相比互相关法和鲁棒 Capon 方法, 本文方法在低输入信噪比条件下, 具有较高的同步误差估计精度。

参考文献 (References)

[1] VAN TREES H L. Optimum array processing: part IV of detection, estimation, and modulation theory[M]. Hoboken: Wiley, 2002.

[2] 鲁耀兵, 高红卫. 分布孔径雷达[M]. 北京: 国防工业出版社, 2017: 216-217.
LU Y B, GAO H W. Distributed aperture radar[M]. Beijing: National Defense Industry Press, 2017: 216-217(in Chinese).

[3] 刘兴华, 王国玉, 徐振海, 等. 分布式孔径相参合成原理、发展与技术实现综述[J]. 雷达学报, 2023, 12(6): 1229-1248.
LIU X H, WANG G Y, XU Z H, et al. Review of principles, development and technical implementation of coherently combining distributed apertures[J]. Journal of Radars, 2023, 12(6): 1229-1248(in Chinese).

[4] 鲁耀兵, 高红卫, 周宝亮. 分布式孔径相参合成雷达技术[J]. 雷达学报, 2017, 6(1): 55-64.
LU Y B, GAO H W, ZHOU B L. Distributed aperture coherence-synthetic radar technology[J]. Journal of Radars, 2017, 6(1): 55-64 (in Chinese).

[5] HEIMILLER R C, BELYEA J E, TOMLINSON P G. Distributed array radar[J]. IEEE Transactions on Aerospace and Electronic Systems, 1983, 19(6): 831-839.

[6] LIN C. Distributed subarray antennas for multifunction phased-array radar[D]. Monterey: Naval Postgraduate School, 2003.

[7] 刘泉华, 张凯翔, 梁振楠, 等. 地基分布式相参雷达技术研究综述[J]. 信号处理, 2022, 38(12): 2443-2459.
LIU Q H, ZHANG K X, LIANG Z N, et al. Research overview of ground-based distributed coherent aperture radar[J]. Journal of Signal Processing, 2022, 38(12): 2443-2459(in Chinese).

[8] BIALER O, JONAS A, TIRER T. Super resolution wide aperture automotive radar[J]. IEEE Sensors Journal, 2021, 21(16): 17846-17858.

[9] GOTTINGER M, HOFFMANN M, CHRISTMANN M, et al. Coherent automotive radar networks: the next generation of radar-based imaging and mapping[J]. IEEE Journal of Microwaves, 2021, 1(1): 149-163.

[10] 刘晓瑜, 吴建新, 王彤, 等. 基于多特显点的无人机分布式相参雷达相位同步误差估计方法[J]. 系统工程与电子技术, 2020, 42(4): 781-791.
LIU X Y, WU J X, WANG T, et al. Phase synchronization error estimation based on prominent scatterers for UAV distributed coherent aperture radar[J]. Systems Engineering and Electronics, 2020, 42(4): 781-791(in Chinese).

[11] SUN P L, TANG J, HE Q, et al. Cramer-Rao bound of parameters estimation and coherence performance for next generation radar[J]. IET Radar, Sonar & Navigation, 2013, 7(5): 553-567.

[12] FRISCHEN A, HASCH J, WALDSCHMIDT C. A cooperative MIMO radar network using highly integrated FMCW radar sensors[J]. IEEE Transactions on Microwave Theory and Techniques, 2017, 65(4): 1355-1366.

[13] NANZER J A, MGHABGHAB S R, ELLISON S M, et al. Distributed phased arrays: challenges and recent advances[J]. IEEE Transactions on Microwave Theory and Techniques, 2021, 69(11): 4893-4907.

[14] 宋靖, 周青松, 张剑云. 基于相关法的分布式全相参雷达相干参数估计及相参性能[J]. 电子与信息学报, 2015, 37(7): 1710-1715.
SONG J, ZHOU Q S, ZHANG J Y. Coherent parameters estimation by cross-correlation for distributed aperture fully coherent radar[J]. Journal of Electronics & Information Technology, 2015, 37(7): 1710-1715(in Chinese).

[15] KUNG S Y, ARUN K S, BHASKAR RAO D V. State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem[J]. JOSA, 1983, 73(12): 1799-1811.

[16] FRISCHEN A, HAKOBYAN G, WALDSCHMIDT C. Coherent measurements with MIMO radar networks of incoherent FMCW sensor nodes[J]. IEEE Microwave and Wireless Components Letters, 2020, 30(7): 721-724.

[17] LEI L, LIE J P, GERSHMAN A B, et al. Robust adaptive beamforming in partly calibrated sparse sensor arrays[J]. IEEE Transactions on Signal Processing, 2010, 58(3): 1661-1667.

[18] TIRER T, BIALER O. Direction of arrival estimation and phase-correction for noncoherent subarrays: a convex optimization approach[J]. IEEE Transactions on Aerospace and Electronic Systems, 2022, 58(6): 5571-5585.

[19] 刘兴华, 徐振海, 王罗胜斌, 等. 基于信号重建的分布式相参雷达相参参数估计算法[J]. 系统工程与电子技术, 2018, 40(9): 1931-1938.
LIU X H, XU Z H, WANG L S B, et al. Coherent parameters estimation algorithm for distributed coherent aperture radar based on signal reconstruction[J]. Systems Engineering and Electronics, 2018, 40(9): 1931-1938(in Chinese).

[20] SUN H B, BRIGUI F, LESTURGIE M. Analysis and comparison of MIMO radar waveforms[C]//Proceedings of the 2014 International Radar Conference. Piscataway: IEEE Press, 2015: 1-6.

[21] SUN S Q, PETROPULU A P, POOR H V. MIMO radar for advanced driver-assistance systems and autonomous driving: advantages and challenges[J]. IEEE Signal Processing Magazine, 2020, 37(4): 98-117.

[22] SOLODKY G, LONGMAN O, VILLEVAL S, et al. CDMA-MIMO radar with the tansec waveform[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(1): 76-89.

[23] BAKR O M, JOHNSON M. Impact of phase and amplitude errors on array performance: UCB/EECS-2009-1[R]. Berkeley: Electrical Engineering and Computer Sciences University of California at Berkeley, 2009: 80.

[24] 王永良. 空间谱估计理论与算法[M]. 北京: 清华大学出版社, 2004.
WANG Y L. Theory and algorithm of spatial spectrum estimation[M]. Beijing: Tsinghua University Press, 2004(in Chinese).

[25] 张泽, 陈辉, 王永良. 空时二维线性预测谱估计算法[J]. 系统工程与电子技术, 2019, 41(9): 1937-1944.
ZHANG Z, CHEN H, WANG Y L. Linear predictive spectrum estimation algorithm based on space-time two-dimensional[J]. Systems Engineering and Electronics, 2019, 41(9): 1937-1944(in Chinese).

[26] CAPON J. High-resolution frequency-wavenumber spectrum analysis[J]. Proceedings of the IEEE, 1969, 57(8): 1408-1418.

[27] POWELL G E, PERCIVAL I C. A spectral entropy method for distinguishing regular and irregular motion of Hamiltonian systems[J]. Journal of Physics A: Mathematical and General, 1979, 12(11): 2053.

[28] ZHANG S X, XING M D, XIA X G, et al. A robust channel-calibration algorithm for multi-channel in azimuth HRWS SAR imaging based on local maximum-likelihood weighted minimum entropy[J]. IEEE Transactions on Image Processing, 2013, 22(12): 5294-5305.

[29] KENNEDY J, EBERHART R. Particle swarm optimization[C]// Proceedings of the International Conference on Neural Networks. Piscataway: IEEE Press, 2002: 1942-1948.

[30] 王雪琦, 涂刚毅, 吴少鹏. 分布式相参雷达多脉冲积累相参参数估计方法[J]. 太赫兹科学与电子信息学报, 2020, 18(6): 1003-1009.
WANG X Q, TU G Y, WU S P. Coherent parameters estimation method for distributed coherent radar based on multi-pulse accumulation[J]. Journal of Terahertz Science and Electronic Information Technology, 2020, 18(6): 1003-1009(in Chinese).

Space-time spectral entropy based synchronization error estimation method for distributed array radar

GENG Xueyin，WANG Jun，YANG Bin*，SUN Jinping

(School of Electronic and Information Engineering，Beihang University，Beijing 100191，China)

Abstract：Distributed array radar (DAR) has the benefits including flexible deployment and excellent spatial resolution. However, the independent clock and oscillator configurations of each radar unit, along with the instability introduced by trigger signals transmitted over the feeder link, introduce errors in time and phase synchronization, thereby reducing the accuracy of coherent synthesis in a DAR. This paper proposes a time and phase synchronization error estimation method based on space-time spectral entropy. The distance-angle space-time two-dimensional spectrum is first constructed by establishing a space-time covariance matrix with time and phase synchronization faults. Then, based on entropy theory, the correspondence between synchronization errors and space-time spectral shape uncertainty is established. The time and phase synchronization errors are estimated by optimizing the space-time spectral entropy to minimize its value. Simulation experiments validate the accuracy of the proposed method, particularly exhibiting good estimation performance under low signal-to-noise ratio.

Keywords：distributed array radar; time and phase synchronization error; space-time spectral entropy; space-time two-dimensional spectrum; synchronization error estimation

Received：2024-03-27；**Accepted**：2024-05-09；**Published Online**：2024-06-12 11：49

URL：link.cnki.net/urlid/11.2625.V.20240612.0919.001

Foundation items：National Natural Science Foundation of China (62131001,62171029)；“Leading Goose”R&D Program of Zhejiang (2023C01148)

* **Corresponding author**. E-mail：young_being@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0178

基于多种微结构的机翼结构/材料多尺度气动弹性优化

黎珂宇[1,2]，杨超[1]，王晓喆[3,*]，万志强[1]，李畅[1]

(1. 北京航空航天大学 航空科学与工程学院，北京 100191；2. 中国运载火箭技术研究院，北京 100076；
3. 北京航空航天大学 无人系统研究院，北京 100191)

摘 要：微结构材料具有质量轻、功能多的特点，在航空航天领域具有广阔的应用前景。然而，单种微结构材料的刚度特性无法适应飞机进一步的减重要求。基于此，针对大展弦比机翼的宏/微结构构型，提出一种基于多种微结构构型的机翼结构和材料的多尺度气动弹性优化方法。由均匀化方法等效微结构材料的力学特性，结合宏观结构单元密度，计算整体刚度，通过求解气动弹性平衡方程获得弹性气动力及位移，基于敏度信息实现机翼结构和材料的协同设计。优化结果表明：在相同载荷下，基于多种微结构材料的整体结构柔顺度较单种微结构材料减小了18.6%。与刚性气动力下的优化结果相比，机翼弹性效应对微结构拓扑构型影响较小，而宏观结构的外段被加强。所提方法可有效优化复杂边界条件下的机翼结构，实现机翼内合理的刚度分布。

关 键 词：微结构材料；大展弦比机翼；多种微结构构型；多尺度气动弹性优化；机翼弹性效应
中图分类号：V224
文献标志码：A **文章编号**：1001-5965(2026)05-1635-12

轻量化是飞行器结构设计的重要目标。随着材料科学的进步，新型材料在飞行器设计中被广泛使用。微结构材料是在微观上采用特定的结构形式，经过周期性排列组成的材料[1]，这种材料具有高比刚度和比强度特性，除此之外，其在散热、减震等方面也具有优秀的特性[2-3]。

微结构材料的特性与航空领域对质量的严苛要求一致，因此，众多学者开展了其在飞行器上应用的相关研究。Jenett等[4]采用微结构材料设计制造了轻质柔性机翼，其中，沿展向布置立方体微结构材料，沿弦向布置刚度更小的Kelvin微结构材料，通过不同构型的微结构布置，实现了机翼局部的密度和刚度设计。Cramer等[5]基于聚醚酰亚胺材料，采用立方体单元，装配形成柔性飞行器的内部结构，该飞行器质量密度甚至低于仿鸟类飞行器，且通过微结构单元的分布，能够实现结构的刚度设计，在小迎角工况下，使后掠机翼产生正扭，提升飞行器的气动升力；其还采用相同构型的微结构填充机翼整个空间，进而设计了商用客机的主承力结构，并针对该机翼开展刚度分析，结果发现，相同翼尖位移即相同弯曲刚度下，该机翼略重于传统布局机翼，这也说明了单种微结构不利于飞行器的刚度设计[6]。Wang等[7]通过变密度法设计了高超声速飞行器的舵面结构，舵面采用实体-微结构混合材料填充，相同质量下，微结构材料舵面的刚度优于实体结构。

随着计算技术的进步，拓扑优化方法在飞机结构设计中得到广泛应用[8]。拓扑优化通过在参考域中寻求最优的材料布局，以满足约束条件，并使目标函数最小化。目前的拓扑优化方法主要有水平集法[9-10]、渐进结构优化法[11-12]和固体各向同性材料惩罚法[13-14]。通过微观尺度上设计微结构构型以优

收稿日期：2024-03-27；录用日期：2024-06-14；网络出版时间：2024-06-21 16:13
网络出版地址：link.cnki.net/urlid/11.2625.V.20240621.1146.001
基金项目：航空科学基金 (2022Z012051001)
*通信作者. E-mail：wangxiaozhemvp@buaa.edu.cn

引用格式：黎珂宇，杨超，王晓喆，等. 基于多种微结构的机翼结构/材料多尺度气动弹性优化[J]. 北京航空航天大学学报，2026，52(5)：1635-1646.
LI K Y，YANG C，WANG X Z，et al. Aeroelastic optimization of wing structure and material using multiple microstructures[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1635-1646 (in Chinese).

化微结构材料的性能，宏观尺度上设计宏观结构构型以获得最优的传力路线，能够达到具有优越特性的整体结构。Rodrigues 等[15]提出同时考虑“结构”和“材料”的结构性能优化，宏观和微结构通过等效弹性模量矩阵联系。通过变密度法进行宏观结构设计，根据各单元密度优化微结构材料的拓扑构型。Gao 等[16]研究了二维和三维结构/材料的多尺度拓扑优化问题，以最小柔顺度为目标，通过变密度法实现了单种微结构和宏观结构拓扑构型的并行设计。Wang 等[17]采用 2 种设计变量描述多尺度问题，包括宏观拓扑变量和微结构参数变量，其中，微结构刚度特性通过插值方法获取，采用非均匀分布的微结构材料设计的整体结构刚度优于单种微结构材料。Zhang 等[18]对宏观和微结构材料采用解耦式设计，变密度法得到的宏观结构单元密度作为微结构优化的体积约束，采用参数化水平集方法优化出多种初始微结构构型，在此基础上，引入代理模型，实现了微结构拓扑的连续渐变，保证了整体结构的连通性。近年来，虽然结构/材料多尺度优化取得了较大进展，但关于飞行器的多尺度刚度设计相关研究较少。Li 等[19]针对大展弦比机翼开展了多尺度气动弹性优化，在刚性气动力和弹性气动力下设计了宏/微结构构型，研究结果发现，单种微结构材料的力学特性无法适应弹性飞机的进一步减重要求。

针对以上问题，本文提出一种面向机翼结构和多种微结构的多尺度气动弹性优化方法，将气动弹性分析模块嵌入到结构/材料多尺度优化框架，以适用于飞行载荷下的双尺度优化，研究宏观和微结构拓扑对机翼刚度设计的影响。其中，宏观结构和微结构材料构型采用参数化水平集法描述，从而根据形状导数高效求解敏度信息。引入多个水平集描述微结构构型，通过均匀化方法计算微结构材料的等效力学特性，进而提高优化的效率。除此之外，采用线性势流理论计算刚性气动力，采用静气动弹性响应方程计算弹性气动力，将刚/弹性气动力作为优化问题的边界条件，研究结构弹性效应对优化问题的影响。

1 结构/材料多尺度优化

1.1 微结构材料均匀化方法

为在线弹性范围内表征微结构材料中的弹性属性，需引入 2 套坐标系，分别描述宏观和微结构 2 个尺度的特性[20]。变量在宏观坐标系 $\boldsymbol{x}$ (具体方向为 x_1、y_1、z_1) 内缓慢变化，而在微观坐标系 $\boldsymbol{y}$ (具体方向为 x_2、y_2、z_2) 内快速振荡。因此，宏观结构内的位移场 $\boldsymbol{u}^{\varepsilon}$ 可由双尺度渐进理论展开，表示为

$$\boldsymbol{u}^{\mu}(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{u}_0(\boldsymbol{x},\boldsymbol{y})+\mu\boldsymbol{u}_1(\boldsymbol{x},\boldsymbol{y})+\mu^2\boldsymbol{u}_2(\boldsymbol{x},\boldsymbol{y})+\cdots \tag{1}$$

式中：μ 为变换参数；$\boldsymbol{u}_0$、$\boldsymbol{u}_1$、$\boldsymbol{u}_2$ 为各阶的位移场。

根据虚功原理，初始应变场 $\boldsymbol{\varepsilon}^0$ 下微结构材料的等效弹性矩阵 $\boldsymbol{E}^{\mathrm{H}}$ 的张量可表示为

$$\boldsymbol{E}^{\mathrm{H}}_{ijkl}=1/|\Omega|\int_{\Omega}\boldsymbol{E}_{pqrs}(\boldsymbol{\varepsilon}^{0(ij)}_{pq}-\boldsymbol{\varepsilon}^{*(ij)}_{pq})(\boldsymbol{\varepsilon}^{0(kl)}_{rs}-\boldsymbol{\varepsilon}^{*(kl)}_{rs})\mathrm{d}\Omega \tag{2}$$

式中：Ω 为微结构的面积或体积；$\boldsymbol{E}_{pqrs}$ 为原材料的弹性张量；$\boldsymbol{\varepsilon}^{0(ij)}_{pq}$、$\boldsymbol{\varepsilon}^{0(kl)}_{rs}$ 为初始给定的单位测试应变场；$\boldsymbol{\varepsilon}^{*(ij)}_{pq}$、$\boldsymbol{\varepsilon}^{*(kl)}_{rs}$ 为微结构响应应变；上下角标 i、j、k、l、p、q、r、s 为索引值。

微结构响应应变通过弹性平衡方程求解：

$$\frac{1}{|\Omega|}\int_{\Omega}\boldsymbol{E}_{ijpq}\boldsymbol{\varepsilon}^{*(kl)}_{pq}\frac{\partial\boldsymbol{v}_i}{\partial\boldsymbol{y}_j}\mathrm{d}\Omega=\frac{1}{|\Omega|}\int_{\Omega}\boldsymbol{E}_{ijpq}\boldsymbol{\varepsilon}^{0(kl)}_{pq}\frac{\partial\boldsymbol{v}_i}{\partial\boldsymbol{y}_j}\mathrm{d}\Omega \tag{3}$$

式中：$\boldsymbol{v}_i$ 为虚位移。

根据式(3)，在微结构上施加初始单位测试应变场，并在满足周期性条件下求解方程[21]。

1.2 参数化水平集

水平集法是一种隐式描述结构边界的方法[22]，基于水平集函数的结构拓扑描述方法表述如下：

$$\begin{cases}\Phi(\boldsymbol{x})>0 & \forall\boldsymbol{x}\in\Omega\backslash\Gamma\\ \Phi(\boldsymbol{x})=0 & \forall\boldsymbol{x}\in\Gamma\cap D\\ \Phi(\boldsymbol{x})<0 & \forall\boldsymbol{x}\in D\backslash\Omega\end{cases} \tag{4}$$

式中：Φ 为水平集函数；D 为固定参考域；Γ 为结构拓扑边界。

结构边界的变化通过 Hamilton-Jacobi 偏微分方程描述，表示为

$$\frac{\partial\Phi}{\partial t}-V|\nabla\Phi|=0 \tag{5}$$

式中：t 为伪时间；V 为结构边界的法向速度。

因此，通过求解式(5)，可以确定形状边界移动的方向和距离，然而，偏微分方程的解析解很难求取，限制了水平集法的应用。为解决偏微分方程求解过程中的数值困难问题，通过插值函数提高求解精度和有效性。本文采用 Wendland[23]提出的紧支径向基函数，表示为

$$\phi_i(\boldsymbol{x})=(1-r)^4(4r+1) \tag{6}$$

式中：ϕ_i 为径向基函数；r 为支撑半径，定义为

$$r=\frac{d_I}{d_{mI}}=\frac{\sqrt{|\boldsymbol{x}-\boldsymbol{x}_i|^2}}{d_{mI}} \tag{7}$$

其中：d_I 为整个参考域内样本点 $\boldsymbol{x}$ 与参考点 $\boldsymbol{x}_i$ 间的欧氏距离；d_{mI} 为影响域大小。

通过 N 个径向基函数对水平集函数进行插值获得参数化水平集[24]，表示为

$$\Phi(\boldsymbol{x},t)=\sum_{i=1}^{N}\alpha_i(t)\phi_i(\boldsymbol{x}) \tag{8}$$

式中: α_i为扩展系数。

因此, 参数化水平集 (式 (8)) 可以采用矩阵形式表示为

$$\Phi(\boldsymbol{x},t)=\boldsymbol{\chi}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{\alpha} \tag{9}$$

$$\boldsymbol{\chi}(\boldsymbol{x})=\left[\begin{array}{cccc}\phi_1(\boldsymbol{x}) & \phi_2(\boldsymbol{x}) & \cdots & \phi_N(\boldsymbol{x})\end{array}\right]^{\mathrm{T}} \tag{10}$$

$$\boldsymbol{\alpha}(t)=\left[\begin{array}{cccc}\alpha_1(t) & \alpha_2(t) & \cdots & \alpha_N(t)\end{array}\right]^{\mathrm{T}} \tag{11}$$

将参数化水平集代入拓扑控制方程中, 获取基于径向基函数的 Hamilton-Jacobi 方程, 表示为

$$\boldsymbol{\chi}^{\mathrm{T}}(\boldsymbol{x})\frac{\mathrm{d}\boldsymbol{\alpha}(t)}{\mathrm{d}t}-V\left|(\nabla\boldsymbol{\chi}(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{\alpha}(t)\right|=0 \tag{12}$$

2　机翼结构/材料多尺度气动弹性优化

2.1　结构/材料多尺度气动弹性优化方法

基于气动力建立针对机翼的结构/材料多尺度优化框架, 体现了气动/结构多学科设计优化方法, 如图 1 所示。其中, 宏微观拓扑通过参数化水平集描述, 微结构材料的力学特性通过均匀化方法等效。虚线框内为均匀化方法的实现流程, 根据微观水平集函数获得微结构的刚度矩阵, 计算等效弹性矩阵。结合该矩阵和宏观水平集函数计算出整体结构的刚度矩阵。刚性气动力通过线性势流理论计算, 弹性气动力结果和弹性变形结果通过气动力和结构耦合分析得到。根据气动弹性分析结果计算得到目标函数及约束函数数值, 并获得目标函数和约束函数对宏微观设计变量的敏度值。通过最优准则方法更新宏微观的设计变量。优化直至达到收敛性条件为止。气动弹性优化中, 由于气动力会引起机翼结构弹性变形, 导致气动力重新分布, 伴随机翼刚度的降低, 也会导致弹性效应更加明显。因此, 与刚性气动力结果相比, 考虑弹性效应的气动力会导致优化问题中的边界条件不断变化。

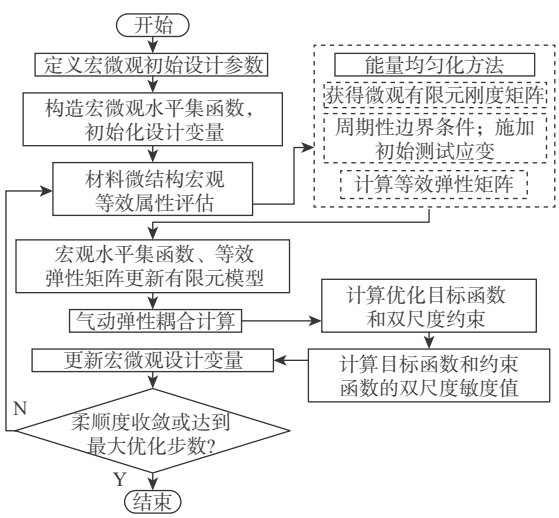


图 1　多尺度气动弹性优化框架

Fig. 1　Framework of multiscale aeroelastic optimization

2.2　优化模型

结构/材料多尺度气动弹性优化按尺度分为 2 个问题, 微观尺度用于设计微结构的拓扑构型。不同的微结构构型有不同的等效弹性矩阵, 从而对整体刚度产生影响。宏观尺度用于整体结构的拓扑设计, 类似于传统的结构拓扑优化, 在宏观参考域内设计更有效的结构。多尺度优化问题同时对宏观和微结构的拓扑构型进行迭代, 从而获取最优的目标函数。多尺度优化有必要定义多个水平集函数来描述这些结构的拓扑结构。水平集函数表示如下:

$$\begin{cases}\Phi_{\mathrm{M}}(\boldsymbol{x})>0 & \forall\boldsymbol{x}\in\Omega_{\mathrm{M}}\backslash\Gamma_{\mathrm{M}}\\ \Phi_{\mathrm{M}}(\boldsymbol{x})=0 & \forall\boldsymbol{x}\in\Gamma_{\mathrm{M}}\\ \Phi_{\mathrm{M}}(\boldsymbol{x})<0 & \forall\boldsymbol{x}\in D_{\mathrm{M}}\backslash\Omega_{\mathrm{M}}\\ \Phi_{\mathrm{m}}^{\lambda}(\boldsymbol{y})>0 & \forall\boldsymbol{y}\in\Omega_{\mathrm{m}}^{\lambda}\backslash\Gamma_{\mathrm{m}}^{\lambda}\\ \Phi_{\mathrm{m}}^{\lambda}(\boldsymbol{y})=0 & \forall\boldsymbol{y}\in\Gamma_{\mathrm{m}}^{\lambda}\\ \Phi_{\mathrm{m}}^{\lambda}(\boldsymbol{y})<0 & \forall\boldsymbol{y}\in D_{\mathrm{m}}^{\lambda}\backslash\Omega_{\mathrm{m}}^{\lambda}\end{cases} \tag{13}$$

式中: Φ_{M}为宏观结构的水平集函数; $\Phi_{\mathrm{m}}^{\lambda}$为微结构的水平集函数; Ω_{M}、Γ_{M}和D_{M}分别为宏观结构的设计域、结构边界和参考域; $\Omega_{\mathrm{m}}^{\lambda}$、$\Gamma_{\mathrm{m}}^{\lambda}$和$D_{\mathrm{m}}^{\lambda}$分别为第$\lambda$个微结构的设计域、结构边界和参考域。

因此, 通过径向基函数对宏观和微结构的水平集函数插值, 可以获得相应的参数化水平集函数:

$$\begin{cases}\Phi_{\mathrm{M}}(\boldsymbol{x},t)=\boldsymbol{\chi}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{\alpha}_{\mathrm{M}}(t)\\ \Phi_{\mathrm{m}}^{\lambda}(\boldsymbol{y},t)=\boldsymbol{\chi}_{\lambda}^{\mathrm{T}}(\boldsymbol{y})\boldsymbol{\alpha}_{\mathrm{m}}^{\lambda}(t)\end{cases} \tag{14}$$

式中: $\boldsymbol{\chi}(\boldsymbol{x})$和$\boldsymbol{\alpha}_{\mathrm{M}}(t)$分别为宏观结构的径向基函数矩阵及扩展系数; $\boldsymbol{\chi}_{\lambda}(\boldsymbol{y})$和$\boldsymbol{\alpha}_{\mathrm{m}}^{\lambda}(t)$分别为第$\lambda$个微结构的径向基函数矩阵及扩展系数。

拓扑优化问题可以简单概括为: 当满足体积约束时, 在给定载荷下获得最小柔顺度的结构。基于参数化水平集的结构/材料多尺度拓扑优化问题可以表述为如下数学模型:

$$\begin{cases}\text{Minimize: } J(\boldsymbol{u},\Phi)=\\ \quad\displaystyle\sum_{\lambda=1}^{K}\int_{D_{\mathrm{M}}}\boldsymbol{\varepsilon}(\boldsymbol{u}_{\mathrm{M}})\boldsymbol{E}_{\lambda}^{\mathrm{H}}(\boldsymbol{u}_{\mathrm{m}}^{\lambda},\Phi_{\mathrm{m}}^{\lambda})\boldsymbol{\varepsilon}(\boldsymbol{u}_{\mathrm{M}})H(\Phi_{\mathrm{M}})\mathrm{d}\Omega_{\mathrm{M}}\\ \text{s.t.}\begin{cases}a(\Phi_{\mathrm{M}},\boldsymbol{u}_{\mathrm{M}},\boldsymbol{v}_{\mathrm{M}},\boldsymbol{E}_{\lambda}^{\mathrm{H}})=l(\Phi_{\mathrm{M}},\boldsymbol{v}_{\mathrm{M}})\\ a(\Phi_{\mathrm{m}}^{\lambda},\boldsymbol{u}_{\mathrm{m}}^{\lambda},\boldsymbol{v}_{\mathrm{m}}^{\lambda})=l(\Phi_{\mathrm{m}}^{\lambda},\boldsymbol{v}_{\mathrm{m}}^{\lambda})\\ G_{\mathrm{M}}(\Phi_{\mathrm{M}})=\displaystyle\sum_{\lambda=1}^{K}\left\{\int_{D_{\mathrm{M}}}H(\Phi_{\mathrm{M}})\mathrm{d}\Omega_{\mathrm{M}}\cdot\right.\\ \qquad\left.\displaystyle\int_{D_{\mathrm{m}}^{\lambda}}H(\Phi_{\mathrm{m}}^{\lambda})\mathrm{d}\Omega_{\mathrm{m}}\right\}\Big/M_0-M_{\mathrm{M}}\leqslant 0\\ G_{\mathrm{m}}^{\lambda}(\Phi_{\mathrm{m}}^{\lambda})=\displaystyle\int_{D_{\mathrm{m}}^{\lambda}}H(\Phi_{\mathrm{m}}^{\lambda})\mathrm{d}\Omega_{\mathrm{m}}/V_{\mathrm{m},0}^{\lambda}-V_{\mathrm{m}}^{\lambda}\leqslant 0\\ \alpha_{\mathrm{M}}^{\min}\leqslant\alpha_{\mathrm{M}}^{i}\leqslant\alpha_{\mathrm{M}}^{\max}\\ \alpha_{\mathrm{m}}^{\lambda,\min}\leqslant\alpha_{\mathrm{m}}^{\lambda,j}\leqslant\alpha_{\mathrm{m}}^{\lambda,\max}\end{cases}\end{cases} \tag{15}$$

式中: J为柔顺度, 为优化问题的目标函数; Φ为水

平集函数；α_{M}^{i}为第 i 个宏观设计变量；$\alpha_{\mathrm{m}}^{\lambda,j}$为第 λ 种微结构的第 j 个设计变量；上标 min 和 max 分别表示设计变量的最小值和最大值；$\boldsymbol{u}$ 为位移向量；$\boldsymbol{\varepsilon}$为应变张量；a 为基于能量的双线性形式；l 为载荷的线性形式；$\boldsymbol{u}_{\mathrm{M}}$和$\boldsymbol{u}_{\mathrm{m}}^{\lambda}$分别为宏观结构和第 λ 种微结构的位移向量；$\boldsymbol{v}_{\mathrm{M}}$和$\boldsymbol{v}_{\mathrm{m}}^{\lambda}$分别为宏观结构和第 λ 种微结构的虚位移向量；G_{M}和G_{m}^{λ}分别为宏观结构和第 λ 种微结构的体积约束；M_0为采用全填充的实体材料时，整体结构的质量；M_{M}为整体结构的最小质量占比，由宏观结构及微结构的体积决定；$V_{\mathrm{m},0}^{\lambda}$为第 λ 种微结构的参考域体积；V_{m}^{λ}为微结构最小体积占比；$\boldsymbol{E}_{\lambda}^{\mathrm{H}}$为第 λ 种微结构的等效弹性矩阵；$H(\Phi_{\mathrm{M}})$和$H(\Phi_{\mathrm{m}}^{\lambda})$为 Heaviside 函数，表征材料的存在性。

基于能量的双线性形式$a(\Phi_{\mathrm{M}},\boldsymbol{u}_{\mathrm{M}},\boldsymbol{v}_{\mathrm{M}},\boldsymbol{E}_{\lambda}^{\mathrm{H}})$和载荷的线性形式$l(\Phi_{\mathrm{M}},\boldsymbol{v}_{\mathrm{M}})$，可以将弹性平衡方程写为弱解形式，表示为

$$\begin{cases} a(\Phi_{\mathrm{M}},\boldsymbol{u}_{\mathrm{M}},\boldsymbol{v}_{\mathrm{M}},\boldsymbol{E}_{\lambda}^{\mathrm{H}}) = \\ \quad \sum_{\lambda=1}^{K}\int_{D_{\mathrm{M}}}\boldsymbol{\varepsilon}(\boldsymbol{u}_{\mathrm{M}})\boldsymbol{E}_{\lambda}^{\mathrm{H}}(\boldsymbol{u}_{\mathrm{m}}^{\lambda},\Phi_{\mathrm{m}}^{\lambda})\boldsymbol{\varepsilon}(\boldsymbol{v}_{\mathrm{M}})H(\Phi_{\mathrm{M}})\mathrm{d}\Omega_{\mathrm{M}} \\ l(\Phi_{\mathrm{M}},\boldsymbol{v}_{\mathrm{M}}) = \int_{D_{\mathrm{M}}}\boldsymbol{f}_{\mathrm{M}}\boldsymbol{v}_{\mathrm{M}}H(\Phi_{\mathrm{M}})\mathrm{d}\Omega_{\mathrm{M}} + \\ \quad \int_{D_{\mathrm{M}}}\boldsymbol{p}_{\mathrm{M}}\boldsymbol{v}_{\mathrm{M}}\delta(\Phi_{\mathrm{M}})|\nabla\Phi_{\mathrm{M}}|\mathrm{d}\Omega_{\mathrm{M}} \end{cases} \tag{16}$$

式中：$\boldsymbol{f}_{\mathrm{M}}$和$\boldsymbol{p}_{\mathrm{M}}$分别为体力和面力，气动弹性优化中体力为惯性力，面力为气动力；$\delta(\Phi_{\mathrm{M}})$为 Dirac 函数。

根据形状导数获得目标函数及宏观体积约束对宏观扩展系数的导数，表示为

$$\begin{cases} \dfrac{\partial J}{\partial \boldsymbol{\alpha}_{\mathrm{M}}} = \displaystyle\int_{D_{\mathrm{M}}}\left(-\sum_{\lambda=1}^{K}\boldsymbol{\varepsilon}(\boldsymbol{u}_{\mathrm{M}})\boldsymbol{E}_{\lambda}^{\mathrm{H}}\boldsymbol{\varepsilon}(\boldsymbol{u}_{\mathrm{M}})\boldsymbol{\chi}^{\mathrm{T}}(\boldsymbol{x})\delta(\Phi_{\mathrm{M}})\right)\mathrm{d}\Omega_{\mathrm{M}} \\ \dfrac{\partial G_{\mathrm{M}}}{\partial \boldsymbol{\alpha}_{\mathrm{M}}} = \dfrac{1}{M_0}\displaystyle\sum_{\lambda=1}^{K}\left(\int_{D_{\mathrm{M}}}\boldsymbol{\chi}^{\mathrm{T}}(\boldsymbol{x})\delta(\Phi_{\mathrm{M}})\mathrm{d}\Omega_{\mathrm{M}}\int_{D_{\mathrm{m}}^{\lambda}}H(\Phi_{\mathrm{m}}^{\lambda})\mathrm{d}\Omega_{\mathrm{m}}\right) \end{cases} \tag{17}$$

同理，计算目标函数及微结构体积分数对微结构扩展系数的导数，表示为

$$\begin{cases} \dfrac{\partial J}{\partial \boldsymbol{\alpha}_{\mathrm{m}}^{\lambda}} = \displaystyle\int_{D_{\mathrm{M}}}\boldsymbol{\varepsilon}(\boldsymbol{u}_{\mathrm{M}})\left[-\frac{1}{|\Omega_{\mathrm{m}}^{\lambda}|}\int_{D_{\mathrm{m}}^{\lambda}}\boldsymbol{E}_{pqrs}(\boldsymbol{\varepsilon}_{pq}^{0(ij)} - \right. \\ \qquad \left.\boldsymbol{\varepsilon}_{pq}(\boldsymbol{u}_{\mathrm{m}}^{ij}))(\boldsymbol{\varepsilon}_{rs}^{0(kl)} - \boldsymbol{\varepsilon}_{rs}(\boldsymbol{u}_{\mathrm{m}}^{kl}))\boldsymbol{\chi}_{\lambda}^{\mathrm{T}}(\boldsymbol{y})\delta(\Phi_{\mathrm{m}}^{\lambda})\mathrm{d}\Omega_{\mathrm{m}}^{\lambda}\right]\cdot \\ \qquad \boldsymbol{\varepsilon}(\boldsymbol{u}_{\mathrm{M}})H(\Phi_{\mathrm{M}})\mathrm{d}\Omega_{\mathrm{M}} \\ \dfrac{\partial G_{\mathrm{m}}}{\partial \boldsymbol{\alpha}_{\mathrm{m}}^{\lambda}} = \dfrac{1}{V_{\mathrm{m},0}^{\lambda}}\displaystyle\int_{D_{\mathrm{m}}^{\lambda}}\boldsymbol{\chi}_{\lambda}^{\mathrm{T}}(\boldsymbol{y})\delta(\Phi_{\mathrm{m}}^{\lambda})\mathrm{d}\Omega_{\mathrm{m}}^{\lambda} \end{cases} \tag{18}$$

根据敏度信息，采用最优准则方法[25]对多尺度设计中的设计变量进行更新。宏观扩展系数的更新方案如下：

$$\alpha_{\mathrm{M}}^{i(k+1)} = \begin{cases} \min(\alpha_{\mathrm{M}}^{i(k)}+\chi_{\mathrm{M}},\alpha_{\mathrm{M}}^{\max})\min(\alpha_{\mathrm{M}}^{i(k)}+\chi_{\mathrm{M}},\alpha_{\mathrm{M}}^{\max}) \leqslant (K_{\mathrm{M}}^{i(k)})^{\zeta_{\mathrm{M}}}\alpha_{\mathrm{M}}^{i(k)} \\ (K_{\mathrm{M}}^{i(k)})^{\zeta_{\mathrm{M}}}\alpha_{\mathrm{M}}^{i(k)}\max(\alpha_{\mathrm{M}}^{i(k)}-\chi_{\mathrm{M}},\alpha_{\mathrm{M}}^{\min}) < (K_{\mathrm{M}}^{i(k)})^{\zeta_{\mathrm{M}}}\alpha_{\mathrm{M}}^{i(k)} < \min(\alpha_{\mathrm{M}}^{i(k)}+\chi_{\mathrm{M}},\alpha_{\mathrm{M}}^{\max}) \\ \max(\alpha_{\mathrm{M}}^{i(k)}-\chi_{\mathrm{M}},\alpha_{\mathrm{M}}^{\min})(K_{\mathrm{M}}^{i(k)})^{\zeta_{\mathrm{M}}}\alpha_{\mathrm{M}}^{i(k)} \leqslant \max(\alpha_{\mathrm{M}}^{i(k)}-\chi_{\mathrm{M}},\alpha_{\mathrm{M}}^{\min}) \end{cases} \tag{19}$$

其中：$\alpha_{\mathrm{M}}^{i(k+1)}$为宏观结构水平集在优化第 k+1 步时，宏观结构的第 i 个设计变量；χ_{M}为宏观设计变量在每一步优化中的步长限制；ζ_{M}为阻尼因子；$K_{\mathrm{M}}^{i(k)}$为优化在第 k 步时，宏观结构第 i 个设计变量的更新因子，表示为

$$K_{\mathrm{M}}^{i(k)} = -\frac{\partial J}{\partial \alpha_{\mathrm{M}}^{i(k)}}\Big/\max\left(\mu,\Lambda_{\mathrm{M}}^{(k)}\frac{\partial G_{\mathrm{M}}}{\partial \alpha_{\mathrm{M}}^{i(k)}}\right) \tag{20}$$

其中：μ为极小值，从而避免更新因子的分母为 0；$\Lambda_{\mathrm{M}}^{(k)}$为宏观结构在第 k 个优化步的拉格朗日系数。

同理，微结构设计变量的优化更新类似。

2.3 气动力计算及静气动弹性响应分析

气动力计算基于线化位势流理论，形式为

$$(1-Ma_{\infty}^{2})\frac{\partial^2\varphi}{\partial\xi^2}+\frac{\partial^2\varphi}{\partial\eta^2}+\frac{\partial^2\varphi}{\partial\zeta^2}=0 \tag{21}$$

式中：Ma_{∞}为马赫数；φ为速度势；(ξ,η,ζ)为描述流场空间位置和流动方向的坐标系。

气动弹性响应方程表示为

$$(\boldsymbol{K}-\bar{q}\boldsymbol{Q}_a)\boldsymbol{u}_{\mathrm{M}} = \bar{q}\boldsymbol{Q}_x\boldsymbol{u}_x+\boldsymbol{F}_{\mathrm{I}}+\boldsymbol{P} \tag{22}$$

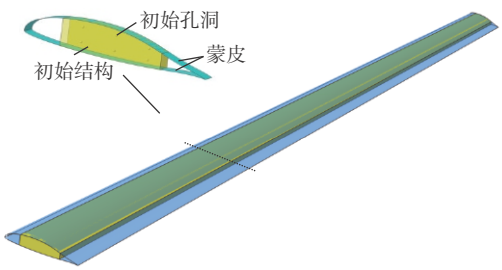
式中：$\boldsymbol{K}$为刚度矩阵，由宏微结构决定；$\bar{q}$为动压；$\bar{q}\boldsymbol{Q}_a\boldsymbol{u}_{\mathrm{M}}$为结构弹性变形引起的气动力增量；$\bar{q}\boldsymbol{Q}_x\boldsymbol{u}_x$为由配平参数(迎角、舵偏角等)引起的气动力，$\boldsymbol{u}_x$为配平参数 (迎角、舵偏角)；$\boldsymbol{F}_{\mathrm{I}}$为刚体运动引起的惯性力；$\boldsymbol{P}$为初始气动模型引起的气动力。

3 大展弦比机翼结构/材料多尺度气动弹性优化

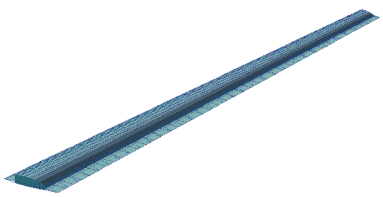
3.1 模型描述

选取大展弦比机翼作为研究对象，模型参考全球鹰无人机，其几何参数及优化中的初始模型如图 2 所示。其中，机翼翼型为 LRN1015，翼根弦长为 1.52 m，前缘后掠角为 5.9°。机翼主要包含蒙皮(蓝色)和翼盒(绿色)。蒙皮为复合材料，不作为优化变量。翼盒部分为宏观结构参考域，在初始模型

的翼盒中, 包含 384 个半径为 5 mm 的孔洞, 初始宏观结构体积为 97.2%。初始微结构完全填充, 体积为 100%。宏观结构拓扑的径向基函数构造中 d_{mI} = 100 mm, 微结构 d_{mI} = 2.5 mm。为方便通过增材制造技术对微结构材料进行加工, 翼盒原材料选用钛合金, 弹性模量为 1.09×10^5 MPa, 泊松比为 0.31。翼盒通过 3 600 个六面体单元进行离散。而微结构材料的参考域为尺寸为 20 mm×20 mm ×20 mm 的正方体结构, 采用 1 mm×1 mm×1 mm 的八节点六面体单元进行离散。



(a) 机翼几何模型



(b) 机翼有限元模型

图 2　无人机机翼模型

Fig. 2　Unmanned aerial vehicle wing model

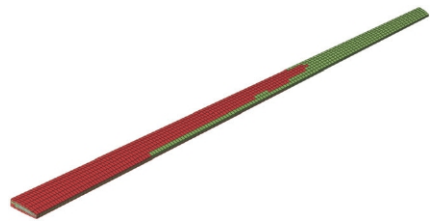
采用单种微结构及多种微结构材料填充机翼翼盒, 如图 3 所示。单种微结构材料优化中, 微结构只有一种构型, 形成的微结构材料填充到整个宏观拓扑中。多种微结构的材料分布共包含 2 个步骤: ①采用变密度法优化机翼结构, 获得单元密度分布, 如图 3(b) 所示, 绿色区域代表密度值为 [0,0.6] 的单元, 红色区域代表密度值为 [0.6,1.0] 的单元。能够看出, 高密度单元主要分布在机翼内段上下区域。②根据材料分布优化结果, 为方便加工, 对离散单元进行处理, 获得用于多尺度优化的模型。将机翼翼盒划分为 4 个区域, 其中, 沿展向和高度方向按照内段-靠里、内段-靠外、外段-靠里、外段-靠外分别填充不同的微结构材料, 内段为 0～9 000 mm(半翼展的 64.3% 以内), 外段为 9 000～14 000 mm, 高度方向距离蒙皮 1/4 剖面厚度以内为靠外段, 以外为靠里段, 如图 3(c) 所示。

3.2　基于单/多种微结构材料的机翼多尺度优化

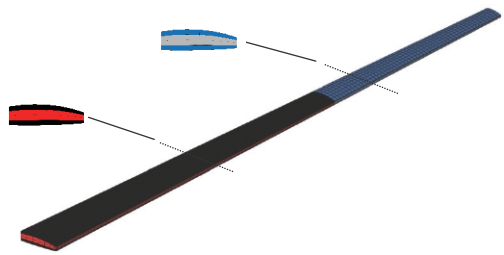
在机翼结构与材料的多尺度气动弹性优化中, 飞行工况高度为 15 000 m, 马赫数为 0.785, 动压为



(a) 单种微结构材料分布



(b) 多种微结构材料分布-变密度法



(c) 多种微结构材料分布-后处理

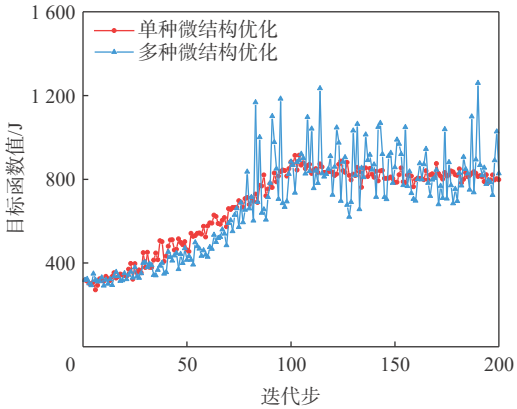
图 3　微结构材料分布

Fig. 3　Distribution of microstructured materials

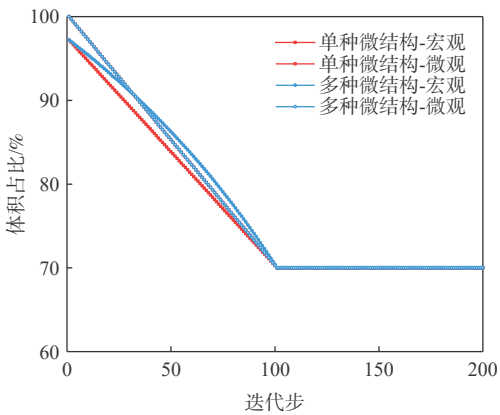
3 051 Pa, 飞行迎角为 4°。单种微结构材料优化中, 宏观结构体积约束为 70%, 微结构材料体积约束为 70%。若宏观及微结构用实体材料全填充的质量为 100%, 则优化后整体结构质量为总质量的 49%。多种微结构材料优化问题(式(15))中质量约束 M_{M}为 49%, 4 个区域的微结构材料体积约束 V_{m}^{λ}均为 70%, 从而对比引入多种微结构材料对多尺度优化结果的影响。采用低阶面元法计算刚性气动力, 并获得柔顺度。

2 种优化的目标函数和约束条件的收敛如图 4 所示。前 100 步结构由于宏观和微结构不断降低, 逼近体积约束, 柔顺度不断增大。而当宏微观体积达到体积约束后, 柔顺度逐渐降低。其中, 由于多种微结构材料优化在每次迭代中都存在 4 种微结构材料的弹性模量矩阵变化, 目标函数比单种微结构材料波动剧烈, 而多种微结构材料最优构型的柔顺度小于单种微结构材料。其中, 单种微结构柔顺度为 762 J, 而多种微结构柔顺度为 620 J, 减小了 18.6%。

单种和多种微结构材料优化得到的宏观最优构型如图 5 所示。在求解静气动弹性响应方程时, 宏观结构中的网格密度建立了 0.001 的阈值, 保证了整体刚度矩阵的非奇异性。而显示中并没有包含阈值密度。



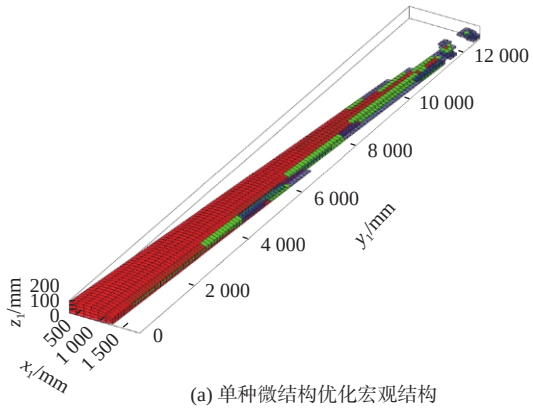
(a) 目标函数收敛



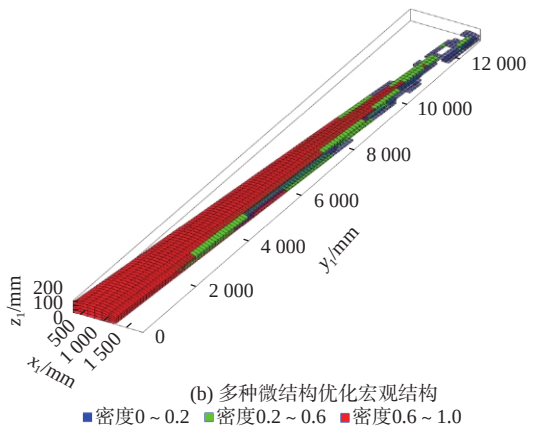
(b) 体积约束收敛

图 4 目标函数和体积约束收敛

Fig. 4 Convergence of the objective functions and volume constraints



(a) 单种微结构优化宏观结构



(b) 多种微结构优化宏观结构

密度0 ~ 0.2 密度0.2 ~ 0.6 密度0.6 ~ 1.0

图 5 优化宏观结构拓扑

Fig. 5 Macrostructural topology of optimization

优化的宏观结构拓扑显示机翼在 12 000 mm 外的单元密度较小，并未完全连接至翼尖。因此，为获得有利于工程实现的宏观结构拓扑，对优化后的机翼外段结构进行后处理。处理方式为选择翼展大于 11 500 mm 的宏观结构单元，将其密度平均到机翼外部的指定结构单元上，结果如图 6 所示。处理前后机翼结构在刚性气动力下的柔顺度如表 1 所示，单种及多种微结构的后处理结构相较于最优结构柔顺度分别增大了 11.2% 及 7.6%，验证了优化结果具有最优解，表明机翼在固定载荷下，翼尖段的刚度对承载影响相对较小。此外，还给出了宏观结构在 4 个展向位置的剖面，阴影越深，表示密度越大。最终设计的宏观拓扑具有相似的趋势，单元密度的变化趋势表现为 3 个特征：①从翼根到翼尖密度逐渐减小；②上下部分单元密度大于中间部分；③翼盒前端密度大于后端密度。其中，密度值 1.0 主要分布在翼根前端的上下 2 部分。除此之外，优化设计的宏观结构体积为 70%，因此，机翼内段拓扑呈现出类似于全填充的构型，随展向位置靠近翼尖位置，结构逐渐呈现出工字梁的构型。

优化最终设计的微结构拓扑如图 7 和图 8 所示。为保证整体刚度矩阵的非奇异性，设置等效弹性模量矩阵的下限为 200 MPa。单种微结构拓扑为正方体参考域内部沿 y_2 轴切割出椭圆柱形孔，其等效弹性模量矩阵为

$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix} 51\,971.6 & 21\,727.2 & 18\,116.2 & 0 & 0 & 0 \\ 21\,727.2 & 91\,994.0 & 28\,905.5 & 0 & 0 & 0 \\ 18\,116.2 & 28\,905.5 & 75\,127.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 18\,571.8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 25\,069.9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 12\,677.0 \end{bmatrix}$$

可看出，微结构材料在 y_2 方向具有较大的刚度，x_2y_2、y_2z_2 平面内的剪切模量远远大于 x_2z_2 平面。

多种微结构优化中的内段-靠外及外段-靠外的微结构拓扑也呈现出类似的构型，微结构材料

的等效弹性模量矩阵呈现出类似的刚度分布, 分别为

$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}59\,762.3 & 24\,588.1 & 19\,554.2 & 0 & 0 & 0\\ 24\,588.1 & 92\,660.1 & 28\,186.1 & 0 & 0 & 0\\ 19\,554.2 & 28\,186.1 & 71\,368.8 & 0 & 0 & 0\\ 0 & 0 & 0 & 20\,711.8 & 0 & 0\\ 0 & 0 & 0 & 0 & 23\,916.5 & 0\\ 0 & 0 & 0 & 0 & 0 & 13\,486.7\end{bmatrix}$$

$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}49\,688.5 & 20\,765.0 & 17\,295.5 & 0 & 0 & 0\\ 20\,765.0 & 91\,856.4 & 29\,416.2 & 0 & 0 & 0\\ 17\,295.5 & 29\,416.2 & 77\,595.7 & 0 & 0 & 0\\ 0 & 0 & 0 & 17\,751.0 & 0 & 0\\ 0 & 0 & 0 & 0 & 25\,648.5 & 0\\ 0 & 0 & 0 & 0 & 0 & 11\,549.6\end{bmatrix}$$

而内段-靠里及外段-靠里的微结构拓扑在 x_2 方向上没有材料联通, 因此, 微结构材料的等效弹性模量矩阵分别为

$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}0 & 0 & 0 & 0 & 0 & 0\\ 0 & 84\,313.2 & 25\,843.2 & 0 & 0 & 0\\ 0 & 25\,843.2 & 83\,365.0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 28\,911.4 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}0 & 0 & 0 & 0 & 0 & 0\\ 0 & 84\,324.9 & 26\,061.0 & 0 & 0 & 0\\ 0 & 26\,061.0 & 84\,118.6 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 29\,060.3 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

呈现出沿 x_2 方向的模量远远小于其他方向, y_2z_2 平面内的剪切模量最大。

不同展向位置剖面的微结构材料如图 9 所示。其中, 宏观拓扑决定了微结构的存在性, 宏观结构拓扑对应实体部分由周期性重复排列的微结构组成, 而孔洞部分没有微结构。

机翼的飞行载荷由气动力和惯性力组成, 沿展向分布如图 10 所示。弯矩和剪切力从翼尖向翼根增大, 翼根处机翼结构的载荷最大。

根据载荷分布可以看出, 机翼外段的载荷最小, 因此, 宏观结构的优化主要削弱了外段。而机翼结构主要承受剪切力及弯矩, 因此, 外段部分呈现出工字梁形式, 以承受由于弯矩产生的 y_2 方向拉压应力。微观尺度方面, 微结构材料普遍在 y_2 方向有较大的弹性模量, 主要是为承受弯曲载荷, 与宏观尺度结构优化趋势一致。除此之外, 微结构材料在展向方向的结构拓扑差异较小, 而在高度方向的结构拓扑有明显差异, 相较于靠外的微结构构型, 靠里微结构的等效弹性模量矩阵主要削弱了 x_2 方向、x_2y_2 及 x_2z_2 平面的刚度, 这与微结构材料所在宏观空间位置有关, 靠里的微结构在 z_2 方向离轴线较近, 因此, 扭转切应力较小, 而在 x_2 方向离轴线较

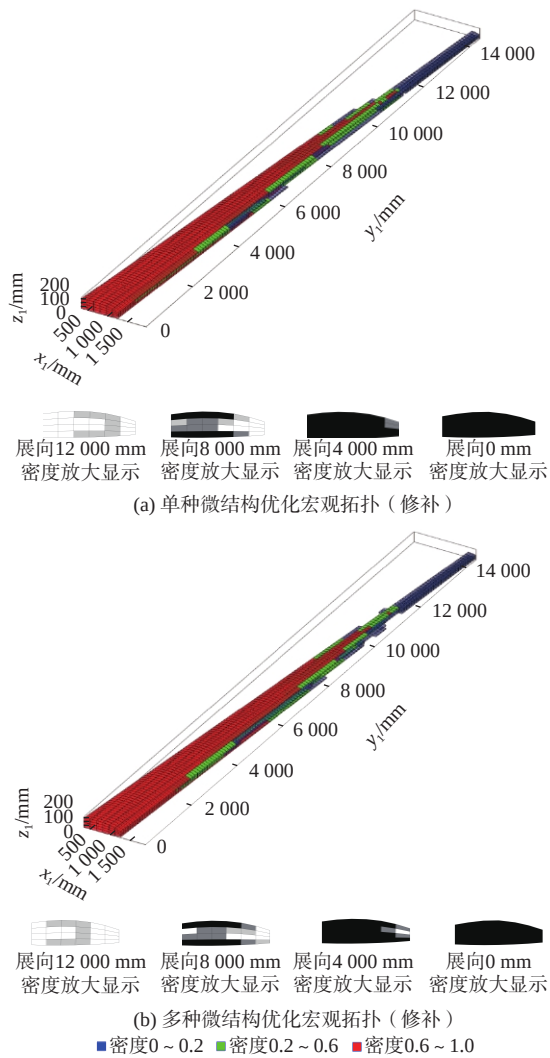


图 6　后处理的宏观结构拓扑

Fig. 6　Macrostructural topology of post-processing

表 1 优化及后处理结果

Table 1 Optimization and post-processing results

结构	单种微结构 柔顺度/J	多种微结构 柔顺度/J
最优结构	762	620
后处理结构	847	667

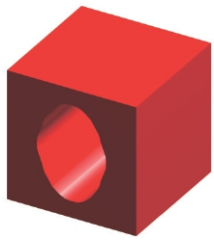


图 7 单种微结构优化拓扑

Fig. 7 Optimization topology for a single microstructure

远，需要承受 y_2z_2 平面上的扭转切应力，因此，优化通过加强 y_2z_2 平面上的材料分布提高剪切模量。

3.3 刚/弹性气动力下机翼结构/材料的多尺度优化

本节的飞行工况与 3.2 节一致。为削弱结构的刚度，增大弹性效应，多种微结构材料优化问题(式(15))中质量约束 M_{M} 减小到 20%，内段-靠外、内段-靠里、外段-靠外及外段-靠里的微结构材料体积约束 V_{m}^{λ} 分别为 60%、50%、40%、30%。其中，气动力分别为低阶面元法计算的刚性气动力及静气动弹性方程计算的弹性气动力。

刚性气动力下设计的结构柔顺度为 938 J，弹性气动力下柔顺度为 1 009 J。弹性柔顺度较刚性柔顺度增大了 7.6%。机翼的气动力沿展向分布如图 11 所示，其中，扭转为机翼 1/4 弦线的扭矩。刚性和弹性气动力显示出近似的分布趋势，剪切力、弯矩及扭矩从翼尖向翼根增大，翼根处机翼结构的载荷最大。而弹性气动力的剪切力和弯矩都大于刚性气动力，这主要是由于机翼在气动力作用下发生了正向扭转变形，会产生额外升力，导致气动力增大。

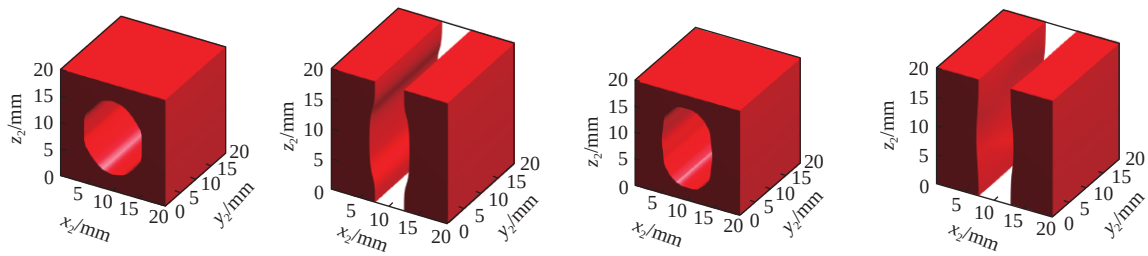
刚性和弹性气动力下优化设计的宏观拓扑如图 12 所示。刚性和弹性气动力下的结构拓扑呈现类似的趋势，随着展向位置靠近翼尖，结构逐渐减少。图 13 为刚/弹性气动力下宏观结构体积占比沿展向的分布。结果表明，在弹性气动力下，机翼外段会分配更多的结构。

优化后的微结构拓扑如图 14 和图 15 所示。图 14(a) 对应的弹性模量矩阵分别为

$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}0&0&0&0&0&0\\0&72\,221.4&22\,004.6&0&0&0\\0&22\,004.6&70\,982.4&0&0&0\\0&0&0&0&0&0\\0&0&0&0&24\,663.3&0\\0&0&0&0&0&0\end{bmatrix},$$

$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}0&0&0&0&0&0\\0&60\,170.5&18\,295.8&0&0&0\\0&18\,295.8&59\,018.7&0&0&0\\0&0&0&0&0&0\\0&0&0&0&20\,535.9&0\\0&0&0&0&0&0\end{bmatrix},$$

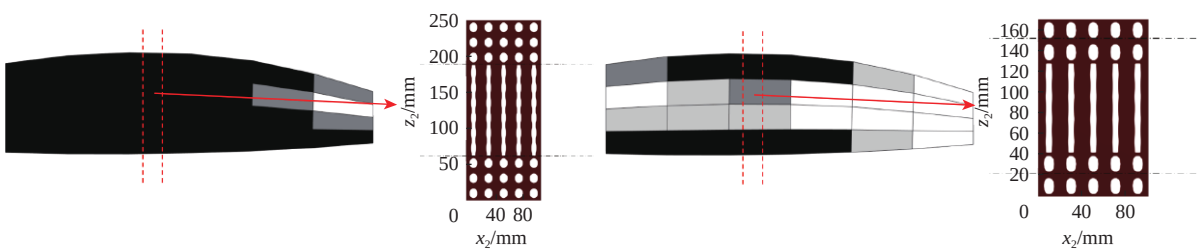
图 14(b) 对应的弹性模量矩阵分别为



(a) 内段-靠外 (b) 内段-靠里 (c) 外段-靠外 (d) 外段-靠里

图 8 多种微结构优化拓扑

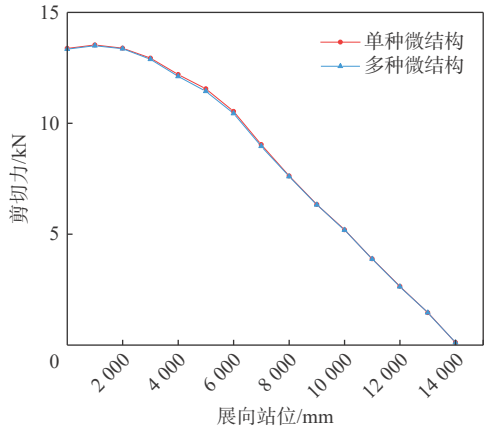
Fig. 8 Optimization topology for multiple microstructure



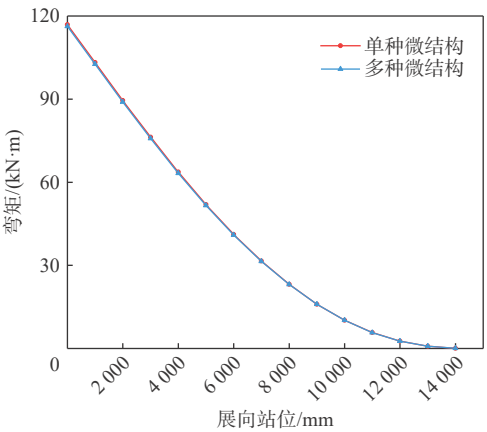
(a) 展向4 000 mm最优多尺度设计 (b) 展向9 000 mm最优多尺度设计

图 9 展向位置的最优多尺度设计

Fig. 9 Optimal multiscale design of span position



(a) 飞行载荷下剪切力分布



(b) 飞行载荷下弯矩分布

图 10　机翼的飞行载荷分布

Fig. 10　Flight load distribution of the wing

$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}0&0&0&0&0&0\\0&48\,223.9&14\,949.4&0&0&0\\0&14\,949.4&48\,223.8&0&0&0\\0&0&0&0&0&0\\0&0&0&0&16\,637.2&0\\0&0&0&0&0&0\end{bmatrix}$$

$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}0&0&0&0&0&0\\0&36\,176.6&11\,214.7&0&0&0\\0&11\,214.7&36\,176.6&0&0&0\\0&0&0&0&0&0\\0&0&0&0&12\,480.9&0\\0&0&0&0&0&0\end{bmatrix}$$

图 15(a) 对应的弹性模量矩阵分别为

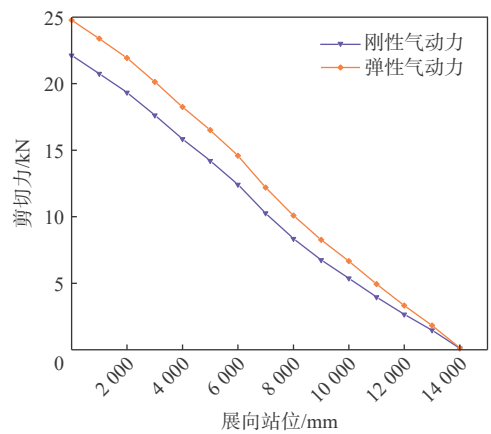
$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}0&0&0&0&0&0\\0&72\,220.7&22\,001.3&0&0&0\\0&22\,001.3&70\,971.9&0&0&0\\0&0&0&0&0&0\\0&0&0&0&24\,660.2&0\\0&0&0&0&0&0\end{bmatrix}$$

$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}0&0&0&0&0&0\\0&60\,200.6&18\,392.0&0&0&0\\0&18\,392.0&59\,329.2&0&0&0\\0&0&0&0&0&0\\0&0&0&0&20\,593.5&0\\0&0&0&0&0&0\end{bmatrix}$$

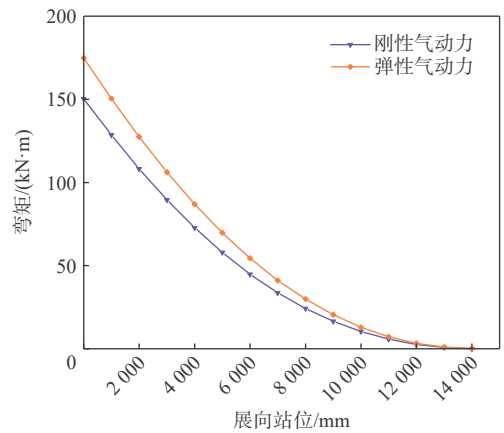
图 15(b) 对应的弹性模量矩阵分别为

$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}0&0&0&0&0&0\\0&48\,235.4&14\,953.0&0&0&0\\0&14\,953.0&48\,235.4&0&0&0\\0&0&0&0&0&0\\0&0&0&0&16\,641.2&0\\0&0&0&0&0&0\end{bmatrix}$$

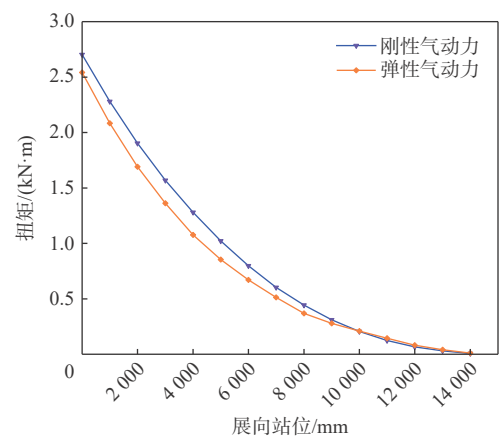
$$\boldsymbol{E}^{\mathrm{H}}=\begin{bmatrix}0&0&0&0&0&0\\0&36\,176.6&11\,214.7&0&0&0\\0&11\,214.7&36\,176.6&0&0&0\\0&0&0&0&0&0\\0&0&0&0&12\,480.9&0\\0&0&0&0&0&0\end{bmatrix}$$



(a) 气动力下剪切力分布



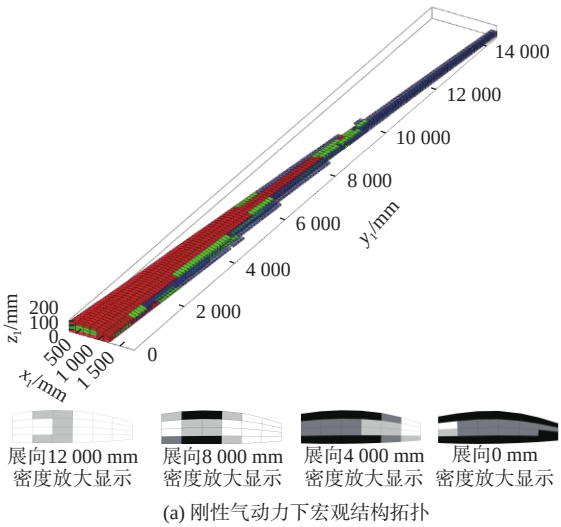
(b) 气动力下弯矩分布



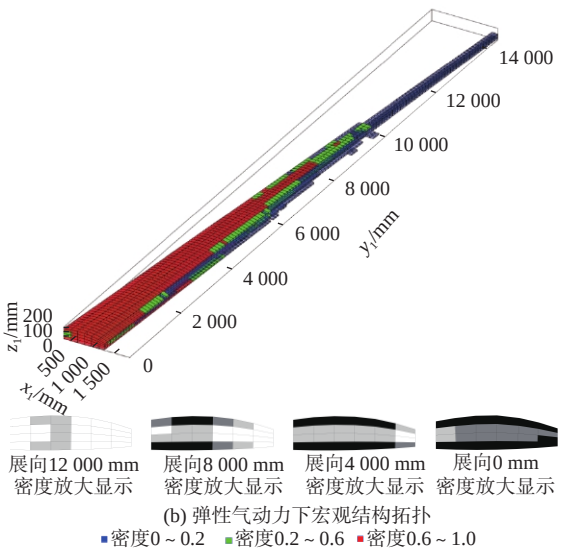
(c) 气动力下扭矩分布

图 11　机翼的气动力分布

Fig. 11　Aerodynamic load distribution of the wing



(a) 刚性气动力下宏观结构拓扑



(b) 弹性气动力下宏观结构拓扑

图 12 刚性和弹性气动力下宏观结构拓扑

Fig. 12 Macrostructural topology under rigid loading and elastic aerodynamic loads

2种边界条件下的优化问题具有相似的拓扑结构，随着体积约束的降低，微结构拓扑趋近于平行 y_2z_2 平面的厚板。微材料结构的弹性矩阵表明，y_2 方向及 z_2 方向模量较大。

根据图 11 结果可知，机翼外段的载荷最小，由于气动力整体趋势类似，优化问题中的宏观结构主要对外段进行了削弱。对于微结构构型，由于气动力导致结构承受的扭矩较弯矩小，弯矩导致的应力应变对优化的影响较大。在弯矩作用下，机翼结构需要沿 y_2 方向承受较大的拉压应力，因此，最终设计的微结构主要平行于 y_2z_2 平面。对于扭矩产生的剪切力，翼盒的上下表面具有蒙皮结构，能够提供 x_2y_2 平面内抵抗扭矩的刚度，而翼盒侧边设计出的宏微观构型能够提供 y_2z_2 平面内抵抗扭矩的刚度。对于后掠角较小的大展弦比机翼，机翼的外段会在气动力下产生正扭转，从而增加气动力。因此，与刚性气动力相比，弹性气动力下的优化结果

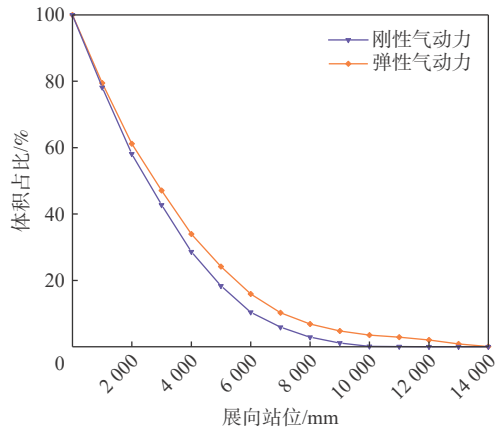
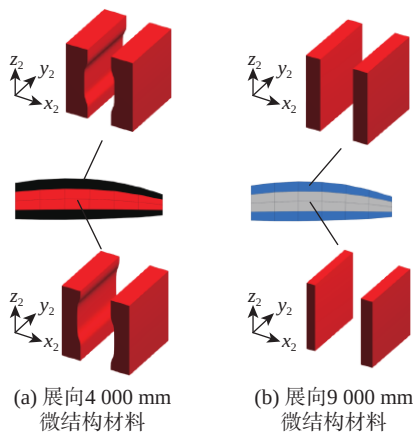


图 13 不同展向位置单元体积对总体积的百分比

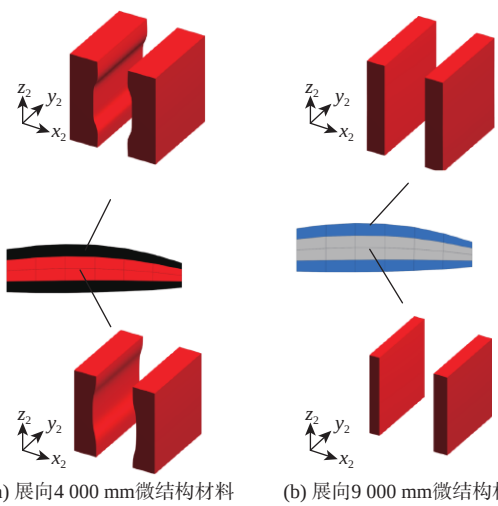
Fig. 13 Ratio of percentage of element volume to total volume at different spanwise stations



(a) 展向4 000 mm 微结构材料　(b) 展向9 000 mm 微结构材料

图 14 刚性气动力下的微结构拓扑

Fig. 14 Microstructural topology under rigid aerodynamic load



(a) 展向4 000 mm微结构材料　(b) 展向9 000 mm微结构材料

图 15 弹性气动力下的微结构拓扑

Fig. 15 Microstructural topology under elastic aerodynamic load

将更多的结构分配到机翼外段，从而限制机翼的弹性变形。

4 结 论

本文提出一种考虑飞行载荷的机翼结构和材

料多尺度气动弹性优化方法。

1）优化方法被应用于大展弦比机翼结构设计中。基于求取的形状导数，采用优化准则方法迭代扩展系数，得到了理想的柔顺度。

2）通过设计宏观和微结构构型，对比了采用单种和多种微结构材料在飞行器刚度设计中的结果。在相同气动力下，基于多种微结构材料的柔顺度较单种微结构材料减小了 18.6%。

3）研究了大展弦比机翼弹性效应对优化结果的影响。由于机翼结构的正扭导致弹性气动力大于刚性气动力，而刚/弹性气动力下的微结构构型类似，弹性气动力下的宏观结构较刚性气动力，主要对外段进行加强，从而抑制机翼扭转。

参考文献（References）

[1] COMPTON B G, LEWIS J A. 3D-printing of lightweight cellular composites[J]. Advanced Materials, 2014, 26(34): 5930-5935.

[2] ZHENG X Y, LEE H, WEISGRABER T H, et al. Ultralight, ultrastiff mechanical metamaterials[J]. Science, 2014, 344(6190): 1373-1377.

[3] 倪维宇, 张横, 姚胜卫. 考虑阻尼性能的复合结构多尺度拓扑优化设计[J]. 航空学报, 2021, 42(3): 332-342.
NI W Y, ZHANG H, YAO S W. Concurrent topology optimization of composite structures for considering structural damping[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(3): 332-342 (in Chinese).

[4] JENETT B, CALISCH S, CELLUCCI D, et al. Digital morphing wing: active wing shaping concept using composite lattice-based cellular structures[J]. Soft Robotics, 2017, 4(1): 33-48.

[5] CRAMER N B, CELLUCCI D W, FORMOSO O B, et al. Elastic shape morphing of ultralight structures by programmable assembly[J]. Smart Materials and Structures, 2019, 28(5): 055006.

[6] CRAMER N B, KIM J, JENETT B, et al. Simulated scalability of discrete lattice materials substructures to commercial scale aircraft [C]//AIAA Aviation 2020 Forum. Reston: AIAA, 2020: 2656.

[7] WANG C, ZHU J H, WU M Q, et al. Multi-scale design and optimization for solid-lattice hybrid structures and their application to aerospace vehicle components[J]. Chinese Journal of Aeronautics, 2021, 34(5): 386-398.

[8] WANG X Z, ZHANG S S, WAN Z Q, et al. Aeroelastic topology optimization of wing structure based on moving boundary meshfree method[J]. Symmetry, 2022, 14(6): 1154.

[9] OTOMORI M, YAMADA T, IZUI K, et al. Matlab code for a level set-based topology optimization method using a reaction diffusion equation[J]. Structural and Multidisciplinary Optimization, 2015, 51(5): 1159-1172.

[10] WANG S Y, WANG M Y. Radial basis functions and level set method for structural topology optimization[J]. International Journal for Numerical Methods in Engineering, 2006, 65(12): 2060-2090.

[11] XIE Y M, STEVEN G P. A simple evolutionary procedure for structural optimization[J]. Computers & Structures, 1993, 49(5): 885-896.

[12] XIA L, XIA Q, HUANG X D, et al. Bi-directional evolutionary structural optimization on advanced structures and materials: a comprehensive review[J]. Archives of Computational Methods in Engineering, 2018, 25(2): 437-478.

[13] BENDSØE M P, SIGMUND O. Material interpolation schemes in topology optimization[J]. Archive of Applied Mechanics, 1999, 69(9): 635-654.

[14] 闫浩, 吴晓明. 载荷敏度抑制下涡轮盘拓扑优化及结构演化[J]. 航空学报, 2022, 43(5): 339-348.
YAN H, WU X M. Topology optimization and structure evolution of turbine disks based on load sensitivity suppression[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(5): 339-348(in Chinese).

[15] RODRIGUES H, GUEDES J M, BENDSOE M P. Hierarchical optimization of material and structure[J]. Structural and Multidisciplinary Optimization, 2002, 24(1): 1-10.

[16] GAO J, LUO Z, XIA L, et al. Concurrent topology optimization of multiscale composite structures in Matlab[J]. Structural and Multidisciplinary Optimization, 2019, 60(6): 2621-2651.

[17] WANG C, ZHU J H, ZHANG W H, et al. Concurrent topology optimization design of structures and non-uniform parameterized lattice microstructures[J]. Structural and Multidisciplinary Optimization, 2018, 58(1): 35-50.

[18] ZHANG Y, LI H, XIAO M, et al. Concurrent topology optimization for cellular structures with nonuniform microstructures based on the Kriging metamodel[J]. Structural and Multidisciplinary Optimization, 2019, 59(4): 1273-1299.

[19] LI K Y, YANG C, WANG X Z, et al. Multiscale aeroelastic optimization method for wing structure and material[J]. Aerospace, 2023, 10(10): 866.

[20] HASSANI B, HINTON E. A review of homogenization and topology optimization I: homogenization theory for media with periodic structure[J]. Computers & Structures, 1998, 69(6): 707-717.

[21] XIA Z H, ZHOU C W, YONG Q L, et al. On selection of repeated unit cell model and application of unified periodic boundary conditions in micro-mechanical analysis of composites[J]. International Journal of Solids and Structures, 2006, 43(2): 266-278.

[22] MEI Y L, WANG X M. A level set method for structural topology optimization and its applications[J]. Advances in Engineering Software, 2004, 35(7): 415-441.

[23] WENDLAND H. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree[J]. Advances in Computational Mathematics, 1995, 4(1): 389-396.

[24] LUO Z, TONG L Y, KANG Z. A level set method for structural shape and topology optimization using radial basis functions[J]. Computers & Structures, 2009, 87(7-8): 425-434.

[25] KIM N H, DONG T, WEINBERG D, et al. Generalized optimality criteria method for topology optimization[J]. Applied Sciences, 2021, 11(7): 3175.

Aeroelastic optimization of wing structure and material using multiple microstructures

LI Keyu[1, 2], YANG Chao[1], WANG Xiaozhe[3, *], WAN Zhiqiang[1], LI Chang[1]

(1. School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China;
2. China Academy of Launch Vehicle Technology, Beijing 100076, China;
3. Institute of Unmanned System, Beihang University, Beijing 100191, China)

Abstract: Microstructured materials are distinguished by their lightweight and multifunctional characteristics, which have a wide variety of applications in the aerospace field. However, the stiffness properties of a single microstructural material are not adaptable to the further weight reduction requirements of an airplane. This paper presents a multiscale aeroelastic optimization method for wing structure and materials based on multiple microstructural configurations. The method is designed for macro/microstructural configurations of wings with large aspect ratios. Homogenization is used to quantify the mechanical properties of microstructured materials, and the density of macroscopic structural elements is combined to determine the total stiffness. The elastic aerodynamic force and displacement are obtained by solving the aeroelastic balance equation. Co-design of the wing structure and material is achieved based on sensitivity information. The optimization results indicate that the overall structural compliance based on multiple microstructured materials is reduced by 18.6% compared to a single microstructured material under the same load. Compared to the optimization results obtained under rigid aerodynamic forces, the effect of wing elasticity has a lesser impact on the microstructure topology configuration, while the outer segments of the macrostructure are strengthened. Furthermore, an acceptable distribution of stiffness throughout the wing may be achieved and the wing structure can be efficiently optimized under complex boundary conditions using the method described in this study.

Keywords: microstructured materials; wing with large aspect ratio; multiple microstructural configurations; multiscale aeroelastic optimization; effect of wing elasticity

Received: 2024-03-27; **Accepted**: 2024-06-14; **Published Online**: 2024-06-21 16:13

URL: link.cnki.net/urlid/11.2625.V.20240621.1146.001

Foundation item: Aeronautical Science Foundation of China (2022Z012051001)

* **Corresponding author**. E-mail: wangxiaozhemvp@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0110

基于数字孪生的健康服役框架及其在永磁同步电机的应用

郭浩宇[1, 2, 3, 4]，王少萍[1, 3, 4, *]，石健[1, 3, 4]，张育玮[1, 3, 4]

(1. 北京航空航天大学 自动化科学与电气工程学院，北京 100191；2. 都灵理工大学 机械与航空航天工程系，都灵 10129；
3. 北京航空航天大学 宁波创新研究院，宁波 315800；4. 天目山实验室，杭州 310023)

摘 要：现代机电系统研制周期短、健康服役要求高，传统的设计方法已无法满足研制需求，尤其无法表征机电系统健康服役状态变化、性能退化及外部干扰和不确定性的影响。针对该问题，聚焦机电系统健康服役，提出一种基于数字孪生的机电系统健康服役框架，旨在高效利用全生命周期数据实现快速迭代。该框架涵盖了设备的整个服役周期，能够利用不同阶段的工况数据更新数字孪生模型，实现对设备实时状态的准确检测。这种基于数字孪生的方法不仅能够在模型中模拟机电系统的服役状态，还能实时评估机电系统的健康状态，更好地满足现代机电系统的研制和服役需求。以航空航天领域常用的机电驱动设备永磁同步电机为例，证明了所提方法的有效性和实用性。

关 键 词：数字孪生；服役工况；迭代更新；健康管理；永磁同步电机
中图分类号：TP391.9
文献标志码：A **文章编号**：1001-5965(2026)05-1647-10

现代机电系统的研制周期越来越短、健康服役时间要求越来越长，传统的设计方法已不足以全面描绘设备在服役期间的健康状态变化及性能退化。这些方法更难以捕捉外部干扰和各种不确定性因素对机电系统的影响，从而导致机电系统执行任务时面临较高的风险，安全性无法得到充分保障。为了有效利用历史数据，根据实际需求及服役工况变化快速迭代设计，数字孪生技术应时而生。数字孪生技术通过在虚拟空间中创建一个与真实机电系统相对应的数字模型，能够模拟和反映实体机电系统在其整个生命周期中的各种状态和过程。通过数字孪生技术，不仅可以在模型中提前模拟机电系统的服役状态，还能够准确捕捉不同工况下的性能退化规律，进而实现对机电系统健康状态的实时评估和监控，极大提升了机电系统的可靠性和安全性。

数字孪生领域主要关注与实物相同的高保真数学模型构建，以实现产品不同条件下虚拟空间对实体对象的准确映射[1]。凭借该优势，数字孪生技术在状态监测与健康管理方面取得了一定的研究成果。

文献 [2] 提出了一种数字孪生方法，用于支持汽车制动系统的热监测和累积磨损预测。文献 [3] 提出了一种基于传感器监测数据的数字孪生液压支架姿态监测方法，并通过三维可视化界面反映支架当前的姿态。文献 [4] 根据运行时间、车辆距离和车载电机的健康状态，开发了一种以神经网络为建模方式的数字孪生系统，以符合电机运行状态。

收稿日期：2024-02-28；**录用日期**：2024-05-10；**网络出版时间**：2024-08-01 09:03
网络出版地址：link.cnki.net/urlid/11.2625.V.20240731.1437.001
基金项目：国家自然科学基金 (U2233212)；北京市自然科学基金 (L221008)；国家留学基金 (202106020001)
* **通信作者**. E-mail：shaopingwang@buaa.edu.cn

引用格式：*郭浩宇，王少萍，石健，等. 基于数字孪生的健康服役框架及其在永磁同步电机的应用 [J]. 北京航空航天大学学报，2026，52（5）：1647-1656. GUO H Y，WANG S P，SHI J，et al. Digital twin-based health service framework and its application in permanent magnet synchronous motors[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（5）：1647-1656（in Chinese）.*

文献 [5] 利用小卫星的在轨和地面多类数据，建立了小卫星健康预测的数字孪生模型，并结合机理知识和性能数据，提出了一种健康评估和在轨剩余寿命预测框架。文献 [6] 通过数字孪生驱动的方式对飞机起落架系统的健康管理技术进行研究，解决传统健康管理方式存在的知识不完备、数据不平衡和模型固化等问题。文献 [7] 建立了海上风电机组扭转动态数字孪生模型，并提出了针对该数字孪生相关的参数更新算法，以实时了解参数的健康状态。文献 [8] 基于数字孪生五维模型提出了对关节机器人的健康状态评估方法，通过使用卷积神经网络对定子电流进行预测以检测轴承故障，并开发了关节机器人的可视化和运维平台。文献 [9] 以有限元形式建立了 500 m 口径球面射电望远镜的数字孪生模型，并通过有限元分析评估安全状态并预测疲劳寿命，以实现有效的状态维护。文献 [10] 基于物理模型进行了机翼性能退化评估，通过动态贝叶斯网络监测机翼系统的健康状态。 文献 [11] 开发了一种基于物理模型的数字孪生系统，能够准确预测飞机复合材料夹层结构的动态行为和分层面积，该数字孪生模型与实验结果高度匹配。文献 [12] 采用基于物理模型的数字孪生仿真方式，计算机械设备的剩余使用寿命，以实现对制造资源的预测性维护。文献 [13] 总结了航空发动机主轴轴承生命周期中基于数字孪生的关键制造技术，同时对航空发动机主轴轴承进行了故障诊断和寿命分析。文献 [14] 提出了应用数字孪生技术预测控制元件驱动机理的概念框架，并通过数据驱动的方法，利用线圈电流曲线进行异常检测，以优化核电站的运行和维护过程。文献 [15] 讨论了基于数字孪生的故障预测与健康管理方面的应用，重点关注了设计、开发和实现的核心方面，并证明了这些方法可以改善液氢储存系统的可靠性。

通过对以上研究工作分析发现，虽然许多工作实现了通过数字孪生对实际机电系统运行全服役过程维护的目标，但对以下方面缺少相应的研究：①实时性和动态性。已有研究大多是在现有软件平台基础上的离线仿真，数字孪生作为能够反映实际机电系统特性的概念技术，要求其能够具备伴随实体状态实时变化的特征和具备自我更新的动态特性。②当前的研究工作已经实现了对特定工况和设备状态下的健康评估，并通过分析多种工况下的结果对整个运行周期进行分析。然而，这些分析尚未充分反映出各个工况或机电系统在服役过程中的变化情况，更全面的数字孪生的服役过程需要考虑历史数据累积与实时的数据融合。

通过考虑服役历史和当前运行阶段的状态数据，能使数字孪生模型具有高保真性和反映退化特性的能力，有助于更好地为复杂系统的健康管理和维护决策提供支持和服务。

针对上述问题，本文提出了基于数字孪生技术的机电系统健康服役框架。该框架覆盖了机电系统整个服役过程，在不同阶段利用工况数据实时更新数字孪生模型，更精确地表征机电系统的运行状态，为机电系统的安全使用奠定了理论基础。同时，以航空航天领域中常用的永磁同步电机进行应用和验证，结果表明，本文方法能够为设备的研制和健康维护提供更为可靠和精确的数据支持。

1 基于数字孪生的健康服役框架

基于数字孪生的健康服役框架核心是准确映射物理实体在复杂工况作用下服役全过程的模型变化，其中，最重要的 2 个要素是：孪生数据和健康评估。图 1 为本文提出的数字孪生的机电系统健康服役框架。

1.1 数字孪生体

1.1.1 虚拟对象

图 1 左侧所示为物理实体及其虚拟对象组成的复合系统。在数字孪生架构中，虚拟对象代表所需描述对象的物理解析模型或运算。在数字孪生的实际应用中，希望能够通过具有明确物理参数含义的模型来获取当前机电系统的运行状态。因此，在构建数字孪生的虚拟对象时，需要遵循“第一性原理”，即建立能够描述实际对象的物理解析模型。这种模拟涵盖了机电系统的各个方面，从基本物理特性到复杂的运行机制，是全面表征物理实体实际表现全面而详尽的虚拟副本。

1.1.2 本体参数与状态数据

虚拟对象用于描述实际对象的物理规律，而孪生数据中的本体参数和状态参数则为在虚拟对象中作用的数据。本体参数（$\theta=\{\theta_1,\theta_2,\cdots,\theta_m\}$）用来表示实际机电系统的固有内部参数，如弹簧的弹性系数 k、电路中的定值电阻 R 等，用于描述实际对象的固有属性。而状态参数（$C=\{C_1,C_2,\cdots,C_k\}$）用于描述外界作用在以本体参数为基础的虚拟对象上时，实际对象的运行状态，如在外力为 F_e 的情况下，弹性系数为 k 的弹簧的伸长量为 $x=F_e/k$，在电压为 V 时，阻值为 R_e 的电阻电流为 $I_e=V/R_e$，如图 2 所示。在以弹簧为例的数字孪生体中，虚拟对象为外部作用力 F_e 与弹性系数 k 的除法运算，本体参数为弹性系数 k，状态参数为弹簧伸长量 x；在以电阻为例的数字孪生体中，虚拟对象为电压 V 与电阻

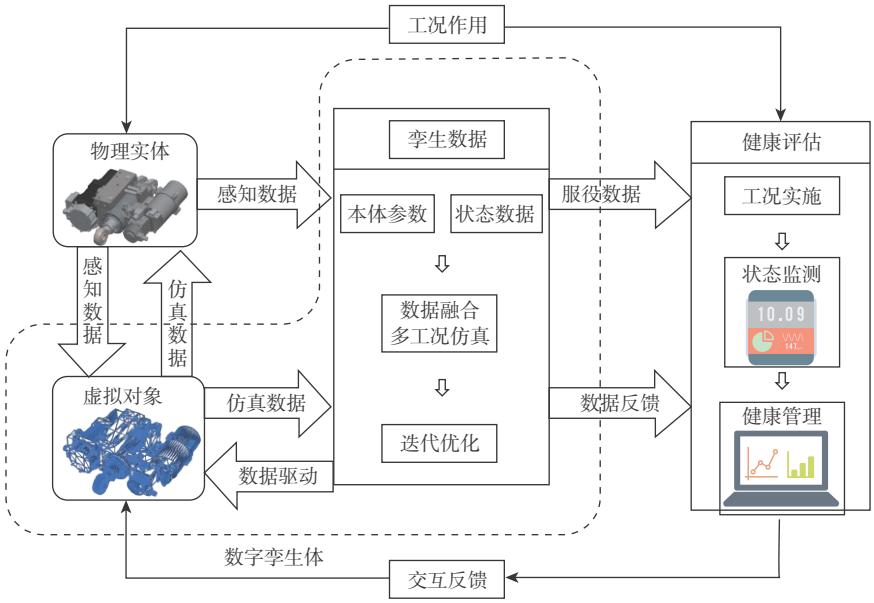


图 1　基于数字孪生的健康服役框架

Fig. 1　Digital twin-based health service framework

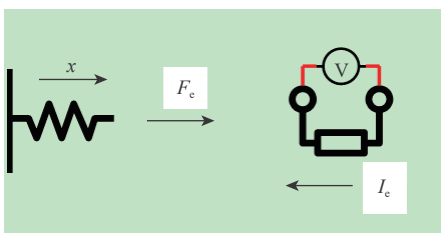


图 2　虚拟对象、本体参数和状态参数示意图

Fig. 2　Schematic diagram of virtual objects, ontology parameters, and state parameters

R_e的除法运算关系, 本体参数为电阻R_e, 状态参数为电流I_e。

但现代的机电系统通常涉及多个学科, 且各学科之间存在着复杂的耦合关系。为了能够更准确地模拟实际机电系统, 需要尽可能地将所有可能影响或反映实际机电系统的学科进行多数据融合仿真, 以确保数字孪生体的保真度。

1.1.3　工况作用下的数据融合

物理实体通常会在某一工况下运行工作, 忽略以往工况对实际机电系统的累计影响会影响数字孪生模型的建模精度。因此, 在数字孪生模型中也要刻画多工况作用的过程。以 1.1.2 节弹簧系统为例, 其伸长量的模型如下:

$$\frac{F_e}{x} = k(t,m,h,\cdots) \tag{1}$$

式中: t、m、h 分别为温度、材料、历史工况, 共同决定弹簧的弹性系数。在高精度的数字孪生建模中, 需要综合考虑这些因素及材料的疲劳, 以精确描述对应的实体对象。

假设物理实体经历的工况为$Q=\{q_1,q_2,\cdots,q_i,\cdots,q_l\}$, 统一的数字孪生模型可以表征为

$$D_0 = f(Q,C,\theta) \tag{2}$$

式中: D_0为多数据融合的数字孪生体。在多工况仿真过程中, 数字孪生模型能够根据不同操作条件和环境变化进行调整, 从而模拟设备在各种可能工况下的行为和性能。

数据融合的目的是将来源于不同源的数据集成到一个可靠的、统一的模型中。对于涉及多参数和多工况的数字孪生模型来说, 以本体参数为例, 通过贝叶斯推理的多传感器融合方法表示为

$$p(\theta_i|z_i) = \frac{p(z_i|\theta_i)p(\theta_i)}{p(z_i)} \tag{3}$$

式中: $p(\theta_i \mid z_i)$为后验概率, 即在给定观测z_i的条件下, 参数θ_i的概率; $p(z_i|\theta_i)$为观测的似然性; $p(\theta_i)$为先验概率, 即在观测之前状态的概率; $p(z_i)$为观测的边缘概率。

还可以通过最简单的加权平均法对具有相同维度的数据进行数据融合, 以工况参数为例, 如下:

$$Q_{\text{fused}} = \sum_{i=1}^{n} \omega_i q_i \Big/ \sum_{i=1}^{n} \omega_i \tag{4}$$

式中: Q_{fused}为融合后的数据; q_i为第i个数据源的数据; ω_i为对应的权重; n为数据源总数。

通过这种基于数据融合的多工况仿真, 能帮助决策者和维护团队优化操作策略, 预防设备故障, 从而提高设备的可靠性和效率。该方法的另一个关键优势是能够在不干扰实际设备运行的情况下进行, 为风险评估和决策支持提供了安全且高效的工具。

1.1.4 优化迭代

数字孪生模型的优化迭代是数字模型根据外部特征进行自我演化的过程，其结合了实时数据分析和当前的健康因子，完成模型更新的闭环，以不断提升模型的准确性和效能。

在该过程中，系统先收集来自实体机电系统的实时运行数据，包括本体参数、状态参数和环境变量等。这些数据被传输至数字孪生模型，并用于校准和调整模型参数。数字孪生更新模型可以表示为

$$D_{t+1}=F(D_t,\theta_t,C_t,H,\eta) \tag{5}$$

式中：D_{t+1}为$t+1$时刻的数字孪生模型；θ_t为t时刻实体数据；H为健康因子；η为参数更新的学习率。F函数内，t时刻的孪生模型、运行工况、模型参数及健康因子决定了新数据如何影响数字孪生模型的更新，F是一个更新函数，其定义了如何根据当前的数字孪生模型、物理设备数据、环境性能数据及模型参数来更新数字孪生模型。

在模型的参数更新过程中，通过健康因子H触发。定义损失函数$f_{\text{loss}}(\theta,C)$衡量模型与实体之间的差异，如下：

$$f_{\text{loss}}(\theta,C)=(D_{t+1}-D_t)^2 \tag{6}$$

再计算差异的变化梯度，来决定模型本体参数θ的变化方向，如下：

$$\begin{cases}\nabla_\theta f_{\text{loss}}(\theta,C)=\dfrac{\partial f_{\text{loss}}}{\partial\theta}\\ \nabla_C f_{\text{loss}}(\theta,C)=\dfrac{\partial f_{\text{loss}}}{\partial C}\end{cases} \tag{7}$$

根据参数的变化方向更新参数，并通过不断迭代得到最后的更新参数结果，如下：

$$\begin{cases}\theta_{\text{new}}=\theta-\eta\nabla_\theta f_{\text{loss}}(\theta,C)\\ C_{\text{new}}=C-\eta\nabla_C f_{\text{loss}}(\theta,C)\end{cases} \tag{8}$$

本文提出的建模及验证方法主要适用于参数易获取且具有可计算表达式的模型，便于进行模型精度之间的对比。

对于难以直接获取数据或难以直接处理进行对比的模型，可以参考以下方法进行模型与实际模型之间的对比：

1）间接测量和推断。利用可获得的相关数据或信息，通过推断或间接测量来更新模型，这可能需要使用统计学方法或利用已知的模型进行推断。

2）利用近似方法和模拟。当直接对比困难时，可以使用近似方法或模拟来尽可能地模拟实际情况，对模型进行更新。

3）利用领域专家知识。寻求领域专家的意见和知识，以辅助更新模型，其可能有关于系统行为的经验知识，可用来指导模型更新。

1.2 基于数字孪生的健康状态评估

在进行数字孪生机电系统的状态评估时，先对其数字孪生状态进行实时监测，以获取关键信息数据，再将这些数据应用于专门的性能预测算法中，对应于实际机电系统的数字孪生体进行细致的性能监控和分析。这意味着结合当前的工况作用和历史服役数据，通过数据反馈的方式，对当前实际机电系统的健康状态进行评估，并相应地调整数字孪生模型。因此，根据健康评估的数据来调整数字孪生模型，以更准确地反映实际机电系统的运行状态和健康状况。在图1中，从孪生数据指向虚拟对象的箭头表示数据驱动，促使数字孪生模型运行；而从健康评估指向孪生数据的箭头则是基于外部检测差异，用于更新孪生数据，两者含义不同。该方法不仅增强了对设备当前状态的理解，还为设备的长期维护和管理提供了重要依据。

1.2.1 基于工况实施的状态监测

区别于传统的状态监测方法，本文提出了一种创新的状态监测方式，其依托于不同状态下工况数据的性能分析。该方法通过分析以往累积的各种工况数据，对数字孪生模型进行持续的更新和调整，使其能够更好地适应各种变化的工况，不仅提高了状态监测的准确性，而且数字孪生历史工况的记录和基于这些历史数据更新参数的过程可以通过数学模型来简化表示。该模型不仅需要记录历史数据，还需要反映如何根据这些数据调整和优化参数。记录数字孪生历史工况的公式可能如下：

$$C_{t+1}=C_t\cup\{\theta_t,D_t,\Delta t\} \tag{9}$$

式中：C_t和C_{t+1}分别为时间点t和t+1的所监测的历史服役工况数据集；Δt为从上一次更新到当前的时间间隔；“∪”表示在t时刻收集的历史数据，经过Δt后作为新的元素添加到历史数据中，形成新的历史数据集C_{t+1}。

1.2.2 健康管理

状态监测在数字孪生系统中扮演着信息交互的关键桥梁角色，其不仅连接了物理实体和虚拟模型之间的数据流动，还确保了两者之间的实时同步与互动。通过精确的状态监测，数字孪生系统能够实时反映物理设备的运行状况。

根据C_t是在时间点t的状态监测结果，这些参数综合了实体机电系统与其数字孪生的数据，反映了在当前本体参数设定下，基于历史工况数据采集的数据。通过状态监测得到的信号特征，可以利用

当前状态和健康状态阈值对比, 对机电系统当前状态进行评估。由于数字孪生系统能够直接获取内部参数情况, 比较关键参数阈值, 即 H_t/H_T, 其中, H_t为当前参数值, H_T为该参数的阈值。H_t/H_T只是一种计算状态的方法, 可以选择其他计算方法, 但需要保证该参数的值的范围能够在 0～1 之间, 其计算形式并不唯一。通过当前时刻的健康状态向量与完好的状态向量进行比对, 得到范围为 0～1 的健康因子 H。本文以本体结构参数为例定义健康度为

$$\begin{cases} \theta_L \leqslant H < 1 & \text{轻度退化} \\ \theta_M \leqslant H < \theta_L & \text{中度退化} \\ \theta_S \leqslant H < \theta_M & \text{重度退化} \\ 0 \leqslant H < \theta_S & \text{失效} \end{cases} \tag{10}$$

式中: 不同的θ_*代表着不同健康度的上下界。当 $H \in [\theta_L, 1)$时, 系统健康状态处于轻度退化状态; 当 $H \in [\theta_M, \theta_L)$时, 系统健康状态处于中度退化状态; 当 $H \in [\theta_S, \theta_M)$时, 系统健康状态处于重度退化状态; 当 $H \in [0, \theta_S)$时, 系统健康状态处于失效状态。此处健康因子仅为示例, 具体的健康度划分方法需要结合实际的机电系统运行情况决定。作为一种交互反馈信息, 基于实际机电系统现有的状态提供维护建议。

基于获得的健康因子 H, 再基于孪生数据中的模型参数进行优化(见式 (5)), 进而对数字孪生模型执行参数更新, 形成基于健康状态更新优化的闭环反馈。

同理, 式 (10) 也可以应用于状态参数的健康状态参数更新。将健康因子 H 作为激励, 在服役全生命周期内对数字孪生模型更新, 进而得到准确反映物理实体服役过程的数字孪生模型。

2　永磁同步电机上的应用

永磁同步电机是一种采用永磁体作为磁场源的同步电机。与传统的感应电机相比, 永磁同步电机具有高效、高功率密度、响应迅速和精确控制等优点, 因此, 在许多应用领域中得到广泛应用, 其结构如图 3 所示。图中, R_a、R_b、R_c分别为电机三相绕组中的电阻, L_a、L_b、L_c分别为电机三相绕组中的自感系数, U_a、U_b、U_c分别为电机三相绕组中的电压, i_a、i_b、i_c分别为电机三相绕组中的定子电流。

永磁同步电机作为重要的动力来源, 其工况十分恶劣。一旦发生失效或故障, 不仅会对整个机电系统造成严重威胁, 还可能危及人员的安全。因此, 对永磁同步电机的健康进行有效的管理显得尤为迫切。

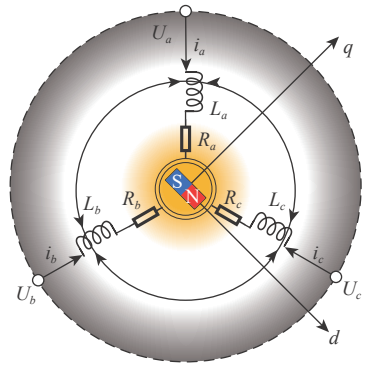


图 3　永磁同步电机的结构

Fig. 3　Structure of permanent magnet synchronous motor

但对于永磁同步电机这类结构紧凑、集成度高、工况环境恶劣的机电系统, 采用传统的状态监测和健康管理方式存在着巨大挑战, 其通常在复杂、恶劣的环境中执行任务, 如高温、振动、过载等极端条件。此外, 一些内部故障可能难以直接检测, 如电机永磁体退磁、轴承损坏等问题。这些故障可能需要更为复杂的监测手段和技术, 甚至需要拆解才能发现。传统的状态监测与健康管理方式在恶劣条件下难以保证获取足够且高质量的数据。因此, 基于传统方式的永磁同步电机的状态监测与健康管理结果的可靠性存在一定的隐患。

基于本文提出的数字孪生健康服役框架利用历史工况数据, 根据不同时间点对设备模型进行动态更新。通过这种方式, 可以从模型中获得更为精确和全面的数据, 从而实现对设备状态的实时监控和高效管理。这种方法不仅提高了数据的准确性, 也为设备维护和决策支持提供了强有力的数据支撑。

2.1　永磁同步电机的服役数字孪生模型

2.1.1　虚拟对象

数字孪生中的虚拟对象是对实际机电系统的行为和运行规则的高度精确刻画。采用 i_d=0 的永磁同步电机的转速与电流闭环控制模型作为说明案例, 其控制框图如图 4 所示。

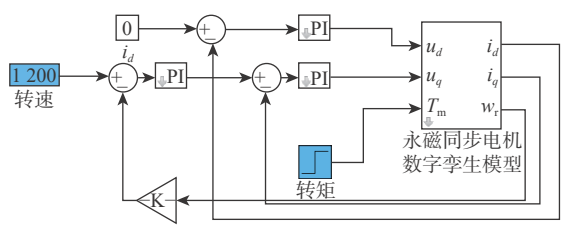


图 4　永磁同步电机的控制框图

Fig. 4　Control block diagram of permanent magnet synchronous motor

在该控制模型中, 蓝色环节为电机的动力学参数。转速与扭矩转速为外部输入, 负载扭矩加载在

电机上的外部工况。因此，通过将实际转速和工况传输至对应的数字孪生模型中，即可获得相应状态下实际电机机电系统中的运行状态。本文将从电机实体中采集到的转速和扭矩分别传入电机的数字孪生模型中，以实现相同工况下电机运行的仿真。这样，可以在数字孪生模型中模拟实际电机的行为，从而更好地理解其在不同工况下的性能表现。通过结合外部实际电机的转速、扭矩和所建立的永磁同步电机的数字孪生体，实现了对实际电机运行行为的详细模拟和深入分析，如图5所示。图中，SVPWM是空间矢量脉冲宽度调制，VSI是电压源型逆变器，DT是电机的数字孪生模型。

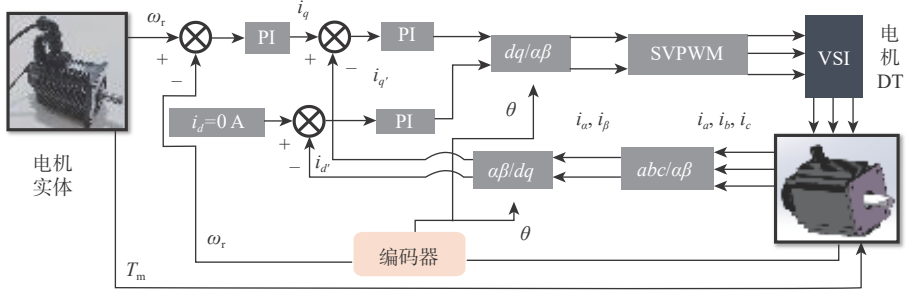


图5 永磁同步电机的数字孪生系统

Fig. 5 Digital twin system of permanent magnet synchronous motors

图5中，i_a、i_b、i_c、i_α、i_β、i_d、i_q分别为对电机进行Clark和Park变化过程中的电流数值。在电机的运行过程中，转速ω_r和扭矩T_m是主要的工作参数。这2个参数不仅关键地描绘了电机的核心行为，而且也是操作者通过虚拟对象最直观、快捷了解设备性能的途径。这一过程可以借助可视化技术实现，使得操作者能够直观地观察和分析电机的运行状态，从而有效地监控和调节设备性能。

永磁同步电机扭矩和转速的虚拟对象如下：

$$\begin{aligned} T_m &= \frac{3}{2}P_n i_q \psi_f \\ \frac{d\omega_r}{dt} &= \frac{1}{J}(T_m - T_{load} - B\omega_r) \end{aligned} \tag{11}$$

式中：T_m为电机输出扭矩；T_{load}为外部负载扭矩；P_n为永磁体极对数；J为转动惯量；B为摩擦系数；i_q为定子电流在q轴上的分量；ψ_f为转子磁链。从式(11)可见，电机外部负载大小不变的情况下，发生退磁现象(ψ_f减小)，需要更大的电流驱动，输入功率会变大。

2.1.2 孪生数据

在数字孪生技术中，孪生数据扮演了至关重要的角色，其不仅作为描述机电系统模型的详细解析模型，还包含了模型的更新机制。这些数据确保数字孪生能够准确反映实际设备的状态和行为。以永磁同步电机为例，其核心的d-q轴电气数学模型为

$$\begin{cases} u_d = Ri_d + L_d\dfrac{di_d}{dt} - \omega_e L_q i_q \\ u_q = Ri_q + L_q\dfrac{di_q}{dt} + \omega_e\left(L_d i_d + \psi_f\right) \end{cases} \tag{12}$$

式中：u_d、u_q为定子电压的d-q分量；i_d、i_q为定子电流的d-q轴分量；R为定子的电阻；L_d、L_q为电感的d-q分量；ω_e为电角速度；ψ_f为永磁体磁通量。

在电机的数字孪生模型中，参数R、L_d、L_q和ψ_f构成了模型的本体参数，这些数据体现了设备的基础和固有属性。本文采用表贴式永磁同步电机作为实验用例，该电机的特点是d轴和q轴的电感值相等，即$L_d = L_q = L$。而u_d、u_q、i_d、i_q则是电机的状态数据，实时反映了电机运行状况。

这种区分确保了数字孪生模型既能准确地描述电机多学科的基本特性，又能灵活地追踪和反映其运行状态的变化。对永磁同步电机中的关键部件进行建模有助于明确其工作机理，同时还能分析各部件的退化因素及其引起的故障现象，为进一步的永磁同步电机故障诊断提供了基础。

2.1.3 迭代优化

数字孪生模型的更新功能是其与传统仿真技术区别的核心特征，这一功能被融入到孪生数据中，从而实现了对模型的持续更新和优化。这种动态更新机制确保数字孪生模型能够实时反映实际机电系统的状态变化，从而为决策支持和设备管理提供更加精准和可靠的数据支撑。

根据式(5)对模型更新的定义，带遗忘因子的递归最小二乘法(recursive least squares, RLS)是一种有效的在线参数估计方法，可以用来辨识永磁同步电机的电压方程参数。选择合适的初始参数估计值$\hat{\theta}(0)$，如$\hat{R}(0)$、$\hat{L}(0)$、$\hat{\psi}_f(0)$和协方差矩阵$\boldsymbol{P}(0)$。RLS流程如下：

首先，对每一个时间步k计算预测误差，如下：

$$\boldsymbol{e}(k) = \boldsymbol{u}(k) - \hat{\boldsymbol{u}}(k) \tag{13}$$

式中：$\boldsymbol{u}$为d和q轴上的电压分量。

其次, 根据包含 ω_e、i_q 等的数据与多工况仿真数据融合, 得到的增益矩阵为

$$\boldsymbol{K}(k)=\frac{\boldsymbol{P}(k-1)\boldsymbol{\phi}(k)}{\lambda+\boldsymbol{\phi}(k)^{\mathrm{T}}\boldsymbol{P}(k-1)\boldsymbol{\phi}(k)} \tag{14}$$

式中: λ 为遗忘因子(通常介于 0.9 和 1 之间)。

最后, 基于增益矩阵, 更新参数和协方差矩阵为

$$\begin{cases}\hat{\boldsymbol{\theta}}(k)=\hat{\boldsymbol{\theta}}(k-1)+\boldsymbol{K}(k)\boldsymbol{e}(k)\\ \boldsymbol{P}(k)=(\boldsymbol{P}(k-1)-\boldsymbol{K}(k)\boldsymbol{\phi}(k)^{\mathrm{T}}\boldsymbol{P}(k-1))/\lambda\end{cases} \tag{15}$$

输出最终的参数估计结果 $\hat{\boldsymbol{\theta}}(k)$, 即 $\hat{R}$、$\hat{L}$、$\hat{\psi}_f$。

参数的辨识方法需要结合实际数字孪生模型的形式。如果模型相对简单, 可以通过简便的运算方法得到当前关注的参数值。例如, 在式 (11) 中, 对摩擦系数 B 进行实时辨识时, 可以结合当前的数据信息进行计算, 具体如下:

$$B=\frac{1}{\omega_{\mathrm{r}}}\left(T_{\mathrm{m}}-T_{\mathrm{load}}-J\frac{\mathrm{d}\omega_{\mathrm{r}}}{\mathrm{d}t}\right) \tag{16}$$

式中: $\mathrm{d}\omega_{\mathrm{r}}/\mathrm{d}t$ 为电机的角加速度; T_{m} 可由式 (11) 进行计算。通过式 (16) 的方法对实际电机的实时状态进行监测, 能够实时地辨识和检测摩擦系数 B。对机械和电气参数状态进行全面计算有助于更充分地了解永磁同步电机的运行状态。

2.2　永磁同步电机数字孪生健康评估

2.2.1　工况实施

在设备的整个生命周期中, 数字孪生利用实际设备在过去运行期间的历史工况数据, 对其模型的本体参数和状态参数进行动态更新。数字孪生不仅能够基于历史运行数据进行更新, 从而实现对设备性能的持续监控和优化, 还能根据当下及历史退化结果预测未来的性能趋势和潜在故障, 为设备维护和决策提供数据支持。表贴式永磁同步电机的关键参数如表 1 所示。

表 1　永磁同步电机关键参数值

Table 1　Key parameter values of permanent magnet synchronous motor

本体参数	数值
磁链 ψ_f/Wb	0.153
电感 $L(L_d=L_q=L)$/mH	1.24
电阻 R/Ω	0.36

根据服役数字孪生在不同服役工况下运行的状态特点, 在历史工况, 即 T_{m}=6 N·m、ω_{r}=1 000 r/min, T_{m}=6 N·m、ω_{r}=1 200 r/min, T_{m}=6 N·m、ω_{r}=1 400 r/min, 实施下对电机的数字孪生模型进行更新验证, 结果如图 6 所示。可以观察到, 基于 RLS 的参数辨识方法在电机的不同外部工况中, 通过状态数据进行的

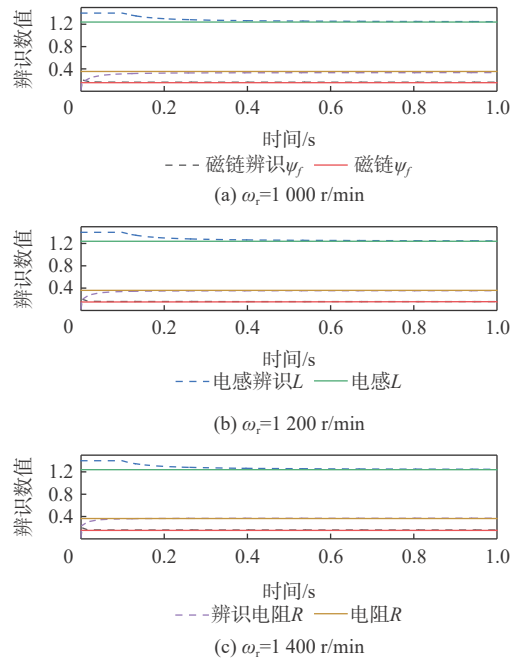


图 6　不同工况下孪生数据的迭代更新结果

Fig. 6　Iterative update results of twin data under different working conditions

参数辨识结果与实际电机本体参数之间展现出了高度的匹配, 保证了数字孪生体的孪生数据环节中模型迭代优化过程的准确性。这种高精度的参数更新不仅反映了算法本身的有效性, 而且突显了数字孪生在模拟和预测电机性能方面的高度可靠性。通过持续迭代和优化调整, 数字孪生能够确保模型始终紧密伴随实际机电系统的最新状态, 从而为性能预测提供了可靠的数据支撑。

2.2.2　状态监测

永磁体退磁是永磁同步电机常见的退化故障。当永磁体退磁后, 同样的扭矩需要更大的定子电流带驱动, 这会导致永磁体电机内部过热, 温度的升高会进一步影响永磁体, 从而形成恶性循环, 导致电机加剧退化。

永磁体的性能状态决定了电机的运行性能。通过结合多传感器状态监测信息, 根据文献 [16] 中的方式, 本文基于实际机电系统的多传感器数据对永磁同步电机的磁链数值进行了实时更新。基于 800 h 的服役历史工况数据, 本文对永磁同步电机中的永磁体磁链数值进行了健康度分析。实验数据测得的电机定子电流数据如图 7 所示。

图 7 表示了电机在不同服役工况(ω_{r} = 1 000, 1 200, 1 400 r/min转速)下, 800 h 内所记录的定子电流变化趋势。由于未改变外部扭矩 T_{m} 的大小, 定子电流的幅值并不会发生巨大改变, 但可以看出, 随着使用时间的延长, 定子电流逐渐增加, 这一现象表示着永磁体的磁链正在逐步减弱。电流的变

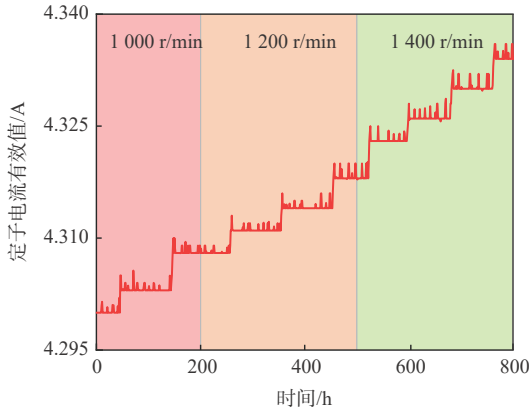


图7 永磁同步电机实测电流

Fig. 7 Measured current of permanent magnet synchronous motor

化是电机性能退化的重要指标，通过对这种趋势的及时识别和分析，可以采取适当的措施来调整运行策略或进行维护，以延长电机的使用寿命，并保持其最佳性能。

本文定义电流的健康状态分为3个阶段：(4.3,4.311)A、(4.311,4.323)A、(4.323,4.334)A。

永磁体退磁会导致同样的扭矩下电流的上升。当电流分别超过4.311 A(一级退化)、4.323 A(二级退化)和4.334 A(三级退化)或更大时，意味着电机的健康因子变化，需要对电机数字孪生模型进行更新。基于式(11)模型与参数更新的定义和式(12)的状态监测定义内容，永磁同步电机数字孪生体的磁链初始值为ψ_{f_0}=0.153 Wb。其根据RLS的方法，对应的健康因子对磁链的辨识数值变化如图8所示。

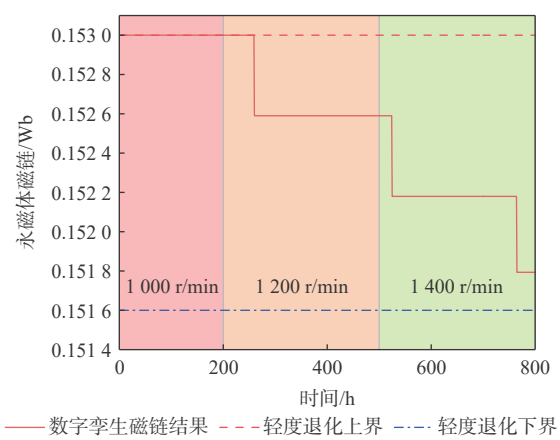


图8 基于数字孪生的永磁体磁链变化

Fig. 8 Digital twin-based flux variation of permanent magnet

根据图8中参数辨识的更新效果，对比了不同服役期的电机定子电流实际拟合曲线与数字孪生模型中模拟退磁磁链得到的电流曲线，结果如图9所示。进而能够直观地评估数字孪生模型在模拟电机性能退化方面的准确性和有效性。

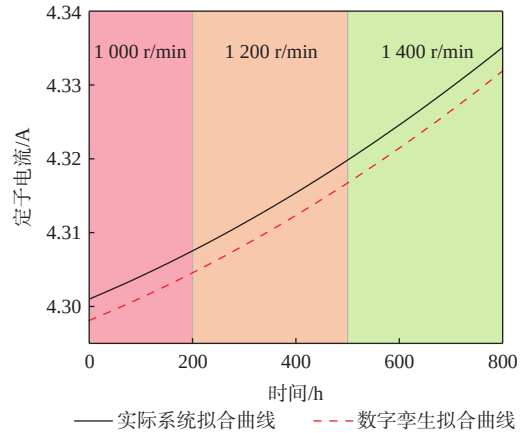


图9 实体电机及其孪生体不同服役期电流对比

Fig. 9 Comparison of current of physical motor and its twin in different service periods

从图9的数据分析可见，经过800 h不同工况服役后，数字孪生模型与实际机电系统之间依然保持了较高的趋势和精确度。根据图9中的数据计算，每一时刻的定子电流的误差率约为0.5%。这表明数字孪生技术在长时间复杂工况下依然能够有效地模拟和预测实体机电系统的性能和状态。因此，基于对电机状态的精准监测和分析，能够对电机执行高效而准确的健康维护策略。

2.2.3 健康管理

本文定义永磁体退化阈值为ψ_{f_τ}=0.1 Wb。永磁体的退磁表现为下降趋势，本文定义永磁体健康因子计算方法如下：

$$H=\frac{\psi_{f_0}-\psi_{f_t}}{\psi_{f_0}-\psi_{f_\tau}} \tag{17}$$

式中：$H\in[0,1]$为永磁体的健康状态；ψ_{f_t}为t时刻时磁链数值。其健康度划分如下：[0.9,1)、[0.7,0.9)、[0,0.7)。当$H\in[0.9,1)$时，电机处于轻度退化状态；当$H\in[0.7,0.9)$时，电机处于中度退化状态；当$H\in[0,0.7)$时，电机处于重度退化状态。基于此，800 h时的永磁体健康度为：$H=(0.153-0.151\,7)/(0.153-0.1)\approx0.975$。此时永磁体的健康状态为轻度退化状态，可以继续使用。针对不同健康状态的电机，可以制定相应的维护策略，永磁同步电机中涉及的永磁体退磁的具体健康维护措施如表2所示。

表2 永磁同步电机健康维护措施

Table 2 Health maintenance measures for permanent magnet synchronous motors

健康状态	维护措施
轻度退化	定期清洁，正常使用
中度退化	调整磁极位置，检查绕组
重度退化	更换永磁体，检修电机其余部件
再次重度退化	更换整个电机

同时, 可以基于当前的参数状态, 并结合专家系统, 对机电系统的退化进行更详细的分析和说明。通过这种方式, 操作或维护人员能够更直观地了解整体机电系统服役期间的健康状况, 从而更便于进行运维管理。

3　结　论

1) 本文提出了一种基于数字孪生的健康服役框架, 该框架通过监测机电系统在不同工况下的表现, 并将这些数据反馈, 不断更新其自身的数字孪生模型, 实现了对机电系统状态的实时伴随和健康评估。通过该框架, 能够有效提升设备的可靠性, 同时降低维护成本和停机时间。

2) 定义了数字孪生健康状态评估体系, 利用健康因子动态更新的数字孪生模型, 综合考虑机电系统在各种工况下的历史数据及其健康状态。这种评估机制使得数字孪生不仅是实时的状态反映器, 更是一个具有全局连续属性的健康管理工具, 其能够为设备维护和性能优化提供科学、精准的数据支持, 从而提高设备的运行效率和可靠性。

3) 以典型的多学科领域中的机电系统永磁同步电机为例, 通过对该电机的实验监测所得的定子电流数据与数字孪生仿真数据进行对比, 进一步定义电机的健康状态指标及相应的维护策略。通过实际案例的深入分析, 充分验证了本文方法在现实应用场景中的有效性。

参考文献（References）

[1] TAO F, ZHANG H, LIU A, et al. Digital twin in industry: state-of-the-art[J]. IEEE Transactions on Industrial Informatics, 2019, 15(4): 2405-2415.

[2] RAJESH P K, MANIKANDAN N, RAMSHANKAR C S, et al. Digital twin of an automotive brake pad for predictive maintenance[C]//Proceedings of the 2nd International Conference on Recent Trends in Advanced Computing. Amsterdam: Elsevier, 2019.

[3] XIE J C, WANG X W, YANG Z J, et al. Virtual monitoring method for hydraulic supports based on digital twin theory[J]. Mining Technology, 2019, 128(2): 77-87.

[4] VENKATESAN S, MANICKAVASAGAM K, TENGENKAI N, et al. Health monitoring and prognosis of electric vehicle motor using intelligent-digital twin[J]. IET Electric Power Applications, 2019, 13(9): 1328-1335.

[5] 曹雪蕊, 张学艺, 彭开香, 等. 基于数字孪生的小卫星健康预测技术[J]. 南京航空航天大学学报, 2022, 54(增刊 1): 35-42.
CAO X R, ZHANG X Y, PENG K X, et al. Small satellite health prediction based on digital twin[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2022, 54(Sup 1): 35-42(in Chinese).

[6] 郭丞皓, 于劲松, 宋悦, 等. 基于数字孪生的飞机起落架健康管理技术[J]. 航空学报, 2023, 44(11): 175-193.
GUO C H, YU J S, SONG Y, et al. Application of digital twin-based aircraft landing gear health management technology[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(11): 175-193 (in Chinese).

[7] MOGHADAM F K, NEJAD A R. Online condition monitoring of floating wind turbines drivetrain by means of digital twin[J]. Mechanical Systems and Signal Processing, 2022, 162: 108087.

[8] WANG Q Y, JIAO W H, ZHANG Y M. Deep learning-empowered digital twin for visualized weld joint growth monitoring and penetration control[J]. Journal of Manufacturing Systems, 2020, 57: 429-439.

[9] LI Q W, JIANG P, LI H. Prognostics and health management of FAST cable-net structure based on digital twin technology[J]. Research in Astronomy and Astrophysics, 2020, 20(5): 67.

[10] LI C Z, MAHADEVAN S, LING Y, et al. Dynamic Bayesian network for aircraft wing health monitoring digital twin[J]. AIAA Journal, 2017, 55(3): 930-941.

[11] GIANNAROS E, KOTZAKOLIOS A, KOSTOPOULOS V, et al. Low- and high-fidelity modeling of sandwich-structured composite response to bird strike, as tools for a digital-twin-assisted damage diagnosis[J]. International Journal of Impact Engineering, 2022, 160: 104058.

[12] AIVALIOTIS P, GEORGOULIAS K, ARKOULI Z, et al. Methodology for enabling digital twin using advanced physics-based modelling in predictive maintenance[J]. Procedia CIRP, 2019, 81: 417-422.

[13] LI Y F, LI M, YAN Z, et al. Application of life cycle of aeroengine mainshaft bearing based on digital twin[J]. Processes, 2023, 11(6): 1768.

[14] OLUWASEGUN A, JUNG J C. The application of machine learning for the prognostics and health management of control element drive system[J]. Nuclear Engineering and Technology, 2020, 52(10): 2262-2273.

[15] CORREA-JULLIAN C, GROTH K M. Opportunities and data requirements for data-driven prognostics and health management in liquid hydrogen storage systems[J]. International Journal of Hydrogen Energy, 2022, 47(43): 18748-18762.

[16] GUO H Y, WANG S P, SHI J, et al. Dynamically updated digital twin for prognostics and health management: application in permanent magnet synchronous motor[J]. Chinese Journal of Aeronautics, 2023, 37(6): 244-261.

Digital twin-based health service framework and its application in permanent magnet synchronous motors

GUO Haoyu[1, 2, 3, 4], WANG Shaoping[1, 3, 4, *], SHI Jian[1, 3, 4], ZHANG Yuwei[1, 3, 4]

(1. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China;
2. Department of Mechanical and Aerospace Engineering, Polytechnic University of Turin, Turin 10129, Italy;
3. Ningbo Institute of Technology, Beihang University, Ningbo 315800, China;
4. Tianmushan Laboratory, Hangzhou 310023, China)

Abstract: The modern equipment development cycle is short, and health service requirements are high; and traditional design methods can no longer meet the needs of equipment development, especially the changes in the health service status of equipment, performance degradation, and the impact of external interference and uncertainty. In order to solve the above problems, this paper focuses on equipment health service and proposes a digital twin-based equipment health service framework, which aims to efficiently use the whole life cycle data to achieve rapid iteration. In order to accurately detect the equipment's current status in real time, the framework can update the digital twin with operating condition data at various times over the equipment's whole service life cycle. This digital twin-based approach not only simulates the service status of equipment in the model, but also enables real-time assessment of the health status of the equipment. Through this method, it is possible to better meet the needs of the development and service of modern equipment. To demonstrate the efficacy and viability of the suggested approach, this study uses the permanent magnet synchronous motor, a widely used electromechanical drive device in the aerospace industry, as an example.

Keywords: digital twin; service conditions; iterative updates; health management; permanent magnet synchronous motors

Received: 2024-02-28; **Accepted**: 2024-05-10; **Published Online**: 2024-08-01 09:03

URL: link.cnki.net/urlid/11.2625.V.20240731.1437.001

Foundation items: National Natural Science Foundation of China (U2233212); Beijing Municipal Natural Science Foundation (L221008); Chinese Government Overseas Study Fund (202106020001)

* **Corresponding author**. E-mail: shaopingwang@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0126

轰炸机挂飞导弹设计突风载荷谱编制方法

卫昆钰[1]，李晨迪[1]，李博文[1]，袁愿[2]，贺小帆[1,*]

(1. 北京航空航天大学 航空科学与工程学院，北京 100191；2. 北京航空工程技术研究中心，北京 100084)

摘 要：突风载荷是轰炸类飞机机腹挂飞导弹疲劳损伤的主要来源，为评定挂飞导弹寿命，需要编制挂飞导弹突风设计载荷谱。对挂飞导弹特点进行阐释，分析挂飞导弹服役突风载荷环境，基于挂飞导弹典型任务剖面对实测突风速度累积超越数曲线进行收集，采用离散突风模型计算突风载荷响应，得到可用于挂飞导弹编谱的突风过载累积超越数曲线族，进行统计分析得到代表平均使用情况的突风过载累积超越数曲线，编制以任务段为基本单位的突风过载5×5谱，由此编制了挂飞导弹飞-续-飞谱作为挂飞导弹突风设计谱，该谱能够反映挂飞导弹的预期使用突风环境，解决了设计阶段缺少实测载荷情况下的挂飞导弹突风谱的编制问题。

关 键 词：挂飞导弹；突风载荷谱；设计谱；5×5谱；突风速度累积超越数曲线

中图分类号：V215.5

文献标志码：A **文章编号**：1001-5965(2026)05-1657-09

轰炸机机腹挂飞导弹(以下简称挂飞导弹)在发射前需多次挂飞训练[1]，在随飞机执行任务的过程中，会受到多种交变载荷的共同作用，从而引发结构疲劳问题，其疲劳问题及寿命问题日益受到重视。在挂飞导弹结构设计阶段，需要进行疲劳分析和疲劳试验来评估寿命。作为疲劳分析和试验的输入，挂飞导弹设计载荷谱的编制具有重要意义。

挂飞导弹作为飞机成品，一直采用振动耐久性的思想进行定寿，将振动载荷谱作为试验和分析的输入进行分析和试验。美国军用规范MIL-M-8856B[2]规定了导弹服役寿命定寿时采用耐久性思想，在发射前的挂飞过程必须考虑振动声载荷、空气动力载荷和惯性载荷。关于挂飞导弹的振动疲劳载荷谱编制问题已有相当多的研究，文献[3]对某模型导弹在某喷气飞机上挂飞时的振动环境，进行了数据处理，分析了振动环境的振源、传递路径、频率结构等特点。文献[4]在激励谱未知的情况下，采用逆虚拟激励方法反求导弹振动谱。

轰炸机飞行速度低，机动动作幅度小、次数少，机械载荷，如突风载荷和机动载荷，对疲劳问题的影响更为显著[5]。文献[6]对挂飞导弹在机动载荷作用下的疲劳寿命分析方法进行了研究，基于国军标GJB67.6A-2008[7]中提供的飞机机动过载，采用任务分析法编制了挂飞疲劳载荷谱，但未考虑突风载荷。事实上，突风载荷是对挂飞导弹疲劳损伤起主要贡献的机械载荷。突风载荷谱编制主要包括2个问题：①突风载荷的实测与计算；②突风载荷谱编制方法。从20世纪30年代起国外航空界就开展了突风载荷的实测与计算，发展了离散突风模型、连续紊流模型[8-9]。目前运输类飞机形成了基于突风速度超越数曲线的标准谱编制方法，其中包括空客公司的TWIST谱[10]和波音公司的的5×5谱[11]。突风载荷分散性大，随机性强[12]，一般采用随机变量模型对实测突风数据进行统计处理并平均化得到，文献[13]对美国联邦航空局(Federal Aviation Administration, FAA)搜集的小飞机速度-突风-高度(velocity-gust-height, VGH)数据进行了统计分析，认为指定过载下超越数服从对数正态分布。

收稿日期：2024-03-05；**录用日期**：2024-05-17；**网络出版时间**：2024-05-27 08:48

网络出版地址：link.cnki.net/urlid/11.2625.V.20240524.0858.001

*通信作者. E-mail：xfhe@buaa.edu.cn

引用格式：卫昆钰，李晨迪，李博文，等. 轰炸机挂飞导弹设计突风载荷谱编制方法[J]. 北京航空航天大学学报，2026，52(5)：1657-1665. WEI K Y，LI C D，LI B W，et al. Developing of tornado missile design gust load spectrum for bomber-mounted air-to-ground missiles[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1657-1665 (in Chinese).

由于结构、位置及基频的不同，突风过载作用下挂飞导弹与飞机响应分析并不相同。MIL-M-8856B[2] 指出可以参考载机挂载导弹时的空气动力系数和过载系数，并计算对应的导弹载荷。但由于大型轰炸类飞机的载荷数据未公开，难以得到载机数据。如何针对挂飞导弹的载荷、结构和使用特点，编制突风载荷谱成为亟待解决的问题。

因此，本文以挂载于轰炸类飞机机腹位置的大质量圆柱形挂飞导弹为研究对象，收集了 20 世纪 70 年代以来实测的多个运输机型号服役期间的突风速度累积超越数曲线，以此为基础计算得到导弹突风过载累积超越数曲线，建立了突风过载累积超越数曲线的分散性模型；采用 5×5 谱法编制挂载导弹突风设计使用过载谱，实现了挂飞导弹服役期间预期承受突风载荷的准确复现。

1 挂飞导弹载荷特点及任务剖面分析

1.1 挂飞导弹特点与载荷分析基础

大型轰炸类飞机挂飞导弹具有尺寸大、质量大、刚度大的特点，这类导弹形状一般为对称圆柱体，长度大于 5 m 之间，直径大于 0.5 m，发射质量超过 4 000 kg。

大质量挂飞导弹一般采用外挂或者半埋式挂载于机腹位置，为减少对轰炸机重心位置以及飞机操纵性能的影响，挂飞导弹重心位置一般与飞机重心位置重合。多采用外挂式或半埋式挂载方式，导弹在挂飞过程中均固连在机身挂架上，可以认为 2 者间不存在相对运动。

导弹在发射前需多次挂飞训练，典型任务剖面下任务段一般包括：起飞滑跑、离场、爬升、巡航、下降、进场、着陆滑跑等。以上任务段导弹均会受到突风载荷，但一般不考虑地面任务段的突风载荷。

1.2 突风设计谱编制方法及任务剖面确定

挂飞导弹突风设计谱编制的流程图如图 1 所示。

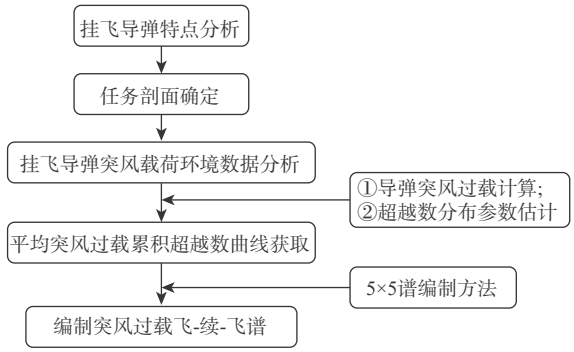


图 1 挂飞导弹突风设计谱编制流程

Fig. 1 Design gust spectrum development process for air-to-ground missile

采用任务分析法作为编制导弹设计谱的总体方法，应根据该型挂飞导弹载机的任务要求，通过聚类分析和归纳法确定载机带弹飞行情况下预期的所有典型任务剖面和具体任务段组成和飞行参数[14-15]。飞行参数主要包括载机带弹质量、燃油质量、飞行高度、飞行距离、当量空速、所在高度空气密度、升力线斜率等。同时应确定飞-续-飞谱基本寿命周期的起落数 n，该值一般取导弹目标寿命的 10%～20%[16]。

2 突风载荷环境数据分析

本文搜集 20 世纪 70 年代以来多个运输类飞机的离散突风速度 U_{de} 超越数数据[17-25]，得到不同海拔高度下的数据信息如表 1 所示[17-24]。离散突风模型认为向上和向下突风出现的概率相同，采用式 (1) 将原始数据中向上和向下突风的超越数合并 (N 为累积超越数)，归并为单位海里的突风速度 U_{de} 累积超越数曲线族，如图 2 所示。

表 1 实测离散突风数据

Table 1 Actual measured discrete gust data

数据来源	飞机类型	飞机数量	飞行次数	总时长/h	总海里数/km
FAA	ERJ-145XR		47 273	88 305	64 351 759
	B777-200ER		10 047	67 000	56 490 728
	A320	56	10 066	30 817	24 173 545
	MD-82/ 83	8	3 978	7 120	4 950 350
	B737-400	17	11 721	19 105	9 713 158
	BE-1900D	28	903	585	252 687
	B767-200ER	10	1 285	9 164	7 035 487
KSSU	B747	52	24 358	121 893	105 096 876
NASA	B727	20	985	1 765	2 080 148

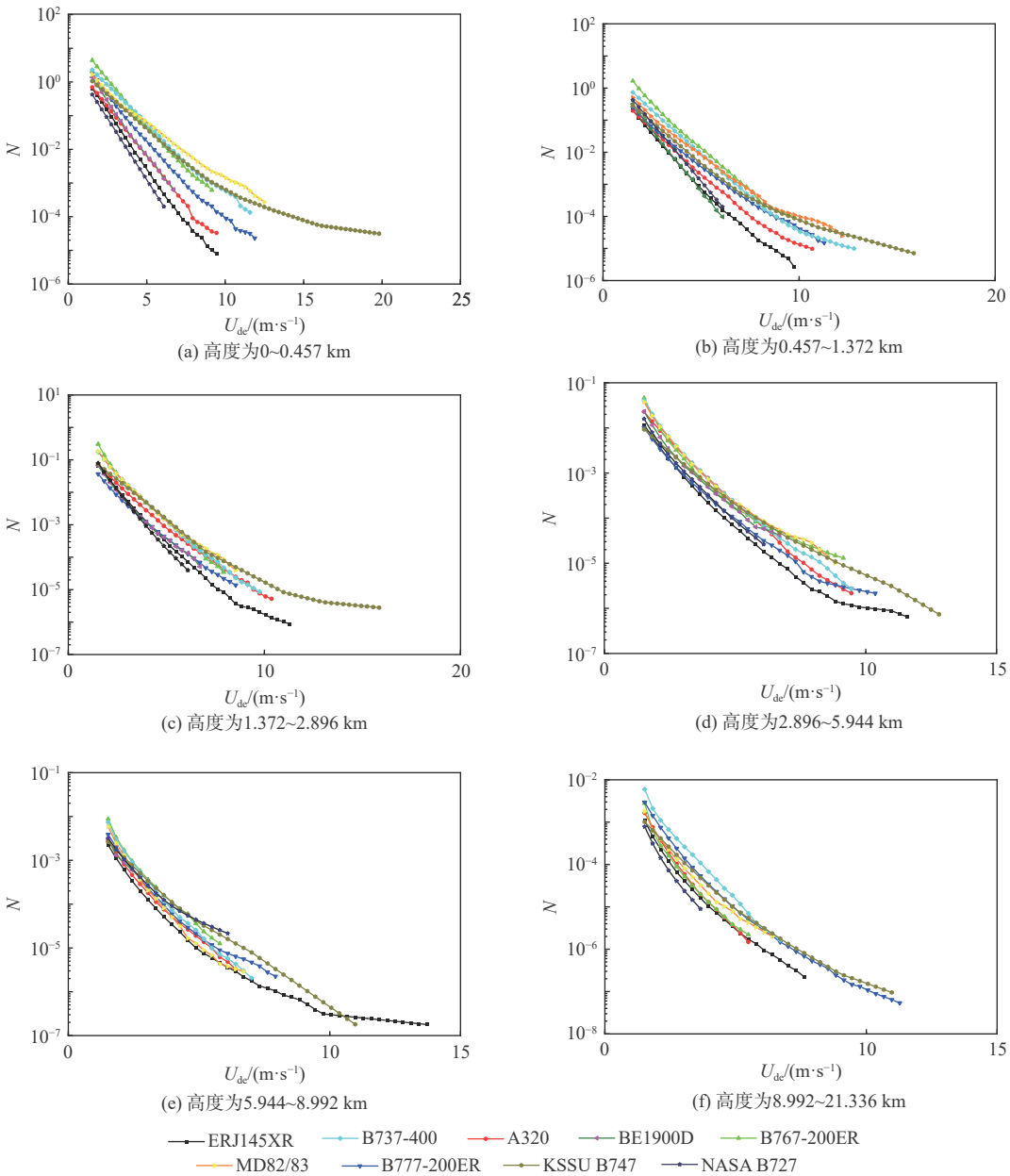


图 2　按高度归并的 U_{de} 实测数据

Fig. 2　Altitude-aggregated U_{de} measured data

$$N(\pm U_{de}) = \sqrt{N(\pm U_{de})N(-U_{de})} \tag{1}$$

式中: 这些数据代表了历史中机群在不同海拔高度下飞行单位海里所遭遇的特定速度突风的平均次数, 且离散突风速度曲线与具体飞机型号无关, 只反映了大气的突风特性。

3　平均突风过载累积超越数曲线

3.1　导弹突风过载计算

3.1.1　载机刚体模型突风过载计算

采用离散突风模型对当量突风速度和载机突风过载进行转换。将机身简化为二维刚体, 假设突风风速远小于飞机飞行当量空速 V, 在突风扰动的作用下只有浮沉运动, 飞机质心受扰动产生的垂直增量过载可表示为

$$\Delta n_{yra} = \ddot{h}/g \tag{2}$$

式中: Δn_{yra} 为刚体模型下飞机突风增量过载; h 为飞机质心的垂直位移; g 为重力加速度, m/s^2。

则飞机的小扰动线化方程为

$$m\ddot{h} = q_\infty S C_L \left(\frac{U_{de}}{V} - \frac{\dot{h}}{V} \right) \tag{3}$$

式中: q_∞为分型动压; S 为机翼面积; C_L 为升力线斜率 (1/rad); U_{de} 为当量垂向突风速度, m/s; V 为飞机当量空速。

式 (3) 为二阶微分方程, 当考虑零初始条件时根据式 (2) 计算飞机 “1-cos” 型 (突风扰动的周期为机翼平均弦长的 25 倍) 离散突风响应最大增量过载为

$$\begin{cases}\Delta n_{\mathrm{yra}}=RU_{\mathrm{de}}\\ R=\dfrac{\rho_0 C_L V}{2G/S}K_{\mathrm{g}}\\ K_{\mathrm{g}}=\dfrac{0.88\mu_{\mathrm{g}}}{5.3+\mu_{\mathrm{g}}}\\ \mu_{\mathrm{g}}=\dfrac{2G/S}{gC_L\rho_h\overline{C}}\end{cases}\tag{4}$$

式中：R为机翼与机身的垂直突风响应参数，s/m；ρ_0为海平面空气密度，kg/ m^3；G/S为翼载，N/m^2；K_{g}为离散突风减缓因子；μ_{g}为飞机质量参数；$\overline{C}$为机翼平均几何弦长，m；ρ_h为所在高度空气密度，kg/m^3。

3.1.2 导弹突风增量过载计算

挂飞导弹的突风增量过载如下：

$$\Delta n_{\mathrm{y}}=K\Delta n_{\mathrm{ym}}\tag{5}$$

式中：Δn_{yrm}为刚体模型下弹体突风增量过载；K为挂飞导弹重心处加速度动响应系数，用于修正载机弹性变形引起的弹体突风增量过载变化。

1）Δn_{yrm}计算

如1.1节所述，可以认为导弹为机体的一部分，并且认为二者的垂向过载相等（垂向载荷是主要疲劳损伤来源），两者误差可以忽略。并且认为二者的垂向过载相等（垂向载荷是主要疲劳损伤来源），见式(6)，两者误差可以忽略。

$$\Delta n_{\mathrm{yrm}}=\Delta n_{\mathrm{yra}}\tag{6}$$

机腹位置挂飞导弹不影响载机的气动外形和升力线斜率，因此，式(4)仍适用于挂载导弹后的载机，但需要修改翼载G/S（挂弹后质量发生变化），以获得导弹的突风过载Δn_{yrm}。

2）K计算

采用NASTRAN软件构建载机和挂飞导弹结构动力学模型，采用偶极子网格法进行计算刚性载机和弹性载机各任务段下挂飞导弹的加速度，挂飞导弹重心处加速度动响应系数K为

$$K=\Delta n_{\mathrm{ye}}/\Delta n_{\mathrm{yr}}\tag{7}$$

式中：Δn_{ye}是弹性飞机模型下的挂飞导弹重心处加速度峰值；Δn_{yr}是刚性飞机模型下的挂飞导弹重心处加速度峰值，弹性飞机和刚性飞机分别指机身、机翼、平尾、垂尾以及弹体定义为弹性和刚性。

将式(6)与式(7)代入式(5)，即可得到挂飞导弹突风增量过载Δn_{y}。

3.2 超越数分布参数估计

将过载累积超越数曲线离散为$(\Delta n_y, N)_i (i=1, 2,\cdots, n)$数据对，得到各级过载对应超越数$\Delta N_i=N_i-N_{i-1}$，对其进行分布特性分析。本文假定指定$\Delta n_y$对应的超越数$\Delta N$服从对数正态分布。将指定$\Delta n_y$下超越数样本按从小到大的顺序排列为$\Delta N_i (i=1, 2,\cdots, n)$，按秩统计方法，$\Delta N_i$对应的概率为

$$P_i=\frac{i-3/8}{n+1/4}\tag{8}$$

式中：i为将ΔN从小到大排列后的第i个样本；n为样本数量。

得到ΔN_i对应的经验概率值P_i及分位点$u_{p,i}$，$(\lg\Delta N_i, u_{p,i})$数据对可线性化为式(9)，将$(\lg\Delta N_i, u_{p,i})$数据对代入该式求解得到对数均值μ_i和对数标准差σ_i。

$$\lg\Delta N=\mu_i+\mu_p\sigma_i\tag{9}$$

3.3 平均突风过载累积超越数曲线确定

以平均突风过载累积超越数曲线反映挂飞导弹平均使用情况。对于每一海拔高度下的各级突风过载，计算覆盖率为50%的超越数，即

$$\Delta N_{50,i}=10^{\mu_i+\mu_{50}\sigma_i}\tag{10}$$

累加得到累积超越数$N_i=\sum_{i=1}^{m}N_{50i}$，采用对数多项式方程拟合各任务段的平均突风过载累积超越数曲线，如下：

$$\lg\left(N\left(\Delta n_{\mathrm{y}}\right)\right)=a_0+a_1\Delta n_y+a_2\Delta n_y^2+a_3\Delta n_y^3\tag{11}$$

4 挂飞导弹突风过载谱编制

4.1 高载截取与低载删除

以挂飞导弹基本寿命周期n次起落遇到一次的高载进行高载截除[16]，将各任务段的突风过载累积超越数曲线方程$N_s(\Delta n_y)$累加后乘以基本寿命周期下的该剖面出现次数n得到基本寿命周期对应的突风过载累积超越数曲线方程（见式(12)），并确定基本寿命周期下出现1次的高载。

$$N_{\mathrm{t}}(\Delta n_y)=n\sum_{s=1}^{5}N_s(\Delta n_y)\tag{12}$$

参考文献[26]取最大载荷幅值的10%～14%作为低载删除载荷。

4.2 5×5谱编制

5×5谱载荷矩阵如表2所示。5×5谱编制需要满足极值载荷对数正态分布准则和载荷相似性准则。

步骤如下：

1）确定各飞行类型出现次数y_i；

2）引入载荷水平比R_i（见式(13)），将过载累积超越数曲线进行离散为4段。

$$R_i=(L_{\mathrm{B}i}-L_{\mathrm{B}i+1})/(L_{\mathrm{B}1}-L_{\mathrm{B}n})\tag{13}$$

式中：$L_{\mathrm{B}i}$为第i段（从右往左数）直线的右端点。

表 2　5×5 谱载荷矩阵

Table 2　5-by-5 spectrum load matrix

飞行类型	出现次数	各级突风过载循环次数					每次飞行合计循环数	每块合计循环数
		Δn_{yd1}	Δn_{yd2}	Δn_{yd3}	Δn_{yd4}	Δn_{yd5}		
A	y_1	1	$B_{1,2}$	$B_{1,3}$	$B_{1,4}$	$B_{1,5}$	$\sum_{i=2}^{5} B_{1,i}$	$\sum_{i=2}^{5} f_1 B_{1,i}$
B	y_2	0	1	$B_{2,3}$	$B_{2,4}$	$B_{2,5}$	$\sum_{i=3}^{5} B_{2,i}$	$\sum_{i=3}^{5} f_2 B_{2,i}$
C	y_3	0	0	1	$B_{3,4}$	$B_{3,5}$	$\sum_{i=4}^{5} B_{3,i}$	$\sum_{i=4}^{5} f_3 B_{3,i}$
D	y_4	0	0	0	1	$B_{4,5}$	$\sum_{i=5}^{5} B_{4,i}$	$\sum_{i=5}^{5} f_4 B_{4,i}$
E	y_5	0	0	0	0	1	1	y_5

采用直线代替该离散段的曲线, 第 i 段的线性方程为

$$\Delta n_y = a_i \lg N + b_i \tag{14}$$

式中: a_i、b_i 为第 i 段载荷谱曲线的拟合常数。

采用当量损伤法求解 5×5 谱下各级载荷及出现次数 $(\Delta n_{yd}, N_{eq})_j (j=1,2,3,4,5)$。假设需进行当量计算的某离散段, 其该级当量过载为 Δn_{yd}, 当量载荷循环次数为 N_{eq}, 则当量载荷为

$$\Delta n_{yd} = \left(-\frac{\ln 10}{n} \times \sum_{i=1}^{m} A_i \right)^{\lg S} \tag{15}$$

式中:

$$A_i = a_i^{-1} 10^{\frac{-b_i}{a_i}} \int_{L_{B_i}}^{L_{B_{i+1}}} \Delta n_{yd}^{\frac{1}{\lg S}} 10^{\frac{\Delta g}{a_i}} \mathrm{d}(\Delta n_{yd}) \tag{16}$$

S 为材料 S-N 曲线斜度参数, 铝合金 S=2.0。

当量载荷循环次数:

$$N_{eq} = -\ln 10 \Delta n_{yd}^{-1/\lg s} \sum_{i=1}^{m} A_i \tag{17}$$

应对载荷水平比 R_i 和 1) 中的各类飞行类型出现次数 y_i 反复进行迭代, 以使各级当量过载 Δn_{yd} 满足极值载荷对数正态分布准则。

3) 根据 2)中 $(\Delta n_{yd}, N_{eq})_j$ 及突风相似性准则确定各飞行类型下各级载荷出现次数 $B_{i,j}$。

4.3　飞-续-飞谱编制

1) 编制任务段谱

将第 j 种剖面第 i 个任务段的第 k 种飞行类型单次起落下各级载荷及其相应频数联系起来, 随机交替选取峰、谷值, 形成任务段特定飞行类型的载荷谱 $\boldsymbol{f}_{i,j,k}$ 记为 $(\Delta n_{ydn}, -\Delta n_{ydm})_h$, 并把该任务段谱表示为以下的载荷对序列的形式:

$$\boldsymbol{f}_{i,j,k} = (\Delta n_{ydn}, -\Delta n_{ydm})_h \tag{18}$$

2) 编制任务剖面谱

记第 $j(j=1,2,\cdots,J)$ 种剖面的第 $k(k=1,2,\cdots,K)$ 种飞行类型的载荷谱为 $F_{j,k}$。表示为 $\boldsymbol{f}_{i,j,k}$ 按该剖面下的任务段顺序 $i(i=1,2,\cdots,m)$ 进行排序。按上述方法编制所有任务剖面下各种飞行类型的任务剖面谱。将所有种类的任务剖面谱用向量形式表示为

$$\boldsymbol{B} = \left[F_{1,1}, \cdots, F_{1,K}, F_{2,1}, \cdots, F_{2,K}, \cdots, F_{J,1}, \cdots, F_{J,K} \right] \tag{19}$$

3) 编制飞-续-飞载荷谱

记 n 次飞行下第 j 种剖面 $(j=1,2,\cdots,J)$ 第 k 种飞行类型 $(k=1,2,\cdots,K)$ 出现 $y_{j,k}$ 次。将所有的 $y_{j,k}$ 用向量形式表示为

$$\boldsymbol{A} = \left[y_{1,1}, \cdots, y_{1,K}, y_{2,1}, \cdots, y_{2,K}, \cdots, y_{J,1}, \cdots, y_{J,K} \right] \tag{20}$$

$\boldsymbol{A}$ 向量和 $\boldsymbol{B}$ 向量是一一对应的关系, 表示该任务剖面谱在单位块谱中共出现多少次。则总谱 $\boldsymbol{G}$ 可以表示为

$$\boldsymbol{G} = \boldsymbol{A} \cdot \boldsymbol{B} = \left[y_{1,1} \times F_{1,1}, \cdots, y_{1,K} \times F_{1,K}, \cdots, y_{J,K} \times F_{J,K} \right] \tag{21}$$

将 $\boldsymbol{G}$ 中每次完整飞行下的任务剖面谱重新排序, 即可得到最终的飞-续-飞谱。

5　示　例

5.1　挂飞导弹载机任务剖面

设某挂飞导弹载机 (基本参数列于表 3) 典型任务剖面各任务段飞行参数如表 4 所示。

表 3　载机参数

Table 3　Aircraft parameters

翼展/m	长度/m	高度/m	机翼面积/m^2	挂飞导弹空质量/kg	机翼平均几何弦长/m
20.4	21.0	7.3	51.2	5 000	2.5

本例中, 该型挂飞导弹目标寿命为 1 000 次挂飞起落, 编制以 200 次起落为单位的飞-续-飞谱。

5.2　平均突风过载累积超越数曲线获取

将表 4 各参数代入式 (4) 中, 得到不同任务段 R 值, 分别为 0.046 9、0.059 8、0.075 7、0.074 2、0.063 3。

表 4 任务段飞行参数

Table 4 Mission segment flight parameters

任务段	载机带弹总质量/kg	高度/m	飞行距离/km	当量空速/(m·s^{-1})	所在高度空气密度/(kg·m^{-3})	升力线斜率
离场	21 604	265.2	6.3	92.5	1.215 0	4.20
爬升	19 737	1 071.3	93.0	102.7	1.112 0	4.43
平飞	18 253	2 439.1	186.7	115.6	0.956 1	4.61
下降	16 241	987.5	89.9	107.0	1.103 0	4.41
进场	16 052	245.7	5.1	95.5	1.207 0	4.18

采用 NASTRAN 求解导弹突风动响应系数，建模时将机身、机翼、尾翼和机身中部外挂物分别简化为单梁，采用广义元与机身连接。机翼、平尾、垂尾的平板气动力模型采用非定常气动力的理论方法建立，机身采用十字交叉气动面[27]，用偶极子涡格法计算刚性载机和弹性载机各任务段下挂飞导弹重心位置的加速度，代入式 (7) 可得动态响应系数 K。按照表 4 要求的空中飞行情况进行离散突风计算，得到不同任务段机腹位置挂飞导弹的动响应系数 K，分别为 1.01、0.97、0.95、0.97、1.0。

采用 3.1.2 节中的方法将突风速度累积超越数曲线转换为不同任务段的导弹突风过载累积超越数曲线。

采用 3.3 节中的方法拟合得到各任务段的平均突风过载累积超越数曲线，如图 3 所示。

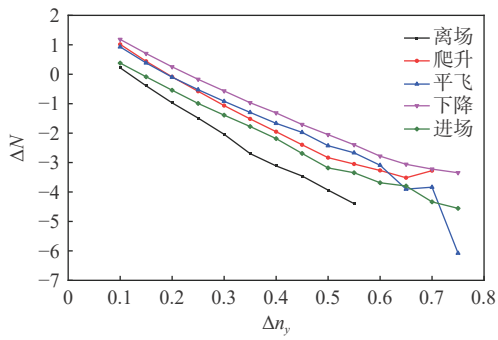


图 3 各任务段的平均突风过载累积超越数曲线

Fig. 3 Mean gust vertical acceleration accumulated exceedance curves for mission segments

5.3 挂飞导弹突风过载谱编制

高载截取水平为 0.627 5 g，低载截除水平为 0.09 g。各任务段的 5×5 谱如表 5～表 9 所示。

表 5 离场任务段突风过载 5×5 谱

Table 5 Gust vertical acceleration 5-by-5 spectrum of outgoing mission segment

飞行类型	次数	各级突风过载循环次数					每次飞行循环次数
		0.340 7g	0.261 1g	0.215 1g	0.167 0g	0.120 7g	
A	1	1	1	1	1	2	6
B	5	0	1	1	1	2	5
C	24	0	0	1	1	3	5
D	50	0	0	0	1	2	3
E	120	0	0	0	0	2	2

表 6 爬升任务段突风过载 5×5 谱

Table 6 Gust vertical acceleration 5-by-5 spectrum of climbing mission segment

飞行类型	次数	各级突风过载循环次数					每次飞行循环次数
		0.486 1g	0.382 3g	0.329 3g	0.267 2g	0.215 6g	
A	1	1	2	2	4	7	16
B	5	0	1	1	3	5	10
C	24	0	0	1	1	3	5
D	50	0	0	0	1	3	4
E	120	0	0	0	0	3	3

表 7　平飞任务段突风过载 5×5 谱

Table 7　Gust vertical acceleration 5-by-5 spectrum of level-flying mission segment

飞行类型	次数	各级突风过载循环次数					每次飞行循环次数
		0.521 1g	0.393 2g	0.324 9g	0.262 2g	0.203 6g	
A	1	1	3	6	11	17	38
B	5	0	1	4	7	13	25
C	24	0	0	1	4	8	13
D	50	0	0	0	1	3	4
E	120	0	0	0	0	1	1

表 8　下降任务段突风过载 5×5 谱

Table 8　Gust vertical acceleration 5-by-5 spectrum of descending mission segment

飞行类型	次数	各级突风过载循环次数					每次飞行循环次数
		0.594 0g	0.507 0g	0.376 4g	0.284 9g	0.192 0g	
A	1	1	3	6	24	80	114
B	5	0	1	4	10	45	60
C	24	0	0	1	3	11	15
D	50	0	0	0	2	7	9
E	120	0	0	0	0	5	5

表 9　进场任务段突风过载 5×5 谱

Table 9　Gust vertical acceleration 5-by-5 spectrum of approach mission segment

飞行类型	次数	各级突风过载循环次数					每次飞行循环次数
		0.443 5g	0.322 6g	0.256 9g	0.201 6g	0.148 1g	
A	1	1	1	2	3	7	14
B	5	0	1	1	3	5	10
C	24	0	0	1	1	3	5
D	50	0	0	0	1	3	4
E	120	0	0	0	0	3	3

挂飞导弹的飞-续-飞谱如图 4 所示。

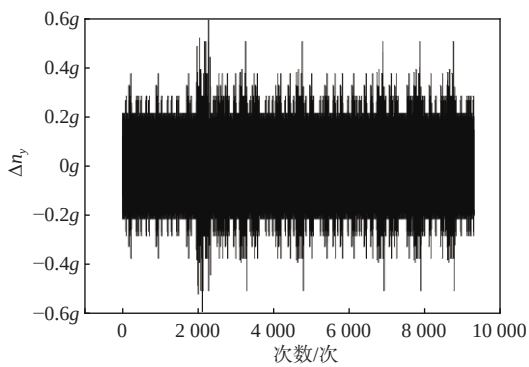


图 4　挂飞导弹飞-续-飞谱

Fig. 4　Flight-by-flight spectrum of the mounted missiles

6　结　论

本文针对轰炸机挂飞导弹, 建立了设计突风载荷谱编制方法, 结论如下:

1) 本文基于搜集得到的多个运输机型号服役期间的突风速度累积超越数曲线历史数据, 针对大型轰炸类飞机挂飞导弹特点, 建立由突风速度到导弹突风过载的计算模型, 得到适用于挂飞导弹的广域海拔高度下的突风过载累积超越数曲线。

2) 本文基于任务分析法和 5×5 谱法, 发展一种挂飞导弹的突风载荷谱编制方法。实现了挂飞导弹服役期间预期承受突风载荷的复现, 对于挂飞导弹寿命评估具有重要意义。

3) 本文结合某型挂飞导弹典型任务剖面, 进行了设计突风载荷谱的编制, 验证了本文编谱方法的有效性和合理性。

参考文献（References）

[1] 许胜刚, 陆家乐, 李劲. 机载导弹贮存寿命评估方法研究[J]. 电子产品可靠性与环境试验, 2023, 41(1): 1-4.
XU S G, LU J L, LI J. Study on evaluation method of airborne missile storage life[J]. Electronic Product Reliability and Environmental Testing, 2023, 41(1): 1-4(in Chinese).

[2] U.S. Department of Defense. Missiles, guided, structural integrity, general specification for: MIL-M-8856B[S]. Washington, D.C.: Department of Defense, 1990.

[3] 张曾昌, 周百里. 飞机外挂挂飞振动环境[C]//第六届全国振动理论与应用学术会议论文集. 武汉: 华中理工大学, 1997: 460-464.
ZHANG Z C, ZHOU B L. Aircraft external store launch vibration environment[C]//Proceedings of the 6th National Conference on Vibration Theory and Applications. Wuhan: Huazhong University of Science and Technology, 1997: 460-464(in Chinese).

[4] 吕卫民, 肖阳, 方登建. 基于逆虚拟激励法的导弹振动谱设计[J]. 北京航空航天大学学报, 2017, 43(5): 872-879.
LYU W M, XIAO Y, FANG D J. Design of missile vibration spectrum based on inverse pseudo-excitation method[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(5): 872-879(in Chinese).

[5] 高阳. 导弹挂飞载荷计算与软件开发[D]. 北京: 北京航空航天大学, 2014.
GAO Y. Calculation of missile launching load and software development[D]. Beijing: Beihang University, 2014(in Chinese).

[6] 邓劭廷, 薛景丹, 吴寒, 等. 机载弹药挂飞疲劳寿命研究[J]. 智能制造, 2023(S1): 30-36.
DENG S T, XUE J D, WU H, et al. Study on fatigue life of airborne ammunition in hanging flight[J]. Intelligent Manufacturing, 2023(S1): 30-36(in Chinese).

[7] 中国人民解放军总装备部. 军用飞机结构强度规范 第6部分: 重复载荷、耐久性和损伤容限: GJB 67.6A—2008[S]. 北京: 中国人民解放军总装备部, 2008.
PLA General Armaments Department. Military airplane structural strength specification part 6: Repeated loads, durability and damage tolerance: GJB 67.6A—2008[S]. Beijing: PLA General Armaments Department, 2008(in Chinese).

[8] BENOY M B, GRAYLEY M E, LAMBERT J A B, et al. Average gust frequencies, Subsonic transport aircraft: ESDU 69023D[R]. London: ESDU, 1989: 13-15.

[9] PRESS H, STEINER R. An approach to the problem of estimating severe and repeated gust loads for missile operations: NACA-TN-4332[R]. Langley Field: National Advisory Committee for Aeronautics, 1958: 29-45.

[10] SIMPSON A R, CHEVERTON K E, GROUCH D, et al. Standard fatigue loading sequences: ESDU 97018[R]. London: ESDU, 1997: 18-22.

[11] FOWLER K R, WATANABE R T. Development of jet transport airframe fatigue test spectra[C]//Proceedings of the Development of Fatigue Loading Spectra. Philadelphia: American Society for Testing and Materials, 1989: 36-64.

[12] 朱俊贤, 贺小帆, 李玉海. 阵风载荷谱的损伤分散性研究[J]. 北京航空航天大学学报, 2021, 47(10): 2129-2137.
ZHU J X, HE X F, LI Y H. Damage variation of gust load spectra[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(10): 2129-2137(in Chinese).

[13] REYER W M. Probability basis of safe-life evaluations in small airplanes[C]//Proceedings of the 9th Joint FAA/DoD/NASA Aging Aircraft Conference. Washington, D. C. : Federal Aviation Administration, 2006.

[14] 杜宇飞, 孙志刚, 陆琪, 等. 基于任务段的航空发动机载荷谱聚类方法[J]. 航空动力学报, 2019, 34(5): 987-996.
DU Y F, SUN Z G, LU Q, et al. Aeroengine load spectrum clustering method based on mission segment[J]. Journal of Aerospace Power, 2019, 34(5): 987-996(in Chinese).

[15] 蒋祖国. 飞机结构载荷/环境谱[M]. 北京: 电子工业出版社, 2012: 153-161.
JIANG Z G. Aircraft structural load/environment spectrum[M]. Beijing: Publishing House of Electronics Industry, 2012: 153-161(in Chinese).

[16] DE JONG J B, SCHUTZ D, LOWAK H, et al. A standardized load sequence for flight simulation tests on transport aircraft wing structures: MR_TR_73029_U[R]. Amsterdam: National Aerospace Laboratory NLR, 1973.

[17] RUSTENBURG J W, SKINN D A, TIPPS D O. Statistical loads data for the Airbus A320 aircraft in commercial operations: DOT/FAA/AR-02/35[R]. Springfield: FAA, 2002: 74-83.

[18] RUSTENBURG J W, SKINN D A, TIPPS D O. Statistical loads data for the Boeing 737-400 aircraft in commercial operations: DOT/FAA/AR-98 /28[R]. Springfield: FAA, 1998: 59-69.

[19] TIPPS D O, RUSTENBURG J W, SKINN D A. Statistical loads data for the B767-200ER aircraft in commercial operations: DOT/FAA/AR-00/10[R]. Springfield: FAA, 2000: 53-60.

[20] TIPPS D O, SKINN D A, RUSTENBURG J W, et al. Statistical loads data for the Boeing 777-200ER aircraft in commercial op erations: DOT/FAA/AR-06/11[R]. Springfield: FAA, 2006: 74-81.

[21] TIPPS D O, SKINN D A, RUSTENBURG J W. Statistical loads data for the BE1900D aircraft in commercial operations: DOT/FAA/AR-00/11[R]. Springfield: FAA, 2000: 57-59.

[22] JONES T, RUSTENBURG J W, SKINN D A. Statistical loads data for the Embraer-145XR aircraft in commercial operations: DOT / FAA/AR-07/61[R]. Springfield: FAA, 2007: 64-70.

[23] SKINN D A, TIPPS D O, RUSTENBURG J W. Statistical loads data for MD82 /83 aircraft in commercial operations: DOT/FAA/AR-98/65[R]. Springfield: FAA, 1999: 51-57.

[24] HOUBOLT J C. Manual on the flight of flexible aircraft in turbulence[M]. London: Specialised Printing Services Limited, 1991: 34-36.

[25] CRABILL N L. The NASA digital VGH program. Exploration of methods and final results. Volume 3: B 727 data 1978-1980: 1765 hours: DOT/FAA-CT-89/36-3[R]. Hampton: Eagle Engineering, Incorporated, Hampton Division, 1989: 110-120.

[26] 姚卫星. 结构疲劳寿命分析[M]. 北京: 国防工业出版社, 2003: 86-87.
YAO W X. Fatigue life prediction of structures[M]. Beijing: National Defense Industry Press, 2003: 86-87(in Chinese).

[27] 吴元太子, 杨红杰. 大展弦比飞机机翼及外挂物阵风动响应研究[J]. 中国高新科技, 2020(24): 78-79.
WU Y T Z, YANG H J. Research on gust response of high-aspect-ratio wing and stores[J]. China High and New Technology, 2020(24): 78-79(in Chinese).

Developing of tornado missile design gust load spectrum for bomber-mounted air-to-ground missiles

WEI Kunyu[1]，LI Chendi[1]，LI Bowen[1]，YUAN Yuan[2]，HE Xiaofan[1,*]

(1. School of Aeronautics and Astronautics，Beihang University，Beijing 100191，China；
2. Beijing Aeronautical Technology Research Center，Beijing 100084，China)

Abstract：The gust load is the primary source of fatigue damage for air-to-ground missiles carried on the fuselage of bomber-type aircraft. To assess the lives of these missiles, it is necessary to develop the gust load spectrum for air-to-ground missile launches. This study provided an explanation of the characteristics of air-to-ground missiles and analyzed the gust load environment during their service. Based on the typical mission profiles of air-to-ground missiles, discrete gust speed exceedance curves were collected from measured data. A discrete gust model was applied to calculate gust load responses, resulting in a family of gust vertical acceleration cumulative exceedance curves applicable to a wide range of altitudes for air-to-ground missile spectrum development. A representative curve for typical usage situations was obtained by statistical analysis. A flight-by-flight spectrum for air-to-ground missile launch was created from a 5-by-5 spectrum of gust vertical acceleration based on mission segments. This spectrum presented herein serves as an indicative reflection of the anticipated gust environment for air-to-ground missiles. This study addresses the challenge of formulating the gust spectrum for air-to-ground missiles during the design phase, particularly in situations where measured loads are lacking.

Keywords：air-to-ground missile；gust load spectrum；design spectrum；five by five spectrum；gust velocity cumulative exceedance curves

Received：2024-03-05；**Accepted**：2024-05-17；**Published Online**：2024-05-27 08：48
URL：link.cnki.net/urlid/11.2625.V.20240524.0858.001

* **Corresponding author**. E-mail：xfhe@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0054

航空柱塞泵缸体倾覆行为仿真及试验研究

谷建宁[1, 2, *]，张小波[1]，夏天翔[1]，王天照[3]

(1. 航空工业南京机电液压工程研究中心，南京 211106；2. 航空机电系统综合航空科技重点实验室，南京 211106；
3. 浙江大学 流体动力基础件与机电系统全国重点实验室，杭州 310027)

摘　　要：航空柱塞泵是航空液压系统的核心，其内部零部件倾覆过大可能会使整个系统性能下降，甚至失效。如何提高航空柱塞泵的稳定性、减小零部件倾覆一直是其高压高速化发展研究的关键问题。为此，从缸体振动位移角度出发，建立一种基于数据驱动的高保真动力学模型，探究柱塞泵出口压力和转速对缸体倾覆的影响，提出减小配流盘腰形槽尺寸的改进措施，并通过数值仿真进行验证。研究结果表明：减小配流盘腰形槽尺寸能使缸体倾覆有所改善；修正后的模型能够反映实际的倾覆状态，与试验测量结果相对误差在5%以内。研究成果为航空柱塞泵的仿真和稳定性研究提供了研究思路，为批产一致性制造提供理论和试验支撑。

关　键　词：航空柱塞泵；缸体倾覆；振动位移；联合仿真；模型修正

中图分类号：TH137; V233.91

文献标志码：A　　**文章编号**：1001-5965(2026)05-1666-14

航空柱塞泵能够将发动机或电机输出的机械能转化为液压能，从而帮助实现飞行姿态调整、起落架收放和刹车等功能，是航空液压系统的“心脏”[1-2]，其具有结构紧凑、质量轻、单位功率体积小、功重比高等优点，在航空领域具有重要且广泛的应用[3-4]。在航空柱塞泵中，最重要的3个摩擦副包括配流副、柱塞副和滑靴副，任一摩擦副的失效都可能引起整个柱塞泵性能的下降和失效[5-6]。其中，旋转的缸体既与固定的配流盘组成配流副，又与不断运动的柱塞组成柱塞副，是整个泵完成吸排油最关键的零部件。随着工业技术的不断进步和计算机仿真技术的发展，航空柱塞泵逐渐向着高压化、高速化、低脉动、高可靠性和智能节能的方向发展[7]；现有液压系统的最佳经济工作压力也已经从35 MPa逐渐向40 MPa发展[8]。当柱塞泵的出口压力增加时，柱塞泵内的摩擦接触件将面临更加严苛的工作环境，如受力情况更严酷、倾覆程度增大、泄漏量增加等。与其他零部件相比，当缸体产生较严重的倾覆时，会对柱塞泵造成较大破坏，导致其在短时间内快速失效。因此，对航空柱塞泵缸体的倾覆行为开展试验和仿真研究具有重要的工程意义和应用价值。

近年来，国内外学者对柱塞泵缸体倾覆行为的研究已经取得了一定的成果。在理论分析和仿真研究方面，Manring[9]通过理论分析研究了柱塞泵中缸体的倾覆现象，用试验结果验证了理论分析，并在此基础上，提出一种基于缸体与配流盘之间反作用力质心位置的倾覆标准。Bergada等[10]在分析配流副油膜厚度和动力学特性的基础上，提出一种新方程来计算压力分布和泄漏流量，并表明油压和油温对油膜厚度、振动、偏心和倾覆等行为有显著影响。Zhang等[11]在建立弹流润滑数值模型的基础上，分析了缸体倾斜惯性矩对缸体微动和油膜厚度的影响，结果表明，惯性矩对缸体微动和油膜厚度分布有显著影响，惯性矩增大会导致微运动振幅增大，油膜厚度波动更严重。在试验测试方面，胡骁等[12]开展配流副油膜测试试验后，将缸体和配流盘实测

收稿日期：2024-01-22；录用日期：2024-04-02；网络出版时间：2024-04-24 20:01
网络出版地址：link.cnki.net/urlid/11.2625.V.20240423.1829.002
基金项目：航空科学基金 (20182876003)
*通信作者. E-mail：www.1198@foxmail.com

*引用格式：谷建宁，张小波，夏天翔，等. 航空柱塞泵缸体倾覆行为仿真及试验研究 [J]. 北京航空航天大学学报，2026，52（5）：1666-1679.
GU J N，ZHANG X B，XIA T X，et al. Simulation and experimental research for overturning behavior of aviation plunger pump cylinder bodies [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（5）：1666-1679（in Chinese）.*

配合面轮廓导入配流副数值计算模型中开展研究, 结果表明, 考虑局部接触模型比全油润滑模型预测的油膜厚度更精准。Wegner 等[13] 搭建试验台, 测量缸体在各个自由度的运动及缸体与配流盘之间的摩擦力矩, 并分析工作压力和转速对缸体运动的影响。闫伟鹏[14] 在配流盘的密封带之外开设传感器通道, 采用微型电涡流位移传感器测量缸体在不同压力、转速和温度下的轴向位置变化, 并运用多场耦合理论构建配流副流-固-热数值耦合计算模型, 得到与试验相同的结论, 揭示了油膜与缸体和配流盘的相互作用机理。Xu 等[15] 提出一种较容易实现、对泵改动较少的测量方案, 并通过试验对不同工作压力下缸体的动态特性进行了详细分析。

综上所述, 对于柱塞泵的缸体倾覆行为, 在理论研究、仿真分析和试验测量等方面均已取得了一定进展, 但大多数研究都是通过配流副的油膜厚度和油膜压力分布等来间接反映缸体倾覆行为, 对缸体本身的倾覆行为研究较少; 且目前通过试验研究的柱塞泵大多是 20 MPa 左右的柱塞泵, 对具有更高功重比的 30 MPa 左右的柱塞泵研究较少。此外, 目前对能够高效且高保真地计算缸体倾覆行为的仿真模型研究还不够成熟。缸体是柱塞泵中连接柱塞和配流盘的核心部件, 其倾覆行为过大时, 会使缸体与配流盘间的流体连通瞬间丧失, 从而导致柱塞泵的瞬时故障, 甚至永久故障。因此, 对柱塞泵缸体倾覆行为的仿真及试验研究具有重要意义。

本文以某型 35 MPa 航空柱塞泵为研究对象, 基于系统仿真和多刚体动力学仿真的联合仿真建立航空柱塞泵主轴-缸体系统刚柔耦合动力学模型, 并在模型中加入传感器简化模型, 通过仿真计算, 初步得到缸体在额定工况时的振动位移。通过设计并开展试验, 在对柱塞泵改动较小的情况下, 利用电涡流位移传感器对缸体 2 个平行截面进行测量, 得到缸体实际的振动位移。通过对试验数据的整理和理论分析, 探究柱塞泵出口压力和转速对缸体倾覆的影响, 并提出一种减小缸体倾覆行为的改进方法。根据试验数据对仿真模型进行修正, 得到相对误差在 5% 以内的高保真仿真模型。本文研究结果有助于柱塞泵抗倾覆稳定性研究, 为柱塞泵的设计和发展提供了一定思路。

1　基于联合仿真的刚柔耦合动力学模型的建立

1.1　工作原理

典型柱塞泵的结构示意图如图 1 所示。传动轴与主轴通过花键连接, 主轴通过花键与缸体连接。柱塞均匀地分布在缸体柱塞孔中, 滑靴安装在柱塞头部, 并在卡盘的限制下贴在斜盘表面。缸体另一端与固定的配流盘形成配流副。在工作过程中, 发动机或电机带动传动轴旋转, 传动轴带动主轴旋转, 进而通过花键带动缸体旋转。旋转过程中, 柱塞在滑靴和卡盘的作用下由下死点(对应 180°位置)开始伸出。随着柱塞腔容积增大, 腔内压力不断减小, 油液进而通过配流盘吸油窗口进入柱塞腔。到达上死点(对应 0°位置)时, 在有一定倾角的斜盘作用下, 柱塞开始缩回腔内。随着柱塞腔容积减小, 油液被压缩, 压力升高, 并通过压油窗口排出。随后, 再次进入吸油过程, 开始循环[16-17]。

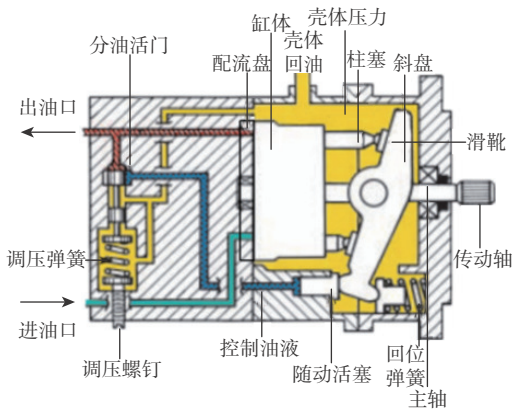


图 1　典型柱塞泵结构示意图

Fig. 1　Schematic diagram of typical plunger pump structure

从柱塞泵的工作原理可以看出, 油液压力在吸排油过程中是不断变化的。因此, 缸体上始终存在一定的倾覆力矩使其产生倾覆, 从而影响主轴-缸体系统在运转过程中的稳定性和平衡性[18]。此外, 由于吸油区和排油区的液压力不同, 缸体与配流盘间的油膜往往是楔形油膜。当柱塞泵设计不合理或工况较为严苛导致缸体受到的倾覆力矩较大时, 楔形油膜的斜率增加, 缸体的倾覆程度增大, 从而产生偏磨和侧翻等现象。这时, 配流副密封效果降低, 泄漏量增加, 柱塞泵的性能下降, 甚至无法正常工作。因此, 通过测量缸体 2 个平行截面的振动位移对其倾覆情况进行分析具有十分重要的意义和价值。

1.2　航空柱塞泵主要部件的受力分析及运动分析

柱塞泵中的柱塞滑靴组件是起着连接作用的关键组件, 既随着缸体一起转动, 又在斜盘胁迫下作往复直线运动使柱塞腔容积改变, 从而实现吸排油功能。因此, 柱塞滑靴组件是受力情况和运动状态最为复杂的组件, 对受力更加恶劣的压油区组件进行受力分析和运动分析有助于对缸体倾覆的研究。如图 2 所示, 对初始位于上死点位置的柱塞滑靴组件进行受力分析, 其主要受作用在柱塞底部的轴向液压力、由向心加速度产生的离心反力、由斜

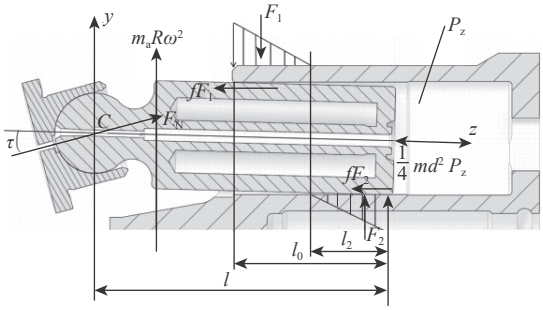


图2 柱塞受力分析示意图

Fig. 2 Schematic diagram of plunger force analysis

盘产生的斜盘支反力、柱塞与柱塞腔壁之间的接触力、柱塞与柱塞腔壁间的摩擦力及组件作往复直线运动时产生的惯性力(由于方向不确定,图中未标出)。在工作过程中,当柱塞滑靴组件处于平衡状态时,其力平衡方程组为[19]

$$\begin{cases} F_N\sin\tau - F_1 + F_2 + m_aR\omega^2 = 0 \\ F_N\cos\tau - f(F_1+F_2) - (\pi d^2P_z)/4 - \\ \quad m_aR\omega^2\tan\tau\cos\varphi = 0 \\ \left[F_1(6l-2l_2-4l_0-3fd) - F_2(6l-2l_2+3fd)\right]/ \\ \quad 6 - m_aR\omega^2l_a = 0 \end{cases} \tag{1}$$

式中:d为柱塞直径;P_z为柱塞腔压力;m_a为柱塞和滑靴的总质量;R为柱塞分布圆半径;ω为缸体旋转角速度;τ为斜盘倾角;φ为缸体旋转角;f为受对偶材料特性影响的摩擦因数;l_0为柱塞在缸体孔中的最小接触长度;l为柱塞名义长度;l_a为柱塞滑靴组件重心到球心C的距离;l_2为柱塞端部在缸体孔内的接触长度;F_N为斜盘支反力;F_1和F_2为柱塞与柱塞腔壁之间的接触力。

由力分布三角形的相似原理可推知

$$F_1/F_2 = (l_0-l_2)^2/l_2^2 \tag{2}$$

由式(1)和式(2)可推出F_N、F_1和F_2的值。本文主要研究缸体倾覆行为,故还需对缸体端面进行受力分析。缸体在其端面主要受柱塞孔中液体产生的轴向压紧力F_P和配流盘对其液压支承力F_s。

轴向压紧力F_P与柱塞泵中的柱塞数z相关。在一般的柱塞泵中,为了降低压力脉动,柱塞数通常设计为奇数。因此,F_P在缸体旋转的前半周期和后半周期不尽相同。为方便计算,通常用平均压紧力F_{Pm}来表征,具体表达式为

$$F_{Pm} = \frac{1}{8}z\pi d^2P_z \tag{3}$$

配流盘的液压支承力由外密封带液压支承力、内密封带液压支承力和压油窗口高压油的液压支承力3部分组成。其中,外密封带中的间隙流动为源流流动,而内密封带中的间隙流动为汇流流动。则当密封带半径为r处的压力为P时,外密封带中的压力分布$p = P_d\ln(R_4/r)/\ln(R_4/R_3)$,内密封带中的压力分布$p = P_d\ln(r/R_1)/\ln(R_2/R_1)$,其中,$P_d$为压油窗口的压力,$R_1$、$R_2$和$R_3$、$R_4$分别为内密封带内、外径和外密封带内、外径[20]。假设配流盘上压力区包角为ϕ,分别对3部分的液压支承力进行积分并求和,可以推出F_s的具体表达式为

$$F_s = \int_{R_2}^{R_3}P_d\phi r\mathrm{d}r + \int_{R_1}^{R_2}P_d\frac{\ln(r/R_1)}{\ln(R_2/R_1)}\phi r\mathrm{d}r + \int_{R_3}^{R_4}P_d\frac{\ln(R_4/r)}{\ln(R_4/R_3)}\phi r\mathrm{d}r \tag{4}$$

进一步求解,给出F_s的表达式为

$$F_s = \frac{\phi}{4}\left[\frac{R_4^2-R_3^2}{\ln(R_4/R_3)} - \frac{R_2^2-R_1^2}{\ln(R_2/R_1)}\right]P_d \tag{5}$$

对柱塞滑靴组件进行运动分析。对于柱塞,其一方面与缸体一起旋转,绕缸体轴线作圆周运动,另一方面又相对缸体作往复直线运动。因此,柱塞轴线上任一点的运动轨迹均呈椭圆状。如图3所示,以上死点位置为起点,当缸体旋转角为φ时,柱塞位移s可表示为

$$s = R(1-\cos\varphi)\tan\tau \tag{6}$$

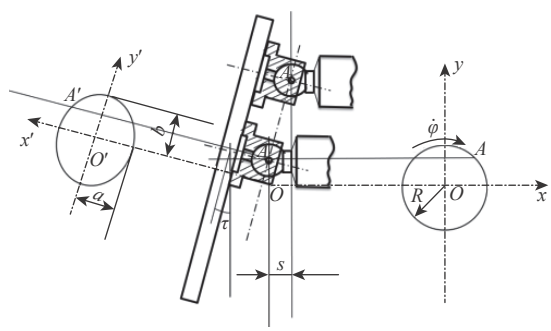


图3 柱塞滑靴组件运动分析简图

Fig. 3 Schematic diagram of piston slide boot assembly motion analysis

将式(6)对时间求导,即可得到柱塞的速度和加速度表达式。对于滑靴,其主要是在斜盘平面上作相对滑动。假设滑靴球窝中心A在缸体平面上的坐标为$x = R\sin\varphi$,$y = R\cos\varphi$,则其在斜盘平面上对应位置的坐标为$x' = R\sin\varphi$,$y' = (R\cos\varphi)/\cos\tau$。说明滑靴相对于斜盘平面的运动轨迹为椭圆,半短轴为R,半长轴为$R/\cos\tau$。

根据航空柱塞泵主要部件的受力分析及运动关系,进一步建立整泵的系统仿真模型、主轴-缸体系统的多刚体动力学仿真模型及刚柔耦合动力学仿真模型,以通过仿真的方式对缸体振动位移进行计算。

1.3 航空柱塞泵系统仿真模型的建立与验证

航空液压系统所采用的柱塞泵大都是恒压变量式柱塞泵,其可以在整个流量范围内提供几乎恒定的压力,且能够根据负载端的需求进行流量调

节, 具有高效、节能等特点。根据航空柱塞泵的调压原理及主要部件的受力分析和运动分析, 利用 Amesim 软件建立航空柱塞泵系统仿真模型, 如图 4 所示, 主要包括柱塞模型、配流盘和缸体吸排油模型及随动活塞和回位弹簧模型等。通过该模型可以对航空柱塞泵的压力流量特性、压力脉动特性和斜盘力矩等动力学特性进行分析。而压力流量特性能够反映泵的整体调压性能, 且较易通过试验进行测量, 因此, 通过柱塞泵的压力流量特性曲线对系统仿真模型进行验证。

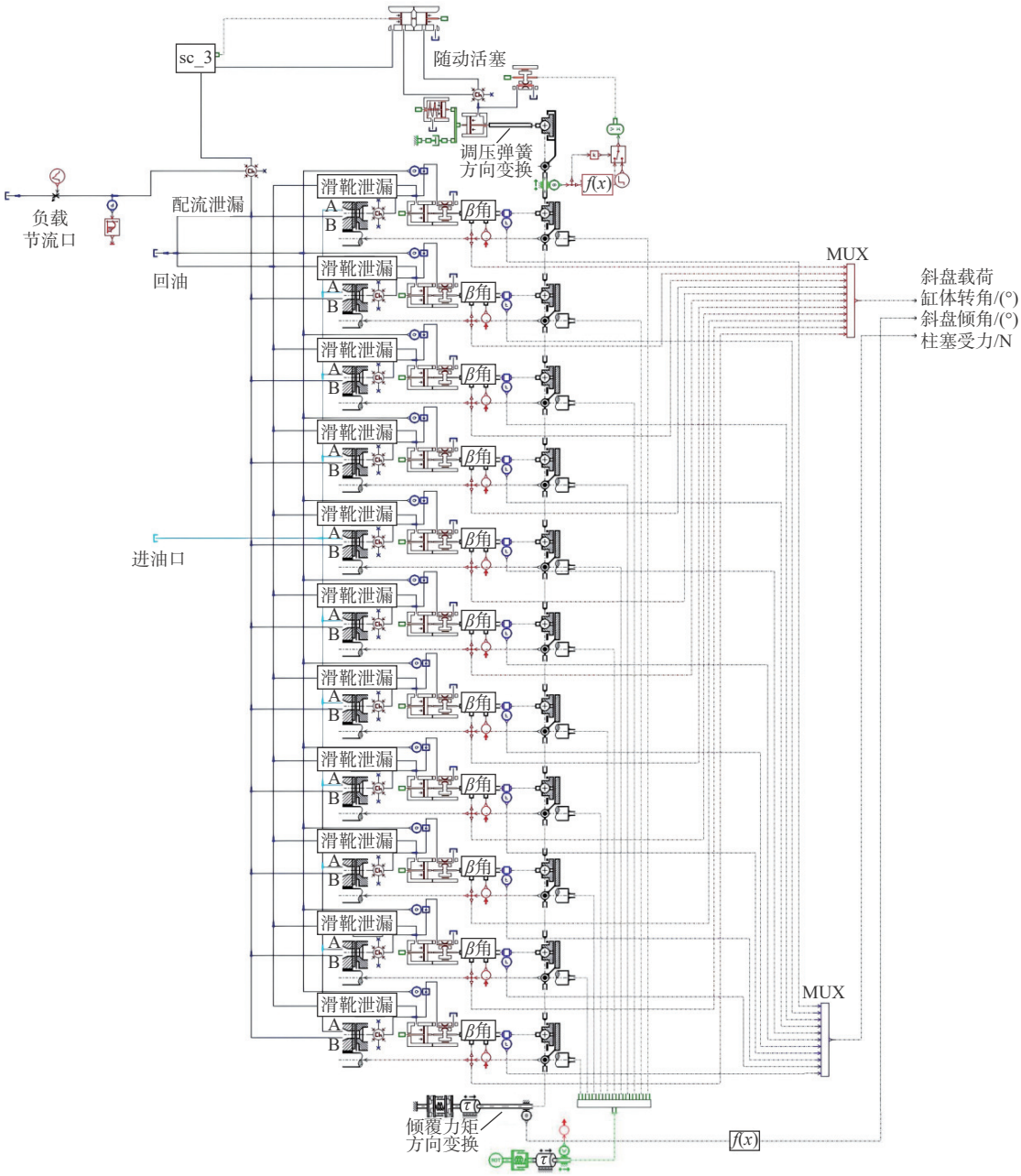


图 4　航空柱塞泵系统仿真模型

Fig. 4　System simulation model of aviation plunger pump

通过调节负载压力使负载流量不断变化的方式测量柱塞泵的压力流量特性。通过改变调节阀流通面积, 使负载压力从 0 开始大范围变化, 发现其流量大幅下降的点在 33 MPa 附近, 该泵设计的额定工作压力为 35 MPa, 在 33～35 MPa 内使负载压力小范围变化, 以寻找最大全流量点和零流量点, 进而得到压力-流量特性曲线。

将试验和仿真得到的负载压力 P_L 与负载流量 Q_L 特性曲线进行对比, 如图 5 所示。可以看出, 在调压前, 随着负载压力的升高, 负载流量几乎保持不变。这是因为此时的负载压力较低, 不足以克服调压弹簧的弹簧力, 则斜盘倾角始终不变, 柱塞泵的输出流量也几乎不变。当负载压力大于弹簧力时, 推动随动活塞右移, 使斜盘倾角减小, 进而使柱塞泵的输出流量降低。这与恒压变量式柱塞泵的工作原理一致。还可以看出, 仿真与试验结果基本相同, 说明航

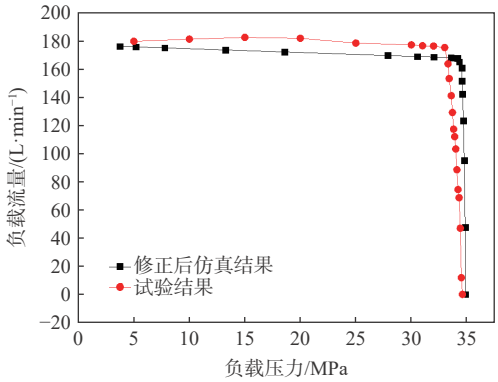


图 5 负载压力-负载流量特性曲线对比

Fig. 5 Comparison of load pressure-load flow characteristic curves

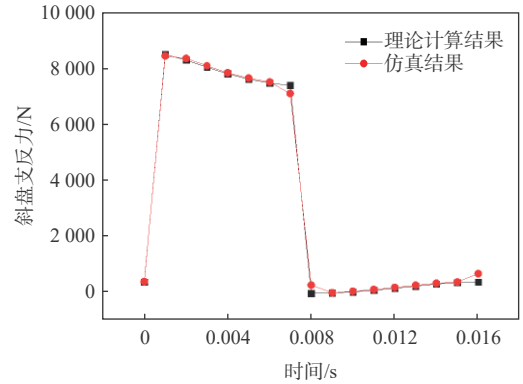


图 7 多刚体动力学仿真结果与理论结果的对比

Fig. 7 The comparison of multi rigid body dynamics simulation results with theoretical results

空柱塞泵系统仿真模型建立无误。其中的误差主要是由于系统仿真建模过程中忽略了柱塞、滑靴和随动活塞的质量,且未考虑球铰副的泄漏和温度的影响。

1.4 主轴-缸体系统多刚体动力学仿真模型的建立与验证

在明确航空柱塞泵主要部件间的运动关系和受力情况后,在 Motion 软件中通过建立运动约束和施加载荷构建出柱塞泵主轴-缸体系统多刚体动力学仿真模型,如图 6 所示。

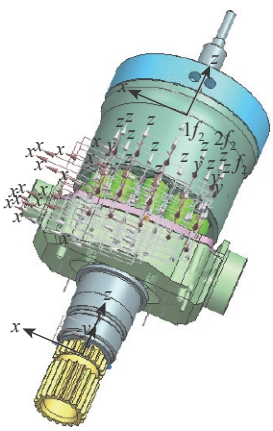


图 6 主轴-缸体系统多刚体动力学仿真模型

Fig. 6 Multi rigid body dynamic simulation model of spindle-cylinder system

通过该模型,可以对各部件间的受力情况和运动状态进行计算。由柱塞滑靴组件的受力分析可知,斜盘支反力 F_N 具有一定代表性,且易于通过理论公式计算。因此,通过对比理论计算和仿真计算得到的斜盘支反力变化曲线来说明多刚体动力学仿真模型的准确性。

提取仿真结果中初始位于上死点位置的滑靴在一个周期内的受力情况,将仿真和理论公式得到的斜盘支反力变化曲线进行对比,如图 7 所示。

从图 7 可以看出,随着初始位于上死点位置的柱塞开始压油,柱塞腔内液压力不断增加,斜盘支反力也随之增大;而后,随着柱塞与缸体间摩擦力增加及油液排出,斜盘支反力有所下降。在完成压油后,初始位于上死点位置的柱塞进入吸油区,斜盘支反力迅速下降,并随着油液吸入再次有所上升。这与柱塞泵实际的吸排油过程基本一致,且仿真结果与理论结果基本相同,说明主轴-缸体系统多刚体动力学仿真模型基本无误。

1.5 主轴-缸体系统刚柔耦合动力学仿真模型的建立

由于系统仿真模型建立过程中没有考虑柱塞和滑靴质量等因素的影响,因此,对柱塞位移和速度的计算与实际情况存在一定差异;而多刚体动力学仿真模型中,施加柱塞腔液压力载荷时,也进行了一定简化,与实际情况不尽相同。因此,将多刚体动力学仿真计算得到的柱塞位移和速度输入到系统仿真模型中,同时,将系统仿真计算得到的柱塞腔液压力输入到多刚体动力学仿真模型中,以建立航空柱塞泵最大全流量状态时的 1D+3D 联合仿真模型。

在联合仿真模型中,对初始位于上死点位置的滑靴在一个周期内的受力情况进行提取,并与理论公式的计算结果进行比较,如图 8 所示。可以看出,联合仿真结果与理论计算结果的趋势基本一致。但在联合仿真结果中,斜盘支反力变化曲线在

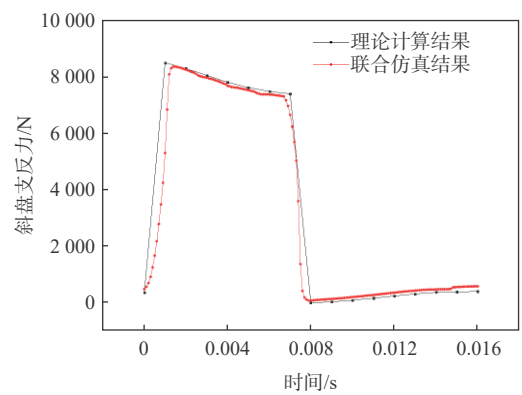


图 8 联合仿真结果与理论结果的对比

Fig. 8 The comparison between joint simulation results and theoretical results

起始点和拐点处存在短暂的平滑过渡; 且最大值比理论计算结果小, 最小值比理论计算结果大, 这与实际情况更加贴合。因为在实际压油过程中, 斜盘支反力的增大是随着油液压力的升高而逐渐增大的, 并非瞬间增大。吸油过程同理。此外, 在理论计算中认为高压区的柱塞腔压力均为 35 MPa, 低压区的柱塞腔压力均为 0。但实际过程中, 由于存在泄漏, 高压区柱塞腔的压力会略小一些, 而低压区柱塞腔由于存在死油, 其液压力不会下降到 0。因此, 实际斜盘支反力的最大值比理论结果小, 而最小值比理论结果大。综上所述, 联合仿真的计算结果更加准确, 能较好地反映实际情况。

进一步通过联合仿真计算每个柱塞传递给缸体的载荷, 进而将柱塞和滑靴等部件简化, 只保留主轴和缸体模型, 以提高仿真的计算效率, 并为后续主轴-缸体系统刚柔耦合动力学仿真模型的建立奠定基础。考虑到主轴前端通过单列向心球轴承与壳体相连, 后端通过单列圆柱滚子轴承与配流盘相连, 因此, 在简化后的模型中加入轴承单元以模拟实际支撑情况。其中, 单列向心球轴承径向刚度的理论计算公式为

$$R_{\mathrm{r}} = 0.118\times10^{4}\left(D_{\mathrm{W}}F_{\mathrm{r}}Z^{2}\cos^{5}\alpha\right)^{1/3} \tag{7}$$

单列圆柱滚子轴承径向刚度的理论计算公式为

$$R_{\mathrm{r}} = 0.34\times10^{4}\left(F_{\mathrm{r}}^{1/10}Z^{9/10}L_{\mathrm{W}}^{8/10}\cos^{19/10}\alpha\right) \tag{8}$$

式中: R_{r}为轴承径向刚度; F_{r}为轴承径向力; Z为轴承滚子数量; α为轴承接触角; D_{W}为向心球轴承的球直径; L_{W}为圆柱滚子轴承的圆柱长度。由于 2 个轴承均属于滚动轴承, 忽略轴向刚度和阻尼的影响。考虑到配流副中的油液和其他螺钉等连接件对缸体也有一定支撑作用, 在与配流盘接触的缸体端面施加 4 个弹簧, 以模拟这种作用。将弹簧的刚度初步取为 10 000 N/mm。此时的模型为多刚体动力学模型, 仍然难以反映实际缸体的倾覆情况, 因此, 进一步利用有限元软件对主轴的模态进行计算, 并将结果导入, 以实现主轴的柔性化, 进而建立主轴-缸体系统刚柔耦合动力学模型, 同时, 加入传感器的简化模型, 如图 9 所示。

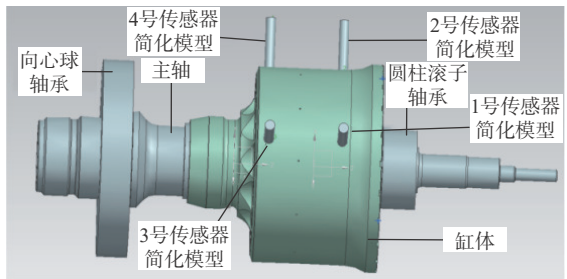
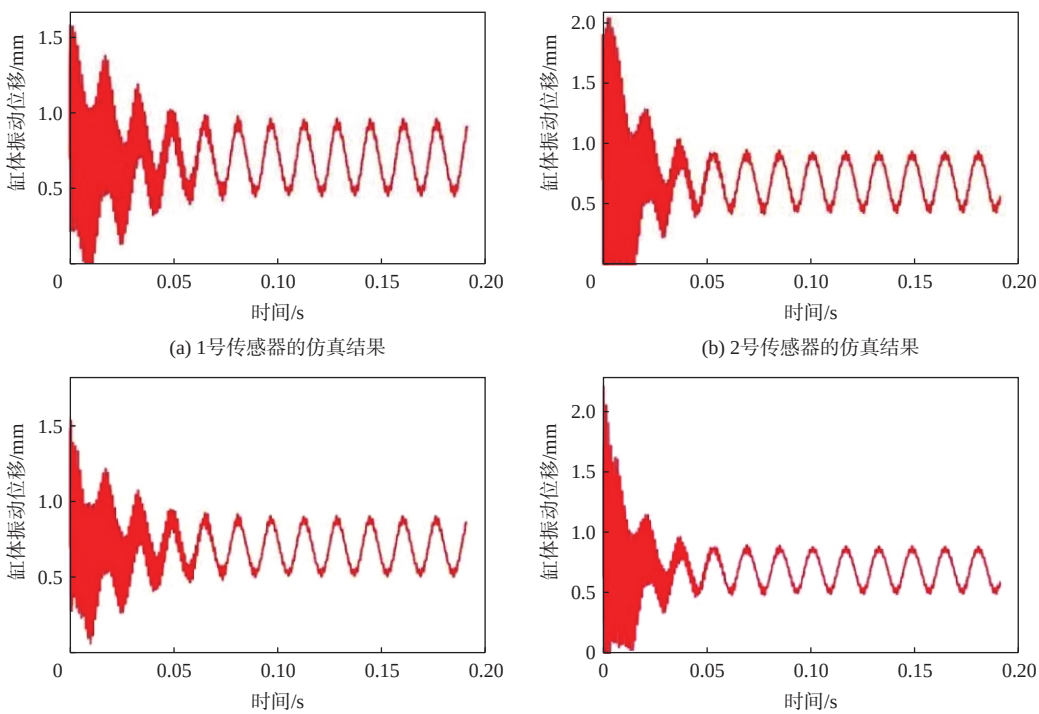


图 9　主轴-缸体系统刚柔耦合动力学模型

Fig. 9　The rigid flexible coupling dynamic model of the spindle-cylinder system

上述模型中, 传感器简化模型与缸体的初始距离为 0.7 mm, 通过 Motion 软件中的动画测量功能获得缸体外表面与传感器端面间的实时动态距离, 以模拟柱塞泵工作过程中缸体的振动位移。通过计算, 初步得到 4 个传感器对应测量点的振动位移信号, 如图 10 所示。



(a) 1号传感器的仿真结果　(b) 2号传感器的仿真结果

(c) 3号传感器的仿真结果　(d) 4号传感器的仿真结果

图 10　仿真得到的缸体振动位移

Fig. 10　Simulated cylinder vibration displacement

基于缸体4个测点的振动位移信号能够对缸体的倾覆情况进行分析，具体分析过程在后续试验环节详述。由于模型的建立进行了简化和等效，虽然整体计算过程效率较高，但计算结果仍与实际存在一定差异，后续将通过试验进行进一步修正。

2 缸体动特性试验研究及倾覆行为优化

2.1 缸体动特性试验的设计与开展

考虑缸体封装在壳体内部，且在工作过程中不断旋转，采用非接触式的电涡流位移传感器进行测量。柱塞泵的传力路径较为复杂，对缸体信号进行直接测量才能较准确地分析其倾覆情况。因此，在壳体上开设螺纹孔，使传感器伸入壳体腔内，对缸体进行直接测量。该测量方式对柱塞泵改动较小、对传感器要求较低，且适用于大多数种类的柱塞泵。电涡流位移传感器采用国产的CWY-DO型传感器，其具体技术指标如表1所示。

由于壳体内部存在约0.3 MPa的回油压力，而螺纹连接难以实现很好的密封，因此，在传感器与壳体的外接触面涂抹HY9-914树脂胶进行密封，以防止工作过程中回油油液的泄漏。在航空柱塞泵综合性能试验台上开展试验，测量全流量状态时出口压力为10 MPa、20 MPa和35 MPa，75%全流量、50%全流量、25%全流量和0流量及20%、40%、60%、80%、100%和110%额定转速下的缸体振动位移。以出口压力为10 MPa、转速为100 r/min时的工况为初始状态，利用能够提供+9 V/±15 V供电电压的DEWETRON便携式一体化数据采集仪进行数据采集。

表1 电涡流位移传感器的技术指标

Table 1 Technical specifications of eddy current displacement sensors

参数值及要求	探头到缸体的距离/mm	零点/mm	灵敏度/(V·mm^{-1})	平均灵敏度误差/%	典型分辨率/μm	工作温度/℃	承受压力/MPa	非线性误差/%	探头直径/mm	频响/kHz	探头长度/mm	校核方式	供电电压/V
CWY-DO电涡流位移传感器参数	0.2～1.2	0.2	8	不超过±5	1	−20～120	0.2～0.4	不超过±1	5	0～10	40	铁磁性校核	12～15
试验环境要求	>0.3				≤1	室温	约0.3		≤6	>6.92	>21.3	不锈钢材质	+9/±15

试验台如图11所示，主要包括进油、出油和回油3条通道，且每条通道均有温度、压力和流量传感器，能够实现对航空柱塞泵功能性能的实时检测和控制。

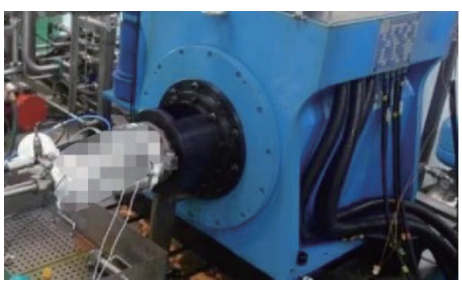


图11 柱塞泵试验台

Fig. 11 Plunger pump test bench

2.2 试验结果分析

对于缸体倾覆行为，从其相对平衡位置的倾覆角γ和径向偏离位移L_q这2个方面进行分析，如图12所示。将图9中3号传感器测量得到的振动位移相对值用x_1表示、4号传感器测量得到的振动位移相对值用y_1表示、1号传感器测量得到的振动位移相对值用x_2表示、2号传感器测量得到的振动位移相对值用y_2表示，将2个测量截面到转子底面的距离分别用z_1和z_2表示，则可以给出缸体倾覆角γ和径向偏离位移L_q的表达式分别为

$$\gamma=\arctan\left(\frac{\sqrt{(y_2-y_1)^2+(x_2-x_1)^2}}{z_2-z_1}\right) \tag{9}$$

$$L_q=(z_2-z_1)\tan\gamma \tag{10}$$

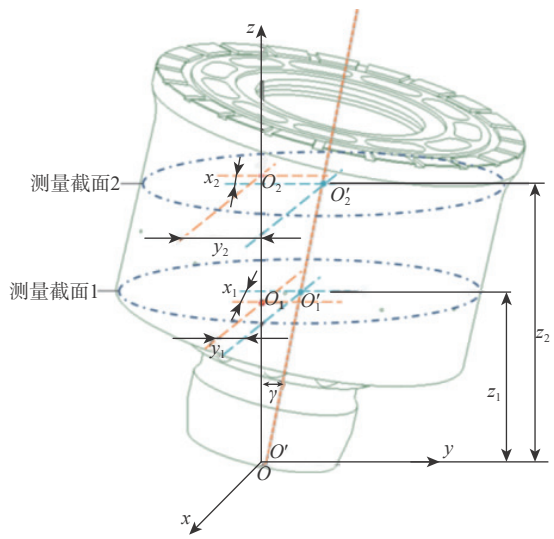


图12 缸体倾覆示意图

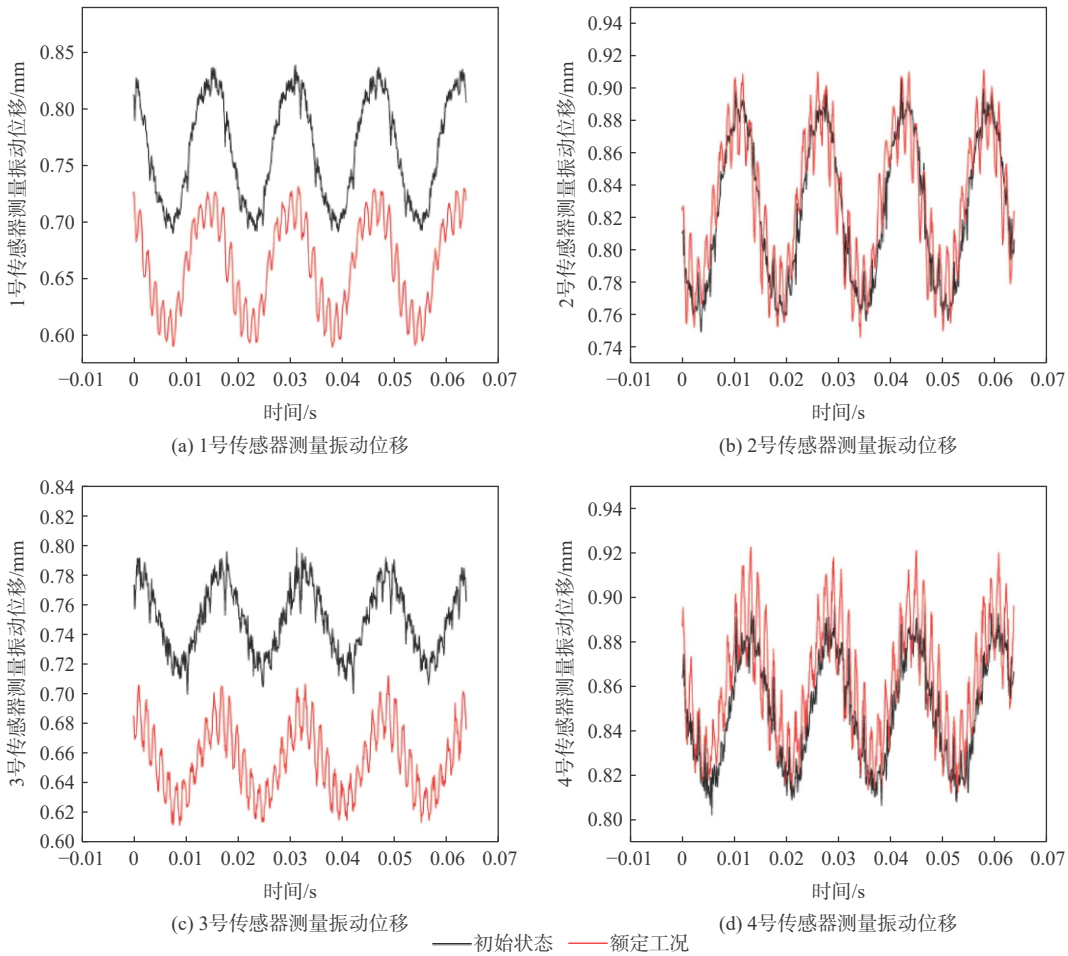
Fig. 12 Schematic diagram of cylinder overturning

传感器直接得到的位移响应曲线频率与转速相关，初始状态的转速为100 r/min，其他工况的转速不尽相同。因此，在研究不同转速的结果时，对初始状态的测量结果进行处理，使其周期缩小，并提取对应研究工况周期的结果。若存在相位差，则令数据平移。以转速为3 775 r/min、出口压力为35 MPa

的工况, 即额定工况为例, 经过处理得到前 4 个周期的测量结果, 如图 13 所示。

从图 13 可以看出, 不同转速工况的位移响应曲线频率相同, 可通过直接相减的方式获得测量相对值, 进而根据式(9)和式(10)得到额定工况时缸体倾覆角γ和径向偏离位移L_q随时间t的变化关系, 如图 14 和图 15 所示。可以看出, 额定工况时, 缸体倾覆角在 0°～0.08°之间呈周期变化, 径向偏离位移在 0～0.06 mm 之间呈周期变化, 整体倾覆情况较小, 不会发生较严重事故。

相比于缸体倾覆角, 径向偏离位移能够更直观地反映缸体倾覆情况。因此, 对径向偏离位移进行进一步分析。考虑均方根值能够表示信号的能量, 常用于评价振动等级或烈度, 用缸体径向偏离位移的均方根值变化情况来表征缸体倾覆程度的变化, 给出其具体计算公式为[21]



(a) 1号传感器测量振动位移　(b) 2号传感器测量振动位移

(c) 3号传感器测量振动位移　(d) 4号传感器测量振动位移

图 13　额定工况和初始状态前 4 个周期的测量结果

Fig. 13　Measurement results for the first four cycles of rated operating conditions and initial state

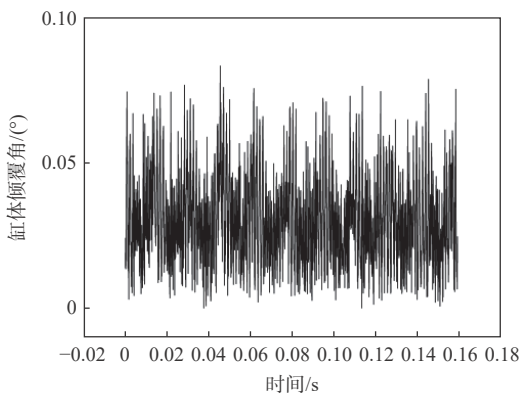


图 14　额定工况时的缸体倾覆角变化曲线

Fig. 14　Cylinder overturning angle variation curve under rated operating conditions

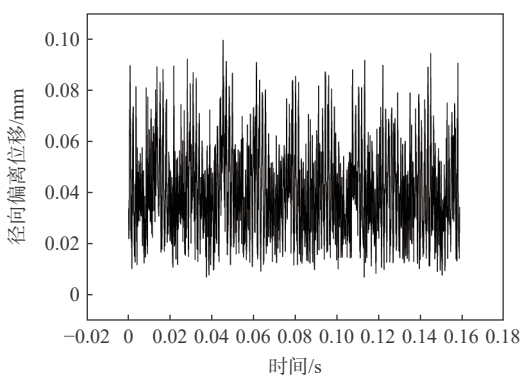


图 15　额定工况时的径向偏离位移变化曲线

Fig. 15　Radial deviation displacement variation curve under rated operating conditions

$$L_{\mathrm{qr}}=\sqrt{\frac{1}{n}\sum_{i=0}^{n-1}Y_i^2} \tag{11}$$

式中：L_{qr}为缸体径向偏离位移的均方根值；n为采样点数；Y_i为时域内的缸体径向偏离位移。此外，对径向偏离位移曲线进行快速傅里叶变换，发现所有工况的1倍频幅值均较大，因此，将径向偏离位移曲线1倍频幅值L_{q1}的变化情况也作为表征缸体倾覆程度变化的参数进行综合评价。

当转速为3 775 r/min时，将全流量状态下出口压力为10 MPa、20 MPa和35 MPa及75%全流量、50%全流量、25%全流量和0流量时的结果按照上述过程进行整理和计算，得到相同转速、不同出口压力时缸体径向偏离位移的均方根值和1倍频幅值，如表2所示。

表2 相同转速、不同出口压力时缸体径向偏离位移的均方根值和1倍频幅值

Table 2 The root mean square value and 1st harmonic amplitude of radial deviation displacement of cylinder body at the same rotational speed and different outlet pressures

测量工况	径向偏离位移均方根值/mm	径向偏离位移1倍频幅值/mm
10 MPa	0.071 04	0.008 71
20 MPa	0.030 66	0.001 56
35 MPa	0.037 41	0.005 46
75%全流量	0.039 31	0.007 68
50%全流量	0.033 47	0.005 79
25%全流量	0.030 22	0.005 17
0流量	0.034 61	0.006 98

从表2可以看出，在全流量时，随着出口压力增加，缸体倾覆先减小后增大，在出口压力较低时，缸体倾覆较为严重。当出口压力较低时，柱塞泵内部各处油膜较薄，柱塞泵整体工作情况并不稳定，因此，缸体倾覆较大。随着出口压力增加，柱塞泵工作状态逐渐稳定，缸体倾覆逐渐减小。但随着出口压力进一步增加，柱塞泵内摩擦接触件的受力开始增加，工作环境越来越恶劣，缸体倾覆情况逐渐变大。调压后，随着输出流量减小，缸体倾覆先增大后逐渐减小，当流量减小到0时，又进一步增大。这是由于当输出流量开始减小时，斜盘倾角变化较小，缸体受力情况虽有所缓解，但出口压力随着输出流量的减小进一步增大，因此，缸体整体受力情况呈增加趋势，缸体倾覆增大。当斜盘倾角进一步减小时，缸体整体受力情况下降，缸体倾覆减小。但当斜盘倾角接近0时，斜盘横向角开始起主导作用，缸体受力再次增加，缸体倾覆也随之增大。其中，斜盘横向角的主要作用是对柱塞泵的闭死压缩和膨胀进行补偿。斜盘横向角一般在1°～3°之间，数值较小。

当出口压力为35 MPa时，将20%、40%、60%、80%、100%和110%额定转速下的结果按照上述过程进行整理和计算，得到如表3所示的结果。

表3 相同出口压力、不同转速时缸体径向偏离位移的均方根值和1倍频幅值

Table 3 The root mean square value and 1st harmonic amplitude of radial deviation displacement of cylinder body at the same outlet pressure and different rotational speeds

测量工况	径向偏离位移均方根值/mm	径向偏离位移1倍频幅值/mm
20%额定转速	0.045 89	0.032 31
40%额定转速	0.040 85	0.014 93
60%额定转速	0.040 62	0.011 10
80%额定转速	0.039 49	0.006 23
100%额定转速	0.037 41	0.005 46
110%额定转速	0.036 13	0.005 96

从表3可以看出，当出口压力不变时，缸体倾覆行为随转速增大整体上不断减小。这是由于转速越高，陀螺效应越明显。具体地，当转速较高时，缸体绕着主轴旋转的同时，在其对称轴上绕一个小的角度自转，从而对外部力矩产生抵抗，使其稳定性增强，倾覆程度减小。

进一步对缸体倾覆角进行分析。当缸体产生倾覆后，缸体与配流盘间的油膜厚度发生改变，而这种变化可以通过缸体倾覆角和倾覆方位角来描述。其中，倾覆方位角θ指的是缸体产生倾覆后，其轴线在配流盘上的投影与x轴正方向所形成的夹角，用来表征缸体与配流盘间油膜厚度最薄的位置。将图12所示情况投影到配流盘端面所在平面，由于缸体在工作过程中先经过1、3号传感器，后经过2、4号传感器，则定义1、3号传感器的测量方向为x正方向，2、4号传感器的测量方向为y正方向，如图16所示。根据图中的几何关系可以给出θ的计算公式为

$$\theta=\begin{cases}\arctan\left(\dfrac{y_1-y_2}{x_1-x_2}\right) & x_1-x_2\geqslant 0\\ \arctan\left(\dfrac{y_1-y_2}{x_1-x_2}\right)+\pi & x_1-x_2<0\end{cases} \tag{12}$$

缸体与配流盘间油膜最薄的位置是缸体径向偏离位移最大的位置，若此时缸体倾覆角也最大，则该位置是缸体和配流盘最容易发生磨损的位置。基于此，通过试验数据找出该型柱塞泵中缸体与配流盘最容易发生磨损的位置区间。

以图14所示曲线中前4个周期的结果为例，取出缸体倾覆角最大值及其对应时刻，结合式(12)的

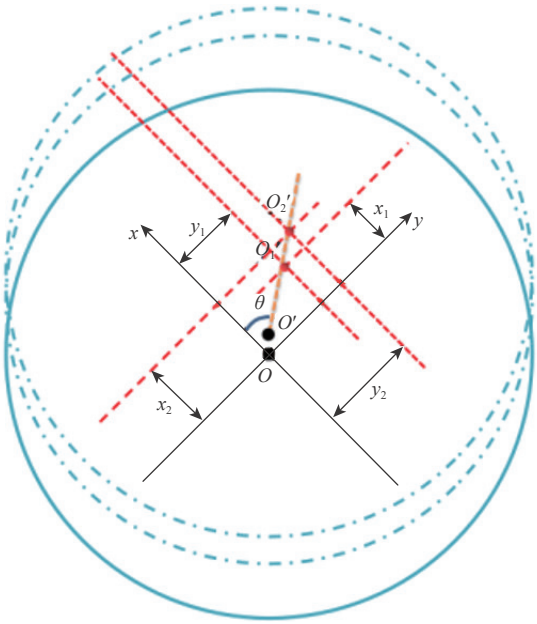


图 16　缸体倾覆方位角示意图

Fig. 16　Schematic diagram of cylinder overturning azimuth angle

计算结果可知, 第 1 周期内缸体倾覆角最大时对应的倾覆方位角为 29.61°, 第 2 周期为 39.39°, 第 3 周期为 56.91°, 第 4 周期为 31.06°。说明在前 4 周期内缸体与配流盘最容易发生磨损的位置在缸体倾覆方位角为 29.61°～56.91°的区间上。对更多的周期进行分析后, 倾覆方位角也在这个区间内, 说明该区间是额定工况下缸体倾覆最剧烈的位置, 如图 17 所示。

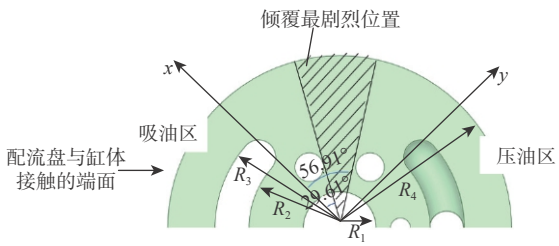


图 17　额定工况时缸体倾覆最剧烈的位置

Fig. 17　The position where the cylinder overturns the most severely under rated operating conditions

该区间位于缸体从吸油区向压油区运动的过渡区域。为了进一步明确缸体倾覆最剧烈的位置, 将转速为 3 775 r/min 时, 全流量状态下出口压力为 10 MPa、20 MPa 和 35 MPa 及 75% 全流量、50% 全流量、25% 全流量和 0 流量时的结果按照上述过程进行整理和计算, 得到相同转速、不同出口压力时缸体倾覆最剧烈的位置, 如表 4 所示。

从表 4 可以看出, 当出口压力较小时, 缸体倾覆最剧烈的位置发生在压油区, 这可能是由于此时压油区油膜较薄而造成的。在其他工况下, 缸体倾覆最剧烈的位置均集中在缸体从吸油区向压油区运动的过渡区域。进一步将出口压力为 35 MPa 时, 20%、40%、60%、80%、100% 和 110% 额定转速下的结果按照上述过程进行整理和计算, 得到相同出口压力、不同转速时缸体倾覆最剧烈的位置, 如表 5 所示。

表 4　相同转速、不同出口压力时缸体倾覆最剧烈的位置

Table 4　The most severe overturning position of the cylinder body at the same rotational speed and different outlet pressures

测量工况	缸体倾覆最剧烈的位置区间/(°)
10 MPa	130.56～167.59
20 MPa	32.41～57.28
35 MPa	29.61～56.91
75%全流量	48.63～84.39
50%全流量	41.43～73.84
25%全流量	16.12～51.39
0流量	49.89～74.80

表 5　相同出口压力、不同转速时缸体倾覆最剧烈的位置

Table 5　The most severe overturning position of the cylinder body at the same outlet pressure and different rotational speeds

测量工况	缸体倾覆最剧烈的位置区间/(°)
20%额定转速	40.63～80.13
40%额定转速	57.43～66.77
60%额定转速	9.26～40.90
80%额定转速	45.92～79.81
100%额定转速	29.61～56.91
110%额定转速	44.45～46.96

从表 5 可以看出, 出口压力相同时, 不同转速下缸体倾覆最剧烈的位置也均集中在缸体从吸油区向压油区运动的过渡区域。说明该型航空柱塞泵配流盘最容易发生磨损的区域是缸体从吸油区向压油区运动的过渡区域。

2.3　减小缸体倾覆措施

结合 2.2 节的分析, 对该型柱塞泵的 2 台实物泵进行拆解分析, 发现 2 台泵的配流盘在吸油区向压油区的过渡区域均发生了明显的磨损, 如图 18 所示。由此可以看出, 上述分析和试验过程基本正确。进一步对造成该现象的原因进行深入分析, 以分析减小缸体倾覆的举措。

由式(3)和式(5)可知, 在缸体端面受到的力中, 与缸体旋转角相关的唯一参数是压力区包角ϕ, 其由配流盘压油区的腰形槽和与其相连通的缸体端面小腰形配流窗口共同组成, 如图 19 所示(该图仅用于对压力区包角进行相关说明, 与产品实际尺

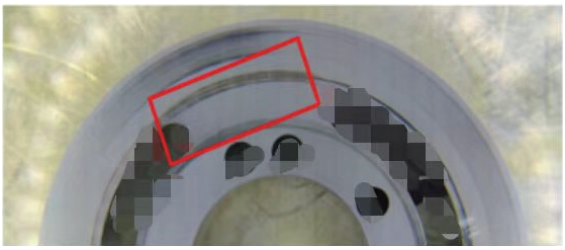


图 18 配流盘端面磨损情况

Fig. 18 Wear condition of the end face of the distribution plate

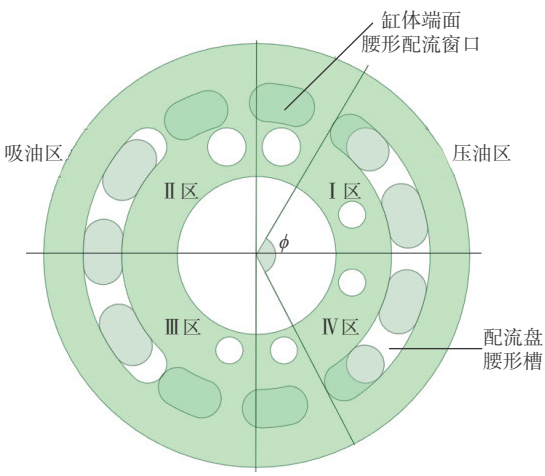


图 19 压力区包角示意图

Fig. 19 The schematic diagram of pressure zone envelope angle

寸无关)。进一步对ϕ的变化进行分析以探究造成偏磨的原因。

通过对仿真动画的逐帧分析发现,压力区包角ϕ随缸体旋转角φ的增加呈周期性变化,且在其一个周期内的变化主要分为5个阶段,如图20所示。由缸体端面的受力分析可知,配流盘对缸体的液压支承力F_s作用在压力区包角在高压区的角分线上,则压力区包角的大小不但影响液压支承力的大小,还影响液压支承力的作用位置。对图20的变化进一步分析发现,压力区包角在一个变化周期内有75.1%的时间处在最大值,此时,缸体高压区一侧受到较大的液压支承力,使缸体向着低压区倾覆。而在该段时间内,压力区包角的角平分线有69.0%的时间处于图19所示的Ⅳ区,使图19所示的Ⅱ区容易出现磨损。而在Ⅱ区中容易产生磨损的位置为低压区向高压区的过渡区域。具体来说,当柱塞泵在额定工况工作1 min时,配流盘上低压区向高压区的过渡区域有31.1 s的时间容易出现磨损。这就解释了配流盘在柱塞泵工作一段时间后出现磨损的原因。因此,可通过改进压力区包角的设计来改善缸体端面的受力情况,从而使缸体倾覆减小。

基于上述结论,将配流盘腰型槽两端减小5°。利用数值仿真得到改进前后液压支承力F_s随缸体旋转角φ的变化曲线,如图21所示。可以看出,将

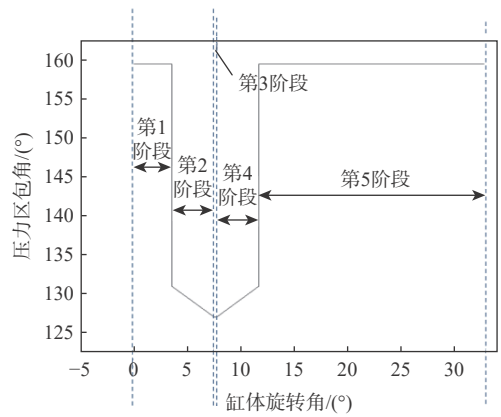


图 20 压力区包角在一个周期内的变化情况

Fig. 20 The variation of pressure zone envelope angle within one cycle

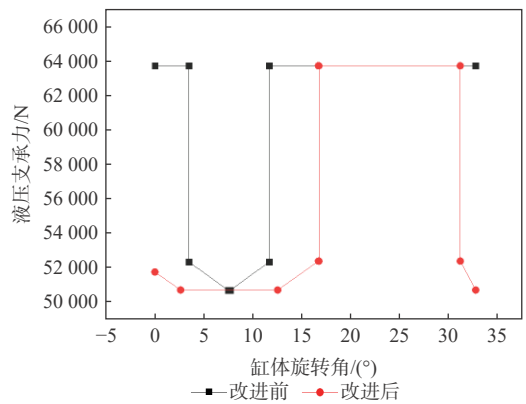


图 21 改进前后数值仿真结果对比

Fig. 21 The comparison of numerical simulation results before and after improvement

配流盘腰形槽尺寸减小后,可以使缸体端面的受力情况得到改善,进而改善缸体的倾覆情况。

3 基于试验数据的模型修正

在1.5节所述的仿真模型中,轴承的径向刚度均通过理论公式计算得到,存在一定假设条件;同时,均未考虑轴承的轴向刚度和阻尼,因此,轴承的刚度设置与实际情况存在一定差异。为了得到高保真仿真模型,对向心球轴承的刚度k_t和圆柱滚子轴承的刚度k_c进行修正。此外,用于模拟支撑作用的缸体端面弹簧的刚度是由经验得到的估计值,因此,同时对端面弹簧刚度k_d进行修正。

将缸体径向偏离位移在额定工况时的均方根值L_{qr}和1倍频幅值L_{q1}作为仿真模型的修正指标,修正表达式为[22]

$$\begin{cases} \min F(\boldsymbol{X}) = \sum_j \alpha_j \left[\left(\hat{f}_j - f_j(\boldsymbol{X}) \right) / f_j \right]^2 & j = 1,2 \\ \text{s.t. } \boldsymbol{X}_{\mathrm{L}} < \boldsymbol{X} < \boldsymbol{X}_{\mathrm{U}} \end{cases} \tag{13}$$

式中: F为目标函数; α_j为权重系数, 取 1/2; $\hat{f}_1$和$\hat{f}_2$分别为试验得到的均方根值和 1 倍频幅值; f_1和f_2分别为仿真计算出的均方根值和 1 倍频幅值; $\boldsymbol{X}$为修正参数; $\boldsymbol{X}_{\mathrm{U}}$、$\boldsymbol{X}_{\mathrm{L}}$分别为修正参数上、下限。

在 3 个修正参数中, k_{t}和k_{c}的初值可通过式(7)和式(8)进行计算。进一步结合供应商提供的误差范围得到k_{t}的取值范围为 3.145×10^5 ～3.15×10^5 N/mm, k_{c}的取值范围为 7.66×10^5～7.665×10^5 N/mm。利用主轴-缸体系统刚柔耦合动力学模型进行初步计算发现, k_{d}在 1.3×10^4～2.4×10^4 N/mm 范围内变化较小, 因此, 将该区间作为k_{d}的取值范围。利用响应面设计方法得到仿真结果与修正参数的函数关系, 如图 22 和图 23 所示。其中, 均方根值的拟合相关系数为 0.999 4, 1 倍频幅值的拟合相关系数为 0.997 8, 说明响应面拟合曲线在修正参数的取值范围内能够较准确地反映均方根值和 1 倍频幅值与修正参数的关系。

得到拟合曲线后, 利用牛顿迭代法进行迭代修正。将迭代表达式定义为[23]

$$q_{k+1} = q_k - \boldsymbol{A}_k^{-1}\boldsymbol{g}_k \tag{14}$$

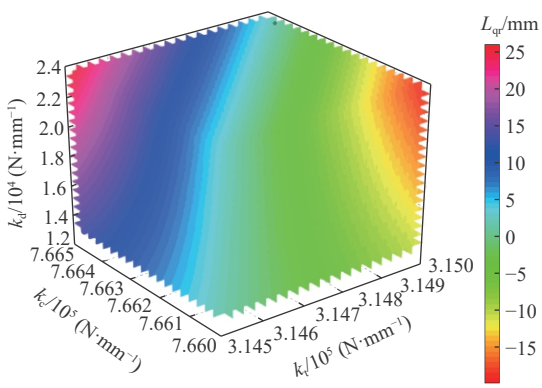


图 22　径向偏离位移均方根值与修正参数的关系

Fig. 22　The relationship between radial deviation displacement root mean square value and correction parameters

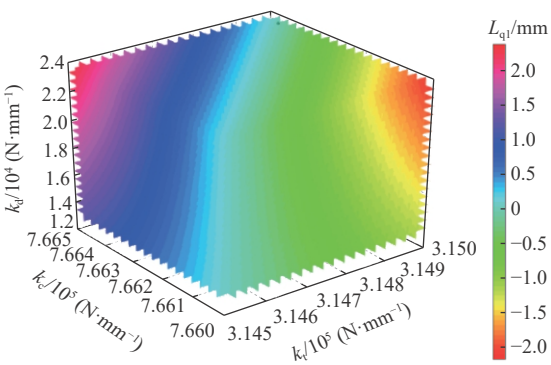


图 23　径向偏离位移 1 倍频幅值与修正参数的关系

Fig. 23　The relationship between radial deviation displacement 1 fold frequency amplitude and correction parameters

式中: q_k为修正参数在第k个迭代步的取值; $\boldsymbol{A}_k$表示目标函数泰勒级数的二次项系数; $\boldsymbol{g}_k$为q在q_k处的梯度矩阵。经过 7 次迭代后, 修正误差收敛, 最终结果为 k_{t}=314 770 N/mm, k_{c}=766 242 N/mm, k_{d}=18 350 N/mm。此时, 仿真结果与试验结果的相对误差最小, 其中, 均方根值的相对误差为 4.07%, 1 倍频幅值的相对误差为 4.61%。可以认为得到了高精度仿真模型。

4　结　论

本文聚焦航空柱塞泵缸体倾覆行为, 通过仿真与试验结合的方法展开研究。通过分析结果, 得出如下结论:

1) 调压前, 随着出口压力增加, 缸体倾覆先减小后增大; 调压后, 随着出口压力增加, 缸体倾覆先增大后减小, 流量逐渐变为 0 时, 缸体倾覆进一步增大。

2) 随着转速增加, 缸体倾覆逐渐减小。

3) 减小配流盘腰形槽的尺寸可以使缸体端面的受力情况得到改善, 进而使缸体倾覆有所减小。

本文仍有一些不足, 对减小缸体倾覆的改进措施没有进一步开展试验验证, 后续将继续进行研究。本文工作为航空柱塞泵抗倾覆和可靠性研究提供了一定的支持和帮助, 为后续航空柱塞泵的设计和发展研究提供了思路和方法。

参考文献（References）

[1] 李玉龙, 何永勇, 雒建斌. 航空柱塞泵关键摩擦副表面改性与性能增强[J]. 清华大学学报 (自然科学版), 2021, 61(12): 1405-1422.
LI Y L, HE Y Y, LUO J B. Surface modifications and performance enhancements of key friction pairs in aviation hydraulic piston pumps[J]. Journal of Tsinghua University (Science and Technology), 2021, 61(12): 1405-1422(in Chinese).

[2] 刘伟. 航空柱塞泵常见故障原因与分析[J]. 航空维修与工程, 2023(9): 59-61.
LIU W. Causes and analysis of common faults in aviation piston pump[J]. Aviation Maintenance & Engineering, 2023(9): 59-61(in Chinese).

[3] 杨毅博, 李运华. 全飞行剖面下的航空液压泵热载荷分析与温度预测[J]. 液压与气动, 2021, 45(8): 34-38.
YANG Y B, LI Y H. Thermal load analysis and temperature prediction of aviation hydraulic pump under full flight profile[J]. Chinese Hydraulics & Pneumatics, 2021, 45(8): 34-38(in Chinese).

[4] 陈远玲, 班成周, 刘银水, 等. 航空柱塞泵配流副磨损模型与寿命预测[J]. 液压与气动, 2020, 44(12): 1-7.
CHEN Y L, BAN C Z, LIU Y S, et al. Wear model and life prediction of port pair in aircraft piston pump[J]. Chinese Hydraulics & Pneumatics, 2020, 44(12): 1-7(in Chinese).

[5] 欧阳小平, 王天照, 方旭. 高速航空柱塞泵研究现状[J]. 液压与气动, 2018, 42(2): 1-8.

OUYANG X P, WANG T Z, FANG X. Research status of the high speed aircraft piston pump[J]. Chinese Hydraulics & Pneumatics, 2018, 42(2): 1-8(in Chinese).

[6] 陈金华, 郭生荣, 卢岳良. 民机发动机驱动泵的长寿命设计技术[J]. 液压与气动, 2021, 45(8): 160-170.
CHEN J H, GUO S R, LU Y L. Long life design technology of civil aircraft engine driven pump[J]. Chinese Hydraulics & Pneumatics, 2021, 45(8): 160-170(in Chinese).

[7] 郭生荣. 飞机机电流体系统综合技术[M]. 上海: 上海科学技术出版社, 2020: 63.
GUO S R. Integrated technology of aircraft electromechanical fluid system[M]. Shanghai: Shanghai Scientific & Technical Publishers, 2020: 63(in Chinese).

[8] 张绍九. 液压密封[M]. 北京: 化学工业出版社, 2012.
ZHANG S J. Hydraulics seals[M]. Beijing: Chemical Industry Press, 2012(in Chinese).

[9] MANRING N D. Tipping the cylinder block of an axial-piston swash-plate type hydrostatic machine[J]. Journal of Dynamic Systems, Measurement, and Control, 2000, 122(1): 216-221.

[10] BERGADA J M, DAVIES D L, KUMAR S, et al. The effect of oil pressure and temperature on barrel film thickness and barrel dynamics of an axial piston pump[J]. Meccanica, 2012, 47(3): 639-654.

[11] ZHANG J H, CHAO Q, XU B. Analysis of the cylinder block tilting inertia moment and its effect on the performance of high-speed electro-hydrostatic actuator pumps of aircraft[J]. Chinese Journal of Aeronautics, 2018, 31(1): 169-177.

[12] 胡骁, 王少萍, 韩磊. 轴向柱塞泵配流副动态压力分布建模与仿真[J]. 液压气动与密封, 2012, 32(8): 68-71.
HU X, WANG S P, HAN L. Modeling and simulation on pressure distribution of plane port pair in axial piston pump[J]. Hydraulics Pneumatics & Seals, 2012, 32(8): 68-71(in Chinese).

[13] WEGNER S, GELS S, JANG D S, et al. Experimental investigation of the cylinder block movement in an axial piston machine[C]//Proceedings of the ASME/BATH 2015 Symposium on Fluid Power and Motion Control. New York: ASME, 2016.

[14] 闫伟鹏. 轴向柱塞泵配流副抗倾覆稳定性及油膜润滑特性研究[D]. 哈尔滨: 哈尔滨工业大学, 2020.
YAN W P. Research on stability against tipping and lubrication characteristics of cylinder block/valve plate in axial piston pumps [D]. Harbin: Harbin Institute of Technology, 2020(in Chinese).

[15] XU H G, ZHANG J H, SUN G M, et al. The direct measurement of the cylinder block dynamic characteristics based on a non-contact method in an axial piston pump[J]. Measurement, 2021, 167: 108279.

[16] 张英臣, 卢岳良, 夏天翔, 等. 恒压式变量柱塞泵建模及特性分析[C]//第八届中国航空学会青年科技论坛. 北京: 中国航空学会, 2018: 610-616.
ZHANG Y C, LU Y L, XIA T X, et al. Modeling and analyzing of the characteristics of the constant-pressure variable plunger pump [C]//Proceedings of the 8th China Aviation Society Youth Science and Technology Forum. Beijing: Chinese Society of Aeronautics and Astronautics, 2018: 610-616(in Chinese).

[17] 辛永发. 基于刚-柔-液耦合的轴向柱塞泵缸体振动特性分析[D]. 兰州: 兰州理工大学, 2022.
XIN Y F. Analysis of vibration characteristics of axial piston pump cylinder based on rigid-flex-liquid coupling[D]. Lanzhou: Lanzhou University of Technology, 2022(in Chinese).

[18] 邓海顺, 黄坤, 王传礼, 等. 平衡式两排轴向柱塞泵缸体倾覆力矩[J]. 吉林大学学报 (工学版), 2015, 45(5): 1468-1473.
DENG H S, HUANG K, WANG C L, et al. Overturning moment of cylinder of balanced two-ring axial piston pump[J]. Journal of Jilin University (Engineering and Technology Edition), 2015, 45(5): 1468-1473(in Chinese).

[19] 李壮云. 液压元件与系统[M]. 3 版. 北京: 机械工业出版社, 2011.
LI Z Y. Hydraulic components and systems[M]. 3rd ed. Beijing: China Mechine Press, 2011(in Chinese).

[20] 李晶, 吴双伟. 轴向柱塞泵配流副楔形油膜温度特性[J]. 中国工程机械学报, 2019, 17(1): 8-14.
LI J, WU S W. Temperature characteristics of plane port pair of axial piston pump[J]. Chinese Journal of Construction Machinery, 2019, 17(1): 8-14(in Chinese).

[21] YANG H J, JIANG J H, CHEN G P, et al. Dynamic load identification based on deep convolution neural network[J]. Mechanical Systems and Signal Processing, 2023, 185: 109757.

[22] 张昊冉, 陈国平, 何欢. C/SiC 复合材料结构件热模态模型建模和修正方法[J]. 振动工程学报, 2023, 36(6): 1546-1554.
ZHANG H R, CHEN G P, HE H. Modeling and modification methods of thermal modal model of C/SiC composite structure[J]. Journal of Vibration Engineering, 2023, 36(6): 1546-1554(in Chinese).

[23] HAGAN M T, DEMUTH H B, BEALE M H. 神经网络设计[M]. 戴葵译. 北京: 机械工业出版社, 2002.
HAGAN M T, PEMUTH H B, BEALE M H. Neural network design[M]. DAI K translated. Beijing: China Mechine Press, 2002(in Chinese).

Simulation and experimental research for overturning behavior of aviation plunger pump cylinder bodies

GU Jianning[1, 2, *], ZHANG Xiaobo[1], XIA Tianxiang[1], WANG Tianzhao[3]

(1. AVIC Nanjing Engineering Institute of Aircraft System, Nanjing 211106, China;
2. Aviation Key Laboratory of Science and Technology on Aero Electromechanical System Integration, Nanjing 211106, China;
3. State Key Laboratory of Fluid Power Fundamentals and Mechatronic System, Zhejiang University, Hangzhou 310027, China)

Abstract: The aviation plunger pump is the core of the aviation hydraulic system, and excessive overturning of its internal parts may cause the performance of the whole system to degrade and fail. How to improve the stability of the aviation plunger pump and reduce the overturning of the parts has been the key issue in the research of its high-pressure and high-speed development. The impact of outlet pressure and plunger pump speed on cylinder overturning was investigated in this article, which created a data-driven high-fidelity dynamics model from the viewpoint of cylinder vibration displacement. Then, an improvement measure to reduce the waist groove size of the distribution plate was proposed and verified through numerical simulation. The results show that reducing the size of the waist groove on the distribution plate can improve the overturning of the cylinder body. The revised model can reflect the actual tilting state, and the relative error with the experimental measurement results is within 5%. The research findings offer theoretical and experimental support for the reliable manufacturing of batch production, as well as research suggestions for the simulation and stability studies of aeronautical piston pumps.

Keywords: aviation plunger pump; cylinder overturning; vibration displacement; co-simulation; model updating

Received: 2024-01-22; **Accepted**: 2024-04-02; **Published Online**: 2024-04-24 20:01
URL: link.cnki.net/urlid/11.2625.V.20240423.1829.002
Foundation item: Aeronautical Science Foundation of China (20182876003)
* **Corresponding author**. E-mail: www.1198@foxmail.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0138

一维阵列多液滴蒸发模型及实验

韩琪炜[1]，金捷[1, 2, 3]，赖根鸿[1]，黄俊琦[1]，王方[1, 2, 3, *]

(1. 北京航空航天大学 能源与动力工程学院，北京 100191；2. 北京航空航天大学江西研究院，南昌 330096；
3. 北京航空航天大学成都航空动力创新研究院，成都 611930)

摘 要：在航空发动机燃烧室内，燃料液滴群蒸发时往往受液滴间相互作用影响，然而单液滴蒸发特性难以表征液滴群蒸发状态，因此，需要进行多液滴蒸发实验，探究多液滴蒸发机理。搭建实验平台，采用挂滴法开展航空煤油/甘油的双液滴和直线排列三液滴蒸发实验，结合液滴蒸发理论和实验数据，建立液滴蒸发模型。结果表明：多液滴蒸发速率随着液滴间距增长而上升，趋近于单液滴蒸发速率，蒸发速率和液滴间距呈指数关系，模型整体精度较高。实验所得实验数据和提出的液滴蒸发模型，有效补充了多液滴蒸发特性的研究内容，并对液滴群蒸发特性的研究和航空发动机燃烧室的设计有着积极的意义。

关 键 词：多液滴直线排列；蒸发特性；液滴蒸发模型；液滴间距；航空煤油
中图分类号：V231.2.3
文献标志码：A **文章编号**：1001-5965(2026)05-1680-11

航空发动机燃烧室中化学燃料的燃烧为飞行器动力的来源，而燃烧过程的排放物是造成环境污染的主要原因[1]。提高燃烧效率，降低污染物排放的重要途径之一便是提高雾化和喷雾的质量[2]。在燃料液体经过喷嘴或甩油盘后会被雾化成大量液滴形成液滴群。以往对喷雾中液滴的蒸发研究主要集中在单液滴蒸发实验和蒸发模型的建立，亦或是液滴群蒸发的仿真计算。

单液滴蒸发模型的研究较为成熟。1953年文献[3-4]提出的D^2模型是最早最基础的单液滴蒸发模型；后续学者在D^2模型上针对液滴内部温度传导速率是否无穷大，液滴内部的扩散是否受限制等问题进行研究，相继出现无限传导模型[5-6]/有限传导模型[7-8]；上述的3种模型是基于球对称假设下适用于无对流液滴蒸发的情况，此外单液滴蒸发模型还有Ranz-Marshall模型[9]，有效导热模型[10]及考虑自然对流的厚交换层模型[11](natural convection thick exchange layer, NC-TEL)等。

燃烧室中距离喷嘴较远处，喷雾中液滴群的液滴间距较大，液滴间相互作用可忽略不计，利用单液滴蒸发特性表征液雾蒸发特性是可行的；但是在喷雾距离喷嘴较近处，液滴间相互作用对于液滴蒸发影响不可忽视[12]。因为直接研究喷雾的蒸发特性在技术上困难较大，所以通过研究多液滴蒸发特性来反映液滴间相互作用，进而解释喷雾蒸发中的特殊现象有着较高的可行性。

对于多液滴蒸发研究的主要手段为计算仿真。在对多液滴蒸发工况进行仿真[13-15]后发现：多液滴蒸发会使液滴质量损失率降低，令气体浓度分布和温度场在每个液滴周围产生波动。不同排列方式、不同组分或不同数量的液滴等均是造成液滴间相互作用的重要影响因子。仿真计算在一定程度上解释了多液滴蒸发机理，但是缺乏实验数据的支撑使其结果可信度大大降低。

多液滴蒸发实验数量较少，且大多集中在双滴亦或是低温条件下的非燃料液滴，难以支撑多液滴

收稿日期：2024-03-11；**录用日期**：2024-04-23；**网络出版时间**：2024-05-06 18:52
网络出版地址：link.cnki.net/urlid/11.2625.V.20240430.1628.003
基金项目：国家自然科学基金 (20232ACB204026,12172345,92041001)
* **通信作者**. E-mail：fwang@buaa.edu.cn

引用格式：*韩琪炜，金捷，赖根鸿，等. 一维阵列多液滴蒸发模型及实验 [J]. 北京航空航天大学学报，2026，52（5）：1680-1690.*
HAN Q W，JIN J，LAI G H，et al. Experimental study of multi-droplet evaporation model of one-dimensional array [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（5）：1680-1690（in Chinese）.

蒸发模型的建立。文献 [16] 将液体排列成有序阵列和无序阵列, 研究其相互之间作用关系, 并提出了瞬态蒸发模型。其仿真结果大致符合文献 [17] 进行的疏水基板上排列的多个水滴蒸发的实验结果。文献 [18] 从理论和实验方面研究了强迫对流下双液滴阵列的蒸发燃烧特性, 实验结果与数值模拟结果大致符合。文献 [19-20] 采用飞滴法对直线液滴阵列在自由落体情况下的蒸发特性进行了研究, 并对实验数据进行了拟合, 得出了经验公式。

上述实验和仿真对液滴间相互作用有了一定研究, 但在液滴组分选择和液滴蒸发环境方面无法反映燃烧室内燃料液滴蒸发状况。为了进一步研究燃烧室内液滴蒸发特性和液滴间相互作用, 本文进行了静止状态下的航空煤油/甘油直线排列多液滴蒸发实验研究。其中, 选择航空煤油可以真实反映燃烧室内液滴蒸发情况; 选择甘油则是因为甘油为单组分液体, 蒸发稳定且蒸发特性明显。将 2 种液滴蒸发特性进行对比, 可避免航空煤油的多组分液体的性质对液滴蒸发规律造成干扰, 使本文所得规律更有普适性和可靠性。此外, 本文从液滴蒸发理论入手, 分析了造成液滴间相互作用的原因, 并结合实验数据进行一维阵列多液滴蒸发模型实验研究。

1　实验装置

本文实验所使用的实验装置如图 1 所示, 包括加热炉、挂滴装置、电偶丝、光源、高速摄影机等设备。实验采用十字交叉挂滴法, 将 2 条 0.04 mm 粗的玻璃纤维交叉缠绕在不锈钢丝边框上, 形成挂滴装置, 其交点用以悬挂液滴。加热炉内填充纳米隔热材料, 内置电阻丝用于加热, 外部有金属外壳用以保护。加热炉侧面正对光源和摄影机, 并有开口用以放置石英玻璃, 起到隔热和观测液滴的作用, 下方开口用以送给液滴。挂滴平移装置由挂滴装置和同步带滑台组成。液滴周边放置 2 根间距为 2 cm 的热电偶用于测量温度, 通过调节加热丝功率调节旋钮调整各电阻丝的功率大小, 确保液滴周围温度稳定在预设值, 2 根热电偶所测温度误差需要小于 3 K。光源选择为高功率 LED 灯, 用来提供足够的光照, 保证拍摄质量。高速摄影机为 Phantom 公司生产的 V1212, 其最大拍摄帧率 10 000 Hz, 实验中综合考虑液滴蒸发拍摄效果和数据存储容量限制, 最终选择拍摄帧率 100 Hz。实验时, 先用微量注射泵挂滴, 再通过挂滴平移装置将液滴途径加热炉下方开口送入加热炉内, 使用高速摄影机拍摄液滴蒸发过程, 最后, 利用自编的数据处理软件进行数据处理。

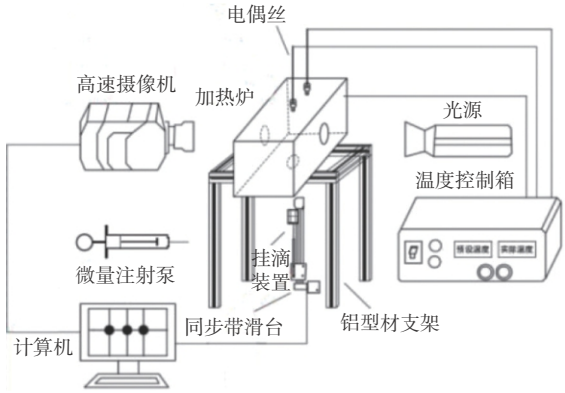


图 1　蒸发实验平台

Fig. 1　Evaporation experimental platform

2　实验内容

实验试剂为“翁江试剂”生产的 RP3-航空煤油和甘油。液滴初始直径 d_0 为 1.1±0.1 mm, 液滴初始温度为室温, 约为 298 K; 环境压力约为 0.1 MPa, 即大气压力; 经过精确标定的热电偶精度为 1 K; 挂滴用不锈钢边框的粗细为 2.00 mm。

本文实验内容为 2 液滴与直线排列 3 液滴在 400～600 ℃ 下的蒸发实验。如图 2 所示, 直线排列 3 液滴临近 2 滴的液滴间距相等, s 为液滴间距, d_0 为液滴初始直径。实验通过改变液滴间距与液滴初始直径之比 s/d_0 和液滴周边温度 T, 根据液滴蒸发特性曲线和蒸发速率的变化, 从而推导出直线排列多液滴蒸发的规律。

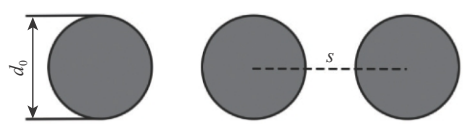


图 2　直线排列 3 液滴

Fig. 2　Three droplets in a straight line

3　结果与讨论

3.1　量、单位和公式

根据 1953 年提出的经典 D^2 定律[3-4], 在液滴稳定蒸发阶段, 液滴直径的平方 d^2 和时间呈线性递减, 其斜率的绝对值即为液滴的蒸发速率 C_v:

$$C_v = \frac{\mathrm{d}\left(d^2\right)}{\mathrm{d}t} \tag{1}$$

根据式 (1) 不难发现, 求得液滴蒸发速率需要 2 个参数, 液滴直径 d 和蒸发时间 t。蒸发时间 t 可由拍摄帧率和图像数量计算得出, 液滴直径 d 则需要对拍摄图像进行处理。实验中的标尺为 2.00 mm 粗的不锈钢边框, 通过测量图像中边框法向方向的像素数量, 得到单位像素对应实际中的尺寸, 将其

乘以液滴直径处像素数得出液滴真实直径。图3展示了图像处理方法：由于实验方法为挂滴法，液滴和玻璃纤维的接触面会不可避免地产生形变，导致其形态在忽略了玻璃纤维引起的变形后呈现椭圆形而非圆形。因此，需要先用Photoshop软件消去玻璃纤维和液滴变形的部分，再通过等面积法求得液滴的当量直径 d_{equ}：

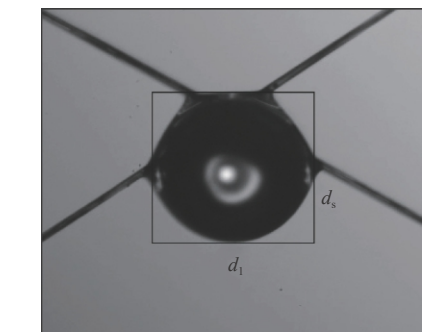
$$S_{\text{Ellipsoid}} = \pi d_l d_s / 4 \tag{2}$$

$$S_{\text{Spheroid}} = \pi d^2 / 4 \tag{3}$$

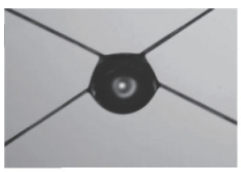
利用等面积法可求得当量直径：

$$d_{equ} = (d_s d_l)^{\frac{1}{2}} \tag{4}$$

式中：d_s 为短轴直径；d_l 为长轴直径；d_{equ} 为当量直径。将图像处理完成后导入自编的程序，便可得到归一化液滴直径的平方 d^2/d_0^2 与归一化时间 t/d_0^2 的关系图，又称液滴蒸发特性曲线图。采用归一化可以忽略初始直径的误差，将多个液滴蒸发特性曲线整合入一张图进行对比，观察其蒸发快慢。此外本文选用液滴初始直径和文献[21]不同，进行归一化可以对两者蒸发快慢进行对比。



(a) 液滴的长轴和短轴



(b1) 原始图像



(b2) 去除玻璃纤维干扰



(b3) 二值化



(b4) 等面积法

(b) 等效直径计算方法

图3 数据处理方法

Fig. 3 Data processing method

本文通过控制微量注射器内液体量来保证液滴初始直径 d_0 相同。如图4所示，本文选定的 d_0 范围为1.1±0.10 mm，在20次挂滴试验中，计算得出 d_0 均在误差允许的范围内。实际操作中微量注射器每次推送液体量为1 μL，根据球体体积公式可计算得出理论上 d_0=1.24 mm。考虑到挂滴过程中液滴的蒸发及注射器针尖残留的液体，导致了实际液滴直径略小于理论值。因此，本文采用的数据处理方法具有的可靠性，计算得到液滴直径较精确，并尽可能保持 d_0 相同。

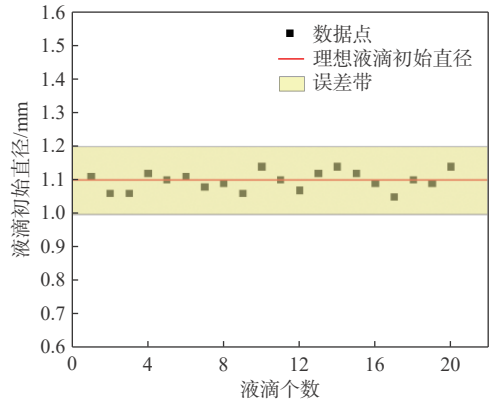


图4 液滴初始直径误差

Fig. 4 Initial droplet diameter error

为验证实验平台可靠性，本文先进行了正庚烷单液滴蒸发实验，并与文献[21]的实验相对比。两者环境压力均为0.1 MPa，液滴周边温度分别为300 ℃、400 ℃、500 ℃和600 ℃。文献[21]选用液滴初始直径 d_0=1.0±0.11 mm，略小于本文实验选用。

根据图5可知，液滴蒸发有2个阶段，第1阶段为初始蒸发阶段，液滴直径先增大后减小，这是由于液滴受热膨胀导致的；第2阶段为稳定蒸发阶段，符合 D^2 蒸发定律，d^2/d_0^2 和 t/d_0^2 呈线性关系，两者斜率的绝对值为蒸发速率 C_v。

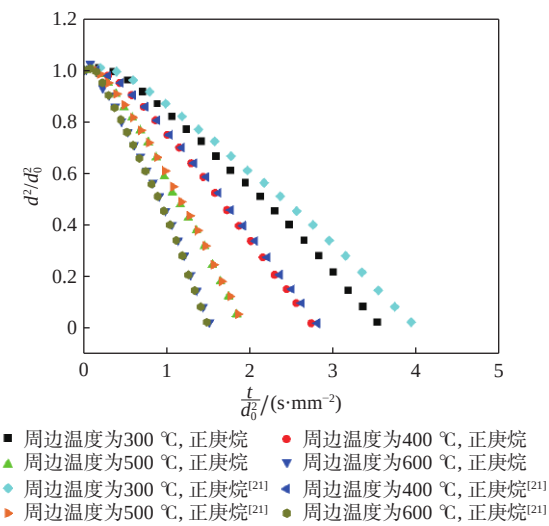


图5 单滴正庚烷蒸发

Fig. 5 A single drop of n-heptane evaporation

文献[21]与本文实验的蒸发特性曲线基本重合，相对误差不高于10%。这表明本文采用的实验平台可靠性较高，实验数据准确。

3.2 实验结果和分析

3.2.1 单液滴蒸发

本文先进行了静止状态下的单液滴蒸发实验，

并将实验数据与传统的 D^2 蒸发模型、R-M 蒸发模型、A-S 蒸发模型和 NC-TEL 模型进行对比。

对于 D^2 模型, 其蒸发速率 C_v 定义为[3-4]

$$C_v=\frac{8\lambda_g}{\rho_l c_{pg}}\ln(1+B_T) \tag{5}$$

$$B_T=c_{pg}\left(T_g-T_s\right)/h_{fg} \tag{6}$$

式中: ρ_l为液体密度; λ_g为周围气体热导率; c_{pg}为周围气体比定压热容; T_s为液体表面温度; T_g为周围气体温度; h_{fg}为蒸发潜热。

在 D^2 模型基础上结合膜理论, 考虑液滴表面的吹风效应。A-S 模型[10] 的蒸发速率 C_v 为

$$\begin{cases}C_v=\dfrac{4\bar{\lambda}_g}{\rho_l\bar{c}_{pg}}Nu^*\ln(1+B_T)\\ Nu^*=2+(Nu_0-2)/F_T\\ Nu_0=2+0.552Re^{1/2}Pr^{1/3}\\ F_T=(1+B_T)^{0.7}\dfrac{\ln(1+B_T)}{B_T}\end{cases} \tag{7}$$

式中: Pr 为普朗特数; Re 为雷诺数; Nu 为努塞尔数; Nu^*为 Nu 的类归一化形式。

R-M 模型[9] 是商用仿真软件 Fluent 的常用液滴蒸发模型, 其蒸发速率 C_v 为

$$C_v=\frac{4\lambda_g}{\rho_l c_{pg}}\left(2+0.6Re^{1/2}Pr^{1/3}\right)\ln(1+B_T) \tag{8}$$

NC-TEL 模型考虑了自然对流对液滴蒸发产生的影响, 在自然对流条件下有着优秀的预测效果。根据文献 [11], 蒸发速率定义 C_v 为

$$C_v=\frac{4\theta_w\lambda_w}{\rho_l c_{pg}}\left[\frac{2\ln(1+\theta_w)}{\theta_w}\sqrt{\frac{T_g-T_s}{2T_w}}+\frac{2}{3\zeta}\cdot\frac{\theta_w^2}{1+\theta_w}Re^{1/3}Pr^{1/3}+\frac{2}{3\delta}\cdot\frac{\theta_w^2}{1+\theta_w}Ra^{1/3}Pr^{1/3}\right] \tag{9}$$

$$T_w=T_g+\left(T_s-T_g\right)/3 \tag{10}$$

$$\theta_w=c_{pg}\left(T_g-T_w\right)/h_{fg} \tag{11}$$

$$\lambda_w=\lambda_g+\left(\lambda_s-\lambda_g\right)/3 \tag{12}$$

式中: λ_s为液体表面热导率; T_s为液体沸点; Ra 为瑞利数; ζ和δ为关于燃料种类的自由参数。

图 6 为蒸发速率与温度关系图。其中, 蒸发模型中所需的物性参数中甘油部分由 ap1700 平台计算得来, RP-3 航空煤油部分参考文献 [22-26]。一般甘油蒸发速率慢于航空煤油, C_v 与 T 正相关。对于航空煤油 NC-TEL 模型对于 C_v 的预测值与实验数据较为贴合, 而其余液滴蒸发模型预测值与实验数据有较大出入。而对于甘油 2 种模型的预测值与实验数据都很贴近。

这是因为甘油是单组份液体, 符合有关模型单组分液体的假设, 而航空煤油的多组分性质导致了

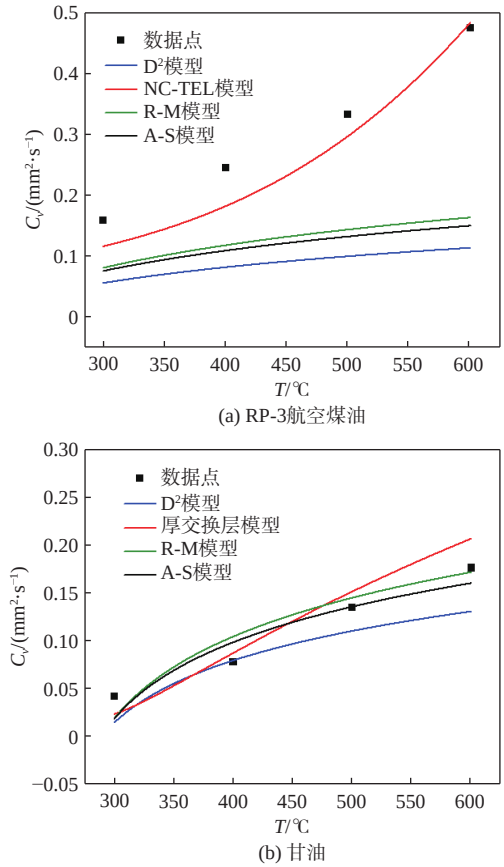


图 6　蒸发速率和温度关系图

Fig. 6　Evaporation rate and temperature diagram

NC-TEL 模型以外模型的预测效果变差。此外, 实验加热炉内空气不是完全静止, 存在自然对流, 所以 NC-TEL 模型能对航空煤油和甘油的单液滴蒸发速率有着良好的预测效果。

3.2.2　多液滴蒸发

本文以 400 ℃ 甘油液滴为例, 展示 2 液滴和直线排列 3 液滴实验的蒸发特性曲线的特征。

图 7 为 400 ℃ 下, s/d_0=1.23 时甘油双液滴蒸发特性曲线图。可以发现, 甘油的初始蒸发阶段持续时间较长, 稳定蒸发阶段十分符合 D^2 定律。双液

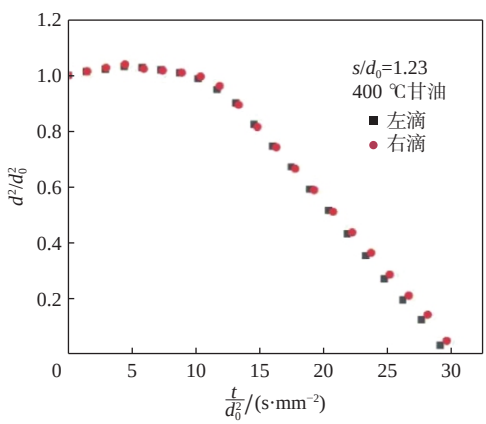
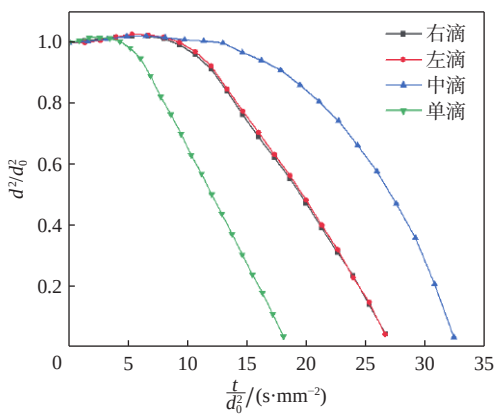


图 7　400 ℃ 甘油双液滴蒸发

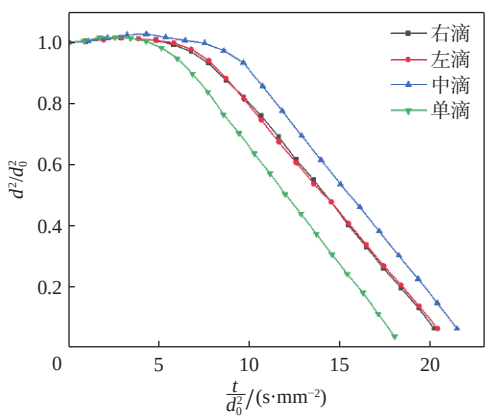
Fig. 7　Glycerol double droplets evaporation at 400 ℃

滴蒸发时，左右液滴互相对称，蒸发环境几乎相同；从实验结果来看，左右液滴蒸发特性曲线基本重合，实验数据相差极小。因此，考虑直接对左右滴直径取均方根，并与蒸发时间相对应，得出平均蒸发特性曲线。后续直接用左右滴的平均蒸发特性曲线来表征双液滴的蒸发特性。此外，航空煤油和其他工况下的甘油左右2滴的蒸发特性曲线重合度也很高，本文不再展示更多图例。

图8显示了400 ℃甘油在s/d_0=1.51和3.96的情况下，左右中滴和单液滴蒸发特性曲线的对比。与2液滴蒸发实验相同，左右滴的蒸发特性曲线基本重合，拥有清晰的初始蒸发阶段和稳定蒸发阶段。而中滴的蒸发特性曲线在不同液滴间距下有着不同的表现。当液滴间距较大时，中滴的蒸发特性曲线接近边滴的蒸发特性曲线，也具有清晰的初始蒸发阶段和稳定蒸发阶段。当液滴间距较小时，中滴的蒸发特性曲线位于边滴的蒸发特性曲线的右侧较远处。在边滴蒸发结束前，中滴蒸发速率很小；当边滴蒸发结束后，中滴蒸发速率迅速变大，最终停留在单液滴和边滴的蒸发速率之间。随着液滴间距升高，左中右3滴的蒸发特性曲线均向左移动，趋近单液滴蒸发特性曲线。



(a) s/d_0=1.51



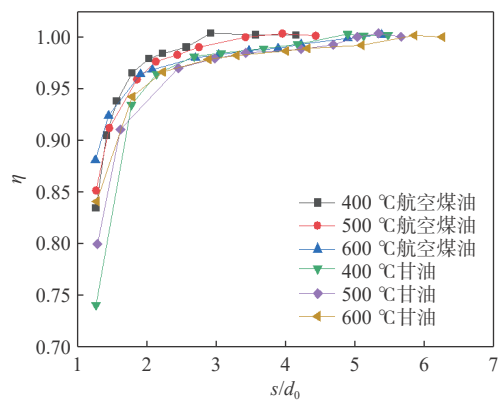
(b) s/d_0=3.96

图8 400 ℃甘油3液滴蒸发

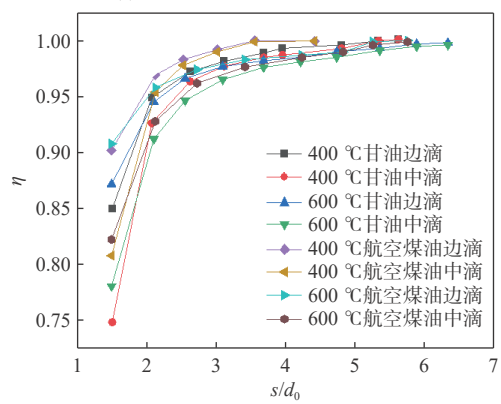
Fig. 8 Evaporation of glycerol triacylglycerol three-droplets at 400 ℃

对于左右滴，处理方法同双液滴蒸发实验，用边滴的蒸发特性曲线表征左右滴的蒸发特性曲线。对于中滴，考虑到初始蒸发阶段产生原因是液滴受热膨胀，而3液滴同时进入加热炉，拥有相同的加热时间。因此，本文认为中滴和边滴进入蒸发稳定阶段的时间是相同的。对中滴的蒸发稳定阶段直接进行线性拟合得出中滴的蒸发速率。

定义归一化蒸发速率η为多液滴蒸发速率和单液滴蒸发速率之比。如图9为双液滴和直线排列3液滴的η和s/d_0的关系图，其中，2液滴蒸发时$s/d_0\in[1.23,6.24]$；3液滴蒸发时$s/d_0\in[1.50,6.39]$。在任何工况下，无论是双液滴抑或是3液滴蒸发实验，η和s/d_0呈正相关并趋近于1，液滴间相互作用随着s增大而减小，直至可以忽略，η和s/d_0的变化规律与指数函数极为类似。此外，对于双液滴和直线排列3液滴蒸发符合以下规律：温度越高，最小液滴间距处的η越高；甘油最小液滴间距的η略高于航空煤油；温度越高η趋近于1的速度越慢；甘油η趋近于1的速度略慢于航空煤油；3液滴蒸发时，中滴蒸发速率总是低于边滴。



(a) 400 ℃、500 ℃和600 ℃下2液滴蒸发



(b) 400 ℃和600 ℃下直线排列3液滴蒸发

图9 η和s/d_0的关系图

Fig. 9 Relationship diagram between η and s/d_0

4 理论分析

文献[27]对液滴蒸发提出了4点假设：①液滴

蒸发环境无强迫对流; ②液滴不存在燃烧, 热辐射和热解离的情况; ③液滴表面存在 Stefan 流, 不同相(即液滴蒸汽和环境气体)之间互相扩散; ④液滴形态总是球对称的; 随后, 对连续方程、动量方程、扩散方程和能量方程进行了处理, 得到:

$$G\left[c_{\mathrm{p}}(T-T_{\mathrm{w}})+q_{\mathrm{e}}\right]=4\pi r^2\lambda\frac{\mathrm{d}T}{\mathrm{d}r} \tag{13}$$

$$G(Y_{\mathrm{F}}-1)=4\pi r^2 D\rho\left(\frac{\mathrm{d}Y_{\mathrm{F}}}{\mathrm{d}r}\right) \tag{14}$$

式中: G 为总质量流; r 为距离液滴中心距离; r_{w} 为液滴半径; c_p 为液体定压比热容; T_{w} 为液滴表面温度; T_∞为无穷远处温度; q_{e} 为单位质量吸热量; λ 为导热系数; Y_{F} 为蒸汽质量分数; Y_{Fw} 为液滴表面蒸汽质量分数; $Y_{\mathrm{F}\infty}$为无穷远处蒸汽质量分数; D 为扩散系数; ρ 为密度 (蒸汽质量分数 Y_{F}<1)。

通过对式 (14) 两端进行处理并积分, 可得液滴周围蒸汽质量分数表达式:

$$\frac{\mathrm{d}Y_{\mathrm{F}}}{(1-Y_{\mathrm{F}})}=-\frac{G\mathrm{d}r}{4\pi r^2 D\rho}\xrightarrow{\text{两边积分}}\ln(1-Y_{\mathrm{F}})=\frac{k_1}{r}\Rightarrow$$
$$Y_F=1-\mathrm{e}^{\frac{k_1}{r}} \tag{15}$$

式中: $k_1=G/(4\pi D\rho)$为正常数。

文献 [27] 也给出了蒸发速率 C_v 的表达式:

$$C_v=8\overline{D}\ln\left(1+\frac{Y_{\mathrm{Fw}}-Y_{\mathrm{F}\infty}}{1-Y_{\mathrm{Fw}}}\right) \tag{16}$$

$$C_v=\frac{8\overline{\lambda}}{\overline{\rho}c_{\mathrm{p}}}\ln\left[1+\frac{c_{\mathrm{p}}(T_\infty-T_{\mathrm{w}})}{q_{\mathrm{e}}}\right] \tag{17}$$

从式 (16) 和式 (17) 不难发现, 蒸发速率 C_v 主要和蒸汽质量分数 Y_{F} 或温度 T 相关, C_v 和 T_∞呈正相关和 $Y_{\mathrm{F}\infty}$呈负相关, 符合实验结论。其中, 其他液滴对受影响液滴的周边温度的改变较难分析, 但是对蒸汽质量分数的影响可以由单液滴周边蒸汽质量分数分布推导, 因此, 本文从蒸汽质量分数 Y_{F} 入手。在多液滴蒸发实验中, $Y_{\mathrm{F}\infty}$的最高点在液滴连线的中心点, 可用式 (15) 来模拟此点的蒸汽质量分数 $Y^1_{\mathrm{F}\infty}=1-\mathrm{e}^{2k_1/s}$, 又因为其余方向的 $Y_{\mathrm{F}\infty}$远低于该位置, 因此, 最终可得 $Y_{\mathrm{F}\infty}=k_2Y^1_{\mathrm{F}\infty}$ ($0<k_2<1$)。$Y_{\mathrm{F}\infty}$和液滴间距负相关, 因此, C_v 和液滴间距正相关, 液滴间距越远, 蒸发速率越高, 且无限趋近于单液滴蒸发速率 $C_v=8\overline{D}\ln[1+Y_{\mathrm{Fw}}/(1-Y_{\mathrm{Fw}})]$, 和实验数据吻合。

此外, 文献 [16] 认为在多液滴蒸发时, 液滴受到的影响中, 距离相近液滴影响强于距离较远的液滴, 并给出了液滴间互相作用加权因子 w。定义 1-η 为受抑制数, 表现液滴间相互作用影响大小。若多液滴中共有 n 个液滴, 第 k 个液滴对第 i 个液滴的加权因子为

$$w_{ik}=w_{ki}=\frac{d_0/s_{ik}}{\sum_{k=1,k\neq i}^{n-1}d_0/s_{ik}}\quad 1\leqslant i\leqslant n,1\leqslant k\leqslant n \tag{18}$$

2 液滴和直线排列 3 液滴排列中, 液滴之间的加权因子 w 如表 1 所示。

表 1　加权因子 w

Table 1　Weighting factor w

排列方式	影响→被影响液滴	加权因子
2液滴		1
3液滴	边滴→中滴	1/2
	边滴→边滴	1/3
	中滴→边滴	2/3

表 1 中, 边滴对中滴和中滴对边滴的间距相同而加权因子数值不同, 因此, 需要进行归一化处理。受抑制数表征两两液滴间的蒸发抑制作用, 若要得到目标液滴蒸发蒸发被抑制的总体大小, 需要将加权因子乘以每 2 个液滴之间的受抑制数 ($1-\eta_i$) 并相加, 最后, 除以加权因子的最大值, 可得其他所有液滴对该液滴的综合受抑制数, 因此, 可以得到某一液滴的归一化蒸发速率 η:

$$\eta=1-\frac{\prod_{k=1,k\neq i}^{n}w_{ik}(1-\eta_k)}{\max\limits_{k=1,k\neq i}(w_{ik})} \tag{19}$$

至此, 理论上可以通过式 (16) 和式 (17) 求得双液滴蒸发时的归一化蒸发速率 η, 代入式 (18) 和式 (19) 得出直线排列多液滴蒸发时各液滴的 η。

5　液滴蒸发模型与实验数据对比

本文于第 3 节对比了多种液滴蒸发模型, 其中, NC-TEL 模型对各种组分各个温度与实验数据贴合最好。随后对双液滴蒸发实验数据的拟合可以得出单个液滴对其他液滴蒸发速率的影响。最后, 将液滴间相互作用扩展至多个液滴, 便可以得到多液滴蒸发模型。

由于式 (16) 和式 (17) 中所需的多液滴蒸发的温度场和蒸汽浓度场很难用数学方法定量表达, 因此, 本文考虑结合实验数据得出液滴蒸发模型。根据第 3 节的分析可得: 液滴的蒸发速率和液滴间距呈指数关系, 因此, 在选择液滴蒸发模型拟合公式时尽量选择指数函数。归一化蒸发速率 η 为多液滴蒸发速率和单液滴蒸发速率之比, 结合式 (15) 和式 (16)。其中, 单液滴蒸发时无穷远处蒸汽浓度 $Y_{\mathrm{F}\infty}$=0, 多液滴蒸发时无穷远处蒸汽浓度由前文所得: $Y_{\mathrm{F}\infty}=k_2Y'_{\mathrm{F}\infty}$, 其具体表达式为

$$\eta=\frac{\ln\left(1+\dfrac{Y_{\mathrm{Fw}}-Y'_{\mathrm{F}\infty}}{1-Y_{\mathrm{Fw}}}\right)}{\ln\left(1+\dfrac{Y_{\mathrm{Fw}}}{1-Y_{\mathrm{Fw}}}\right)}\approx\frac{\dfrac{Y_{\mathrm{Fw}}-Y'_{\mathrm{F}\infty}}{1-Y_{\mathrm{Fw}}}}{\ln\left(1+\dfrac{Y_{\mathrm{Fw}}}{1-Y_{\mathrm{Fw}}}\right)}=\frac{Y_{\mathrm{Fw}}-k_2\left(1-\mathrm{e}^{\frac{2k_1}{s}}\right)}{\ln\left(1+\dfrac{Y_{\mathrm{Fw}}}{1-Y_{\mathrm{Fw}}}\right)(1-Y_{\mathrm{Fw}})} \tag{20}$$

式中：s、Y_{Fw}、k_1 和 k_2 均为常数，因此，可以得到关于 $\eta(y)$ 和 $s/d_0(x)$ 的关系式：

$$y=a-b\mathrm{e}^{c/x} \tag{21}$$

式中：参数 a,b,c 均为常数。理论上，液滴间距越大，η 越趋近于 1，因此，式 (21) 存在 y=1 的上渐近线，所以可令 a=1+b，b>0。

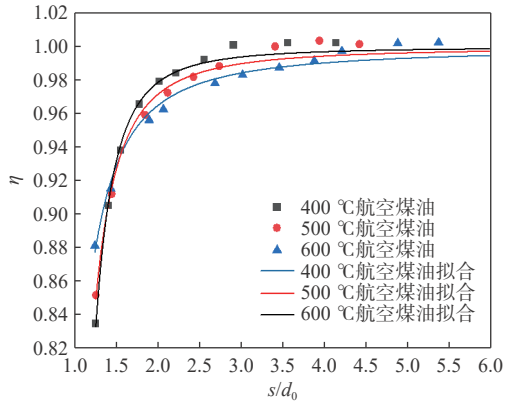
图 10 展示了 400 ℃、500 ℃ 和 600 ℃ 条件下航空煤油和甘油双液滴蒸发的 η 与 s/d_0 的关系。不难发现，拟合曲线的 η 随着 s/d_0 增大而增大并趋近于 1，与数据点变化趋势相同，拟合效果良好。表 2 展示了各种工况下双液滴蒸发的拟合公式，并给出了拟合优度 R^2。

拟合优度 R^2 是一种评估回归模型拟合程度的指标，其取值范围在 0 到 1 之间。R^2 越高证明回归模型与数据点的拟合程度越高，反之则证明模型对数据的拟合程度越差。拟合优度 R^2 的定义如式 (22) 所示。在横坐标为 x，纵坐标为 y 的二维坐标系中，其中，y_i 表示数据点的值，$\hat{y}_i$ 表示数据点与数据点有着相同横坐标的拟合曲线上值的差值，$\bar{y}$ 表示所有数据点的平均值。

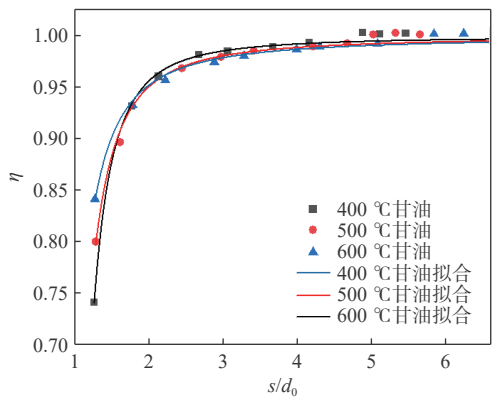
$$\begin{cases}R^2=1-\dfrac{S_{\mathrm{SR}}}{S_{\mathrm{ST}}}\\ S_{\mathrm{SR}}=\displaystyle\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2\\ S_{\mathrm{ST}}=\displaystyle\sum_{i=1}^{n}(y_i-\bar{y})^2\end{cases} \tag{22}$$

式中：S_{SR} 为残差平方和；S_{ST} 为总平方和。

由表 2 可见，对于双液滴蒸发，各工况下拟合优度均大于 0.978 70，可见，拟合公式对航空煤油和甘油双液滴蒸发的数据点拟合效果良好。



(a) 400 ℃～600 ℃航空煤油2液滴



(b) 400 ℃～600 ℃甘油2液滴

图 10 实验数据和拟合曲线

Fig. 10 Experimental data and fitting curves

2 液滴蒸发的 η 和 s/d_0 关系式反映了两两液滴间的相互作用，结合式 (19) 和式 (21) 可以得到等液滴间距的直线排列多液滴蒸发模型。本文进行了直线排列 3 液滴蒸发实验，可用其实验数据验证该模型准确性。

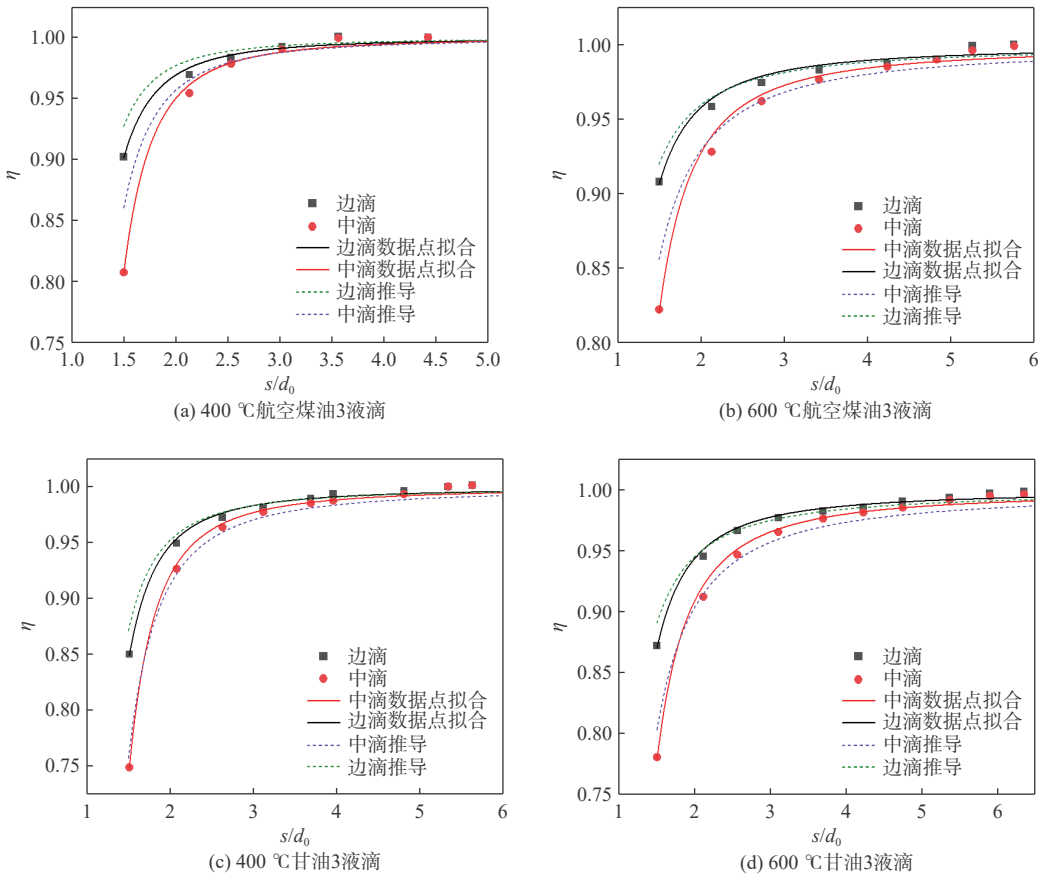
直线排列的 3 液滴的拟合公式同样可以根据式 (21) 对实验数据进行拟合得出。将经 2 液滴蒸发拟合公式推导所得公式与拟合数据点所得公式对比，可对液滴蒸发模型的合理性进行判断。

图 11 为 400 ℃ 和 600 ℃ 下航空煤油和甘油的 η 与 s/d_0 关系，表 3 列出了拟合公式和拟合优度。不难看出，直线排列 3 液滴的拟合公式的拟合优度 R^2 在 0.971 02 之上，与实验数据点有着不错的拟合效果。从 2 液滴蒸发拟合公式推导得来的公式与数

表 2 2 液滴蒸发拟合公式和拟合优度

Table 2 Two-droplet evaporation fitting formula and goodness-of-fit

组分	温度/℃	拟合公式	拟合优度R^2
航空煤油	400	$f(x)=1.000\,733-0.000\,733\mathrm{e}^{6.843\,9/x}$	0.978 70
	500	$f(x)=1.002\,04-0.002\,04\mathrm{e}^{5.434\,0/x}$	0.988 69
	600	$f(x)=1.006\,42-0.006\,42\mathrm{e}^{3.750\,1/x}$	0.994 91
甘油	400	$f(x)=1.002\,43-0.002\,43\mathrm{e}^{5.895\,5/x}$	0.996 14
	500	$f(x)=1.005\,06-0.005\,06\mathrm{e}^{4.761\,6/x}$	0.988 04
	600	$f(x)=1.008\,38-0.008\,38\mathrm{e}^{3.814\,4/x}$	0.985 53



(a) 400 ℃航空煤油3液滴　(b) 600 ℃航空煤油3液滴

(c) 400 ℃甘油3液滴　(d) 600 ℃甘油3液滴

图 11　实验数据和拟合曲线

Fig. 11　Experimental data and fitting curves

据点拟合公式重合度较高, 无较大的偏差。从 2 液滴推导而来的公式在变化率方面略低, 2 类公式具体表达式如表 3 和表 4。

为定量分析 2 类公式的重合度, 本文定义了 2 类公式间相对误差 δ: $\delta=(\eta_1-\eta_2)/\eta_1$($\eta_1$为数据点对应的$\eta$, η_2为拟合曲线上对应的η), $|\delta|<0.01$ 表示低误差。其中, η_1代表针对数据点拟合的η值, η_2表示从双液滴推导出的η值。

如图 12 所示, 经 2 液滴蒸发拟合公式推导所得公式与拟合数据点所得公式的相对误差小于 0.066 4, 曲线非常接近。在最小液滴间距处, 400 ℃ 航空煤油中滴的误差相对较高, 为 0.066 4; 600 ℃ 航空煤油中滴在最小液滴间距下误差最低, 为 0.016 1。相对误差随着液滴间距增大而急速下降至低误差率范围内 (浅黄色带)。在 $s/d_0\in[1.5,10]$ 范围内, 据表 4 可得, 95% 以上的部分$|\delta|<0.01$。因此, 可以得到: 从 2 液滴蒸发推导所得表 4 中公式在中高液滴间距情况下与实验数据的误差几乎为 0, 但在小液滴间距情况下对于液滴蒸发速率的预测同样具有不错的精度。

据第 3 节可知 NC-TEL 模型在预测甘油和航空煤油单液滴蒸发速率方面有着高精度。结合 NC-TEL 模型和归一化蒸发速率 η 计算式, 可以得出一维阵列多液滴蒸发模型, 其表达式如下:

$$C_v=N(T)\cdot\eta(s/d_0) \tag{23}$$

表 3　3 液滴蒸发拟合公式和拟合优度

Table 3　Three-droplet evaporation fitting formula and goodness-of-fit

蒸发条件	拟合公式	拟合优度R^2
400 ℃甘油边滴	$f(x)=1.002\,53-0.002\,53e^{6.959\,77/x}$	0.996 20
400 ℃甘油中滴	$f(x)=1.002\,37-0.002\,37e^{6.295\,22/x}$	0.991 81
600 ℃甘油边滴	$f(x)=1.006-0.006e^{4.676\,51/x}$	0.986 43
600 ℃甘油中滴	$f(x)=1.007\,85-0.007\,85e^{5.060\,98/x}$	0.990 28
400 ℃航空煤油边滴	$f(x)=1.000\,935-0.000\,935e^{7.996\,69/x}$	0.995 57
400 ℃航空煤油中滴	$f(x)=1.001\,08-0.001\,08e^{6.787\,82/x}$	0.987 73
600 ℃航空煤油边滴	$f(x)=1.006\,01-0.006\,01e^{5.146\,85/x}$	0.985 46
600 ℃航空煤油中滴	$f(x)=1.005\,31-0.005\,31e^{4.384\,85/x}$	0.971 02

表4 公式和低误差率

Table 4 formulas and low error rates

蒸发条件	推导公式	低误差率/%
400 ℃甘油边滴	$f(x)=1.001\,099\,5-0.000\,733e^{6.843\,9/x}-0.000\,366\,5e^{6.843\,9/(2x)}$	96.5
400 ℃甘油中滴	$f(x)=1.004\,86-0.004\,86e^{5.895\,5/x}$	99.1
600 ℃甘油边滴	$f(x)=1.003\,645-0.002\,43e^{5.895\,5/x}-0.001\,215e^{5.895\,5/(2x)}$	97.6
600 ℃甘油中滴	$f(x)=1.016\,76-0.016\,76e^{3.814\,4/x}$	98.2
400 ℃航空煤油边滴	$f(x)=1.012\,57-0.008\,38e^{3.814\,4/x}-0.004\,19e^{3.814\,4/(2x)}$	95.0
400 ℃航空煤油中滴	$f(x)=1.001\,766-0.001\,766e^{6.843\,9/x}$	96.2
600 ℃航空煤油边滴	$f(x)=1.009\,63-0.006\,42e^{3.750\,1/x}-0.003\,21e^{3.750\,1/(2x)}$	98.9
600 ℃航空煤油中滴	$f(x)=1.012\,84-0.012\,84e^{3.750\,1/x}$	97.8

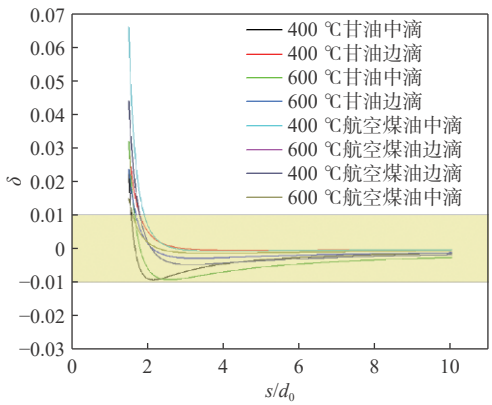


图12 表4中公式相对误差

Fig. 12 Relative error of the formula of table 4

式中：$N(T)$表示以温度 T(即 T_s) 为变量的 NC-TEL 液滴蒸发模型；$\eta(s/d_0)$为本文主要探讨的多液滴蒸发下的归一化蒸发速率。实际应用时先用 NC-TEL 模型求出单液滴蒸发速率，再根据液滴个数和液滴间距计算出其归一化蒸发速率，相乘即可得到多液滴蒸发速率。无论是 NC-TEL 模型还是本文提出的 η 计算公式的预测值与实验数据均十分接近，预测效果良好。

综上所述，本文提出一维阵列多液滴蒸发模型与2液滴和3液滴的蒸发实验数据拟合较好，在预测多液滴蒸发速率方面有一定的可靠性。一定程度上解释了液滴间相互作用的机理，和多液滴中不同液滴间距下液滴之间影响的强弱。

6 结 论

1）对于单液滴蒸发速率，NC-TEL 模型对航空煤油和甘油均有着不错的预测效果，而其余模型在预测甘油时误差尚可，预测航空煤油效果很差。

2）双液滴蒸发左滴和右滴蒸发特性曲线相差很小，与单液滴蒸发特性曲线十分相似，稳定蒸发阶段符合 D^2 定律。直线排列3液滴蒸发中，左滴右滴蒸发特性曲线十分接近，稳定蒸发阶段符合 D^2 定律，但中滴的稳定蒸发阶段蒸发特性曲线斜率的绝对值在边滴蒸发结束后上升。

3）随着液滴间距增长，2液滴和3液滴蒸发速率趋近于单液滴蒸发速率。其中，中滴蒸发速率总是低于左右2滴。温度越高蒸发速率增长速率越慢。

4）本文提出的半经验半理论多液滴蒸发模型，对2液滴和直线排列3液滴蒸发实验数据均有着不错的拟合效果，拟合优度均在0.97以上。

5）根据双液滴推导得来的公式与直接拟合3液滴蒸发数据点公式的相对误差，随着液滴间距增大迅速减小，相对误差控制在0.064以下，其中超过95%的范围相对误差小于1%。

参考文献（References）

[1] 郑晓峰, 冯耀勋, 贾明生. 富氧燃烧的节能特性及其对环境的影响[J]. 节能, 2006, 25(7): 26-28.
ZHENG X F, FENG Y X, JIA M S. Environmental influence and energy-saving characteristic of oxygen-enriched combustion[J]. Energy Conservation, 2006, 25(7): 26-28 (in Chinese).

[2] JIANG X, SIAMAS G A, JAGUS K, et al. Physical modelling and advanced simulations of gas–liquid two-phase jet flows in atomization and sprays[J]. Progress in Energy and Combustion Science, 2010, 36(2): 131-167.

[3] GODSAVE G A E. Studies of the combustion of drops in a fuel spray: the burning of single drops of fuel[J]. Symposium (International) on Combustion, 1953, 4(1): 818-830.

[4] SPALDING D B. Combustion of liquid fuels[J]. Nature, 1950, 165(4187): 160.

[5] LAW C K. Unsteady droplet combustion with droplet heating[J]. Combustion and Flame, 1976, 26: 17-22.

[6] LAW C K, SIRIGNANO W A. Unsteady droplet combustion with droplet heating: II: Conduction limit[J]. Combustion and Flame, 1977, 28: 175-186.

[7] PRAKASH S, SIRIGNANO W A. Theory of convective droplet vaporization with unsteady heat transfer in the circulating liquid phase[J]. International Journal of Heat and Mass Transfer, 1980, 23(3): 253-268.

[8] PRAKASH S, SIRIGNANO W A. Liquid fuel droplet heating with internal circulation[J]. International Journal of Heat and Mass

Transfer, 1978, 21(7): 885-895.

[9] RANZ W E, MARSHALL W R. Evaporation from drops[J]. Chemical Engineering Progress, 1952, 48(173): 141-146.

[10] ABRAMZON B, SIRIGNANO W A. Droplet vaporization model for spray combustion calculations[J]. International Journal of Heat and Mass Transfer, 1989, 32(9): 1605-1618.

[11] WANG F, LIU R, LI M, et al. Kerosene evaporation rate in high temperature air stationary and convective environment[J]. Fuel, 2018, 211: 582-590.

[12] BRZUSTOWSKI T A, TWARDUS E M, WOJCICKI S, et al. Interaction of two burning fuel droplets of arbitrary size[J]. AIAA Journal, 1979, 17(11): 1234-1242.

[13] MIN J L, BAO J R, CHEN W, et al. Numerical simulation of liquid hydrogen droplets "group" evaporation and combustion[J]. Cryogenics, 2020, 108: 103091.

[14] JIANG T L, CHIANG W T. Effects of multiple droplet interaction on droplet vaporization in subcritical and supercritical pressure environments[J]. Combustion and Flame, 1994, 97(1): 17-34.

[15] WRAY A W, DUFFY B R, WILSON S K. Competitive evaporation of multiple sessile droplets[J]. Journal of Fluid Mechanics, 2020, 884: A45.

[16] LIU Q Z, LU R B, QIAO Y L, et al. Analysis of the correction factors and coupling characteristics of multi-droplet evaporation[J]. International Journal of Heat and Mass Transfer, 2022, 195: 123138.

[17] KHILIFI D, FOUDHIL W, FAHEM K, et al. Study of the phenomenon of the interaction between sessile drops during evaporation[J]. Thermal Science, 2019, 23(2 Part B): 1105-1114.

[18] BALAKRISHNAN P, SUNDARARAJAN T, NATARAJAN R. Interference effects during burning of tandem porous spheres in mixed convective environment[J]. AIAA Journal, 2000, 38(10): 1889-1898.

[19] ELKOTB M M, ALY S L, ELSALMAWY H A. Evaporation characteristics of fuel and multifuel droplets[J]. Combustion and Flame, 1991, 85(3-4): 300-308.

[20] MAQUA C, CASTANET G, GRISCH F, et al. Monodisperse droplet heating and evaporation: Experimental study and modelling[J]. International Journal of Heat and Mass Transfer, 2008, 51(15-16): 3932-3945.

[21] JAVED I, BAEK S W, WAHEED K. Evaporation characteristics of heptane droplets with the addition of aluminum nanoparticles at elevated temperatures[J]. Combustion and Flame, 2013, 160(1): 170-183.

[22] DENG H W, ZHANG C B, XU G Q, et al. Density measurements of endothermic hydrocarbon fuel at sub-and supercritical conditions[J]. Journal of Chemical & Engineering Data, 2011, 56(6): 2980-2986.

[23] DENG H W, ZHANG C B, XU G Q, et al. Viscosity measurements of endothermic hydrocarbon fuel from (298 to 788) K under supercritical pressure conditions[J]. Journal of Chemical & Engineering Data, 2012, 57(2): 358-365.

[24] DENG H W, ZHU K, XU G Q, et al. Isobaric specific heat capacity measurement for kerosene RP-3 in the near-critical and supercritical regions[J]. Journal of Chemical & Engineering Data, 2012, 57(2): 263-268.

[25] XU G Q, JIA Z X, WEN J, et al. Thermal-conductivity measurements of aviation kerosene RP-3 from (285 to 513)K at sub- and supercritical pressures[J]. International Journal of Thermophysics, 2015, 36(4): 620-632.

[26] 张志强, 郝毓雅, 陈战斌. 航空喷气燃料 RP-3 运动粘度特性研究[J]. 现代机械, 2020(1): 38-40.
ZHANG Z Q, HAO Y Y, CHEN Z B. Kinematic viscosity characteristics of aviation jet fuel RP-3[J]. Modern Machinery, 2020(1): 38-40 (in Chinese).

[27] 周力行. 燃烧理论和化学流体力学[M]. 北京: 科学普及出版社, 1986.
ZHOU L X. Combustion theory and chemical fluid mechanics[M]. Beijing: Popular Science Press, 1986 (in Chinese).

Experimental study of multi-droplet evaporation model of one-dimensional array

HAN Qiwei[1], JIN Jie[1, 2, 3], LAI Genhong[1], HUANG Junqi[1], WANG Fang[1, 2, 3, *]

(1. School of Energy and Engineering, Beihang University, Beijing 100191, China;
2. Jiangxi Research Institute of Beihang University, Nanchang 330096, China;
3. Chengdu Aerodynamics Institute of Beihang University, Chengdu 611930, China)

Abstract: The evaporation of fuel droplet groups in the combustion chamber of aero engines is frequently influenced by droplet interactions; yet, it is challenging to describe the evaporation state of droplet groups based on the evaporation characteristics of individual droplets. Therefore, it is necessary to carry out experimental research on multi-droplet evaporation to explore the evaporation mechanism of multi-droplets. Firstly, an experimental platform was built and the evaporation experiment of two droplets and three droplets of aviation kerosene/glycerin arranged in a straight line was carried out by the hanging drop method. Then, combined with the theoretical and experimental data of droplet evaporation, the droplet evaporation model was established. The findings demonstrate that the rate at which numerous droplets evaporate increases as droplet spacing increases, approaching the rate at which single droplets evaporate. The evaporation rate and droplet spacing have an exponential relationship, and the overall accuracy of the model is high. In conclusion, the experimental data obtained from the experiment and the proposed droplet evaporation model effectively supplement the research content of multi-droplet evaporation characteristics,which has positive significance for the study of droplet group evaporation characteristics and the design of aero engine combustion chambers.

Keywords: multiple liquid droplets arranged in a straight line; evaporation characteristics; droplet evaporation model; droplet spacing; aviation kerosene

Received: 2024-03-11; **Accepted**: 2024-04-23; **Published Online**: 2024-05-06 18:52

URL: link.cnki.net/urlid/11.2625.V.20240430.1628.003

Foundation items: National Natural Science Foundation of China (20232ACB204026,12172345,92041001)

* **Corresponding author**. E-mail: fwang@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0167

基于平方马氏距离的空间目标碰撞风险分析算法

黄梦蝶，王禄丰，黄旭星，李爽*

(南京航空航天大学 航天学院，南京 211106)

摘　　要：针对在空间目标碰撞风险预报中概率稀释现象导致碰撞概率的判断结果可靠性降低的问题，提出了基于平方马氏距离的碰撞风险评估方法，以提高碰撞风险预报结果的可靠性。基于3σ法则建立平方马氏距离阈值模型，实现碰撞风险的快速判别；对联合球体半径、位置误差、误差椭圆长短轴比及旋转角等因素进行建模和分析，优化平方马氏距离的模型参数，提高平方马氏距离的计算效率；在此基础上，基于平方马氏距离发展碰撞风险预报评估模型，解决概率稀释问题。仿真结果表明：所提方法可有效降低高风险事件的漏检率。

关 键 词：空间目标；碰撞风险分析；平方马氏距离；碰撞概率；3σ法则

中图分类号：V528

文献标志码：A　　**文章编号**：1001-5965(2026)05-1691-10

随着人类航天活动的不断增加，太空中出现了越来越多的飞行器，地外空间正变得越来越拥挤。UCS(Union of Concerned Scientists)卫星数据库的数据显示，截至2023年5月1日，空间中运行的卫星总数已经达到7 560颗，而伴随着多个巨型星座计划的实施，这一数字仍将呈现持续且快速的增长[1]。同时由于航天器碰撞或解体等事件产生了材料、大小和形状各异的碎片，导致空间碎片遍布[2]。在轨航天器时刻面临着与其他空间目标接近甚至碰撞的风险，国际空间站(International Space Station, ISS)于2023年8月6日为避免与宇宙1408碎片相撞，执行了一次避碰机动，这已经是自1999年以来ISS进行的第37次避碰规避[3]。

对未来可能发生的危险碰撞事件进行评估和预测，以提前进行规避，是维护太空安全中非常关键的一环。碰撞概率是在进行碰撞风险评估时的一个常用判据，并被应用于地月轨道、航天器编队，碎片云短期或长期的碰撞风险分析等[4-6]。碰撞概率通过将描述空间目标位置误差的三维高斯椭球向某二维平面投影，并计算该二维密度函数在拟合球内的二重积分来获取[7]。然而，Núñez等[8]证明对于连续的高斯分布概率密度函数，无论向哪个平面投影，计算得到的碰撞概率都会高于真实值，这可能导致高误报率。由于碰撞概率是基于不确定性计算的，因此求取其精确值本身就是不可能的[9]。此外，当误差超过一定水平时，可能出现概率稀释现象，这时因为碰撞概率很小而往往认为相遇是安全的，这样的结论可靠性较低[10]。Stef等[4]在研究近月球轨道的碰撞风险情况时发现，协方差矩阵越大，超过碰撞概率阈值的危险接近事件的数量越多，导致避碰机动频率的增加。

除了碰撞概率外，另一种有效的碰撞风险判据是马氏距离。马氏距离最初由印度统计学家P. C. Mahalanobis提出，其能排除变量之间相关性的干扰，且不具备量纲。在本文背景下，可将其物理意义描述为联合球体中心与协方差椭圆中心的距离相当于多少个标准差[11]。为了研究方便，通常选择马氏距离的平方来进行分析判断。Phillips[12]对比了马氏距离、投影距离、碰撞概率等多种指标，证明了马氏距离的可行性及高计算效率。通常将交

收稿日期：2024-03-26；录用日期：2024-06-25；网络出版时间：2024-07-10 15:15
网络出版地址：link.cnki.net/urlid/11.2625.V.20240708.1109.002
基金项目：空间智能控制技术全国重点实验室开放基金 (HTKJ2024KL502001,HTKJ2024KL502031)
*通信作者. E-mail：lishuang@nuaa.edu.cn

引用格式：黄梦蝶，王禄丰，黄旭星，等. 基于平方马氏距离的空间目标碰撞风险分析算法 [J]. 北京航空航天大学学报，2026，52（5）：1691-1700. HUANG M D，WANG L F，HUANG X X，et al. Space target collision risk analysis algorithm based on the square Mahalanobis distance [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（5）：1691-1700（in Chinese）.

会没有风险却预报有风险的情况称为虚警，将交会有风险却没有被检测到的情况称为漏检，Alfano 等[10]得到了不同半径下，不同马氏距离对应的虚警和漏检概率边界。Modenini 等[13]探讨了碰撞概率、置信区间和平方马氏距离这 3 个度量之间的关系，说明平方马氏距离的最大值和最小值分别对应碰撞概率的最小值和最大值，并给出了求最值的解析表达式。Núñez 等[14]利用马氏距离开发了马氏壳体采样 (Mahalanobis shell sampling, MSS) 算法，对样本的均值和协方差进行无偏估计，使其更接近原始分布，用以计算碰撞概率。然而，这些研究中对平方马氏距离所代表的风险等级的划分方法不够有效和直接，也并未对是否在所有情况下都需要计算平方马氏距离的最值做更多分析。

本文针对碰撞概率稀释导致的漏检问题，基于平方马氏距离提出空间目标碰撞风险分析方法，主要贡献归纳如下：①从实现平方马氏距离风险等级的快速划分出发，根据3σ法则，计算得到了平方马氏距离的安全阈值；②通过对 5 个影响因素进行分析，获得了不需要计算最小值的各因素取值区间，对平方马氏距离的计算方法进行了优化，提高了计算效率；③基于平方马氏距离提出了一套碰撞风险预报评估方法，并对 2 个空间目标的碰撞风险进行了预报分析，验证了该方法的可行性。

1 平方马氏距离基本理论

1.1 平方马氏距离定义

对于空间目标间碰撞概率的计算问题，一般假设空间目标在相遇时存在较大的相对速度，相遇过程非常短暂，相对运动轨迹可以被近似为直线[15-16]。同时，空间目标的位置误差满足均值为零的高斯分布，并且相互独立。空间目标的外形可抽象为半径已知的球体，2 个球体结合形成 1 个大包络球，称为联合球体，其半径为 R。

在这些假设的基础上，可以将 2 个空间目标在地心惯性坐标系下的误差协方差矩阵直接相加，得到表示联合球体位置误差分布的三维误差椭球。将该误差椭球投影到最接近时刻(closest approach, CA)时与相对速度矢量垂直的平面(称为相遇平面)，得到如图 1 所示的二维几何模型。其中，以 t 为中心的圆是联合球体在相遇平面的投影圆 B，以 c 为中心的椭圆代表二维误差椭圆。为了便于计算，在该平面内采用由二维误差椭圆的主轴定义的坐标系，坐标原点位于目标 t，x 轴与椭圆长轴平行。

二维误差椭圆可以用高斯分布 $N(\boldsymbol{d},\boldsymbol{C})$来描述，其中，$\boldsymbol{d}$ 为相对位置矢量，其长度为Δ，$\boldsymbol{C}$ 为联合误差协方差矩阵。将 x 轴与 $\boldsymbol{d}$ 之间的夹角记为θ，定义为旋转角。碰撞概率(collision probability, P_c)为该二维高斯分布概率密度函数在投影圆 B 上的积分：

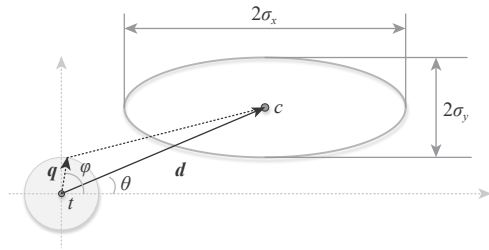


图 1 碰撞事件的二维几何模型

Fig. 1 2-D geometric model of collision event

$$P_c = \frac{1}{2\pi|\boldsymbol{C}|^{1/2}} \iint_B \mathrm{e}^{-(1/2)(\boldsymbol{x}-\boldsymbol{d})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{d})}\mathrm{d}x\mathrm{d}y \tag{1}$$

选取的坐标系使得联合协方差矩阵为对角矩阵：

$$\boldsymbol{C} = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} \tag{2}$$

为了提高碰撞概率的计算速度，已有多位学者对式 (1) 的化简方法进行了研究，应用比较广的是将其化为无穷级数[17-18]。碰撞概率的计算不是本文重点，具体计算方法这里不再赘述。

式 (1) 中被积函数的指数即为投影圆表面的一点到误差椭圆中心的平方马氏距离，即

$$M^2(\boldsymbol{q}) = (\boldsymbol{q}-\boldsymbol{d})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{q}-\boldsymbol{d}) \tag{3}$$

式中：$\boldsymbol{q}$ 为从联合球体中心到投影圆表面任意一点的矢量。当忽略联合球体半径时有

$$M^2 = \boldsymbol{d}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{d} \tag{4}$$

随着 $\boldsymbol{q}$ 的位置变化，平方马氏距离会出现极大值和极小值。文献 [13] 给出了求极值的解析表达式，首先通过解一个一元四次方程求得 k。

$$p(k) = ak^4 + bk^3 + ck^2 + dk + e = 0 \tag{5}$$

其中：

$$\begin{cases} a = (\Delta/R)^2 - 1 \\ b = 2\sigma_y^2\left[(\Delta/R)^2(\cos^2\theta + \sin^2\theta A_{\mathrm{r}}^2) - (1+A_{\mathrm{r}}^2)\right] \\ c = \sigma_y^4\left[(\Delta/R)^2(\cos^2\theta + \sin^2\theta A_{\mathrm{r}}^4) - (1+A_{\mathrm{r}}^2)^2 - 2A_{\mathrm{r}}^2\right] \\ d = -2\sigma_y^6(1+A_{\mathrm{r}}^2)A_{\mathrm{r}}^2 \\ e = -\sigma_y^8 A_{\mathrm{r}}^4 \end{cases} \tag{6}$$

式中：A_{r}为误差椭圆的长短轴比，$A_{\mathrm{r}} = \sigma_x/\sigma_y$。将式 (5) 的实数解代入式 (7) 中，可分别求得平方马氏距离的极大值和极小值。

$$M^2(\boldsymbol{q}^*) = (\Delta\sigma_y)^2\left[\left(\frac{\sin\theta}{k+\sigma_y^2}\right)^2 + \left(\frac{A_{\mathrm{r}}\cos\theta}{k+A_{\mathrm{r}}^2\sigma_y^2}\right)^2\right] \tag{7}$$

为了降低漏检率，后文选择平方马氏距离极值中的最小值(记为$M_{\min}^2$)进行分析。

1.2 平方马氏距离安全阈值

碰撞概率若超过黄色阈值，表示两空间目标的

接近比较危险, 需要引起操作员注意; 碰撞概率若超过红色阈值, 表示情况非常危险, 建议立即采取规避措施[19]。一般碰撞概率的黄色阈值为 10^{-5}, 红色阈值为 10^{-4}。

同样, 为了定量评估平方马氏距离所指征的风险等级, 需要求得平方马氏距离的安全阈值。本文根据3σ法则来计算: 对于均值为μ, 标准差为σ的正态分布, 变量在$(\mu-3\sigma,\mu+3\sigma)$区间内的概率为 99.74%, 分布在该区间外的概率不到 0.3%, 被视为极端不可能事件。

由于测量误差等的影响, 空间目标状态量的真实值在预报值附近呈正态分布。由3σ法则可知, 目标 c 的分布范围几乎不可能超过其3σ误差椭圆, 则当目标 t 所处位置在该3σ椭圆外时, 就可以认为两者不会相撞。基于这一思路, 对平方马氏距离的阈值进行求取。

设目标 c 当前的位置状态矢量为$\boldsymbol{x}_c$, 误差协方差矩阵为 $\boldsymbol{C}$, 则目标 c 的误差椭圆的一般方程为

$$(\boldsymbol{x}-\boldsymbol{x}_c)^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{x}_c)=1 \tag{8}$$

设 3σ法则下的误差协方差矩阵为$\boldsymbol{C}_{3\sigma}$, $\boldsymbol{C}_{3\sigma}$与 $\boldsymbol{C}$ 的关系为

$$\boldsymbol{C}_{3\sigma}=3^2\boldsymbol{C} \tag{9}$$

进而可得3σ法则下目标 c 的误差椭圆一般方程$(\boldsymbol{x}-\boldsymbol{x}_c)^{\mathrm{T}}\boldsymbol{C}_{3\sigma}^{-1}(\boldsymbol{x}-\boldsymbol{x}_c)=1$可以转换为

$$(\boldsymbol{x}-\boldsymbol{x}_c)^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{x}_c)=9 \tag{10}$$

设目标 t 的位置状态矢量为$\boldsymbol{x}_t$, 若满足式(11), 则表示目标 t 在目标 c 的3σ椭圆外, 说明两空间目标没有发生碰撞。

$$(\boldsymbol{x}_t-\boldsymbol{x}_c)^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{x}_t-\boldsymbol{x}_c)>9 \tag{11}$$

由于$\boldsymbol{x}_c-\boldsymbol{x}_t=\boldsymbol{d}$, 将其代入式 (11), 得

$$\boldsymbol{d}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{d}>9 \tag{12}$$

即平方马氏距离的阈值为 9。若平方马氏距离大于 9, 则判定两空间目标没有碰撞风险, 反之则存在危险。需要注意的是, 在应用平方马氏距离判据时, 由于两空间目标均存在位置误差, 所以误差协方差矩阵 $\boldsymbol{C}$ 应为两空间目标的联合误差协方差矩阵。

由式 (1) 可知, 计算 P_c需要的 3 个物理量为联合球体半径、误差协方差矩阵和两空间目标的相对位置信息。设联合球体半径为 10 m, 相对位置矢量为 [0.305 8, −0.217 5] km, 改变误差协方差矩阵(即改变误差椭圆的形状A_{r}和大小σ_x), 得到P_c和M^2的变化如图 2 所示。图中的透明截面分别表示P_c的红色阈值和黄色阈值, 以及M^2的阈值。

随着误差椭圆的改变, P_c呈现不同幅度的增减变化, M^2随误差增大单调减小。为了便于将 2 个判据进行对比, 取$A_{\mathrm{r}}=\sigma_x/\sigma_y=2.378\,9$, 得到二维截面图, 如图 3 所示。由图 3(a) 可以看到, 对于固定的

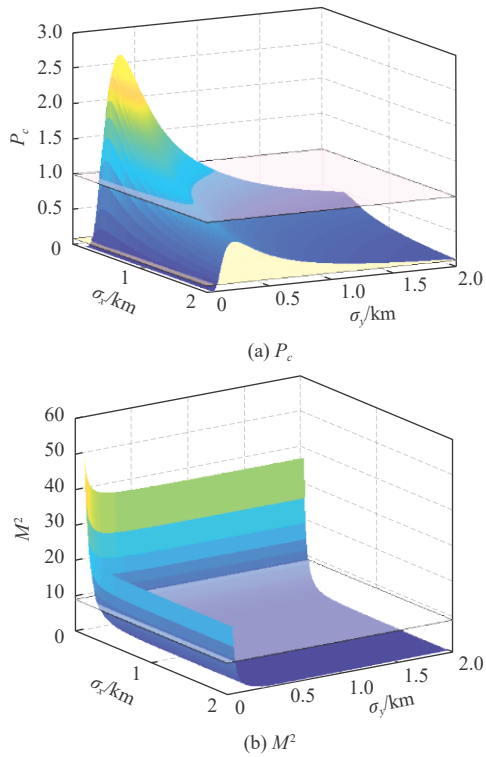


图 2　P_c和M^2随误差椭圆的变化

Fig. 2　P_c and M^2 vary with the error ellipse

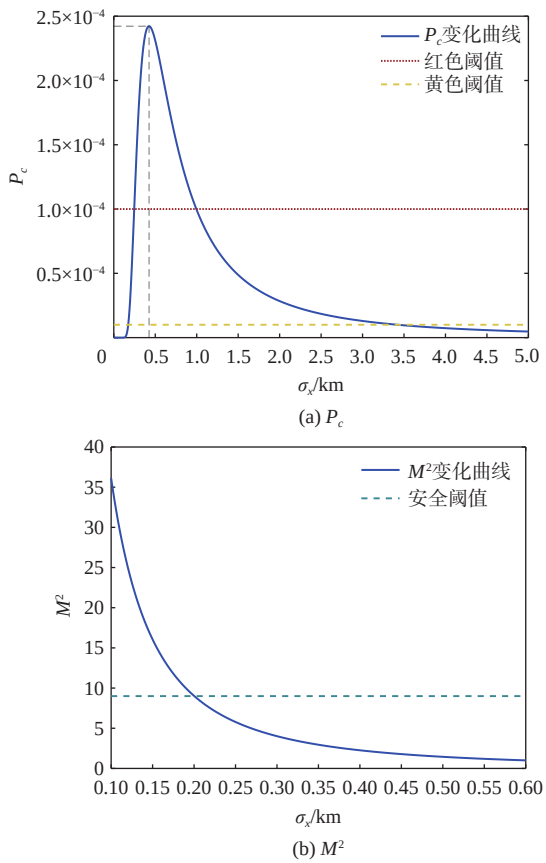


图 3　P_c和M^2随位置误差变化 ($A_{\mathrm{r}}=2.378\,9$)

Fig. 3　P_c and M^2 vary with the position error ($A_{\mathrm{r}}=2.378\,9$)

R 和 $\boldsymbol{d}$, 碰撞概率作为位置误差的函数呈现先增后减的变化趋势, 存在最大值, 这称为概率稀释现

象。图3(a)中，当σ_x=0.425 1 km时，P_c取得最大值2.422×10^{-4}，该值大于红色阈值，说明此次相遇十分危险。然而，如果真实误差大于3.422 0 km，碰撞概率将小于黄色阈值，显示没有碰撞风险，这就导致了漏检的发生。而此时平方马氏距离小于阈值，显示2个空间目标很可能碰撞，这个结论更接近真实情况。说明引入平方马氏距离可以避免仅依靠碰撞概率这一判据导致的漏检。

2 基于平方马氏距离的碰撞风险预报评估方法

2.1 平方马氏距离计算方法优化

如1.1节中所述，平方马氏距离在联合球体半径R>0时，存在最小值$M^2_{\min}$，当忽略半径，即$R\to0$时，为常值M^2。对比两者的计算过程不难发现，计算$M^2_{\min}$需要求解一个一元四次方程，而计算M^2只是进行一个普通的矩阵乘法运算，因此两者在计算时间上差异明显。对M^2与$M^2_{\min}$分别进行了1 000次模拟运算，发现求解$M^2_{\min}$所需的总时间约为计算M^2的总时间的3万倍。

随着传播时间的增加，误差椭球会逐渐拉长变扁，形态发生较大的变化，这时半径的存在将导致M^2与$M^2_{\min}$产生巨大的差异。以表1中的相遇事件为例，这时θ很小，误差椭圆中心接近x轴，同时A_r较大，误差椭圆变形严重。计算得到M^2=52.923 1大于阈值，而$M^2_{\min}$=0.115 7小于阈值，两者差异明显，对碰撞风险的判断结果也截然相反。

表1 M^2和$M^2_{\min}$之间的偏差

Table 1 Deviation between M^2 and $M^2_{\min}$

R/m	Δ/m	σ_x/m	θ/rad	A_r
10	50	120	π/36	200

由此可见，计算$M^2_{\min}$得到的碰撞风险结果更加全面和可靠。然而，并不总是需要计算$M^2_{\min}$，在很多情况下两者近似相等，可以用M^2代替$M^2_{\min}$来进行风险分析，进而减少计算消耗。为了明确什么情况下可以进行代替，本文定义了2个变量：V_r和分界点U_{vr}。V_r为M^2和$M^2_{\min}$的比值：

$$V_r = M^2/M^2_{\min} \tag{13}$$

当$V_r<U_{vr}$时，认为M^2与$M^2_{\min}$近似相等，只计算M^2；而当$V_r>U_{vr}$时，则计算更为可靠的$M^2_{\min}$。

此外，还存在一些特殊情况，例如V_r=1.05，M^2=9.345>9，$M^2_{\min}=M^2/V_r$=8.9<9。这时V_r已经非常小，但是两者恰好在9附近体现出差别，一个大于阈值，另一个小于阈值，因而导致判断结果相反。这说明即便满足$V_r<U_{vr}$，也可能出现判断矛盾的情况。为了避免这类问题发生，需要引入一个允许误差E_{ac}。当$V_r<U_{vr}$时，还需满足M^2大于E_{ac}，才能同时保证$M^2_{\min}$大于阈值，否则仍然需要计算$M^2_{\min}$。

$$E_{ac} = 9\times U_{vr} \tag{14}$$

由式(7)可知，平方马氏距离与相对距离Δ、半径R、位置误差σ_x、误差椭圆长短轴比A_r、旋转角θ这5个因素有关。当这些因素落在某些特定区间时，可确保$V_r<U_{vr}$。接下来，将对这些影响因素逐一进行分析，以求解满足$V_r<U_{vr}$的区间范围。平方马氏距离与误差椭圆所在的象限无关，因此在分析过程中限定θ的大小为$[0,\pi/2]$。

1)位置误差。文献[13]中证明，在投影圆上平方马氏距离取得极值的点为投影圆与误差椭圆的切点。如图1所示，记切点和圆心的连线与x轴的夹角为φ，则$\boldsymbol{q}^*=(R\cos\varphi,R\sin\varphi)$。已知相对距离矢量为$\boldsymbol{d}=(\Delta\cos\theta,\Delta\sin\theta)$，$\sigma_x=\sigma_yA_r$，由式(7)可得平方马氏距离最小值为

$$M^2(\boldsymbol{q}^*)=\frac{(\Delta\cos\theta-R\cos\varphi)^2}{\sigma_y^2A_r^2}+\frac{(\Delta\sin\theta-R\sin\varphi)^2}{\sigma_y^2} \tag{15}$$

不考虑半径时，由式(2)及式(4)得平方马氏距离常值为

$$M^2=\frac{(\Delta\cos\theta)^2}{\sigma_y^2A_r^2}+\frac{(\Delta\sin\theta)^2}{\sigma_y^2} \tag{16}$$

式(16)与式(15)的比值为

$$V_r=\frac{M^2}{M^2(\boldsymbol{q}^*)}=\frac{(\Delta/R\cos\theta)^2+(\Delta/R\sin\theta A_r)^2}{(\Delta/R\cos\theta-\cos\varphi)^2+(\Delta/R\sin\theta-\sin\varphi)^2A_r^2} \tag{17}$$

式(17)中不含位置误差σ_x或σ_y，说明该误差对V_r没有影响。后文将Δ/R作为一个影响因素来分析，设临界点为U_{DR}，当$\Delta/R\leqslant U_{DR}$时，表示相对距离过小使得联合球体半径不能忽略，必须计算最小值$M^2_{\min}$，而不能计算M^2。

2)误差椭圆长短轴比。误差椭圆长短轴比A_r的不同会影响极值点的位置。如图4所示，当A_r非常小时，误差椭圆几乎变成了一条与x轴垂直的直线，此时极值点为投影圆与x轴正半轴的交点；随着A_r增大，极值点逐渐上移；当A_r非常大时，误差椭圆几乎变成一条与x轴平行的直线，此时φ为90°，极值点为投影圆与y轴正半轴的交点。

通过比较式(3)和式(4)，M^2与$M^2_{\min}$的差别在于所选用的与协方差矩阵相乘的向量不同。设M^2中采用的向量为$\boldsymbol{a}$，$M^2_{\min}$采用的向量为$\boldsymbol{b}$，设$\boldsymbol{a}$与$\boldsymbol{b}$的夹角为α。当α为0时，$\boldsymbol{a}$与$\boldsymbol{b}$平行，说明此时M^2与

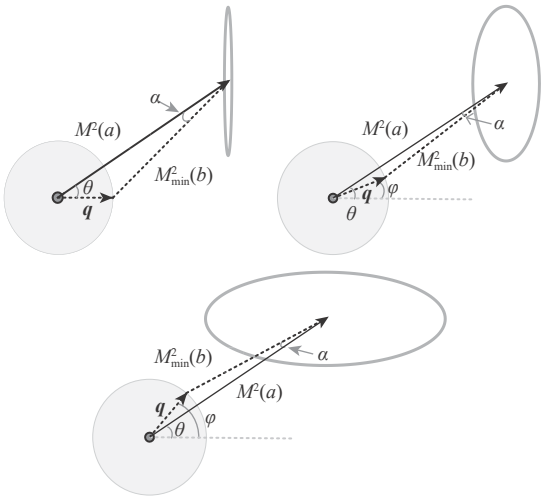


图 4　A_r不同, 极值点位置随之变化

Fig. 4　Position of the extreme point changes with different A_r

$M^2_{\min}$的差别最小, V_r获得最小值。这时, φ等于θ, 切点在投影圆圆心与误差椭圆中心的连线上, 说明误差椭圆为圆形, A_r等于 1。随着A_r的增大, 极值点沿着投影圆逐渐上移, α由某个值先减小为 0, 然后继续增大, 导致V_r呈现先减小后增大的趋势。设满足$V_r < U_{vr}$的A_r的闭区间为$[A_{r\min}\ \ A_{r\max}]$。

3)旋转角。θ变化时, 极值点的位置也会相应改变。如图 5 所示, 当θ等于 0 时, 误差椭圆中心落在x轴上, 极值点为投影圆与x轴正半轴的交点; 随着θ增大, 极值点在第一象限内开始沿投影圆外围变化, 逐渐上移; 当θ等于 $\pi/2$ 时, 误差椭圆中心落于y轴上, 此时极值点为投影圆与y轴正半轴的交点。

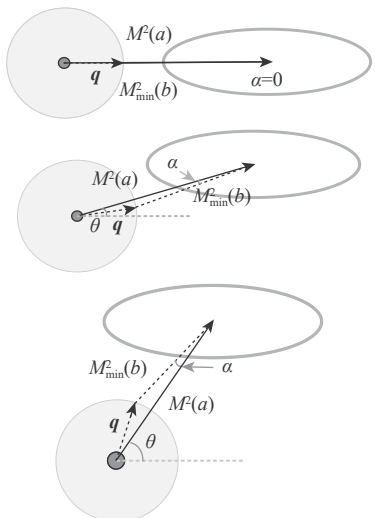


图 5　θ变化导致极值点变化

Fig. 5　Position of the extreme point changes with different θ

当θ为 0 时, $\boldsymbol{a}$与$\boldsymbol{b}$平行, α为 0; 当θ为 $\pi/2$ 时, $\boldsymbol{a}$与$\boldsymbol{b}$同样平行, α也为 0。随着θ的增大, α从 0 开始增大, 达到最大值后又继续减小到 0。与上节同理, 可推知随着θ的增大, 由于α先增大后减小, V_r也先增大后减小。设满足$V_r<U_{vr}$的θ的闭区间为$[\theta_{\min}\ \ \theta_{\max}]$。

以上, 求解了满足$V_r<U_{vr}$的各因素的区间范围, 可以对计算M^2或者$M^2_{\min}$的不同情况进行分类。从式 (3) 和式 (4) 中不难发现$M^2_{\min}$始终小于M^2, 若M^2小于 9, 则$M^2_{\min}$同时也小于 9, 两者判断结果不会矛盾。这说明, 如果M^2小于 9, 也无需再计算$M^2_{\min}$。

综上所述, 平方马氏距离的计算方法总结如下, 计算流程图如图 6 所示。

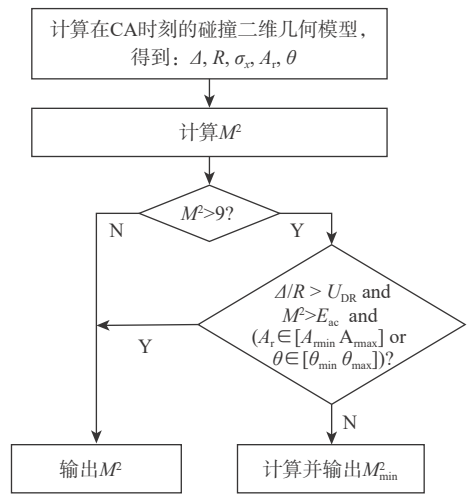


图 6　平方马氏距离的计算方法

Fig. 6　Calculation method of the square Mahalanobis distance

1) 对相遇事件在 CA 时刻的碰撞二维模型进行求解, 计算得到: $\Delta, R, \sigma_x, A_r, \theta$;

2) 计算M^2, 并判断$M^2>9$是否成立, 若成立则进入下一步, 不成立则直接输出M^2;

3) 判断①$\Delta/R > U_{DR}$, ②$M^2>E_{ac}$, ③$A_r \in [A_{r\min}\ A_{r\max}]$或$\theta \in [\theta_{\min}\ \ \theta_{\max}]$是否成立;

4) 若①②③均满足, 则直接输出M^2; 若有一个不满足, 则需要计算$M^2_{\min}$。

2.2　碰撞风险预报评估方法

在最早期的空间目标碰撞预警问题中, 采用的是最小距离法[20](BOX 区域方法)。该方法只考虑相对距离因素: 在空间目标周围定义一个 50 km×20 km×20 km 的区域为预警区域, 定义一个 10 km×4 km×4 km 的区域为规避区, 通过判断 2 个空间目标之间的位置关系来做出相应的预警或规避的举措。

本文以该方法为基础, 选择通过相对距离来进行初步筛选, 以 BOX 区域方法中预警区域的一半, 即 $\sqrt{25^2+10^2+10^2}=28.722\ 8$ km为筛选阈值。当 2 个空间目标的相对距离小于 28.722 8 km 时, 初步认为其存在碰撞风险, 为进一步的预警分析提供基础。

根据以上分析, 本文提出了一种基于平方马氏距离的碰撞风险预报评估方法:

1）计算相对距离进行初步筛选，获取危险持续区间：对于需要进行碰撞风险分析的 2 个空间目标，通过轨道动力学演化分别获取它们在未来两天内的位置速度状态矢量信息，并计算每个时刻的相对距离，记录相对距离小于 28.722 8 km 的开始时刻和结束时刻，生成危险持续区间。

2）计算每段危险持续区间内的 CA 时刻和碰撞二维几何模型：缩小递推步长，重新计算每段危险持续区间内相对距离的变化情况，得到各个区间内的 CA 时刻及对应的二维碰撞模型，求得 Δ, R, σ_x, A_r, θ。

3）计算 CA 时刻的平方马氏距离：计算方法如图 6 所示，最后输出平方马氏距离的常值 M^2 或最小值 $M^2_{\min}$。

4）做出风险判断：如果平方马氏距离大于阈值 9，表明此次相遇安全，否则此次相遇危险。

具体的流程图如图 7 所示。

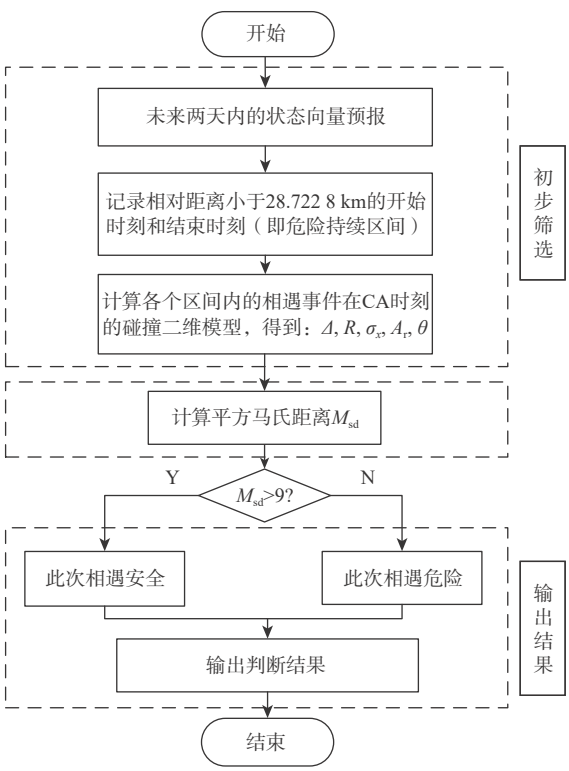


图 7 碰撞风险预报评估流程图

Fig. 7 Flow chart of collision risk prediction and evaluation

3 数值仿真

3.1 碰撞风险预报评估方法分类条件数值求解

本节将对第 2 节预报评估方法中的各变量进行求解。美国联合空间作战中心[21]发出碰撞警告的阈值是 1 km，因此选择将 U_{DR}=100 作为合适的临界点。从允许误差 E_{ac} 与阈值 9 的量级出发，取 U_{vr} 为 1.2。本文算法部署在 Windows10 电脑上，处理器为 Intel(R) Core(TM) i5-9300H @ 2.40 GHz，内存为 16.0 GB，计算软件采用 MATLAB2018b。

1）位置误差。分析位置误差的影响时，设其他变量为定值。保持 A_r=3, θ=π/4 为常值，只改变 σ_x 的大小，计算不同相对距离 Δ 下的 M^2，将其与 $M^2_{\min}$ 作比，比值结果如图 8 所示。

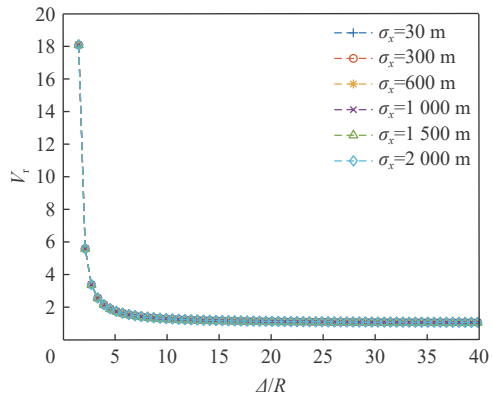


图 8 σ_x 对 V_r 的影响

Fig. 8 Influence of σ_x on V_r

图中 6 条曲线度高度重合在一起，说明当 A_r 和 θ 固定时，取相同的相对距离 Δ，不同的 σ_x 下对应的 V_r 一直相等，即 σ_x 的大小不会对 V_r 产生影响。

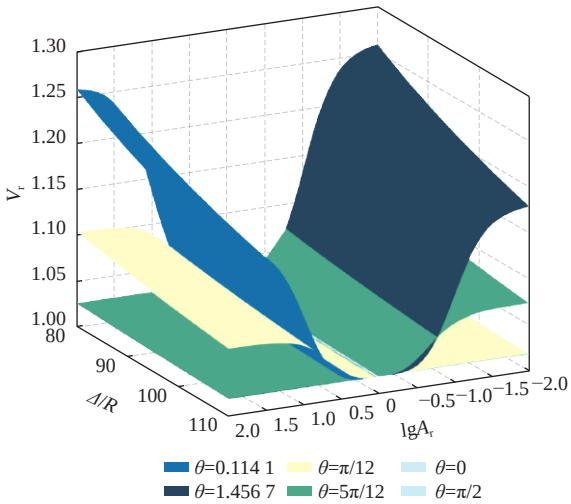
2）误差椭圆长短轴比。取 θ 为 [0, 0.114 1, π/12, 5π/12, 1.456 7, π/2], σ_x=120 m，改变 A_r，计算不同 θ 下对应的 V_r 随 A_r 的变化图，结果见图 9(a)。取 Δ/R = 100，得到图 9(a) 中的横截面，如图 9(b) 所示。

从图 9(a) 中可以观察到，随着 Δ/R 的增大，V_r 逐渐降低，表明相对距离的增大会减弱半径对平方马氏距离计算的影响。如图 9(b) 所示，当 A_r 由小到大或由大到小变化时，V_r 先减小，在 A_r=1 时取得最小值，然后又继续增大，这与 2.1 节的推导一致。当 2 个 θ 关于 π/4 对称时，其对应的变化图关于平面 A_r=1 对称。

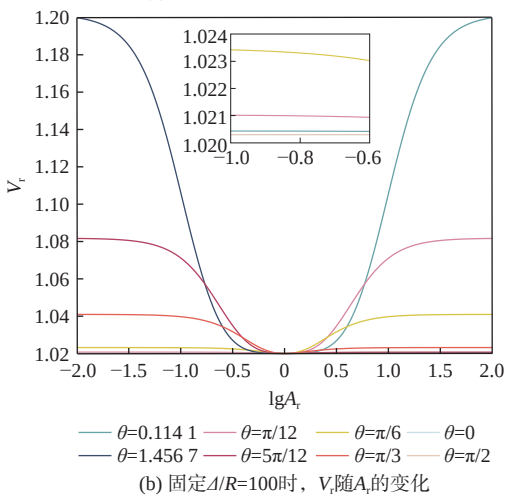
在图 9(b) 中，θ 越接近 0 或 π/2，V_r 随 A_r 的变化图形就越陡峭，表明 θ 越靠近坐标轴，M^2 与 $M^2_{\min}$ 的差异就越大。但是当 θ 等于 0 或 π/2 时，V_r 的变化反而非常平缓，仅与 Δ/R 有关。这是因为在这两类情况下，A_r 的变化不会影响极值点的位置，极值点始终位于投影圆与坐标轴的交点。

经过计算，当 θ 等于 0.114 1 和 1.456 7 时，在 Δ/R 等于 100 处，V_r 的最大值等于 1.2。同时由于二维误差椭圆圆心在第一象限或第四象限[15]，θ 满足 $-\pi/2 \leqslant \theta \leqslant \pi/2$，因此，可以得出结论：当 Δ/R 大于 100 时，V_r 小于 1.2 在 $\theta \in [0.114\,1 \quad 1.456\,7] + k\pi/2 (k=0, -1)$ (rad) 时均成立。

3）旋转角。取 A_r 为 [0.04, 0.055 0, 0.1, 1, 10, 18.365 4, 25], σ_x 为 120 m，改变 θ，计算不同 A_r 时对



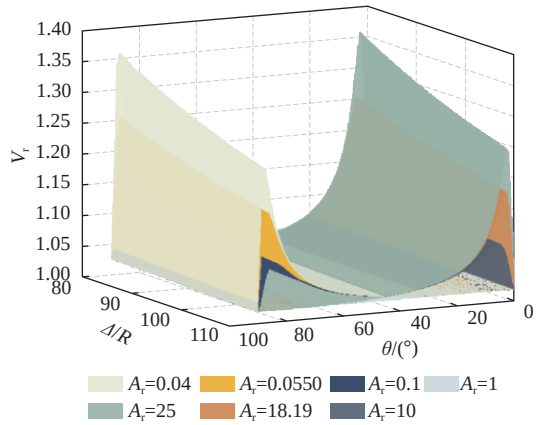
(a) 不同θ下，V_r随A_r和Δ/R的变化



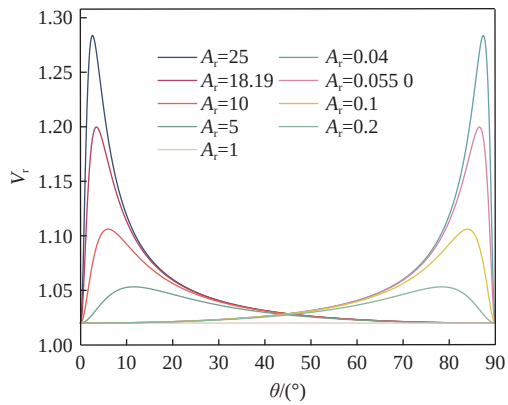
(b) 固定Δ/R=100时，V_r随A_r的变化

图 9　固定σ_x，不同θ下V_r随A_r和Δ/R的变化

Fig. 9　Change of V_r with A_r and Δ/R in different θ (σ_x= 120 m)



(a) 不同A_r下，V_r随θ和Δ/R的变化



(b) 固定Δ/R=100时，V_r随θ的变化

图 10　固定σ_x，不同A_r下V_r随θ和Δ/R的变化

Fig. 10　Change of V_r with θ and Δ/R in different A_r (σ_x= 120 m)

应的V_r随θ的变化图，结果如图 10(a) 所示。固定Δ/R=100 时，变化图如图 10(b) 所示。

从图 10 中可知，随着θ的增大，V_r先增大后减小，与理论分析得到的变化规律相符。当A_r = 0.2 时，V_r在θ = 78.15°处取得最大值，这时φ = 11.85°，α为最大值 66.300 0°；当A_r = 5时，V_r在θ=11.85°处取得最大值，这时φ = 78.149 0°，α为最大值 66.299 0°；当A_r互为倒数时，其对应的变化图关于平面θ=π/4 对称。

得到以下结论：①A_r越远离 1，V_r随θ的变化越剧烈；②当A_r等于 1 时，误差椭圆为圆形，由于φ始终等于θ，极值点的位置保持不变，V_r的大小只与Δ/R有关；③经过计算，当A_r等于 0.055 0 和 18.190 0 时，在Δ/R等于 100 处，V_r随θ变化的最大值等于 1.2。因此可以得出结论：当Δ/R>100 时，V_r<1.2 在$A_r \in$ [0.055 0　18.190 0]时均成立。

本节对平方马氏距离的各影响因素进行了分析，进一步完善了平方马氏距离的计算方法及碰撞风险预报评估方法，求解得到各变量的取值如表 2 所示。

表 2　碰撞风险预报评估方法中的参数

Table 2　Parameters in collision risk prediction and evaluation method

参数	数值
U_{DR}	100
U_{vr}	1.2
E_{ac}	10.8
[A_{rmin}　A_{rmax}]	[0.055 0　18.190 0]
[θ_{min}　θ_{max}]/ rad	[0.114 1　1.456 7]+ $k\pi/2$(k = 0, −1)

3.2　算例分析

本文基于 CelesTrak 网站提供的两行轨道根数 (two-line element, TLE)，以卫星 FLOCK 4Y-25 和碎片 SL-3 DEB 为研究对象，对其在轨运行中的碰撞风险进行了计算。

卫星 FLOCK 4Y-25 属于鸽群星座，是一个立方体卫星，于 2023 年 1 月 3 日发射。SL-3 DEB 为一卫星产生的空间碎片，该卫星由独立国家联合体 (Commonwealth of Independent States, CIS) 于 1983 年 7 月 24 日发射。两者的 TLE 数据如表 3 所示。

表3 FLOCK 4Y-25和SL-3 DEB的TLE

Table 3 TLE of FLOCK 4Y-25 and SL-3 DEB

卫星/碎片	两行轨道根数
FLOCK 4Y-25	1 55027U 23001U 24010.16624038 .00012377 00000+0 52739-3 0 9990
	2 55027 97.4492 72.8718 0010226 14.6525 345.5006 15.23053222 56530
SL-3 DEB	1 14928U 83075F 24010.30073432 .00006854 00000+0 32096-3 0 9991
	2 14928 97.7779 279.1809 0026055 294.6725 65.1796 15.19768913182859

本文以12 Jan 2024 10:05:36.000 UTC为起点，通过SGP4模型对卫星和碎片进行轨道演化，采用无迹变换传播协方差信息[22-23]。在仿真计算中，由于危险持续区间通常较短，所以积分步长不能大于1 s。获得2个空间目标在2天内的位置速度状态矢量后，按照图7的方法对其碰撞风险情况进行求解分析。

3.2.1 计算结果

设2个空间目标的三轴初始位置误差均为10^{-1} km，三轴初始速度误差均为10^{-5} km/s，拟合半径均为5 m。上述卫星与空间碎片之间的碰撞风险分析结果如表4所示。

表4 卫星FLOCK 4Y-25和碎片SL-3 DEB间的碰撞风险分析

Table 4 Collision risk analysis between Satellite FLOCK 4Y-25 and Debris SL-3 DEB

危险区间	区间起止时间	危险持续时长/s	最接近时刻	最接近距离/km	M^2	$M^2_{\min}$	是否计算最小值	θ/rad	A_r	P_c(2D)	P_c(max)	σ_d/σ_x	$R/\lvert\boldsymbol{C}\rvert^{1/4}$	评估结果
第①段	01-13T18:26:40.000—01-13T18:26:41.000	1	01-13T18:26:40.405	24.794 8	10 817	10 808	是	−0.023 5 (N)	0.016 7 (N)	−∞	−∞	2.937 5	0.005 4	安全
第②段	01-13T19:13:56.000—01-13T19:13:57.000	1	01-13T19:13:56.765	22.828 8	103.231 6	103.141 1	否	−0.017 7 (N)	0.157 4 (Y)	−∞	−∞	2.991 3	0.001 8	安全
第③段	01-13T20:01:21.000—01-13T20:01:24.000	3	01-13T20:01:22.701	0.025 5	$4.480\,4\times10^{-7}$	$1.633\,8\times10^{-7}$	否	0.229 1 (Y)	0.521 4 (Y)	−7.730 4	−1.529 9	$5.342\,9\times10^{-4}$	$1.928\,9\times10^{-4}$	危险
第④段	01-13T20:48:39.000—01-13T20:48:40.000	1	01-13T22:48:39.056	24.421 8	0.020 6	0.020 6	否	−0.077 5 (N)	0.441 5 (Y)	−9.119 3	−7.562 8	0.101 6	$3.918\,2\times10^{-5}$	安全

表4包括两目标接近的危险持续区间、危险持续时长、最接近时刻，以及该时刻对应的最小接近距离。表4中，θ和A_r后面的括号内，“Y”表示数值在表2中所求区间内，“N”表示不在；P_c(2D)为经过投影后由二重积分计算的碰撞概率，P_c(max)为在概率稀释现象中的碰撞概率最大值(表格中的碰撞概率均取以10为底的对数)。

表4中，σ_x为协方差矩阵在x轴方向的误差，σ_d为碰撞概率取得最大值时的误差，σ_d/σ_x越小说明概率稀释现象越严重。$R/\lvert\boldsymbol{C}\rvert^{1/4}$反映了联合球体半径与协方差矩阵之间的大小关系。表中$R/\lvert\boldsymbol{C}\rvert^{1/4}$单调递减，由于$R$不变，这表明随着时间的延长，误差一直在连续发散，直至远远大于联合球体半径和最接近距离。最大碰撞概率的判断结果更接近相遇事件真实的危险等级，当最大碰撞概率小于10^{-5}时认为此次相遇安全，否则危险，因此仅第③段区间的结果为“危险”。

3.2.2 结果讨论

1) 在表4中，平方马氏距离的最小值和常值同时大于或小于阈值，表明在大多数情况下，两者的差异都很小，可以用常值替代最小值进行风险分析，有效避免计算最小值，进而提高计算效率。本文方法共用时为0.040 3 s。而常规方法选择计算4次最小值，仿真用时为0.113 4 s，时间节约了64.46%。在需要进行大规模碰撞风险筛选时，计算效率的提高将进一步体现。

2) 第③段区间属于危险相遇，P_c(2D)小于10^{-5}导致风险漏报，而平方马氏距离小于阈值指示危险，很好地避免了这次漏检。初步分析，其原因是概率稀释的作用，此时σ_d/σ_x处于10^{-4}量级，误差σ_x已远远大于峰值误差σ_d，导致碰撞概率处于很低的水平。

3) 第④段相遇过程是安全的，但平方马氏距离仍然小于阈值，指示这一段是危险的，造成虚警。根据马氏距离可以简单地理解为相对距离除以误差进行初步分析，其原因是分子(相对距离)是有限的，小于28.722 8 km，而预报时间过长时误差发散太大，即分母过大，从而导致马氏距离不可避免地变小。

4 结 论

针对空间目标间碰撞风险评估的问题，本文提出了基于平方马氏距离的风险分析算法进行判

断。理论分析和仿真实验表明:

1) 本文算法能对平方马氏距离最小值和常值的计算进行合理分类, 从而有效节省了计算成本, 提高了计算效率。

2) 本文算法相较于碰撞概率算法能有效减少高风险事件的漏检, 提高风险评估结果的可靠性。

3) 本文算法虽然可能导致虚警, 但能减少漏检率, 而漏检比虚警的危害更大。同时, 因为操作者对于危险相遇一般不会立即采取规避措施, 而是根据事件的风险等级实时监测相关目标, 多次更新目标的轨道信息以进行进一步的风险评估, 虚警的影响会被削弱。

参考文献 (References)

[1] 段玉瑞, 孙振江, 王建伟, 等. 地球轨道空间碰撞风险区域态势分析[J]. 力学与实践, 2022, 44(6): 1279-1285.
DUAN Y R, SUN Z J, WANG J W, et al. Collision risk analysis of orbital areas in earth space[J]. Mechanics in Engineering, 2022, 44(6): 1279-1285(in Chinese).

[2] ANZ-MEADOR P, KING A, MATNEY M. Evolution of major debris clouds in low earth orbit[J]. The Orbital Debris Quarterly News, 2023, 27(3): 7.

[3] NASA. ISS maneuvers to avoid potential collisions twice in august 2023[J]. The Orbital Debris Quarterly News, 2023, 27(4): 1.

[4] STEF C, MARIEL B, Brain C G, et al. Cislunar orbit collision probability analysis[C]//AIAA SCITECH 2024 Forum. Reston: AIAA, 2024.

[5] PARIGINI C, ALGETHAMIE R, ARMELLIN R. Short-term collision probability caused by debris cloud[J]. Journal of Guidance, Control, and Dynamics, 2024, 47(5): 874-886.

[6] CARPINE M A, PAUL Y, DELANDE E, et al. Point process formulation of long-term collision risk statistics from orbital fragmentation[J]. Journal of Guidance, Control, and Dynamics, 2024, 47(7): 1309-1326.

[7] CHAN K. Collision probability analyses for earth-orbiting satellites[J]. Advances in the Astronautical Sciences, 1997, 96: 1033-1048.

[8] NÚÑEZ GARZÓN U E, LIGHTSEY E G. Relating collision probability and separation indicators in spacecraft formation collision risk analysis[J]. Journal of Guidance, Control, and Dynamics, 2022, 45(3): 517-532.

[9] CHAN K. Comparison of methods for spacecraft collision probability computations[C]//Proceedings of the 15th International Space Conference of Pacific-basin Societies. San Diego: Univelt, Inc., 1997.

[10] ALFANO S, OLTROGGE D. Probability of collision: valuation, variability, visualization, and validity[C]//Proceedings of the AIAA/AAS Astrodynamics Specialist Conference 2016. Reston: AIAA, 2016: 5654.

[11] DE MAESSCHALCK R, JOUAN-RIMBAUD D, MASSART D L. The mahalanobis distance[J]. Chemometrics and Intelligent Laboratory Systems, 2000, 50(1): 1-18.

[12] PHILLIPS M R. Spacecraft collision probability estimation for rendezvous and proximity operations[D]. Logan: Utah State University, 2012.

[13] MODENINI D, CURZI G, LOCARINI A. Relations between collision probability, mahalanobis distance, and confidence intervals for conjunction assessment[J]. Journal of Spacecraft and Rockets, 2022, 59(4): 1125-1134.

[14] NÚÑEZ GARZÓN U E, LIGHTSEY E G. Stochastic convergence of sobol-based mahalanobis shell sampling collision probability computation[J]. Journal of Guidance, Control, and Dynamics, 2023, 46(5): 796-814.

[15] 白显宗, 陈磊. 空间目标碰撞概率计算方法研究[J]. 宇航学报, 2008, 29(4): 1435-1442.
BAI X Z, CHEN L. Research on calculational method of collision probability between space objects[J]. Journal of Astronautics, 2008, 29(4): 1435-1442(in Chinese).

[16] DE VITTORI A, PALERMO M F, DI LIZIA P, et al. Low-thrust collision avoidance maneuver optimization[J]. Journal of Guidance, Control, and Dynamics, 2022, 45(10): 1815-1829.

[17] LI J S, YANG Z, LUO Y Z. A review of space-object collision probability computation methods[J]. Astrodynamics, 2022, 6(2): 95-120.

[18] MASSON M, ARZELIER D, BRÉHARD F, et al. Fast and reliable computation of instantaneous orbital collision probability[J]. Journal of Guidance, Control, and Dynamics, 2024, 47(8): 1660-1673.

[19] 李翠兰, 欧阳琦, 陈明, 等. 大型低轨航天器与星座卫星的碰撞风险研究[J]. 宇航学报, 2020, 41(9): 1158-1165.
LI C L, OUYANG Q, CHEN M, et al. Analysis of collision risk between constellation satellites and large low-orbit spacecraft[J]. Journal of Astronautics, 2020, 41(9): 1158-1165(in Chinese).

[20] LELEUX D, SPENCER R, ZIMMERMAN P, et al. Probability-based space shuttle collision avoidance[C]//Proceedings of the SpaceOps 2002 Conference. Reston: AIAA, 2002.

[21] MISHNE D, EDLERMAN E. Collision-avoidance maneuver of satellites using drag and solar radiation pressure[J]. Journal of Guidance, Control, and Dynamics, 2017, 40(5): 1191-1205.

[22] LUO Y Z, YANG Z. A review of uncertainty propagation in orbital mechanics[J]. Progress in Aerospace Sciences, 2017, 89: 23-39.

[23] 王涛, 李正良, 范文亮. 基于扩展型共轭无迹变换的随机不确定性传播分析方法[J]. 工程力学, 2021, 38(8): 15-23.
WANG T, LI Z L, FAN W L. An uncertainty propagation analysis method for stochastic system based on extended conjugate unscented transform[J]. Engineering Mechanics, 2021, 38(8): 15-23(in Chinese).

Space target collision risk analysis algorithm based on the square Mahalanobis distance

HUANG Mengdie，WANG Lufeng，HUANG Xuxing，LI Shuang*

(College of Astronautics，Nanjing University of Aeronautics and Astronautics，Nanjing 211106，China)

Abstract：A collision risk assessment method based on the square Mahalanobis distance was proposed to increase the reliability of collision risk prediction results in order to address the issue that probability dilution in space target collision risk prediction reduces the reliability of collision probability. Firstly, the threshold of the square Mahalanobis distance was obtained according to the 3σ rule to determine whether an event has collision risk. Then, the factors such as combined hard-body radius, position error, the aspect ratio of the error ellipse, and the rotation angle were modeled and analyzed, and the model parameters of the square Mahalanobis distance were optimized to improve the computational efficiency. Finally, a collision risk prediction and evaluation model based on the square Mahalanobis distance was developed to solve the probability dilution problem. The outcome of the simulation demonstrates that the suggested approach can successfully lower the high-risk event missed detection rate.

Keywords：space targets；collision risk analysis；square Mahalanobis distance；collision probability；3σ rule

Received：2024-03-26；**Accepted**：2024-06-25；**Published Online**：2024-07-10 15：15

URL：link.cnki.net/urlid/11.2625.V.20240708.1109.002

Foundation items：Funded by National Key Laboratory of Space Intelligent Control (HTKJ2024KL502001, HTKJ2024KL502031)

* **Corresponding author**. E-mail：lishuang@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0169

基于三维格网误差建模的GNSS/IMU抗差自适应定位算法

盛琪，孙蕊*，何雨霖，张珩瑜
(南京航空航天大学 民航学院，南京 211106)

摘 要：城市复杂环境中高大建筑会造成全球卫星导航系统（GNSS）、信号非视距传播（NLOS）和多径干扰（MI），影响智能交通定位精度；现有二维格网多径建模方法存在高程精度不足、测量噪声协方差调参策略简单的缺陷。提出一种基于三维格网误差建模的GNSS/惯性测量单元（IMU）抗差自适应定位算法：在现有二维格网的基础上，划分高度空间，进一步实现精细化建模，在多径误差预测阶段，通过格网中心匹配方法缓解错误匹配导致的模型预测误差；基于多径误差预测值，提出滤波模型选择策略，并结合抗差理论提出抗差阈值动态调节策略，达到对测量噪声协方差的环境自适应更新，能有效提升城市复杂场景GNSS/IMU组合导航定位精度。车载实验结果表明：所提算法相较传统GNSS/IMU紧组合算法和传统GNSS/IMU抗差自适应算法，定位精度分别提升了48.43%和31.48%，相较二维格网辅助的GNSS/IMU抗差自适应算法提升了27.57%。

关 键 词：多径误差建模；抗差滤波；自适应滤波；GNSS/IMU组合导航；非视距传播
中图分类号：V324; P228; V249.3
文献标志码：A **文章编号**：1001-5965(2026)05-1701-11

智能交通系统(intelligent traffic system, ITS)的建设能够缓解城市道路拥堵、运输效率不足等问题,对加速城市建设、保障城市安全具有重要现实意义。全球卫星导航系统(Global Navigation Satellite System, GNSS)是ITS位置服务的重要支撑,能够提供高精度定位、导航、授时(positioning, navigation, timing, PNT)信息。然而在城市复杂环境下,GNSS信号易被树木及高大建筑物遮挡,产生的多径效应严重制约了GNSS提供的PNT信息精度[1]。

城市多径效应主要包含非视距(non-line-of-sight, NLOS)接收和多路径干扰(multipath interference, MI),解决NLOS/MI的方法主要分为3类:天线设计方法、信号处理方法和基于观测值建模的方法。天线设计方法包括使用扼流圈天线[2-3]、右旋极化天线[4]、双极化天线[5]及多天线阵列[6]等抑制接收低仰角信号,但其均存在体积过大、价格昂贵的问题,不适用于城市ITS。信号处理方法是利用MI信号的相关函数特性,通过检测干扰分量,以减小多径误差[7],但由于信号处理方法无法处理短程多径,无法满足城市峡谷环境下ITS用户的使用需求。

基于观测值建模的方法是基于数理统计、机器学习等手段,建立GNSS原始观测量与多径误差之间的数学模型,以此缓解多径效应,提升定位精度。模型输入的GNSS原始观测量一般为信号强度、卫星高度角、卫星方位角、位置精度因子(position dilution of precision, PDOP)和伪距残差等。静态环境下,SUN等利用梯度提升决策树(gradient boosting decision tree, GBDT)来预测并直接修正多径误差,三维定位精度比传统组合导航提高了70%以上[8]。QIN等[9]提出一种基于袋装树的

收稿日期：2024-03-26；录用日期：2024-05-31；网络出版时间：2024-08-30 10:47
网络出版地址：link.cnki.net/urlid/11.2625.V.20240829.1453.002
基金项目：江苏省自然科学基金 (BK20211569)
*通信作者. E-mail：rui.sun@nuaa.edu.cn

引用格式：盛琪，孙蕊，何雨霖，等. 基于三维格网误差建模的GNSS/IMU抗差自适应定位算法[J]. 北京航空航天大学学报，2026，52(5)：1701-1711. SHENG Q，SUN R，HE Y L，et al. A robust adaptive positioning algorithm for GNSS/IMU based on 3D grid error modeling[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1701-1711 (in Chinese).

多径误差预测算法来提高树叶遮挡环境下的GNSS定位精度，三维定位精度能够提高80%以上。动态环境下，SUN等提出了基于基准点（point-based correction, PBC）的多径建模算法，采用预测的多径误差迭代补偿伪距，相比传统算法3D定位精度提升60%[10]，还提出了基于区域格网（grid-based correction, GBC）的多径建模算法对城市区域进行精密建模，采用六边形格网划分城市区域，逐个格网建立多径误差模型，三维定位精度相比传统组合导航提升63%[10]。除了进行伪距修正，还可利用多径误差预测值动态调整卡尔曼滤波参数，达到更高的定位精度。传统卡尔曼滤波器的GNSS测量噪声协方差不能动态地根据GNSS数据质量自适应调整。ZHANG等在GNSS/惯性测量单位（inertial measurement unit, IMU）松组合架构上提出了一种基于多径误差预测的GNSS测量噪声协方差更新策略，该方法的三维定位精度相比传统GNSS/IMU松组合算法提高了43.10%[11]。SUN根据多径误差预测值构建自适应因子，提出了一种GNSS/IMU紧组合的测量噪声协方差更新方案，该算法的定位精度相比传统的GNSS/IMU紧组合算法提高了55%，相比利用多径误差预测修正伪距的方法提高了15%，证明了基于多径误差预测的自适应滤波算法对GNSS/IMU组合导航定位精度提升的有效性[12]。然而，上述方法仍然存在3个问题：①多径误差模型不够精密。现有格网划分方案未对高度进行划分，对于高架、斜坡等高度存在复杂变化的区域，所建立的二维格网模型可用性大大降低；② 惯导推算当前历元的初始定位解往往不准确，若匹配至错误格网，会降低定位精度；③ 现有方法令测量噪声协方差等于多径误差预测值的平方，这种调整策略过于简单，且理论依据不充足。

抗差滤波[13]和自适应滤波[14]能够在一定程度上缓解多径效应，将两者结合的抗差自适应滤波能有效抵抗GNSS观测及IMU推算误差[15-16]，但需要结合具体场景合理选择滤波模型，使GNSS观测值和IMU推算值在组合导航系统中的受信任程度得到合理分配。NING等[17]提出利用最优径向基函数神经网络来进行滤波模型选择，区分并消除模拟的GNSS观测误差和IMU运动学推算误差成功率高达92%。GAO等[18]通过给标准卡尔曼滤波、自适应滤波和抗差滤波的结果分配不同的权重，使模型同时具有抗差性和自适应性，相比传统扩展卡尔曼滤波的定位精度提升了38.40%。但上述方法未能正确区分出误差来源，且加权方法不能最大程度地抑制观测信息误差和动力学模型误差。同时，现有抗差滤波中判断是否抗差的阈值，即抗差因子权函数包含的阈值，其大小需要在经验值范围内反复尝试，且确定后即为固定值，无法根据环境动态调节。因此，现有抗差自适应滤波存在的问题主要有3个：①执行滤波模型选择时，不能直观区分出误差来源；②抗差因子权函数中判断是否抗差的阈值难以确定；③固定的抗差阈值无法在快速变化的城市环境下动态调整，导致抗差性能降低。

对此，本文提出三维格网误差建模辅助的GNSS/IMU抗差自适应定位算法：

1）在现有二维格网的基础上，进一步划分城市高度空间，构建顾及多径环境立体特征的区域三维格网布局。为缓解惯导推算的初始定位解不准确导致的格网错误匹配，仅对落在每个立体格网中心区域的GNSS数据进行多径误差精密建模。

2）三维格网误差模型的预测值能够直观反映当前GNSS数据质量，本文算法根据该多径误差预测值的大小选择合适的滤波模型。

3）基于抗差理论，采用多径误差预测值动态插值调节抗差阈值，从而达到调节调整测量噪声协方差的目的，提升滤波算法的环境适应能力。

1 算法框架

本文算法如图1所示，具体步骤如下：

步骤1 三维格网布局构建。基于地理信息系统（geographic information system, GIS）数据库及城市区域先验信息划定三维格网布局范围，通过六边形格网紧密排布对该区域进行平面划分，继而将所得二维格网布局拓展高度维度，以选定的高度间隔对高度空间均匀划分。

步骤2 多径误差预测模型离线训练。基于卫星运动的时空重复性，对采集的GNSS原始观测量进行特征提取、样本标定、数据集划分和模型训练。沿同一路线多日重复行驶，分星座提取载噪比C/N_0、卫星仰角和卫星高度角，并利用高精度GNSS/IMU组合导航反演伪距得到的多径误差值进行标定，经三维格网划分后，采用袋装树算法逐个格网进行回归训练，建立GNSS多径误差预测模型。

步骤3 抗差自适应滤波算法在线测试。计算当前历元惯导推算的位置，以此算出卫星仰角和方位角，并根据惯导推算结果匹配所在格网，并调用相应位置、相应星座的多径误差预测模型。为缓解格网错误匹配导致的定位误差，提出格网中心匹配方法，仅当惯导推算位置落在格网中心区域时认为格网正确匹配，调用相应模型进行多径误差预测。输入当前C/N_0、卫星仰角和卫星高度角，即可得当

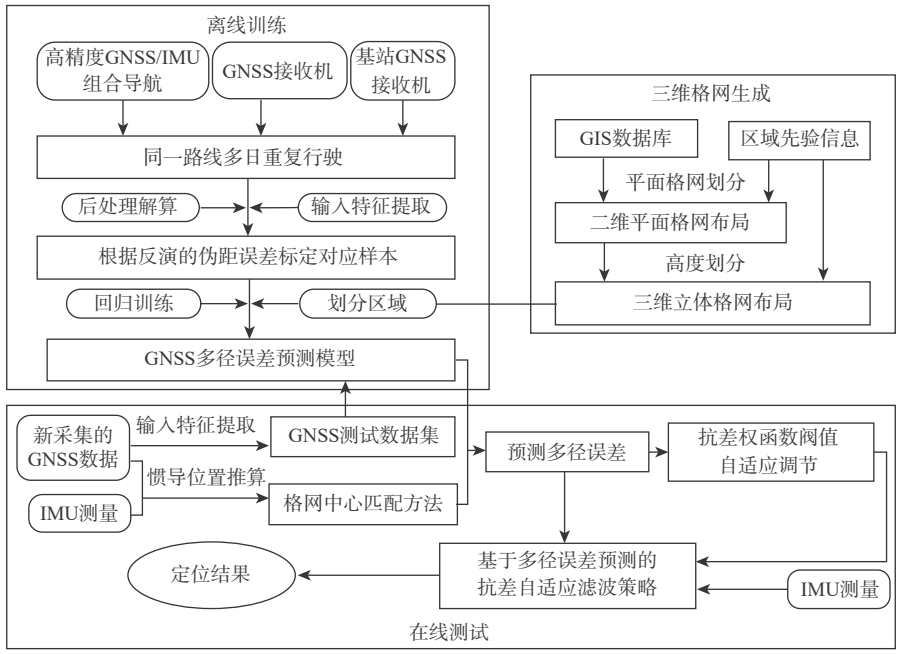


图 1　算法框架

Fig. 1　Algorithm framework

前历元各卫星的多径误差预测值, 用于抗差自适应滤波模型选择及抗差权函数阈值自适应调节。

2　基于三维格网的多径误差模型

卫星运动具有周期性和时空重复性, 当其运动经过同一位置时, 卫星星座呈现相同分布[19]。在卫星运动不同周期的对应时段, 若地面区域的建筑环境变化不大, 则能够认为该区域内的多径环境也具有一定的周期性。但由于多径误差的产生同时具有随机性, 多径误差并不能表现出规律的周期变化。为解决这一问题, 在 GNSS 原始观测量中提取能够反映当前多径环境的特征变量, 采用袋装树算法训练回归模型, 即可得到一个卫星运动周期内的多径误差预测规则。但由于城市环境复杂多变, 不同场景下的多径环境一般存在较大差异, 故而用三维格网划分城市区域及相应地理位置的 GNSS 训练数据, 利用袋装回归算法对各格网内包含的 GNSS 数据进行训练, 能够实现多径误差的精细化建模。

2.1　基于三维格网的 GNSS 训练数据集构建

2.1.1　特征变量选取与标定

GNSS 原始观测量中很多变量都能够指示多径误差值, 如 PDOP、C/N_0、伪距残差、卫星仰角和卫星方位角等。输入变量太多将增加算法时间成本, 输入变量过少将导致模型可靠性低。综合考虑算法效率和模型精度, 本文挑选C/N_0、卫星仰角和卫星方位角作为输入特征[8]:

1) 载噪比C/N_0: GNSS 信号被反射后, 信号强度一般会降低, 信号强度可通过C/N_0来衡量。因此, 利用 NLOS/MI 信号强度通常低于视距(line-of-sight, LOS)信号这一特点[20], 可以通过C/N_0数值来判断其类型。

2) 卫星仰角θ_e: θ_e能够描述当前位置的多径环境。可以认为仰角越低的卫星, 其信号就越容易被城市环境中的建筑物遮挡。

3) 卫星方位角θ_a: θ_a也能够反映当前位置的多径环境。对于城市区域来说, 周边的高大建筑物的方位是固定的, 处于这些方位的卫星信号被遮挡或反射的概率非常高。

利用高精度 GNSS/IMU 组合导航的解算结果反演伪距得到伪距真实值, 与测量伪距作差即可得到伪距误差值。以此作为标签标定样本。设R为接收机到卫星之间的几何距离, ρ为伪距, 经接收机钟差、卫星钟差、电离层及对流层延迟改正后的伪距ρ^c为

$$\rho^c = R + c(\Delta\delta t_s - \Delta\delta t_r) - \Delta I - \Delta T + \varepsilon \tag{1}$$

$$R = \sqrt{(x_s - x_r)^2 + (y_s - y_r)^2 + (z_s - z_r)^2} \tag{2}$$

式中: (x_s, y_s, z_s)和(x_r, y_r, z_r)分别为卫星和接收机在地心地固(earth-centered earth-fixed, ECEF)坐标系中的坐标; $\Delta\delta t_s$和$\Delta\delta t_r$分别为卫星和接收机钟差的残差; c为真空中的光速; ΔI和ΔT分别为电离层和对流层误差的残差; ε为观测噪声。伪距误差$\Delta\rho$可以表示为

$$\Delta\rho = \rho^c - R = c(\Delta\delta t_s - \Delta\delta t_r) - \Delta I - \Delta T + \varepsilon \tag{3}$$

虽然$\Delta\rho$包含多方面的影响，但在城市复杂环境中，多径误差的影响最为显著。故而可认为$\Delta\rho$为多径误差，以此标定训练样本，每个样本的训练数据为$[C/N_0,\theta_e,\theta_a,\Delta\rho]$，其中，$\Delta\rho$为样本标签。

2.1.2 基于区域三维格网布局的训练数据划分

三维网格布局的本质是对每个二维网格所属区域的高度空间进行分层，并在每层使用不同模型进行多径误差预测，实现对不同高度场景下多径误差的精细化建模。构建三维网格布局的步骤如下：

步骤 1 二维格网布局。通过六边形格网的紧密排布，对实验所在的城市区域进行划分。基于GIS提供的城市区域地理信息和GNSS先验位置信息，选定所需划分的城市区域，并根据选定区域的实际情况确定六边形格网边长。二维格网划分方法如图2所示。



图2 城市区域二维格网划分示意图

Fig. 2 2D grid division of urban areas diagram

步骤 2 高度划分。城市复杂环境下，同一平面不同高度上的多径环境往往存在差异，对城市区域高度空间的进一步划分是实现精细化建模的关键。根据GNSS先验位置信息中包含的高度信息，以先验位置中的最低点作为起始平面，选定合适的高度间隔，对每个二维格网对应的高度区域进行均匀划分。如图3所示，对于先验位置P_1和P_2，在二维格网布局下属于同一平面格网，经高度划分后，两者分别归属不同高度的立体格网，在定位过程中调用各自格网的相应模型进行多径误差预测。

格网的时空分辨率对定位结果影响显著，时空分辨率过大会导致模型精细化程度降低，过小则会导致计算效率降低。实验区域为高架场景，为将高架上下层划分到不同格网层，将格网高度设为4 m；考虑到实验区域的横向道路宽度和纵向场景变化情况，相隔20 m以内的场景大致可以认为多径环境相同。本文建立的三维六边形格网，时间分辨率为2 h，空间分辨率为边长20 m，高度4 m。由于全球定位系统（global positioning system, GPS）和北斗卫星导航系统（BeiDou navigation satellite system, BDS）的卫星轨道周期和卫星星座分布不同，本文对2个卫星星座分别进行训练数据划分。基于构建的三维格网布局及先验位置信息，将训练样本与其所处位置的格网相匹配，对每个格网内的不同星座样本建立不同的多径误差预测规则。

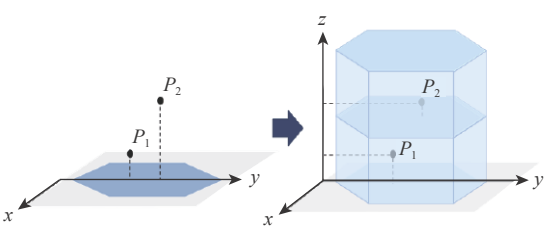


图3 高度划分示意图

Fig. 3 Diagram of the division in height

2.2 基于袋装回归树的多径误差预测模型

袋装树通过组合许多决策树的预测来做出最终决策，可以显著提高分类性能[21-22]。训练过程如图4所示。基于集成袋装回归树的多径误差预测过程描述如下：

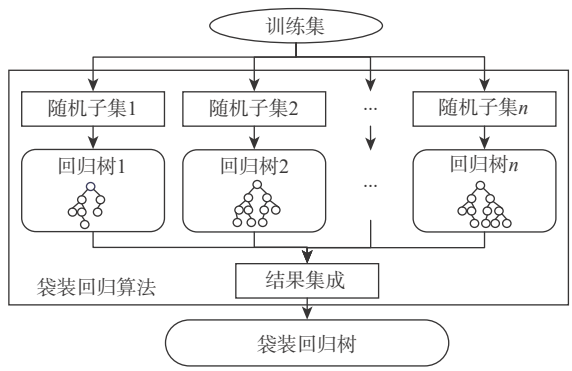


图4 袋装回归树算法流程

Fig. 4 Flowchart of the ensemble bagged regression tree algorithm

步骤 1 基于单个回归树的模型建立。

a）选择最佳分割变量j和分割点s来分割区域：

$$\begin{cases} R_1(j,s)=\{x|x_j<s\} \\ R_2(j,s)=\{x|x_j\geqslant s\} \end{cases} \tag{4}$$

$$A_1=\sum_{x_i\in R_1(j,s)}(\Delta\rho_i-\tilde{y}_{R_1})^2 \tag{5}$$

$$A_2=\sum_{x_i\in R_2(j,s)}(\Delta\rho_i-\tilde{y}_{R_2})^2 \tag{6}$$

$$r_{\mathrm{RSS}}=\min_{j,s}(A_1+A_2) \tag{7}$$

$$\tilde{y}_{R_1}=\mathrm{average}(y_i|x_i\in R_1(j,s)) \tag{8}$$

$$\tilde{y}_{R_2}=\mathrm{average}(y_i|x_i\in R_2(j,s)) \tag{9}$$

式中：$R_1(j,s)$和$R_2(j,s)$为分割后的2个区域；A_1和A_2分别为两个区域的残差平方和（residual sum of squares, RSS）；r_{RSS}为残差平方和，遍历最佳分割变量j，找到每个最佳分割变量j下对应的分割点s使

r_{RSS} 最小; average 表示取均值; $\tilde{y}_{R_1}$ 和 $\tilde{y}_{R_2}$ 分别为 2 个划分区域预测值的平均值; y_i 为任一预测值; x_i 为任一元素。2 个区域输出预测值 $\hat{c}_q$ 为

$$\hat{c}_q = \frac{1}{N_q}\sum_{x_i \in R_q(j,s)} \Delta\rho_i \qquad x \in R_q, q = 1,2 \tag{10}$$

b) 对于 2 个子区域, 循环步骤 a) 直到满足停止迭代的条件。随后建立回归树:

$$h(x) = \sum_{l=1}^{L} \hat{c}_l I(x \in R_l) \qquad l = 1,2,3,\cdots,L \tag{11}$$

式中: $h(x)$为回归树的预测函数; l为区域序号; I为特征函数。

步骤 2　基于集成袋装树的模型建立。融合步骤 1 中获得的所有基于单个回归树的模型, 将其预测平均值作为袋装回归树最终输出:

$$H_{\mathrm{BAG}}(x) = \frac{1}{N}\sum_{n=1}^{N} h_n(x) \tag{12}$$

式中: $h_n(x)$为第n颗回归树的预测函数。

3　基于多径误差预测的 GNSS/IMU 抗差自适应滤波算法

基于三维格网的位置匹配, 调用对应环境的多径误差预测模型, 并基于该预测值$\Delta\hat{\rho}$进行滤波模型的环境自适应选择。为解决传统抗差自适应滤波模型中抗差权函数阈值难以确定的问题, 利用多径误差预测值$\Delta\tilde{\rho}$构建权函数动态阈值的抗差滤波算法, 如图 5 所示。

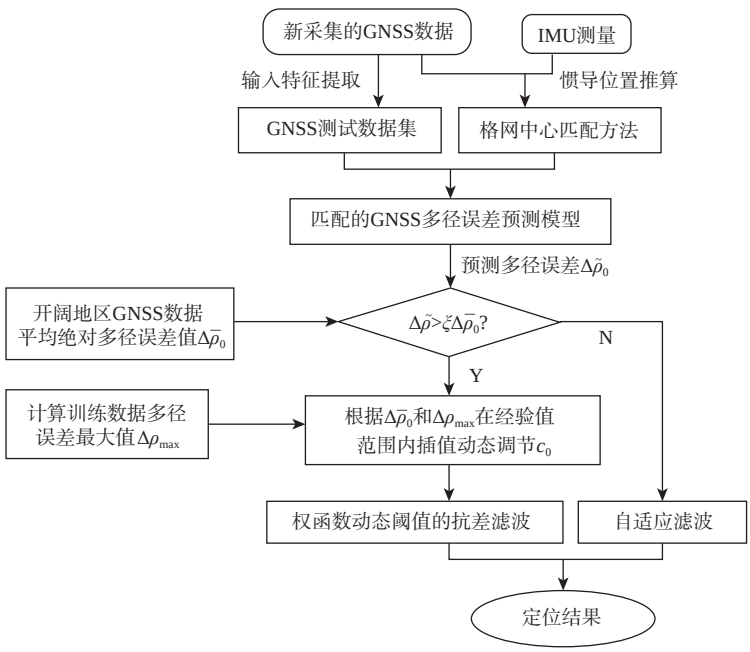


图 5　三维格网辅助的抗差自适应滤波算法框图

Fig. 5　Flowchart of the 3D grid-aided robust adaptive filtering algorithm

3.1　基于三维格网中心匹配的多径误差预测

对于采集的测试数据, 将当前历元惯导推算位置作为当前历元初步定位结果。先根据初步定位结果进行二维格网匹配, 设六边形格网中心点为 O, 东北天坐标系下的 E 轴和 N 轴分别对应 x 轴和 y 轴, 取某训练样本的位置为 P_0, 判定其是否属于该格网:

$$\begin{cases} y_0 \leqslant a \\ x_0 \leqslant \dfrac{\sqrt{3}a}{2} \end{cases} \tag{13}$$

$$a - y_0 \leqslant \frac{\sqrt{3}x_0}{2} \tag{14}$$

式中: a 为格网边长; x_0和y_0分别为P_0到 x 轴和 y 轴的距离, 如图 6 所示。

仅当式(13)、式(14)都满足时, 才能判定P_0在该二维格网内, 否则, 需对下一个二维格网执行判断。为缓解惯导推算结果不准确导致的格网错误匹配, 判断P_0是否落在二维格网中心区域:

$$\sqrt{x_0^2 + y_0^2} < \lambda a \tag{15}$$

式中: λ为平面匹配系数, $0 < \lambda < \dfrac{\sqrt{3}}{2}a$。本文中$\lambda$取 0.6。

平面匹配完成后, 继续进行高度匹配, 由于最底层格网的起始平面高度为先验数据中的最低高度, 且每一层格网高度均匀, 故而每一层格网的高度范围是确定且已知的。根据惯导推算的高度信息匹配到相应层数的格网, 判断是否位于格网高度中心:

$$h_1 + \mu < z_0 < h_2 - \mu \tag{16}$$

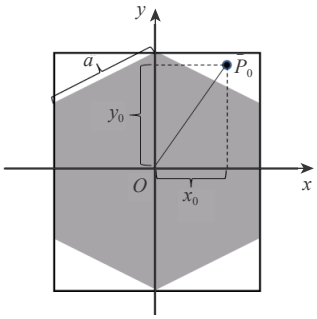


图6 二维格网匹配示意图

Fig. 6 Diagram of 2D grid matching

式中：z_0为P_0的高度坐标；h_1和h_2分别为P_0所属立体格网的底层和顶层高度；μ为高度匹配因子，本文中μ取1 m。

若满足式(16)，则调用该格网的多径误差预测模型：

$$\Delta\hat{\rho}=\begin{cases}H_{\mathrm{BAG}}^{\mathrm{GPS}}(x) & \mathrm{SV}<33\\ H_{\mathrm{BAG}}^{\mathrm{BDS}}(x) & \mathrm{SV}\geqslant 33\end{cases} \tag{17}$$

式中：$\Delta\hat{\rho}$为当前历元当前卫星的多径误差预测值，本文仅使用GPS和BDS这个星座的数据，不考虑其他星座，故以卫星号SV是否小于33作为星座类别判断；$H_{\mathrm{BAG}}^{\mathrm{GPS}}$和$H_{\mathrm{BAG}}^{\mathrm{BDS}}$分别为GPS卫星和BDS卫星的多径误差预测函数；$X$为当前历元当前卫星的输入特征，$X=\{C/N_0,\theta_{\mathrm{e}},\theta_{\mathrm{a}}\}$。

3.2 基于多径误差建模的抗差自适应滤波算法

3.2.1 GNSS/IMU组合导航抗差自适应滤波模型

本文GNSS/IMU组合方式为紧组合，借助卡尔曼滤波框架，构建系统误差状态向量：

$$\boldsymbol{X}=\begin{bmatrix}\delta\boldsymbol{r}_{\mathrm{IMU}}^{\mathrm{T}} & \delta\boldsymbol{v}_{\mathrm{IMU}}^{\mathrm{T}} & \delta\boldsymbol{\phi}_{\mathrm{IMU}}^{\mathrm{T}} & \boldsymbol{b}_{\mathrm{g}}^{\mathrm{T}} & \boldsymbol{b}_{\mathrm{a}}^{\mathrm{T}}\\ \boldsymbol{s}_{\mathrm{g}}^{\mathrm{T}} & \boldsymbol{s}_{\mathrm{a}}^{\mathrm{T}} & t_{\mathrm{b}} & \delta t_{\mathrm{b}}\end{bmatrix}^{\mathrm{T}} \tag{18}$$

式中：$\delta\boldsymbol{r}_{\mathrm{IMU}}$、$\delta\boldsymbol{v}_{\mathrm{IMU}}$和$\delta\boldsymbol{\phi}_{\mathrm{IMU}}$分别为IMU的三维位置、速度和姿态误差向量；$\boldsymbol{b}_{\mathrm{g}}$和$\boldsymbol{b}_{\mathrm{a}}$分别为陀螺仪和加速度计三轴零偏；$\boldsymbol{s}_{\mathrm{g}}$和$\boldsymbol{s}_{\mathrm{a}}$分别为陀螺仪和加速度计三轴比例因子误差向量；$t_{\mathrm{b}}$和$\delta t_{\mathrm{b}}$分别为GNSS接收机钟差和钟漂。

采用卡尔曼滤波器进行数据融合：

$$\hat{\boldsymbol{X}}_{k,k-1}=\boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{X}}_{k-1} \tag{19}$$

$$\boldsymbol{\Sigma}_{k,k-1}=\boldsymbol{\Phi}_{k,k-1}\boldsymbol{\Sigma}_{k-1}\boldsymbol{\Sigma}_{k,k-1}^{\mathrm{T}}+\boldsymbol{Q}_{k-1} \tag{20}$$

$$\boldsymbol{K}_k=\boldsymbol{\Sigma}_{k,k-1}\boldsymbol{H}_k^{\mathrm{T}}\left(\boldsymbol{H}_k\boldsymbol{\Sigma}_{k,k-1}\boldsymbol{H}_k^{\mathrm{T}}+\boldsymbol{R}_k\right)^{-1} \tag{21}$$

$$\boldsymbol{V}_k=\boldsymbol{Z}_k-\boldsymbol{H}_k\hat{\boldsymbol{X}}_{k,k-1} \tag{22}$$

$$\hat{\boldsymbol{X}}_k=\hat{\boldsymbol{X}}_{k,k-1}+\boldsymbol{K}_k\boldsymbol{V}_k \tag{23}$$

$$\boldsymbol{\Sigma}_k=(\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\Sigma}_{k,k-1}(\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k)^{\mathrm{T}}+\boldsymbol{K}_k\boldsymbol{R}_k\boldsymbol{K}_k^{\mathrm{T}} \tag{24}$$

式中：$\hat{\boldsymbol{X}}_{k,k-1}$为一步预测状态向量；$\boldsymbol{\Phi}_{k,k-1}$为状态转移矩阵；$\boldsymbol{\Sigma}_{k,k-1}$为一步预测状态协方差矩阵；$\boldsymbol{Z}_k$为观测向量；$\boldsymbol{H}_k$为观测矩阵；$\boldsymbol{V}_k$为新息；$\hat{\boldsymbol{X}}_k$为估计状态向量；$\boldsymbol{I}$为单位矩阵；$\boldsymbol{K}_k$为滤波增益矩阵；$\boldsymbol{R}_k$为观测噪声对称正定方差矩阵；$\boldsymbol{Q}_{k-1}$为动力学模型误差非负定方差矩阵；$\boldsymbol{\Sigma}_k$为估计状态协方差阵。

为提高卡尔曼滤波的抗差性能和自适应性能，引入抗差估计和自适应滤波分别抑制观测异常和动力学模型误差对卡尔曼滤波的影响，$\hat{\boldsymbol{X}}_k$可写为[16]

$$\hat{\boldsymbol{X}}_k=\left(\boldsymbol{H}_k^{\mathrm{T}}\overline{\boldsymbol{P}}_k\boldsymbol{H}_k+\alpha_k\boldsymbol{P}_{\hat{\boldsymbol{X}}_{k-1}}\right)^{-1}\left(\boldsymbol{H}_k^{\mathrm{T}}\overline{\boldsymbol{P}}_k\boldsymbol{Z}_k+\alpha_k\boldsymbol{P}_{\hat{\boldsymbol{X}}_{k-1}}\hat{\boldsymbol{X}}_{k-1}\right) \tag{25}$$

$$\overline{\boldsymbol{P}}_k=\left(\overline{\boldsymbol{R}}_k\right)^{-1}=\left(\frac{\boldsymbol{R}_k}{\gamma_k}\right)^{-1} \tag{26}$$

$$\boldsymbol{P}_{\hat{\boldsymbol{X}}_{k-1}}=(\boldsymbol{\Sigma}_{k,k-1})^{-1} \tag{27}$$

式中：$\overline{\boldsymbol{P}}_k$为观测向量的权矩阵；$\boldsymbol{P}_{\hat{\boldsymbol{X}}_{k-1}}$为一步预测状态向量$\hat{\boldsymbol{X}}_{k-1}$的权矩阵；$\gamma_k$和$\alpha_k$分别为抗差因子矩阵和自适应因子。抗差因子矩阵权函数矩阵可通过huber函数构建：

$$\gamma_k^i=\begin{cases}1 & \left|\tilde{V}_i^k\right|\leqslant c_0\\ \dfrac{c_0}{\left|\tilde{V}_i^k\right|} & \left|\tilde{V}_i^k\right|>c_0\end{cases} \tag{28}$$

式中：γ_k^i为抗差因子矩阵的第i个对角线元素；$\tilde{V}_k^i$为标准化预测残差，$\tilde{V}_k^i=\dfrac{V_k^i}{\sqrt{\Sigma_{\boldsymbol{V}_k}^{ii}}}$，其中，$i$表示第$i$个对角线元素，$\Sigma_{\boldsymbol{V}_k}^{ii}$为预测残差协方差矩阵的第$i$个对角线元素；$c_0$为常数，取值范围通常为1.5～2.0。

构建基于预测残差统计量$\Delta\overline{V}_k$的自适应因子如下[16]：

$$\alpha_k=\begin{cases}1 & \Delta\overline{V}_k\leqslant c_1\\ \dfrac{c_1}{\Delta\overline{V}_k} & \Delta\overline{V}_k>c_1\end{cases} \tag{29}$$

$$\Delta\overline{V}_k=\left(\frac{\boldsymbol{V}_k^{\mathrm{T}}\boldsymbol{V}_k}{\mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{V}_k})}\right)^{\frac{1}{2}} \tag{30}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{V}_k}=\boldsymbol{H}_k\boldsymbol{\Sigma}_{k,k-1}\boldsymbol{H}_k^{\mathrm{T}}+\boldsymbol{R}_k \tag{31}$$

式中：tr(·)表示求矩阵的迹；c_1为常数，通常取1～1.5。

3.2.2 基于多径预测值的抗差自适应滤波策略

传统的抗差自适应算法常用PDOP指示GNSS数据质量，从而进行滤波模型选择。但在城市复杂环境中，PDOP并不能有效反映GNSS数据质量[23-24]。本文基于三维格网误差模型输出的多径误差预测值$\Delta\hat{\rho}$来进行滤波模型选择，$\Delta\hat{\rho}$相比PDOP能够更直观地反映GNSS数据质量，基于$\Delta\hat{\rho}$的滤波选择方法能够进一步提升滤波性能。

通常开阔地区的GNSS数据质量较好，借助开阔地区GNSS数据的平均绝对多径误差值$|\Delta\bar{\rho}|$进行

GNSS 数据质量评估:

$$|\Delta\tilde{\rho}| > \xi|\Delta\bar{\rho}| \tag{32}$$

式中: $|\Delta\tilde{\rho}|$为当前历元当前卫星的绝对多径误差预测值; ξ为调节系数, 本文取 1.5。

若式(32)不成立, 则认为当前 GNSS 数据质量好, 选择自适应滤波进行数据融合。若预测残差统计量$\Delta\bar{V}_k$大于c_2, 构造自适应因子缓解 IMU 推算误差; 反之, 则评估结果为当前 GNSS 数据质量差, 采用动态阈值的抗差滤波进行组合导航解算。

传统算法通过 huber 函数构建抗差因子矩阵权函数矩阵, 但其阈值c_0难以确定, 且固定的阈值无法根据环境自适应调节, 造成抗差性能下降。根据抗差理论, 若 GNSS 数据质量差, 则测量协方差R应当变大。抗差过程中, R通过式(33)更新:

$$R_k^i = \left(\gamma_k^i\right)^{-1} * R_k^i \tag{33}$$

结合式(26), 当 GNSS 数据质量差时, 在经验值 1.5～2.0 范围内适当调低阈值c_0, 反之, 则调高c_0, 能够达到更好的抗差性能。GNSS 数据质量可由$\Delta\tilde{\rho}$指示, 基于多径误差预测值$\Delta\tilde{\rho}$、开阔地区 GNSS 数据平均绝对多径误差值$|\Delta\bar{\rho}|$和训练数据多径误差最大值插值动态调节阈值c_0, 如图 7 所示。

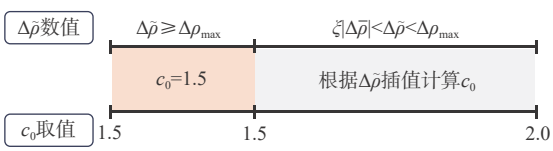


图 7　动态调节c_0示意图

Fig. 7　Diagram of dynamic adjustment c_0

对应的c_0插值调节计算方法如下:

$$c_0 = 1.5 + 0.5\left(\frac{\Delta\tilde{\rho} - \Delta\rho_{\max}}{\xi|\Delta\bar{\rho}| - \Delta\rho_{\max}}\right) \tag{34}$$

4　实验验证与结果分析

4.1　实验设备及场景

为验证本文算法的有效性, 连续多日 15:00—17:00 在南京市南京南站附近沿同一路线进行车载实验。实验路段包含严重城市峡谷区域和高架区域, 如图 8 所示。实验设备如图 9 所示, IMU 为 STIM300, 采样率为 125 Hz; GNSS 接收机为 BDStar Navigation C520-AT, 采样率为 10 Hz, STIM300 的加速度计零偏不稳定性为 4.9×10^{-4} m/s^2, 速度随机游走为 1×10^{-3} m·s^{-2}·Hz$^{-\frac{1}{2}}$; 陀螺仪零偏不稳定性为 2.42×10^{-6} rad/s, 角度随机游走为 4.36×10^{-5} rad·s$^{-\frac{1}{2}}$。参考值由高精度组合导航 HGuide N580 进行数据采集并经过后处理得到, HGuide N580 的加速度计零偏不稳定性为 2.45×10^{-4} m/s^2, 速度随机游走为 5×10^{-4} m·s^{-2}·Hz$^{-\frac{1}{2}}$; 陀螺仪零偏不稳定性为 1.21×10^{-6} rad/s, 角度随机游走为 1.16×10^{-5} rad·s$^{-\frac{1}{2}}$。天线型号均为 ZYACF-S806。本文实验共训练了 GPS 和 BDS 格网模型各 763 个。模型训练样本数量情况如表 1所示。为保证模型精度, 仅对样本数量大于 200 的数据集进行训练。



图 8　实验环境

Fig. 8　Experiment environment

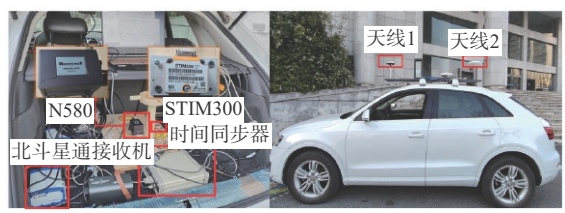


图 9　车载实验平台

Fig. 9　Vehicle test platform

表 1　模型训练样本数量

Table 1　Number of training samples of the models

星座	平均样本数量	最大样本数量	最少样本数量
GPS	979	20 563	200
BDS	1 429	31 901	200

4.2　实验结果及分析

将本文算法与 3 种已有算法进行对比, 分别为算法 1(传统 GNSS/IMU 组合导航算法)、算法 2(GNSS/IMU 抗差自适应滤波算法)、算法 3(二维格网辅助的 GNSS/IMU 抗差自适应滤波算法)、本文算法(三维格网辅助的 GNSS/IMU 抗差自适应算法)。3 种对照算法及本文算法内容如表 2 所示。

卫星数如图 10 所示, PDOP 值如图 11 所示。可知, 第 100 s、400～800 s 和 1 000～1 200 s 的部分历元卫星数量较少, 且 PDOP 过大。

图 12 为参考的真实实验路线及 4 种算法定位轨迹。能够看出, 与其他组合导航算法相比, 本文算法解算结果更接近参考轨迹。场景 a 和场景 b 为高架区域, 其定位误差曲线对应图 13 和图 14 的

表 2 算法描述

Table 2 Algorithm description

算法	算法描述
算法1	1)传统卡尔曼滤波 2)GNSS/IMU紧组合
算法2	1)传统抗差自适应滤波 2)GNSS/IMU紧组合
算法3	1)二维格网误差建模辅助抗差自适应滤波选择 2)二维格网误差建模辅助权函数阈值动态调节 3)GNSS/IMU紧组合
本文算法	1)三维格网误差建模辅助抗差自适应滤波选择 2)三维格网误差建模辅助权函数阈值动态调节 3)GNSS/IMU紧组合

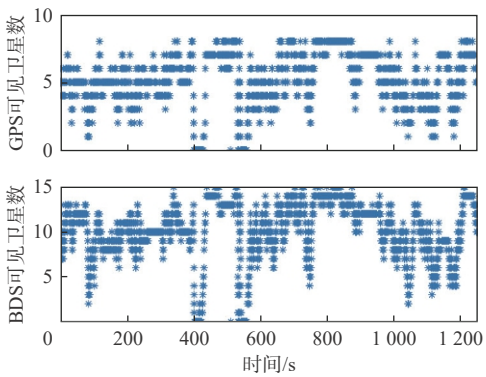


图 10 卫星数

Fig. 10 Number of satellites

1 000～1 250 s 左右，该段时间内本文算法的三维定位误差为 11.81 m，相比于算法 3 的 14.82 m 提升了 20.31%，相比于传统的算法 1 和算法 2 的 15.82 m 和 20.23 m 分别提升了 25.35% 和 41.62%。这是由于算法 3 和本文算法利用格网多径误差建模方法合理选择了滤波模型，并根据多径误差预测值进行了抗差权函数阈值的环境自适应调节。但由于算法 3 采用的二维格网在高架区域不能区分高架上下环境，故而定位表现差于本文算法。场景 c 包含城市复杂环境和一小段隧道，4 种算法在这种场景下的定位误差都相对有所增大，定位误差曲线对应图 13 和图 14 的 200～600 s 左右。在这种场景下，

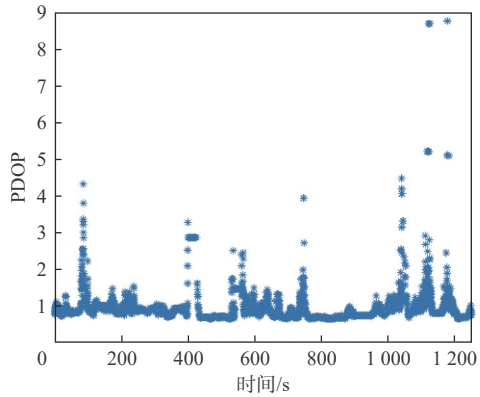


图 11 PDOP 值

Fig. 11 PDOP values

本文算法的三维定位误差为 10.21 m，略高于算法 3 的 10.27 m。相比于传统的算法 1 和算法 2，本文算法的三维定位误差提升率分别为 62.84% 和 26.45%。这是因为在隧道中组合导航的定位结果大多由 IMU 推算，算法 3 和本文算法能够通过误差建模提升进入隧道前的定位和姿态解算精度，在长度较短的隧道内一定程度上能够缓解 IMU 的误差发散。又由于三维格网对城市环境的划分比二维格网更加精密，本文算法相比算法 3 在进入隧道前的定位误差更小。场景 d 的 4 种算法定位误差都相对较小，定位误差曲线对应图 13 和图 14 的 800～1 000 s 左右。在这种情况下，本文算法仍然有最高的三维定位误差，达到 7.76 m，算法 1、算法 2 和算法 3 的三维定位误差分别为 8.52 m、8.47 m 和 7.99 m，进一步证实了城市环境精密划分的有效性。

如表 3 所示，算法 1 的定位误差较大，水平和 3D 定位误差分别为 15.75 m 和 19.16 m。基于传统 GNSS/IMU 抗差自适应滤波的算法 2 的水平和 3D 位置误差分别提升了 32.44% 和 24.74%，说明在城市复杂环境下通过抵抗 GNSS 观测误差和 IMU 推算误差能够明显提升组合导航定位精度。算法 3 采用二维格网多径建模辅助 GNSS/IMU 抗差自适

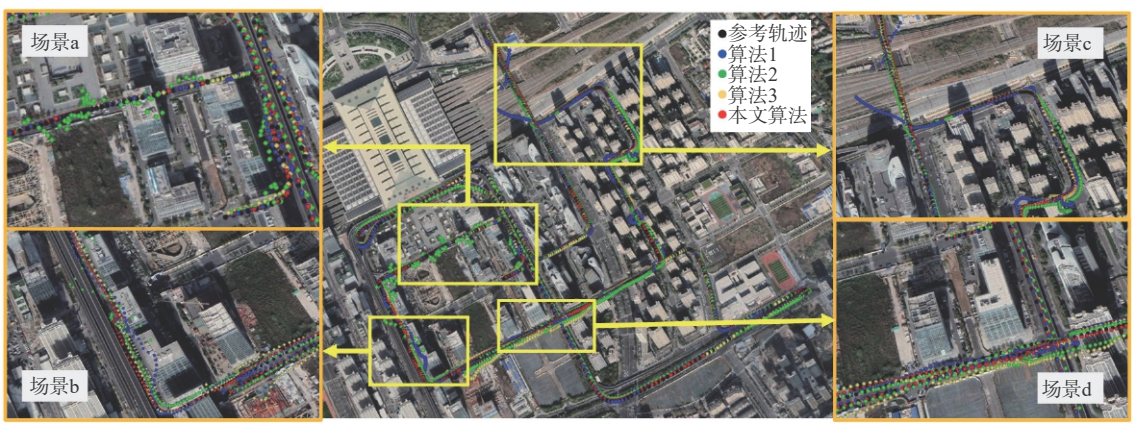


图 12 4 种算法定位轨迹对比

Fig. 12 Comparison of positioning trajectories of the four algorithms

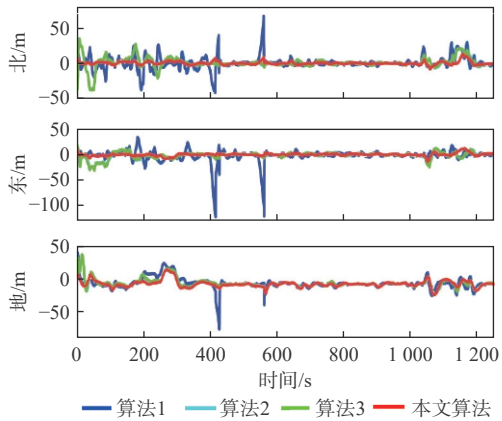


图 13　北东地定位误差

Fig. 13　Positioning error in NED

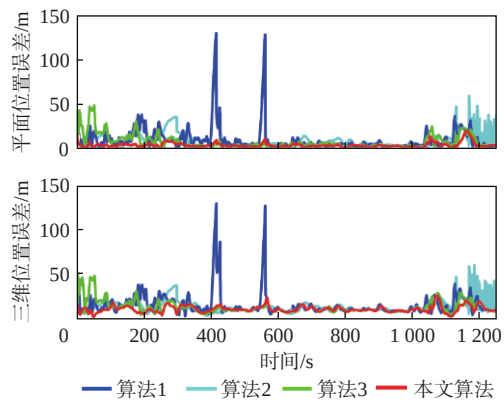


图 14　水平和三维定位误差

Fig. 14　Horizontal and 3D positioning error

应滤波, 相较于算法 1 的水平和 3D 位置误差分别提升了 36.76% 和 28.81%, 相较于算法 2 能够进一步提升三维定位精度至 13.64 m。这充分证明了格网多径建模辅助滤波模型选择和抗差权函数阈值动态调节方法在城市复杂场景下的定位效果。本文算法在北、东和地方向的定位精度分别为 2.62 m、2.87 m 和 9.08 m, 相较算法 3 定位精度进一步提升。这是由于本文算法进一步在高度方向上对城市环境进行了划分, 对于高架、斜坡等在高度上存在一定复杂性的区域, 三维格网能够对不同高度上的不同多径环境分别建立误差预测模型, 从而实现针对不同复杂环境的精细化建模。

表 3　4 种算法定位精度对比

Table 3　Positioning accuracy comparison of the four algorithms

方向	均方根误差/m				提升比例/%				
	算法1	算法2	算法3	本文算法	算法2较算法1	算法3较算法1	本文算法较算法1	本文算法较算法2	本文算法较算法3
北	9.35	6.97	7.76	2.62	25.45	17	71.98	62.41	66.24
东	12.67	8.05	6.24	2.87	36.464	50.75	77.35	64.35	54.01
地	10.91	9.74	9.31	9.08	10.72	14.7	16.77	6.78	2.47
水平	15.75	10.64	9.96	3.88	32.40	36.76	75.37	63.53	61.04
3D	19.16	14.42	13.64	9.88	24.74	28.81	48.43	31.48	27.57

实验整体定位误差如图 13 和图 14 所示。在 0～100 s 内, 由于二维格网误差模型预测不准确, 导致算法 3 误差大于传统算法, 而本文算法采用的三维格网误差模型仍然有较高的预测精度, 能够实现定位精度提升。在后续的 1 150 s, 由于算法 3 和本文算法采用格网误差建模辅助 GNSS/IMU 抗差自适应滤波, 2 种算法的定位精度均高于传统算法 1 和算法 2。本文算法在 200 s 和 1 100 s 左右相较于算法 3, 能够看出明显的定位误差降低, 这是基于三维格网划分高度空间的精密建模带来的定位精度提升。

5　结　论

1) 本文所提基于三维格网布局的多径误差模型细分了高度空间, 并通过格网中心匹配方法缓解了错误匹配带来的定位误差, 与现有建模方法相比, 能够有效提升多径误差预测精度。

2) 基于抗差理论, 本文采用多径误差预测值动态调节抗差滤波中抗差因子权函数阈值, 从而使测量噪声协方差根据所处环境进行自适应调节, 能够进一步提升滤波性能。

3) 从实验结果的 3D 定位精度来看, 本文算法相较传统 GNSS/IMU 紧组合算法提高了 48.43%, 相较传统 GNSS/IMU 抗差自适应算法, 3D 定位精度提升了 31.48%; 相较基于二维格网误差建模 GNSS/IMU 组合导航算法, 3D 定位精度提升了 27.57%。

现有的格网模型均存在一定的时空局限性, 泛化性和迁移性不强。后续研究中, 可以通过使用激光雷达、相机建图等将城市环境作为输入特征之一, 并通过增强学习等方法不断迭代更新模型规则, 使模型能够适配任意时间、任意场景, 提升本文算法的可用性和时空迁移性。

参考文献 (References)

[1] MACGOUGAN G, LACHAPELLE G, KLUKAS R, et al. Performance analysis of a stand-alone high-sensitivity receiver[J]. GPS Solutions, 2002, 6(3): 179-195.

[2] 赵泽方, 杜平, 邵玉. 小型化 3D 扼流圈天线设计[J]. 中国科技论

文, 2016, 11(14): 1605-1608.
ZHAO Z F, DU P, SHAO Y. Design of a miniaturized 3D choke ring antenna[J]. China Sciencepaper, 2016, 11(14): 1605-1608(in Chinese).

[3] 陆思丞, 王昕, 张胜辉. 一种小型化低剖面天线扼流圈的设计[C]//2019年全国微波毫米波会议论文集. 2019: 191-194.
LU S C, WANG X, ZHANG S H. Design of a miniaturized low-profile antenna choke[C]//Proceedings of the 2019 National Conference on Microwave and Millimeter Waves. Guangzhou: Chinese Institute of Electronics, 2019: 191-194.

[4] BRENNEMAN M, MORTON J, YANG C, et al. Mitigation of GPS multipath using polarization and spatial diversities[C]//Proceedings of the 20th International Technical Meeting of the Satellite Division of The Institute of Navigation. 2007: 1221-1229.

[5] JIANG Z Y, GROVES P D. NLOS GPS signal detection using a dual-polarisation antenna[J]. GPS solutions, 2014, 18(1): 15-26.

[6] TRANQUILLA J M, CARR J P, AL-RIZZO H M. Analysis of a choke ring groundplane for multipath control in Global Positioning System (GPS) applications[J]. IEEE Transactions on Antennas and Propagation, 1994, 42(7): 905-911.

[7] 王尔申, 张淑芳, 张芝贤. GPS接收机抗多径技术研究现状与趋势[J]. 电讯技术, 2011, 51(1): 114-119.
WANG E S, ZHANG S F, ZHANG Z X. Research status and trend of mitigating multipath effects technology for GPS receiver[J]. Telecommunication Engineering, 2011, 51(1): 114-119(in Chinese).

[8] SUN R, WANG G, ZHANG W, et al. A gradient boosting decision tree based GPS signal reception classification algorithm[J] Applied Soft Computing, 2020, 86: 105942.

[9] QIN F, FU L X, WANG Y Q, et al. A bagging tree-based pseudorange correction algorithm for global navigation satellite system positioning in foliage canyons[J]. International Journal of Distributed Sensor Networks, 2021, 17(5): 155014772110167.

[10] SUN R, FU L, CHENG Q, et al. Resilient pseudorange error prediction and correction for GNSS positioning in urban areas[J]. IEEE Internet of Things Journal, 2023, 10(11): 9979-9988.

[11] ZHANG Z X, SUN R. Improving loosely coupled GNSS/IMU fusion performance with pseudorange error prediction in urban areas[M]//Advances in Guidance, Navigation and Control. Singapore: Springer Nature, 2023: 2850-2860.

[12] SUN R, ZHANG Z X, CHENG Q, et al. Pseudorange error prediction for adaptive tightly coupled GNSS/IMU navigation in urban areas[J]. GPS Solutions, 2021, 26(1): 28.

[13] 王坚, 刘超, 高井祥, 等. 基于抗差EKF的GNSS/INS紧组合算法研究[J]. 武汉大学学报(信息科学版), 2011, 36(5): 596-600.
WANG J, LIU C, GAO J X, et al. GNSS/INS tightly coupled navigation model based on robust EKF[J]. Geomatics and Information Science of Wuhan University, 2011, 36(5): 596-600(in Chinese).

[14] 吴富梅, 杨元喜. 基于小波阈值消噪自适应滤波的GPS/INS组合导航[J]. 测绘学报, 2007, 36(2): 124-128.
WU F M, YANG Y X. GPS/INS integrated navigation by adaptive filtering based on wavelet threshold de-noising[J]. Acta Geodaetica et Cartographica Sinica, 2007, 36(2): 124-128(in Chinese).

[15] YANG Y, HE H, XU G. Adaptively robust filtering for kinematic geodetic positioning[J]. Journal of Geodesy, 2001, 75(2): 109-116.

[16] 杨元喜, 任夏, 许艳. 自适应抗差滤波理论及应用的主要进展[J]. 导航定位学报, 2013, 1(1): 9-15.
YANG Y X, REN X, XU Y. Main progress of adaptively robust filter with applications in navigation[J]. Journal of Navigation and Positioning, 2013, 1(1): 9-15(in Chinese).

[17] NING Y, WANG J, HAN H, et al. An optimal radial basis function neural network enhanced adaptive robust Kalman filter for GNSS/INS integrated systems in complex urban areas[J]. Sensors, 2018, 18(9): 3091.

[18] GAO B B, HU G G, ZHONG Y M, et al. Cubature Kalman filter with both adaptability and robustness for tightly-coupled GNSS/INS integration[J]. IEEE Sensors Journal, 2021, 21(13): 14997-15011.

[19] SONG B, MA X D, ZHANG S B. Analysis of BDS Satellite Orbit Periodicity[J]. GNSS World of China, 2016, 41(5): 47-50.

[20] YOZEVITCH R, BEN MOSHE B, WEISSMAN A. A robust GNSS LOS/NLOS signal classifier[J]. Navigation, 2016, 63(4): 429-442.

[21] SAEED M S, MUSTAFA M W, SHEIKH U U, et al. Ensemble bagged tree based classification for reducing non-technical losses in multan electric power company of Pakistan[J]. Electronics, 2019, 8(8): 860.

[22] MOUSAVI R, EFTEKHARI M. A new ensemble learning methodology based on hybridization of classifier ensemble selection approaches[J]. Applied Soft Computing, 2015, 37: 652-666.

[23] ATIA M M, WASLANDER S L. Map-aided adaptive GNSS/IMU sensor fusion scheme for robust urban navigation[J]. Measurement, 2019, 131: 615-627.

[24] WANG L, GROVES P D, ZIEBART M. Multi-constellation GNSS performance evaluation for urban canyons using large virtual reality city models[J]. Journal of Navigation, 2012, 65(3): 459-476.

A robust adaptive positioning algorithm for GNSS/IMU based on 3D grid error modeling

SHENG Qi，SUN Rui*，HE Yulin，ZHANG Hengyu

(College of Civil Aviation，Nanjing University of Aeronautics and Astronautics，Nanjing 211106，China)

Abstract: In complex urban environments, tall buildings cause Global Navigation Satellite System (GNSS) signals to suffer from Non-Line-of-Sight (NLOS) propagation and Multipath Interference (MI), which degrade the positioning accuracy of intelligent transportation systems. The existing two-dimensional grid-based multipath modeling method has shortcomings including insufficient precision in the height direction and an overly simplistic adjustment strategy for measurement noise covariance. This article proposes a GNSS/inertial measurement unit (IMU) robust adaptive filter algorithm based on 3D grid error modeling. By dividing the height space on the basis of the existing 2D grid, fine modeling can be further achieved. In the stage of multipath error prediction, the grid-center-matching method is used to alleviate the model prediction error caused by incorrect matching. Then, we propose a filtering model selection strategy based on the multipath error predictions. Moreover, according to the robust theory, we propose a robust threshold dynamic adjustment strategy to update the measurement noise covariance adaptively. The positioning performance of GNSS/IMU integrated navigation in complex urban environments can be significantly improved by the proposed algorithm. The 3D positioning accuracy of the suggested algorithm has improved by 27.57% when compared to the 2D grid assisted GNSS/IMU robust adaptive algorithm and by 48.44% and 31.51%, respectively, when compared to the traditional GNSS/IMU tight combination algorithm and the traditional GNSS/IMU robust adaptive algorithm, according to the results of urban environment vehicle experiments.

Keywords: pseudorange error modeling; robust filtering; adaptive filtering; GNSS/IMU integrated navigation; non-line-of-sight

Received：2024-03-26；**Accepted**：2024-05-31；**Published Online**：2024-08-30 10：47

URL：link.cnki.net/urlid/11.2625.V.20240829.1453.002

Foundation item：Natural Science Foundation of Jiangsu Province (BK20211569)

* **Corresponding author**. E-mail：rui.sun@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0136

考虑多环节间隙非线性的电动舵机多学科耦合动力学建模

杨石琳[1]，乔凯俐[1,2]，万晓忠[1]，段绵超[1,3]，郭宁[1,*]，徐超[1]

(1. 西北工业大学 航天学院，西安 710072；2. 中国飞行试验研究院，西安 710089；
3. 湖北航天技术研究院总体设计所，武汉 430040)

摘 要：由于设计装配不确定性，电动舵机传动系统多环节存在间隙等非线性因素，严重影响工作品质，需建立非线性动力学模型。同时，其涵盖伺服控制、电机传动、减速机构、结构弹性等多学科，建模极具挑战。本研究以典型的滚珠丝杠式电动舵机为研究对象，考虑舵机传动元件的弹性刚度及齿轮副、支撑轴承和拨叉副处的间隙非线性，对电机动力学、伺服控制器进行精细化建模，建立了一种能考虑多环节间隙非线性的电动舵机系统高保真多学科动力学模型，并开展了电动舵机系统性能仿真分析。仿真结果与舵机-舵面系统地面振动试验结果进行了对比。结果表明：所建立的电动舵机多学科耦合动力学模型能够反映机构不同环节间隙非线性的影响和影响程度，仿真和试验结果吻合较好，最大误差小于10%。研究结果为航空航天复杂电动伺服系统的数字孪生研究奠定了基础。

关 键 词：电动舵机；非线性动力学；多学科耦合；间隙；地面振动试验
中图分类号：V19; O322
文献标志码：A **文章编号**：1001-5965(2026)05-1712-08

电动舵机由于其结构简单、控制精度高、使用和维护方便等优点，被广泛应用于各类中小型智能弹药[1]。电动舵机根据控制器给出的指令带动舵面偏转，达到改变飞行姿态及航迹的目的。机械传动机构是电动舵机的重要组成部分，其动力学特性直接影响舵机输出的动态性能，决定着飞行任务的成败[2]。由于加工、制造和装配等，机械传动机构内部存在的间隙、摩擦、饱和等非线性因素会导致电动舵机动态性能下降，造成电动舵机系统不稳定，甚至影响飞行器的飞行品质[3]。因此，详细研究上述非线性因素对电动舵机系统动态性能的影响，进行准确的建模分析是十分重要的工程科学问题。

传统的电动舵机建模往往忽略了电动舵机内部的非线性因素，进行线性化近似。线性化建模方法主要有2类：①采用理想传动比和传递函数[4-7]进行建模，其研究重点在于控制器的选取和舵回路的设计[8]，多用于舵机系统的控制分析，但无法准确考虑间隙、摩擦等非线性特性，导致仿真结果与实际不符[9]；②使用一个刚度为常数的线性扭转弹簧[10-12]来等效替代机械传动机构，这种方法仅适用于传动系统刚度较大的情况[13]，但多项实验结果表明，电动舵机本身具有动态特性，其刚度并非是恒定的常数[14]。

近年来，在舵机系统动力学建模方面，国内外研究逐渐从线性化建模过渡到非线性建模，Ball-

收稿日期：2024-03-08；录用日期：2024-03-29；网络出版时间：2024-04-22 11:45
网络出版地址：link.cnki.net/urlid/11.2625.V.20240418.1320.002
基金项目：国家自然科学基金 (12072268)；北京强度环境研究所联合创新基金 (BQJJ202306)
* 通信作者. E-mail：guoning@nwpu.edu.cn

引用格式：杨石琳，乔凯俐，万晓忠，等. 考虑多环节间隙非线性的电动舵机多学科耦合动力学建模 [J]. 北京航空航天大学学报，2026，52（5）：1712-1719. YANG S L，QIAO K L，WAN X Z，et al. Multidisciplinary coupled dynamics modeling of electro-mechanical actuator considering nonlinear clearances in multiple stages[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（5）：1712-1719（in Chinese）.

esteros 等[15] 使用非线性微分方程来建立舵机的非线性模型, 从而提高电液舵机的动刚度。

Kadiyala 等[16] 建立了电动舵机的齿轮间隙、摩擦非线性与电流饱和引起的非线性模型, 对控制器进行优化设计。卢晋[17] 和张仁嘉[18] 等考虑了多种非线性因素对电动舵机动刚度特性的影响, 但在间隙非线性方面仅考虑了丝杠-拨叉副环节的间隙。张程等[19] 建立了含间隙非线性的折叠舵动力学模型, 使用 Newmark 方法和牛顿迭代法求解了折叠舵的动力学响应。因此, 目前的研究在非线性环节的考虑上多依赖于工程经验, 未全面考虑机械传动机构内部的非线性环节。

本文以典型的航空航天滚珠丝杠式电动舵机系统为研究对象, 建立考虑电动舵机系统多环节间隙因素的动力学模型, 进一步考虑电动舵机控制小回路的影响, 搭建电动舵机多学科耦合模型, 进而探究间隙非线性对电动舵机系统动力学响应的影响, 为电动舵机系统机械设计优化、控制回路设计、舵系统气弹问题[20] 等研究提供更完备的电动舵机系统动力学模型。

1　电动舵机系统机电耦合非线性动力学建模

电动舵机是一类典型的机电一体化系统, 一般主要分为 3 个部分, 分别为伺服控制器、电机和机械传动系统。本文首先分别建立上述子系统模型, 再根据子系统模型之间的输入输出传递关系来集合构建总体电动舵机机电耦合动力学模型。

1.1　伺服控制系统建模

本文研究的电动舵机系统采用比例-微分 (proportional-derivative, PD) 控制, 伺服控制策略采用位置反馈控制, 构建的伺服控制系统框图模型如图 1 所示。

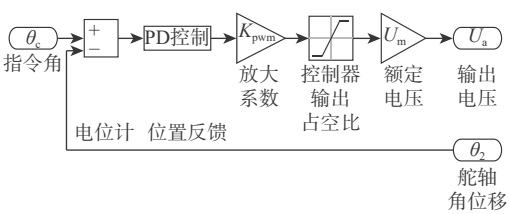


图 1　伺服控制系统框图模型

Fig. 1　Servo control system block diagram model

1.2　电机系统建模

本文研究的电动舵机系统采用无刷直流电机。结合电机电压方程(式 (1))和力矩平衡方程(式 (2)～式 (3))来建立电机模型。

$$U_a = L\frac{di}{dt} + Ri + K_e\frac{d\theta_0}{dt} \tag{1}$$

$$T_e = K_T i \tag{2}$$

$$T_e = J_0\frac{d^2\theta_0}{dt^2} + B\frac{d\theta_0}{dt} + T_{gear} \tag{3}$$

式中: L为电枢相电感; R为电枢相电阻; K_e为反电势常数; i为电枢电流; θ_0为电机转子角位移; t为时间变量; K_T为电磁力矩常数; T_e为电机电磁转矩; J_0为电机转子转动惯量; B为电机阻尼系数; T_{gear}为整体齿轮副对电机轴的反作用力矩。

根据式 (1)～式 (3) 搭建的电机框图模型如图 2 所示, 图中, s 为拉普拉斯变换中的复频域变量。

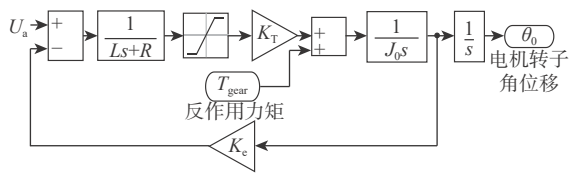


图 2　无刷直流电机框图模型

Fig. 2　Brushless DC motor block diagram model

1.3　含多环节间隙型非线性机械传动系统动力学建模

本文所研究的电动舵机的机械传动系统主要包含齿轮减速系统、滚珠丝杠减速系统、丝杠支撑轴承和拨叉副等部件。考虑上述关键部件均为弹性元件, 利用质量块、弹簧等元件对传动系统进行简化, 质量块视为刚体, 弹簧视为柔性连接。使用有限元软件进行高保真建模, 分别计算齿轮副、滚珠丝杠、支撑轴承及拨叉副的刚度, 采用集中参数建模方法建立电动舵机系统的动力学模型, 如图 3 所示, 其中, x_{zc} 为轴承广义坐标, x_{nut} 为滚珠丝杠螺母的广义坐标, K_{zc} 为轴承的轴向刚度, K_{sg} 为滚珠丝杠的轴向刚度。动力学模型中的广义坐标为齿轮副的输出转角 θ_1、滚珠丝杠螺母的轴线位移 x、轴套舵轴的转角 θ_2 和舵面转角 θ_3。

首先, 假设电机转子为锁死状态, 且忽略机械

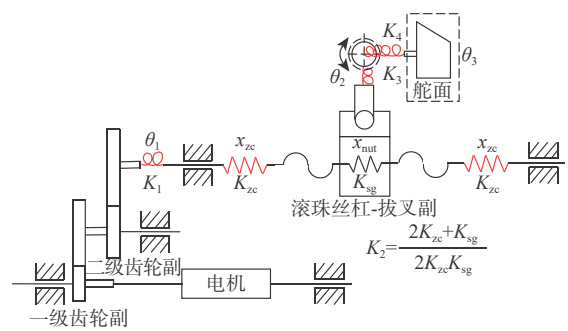


图 3　电动舵机机械传动系统动力学模型

Fig. 3　Dynamic model of electro-mechanical actuator transmission system

传动系统中的非线性环节，则可以列出机械传动系统动力学模型的势能方程和动能方程，即

$$V=\frac{1}{2}K_1\theta_1^2+\frac{1}{2}K_2\left(x-\frac{\theta_1}{i_3}\right)^2+\frac{1}{2}K_3\left(\theta_2-\frac{x}{L_{\mathrm{bc}}}\right)^2+\frac{1}{2}K_4(\theta_3-\theta_2)^2 \tag{4}$$

$$T=\frac{1}{2}\left(J_1i_{13}^2+J_2i_{23}^2+J_3+J_{\mathrm{sg}}\right)\dot{\theta}_1^2+\frac{1}{2}m_{\mathrm{nut}}\dot{x}^2+\frac{1}{2}J_{\mathrm{zt}}\dot{\theta}_2^2+\frac{1}{2}J_{\mathrm{dm}}\dot{\theta}_3^2 \tag{5}$$

式中：T为势能；V为动能；K_1、K_2、K_3和K_4分别为齿轮副整体扭转刚度、滚珠丝杠-轴承串联等效刚度、拨叉副扭转刚度和舵面扭转刚度；i_3为滚珠丝杠传动比；i_{13}^2为第一级齿轮副传动比；i_{23}^2为第二级齿轮副传动比；J_1、J_2、J_3、J_{sg}、J_{zt}和J_{dm}分别为电机齿轮转动惯量、传动齿轮转动惯量、丝杠齿轮转动惯量、丝杠转动惯量、拨叉轴套转动惯量和舵面转动惯量；m_{nut}为螺母质量；L_{bc}为拨叉的力臂长度。

令

$$\begin{cases}J_{\mathrm{A}}=J_1i_{13}^2+J_2i_{23}^2+J_3+J_{\mathrm{sg}}\\M_{\mathrm{B}}=m_{\mathrm{nut}}\\J_{\mathrm{C}}=J_{\mathrm{zt}}\\J_{\mathrm{D}}=J_{\mathrm{dm}}\end{cases} \tag{6}$$

则动能方程可表示为

$$T=\frac{1}{2}J_{\mathrm{A}}\dot{\theta}_1^2+\frac{1}{2}M_{\mathrm{B}}\dot{x}^2+\frac{1}{2}J_{\mathrm{C}}\dot{\theta}_2^2+\frac{1}{2}J_{\mathrm{D}}\dot{\theta}_3^2 \tag{7}$$

将式(4)和式(7)代入第二类拉格朗日方程，可得

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial(T-V)}{\partial\dot{\varphi}_i}-\frac{\partial(T-V)}{\partial\varphi_i}=0 \tag{8}$$

式中：φ_i为广义坐标。

整理后可得线性机械传动系统的动力学方程，具体表示为

$$\begin{cases}J_{\mathrm{A}}\ddot{\theta}_1+\left(K_1+\dfrac{K_2}{i_3^2}\right)\theta_1-\dfrac{K_2}{i_3}x=0\\M_{\mathrm{B}}\ddot{x}-\dfrac{K_2}{i_3}\theta_1+\left(K_2+\dfrac{K_3}{L_{\mathrm{bc}}^2}\right)x-\dfrac{K_3}{L_{\mathrm{bc}}}\theta_2=0\\J_{\mathrm{C}}\ddot{\theta}_2-\dfrac{K_3}{L_{\mathrm{bc}}}x+(K_3+K_4)\theta_2-K_4\theta_3=0\\J_{\mathrm{D}}\ddot{\theta}_3-K_4\theta_2+K_4\theta_3=M_0\sin(\omega t)\end{cases} \tag{9}$$

式中：M_0为舵面负载力矩。

继而，根据机械传动系统中各部件的机构运动特性来考虑传动系统中的间隙非线性环节，具体包含齿轮减速系统中的间隙、滚珠丝杠支撑轴承中的间隙及拨叉副中的间隙。其中，间隙模型采用中心间隙型非线性刚度模型[21]来描述，具体为

$$f_{\mathrm{nl}}(\Delta x_{ij})=\begin{cases}k(\Delta x_{ij}-\delta) & \Delta x_{ij}\geqslant\delta\\0 & -\delta\leqslant\Delta x_{ij}\leqslant\delta\\k(\Delta x_{ij}-\delta) & \Delta x_{ij}\leqslant-\delta\end{cases} \tag{10}$$

式中：Δx_{ij}为第i个自由度和第j个自由度位移之间的差值；δ为间隙值；k为间隙区间外的刚度值。

将式(10)所示的非线性弹性恢复力进一步整理为线性部分和非线性部分的叠加，即

$$f_{\mathrm{nl}}(\Delta x_{ij})=f_{\mathrm{L}}(\Delta x_{ij})+f_{\mathrm{N}}(\Delta x_{ij}) \tag{11}$$

式中：$f_{\mathrm{L}}(\Delta x_{ij})$为非线性弹性恢复力中的线性部分；$f_{\mathrm{N}}(\Delta x_{ij})$为非线性弹性恢复力中的非线性部分，具体可表示为

$$f_{\mathrm{L}}(\Delta x_{ij})=k\Delta x_{ij} \tag{12}$$

$$f_{\mathrm{N}}(\Delta x_{ij})=\begin{cases}-k\delta & \Delta x_{ij}\geqslant\delta\\-k\Delta x_{ij} & -\delta<\Delta x_{ij}<\delta\\k\delta & \Delta x_{ij}\leqslant-\delta\end{cases} \tag{13}$$

式(9)所示的动力学方程中已经包含了非线性刚度中的线性恢复力部分，因此，只需将不同间隙非线性环节的非线性恢复力部分添加在动力学方程(式(9))的右端，整理可得

$$\begin{cases}J_{\mathrm{A}}\ddot{\theta}_1+\left(K_1+\dfrac{K_2}{i_3^2}\right)\theta_1-\dfrac{K_2}{i_3}x=\dfrac{f_{\mathrm{N}}(\theta_1,x)}{i_3}-f_{\mathrm{N}}(\theta_1)\\M_{\mathrm{B}}\ddot{x}+\dfrac{K_3}{L_{\mathrm{bc}}}\left(\dfrac{x}{L_{\mathrm{bc}}}-\theta_2\right)+K_2\left(x-\dfrac{\theta_1}{i_3}\right)=\dfrac{f_{\mathrm{N}}(x,\theta_2)}{L_{\mathrm{bc}}}-f_{\mathrm{N}}(\theta_1,x)\\J_{\mathrm{C}}\ddot{\theta}_2+K_3\left(\theta_2-\dfrac{x}{L_{\mathrm{bc}}}\right)=-f_{\mathrm{N}}(x,\theta_2)\\J_{\mathrm{D}}\ddot{\theta}_3-K_4\theta_2+K_4\theta_3=M_0\sin(\omega t)\end{cases} \tag{14}$$

式中：$f_{\mathrm{N}}(\theta_1)$、$f_{\mathrm{N}}(\theta_1,x)$和$f_{\mathrm{N}}(x,\theta_2)$分别为齿轮减速系统间隙、滚珠丝杠支撑轴承间隙和拨叉副间隙的非线性恢复力部分，其表达式分别为

$$f_{\mathrm{N}}(\theta_1)=\begin{cases}-K_1\delta_1 & \theta_1\geqslant\delta_1\\-K_1\theta_1 & -\delta_1<\theta_1<\delta_1\\K_1\delta_1 & \theta_1\leqslant-\delta_1\end{cases} \tag{15}$$

$$f_{\mathrm{N}}(\theta_1,x)=\begin{cases}-K_2\delta_2 & \Delta x_{12}\geqslant\delta_2\\-K_2\Delta x_{12} & -\delta_2<\Delta x_{12}<\delta_2\\K_2\delta_2 & \Delta x_{12}\leqslant-\delta_2\end{cases} \tag{16}$$

$$f_{\mathrm{N}}(x,\theta_2)=\begin{cases}-K_3\delta_3 & \Delta x_{23}\geqslant\delta_3\\-K_3\Delta x_{23} & -\delta_3<\Delta x_{23}<\delta_3\\K_3\delta_3 & \Delta x_{23}\leqslant-\delta_3\end{cases} \tag{17}$$

式中：

$$\Delta x_{12}=x-\frac{\theta_1}{i_3} \tag{18}$$

$$\Delta x_{23}=\theta_2-\frac{x}{L_{\mathrm{bc}}} \tag{19}$$

式中: δ_1、δ_2和δ_3分别为存在于齿轮减速系统、轴承和拨叉副中的间隙值。

根据式 (14) 搭建电动舵机的机械传动系统框图模型, 如图 4 所示其中, T_{bc} 为拨叉副作用于滚珠丝杠的反作用力矩, T_{dm} 为舵面作用于拨叉副-轴套的反作用力矩。

1.4　耦合动力学整体模型

将上述建立的伺服控制系统、电机系统和机械传动系统模型按照对应的输入输出关系连接起来, 则可构建滚珠丝杠式电动舵机整体机电耦合动力学模型, 如图 5 所示。其中, 输入为指令角θ_{in}和舵面所受的外力矩T_{load}, 输出为舵面转角θ_3, T_0 为齿轮副作用于电机的反作用力矩。

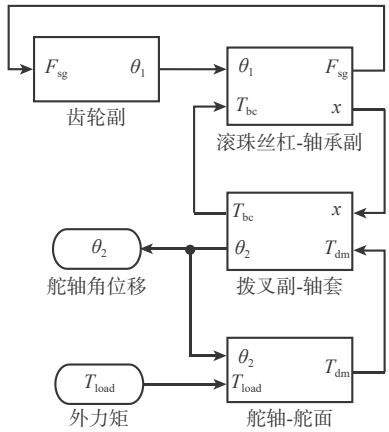


图 4　电动舵机机械传动系统框图模型

Fig. 4　Block diagram model of electro-mechanical actuator transmission system

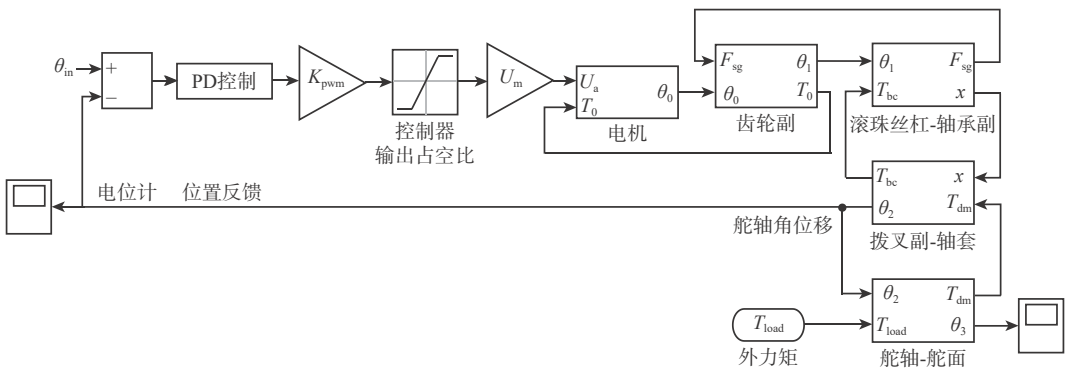


图 5　电动舵机整体机电耦合框图模型

Fig. 5　Electromechanical coupling block diagram model of electro-mechanical actuator

2　电动舵机机电耦合系统动态性能分析

为验证建模的准确性和有效性, 基于 MATLAB/Simulink 搭建了 3 种电动舵机机电耦合系统模型, 分别为: ①控制系统设计常用的电动舵机简化模型, 即采用传动比等效代替机械传动系统; ②基于图 5 建立的电动舵机精细化模型, 但不考虑其中的间隙非线性环节; ③非线性电动舵机精细化模型, 即在第 2 种模型的基础上考虑了齿轮副间隙、轴承间隙和拨叉副间隙, 根据实际设计公差范围确定的各环节间隙值见表 1。

表 1　非线性电动舵机系统各环节间隙设置情况

Table 1　Clearance settings for various sectors in nonlinear electro-mechanical actuator system

工况	齿轮副间隙/rad	轴承间隙/mm	拨叉副间隙/rad
间隙最小组合	0.001	0.001	0.000 2
间隙最大组合	0.01	0.01	0.000 4

通过对上述 3 种模型进行性能仿真分析, 验证其输出是否满足电动舵机系统性能指标要求, 并与电动舵机试验阶跃响应数据进行对比研究。

在空载情况下, 分别对 3 种电动舵机模型施加幅值为 1°的典型阶跃信号, 通过仿真获得的舵偏角阶跃响应曲线与试验结果对比情况如图 6 所示。

不同电动舵机模型的舵偏角阶跃响应分析结果见表 2。

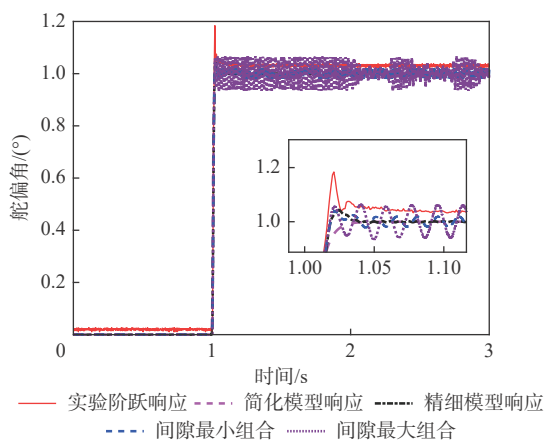


图 6　1°阶跃激励下不同舵机仿真模型的结果对比

Fig. 6　Comparison of the results of different actuator simulation models under a 1° step excitation

表2 1°阶跃激励下电动舵机模型舵偏角阶跃响应仿真与试验结果

Table 2 Simulation and experimental results of step response of rudder angle in electro-mechanical actuator model under a 1° step excitation

工况	超调量/%	过渡过程时间/ms
指标	≤25	≤40
试验	15.06	16.00
简化模型	0.00	20.43
精细模型	4.40	16.83
间隙最小模型	4.20	16.83
间隙最大模型	6.70	16.45

综合图6和表2可以看出，在1°阶跃信号下，3种电动舵机机电耦合模型分析结果的超调量和过渡过程时间均满足性能指标要求，说明上述电动舵机模型的建模均是合理可行的。

综合比较线性简化模型和精细模型的阶跃响应曲线与试验阶跃响应曲线，可以看出，简化模型阶跃响应曲线无超调量，精细模型阶跃响应曲线和试验阶跃响应曲线均先超调后逐渐趋于稳定，稳定后呈现小幅度的周期性振荡。由于考虑了机械传动系统弹性刚度，在阶跃等宽带输入条件下，电动舵机精细模型能够预测宽频激励引起的弹性动力学现象，因而，相比于传统的电动舵机简化模型，描述能力更强，也更为完备。

综合线性电动舵机模型和非线性电动舵机模型的阶跃响应曲线与试验阶跃响应曲线可以看出，由于考虑了机械传动系统各部件中的间隙环节，非线性电动舵机模型的阶跃响应能够预测由于间隙存在导致电动舵机系统自激振荡而产生的振荡现象[22-26]，因此，其响应曲线在稳态段出现波动现象，且随着间隙值的增大，波动值也逐渐增大。结果表明，间隙最小组合对应的阶跃响应波动为等幅波动，幅值为0.02°；间隙最大组合对应的阶跃响应波动幅值不断变化，最大幅值为0.07°。

3 零位控制下电动舵机机电耦合系统动力学分析

通过第2节性能仿真分析验证了本文建立的考虑非线性环节和机械传动系统弹性刚度的电动舵机模型的有效性和建模方法的正确性。本节主要在上述模型的基础上探究传动系统各个部件中的间隙环节对电动舵机系统动力学特性的影响。

工况设置为给定电动舵机系统零位保持的控制指令，同时，采用步进正弦扫频信号激励舵面扭转振动，扫频范围为100～200 Hz，分别设置激励力幅值为10 N、20 N、30 N、40 N和50 N，力臂长度为135 mm，通过动力学分析探究不同环节的间隙非线性对电动舵机系统动态特性的影响。

3.1 考虑多环节间隙非线性电动舵机动力学分析

本节探究考虑多环节间隙非线性下的电动舵机系统的动态响应，其中，非线性电动舵机间隙值见表1。绘制考虑多环节间隙的电动舵机系统在间隙边界工况下的共振频率-共振幅值曲线，如图7所示。

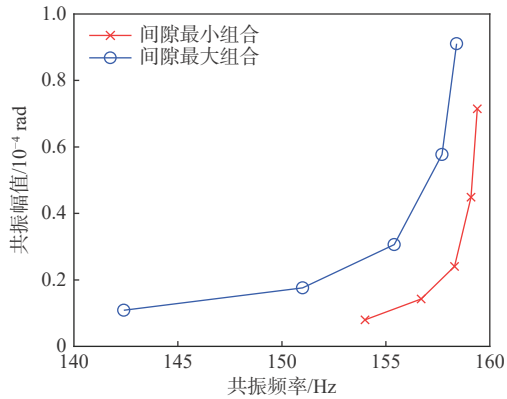


图7 考虑多环节间隙的电动舵机舵偏角共振频率-共振幅值曲线

Fig. 7 Resonance frequency-amplitude curve of rudder angle for electro-mechanical actuator considering various sectors clearances

利用实验室现有条件进行非线性动力学试验，提取电动舵系统在不同激振力下的一阶扭转共振频率，并与相同工况下的舵机机电耦合模型在间隙上、下边界工况下的共振频率变化进行对比，绘制激励力-共振频率曲线，如图8所示。

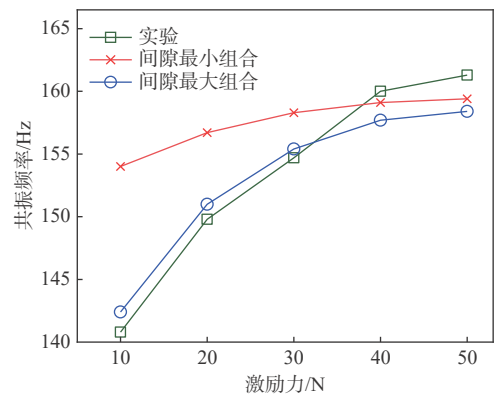


图8 考虑多环节间隙的电动舵机舵偏角激励力-共振频率曲线

Fig. 8 Excitation force-resonance frequency curve of rudder angle for electro-mechanical actuator considering various sectors clearances

由图7可知，当设置间隙最小组合时，共振频率从154 Hz增加到159.4 Hz，增加了3.5%；当设置间隙最大组合时，其共振频率从142.4 Hz增加到158.4 Hz，增加了11.2%。随着激励力的增大，间隙最小组合与间隙最大组合的共振频率逐渐靠近。

综合图8和表3可以看出，与试验结果相比，在各激励力幅值下，间隙最小组合仿真结果和间隙最大组合仿真结果的共振频率与试验结果误差均不超过10%，表明本文建立的电动舵机系统机电耦合模型能

表 3　考虑多环节间隙的电动舵机舵偏角共振频率数据

Table 3　Resonance frequency data of rudder angle for electro-mechanical actuator considering various sectors clearances

激励力/N	试验共振频率/Hz	间隙最小组合共振频率/Hz	与试验结果误差/%	间隙最大组合共振频率/Hz	与试验结果误差/%
10	140.8	154.0	9.4	142.4	1.1
20	149.8	156.7	4.6	151.0	0.8
30	154.7	158.3	2.3	155.4	0.5
40	160.0	159.1	2.0	157.7	1.4
50	161.3	159.4	1.2	158.4	1.8

够实现正向的高精度系统性能预测。

3.2　考虑单环节间隙非线性电动舵机动力学分析

在 3.1 节基础上, 本节分别在齿轮副、轴承和拨叉副中添加间隙环节, 间隙工况设置同表 1。绘制考虑单环节间隙的电动舵机系统在间隙边界工况下的共振频率-共振幅值曲线, 如图 9 所示。

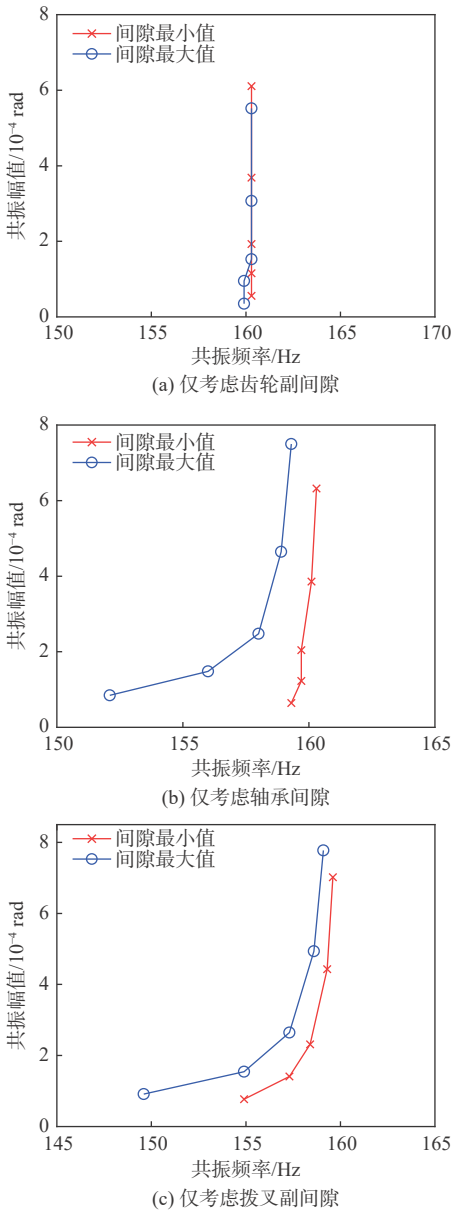


图 9　考虑单环节间隙的电动舵机舵偏角共振频率-共振幅值曲线

Fig. 9　Resonance frequency-amplitude curve of rudder angle for electro-mechanical actuator considering single sector clearance

综合图 9(a) 和表 4 可以看出, 当齿轮副间隙取最小值时, 随着激励力幅值的增大, 共振频率基本保持不变, 具体为 160.3 Hz; 当齿轮副间隙取最大值时, 随着激励力幅值的增大, 共振频率从 159.9 Hz 增加到 160.3 Hz, 增加了 0.3%。

表 4　不同激励工况下考虑单环节间隙的电动舵机舵偏角共振频率和共振幅值数据

Table 4　Resonance frequency and amplitude data of rudder angle for electro-mechanical actuator considering single sector clearance under various excitation conditions

工况	激励力/N	间隙最小共振频率/Hz	间隙最小共振幅值/rad	间隙最大共振频率/Hz	间隙最大共振幅值/rad
齿轮副间隙	10	160.3	5.589×10^{-5}	159.9	3.494×10^{-5}
	20	160.3	1.154×10^{-4}	159.9	9.493×10^{-5}
	30	160.3	1.928×10^{-4}	160.3	1.525×10^{-4}
	40	160.3	3.685×10^{-4}	160.3	3.069×10^{-4}
	50	160.3	6.108×10^{-4}	160.3	5.518×10^{-4}
轴承间隙	10	159.3	6.438×10^{-5}	152.1	8.463×10^{-5}
	20	159.7	1.227×10^{-4}	156.0	1.481×10^{-4}
	30	159.7	2.040×10^{-4}	158.0	2.478×10^{-4}
	40	160.1	3.856×10^{-4}	158.9	4.646×10^{-4}
	50	160.3	6.321×10^{-4}	159.3	7.499×10^{-4}
拨叉副间隙	10	154.9	7.710×10^{-5}	149.6	9.128×10^{-5}
	20	157.3	1.408×10^{-4}	154.9	1.542×10^{-4}
	30	158.4	2.314×10^{-4}	157.3	2.647×10^{-4}
	40	159.3	4.428×10^{-4}	158.6	4.937×10^{-4}
	50	159.6	7.020×10^{-4}	159.1	7.772×10^{-4}

综合图 9(b) 和表 4 可以看出, 当轴承间隙取最小值时, 共振频率从 159.3 Hz 增加到 160.3 Hz, 增加了 0.6%; 当轴承间隙取最大值时, 共振频率从 152.1 Hz 增加到 159.3 Hz, 增加了 4.7%。

综合图 9(c) 和表 4 可以看出, 当拨叉副间隙取最小值时, 共振频率从 154.9 Hz 增加到 159.6Hz, 增加了 3.0%; 当拨叉副间隙取最大值时, 共振频率从 149.6 Hz 增加到 159.1 Hz, 增加了 6.4%。

由上述分析可得, 对于电动舵机系统而言, 拨叉副间隙对其动态特性影响最大, 轴承间隙次之, 齿轮副间隙影响最小。考虑单环节间隙的电动舵机舵偏角共振频率-共振幅值曲线的趋势与 3.1 节仿真分析结果类似。该结论与传统电动舵机建模中依赖的工程经验相符合, 即越靠近舵轴部件的间

隙环节对电动舵机系统动态特性的影响越大。仿真结论为电动舵机非线性动力学建模与实际工程经验提供了理论支撑。

4 结 论

本文以滚珠丝杠式电动舵机为研究对象，从电动舵机传动系统动力学建模、电动舵机机电耦合建模和动力学分析等方面展开了研究工作，主要研究结论如下：

1）针对传统舵机动力学建模中非线性环节考虑不全的问题，本文模型考虑了传动机构中的多环节间隙非线性，并对电机和伺服控制器进行了精细建模，所建立的模型是多学科机电耦合动力学模型。

2）通过仿真分析与试验结果对比，验证了本文模型的有效性和正确性。本文模型能够反映机构不同环节间隙非线性的影响和影响程度，仿真和试验结果吻合较好，最大误差小于10%。

3）不同环节间隙非线性对比研究表明，拨叉副间隙对电动舵机机电耦合模型动态特性影响最大，轴承间隙次之，齿轮副间隙影响最小。

参考文献（References）

[1] YOO C H, LEE Y C, LEE S Y. A robust controller for an electromechanical fin actuator[C]//Proceedings of the 2004 American Control Conference. Piscataway: IEEE Press, 2005, 5: 4010-4015.

[2] 黄璐, 文军, 姜杰. 电动舵机系统建模与仿真研究[J]. 自动化与仪器仪表, 2020(12): 237-239.
HUANG L, WEN J, JIANG J. Research on modeling and simulation of electric servo system[J]. Automation & Instrumentation, 2020(12): 237-239(in Chinese).

[3] KIM S H, TAHK M J. Dynamic stiffness transfer function of an electromechanical actuator using system identification[J]. International Journal of Aeronautical and Space Sciences, 2018, 19(1): 208-216.

[4] LU J, WU Z G, YANG C. Active flutter suppression of nonlinear fin-actuator system via inverse system and immersion and invariance method[J]. Chinese Journal of Aeronautics, 2024, 37(4): 343-362.

[5] RISTANOVIĆ M, LAZIĆ D, INĐIN I. Experimental validation of improved performances of an electromechanical aerofin control system with a PWM controlled DC motor[J]. FME Transactions, 2006, 34(1): 15-20.

[6] ÖZKAN B. Control of an electromechanical control actuation system using a fractional order proportional, integral, and derivative-type controller[J]. IFAC Proceedings Volumes, 2014, 47(3): 4493-4498.

[7] 石忠佼, 朱化杰, 赵良玉, 等. 考虑舵机动力学的旋转弹自适应解耦控制[J]. 航空学报, 2022, 43(3): 333-343.
SHI Z J, ZHU H J, ZHAO L Y, et al. Adaptive decoupling control for a class of spinning rockets considering actuator dynamics[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(3): 333-343(in Chinese).

[8] 黄健伟, 覃泽龙, 徐新, 等. 基于ESO的无人机舵机系统控制策略研究[J]. 兵器装备工程学报, 2023, 44(7): 257-263.
HUANG J W, QIN Z L, XU X, et al. Research on the control strategy of UAV electromechanical actuator system based on ESO[J]. Journal of Ordnance Equipment Engineering, 2023, 44(7): 257-263(in Chinese).

[9] 李朋, 高智刚, 周军, 等. 高性能电动舵机系统高保真建模方法研究[J]. 西北工业大学学报, 2018, 36(1): 7-12.
LI P, GAO Z G, ZHOU J, et al. High fidelity modeling method of high performance electromechanical actuator[J]. Journal of Northwestern Polytechnical University, 2018, 36(1): 7-12(in Chinese).

[10] KAYRAN A, NALCI M O. Aeroservoelastic modeling and analysis of a missile control surface with a nonlinear electromechanical actuator[C]//Proceedings of the AIAA Atmospheric Flight Mechanics Conference. Reston: AIAA, 2014: 2055.

[11] 许行之, 高亚奎, 章卫国. 考虑舵机动力学的舵回路系统颤振特性分析[J]. 计算机仿真, 2014, 31(10): 80-85.
XU X Z, GAO Y K, ZHANG W G. Flutter characteristics analysis for actuator-fin system with dynamic stiffness[J]. Computer Simulation, 2014, 31(10): 80-85(in Chinese).

[12] 张新替, 吴志刚, 杨超. 考虑舵机动力学的舵系统颤振特性分析[J]. 北京航空航天大学学报, 2011, 37(8): 927-932.
ZHANG X T, WU Z G, YANG C. Flutter analysis for rudder considering actuator's dynamics[J]. Journal of Beijing University of Aeronautics and Astronautics, 2011, 37(8): 927-932(in Chinese).

[13] ZHANG X T, WU Z G, YANG C. New flutter-suppression method for a missile fin with an actuator[J]. Journal of Aircraft, 2013, 50(3): 989-994.

[14] LAYTON D, GAINES V. F-22 actuator dynamic stiffness (impedance) testing[C]//Proceedings of the 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston: AIAA, 2007: 1792.

[15] SANTOS BALLESTEROS H M, DAS NEVES CALVO R, FILHO A A. Dynamic stiffness enhancement of a flight control actuator using control techniques[C]//Proceedings of the 2017 IEEE International Conference on Mechatronics. Piscataway: IEEE, 2017: 260-265.

[16] KADIYALA V K, JATOTH R K, POTHALAIAH S. Design and implementation of fractional order PID controller for aerofin control system[C]//Proceedings of the 2009 World Congress on Nature & Biologically Inspired Computing. Piscataway: IEEE Press, 2010: 696-701.

[17] 卢晋, 吴志刚, 杨超. 电动舵机模块化建模及动刚度仿真[J]. 北京航空航天大学学报, 2021, 47(4): 765-778.
LU J, WU Z G, YANG C. Modular modeling and dynamic stiffness simulation of electromechanical actuator[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(4): 765-778(in Chinese).

[18] 张仁嘉, 吴志刚, 杨超. 电动伺服舵系统动力学建模及颤振分析[J]. 北京航空航天大学学报, 2016, 42(7): 1368-1376.
ZHANG R J, WU Z G, YANG C. Dynamic modeling and flutter analysis of a fin-actuator system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(7): 1368-1376(in Chinese).

[19] 张程, 任浩源, 史泰龙, 等. 含非线性连接的折叠舵全时域多学科耦合分析方法及应用[J]. 航空学报, 2023, 44(增刊2): 729461.
ZHANG C, REN H Y, SHI T L, et al. Multidisciplinary full-time coupling methods and applications of folding fin containing non-linear connections[J]. Acta Aeronautica et Astronautica Sinica, 2023,

44(Sup2): 729461.

[20] HE H N, TANG H, YU K P, et al. Nonlinear aeroelastic analysis of the folding fin with freeplay under thermal environment[J]. Chinese Journal of Aeronautics, 2020, 33(9): 2357-2371.

[21] 王强, 马志赛, 张欣, 等. 基于模态综合法的含间隙折叠舵面动态特性分析[J]. 航空学报, 2020, 41(5): 197-205.
WANG Q, MA Z S, ZHANG X, et al. Dynamic characteristic analysis for a folding fin with freeplay nonlinearities based on mode synthesis method[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(5): 197-205(in Chinese).

[22] 白刘月, 吴志刚, 杨超. 含间隙全动舵面的非线性颤振模式及被动抑制方法[J]. 北京航空航天大学学报, 2023, 49(9): 2361-2373.
BAI L Y, WU Z G, YANG C. Nonlinear flutter modes and flutter suppression of an all-movable fin with freeplay[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(9): 2361-2373(in Chinese).

[23] 李友年, 陈星阳. 舵机间隙环节对控制系统的影响分析[J]. 航空兵器, 2012, 19(1): 25-27.
LI Y N, CHEN X Y. Influence analysis of actuator's gap on control system[J]. Aero Weaponry, 2012, 19(1): 25-27(in Chinese).

[24] 胡江涛, 曹云峰. 滚珠丝杠式电动舵机系统非线性特性分析[J]. 计算机测量与控制, 2018, 26(2): 190-194.
HU J T, CAO Y F. Nonlinear factors analysis of EMA with ball screw drive[J]. Computer Measurement & Control, 2018, 26(2): 190-194(in Chinese).

[25] 刘鹏, 李怀兵, 杨超凡, 等. 传动间隙引起的电动舵机自振荡现象研究[J]. 微电机, 2022, 55(7): 89-93.
LIU P, LI H B, YANG C F, et al. Research on self oscillation of electric actuator caused by transmission clearance[J]. Micromotors, 2022, 55(7): 89-93(in Chinese).

[26] 张新华, 黄建, 张兆凯, 等. 大惯量下电动伺服机构非线性特性与控制方法研究[J]. 导航定位与授时, 2017, 4(2): 41-47.
ZHANG X H, HUANG J, ZHANG Z K, et al. Research on nonlinear factor characteristics and control method on large inertia electromechanical actuator servo mechanism[J]. Navigation Positioning and Timing, 2017, 4(2): 41-47(in Chinese).

Multidisciplinary coupled dynamics modeling of electro-mechanical actuator considering nonlinear clearances in multiple stages

YANG Shilin[1], QIAO Kaili[1, 2], WAN Xiaozhong[1], DUAN Mianchao[1, 3], GUO Ning[1, *], XU Chao[1]

(1. School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China;
2. Chinese Flight Test Establishment, Xi'an 710089, China;
3. System Design Institute of Hubei Aerospace Technology Academy, Wuhan 430040, China)

Abstract: Due to uncertainties in the design and assembly process, nonlinear factors such as clearances often exist in multiple sectors of the aerospace electro-mechanical actuator transmission system, significantly affecting the working quality of the electro-mechanical actuator. Therefore, it is essential to accurately establish a nonlinear dynamic model for the electro-mechanical actuator accurately. Moreover, electro-mechanical actuators involve multiple disciplines such as servo control, motor transmission, reduction mechanism, structural elasticity, etc., making the modeling process highly challenging. The research object in this paper is a standard ball screw-type electro-mechanical actuator. Considering the elastic stiffness of actuator transmission components and the nonlinearity of clearances at gear pairs, supporting bearings, and fork pairs, the motor dynamics and servo controllers are finely modeled. The electro-mechanical actuator system is represented by a high-fidelity multidisciplinary dynamic model that can account for nonlinearities in different sector clearances. Furthermore, simulation analysis of the electro-mechanical actuator system performance is conducted. The simulation results are compared with ground vibration test results of the actuator-rudder system. The results indicate that the multidisciplinary coupled dynamic model of the electro-mechanical actuator established in this paper can reflect the influence and degree of influence of the clearances in different parts of the mechanism. The simulation and experimental results are in good agreement, with a maximum error of less than 10%. This research lays the foundation for the digital twin study of complex aerospace electric servo systems.

Keywords: electro-mechanical actuator; nonlinear dynamics; multidisciplinary coupling; clearance; ground vibration test

Received: 2024-03-08; **Accepted**: 2024-03-29; **Published Online**: 2024-04-22 11:45

URL: link.cnki.net/urlid/11.2625.V.20240418.1320.002

Foundation items: National Natural Science Foundation of China (12072268); Beijing Institute of Strength and Environment Engineering Joint Innovation Foundation(BQJJ202306)

* **Corresponding author**. E-mail: guoning@nwpu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0144

飞翼布局飞机颤振主动抑制的多输入/多输出控制律设计

冯宇轩，霍应元*，李俊杰
(航空工业第一飞机设计研究院 强度设计研究所，西安 710089)

摘 要：针对飞翼布局飞机刚体模态与弹性模态耦合严重，体自由度颤振问题突出的特点，提出一种基于鲁棒控制理论的飞翼布局飞机体自由度颤振主动抑制控制律设计方法。该控制律设计方法将受控对象所受干扰及建模过程中存在的不确定参数统一视为“未知扰动”，通过分析其输入/输出关系对其进行状态估计，最后，综合给出反馈控制律。为验证所提控制律的颤振抑制效果，以一飞翼布局无人机标模为分析对象，分别以机身内侧2副翼舵面作为控制输入，飞机刚体俯仰速度和翼尖运动速度为反馈信号来设计控制器。对闭环系统 V-g 曲线、闭环时域状态方程等进行了数值仿真，结果表明：该控制律设计方法能有效提高飞机颤振临界速度。

关 键 词：气动弹性；飞翼布局飞机；颤振主动抑制；体自由度颤振；MIMO输出反馈控制器
中图分类号：V215.3+4
文献标志码：A **文章编号**：1001-5965(2026)05-1720-08

飞翼布局飞机一般采用大展弦比机翼设计，以获得相较于传统布局飞机更好的升阻特性，同时，采用轻质材料以减轻机体结构质量，以达到提高飞机高空长航时的总体设计目标。由于大展弦比柔性机翼的结构质量较轻，且刚度较低，飞翼布局飞机的机翼在受到气动载荷作用后会产生较大的弹性变形，而弹性变形进一步影响气动载荷的分布，这种气动与结构间的相互耦合作用会产生较为复杂的气动弹性问题，其中，较为常见的为机翼的弯曲运动模态与扭转运动模态耦合所导致的弯扭耦合颤振。而飞翼布局飞机的情况有所不同，由于飞翼布局飞机的机身较短，其刚体俯仰运动频率较传统布局飞机有所提高，同时，大展弦比柔性机翼的低阶弹性模态频率往往比较低，因此，会出现机翼的低阶弯曲振动模态与短周期的刚体模态(多为机体俯仰运动刚体模态)相耦合，诱发一种不同于传统弯扭耦合颤振的颤振形态，即体自由度颤振。由于刚体模态参与耦合导致的气动弹性不稳定性，除了要考虑机翼质量、刚度的分布外，还要考虑全机刚体模态及飞机飞行动力学因素的影响。

飞机设计要求在设定的飞行包线内，必须保证不会发生颤振和发散，并保有一定的安全余量。在传统的飞机设计过程中，气动弹性设计最常用且有效的方法是提高飞机的结构刚度，或优化质量、刚度的分布，从而削弱惯性与气动耦合，提高临界颤振速度，保证在设计的飞行包线内不发生颤振现象。但提高结构刚度必然会导致结构质量的增加，这与总体设计想达到的高空长航时的飞机整体性能相互矛盾。随着现代控制技术的发展，为解决上述矛盾提供了新思路：通过在飞机机翼表面布置多个可操纵的控制面，并设计主动控制系统，依据传感器测得的结构运动状态，驱动飞机操纵面发生偏转，主动调节气动力分布，从而削弱惯性与气动耦合，以此达到提高颤振临界速度、抑制颤振的目的[1]。

20世纪90年代，美国开展的“主动气动弹性机

收稿日期：2024-03-14；录用日期：2024-05-09；网络出版时间：2024-05-20 16:15
网络出版地址：link.cnki.net/urlid/11.2625.V.20240517.1909.002
*通信作者. E-mail：huoyy@sina.com

引用格式：冯宇轩，霍应元，李俊杰. 飞翼布局飞机颤振主动抑制的多输入/多输出控制律设计[J]. 北京航空航天大学学报，2026，52(5)：1720-1727.
FENG Y X，HUO Y Y，LI J J. Design of multiple-input/multiple-output control law for active flutter suppression of flying-wing aircraft[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1720-1727 (in Chinese).

翼”计划首次验证了颤振主动抑制(active flutter suppression, AFS)理念[2]。近年来, 关于颤振主动抑制技术的研究也取得了重要进展[3-14]。于明礼等采用超声电机为作动器, 研究分析了带控制面的二元机翼模型的颤振主动抑制问题, 设计了H_∞控制器和μ控制器, 并进行了试验验证[15-16]。Huang 等进行了小展弦比机翼模型的颤振主动抑制设计, 利用极点配置优化设计的线性反馈控制抑制颤振失稳, 并在风洞试验中开展了机翼模型闭环颤振模态的在线辨识技术的研究[17]。李星设计开展了多操纵面大展弦比机翼的颤振主动抑制技术研究, 并进行了风洞试验[18]。除上述对于二元翼段和机翼模型开展的颤振主动抑制研究外, 还有少部分针对全机模型的颤振主动抑制技术研究, 由于全机的气动/结构动力学建模计算不完全准确, 模型不确定性及控制律降阶需重点考虑[19]。杨超等以仿 F/A-18A 的全机模型为研究对象, 采用线性二次型高斯方法, 研究了常规布局飞机颤振主动抑制技术[20]。沐旭升等设计了针对体自由度颤振主动抑制的自抗扰控制律[21]。洛克希德·马丁公司设计并制作了验证主动抑制体自由度颤振的验证机[22], 并成功提高了该机体自由度颤振速度。相较于常规布局飞机, 飞翼布局飞机需重点考虑全机刚体模态对颤振主动抑制的影响。

本文的研究目标是对颤振主动抑制存在的建模及控制率设计问题进行初步探索。主要研究内容是以飞翼布局无人机标模为研究对象, 应用H_∞鲁棒控制理论, 设计颤振主动抑制控制器, 并通过开/闭环颤振分析结果对比, 验证颤振主动抑制控制器的有效性。

1　气动伺服弹性模型建模

1.1　飞翼布局飞机模型

本文研究对象为大展弦比飞翼布局无人机模型(见图 1), 其主要几何尺寸如表 1 所示, 在此基础上, 建立结构梁架模型及气动模型, 利用 NASTRAN 的自由振动分析模块可得到全机模型的振动特性及模态信息。模型的前 6 阶模态为刚体运动模态, 部分弹性模态的固有频率如表 2 所示, 振型如图 2 所示。

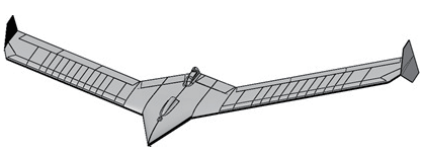


图 1　飞翼布局飞行器模型

Fig. 1　Flying wing configuration aircraft model

表 1　模型主要几何尺寸

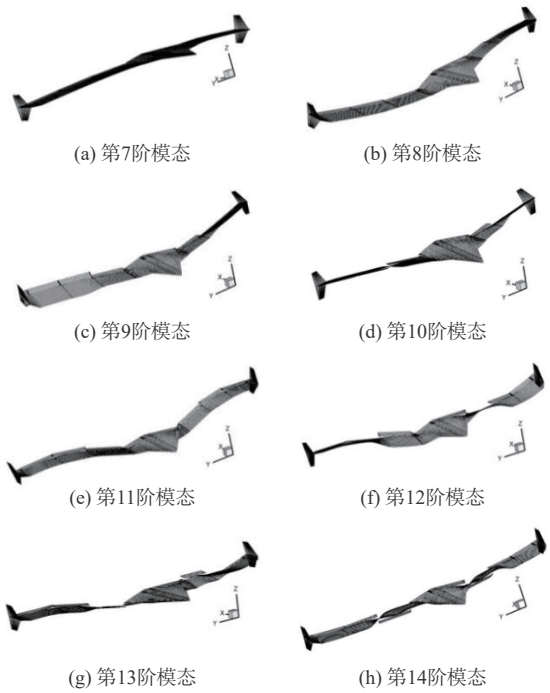
Table 1　Principal geometric dimensions of the model

展长b/m	机身长度L/m	展弦比	后掠角λ/(°)	翼根弦长c_{root}/m
3.05	0.85	10.2	22	0.299

表 2　模态参数

Table 2　Modal parameters

模态阶数	频率/Hz	模态振型
7	5.377	一阶对称弯曲
8	9.471	一阶反对称弯曲
9	12.783	一阶反对称扭转
10	14.411	一阶对称扭转
11	19.345	二阶对称弯曲
12	27.260	二阶反对称扭转
13	30.660	二阶反对称弯曲
14	31.654	二阶对称扭转



(a) 第7阶模态　(b) 第8阶模态

(c) 第9阶模态　(d) 第10阶模态

(e) 第11阶模态　(f) 第12阶模态

(g) 第13阶模态　(h) 第14阶模态

图 2　前 8 阶弹性模态振型

Fig. 2　First eight elastic modal shapes

1.2　全机气动伺服弹性模型建模

气动伺服弹性系统建模是进行闭环颤振分析及颤振主动抑制控制律设计的第 1 步, 需考虑对整个气动伺服弹性闭环系统中的传感器、作动器及操纵面等各个环节进行动力学建模, 最终合并建立一套完整的适用于颤振主动抑制控制律设计的气动伺服弹性(aeroservoelastic, ASE)分析模型。

应用传统动力学分析方法, 从气动弹性动力学系统的功能原理关系分析, 利用拉格朗日方程, 可得到气动弹性系统的运动微分方程, 对于真实飞机

结构，由于其状态量阶数较高，其结构参数无法通过简单的解析分析获得，因此，需要进行结构离散化，将整机模型简化为状态量较少的板杆模型或梁架模型，采用有限元方法得到结构的模态信息。将有限元分析得到的离散形式的结构动能、势能及气动力代入拉格朗日方程中，可得到气动弹性动力学方程为[23]

$$\begin{bmatrix}\boldsymbol{M}_{\xi\xi} & \boldsymbol{M}_{\xi\delta}\\ \boldsymbol{M}_{\delta\xi} & \boldsymbol{M}_{\delta\delta}\end{bmatrix}\begin{bmatrix}\ddot{\boldsymbol{\xi}}\\ \ddot{\boldsymbol{\delta}}\end{bmatrix}+\begin{bmatrix}\boldsymbol{D}_{\xi\xi} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\dot{\boldsymbol{\xi}}\\ \dot{\boldsymbol{\delta}}\end{bmatrix}+\begin{bmatrix}\boldsymbol{K}_{\xi\xi} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\delta}\end{bmatrix}=q_\infty\begin{bmatrix}\boldsymbol{Q}_{\xi\xi} & \boldsymbol{Q}_{\xi\delta}\\ \boldsymbol{Q}_{\delta\xi} & \boldsymbol{Q}_{\delta\delta}\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\delta}\end{bmatrix} \tag{1}$$

式中：q_∞为来流动压；$\boldsymbol{\xi}$、$\dot{\boldsymbol{\xi}}$、$\ddot{\boldsymbol{\xi}}$分别为各阶模态的模态位移、模态速度、模态加速度；$\boldsymbol{\delta}$、$\dot{\boldsymbol{\delta}}$、$\ddot{\boldsymbol{\delta}}$分别为各个操纵舵面的偏转角、偏转角速度、偏转角加速度；$\boldsymbol{M}_{\xi\xi}$、$\boldsymbol{D}_{\xi\xi}$、$\boldsymbol{K}_{\xi\xi}$分别为广义质量、广义阻尼、广义刚度矩阵；$\boldsymbol{M}_{\xi\delta}$、$\boldsymbol{M}_{\delta\xi}$、$\boldsymbol{M}_{\delta\delta}$为惯性耦合项；$\boldsymbol{Q}_{\xi\xi}$、$\boldsymbol{Q}_{\xi\delta}$、$\boldsymbol{Q}_{\delta\xi}$、$\boldsymbol{Q}_{\delta\delta}$为对应的广义气动力矩阵求和项。

采用最小状态法对非定常广义气动力矩阵进行时域拟合，分离状态变量，可将气动伺服弹性动力学方程转化为如下状态空间模型：

$$\begin{cases}\dot{\boldsymbol{x}}_{\mathrm{ae}}=\boldsymbol{A}_{\mathrm{ae}}\boldsymbol{x}_{\mathrm{ae}}+\boldsymbol{B}_{\mathrm{ae}}\boldsymbol{u}_{\mathrm{ae}}\\ \boldsymbol{y}_{\mathrm{ae}}=\boldsymbol{C}_{\mathrm{ae}}\boldsymbol{x}_{\mathrm{ae}}+\boldsymbol{D}_{\mathrm{ae}}\boldsymbol{u}_{\mathrm{ae}}\end{cases} \tag{2}$$

式中：$\dot{\boldsymbol{x}}_{\mathrm{ae}}$为状态变量；$\boldsymbol{A}_{\mathrm{ae}}$为状态矩阵；$\boldsymbol{B}_{\mathrm{ae}}$为输入矩阵；$\boldsymbol{C}_{\mathrm{ae}}$为输出矩阵；$\boldsymbol{D}_{\mathrm{ae}}$为前馈矩阵。

其中

$$\begin{cases}\boldsymbol{x}_{\mathrm{ae}}=\begin{bmatrix}\boldsymbol{\xi}\\ \dot{\boldsymbol{\xi}}\\ \boldsymbol{x}_{\mathrm{a}}\end{bmatrix}\\ \boldsymbol{u}_{\mathrm{ae}}=\begin{bmatrix}\boldsymbol{\delta}\\ \dot{\boldsymbol{\delta}}\\ \ddot{\boldsymbol{\delta}}\end{bmatrix}\end{cases} \tag{3}$$

式中：$\boldsymbol{x}_{\mathrm{a}}$为气动力滞后根矢量；$\boldsymbol{y}_{\mathrm{ae}}$为各传感器的输出量。

飞机的作动器环节是控制操纵面偏转的输入环节，根据控制系统所输出的操纵指令，作动器将驱动操纵面偏转运动，使操纵面达到预期的偏转。由于作动器本身存在质量及摩擦，存在时滞，在建模过程中，需要对作动器进行动力学分析。例如第i个操纵面的舵机动力学模型可以简化为如下传递函数形式：

$$\frac{\boldsymbol{\delta}_{ci}(s)}{\boldsymbol{u}_{\mathrm{aci}}(s)}=\frac{a_3}{s^3+a_1s^2+a_2s+a_3} \tag{4}$$

式中：$\boldsymbol{u}_{\mathrm{aci}}$为飞行控制系统反馈输入舵机的操纵面偏转指令；$\boldsymbol{\delta}_{ci}$为实际输出的操纵面偏转角；$s$为复变量，表示拉普拉斯变换域中的频率；$a_1$、$a_2$和$a_3$为特征多项式系数。舵机环节一般为三阶环节，可视为作动器驱动舵面偏转的二阶振荡环节和代表舵机带宽限制的一阶延迟环节的串联模型。

对于式(4)，需要引入3个状态变量$\boldsymbol{\delta}$、$\dot{\boldsymbol{\delta}}$、$\ddot{\boldsymbol{\delta}}$来建立对应的状态空间方程表达式：

$$\dot{\boldsymbol{x}}_{\mathrm{aci}}=\boldsymbol{A}_{\mathrm{aci}}\boldsymbol{x}_{\mathrm{aci}}+\boldsymbol{B}_{\mathrm{aci}}\boldsymbol{u}_{\mathrm{aci}} \tag{5}$$

其中，

$$\begin{cases}\boldsymbol{x}_{\mathrm{aci}}=\begin{bmatrix}\boldsymbol{\delta}\\ \dot{\boldsymbol{\delta}}\\ \ddot{\boldsymbol{\delta}}\end{bmatrix}\\ \boldsymbol{A}_{\mathrm{aci}}=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ -a_3 & -a_2 & -a_1\end{bmatrix}\\ \boldsymbol{B}_{\mathrm{aci}}=\begin{bmatrix}0\\ 0\\ a_3\end{bmatrix}\end{cases} \tag{6}$$

各个舵机的输入量$\boldsymbol{u}_{\mathrm{ac}}$与飞行控制系统的输出量$\boldsymbol{y}_{\mathrm{c}}$有如下增益关系：

$$\boldsymbol{u}_{\mathrm{ac}}=\boldsymbol{G}\boldsymbol{y}_{\mathrm{c}} \tag{7}$$

式中：$\boldsymbol{G}$为增益矩阵。

对于多个操纵面的情况，将各个$\boldsymbol{x}_{\mathrm{aci}}$按顺序组装起来，即可得到相应的操纵面状态向量$\boldsymbol{x}_{\mathrm{ac}}$，且其对应了式(7)中的$\boldsymbol{u}_{\mathrm{ac}}$，即

$$\boldsymbol{u}_{\mathrm{ac}}=\boldsymbol{x}_{\mathrm{ac}} \tag{8}$$

由此，可得到包含传感器、作动器等环节的气动伺服弹性模型的状态空间方程，形式为

$$\begin{cases}\dot{\boldsymbol{x}}_{\mathrm{p}}=\boldsymbol{A}\boldsymbol{x}_{\mathrm{p}}+\boldsymbol{B}\boldsymbol{u}_{\mathrm{p}}\\ \boldsymbol{y}_{\mathrm{p}}=\boldsymbol{C}\boldsymbol{x}_{\mathrm{p}}\end{cases} \tag{9}$$

其中，

$$\begin{cases}\boldsymbol{x}_{\mathrm{p}}=\begin{bmatrix}\boldsymbol{x}_{\mathrm{ae}}\\ \boldsymbol{x}_{\mathrm{ac}}\end{bmatrix}\\ \boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_{\mathrm{ae}} & \boldsymbol{B}_{\mathrm{ae}}\\ \boldsymbol{0} & \boldsymbol{A}_{\mathrm{ac}}\end{bmatrix}\\ \boldsymbol{B}=\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{B}_{\mathrm{ac}}\end{bmatrix}\\ \boldsymbol{u}_{\mathrm{p}}=\boldsymbol{u}_{\mathrm{ac}}\\ \boldsymbol{x}_{\mathrm{p}}=\boldsymbol{y}_{\mathrm{ae}}\\ \boldsymbol{C}=\begin{bmatrix}\boldsymbol{C}_{\mathrm{ae}} & \boldsymbol{D}_{\mathrm{ae}}\end{bmatrix}\end{cases} \tag{10}$$

式中的状态变量、输入变量和输出变量定义为

$$\begin{cases}\boldsymbol{x}_{\mathrm{p}}=\begin{bmatrix}\boldsymbol{\xi}\\ \dot{\boldsymbol{\xi}}\\ \boldsymbol{x}_{\mathrm{a}}\\ \boldsymbol{\delta}\\ \dot{\boldsymbol{\delta}}\\ \ddot{\boldsymbol{\delta}}\end{bmatrix}\\ \boldsymbol{u}_{\mathrm{p}}=\begin{bmatrix}\boldsymbol{u}_{\mathrm{ac1}}\\ \boldsymbol{u}_{\mathrm{ac2}}\end{bmatrix}\end{cases} \tag{11}$$

式中: $\boldsymbol{u}_{ac1}$、$\boldsymbol{u}_{ac2}$分别为操纵的内侧 2 个副翼作动器的偏角。本文所建立的气动伺服弹性模型的状态变量为 24 阶, 包括 6 阶结构模态的模态位移、模态速度, 6 阶气动力滞后根矢量和 2 个作动器的状态变量。

1.3　模型开环颤振分析

得到气动伺服弹性状态方程后, 可求解得到该分析模型的临界颤振速度。图 3 为开环气动伺服弹性系统模态频率及模态阻尼随飞行速度的变化曲线。由模态阻尼随飞行速度变化 *V*-*g* 曲线变化可以看出, 当来流速度接近 24.45 m/s 时, 一阶对称弯曲模态穿越, 发生颤振失稳, 颤振频率为 4.18 Hz。由模态频率变化的趋势可以看出, 随着来流风速的增加, 刚体俯仰运动的频率逐渐增高, 而一阶对称弯曲模态的频率逐渐降低, 在接近颤振速度时, 两者频率接近产生耦合。

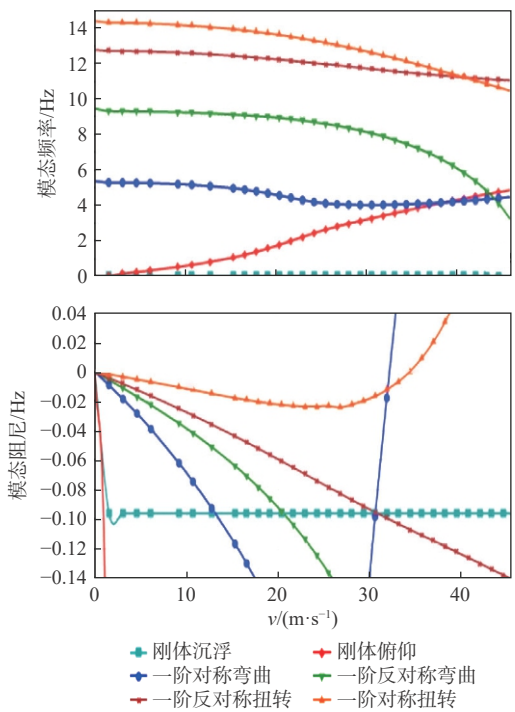


图 3　开环颤振分析结果

Fig. 3　Open-loop flutter analysis results

图 4 显示了主要低频模态对颤振的贡献量。根据贡献量大小, 依次为一阶对称扭转模态、刚体俯仰模态及刚体沉浮模态。显然, 一阶对称弯曲模态起主要作用。通过对颤振速度下各阶模态所对应的特征向量(表 3)分析可得, 在颤振速度下, 刚体俯仰运动对气动力变化的贡献量最大, 因为俯仰运动引起的飞机攻角变化严重影响飞机整体气动力分布。综上所述, 诱发第 1 支颤振失稳的原因是该飞机的一阶对称弯曲模态与刚体俯仰模态耦合, 进而引发体自由度颤振。

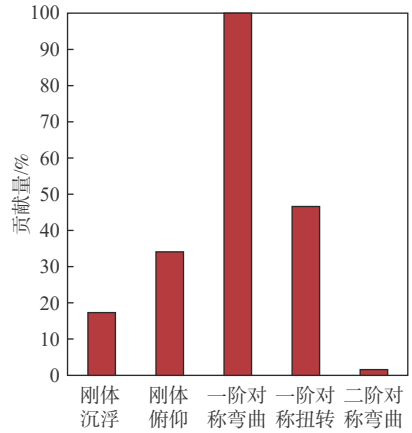


图 4　各阶模态颤振贡献量

Fig. 4　Contribution of each mode to modal flutter

表 3　各阶模态对应特征向量

Table 3　Feature vectors corresponding to each mode

模态阶数	频率/Hz	模态振型	特征向量
3	0	刚体沉浮	0.054−0.048i
4	0	刚体俯仰	1.00
7	5.377	一阶对称弯曲	−0.58+0.30i
10	14.411	一阶对称扭转	0.036−0.013i
11	19.345	二阶对称弯曲	0.021−0.004i

2　多输入/多输出控制律设计

颤振主动抑制设计是通过设计相应的反馈控制系统, 实现对应的操纵面偏转, 改变机翼气动力分布, 提高颤振主模态的气动阻尼, 进而使颤振主模态在更高的来流速度范围内实现收敛, 达到在飞行包线内抑制颤振发生、提高颤振临界速度的目的; 同时, 为保证在各种飞行状态下颤振主动抑制系统的可靠性, 需降低高频结构模态与低频飞行控制系统和颤振主动抑制系统的相互影响。

2.1　H_∞控制基本理论

本文采用H_∞鲁棒控制来设计飞行控制器。H_∞优化的目的在于使某些闭环频响函数的峰值最小化, 例如灵敏度函数。灵敏度函数 S 描述的是控制系统输出对干扰信号的灵敏度, 在理想情况下, 灵敏度函数值 S=0。灵敏度最小化问题指, 对于属于有限能量信号集的干扰信号, 设计一个控制器, 使闭环系统稳定, 并使系统外部干扰对系统的控制输出影响最小, 本质上是寻找一个控制器 K, 该控制器在使系统稳定的同时, 使灵敏度函数的峰值最小。该峰值定义为

$$\|S\|_\infty = \max_{\omega \in R} |S(\mathrm{j}\omega)| \tag{12}$$

式中: j 为虚数单位; ω为角频率。

如果灵敏度函数峰值不存在，可以用灵敏度函数的上界代替式(12)的最大值，即

$$\|S\|_\infty = \sup_{\omega\in R}\left|S\left(\mathrm{j}\omega\right)\right| \tag{13}$$

如果能保证灵敏度函数 S 的峰值较小，则在所有频率上，S 的幅值都较小，即干扰输入对应的系统输出衰减，因此，该方法相当于极小化干扰对输出的影响。即使在最差的情况下，也希望可以找到一个控制器 K，能抑制存在的干扰。

鲁棒控制律设计的第 1 步是构建广义受控对象 P，该广义受控对象是由 ASE 模型、根据频率选择的输入输出加权参数和对闭环系统鲁棒性有影响的外部干扰输入连接构成。以 $\boldsymbol{v}$ 为输入、$\boldsymbol{z}$ 为输出的线性时不变动力学系统 $G(s)$ 的 H_∞ 范数定义为

$$\|G(s)\|_\infty = \sup_{\omega\in R}\bar{\sigma}\left[G\left(\mathrm{j}\omega\right)\right] = \sup_{\|\boldsymbol{v}\|_2<\infty}\frac{\|\boldsymbol{z}\|_2}{\|\boldsymbol{v}\|_2} \tag{14}$$

式中：$\bar{\sigma}$ 为最大奇异值。

其中，

$$\|\boldsymbol{v}\|_2 = \sqrt{\int_{-\infty}^{\infty}\boldsymbol{v}^2(t)\,\mathrm{d}t} \tag{15}$$

该 H_∞ 范数描述了传递函数 $G(s)$ 的峰值，也表示 $G(s)$ 在所有频率范围内给定输入信号的最大峰值，可以用来描述广义受控对象的动态性能。于是在广义受控对象 P 构造完成后，可引入一个性能指标 $\gamma>0$，该指标既可以描述广义受控对象 P 的 H_∞ 范数上界极限，也可以描述广义受控对象的闭环系统性能，对于灵敏度函数极小化问题，通常需要满足

$$\|WS\|_\infty = \sup_{\omega\in R}\left|W\left(\mathrm{j}\omega\right)S\left(\mathrm{j}\omega\right)\right| < \gamma \tag{16}$$

式中：W 为频率加权函数。

如果加权后灵敏度函数 S 的峰值 $\|WS\|_\infty$ 很小，则说明在参考频率内 S 的幅值很小，则所有干扰信号在此系统中是衰减的，接近理想目标效果，进而可通过求解 2 个 Riccati 方程得到鲁棒 H_∞ 控制器。

2.2 H_∞ 控制器设计

为保证 H_∞ 控制器的控制效果，并能对存在一定不确定性的模型有较好的鲁棒性，一种比较适合的广义受控对象模型如图 5 所示，其能在完成稳定性设计的同时，还可以通过设计加权参数，使广义受控对象达到将闭环系统的灵敏度函数的 H_∞ 范数最小化的要求，图中，$K(s)$ 即为需设计的控制器环节。

完整的受控对象 $G(s)$ 分为 2 部分，记为

$$G(s) = \begin{bmatrix} G_z(s) \\ G_y(s) \end{bmatrix} \tag{17}$$

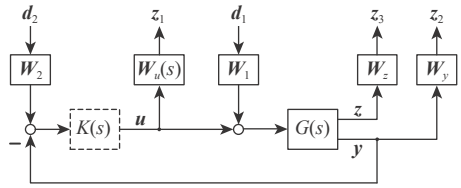


图 5 控制系统框图

Fig. 5 Control system block diagram

第 1 部分 $G_z(s)$ 用来计算广义受控对象的性能，其输出的不可观测量 $\boldsymbol{z}$ 为受控对象的指标量，此处选为关键模态的广义模态速度 $\dot{\boldsymbol{\xi}}$。第 2 部分 $G_y(s)$ 为之前建立的气动伺服弹性模型，其中包括气动弹性模型和作动器模型等，可观测的输出量 $\boldsymbol{y}$ 为 ASE 模型的输出。广义受控对象的输入，即输入端干扰和输出端干扰分别定义为 $\boldsymbol{d}_1$ 和 $\boldsymbol{d}_2$，其权重系数矩阵分别为 $\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$；广义受控对象的输出分别为控制器输出信号 $\boldsymbol{u}$、气动伺服弹性模型的可观测输出 $\boldsymbol{y}$ 和不可观测输出 $\boldsymbol{z}$，经过权重系数矩阵 $\boldsymbol{W}_u$、$\boldsymbol{W}_y$ 和 $\boldsymbol{W}_z$ 处理后的输出即输出量 $\boldsymbol{z}_1$、$\boldsymbol{z}_2$ 和 $\boldsymbol{z}_3$。由上述定义的输入输出信号即可得到广义受控对象的输入与输出间的传递函数方程：

$$\begin{bmatrix}\boldsymbol{z}_1\\ \boldsymbol{z}_2\\ \boldsymbol{z}_3\end{bmatrix} = \begin{bmatrix}\boldsymbol{W}_u & & \\ & \boldsymbol{W}_y & \\ & & \boldsymbol{W}_z\end{bmatrix}\begin{bmatrix}-T_\mathrm{i} & KS_\mathrm{i}\\ G_yS_\mathrm{i} & T_\mathrm{o}\\ G_zS_\mathrm{i} & G_zKS_\mathrm{i}\end{bmatrix}\cdot \begin{bmatrix}\boldsymbol{W}_1 & \\ & \boldsymbol{W}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{d}_1\\ \boldsymbol{d}_2\end{bmatrix} \tag{18}$$

其中，

$$\begin{cases} T_\mathrm{i} = KG_y\left(1+KG_y\right)^{-1}\\ T_\mathrm{o} = G_yK\left(1+G_yK\right)^{-1}\\ S_\mathrm{i} = \left(1+KG_y\right)^{-1}\end{cases} \tag{19}$$

式中：传递函数 T_o、T_i 和 S_i 分别定义为输出补偿灵敏度、输入补偿灵敏度和输入灵敏度；传递函数 KS_i 和 G_yS_i 分别为控制灵敏度和干扰灵敏度；G_zS_i 和 G_zKS_i 分别为输入端干扰 $\boldsymbol{d}_1$ 和输出端干扰 $\boldsymbol{d}_2$ 与关键模态的广义模态速度 $\dot{\boldsymbol{\xi}}$ 的传递函数，减小这 2 个传递函数的幅值，就可以减弱颤振关键模态对于干扰信号的响应。

输出的权重系数 $\boldsymbol{W}_z$ 可以调整闭环颤振裕度，较大的权重量会产生更高的阻尼增益，以此降低广义受控对象的频响峰值；输入的权重控制系数 $\boldsymbol{W}_u$ 可以使控制器在特定的频率范围内更有效地工作。为降低颤振主动抑制控制系统对飞行动力学操纵(小于 1 Hz)和高频结构模态(大于 30 Hz)的影响，选用 $\boldsymbol{W}_u$ 为带阻滤波器。

全机模型的传感器及操纵面配置如图 6 所示，该全机模型共有 4 组舵面，机翼上有 3 组操纵面，机身上有 1 组操纵面。通过 1.3 节中对开环颤

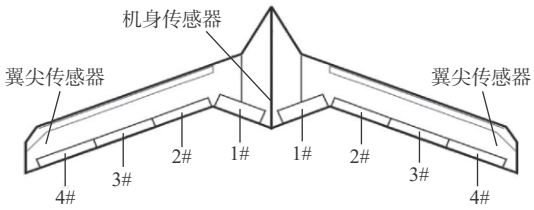


图 6　传感器位置示意图

Fig. 6　Schematic diagram of sensor positioning

振机理的分析, 可以得到, 该模型的颤振主动抑制的关键模态为第一阶弯曲模态和刚体俯仰模态。因此, 选取机翼翼尖和机身重心附近的速度传感器作为模型的输出反馈信号 $\boldsymbol{y}$, 分别作为结构弯曲及俯仰运动的测量值; 选取气动弹性系统模型的刚体俯仰模态、第一阶对称弯曲模态和第一阶对称扭转模态的模态位移计算解作为不可观测输出 $\boldsymbol{z}$; 选取飞机机翼操纵面为控制输入, 以有效抑制颤振主模态——第一阶弯曲模态。同时, 为保证颤振主动抑制控制器对全机飞行操稳控制的影响较小, 本文在分析过程中, 优先选择对于刚体操纵工作占比较小的机翼内侧 2 组操纵面(2#和 3#)作为模型的控制输入 $\boldsymbol{u}$, 且因开环颤振分析得到的颤振关键模态均为对称模态, 故机翼两侧对称操纵面采取统一通道, 2#和 3#操纵面各自采取单独通道。

选择 26.5 m/s 来流速度下气动伺服弹性模型的状态空间方程作为初始设计状态进行控制系统设计, 最终确定的权重系数分别为

$$\begin{cases} \boldsymbol{W}_z = 0.01\mathrm{diag}(10,1,1) \\ \boldsymbol{W}_u = \dfrac{100s^2+7\,506s+137\,500}{s^2+12\,700s+1\,375} \\ \boldsymbol{W}_1 = 0.005 \\ \boldsymbol{W}_2 = 0.01\mathrm{diag}(1,1,2,2) \\ \boldsymbol{W}_y = 0.001\mathrm{diag}(1,1,70,100) \end{cases} \tag{20}$$

将上述权重系数代入求解, 即可得到满足要求的控制器 $K(s)$环节。在实践中, 由于 H_∞控制理论计算得到的控制器阶数与原受控对象模型阶数相当, 均为高阶模型, 难以在实际场景中完成设计, 因此, 需对反馈控制器进行降阶处理, 本文采用平衡截断法将反馈控制器降阶为 10 阶。

3　数值仿真与验证

对带反馈控制器的闭环气动伺服弹性模型进行闭环颤振计算, 结果如图 7 所示, 为闭环气动伺服弹性系统模态频率及模态阻尼随飞行速度的变化曲线。由 *V-g* 曲线变化可以看出, 当来流速度接近 32.13 m/s 时, 一阶对称弯曲模态穿越, 颤振频率

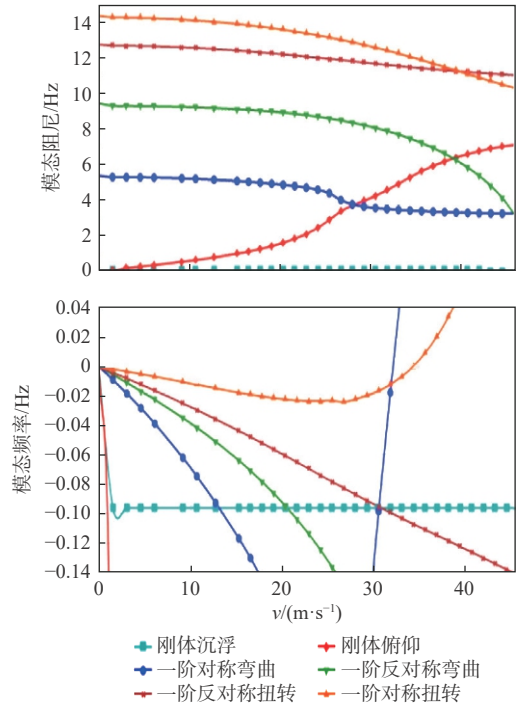


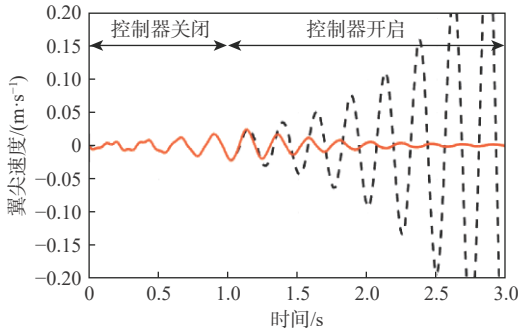
图 7　闭环颤振分析结果

Fig. 7　Closed-loop flutter analysis results

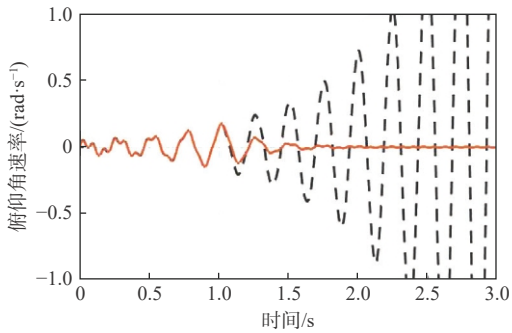
为 3.48 Hz。对比开闭环计算结果可知, 该反馈控制器可将颤振速度由 24.45 m/s 提升到约 32.13 m/s, 提升了约 31.3%。由频率变化及阻尼变化曲线分析可知, 诱发颤振失稳的原因是结构一阶对称弯曲模态与刚体俯仰模态耦合, 颤振机理并未发生改变。同时对比图 3 与图 7 可以看出, 在系统闭环后, 除颤振主模态外, 各阶模态的模态频率和阻尼均未发生较大变化, 说明该控制系统对非关键模态的影响较小, 不会改变颤振机理。

对开/闭环气动伺服弹性模型分别进行时域仿真分析, 以 26.5 m/s 的速度状态为时域分析的基准状态。图 8 中虚线为开环系统的翼尖速度和机身俯仰角速度的时域响应曲线, 在初始时刻(0 s)对系统加入微小扰动, 可以看到, 在扰动作用下, 弯曲振动和俯仰运动均快速发散, 开环系统在该速度下不稳定。图中红线为闭环系统的时域响应, 为使分析结果更直观, 闭环系统在时间为 1 s 时开启反馈控制器, 可以看到, 在控制器开启前, 弯曲振动和俯仰运动均快速发散, 而在控制器闭合后, 操纵面发生偏转, 飞机的翼尖速度和俯仰角速度很快趋于收敛, 同时可以看到, 舵偏角值均在有效范围内。

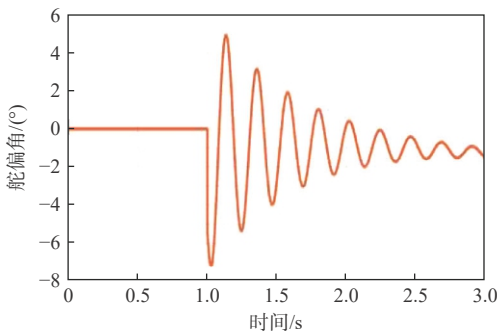
综上所述, 该颤振主动抑制控制器能够准确识别颤振的主模态, 并有效地针对运动状态施加控制, 达到主动抑制颤振、提高临界颤振速度的目的。



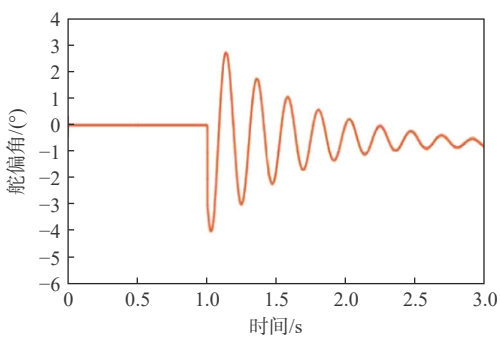
(a) 翼尖速度时域响应



(b) 俯仰角速度时域响应



(c) 2#操纵面时域响应



(d) 3#操纵面舵面时域响应

图 8 时域响应分析结果

Fig. 8 Time-domain response analysis results

4 结 论

本文针对飞翼布局飞机的体自由度颤振问题，选取一飞翼布局无人机标模为分析对象，设计了多输入/多输出的输出反馈控制器，并对系统进行了开/闭环颤振计算及时域仿真分析，对比分析了控制器的控制效果，验证了设计方法的有效性和可行性，主要结论如下：

1）建立了考虑外界扰动的广义气动弹性受控模型，可用于颤振主动抑制的H_∞控制器设计。

2）飞翼布局飞机易发生体自由度颤振，相较于传统弯扭耦合颤振，还需将机体纵向运动量作为输出反馈量，可有效识别颤振的主模态，保证颤振主动抑制控制器的有效性。

3）由 *V-g* 曲线及时域仿真结果可以分析得出，本文提出的颤振主动抑制多输入/多输出控制器具有较好的颤振抑制效果，可将临界颤振速度提升约31.3%。

此外，计划后期建立综合考虑飞控系统的ASE分析模型，进一步优化颤振主动抑制控制律的设计技术。

参考文献（References）

[1] 黄锐, 胡海岩. 飞行器非线性气动伺服弹性力学[J]. 力学进展, 2021, 51(3): 428-466.
HUANG R, HU H Y. Nonlinear aeroservoelasticity of aircraft[J]. Advances in Mechanics, 2021, 51(3): 428-466(in Chinese).

[2] PENDLETON E W, BESSETTE D, FIELD P B, et al. Active aeroelastic wing flight research program: Technical program and model analytical development[J]. Journal of Aircraft, 2000, 37(4): 554-561.

[3] ZHANG R, SINGH S N. Adaptive output feedback control of an aeroelastic system with unstructured uncertainties[J]. Journal of Guidance, Control, and Dynamics, 2001, 24(3): 502-509.

[4] BEHAL A, RAO V M, MARZOCCA P, et al. Adaptive control for a nonlinear wing section with multiple flaps[J]. Journal of Guidance, Control, and Dynamics, 2006, 29(3): 744-749.

[5] ZENG J, KUKREJA S L, MOULIN B. Experimental model-based aeroelastic control for flutter suppression and gust-load alleviation [J]. Journal of Guidance, Control, and Dynamics, 2012, 35(5): 1377-1390.

[6] LIVNE E. Aircraft active flutter suppression: state of the art and technology maturation needs[J]. Journal of Aircraft, 2017, 55(1): 410-452.

[7] LIBRESCU L, NA S, QIN Z M, et al. Active aeroelastic control of aircraft composite wings impacted by explosive blasts[J]. Journal of Sound and Vibration, 2008, 318(1-2): 74-92.

[8] CAVAGNA L, RICCI S, SCOTTI A. Active aeroelastic control over a four control surface wing model[J]. Aerospace Science and Technology, 2009, 13(7): 374-382.

[9] 杨发友, 顾仲权. 飞机机翼颤振自适应抑制研究[J]. 南京航空航天大学学报, 2004, 36(6): 708-712.
YANG F Y, GU Z Q. Adaptive flutter suppression for aircraft wing[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2004, 36(6): 708-712(in Chinese).

[10] 宋晨, 吴志刚, 杨超. 基于PID控制器的颤振主动抑制控制律设计[C]//第十一届全国空气弹性学术交流会论文集. 昆明: 中国空气动力学会, 2009: 230-235.
SONG C, WU Z G, YANG C. Design of active flutter suppression control law based on PID controller[C]//Proceedings of the 11th National Conference on Aeroelasticity. Kunming: CARS, 2009: 230-235.

[11] 张子健, 徐敏, 曾晓彬, 等. LQG/LTR鲁棒控制器在机翼主动颤振抑制中的应用[J]. 西北工业大学学报, 2009, 27(2): 190-194.
ZHANG Z J, XU M, ZENG X B, et al. Applying LQG/LTR theory

to designing robust controller for active flutter suppression[J]. Journal of Northwestern Polytechnical University, 2009, 27(2): 190-194(in Chinese).

[12] 张子健, 徐敏, 陈士橹. 机翼颤振的混合灵敏度 H_∞鲁棒控制器设计[J]. 计算力学学报, 2010, 27(4): 661-666.
ZHANG Z J, XU M, CHEN S L. Mixed sensitivity H_∞ controller designing for active flutter suppression[J]. Chinese Journal of Computational Mechanics, 2010, 27(4): 661-666(in Chinese).

[13] 刘祥, 孙秦. 一种弹性机翼的颤振主动抑制与阵风减缓方法[J]. 西北工业大学学报, 2015, 33(5): 804-810.
LIU X, SUN Q. A robust active flutter suppression and gust alleviation method for flexible wing[J]. Journal of Northwestern Polytechnical University, 2015, 33(5): 804-810(in Chinese).

[14] 郑晓珂, 唐炜, 王立博, 等. 颤振主动抑制的 LPV 控制设计[J]. 振动工程学报, 2018, 31(3): 411-416.
ZHENG X K, TANG W, WANG L B, et al. LPV control design for active flutter suppression[J]. Journal of Vibration Engineering, 2018, 31(3): 411-416(in Chinese).

[15] 于明礼, 文浩, 胡海岩. 二维翼段颤振的 H_∞控制[J]. 振动工程学报, 2006, 19(3): 326-330.
YU M L, WEN H, HU H Y. Active flutter suppression of a two-dimensional airfoil using H_∞ synthesis[J]. Journal of Vibration Engineering, 2006, 19(3): 326-330(in Chinese).

[16] 于明礼, 文浩, 胡海岩, 等. 二维翼段颤振的 μ 控制[J]. 航空学报, 2007, 28(2): 340-343.
YU M L, WEN H, HU H Y, et al. Active flutter suppression of a two-dimensional airfoil section using μ synthesis[J]. Acta Aeronautica et Astronautica Sinica, 2007, 28(2): 340-343(in Chinese).

[17] HUANG R, ZHAO Y H, HU H Y. Wind-tunnel tests for active flutter control and closed-loop flutter identification[J]. AIAA Journal, 2016, 54(7): 2089-2099.

[18] 李星. MIMO 几何非线性机翼气动伺服弹性分析与试验[D]. 大连: 大连理工大学, 2020.
LI X. Aeroservoealstic analysis and experiment of MIMO wings with geometrically nonlinear wings[D]. Dalian: Dalian University of Technology, 2020(in Chinese).

[19] HASHEMI K E, PAK C G, AKELLA M R. Akella, Delta adaptive flexible motion control for the X-56A aircraft[C]//Proceedings of the AIAA Atmospheric Flight Mechanics Conference. Reston: AIAA, 2015: 2244.

[20] 杨超, 宋晨, 吴志刚, 等. 多控制面飞机的全机颤振主动抑制设计[J]. 航空学报, 2010, 31(8): 1501-1508.
YANG C, SONG C, WU Z G, et al. Active flutter suppression of airplane configuration with multiple control surfaces[J]. Acta Aeronautica et Astronautica Sinica, 2010, 31(8): 1501-1508(in Chinese).

[21] 沐旭升, 邹奇彤, 黄锐, 等. 体自由度颤振主动抑制的多输入/多输出自抗扰控制律设计[J]. 振动工程学报, 2020, 33(5): 910-920.
MU X S, ZOU Q T, HUANG R, et al. Design of multiple-input/multiple-output active disturbance rejection controller for body-freedom flutter suppression[J]. Journal of Vibration Engineering, 2020, 33(5): 910-920(in Chinese).

[22] BURNETT E, ATKINSON C, BERANEK J, et al. NDOF simulation model for flight control development with flight test correlation[C]// Proceedings of the AIAA Modeling and Simulation Technologies Conference. Reston: AIAA, 2010: 7780.

[23] 赵永辉. 气动弹性力学与控制[M]. 北京: 科学出版社, 2007.
ZHAO Y H. Aeroelastic mechanics and control[M]. Beijing: Science Press, 2007(in Chinese).

Design of multiple-input/multiple-output control law for active flutter suppression of flying-wing aircraft

FENG Yuxuan，HUO Yingyuan*，LI Junjie

(Research Institute of Strength Design，the First Aircraft Design and Research Institute of Aviation Industry Corporation，Xi'an 710089，China)

Abstract: In this paper, a multiple-input/multiple-output control law design method for active flutter suppression of flying-wing aircraft is proposed. The input/output relationship of the plant is used to estimate and adjust the "unknown perturbation" caused by the uncertain factors of the controlled plant. The output feedback control law is then designed. In order to validate the effectiveness of the proposed control law on flutter suppression, a high aspect ratio flying-wing layout aircraft was chosen as the research object of this paper. For the controller synthesis, the inboard ailerons served as the control surfaces, and the rigid-body pitch rate combined with the wing tip motion velocity provided the feedback. A high aspect ratio flying-wing layout aircraft was selected as the research object for this paper in order to verify the efficacy of the suggested control law on flutter suppression. The two aileron rudder surfaces on the inside of the fuselage were used as control inputs, and the rigid body pitch speed and wing tip motion speed were used as feedback signals to design the control input. The relation between velocity and damping of the closed-loop system and the closed-loop time-domain state equation is carried out. The results show that the control law design method can effectively improve the critical flutter velocity of the aircraft.

Keywords: aeroelasticity; flying-wing aircraft; active flutter suppression; body freedom flutter; MIMO output feedback controller

Received：2024-03-14；**Accepted**：2024-05-09；**Published Online**：2024-05-20 16：15

URL：link.cnki.net/urlid/11.2625.V.20240517.1909.002

* **Corresponding author**. E-mail：huoyy@sina.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0155

吊舱组合导航系统弹性杆臂误差补偿方法

王海鹏[1]，刘海娜[2]，王振东[3]，缪存孝[1,*]，张贺[1]，叶文[4]

(1. 北京科技大学 机械工程学院，北京 100083；2. 北京航天控制仪器研究所，北京 100048；3. 航空工业成都飞机工业（集团）有限责任公司，成都 610073；4. 中国计量科学研究院，北京 100029)

摘　　要：针对机载光电吊舱受迫振动和框架角动态变化导致的动态弹性杆臂效应问题，提出一种补偿弹性多杆臂误差的组合导航滤波方法。通过分析吊舱受迫振动的主要形式，建立吊舱线振动和角振动的动力学模型及由吊舱振动导致的弹性外杆臂模型；在多杆臂误差模型的基础上，对吊舱振动导致的速度、位置误差及振动角误差进行推导，分别建立弹性外杆臂误差模型和振动角误差模型，并将其引入状态向量中，通过对不同机动状态下的系统状态向量进行可观测度分析，得到各状态向量可观测的充分条件，为规划机动方案提供了基于可观测度的自适应调整依据。仿真实验结果表明：相较于刚性杆臂误差补偿，弹性多杆臂误差补偿后的速度误差均方根误差最大减少了73.4%，位置误差均方根误差最大减少了77.4%，有效提升了吊舱组合导航精度。

关　键　词：光电吊舱；受迫振动；杆臂误差；卡尔曼滤波；可观测度分析

中图分类号：V249.32

文献标志码：A　　**文章编号**：1001-5965(2026)05-1728-11

光电吊舱是无人机系统中重要的任务载荷，通常作为无人机的“眼睛”[1]。光电吊舱利用其内部搭载的可见光相机、红外成像仪、激光测距仪等高精度测量设备，实现对目标的探测、跟踪和定位等功能，在对敌侦察、军事打击、航空测量等军用及民用领域发挥着重要作用[2-3]。

随着机载光电吊舱技术的持续发展，对获取光电吊舱的速度、位置及姿态的实时精度要求也在不断提高。目前，有两轴或三轴稳定平台的光电侦察/瞄准吊舱，可以实现大范围的目标搜索、捕获和跟踪，光电吊舱中的光电探测系统在确定目标的相对方位和距离后，通过自身精确位置、速度、姿态信息可计算出监测目标的准确位置，显著提升了无人机攻击与侦察能力[4]。为使吊舱组合导航系统具备较高的实时精度，通常将子惯导安装于吊舱内部，与机载全球卫星导航系统(Global Navigation Satellite System, GNSS)进行组合。在光电吊舱执行任务时，其框架角处于不断转动状态，由于子惯导并非安装于吊舱几何中心，且与机载GNSS天线的安装位置存在一定的距离，必然会导致多杆臂效应误差，而且在载机机动过程中，受机动冲击、高速气流、机体振动等因素的影响[5]，吊舱会产生受迫振动，使吊舱实时姿态存在振动角误差及其几何中心与GNSS天线之间的相对位置(即吊舱外杆臂矢量)产生动态弹性变化，将严重影响吊舱组合导航系统的速度和定位定姿精度。

目前，通常将杆臂误差作为状态向量引入组合导航卡尔曼滤波器中，通过在线估计的方式补偿其导致的速度和位置误差[6-7]。文献[8]分析了载体处于不同机动状态时杆臂误差估计结果所受的影响，并将杆臂误差估计精度控制在厘米级。文献[9]通过分析杆臂误差产生的机理，将杆臂误差及加速度

收稿日期：2024-03-19；**录用日期**：2024-05-09；**网络出版时间**：2024-06-03 14:35
网络出版地址：link.cnki.net/urlid/11.2625.V.20240531.1831.001
基金项目：国家自然科学基金 (61603035)
* **通信作者**. E-mail：miaocx@ustb.edu.cn

引用格式：王海鹏，刘海娜，王振东，等. 吊舱组合导航系统弹性杆臂误差补偿方法[J]. 北京航空航天大学学报，2026，52(5)：1728-1738.
WANG H P，LIU H N，WANG Z D，et al. Pod integrated navigation system elastic lever arm error compensation method[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1728-1738 (in Chinese).

计和陀螺的标度因数误差引入状态变量中, 从而实现杆臂误差的在线估计和补偿。但在吊舱组合导航系统中, 由于子惯导安装于吊舱内部及吊舱框架角不断转动[10-11], 将产生多杆臂效应误差, 传统的补偿方法将不再适用, 因此, 文献 [12] 在考虑 GNSS 天线与稳定平台的长杆臂误差的同时, 引入稳定平台与惯导之间的短杆臂误差, 建立了更为准确的多杆臂误差模型, 文献 [13] 提出了一种适用于光电吊舱的多杆臂效应补偿算法, 能够较为准确地估计出吊舱内、外杆臂向量。文献 [14-16] 对机身挠曲变形所产生的动态杆臂误差进行了较为深入的研究。文献 [17] 通过建立机翼挠曲模型, 得到了杆臂在挠曲变形下的标称动态矢量, 并推导了杆臂动态补偿理论。但吊舱受迫振动导致的弹性外杆臂误差及振动角误差一般较少被提及, 若不对其进行相应补偿, 则其引入的速度、位置、姿态误差将对吊舱组合导航结果造成影响。

本文针对吊舱组合导航系统中的弹性多杆臂误差和振动角误差问题, 从建立吊舱振动动力学模型出发, 在多杆臂误差模型的基础上, 推导弹性外杆臂误差及振动角误差, 并将其引入状态向量中, 进行滤波估计补偿, 并采用基于奇异值分解的可观测度分析方法对不同机动条件下的状态向量进行定量分析, 为合理规划机动方案提供依据。最后, 将弹性多杆臂补偿方法与传统刚性多杆臂补偿方法进行仿真实验对比。

1　弹性多杆臂误差模型构建

1.1　吊舱振动动力学模型

吊舱通常安装于无人机腹部, 如图 1 所示, 与载机之间的连接通常布置有隔振器。在实际应用中, 吊舱隔振器通常为对称布置, 本文主要分析 4 个隔振器对称布置时的动力学模型, 布置方式如图 2 所示。

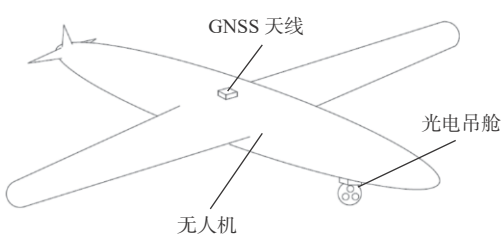


图 1　吊舱安装位置示意图

Fig. 1　Installation position diagram of pod

建立 4 个隔振器对称布置时的动力学模型有以下简化原则:

1) 光电吊舱的质量和刚度要远大于隔振器, 故将光电吊舱视作刚体, 忽略隔振器的质量。

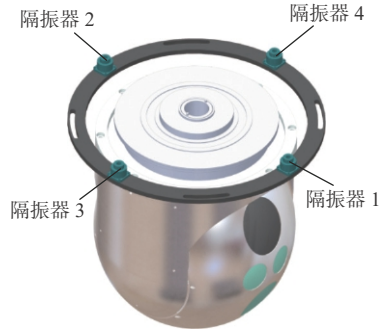


图 2　隔振器布置方式示意图

Fig. 2　Schematic diagram of vibration isolator layout

2) 隔振器阻尼力与速度的一次方成正比。

3) 根据隔振器的布置方式, 考虑受振动激励影响, 光电吊舱产生的受迫振动响应为横向倾覆、纵向倾覆和垂直位移。

经简化后, 吊舱的弹性振动可视为三自由度振动系统, 如图 3 所示。

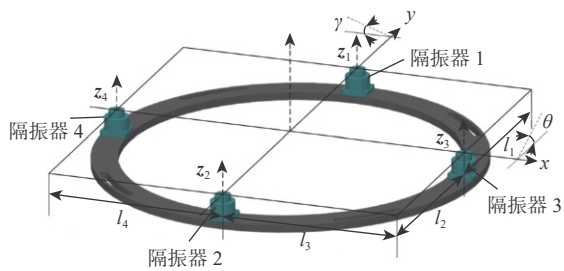


图 3　吊舱振动系统示意图

Fig. 3　Schematic diagram of the pod vibration system

设光电吊舱质量为m, 4 个隔振器的刚度为k_1、k_2、k_3、k_4, 阻尼系数为b_1、b_2、b_3、b_4, 隔振器安装半径为r; 设外部环境影响所引起的 4 个隔振器的位移激励为$\boldsymbol{u}_1$、$\boldsymbol{u}_2$、$\boldsymbol{u}_3$、$\boldsymbol{u}_4$, 4 个隔振器连接点的位移为$\boldsymbol{z}_1$、$\boldsymbol{z}_2$、$\boldsymbol{z}_3$、$\boldsymbol{z}_4$, 光电吊舱安装面中心位移为$\boldsymbol{z}$; 中心轴x穿过隔振器 3 和隔振器 4, 两侧的隔振器 1 和隔振器 2 到中心轴x的距离分别为l_1和l_2, 中心轴y穿过隔振器 1 和隔振器 2, 两侧的隔振器 3 和隔振器 4 到中心轴y的距离分别为l_3和l_4, 且$l_1=l_2=l_3=l_4=r$; 光电吊舱绕中心轴y的振动角为γ, 转动惯量为I_γ, 光电吊舱绕中心轴x的振动角为θ, 转动惯量为I_θ。

根据动力学定律建立吊舱弹性振动微分方程如下:

$$m\ddot{\boldsymbol{z}}+\sum_{i=1}^{4}b_i\dot{\boldsymbol{z}}_i+\sum_{i=1}^{4}k_i\boldsymbol{z}_i=\sum_{i=1}^{4}b_i\dot{\boldsymbol{u}}_i+\sum_{i=1}^{4}k_i\boldsymbol{u}_i \tag{1}$$

$$I_\gamma\ddot{\gamma}+r\sum_{i=3}^{4}(-1)^i b_i\dot{\boldsymbol{z}}_i+r\sum_{i=3}^{4}(-1)^i k_i\boldsymbol{z}_i=r\sum_{i=3}^{4}(-1)^i b_i\dot{\boldsymbol{u}}_i+r\sum_{i=3}^{4}(-1)^i k_i\boldsymbol{u}_i \tag{2}$$

$$I_\theta\ddot{\theta}+r\sum_{i=1}^{2}(-1)^i b_i\ddot{z}_i+r\sum_{i=1}^{2}(-1)^i k_i z_i=$$
$$r\sum_{i=1}^{2}(-1)^i b_i\ddot{u}_i+r\sum_{i=1}^{2}(-1)^i k_i u_i \tag{3}$$

考虑振动角γ和θ为小角度，由三角函数泰勒展开忽略二阶以上小量后，光电吊舱安装面中心的位移z与4个隔振器连接点的位移z_i之间几何关系满足

$$\begin{cases} z_1=z+r\theta \\ z_2=z-r\theta \\ z_3=z-r\gamma \\ z_4=z+r\gamma \end{cases} \tag{4}$$

将式(4)代入式(1)～式(3)中，进而整理为矩阵形式，得到改写后的吊舱弹性振动微分方程为

$$\boldsymbol{M}\ddot{\boldsymbol{x}}+\boldsymbol{B}\dot{\boldsymbol{x}}+\boldsymbol{K}\boldsymbol{x}=\boldsymbol{E}\dot{\boldsymbol{u}}+\boldsymbol{F}\boldsymbol{u} \tag{5}$$

式中：

$$\boldsymbol{x}=\begin{bmatrix} z & \gamma & \theta \end{bmatrix}^{\mathrm{T}};\ \boldsymbol{u}=\begin{bmatrix} u_1 & u_2 & u_3 & u_4 \end{bmatrix}^{\mathrm{T}};$$

$$\boldsymbol{M}=\begin{bmatrix} m & 0 & 0 \\ 0 & I_\gamma & 0 \\ 0 & 0 & I_\theta \end{bmatrix};\ \boldsymbol{E}=\begin{bmatrix} b_1 & b_2 & b_3 & b_4 \\ 0 & 0 & -rb_3 & rb_4 \\ -rb_1 & rb_2 & 0 & 0 \end{bmatrix};$$

$$\boldsymbol{F}=\begin{bmatrix} k_1 & k_2 & k_3 & k_4 \\ 0 & 0 & -rk_3 & rk_4 \\ -rk_1 & rk_2 & 0 & 0 \end{bmatrix};$$

$$\boldsymbol{B}=\begin{bmatrix} \sum_{i=1}^{4}b_i & r\sum_{i=3}^{4}(-1)^i b_i & r\sum_{i=1}^{2}(-1)^{i-1}b_i \\ r\sum_{i=3}^{4}(-1)^i b_i & r^2\sum_{i=3}^{4}b_i & 0 \\ r\sum_{i=1}^{2}(-1)^i b_i & 0 & r^2\sum_{i=1}^{2}b_i \end{bmatrix};$$

$$\boldsymbol{K}=\begin{bmatrix} \sum_{i=1}^{4}k_i & r\sum_{i=3}^{4}(-1)^i k_i & r\sum_{i=1}^{2}(-1)^{i-1}k_i \\ r\sum_{i=3}^{4}(-1)^i k_i & r^2\sum_{i=3}^{4}k_i & 0 \\ r\sum_{i=1}^{2}(-1)^i k_i & 0 & r^2\sum_{i=1}^{2}k_i \end{bmatrix}$$

由于阻尼矩阵$\boldsymbol{B}$和刚度矩阵$\boldsymbol{K}$为非对角阵，各自由度的振动将存在耦合现象，因此，在进行光电吊舱隔振器设计时，应对其解耦，即满足$b_1=b_2$、$b_3=b_4$、$k_1=k_2$、$k_3=k_4$条件，此时，阻尼矩阵$\boldsymbol{B}$和刚度矩阵$\boldsymbol{K}$由非对角矩阵转化为对角矩阵，沿z轴的线振动激励只会引起光电吊舱沿z轴的线振动，而不会耦合为光电吊舱绕x轴和y轴的角振动，即光电吊舱的线振动和角振动没有关联，解耦后的吊舱弹性振动微分方程为

$$\boldsymbol{M}\begin{bmatrix} \ddot{z} \\ \ddot{\gamma} \\ \ddot{\theta} \end{bmatrix}+\begin{bmatrix} \sum_{i=1}^{4}b_i & 0 & 0 \\ 0 & r^2\sum_{i=3}^{4}b_i & 0 \\ 0 & 0 & r^2\sum_{i=1}^{2}b_i \end{bmatrix}\begin{bmatrix} \dot{z} \\ \dot{\gamma} \\ \dot{\theta} \end{bmatrix}+$$
$$\begin{bmatrix} \sum_{i=1}^{4}k_i & 0 & 0 \\ 0 & r^2\sum_{i=3}^{4}k_i & 0 \\ 0 & 0 & r^2\sum_{i=1}^{2}k_i \end{bmatrix}\begin{bmatrix} z \\ \gamma \\ \theta \end{bmatrix}=\boldsymbol{E}\dot{\boldsymbol{u}}+\boldsymbol{F}\boldsymbol{u} \tag{6}$$

当吊舱处于高空环境中时，受迫振动导致的吊舱几何中心位移使吊舱几何中心与卫星天线之间的杆臂向量发生动态变化，且吊舱振动角将影响吊舱对目标的定位精度。

1.2 弹性外杆臂误差及振动角误差

光电吊舱弹性振动将使吊舱外杆臂$\delta\boldsymbol{l}_{\mathrm{dg}}$的长度发生变化，即吊舱外杆臂由刚性变为弹性，受动态振动影响后的弹性外杆臂$\delta\boldsymbol{L}_{\mathrm{dg}}$可表示为

$$\delta\boldsymbol{L}_{\mathrm{dg}}=\delta\boldsymbol{l}_{\mathrm{dg}}+\Delta\boldsymbol{l}_{\mathrm{dg}}=\delta\boldsymbol{l}_{\mathrm{dg}}+\Delta z+h\delta\boldsymbol{a}_{\mathrm{dr}} \tag{7}$$

式中：$\delta\boldsymbol{l}_{\mathrm{dg}}$为刚性外杆臂矢量；$\Delta\boldsymbol{l}_{\mathrm{dg}}$为振动导致的吊舱质心位移矢量；$\Delta\boldsymbol{z}=\begin{bmatrix}0 & 0 & z\end{bmatrix}^{\mathrm{T}}$为吊舱竖直振动位移量；$h$为吊舱安装面中心到光电吊舱几何中心的距离；$\delta\boldsymbol{a}_{\mathrm{dr}}=\begin{bmatrix}\gamma_{\mathrm{dr}} & \theta_{\mathrm{dr}} & 0\end{bmatrix}^{\mathrm{T}}$为吊舱振动倾覆角，其中，$\gamma_{\mathrm{dr}}$和$\theta_{\mathrm{dr}}$分别为吊舱振动横滚倾覆角和俯仰倾覆角。

如图4所示，O为地心，光电吊舱几何中心O_{d}与GNSS天线中心O_{G}之间的弹性外杆臂矢量为$\delta\boldsymbol{L}_{\mathrm{dg}}$，$\boldsymbol{r}_{\mathrm{g}}$为GNSS天线中心相对地心的矢量；$\boldsymbol{r}_{\mathrm{d}}$为吊舱几何中心相对地心的矢量，三者之间的矢量关系满足

$$\boldsymbol{r}_{\mathrm{g}}=\boldsymbol{r}_{\mathrm{d}}+\delta\boldsymbol{L}_{\mathrm{dg}}=\boldsymbol{r}_{\mathrm{d}}+\delta\boldsymbol{l}_{\mathrm{dg}}+\Delta\boldsymbol{l}_{\mathrm{dg}} \tag{8}$$

弹性外杆臂导致吊舱几何中心与GNSS天线中心之间的速度误差，将式(8)两边相对地球坐标

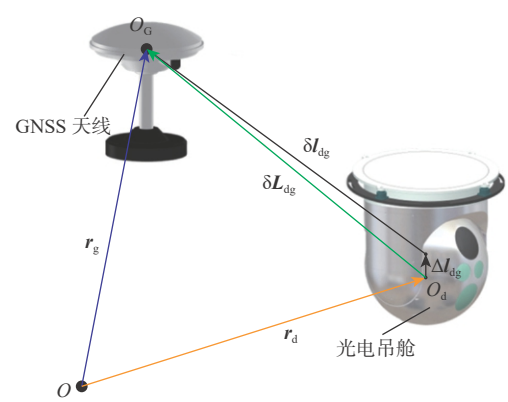


图4 弹性外杆臂示意图

Fig. 4 Elastic rod arm schematic diagram

系e系求导, 可得

$$\left.\frac{\mathrm{d}\boldsymbol{r}_{\mathrm{g}}}{\mathrm{d}t}\right|_{\mathrm{e}}=\left.\frac{\mathrm{d}\boldsymbol{r}_{\mathrm{d}}}{\mathrm{d}t}\right|_{\mathrm{e}}+\left.\frac{\mathrm{d}(\delta\boldsymbol{l}_{\mathrm{dg}})}{\mathrm{d}t}\right|_{\mathrm{e}}+\left.\frac{\mathrm{d}(\Delta\boldsymbol{l}_{\mathrm{dg}})}{\mathrm{d}t}\right|_{\mathrm{e}}=\left.\frac{\mathrm{d}\boldsymbol{r}_{\mathrm{d}}}{\mathrm{d}t}\right|_{\mathrm{e}}+\left.\frac{\mathrm{d}(\delta\boldsymbol{l}_{\mathrm{dg}})}{\mathrm{d}t}\right|_{\mathrm{d0}}+\boldsymbol{\omega}_{\mathrm{ed0}}\times\delta\boldsymbol{l}_{\mathrm{dg}}+\left.\frac{\mathrm{d}(\Delta\boldsymbol{l}_{\mathrm{dg}})}{\mathrm{d}t}\right|_{\mathrm{dr0}}+\boldsymbol{\omega}_{\mathrm{edr0}}\times(\Delta\boldsymbol{z}+h\delta\boldsymbol{a}_{\mathrm{dr}})=\left.\frac{\mathrm{d}\boldsymbol{r}_{\mathrm{d}}}{\mathrm{d}t}\right|_{\mathrm{e}}+\boldsymbol{\omega}_{\mathrm{ed0}}\times\delta\boldsymbol{l}_{\mathrm{dg}}+\boldsymbol{\omega}_{\mathrm{edr0}}\times(\Delta\boldsymbol{z}+h\delta\boldsymbol{a}_{\mathrm{dr}})\tag{9}$$

即

$$\boldsymbol{v}_{\mathrm{g}}=\boldsymbol{v}_{\mathrm{d}}+\boldsymbol{\omega}_{\mathrm{ed0}}\times\delta\boldsymbol{l}_{\mathrm{dg}}+\boldsymbol{\omega}_{\mathrm{edr0}}\times(\Delta\boldsymbol{z}+h\delta\boldsymbol{a}_{\mathrm{dr}})\tag{10}$$

式中: $\boldsymbol{v}_{\mathrm{g}}$为 GNSS 天线的地速; $\boldsymbol{v}_{\mathrm{d}}$为吊舱几何中心的地速; $\boldsymbol{\omega}_{\mathrm{ed0}}$为振动位移前吊舱几何中心在地球坐标系下的角速率; $\boldsymbol{\omega}_{\mathrm{edr0}}$为振动位移后吊舱几何中心在地球坐标系下的角速率。

将式(10)投影至导航坐标系n系后, 得到弹性外杆臂$\delta\boldsymbol{L}_{\mathrm{dg}}$导致的速度误差$\delta\boldsymbol{v}^{\mathrm{n}}_{\mathrm{Ldg}}$为

$$\delta\boldsymbol{v}^{\mathrm{n}}_{\mathrm{Ldg}}=\boldsymbol{v}^{\mathrm{n}}_{\mathrm{d}}-\boldsymbol{v}^{\mathrm{n}}_{\mathrm{g}}=-\boldsymbol{C}^{\mathrm{n}}_{\mathrm{d0}}(\boldsymbol{\omega}^{\mathrm{d0}}_{\mathrm{ed0}}\times)\delta\boldsymbol{l}^{\mathrm{d0}}_{\mathrm{dg}}-\boldsymbol{C}^{\mathrm{n}}_{\mathrm{dr0}}(\boldsymbol{\omega}^{\mathrm{dr0}}_{\mathrm{edr0}}\times)(\Delta\boldsymbol{z}^{\mathrm{dr0}}+h\delta\boldsymbol{a}^{\mathrm{dr0}}_{\mathrm{dr}})\tag{11}$$

式中: $\boldsymbol{v}^{\mathrm{n}}_{\mathrm{d}}$和$\boldsymbol{v}^{\mathrm{n}}_{\mathrm{g}}$分别为投影至$n$系后吊舱和 GNSS 天线的速度; $\boldsymbol{\omega}^{\mathrm{d0}}_{\mathrm{ed0}}$为振前吊舱角速率在吊舱初始坐标系$d_0$系下的投影; $\boldsymbol{\omega}^{\mathrm{dr0}}_{\mathrm{edr0}}$为振后吊舱角速率在吊舱振动位移后初始坐标系$d_{\mathrm{r}0}$系下的投影; $\boldsymbol{C}^{\mathrm{n}}_{\mathrm{d0}}$为$d_0$系到$n$系的转换矩阵; $\boldsymbol{C}^{\mathrm{n}}_{\mathrm{dr0}}$为$d_{\mathrm{r}0}$系到$n$系的转换矩阵; $\delta\boldsymbol{l}^{\mathrm{d0}}_{\mathrm{dg}}$为吊舱外杆臂在$d_0$系下的投影; $\Delta\boldsymbol{z}^{\mathrm{dr0}}$为振动导致的吊舱中心竖直位移量在$d_{\mathrm{r}0}$系下的投影; $\delta\boldsymbol{a}^{\mathrm{dr0}}_{\mathrm{dr}}$为吊舱振动倾覆角在$d_{\mathrm{r}0}$系下的投影。

本文所考虑两轴吊舱在吊舱框架角转动后, d_0系与吊舱实时坐标系d系不重合, 故$\boldsymbol{C}^{\mathrm{n}}_{\mathrm{d0}}$和$\boldsymbol{\omega}^{\mathrm{d0}}_{\mathrm{ed0}}$无法直接得到, 且吊舱发生振动后, $d_{\mathrm{r}0}$系不与d_0系重合, 故$\boldsymbol{C}^{\mathrm{n}}_{\mathrm{dr0}}$和$\boldsymbol{\omega}^{\mathrm{dr0}}_{\mathrm{edr0}}$也无法直接得到, 但可通过吊舱框架实时框架角、吊舱子惯导姿态角、载机主惯导姿态求得, 即

$$\boldsymbol{\omega}^{\mathrm{d0}}_{\mathrm{ed0}}=\begin{bmatrix}\mathrm{d}(\theta_{\mathrm{s}}-\theta_{\mathrm{d}})/\mathrm{d}t\\ \mathrm{d}\gamma_{\mathrm{s}}/\mathrm{d}t\\ \mathrm{d}(\psi_{\mathrm{s}}-\psi_{\mathrm{d}})/\mathrm{d}t\end{bmatrix}\tag{12}$$

$$\boldsymbol{C}^{\mathrm{n}}_{\mathrm{d0}}=\boldsymbol{C}^{\mathrm{n}}_{\mathrm{s}}\boldsymbol{C}^{\mathrm{s}}_{\mathrm{d}}\boldsymbol{C}^{\mathrm{d}}_{\mathrm{d0}}=\boldsymbol{C}^{\mathrm{n}}_{\mathrm{s}}\boldsymbol{C}^{\mathrm{s}}_{\mathrm{d}}\begin{bmatrix}\cos\psi_{\mathrm{d}} & \sin\psi_{\mathrm{d}} & 0\\ -\sin\psi_{\mathrm{d}}\cos\theta_{\mathrm{d}} & \cos\psi_{\mathrm{d}}\cos\theta_{\mathrm{d}} & \sin\theta_{\mathrm{d}}\\ -\sin\theta_{\mathrm{d}}\sin\psi_{\mathrm{d}} & -\sin\theta_{\mathrm{d}}\cos\psi_{\mathrm{d}} & \cos\theta_{\mathrm{d}}\end{bmatrix}\tag{13}$$

$$\boldsymbol{\omega}^{\mathrm{dr0}}_{\mathrm{edr0}}=\begin{bmatrix}\mathrm{d}(\theta_{\mathrm{s}}-\theta_{\mathrm{d}}-\theta_{\mathrm{m}})/\mathrm{d}t\\ \mathrm{d}(\gamma_{\mathrm{s}}-\gamma_{\mathrm{m}})/\mathrm{d}t\\ \mathrm{d}(\psi_{\mathrm{s}}-\psi_{\mathrm{d}}-\gamma_{\mathrm{m}})/\mathrm{d}t\end{bmatrix}\tag{14}$$

$$\boldsymbol{C}^{\mathrm{n}}_{\mathrm{dr0}}=\boldsymbol{C}^{\mathrm{n}}_{\mathrm{d0}}\boldsymbol{C}^{\mathrm{d0}}_{\mathrm{dr0}}=\boldsymbol{C}^{\mathrm{n}}_{\mathrm{d0}}\begin{bmatrix}\cos\gamma_{\mathrm{dr}} & -\sin\theta_{\mathrm{dr}}\sin\gamma_{\mathrm{dr}} & -\sin\gamma_{\mathrm{dr}}\cos\theta_{\mathrm{dr}}\\ 0 & \cos\theta_{\mathrm{dr}} & \sin\theta_{\mathrm{dr}}\\ \sin\gamma_{\mathrm{dr}} & -\sin\theta_{\mathrm{dr}}\cos\gamma_{\mathrm{dr}} & \cos\theta_{\mathrm{dr}}\cos\gamma_{\mathrm{dr}}\end{bmatrix}\tag{15}$$

式中: $\boldsymbol{C}^{\mathrm{n}}_{\mathrm{s}}$为子惯导坐标系$s$系到$n$系的转换矩阵; $\boldsymbol{C}^{\mathrm{s}}_{\mathrm{d}}$为$d$系到$s$系的转换矩阵, 由于子惯导与吊舱固连, 故为常值矩阵; $\boldsymbol{C}^{\mathrm{d}}_{\mathrm{d0}}$为$d_0$系到$d$系的转换矩阵; $\boldsymbol{C}^{\mathrm{d0}}_{\mathrm{dr0}}$为$d_{\mathrm{r}0}$系到$d_0$系的转换矩阵; ψ_{d}、θ_{d}分别为吊舱框架偏航角和俯仰角; ψ_{s}、θ_{s}、γ_{s}分别为子惯导的偏航角、俯仰角和横滚角; θ_{m}、γ_{m}分别为载机主惯导的俯仰角和横滚角。

如图 4 所示, 弹性外杆臂导致吊舱几何中心与 GNSS 天线之间的位置误差。将弹性外杆臂分别在东向(E)、北向 (N)、天向 (U) 投影, 可得

$$\delta\boldsymbol{L}^{\mathrm{n}}_{\mathrm{dg}}=\begin{bmatrix}\delta L_{\mathrm{E}} & \delta L_{\mathrm{N}} & \delta L_{\mathrm{U}}\end{bmatrix}^{\mathrm{T}}=\boldsymbol{C}^{\mathrm{n}}_{\mathrm{d0}}\delta\boldsymbol{l}^{\mathrm{d0}}_{\mathrm{dg}}+\boldsymbol{C}^{\mathrm{n}}_{\mathrm{dr0}}\Delta\boldsymbol{l}^{\mathrm{dr0}}_{\mathrm{dg}}=\boldsymbol{C}^{\mathrm{n}}_{\mathrm{d0}}\delta\boldsymbol{l}^{\mathrm{d0}}_{\mathrm{dg}}+\boldsymbol{C}^{\mathrm{n}}_{\mathrm{dr0}}(\Delta\mathrm{z}^{\mathrm{dr0}}+h\delta\boldsymbol{a}^{\mathrm{dr0}}_{\mathrm{dr}})\tag{16}$$

式中: $\boldsymbol{L}^{\mathrm{n}}_{\mathrm{dg}}$为弹性外杆臂在导航坐标系 n 系下投影; L_{E}为弹性外杆臂在导航坐标系系 n 系东向 E 的投影; L_{N}为弹性外杆臂在导航坐标系系 n 系北向 N 的投影; $\boldsymbol{L}_{\mathrm{U}}$为弹性外杆臂在导航坐标系系 n 系天向 U 的投影; $\boldsymbol{l}^{\mathrm{dr0}}_{\mathrm{dg}}$为振动引起的吊舱质心位移矢量在吊舱震动位移后初始坐标系 dr0 系下的投影。

若记 GNSS 天线中心和振动后吊舱几何中心的位置为

$$\begin{cases}\boldsymbol{p}_{\mathrm{g}}=\begin{bmatrix}L_{\mathrm{g}} & \lambda_{\mathrm{g}} & h_{\mathrm{g}}\end{bmatrix}^{\mathrm{T}}\\ \boldsymbol{p}_{\mathrm{d}}=\begin{bmatrix}L_{\mathrm{d}} & \lambda_{\mathrm{d}} & h_{\mathrm{d}}\end{bmatrix}^{\mathrm{T}}\end{cases}\tag{17}$$

式中: L_{g}、λ_{g}、h_{g}分别为 GNSS 天线的经度、纬度、高度; L_{d}、λ_{d}、h_{d}分别为振动后吊舱几何中心的经度、纬度、高度。

则 GNSS 天线与吊舱几何中心的位置误差为

$$\begin{cases}L_{\mathrm{d}}-L_{\mathrm{g}}=-\delta L_{\mathrm{N}}/R_{\mathrm{Mh}}\\ \lambda_{\mathrm{d}}-\lambda_{\mathrm{g}}=-\delta L_{\mathrm{E}}\sec L_{\mathrm{d}}/R_{\mathrm{Nh}}\\ h_{\mathrm{d}}-h_{\mathrm{g}}=-\delta L_{\mathrm{U}}\end{cases}\tag{18}$$

式中: R_{Mh}和R_{Nh}分别为子午圈主曲率半径和卯酉圈主曲率半径[18]。

联立式(16)～式(18), 可得由弹性外杆臂导致的位置误差为

$$\delta\boldsymbol{p}^{\mathrm{n}}_{l_{\mathrm{dg}}}=\boldsymbol{p}_{\mathrm{d}}-\boldsymbol{p}_{\mathrm{g}}=-\boldsymbol{M}_{\mathrm{pv}}\boldsymbol{C}^{\mathrm{n}}_{\mathrm{d0}}\delta\boldsymbol{l}^{\mathrm{d0}}_{\mathrm{dg}}-\boldsymbol{M}_{\mathrm{pv}}\boldsymbol{C}^{\mathrm{n}}_{\mathrm{dr0}}\Delta\boldsymbol{l}^{\mathrm{dr0}}_{\mathrm{dg}}=-\boldsymbol{M}_{\mathrm{pv}}\boldsymbol{C}^{\mathrm{n}}_{\mathrm{d0}}\delta\boldsymbol{l}^{\mathrm{d0}}_{\mathrm{dg}}-\boldsymbol{M}_{\mathrm{pv}}\boldsymbol{C}^{\mathrm{n}}_{\mathrm{dr0}}(\Delta\boldsymbol{z}^{\mathrm{dr0}}+h\delta\boldsymbol{a}^{\mathrm{dr0}}_{\mathrm{dr}})\tag{19}$$

式中: $\boldsymbol{M}_{\mathrm{pv}}$为$n$系下的位置误差到经度、纬度和高度误差的转换矩阵[18]。

由吊舱振动形式分析可知, 吊舱振动将产生纵向倾覆振动角和横向倾覆振动角。若记主、子惯导和吊舱框架角姿态为

$$\begin{cases}\boldsymbol{a}^{\mathrm{n}}_{\mathrm{m}}=\begin{bmatrix}\gamma_{\mathrm{m}} & \theta_{\mathrm{m}} & \psi_{\mathrm{m}}\end{bmatrix}^{\mathrm{T}}\\ \boldsymbol{a}^{\mathrm{n}}_{\mathrm{s}}=\begin{bmatrix}\gamma_{\mathrm{s}} & \theta_{\mathrm{s}} & \psi_{\mathrm{s}}\end{bmatrix}^{\mathrm{T}}\\ \boldsymbol{a}^{\mathrm{n}}_{\mathrm{d}}=\begin{bmatrix}0 & \theta_{\mathrm{d}} & \psi_{\mathrm{d}}\end{bmatrix}^{\mathrm{T}}\end{cases}\tag{20}$$

式中: $\boldsymbol{a}^{\mathrm{n}}_{\mathrm{m}}$、$\boldsymbol{a}^{\mathrm{n}}_{\mathrm{s}}$和$\boldsymbol{a}^{\mathrm{n}}_{\mathrm{d}}$分别为主、子惯导姿态和吊舱框

架角，ψ_m为载机主惯导偏航角。

则吊舱振动角误差为

$$\delta\boldsymbol{a}_{dr}=\boldsymbol{a}_s^n-\boldsymbol{a}_d^n-\boldsymbol{a}_m^n=\begin{bmatrix}\gamma_s-\gamma_m\\ \theta_s-\theta_d-\theta_m\\ \psi_s-\psi_d-\psi_m\end{bmatrix} \tag{21}$$

1.3 吊舱内杆臂误差

如图5所示，O为地心；子惯导中心O_s与光电吊舱几何中心O_d之间的杆臂为$\delta\boldsymbol{l}_{sd}$；$\boldsymbol{r}_s$为子惯导中心相对地心的矢量；$\boldsymbol{r}_d$为吊舱几何中心相对地心的矢量，三者矢量关系满足

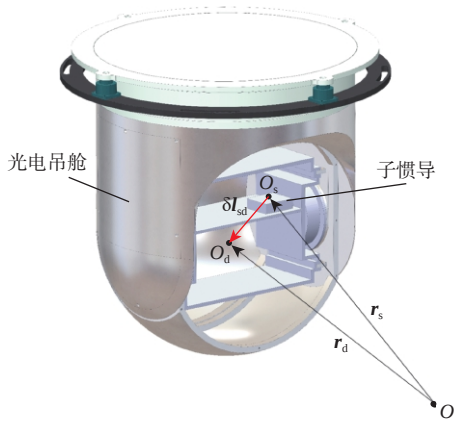


图5 吊舱内杆臂示意图

Fig. 5 Schematic diagram of rod arm in pod

$$\boldsymbol{r}_d=\boldsymbol{r}_s+\delta\boldsymbol{l}_{sd} \tag{22}$$

吊舱内杆臂导致吊舱几何中心与子惯导中心之间的速度误差，将式(22)两边相对地球坐标系即e系求导，可得

$$\left.\frac{\mathrm{d}\boldsymbol{r}_d}{\mathrm{d}t}\right|_e=\left.\frac{\mathrm{d}\boldsymbol{r}_s}{\mathrm{d}t}\right|_e+\left.\frac{\mathrm{d}(\delta\boldsymbol{l}_{sd})}{\mathrm{d}t}\right|_e=\left.\frac{\mathrm{d}\boldsymbol{r}_s}{\mathrm{d}t}\right|_e+\left.\frac{\mathrm{d}(\delta\boldsymbol{l}_{sd})}{\mathrm{d}t}\right|_s+\boldsymbol{\omega}_{es}\times\delta\boldsymbol{l}_{sd}=\left.\frac{\mathrm{d}\boldsymbol{r}_s}{\mathrm{d}t}\right|_e+\boldsymbol{\omega}_{es}\times\delta\boldsymbol{l}_{sd} \tag{23}$$

即

$$\boldsymbol{v}_d=\boldsymbol{v}_s+\boldsymbol{\omega}_{es}\times\delta\boldsymbol{l}_{sd} \tag{24}$$

式中：$\boldsymbol{v}_s$为子惯导中心的地速；$\boldsymbol{\omega}_{es}$为子惯导在地球坐标系e系下的角速率。

将式(24)投影至导航坐标系n系后，得到吊舱内杆臂$\delta\boldsymbol{l}_{sd}$导致的速度误差$\delta\boldsymbol{v}_{lsd}^n$为

$$\delta\boldsymbol{v}_{lsd}^n=\boldsymbol{v}_s^n-\boldsymbol{v}_d^n=-\boldsymbol{C}_s^n(\boldsymbol{\omega}_{es}^s\times)\delta\boldsymbol{l}_{sd}^s \tag{25}$$

式中：$\boldsymbol{v}_s^n$为投影至导航系后子惯导的速度；$\boldsymbol{\omega}_{es}^s$为子惯导角速率在$s$系下的投影；$\delta\boldsymbol{l}_{sd}^s$为吊舱内杆臂在$s$系下的投影。

如图5所示，吊舱内杆臂导致子惯导与吊舱几何中心的位置误差。由于子惯导刚性连接于吊舱内部安装面，内杆臂在子惯导坐标系下的投影$\delta\boldsymbol{l}_{sd}^s$为常值向量，用以下分量形式表示：

$$\delta\boldsymbol{l}_{sd}^s=\begin{bmatrix}\delta l_{sdx}^s & \delta l_{sdy}^s & \delta l_{sdz}^s\end{bmatrix}^{\mathrm{T}} \tag{26}$$

将内杆臂$\delta\boldsymbol{l}_{sd}$分别在东向(E)、北向(N)、天向(U)投影，可得

$$\delta\boldsymbol{l}_{sd}^n=\begin{bmatrix}\delta l_E & \delta l_N & \delta l_U\end{bmatrix}^{\mathrm{T}}=\boldsymbol{C}_s^n\delta\boldsymbol{l}_{sd}^s \tag{27}$$

若记光电吊舱几何中心和子惯导中心的位置为

$$\begin{cases}\boldsymbol{p}_d=\begin{bmatrix}L_d & \lambda_d & h_d\end{bmatrix}^{\mathrm{T}}\\ \boldsymbol{p}_s=\begin{bmatrix}L_s & \lambda_s & h_s\end{bmatrix}^{\mathrm{T}}\end{cases} \tag{28}$$

式中：L_s、λ_s、h_s分别为子惯导的经度、纬度、高度。

则吊舱几何中心与子惯导中心的位置误差为

$$\begin{cases}L_s-L_d=-\delta l_N/R_{Mh}\\ \lambda_s-\lambda_d=-\delta l_E\sec L_s/R_{Nh}\\ h_s-h_d=-\delta l_U\end{cases} \tag{29}$$

联立式(16)～式(18)，可得由吊舱内杆臂导致的位置误差为

$$\delta\boldsymbol{p}_{lsd}^n=\boldsymbol{p}_s-\boldsymbol{p}_d=-\boldsymbol{M}_{pv}\boldsymbol{C}_s^n\delta\boldsymbol{l}_{sd}^s \tag{30}$$

2 弹性多杆臂补偿滤波方法

为建立针对弹性多杆臂误差的系统状态方程，选取27维状态向量如下：

$$\boldsymbol{X}=\begin{bmatrix}(\boldsymbol{\phi}^n)^{\mathrm{T}} & (\delta\boldsymbol{v}^n)^{\mathrm{T}} & (\delta\boldsymbol{p}^n)^{\mathrm{T}} & (\boldsymbol{\varepsilon}^s)^{\mathrm{T}} & (\nabla^s)^{\mathrm{T}}\\ (\delta\boldsymbol{l}_{sd}^s)^{\mathrm{T}} & (\delta\boldsymbol{l}_{dg}^{d0})^{\mathrm{T}} & (\Delta\boldsymbol{z}^{dr0})^{\mathrm{T}} & (\delta\boldsymbol{a}_{dr}^{dr0})^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \tag{31}$$

式中：$\boldsymbol{\phi}^n=\begin{bmatrix}\phi_E & \phi_N & \phi_U\end{bmatrix}^{\mathrm{T}}$为子惯导姿态误差；$\delta\boldsymbol{v}^n=\begin{bmatrix}\delta v_E & \delta v_N & \delta v_U\end{bmatrix}^{\mathrm{T}}$为子惯导速度误差；$\delta\boldsymbol{p}^n=\begin{bmatrix}\delta L & \delta\lambda & \delta h\end{bmatrix}^{\mathrm{T}}$为子惯导位置误差；$\boldsymbol{\varepsilon}^s=\begin{bmatrix}\varepsilon_x & \varepsilon_y & \varepsilon_z\end{bmatrix}^{\mathrm{T}}$为子惯导陀螺零偏误差；$\nabla^s=\begin{bmatrix}\nabla_x & \nabla_y & \nabla_z\end{bmatrix}^{\mathrm{T}}$为子惯导加速度计零偏误差。

状态方程为

$$\dot{\boldsymbol{X}}=\boldsymbol{F}\boldsymbol{X}+\boldsymbol{G}\boldsymbol{W}^s \tag{32}$$

式中：$\boldsymbol{X}$为系统状态向量；$\boldsymbol{F}$为状态转移矩阵；$\boldsymbol{W}^s$为系统噪声向量；$\boldsymbol{G}$为系统噪声分配矩阵。

状态转移矩阵为

$$\boldsymbol{F}=\begin{bmatrix}\boldsymbol{M}_{aa} & \boldsymbol{M}_{av} & \boldsymbol{M}_{ap} & -\boldsymbol{C}_s^n & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3}\\ \boldsymbol{M}_{va} & \boldsymbol{M}_{vv} & \boldsymbol{M}_{vp} & \boldsymbol{0}_{3\times3} & \boldsymbol{C}_s^n & \boldsymbol{0}_{3\times3}\\ \boldsymbol{0}_{3\times3} & \boldsymbol{M}_{pv} & \boldsymbol{M}_{pp} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3}\\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{M}_g & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{12\times3}\\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{M}_a & \boldsymbol{0}_{12\times3}\\ \boldsymbol{0}_{12\times3} & \boldsymbol{0}_{12\times3} & \boldsymbol{0}_{12\times3} & \boldsymbol{0}_{12\times3} & \boldsymbol{0}_{12\times3} & \boldsymbol{0}_{12\times3}\end{bmatrix} \tag{33}$$

式中：$\boldsymbol{M}_{aa}$、$\boldsymbol{M}_{av}$、$\boldsymbol{M}_{ap}$、$\boldsymbol{M}_{va}$、$\boldsymbol{M}_{vv}$、$\boldsymbol{M}_{vp}$、$\boldsymbol{M}_{pv}$、$\boldsymbol{M}_{pp}$为捷联惯导误差方程系数矩阵[18]；

$$\boldsymbol{M}_g=\begin{bmatrix}-1/\tau_g & 0 & 0\\ 0 & -1/\tau_g & 0\\ 0 & 0 & -1/\tau_g\end{bmatrix};$$

$$\boldsymbol{M}_{\mathrm{a}}=\begin{bmatrix}-1/\tau_{\mathrm{a}} & 0 & 0\\ 0 & -1/\tau_{\mathrm{a}} & 0\\ 0 & 0 & -1/\tau_{\mathrm{a}}\end{bmatrix}。$$

式中: τ_{g}和τ_{a}分别为陀螺仪和加速度计的相关时间常数。

系统噪声向量$\boldsymbol{W}^{\mathrm{s}}$为

$$\boldsymbol{W}^{\mathrm{s}}=\begin{bmatrix}w_{\mathrm{g}}^{\mathrm{s}} & w_{\mathrm{a}}^{\mathrm{s}}\end{bmatrix}^{\mathrm{T}} \tag{34}$$

式中, $w_{\mathrm{g}}^{\mathrm{s}}$和$w_{\mathrm{a}}^{\mathrm{s}}$分别为陀螺角速度测量和加速度计比力测量白噪声。

系统噪声分配矩阵$\boldsymbol{G}$为

$$\boldsymbol{G}=\begin{bmatrix}-\boldsymbol{C}_{\mathrm{s}}^{\mathrm{n}} & \boldsymbol{0}_{3\times3}\\ \boldsymbol{0}_{3\times3} & \boldsymbol{C}_{\mathrm{s}}^{\mathrm{n}}\\ \boldsymbol{0}_{21\times3} & \boldsymbol{0}_{21\times3}\end{bmatrix} \tag{35}$$

根据 1.2 节和 1.3 节中推导得到的速度误差模型, 得到速度量测方程:

$$\begin{aligned}\boldsymbol{Z}_{\boldsymbol{v}}=\hat{\boldsymbol{v}}_{\mathrm{s}}^{\mathrm{n}}-\hat{\boldsymbol{v}}_{\mathrm{g}}^{\mathrm{n}}=\delta\boldsymbol{v}^{\mathrm{n}}-\boldsymbol{C}_{\mathrm{s}}^{\mathrm{n}}(\boldsymbol{\omega}_{\mathrm{es}}^{\mathrm{s}}\times)\delta\boldsymbol{l}_{\mathrm{sd}}^{\mathrm{s}}-\\ \boldsymbol{C}_{\mathrm{d0}}^{\mathrm{n}}(\boldsymbol{\omega}_{\mathrm{ed0}}^{\mathrm{d0}}\times)\delta\boldsymbol{l}_{\mathrm{dg}}^{\mathrm{d0}}-\boldsymbol{C}_{\mathrm{dr0}}^{\mathrm{n}}(\boldsymbol{\omega}_{\mathrm{edr0}}^{\mathrm{dr0}}\times)(\Delta\boldsymbol{z}^{\mathrm{dr0}}+h\delta\boldsymbol{a}_{\mathrm{dr}}^{\mathrm{dr0}})=\boldsymbol{H}_{\boldsymbol{v}}\boldsymbol{X}\end{aligned} \tag{36}$$

式中: $\boldsymbol{Z}_{\boldsymbol{v}}$为速度量测; $\boldsymbol{H}_{\boldsymbol{v}}$为速度量测矩阵。

$$\begin{aligned}\boldsymbol{H}_{\boldsymbol{v}}=\begin{bmatrix}\boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times9} & -\boldsymbol{C}_{\mathrm{s}}^{\mathrm{n}}(\boldsymbol{\omega}_{\mathrm{es}}^{\mathrm{s}}\times) & -\boldsymbol{C}_{\mathrm{d0}}^{\mathrm{n}}(\boldsymbol{\omega}_{\mathrm{ed0}}^{\mathrm{d0}}\times)\\ -\boldsymbol{C}_{\mathrm{dr0}}^{\mathrm{n}}(\boldsymbol{\omega}_{\mathrm{edr0}}^{\mathrm{dr0}}\times) & -\boldsymbol{C}_{\mathrm{dr0}}^{\mathrm{n}}(\boldsymbol{\omega}_{\mathrm{edr0}}^{\mathrm{dr0}}\times)h\end{bmatrix}\end{aligned} \tag{37}$$

根据 1.2 节和 1.3 节中推导得到的位置误差模型, 得到位置量测方程:

$$\begin{aligned}\boldsymbol{Z}_{\boldsymbol{p}}=\hat{\boldsymbol{p}}_{\mathrm{s}}^{\mathrm{n}}-\hat{\boldsymbol{p}}_{\mathrm{g}}^{\mathrm{n}}=\delta\boldsymbol{p}^{\mathrm{n}}-\boldsymbol{M}_{\mathrm{pv}}\boldsymbol{C}_{\mathrm{s}}^{\mathrm{n}}\delta\boldsymbol{l}_{\mathrm{sd}}^{\mathrm{s}}-\\ \boldsymbol{M}_{\mathrm{pv}}\boldsymbol{C}_{\mathrm{d0}}^{\mathrm{n}}\delta\boldsymbol{l}_{\mathrm{dg}}^{\mathrm{d0}}-\boldsymbol{M}_{\mathrm{pv}}\boldsymbol{C}_{\mathrm{dr0}}^{\mathrm{n}}(\Delta\boldsymbol{z}^{\mathrm{dr0}}+h\delta\boldsymbol{a}_{\mathrm{dr}}^{\mathrm{dr0}})=\boldsymbol{H}_{\boldsymbol{p}}\boldsymbol{X}\end{aligned} \tag{38}$$

式中: $\boldsymbol{Z}_{\boldsymbol{p}}$为位置量测; $\boldsymbol{H}_{\boldsymbol{p}}$为位置量测矩阵; $\hat{\boldsymbol{p}}_{\mathrm{s}}^{\mathrm{n}}$为子惯导位置量测信息; $\hat{\boldsymbol{p}}_{\mathrm{g}}^{\mathrm{n}}$为 GNSS 位置测量信息。

$$\begin{aligned}\boldsymbol{H}_{\boldsymbol{p}}=\begin{bmatrix}\boldsymbol{0}_{3\times6} & \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times6} & -\boldsymbol{M}_{\mathrm{pv}}\boldsymbol{C}_{\mathrm{s}}^{\mathrm{n}} & -\boldsymbol{M}_{\mathrm{pv}}\boldsymbol{C}_{\mathrm{d0}}^{\mathrm{n}}\\ -\boldsymbol{M}_{\mathrm{pv}}\boldsymbol{C}_{\mathrm{dr0}}^{\mathrm{n}} & -\boldsymbol{M}_{\mathrm{pv}}\boldsymbol{C}_{\mathrm{dr0}}^{\mathrm{n}}h\end{bmatrix}\end{aligned} \tag{39}$$

根据 1.2 节中推导得到的振动角误差模型, 得到姿态量测方程:

$$\boldsymbol{Z}_{\boldsymbol{a}}=\hat{\boldsymbol{a}}_{\mathrm{s}}^{\mathrm{n}}-\hat{\boldsymbol{a}}_{\mathrm{d}}^{\mathrm{n}}-\hat{\boldsymbol{a}}_{\mathrm{m}}^{\mathrm{n}}=\boldsymbol{\phi}^{\mathrm{n}}-\delta\boldsymbol{a}_{\mathrm{dr}}^{\mathrm{dr0}}=\boldsymbol{H}_{\boldsymbol{a}}\boldsymbol{X} \tag{40}$$

式中: $\boldsymbol{Z}_{\boldsymbol{a}}$为姿态量测; $\boldsymbol{H}_{\boldsymbol{a}}$为姿态量测矩阵。

$$\boldsymbol{H}_{\boldsymbol{a}}=\begin{bmatrix}\boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3}\end{bmatrix} \tag{41}$$

结合式(36)～式(41), 建立系统量测方程为

$$\boldsymbol{Z}=\begin{bmatrix}\hat{\boldsymbol{v}}_{\mathrm{s}}^{\mathrm{n}}-\hat{\boldsymbol{v}}_{\mathrm{g}}^{\mathrm{n}}\\ \hat{\boldsymbol{p}}_{\mathrm{s}}^{\mathrm{n}}-\hat{\boldsymbol{p}}_{\mathrm{g}}^{\mathrm{n}}\\ \hat{\boldsymbol{a}}_{\mathrm{s}}^{\mathrm{n}}-\hat{\boldsymbol{a}}_{\mathrm{d}}^{\mathrm{n}}-\hat{\boldsymbol{a}}_{\mathrm{m}}^{\mathrm{n}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{H}_{\boldsymbol{v}}\\ \boldsymbol{H}_{\boldsymbol{p}}\\ \boldsymbol{H}_{\boldsymbol{a}}\end{bmatrix}\boldsymbol{X}+\boldsymbol{V} \tag{42}$$

式中: $\boldsymbol{V}$为系统量测白噪声。

3　仿真验证

3.1　仿真飞行条件

以东-北-天地理坐标系为导航坐标系, 载机初始位置为L: 34.144 5°、λ: 108.543 4°、h: 380 m; 姿态误差为 0.05°; 速度误差为 0.05 m/s; GNSS 差分定位误差为 0.01 m; 子惯导陀螺零偏为0.3°/h, 随机游走为0.3°/$\sqrt{\mathrm{h}}$; 子惯导加表零偏为100 μg, 随机游走为10 μg/$\sqrt{\mathrm{Hz}}$; 光电吊舱框架偏航角转动范围为−45°～45°, 俯仰角转动范围为−30°～30°, 均以频率为0.01Hz的正弦波方式持续转动; 吊舱内杆臂长度$\delta\boldsymbol{l}_{\mathrm{sd}}^{\mathrm{s}}=\begin{bmatrix}0.1 & 0.1 & 0.05\end{bmatrix}$m, 吊舱外杆臂长度$\delta\boldsymbol{l}_{\mathrm{dg}}^{\mathrm{d0}}=\begin{bmatrix}0.1 & 2 & 0.8\end{bmatrix}$m; 仿真时长为 1 500 s。模拟隔振器位移激励作用下的吊舱受迫振动量, 如图 6 所示。

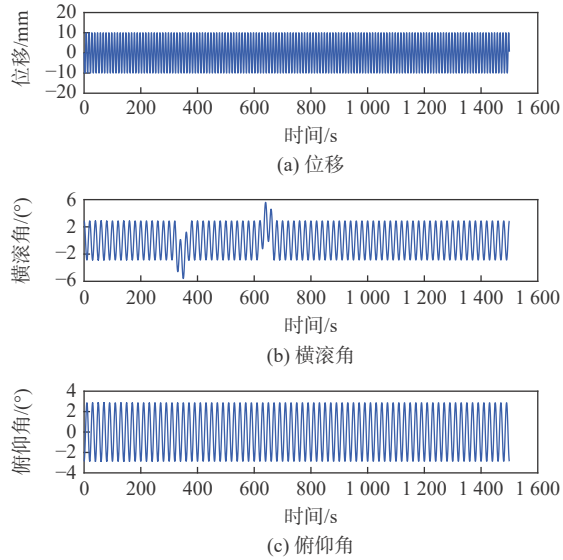


图 6　吊舱受迫振动量

Fig. 6　Forced vibration amplitude of pod

3.2　系统可观测度分析

在光电吊舱实际应用中, 载机运动方式和吊舱框架角处于不断变化状态, 即光电吊舱组合导航系统为时变系统, 若直接对该系统进行可观测性分析将变得十分复杂, 为实现定量分析系统可观测性, 采用基于奇异值分解的可观测性分析方法对光电吊舱组合导航系统在载机静止、直飞、变向机动及吊舱框架角转动与否条件下的可观测性进行分析, 即对以下 5 个状态下的组合导航系统中各状态向量的可观测度进行定量分析, 仿真参数设置同 3.1 节。

1) 全部静止状态: 此时, 载机和光电吊舱均不存在机动, 系统为时不变系统。

2) 载机直飞吊舱静止状态: 此时, 载机匀速平飞, 吊舱框架角保持静止, 但吊舱存在如图 6 所示的频率为 0.1 Hz、振幅为 10 mm 的竖直方向线振动和频率为 0.05 Hz、最大振幅分别为 5.55°和 2.86°的横滚倾覆角振动和俯仰倾覆角振动。

3) 载机直飞吊舱转动状态: 此时, 载机在匀速平飞的同时, 吊舱框架角均以频率为0.01 Hz 的正弦波的方式转动, 且吊舱存在机动方式 2 中所提到

的振动，偏航角转动范围为−45°～45°，俯仰角转动范围为−30°～30°。

4）载机变向吊舱静止状态：此时，载机向左90°转向机动，吊舱框架角保持静止，但吊舱存在机动方式2中所提到的振动。

5）载机变向吊舱转动状态：载机向左90°转向机动的同时，吊舱框架角均以频率为0.01 Hz的正弦波方式转动，且吊舱存在机动方式2中所提到的振动，偏航角转动范围为−45°～45°，俯仰角转动范围为−30°～30°。

每个状态的时间段取20 s，不同机动方式下系统状态向量可观测度分析结果如表1所示。

表1 状态变量对应的奇异值分析结果

Table 1 Singular value analysis results corresponding to state variables

机动方式	奇异值											
	ϕ_E	ϕ_N	ϕ_U	δv_E	δv_N	δv_U	δL	$\delta\lambda$	δh	ε_x	ε_y	ε_z
1	26.81	26.81	4.38×10^{-4}	4.73	4.73	8.42	13.41	13.41	13.41	33.62	33.62	2.11×10^{-4}
2	39.69	39.69	2.23×10^{-3}	6.62	6.62	17.74	22.31	22.31	22.31	40.09	40.09	2.46×10^{-3}
3	47.73	47.73	21.17	6.62	6.62	17.74	22.31	22.31	22.31	50.29	50.29	5.92
4	57.45	57.45	15.83	6.62	6.62	17.74	22.31	22.31	22.31	48.11	48.11	4.59
5	65.78	65.78	20.68	6.62	6.62	17.74	22.31	22.31	22.31	59.17	59.17	8.76

机动方式	奇异值											
	V_x	V_y	V_z	δl^s_{sdx}	δl^s_{sdy}	δl^s_{sdz}	δl^{d0}_{dgx}	δl^{d0}_{dgy}	δl^{d0}_{dgz}	z	γ_{dr}	θ_{dr}
1	2.30×10^{-4}	2.30×10^{-4}	4.57	4.01×10^{-5}	4.01×10^{-5}	4.01×10^{-5}	4.33×10^{-5}	4.33×10^{-5}	4.33×10^{-5}	4.33×10^{-5}	3.72×10^{-5}	3.72×10^{-5}
2	1.97×10^{-3}	1.97×10^{-3}	7.31	0.93×10^{-4}	0.93×10^{-4}	0.93×10^{-4}	1.22×10^{-4}	1.22×10^{-4}	1.22×10^{-4}	1.22×10^{-4}	1.14	1.14
3	1.44	1.44	2.16	1.06	1.06	1.27	1.22×10^{-4}	1.22×10^{-4}	1.22×10^{-4}	1.22×10^{-4}	1.55	1.55
4	1.21	1.21	1.49	2.37×10^{-3}	2.37×10^{-3}	2.37×10^{-3}	0.94	0.94	0.94	0.94	1.36	1.36
5	2.65	2.65	3.16	1.89	1.89	1.89	1.27	1.27	1.27	1.27	1.93	1.93

由表1可知，在机动方式1条件下，杆臂相关向量均不可观测，机动方式2、3、4、5状态下的杆臂相关状态向量的可观测度如图7所示。

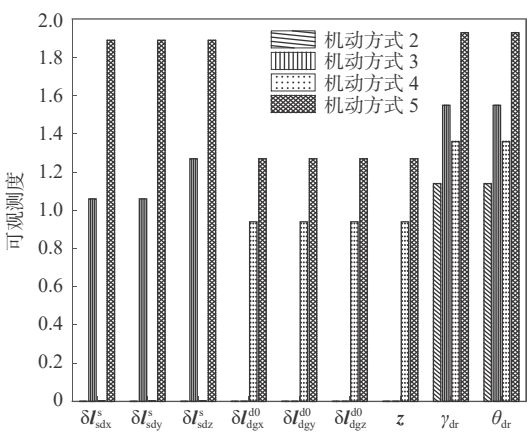


图7 不同机动条件下杆臂相关状态向量的可观测度

Fig. 7 The observable degree of the arm-related state vector under different maneuvering conditions

由表1和图7可知，在机动方式2条件下，只有吊舱振动倾覆角可观测，而内外杆臂向量可观测度极小，可认为其不可观测；在机动方式3条件下，吊舱内杆臂和振动倾覆角可观测，弹性外杆臂可观测度极小，可认为其不可观测；在机动方式4条件下，弹性外杆臂和振动倾覆角可观测，而吊舱内杆臂可观测度极小，可认为其不可观测；在机动方式5条件下，系统状态变量均可观测，且具有较高的可观测度。因此，在估计补偿吊舱弹性多杆臂误差时，应依据状态变量可观测度自适应调整机动方案，以满足吊舱组合导航系统状态变量可观。

3.3 仿真结果与分析

机载光电吊舱的飞行轨迹如图8和图9所示。

由图8和图9可以看出，在光电吊舱振动条件下仅补偿刚性多杆臂后，轨迹受吊舱振动的影响，与轨迹真值存在不容忽视的偏移，在载机变向时尤为明显；相比之下，弹性多杆臂补偿后的光电吊舱飞行轨迹位置精度有显著提升。

吊舱内、外杆臂估计结果如图10和图11所

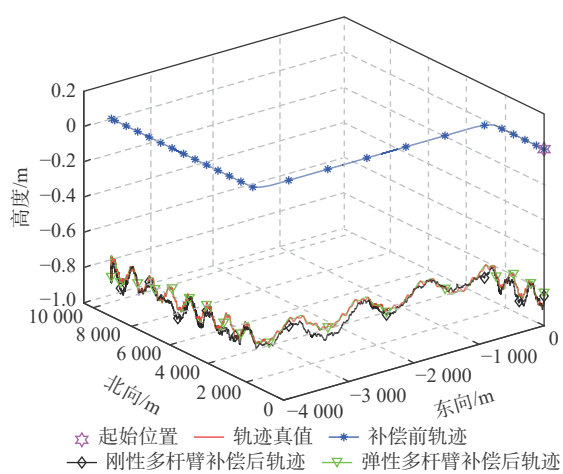


图8 仿真三维飞行轨迹

Fig. 8 Simulation of three-dimensional flight trajectory

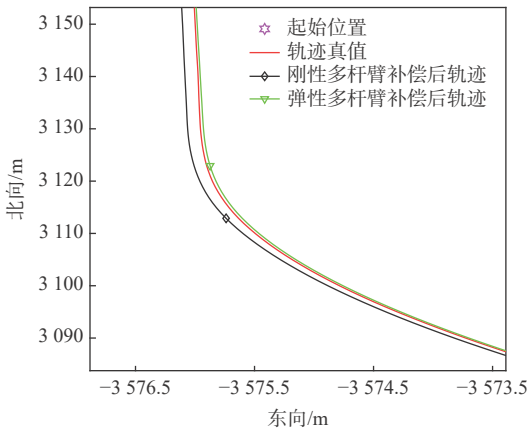


图 9　仿真二维飞行轨迹局部放大图

Fig. 9　Simulation of two-dimensional flight trajectory local amplification diagram

示, 其中, 内杆臂的 L_x, L_y, L_z 分别为内杆臂在子惯导坐标系 x, y, z 轴的投影, 外杆臂的 L_x, L_y, L_z 分别为外杆臂在光电吊舱坐标系 x, y, z 轴的投影。

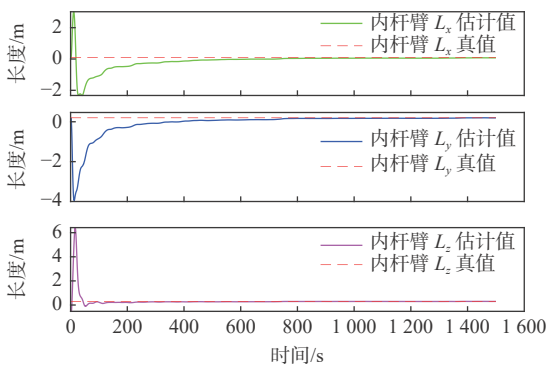


图 10　吊舱内杆臂估计结果

Fig. 10　Pod inner lever arm estimation results

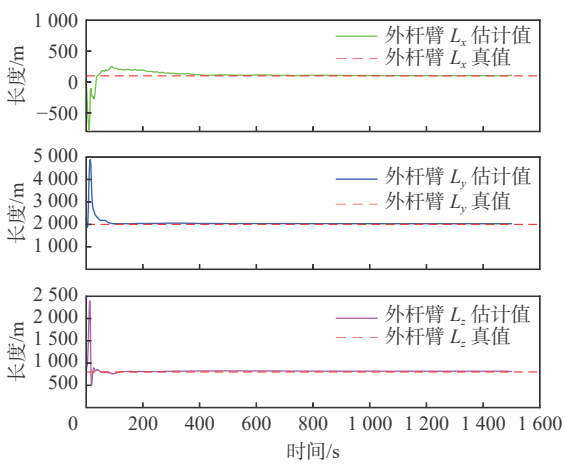


图 11　吊舱外杆臂估计结果

Fig. 11　Pod outer lever arm estimation results

由图 10 和图 11 可以看出, 随着载机机动和吊舱框架角的充分转动, 在 300 s 附近时, 吊舱内、外杆臂 3 个轴向分量估计结果均收敛至内、外杆臂设置真值附近, 并保持稳态, 经计算, 内杆臂估计差值比在 4% 以内, 外杆臂估计差值比在 3% 以内。

吊舱竖直振动位移估计误差和吊舱振动角估计误差如图 12 和图 13 所示。可以看出, 本文方法可有效估计吊舱竖直振动位移和振动倾覆角, 估计误差曲线收敛较快, 且收敛值稳定在理论 0 误差附近。

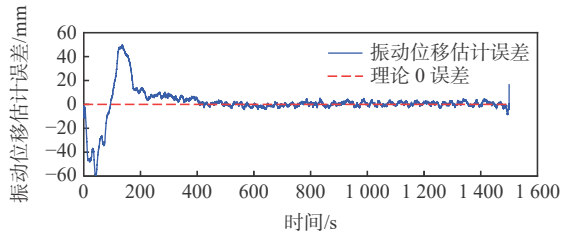


图 12　振动位移估计误差

Fig. 12　Vibration displacement estimation error

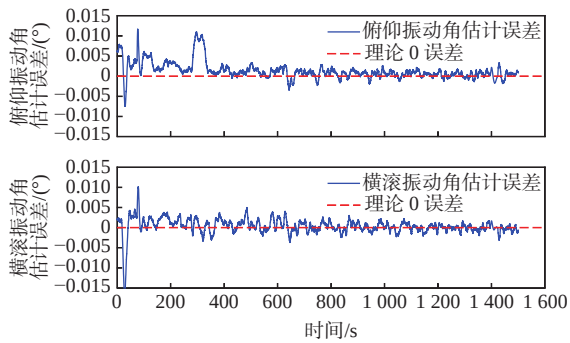


图 13　振动角估计误差

Fig. 13　Vibration angle estimation error

弹性多杆臂补偿后和刚性多杆臂补偿后吊舱速度误差对比如图 14～图 16 所示。

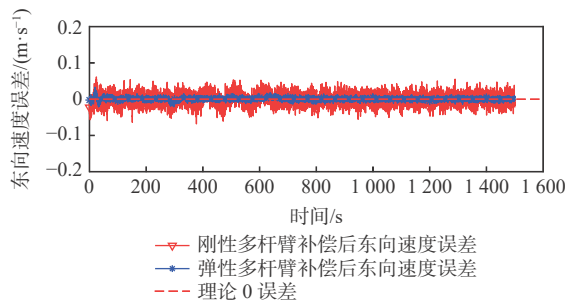


图 14　东向速度误差对比

Fig. 14　Comparison of east velocity error

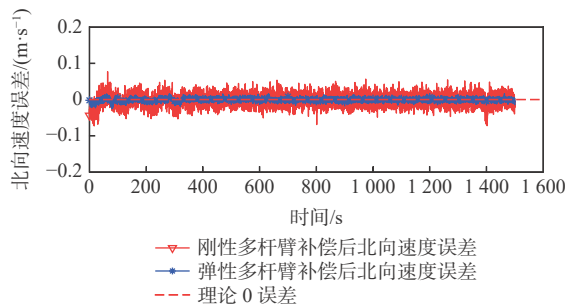


图 15　北向速度误差对比

Fig. 15　Comparison of north velocity error

由图 14～图 16 可以看出, 受吊舱振动的影响, 在刚性多杆臂误差补偿后, 3 个方向上的速度误差

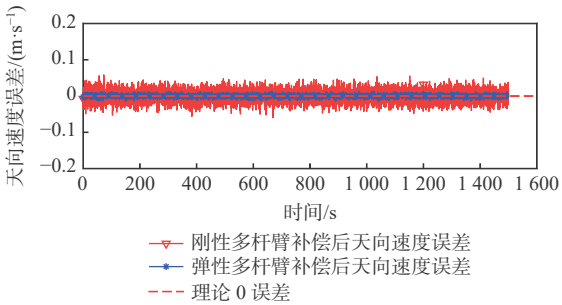


图 16 天向速度误差对比

Fig. 16 Comparison of radial velocity error

曲线仍处于持续波动或跳动状态，且幅值相对较大，而经过弹性多杆臂误差补偿后，3 个方向上的速度误差曲线波动或跳动幅值明显减小，且较快地收敛至理论 0 误差值附近，并趋于稳定。

弹性多杆臂补偿后和刚性多杆臂补偿后吊舱位置误差对比如图 17～图 19 所示。

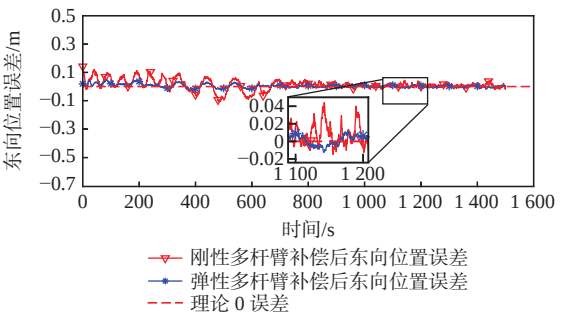


图 17 东向位置误差对比

Fig. 17 Comparison of east position error

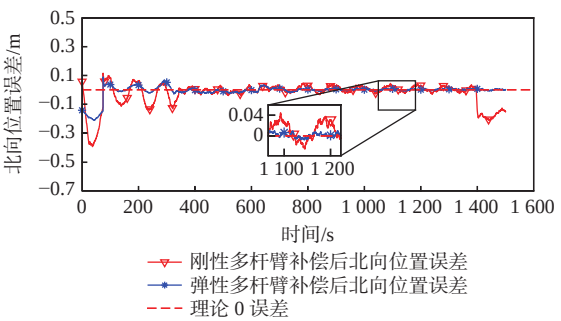


图 18 北向位置误差对比

Fig. 18 Comparison of north position error

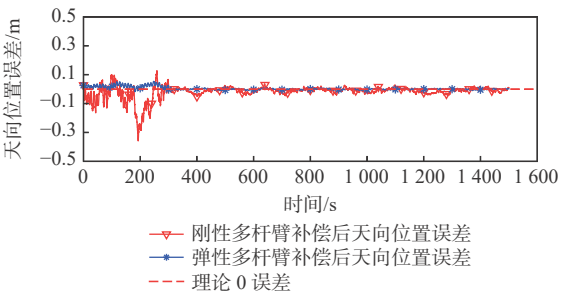


图 19 天向位置误差对比

Fig. 19 Comparison of radial position error

由图 17～图 19 可以看出，受吊舱振动的影响，在刚性多杆臂误差补偿后，3 个方向上的位置误差曲线仍存在较为明显的波动，而相较于刚性多杆臂补偿，弹性多杆臂补偿后，3 个方向上的位置误差曲线波动明显变小，且收敛值更接近理论 0 误差。

弹性多杆臂补偿后和刚性多杆臂补偿后，吊舱 3 个方向上的速度、位置误差的均方根误差如表 2、表 3 所示。可以看出，相较于刚性补偿，弹性补偿后的速度、位置误差的均方根误差均明显减小，其中，速度误差均方根误差在东、北、天 3 个方向上分别减少 67.6%、62.7%、73.4%，位置误差均方根误差在东、北、天 3 个方向上分别减少 67.1%、49.9%、77.4%，说明吊舱弹性振动导致的弹性多杆臂误差严重影响光电吊舱组合导航系统速度和位置解算精度，同时，也说明了弹性多杆臂误差补偿方法对提升光电吊舱组合导航系统速度和位置解算精度的有效性。

表 2 速度误差均方根误差

Table 2 The root mean square error of velocity error

数据类别	速度误差均方根误差/(m·s⁻¹)		
	东向	北向	天向
刚性补偿后	0.014 8	0.015 8	0.014 7
弹性补偿后	0.004 8	0.005 9	0.003 9

表 3 位置误差均方根误差

Table 3 The root mean square error of position error

数据类别	位置误差均方根误差/m		
	东向	北向	天向
刚性补偿后	0.038 9	0.085 8	0.050 0
弹性补偿后	0.012 8	0.043 0	0.011 3

4 结 论

本文针对光电吊舱组合导航系统中的弹性多杆臂误差问题，提出了弹性多杆臂误差估计补偿方法，显著提升了吊舱组合导航精度。

1）通过建立并分析吊舱振动动力学模型，得到吊舱受迫振动主要表现形式为质心位移和俯仰、横滚角振动，引起的弹性杆臂和姿态振动将影响组合导航系统速度、位置和姿态精度。

2）通过推导吊舱内、外杆臂误差及振动角误差，建立了吊舱多杆臂补偿滤波模型，并利用基于奇异值分解的可观测性分析方法对 5 个不同机动条件下的状态向量可观测度进行了定量分析，得到了系统可观测的充分条件，为仿真机动方案提供了自适应调整依据。

3）通过仿真验证了本文提出的弹性多杆臂误

差估计补偿方法能有效提升吊舱组合导航精度, 相较于传统刚性多杆臂补偿, 弹性多杆臂补偿后, 速度、位置误差显著降低, 其中, 速度误差均方根误差最大减少了 73.4%, 位置误差均方根误差最大减少了 77.4%。

参考文献（References）

[1] 梁卫清, 魏志强, 袁红伟, 等. 小型高性能无人机载光电吊舱的发展现状与方向[J]. 电视技术, 2022, 46(7): 65-68.
LIANG W Q, WEI Z Q, YUAN H W, et al. Development status and direction of small-scale high-performance UAV on-board optoelectronic pod[J]. Video Engineering, 2022, 46(7): 65-68(in Chinese).

[2] 高卓, 江泽, 邓麟. 机载光电吊舱目标定位技术研究[J]. 导航定位学报, 2013, 1(4): 74-78.
GAO Z, JIANG Z, DENG L. Research on airborne optoelectronic pod target location method[J]. Journal of Navigation and Positioning, 2013, 1(4): 74-78(in Chinese).

[3] 甄云卉, 路平. 无人机相关技术与发展趋势[J]. 兵工自动化, 2009, 28(1): 14-16.
ZHEN Y H, LU P. Related technologies and development of UAV[J]. Ordnance Industry Automation, 2009, 28(1): 14-16(in C-hinese).

[4] 杨小康, 严恭敏, 崔涛, 等. 用于机载光电吊舱的 MEMS-INS/GNS-S 组合导航系统杆臂误差精确建模与补偿方法[C]//中国惯性技术学会 2021 科技工作者研讨会论文集——前沿技术与惯性技术的融合与应用论文集. 大连: 中国惯性技术学会, 2021: 125-131.
YANG X K, YAN G M, CUI T, et al. Accurate modeling and compensation method of lever arm error of MEMS-INS/GNSS integrated navigation system for airborne photoelectric pod[C]//Proceedings of the Symposium on Dynamic Development Direction of Inertial Technology Development-Integration and Application of Frontier Technology and Inertial Technology. Dalian: Chinese Society of Inertial Technology, 2021: 125-131(in Chinese).

[5] 张兰霜. 某型光电吊舱的隔振研究[D]. 哈尔滨: 哈尔滨工业大学, 2021: 10-23.
ZHANG L S. Research on vibration isolation of a photoelectric pod[D]. Harbin: Harbin Institute of Technology, 2021: 10-23(in Chinese).

[6] CHEN J Y, ZHANG D L, HAN G T, et al. A method for lever arm estimation in INS/GPS integration using direct unscented Kalman filter[C]//Proceedings of the 2020 IEEE 6th International Conference on Computer and Communications. Piscataway: IEEE Press, 2020: 985-990.

[7] STOVNER B N, JOHANSEN T A. GNSS-antenna lever arm compensation in aided inertial navigation of UAVs[C]//Proceedings of the 2019 18th European Control Conference. Piscataway: IEEE Press, 2019: 4040-4046.

[8] HONG S, LEE M H, CHUN H H, et al. Experimental study on the estimation of lever arm in GPS/INS[J]. IEEE Transactions on Vehicular Technology, 2006, 55(2): 431-448.

[9] 徐景硕, 王勇军, 刘亚. 大失准角情形下 UKF 与 CKF 的比较研究[J]. 电光与控制, 2017, 24(9): 42-46.
XU J S, WANG Y J, LIU Y. Comparative study on UKF and CKF in large azimuth misalignment for SINS[J]. Electronics Optics & Control, 2017, 24(9): 42-46 (in Chinese).

[10] CONG M Y, LU H, CHENG X, et al. Transfer alignment method and realization of SINS on moving base based on Kalman filter[C]//Proceedings of the 2019 IEEE 1st International Conference on Civil Aviation Safety and Information Technology. Piscataway: IEEE Press, 2020: 329-333.

[11] WANG B, YE W, LIU Y. An improved real-time transfer alignment algorithm based on adaptive noise estimation for distributed POS[J]. IEEE Access, 2020, 8: 102119-102127.

[12] CAO Q, ZHONG M, ZHAO Y. Dynamic lever arm compensation of SINS/GPS integrated system for aerial mapping[J]. Measurement, 2015, 60: 39-49.

[13] 翁浚, 刘健宁, 寇科, 等. 吊舱 SINS/GNSS 组合导航多杆臂效应在线估计算法[J]. 中国惯性技术学报, 2021, 29(2): 184-190.
WENG J, LIU J N, KOU K, et al. Online estimation algorithm of multi-lever arm effect for integrated pod SINS/GNSS navigation[J]. Journal of Chinese Inertial Technology, 2021, 29(2): 184-190(in Chinese).

[14] 徐晓苏, 邹海军, 刘义亭, 等. 基于鲁棒滤波的挠曲变形和动态杆臂补偿算法[J]. 中国惯性技术学报, 2015(1): 9-13.
XU X S, ZOU H J, LIU Y T, et al. Compensation algorithm of flexural deformation and dynamic lever-arm based on robust filtering [J]. Journal of Chinese Inertial of Technology, 2015(1): 9-13(in Chinese).

[15] GAO Q, ZHAO G, WANG X. Transfer alignment error compensator design for flexure and lever-arm effect[C]//Proceedings of the IEEE Conference on Industrial Electronics and Applications. Xi'an: ICIEA, 2009.

[16] CAO Q, ZHONG M, GUO J. Non-linear estimation of the flexural lever arm for transfer alignment of airborne distributed position and orientation system[J]. IET Radar, Sonar & Navigation, 2017, 11(1): 41-51.

[17] 谷雨, 司帆, 赵剡, 等. 一种改进的机载武器传递对准中杆臂效应动态补偿方法[J]. 弹箭与制导学报, 2018, 38(1): 41-44.
GU Y, SI F, ZHAO Y, et al. An improved dynamic compensation method for lever arm effect in transfer alignment of airborne weapon[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2018, 38(1): 41-44(in Chinese).

[18] 严恭敏, 翁浚. 捷联惯导算法与组合导航原理[M]. 西安: 西北工业大学出版社, 2019: 204-207.
YAN G M, WENG J. Strapdown inertial navigation algorithm and integrated navigation principle[M]. Xi'an: Northwestern Polytechnical University Press, 2019: 204-207(in Chinese).

Pod integrated navigation system elastic lever arm error compensation method

WANG Haipeng[1], LIU Haina[2], WANG Zhendong[3], MIAO Cunxiao[1,*], ZHANG He[1], YE Wen[4]

(1. School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China;
2. Beijing Aerospace Control Device Institute, Beijing 100048, China;
3. AVIC Chengdu Aircraft Industrial (Group) Co. Ltd., Chengdu 610073, China;
4. National Institute of Metrology, Beijing 100029, China)

Abstract: Aiming at the problem of the dynamic elastic lever arm effect caused by the forced vibration of the airborne photoelectric pod and the dynamic change of frame angle, an integrated navigation filtering method for online compensation of elastic multi lever arm error is proposed. Firstly, the dynamic model of the linear and angular vibrations of the pod as well as the elastic outer arm model induced by the vibration of the pod are established by examining the primary forms of the forced vibration of the pod. Secondly, based on the multi-arm error model, the velocity, position error and vibration angle error caused by the vibration of the pod are deduced. The elastic outer arm error model and the vibration angle error model are established and introduced into the state vector. The required requirements for each state vector's observability are found by examining the observability of the system state vectors under various maneuvering states. This provides an adaptive adjustment foundation based on the observability for the planning maneuver scheme. The simulation results show that the root mean square error of the velocity error is reduced by 73.4% and the root mean square error of the position error is reduced by 77.4% after the elastic multi-arm error compensation, compared with the rigid arm error compensation, which effectively improves the accuracy of the pod integrated navigation.

Keywords: photoelectric pod; forced vibration; lever arm error; Kalman filtering; observability analysis

Received: 2024-03-19; **Accepted**: 2024-05-09; **Published Online**: 2024-06-03 14:35
URL: link.cnki.net/urlid/11.2625.V.20240531.1831.001
Foundation item: National Natural Science Foundation of China (61603035)
* **Corresponding author**. E-mail: miaocx@ustb.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0164

机载中空纤维膜分离特性的预测分析

柴婷[1]，付子祺[1]，刘冠男[2]，白文涛[3]，胡万俊[4]，冯诗愚[1,*]

(1. 南京航空航天大学 航空学院，南京 210016；2. 西安交通大学 化学工程与技术学院，西安 710049；3. 中航工业南京机电科技有限公司产品研发部，南京 210000；4. 四川泛华航空仪表电器有限公司质量管理部，成都 610599)

摘 要：机载中空纤维膜分离性能实验周期长，操作工况复杂，有必要研究使用较少的数据进行膜性能的精准预测。因此，建立了一种基于人工神经网络的中空纤维膜分离性能预测模型，搭建误差反向传播 (BP) 神经网络、Elman 神经网络和遗传算法优化的 BP(GA-BP) 神经网络，针对机载中空纤维膜分离性能随引气压力、飞行高度和引气流量的变化规律数据，对比预测不同算法下中空纤维膜的性能预测情况。研究结果表明：搭建的 3 种神经网络算法均能精准预测中空纤维膜的分离特性，其平均百分比误差 (MAPE) 均小于 1%；分别选取全体数据集的 5%、10% 和 20%，GA-BP 神经网络的预测结果最佳，其预测准确率比其他 2 种神经网络平均高 1.43%。基于研究结果，在实际工作中，可以选择合适的算法进行膜实验数据的处理与性能的高效分析。

关 键 词：机载油箱惰化系统；中空纤维膜；分离性能；人工神经网络；预测
中图分类号：V228; TQ021.4
文献标志码：A **文章编号**：1001-5965(2026)05-1739-07

飞机在飞行过程中，燃油箱上部空间充满了易燃易爆的气体混合物，遇到火花时，极易发生爆炸事故，严重威胁飞机飞行安全。为降低油箱发生燃爆的概率，必须采取有效措施来抑制油箱燃爆。研究表明[1]，目前最为有效的措施是机载油箱惰化技术，即采用机载空气分离技术制取惰性气体，以控制飞机油箱上部空间的氧体积分数，使之满足适航状态。通常，可以采用分子筛吸附和中空纤维膜分离 2 类技术来制取惰性气体。其中，中空纤维膜分离技术凭借其可靠性高、制氮效果好、经济性好等特点，已成为油箱惰化技术中的首选方案。

近年来，国内外很多学者针对中空纤维膜已经开展了大量理论研究和实验工作，并取得了一定成果。在理论研究方面，贺高红等[2]对气体分离膜丝内压降规律进行了研究，验证了丝内压降服从微分 Hagen-Poiseuille 方程，为后续建立机载中空纤维膜分离的数学模型奠定了坚实的基础。冯诗愚等[3]采用微元方法建立了机载制氮系统中空纤维膜数学模型，对中空纤维膜分离特性进行了研究。蔡琰等[4]在其研究成果基础上，考虑浓差极化问题对膜性能的影响，对中空纤维膜模型进行数学分析，并建立微分方程，满足了中空纤维膜更进一步研究的需要。Yang 等[5]采用耦合数值迭代解法建立了温度压力控制分系统和中空纤维膜的数学模型，对比在温度和压力控制与不控制的情况下，整个飞行包线中空纤维膜设计面积、产氮效率和氧气体积分数的变化情况。耿雷铭等[6]以N_2/O_2分离为研究对象，对中空纤维膜极化系数影响因素进行了分析。在实验研究方面，薛勇等[7]采用大气环境模拟舱和地面试验平台，对机载膜制氮系统的分离性能进行了全面实验研究。刘小芳等[8]通过建立机载中空纤维膜环境模拟性能试验平台，对机载中空纤维膜进行

收稿日期：2024-03-25；**录用日期**：2024-05-03；**网络出版时间**：2024-07-25 14:33
网络出版地址：link.cnki.net/urlid/11.2625.V.20240725.0911.002
基金项目：国家自然科学基金 (52372331)；中央高校基本科研业务费专项资金；江苏高校优势学科建设工程；南京航空航天大学研究生科研与实践创新计划项目 (xcxjh20230103)
* **通信作者**. E-mail：shiyuf@nuaa.edu.cn

引用格式：柴婷，付子祺，刘冠男，等. 机载中空纤维膜分离特性的预测分析 [J]. 北京航空航天大学学报，2026，52（5）：1739-1745.
CHAI T，FU Z Q，LIU G N，et al. Prediction and analysis of separation performance of on-board hollow fiber membrane [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（5）：1739-1745（in Chinese）.

了系统的环境模拟性能试验研究和理论分析。Burns等[9]在A320上对飞行包线下的中空纤维膜性能进行了试验研究。美国联邦航空管理局(Federal Aviation Administration, FAA)与航空公司合作,在B737、B747和A320飞机上安装由渗透纤维膜空气分离器和其他设备组成的系统,进行飞行试验,在飞行测试成功后,FAA颁布一系列修正案和适航条例,提高油箱安全性[10-11]。

由于中空纤维膜非理想化现象非常严重,单纯依靠数学建模理论分析,与实际情况有较大差别。在现阶段的研究中,通过实验方法获取中空纤维膜的性能仍是最主要且常见的方法,但因为膜的分离性能受多种变量的影响,如引气温度、引气压力和引气流量等,操作工况复杂,如果单纯依靠实验来获取每个变量工况下的分离性能参数,实验周期长,时间成本高。因此,如何用较少的实验数据预测对应工况下的膜分离性能,从而节省时间,降低实验成本,是一项值得深入研究的课题。

人工神经网络凭借其强大的自学习能力,可以实现预测功能,在航空航天领域及机载燃油系统有着广泛的应用。李振环等[12]提出了一种基于径向基神经网络的递归预测模型,实现了对航空发动机空气系统盘腔瞬态壁温的动态预测。薛勇[13]利用反向传播(back propagation, BP)神经网络实现了对机载中空纤维膜的性能预测。这些研究都表明人工神经网络具有较高的预测精度,因此,可以进一步研究不同神经网络模型的预测效果,并在训练集数据样本的数量不同时,对比分析预测算法模型的准确率。

本文介绍3种不同的神经网络算法,分别搭建中空纤维膜分离性能预测模型,通过与数值结果的对比,验证3种预测模型的准确性。在数据样本较少时,对比不同算法的预测精度。对于不同数据集,选择合适的算法进行中空纤维膜的性能预测具有一定意义。

1 人工神经网络预测模型建立

1.1 样本数据选择

针对中空纤维膜分离性能,通过仿真软件计算出当引气温度在80℃,引气压力分别控制在0.3、0.35、0.4、0.45、0.5 MPa,飞行高度控制在0、10、15 km时的共1 020组数据。

在进行预测训练时,为避免训练样本数据过于集中或过于规律,从而影响网络性能,采用随机选取的方法,在1 020组数据中随机选择100组数据作为训练样本,测试样本由剩余920组数据中随机抽取出的20组数据组成。

1.2 BP神经网络模型建立

1.2.1 建模方法

研究构造一个3层BP神经网络,由于在实际工作中,机载中空纤维膜分离的引气温度变化范围很小,在本次预测中,输入层只设计3个参数:引气压力、飞行高度和引气流量;输出层设计2个参数:富氮气体氧体积分数$x_{\mathrm{O,NEA}}$和制氮效率η。

其中,制氮效率定义为

$$\eta=\frac{m_{\mathrm{NEA}}}{m_{\mathrm{air}}} \tag{1}$$

式中:m_{NEA}、m_{air}分别为富氮气体流量、引气流量,单位均为kg/h。

富氮气体氧体积分数定义为

$$x_{\mathrm{O,NEA}}=\frac{n_{\mathrm{O,NEA}}}{n_{\mathrm{O,NEA}}+n_{\mathrm{N,NEA}}} \tag{2}$$

式中:$n_{\mathrm{O,NEA}}$、$n_{\mathrm{N,NEA}}$分别为富氮气体氧摩尔流量、富氮气体氮摩尔流量,单位均为mol/s。

基于Matlab语言,采用Levenberg-Marquardt算法对BP神经网络进行训练,均方误差为误差函数。

1.2.2 参数预处理

神经网络隐含层采用的传递函数为Tan-Sigmoid函数:

$$f(x)=\frac{\mathrm{e}^{x}-\mathrm{e}^{-x}}{\mathrm{e}^{x}+\mathrm{e}^{-x}} \tag{3}$$

Tan-Sigmoid函数是双曲正切函数,其特点是$(-\infty,+\infty)$范围的数据被映射到区间$(-1,1)$内,即对输入数据进行归一化处理[14]。

输出层的函数采用purelin函数:

$$f(x)=x \tag{4}$$

1.2.3 隐含层节点数设计

当隐含层节点数过少时,会导致网络学习能力和信息处理能力不足,出现“欠拟合”,而当节点数过多时,不仅会导致网络结构复杂化,从而增加网络的训练时间,而且会使局部极小点在学习过程中更易出现“过拟合”的倾向。显然,选择合适的隐含层节点数至关重要。因此,建立算法模型首先要改变隐含层节点数,并保证模型内其他参数不变,采用控制单一变量的方法进行计算,根据计算结果确定最佳隐含层节点数。

最佳隐含层节点数的确定公式[15]为

$$b=\sqrt{q+p}+a \tag{5}$$

式中:b为隐含层节点数;q为输出层节点数;p为输入层节点数;a的取值范围为[1,10]。

即依次对BP神经网络设置10个隐含层节点,并计算训练集均方误差(mean squared error, MSE),MSE最小的节点数即最佳隐含层节点数。由表1

表 1　隐含层节点数确定结果

Table 1　Result of determining the number of hidden layer nodes

隐含层节点数	MSE
3	$2.390\,2\times10^{-3}$
4	$4.271\,6\times10^{-2}$
5	$2.924\,6\times10^{-4}$
6	$7.785\,1\times10^{-4}$
7	$2.052\,9\times10^{-2}$
8	$9.288\,6\times10^{-5}$
9	$1.471\,4\times10^{-2}$
10	$8.544\,9\times10^{-5}$
11	$1.022\,9\times10^{-3}$
12	$1.450\,9\times10^{-4}$

可知, 最佳隐含层节点数为 10 个。

1.2.4　精度评价指标

本文评价测试集与模型计算结果的误差主要有 4 个指标[16]: 平均绝对误差(mean absolute error, MAE)、MSE、均方根误差(root mean square error, RMSE)和平均绝对百分比误差(mean absolute percentage error, MAPE)。

$$M_{\mathrm{AE}}=\frac{1}{m}\sum_{i=1}^{m}\left|h(x^{(i)})-y^{(i)}\right| \tag{6}$$

$$M_{\mathrm{SE}}=\frac{1}{m}\sum_{i=1}^{m}(h(x^{(i)})-y^{(i)})^2 \tag{7}$$

$$R_{\mathrm{MSE}}=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(h(x^{(i)})-y^{(i)})^2} \tag{8}$$

$$M_{\mathrm{APE}}=\frac{100\%}{m}\sum_{i=1}^{m}\left|\frac{h(x^{(i)})-y^{(i)}}{y^{(i)}}\right| \tag{9}$$

式中: M_{AE}、M_{SE}、R_{MSE}、M_{APE}分别为平均绝对误差、均方误差、均方根误差、平均绝对百分比误差; m为测试集样本数量; $h(x^{(i)})$为测试集第i组的模型计算结果; $y^{(i)}$为该组数据的真实结果。

1.2.5　预测结果

如图 1 所示, 测试集标准值与模型计算预测结果相差不大, 富氮气体氧体积分数的 MAPE 为 0.98%, 即预测准确率为 99.02%; 制氮效率的 MAPE 为 0.82%, 预测准确率为 99.18%。这说明 BP 神经网络预测模型的准确率较高。

1.3　Elman 神经网络模型建立

1.3.1　建模方法

Elman 神经网络是一种典型的动态递归神经网络, 在 BP 神经网络隐含层基础上增加一个承接层, 作为一步延时算子, 达到记忆的目的, 从而增强网络的全局稳定性。

为达到对比效果, Elman 神经网络的输入参数、输出参数、训练样本、训练算法、传递函数及隐含层设计都与 BP 神经网络相同, Elman 神经网络的承接层与隐含层节点数相同。

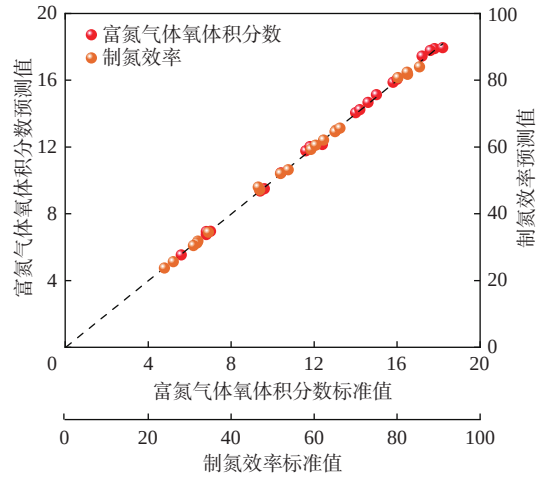


图 1　BP 神经网络模型预测结果与测试集结果比较

Fig. 1　Comparison of BP neural network between model prediction results and test set results

1.3.2　预测结果

如图 2 所示, 测试集标准值与模型计算预测值的数值非常接近, 富氮气体氧体积分数的 MAPE 为 0.17%, 即预测准确率为 99.83%; 制氮效率的 MAPE 为 0.22%, 预测准确率为 99.78%。与 BP 神经网络相比, 富氮气体氧体积分数和制氮效率的 MAPE 均明显降低。这说明 Elman 神经网络比 BP 神经网络的预测准确率更高。

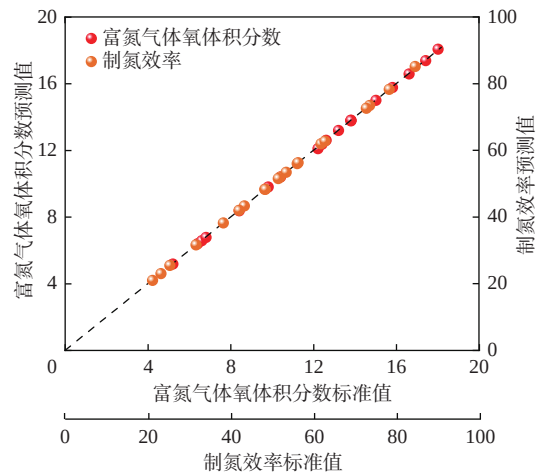


图 2　Elman 神经网络模型预测结果与测试集结果比较

Fig. 2　Comparison of Elman neural network between model prediction results and test set results

1.4　GA-BP 神经网络模型建立

遗传算法(genetic algorithm, GA)是受生物遗传学中自然选择、优胜劣汰机制启发的一种算法, 模拟生物和人类进化的方式来求解复杂问题。

BP 神经网络算法通过给定初始的权值和阈值, 前向传播数据与误差反向传递来更新权重阈值。由于随机初始的参数未必是最优的起点, 会对训练模型的可靠性和稳定性产生较大的干扰[15]。而遗传算法优化的 BP(genetic algorithms back propagation,

GA-BP)神经网络具体很好的全局搜索能力,通过GA优化出最佳的初始权值和阈值,使训练准确,预测结果可靠。

1.4.1 建模方法

GA-BP神经网络的实现过程可以分为3个部分:BP神经网络结构确定、GA优化权值和阈值、BP神经网络训练及预测。其中,BP神经网络结构确定与1.2节步骤相同,训练算法选择Levenberg-Marquardt算法。

GA通过计算适应度值,即MSE来优化权值和阈值,通过选择、交叉、变异操作寻找最优个体。其中,选择算子采用轮盘赌算法,交叉算子采用两点交叉算子,变异算子采用高斯变异算子。其他参数设置见表2。

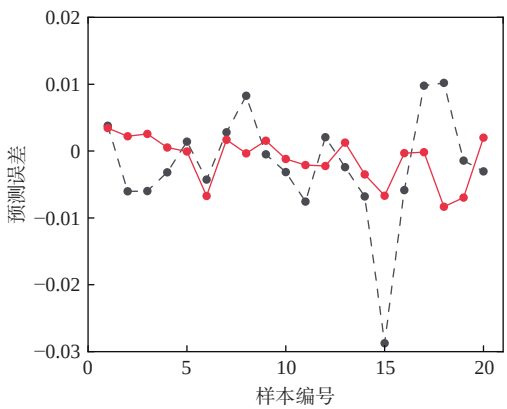
表2 遗传算法参数设置

Table 2 Genetic algorithm parameter settings

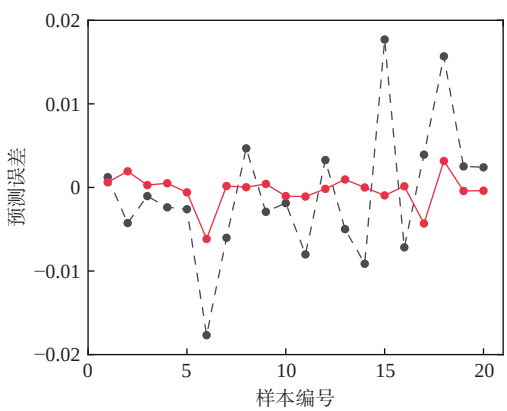
种群规模	最大迭代次数	交叉概率	变异概率	各基因取值范围
30	30	0.8	0.2	(−3, 3)

1.4.2 预测结果

图3给出了GA-BP神经网络模型下的预测误差与不使用GA优化的标准BP神经网络的预测误差对比。



(a) 富氮气体氧体积分数的预测误差对比



(b) 制氮效率的预测误差对比



图3 2种神经网络预测值与实际值误差对比

Fig. 3 Comparison of error between two neural network predictions and actual values

使用标准BP神经网络时,测试集富氮气体氧体积分数的MAPE为0.58%,制氮效率的MAPE为0.60%。使用GA-BP神经网络时,测试集富氮气体氧体积分数的MAPE为0.27%,制氮效率的MAPE为0.12%。

由图3可见,GA-BP神经网络较传统BP神经网络,误差显著减小,这反映了GA-BP神经网络的优化效果显著,预测效果比标准BP神经网络更精准。

2 人工神经网络预测模型对比

由第1节可知,3种预测模型对于中空纤维膜的分离性能均有良好的预测准确性。本节探讨对比3种预测模型分别用较少的数据获得较高预测算法模型的准确率。

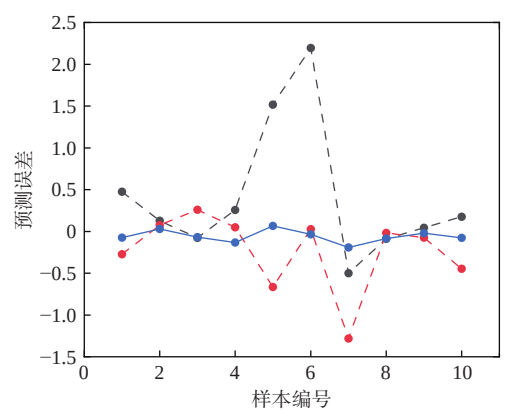
2.1 数据集的5%

2.1.1 训练样本选择

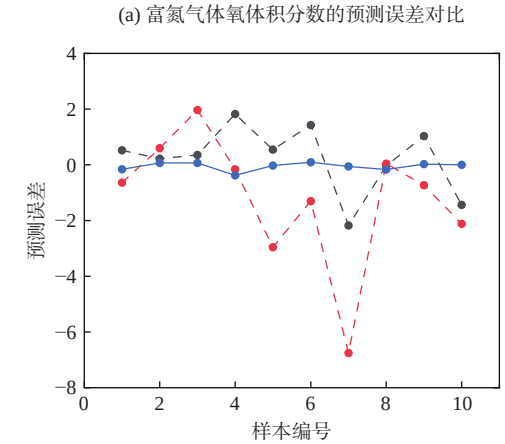
采用随机选取的原则从全体数据集提取5%,即50组数据进行训练,并从剩下的数据集中随机选择10组数据组成测试集。

2.1.2 预测结果

由图4可知,在5%数据集时,GA-BP神经网络



(a) 富氮气体氧体积分数的预测误差对比



(b) 制氮效率的预测误差对比

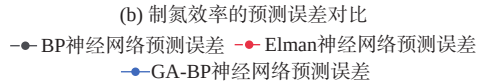


图4 5%数据集预测误差对比

Fig. 4 Comparison of prediction errors on a 5% dataset

的预测效果最好, 预测误差最小, 也具有较好的稳定性。其中, 标准 BP 神经网络的富氮气体氧体积分数的预测准确率为 95.53%, 制氮效率的预测准确率为 97.64%。Elman 神经网络的富氮气体氧体积分数的预测准确率为 96.39%, 制氮效率的预测准确率为 95.53%。GA-BP 神经网络的预测准确率最高, 富氮气体氧体积分数准确率为 99.17%, 制氮效率的准确率为 99.77%。对比 1.2 节与 1.3 节, 当训练集样本数量较少时, 标准 BP 神经网络和 Elman 神经网络的预测效果会明显下降, 这不仅反映了数据量对预测结果有很大影响, 也反映了 BP 神经网络和 Elman 神经网络的不稳定性。

2.2　数据集的 10%

2.2.1　训练样本选择

同 5% 数据集相同, 采用随机选取的原则从全体数据集提取 10%, 即 100 组数据进行训练, 并从剩下的数据集中随机选择 10 组数据组成测试集。

2.2.2　预测结果

由图 5 可知, 在 10% 数据集时, GA-BP 神经网络的预测效果最好, 预测误差最小。其富氮气体氧体积分数的预测准确率为 99.73%, 制氮效率的预测准确率为 99.88%。Elman 神经网络的富氮气体氧体积分数的预测准确率为 99.24%, 制氮效率的预测准确率为 99.24%。标准 BP 神经网络的预测效果最差, 富氮气体氧体积分数的预测准确率为 98.94%, 制氮效率的预测准确率为 98.65%。

对比 5% 数据集和 10% 数据集发现, 当数据量增多时, 每种预测模型的预测准确率均会升高。

2.3　数据集的 20%

2.3.1　训练样本选择

采用随机选取的原则从全体数据集提取 20%, 即 200 组数据进行训练, 并从剩下的数据集中随机选择 10 组数据组成测试集。

2.3.2　预测结果

由图 6 可知, 在 20% 数据集时, 3 种神经网络都有较好的预测效果, 其中, Elman 神经网络与 GA-BP 神经网络的预测效果不相上下。标准 BP 神经网络的富氮气体氧体积分数的预测准确率为 99.33%, 制氮效率的预测准确率为 99.41%。Elman 神经网络的富氮气体氧体积分数的预测准确率为 99.93%, 制氮效率的预测准确率为 99.87%。GA-BP 网络的富氮气体氧体积分数的预测准确率为 99.93%, 制氮

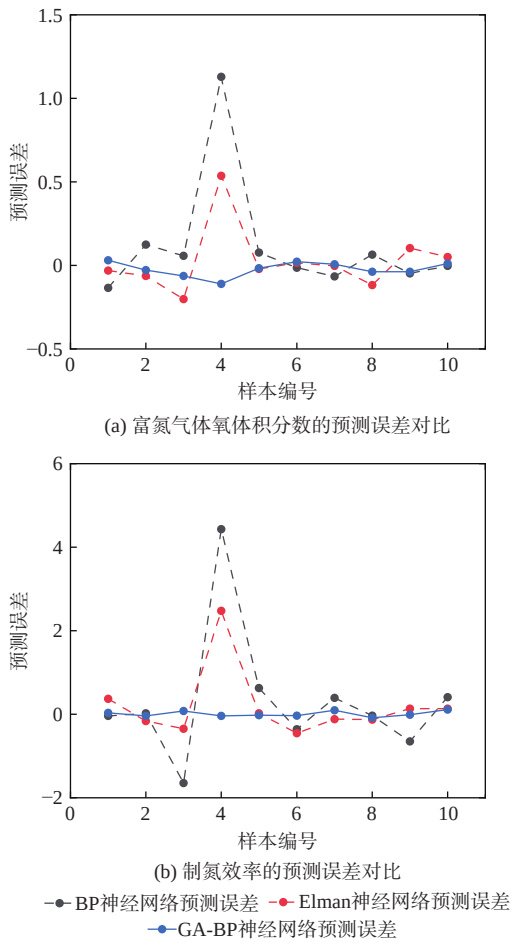


图 5　10% 数据集预测误差对比

Fig. 5　Comparison of prediction errors on a 10% dataset

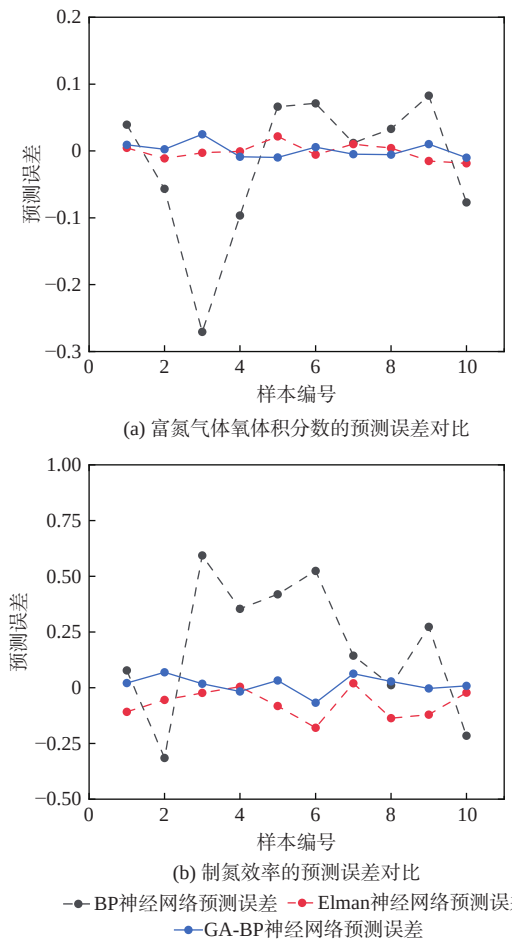


图 6　20% 数据集预测误差对比

Fig. 6　Comparison of prediction errors on a 20% dataset

效率的预测准确率为99.94%。

在20%数据集时，3种神经网络的预测准确率均能达到99%以上，且对比10%数据集，每种预测模型的预测准确率均会升高。

3 结 论

本文针对机载中空纤维膜的分离特性，应用3种神经网络预测技术分别预测分离性能。研究结果表明：

1）标准BP神经网络、Elman神经网络和GA-BP神经网络均能较好地预测中空纤维膜的分离特性，预测准确率均能达到99%以上。

2）在研究数据量不同时，数据量越多，预测准确率越高，可见，提高网络算法的预测精度可以通过增加训练集数据量来实现。从5%数据集、10%数据集及20%数据集的综合表现来看，无论数据量多少，GA-BP神经网络总有较好的预测效果，其预测准确率相比于其他2种神经网络平均高1.43%。

3）膜分离性能分析实验的工况复杂性导致实验周期长、时间成本高。本文建立了中空纤维膜分离性能预测模型，为不同工况环境下的数据分析提供了模型基础，能够快速高效地向数据集提供算法选择的支持。

参考文献（References）

[1] 刘卫华, 冯诗愚. 飞机燃油箱惰化技术[M]. 北京: 科学出版社, 2018: 2-3.
LIU W H, FENG S Y. Aircraft fuel tank inerting technology[M]. Beijing: Science Press, 2018: 2-3(in Chinese).

[2] 贺高红, 邹越, 徐仁贤, 等. 中空纤维气体分离膜丝内压降规律的研究[J]. 膜科学与技术, 1994, 14(2): 22-28.
HE G H, ZOU Y, XU R X , et al. A study on pressure drop in hollow-fiber membrane for gas separation[J]. Membrane Science and Technology, 1994, 14(2): 22-28(in Chinese).

[3] 冯诗愚, 卢吉, 刘卫华, 等. 机载制氮系统中空纤维膜分离特性[J]. 航空动力学报, 2012, 27(6): 1332-1339.
FENG S Y, LU J, LIU W H, et al. Separation performance of hollow fiber membrane for on-board inerting gas generating system[J]. Journal of Aerospace Power, 2012, 27(6): 1332-1339(in Chinese).

[4] 蔡琰, 林贵平, 曾宇, 等. 中空纤维膜机载制氮装置的数学建模分析[J]. 航空动力学报, 2015, 30(9): 2100-2107.
CAI Y, LIN G P, ZENG Y, et al. Mathematical modeling analysis of hollow fiber membrane onboard inert gas generation system[J]. Journal of Aerospace Power, 2015, 30(9): 2100-2107(in Chinese).

[5] YANG J Y, TAI B T, CHAO W C, et al. Analysis of the influence of on-board temperature and pressure control system on inert gas generating performance of hollow fiber membrane[J]. IOP Conference Series: Materials Science and Engineering, 2020, 751(1): 012048.

[6] 耿雷铭, 张瑞华, 刘卫华. 中空纤维膜极化系数影响因素分析[J]. 航空动力学报, 2025, 40(3): 14-20.
GENG L M, ZHANG R H, LIU W H. Influencing factors of polarization coefficient of hollow fiber membrane[J]. Journal of Aerospace Power, 2025, 40(3): 14-20(in Chinese).

[7] 薛勇, 刘卫华, 高秀峰, 等. 机载惰化系统中空纤维膜分离性能的实验研究[J]. 西安交通大学学报, 2011, 45(3): 107-111.
XUE Y, LIU W H, GAO X F, et al. Experimental study on separation performance of hollow fiber membrane for onboard inert gas generating system[J]. Journal of Xi'an Jiaotong University, 2011, 45(3): 107-111(in Chinese).

[8] 刘小芳, 刘卫华, 钱国诚, 等. 机载中空纤维膜富氮性能实验[J]. 航空动力学报, 2012, 27(5): 975-980.
LIU X F, LIU W H, QIAN G C, et al. Experimentation on nitrogen-enriched characteristics of on-board hollow fibre membrane[J]. Journal of Aerospace Power, 2012, 27(5): 975-980(in Chinese).

[9] BURNS M, CAVAGE W M, HILL R, et al. Flight-testing of the FAA onboard inert gas generation system on an airbus A320: DOT/FAA/AR-03/58[R]. New Jersey: FAA, 2004.

[10] MICHAEL B. Evaluation of fuel tank flammability and the FAA inerting system on the NASA 747 SCA: DOT/FAA/AR-04/41[R]. New Jersey: FAA, 2004.

[11] MICHAEL B, WILLIAM M C, RICHARD H, et al. Flight-testing of the FAA onboard inert gas generation system on an airbus A320: DOT/FAA/AR- 03/58[R]. New Jersey: FAA, 2004.

[12] 李振环, 王海, 丁小飞, 等. 基于递归径向基神经网络的航空发动机盘腔瞬态壁温预测[J]. 推进技术, 2023, 44(12): 187-195.
LI Z H, WANG H, DING X F, et al. Transient wall temperature prediction of aero-engine disk cavity based on recurrent radial basis function neural network[J]. Journal of Propulsion Technology, 2023, 44(12): 187-195.

[13] 薛勇. 机载中空纤维膜分离性能及民机燃油箱冲洗惰化研究[D]. 南京: 南京航空航天大学, 2010.
XUE Y. Study on separation performance of onboard hollow fiber membrane and flush inerting for fuel tank of civil aircraft[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2010(in Chinese).

[14] 倪志伟. BP网络中激活函数的深入研究[J]. 安徽大学学报 (自然科学版), 1997, 21(3): 48-51.
NI Z W. Deep study on activation function in BP network[J]. Journal of Anhui University (Natural Sciences), 1997, 21(3): 48-51(in Chinese).

[15] 张学锋, 朱梅, 汤亚玲, 等. 基于GA-BP的烧结风箱阀门预测模型[J]. 西华大学学报 (自然科学版), 2025, 44(2): 113-119.
ZHANG X F, ZHU M, TANG Y L, et al. GA-BP based sintering airbox valve prediction model[J]. Journal of Xihua University (Natural Science Edition), 2025, 44(2): 113-119(in Chinese).

[16] 陈亚琴, 彭浩, 冯诗愚. 基于人工神经网络的微重力流动冷凝换热预测[J]. 南京航空航天大学学报, 2021, 53(6): 989-995.
CHEN Y Q, PENG H, FENG S Y. Prediction of flow condensation heat transfer under microgravity flow based on artificial neural network[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2021, 53(6): 989-995(in Chinese).

Prediction and analysis of separation performance of on-board hollow fiber membrane

CHAI Ting[1], FU Ziqi[1], LIU Guannan[2], BAI Wentao[3], HU Wanjun[4], FENG Shiyu[1,*]

(1. School of Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;
2. School of Chemical Engineering and Technology, Xi'an Jiaotong University, Xi'an 710049, China;
3. Product Research and Development Department of AVIC Nanjing Electromechanical Technology Co., Ltd., Nanjing 210000, China;
4. Quality Management Department of Sichuan Fanhua Aviation Instrument Electric Appliance Co., Ltd., Chengdu 610599, China)

Abstract: Due to the long experimental period and complex operating conditions, it is necessary to study the precise prediction of membrane performance using limited data in airborne hollow fiber membrane separation performance experiments. An artificial neural network-based prediction model for hollow fiber membrane separation performance has been developed. Error back propagation (BP) neural networks, Elman neural networks, and genetic algorithm optimized BP (GA-BP) neural networks have been constructed to investigate the variation of airborne hollow fiber membrane separation performance with bleed pressure, flight altitude, and bleed flow rate. Compare and predict the performance of hollow fiber membranes under different algorithms. The research results indicate that the three network algorithms constructed can accurately predict the separation characteristics of hollow fiber membranes, with an mean absolute percentage error (MAPE) of less than 1%. With an average prediction accuracy 1.43% greater than the other two, GA-BP neural networks typically produce the best prediction results when selecting 5%, 10%, and 20% of the complete dataset. Based on the research results, suitable algorithms can be selected in practical work to process membrane experimental data and efficiently analyze performance.

Keywords: airborne inerting system; hollow fiber membrane; separation performance; artificial neural network; prediction

Received: 2024-03-25; **Accepted**: 2024-05-03; **Published Online**: 2024-07-25 14:33

URL: link.cnki.net/urlid/11.2625.V.20240725.0911.002

Foundation items: National Natural Science Foundation of China (52372331); the Fundamental Research Funds for the Central Universities; Priority Academic Program Development of Jiangsu Higher Education Institutions; Postgraduate Research & Practice Innovation Program of NUAA (xcxjh20230103)

* **Corresponding author**. E-mail: shiyuf@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0170

基于桥臂共享型多电平变换器的SRM瞬时转矩控制

陈凡强[1]，李存贺[1,*]，杜钦君[1]，柳健[2]，焦提操[1]

(1. 山东理工大学 电气与电子工程学院，淄博 255022；2. 大连海事大学 船舶电气工程学院，大连 116026)

摘 要：针对开关磁阻电机(SRM)现有不对称半桥功率变换器电平等级单一、转矩脉动抑制效果差的问题，提出一种基于新型多电平功率变换器的直接瞬时转矩控制(DITC)方法。首先，构造SRM桥臂共享型五电平功率变换器拓扑结构，通过引入横向与纵向共享桥臂，以最少功率器件数量实现多种电平组合，增加控制灵活度。其次，为消除桥臂共享导致的换相区开关状态紊乱，设计一种开关状态受限下的DITC策略，综合转矩误差、转子位置及电容电压平衡等多种因素，制定各相导通规则，最小化电机转矩脉动。与现有不对称半桥功率变换器的对比仿真和实验结果表明，所提SRM桥臂共享型多电平转矩控制不仅降低了电机转矩脉动，而且减小了铜耗，改善了电机运行效率。

关 键 词：开关磁阻电机；桥臂共享；五电平功率变换器；转矩脉动；直接瞬时转矩控制

中图分类号：TM352

文献标志码：A **文章编号**：1001-5965(2026)05-1746-10

开关磁阻电机(switched reluctance motor, SRM)具有成本低、结构简单、调速性能良好等优点，在航空航天、锻压机械、电动汽车等多种领域具有广阔的发展前景[1-2]。但SRM定转子特有的双凸极结构及高度非线性的电磁特性导致其转矩脉动严重，限制了SRM在高性能场合下的应用与发展[3-5]。近年来，国内外学者针对SRM转矩脉动抑制做了大量研究。

近年来，在降低SRM转矩脉动的研究中，主要从2个方面入手：①对电机本体的改进，通过优化电机的本体结构实现减小转矩脉动的目的[6]；②优化电机功率变换拓扑及控制策略，以降低转矩脉动[7]。文献[8]设计了一种转子齿两侧开槽的转子结构，改变了其表面气隙磁密的方向，通过削弱径向波来减小转矩脉动。文献[9]提出双定子结构，内外定子采用U型分块结构改善电磁路径，并通过改变转子内外齿角度降低转矩脉动。然而，从SRM本体角度抑制转矩脉动可能会导致系统效率降低，且耗时较长。

相比SRM本体结构改进，通过设计精巧的功率变换电路及控制策略能够以较低的代价实现电机转矩脉动最小化。文献[10-11]采用基于传统不对称半桥功率变换器的直接瞬时转矩控制(direct instantaneous torque control, DITC)方法来减少SRM转矩脉动，该方法直接对电磁转矩进行控制，通过转矩滞环使实际转矩和给定转矩的误差处于可控状态，结构简单可靠。然而，不对称半桥功率变换器滞环控制下仅具有 u_{dc}、0、$-u_{dc}$ 这3种电平状态，转矩脉动抑制效果有限。文献[12]通过引入boost电容，设计了具有可升压功能的SRM四电平功率拓扑，并将转子位置划分为3个区间，每个区间分别使用不同的滞环控制，形成了新的DITC策略。文献[13]针对SRM高速驱动系统，提出了一种附加电容和功率开关的新型四电平功率变换器和瞬时开关角检测器，不仅能够对绕组快速充放电，还能精准控制开关角。文献[14]提出一种改进的四

收稿日期：2024-03-26；**录用日期**：2024-06-21；**网络出版时间**：2024-07-04 11:00

网络出版地址：link.cnki.net/urlid/11.2625.V.20240703.1324.003

基金项目：国家自然科学基金面上项目(62076152)；山东省泰山学者项目(TSQN202306191)；山东省自然科学基金面上项目(ZR2020MF096)

* **通信作者**. E-mail：licunhe@sdut.edu.cn

引用格式：陈凡强，李存贺，杜钦君，等. 基于桥臂共享型多电平变换器的SRM瞬时转矩控制[J]. 北京航空航天大学学报，2026，52(5)：1746-1755. CHEN F Q，LI C H，DU Q J，et al. Instantaneous torque control of SRM based on bridge arm shared multilevel converter[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1746-1755 (in Chinese).

电平功率变换器，根据前端电容器连接方式的不同，提供不同的工作状态以实现高电压或大电流输出。四电平拓扑实现了绕组的快速励磁，但功率变换器中的泵压电容不易控制，且输出波形中谐波含量较高。

为进一步改善转矩脉动抑制效果，文献 [15] 设计了一种四相 SRM 五电平拓扑结构，结合所设计的单相和换相导通滞环控制规则，能够改善转矩动态性能。文献 [16] 提出一种基于五电平功率变换器的容错控制策略，面对不同的电机转速，提供不同的励磁和退磁电压，当电机驱动开路故障时，能较好地减轻转矩波动。文献 [17] 提出一种基于二极管钳位型五电平拓扑的 DITC 策略，划分了导通区间，并设计了低速与高速下的 DITC 导通规则，实现了宽速域下的转矩抑制。然而，上述五电平功率变换器拓扑复杂，且所用功率器件数量较多，不利于驱动系统成本控制。

本文提出一种基于桥臂共享的新型五电平功率变换器拓扑结构，并设计了开关状态受限下的 DITC 策略。新型功率拓扑在传统不对称半桥功率变换器的基础上，通过增加少量共享开关器件，实现五电平励磁和去磁的功能。在此基础上，提出改进的五电平 DITC 方法，制定开关状态受限下的转矩滞环控制规则，消除桥臂共享带来的换相区开关状态冲突，减小 SRM 换相区电流尖峰，最小化转矩脉动。仿真和实验验证了本文提出的新型功率变换拓扑及转矩控制方法的有效性。

1　SRM 瞬时转矩计算

SRM 独特的双凸极结构和脉冲供电方式使其具有高度的非线性[18]。根据虚位移原理，SRM 瞬时转矩计算公式为

$$\begin{cases} T_k(i_k,\theta)=\dfrac{\partial W'_{\mathrm{m}}(i_k,\theta)}{\partial\theta}\Bigg|_{i_k=\mathrm{const}} \\ T_{\mathrm{e}}=\sum T_k(i_k,\theta) \end{cases} \tag{1}$$

式中：T_k 和 T_{e} 分别为第 k 相输出转矩和总输出转矩；i_k 为第 k 相电流；θ 为转子位置角；W'_{m} 为电机磁共能，可进一步表示为

$$W'_{\mathrm{m}}(i_k,\theta)=\int_0^i \psi_k(i_k,\theta)\,\mathrm{d}i \tag{2}$$

式中：ψ_k 为第 k 相磁链。

为保证 SRM 瞬时转矩估计的实时性和准确性，采用实验测量法[19] 建立电机转矩模型。以三相 12/8 极 SRM 为例，定义定转子凸极完全对齐位置为 0°，由于 SRM 一个电周期内的电磁特性具有对称性，故测量 0～22.5° 范围内电磁特性即可获取 SRM 全周期模型。SRM 瞬时转矩建模步骤如下：

1）以 0.9° 为步长，依次固定转子位置，对相绕组通直流脉冲电压，采集相电流和电压数据。实验测量获得的 2 个极端位置处相电流和电压波形，如图 1 所示。

2）将测量的相电流和电压数据依次代入 SRM 磁链表达式(3)，计算转子不同位置处磁链特性，进而获得 SRM 整个周期磁链特性，如图 2(a)所示。

$$\psi_k(m)=\sum_{j=1}^{m}[u_k(j)-Ri_k(j)]T_{\mathrm{s}}|_{\theta=\mathrm{const}} \tag{3}$$

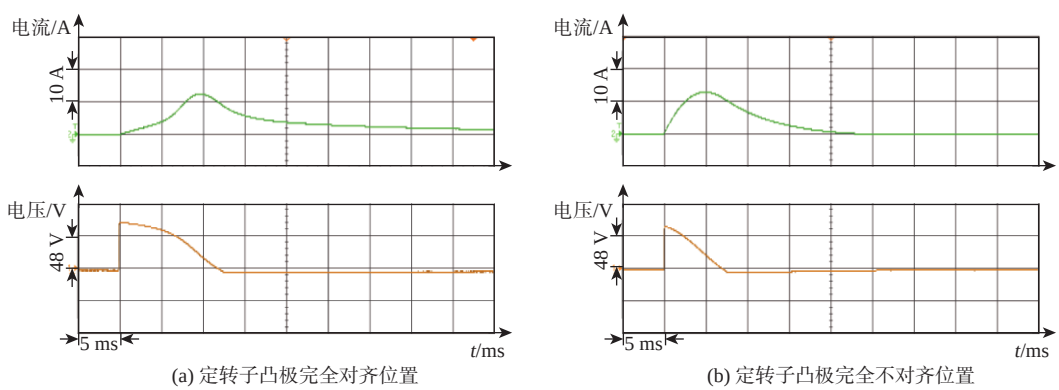
式中：u_k 为第 k 相电压；R 为绕组电阻；m 为采样点数；T_{s} 为采样周期。

3）将磁链数据代入式(1)和式(2)，计算可得 SRM 瞬时转矩特性，如图 2(b)所示。

为兼顾转矩计算精度和实时性，将转矩模型数据以二维表形式存储在微控制器中，每个周期对相电流和转子位置进行实时采样，采用双线性插值法计算瞬时转矩输出，具体为

$$\begin{aligned} f(x,y)=&\frac{1}{(x_2-x_1)(y_2-y_1)}[f(Q_{11})(x_2-x)(y_2-y)+\\ &f(Q_{21})(x-x_1)(y_2-y)+f(Q_{12})(x_2-x)(y-y_1)+\\ &f(Q_{22})(x-x_1)(y-y_1)] \end{aligned} \tag{4}$$

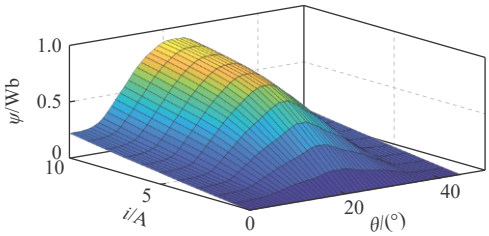
式中：$Q_{11}=(x_1,y_1)$、$Q_{12}=(x_1,y_2)$、$Q_{21}=(x_2,y_1)$、$Q_{22}=(x_2,y_2)$ 为二维空间内已知数据点。



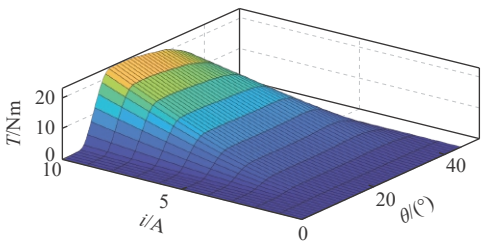
(a) 定转子凸极完全对齐位置　　(b) 定转子凸极完全不对齐位置

图 1　实验测量相电压和电流波形

Fig. 1　Experimental measurement of phase voltage and current waveforms



(a) 磁链特性



(b) 转矩特性

图 2 SRM 的磁链与转矩特性图

Fig. 2 Flux-linkage and torque characteristic of SRM

2 传统 DITC

2.1 DITC 系统结构

基于 DITC 的 SRM 系统框图如图 3 所示，系统由速度控制单元、转速计算单元、DITC 单元、功率变换器和 SRM 组成。其中，W_{ref} 为参考转速，T_{ref} 为参考转矩，W 为实际转速。转矩计算单元根据测量获得的电流和位置信号实时估算瞬时转矩，DITC 单元根据转矩误差和位置信息生成功率变换器控制信号，进而将转矩误差限定在一定的范围内。

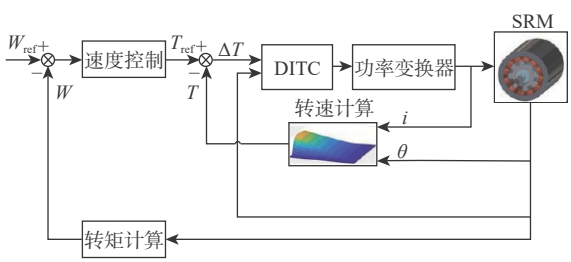


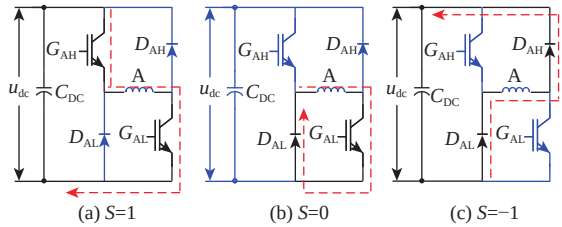
图 3 DITC 系统框图

Fig. 3 DITC system block diagram

2.2 不对称半桥功率变换器分析

SRM 传统功率变换器通常采用不对称半桥功率变换器，以 A 相为例，不对称半桥功率变换器 1、0、−1 这 3 种开关状态如图 4 所示。其中，C_{DC} 为母线电容，D_{AH} 为 A 相上桥臂续流二极管，D_{AL} 为 A 相下桥臂续流二极管，G_{AH} 为 A 相高处的开关器件，G_{AL} 为 A 相下桥臂开关器件。

不对称半桥功率变换器 3 种开关状态可以提供 u_{dc}、0、$-u_{dc}$ 这 3 种相电压等级，SRM 各相电压和转矩的关系可表示为



(a) S=1 (b) S=0 (c) S=−1

图 4 不对称半桥功率变换器开关状态

Fig. 4 Switching states of asymmetric half-bridge power converter

$$\begin{cases} u_k = L_k(i_k,\theta)\dfrac{\mathrm{d}i_k}{\mathrm{d}t} + i_k\omega\dfrac{\partial L_k(i_k,\theta)}{\partial\theta} + i_kR \\ T_k = \dfrac{\partial\int_0^i i_kL_k(i_k,\theta)\,\mathrm{d}i_k}{\partial\theta} = \int_0^i i_k\dfrac{\partial L_k(i_k,\theta)}{\partial\theta}\mathrm{d}i_k \end{cases} \tag{5}$$

式中，L_k 为第 k 相电感。

2.3 三电平 DITC 策略

DITC 各相导通状态可以划分为单相导通区和换相重叠区，具体如图 5 所示。其中，S_A 为 A 相开关状态，S_B 为 B 相开关状态。在单相导通区，转矩控制采用双滞环的方法，能够将转矩控制在合理范围内。在换相重叠区，要使 B 相逐渐代替 A 相进行转矩输出，A 相使用转矩外环±T_H 控制，B 相使用转矩内环±T_L 控制，A、B 相辅相成，完成换相操作。然而，传统不对称半桥功率变换器电平状态较少，对应 1、0、−1 这 3 种开关状态，仅具有 u_{dc}、0、$-u_{dc}$ 这 3 种电压等级。由式(5)可知，在换相重叠区电感小，单一的励磁电压会使电流变化速率过大，导致出现电流尖峰，恶化 DITC 转矩脉动抑制效果。

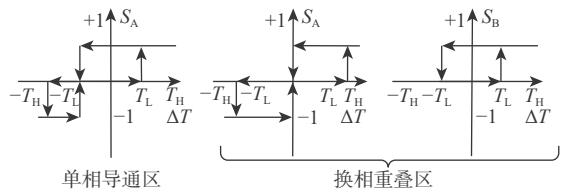


图 5 DITC 滞环控制

Fig. 5 Hysteresis control of DITC

3 桥臂共享型五电平 DITC

3.1 新型五电平功率变换器

为进一步改善转矩控制性能，提出一种桥臂共享型五电平功率变换器，具体如图 6 所示。该电路拓扑在传统不对称半桥功率变换器的基础上，通过增加分压电容 C_H、C_L，横向共享桥臂 G_{U1}、G_{U2} 和纵向共享桥臂 G_H、G_L，以最少器件数量实现 SRM 五种电压等级变换。如图 6 所示，桥臂共享型五电平功率变换器具有+2、+1、0、−1、−2 这 5 种开关状态。其中，G_{BH} 为 B 相上桥臂开关器件，D_{BH} 为 B 相上桥臂续流二极管，D_{BL} 为 B 相下桥臂续流二极管，

G_{BL} 为 B 相下桥臂开关器件，G_{CH} 为 C 相上桥臂开关器件，D_{CH} 为 C 相上桥臂续流二极管，D_{CL} 为 C 相下桥臂续流二极管，G_{CL} 为 C 相下桥臂开关器件。

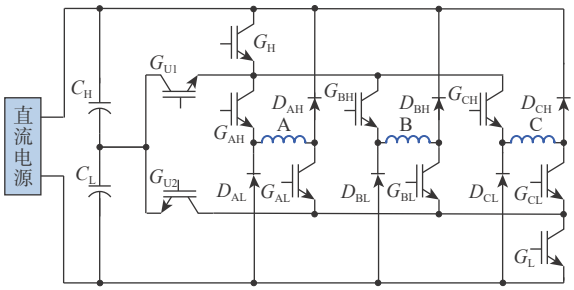


图 6　桥臂共享型五电平功率变换器

Fig. 6　Bridge arm shared five-level power converter

以 A 相为例，分析其励磁和去磁机理。如图 7(a)所示，开关状态为+2 时，相电压为 u_{dc}，绕组处于快速励磁状态，能够迅速建立起绕组电流；如图 7(b)和图 7(c)所示，开关状态为+1 时，相电压为 $u_{dc}/2$，由于需要考虑电容平衡的问题，根据电容电压不同，形成 2 种导通回路，均可缓慢励磁，建立起绕组电流；如图 7(d)和图 7(e)所示，开关状态为 0 时，相电压为 0，依据开关管不同，导通状态形成 2 种导通回路，实现续流的功能；如图 7(f)和图 7(g)所示，开关状态为−1 时，相电压为$-u_{dc}/2$，根据电容电压大小，形成 2 种导通回路，均可完成绕组缓慢去磁；如图 7(h)所示，开关状态为−2 时，所有开关管关闭，相电压为$-u_{dc}$，处于快速去磁状态，能够使绕组电流迅速下降。A 相开关状态和开关管导通状态如表 1 所示。

对于 n 相 SRM，桥臂共享型五电平功率变换器与现有五电平功率变换器所用功率器件数量对比如表 2 所示。可以看出，相比现有 SRM 二极管钳位型[20]、飞跨电容型[21]、H 桥级联型[22]和 T 型[23]等五电平功率变换器，本文提出的桥臂共享型五电平功率变换器的功率器件数量最少，可以有效节省系统成本。

3.2　桥臂共享型五电平 DITC 策略

在 DITC 策略中，为最小化 SRM 转矩脉动，需要根据转子的位置为各相选择励磁、续流或去磁状态。以 B 相为例，取 B 相的开通角 θ_{on_B} 到 C 相开通角 θ_{on_C} 作为一个周期，该周期内的相电感、相电流和转矩曲线如图 8 和图 9 所示。

如图 8 和图 9 所示，在区域 1 中，B 相电感 L_B 未发生大量变化，电流处于刚开始建立阶段，无转矩输出，此时，转矩 T 可看作由 A 相输出；在区域 2 中，A 相和 B 相电感都急剧变化，A 相电流 I_A 和 B 相电流 I_B 同时存在，此时，转矩由两相共同输出；在区域 3 中，B 相电感正向变化，A 相电感反

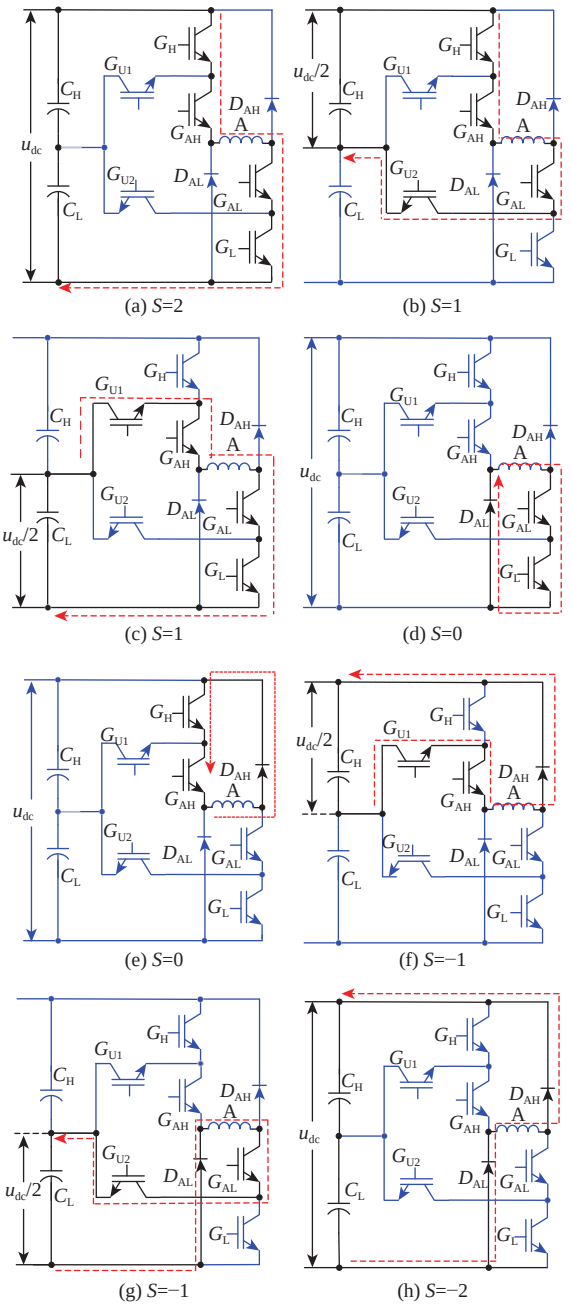


图 7　五电平功率变换器的开关状态

Fig. 7　Switching states of five-level power converter

表 1　A 相开关状态与开关管导通状态

Table 1　A-phase switching state and switching tube conduction state

电平状态	开关管状态					
	G_{U1}	G_{U2}	G_H	G_L	G_{AH}	G_{AL}
+2	0	0	1	1	1	1
+1	0	1	1	0	1	1
	1	0	0	1	1	1
0	0	0	0	1	0	1
	0	0	1	0	1	0
−1	1	0	0	0	1	0
	0	1	0	0	0	1
−2	0	0	0	0	0	0

表 2 *n* 相 SRM 五电平功率拓扑器件数量对比

Table 2 Comparison of device counts for *n*-phase SRM five-level drive systems

五电平类型	开关管数量	二极管数量	电容数量
桥臂共享型	2*n*+4	2*n*	2
二极管钳位型[20]	4*n*	4*n*	2
飞跨电容型[21]	3*n*	3*n*	*n*
H桥级联型[22]	5*n*	*n*	2*n*
T型[23]	4*n*	2*n*	2

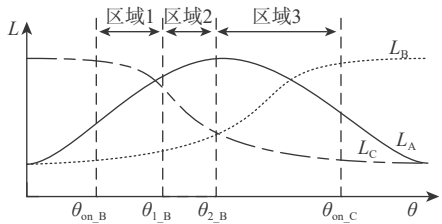


图 8 SRM 相电感曲线

Fig. 8 Phase inductance curves of SRM

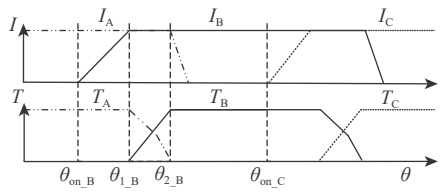


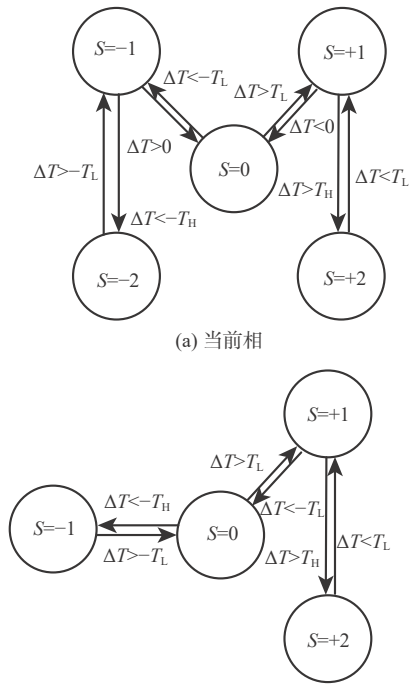
图 9 SRM 相电流和转矩曲线

Fig. 9 Phase current and torque curves of SRM

向变化，此时，希望 A 相电流急剧变化至零，转矩全部由 B 相输出。由上述可知，转矩的输出经历了单相输出、换相输出再到单相输出的阶段。根据 SRM 相电感、电流及转矩特性，综合考虑五电平功率变换器拓扑及开关状态，制定 DITC 滞环控制策略，如图 10 所示。

如图 10(a)所示，如果转矩误差ΔT增长，且超过滞环内环 T_L 时，A 相绕组处于正常导通状态 S_A=1，绕组电流和转矩增加；当ΔT增加超过滞环外环 T_H 时，为使转矩快速满足需求，A 相绕组需要迅速建立电流，S_A=2；当ΔT减小并小于零时，A 相的实际电流与理想电流接近，但此时误差有减小趋势，实际电流增加，此时 A 相绕组需要缓慢减小电流，S_A=0；当ΔT减小且小于滞环内环$-T_L$时，实际电流大于理想电流，A 相处于去磁状态，S_A=−1；当ΔT小于滞环外环$-T_H$时，A 相绕组需要迅速将电流减小，S_A=−2。

如图 10(b)所示，在换相重叠区中，以 B 相逐渐代替 A 相成为转矩输出相为例，若有降低转矩的需求，应首先令 A 相处于去磁状态，B 相尽可能处于励磁状态。A 相的开关状态和单相导通区时保持一致，B 相的开关状态与单相导通区有所差异，当



(a) 当前相

(b) 下一相

图 10 五电平 DITC 的滞环控制

Fig. 10 Hysteresis control of five-level DITC

ΔT小于滞环内环$-T_L$时，B 相处于缓慢退磁状态，S_B=0；当ΔT减小且小于$-T_H$时，B 相处于正常退磁状态，S_B=−1。

3.3 状态受限下的 DITC 策略

在单相导通区，SRM 各相开关状态互不影响，5 种开关状态都可以正常使用。但在换相重叠区，由于各相之间存在 G_H、G_L、G_{U1}、G_{U2} 这 4 个共享开关管，使换相重叠区相邻两相部分开关状态因共享开关管互相矛盾，导致电压状态紊乱，进而造成转矩输出失控。此外，五电平功率变换器中存在分压电容，需保证运行过程中的电位平衡。因此，综合考虑上述因素对开关状态选取带来的影响，制定开关状态受限下的 DTIC 规则，具体流程如图 11 所示。其中，U_{CH} 和 U_{CL} 分别为电容 CH 和 CL 两端电压。

由图 11 可以看出，当 U_{CH}>U_{CL} 时，为保证直流侧电容电压平衡，励磁状态+1 需要对电容 C_H 放电，退磁状态−1 需要对电容 C_L 充电，此时开关状态分别采用图 7(b)和图 7(g)导通回路。同理，当 U_{CL}>U_{CH} 时，需要对电容 C_L 放电，对电容 C_H 充电，分别选用图 7(c)和图 7(f)作为励磁和去磁回路。在换相重叠区，当+2 与+1 同时存在时，若ΔT>T_L，则相邻两相均选择状态+1，若ΔT>T_H，则相邻两相均选择状态+2，以避免状态+2 和状态+1 同时出现，消除桥臂共享带来的状态紊乱。当状态 0 和状态+1 同时存在时，考虑到状态+1 只需共享开关管 G_H 或 G_L 导通，而部分状态 0 的导通回路会使状态+1 中共享开

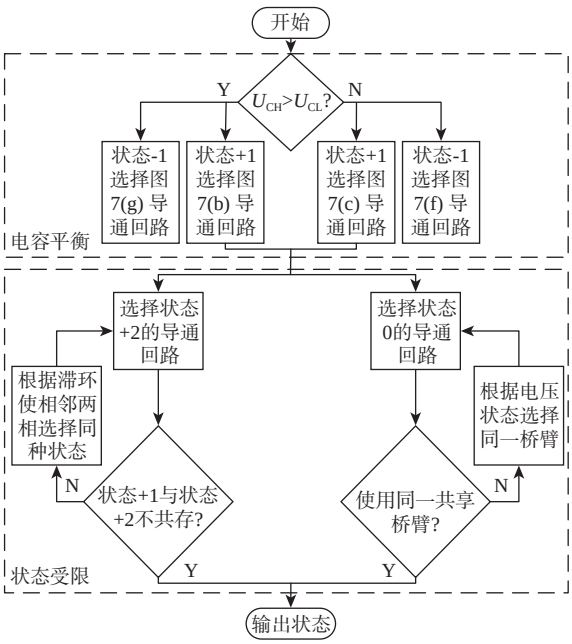


图 11　状态受限下 DITC 流程图

Fig. 11　Flowchart of DITC under state constraints

关管 G_H 和 G_L 同时导通，影响状态+1 的正常工作。为消除桥臂共享带来的状态冲突，必须保证状态 0 和状态+1 使用同一共享桥臂。由于状态+1 已根据电容电压平衡选定导通回路，因此，采用同种方法选择状态 0 的导通回路，即当 U_{CH}>U_{CL} 时，状态 0 选择图 7(e)作为导通回路，当 U_{CL}>U_{CH} 时，状态 0 选择图 7(d)作为导通回路。

4　仿真分析

为验证本文提出的新型五电平功率变换器及其改进 DITC 方法，建立三相 12/8 极 SRM 驱动系统仿真模型，SRM 样机参数如表 3 所示。仿真试验中转矩控制内滞环设置为±0.1，转矩外滞环设置为±0.5。转速环采用比例积分控制，比例系数 k_p 设定为 0.5，积分系数 k_i 设定为 0.1。

表 3　SRM 样机参数

Table 3　The parameters of SRM

参数	数值
额定功率/kW	1.5
额定转速/(r·min^{-1})	1 500
定子电阻/Ω	0.9
转动惯量/(kg·m^2)	0.01
摩擦系数/((Nm)·(rad·s^{-1})$^{-1}$)	0.005

首先，在转速 ω_n=800 r/min、负载转矩 T_L=5 Nm 时，对 SRM 不对称半桥及五电平功率变换器 DITC 控制下的相电流与开关状态进行仿真验证，仿真结果如图 12 和图 13 所示。可知，2 种方法均能稳定控制 SRM 相电流，但本文提出的五电平功率变换器 DITC 电流波形在前期更加平滑，在励磁初期的尖峰电流小于不对称半桥功率变换器，这有助于减小 SRM 转矩脉动与铜损耗。

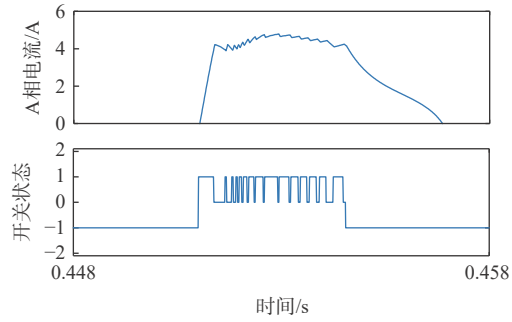


图 12　三电平 DITC 相电流及开关状态

Fig. 12　Three-level DITC phase current switching state

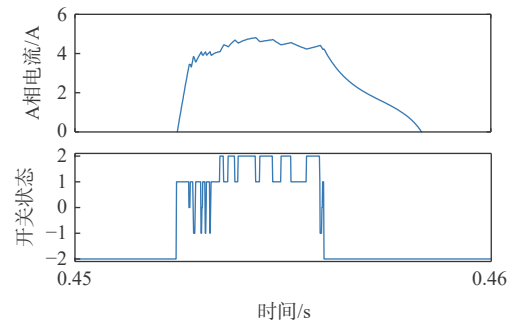


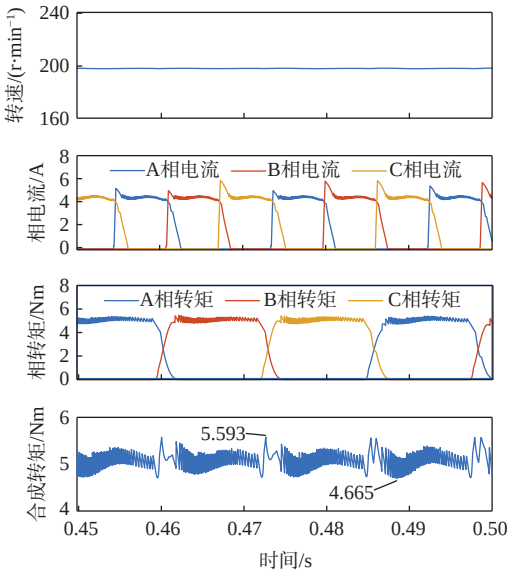
图 13　五电平 DITC 相电流及开关状态

Fig. 13　Five-level DITC phase current switching state

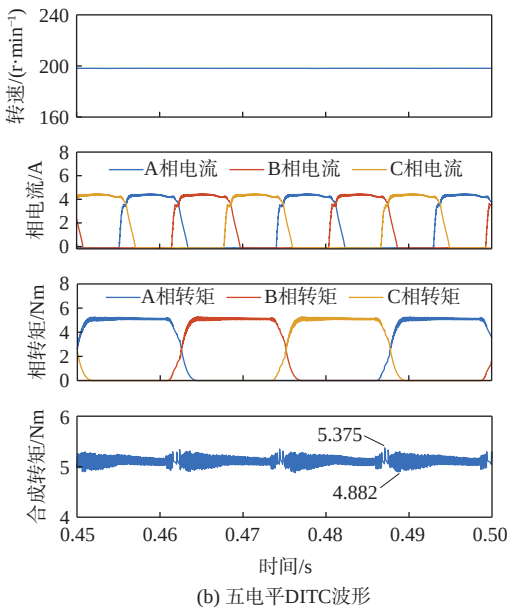
其次，分别在转速 ω_n=200 r/min 和 ω_n=800 r/min、负载转矩 T_L=5 Nm 时，对 SRM 不对称半桥及五电平功率变换器 DITC 转矩抑制能力进行仿真验证，仿真结果如图 14 和图 15 所示。从图 14 可以看出，200 r/min 转速下，三电平 DITC 存在较大的尖峰电流，导致其换相期间转矩波动严重，转矩最大和最小值分别为 5.593 Nm 和 4.665 Nm。本文提出的五电平 DITC 相电流更加平滑，输出转矩脉动更小，转矩最大和最小值分别为 5.375 Nm 和 4.882 Nm。从图 15 可以看出，在 800 r/min 转速下，五电平 DITC 也可以降低转矩脉动，转矩最大和最小值由 6.016 Nm 和 4.88 Nm 改善至 5.931 Nm 和 5.002 Nm。

为定量评估 DITC 下的控制性能，引入转矩脉动系数 K_T、铜耗 P_{cu} 和电机运行效率 η 作为性能指标，转矩脉动系数 K_T 越小，表示 SRM 转矩脉动抑制能力越好；铜耗 P_{cu} 越小，表示电机损耗越小，运行效率 η 越高，表示能量利用率越高。K_T、P_{cu} 和 η 定义为

$$\begin{cases} K_T = \dfrac{T_{max} - T_{min}}{T_{av}} \\ P_{cu} = nI_{rms}^2 R \\ \eta = \dfrac{P_{out}}{P_{in}} \times 100\% = \dfrac{T_{av}\beta}{U_{bus}I_{bus}} \times 100\% \end{cases} \tag{6}$$



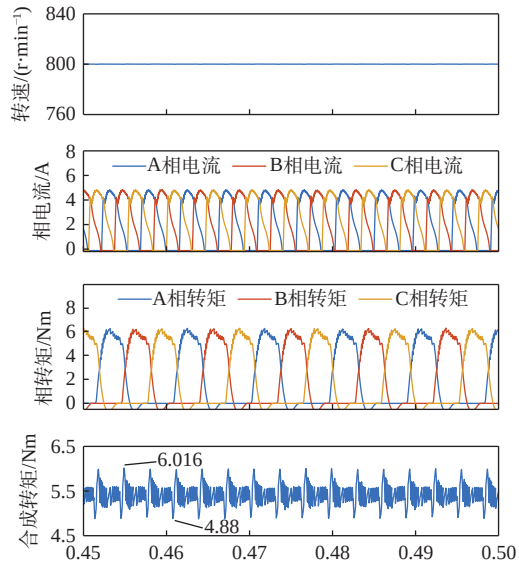
(a) 三电平DITC波形



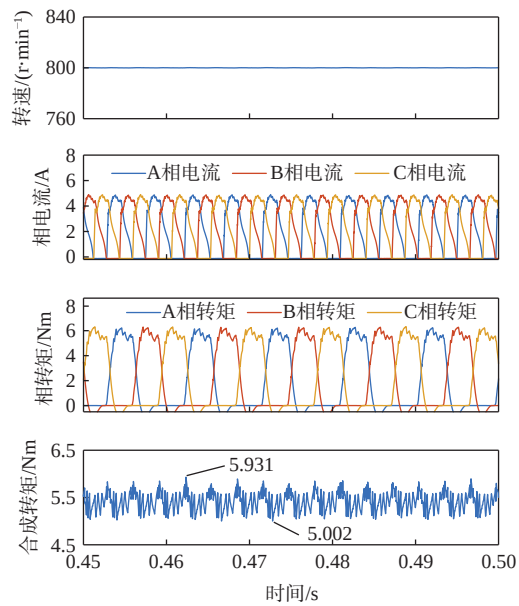
(b) 五电平DITC波形

图 14 转速 ω_n=200 r/min、负载 T_L=5 Nm 时仿真结果

Fig. 14 Simulation results at rotational speed 200 r/min with 5 Nm load



(a) 三电平DITC波形



(b) 五电平DITC波形

图 15 转速 ω_n=800 r/min、负载 T_L=5 Nm 时仿真结果

Fig. 15 Simulation results at rotational speed 800 r/min with 5 Nm load

式中：T_{max}、T_{min}、T_{av} 分别为转矩的最大值、最小值、平均值；P_{out} 和 P_{in} 分别为输出功率和输入功率；U_{bus}、I_{bus} 分别为母线电压和母线电流；β 为电机角速度；I_{rms} 为电流有效值，计算公式为

$$I_{rms}=\sqrt{\frac{1}{P}\int_{\theta_{on}}^{\theta_{on}+T} i^2(t)\,\mathrm{d}t} \tag{7}$$

式中：θ_{on} 为开通角；P 为相电流周期。

表 4 给出了本文提出的五电平 DITC 与传统三电平 DITC 的性能指标对比。可知，转速为 200 r/min 时，三电平 DITC 转矩脉动系数 K_T 和铜耗 P_{cu} 分别为 18.18% 和 19.77 W，五电平 DITC 的 K_T 和 P_{cu} 分别减小至 9.66% 和 18.06 W，效率 η 从 17.14% 提升至 18.74%。在转速为 800 r/min 时，三电平 DITC 转矩脉动系数 K_T 和铜耗 P_{cu} 分别为 20.96% 和 22.35 W，五电平 DITC 的 K_T 和 P_{cu} 分别减小至 17.15% 和 20.7 W，效率 η 从 48.03% 提升至 48.89%。

为详细比较本文方案对电机运行效率的改善

表 4 三电平和五电平 DITC 性能指标对比

Table 4 Comparison of performance indicators between three-level and five-level DITC

DITC	转速/(r·min⁻¹)	脉动系数/%	铜耗/W	效率/%
三电平DITC	200	18.18	19.77	17.14
	800	20.96	22.35	48.03
五电平DITC	200	9.66	18.06	18.74
	800	17.15	20.7	48.89

作用，图 16 展示了负载转矩 T_L=5 Nm 时，SRM 运行效率 η 随转速的变化曲线。可以看出，SRM 转速由 200 r/min 上升至 1 500 r/min，电机运行效率逐渐提高，本文提出的五电平 DITC 方法在低速范围内效率优于传统三电平 DITC，而在高速区，2 种方法效率逐渐趋于一致。这是由于 SRM 尖峰电流随转速增加也在缓慢减小，电流尖峰可抑制空间下降，故效率提升程度逐渐被削弱。

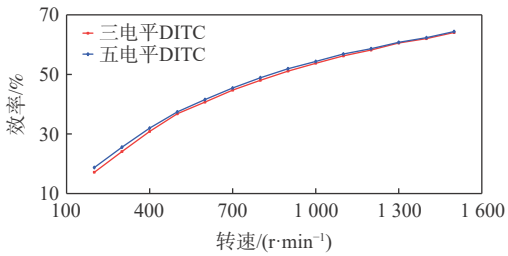


图 16　SRM 运行效率曲线

Fig. 16　Operational efficiency curve of the SRM

仿真结果表明，本文提出的五电平 DITC 方法相较于三电平 DITC 具有更优秀的转矩脉动抑制能力，且可以降低铜耗，提升 SRM 运行效率。

5　实验验证

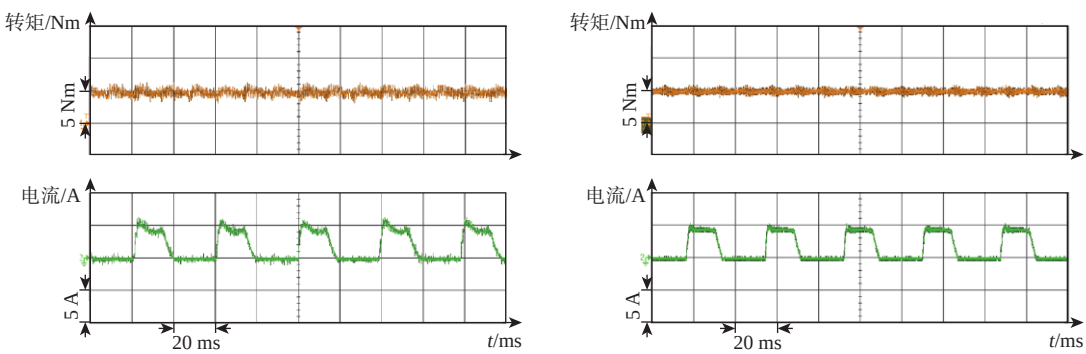
为进一步验证本文方法的有效性，搭建如图 17 所示的实验平台，对桥臂共享型五电平功率变换器及其 DITC 方法进行实验验证。实验平台机械部分由 SRM 样机、联轴器、转矩传感器和磁粉制动器组成。电气部分由上位机、dSPACE DS1104、调压器、功率变换器、接口电路等组成。实验中电机转速环与转矩环采样频率分别设定为 1 kHz 和 10 kHz。SRM 样机参数如表 3 所示。



图 17　SRM 实验平台

Fig. 17　Experimental bench for SRM

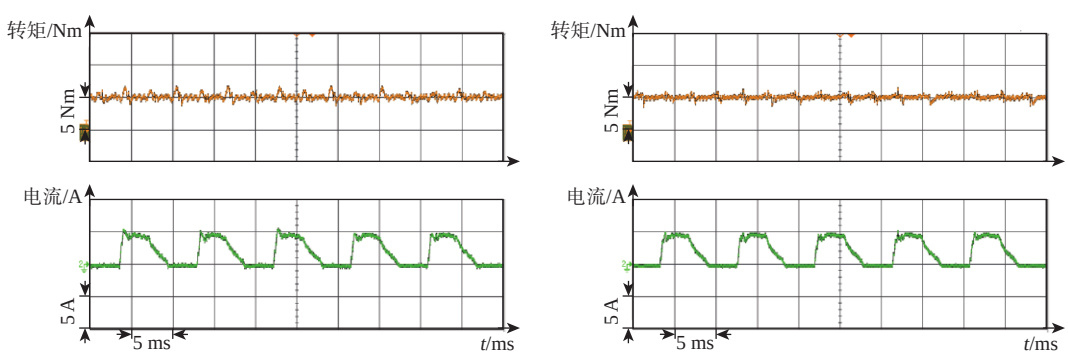
图 18 和图 19 分别为转速 200 r/min 和 800 r/min、负载 5 Nm 时，本文提出的五电平 DITC 与传统三电平 DITC 的对比实验结果。可见，本文提出的五电平 DITC 与传统三电平 DITC 均能将 SRM 瞬时转矩脉动抑制在一定范围内。然而，本文提出的五电平 DITC 下 SRM 尖峰电流和转矩波动明显小于传统三电平 DITC。



(a) 三电平DITC实验波形　　(b) 五电平DITC实验波形

图 18　转速 ω_n=200 r/min、负载 T_L=5 Nm 时实验结果

Fig. 18　Experimental results at rotational speed 200 r/min with 5 Nm load



(a) 三电平DITC实验波形　　(b) 五电平DITC实验波形

图 19　转速 ω_n=800 r/min、负载 T_L=5 Nm 时实验结果

Fig. 19　Experimental results at rotational speed 800 r/min with 5 Nm load

根据式(6)计算可知,转速为200 r/min时,本文提出的五电平DITC下转矩脉动系数K_T、铜耗P_{cu}和效率η分别为43.75%、19.86 W和16.7%,传统三电平DITC下转矩脉动系数K_T、铜耗P_{cu}和效率η分别为50.92%、21.53 W和15.3%。相比传统三电平DITC,本文提出的五电平DITC转矩脉动降低了14.1%,铜损耗降低了7.8%,运行效率提升了1.4%。转速为800 r/min时,本文提出的五电平与传统三电平DITC下转矩脉动系数K_T分别为55.17%和58.86%,铜损耗P_{cu}分别为21.34 W和22.93 W,效率η分别为48.1%和47.5%。与传统三电平DITC相比,本文提出的五电平DITC转矩脉动和铜耗分别降低了6.3%和6.9%,运行效率提升了0.6%。

由实验结果分析可知,本文提出的五电平DITC相较于传统三电平DITC,不仅具有更低的转矩脉动,而且减小了铜耗,提高了电机运行效率。这是因为本文提出的五电平DITC通过引入$u_{dc}/2$和$-u_{dc}/2$这2种额外的励磁和去磁状态,使电流和转矩调节更加精细,避免了转矩大幅波动。尤其是在励磁初期,通过$u_{dc}/2$和$-u_{dc}/2$的引入,消除了励磁电压过大导致的相电流尖峰,进而保证SRM瞬时转矩输出更加平滑。

6 结 论

本文提出了一种适用于SRM的桥臂共享型五电平功率变换器及其改进DITC策略,与现有三电平DITC方法相比,主要贡献如下:

1)在深入分析SRM励磁和去磁机理的基础上,通过引入横向和纵向共享桥臂,构造了SRM新型五电平功率变换器,使SRM驱动系统的开关状态组合更加灵活,电流和转矩调控更加精细。

2)提出了适用于SRM桥臂共享型多电平功率变换器的DITC方法,制定了受限状态下的转矩滞环控制规则和电容均压策略,消除了桥臂共享带来的换相重叠区开关状态紊乱。

3)设计了五电平DITC与传统三电平DITC的对比仿真和实验。仿真和实验结果表明,本文提出的五电平DITC不仅降低了SRM转矩脉动,而且减小了电机铜耗,提升了SRM运行效率。

参考文献(References)

[1] GAAFAR M A, ABDELMAKSOUD A, ORABI M, et al. Switched reluctance motor converters for electric vehicles applications: comparative review[J]. IEEE Transactions on Transportation Electrification, 2023, 9(3): 3526-3544.

[2] KIYOTA K, SUGIMOTO H, CHIBA A. Comparing electric motors: an analysis using four standard driving schedules[J]. IEEE Industry Applications Magazine, 2014, 20(4): 12-20.

[3] BOSTANCI E, MOALLEM M, PARSAPOUR A, et al. Opportunities and challenges of switched reluctance motor drives for electric propulsion: a comparative study[J]. IEEE Transactions on Transportation Electrification, 2017, 3(1): 58-75.

[4] KUMAR A, SINGH B, SINGH G. MPC based position sensorless SRM drive for light electric vehicles with optimized voltage vectors and reduced torque ripple[J]. IEEE Transactions on Industry Applications, 2024, 60(2): 3237-3246.

[5] FANG G L, SCALCON F P, XIAO D X, et al. Advanced control of switched reluctance motors (SRMs): a review on current regulation, torque control and vibration suppression[J]. IEEE Open Journal of the Industrial Electronics Society, 2021, 2: 280-301.

[6] DING W, BIAN H, SONG K D, et al. Enhancement of a 12/4 hybrid-excitation switched reluctance machine with both segmented-stator and rotor[J]. IEEE Transactions on Industrial Electronics, 2021, 68(10): 9229-9241.

[7] 陈越, 蒋启龙, 王金锁, 等. 基于电感特征的开关磁阻电机电流斩波控制策略[J]. 北京航空航天大学学报, 2023, 49(3): 647-656.
CHEN Y, JIANG Q L, WANG J S, et al. Current chopping control strategy of switched reluctance motor based on inductance characteristics[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(3): 647-656(in Chinese).

[8] 张鑫, 王秀和, 杨玉波, 等. 基于转子齿两侧开槽的开关磁阻电机振动抑制方法研究[J]中国电机工程学报, 2015, 35(6): 1508-1515.
ZHANG X, WANG X H, YANG Y B, et al. Vibration reduction of a switched reluctance motor using new rotor tooth with slot on each side[J]. Proceedings of the CSEE, 2015, 35(6): 1508-1515(in Chinese).

[9] 闫文举, 陈昊, 马小平, 等. 不同转子极数下磁场解耦型双定子开关磁阻电机的研究[J]. 电工技术学报, 2021, 36(14): 2945-2956.
YAN W J, CHEN H, MA X P, et al. Development and investigation on magnetic field decoupling double stator switched reluctance machine with different rotor pole numbers[J]. Transactions of China Electrotechnical Society, 2021, 36(14): 2945-2956(in Chinese).

[10] YANG Z B, SUN J F, SUN X D, et al. Direct instantaneous torque control for six-phase SRM with nonsingular fast terminal sliding mode controller[J]. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2024, 12(1): 505-515.

[11] 刘勇智, 李杰, 鄯成龙. 基于最优角度自适应TSF的SRM直接瞬时转矩控制[J]. 北京航空航天大学学报, 2019, 45(11): 2152-2159.
LIU Y Z, LI J, SHAN C L. Direct instantaneous torque control of switched reluctance motor based on optimal angle adaptive TSF[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(11): 2152-2159(in Chinese).

[12] 李珍国, 魏艳君, 阚志忠, 等. 基于四电平功率变换电路的开关磁阻电机瞬时转矩控制[J]. 电工技术学报, 2007, 22(8): 144-149.
LI Z G, WEI Y J, KAN Z Z, et al. Direct instantaneous torque control of SRM using four-level converter[J]. Transactions of China Electrotechnical Society, 2007, 22(8): 144-149(in Chinese).

[13] LEE D H, AHN J W. A novel Four-level converter and instantaneous switching angle detector for high speed SRM drive[J]. IEEE Transactions on Power Electronics, 2007, 22(5): 2034-2041.

[14] MA M Y, CHANG Z Y, HU Y H, et al. An integrated switched reluctance motor drive topology with voltage-boosting and on-board charging capabilities for plug-in hybrid electric vehicles

(PHEVs)[J]. IEEE Access, 2018, 6(1): 1550-1559.

[15] 程勇, 林辉. 五电平拓扑下开关磁阻电机直接瞬时转矩控制[J]. 电机与控制学报, 2011, 15(4): 18-22.
CHENG Y, LIN H. Direct instantaneous torque control of switched reluctance motor with five level converter[J]. Electric Machines and Control, 2011, 15(4): 18-22(in Chinese).

[16] MA M Y, YUAN K X, YANG Q Q, et al. Open-circuit fault-tolerant control strategy based on five-level power converter for SRM system[J]. CES Transactions on Electrical Machines and Systems, 2019, 3(2): 178-186.

[17] 凌辉, 杜钦君, 冯晗, 等. 基于五电平变换器的开关磁阻电动机转矩脉动抑制方法[J]. 上海交通大学学报, 2022, 56(12): 1608-1618.
LING H, DU Q J, FENG H, et al. Torque ripple reduction method of switched reluctance motor based on five-level converter[J]. Journal of Shanghai Jiao Tong University, 2022, 56(12): 1608-1618(in Chinese).

[18] WANG Z L, CAO X, DENG Z Q, et al. Modeling and characteristic investigation of axial reluctance force for bearingless switched reluctance motor[J]. IEEE Transactions on Industry Applications, 2021, 57(5): 5215-5226.

[19] GOBBI R, SAHOO N C, VEJIAN R. Experimental investigations on computer-based methods for determination of static electromagnetic characteristics of switched reluctance motors[J]. IEEE Transactions on Instrumentation and Measurement, 2008, 57(10): 2196-2211.

[20] PENG F, YE J, EMADI A. An asymmetric three-level neutral point diode clamped converter for switched reluctance motor drives[J]. IEEE Transactions on Power Electronics, 2017, 32(11): 8618-8631.

[21] YAMADA N, HOSHI N. Experimental verification on a switched reluctance motor driven by asymmetric flying capacitor multilevel h-bridge inverter[C]//Proceedings of the 2017 IEEE 6th International Conference on Renewable Energy Research and Applications. Piscataway: IEEE Press, 2017: 971-976.

[22] KORKOSZ M, PAKLA B. Multilevel converter for high-voltage high-speed switched reluctance motor[C]//Proceedings of the 2018 Innovative Materials and Technologies in Electrical Engineering. Piscataway: IEEE Press, 2018: 1-4.

[23] MA M, WANG Z, WANG R, et al. A new five-level power converter for switched reluctance motor drive[C]//Proceedings of the IEEE International Power Electronics and Application Conference and Exposition. Piscataway: IEEE Press, 2018: 1-6.

Instantaneous torque control of SRM based on bridge arm shared multilevel converter

CHEN Fanqiang[1], LI Cunhe[1,*], DU Qinjun[1], LIU Jian[2], JIAO Ticao[1]

(1. School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255022, China;
2. College of Marine Electrical Engineering, Dalian Maritime University, Dalian 116026, China)

Abstract: To address the issues of a low number of levels and poor torque ripple suppression effect in existing asymmetric bridge power converters for switched reluctance motors (SRM), a direct instantaneous torque control (DITC) based on a novel multi-level power converter is proposed. Firstly, a bridge arm shared five-level power conversion topology was derived for the SRM. Control flexibility was increased by achieving multiple-level combinations with the least number of power devices by providing shared bridge arms both vertically and horizontally. Secondly, an improved DITC strategy with limited switching states is developed to eliminate the switching state disorder in the commutation overlap zone caused by shared bridge arms. On this basis, various factors such as torque error, rotor position, and capacitor voltage balance are comprehensively considered to formulate conduction rules for each phase to minimize motor torque ripple. The suggested bridge arm shared multilevel torque control not only lowers current and torque ripple but also enhances motor copper loss and operating efficiency, according to comparative simulation and experimental results with traditional asymmetrical bridge power converters.

Keywords: switched reluctance motor; shared bridge arms; five-level power converter; torque ripple; direct instantaneous torque control

Received: 2024-03-26; **Accepted**: 2024-06-21; **Published Online**: 2024-07-04 11:00

URL: link.cnki.net/urlid/11.2625.V.20240703.1324.003

Foundation items: The General Program of National Natural Science Foundation of China (62076152); Taishan Scholar Project of Shandong Province (TSQN202306191); Natural Science Foundation of Shandong Province (ZR2020MF096)

* **Corresponding author**. E-mail: licunhe@sdut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0173

基于卷积注意力与特征融合的火灾检测算法

田佳麒，秦国轩，张为*
(天津大学 微电子学院，天津 300072)

摘 要：针对现实场景下火灾检测精度与速度不平衡的情况，提出一种加强空间特征提取和多尺度特征融合的火灾检测算法。对主干网络的高层语义信息提取进行改进，将感受野卷积注意力模块嵌入主干网络中，提升模型的特征提取能力；引入改进后的强特征融合网络，将低层空间信息和高层语义信息进一步加强融合，提升模型精度；利用局部卷积(PConv)模块对主干网络和检测头进行轻量化改进，在不损失精度的前提下，降低模型的参数量和内存访问；调整回归损失函数，提升模型的检测能力。实验结果表明，改进算法在自建的火灾数据集上的0.5阈值下平均精度均值(mAP50)和0.5:0.95阈值下平均精度均值(mAP50:95)分别提高了2.1%和2.9%，证明了所提算法在火灾检测领域的优越性；在Pascal VOC 07+12公开数据集上的mAP50和mAP50:95分别提高了1.4%和2.4%，证明了所提算法具有较强的泛化性能。

关 键 词：火灾检测；多尺度特征；特征融合；轻量化检测头；损失函数
中图分类号：TP391.41
文献标志码：A **文章编号**：1001-5965(2026)05-1756-11

火灾是日常容易发生的灾害之一，国家消防救援局统计，2023年上半年，我国日均火灾超3 000起，共死亡959人，直接财产损失39.4亿元，造成了严重的生命财产损失。因此，及时发现并扑灭初期火灾具有重大意义。

由于传感器对火灾的检测具有一定的局限性，无法及时发现初期的火灾，又由于近年来深度学习的发展，尤其是目标检测领域的迅速发展，使基于深度学习的火灾检测算法研究开始受到大量关注。Sharma等[1]使用预训练好的VGG16和ResNet50进行火灾检测，提高了火灾检测精度。Hosseini等[2]将卷积神经网络(convolutional neural network, CNN)引入火灾检测中，利用经过改进的UFS-Net对视频帧中的火焰和烟雾进行逐帧检测，并引入投票机制，进而降低了误报率。张融等[3]提出了GhostNet-FCOS火灾检测算法，将FCOS网络进行轻量化，且优化了训练策略，在自建数据集上进行测试后，发现效果优于已知算法。秦瑞等[4]提出了一种无锚框结构的火灾检测算法，引入BFP模块和动态采样策略，强化了网络的特征融合能力，在自制数据集和公开数据集上都取得了更好的效果。

随着计算机视觉技术的发展，目标检测算法逐渐分为基于Transformer的目标检测算法和基于CNN的目标检测算法。基于Transformer的目标检测算法以DETR为代表，并衍生出了诸如Deformable-DETR、DAB-DETR、Sparse-DETR等新型的目标检测算法，但其推理速度慢、部署困难的缺点导致其难以在现实环境中应用。基于CNN的目标检测算法可分为以Fast R-CNN为代表的二阶段算法和以YOLO系列为代表的单阶段算法。二阶段算法将检测分为2步：①生成候选框；②是将候选框进行分类和回归。而单阶段算法是不生成候选框，而

收稿日期：2024-03-26；录用日期：2024-07-30；网络出版时间：2024-08-08 09:08
网络出版地址：link.cnki.net/urlid/11.2625.V.20240807.1402.002
基金项目：国家重点研发计划(2022YFC3006302)
*通信作者. E-mail：tjuzhangwei@tju.edu.cn

引用格式：田佳麒，秦国轩，张为. 基于卷积注意力与特征融合的火灾检测算法[J]. 北京航空航天大学学报，2026，52(5)：1756-1766.
TIAN J Q，QIN G X，ZHANG W. Fire-and-smoke detection algorithm based on convolutional attention and feature fusion[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1756-1766 (in Chinese).

直接进行预测。虽然二阶段算法相比于单阶段算法精度更高, 但单阶段算法的检测效率更高, 所需的计算资源更少, 推理速度更快, 因此, 对于实际应用来说, YOLO 系列算法的应用更为广泛。

本文选用 YOLOv8 模型作为基准网络。YOLOv8 算法吸取了之前 YOLO 算法的成功经验, 其将 C2f 模块引入主干网络和颈部网络中, 以获取更丰富的梯度流; 采用目前主流的解耦头结构, 使分类和回归分离, 提升了模型性能; 引入 Task Aligned Assigner 的动态正样本回归策略, 提高了模型的回归能力。YOLOv8 模型有 n、s、m、l、x 这 5 种, 考虑到在火灾检测任务中需要同时保证高精度和实时性, 本文最终选用 YOLOv8s 为基准网络, 针对早期火灾中火焰和烟雾形态变化大、特征不明显的特点进行研究。本文在深层特征提取网络中引入感受野卷积注意力模块, 将原先的颈部特征融合网络替换为强特征融合网络, 并融入原特征网络的思想以加强特征的融合。在主干网络的浅层 C2f 和检测头部分引入局部卷积 (partial convolution, PConv) 倒残差模块, 减少整体网络的计算量和内存访问。对损失函数进行针对性的改进, 考虑到预测框和真实框宽高比相同, 但框大小不同的情况, 将回归损失函数替换为 MPDIOU。

1　本文算法

以 YOLOv8s 为基准网络, 本节详细描述改进后的算法结构。整体网络结构如图 1 所示, 由于颈部网络结构较为复杂, 在 1.2 节对其作具体阐述。网络中的部分模块有 K 和 S 的字样, 其后的数字分别代表卷积核大小和卷积步长。整体网络由特征提取、特征重融合和轻量化检测头 3 部分组成, 针对火灾现场中烟雾和火焰形状不清晰、尺度差异大的特点, 着重提升网络的多尺度特征提取和融合能力, 旨在更好地提升实际火灾现场的检测性能。

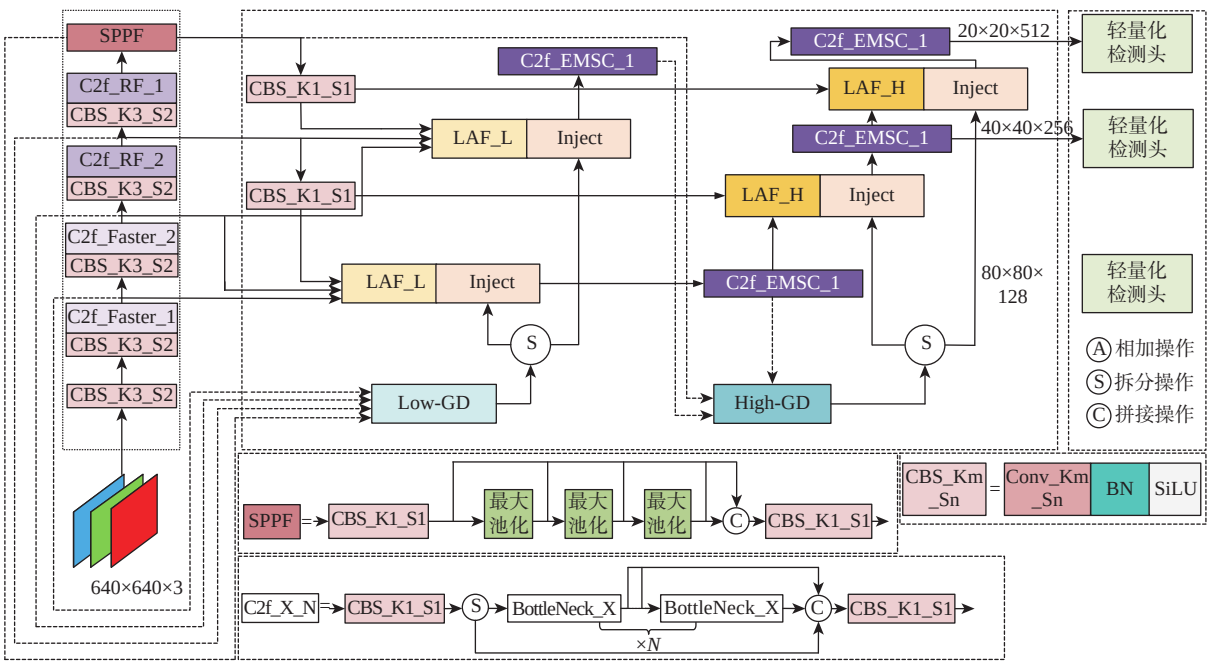


图 1　本文算法整体结构

Fig. 1　The overall structure of the algorithm

1.1　感受野卷积注意力模块

由于火焰和烟雾形状易变化, 且两者尺度差异很大, 需要网络能够尽可能地提取更多的特征。在主干网络部分, 引入了感受野卷积注意力模块, 其由感受野 (reception field, RF)[5] 的空间特征思想改进而来。RF 的思想是通过将原始特征图的空间特征进行特征映射, 将特征放大到原来的 K^2 倍, 因此, 每个 $K\times K$ 的滑块在计算过程中会提取初始的特征图单元特征, 彻底解决参数共享的问题, 进而提取更多特征, 增强网络性能。

基于该思想, 本节通道注意力模块更换为全局注意力机制 (global attention mechanism, GAM) 的通道注意力模块, 以放大通道-空间的依赖性。该模块的整体结构如图 2 所示。

该模块不同于传统卷积注意力模块先进行通道加权再进行空间加权的思想, 其空间注意力和通道注意力是并行的。首先, 输入特征经过维度变换后生成包含通道和空间依赖关系的向量, 将其经过 2 组重加权后生成包含通道-空间依赖性的特征图; 然后, 将该特征图进行反卷积操作, 以获得匹配大小的通道注意力。将调整大小后的特征图进行平均池化和全局池化并得到 2 个特征图, 将拼接后的

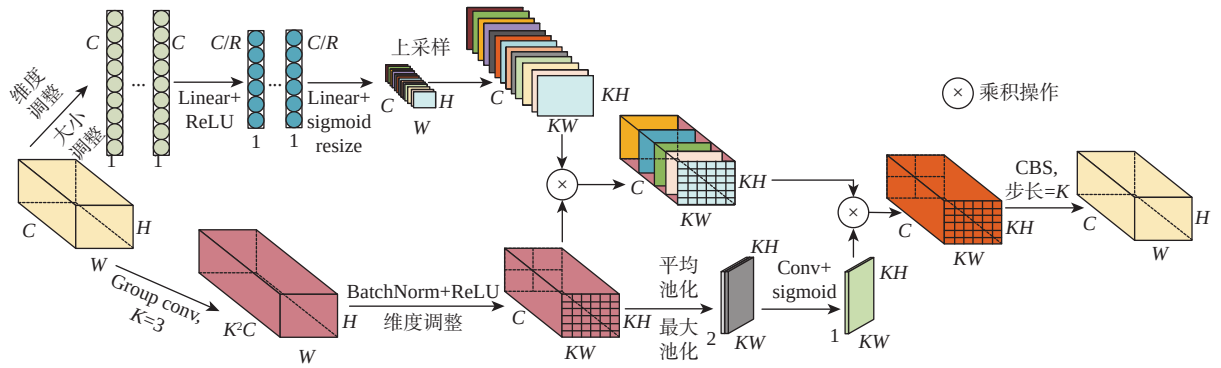


图2 感受野卷积注意力模块

Fig. 2 Receptive field convolutional attention module

特征图进行卷积和激活函数处理，并与通道加权后的特征图进行 Hadamard 乘积操作，得到 $C\times KH\times KW$ 的特征图；最后，将该特征图经过卷积、批归一化和激活函数处理后，输出 $C\times H\times W$ 的特征图。为尽可能少地增加参数量和计算量，同时提取更加丰富的语义信息，本文将感受野卷积注意力模块引入主干网络中的后 2 个 C2f 模块，以替换其中的简单卷积操作，替换后的 BottleNeck-RF 结构如图 3 所示。

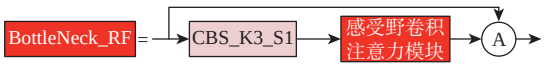


图3 BottleNeck_RF 结构

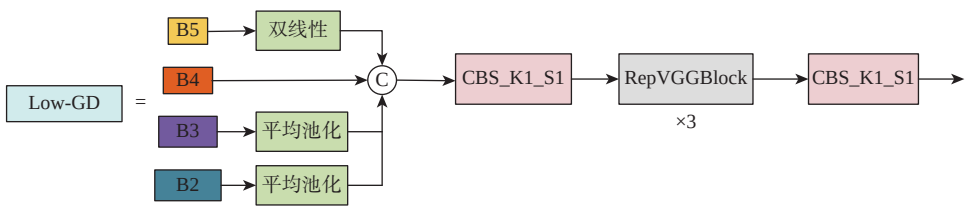
Fig. 3 The structure of BottleNeck_RF

1.2 强特征融合网络

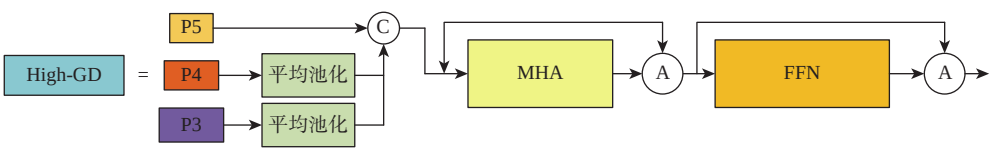
本文基准网络的颈部网络为 PANet[6]，其由特征金字塔网络 (feature pyramid network, FPN) 改进而来，通过加入自底向上的路径拓展进一步加强不同特征图的融合来提升网络性能。虽然 PANet 的性能相较于 FPN 有较大提升，但其也有一定的局限性。PANet 在进行通道融合时会丢失大量的细节信息，同时，其往往是局部特征的融合，且每一层的信息只能与相邻层融合，不能考虑全局的空间和语义信息，这限制了网络性能的进一步提升。本文用文献 [7] 提出的颈部网络结构对 PANet进行替换，并对其进行了进一步的改进，形成强特征融合网络。

文献 [7] 提出的颈部网络构建了一种新的特征收集与分发(gather-distribute, GD)机制，且分为 Low-GD 和 High-GD 这 2 个模块，用以区分融合低层信息和高层信息所用的模块，其结构如图 4 所示。Low-GD 将主干网络的输出特征层按由深到浅的顺序进行输入，通过上采样和平均池化操作统一特征图大小后，将所有输入特征层进行拼接，所有输入特征图与倒数第 2 层特征图大小保持一致，并将拼接后的特征图通过重参数化卷积块[8] 进行通道融合。High-GD 将经过低层融合后的输出特征图作为输入，输入经过统一大小操作后进入 Transformer 模块，经过多头注意力 (multi-head attention, MHA) 和前馈神经网络 (feed-forward network, FFN) 获得更多的高层全局信息，得到的特征图在调整通道维度后进行输出。

在将局部特征进行特征对齐和融合的同时，文献 [7] 提出的颈部网络引入了轻量化相邻层融合 (lightweight adjacent layer fusion, LAF) 模块和信息注入 (information injection, Inject) 模块，其结构如图 5 所示。LAF 模块又分为 Low-LAF 和 High-LAF，分



(a) Low-GD结构



(b) High-GD结构

图4 特征收集与分发结构

Fig. 4 Gather-distribute structure

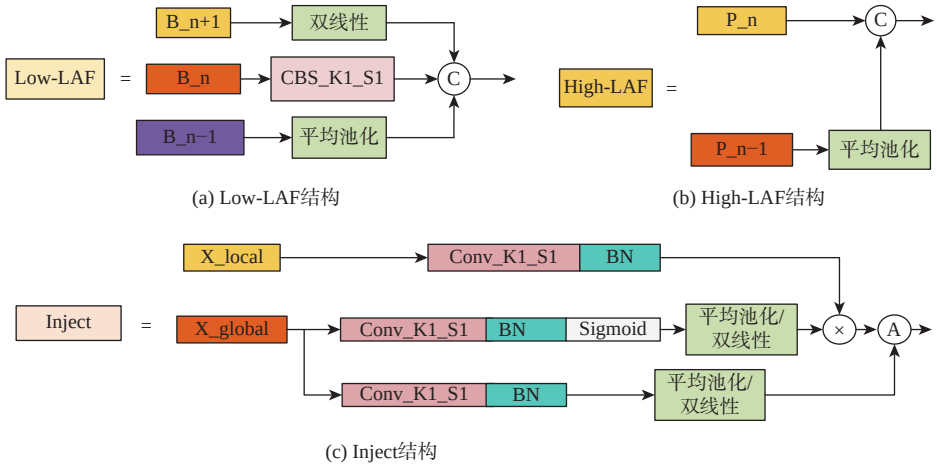


图 5　相邻层融合模块和信息注入模块结构

Fig. 5　Structure of lightweight adjacent layer fusion module and information injection module

别用来融合与输入特征图相邻的 3 层特征和 2 层特征。LAF 模块经过多分支信息融合后, 与经过 GD 机制后输出和分割后的特征图一起输入 Inject 模块中, 即将局部信息和全局信息再一次进行特征融合, 丰富语义特征。Inject 模块中采用在文献 [9] 中进一步改进后的融合模块, 将经过 LAF 模块对齐后的局部特征信息与经过 GD 模块并分割后的全局特征信息进行重加权操作, 最后, 与全局特征相融合, 获得更丰富的梯度流。

虽然 Inject 模块将局部特征和部分全局特征进行了融合, 但其对不同尺度目标的变化并不敏感, 只是进一步强化了对目标的感知能力。基于此, 本文进行进一步改进, 将原先上下整合的重参数卷积模块替换为 C2f 模块, 借鉴文献 [10] 中提出的金字塔注意力机制, 将 BottleNeck 中的第 2 个卷积替换为高效多尺度卷积 (efficient multi-scale convolution, EMSC) 模块, 该模块和替换后的 BottleNeck 结构如图 6 所示。该模块将输入特征图沿通道维度切分为 S 部分, 这里 S=4, 利用多尺度卷积和分组卷积分别对特征图进行卷积操作, 增大了 RF, 在不引入过多计算量的同时, 提升网络对多尺度目标的敏感度, 进一步提升网络性能。

1.3　局部卷积模块

为提升检测速度, 达到实时检测的要求, 引入

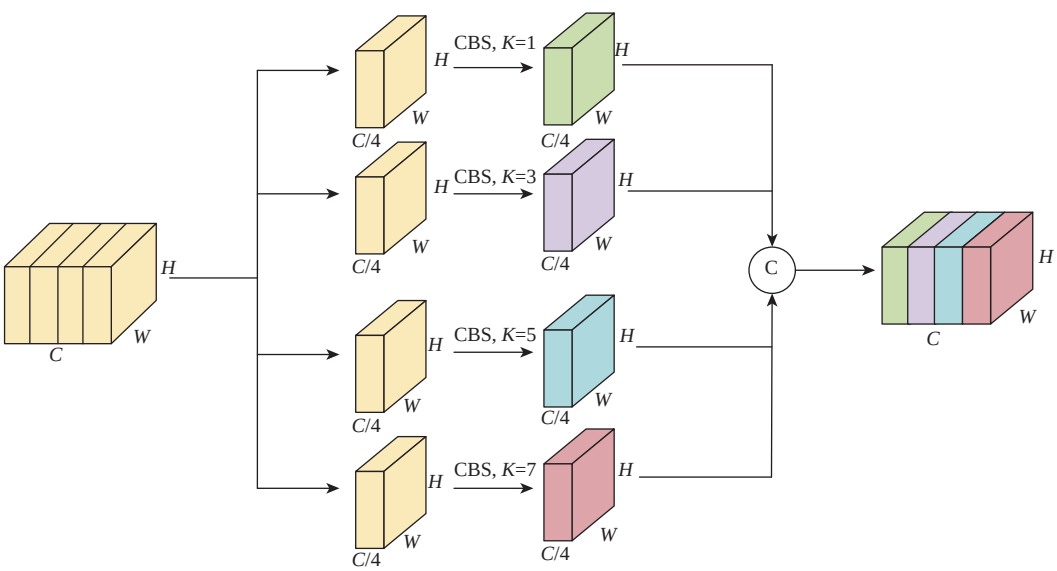


图 6　EMSC 模块和 BottleNeck_EMSC 结构

Fig. 6　EMSC module and BottleNeck_EMSC structure

PConv模块。PConv[11]是一种新型卷积模块，能够使网络在几乎不损失性能的情况下进一步轻量化，其结构如图7(a)所示。文献[11]认为高层网络输出的特征图具有大量的冗余信息，而这些信息会对网络最终的性能产生负面影响，因此，提出利用PConv模块在通道维度上进行部分卷积来进一步提取特征，同时与剩余的通道维度进行拼接，在尽量保证特征信息不丢失的同时，降低模型的参数量和内存访问。文献[11]还提出一种硬件友好的Faster模块，如图7(b)所示，在经过PConv模块特征提取后，接2个卷积操作，为降低计算量，只在第1个卷积后进行批归一化和激活函数处理。本文将Faster模块引入主干网络中的前两个C2f中，在浅层网络中不损失空间信息的情况下，提升了检测速度。本文将Faster模块引入检测头中，如图7(c)所示，在不损失性能的情况下，将双分支中的前2个卷积操作进行合并，最后，分别对目标位置和类别计算损失，大大减少了模型的参数量和计算量。

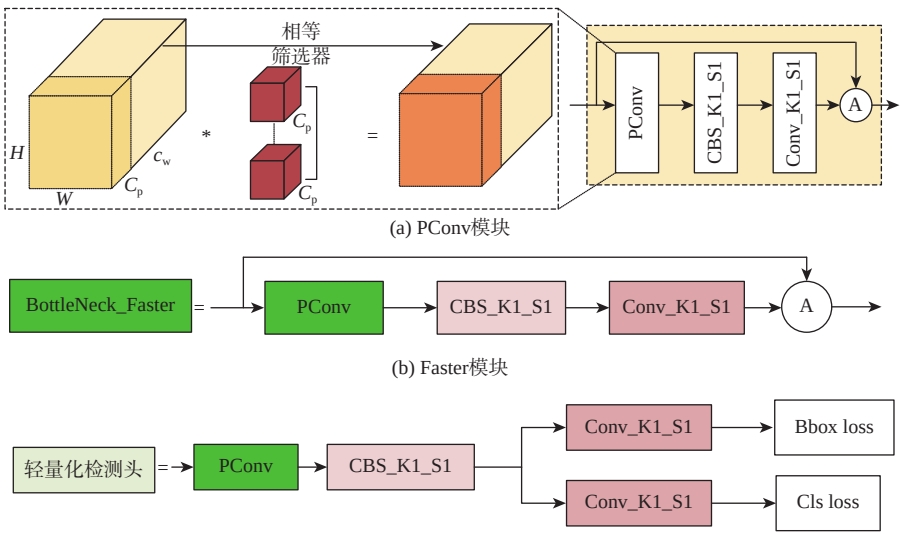


图7 PConv模块、Faster模块和轻量化检测头结构

Fig. 7 PConv module, Faster module and lightweight detection head structure

1.4 优化损失函数

基准网络采用类别损失和回归损失加权的方式来计算总的损失，其中，分类损失采用的是二值交叉熵损失，回归损失采用DFL+CIOU联合计算。对于分类问题，二值交叉熵损失能够很好地对多类别分类问题进行判断，并最终计算损失，因此，本文对回归损失进行了改进。DFL针对遮挡或模糊场景下预测框的边界回归进行了优化，使预测框快速聚焦于与目标位置距离近的分布，通过CIOU对预测框和真实框进行损失计算，使预测框最终拟合。CIOU的相关公式为

$$P_{\mathrm{CIOU}}=P_{\mathrm{IOU}}-\frac{\rho^2(p,p^{\mathrm{gt}})}{c^2}-\alpha v \tag{1}$$

$$L_{\mathrm{CIOU}}=1-P_{\mathrm{IOU}}+\frac{\rho^2(p,p^{\mathrm{gt}})}{c^2}+\alpha v \tag{2}$$

式中：P_{CIOU}为预测框与真实框的交并比；ρ为两框中心点的欧氏距离；p和p^{gt}为两框的中心点；c为包含两框的最小外接矩形的对角线距离；α为惩罚项的参数；v为用来衡量长宽比一致性的参数；L_{CIOU}为CIOU损失函数的最终值。

$$\alpha=\frac{v}{(1-J_{\mathrm{IOU}})+v} \tag{3}$$

$$v=\frac{4}{\pi^2}\left(\arctan\frac{w_{\mathrm{gt}}}{h_{\mathrm{gt}}}-\arctan\frac{w}{h}\right) \tag{4}$$

式中：J_{IOU}为交并比；w_{gt}、h_{gt}为真实框的宽和长；w、h为预测框的宽和高。

虽然CIOU考虑了预测框和真实框的横纵比，但若是真实框和预测框的横纵比相同但大小不同时，CIOU中的惩罚项没有发挥作用，将限制模型的收敛速度和精度。因此，本文引入一种新的损失函数，称为MPDIOU损失[12]，其直接最小化预测框边界与真实框边界之间的左上和右下点的距离，如果预测框和真实框具有相同的宽高比，L_{MPDIOU}值会出现预测框位于真实框内部时低于预测框位于真实框外部的情况，即该损失函数倾向于具有较少冗余的预测框，从而解决了上述问题，提升了精度。此外，文献[12]指出所有的交并比(intersection overunion, IOU)损失都可通过框的左上点坐标和右下点坐标确定，因此，MPDIOU在一定程度上简化了计算过程，加快了收敛速度，其相关公式为

$$d_1=(x_1^{\mathrm{B}}-x_1^{\mathrm{A}})^2+(y_1^{\mathrm{B}}-y_1^{\mathrm{A}})^2 \tag{5}$$

$$d_2=(x_2^{\mathrm{B}}-x_2^{\mathrm{A}})^2+(y_2^{\mathrm{B}}-y_2^{\mathrm{A}})^2 \tag{6}$$

$$P_{\mathrm{MPDIOU}}=\frac{A\cap B}{A\cup B}-\frac{d_1^2}{w^2+h^2}-\frac{d_2^2}{w^2+h^2} \tag{7}$$

$$L_{\mathrm{MPDIOU}}=1-P_{\mathrm{MPDIOU}} \tag{8}$$

式中: $(x_1^{\mathrm{A}},y_1^{\mathrm{A}})$、$(x_2^{\mathrm{A}},y_2^{\mathrm{A}})$为真实框的左上角和右下角坐标; $(x_1^{\mathrm{B}},y_1^{\mathrm{B}})$、$(x_2^{\mathrm{B}},y_2^{\mathrm{B}})$为预测框的左上角和右下角坐标; w、h 为输入原始图像的宽、高; L_{MPDIOU} 为 MPDIOU 损失函数的最终值。根据式 (5)～式 (6), MPDIOU 的示意图如图 8 所示。

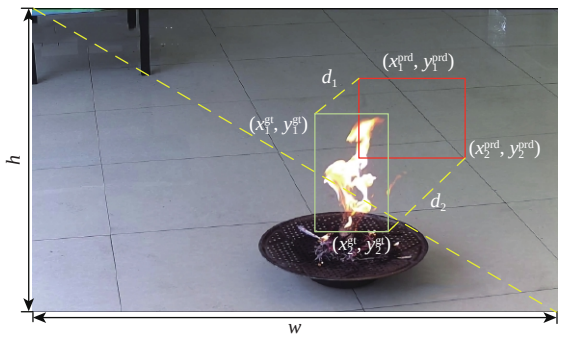


图 8　MPDIOU 示意图

Fig. 8　Schematic diagram of MPDIOU

2　实验与分析

2.1　实验数据集

由于目前缺乏高质量的大型火灾数据集, 本文广泛收集互联网上公开的火灾现场图像和现实模拟火灾场景图像, 建立了一个高多样性的火灾数据集, 共包含火灾现场图像 22 286 张, 其中, 训练集 20 057 张, 测试集 2 229 张, 包含火焰和烟雾 2 类。数据集包含了停车场、地铁站、供电室、学校、工厂等场景下的真实和模拟火灾图像, 部分数据集图像如图 9 所示。

数据集包含大量大、中、小尺度的火焰和烟雾以供训练和测试, 根据情境中的常见物体 (common objectsin content, COCO) 对尺度的划分原则进行尺度划分, 即目标像素在 32 像素×32 像素以下时为小尺度目标, 目标像素在 32 像素×32 像素和 96 像素×96 像素之间时为中尺度目标, 目标像素在 96 像素×96 像素以上时为大尺度目标。训练集和测试集包含的目标尺度如图 10 所示, 训练集下的小尺度目标为 4 857 个, 中尺度目标为 8 994 个, 大尺度目标为 6 836 个。测试集下的小尺度目标为 529 个, 中尺度目标为 871 个, 大尺度目标为 885 个。为验证模型的泛化能力, 在划分训练集与测试集时选择不同场景。



(a) 烟雾场景



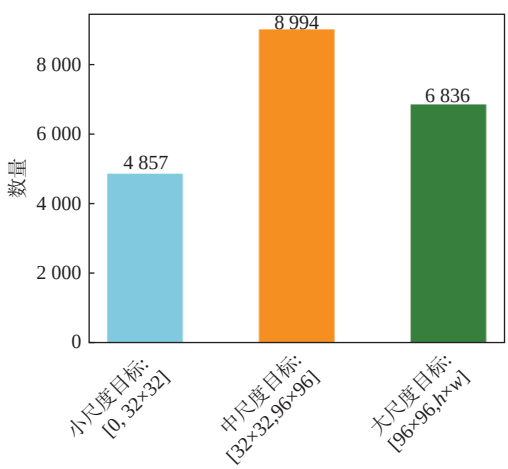
(b) 火焰场景



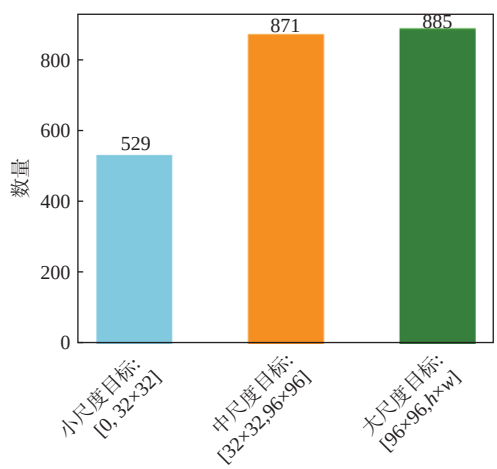
(c) 火焰+烟雾场景

图 9　数据集部分图像

Fig. 9　Partial images of the dataset



(a) 训练集目标分布情况



(b) 测试集目标分布情况

图 10　目标分布情况

Fig. 10　Distribution of targets

2.2　实验环境和超参数配置

本文训练时的实验环境如表 1 所示。采用随机梯度下降优化器进行迭代训练, 图像尺度为 640 像素, 批数量为 16, 训练轮数为 100 轮, 动量和

权重衰减因子分别为0.937和0.000 5，初始学习率为0.01，采用余弦退火策略使最终学习率为初始的10%，其余超参数采用基准网络的默认参数。训练过程中采用亮度变化和Mosaic-4等数据增强方式，以解决过拟合和样本不平衡的问题。

表1 实验环境配置

Table 1 Experiment environment configuration

实验环境	环境条件
CPU	Intel Xeon Silver 4310
GPU	NVIDIA A40
操作系统	Ubuntu 18.04.6
编程语言	Python3.11.4
深度学习框架	Pytorch2.0.1

2.3 评价指标

本文采用mAP50、mAP50:95、参数量、浮点运算速度/($10^9 \cdot s^{-1}$)来评价模型性能。mAP50、mAP50:95的计算需要引入N_{TP}、N_{TN}、N_{FP}、N_{FN}这4个概念，其中，mAP50表示在阈值为0.5时的平均精度均值，mAP50:95表示阈值从0.5以步长0.05逐步上升为0.95时的平均精度均值。相较于mAP50，mAP50:95更加关注整体的边界回归情况，mAP50:95值越高，预测框与真实框的拟合越精确。N_{TP}为将正样本检测为正样本的数量，N_{TN}为将负样本检测为负样本的数量，N_{FP}为将负样本检测为正样本的数量，N_{FN}为将正样本检测为负样本的数量。由此可以得到精确率和召回率的定义为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{9}$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{10}$$

式中：P为精确率；R为召回率。

2.4 消融实验与模块对比分析

2.4.1 消融实验与分析

为验证本文提出的各个改进模块对整体网络的贡献，设计了对应的消融实验，实验结果如表2所示。改进1是在主干网络中加入了感受野卷积注意力模块并组成了C2f_RF模块，mAP50、mAP50:95分别提升了1%、1.4%，证明了加入感受野卷积注意力模块后的C2f拥有更强的特征提取能力，且计算量和参数量只是略微增加；改进2是在改进1的基础上将PANet替换为本文改进的强特征融合网络，mAP50、mAP50:95分别提升了0.5%、0.9%，说明本文改进的强特征融合网络能够更加充分地融合全局和局部特征，引入EMSC模块后，能够更加充分地融合通道信息，综合提高模型的特征融合能力；改进3是在改进2的基础上引入PConv模块来改进主干网络和检测头，使其参数量和计算量相比于改进2减少了11%和27%，同时，相比于基准算法，其检测性能仍然得到提升；本文算法在改进3的基础上将IOU损失由CIOU替换为MPDIOU，mAP50、mAP50:95分别提升了0.6%、0.6%，在加快收敛速度的同时，提升了模型的检测能力。

表2 消融实验比较分析

Table 2 Comparative analysis of ablation experiment

模型	感受野卷积注意力模块	强特征融合网络	PConv	MPDIOU	mAP50/%	mAP50:95/%	参数量	浮点运算速度/$10^9 \cdot s^{-1}$
标准算法					95.2	69.1	11.1×10^6	28.6
改进1	√				96.2	70.5	11.2×10^6	28.8
改进2	√	√			96.7	71.4	15.1×10^6	33.0
改进3	√	√	√		96.7	71.4	13.5×10^6	24.2
本文算法	√	√	√	√	97.3	72.0	13.5×10^6	24.2

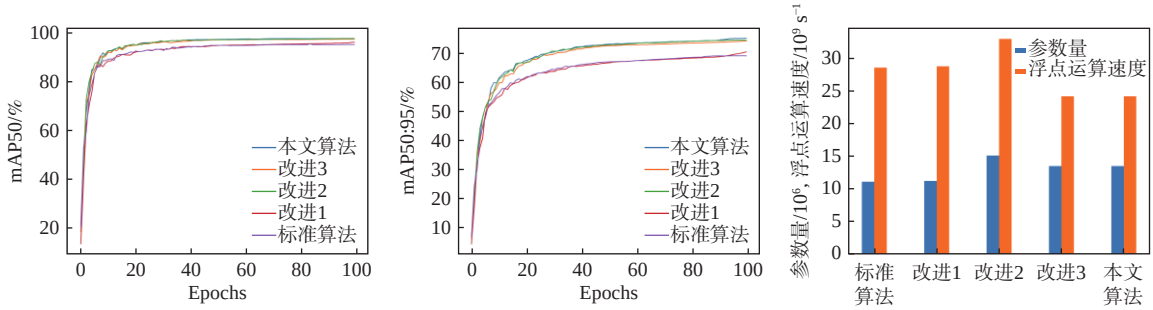
进一步，将消融实验进行可视化对比，对比结果如图11所示。可以看出，相比于标准算法，本文算法取得了明显提升，mAP50总体提升了2.1%，mAP50:95总体提升了2.9%，在参数量略微增加的情况下，计算量有所下降。

2.4.2 感受野卷积注意力模块比较分析

为验证本文感受野卷积注意力模块的有效性，分别将GAM、CBAM和感受野卷积注意力嵌入C2f的BottleNeck中，替换BottleNeck中的第2个卷积，并进行实验，保持实验环境配置和其余网络结构不变，实验结果如表3所示[13-14]。结果表明，本文提出的感受野卷积注意力模块相比于GAM和CBAM，能够更好地提取目标特征，在解决参数共享的问题后，进一步提升了模型的性能。

2.4.3 颈部网络比较分析

为验证本文改进的强特征融合网络的有效性，将基准算法的颈部网络分别替换为文献[7]的融合网络和本文改进的强特征融合网络，并与基准算法在自建火灾数据集上进行对比实验，保持实验环境配置和其余网络结构不变，实验结果如表4所示。由此可以证明本文的强特征融合网络在性能上的优越性。



(a) mAP50变化情况　(b) mAP50:95变化情况　(c) 参数量和计算量对比

图 11　消融实验可视化对比

Fig. 11　Visualization comparison of ablation experiments

表 3　注意力模块比较结果

Table 3　Attention module comparison results

模型	mAP50/%	mAP50:95/%
标准算法	95.2	69.1
+GAM[13]	96.0	70.0
+ CBAM[14]	95.7	69.7
+感受野卷积注意力	96.2	70.5

表 4　颈部网络比较结果

Table 4　Comparison results of the neck network

模型	mAP50/%	mAP50:95/%
标准算法	95.2	69.1
+文献[7]	96.2	71.2
+强特征融合网络	96.5	71.3

2.4.4　损失函数比较分析

保持实验环境配置和其余的网络结构不变, 针对回归损失函数对本文算法精度影响的问题, 将基准网络的回归损失函数分别替换为焦点与高效交并比 (focal and efficient inter section over union, Focal EIOU)、形状感知交并比 (shape-aware intersection over union, SIOU) 和最小点距离交并比 (minimum point distance intersection over union, MPDIOU), 并在自建火灾数据集上进行实验, 结果如表 5 所示[15-16]。可以看出, MPDIOU 的 mAP50 和 mAP50:95 指标更高, 模型性能更好, 因此, 选用 MPDIOU 为本文算法的回归损失函数。

表 5　损失函数对比实验结果

Table 5　Comparison experiment results of loss function

损失函数	mAP50/%	mAP50:95/%
标准算法(CIOU)	95.2	69.1
+Focal EIOU[15]	95.5	69.1
+SIoU[16]	95.8	69.0
+MPDIOU[11]	96.3	69.7

2.5　不同算法对比实验与分析

为验证本文算法的有效性, 在本文建立的数据集上将本文算法与经典目标检测算法进行对比实验, 其他算法的结果均通过其官方代码实验得到, 结果如表 6 所示[17-24]。可以看出, 本文算法的检测精度相比于单阶段和双阶段的经典目标检测算法更高, mAP50 达到了 97.3%, mAP50:95 达到了 72.0%, 检测速度可达到 40.9 帧/s, 满足实时性的要求, 可以更好地适应多尺度的火灾检测。

表 6　不同算法模型的检测结果对比

Table 6　Comparison of detection results of different algorithm models

算法	mAP50/%	mAP50:95/%	参数量	浮点运算速度/($10^9 \cdot s^{-1}$)
Faster R-CNN[17]	88.2	51.5	41.1×10^6	91.0
SSD512[18]	95.7	62.1	24.5×10^6	87.9
YOLOv3-tiny[19]	94.4	61.0	8.7×10^6	13.0
YOLOv5m	96.2	68.8	21.2×10^6	49.0
YOLOv6s[20]	96.7	71.0	18.5×10^6	45.3
YOLOv7-tiny[21]	96.2	69.4	6.1×10^6	13.2
YOLOv8s	95.2	69.1	11.1×10^6	28.6
YOLOv9-T[22]	96.6	70.9	2.7×10^6	11.0
Deformable-DETR[23]	95.7	60.1	40.0×10^6	173.0
RT-DETR-Res18[24]	96.1	70.4	20.0×10^6	60.5
本文算法	97.3	72.0	13.5×10^6	24.2

为验证本文算法的效果, 将本文算法在数据集上不同尺度的实验结果进行可视化, 结果如图 12 所示。可以看出, 对于多种尺度的烟雾和火焰, 无论是在白天和夜晚、有遮挡和无遮挡的情况下, 本文算法均拥有优异的检测性能, 能够高效地进行多尺度的火灾检测。

2.6　公共数据集对比实验与分析

在 Pascal VOC 07+12 数据集上进行进一步的对比实验, Pascal VOC 07+12 数据集共有 20 个类别, 训练集由 Pascal VOC 07 的训练集和验证集、Pascal VOC 12 的训练集和验证集组成, 共 16 651 张图像。该数据集的测试集为 Pascal VOC 07 的测试集, 共 4 952 张图像。由于火焰和烟雾具有形状多变、尺度变化大等特点, 本文的自建火灾数据集对检测算法的多尺度提取和融合能力提出了更高要求, 而 Pascal VOC 07+12 数据集虽然是固定物体,



(a) 烟雾场景



(b) 火焰场景



(c) 火焰+烟雾场景

图 12 算法检测效果可视化对比

Fig. 12 Visualization comparison of the detection effects of algorithm

形状变化不大，但更多的类别和同一图像中目标的多样性同样满足了多尺度的要求。相比于本文自建的火灾数据集，Pascal VOC 07+12 数据集的类别更多，不同类别的尺度差异更大，能更好地衡量算法对不同目标、不同尺度的检测能力。

本文算法与其他算法的实验结果对比如表 7 所示[25]，与基准算法 YOLOv8s 相比，本文算法的 mAP50 提升了 1.4%，mAP50:95 提升了 2.4%，达到了 78.0% 和 57.8%，相比于其他的 CNN 和 Transformer 目标检测算法，本文算法在保持较低参数量和计算量的同时，拥有最好的精度指标，具有较强的泛化性能。

表 7 不同算法模型在公共数据集上的检测结果对比

Table 7 Comparison of detection results of different algorithm models on public datasets

算法	mAP50/%	mAP50:95/%	参数量	浮点运算速度/($10^9 \cdot s^{-1}$)
Faster-RCNN[17]	65.9	36.5	41.2×10^6	91.1
SSD512[18]	65.4	37.6	27.2×10^6	90.4
YOLOv3-tiny[19]	54.4	27.9	8.7×10^6	13.1
YOLOv5m	77.5	53.5	21.2×10^6	49.1
YOLOv6s[20]	75.8	53.5	18.5×10^6	45.4
YOLOv7-tiny[21]	70.1	43.4	6.1×10^6	13.3
YOLOX-s[25]	72.3	44.7	8.9×10^6	26.7
YOLOv8s	76.6	55.4	11.1×10^6	28.7
YOLOv9-T[22]	71.8	52.2	2.7×10^6	11.1
Deformable-DETR[23]	77.5	52.7	40.0×10^6	173.0
RT-DETR-Res18[24]	72.4	53.1	20.0×10^6	60.5
本文算法	78.0	57.8	13.5×10^6	24.3

3 结 论

本文提出了一种空间特征优化和多尺度特征融合的火灾检测算法，具有较强的泛化性。主要工作为：

1）提出了感受野卷积注意力模块，解决卷积参数共享的问题，提取到更多的特征，提升模型性能；将颈部网络替换为本文改进的强特征融合网络，将全局特征和局部特征充分融合，提升了网络对多尺度目标的敏感度。

2）引入 PConv，在几乎不损失性能的情况下，使整体网络轻量化，使网络满足实时性的要求；优化回归损失函数，将 CIOU 替换为 MPDIOU，解决了预测框和真实框宽高比一致但大小不一致时的问题，加快了收敛速度，提升模型性能。

3）将本文改进算法在自建的火灾数据集上进行性能评估，其 mAP50、mAP50:95 高于经典的目标检测算法，相比于原始的 YOLOv8s 基准算法，mAP50 提升了 2.1%，mAP50:95 提升了 2.9%，在满足实时性的同时，拥有更好的精度，能够更好地进行火灾检测；将本文改进算法在 Pascal VOC 07+12 公共数据集进行评估，其性能仍然优于其他算法，相比于原始的 YOLOv8s 基准算法，mAP50 提升了 1.4%，mAP50:95 提升了 2.4%，证明了本文改进算法的泛化性。

参考文献（References）

[1] SHARMA J, GRANMO O C, GOODWIN M, et al. Deep convolutional neural networks for fire detection in images[C]//Proceedings of the Engineering Applications of Neural Networks. Berlin: Springer, 2017: 183-193.

[2] HOSSEINI A, HASHEMZADEH M, FARAJZADEH N. UFS-Net: a unified flame and smoke detection method for early detection of

fire in video surveillance applications using CNNs[J]. Journal of Computational Science, 2022, 61: 101638.

[3] 张融, 张为. 基于改进 GhostNet-FCOS 的火灾检测算法[J]. 浙江大学学报 (工学版), 2022, 56(10): 1891-1899.
ZHANG R, ZHANG W. Fire detection algorithm based on improved GhostNet-FCOS[J]. Journal of Zhejiang University (Engineering Science), 2022, 56(10): 1891-1899(in Chinese).

[4] 秦瑞, 张为. 一种无锚框结构的多尺度火灾检测算法[J]. 西安电子科技大学学报 (自然科学版), 2022, 49(6): 111-119.
QIN R, ZHANG W. Multi-scale fire detection algorithm with an anchor free structure[J]. Journal of Xidian University (Natural Science) and Technology, 2022, 49(6): 111-119(in Chinese).

[5] ZHANG X, LIU C, YANG D G, et al. RFAconv: innovating spatial attention and standard convolutional operation[EB/OL]. (2023-04-06)[2024-01-09]. https://doi.org/10.48550/arXiv.2304.03198.

[6] WANG K X, LIEW J H, ZOU Y T, et al. PANet: few-shot image semantic segmentation with prototype alignment[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 9196-9205.

[7] WANG C C, HE W, NIE Y, et al. Gold-YOLO: efficient object detector via gather-and-distribute mechanism[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. New York: ACM, 2023: 51094-51112.

[8] DING X H, ZHANG X Y, MA N N, et al. RepVGG: making VGG-style ConvNets great again[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 13728-13737.

[9] WAN Q, HUANG Z L, LU J C, et al. Seaformer: squeeze-enhanced axial transformer for mobile visual recognition[EB/OL]. (2023-01-30)[2024-01-11]. https://arxiv.org/abs/2301.13156.

[10] ZHANG H, ZU K K, LU J, et al. EPSANet: an efficient pyramid squeeze attention block on convolutional neural network[C]//Proceedings of the Computer Vision –ACCV 2022. Cham: Springer, 2023: 541-557.

[11] CHEN J R, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 12021-12031.

[12] MA S L, XU Y. MPDIoU: a loss for efficient and accurate bounding box regression[EB/OL]. (2023-07-14)[2024-04-15]. https://arxiv.org/abs/2307.07662.

[13] LIU Y C, SHAO Z R, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions[EB/OL]. (2024-12-10)[2024-01-26]. https://arxiv.org/abs/2112.05561.

[14] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the Computer Vision–ECCV 2018. Cham: Springer, 2018: 3-19.

[15] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.

[16] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression[EB/OL]. (2022-05-25)[2024-01-30]. https://arxiv.org/abs/2205.12740.

[17] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE Press, 2016: 1137-1149.

[18] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the Computer Vision–ECCV 2016. Cham: Springer, 2016: 21-37.

[19] ADARSH P, RATHI P, KUMAR M. YOLOv3-tiny: object detection and recognition using one stage improved model[C]//Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems. Piscataway: IEEE Press, 2020: 687-694.

[20] LI C Y, LI L L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07)[2024-02-03]. https://arxiv.org/abs/2209.02976.

[21] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 7464-7475.

[22] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information[EB/OL]. (2024-02-29)[2024-03-05]. https://arxiv.org/abs/2402.13616.

[23] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable transformers for end-to-end object detection[EB/OL]. (2021-03-18)[2024-03-06]. https://arxiv.org/abs/2010.04159.

[24] ZHAO Y A, LV W Y, XU S L, et al. DETR beat YOLOs on real-time object detection[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 16965-16974.

[25] WANG H, WANG L H, CHEN H, et al. Waste-YOLO: towards high accuracy real-time abnormal waste detection in waste-to-energy power plant for production safety[J]. Measurement Science and Technology, 2024, 35(1): 016001.

Fire-and-smoke detection algorithm based on convolutional attention and feature fusion

TIAN Jiaqi，QIN Guoxuan，ZHANG Wei*

(School of Microelectronics，Tianjin University，Tianjin 300072，China)

Abstract：Aiming at the imbalance between the accuracy and speed of fire-and-smoke detection in real-world scenarios, this paper proposes a fire-and-smoke detection algorithm that strengthens spatial feature extraction and multi-scale feature fusion. We improved the extraction of high-level semantic information from the backbone network by embedding the receptive field convolutional attention module into it, enhancing the model's feature extraction capability. Additionally, we introduced an enhanced strong feature fusion network to further strengthen the fusion of low-level spatial information and high-level semantic information, thereby improving model accuracy. In order to satisfy real-time needs, we additionally used the partial convolution (PConv) module to lightweightly enhance the detection head and backbone network, lowering the model's parameter count and memory access without compromising accuracy. Furthermore, we adjusted the regression loss function to enhance the model's detection capabilities. According to experimental results, the suggested algorithm improves the mean average precision at IoU threshold 0.5 (mAP50) and 0.5:0.95 (mAP50:95) by 2.1% and 2.9%, respectively, demonstrating its superiority in the field of fire-and-smoke detection. Additionally, the mAP50 and mAP50:95 on the Pascal VOC 07+12 public dataset are increased by 1.4% and 2.4%, respectively, demonstrating the algorithm's good generalization performance.

Keywords：fire-and-smoke detection；multi-scale features；feature fusion；lightweight detection head；loss function

Received：2024-03-26；**Accepted**：2024-07-30；**Published Online**：2024-08-08 09：08

URL：link.cnki.net/urlid/11.2625.V.20240807.1402.002

Foundation item：National Key Research and Development Program of China (2022YFC3006302)

* **Corresponding author**. E-mail：tjuzhangwei@tju.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0188

考虑动态重构机制的IMA可靠性建模和评估方法

胡耕硕，焦健*，胡朗霄，景永峰
(北京航空航天大学 可靠性与系统工程学院，北京 100191)

摘　　要：为解决综合模块化航空电子系统（IMA）动态重构机制导致其可靠性难以建模和评估的问题，分析了IMA的组成结构及资源备份和资源抢占2种动态重构机制。根据分析结果，在现有总线性能共享多状态系统模型的基础上，考虑动态重构机制下的计算资源再分配问题，建立适用于动态重构机制下IMA的可靠性模型。同时，提出一种基于通用生成函数（UGF）的动态重构机制下IMA可靠性评估算法，对不同动态重构机制下系统可靠度进行量化评估。以某型直升机IMA中综合任务处理系统为对象进行案例应用研究，分析其备用通用处理单元数量变化对系统可靠度的影响规律，并提出最优化设计方案，以验证所提可靠性建模和评估方法的可行性。

关　键　词：直升机综合模块化航空电子系统；动态重构；多状态系统；通用生成函数；可靠性
中图分类号：TB114.3
文献标志码：A　　**文章编号：**1001-5965(2026)05-1767-10

航电系统是飞机的核心部分，为飞机提供人机交互、环境感知等不可或缺的功能。目前航电系统的成本占飞机设计制造总成本的比例越来越大，民用飞机占比35%～40%，军用飞机占比50%以上[1]。自20世纪80年代以来，综合模块化航空电子系统(integrated modular aviorics, IMA)因具备降低生产成本、提高性能等优势，逐渐成为航电系统设计的主流方向，并在波音787、空客380、F-22等飞机上得到了广泛应用[2]。但是，由于其功能结构复杂，在任务执行过程中，对IMA的高可靠性、高容错能力等方面的需求不断增加[3]。

目前，许多学者致力于动态重构过程的可靠性建模和评估方法研究。例如，丁明[4]基于系统建模语言(systems modeling language, SysML)和架构分析与设计语言(architecture analysis and design language, AADL)提出一种基于模型的重构过程安全性评估方法。研究人员将故障树与Petri网相结合的方法[5]和有限状态机方法[6]应用于重构过程可靠性建模评估工作。此外，也有研究人员采用AADL与Petri网相结合的方法[7]和AADL[8]方法来评估动态重构过程的可靠性和安全性。然而，目前研究大多集中于动态重构过程，缺乏有关IMA可靠性方面的研究。

由于IMA集成性和模块化的特点[9]，系统中各通用处理单元可以通过通信系统总线实现任务的再分配，以应对通用处理单元性能下降或故障的情况，进而提高系统的可靠性。在IMA工作过程中，当系统中通用处理单元的数据处理与运算能力无法满足其数据处理与运算需求时，该通用处理单元可以将需要处理的应用配置数据通过总线传输至数据处理与运算速度快或任务需求低的通用处理单元进行处理，以确保系统可以正常运行。如果在系统可靠性建模过程中，忽略这一过程，评估出来的系统可靠性可能偏小于真实结果。

动态重构机制由于其可以有效提升IMA安全性和可靠性的优点，成为系统设计过程中不可或缺的关键技术。因此，在进行IMA可靠性建模研究过程中，考虑动态重构机制是有必要的。

由于IMA的结构和工作原理与总线性能共享

收稿日期：2024-04-01；录用日期：2024-06-07；网络出版时间：2024-06-29 17:49
网络出版地址：link.cnki.net/urlid/11.2625.V.20240628.0920.001
*通信作者. E-mail：jiaojian@buaa.edu.cn

引用格式：胡耕硕，焦健，胡朗霄，等. 考虑动态重构机制的IMA可靠性建模和评估方法[J]. 北京航空航天大学学报，2026，52(5)：1767-1776. HU G S，JIAO J，HU L X，et al. Reliability modeling and evaluation method of IMA under dynamic reconfiguration[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(5)：1767-1776 (in Chinese).

多状态系统相似，且两者都涉及单元间通过总线进行资源重新分配的过程。因此，可以通过扩展总线性能共享多状态系统模型，构建适用于动态重构机制下 IMA 的可靠性模型。

总线性能共享多状态系统最初是由 Lisnianski 等[10]提出的，研究了由一个主单元和一个备用单元组成的多态系统的可靠性问题，其中，备用单元的剩余性能可以通过单向传输系统传输至主单元。随后，Levitin[11]将性能共享的概念扩展到多单元系统，允许系统中单元的剩余性能通过传输系统共享给任何存在性能不足的单元。近年来，总线性能共享多状态系统的可靠性建模引起了业界的广泛关注。研究人员对考虑性能共享机制多状态模型的扩展进行了广泛研究[12-17]。这些研究覆盖了多状态系统、可靠性建模评估、可用性建模评估[18]及系统优化分配和维护问题[19]等多个方面。在相关文献研究中，目前少有考虑动态重构机制和计算资源再分配顺序问题。然而，在 IMA 建模和评估过程中，动态重构机制和计算资源分配顺序[18]问题是不可忽略的关键因素。

因此，针对 IMA 可靠性建模和评估问题，本文的研究内容包括 2 个方面：①在现有总线性能共享机制多状态系统模型的基础上，考虑动态重构机制和计算资源分配顺序，对其进行扩展研究，提出一种适用于动态重构机制下 IMA 的可靠性模型；②基于上述模型，运用通用生成函数 (universal generating function,UGF) 技术，提出基于 UGF 的系统可靠性评估算法，实现系统可靠性评估。

本文在对 IMA 物理结构和重构机制进行分析的基础上，扩展了总线性能共享多状态系统模型，在考虑计算资源分配顺序和动态重构机制的基础上，建立适用于不同动态重构机制下 IMA 可靠性模型。基于可靠性模型，运用 UGF 技术，评估不同动态重构机制下 IMA 的可靠性。以某直升机的 IMA 为例，对其进行可靠性及变化规律研究，并提出备用通用处理单元的优化设计方案。

1 系统结构和动态重构策略分析

以 IMA 为代表的新型航电系统，将人机交互和显示等功能集成到一个由多种共享计算资源组成的平台，且平台中的计算资源可以支持多个应用程序共同运行。IMA 中各部分的交联逻辑关系如图 1 所示。通用处理单元上的计算资源支持应用程序的运行。根据 ARINC653[20]和 ASAAC 标准[21]，通用处理单元的体系架构可以分为 2 层：上层为应用程序层，用于满足各项飞行任务中的功能需求；下层为实时操作层，主要负责计算资源的管理，并为应用程序提供执行平台[22-21]。同时，应用程序的分区设计可以防止故障的蔓延和转移。针对一组涵盖不同功能的飞行任务，需要通过预定配置方案来指定这些应用程序如何在平台上进行部署。系统中的通用处理单元分为主通用处理单元和普通通用处理单元，重要程度高且影响系统安全的应用程序都驻留在主通用处理单元上。

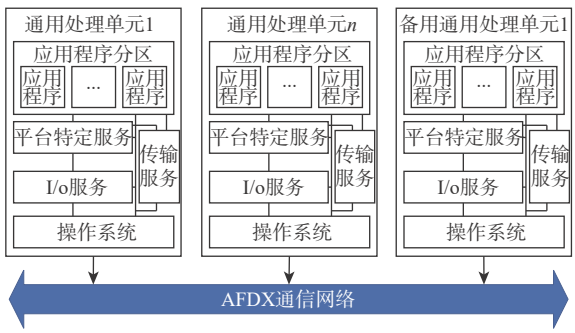


图 1 IMA 交联关系

Fig. 1 The link relationship of IMA

IMA 通过调配通用处理单元的计算资源，实现预设应用程序的正常运行。当通用处理单元出现故障时，IMA 可通过资源备份和资源抢占 2 种重构机制，将系统恢复至正常运行状态或安全状态。资源备份机制是指当通用处理单元出现故障，且备用通用处理单元数量充足时，其可以将数据处理与运算任务能力需求未满足的应用程序相关配置数据通过系统总线传输至备用通用处理单元进行计算处理，以确保系统可以正常执行飞行任务。

资源抢占机制是指当通用处理单元出现故障，但可用备用通用处理单元数量不足时，IMA 已无法正常执行任务，但仍需确保飞行安全，系统将停止通用处理单元中重要程度最低且不影响系统安全的应用程序的运行，而驻留应用程序中重要程度较高且存在数据处理与运算能力需求未被满足的应用程序，将对其计算资源进行抢占，以确保影响飞机安全的应用程序可以正常运行[23]。

动态重构机制显然可以提高 IMA 的任务可靠性和安全性，但仍需建立系统可靠性模型，提出可靠性评估方法，定量评估重构机制下 IMA 的可靠性水平，从而为进一步优化系统设计提供支撑。

2 IMA 可靠性建模研究

2.1 模型结构描述

由于 IMA 集成化和模块化的特点，IMA 本质上可以被视为一个动态重构机制下总线性能共享多状态系统，系统通过通用处理单元之间计算资源

的共享, 实现应用程序的正常运行。因此, 本文对现有总线性能共享多状态系统模型进行扩展, 建立适用于不同动态重构机制下 IMA 的可靠性模型, 如图 2 所示。

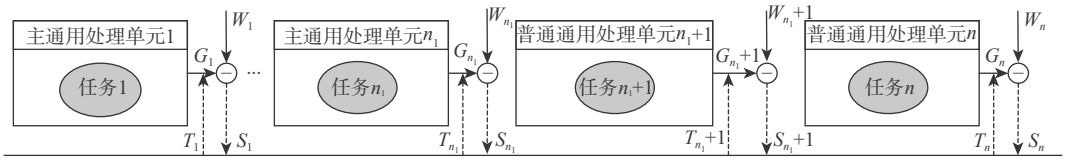


图 2　IMA 可靠性模型

Fig. 2　Reliability model of IMA

该模型由 n 个通用处理单元和备用处理单元组成, 其中, 通用处理单元 1 至通用处理单元 n_1 为主通用处理单元, 其余通用处理单元为普通通用处理单元。驻留在通用处理单元上的应用程序根据重要程序由高到低, 按照通用处理单元的序号由小到大依次配置在通用处理单元上, 其中, 任务 1 为重要程度最高的应用程序, 任务 n 为重要程度最低的应用程序。同时, 影响系统安全的应用程序均驻留在主通用处理单元上。

每个通用处理单元具有多种状态, 通用处理单元的数据处理与运算能力 G_i 从集合 $\{g_{i,1}, g_{i,2}, \cdots, g_{i,H_i}\}$ 中取值; 通用处理单元的数据处理与运算能力需求 W_i 从集合 $\{w_{i,1}, w_{i,2}, \cdots, w_{i,P_i}\}$ 中取值。当系统中各个通用处理单元的数据处理与运算能力满足任务能力需求 $W_i (G_i \geqslant W_i)$ 时, IMA 处于正常工作状态。

当系统通用处理单元自身的数据处理与运算能力无法满足任务能力需求, 即 $G_i < W_i$ 时, 系统将通过触发动态重构机制, 以使系统恢复至任务状态或安全运行状态。系统将按照存在任务需求未被满足的通用处理单元承担应用程序的重要程度由高至低, 依次将未完成的数据处理与运算任务 (需求) 通过通信系统总线分配给存在数据处理与运行能力剩余的通用处理单元。通信系统的传输能力 C 从集合 $\{c_1, c_2, \cdots, c_k\}$ 中取值。由于受通信系统总线带宽的限制, 通用处理单元之间只能对通信总线传输能力 C 范围内的数据处理与运算任务进行再分配。

当通用处理单元出现故障, 且备用通用处理单元数量充足时, 这些通用处理单元将按照其驻留应用程序的重要程度由高至低, 依次利用传输能力有限的总线向备用通用处理单元寻求帮助。当备用通用处理单元的数据处理与运算能力满足故障通用处理单元的任务需求, 且处于正常工作状态通用处理单元的数量满足系统规定数量时, 系统恢复至正常执行任务状态; 在这种动态重构机制中, 由于不同类型的通用处理单元在数据处理与运算能力上存在差异, 主通用处理单元通常可以替代 1～2 个普通通用处理单元, 2 个普通通用处理单元可以代替 1 个主通用处理单元。

当备用通用处理单元数量不足时, IMA 将终止重要程度最低且不影响系统安全的应用程序。存在任务需求未被满足的通用处理单元将根据自身驻留应用程序重要程度由高至低, 依次对该部分节省下来的计算资源进行抢占。当驻留影响系统安全应用程序的通用处理单元的数据处理与运算需求均被满足, 且处于工作状态通用处理单元的数量满足系统安全的最低要求时, 系统恢复至安全状态。资源备份过程和资源抢占过程中, 通用处理单元之间的数据传输会受到总线传输能力的限制。

根据系统中通用处理单元的数据处理与运算任务需求满足情况及处于工作状态通用处理单元的数量, IMA 可以分别处于正常工作、任务失败但安全及危险等不同状态, 进而可以评估其可靠性。

2.2　IMA 可靠性模型研究

基于系统中通用处理单元的数据处理与运算能力必须满足任务需求的数学逻辑关系。每个通用处理单元的数据处理与运算能力剩余 S_i 和能力不足 D_i 为

$$S_i = \max(G_i - W_i, 0) \tag{1}$$

$$D_i = \max(W_i - G_i, 0) \tag{2}$$

系统中各通用处理单元数据处理与运算能力剩余的集合 S 和数据处理与运算能力不足的集合 D 为

$$S = \{S_1, S_2, \cdots, S_n\} \tag{3}$$

$$D = \{D_1, D_2, \cdots, D_n\} \tag{4}$$

由于通信系统总线传输能力 C 有限, 可被重新分配的数据处理与运算能力剩余为

$$T = \min\left(\sum_{i=1}^{n} S_i, C\right) = \min\left(\sum_{i=1}^{n} \max(G_i - W_i, 0), C\right) \tag{5}$$

可用于重新分配的数据处理与运算能力剩余 (各通用处理单元数据处理与运算能力剩余和备用

处理单元数据处理与运算能力S_R)将按照通用处理单元上驻留应用程序的重要程度由大到小依次进行再分配。数据处理与运算能力剩余的再分配法则如图3所示。

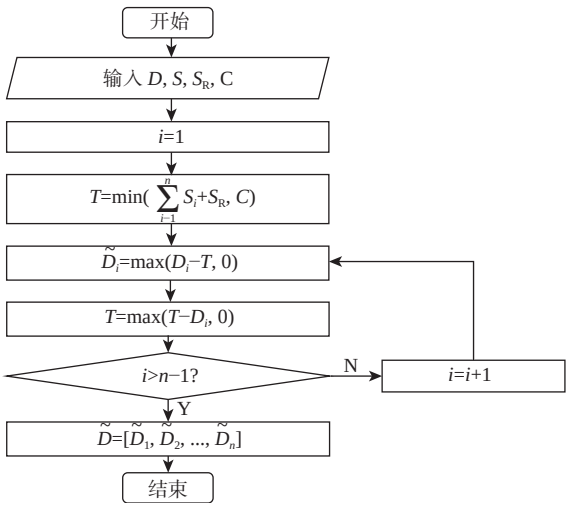


图3 考虑重要度的资源备份机制下数据处理与运算能力剩余分配法则

Fig. 3 Performance surplus allocation rule under resource backup mechanism considering importance

根据上述分配法则获得,资源备份机制下系统中各通用处理单元仍存在的数据处理与运算能力不足集合$\tilde{D}$为

$$\tilde{D}=\{\tilde{D}_1,\tilde{D}_2,\cdots,\tilde{D}_n\} \tag{6}$$

当系统触发资源备份机制成功,系统将恢复至正常执行飞行任务状态。因此,在资源备份机制下,IMA的任务可靠度R_1可以表示为系统的任务需求都可以被满足,且处于可工作状态下的主通用处理单元和普通通用处理单元的数量都满足系统规定数量时的概率,即

$$R_1=\Pr\left\{\sum\tilde{D}=0,\tilde{N}_1\geqslant n_1,\tilde{N}_2\geqslant n_2\right\} \tag{7}$$

式中:$\tilde{N}_1$为资源备份机制下,系统中处于正常工作状态的主通用处理单元及其调用的备用通用处理单元数量;$\tilde{N}_2$为资源备份机制下,系统中处于正常工作状态的普通通用处理单元及其调用的备用通用处理单元数量;n_1为主通用处理单元数量;n_2为普通通用处理单元数量。

在资源抢占机制下,重要程度低且不影响系统安全的应用程序将会被终止,以确保影响安全且存在数据处理与运算任务需求未被满足的应用程序可以通过抢占该部分节省下来的计算资源而保持正常运行状态。资源抢占机制下系统的数据处理与运算能力剩余分配法则如图4所示。

根据上述数据处理与运算能力剩余分配法则,资源抢占机制下系统中各通用处理单元存在的数据处理与运算能力不足的集合$\hat{D}$可以表示为

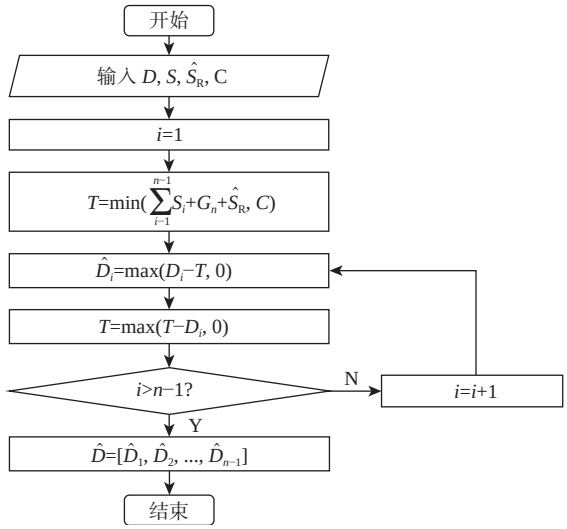


图4 考虑重要度的资源抢占机制下数据处理与运算能力剩余分配法则

Fig. 4 Performance surplus allocation rule under resource preemption mechanism considering importance

$$\hat{D}=\{\hat{D}_1,\hat{D}_2,\cdots,\hat{D}_{n-1}\} \tag{8}$$

资源抢占机制下IMA的安全可靠度R_2可以表示为支持影响系统安全应用程序的任务需求都可以被满足,且处于可工作状态下的主通用处理单元和普通通用处理单元的数量满足保证系统处于安全状态而规定数量时的概率。

$$R_2=\Pr\left\{\sum\hat{D}=0,\hat{N}_1\geqslant\hat{n}_1,\hat{N}_2\geqslant\hat{n}_2\right\} \tag{9}$$

式中:$\hat{N}_1$为资源抢占机制下,系统中处于正常工作状态的主通用处理单元及其调用的备用通用处理单元数量;$\hat{N}_2$为资源抢占机制下,系统处于正常工作状态的普通通用处理单元及其调用的备用通用处理数量;$\hat{n}_1$为安全状态下,系统中处于正常工作的主通用处理单元的最低数量;$\hat{n}_2$为安全状态下,系统中处于正常工作的普通通用处理单元的最低数量。

3 基于UGF的可靠性评估方法研究

在求解多状态系统状态概率问题上,与蒙特卡罗仿真模型等方法相比,Ushakov提出的UGF方法展现出显著优势。UGF可以通过简单直观的递归过程来枚举表示系统数据处理与运算能力水平,短时间内即可获得多状态系统能力水平分布[24]。本文运用UGF,对每个通用处理单元的数据处理与运算能力、任务需求及相关概率进行表达,对UGF的组合进行计算,以表示整个系统的能力分布,从而计算出系统在特定能力水平下运行的概率,即为系统的可靠度。

由于系统的工作状态与通用处理单元的工作状态、数据处理和运算需求的满足情况有关, 本文利用二变量 UGF 对通用处理单元 i 的数据处理与运算能力 G_i 和任务需求 W_i 的概率密度函数进行表达:

$$u_i(z_1,z_2)=\sum_{h_i=1}^{H_i}\alpha_{i,h_i}z_1^{g_{i,h_i}}z_2^{u_{i,h_i}} \tag{10}$$

$$\omega_i(z_1,z_2)=\sum_{p_i=1}^{P_i}\beta_{i,p_i}z_1^{w_{i,p_i}}z_2^{\omega_{i,p_i}} \tag{11}$$

式中: g_{i,h_i} 为通用处理单元的数据处理与运算能力; w_{i,p_i} 为通用处理单元数据处理与运算任务需求; $\alpha_{i,h_i}=\Pr\{G_i=g_{i,h_i}\}$, $\beta_{i,p_i}=\Pr\{W_i=w_{i,p_i}\}$; u_{i,h_i} 为通用处理单元 i 的表现状态, 当 $g_{i,h_i}>0$ 时, 记作 $u_{i,h_i}=1$, 当 $g_{i,h_i}\leqslant 0$ 时, 记作 $u_{i,h_i}=0$; ω_{i,p_i} 为系统对通用处理单元 i 工作状态的需求, 当 $w_{i,p_i}\geqslant 0$ 时, 记作 $\omega_{i,p_i}=-1$。

基于式 (1) 和式 (2) 获得通用处理单元的数据处理与运算能力剩余和计算不足。随后, 通过组合算子 "$\overset{\otimes}{\Leftrightarrow}$" 获得通用处理单元 i 数据处理与运算能力剩余 S_i、数据处理与运算能力不足 D_i 及工作状态 F_i 的联合 UGF。

$$\begin{aligned}\Delta_i(z_1,z_2)&=u_i(z_1,z_2)\overset{\otimes}{\Leftrightarrow}\omega_i(z_1,z_2)=\\&\sum_{v_i=1}^{V_i}\zeta_{i,v_i}z_1^{s_{i,v_i},d_{i,v_i}}z_2^{u_{i,h_i}+\omega_{i,p_i}}=\sum_{v_i=1}^{V_i}\zeta_{i,v_i}z_1^{s_{i,v_i},d_{i,v_i}}z_2^{F_{i,v_i}}\end{aligned} \tag{12}$$

式中: V_i 为收集同类项后 UGF 的项数; s_{i,v_i} 和 d_{i,v_i} 分别为通用处理单元 i 的数据处理与运算能力剩余和数据处理与运算能力不足; ζ_{i,v_i} 为联合事件概率, 其可以表示为 $\zeta_{i,v_i}=\Pr\{S_i=s_{i,v_i}\cap D_i=d_{i,v_i}\cap F_i=F_{i,v_i}\}$; F_{i,v_i} 为通用处理单元 i 的工作状态, 若 $u_{i,h_i}+\omega_{i,p_i}=0$ 时, 通用处理单元 i 处于工作状态, F_{i,v_i} 记作 1, 否则, 通用处理单元 i 处于故障状态, F_{i,v_i} 记作 0。

运用算子 "$\overset{+}{\oplus}$" 迭代计算获得系统中各通用处理单元的数据处理与运算能力剩余集合 S 和数据处理与运算能力不足集合 D 概率密度函数的联合 UGF 表现形式为

$$\begin{aligned}U_\Omega(z_1,z_2)&=\Delta_1(z_1,z_2)\overset{+}{\oplus}\Delta_2(z_1,z_2)\overset{+}{\oplus}\cdots\overset{+}{\oplus}\Delta_i(z_1,z_2)=\\&\sum_{v_1=1}^{V_1}\sum_{v_2=1}^{V_2}\cdots\sum_{v_m=1}^{V_m}\left(\prod_{i=1}^{m}\zeta_{i,v_i}\right)z_1^{\cup s_{i,v_i},\cup d_{i,v_i}}z_2^{\cup(F_{i,v_i})}=\\&\sum_{j=1}^{A_j}\pi_{A,j}z_1^{s_{A,j},d_{A,j}}z_2^{F_{A,j}}\end{aligned} \tag{13}$$

式中: A_j 为多项式合并同类项之后的项数; $s_{A,j}$ 为系统中各通用处理单元数据处理与运算能力剩余的集合; $d_{A,j}$ 为系统中各通用处理单元数据处理与运算能力不足的集合; $F_{A,j}$ 为系统中各通用处理单元工作状态的集合; $\pi_{A,j}=\Pr\{S=s_{A,j},D=d_{A,j},F=F_{A,j}\}$ 为联合事件发生的概率。

系统总线传输能力 C 概率密度函数的 UGF 表现形式为

$$\eta(z_1)=\sum_{k=1}^{K}\delta_k z_1^{c_k} \tag{14}$$

式中: $\delta_k=\Pr\{C_i=c_k\}$。

3.1　资源备份机制下系统可靠性评估方法研究

基于 UGF 的资源备份机制下数据处理与运算能力剩余分配算法, 运用组合算子 "$\odot$", 获得在资源备份机制下 IMA 中各通用处理单元可能存在的数据处理与运算能力不足 $\tilde{D}$ 概率密度函数的 UGF 表现形式为

$$\begin{aligned}\overline{U}_A(z_1,z_2)&=U_\Omega(z_1,z_2)\odot\eta(z_1)=\\&\sum_{j=1}^{A_j}\sum_{k=1}^{K}\pi_{A,j}\delta_k z_1^{\overline{d}_{A,j}}z_2^{F_{A,j}}=\sum_{j_a=1}^{J_A}p_{j_a}z_1^{\overline{d}_{j_a}}z_2^{F_{j_a}}\end{aligned} \tag{15}$$

式中: J_A 为合并同类项之后的项数; $\overline{d}_{j_a}$ 为资源备份机制下系统中各通用处理单元仍存在的数据处理与运算能力不足集合; $P_{j_a}=\Pr\{\tilde{D}=\overline{d}_{j_a},F=F_{j_a}\}$ 为系统中各通用处理单元数据处理与运算能力不足集合 $\overline{D}$ 为 $\overline{d}_{j_a}$、系统中各通用处理单元工作状态集合为 F_{j_a} 时的联合概率。$\overline{d}_{j_a}$ 根据以下计算法则获得, 如图 5 所示。

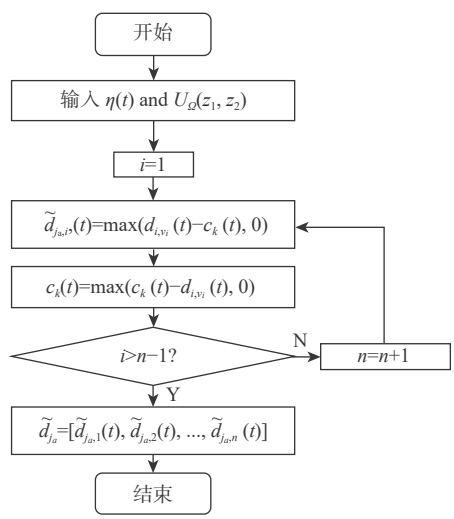


图 5　基于 UGF 的资源备份机制下数据处理与运算能力剩余分配算法

Fig. 5　The algorithm about performance surplus allocation in resource backup mechanism by using UGF

F_{j_a} 由 F_{j_a1} 和 F_{j_a2} 组成, F_{j_a1} 为主通用处理单元及其调用的备用通用处理单元工作状态集合; F_{j_a2} 为普通通用处理单元及其调用备用通用处理单元工作状态集合, 可以分别表示为

$$F_{j_{a1}}=\left\{F_{j_{a1},1},F_{j_{a1},2},\cdots,F_{j_{a1},L_1}\right\} \tag{16}$$

$$F_{j_{a2}}=\left\{F_{j_{a2},1},F_{j_{a2},2},\cdots,F_{j_{a2},L_2}\right\} \tag{17}$$

根据式(7)和式(15),获得系统在资源备份机制下的任务可靠性R_1,可以表示为

$$R_1=\Pr\left\{\sum\overline{d}_{j_a}=0,\sum_{l_1=1}^{L_1}I\left(F_{j_{a1},l_1}=1\right)\geqslant n_1,\ \sum_{l_2=1}^{L_2}I\left(F_{j_{a2},l_2}=1\right)\geqslant n_2\right\} \tag{18}$$

式中:I为示性函数,其可以表示为

$$\begin{cases}I(F_i=1)=1\\ I(F_i\neq 1)=0\end{cases} \tag{19}$$

3.2 资源抢占机制下系统可靠性评估方法研究

运用与前述方法相似的方法,系统中各通用处理单元仍存在的数据处理与运算能力不足集合$\hat{D}$概率密度函数的UGF表现形式为

$$\hat{U}_E(z_1,z_2)=\sum_{j_e=1}^{J_E}p_{j_E}z_1^{\hat{d}_{j_e}}z_2^{F_{j_e}} \tag{20}$$

式中:$\hat{d}_{j_e}$为系统中驻留影响系统安全应用程序的通用处理单元在资源抢占机制下的数据处理与运算能力不足集合;F_{j_e}由主通用处理单元工作状态集合$F_{j_{e1}}$和普通通用处理单元工作状态集合$F_{j_{e2}}$组成。

根据式(9)和式(20),可得系统在资源抢占机制下的安全可靠性R_2为

$$R_2=\Pr\left\{\sum\hat{d}_{j_e}=0,\sum_{l_3=1}^{L_3}I\left(F_{j_{e1},l_3}=1\right)\geqslant\hat{n}_1,\ \sum_{l_4=1}^{L_4}I\left(F_{j_{e2},l_4}=1\right)\geqslant\hat{n}_2\right\} \tag{21}$$

4 案例研究

该型直升机的综合模块化航空电子系统由2个主通用处理单元(通用处理单元Ⅰ)和2个普通通用处理单元(通用处理单元Ⅱ)组成。同时,为保证系统处于安全状态,系统需要2个通用处理单元Ⅰ和1个普通通用处理单元Ⅱ一直处于工作状态,且由于功率和质量的限制,2个型号的备用通用处理单元的数量最多为2。

该IMA的综合任务处理系统由3个通用处理单元组成,其中包括1个主通用处理单元(通用处理单元Ⅰ)、1个普通通用处理单元(通用处理单元Ⅱ)及1个与主通用处理单元各参数一致的备用通用处理单元。重要度较高的应用程序皆驻留在主通用处理单元(通用处理单元Ⅰ)上。系统中各通用处理单元的数据处理与运算能力、数据处理与运算能力任务需求及总线的传输能力如表1和表2所示。

表1 通用处理单元能力和任务需求的参数

Table 1 Parameters of the performance and the demand of general processing units

通用处理单元型号	数据处理与运算能力	发生概率	任务需求	发生概率	价格/元
Ⅰ	4	0.9	2	0.4	0.4
	0	0.1	1	0.6	
Ⅱ	2	0.9	2	0.4	0.2
	0	0.1	0	0.6	

表2 总线传输能力的参数

Table 2 Parameters of performance of common bus

传输能力	发生概率
2	0.8
1	0.1
0	0.1

在飞行任务过程中,主通用处理单元需要一直保持正常工作状态,普通通用处理单元可以由于故障或资源抢占机制的触发而终止其驻留的应用程序。

以某直升机IMA的综合任务处理系统为研究对象,验证本文提出的可靠性建模评估方法的可行性。同时,运用所提方法,实现对IMA备份资源配置方案的优化设计。

1)每个通用处理单元的数据处理与运算能力G_i和数据处理与运算能力需求W_i概率密度函数的UGF形式为

通用处理单元Ⅰ:

$$u_1(z_1,z_2)=0.9z_1^4z_2^{F_1[1]}+0.1z_1^0z_2^{F_1[0]}$$

$$\omega_1(z_1,z_2)=0.4z_1^2z_2^{F_1[-1]}+0.6z_1^1z_2^{F_1[-1]}$$

通用处理单元Ⅱ:

$$u_2(z_1,z_2)=0.9z_1^2z_2^{F_2[1]}+0.1z_1^0z_2^{F_2[0]}$$

$$\omega_2(z_1,z_2)=0.4z_1^1z_2^{F_2[-1]}+0.6z_1^0z_2^{F_2[-1]}$$

备用通用处理单元Ⅰ:

$$u_{R_1}(z_1,z_2)=0.9z_1^4z_2^{F_R[1]}+0.1z_1^0z_2^{F_R[0]}$$

2)通信系统总线传输能力概率密度函数的UGF形式为

$$\eta(z_1)=0.1z_1^0+0.1z_1^1+0.8z_1^2$$

3)运用式(12),计算获得通用处理单元Ⅰ的数

据处理与运算能力剩余S_1和能力不足D_1概率密度函数的 UGF 表现形式为

$$\Delta_1(z_1,z_2)=u_1(z_1,z_2)\overset{\otimes}{\Leftrightarrow}\omega_1(z_1,z_2)=0.54z_1^{[3],[0]}z_2^{[1]}+0.36z_1^{[2],[0]}z_2^{[1]}+0.06z_1^{[0],[1]}z_2^{[0]}+0.04z_1^{[0],[2]}z_2^{[0]}$$

通用处理单元Ⅱ数据处理与运算能力剩余S_2和能力不足D_2概率密度函数的 UGF 表现形式为

$$\Delta_2(z_1,z_2)=u_2(z_1,z_2)\overset{\otimes}{\Leftrightarrow}\omega_2(z_1,z_2)=0.54z_1^{[2],[0]}z_2^{[1]}+0.36z_1^{[1],[0]}z_2^{[1]}+0.06z_1^{[0],[0]}z_2^{[0]}+0.04z_1^{[0],[1]}z_2^{[0]}$$

备用通用处理单元Ⅰ概率密度函数的 UGF 表现形式为

$$\Delta_{R1}(z_1,z_2)=0.9z_1^{[4],[0]}z_2^{[1]}+0.1z_1^{[0],[0]}z_2^{[0]}$$

4) 运用式 (13), 运用组合算子“$\overset{+}{\oplus}$”, 计算系统中各通用处理单元的数据处理与运算能力剩余集合S和能力不足集合D概率密度函数的 UGF 表现形式为

$$\begin{aligned}U_{\Omega}(z_1,z_2)=&\Delta_1(z_1,z_2)\overset{+}{\oplus}\Delta_2(z_1,z_2)=0.003\,6z_1^{[0,0],[1,0]}z_2^{[0],[0]}+\\&0.002\,4z_1^{[0,0],[1,1]}z_2^{[0],[0]}+0.021\,6z_1^{[2,0],[0,0]}z_2^{[1],[0]}+\\&0.032\,4z_1^{[3,0],[0,0]}z_2^{[1],[0]}+0.001\,6z_1^{[0,0],[2,1]}z_2^{[0],[0]}+\\&0.002\,4z_1^{[0,0],[2,0]}z_2^{[0],[0]}+0.014\,4z_1^{[2,0],[0,1]}z_2^{[1],[0]}+\\&0.014\,4z_1^{[0,1],[2,0]}z_2^{[0],[1]}+0.021\,6z_1^{[3,0],[0,1]}z_2^{[1],[0]}+\\&0.032\,4z_1^{[0,2],[1,0]}z_2^{[0],[1]}+0.021\,6z_1^{[0,2],[2,0]}z_2^{[0],[1]}+\\&0.021\,6z_1^{[0,2],[1,0]}z_2^{[0],[1]}+0.194\,4z_1^{[2,2],[0,0]}z_2^{[1],[1]}+\\&0.291\,6z_1^{[3,2],[0,0]}z_2^{[1],[1]}+0.129\,6z_1^{[2,1],[0,0]}z_2^{[1],[1]}+\\&0.194\,4z_1^{[3,1],[0,0]}z_2^{[1],[1]}\end{aligned}$$

5) 资源备份机制下的系统中, 各通用处理单元仍存在的数据运算与处理能力不足集合D概率密度函数的 UGF 表现形式为

$$\begin{aligned}U_{\Omega_1}(z_1,z_2)=&U_{\Omega}(z_1,z_2)\overset{+}{\oplus}\Delta_{R_1}(z_1,z_2)=\\&0.019\,44z_1^{[0,2,4],[2,0,0]}z_2^{[0,1],[1]}+0.029\,16z_1^{[0,2,4],[1,0,0]}z_2^{[0,1],[1]}+\\&0.019\,44z_1^{[3,0,4],[0,1,0]}z_2^{[1],[0,1]}+0.012\,96z_1^{[0,1,4],[2,0,0]}z_2^{[0,1],[1]}+\\&0.012\,96z_1^{[2,0,4],[0,1,0]}z_2^{[1],[0,1]}+0.019\,44z_1^{[0,1,4],[1,0,0]}z_2^{[0,1],[1]}+\\&0.002\,16z_1^{[0,0,4],[1,1,0]}z_2^{[0,1],[0]}+0.014\,4z_1^{[0,0,4],[2,1,0]}z_2^{[0,1],[0]}+\\&0.029\,16z_1^{[3,0,4],[0,0,0]}z_2^{[1],[0,1]}+0.019\,44z_1^{[2,0,4],[0,1,0]}z_2^{[1],[0,1]}+\\&0.002\,16z_1^{[0,1,0],[1,0,0]}z_2^{[0,0],[1]}+0.002\,16z_1^{[0,0,4],[2,0,0]}z_2^{[0,1],[0]}+\\&0.002\,16z_1^{[0,2,0],[2,0,0]}z_2^{[0,0],[1]}+0.00324z_1^{[0,0,4],[1,0,0]}z_2^{[0,1],[0]}+\\&0.00324z_1^{[0,2,0],[1,0,0]}z_2^{[0,0],[1]}+0.002\,16z_1^{[3,0,0],[0,1,0]}z_2^{[1],[0,0]}+\\&0.001\,44z_1^{[0,1,0],[2,0,0]}z_2^{[0,0],[1]}+0.001\,44z_1^{[2,0,0],[0,1,0]}z_2^{[1],[0,0]}+\\&0.000\,24z_1^{[0,0,0],[2,0,0]}z_2^{[0,0],[0]}+0.000\,16z_1^{[0,0,0],[2,1,0]}z_2^{[0,0],[0]}+\\&0.003\,24z_1^{[3,0,0],[0,0,0]}z_2^{[1],[0,0]}+0.002\,16z_1^{[2,0,0],[0,0,0]}z_2^{[1],[0,0]}+\\&0.000\,24z_1^{[0,0,0],[1,1,0]}z_2^{[0,0],[0]}+0.000\,36z_1^{[0,0,0],[1,0,0]}z_2^{[0,0],[0]}\end{aligned}$$

6) 综合任务处理系统的系统数据处理与运算能力不足$\tilde{D}$概率密度函数的 UGF 为

$$\begin{aligned}\overline{U}_A(z_1,z_2)=&U_{\Omega_1}(z_1,z_2)\odot\eta(z_1)0.942\,516+\\&0.000\,36z_1^{[0,0,0],[1,0,0]}z_2^{[0,0],[0]}+0.000\,24z_1^{[0,0,0],[1,1,0]}z_2^{[0,0],[0]}+\\&0.002\,16z_1^{[2,0,0],[0,0,0]}z_2^{[1],[0,0]}+0.003\,24z_1^{[3,0,0],[0,0,0]}z_2^{[1],[0,0]}+\\&0.000\,16z_1^{[0,0,0],[2,1,0]}z_2^{[0,0],[0]}+0.000\,24z_1^{[0,0,0],[2,0,0]}z_2^{[0,0],[0]}+\\&0.000\,144z_1^{[2,0,0],[0,1,0]}z_2^{[1],[0,0]}+0.001\,296z_1^{[1,0,0],[0,0,0]}z_2^{[1],[0,0]}+\\&0.000\,144z_1^{[0,1,0],[2,0,0]}z_2^{[0,0],[1]}+0.001\,296z_1^{[0,0,0],[1,0,0]}z_2^{[0,0],[1]}+\\&0.000\,216z_1^{[3,0,0],[0,1,0]}z_2^{[1],[0,0]}+0.001\,944z_1^{[2,0,0],[0,0,0]}z_2^{[1],[0,0]}+\\&0.000\,324z_1^{[0,2,0],[1,0,0]}z_2^{[0,0],[1]}+0.002\,916z_1^{[0,1,0],[0,0,0]}z_2^{[0,0],[1]}+\\&0.000\,324z_1^{[0,0,4],[1,0,0]}z_2^{[0,1],[0]}+0.002\,916z_1^{[0,0,3],[0,0,0]}z_2^{[0,1],[0]}+\\&0.000\,216z_1^{[0,2,0],[2,0,0]}z_2^{[0,0],[1]}+0.000\,216z_1^{[0,1,0],[1,0,0]}z_2^{[0,0],[1]}+\\&0.001\,728z_1^{[0,0,0],[0,0,0]}z_2^{[0,0],[1]}+0.000\,216z_1^{[0,0,4],[2,0,0]}z_2^{[0,1],[0]}+\\&0.000\,216z_1^{[0,0,3],[1,0,0]}z_2^{[0,1],[0]}+0.001\,728z_1^{[0,0,2],[0,0,0]}z_2^{[0,1],[0]}+\\&0.000\,216z_1^{[0,1,0],[1,0,0]}z_2^{[0,0],[1]}+0.001\,944z_1^{[0,0,0],[0,0,0]}z_2^{[0,0],[1]}+\\&0.001\,944z_1^{[2,0,4],[0,1,0]}z_2^{[1],[0,1]}+0.001\,44z_1^{[0,0,4],[2,1,0]}z_2^{[0,1],[0]}+\\&0.001\,44z_1^{[0,0,3],[1,1,0]}z_2^{[0,1],[0]}+0.011\,52z_1^{[0,0,2],[0,1,0]}z_2^{[0,1],[0]}+\\&0.000\,216z_1^{[0,0,4],[1,1,0]}z_2^{[0,1],[0]}+0.000\,216z_1^{[0,0,3],[0,1,0]}z_2^{[0,1],[0]}+\\&0.001\,728z_1^{[0,0,2],[0,0,0]}z_2^{[0,1],[0]}+0.001\,944z_1^{[0,1,4],[1,0,0]}z_2^{[0,1],[1]}+\\&0.001\,296z_1^{[2,0,4],[0,1,0]}z_2^{[1],[0,1]}+0.001\,296z_1^{[0,1,4],[1,0,0]}z_2^{[0,1],[1]}+\\&0.001\,296z_1^{[0,1,3],[1,0,0]}z_2^{[0,1],[1]}+0.001\,944z_1^{[3,0,4],[0,1,0]}z_2^{[1],[0,1]}+\\&0.002\,916z_1^{[0,2,4],[1,0,0]}z_2^{[0,1],[1]}+0.001\,944z_1^{[0,2,4],[2,0,0]}z_2^{[0,1],[1]}+\\&0.001\,944z_1^{[0,2,3],[1,0,0]}z_2^{[0,1],[1]}\end{aligned}$$

由此可知, 考虑动态重构机制下综合任务处理系统的任务可靠度为 0.942 516。

对于求解动态重构机制下综合任务处理系统的可靠度问题, 当运用可靠性框图方法对该系统的可靠性进行评估时, 其可靠性框图如图 6 所示, 系统的可靠度为 0.891。但是, 由于可靠性框图的方法无法考虑系统的动态重构机制, 导致其可靠性评估结果小于本文方法。因此, 相比于可靠性框图方法, 本文所提可靠性模型在评估系统可靠度时, 考虑更全面, 评估结果更准确。

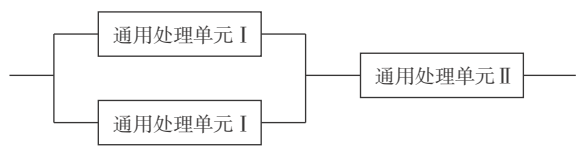


图 6　系统可靠性框图

Fig. 6　The reliability block diagram of system

本文对该型直升机的 IMA 可靠性与备用通用处理单元数量的关系进行研究。在不同备用通用处理单元数量的情况下, 系统可靠度的变化趋势如图 7 所示。可知, 在动态重构机制下, 随着备用通用处理单元数量的增加, 系统的可靠度先快速上升, 后缓慢上升, 最后逐渐趋于平稳。同时, 增加备用通用处理单元Ⅰ的数量对可靠性增长律的影响大于增加备份通用处理单元Ⅱ的数量。

由于在设计过程中, 成本是不可忽略的关键因

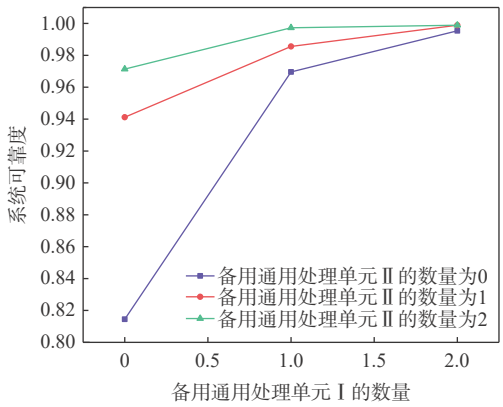


图7 系统可靠性变化趋势

Fig. 7 The tendency of the reliability of system

素。随着备用通用处理单元数量的增加，其成本也随之增加。因此，寻找备用通用处理单元的数量与成本之间的平衡是非常重要的。本文引入效费比，通过研究效费比随着备用通用处理单元数量增加而变化的规律，从而获得最合理的IMA备用通用处理单元配置方案，效费比θ可以表示为

$$\theta = \frac{R}{C} \tag{22}$$

式中：R为动态重构机制下IMA的可靠度；C为备用通用处理单元所需的费用。

备用通用处理单元所需的费用可以表示为

$$C = \cos t_1 \times x_1 + \cos t_2 \times x_2 \tag{23}$$

式中：$\cos t_1$和$\cos t_2$分别为备用通用处理单元Ⅰ和备用通用处理单元Ⅱ的成本；x_1和x_2分别为备用通用处理单元Ⅰ和备用通用处理单元Ⅱ的数量。

由图8可知，在考虑成本的情况下，该系统的最优化方案为备用通用处理单元Ⅱ数量为1的方案。该方案的综合任务的可靠度为0.941 2，效费比为4.706。

相比于其他可靠性评估模型，本模型更加符合IMA动态重构机制的工作原理，获得的评估结果更加全面，可以避免过度的裕度设计。通过案例分析可知，系统的可靠度随着备用通用处理单元数量的增加，先快速上升，后缓慢上升，最终逐渐趋于平稳；增加通用处理单元Ⅰ的数量对系统可靠性增长律的影响更大，系统的效费比随着备份通用处理单元数量的增加而逐渐下降。在相同成本的条件下，优先选择备用通用处理单元Ⅱ数量多的设计方案可以有效地提升效费比。

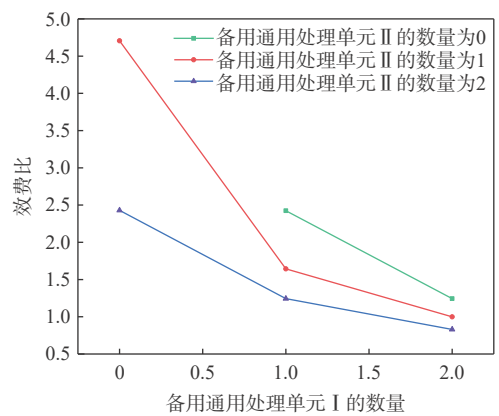


图8 系统效费比变化趋势

Fig. 8 The tendency of benefit cost ratio of the system

5 结 论

1）面向动态重构机制下的IMA，在现有总线性能共享多状态系统模型的基础上，构建适用于不同动态重构机制的可靠性建模与评估方法，实现系统可靠性的定量计算，为IMA的可靠性设计提供参考。

2）以某型直升机IMA为对象开展案例验证，通过对不同备用通用处理单元配置方案进行评估，研究备用通用处理单元数量和类型变化对系统可靠度和效费比的影响规律。

3）案例分析结果表明：随着备用通用处理单元数量增加，系统可靠度先快速提升，后增幅逐步减小并趋于稳定，而系统效费比随备用通用处理单元数量增加呈下降趋势。同时，备用通用处理单元I对可靠度提升更为敏感，在总成本相同条件下，优先配置更多备用通用处理单元II有助于提高系统效费比。

本文仍存在一些不足之处，本文在资源配置方案优化过程中，未考虑带宽等物理资源的约束限制，因此，所获得的备份资源配置方案存在一定的局限性。未来将在更全面考虑物理资源限制的前提下，对系统的资源配置方案优化问题展开研究，为系统设计提供更有效的指导。

参考文献（References）

[1] BIRBER P, BONIOL F, BOYER M, et al. New challenges for future avionic architectures[J]. New Aerospace Lab, 2012(4): 1-10.

[2] GASKA T, WATKIN C, CHEN Y. Integrated modular avionics-past, present, and future[J]. IEEE Aerospace and Electronic Systems Magazine, 2015, 30(9): 12-23.

[3] 詹志娟, 周庆, 洪蓉. DIMA动态重构管理方法研究[C]//2016航空试验测试技术学术交流会论文集. 北京: 测控技术编辑部, 2016: 201-203.
ZHAN Z J, ZHOU Q, HONG R. Research on dynamic reconfiguration management method for distributed integrated modular avionics[C]//Proceedings of the 2016 Academic Conference on Aviation Test and Measurement Technology. Beijing: Measurement & Control Technology Editorial office, 2016: 201-203.

[4] 丁明. 综合模块化航空电子系统重构与验证方法研究[D]. 西安: 西北大学, 2019: 1-5.
DING M. Research on reconfiguration and verification methods for

integrated modular avionics[D]. Xi'an: Northwestern University, 2019: 1-5(in Chinese).

[5] 张安莉. 基于随机 Petri 网的航空电子综合化系统可靠性建模与分析[D]. 西安: 西北工业大学, 2006: 1-5.
ZHANG A L. Reliability modeling and analysis of integrated avionics systems based on stochastic Petri nets [D]. Xi'an: Northwestern Polytechnical University, 2006: 1–5(in Chinese).

[6] WANG R P, LU W T, ZENG C H, et al. Reliability modeling and verification method for dynamic reconfiguration system[C]//Proceedings of the 2018 Prognostics and System Health Management Conference. Piscataway: IEEE Press, 2019: 941-947.

[7] JIANG Z Y, ZHAO T D, WANG S H, et al. New model-based analysis method with multiple constraints for integrated modular avionics dynamic reconfiguration process[J]. Processes, 2020, 8(5): 574.

[8] JU H Y, WANG S H, ZHAO T D. A modeling method of IMA dynamic reconfiguration based on AADL[C]//Proceedings of the 2015 First International Conference on Reliability Systems Engineering. Piscataway: IEEE Press, 2016: 1-5.

[9] 李伯宁. 综合模块化航电系统分区资源分配方法研究[D]. 天津: 中国民航大学, 2022: 8-11.
LI B N. Research on partition resource allocation method of integrated modular avionics system [D]. Tianjin: Civil Aviation University of China, 2022: 8-11(in Chinese).

[10] LISNIANSKI A, DING Y. Redundancy analysis for repairable multi-state system by using combined stochastic processes methods and universal generating function technique[J]. Reliability Engineering & System Safety, 2009, 94(11): 1788-1795.

[11] LEVITIN G. Reliability of multi-state systems with common bus performance sharing[J]. IIE Transactions, 2011, 43(7): 518-524.

[12] CHENG C, YANG J, LI L. Reliability assessment of multi-state phased mission systems with common bus performance sharing considering transmission loss and performance storage[J]. Reliability Engineering & System Safety, 2020, 199: 106917.

[13] ZHAO X, WU C S, WANG S Q, et al. Reliability analysis of multi-state k-out-of-n: G system with common bus performance sharing [J]. Computers & Industrial Engineering, 2018, 124: 359-369.

[14] WU D, CHI Y Y, PENG R, et al. Reliability of capacitated systems with performance sharing mechanism[J]. Reliability Engineering & System Safety, 2019, 189: 335-344.

[15] SU P, WANG G J, DUAN F J. Reliability evaluation of a k-out-of-n(G)-subsystem based multi-state system with common bus performance sharing[J]. Reliability Engineering & System Safety, 2020, 198: 106884.

[16] ZHAI Q Q, YE Z S, PENG R, et al. Defense and attack of performance-sharing common bus systems[J]. European Journal of Operational Research, 2017, 256(3): 962-975.

[17] QIU S, MING H X G. Reliability analysis of multi-state series systems with performance sharing mechanism under epistemic uncertainty[J]. Quality and Reliability Engineering International, 2019, 35(6): 1998-2015.

[18] YU H, YANG J, MO H D. Reliability analysis of repairable multi-state system with common bus performance sharing[J]. Reliability Engineering & System Safety, 2014, 132: 90-96.

[19] XIAO H, PENG R. Optimal allocation and maintenance of multi-state elements in series-parallel systems with common bus performance sharing[J]. Computers & Industrial Engineering, 2014, 72: 143-151.

[20] AERONAUTICAL RADIO, INC. ARINC specification 653: avionics application software standard interface, part Ⅰ−required services [S]. 2015.

[21] ASAAC. Final draft of proposed guidelines for system issues-volume 1-2: system configuration/reconfiguration: ASAAC-STANAG 4626(Part VI)[S]. Belgium Brussels: North Atlantic Treaty Organization, 2004.

[22] 阎芳, 邢培培, 赵长啸, 等. 基于联合 k/n(G) 模型的 DIMA 系统可靠性建模与分析[J]. 航空学报, 2018, 39(6): 185-193.
YAN F, XING P P, ZHAO C X, et al. Reliability modeling and analysis of DIMA system based on joint k/n(G) model[J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(6): 185-193(in Chinese).

[23] 王鹏, 刘嘉琛, 董磊, 等. 面向任务的民机 DIMA 动态重构策略[J]. 系统工程与电子技术, 2021, 43(6): 1618-1627.
WANG P, LIU J C, DONG L, et al. Task oriented DIMA dynamic reconfiguration strategy for civil aircraft[J]. Systems Engineering and Electronics, 2021, 43(6): 1618-1627(in Chinese).

[24] LEVITIN G. The universal generating function in reliability analysis and optimization[M]. London: Springer-Verlag, 2005.

Reliability modeling and evaluation method of IMA under dynamic reconfiguration

HU Gengshuo，JIAO Jian*，HU Langxiao，JING Yongfeng

(School of Reliability and System Engineering，Beihang University，Beijing 100191，China)

Abstract：This study starts by examining the interconnection topology of integrated modular avionics (IMA) and two dynamic reconfiguration mechanisms: resource backup and resource preemption, in order to address the challenge of modeling and evaluating the dependability of IMA under dynamic reconfiguration. Subsequently, based on the operational principles of IMA and dynamic reconfiguration mechanisms, this paper extends a multi-state system with common bus performance sharing while considering the performance allocation problem for IMA under dynamic reconfiguration. Simultaneously, an algorithm about IMA reliability assessment under dynamic reconfiguration mechanisms based on the universal generating function (UGF) is proposed to quantitatively evaluate the reliability of IMA under different dynamic reconfiguration mechanisms. In order to confirm the viability of the suggested extended model and reliability assessment method, a case study is finally carried out on the IMA in the helicopter, examining the effects of variations in the number of standby processing units on the system's dependability and suggesting an ideal design solution.

Keywords：helicopter integrated modular avionics；dynamic reconfiguration；multi-state system；universal generating function；reliability

Received：2024-04-01；**Accepted**：2024-06-07；**Published Online**：2024-06-29 17：49

URL：link.cnki.net/urlid/11.2625.V.20240628.0920.001

* **Corresponding author**. E-mail：jiaojian@buaa.edu.cn

《北京航空航天大学学报》征稿简则

《北京航空航天大学学报》是北京航空航天大学主办的以航空航天科学技术为特色的综合性自然科学学术期刊(月刊)。本刊以反映航空航天领域研究成果与动态、促进学术交流、培养科技人才和推动科技成果向社会生产力转化为办刊宗旨。本刊为中国自然科学技术核心期刊,并被 Ei Compendex 等国内外权威文献检索系统收录。本刊向国内外公开发行,为进一步提高办刊质量和所刊出文章的学术水平,特制定本简则。

1 论文作者及内容

1.1 本刊面向海内外所有学者。

1.2 主要刊载与航空航天科学技术有关的材料科学及工程、飞行器设计与制造、宇航科学与工程、信息与电子技术、控制技术和自动化工程、流体力学和动力工程、计算机科学及应用技术、可靠性工程与失效分析等领域的研究文章。航空航天科学技术民用方面以及具有航空航天工程背景的应用数学、应用物理、应用力学和工程管理等方面的文章也在本刊优先考虑之列。

2 来稿要求

2.1 论文应具有创新性、科学性、学术性和可读性。

2.2 论文为原创作品,尚未公开发表过,并且不涉及泄密问题。若发生侵权或泄密问题,一切责任由作者承担。

2.3 主题明确,数据可靠,图表清晰,逻辑严谨,文字精练,公式符号标准规范,标点符号正确。

2.4 文稿撰写顺序:中文题名(一般不超过 20 个汉字),作者中文姓名、单位、所在城市、邮政编码(作者顺序始终不得更改),中文摘要(包括目的、方法、结果及结论),中文关键词(5 ~ 8 个),中图分类号,引言,正文,参考文献,英文题名,作者英文姓名、单位、所在城市、邮政编码、国别,英文摘要,英文关键词。首、末页下角注明基金项目名称及编号,作者信息。

2.5 作者请登录本刊网页进行在线投稿。

3 稿件的审核、录用与版权

3.1 来稿须经专家两审和主编、编委讨论后决定刊用与否。

3.2 若来稿经审查后认定不宜在本刊发表,将及时告知作者。如果在投稿满 3 个月后仍未收到本刊任何通知,作者有权改投其他刊。在此之前,请勿一稿多投,否则一切后果自负。

3.3 来稿一经作者签字并在本刊刊出,即表明所有作者都已经认可其版权转至本刊编辑部。本刊在与国内外文献数据库或检索系统进行交流及合作时,不再征询作者意见。

邮寄地址:100191 北京市海淀区学院路 37 号 《北京航空航天大学学报》编辑部
办公地点:北京航空航天大学办公楼西配楼412~416房间
电　　话:(010)82316698,82316981,82315499,82317448
E-mail: jbuaa@ buaa. edu. cn
http://bhxb. buaa. edu. cn
http://www. buaa. edu. cn

集群成员期刊

《北京航空航天大学学报》
第五届编辑委员会

主　任（主　编）：赵沁平

（以下按姓氏笔画为序）

副主任（副主编）：丁希仑　王少萍　孙志梅　李秋实　李焕喜　杨嘉陵
苗俊刚　相　艳　徐立军　钱德沛　曹晋滨

编　委：马殿富　王　琪　王　聪　邓小燕　王青云　王荣明　刘　宇
刘　红　江　洁　刘　强　闫　鹏　朱天乐　刘铁钢　齐铂金
陈万春　邹正平　苏东林　杨世春　沈成平　邱志平　宋知人
杨树斌　张晓林　杨晓奕　杨继萍　李惠峰　吴新开　张瑞丰
杨照华　宋凝芳　周　锐　林宇震　林贵平　战　强　姚仰平
胡庆雷　赵秋红　段海滨　赵巍胜　席　平　郭　宏　徐　洁
徐世杰　郭洪波　康　锐　翟　锦　熊华钢

编辑部主任：张　嵘

编　　辑：孙　芳　万玉敏　王　茜　卞欢欢　李亚泰

北京航空航天大学学报

Beijing Hangkong Hangtian Daxue Xuebao

（原《北京航空学院学报》）

（月刊　1956 年创刊）

第52卷 第5期　2026年5月

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)

(Monthly, Started in 1956)

Vol.52 No.5　May 2026

主管单位	中华人民共和国工业和信息化部
主办单位	北京航空航天大学
主　　编	赵沁平
编辑出版	《北京航空航天大学学报》编辑部
邮　　编	100191
地　　址	北京市海淀区学院路37号
印　　刷	北京科信印刷有限公司
发　　行	《北京航空航天大学学报》编辑部
发行范围	国内外发行
联系电话	(010) 82316698 82316981 82315499 82317448
电子信箱	jbuaa@ buaa.edu.cn

Administrated by	Ministry of Industry and Information Technology of the People's Republic of China
Sponsored by	Beijing University of Aeronautics and Astronautics (BUAA) (Beijing 100191, P. R. China)
Chief Editor	ZHAO Qinping
Edited and Published by	Editorial Board of JBUAA
Printed by	Beijing Kexin Printing Co., Ltd.
Distributed by	Editorial Board of JBUAA
Telephone	(010) 82316698 82316981 82315499 82317448
E-mail	jbuaa@buaa.edu.cn
	http: //bhxb.buaa.edu.cn

中国标准连续出版物号：ISSN 1001-5965 / CN 11-2625/V

ISSN 1001-5965

国内定价：50.00元／期

微信公众号　超星电子书