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

低空智能网联体系

低空智能网联体系

张学军　刘法旺　张祖耀　田　野

6 | 2025
Vol.51

北京航空航天大学学报

第 51 卷　第 6 期　（总第 388 期）　2025 年 6 月

目　　次

低空智能网联体系 …… 张学军，刘法旺，张祖耀，田野（1793）
飞机可动翼面缝隙封严结构研究进展 …… 程小全，蔡墨泉，王松伟（1816）
基于功率损失特性分析的特种车辆传动性能研究进展 ……
…… 高钦和，高蕾，刘志浩，王冬，马栋，章一博（1824）
基于迭代最优阻抗的人机物理交互控制方法 ……
…… 刘微容，魏子丰，晋振兵，孟家豪，王星琨，张浩琛（1843）
边缘引导的双分支网络 SAR 图像相干斑抑制方法 ……
…… 朱磊，姚同钰，车晨洁，姚丽娜，张博，潘杨（1852）
融合聚类特征的改进 MRF 轨面缺陷分割方法 …… 闵永智，刘洋（1863）
基于飞联网运行的航空器多速度差跟驰模型 …… 王莉莉，赵云飞，郭微萌（1873）
基于双重随机扰动的人工大猩猩部队优化算法及工程应用 ……
…… 杜晓昕，郝田茹，王波，王振飞，张剑飞，金梅（1882）
基于高增益扰动观测器的磁悬浮球系统滑模控制 …… 林俊亭，陈信州（1897）
基于 QAR 数据与关联规则的着陆阶段超限风险评估 …… 汪磊，张楠，高杉（1907）
基于内聚力模型的透波罩黏接强度性能研究 …… 侯保江，王姣，邢誉峰，张飞，李焱喜（1916）
考虑多车型协同的机坪保障车辆双阶段调度 …… 姜雨，张诗淼，陈名扬，张文静，吴薇薇（1926）
基于部件特性图优化的民航发动机性能退化建模 …… 郭庆，黄启廉，陈金亮（1935）
基于几何蒸馏和特征自适应的磁共振成像重建 …… 朵琳，任勇，许渤雨，杨新（1946）
基于多注意力与双模板更新的视觉跟踪算法 …… 马素刚，孙思维，侯志强，余旺盛，蒲磊（1955）
EHL 下考虑冲击力的局部缺陷角接触球轴承动力学建模 ……
…… 雷春丽，薛伟，樊高峰，宋瑞哲，刘凯（1965）
基于 K 互近邻与核密度估计的 DPC 算法 …… 周玉，夏浩，刘虹瑜，白磊（1978）
冰晶颗粒在低压压气机中的融化特性 …… 贾惟，张峰，杨博文，郑淼（1991）
飞行器大气数据系统构型与解算演变机理 …… 熊亮，张睿，许斌，黄巧平（2004）
星载准绝对式光电编码器抗污方法 …… 韩庆阳，沈宏海，马天翔，梁超，王志冲（2014）
基于深度图卷积网络的自监督群体发现模型 …… 王栽胜，汪晓锋，沈国栋，张增杰，全大英（2022）
基于滑模控制律的机翼极限环主动抑制 …… 王诗其，张征，宋晨，杨超，郑焱午（2033）
热隔膜成形中隔膜材料性能测试及本构模型选择 …… 赵月青，林德志，陈汇，汤家力，陈萍（2041）
基于自适应积分电容的高动态像素结构 …… 刘绥阳，郭仲杰，余宁梅，许睿明（2051）
针对高机动目标的深度强化学习智能拦截制导 …… 张豪，朱建文，李小平，包为民（2060）
VSV 调节机构运动精度可靠度计算及误差分析 …… 杨奕凤，王艺，刘傲宇，李佳，谢里阳（2070）
多元件相关的列车牵引系统可靠性评估与寿命预测 …… 田贵双，王少萍，石健（2081）
中心分级环形燃烧室液雾点熄火边界试验 …… 王智辉，张弛，甘志超，高安雯，韩啸，陶雯婕（2091）

人-车-智共享道路空间行人安全行为建模方法 …… 仵若玮，杨德真，任羿，贾露露，李晓宾，王自力（2099）
星载 GNSS 反射信号台风检测及风眼位置估计 …… 甄佳欢，朱云龙，杨东凯，张国栋，王峰（2106）
μHT-1 推力器工作参数宽范围调节下束流特性实验研究 …… 黄丹，龙建飞，成烨，王嘉彬，徐禄祥，杨威（2119）
等离子体激励对火星条件下翼型气动特性影响 …… 杨香港，高永新，汪忠明，李益文，姚程（2129）
双参数退化型新研产品质保期设计与优化 …… 高帅，李艳宏，孙富强（2137）
曲线梁肋机翼结构气动弹性优化设计 …… 周泉知，杨佑绪，孙录斌，张兴翠，吴逸飞，霍梦文（2148）
基于改进卡尔曼滤波的星内校时守时方法 …… 黄跃，王慧泉，涂实磊，金仲和（2157）
基于强化学习的舰载机着舰直接升力控制技术 …… 柳仁地，江驹，张哲，刘翔（2165）
基于 ISCA-DBN 的飞机地面空调能耗预测 …… 刘涵，林家泉（2176）

期刊基本参数：CN 11-2625/V * 1956 * m * A4 * 392 * zh * P * ¥ 50.00 * 400 * 37 * 2025-06

（**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. 51 No. 6 （Sum 388） June 2025

CONTENTS

Overview of low-altitude intelligent networked system
…… ZHANG Xuejun，LIU Fawang，ZHANG Zuyao，TIAN Ye （1793）
Study progress of gap sealing structure for aircraft movable wing
…… CHENG Xiaoquan，CAI Moquan，WANG Songwei （1816）
Research progress on transmission performance of special vehicle based on power loss characteristics analysis
…… GAO Qinhe，GAO Lei，LIU Zhihao，WANG Dong，MA Dong，ZHANG Yibo （1824）
Human-robot physical interaction control method based on iterative optimal impedance
…… LIU Weirong，WEI Zifeng，JIN Zhenbing，MENG Jiahao，WANG Xingkun，ZHANG Haochen （1843）
SAR image coherence speckle suppression method based on edge-guided dual-branch network
…… ZHU Lei，YAO Tongyu，CHE Chenjie，YAO Lina，ZHANG Bo，PAN Yang （1852）
Improved MRF rail surface defect segmentation method based on fusion of clustering features
…… MIN Yongzhi，LIU Yang （1863）
Aircraft multi-velocity difference car-following model based on flight networking operation
…… WANG Lili，ZHAO Yunfei，GUO Weimeng （1873）
Artificial gorilla troops optimizer based on double random disturbance and its application of engineering problem
…… DU Xiaoxin，HAO Tianru，WANG Bo，WANG Zhenfei，ZHANG Jianfei，JIN Mei （1882）
Sliding mode control of magnetic levitation ball systems based on a high-gain disturbance observer
…… LIN Junting，CHEN Xinzhou （1897）
Risk assessment of landing phase exceedance based on QAR data and association rule
…… WANG Lei，ZHANG Nan，GAO Shan （1907）
Performance study on bonding strength of radome based on cohesive zone model
…… HOU Baojiang，WANG Jiao，XING Yufeng，ZHANG Fei，LI Yanxi （1916）
Dual-phase scheduling of apron support vehicles considering multi-vehicle coordination
…… JIANG Yu，ZHANG Shimiao，CHEN Mingyang，ZHANG Wenjing，WU Weiwei （1926）
Performance degradation modelling of civil aviation engines based on component characteristic map optimization
…… GUO Qing，HUANG Qilian，CHEN Jinliang （1935）
MRI reconstruction based on geometry distillation and feature adaptation
…… DUO Lin，REN Yong，XU Boyu，YANG Xin （1946）
Visual tracking algorithm based on multi-attention and dual-template update
…… MA Sugang，SUN Siwei，HOU Zhiqiang，YU Wangsheng，PU Lei （1955）
Dynamic modelling of angular contact ball bearings with local defects under EHL considering impact force
…… LEI Chunli，XUE Wei，FAN Gaofeng，SONG Ruizhe，LIU Kai （1965）
DPC algorithm based on K-reciprocal neighbors and kernel density estimation
…… ZHOU Yu，XIA Hao，LIU Hongyu，BAI Lei （1978）
Ice crystal particles melting characteristics in a low-pressure compressor
…… JIA Wei，ZHANG Feng，YANG Bowen，ZHENG Miao （1991）
Configuration and calculation evolution mechanism of air data system for aircraft
…… XIONG Liang，ZHANG Rui，XU Bin，HUANG Qiaoping （2004）
Anti-stain method of satellite-borne absolute-to-be optical encoder
…… HAN Qingyang，SHEN Honghai，MA Tianxiang，LIANG Chao，WANG Zhichong （2014）
Self-supervised learning for community detection based on deep graph convolutional networks
…… WANG Zaisheng，WANG Xiaofeng，SHEN Guodong，ZHANG Zengjie，QUAN Daying （2022）
Limit cycle oscillation suppression of an airfoil based on sliding mode control law
…… WANG Shiqi，ZHANG Zheng，SONG Chen，YANG Chao，ZHENG Yanwu （2033）
Performance test and constitutive model selection of diaphragm materials in hot diaphragm forming
…… ZHAO Yueqing，LIN Dezhi，CHEN Hui，TANG Jiali，CHEN Ping （2041）
Structure of highly dynamic pixel based on adaptive integral capacitance
…… LIU Suiyang，GUO Zhongjie，YU Ningmei，XU Ruiming （2051）
Deep reinforcement learning intelligent guidance for intercepting high maneuvering targets
…… ZHANG Hao，ZHU Jianwen，LI Xiaoping，BAO Weimin （2060）
Calculation and error analysis of kinematic accuracy reliability of VSV adjustment mechanism
…… YANG Yifeng，WANG Yi，LIU Aoyu，LI Jia，XIE Liyang （2070）

Reliability assessment and lifetime prediction for train traction system considering multiple dependent components
…… TIAN Guishuang，WANG Shaoping，SHI Jian （2081）
Spray ignition and blow-out boundary of centrally staged annular combustor
…… WANG Zhihui，ZHANG Chi，GAN Zhichao，GAO Anwen，HAN Xiao，TAO Wenjie （2091）
Modelling method for pedestrian safety behavior in shared road spaces of pedestrian-traditional vehicle-autonomous vehicle
…… WU Ruoyu，YANG Dezhen，REN Yi，JIA Lulu，LI Xiaobin，WANG Zili （2099）
Detection of typhoons and estimation of eye position using satellite-based GNSS-reflectometry
…… ZHEN Jiahuan，ZHU Yunlong，YANG Dongkai，ZHANG Guodong，WANG Feng （2106）
Experimental study on beam characteristics of μHT-1 thruster under wide range adjustment of operating parameters
…… HUANG Dan，LONG Jianfei，CHENG Ye，WANG Jiabin，XU Luxiang，YANG Wei （2119）
Effect of plasma excitation on aerodynamic characteristics of airfoil in Martian atmosphere
…… YANG Xianggang，GAO Yongxin，WANG Zhongming，LI Yiwen，YAO Cheng （2129）
Design and optimization of warranty period of new products with two-parameter degradation
…… GAO Shuai，LI Yanhong，SUN Fuqiang （2137）
Aeroelastic optimization design of SpaRibs wing structure
…… ZHOU Quanzhi，YANG Youxu，SUN Lubin，ZHANG Xingcui，WU Yifei，HUO Mengwen （2148）
On-board timekeeping method based on improved Kalman filter
…… HUANG Yue，WANG Huiquan，TU Shilei，JIN Zhonghe （2157）
Direct lift control technology of carrier aircraft landing based on reinforcement learning
…… LIU Rendi，JIANG Ju，ZHANG Zhe，LIU Xiang （2165）
Energy consumption prediction of aircraft ground air conditioning based on ISCA-DBN
…… LIU Han，LIN Jiaquan （2176）

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2025.0060

低空智能网联体系

张学军[1,2,*]，刘法旺[3]，张祖耀[1,4]，田野[3]

(1. 北京航空航天大学 电子信息工程学院，北京 100191；2. 空地一体新航行系统技术全国重点实验室，北京 100191；
3. 工业和信息化部装备工业发展中心，北京 100846；4. 网络化协同空管技术北京市重点实验室，北京 100191)

摘　　要：近日，低空产业联盟对外发布《低空智能网联体系参考架构(2024版)》报告，该报告作为框架性的文件，旨在用最精确简短的内容体现低空智能网联体系的关键构成，但缺乏对体系背后隐含的科学方法、理论基础、实现途径等内容的详细描述。基于此，围绕该报告的内容，对低空智能网联体系发展现状、架构设计思路、关键技术等进行了全面的阐述，旨在对报告涉及的内容做进一步的分析和解读，为后续围绕低空智能网联体系的开发和建设工作提供科学的理论参考。

关 键 词：低空智能网联体系运行概念；低空智能网联体系架构设计；低空智能网联应用服务系统；低空智能网联数据与服务支撑网络；低空智能网联机载终端与基础设施；低空智能网联体系关键技术

中图分类号：V2; V24; V247.1

文献标志码：A　　**文章编号**：1001-5965(2025)06-1793-23

低空经济方兴未艾，低空领域正成为中国经济建设和产业发展的重要方向。随着低空装备制造的发展升级及各试点地区空域的进一步开放，未来的低空空域发展将呈现出高密度、高频次、大规模、复杂异构等特点，载人出行、低空物流、公共服务、文体旅培等应用场景层出不穷，不同的应用需求将推动系统建设、应用服务、装备制造等低空业态的平稳落地。

近日，在2024低空装备产业创新发展大会上，低空产业联盟对外发布《低空智能网联体系参考架构(2024版)》报告[1]。该报告提出，低空智能网联体系作为低空产业结构转型的核心，旨在整合多种技术手段和服务能力，满足不同运行场景对应用服务、资源调度、信息共享及设施设备的需求。要实现这一目标，不仅需要解决技术层面的难题，还需从体系架构和协同机制的角度进行深入研究。

结合未来发展需求和技术趋势，本文针对低空智能网联体系的关键问题，作进一步分析和探讨，旨在阐述该体系的构建依据，剖析体系层次结构的内在逻辑，为后续发展完善提供理论支持。通过对低空智能网联体系的系统性研究，本文力求为相关领域科研人员和产业从业者提供有价值的参考，助力低空运行体系化建设，推动低空经济全面发展。

1 低空智能网联体系概述

低空智能网联体系是由数字化、网联化、智能化的新型飞行器及机载设备、信息物理基础设施、数据信息网络及应用服务系统构成的综合性体系，可实现对低空空域的泛在感知、广域互联和智能管服应用，支撑低空典型应用场景的安全高效运行。该体系是低空经济发展的重要基础，是低空产业结构转型的关键支点，是低空空中交通管理(air traffic manage, ATM)的核心引擎。

1.1 低空智能网联体系发展现状

欧美国家和地区传统航空领域特别是通用航空尤为发达，在运行概念、平台构建、商业模式开发等方面有具有深厚的积累，其低空无人机产业经过十余年理论沉淀、技术开发和探索实践，出台了

收稿日期：2025-01-23；录用日期：2025-03-18；网络出版时间：2025-04-11 10:47
网络出版地址：link.cnki.net/urlid/11.2625.V.20250411.0921.002
*通信作者. E-mail：zhxj@buaa.edu.cn

引用格式：张学军，刘法旺，张祖耀，等. 低空智能网联体系[J]. 北京航空航天大学学报，2025，51(6)：1793-1815.
ZHANG X J，LIU F W，ZHANG Z Y，et al. Overview of low-altitude intelligent networked system[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1793-1815 (in Chinese).

一系列政策法规、标准文件及技术方案。而中国对低空的开发起步较晚，近年来结合自身需求进行了一系列探索和实践，形成了相关的政策标准、技术报告等，以支撑低空产业的持续发展。

1.1.1 国外发展现状

1）美国

围绕无人机发展，美国低空交通运行管控主要分为以下3个阶段：特殊用途无人机融入美国国家空域系统（national airspace system, NAS）建设、无人机交通管理系统（unmanned aircraft system traffic management, UTM）建设、先进空中交通（advanced air mobility, AAM）体系构建，如图1所示。

① 阶段1：特殊用途无人机融入美国NAS建设。20世纪末，由于民用无人机技术发展和市场化条件均未成熟，无人机应用局限于军用或其他特殊用途无人机，用空需求单一，无人机的运行没有对美国NAS构成安全威胁。因此，此阶段美国对无人机的空中交通管理的重点，主要在如何使无人机具备与有人机同等的安全水平方面，使得这类执行特殊任务的中大型无人机能够基于传统民航的通信、导航、监视及信息（communication, navigation, surveillance and information, CNSI）基础设施体系，安全融入国家空域。

2012年开始，美国国会和美国联邦航空管理局（Federal Aviation Administration, FAA）等相继出台了一系列法案和文件[2-3]，对无人机的技术研究、研发制造和运行管理标准等进行了设计，以保证在不降低现有航空系统效率和安全性的前提下，实现无人机系统（unmanned aircraft systems, UAS）安全融入NAS。

② 阶段2：UTM建设。随着无人机技术的快速发展，无人机的应用场景得到很大扩展，FAA预见到无人机的运行规模、飞行密度、用空需求将急剧增加，但现行的ATM系统的CNSI基础设施体系和管理方式无法为无人机安全、高效、经济的运行提供交通管理和服务。因此，FAA开始着手构建一种为无人机的安全运行提供管理和服务的系统。

2015年，FAA与美国航空航天局（National Aeronautics and Space Administration, NASA）联合成立了UTM研究转化小组，旨在为FAA提供量化和可落地的UTM研究成果和技术方案[4]。美国各单位陆续发布了相关的法案标准[5-6]，启动了一系列项目（如UTM Pilot Program、UAS Integration Pilot Program等），为UAS技术方案和应用场景的测试和评估提供了保障。经过对多年的试点和论证进行梳理，FAA的NextGen办公室发布了《UTM运行概念1.0版》[7]，明确了面向真高120 m以下的UAS的运行概念，并在2020年发布了《UTM运行概念2.0版》[8]，对UTM运行概念进一步完善。

③ 阶段3：AAM体系构建。美国的无人机运行管理系统经过多年技术验证和运行概念的梳理，再结合其在低空产业上商业运营的经验，基于需求最明确的载人和载货运输场景，着力在城市区域和城郊及以外区域构建AAM体系，实现对现有城市交通运输体系的扩展。该体系需要新型的智能化飞行器及CNSI基础设施的支撑，并可通过信息网络实现数据的广域互联，为低空的安全运行提供智能化的服务和保障，持续推动现有ATM系统的变革性发展。

早在20世纪中期，城市空中交通（urban air mobility, UAM）已经作为一种城市出行概念开始出现，但由于成本、安全性等因素，并未成为大规模商业化的出行方式。随着无人机、电动垂直起降（electric vertical take-off and landing, eVTOL）飞行器等新型智能飞行载具问世，UAM作为日常运输方式成为了可能[9-10]。

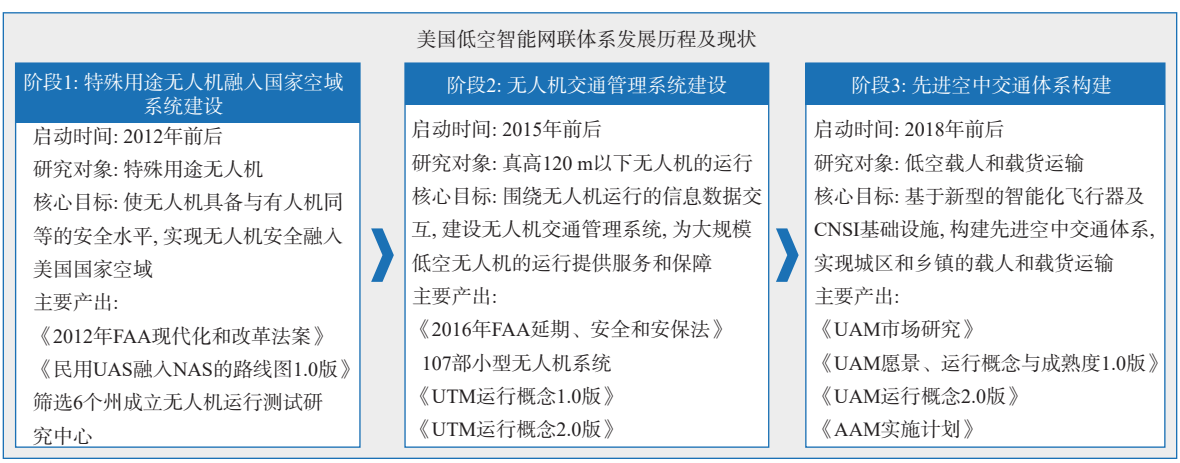


图1 美国低空发展历程及现状

Fig. 1 The development process and current situation of low-altitude development in the United States

2018 年开始, NASA 发布了一系列报告、文件, 分析了未来 UAM 相关技术和市场的可行性[11], 定义了 UAM 的发展阶段、运行环境、空域设计、CNSI 基础设施需求等概念[12-19], 明确了低空交通管理与传统 ATM 在运行概念、所需能力上的差异性, 对低空交通管理的技术手段、信息物理基础设施等提出了更高的要求, 同时将面向城市载人和载货运输的 UAM 定义为 AAM 的一个子集, 并且近年一直在对 UAM 和 AAM 的运行概念进行完善[20-21]。

然而, 美国作为传统航空强国, 通用航空产业非常发达, 其拥有世界领先的飞行器研制水平、完整的基础设施和服务保障产业、成熟的商业模式、深度的行业认可及完善的政策法规支撑。因此, 对于无人机、eVTOL 等新型低空飞行器的低空交通管理上, 传统行业运营商希望尽可能地利用现有基础设施实现短期盈利, 这在一定程度上限制了新技术、新平台的研发和建设进程。

2) 欧洲

与美国 FAA 同时负责美国航空安全监管和空中交通管理不同, 欧洲各国由于面积较小, 空域碎片化问题突出, 需要跨国性质的机构或组织对其空域、航空安全及空中交通管理等进行统一监管。

① 阶段 1: 无人机基础安全管控。20 世纪初, 欧洲各国主要通过无人机执行相对特殊且专业的任务, 无人机数量有限且飞行环境较为简单, 监管的关注点更多放在如何确保这些无人机不干扰公共安全、不影响民众日常生活方面, 进而建立起基础的法律框架和安全准则。

从 2009 年开始, 欧洲单一天空空中交通管理研究(Single European Sky ATM Research, SESAR)、欧盟等机构和组织就已经开始考虑无人机与现有空中交通管制(air traffic control, ATC)交互的需求[22], 并陆续发布了一系列文件, 对欧洲民用无人机, 特别是对遥控无人机(remotely piloted aircraft systems, RPAS)的应用前景、市场、现行标准、关键技术、监管框架等进行了阐述和探讨[23], 同时由欧盟航空安全局(European Union Aviation Safety Agency, EASA)为无人机的安全引入和运营制定法律框架、详细的运行规则、行业标准等[24]。

此阶段的核心工作是为初步进入民用领域的无人机建立安全运营的制度基础, 也使监管部门和公众逐渐形成“必须对无人机进行立法管控”的共识。

② 阶段 2: 多类型、多场景无人机交通管理。随着无人机技术的突破和市场需求[25] 的飙升, 无人机不再只是小型多旋翼或用于特定场景的 RPAS, 各种固定翼、大型无人机及可执行载人运输、物流配送、公共服务等复杂任务的机型不断出现, 给传统 ATM 体系带来了前所未有的挑战。

为应对复杂多变的低空运行需求, 欧洲在立法和技术层面“双管齐下”, 一方面通过大量的修订和新颁布条例(如 2018/1139 号、2019/947 号、2019/945 号等[26-28]), 针对无人机设计、生产、运营、人员资质、风险评估等环节做了系统化的规定; 另一方面, 着手构建并完善无人机交通管理系统 U-space, 该系统面向未来欧洲无人机的大规模运行, 能够为无人机提供远程识别、动态空域管理、冲突探测与回避等核心服务[29], 并持续发布了多个版本的《U-space 运行概念》[30], 对运行环境、空域管理、运行服务、飞行规则、系统架构等内容进行了全面的分析, 同时基于现有基础设施和技术水平, 对 U-space 后续实施阶段进行了阐述。

在这一过程中, 欧洲逐步确立了针对无人机运行的风险分级管理, 并明确了数据交互标准等关键技术要求, 为后续大规模无人机空中交通管理及示范应用提供了可落地的监管和技术框架。

③ 阶段 3: 欧洲“单一天空”的全面融合。随着无人机技术和 eVTOL 等新型飞行器的发展, 欧洲在低空交通管理问题上的关注重点进一步转向“大规模无人机与新型飞行器的协同运行”, 并在单一天空的整体愿景下, 将 U-space 系统和传统 ATM 体系更为紧密地结合, 强调通过统一的法规、标准和基于风险的安全监管体系, 确保欧盟范围内大规模无人机的安全运行, 并通过 U-space 实现数字化空域管理, 进而支撑 eVTOL 等新型飞行器的高密度、自动化、可持续安全运行, 为欧洲有人机和无人机的融合运行奠定基础[31-32]。

该阶段的核心工作是将 U-space 下的无人机及其他低空飞行器交通管理体系真正融入到欧洲单一天空的主干架构中, 实现无人机与有人机在统一规则和空域资源下的高效协同。根据《欧洲 ATM 总体规划 4.0 版》[33] 的规划, 欧洲将继续完善对 eVTOL、无人机等新型飞行器在低空出行场景的顶层设计, 使低空交通管理真正成为未来航空生态不可或缺的一部分。

欧洲在构建低空交通管理体系的过程中(见图 2), 围绕 U-space 框架的开发和实施, 在动态空域管理、无人机识别及轨迹管理等技术领域进行了深入探索和实践, 为政策调整和技术优化提供了宝贵的数据支持, 但仍然面临各成员国监管标准不统一、传统 ATM 体系与新型低空飞行器交通管理衔

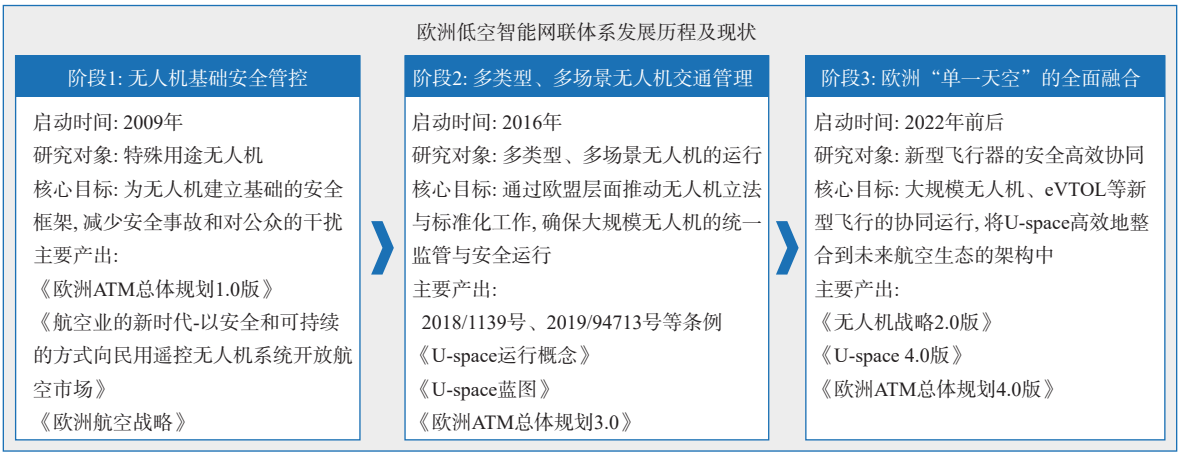


图 2 欧洲低空发展历程及现状

Fig. 2 The development process and current situation of low-altitude development in Europe

接困难、数据信息协同能力低、社会接受度不足等诸多问题。

1.1.2 国内发展现状

2010年, 国务院、中央军委印发《关于深化我国低空空域管理改革的意见》[34]。随后, 多个省份陆续开展了低空空域管理改革相关试点, 进行了多种低空运行管理模式与技术体系的探索与尝试, 相关职能部门围绕低空运行管理和技术实现发布了一系列法律法规、国家标准、规划建议和实施方案等文件, 为中国低空领域的发展积累了许多经验。

2023年12月的中央经济工作会议明确提出, 打造生物制造、商业航天、低空经济等若干战略性新兴产业。2024年全国两会期间, 低空经济首次被写入政府工作报告, 报告提及要加快发展新质生产力, 打造低空经济等新增长引擎[35-36]。随后, 全国各地响应积极, 纷纷出台各类举措支持低空经济的发展和智慧低空基础设施的建设。当前, 已有多家单位、组织和机构, 结合自身业务需求和技术理解, 围绕低空运行管理, 发表了报告、白皮书, 对低空运行概念、低空智能网联技术及未来低空发展思路等做了详细阐述。

1) 中国电子科技集团

中国电子科技集团基于自身在航空领域系统和装备的特有优势, 总结发布了白皮书《低空航行系统: 拥抱低空经济, 安全智慧出行》[37], 主要围绕低空经济时代背景、低空航行面临的挑战等方面展开, 旨在为低空经济发展提供系统解决方案, 推动低空航行安全智慧发展。首先, 从运行概念出发, 对各类型飞行器及用空需求、起降场及CNSI基础设施、飞行监管服务系统等的概念进行了阐述。其次, 围绕空域管理、飞行服务、CNS技术、自主飞行、安全监管、信息服务六大关键能力进行了详细分析, 并在此基础上对低空航行系统的总体发展阶段进行了规划, 如图3所示。

图4为中国电子科技集团低空航行系统的5层体系架构。该架构对整个低空航行系统的系统组成、发展演进做了详细的规划和分析, 从应用场景对应到具体业务能力, 再深入到各业务能力和服务供给, 这些服务通过中间层的业务网、通信网对基础的资源进行调度和管理, 最终实现对整个体系的支撑。

该白皮书以低空航行系统的运行概念的角度, 描述清楚了低空运行、监管及服务的范围和对象, 明确了从运行概念到业务能力需求和服务, 再映射到具体技术实现的能力需求这样一个科学的从理论到实践的方法路径, 可以为低空智能网联体系的构建方法提供指导。虽然该白皮书逻辑清晰、见解独到、内容全面, 但是低空航行系统是一个庞大的概念, 报告中的描述大多较为宏观, 仍有较多问题和内容未能详尽, 对现有低空运行体系的发展和建设提供的帮助有限。

2) 通信网络运营商

作为未来低空运行的参与方, 三大运营商出于自身的业务发展需求, 发布了若干相关报告或白皮书。

① 中国移动。中国移动发布《低空智联网技术体系白皮书(2024)》[38], 分析了低空经济产业链的分布和需求情况, 尤其是对低空信息基础设施的CNS需求进行了详细的分析, 明确了CNS能力在低空发展中的重要意义。该报告基于5G-A低空智联网端到端的技术架构, 围绕空地协同、通感融合及安全可信的理念, 梳理出了包含网络、平台、应用3个层次的低空智联网技术体系, 如图5所示。

该体系明确了CNS在低空信息基础设施上的核心地位, 对涉及的技术从所需能力到技术指标进行了梳理, 并利用现有设施对网络进行了测试分析。但该白皮书中提到的技术手段较为单薄, 对低

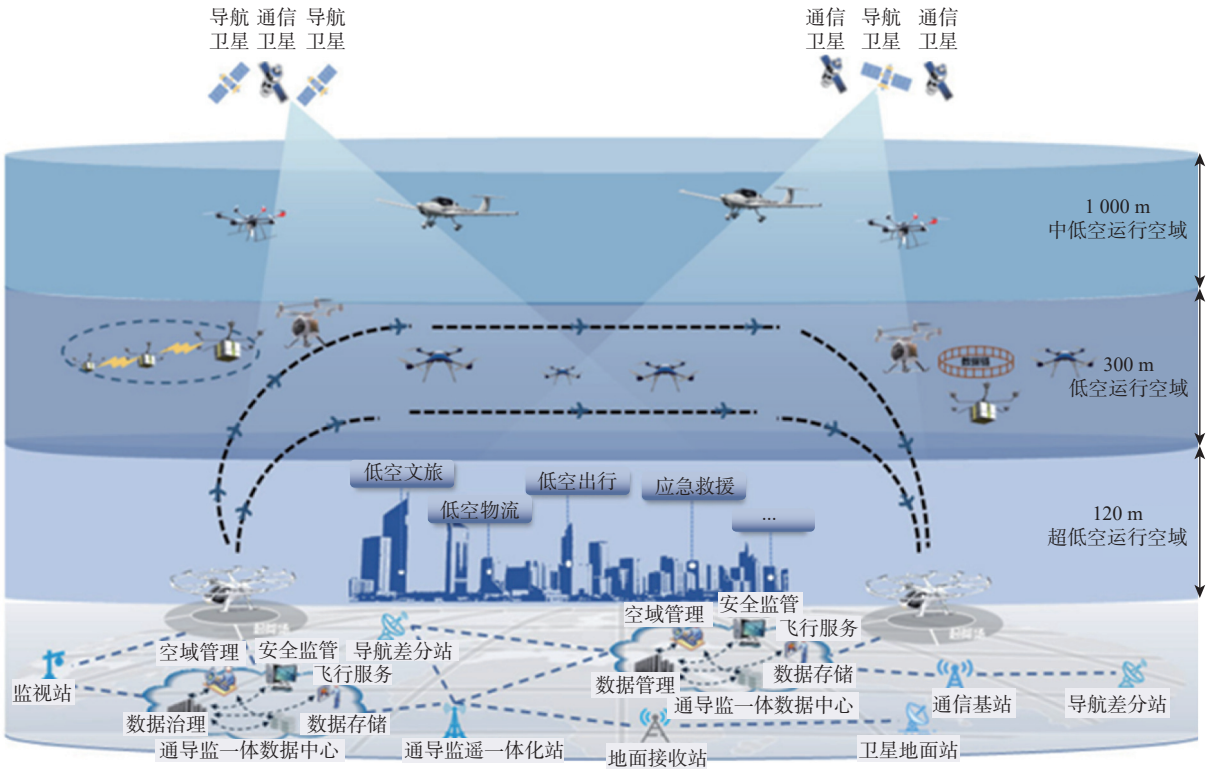


图 3　中国电子科技集团低空飞行运行概念示意图

Fig. 3　Operational concepts of low-altitude navigation by CETC

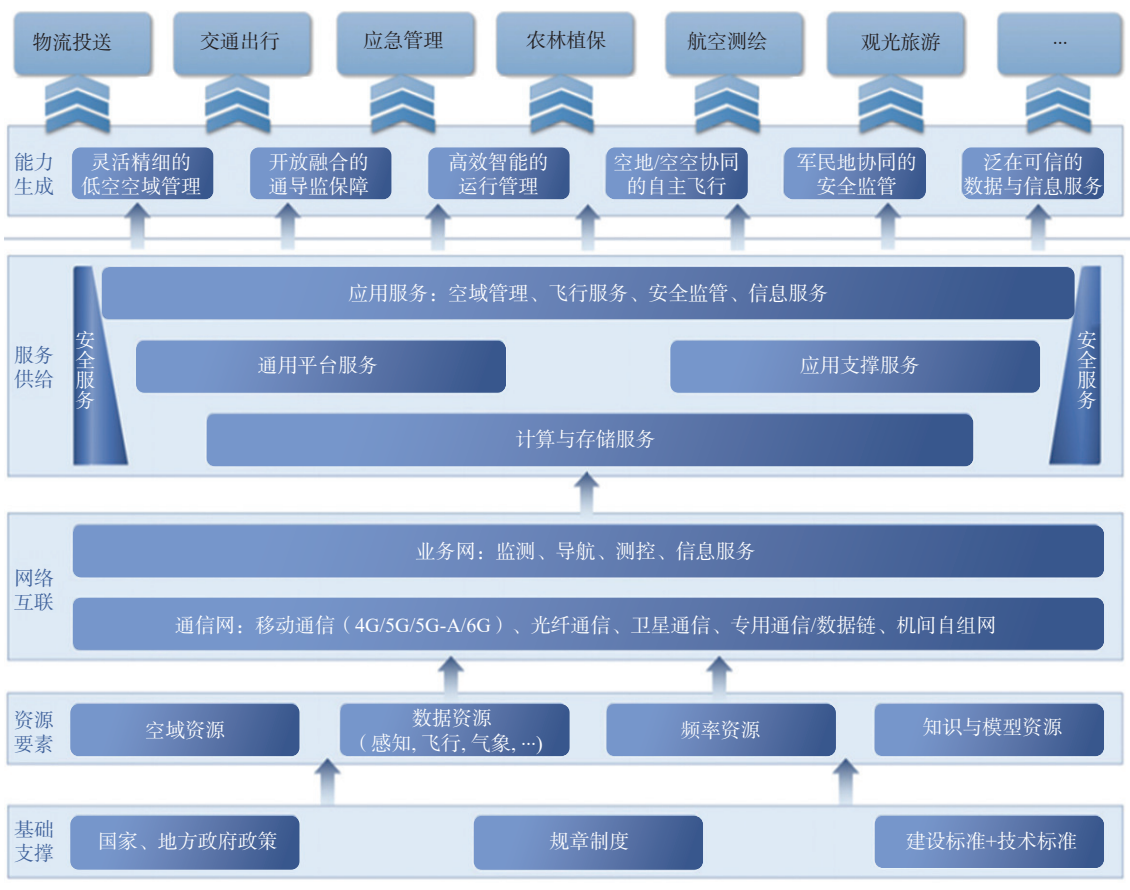


图 4　中国电子科技集团低空航行系统体系架构

Fig. 4　System architecture of the low-altitude navigation system by CETC

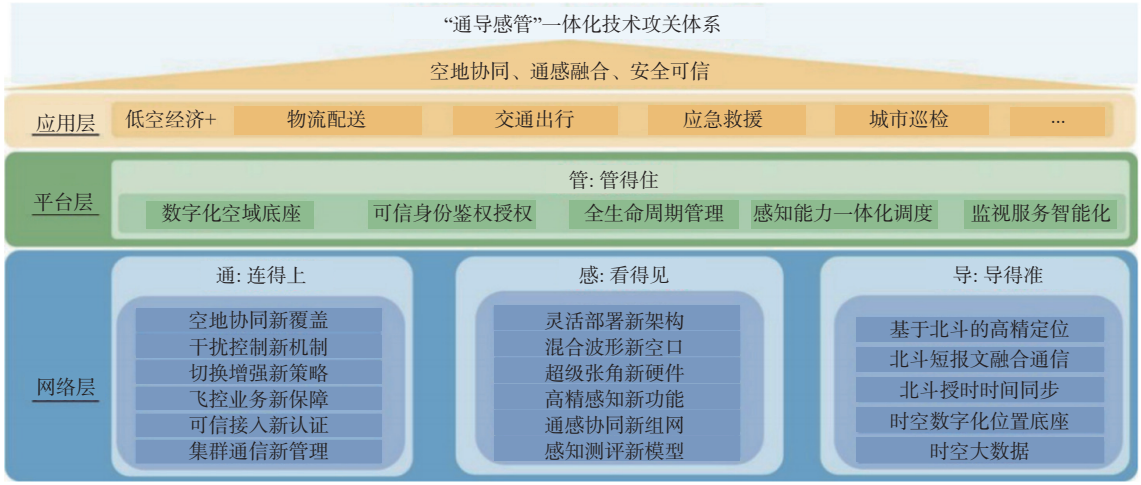


图5 中国移动5G-A低空智联网技术体系

Fig. 5 Technical framework of 5G-A low-altitude intelligent network by CMCC

空运行相关业务的梳理不够深入和全面，因此，仅凭该架构提及的内容不能很好支撑未来低空运行的开展。

② 中国联通。中国联通与广东通信学会、中国信息通信研究院等机构合作发布了《低空智联网发展研究报告(2024年)》[39]，对各国低空产业的战略规划和发展现状进行了分析，提出了以通信网、感知网、导航网、气象网、算力网为核心的低空智联网整体架构(见图6)，对“五张网”涉及到的发展趋势、现有问题、技术方向等内容进行了详细阐述，并且在全国范围内进行了技术验证和部署试点，获取了宝贵的实验数据。此外，该报告还对低空智联网下典型应用场景的业务特点、发展阶段、技术需求等做了详细分析，为低空智能网联的体系化发展和落地应用提供了良好的基础和支撑。

③ 中国电信。中国电信牵头发布了《通感一体低空网络白皮书》[40]，提出面向低空物流、地理测绘、应急救援等典型场景，依托通信、感知、算力等基础设施，构建通信、感知、智能计算一体化的低空智能互联数字化服务体系。该白皮书针对低空网络和监视服务业务，分别构建了低空网络功能架构和低空无人机监管服务平台架构，如图7所示。其中，低空网络功能架构主要包含应用层和功能层，应用层主要包含各典型低空应用场景，功能层基于通信感知网络、算力存储资源，为各应用场景提供通信、感知及智能计算能力。

而低空无人机监视、服务平台架构则分为了访问层、业务应用层、应用支撑层、数据资源层及基础设施层5个层次，如图8所示，围绕气象、空域、情报、飞行计划、监视信息、综合计算6类核心服

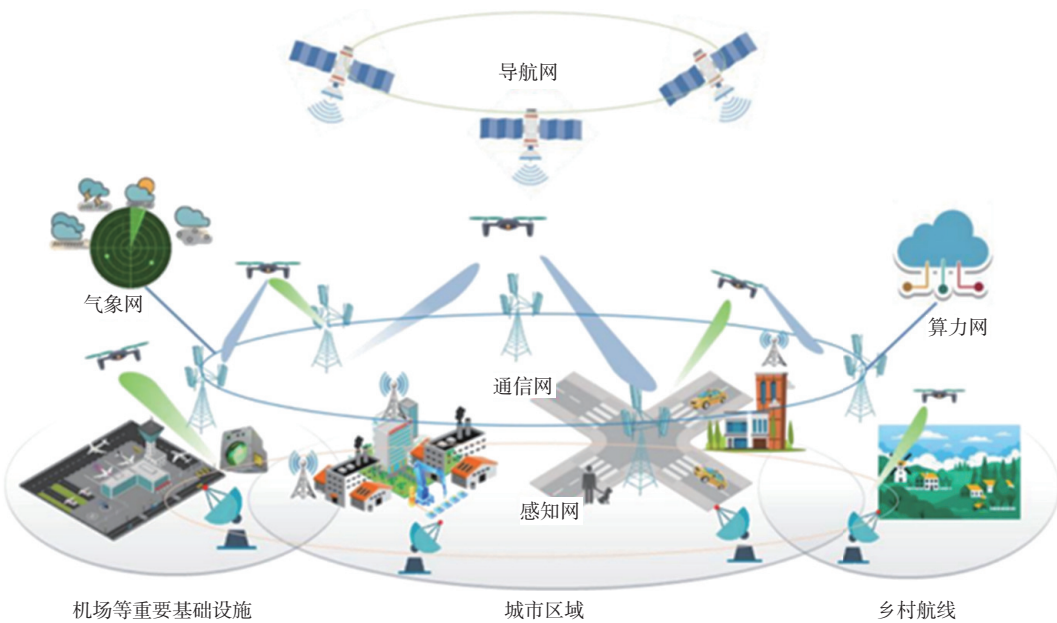


图6 中国联通低空智联网整体架构视图

Fig. 6 Architecture of the low-altitude intelligent network by CUCC

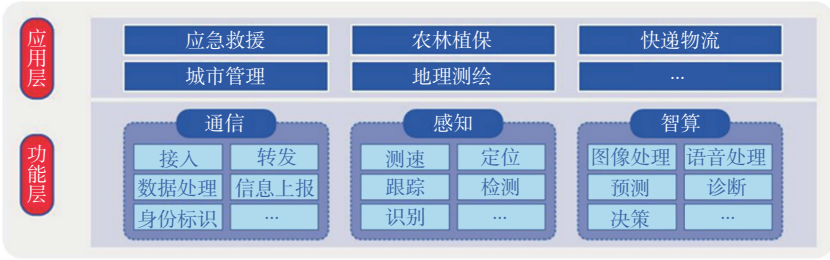


图 7　中国电信低空网络功能架构

Fig. 7　Functional architecture of the low-altitude network by CTC

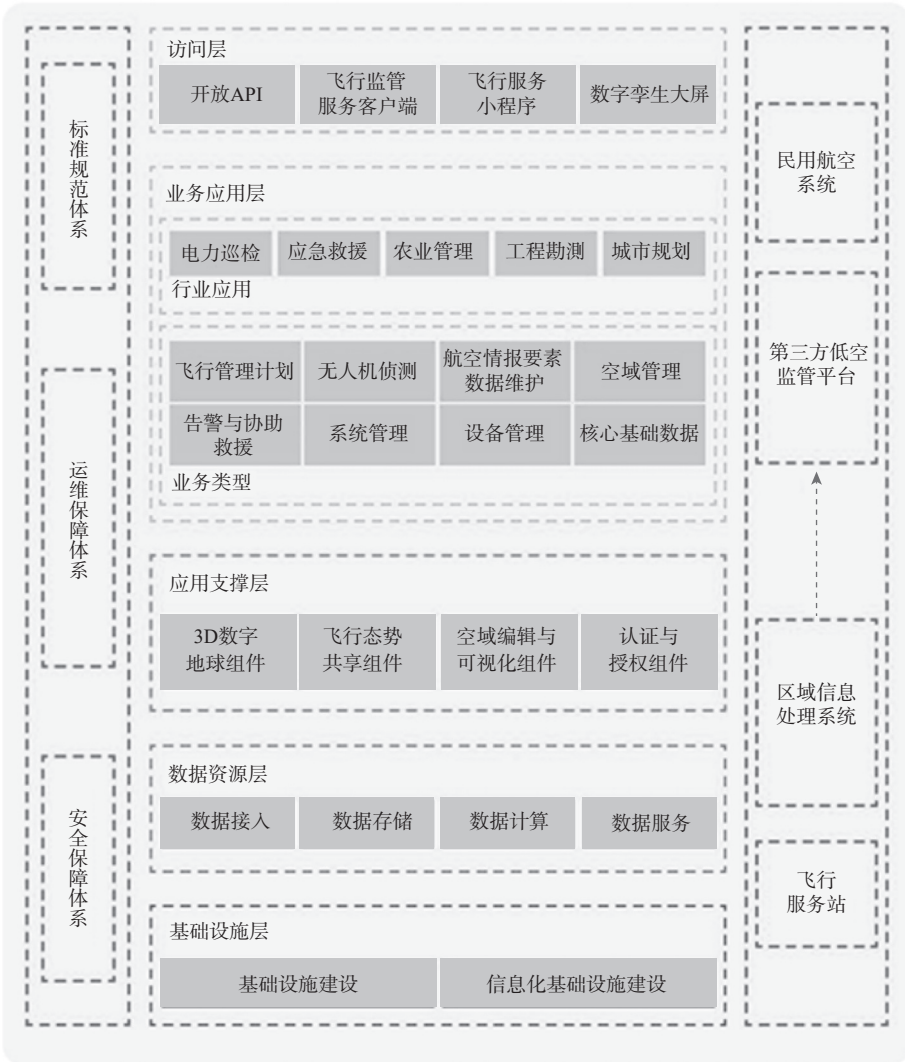


图 8　中国电信低空无人机监管、服务平台架构

Fig. 8　Architecture of the low-altitude UAV supervision and service platform by CTC

务, 以保障低空无人机运行的可视化、可调度和可监控。

在上述 2 个架构中, 低空网络功能的架构要素过少, 很难满足低空运行对网络能力的需求; 低空无人机监管、服务平台的整体框架虽比较完整、思路清晰, 但是层级间的功能划分有待商榷, 某些涉及低空交通管理业务的内容有所欠缺, 需要对相关内容进一步整合优化。

通信网络运营商发布的一系列报告或白皮书, 内容丰富、技术扎实, 对低空智能网联体系涉及到的核心技术, 特别是通信网络相关的技术内容进行了详细的理论阐述和测试验证, 为低空智能网联体系的技术能力分析和评估提供了宝贵的数据支撑。

3) 粤港澳大湾区数字经济研究院

粤港澳大湾区数字经济研究院(International Digital Economy Academy, IDEA)沈向洋教授团队针对深圳市发展现实需求, 在 2022 年和 2023 年发布了两版《低空经济发展白皮书》[41-42], 提供了未来

“异构、高密度、高频次、高复杂度”条件下提升飞行安全和效率的部分解决思路。该白皮书聚焦低空经济的发展，阐述了其发展现状，提出要构建全数字化低空智能融合系统（smart integrated lower airspace system, SILAS），如图 9 所示。

针对低空智能融合基础设施的建设，IDEA 提出要从物理设施、低空感知通信、空域共享、数字化管服 4 个角度，构建低空智能融合基础设施“四张网”，以解决资源共享、安全、效率、成本及协同运行等问题，如图 10 所示。

IDEA 提出的 SILAS 架构，强调利用智能化技术来实现整个低空运行的管控和服务，但与当前低空智能网联体系发展阶段的实际需求不太匹配。当前技术成熟度决定了低空智能网联体系的智能化水平发展是一个由低到高的过程，不是一蹴而就，整个体系的构建需要不断发展演变，持续优化。此外，提出的“四张网”也存在值得商榷的问题。首先，低空智能网联体系下的服务包括了通信、导航、监视、气象、起降场等各个维度的服务，是由整个低空运行的能力需求所决定，而 SILAS 中的“服务网”概念类似于低空飞行服务中心，其主要围绕低空 ATM 开展业务，显然不能满足低空运行的服务需求。其次，“空联网”中对 CNS 等设施设备的描述也较模糊，并未体现机载终端与信息物

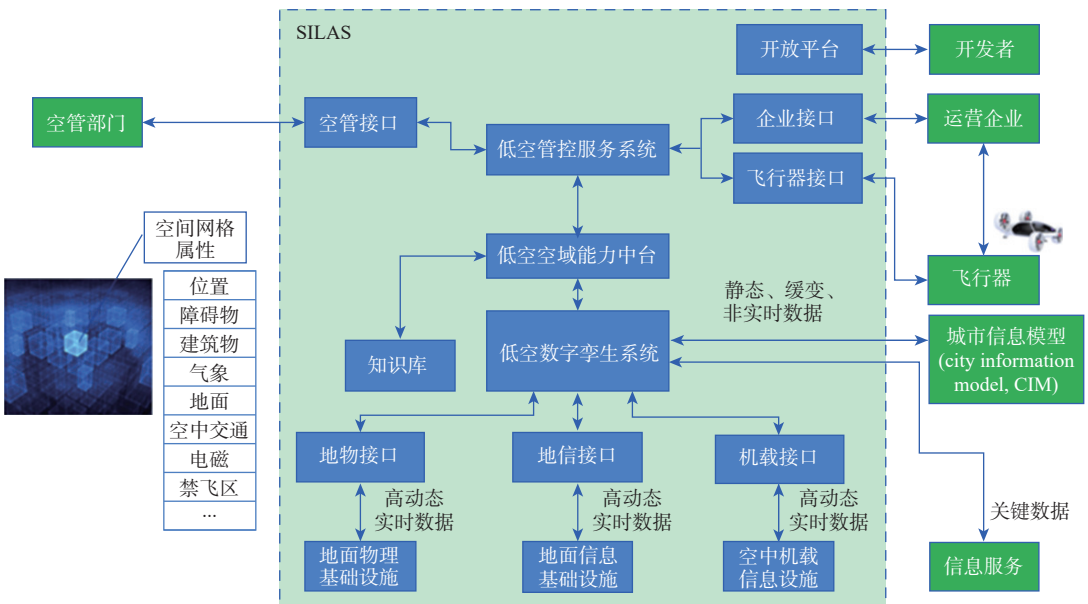


图 9 粤港澳大湾区数字经济研究院 SILAS 框架体系

Fig. 9 Framework of SILAS by IDEA

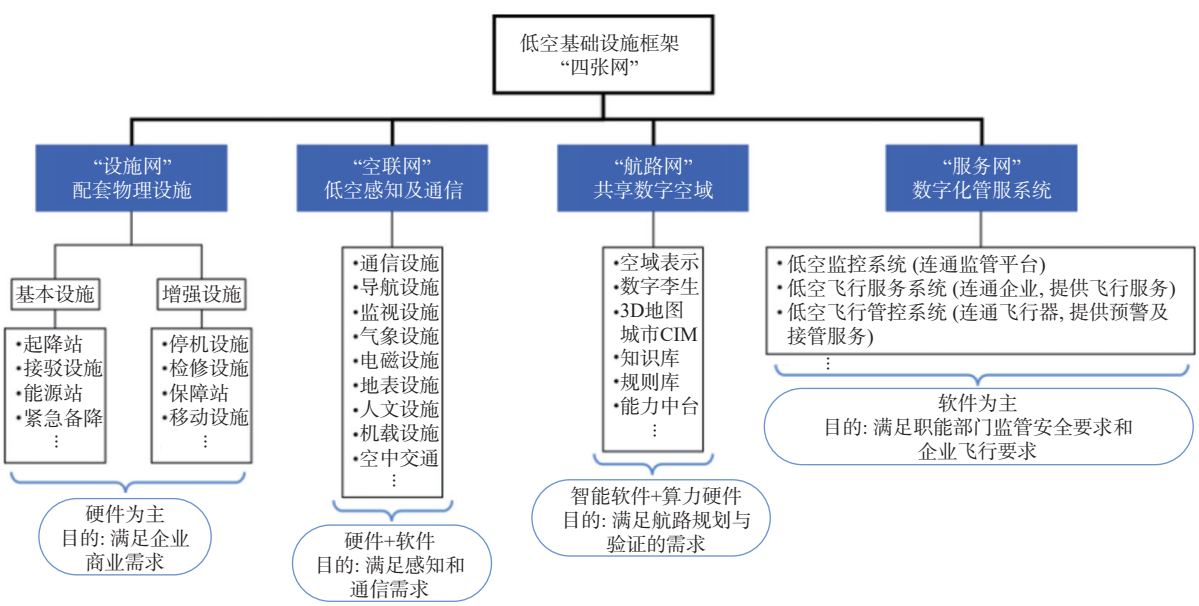


图 10 粤港澳大湾区数字经济研究院低空智能融合基础设施“四张网”

Fig. 10 “Four Networks” of SILAS infrastructure by IDEA

理基础设施的区别, 并且杂糅了一些外围的设施设备, 不适用于现阶段整个体系的装备制造和 基础设施建设, 将会影响低空智能网联体系的发展。

上述各方提出的报告或白皮书, 均从各自领域对低空交通管理的“是什么”、“管什么”及“怎么管”等问题做出深入分析和阐述, 但对低空智能网联体系缺乏一个贯穿应用场景、运行概念、业务需求、业务服务、所需能力、技术实现等方面的整体科学方法的梳理, 并且现有方案对低空智能网联体系如何融入现有 ATM 系统考虑较少, 极大降低了对现有资源的利用率。同时, 由于各地方政府认知、区域资源、支撑机构能力等方面存有参差, 给其政策制定、总体规划、技术体系选择、基础设施建设等带来了较大挑战。因此, 亟需一个顶层的低空智能网联体系框架来对各地方的具体实施边界进行约束, 解释低空发展的方向性问题, 保障未来低空运行活动的顺利开展。

1.1.3　发展现状总结

纵观欧美低空领域战略发展规划, 初始阶段由于需求的限制, 其研究的重点均为如何为特殊用途无人机提供基础的管控服务, 让无人机安全融入现有系统; 但随着无人机技术的不断发展, 其职能机构发现现有 ATM 体系远远无法满足未来无人机的运行需求, 因此, 开始基于各类低空飞行器特点, 梳理运行概念的差异, 挖掘运行模式和能力需求。同时, 分析如何利用现有的技术手段和基础设施, 满足越来越多的无人机、eVTOL 等新型低空飞行器的安全运行。

而中国对低空发展的迫切需求与新型低空飞行器技术的突破几乎处于同一节奏, 与欧美国家在基本国情、顶层战略、管理体制、发展规划、应用场景、市场需求、技术储备及基础设施条件等方面均存在较大差异, 因此, 不能对欧美的发展经验生搬硬套, 而是要构建结合中国低空行业的实际情况, 构建一套适合自身发展和产业特点的低空智能网联体系, 在新的赛道, “换道超车”。

1.2　低空智能网联体系设计思路

任何运行体系的构建都是一项系统工程, 既需要多方合力协同, 又需要在实践中逐步推进。运行体系的建设需要遵循科学的实践方法。运行概念的明确是整个体系构建的起点, 需要对目标任务、运行场景、角色分工及关联因素进行全面分析, 从中梳理核心理念和基本运行规则。而对能力需求的分析是连接运行概念与技术实现的关键环节, 通过对运行规则进行提炼, 形成运行模式, 以挖掘不同运行模式下需要满足的功能、性能指标和适用条件, 能够有效指导技术方案的选择和资源的合理配置。最终, 在技术实现的过程中完成运行体系的落地发展, 并在持续的应用实践中对运行概念、能力需求及技术手段进行迭代和优化, 将概念和需求转化为实际可用的系统能力, 如图 11 所示。。

构建低空智能网联体系是一个系统性工程, 需要多方协同、逐步推进。作为一项复杂系统, 低空智能网联体系的建设须遵循从运行概念到能力需求, 再到技术实现的科学实践路径, 以确保发展方向的正确性和可行性, 并在面对复杂环境和技术挑



图 11　低空发展的方法论

Fig. 11　Methodology for low-altitude development

战时可具有更高的适应性和创新性。

低空智能网联体系的发展演进过程如图 12 所示。运行概念是对低空运行场景、运行目标、角色分工和基本规则的整体描述，其为各方提供了共同的工作框架和发展方向。低空运行涉及多个参与方，包括低空飞行器制造方、低空交通管理和服务提供方、低空行业监管方、低空运营参与方、低空基础设施保障和服务提供方等，为实现低空的高效和安全运行，各参与方需在明确的运行概念指引下协同工作。

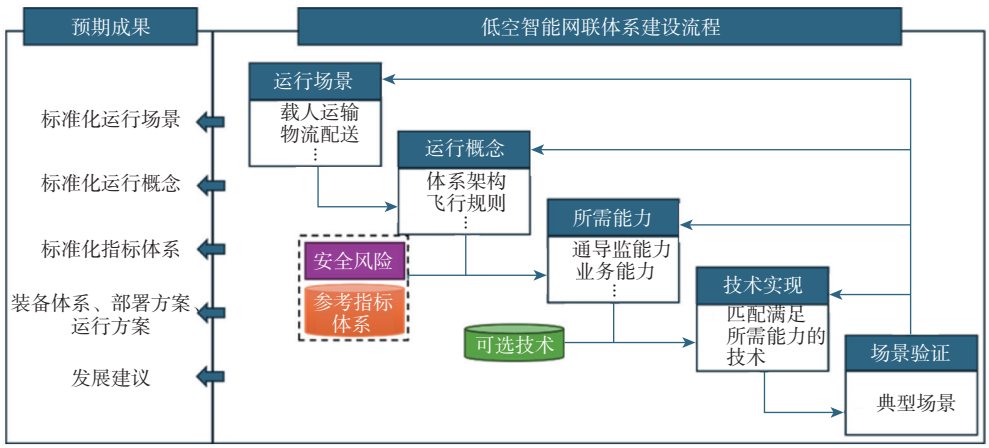


图 12 低空智能网联体系的发展演进过程

Fig. 12 Development and evolution of the low-altitude intelligent networked system

运行概念及体系框架的构建是一个持续发展、迭代优化的过程。当前，城市低空运行概念仍在孕育之中。虽然整体运行思路与体系框架已初步构建，但顶层飞行规则、安全保障框架等关键要素尚未完全明确，仍需在后续试点过程中进一步细化与验证。随着试点的推进，各类运行场景、角色分工、运行模式及飞行规则将不断在实际运行中校准与提升。在此过程中，通过问题发现与经验积累的闭环迭代，将为性能指标的合理设定和能力框架的持续优化提供科学支撑，进而推动低空智能网联体系由初步构想向成熟运行稳步迈进。

随着低空运行场景的不断丰富，应用系统和服务应更加多元化，以满足低空运行的多样化需求，实现经济价值的转化。同时，数据与服务支撑网络应当具备更多应用系统和多源异构数据接入能力，这是保证整个体系兼容性与可扩展性的核心载体和基本保障，为体系的持续发展奠定基础。

2 低空智能网联体系架构设计

低空智能网联体系的发展应当从运行概念出发，紧密围绕低空运行和装备制造的实际需求展开，明确技术体系的核心组成要素、功能需求及发展路径。首先，需要明确低空运行所需的核心组成部分，包括应用系统、数据与服务支撑、装备制造与基础设施保障。在此基础上，考虑载人出行、低空物流、公共服务等应用场景，可将系统分为 3 个层次：机载终端与基础设施层、数据与服务支撑层、应用服务层。机载终端与基础设施层为最底层，主要涉及机载设备与基础设施，包括机载通信设备、导航设备和信息物理基础设施等；数据与服务支撑层为中层，由接入网、交换网和服务网组成，负责数据交换、分发服务；应用服务层为最高层，包含各类航空管理与监管系统，为智能网联体系提供完善的服务和能力支撑。

为保障整个框架体系在发展过程中的规范性和安全性，参考架构还应当确立低空智能网联标准和安全 2 个体系。标准体系通过统一技术规范、接口协议等，以保障各设施设备、数据信息和系统应用的互联互通与兼容性；安全体系则聚焦于装备体系、网络数据、电磁频谱等关键安全问题，共同为低空运行提供可靠保障。

2.1 机载终端与基础设施层

机载终端与基础设施层主要参与方为低空飞行器制造方与低空基础设施保障和服务提供方。作为低空智能网联系架构的最底层，本层主要功能是使飞行器与相关基础设施具备一定智能，具有参与数据交换活动的能力，并可接受上层指挥进行智慧运行。

2.1.1 低空飞行器制造方

低空飞行器制造方主要负责制造具有智能的低空载具，以及服务低空智能载具的机载设备系统。低空智能载具依托智能航电系统和部件，实现飞行器的智能化，其核心能力包括可被感知、可被管理、具有智能航行能力 3 点。

服务低空智能载具的机载设备系统主要包括机载通信设备、多元导航设备、机载感知与识别设备、智能航行设备等。机载通信设备实现远程控制和数据传输,应具备高可靠性、低延迟、抗干扰、多模兼容、高速率、数据安全等特点;多元导航设备是指集成了多种导航传感器和技术的航空器系统,能采取不同的技术并利用不同的数据来源获取定位信息,以实现低空飞行器在空中飞行期间持续的导航准确、可靠和完整;机载感知与识别设备用于实现对地面的感知设备发送飞行器信息,并实现对周边空中交通态势的感知;智能航行设备是集成在低空飞行器上的系统,旨在实现自动化的飞行控制和管理功能。

2.1.2　低空基础设施保障和服务提供方

低空基础设施保障和服务提供方为体系运行提供基础设施保障和服务支持,具体内容包括地空通信基础设施、低空导航基础设施、低空监视基础设施、算力存储基础设施、反制设备、起降基础设施等。利用多种技术手段,依托地基和星基的现有资源,保持低空飞行器随时与智能网络链接,为低空飞行器提供导航、通信、感知、决策等能力。

2.2　数据与服务支撑层

作为整个架构体系的中间层,数据与服务支撑层将机载终端与基础设施层生成的数据供给至应用服务层。数据与服务支撑层由数据接入网、数据交换网、信息服务网构成,其主要构成体现在以下 3 个方面:

1) 数据接入网。数据接入网负责中层与下层的基本信息链接,通过通信接入、导航接入、监视介入等手段,将通信、导航、监视等终端设备接入到通信网络。通过标准化的数据格式和信息交换模型,加之标准化接口和服务框架,确保不同设备间数据无缝对接与统一管理,提升了系统内部的兼容性,并且为不同厂商设备和异构系统的接入提供了便利条件。

2) 数据交换网。数据交换网负责从下层反馈至中层的信息的处理,其包含通信资源管理、导航资源优化、监视数据融合、系统互操作、系统集成化管理等内容。宗旨:对系统中的通信、导航、监视等各类资源进行统一调度和优化分配,保证低空飞行环境中的资源得到充分利用,支撑系统在高负载和多任务情况下依然能够稳定运行,确保系统的鲁棒性、高并发性。

3) 信息服务网。信息服务网负责资源分配和对用户信息服务的支撑。这一部分通过发布-订阅模式和智能分析功能,为用户提供高效的数据分发和决策支持。信息服务网提供智能化资源调度与保障系统,分配通信、导航、监视等资源,并提供智能分析与服务能力。

2.3　应用服务层

作为体系架构的最上层,应用服务层根据中层的数据信息负责智能的计算、规划、管理,并将相关管理、调度命令反馈给数据与服务支撑层,应用于机载终端与基础设施层。应用服务层的参与方有低空运营参与方、低空交通管理与服务提供方、低空行业监管方。

2.3.1　低空运营参与方

低空运营参与方是低空智能网联体系中的直接应用者,基于低空运营管理系统开展业务。该参与方的核心业务包括飞行计划的制定和申报、飞行过程和运行状态的实时跟踪监控、飞行作业过程的统计与控制、任务和资源的动态调度与协同及信息的发布等,为低空智能飞行器的飞行提供基本的飞行任务规划、指导与申报。这些业务能力需要高可靠、低延迟的通信网络以保障飞行器与地面系统之间业务、指控数据的持续交互;高精度的导航技术支持复杂环境下的路径规划与动态调整;可靠的监视技术提供运行状态的实时反馈;要求飞行器具备自主飞行能力以应对突发状况。

2.3.2　低空交通管理与服务提供方

低空交通管理与服务提供方由进行低空交通管理和提供基础服务的企业、机构等组成,主要基于低空交通管理和服务系统开展业务。该方核心业务包括空域资源的分配与优化、飞行安全风险的评估与管控、空域容流量测算与管理、低空交通管制、低空气象和数据信息等服务的提供等。通过收集分析实时空域信息,为低空飞行器提供飞行过程的管理、调度。这些能力依赖于实时精准的飞行器位置和状态信息获取、广泛的空域态势感知、精细化的空域管控、高效的飞行冲突检测与消解、低延迟的数据传输等,为安全高效的低空空域运行提供保障。

2.3.3　低空行业监管方

低空行业监管方由低空监管相关政府职能部门组成,基于低空监管系统开展业务。该方利用低空基础设施提供方和数据与服务支撑层实现安全可靠的低空运行监管。其主要功能包括飞行器身份的识别认证、对运行异常或风险情况的应急处置、运行过程的违规监控与处置及运行事故的分析与追溯。可帮助维持空域安全有效运行、及时处理突发事件并有效打击违法事件。这些业务需要构建全面的数据获取与存储能力,通过对运行数据的

实时采集和安全管理为监管提供数据支撑；同时，还要求系统具备智能自主的分析和管理能力。

3 低空智能网联体系关键技术

3.1 应用服务系统

低空运行监管和服务的各参与方，基于运营管理系统、低空交通管理和服务系统及低空监管系统等多个应用服务云系统，基于数据驱动的低空智能航行、低空数字化空域管理技术、低空空域容流自适应匹配、风险评估等关键技术，围绕不同的低空运行场景、低空飞行器特性和任务特点，开展各类低空业务。表1为各应用服务系统所需关键技术及相应的业务能力[43-48]。

表1 低空智能网联体系应用服务关键技术

Table 1 Key technologies of the application service for the low-altitude intelligent networked system

应用服务系统	关键技术	业务能力[43-48]
运营管理系统	数据驱动的低空智能航行管理技术	通过请求和处理飞行计划、气象信息和容流信息等，实现无人机智能飞行计划生成、航迹动态优化和多无人机协同管理
低空交通管理和服务系统	低空数字化空域管理技术	基于低空空域数字化模型，支持航行方案的仿真验证、空域资源和运行态势的可视化，为实时决策提供数据支撑
	空域容流自适应匹配技术	根据实时飞行需求和空域负载动态调整容量，优化流量分配，避免区域性的空域拥堵和冲突风险
低空监管系统	实时安全风险评估	构建低空运行风险动静态指标体系，实时评估当前空域安全风险；在突发事件中快速生成风险报告，辅助应急决策
	身份识别与认证	系统支持无人机提供身份识别和权限认证、飞行活动记录溯源、防止非法设备入侵等业务，保障空域内低空飞行器的合法运行

3.2 数据与服务支撑网络

数据与服务支撑网络是整个低空智能网联体系的中枢，通过统一的标准化数据接口接入机载终端及基础设施的数据信息，并对异构数据进行信息模型构建，实现各类数据信息的高效交换与数据服务能力生成。

在传统民航空管领域，空管、航司、机场等各方的数据信息数量庞大、结构各异，信息交互需求日益增加，信息“孤岛”问题愈加显著，严重影响了民航业务的正常开展。为高效地开展信息共享和协同决策，国际民航组织（International Civil Aviation Organization, ICAO）提出了广域信息管理系统（system wide information management, SWIM）的概念，旨在建立统一的标准化数据接口和共享平台，使空中交通运输的各参与方能够在松耦合框架下进行信息交换与协同[49-52]。表2为对SWIM关键技术进行梳理的结果。

表2 SWIM关键技术

Table 2 Key technologies in SWIM

关键技术	功能描述	指标参考
数据标准化	统一航班信息、飞行情报、气象信息等异构数据接口	≥7种异构数据
信息交换模型构建	标准化数据交换模型（如AIXM、FIXM、WXXM等）	数据吞吐量≥10 Gbit/s
面向服务架构设计	提供态势感知、运行控制、飞行计划更新等服务支持	服务数据延迟≤100 ms

注：假设一个典型的全国性SWIM系统服务约5 000个终端（机场、航空公司、空管站等），每个终端每秒传输数据200 KB，吞吐量可达5 000 × 200 KB/s ≈ 8 Gbit/s。

面向服务的架构（service-oriented architecture, SOA）通过标准化接口进行信息交互，各服务提供者将各自的功能组件进行封装为服务接口，以供其他参与者按需通过网络进行请求/调用。中国民航系统基于现有基础设施，结合国外SWIM的建设经验，将面向服务架构分为空管业务信息系统、SOA核心服务，数据交互服务及技术基础设施4个层次，如图13所示[51]。

为满足低空ATM系统的网联化、智能化需求，特别是针对无人机数据信息的交互和智能决策，NASA发起了建设新型信息基础设施的数据和推理网络（data and reasoning fabric, DRF）项目[53]。DFR关键技术梳理如表3所示，DRF架构如图14所示。DRF凭借一个去中心化的信息网络，高度互联、相互交织，能够实现未来低空飞行环境中大规模数据与智能推理决策的高度集成与无缝接入，以确保未来低空场景的可靠运行。

根据上述民航SWIM的设计思路和国外相关低空数据服务网络开发经验，可以将低空智能网联体系下的数据与服务支撑网络的核心架构梳理为3层结构，即数据接入、数据交换及服务能力生成3个层级，如图15所示。

同时，定义低空智能网联体系下，表4为数据与服务支撑网络的所需关键技术，主要为数据标准化、信息融合与建模、数据交换及服务能力生成等技术。

3.3 机载终端与基础设施

3.3.1 空地协同新航电技术

为满足未来低空无人化和大规模运行的需求，新一代低空飞行器的航电系统需具备智能化的稳定通信、自主导航、全域感知、协同作业、智能决策等关键功能，以确保飞行器能够高效、安全地应对动态变化的空域和复杂的任务要求，实现复杂环

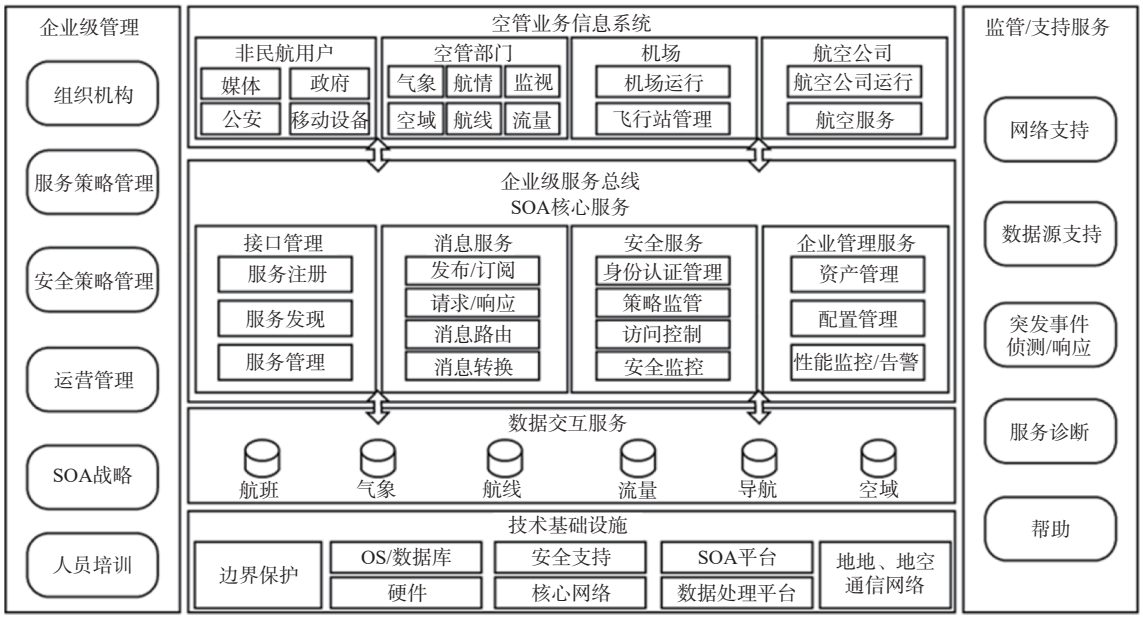


图 13　面向服务架构的空管数据共享平台[51]

Fig. 13　Service-oriented architecture for air traffic management data sharing platform[51]

表 3　DRF 关键技术梳理

Table 3　Analysis of key technologies in DRF

关键技术	功能描述	指标参考
数据融合	动态数据融合与推理生成	终端数量 ⩾ 10 000; 数据吞吐量 ⩾ 10 Gbit/s
智能算法	支持实时态势感知与任务规划服务	算法响应时间 ⩽ 200 ms
分布式服务架构	动态生成分发任务相关的决策支持服务	动态服务生成时间⩽1 s

注: 假设 10 000 架无人机, 每秒上报 100 KB 数据, 同时接受推理结果, 吞吐量可达 10 000 × 100 KB/s ≈ 8 Gbit/s。

境下的自主运行。

1) 一体化 CNSI 集成技术。CNSI 集成技术通过一体化系统架构, 将通信、导航、监视和情报服务模块紧密集成, 提供全面的机载航电解决方案。该技术实现了各个功能模块的协同工作, 形成一个统一的平台, 以简化飞行器的系统设计和操作流程。通信模块支持多种网络形式的融合(如卫星通信、5G 和 VHF/UHF 通信), 确保飞行器与地面控制和其他飞行器之间的实时数据交换。导航模块通过集成全球导航卫星系统(global navigation satellite system, GNSS)、惯性导航和地面辅助定位等技术, 提供高精度的定位与路径规划。监视模块融合 ADS-B 和雷达技术, 实时监控空域状况, 提供精准的碰撞预警和空域管理。情报服务模块利用大数据分析和云计算技术, 提供飞行决策支持和风险预警。通过这一高度集成的架构, CNSI 技术有效提升了低空飞行器在复杂环境中的协同能力和操作效率, 确保了飞行安全和任务的顺利执行。

2) 机载感知与避撞技术。机载感知与避撞(detect and avoid system, DAA)系统是防止空中相撞的最后一道防线, 通过集成多种传感器和智能算法, 实时感知飞行器周围环境, 确保飞行器在低空复杂空域中的安全运行。该技术涉及多源航迹管理、自主安全间隔丢失预警与恢复引导、冲突解脱



图 14　DRF 架构

Fig. 14　The DRF architecture

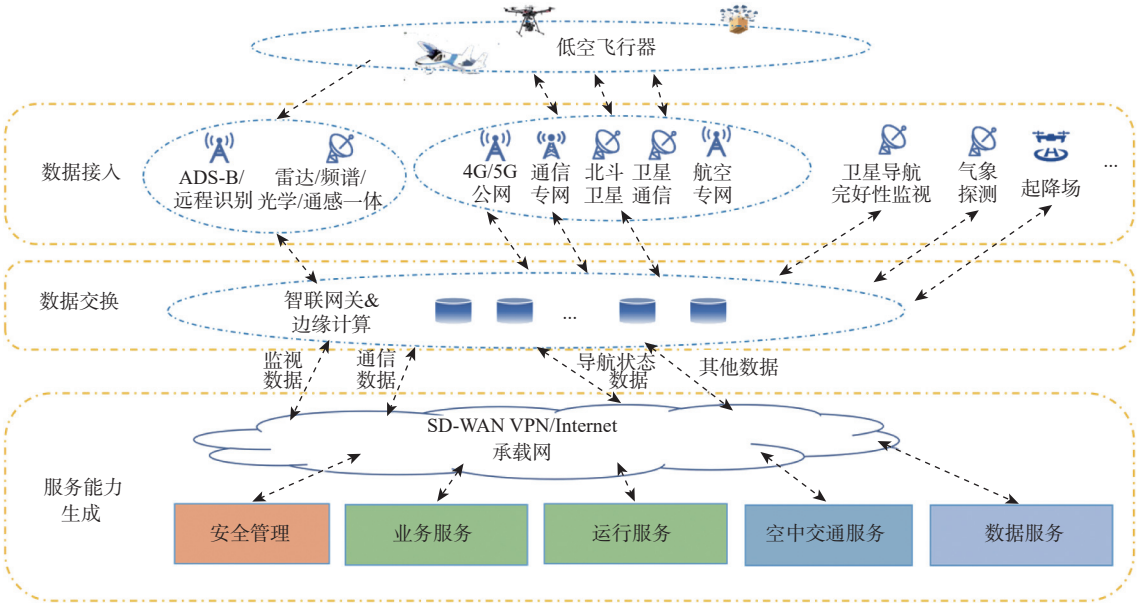


图 15 数据与服务支撑网络架构示意图

Fig. 15 Architecture of data and service support network

表 4 数据与服务支撑网络所需关键技术

Table 4 Key technologies required for the data and service support network

关键技术	功能描述	指标参考
数据标准化	接入机载终端与地面基础设施数据,统一标准化数据接口	≥15种异构数据
信息融合与建模	基于接入的融合数据扩展建模,实现多源数据的智能决策模型构建	单次建模时间 ≤ 10 s;融合误差 ≤ 1%
数据交换	实现城市级规模的低空数据可靠传输与动态交换	数据传输速率 ≥ 10 Gbit/s;延迟 ≤ 50 ms
服务能力生成	生成和分发飞行态势感知、冲突检测、路径优化等业务服务所需的底层数据和决策信息。	支持终端 ≥ 10 000;动态服务生成时间 ≤ 1 s

等关键研究,为机载系统提供多阶段的冲突告警与缓解措施,使无人机具有与有人机同等的“看见与避撞”能力。机载 DAA 技术还能够与空域管理系统、地面控制中心进行数据交换,提供更加全面的空域动态信息,从而实现飞行器之间的协同避碰,减少飞行安全风险。

3)自主飞行与智能决策技术。自主飞行与智能决策技术使无人机能够在无需人工干预的情况下,根据实时感知数据和任务需求自主执行飞行任务。通过集成的传感器系统、自动飞行控制和机载稳定性增强算法,无人机能够感知和分析周围环境,识别障碍物,实现无人机姿态控制与飞行路径动态管理,提高机身稳定性。智能决策系统是该技术的核心,其结合实时环境信息与任务目标,基于复杂的算法模型进行决策,自动优化飞行器的航向、速度和高度,确保飞行路径的灵活性与精确性,同时应对突发情况或环境变化,确保飞行的稳定性和安全性。

3.3.2 低空智能网联体系通信技术

1)能力需求

在复杂的城市环境和动态空域条件下,低空飞行器需要稳定、实时的通信链路来确保飞行安全、任务精准执行和多机协同合作。传统的通信网络难以满足低空飞行器对高带宽、低延迟和高可靠性等要求。表 5 为低空智能网联体系 5 方面通信能力需求的含义。在降低通信时延、提升数据传输速率、扩展通信网络覆盖范围的同时,保障连续性与完好性,成为低空智能网联体系通信能力的核心需求。

表 5 低空智能网联体系通信能力需求

Table 5 Communication capability requirements of the low-altitude intelligent networked system

能力需求	具体含义
传输时延	从发送端发送数据到接收端接收到数据所需的时间间隔
传输速率	在单位时间内可以传输的数据量,通常以bit/s为单位
覆盖率	通信网络能够有效提供服务的区域占总区域的比例
连续性	通信服务在一段时间内不间断的能力
完好性	在未检验出错误情况下,通信处理能够在一个通信处理时间内完成的概率

2)可选技术方案

为满足低空智能网联体系下通信能力的需求,需整合现有通信技术,为低空智能网联体系的发展,提出现实可行的技术方案。表 6 为现有通信技术手段及特点、需解决的问题、关键指标参考值等内容。

表 6　通信可选技术方案

Table 6　Optional communication technology solutions

技术手段	需要解决问题/缺点	指标参考
5G蜂窝通信	需要解决航空优先级、天线指向、高密度服务拥堵相关问题	带宽≥10 Gbit/s 空口时延≤1 ms
LEO卫星通信	可能具有双重功能（通信和导航增强）	通信延迟≤100 ms 上行速率≥10 Mbit/s
低空通信专网和数据链	需要部署多基站或中继设备	数据链: 传输距离≥40 km 传输速率≥300 Kbit/s 专网: 覆盖高度≥1 km 传输速率 100～1.024×10^5 Kbit/s (不同技术路径有差异)
北斗短报文通信	引入数据压缩技术, 减少单次通信的数据量	单条报文1 000Byte （北斗三）
VDL模式2	成本、容量和延迟方面存在问题	通信速率31.5 Kbit/s
VDL模式3	适当的带宽	通信速率31.5 Kbit/s
自组网电台	需解决跳转增加的网络传输延迟	峰值速率≥10 Mbit/s 单台距离≥30 km 组网能力≥64节点 跳数≥10 （不同产品差异较大）
激光通信	成熟度不足	
WiFi/蓝牙	短距离通信需要太多的地面站点以确保可行性; 易受发射器干扰; 适用低速合作目标	最大传输速率: 9.6 Gbit/s 24 Mbit/s(蓝牙) 传输距离: 100 m(WiFi)、 10 m(蓝牙)

① 5G 蜂窝通信。5G 通信技术在不同阶段的演进为低空飞行器的应用提供了逐步增强的技术支撑。从 Release 15 起, 5G 引入的增强型移动宽带(enhanced mobile broadband, eMBB)技术显著提升了低空飞行器的数据传输能力。随着 Release 16 的发布, 低时延高可靠通信(ultra-reliable low latency, URLLC)和大规模机器型通信(massive machine type communication, mMTC)得到优化, 重点解决了实时应用对低时延和高可靠性的需求, 支撑低功耗、广覆盖的无线网络实时交换数据, 为低空飞行器在复杂城市环境中的自动驾驶及大规模协同作业提供了必要的技术保障。Release 17 为上行控制和数据信道引入多个增强特性, 初步拓展空天地覆盖能力。Release 18 及后续版本, 5G 技术进入了更为先进的阶段 5G-A, 聚焦于增强网络的智能化和自适应能力, 特别是通过边缘计算、AI 驱动的网络优化和跨域协同等技术, 使低空飞行器能够在更加复杂和动态的环境中实现更加精确的实时导航、路径规划和空域管理。随着 5G-A 的发展, 低空飞行器将能借助更高带宽、更低时延及更强的网络覆盖, 实现更加智能化的 UAM 系统。

在低空运营场景中, 5G 蜂窝通信因其高带宽和低时延优势, 成为主要通信技术之一。其理论带宽可高于 10 Gbit/s, 传输延迟可降低至 1 ms, 能够支持高清视频传输、航迹管理和实时避撞等高频更新需求。现有 5G 基础设施日臻成熟, 可靠性较高, 适合在城市区域进行高密度部署。5G 基站的有效覆盖半径一般在 100～300 m 之间, 在空旷地区可能有所扩大, 而在城市中心区域则可能有所减小。为了弥补高频段信号在城市区域穿透力弱的不足, 可能需要更密集的 5G 基站部署, 而这同时也需要更高的成本。

② 基于低轨卫星通信网络和 6G 的天地一体化融合网络技术。国内低轨(low Earth orbit, LEO)卫星互联网星座主要包括了 GW 星座(星网)、G60 星座(垣信卫星等)及近期公布的鸿鹄-3 星座(上海蓝箭鸿擎科技有限公司), 规划合计超 3.6 万颗星, 3 个星座是未来国内 LEO 卫星互联网组网核心载体。

LEO 卫星覆盖范围可实现全球连接, 在地面基础设施不足的区域也能提供可靠的通信服务。小于 100 ms 的双向通信延迟可以满足低空运营的动态数据需求, 上行速率约 10 Mbit/s, 适合航迹管理和飞行计划更新等任务。然而, LEO 通信设备成本较高, 且信号在城市高楼环境中容易被遮挡, 在低空高密度应用场景(如无人机集群、城市空中交通), 大量用户同时接入 LEO 卫星网络会导致负载过高, 影响通信质量。

随着 6G 技术的发展, 将实现卫星与地面通信的深度整合。基于 6G 的天地一体融合网络利用 LEO 卫星网络, 面向偏远地区、海洋、航空等全域立体空间实现“泛在连接”和低时延服务, 为地面网络无法部署或建网维护成本高昂的区域填补覆盖空洞; 同时, 发展动态灵活的网络资源管理技术以适应高速移动的网络拓扑。此外, 6G 还将统一空中接口标准和频率, 确保不同网络环境下的兼容性和高效接入, 并通过星地智能动态协作网元优化覆盖和资源调度, 实现星地资源的高效协同。

③ 低空通信专网和数据链。低空通信专网适用于区域性低空运营, 专网能够根据低空无人机的需求进行优化设计, 具有更高的抗干扰能力和数据安全性, 其适用于飞行控制、监控和特定任务场景(如无人机物流配送)的通信。然而, 专网的覆盖范围有限(典型覆盖半径为几十公里), 需要通过部署多基站或中继设备扩展覆盖范围, 主要适用于区域性应用或关键任务需求。

低空数据链单台传输距离可达 40 km(组成专网后可达百公里), 传输速率在 300 Kbit/s 以上, 常用频点为 800～900 MHz。低空通信专网视技术路

径不同，通信速率在100～1.024×10^5 Kbit/s不等，覆盖高度可达千米级。

④ 北斗短报文通信。在偏远地区或紧急情况下，北斗短报文通信提供了低功耗和广覆盖的通信能力，其具有极高的抗干扰性和可靠性，适合在飞行器失联或地面信号中断时作为备份。但北斗三号14 Kbit的报文长度限制和秒级的通信时延使其不适合高清视频传输或实时控制需求。北斗短报文更适合关键状态信息回传和应急通信。

⑤ VDL模式2/模式3。最高频(very high frequency, VHF)通信作为航空领域的成熟技术，因其稳定性和多信道支持能力被广泛采用。使用的无线电频率须在117.975～137 MHz之间选择。最低的可分配频率须为118.000 MHz，最高的可分配频率须为136.975 MHz。各信道之间的间隔须为25 kHz。模式2仅支持数据通信，模式3支持数据通信和话音通信，其覆盖范围为50～200 km，适合城市低空空域的任务。VDL模式更适合作为飞行状态和计划数据传输的基础链路，并在必要时提供应急通信保障。

⑥ 低空自组网电台。低空自组网能够在无需固定基站的情况下，通过节点间的无线连接自动建立和维护网络，适应动态变化的环境(如无人机集群)。自组网支持中高速数据传输(10～100 Mbit/s)，无障碍干扰下通信距离超过30 km，具备较强的抗干扰能力和端到端加密通信的安全性，其分布式路由机制可在节点失效时自动重组网络，确保通信的连续性和可靠性。然而，由于多跳传输可能导致延迟累积，且能耗较高，自组网电台更适合用于无人机协同作业、灾害应急通信及城市物流等需要灵活性和覆盖广泛的任务场景。

⑦ 蓝牙/WiFi。蓝牙和WiFi作为主要的短距离无线通信技术，其具备独特的物理层和协议层特性，通常能支持100 m范围的数据通信。当前，蓝牙技术拥有毫秒级的唤醒延迟和极低功耗，数据传输率可达2 Mbit/s，适合小型设备的周期性广播和设备发现。WiFi通常能提供数Gbit/s的数据传输速率，支持高带宽、低延迟但功耗较高的场景。但由于技术特点和广播模式的发射功率受限，该类技术有效覆盖距离仅在百米范围，难以支持超视距(beyond visual line of sight, BVLOS)的无人机通信和监视需求；同时，广播消息易受多径和同频干扰，流量密集时碰撞概率上升；且广播链路本身不含加密或身份认证，单包最大承载数据量有限，存在被嗅探和伪造的安全隐患。

中国民用航空局发布的《民用无人驾驶航空器系统运行识别规范(征求意见稿)》[54]中将蓝牙和WiFi作为无人驾驶航空器广播式运行识别链路的支撑技术，定期向周边接收设备广播无人机身份、位置、状态等信息，实现对无人机全过程的连续、实态势感知。为保证短距内的实时高可靠通信，对广播周期、信息安全、接收端天线灵敏度、解码概率、系统数据处理能力，完成时延等内容做了明确要求，以确保蓝牙/WiFi广播在低空密集场景下的短距离、低时延、高可靠性运行识别能力。

3.3.3 低空智能网联体系导航技术

1) 能力需求

由于低空空域存在建筑密集、地形复杂、气象多变、信号干扰严重等影响，对导航系统的稳定性和可靠性提出了更高要求，必须采取多元化的技术手段来确保飞行器在复杂环境下的精确定位和安全飞行。表7为低空智能网联体系5方面导航能力需求的含义。在提升定位精度、维持稳定的更新频率的同时，保障可用性、连续性和完好性水平，成为低空智能网联体系导航能力的核心需求。

表7 低空智能网联体系导航能力需求

Table 7 Navigation capability requirements of the low-altitude intelligent networked system

能力需求	具体含义
定位精度	系统测量或计算目标位置的准确度
更新频率	系统在覆盖区域内，2次位置信息更新的时间间隔
可用性	在指定的覆盖区域内，多技术手段融合导航系统能够提供有效服务时间的百分比
连续性	导航系统在覆盖区域一定时间范围内持续提供位置信息的能力
完好性	系统出现故障并超出误差限制时，在规定的时间阈值内向用户发出告警的能力

2) 可选技术方案

为满足低空智能网联体系下导航能力的需求，需要整合现有的导航技术，为低空智能网联体系的发展，提出现实可行的技术方案。表8为现有导航技术手段及特点、需要解决的问题、关键指标参考值等内容。

① 惯性导航。惯性导航系统(inertial navigation system, INS)主要以MEMS惯性传感器和光纤陀螺仪广泛用于旋翼无人机、直升机等领域，提供轻量化、高动态下的自主导航能力。在低空物流等场景中，INS具备短时独立导航、抗遮挡和高动态适应能力，但受制于漂移误差和环境振动影响，需与GNSS或视觉导航融合。根据不同场景，其需满足高位置与姿态精度(误差小于10 m或0.2°)、宽动态范围(加速度5g～10g)、长连续导航时间(30 min至数小时)和强抗干扰性能，以确保复杂环境中的

表 8　导航可选技术方案

Table 8　Optional navigation technology solutions

技术手段	需要解决问题/缺点	指标参考
惯性导航	漂移误差累积; 高精度成本高	角速度精度 ≤0.01 (°)/s; 加速度精度 ≤0.01 m/s²; 累计位置误差 ≤ 0.1%×距离
视觉导航	需要与环境地图进行匹配(如SLAM)	导航精度 ≤ 0.2 m; 特征点匹配成功率 ≥95%; 算法延迟 ≤50 ms
卫星导航及增强	需与RTK、增强系统、完好性监测共同适用; 多径效应; 抗干扰	定位精度: ≤1 m(普通), ≤0.1 m(RTK增强) 信号丢失恢复时间 ≤5 s
高度计	城市环境中的气压高度计可能会有高达 152.4 m的高度误差	高度测量精度 ≤0.1 m 响应时间 ≤10 ms
LiDAR	高分辨率LiDAR设备成本较高	探测距离 ≤200 m 探测精度 ≤0.01 m 点云刷新率 ≥10 Hz

注: 不同制造商设备数据存在差异。

可靠运行。

② 视觉导航。视觉导航通过机载设备(如高分辨率 RGB 摄像头、深度摄像头、双目立体摄像头等)采集环境图像, 并结合地面设施(如视觉信标等)实现精准定位和路径规划。由于视觉导航不依赖外部信号(如卫星、地面基站等), 具有较强的自主性, 广泛应用于 GPS 信号弱或无法接收的环境下工作, 如城市峡谷、室内等。其优势在于结合信标后可显著提升定位精度(可达 0.1 m 以内), 降低对自然环境特征的依赖, 且信标成本低、易于部署; 缺点是需要提前开展背景数据提取, 适应性受限。受环境和设备精度影响, 其定位精度在 0.1～1 m, 复杂环境中需要<0.5 m; 信标检测范围 1～10 m, 响应时间小于 50 ms; 支持高分辨率图像处理, 帧率要求≥30 帧/s。

③ 卫星导航及增强。GNSS 作为适应性最广的导航技术, 在低空运行环境下受到多种因素制约, 其导航精度和稳定性受到严峻的考验, 因此, 需要配合多种技术, 以满足 GNSS 技术在低空各类场景下的高精度、高鲁棒性需求及对完好性的监测。

GNSS 通过机载设备(如 GNSS 接收机、RTK 机载模块等)和地面设备(如 RTK 基站、星基增强系统(satellite based angmentation system, SBAS)/地基增强设备(ground based angmentation system, GBAS)、完好性监测设备等)协同工作, 实现高精度位置信息的实时获取, 广泛应用于无人机低空运行中的自主起降、路径规划、精准任务执行和航迹跟踪等场景。SBAS/GBAS 是 GNSS 的有效补充, 其中, 单频 GBAS 服务可以提供水平/垂直方向 95% 以上的米级导航精度和 2×10^{-7} 的完好性能力水平, 而完好性监视设备提高了 RTK 系统的定位可靠性和完好性, 弥补了 RTK 系统本身无法提供完好性检测的不足, 其导航精度能达到厘米级, 数据更新率为 1～20 Hz, 具备抗干扰能力和多模多频接收能力, 以充分保障低空飞行高精度导航的稳定性和可靠性。

在 GNSS 的应用中, RTK 组网和增强网络是确保低空运行导航精度和可靠性的关键技术。RTK 网络通过多个基站的分布式部署形成网络化结构, 每个基站能够接收来自 GNSS 卫星的信号, 并将校正信息发送至无人机的 RTK 接收机。典型的 RTK 基站组网覆盖范围可达几十公里, 基站间距通常为 10～20 km, 以确保校正信号的连续性和精度。同时, 为了增强系统的覆盖范围和冗余性, 地面 RTK 网络可结合区域性的 GBAS 实现更大范围的导航增强支持。此外, 完好性监测系统同样需要通过组网形式进行部署, 通常在关键节点(如机场、交通枢纽)设置完好性监测设备, 实时监控 GNSS 信号的质量和完好性数据, 发现异常时及时发出告警信号。

④ 高度计。高度计通过机载设备(如气压高度计、激光高度计、超声波高度计)实时测量飞行器的高度信息, 适用于无人机垂直起降、低空悬停和近地飞行等场景。其优势在于垂直精度高、实时性强, 可有效保障低空飞行安全; 但气压高度计易受环境气压变化影响, 激光和超声波高度计则易受障碍物和天气条件限制。关键性能指标包括测量范围 0.1～10 000 m, 测量精度气压高度计为±1 m, 激光和超声波为±0.01 m, 响应时间小于 20 ms, 激光高度计需具备在雨雾天气下的适应能力。

⑤ 激光雷达。激光雷达 LiDAR 一般通过机载设备(如旋转式 LiDAR、固态 LiDAR、MEMS 扫描 LiDAR)实现三维环境感知、路径规划和精准避障, 特别适用于复杂未知地形和动态环境的低空飞行任务。其优势在于定位精度高(可达±1 cm), 视场范围大(水平 90°～360°, 垂直 30°～90°), 测量范围为 0.1～300 m, 数据更新率为 100～300 Hz, 可生成高密度三维点云, 适合复杂地形和动态场景, 同时具备较强的抗干扰能力以确保低空飞行的稳定性和可靠性; 但设备成本较高, 雨雾天气下性能下降, 且机载设备的质量和功耗往往限制了其使用范围。

3.3.4　低空智能网联体系监视技术

1) 能力需求

低空飞行目标根据能否接入低空网络化监管系统、接受通信和管控指令, 分为合作和非合作 2 种类别。针对合作目标的监视技术需重点解决低空各类航空器的协同感知、入网监测及身份识别等问题。非合作目标不主动向低空智能网联监视系

统报送信息，并且低空飞行器具有飞行高度低、反射截面小、飞行速度慢、周围环境复杂等特点，因此，单一的雷达或光电监视技术对非合作目标的探测较为困难。同时，由于无法接受监管系统的管控指令，非合作目标对空域内其他飞行器、地面的人身财产和关键基础设施等具有巨大的安全威胁。表9为低空智能网联体系7方面监视能力需求的含义。监视目标识别率、监视信息更新频率、监视告警时延、误警率、虚警率、漏警率、完整率等，共同构成了低空智能网联体系监视能力的核心需求。

表9 低空智能网联体系监视能力需求

Table 9 Surveillance capability requirements of the low-altitude intelligent networked system

能力需求	具体含义
监视识别率	系统发现并准确识别监视目标的概率
监视信息更新频率	系统在时间范围内更新监视信息的次数
监视告警时延	从事件发生到监视系统发出告警的时间间隔
误警率	系统发出错误告警的概率
虚警率	系统错误的报告监视目标的概率
漏警率	系统未能发出应有告警的概率
完整率	系统在未检验出错误情况下，能够在一个监视周期内完成监视任务的概率

2）可选技术方案

为满足低空智能网联体系下监视能力的需求，需要整合现有的监视技术，为低空智能网联体系的发展提出现实可行的技术方案。表10为现有监视技术手段及特点、需要解决的问题、关键指标参考值等内容。

① 合作目标监视。

a. 远程识别。远程识别是指通过无线通信技术（如WiFi、LTE、5G等），将飞行器的身份信息（如飞行器的唯一标识符、位置信息、速度、航向等）远程广播给地面接收设备或其他飞行器，从而实现对飞行器的远程识别与监视。其单个接收机最远监视范围5 km，最大支持同时监控100架无人机目标，实时性强，适合无人机运行密集的空域管理。然而，远程识别技术需要标准化协议支持，同时可能受到信号覆盖范围和多目标监测能力的限制，适合作为城市无人机监管的辅助技术。《民用无人驾驶航空器系统安全要求》[55]中规定轻型、小型无人机应通过远程识别报送信息。针对所有类型无人机的远程识别报送强制性标准正在制定中。

b. 广播式自动相关监视。通过飞行器广播其位置信息，无需主动请求，地面站和其他飞行器即可接收数据。ADS-B具有实时性强、部署成本低的特点，适合高密度低空空域的航迹管理和动态避撞，接收机空地最大作用距离大于100 km（无遮挡），空空最大作用距离大于200 km，分辨率能达到10 m量级。然而，其信号覆盖受限，可能受到建筑物遮挡和电磁干扰的影响。中国民用航空局发布的《民用微轻小型无人驾驶航空器系统运行识别概念（暂行）》[56]中规定所有民用轻小微无人机禁止在任何空域使用ADS-B信号发射机广播运行识别数据，但允许在无人驾驶航空器上加装ADS-B接收机来实现无人驾驶航空器对有人驾驶航空器的感知和识别。

表10 监视可选技术方案

Table 10 Optional surveillance technology solutions

目标类型	技术手段	需要解决问题/缺点	指标参考
合作目标	远程识别	强制要求轻小微无人机通过远程识别报送信息，最低性能要求（试行）已发布	探测距离1～2 km 频率2.4 GHz/5.8 GHz
	广播式自动相关监视	ADS-B设备可用于中大型无人机；解决1 090 MHz频段拥堵问题	探测距离≥200 km（有人机） 监测识别无人机≥100架 处理速度≥500报文/s
非合作目标	低空监视雷达	减少虚假目标和误警；提高对低反射目标（如微型无人机）的探测能力	探测范围达到20 km 探测高度1 000 m 跟踪目标数量≥500架 识别时间≤10 s
	5G-A通感一体	目前测试验证不充分；昂贵	面向低空场景单站覆盖距离≥1 200 m 感知高度≥300 m
	光电/红外探测	作用距离短，识别精度受限	跟踪距离≥2 km 跟踪误差≤0.02° 识别距离≥1 km
	频谱探测	高密度电磁环境中，容易受到干扰，信号分离困难；对静默无人机探测能力有限	探测距离≥10 km；覆盖频段0.02～8 GHz；测向精度≤3°；目标识别优于20 mm

② 非合作目标监视。

a. 低空监视雷达。低空监视雷达采用微波技术探测低空飞行器的位置和速度，其特点是探测精度高、全天候运行能力强，适合城市低空空域的动态监视，监视范围通常在3～5 km（消费级无人机）、30～50 km（通航飞机）。低空监视雷达存在一些局限性，包括对复杂背景的分辨能力较弱，容易受到地面杂波、植被反射等因素影响，产生虚假目标或误警。同时，在高密度飞行器运行环境中，可能存在目标分离困难的问题。此外，设备成本较高，部署需要预留较大的安装空间和电力资源。为了增强监视效果，低空监视雷达通常需要结合ADS-B等其他技术使用，以提高目标识别的可靠性和准确性。针对复杂应用环境，可通过分布式组网探测提升目标发现概率和定位精度。

b. 5G-A 通感一体。将 5G 通信和感知功能相结合, 利用 5G 网络实现飞行器的实时位置监测和数据传输。5G-A 通感一体技术在低空运营中展现了优越的感知性能, 其感知距离可达 1～20 km, 位置精度为 10～30 m, 速度分辨率为 5 m/s, 刷新率为 1 s, 检测率达 95%, 虚警率低至 5%。同时, 技术支持对 0.01～40 m^2 雷达散射截面积(radar cross section, RCS)目标的探测, 能够在 0～300 m 的低空范围内实现目标识别和航迹跟踪。但目前技术测试验证不充分; 加之 5G 基站部署昂贵, 只适合在城市重点区域部署对非合作目标进行监测。

c. 光电/红外探测。光电和红外均属于光学探测技术, 光学探测设备的成本较高, 处理大规模图像数据也需要较强的计算能力。光学探测非常适合城市核心区域的低空监视任务, 尤其是在需要精确目标识别的场景中。

光电探测利用高分辨率光学摄像头捕捉飞行器图像, 进行目标识别和定位, 其特点是分辨率高, 能够提供详细的目标特征信息, 适用于无人机行为分析和精细化监控任务。光学探测在光线条件良好的情况下探测效果最佳, 典型探测距离可达数百米至数公里, 具体取决于设备参数, 但其性能受天气条件(如雨雾)和环境光照的限制, 可能在低能见度环境中表现不佳。

而红外探测通过红外传感器捕捉飞行器发出的热辐射信号, 定位并监测目标, 能够在低光照或无光环境下进行有效探测, 适合夜间或恶劣天气条件下的飞行器监控。典型红外探测距离为数百米至数公里, 具体视传感器性能而定。红外探测对目标的热辐射特征较为敏感, 能够在复杂背景中分辨飞行器, 但在极端高温或热源干扰严重的环境中, 可能会出现误报。红外传感器的精度和分辨率通常低于光学摄像头, 无法提供精细的目标特征信息。红外探测非常适合用于夜间监控、低光照区域及应急响应中的飞行器监视任务。

d. 频谱探测。频谱探测基于到达时间差(time difference of arrival, TDOA)和到达角度(angle of arrival, AOA)等技术实现对低空飞行器的定位, 探测的频谱能覆盖 0.07～6 GHz 的范围, 包括了目前常见无人机 C2 链路的 2.4 GHz 和 5.8 GHz 频段, 测向精度能达到 3°以内, 是非合作目标监视的有效手段之一, 其特点是隐蔽性强、无源被动探测、绿色环保。但在复杂的电磁环境下, 如城市建筑物反射和各类电磁干扰, 会影响探测准确性, 需要进一步开展研究提高技术能力。

4　低空智能网联体系未来展望

4.1　运行概念与体系框架有待完善

低空智能网联体系的发展是一个逐步推进、持续完善的过程。当前阶段, 低空运行场景和运行概念还未明确, 空域密度和运行风险相对较低。因此, 现阶段飞行器的设计制造主要用于满足单机独立运行的目标, 具备自主导航和简单的任务执行能力, 智能化水平较低, 机载设备以满足基础通信导航监视能力为主, 制造重点在于提高单机可靠性和经济性; 空域仍以隔离运行为主, 通过廊道式的运行模式确保业务的正常开展; 服务能力和监管自动化水平较低, 仅能提供飞行任务的简单支持, 运行数据的实时性和全面性较低, 尚未实现动态信息共享和实时协同管理, 仍需要通过人在回路中的方式进行飞行计划的制定审批、空域态势的监管及风险的安全管理等。

随着相关技术的成熟和应用场景、应用边界的逐渐明确, 低空运行将逐步向中等密度、可控风险的方向发展。飞行器制造开始向网联化和业务驱动智能化方向发展, 网联化单体智能飞行器逐渐普及, 机载设备将集成更多高精度、智能化设备, 以支持多机协同和复杂任务执行, 同时基础设施建设也同步升级, 为飞行器提供动态感知能力, 确保安全高效的运行; 空域管理由静态隔离运行模式向动态网格化运行模式转变, 逐渐具备多运营人共用航路及动态调整飞行计划的能力。依托智能航电系统和实时数据传输实现飞行器之间的协同感知和动态避障; 基于低空智能网联体系的服务支撑平台和应用系统逐步投入使用, 服务和监管的自动化水平得到较大提升, 逐步实现人在回路上的空域容流管理、飞行计划调整、冲突风险告警解脱等能力。

随着低空智能网联体系的进一步扩展和完善, 低空运行将全面进入高密度、大规模、复杂异构的运行阶段。在这一阶段, 飞行器以高速可靠的网络协同为核心, 智能化由单机智能向体系智能转变, 形成高度模块化、低成本制造体系; 空域管理从动态网格化运行进一步演进至灵活低空自由运行, 飞行器能够在无固定航线约束的情况下自主完成复杂任务; 低空监管和服务向高度智能化和无人化的方向发展, 基于 AI 的监管服务平台能够实时监控飞行器运行状态并进行智能化安全管理和调度, 实现人在回路外的低空运行精准管理与高效调控。

4.2　关键技术尚未成熟, 需要持续突破

低空智能网联体系的建设与完善有赖于相关

技术的支持。目前，许多关键技术（如通感一体、卫星互联网、复杂环境精准导航、低空全面可信监视等）尚未成熟，仍需要持续的创新和突破。以5G-A通感一体技术和机载空空监视雷达为例，现有的5G-A技术相比传统5G在通信速率，时延、感知能力精度及对低空的适应性改造上具有一定的优势，但由于其技术特点，发射信号强度一般远大于接收信号，从而会对接收链路造成严重的同频干扰，并且天线间的自干扰信号等与探测目标的回波信号相互混合，会产生海量背景噪声，降低信噪比，增加对目标探测和感知的难度。此外，低空场景下对定位精度的需求至少要达到米级或者亚米级，这要求通感一体化基站之间的同步误差控制在纳秒级别，但当前5G通信系统基站之间同步误差还处在毫秒级，尚无法满足广泛的低空场景应用需求。表11为5G与5G-A技术发展指标典型值对比。

表11 5G-A技术指标发展

Table 11 Development of 5G-A technical indicators

性能参数	5G-A级别	5G级别
下行速率	10 Gbit/s	1 Gbit/s
上行速率	1 Gbit/s	200 Mbit/s
时延	5～10 ms	20 ms
位置精度	厘米级	米级
感知能力	距离、速度	
现存问题	存在同频干扰，时间同步误差不能满足精度要求，网络容量增长限制	无法针对低空场景提供网络支持

受限于当前机载平台的质量和功耗限制，全球仅有少数几家公司（如美国通用原子公司、霍尼韦尔公司）的ATAR设备能满足RTCA DO-366B[57]标准，中国目前还未开发出满足相应标准的产品。其中霍尼韦尔公司的IntuVue RDR-84K雷达，其体积226 mm×125 mm×43 mm，质量仅为0.7 kg，视场达到100°（水平）/30°（垂直），探测距离能达到3 km，但目前对中国禁运，并且功耗高达60 W，不适用于有功耗限制的低空飞行器。而目前车规级毫米波雷达，虽然功耗能满足部分应用需求，但探测距离一般在百米级别，无法满足低空飞行器的安全监视需求[58]。

因此，CNSI技术需要进一步向高精度、低成本方向发展，以适应复杂场景的需求；人工智能技术将在飞行优化、运行调度、安全监测等领域为系统智能决策深度赋能；信息安全技术的升级将提升网络数据安全性和信任管理水平；能源技术的突破将有助于提升飞行器运营效率和经济效益。技术水平的不断升级和成熟，将使低空智能网联体系具备更强的适应性，为未来低空运行、监管和服务保障提供牢固基础和有力支撑，为低空智能网联体系的稳健发展和多元化应用奠定坚实的基础。

4.3 技术方案应当与运行需求相匹配

随着低空飞行量的不断增加，多样化的应用场景与运行需求将不断涌现，这既是推动技术创新的动力，也是对体系适应性提出的新考验。未来，低空运行需要在通信、导航、监视及管控服务等多个环节协同进化，以满足快速增长的运行需求。通过反复试点与应用反馈，将逐步明确最优的技术组合方案、基础设施布局及资源调度方式。此外，数据与服务支撑网络的兼容性和扩展性将持续提升，从而为多源数据融合、异构信息处理及多样化服务能力的迭代发展提供有力支持。通过这一系列的动态调整与优化，低空智能网联体系将逐步具备从小规模试点走向大规模常态化运行的能力，并在未来城市的空中出行与产业生态构建中发挥关键作用。

5 结束语

本文系统分析了低空智能网联体系的国内外发展现状，重点探讨了国外低空建设的发展思路及中国在该领域的相关框架和技术方案。通过对“运行概念-能力需求-技术实现”这一科学方法的深入解读，本文提出了低空智能网联体系的实践路径和发展方向。本文进一步剖析了《低空智能网联体系参考架构（2024版）》中体系架构的内在逻辑，详细阐述了“五方三层”体系的核心内涵、业务需求与关键能力，结合现阶段技术瓶颈，进行了关键技术的全面论证与分析，为低空智能网联体系的理论发展与技术创新提供了参考依据。本文总结了低空智能网联体系发展过程中面临的主要问题，并结合现有研究成果提出了未来发展过程中应遵循的基本原则及优化建议，以期为低空智能网联体系的持续发展提供理论支持与实践指导。

参考文献（References）

[1] 低空产业联盟. 低空智能网联体系参考架构 (2024 版)[R/OL]. (2024-11-11)[2025-01-23]. https://www.miit-eidc.org.cn/art/2024/11/11/art_1644_11070.html.
Low-Altitude Industry Alliance. The framwork of low-altitude intelligent network system(version 2024)[R/OL]. (2024-11-11)[2025-01-23]. https://www.miit-eidc.org.cn/art/2024/11/11/art_1644_11070.html(in Chinese).

[2] U. S. Congress. Faa Modernization and reform act of 2012[Z/OL]//112TH Congress 2d Session. House of Representatives, Report112-381. (2012-05-10) [2025-01-23]. https://www.congress.gov/bill/112th-congress/house-bill/658/text.

[3] Federal Aviation Administration. Integration of civil unmanned air-

craft systems (UAS) in the national airspace system (NAS) roadmap (First Edition-2013)[R/OL]. (2013-11-07)[2025-01-23]. https://www.faa.gov/sites/faa.gov/files/uas/resources/policy_library/uas_roadmap_2013.pdf.

[4] Federal Aviation Administration, National Aeronautics and Space Administration. UAS traffic management research plan (Version1.0) [Z/OL]. (2017-01-31)[2025-01-23]. https://www.faa.gov/sites/faa.gov/files/2022-08/FAA_NASA_UAS_Traffic_Management_Research_Plan.pdf.

[5] U. S. Congress. Public law 114–190—FAA extension, safety, and security act of 2016[Z/OL]. (2016-06-15) [2025-01-23]. https://www.congress.gov/114/plaws/publ190/PLAW-114publ190.pdf.

[6] Federal Aviation Administration. 14 CFR Part 107: Small unmanned aircraft systems [Z/OL]. [2025-01-23]. https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-107.

[7] Federal Aviation Administration. Concept of operations v1.0: unmanned aircraft system (UAS) traffic management (UTM) [R/OL]. (2018-05-18)[2025-01-23]. https://www.nasa.gov/wp-content/uploads/2024/04/2018-utm-conops-v1-0-508.pdf?emrc=94dcbb.

[8] Federal Aviation Administration. Concept of operations v2.0: unmanned aircraft system (UAS) traffic management (UTM) [R/OL]. (2020-03-02)[2025-01-23]. https://www.faa.gov/sites/faa.gov/files/2022-08/UTM_ConOps_v2.pdf.

[9] GERMAN B J, DASKILEWICZ M J, TRANI A, et al. Aviation global demand forecast: model development and ISAAC studies-Task 3.7: concept of operations for ODM VTOL aircraft package delivery[J/OL]. [2025-01-23]. https://ntrs.nasa.gov/api/citations/20205005895/downloads/GeorgiaTech_eVTOL_Cargo_Final_Report_July2018.pdf.

[10] RIMJHA M, ADE M, TARAFDAR S, et al. Aviation global demand forecast model development and ISAAC studies: UAS-VTOL Cargo study[J/OL]. [2025-01-23]. https://ntrs.nasa.gov/api/citations/20205005891/downloads/VirginiaTech_eVTOL_Cargo_Final_Report_July_2018.pdf.

[11] GOYAL R, REICHE C, FERNANDO C, et al. Urban air mobility market study[R/OL]. [2025-01-23]. https://ntrs.nasa.gov/api/citations/20190001472/downloads/20190001472.pdf.

[12] PATTERSON M D, ANTCLIFF K R, KOHLMAN L W. A proposed approach to studying urban air mobility missions including an initial exploration of mission requirements[C/OL]//Proceedings of the Annual Forum and Technology Display. [2025-01-23]. https://ntrs.nasa.gov/api/citations/20190000991/downloads/20190000991.pdf.

[13] KOHLMAN L W, PATTERSON M D, RAABE B E. Urban air mobility network and vehicle type-modeling and assessment[R/OL]. [2025-01-23]. https://ntrs.nasa.gov/api/citations/20190001282/downloads/20190001282.pdf.

[14] PATTERSON M D, ISAACSON D R, Mendonca N L, et al. An initial concept for intermediate-state, passenger-carrying urban air mobility operations[C/OL]//Proceedings of the AIAA Scitech 2021 Forum. [2025-01-23]. https://ntrs.nasa.gov/api/citations/20205010104/downloads/UAM_ConOps_SciTech2021_STRIVESsubmit.pdf.

[15] Hill B P, DeCarme D, Metcalfe M, et al.UAM vision concept of operations (ConOps) UAM Maturity Level (UML) 4 [R/OL]. (2020-12-2)[2025-01-23]. https://ntrs.nasa.gov/citations/20205011091.

[16] Hill B, DeCarme D.Urban Air Mobility (UAM) vision concept of operations (ConOps) UAM maturity level (UML)-4 overview [R/OL]. (2020-12-15)[2025-01-23]. https://ntrs.nasa.gov/citations/20210010443.

[17] RICE G, HELTON D, JENKINS K, et al. Urban air mobility operational concept (opscon) passenger-carrying operations[R/OL]. (2020-05-01)[2025-01-23]. https://ntrs.nasa.gov/citations/20205001587.

[18] STOUFFER V L, COTTON W B, DeAngelis R A, et al. Reliable, secure, and scalable communications, navigation, and surveillance (CNS) options for urban air mobility (UAM)[R/OL]. (2020-08-21)[2025-01-23]. https://ntrs.nasa.gov/citations/20205006661.

[19] LEVITT I, PHOJANAMONGKOLKIJ N, HORN A, et al. UAM airspace research roadmap-Rev. 2.0[R/OL]. [2025-01-23]. https://ntrs.nasa.gov/api/citations/20230002647/downloads/NASA-TM-20230002647_Final.pdf.

[20] Federal Aviation Administration. Advanced air mobility (AAM) implementation plan[R/OL]. [2025-01-23]. https://www.faa.gov/sites/faa.gov/files/AAM-I28-Implementation-Plan.pdf.

[21] Federal Aviation Administration. Concept of operations v2.0: urban air mobility(UAM)[R/OL]. (2023-08-31)[2025-01-23]. https://www.faa.gov/air-taxis/uam_blueprint.

[22] SESAR Joint Undertaking. European ATM master plan: Edition 1 [R/OL]. (2009-03-30)[2025-01-23]. https://www.sesarju.eu/sites/default/files/European_ATM_Master_Plan.pdf.

[23] European Commission. A new era for aviation-opening the aviation market to the civil use of remotely piloted aircraft systems in a safe and sustainable manner[R/OL]. [2025-01-23]. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014AE3189.

[24] European Commission. An aviation strategy for Europe[R/OL]. [2025-01-23]. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52015DC0598.

[25] SESAR Joint Undertaking. European drones outlook study: unlocking the value for Europe[R/OL]. [2025-01-23]. https://www.sesarju.eu/sites/default/files/documents/reports/European_Drones_Outlook_Study_2016.pdf.

[26] European Parliament and Council. Regulation (EU) 2018/1139 of the European Parliament and of the Council of 4 July 2018 on common rules in the field of civil aviation and establishing a European Union Aviation Safety Agency[R/OL]. [2025-01-23]. https://eur-lex.europa.eu/eli/reg/2018/1139/oj/eng.

[27] European Commission. Commission implementing regulation (EU) 2019/947 of 24 May 2019 on the rules and procedures for the operation of unmanned aircraft[R/OL]. [2025-01-23]. https://eur-lex.europa.eu/eli/reg_impl/2019/947/oj/eng.

[28] European Commission. Commission delegated regulation (EU) 2019/945 of 12 March 2019 on unmanned aircraft systems and on third-country operators of unmanned aircraft systems[R/OL]. [2025-01-23]. https://eur-lex.europa.eu/eli/reg_del/2019/945/oj/eng.

[29] SESAR Joint Undertaking. U-space: blueprint[R/OL]. (2017-06-09) [2025-01-23]. https://www.sesarju.eu/u-space-blueprint.

[30] BARRADO C, BOYERO M, BRUCCULERI L, et al. U-space concept of operations: a key enabler for opening airspace to emerging low-altitude operations[J]. Aerospace, 2020, 7(3): 24.

[31] SCOTT B I, ANDRITSOS K I. A drone strategy 2.0 for a smart and sustainable unmanned aircraft eco-system in Europe[J]. Air and

Space Law, 2023, 48(3): 273-296.

[32] SESAR Joint Undertaking. Fourth edition of the U-space concept of operations (ConOps)[R/OL]. [2025-01-23]. https://www.sesarju.eu/node/4544.

[33] SESAR Joint Undertaking. 2025 edition of the European ATM master plan[R/OL]. [2025-01-23]. https://www.sesarju.eu/MasterPlan2025.

[34] 国务院,中央军委. 关于深化我国低空空域管理改革的意见[N]. 光明日报,2010-11-15(2).
State Council of the People's Republic of China, Central Military Commission of the Communist Party of China. Opinions on deepening the reform of low-altitude airspace management in China[N]. Guangming Daily, 2010-11-15(2)(in Chinese).

[35] 中央经济工作会议.2023年12月中央经济工作会议纪要[C/OL]. [2025-01-23]. https://www.gov.cn/yaowen/liebiao/202312/content_6919834.htm?fromModule=lemma_middle-info.
Central Economic Work Conference. 2023 December Central Economic Work Conference Summary [C/OL]. [2025-01-23] https://www.gov.cn/yaowen/liebiao/202312/content_6919834.htm?fromModule=lemma_middle-info(in Chinese).

[36] 全国人民代表大会和中国人民政治协商会.2024年政府工作报告[R/OL]. [2025-01-23]. https://www.gov.cn/yaowen/liebiao/202403/content_6939153.htm.
The National People's Congress and Chinese People's Political Consultative Conference. Government work report[R/OL]. [2025-01-23]. https://www.gov.cn/yaowen/liebiao/202403/content_6939153.htm(in Chinese).

[37] 中国电子科技集团有限公司. 低空航行系统: 拥抱低空经济, 安全智慧出行[R/OL]. [2025-01-23]. https://cetcam.cetc.com.cn/zgdk/1593022/1592495/1816065/20240906084646312229.pdf.
China Electronics Technology Group Corporation. Low-altitude navigation system: embracing low-altitude economy, safe and smart travel [R/OL]. [2025-01-23]. https://cetcam.cetc.com.cn/zgdk/1593022/1592495/1816065/20240906084646312229.pdf(in Chinese).

[38] 中国移动通信集团有限公司. 低空智联网技术体系白皮书[R/OL].[2025-01-23].https://www.dtinsight.com.cn/nd.jsp?id=2826.
China Mobile Communications Group Co. , Ltd. Low-altitude intelligent networked technology system white paper[R/OL]. [2025-01-23]. https://www.dtinsight.com.cn/nd.jsp?id=2826 (in Chinese).

[39] 广东省通信学会, 中国信息通信研究院, 中国联合网络通信有限公司广东省分公司, 等. 低空智联网发展研究报告 (2024年) [R/OL]. [2025-01-23]. https://hanglian.feishu.cn/file/BWBgbyPLxo39sqxmBktcGzX8nXe.
Guangdong Institue of Communications, China Academy of Information and Communications Technology, China United Network Communications Limited Guangdong Branch, et al. Research report on the development of low-altitude intelligent networked [R/OL]. [2025-01-23]. https://hanglian.feishu.cn/file/BWBgbyPLxo39sqxmBktcGzX8nXe(in Chinese).

[40] 中国电信集团有限公司, 爱立信, 诺基亚, 等. 通感一体低空网络白皮书[R/OL]. [2025-01-23]. http://claei.com.cn/files/file/20241119/1731987437969932.pdf
China Telecom Corp Ltd., Ericsson, Nokia, et al. Integrated sensing and communication low-altitude network white paper [R/OL]. [2025-01-23]. http://claei.com.cn/files/file/20241119/1731987437969932.pdf(in Chinese).

[41] 粤港澳大湾区数字经济研究院. 低空经济发展白皮书: 深圳方案[R/OL]. [2025-01-23]. http://www.cuaer.com/plus/view.php?aid=11756
International Digital Economy Academy. Low-altitude economy development white paper: Shenzhen solutions[R/OL]. [2025-01-23]. http://www.cuaer.com/plus/view.php?aid=11756(in Chinese).

[42] 粤港澳大湾区数字经济研究院. 低空经济发展白皮书: 全数字化方案[R/OL]. [2025-01-23]. https://www.hulianhutongshequ.cn/upload/tank/report/2024/202404/1/b47f07c17faa4b7ab3fe915f22b11a18.pdf.
International Digital Economy Academy. Low-altitude economy development white paper: all digital solutions[R/OL]. [2025-01-23]. https://www.hulianhutongshequ.cn/upload/tank/report/2024/202404/1/b47f07c17faa4b7ab3fe915f22b11a18.pdf (in Chinese).

[43] PUENTE-CASTRO A, RIVERO D, PAZOS A, et al. A review of artificial intelligence applied to path planning in UAV swarms[J]. Neural Computing and Applications, 2022, 34(1): 153-170.

[44] SHRESTHA R, OH I, KIM S. A survey on operation concept, advancements, and challenging issues of urban air traffic management[J]. Frontiers in Future Transportation, 2021, 2: 626935.

[45] Yan C, Xiang X, Wang C. Towards real-time path planning through deep reinforcement learning for a UAV in dynamic environments[J]. Journal of Intelligent & Robotic Systems, 2020, 98: 297-309.

[46] 陈志杰, 朱永文, 刘杨. 基于数字化空域系统的城市无人机管理对策研究[J]. 中国民航大学学报, 2023, 41(3): 8-12.
CHEN Z J, ZHU Y W, LIU Y. Research on urban drone management strategies based on digital airspace system[J]. Journal of Civil Aviation University of China, 2023, 41(3): 8-12(in Chinese).

[47] 陈志杰, 汤锦辉, 王冲, 等. 人工智能赋能空域系统, 提升空域分层治理能力[J]. 航空学报, 2021, 42(4): 525018.
CHEN Z J, TANG J H, WANG C, et al. Artificial intelligence empowering airspace systems to enhance layered airspace governance capabilities[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(4): 525018(in Chinese).

[48] 刘泉, 陈瑶瑶, 洪晓苇, 等. 面向无人机的城市低空空域规划的国际经验[J]. 城市规划学刊, 2024(5): 64-70.
LIU Q, CHEN Y Y, HONG X W, et al. International experience in urban low-altitude airspace planning for drones[J]. Urban Planning Forum, 2024(5): 64-70(in Chinese).

[49] International Civil Aviation Organization. Manual on the system-wide information management (SWIM) concept[R]. Montreal: ICAO, 2024.

[50] International Civil Aviation Organization. Procedures for air navigation services-information management[R]. Montreal: ICAO, 2024.

[51] 张军. 现代空中交通管理[M]. 2版. 北京: 北京航空航天大学出版社, 2024.
ZHANG J. Modern air traffic management[M]. 2nd ed. Beijing: Beihang University Press, 2024(in Chinese).

[52] 任冰洁. SWIM数据模型的设计与实现[D]. 天津: 中国民航大学, 2022.
REN B J. Design and implementation of the SWIM data model[D]. Tianjin: Civil Aviation University of China, 2022(in Chinese).

[53] FREEMAN K, SHETYE S, WONG C, et al. NASA data and reas-

oning fabric (DRF)[R/OL]. [2025-01-23]. https://ntrs.nasa.gov/api/citations/20220010954/downloads/DRF_External_Outreach_presentation_19.07.22%20%20-%20%20Repairedr.pdf.

[54] 国家市场监督管理总局, 国家标准化管理委员会.民用无人驾驶航空器系统运行识别规范(征求意见稿)[S/OL]. (2025-01-09)[2025-01-23]. https://www.caac.gov.cn/HDJL/YJZJ/202502/t20250212_226676.html.
State Administration for Market Regulation, National Standardization Administration.Specification for civil unmanned aircraft system operational identification[S/OL]. (2025-01-09)[2025-01-23]. https://www.caac.gov.cn/HDJL/YJZJ/202502/P020250212316294948285.pdf (in Chinese).

[55] 中国民用航空局空管行业管理办公室.民用微轻小型无人驾驶航空器系统运行识别概念 (暂行)[S/OL]. (2022-03-11)[2025-01-23]. https://www.caac.gov.cn/XXGK/XXGK/GFXWJ/202203/t20220311_212290.html.
Civil Aviation Administration of China. Concept of operation identification for civil micro, light, and small unmanned aircraft systems (Interim)[S/OL]. (2022-03-11)[2025-01-23]. https://www.caac.gov.cn/XXGK/XXGK/GFXWJ/202203/t20220311_212290.html(in Chinese).

[56] 国家市场监督管理总局,国家标准化管理委员会. 民用无人驾驶航空器系统安全要求[S]. (2023-05-23). https://std.samr.gov.cn/gb/search/gbDetailed?id=FC83293D549DB452E05397BE0A0A9309.
State Administration for Market Regulation, National Standardization Administration.Safety requirements for civil unmanned aircraft system[S]. (2023-05-23). https://std.samr.gov.cn/gb/search/gbDetailed?id=FC83293D549DB452E05397BE0A0A9309 (in Chinese).

[57] Radio Technical Commission for Aeronautics Special Committee-228. Minimum operational performance standards (MOPS) for air-to-air radar for traffic surveillance[S/OL]. (2020-09-10)[2025-01-23]. https://products.rtca.org/21djqiu/.

[58] 汤新民, 顾俊伟, 刘冰, 等. 低空监视技术及其发展趋势综述[J]. 南京航空航天大学学报, 2024, 56(6): 973-993.
TANG X M, GU J W, LIU B, et al. Overview of low-altitude surveillance technology and its development trends[J]. Journal of Nanjing University of Aeronautics and Astronautics, 2024, 56(6): 973-993(in Chinese).

Overview of low-altitude intelligent networked system

ZHANG Xuejun[1, 2, *], LIU Fawang[3], ZHANG Zuyao[1, 4], TIAN Ye[3]

(1. School of Electronic Information Engineering, Beihang University, Beijing 100191, China;
2. State Key Laboratory of CNS/ATM, Beihang University, Beijing 100191, China;
3. Ministry of Industry and Information Technology Equipment Industry Development Center, Beijing 100846, China;
4. Beijing Key Laboratory for Network-Based Cooperative Air Traffic Management, Beihang University, Beijing 100191, China)

Abstract: Recently, the Low-Altitude Industry Alliance released the *Reference Architecture of the Low-Altitude Intelligent Networked System (2024 Edition)* report, which outlines the basic content of the developmental evolution stages, components, and system framework of the low-altitude intelligent networked system. This document provides a reliable reference and foundation for the development of the low-altitude intelligent networked system. However, as a framework-based report, the report focuses on presenting the key components of the low-altitude intelligent networked system in the most concise and precise manner, lacking detailed descriptions of the underlying scientific methods, theoretical foundations, and implementation approaches. This paper comprehensively elaborates on the current state of development, design concepts, system logic, and key technologies of the low-altitude intelligent networked system based on the report. It aims to further analyze and interpret the content of the report, providing a scientific theoretical reference for the subsequent development and construction of the low-altitude intelligent networked system.

Keywords: concept of operation for low-altitude intelligent interconnected system; design of low-altitude intelligent interconnected system architecture; low-altitude intelligent interconnected application service system; low-altitude intelligent interconnected data and service support network; low-altitude intelligent interconnected airborne terminals and infrastructure; key technologies for low-altitude intelligent interconnected system

Received: 2025-01-23; **Accepted**: 2025-03-18; **Published Online**: 2025-04-11 10:47
URL: link.cnki.net/urlid/11.2625.V.20250411.0921.002

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0397

飞机可动翼面缝隙封严结构研究进展

程小全[1,*]，蔡墨泉[1]，王松伟[2]

(1. 北京航空航天大学 航空科学与工程学院，北京 100191；2. 空装驻石家庄地区军事代表室，石家庄 050081)

摘　　要：主翼面与可动翼面之间的缝隙会对飞机飞行性能造成一定影响，缝隙封严结构的应用会改善机翼的表面光滑性，同时达到增升减阻、优化操纵等效果。基于此，对5种封严结构设计形式进行对比，并将封严结构的设计分为依靠机构和依靠材料2类，分析了其优缺点。比较了4家航空配件公司的封严结构产品，分析了耐磨自由端细节设计。给出了快速拆装修理、限制结构刚度范围、设计时仿真与试验相结合等设计建议。

关 键 词：飞机封严结构；缝隙密封；飞行性能提升；弹性大变形；疲劳性能

中图分类号：V224+.4

文献标志码：A　　**文章编号**：1001-5965(2025)06-1816-08

气动性能和飞行控制性能是飞机设计关注的重点。机翼前后缘的襟翼、缝翼等增升装置，可提高起飞重量，降低巡航油耗，缩短起降距离，增强该机型的市场竞争力[1]。军机中，机动性能高的战斗机在空战中能够占据主动，改进方向舵、升降舵等控制面的设计，提高战斗机的机动性能，F-22和J-20等先进战机已采用了全动V尾设计[2]。

增升装置和控制面都属于可动翼面，在操纵偏转的过程中，其与机身、机翼等固定部件之间会产生缝隙。上下翼面压强差会使气流穿过缝隙到达另一侧表面，导致气流分离，提前由层流转捩为湍流。气流转捩会增加飞行阻力，增大油耗；严重时会降低机翼升力系数，造成失速，威胁机组人员的生命财产安全[3]。此外，水分、杂质等进入缝隙，会对机翼内的连接、传动结构造成腐蚀[4]；缝隙也会削减飞机表面的光滑性，降低整机隐身性能[5]。

因此，对于机翼可动翼面与主翼面之间的缝隙，需要使用封严结构来进行气动密封，以改善机翼表面的连续性和光滑性。由于翼面存在偏转和弯曲，这种缝隙密封属于动态密封，其技术难度较大[6]。航模爱好者将弹性胶带粘贴在缝隙处，以提高航模的操纵性能，胶带起到气动密封作用，在原理上与封严结构相似。

国内外对飞机翼面封严结构相关研究成果的公开报道较少，波音、空客等公司官网上可查到的具体应用也不多。本文对国内外已发表的相关专利和应用情况进行总结，分析这些封严结构的设计特点及可能存在的问题，归纳出封严结构设计时的注意事项，旨在为设计人员提供一定的思路和参考依据。

1 封严结构的应用情况

1.1 封严结构工作原理

如图1所示[1]，可动翼面与主翼面之间存在缝隙，易导致上翼面气流分离，需要加装封严结构进行气流密封。

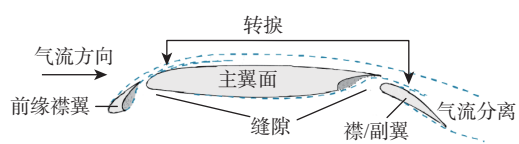


图1 机翼可动翼面与主翼面间的缝隙[1]

Fig. 1 Gap between movable wing surface and main wing surface[1]

收稿日期：2023-06-19；录用日期：2023-09-08；网络出版时间：2023-09-15 16:24
网络出版地址：link.cnki.net/urlid/11.2625.V.20230915.0957.001

*通信作者. E-mail：xiaoquan_cheng@buaa.edu.cn

引用格式：程小全，蔡墨泉，王松伟. 飞机可动翼面缝隙封严结构研究进展[J]. 北京航空航天大学学报，2025，51(6)：1816-1823.
CHENG X Q，CAI M Q，WANG S W. Study progress of gap sealing structure for aircraft movable wing[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1816-1823 (in Chinese).

传统翼面封严结构(见图 2[7])是一块随动板, 一端连接到主翼面上, 另一端搭接在可动翼面上的自由端, 自由端与可动翼面表面相对滑动。随着可动翼面的偏转, 封严结构会转动或弯曲变形, 在飞行过程中实现缝隙密封[7]。

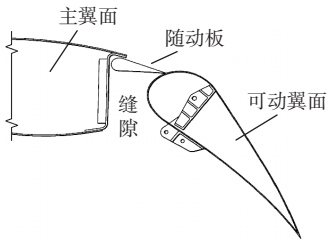


图 2　传统翼面封严结构[7]

Fig. 2　Traditional wing sealing structure[7]

1937 年, NACA 在其技术笔记中提到使用封严结构来提高飞机的操纵性能。如图 3 所示, 美国海军舰载机 F-18 使用了封严结构来密封副翼与主翼面之间的缝隙, 而中国舰载机 J-15 则未进行密封, 这对 J-15 在航母上的起降是极为不利的。

经过长期试验研究, 关于封严结构对飞行性能产生的有利影响, 技术人员总结出了一些规律: 封严结构对飞机巡航速度及爬升率的提升有限, 至多在原有基础上增加约 2%, 起降时的失速速度会降低 2～3 km/h; 封严结构对飞机的滚转特性有很大改善, 可以提高滚转速率, 同时优化飞机驾驶员的操纵体验。



(a) F-18



(b) J-15

图 3　舰载机 F-18 与 J-15 对比

Fig. 3　Comparison between shipboard aircraft F-18 and J-15

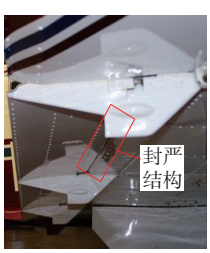
1.2　封严结构在飞机上的应用情况

产品化封严结构的应用, 目前只出现在小型飞机上。美国的 Lasar Aero Styling & Repair 公司、Whelen 航空航天技术公司(WAT)、Knots 2U 公司、Horton 工业公司为不同型号的通用飞机提供封严结构产品。这 4 家公司的相关产品信息如表 1 所示。Lasar Aero Styling & Repair 公司的产品如图 4 所示。

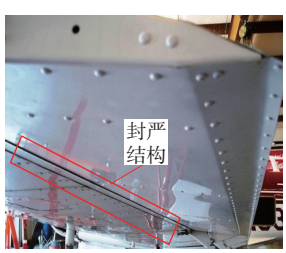
表 1　4 家公司封严结构产品信息对比

Table 1　Comparison of product information for four companies with strict sealing structures

公司名称	适用机型	性能提升	选材和安装	产品价格/美元
Lasar Aero Styling & Repair	Mooney M20系列	巡航速度增加2～4 km/h, 爬升速度显著增加, 操纵性能改善	铝制板材, 抽芯铆钉安装, 襟翼部位安装需12 h, 副翼部位为4 h	180～325
Whelen	Comanche PA-24, PA-28, PA-32, PA-34, PA-39, PA-44系列	巡航速度提升4～5 km/h , 着陆速度更低	高性能环氧树脂层合板, 3～4 h安装时间	999～1 799
Knots 2U	Piper、Beech、Cessna等系列	将Piper Twin Comanche型号的巡航速度提升32 km/h, 其余型号提升6～8 km/h	铝制板材、超高分子量塑料防摩擦条, 用螺钉安装于可动翼面内部, 安装时间4 h	250～900
Horton	Cessna 182 Skylane	巡航速度提升3 km/h, 着陆时的滚转速率和低速状态下的操纵性能明显改善	铝制板材或玻璃纤维层合板, 带有硅树脂黏合剂, 使用空心铆钉固定, 安装时间为8～12 h	350



(a) 襟翼封严结构
(适用于M20B~M20G)

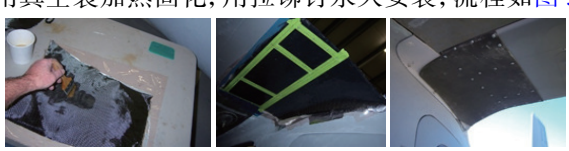


(b) 副翼下翼面封严结构
(适用于M20B~M20K)

图 4　Lasar Aero Styling & Repair 公司的封严结构产品安装示意图

Fig. 4　Installation diagram of Lasar Aero Styling & Repair Company's sealing structure products

飞行爱好者 Setzer 将 6 层碳纤维织物和 2 层玻璃纤维织物剪裁成所需封严结构的形状, 刷涂树脂, 用自攻螺钉将这 8 层铺层固定在机翼根部, 再用真空袋加热固化, 用拉铆钉永久安装, 流程如图 5



(a) 手工刷涂树脂　(b)自攻螺钉固定, 加热固化　(c) 拉铆钉永久固定

图 5　自制层合板封严结构

Fig. 5　Self-made laminated plate sealing structure

所示。该封严结构用于密封机翼根部和机身之间的缝隙,可以使起降失速速度降低 5 km/h。

2 翼面封严结构形式现状

缝隙内气动力较大时,传统翼面封严结构的随动板易被气流冲开(见图 6[7]),密封效果不好。为满足现代飞机的高速、大机动等性能需求,需设计新的结构形式。目前,封严结构设计还没有形成相应的规范,现有的每种设计都有其优缺点。本文对已有的结构设计进行总结分析,发现主要有依靠机构和依靠材料 2 类设计形式。

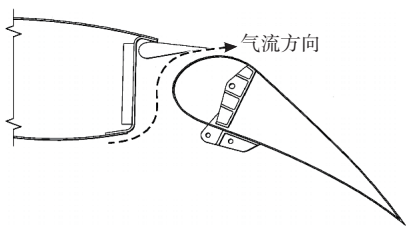


图 6 传统翼面封严结构[7]

Fig. 6 Traditional wing sealing structures[7]

2.1 依靠机构的设计形式

2.1.1 铰链连杆传动式封严结构

潘立新等[8]设计了一种使用铰链安装、连杆传动的封严结构。如图 7(a) 所示,使用特殊的 L 型连杆(见图 7(b)),将封严结构面板(简称封严板)内壁与飞机主翼面后缘上的铰链支座连接。L 型连杆的中空设计可使舵面操纵机构从其内部通过,各部件转动时不会互相干扰。舵面偏转时,L 型连杆能给封严板提供指向翼面内部的拉力,确保自由端能紧密贴合在舵面前缘蒙皮上,实现缝隙封严。

铰链连杆传动式封严结构无需额外驱动,仅靠舵面操纵机构带动 L 型连杆就可实现翼面封严,且适用于复杂翼型。舵面偏转不会造成封严板大幅度弯曲变形,降低了对封严板材料疲劳、弹性大变形等性能的要求。此外,该结构还可以采用热防护材料来对空天往返飞行器进行隔热密封。

铰链连杆传动式封严结构传动零件数量较多,用于大展长机翼时,需分块设计不同翼段的封严结构,会大幅度增加机翼后缘的结构质量,这与飞机设计的减重要求相悖,且大量零件给狭窄缝隙内的空间布局设计带来了挑战。

2.1.2 扭簧预加载式封严结构

胡利等[9]为某型先进战斗机设计了一种用于前缘襟翼的封严结构,如图 8(a) 所示。该结构分为扭簧、转轴、固定板组件、转动板组件,安装时需先将 4 个部分组装完成(见图 8(b)),再将固定板固支连接在前缘襟翼内壁,转动板旋转一定角度后压在

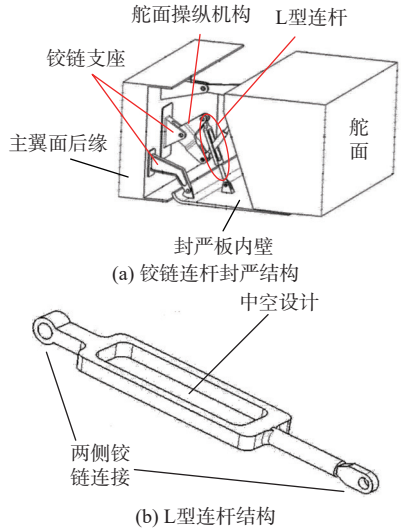


图 7 飞行舵面前缘随动封严结构[8]

Fig. 7 A kind of Follow-up sealing structure for leading edge of aircraft rudder surface[8]

主翼面伸出的弧形导流片上(见图 8(c))。扭簧扭矩随着襟翼偏转角度的增大而增大,可以保证转动板迅速回弹,避免气动力较强时贴合不紧密,出现无法封严甚至振动发散的现象。整个结构在机翼内部占据了较大空间,可能会与襟翼操纵机构及导流片发生干涉。

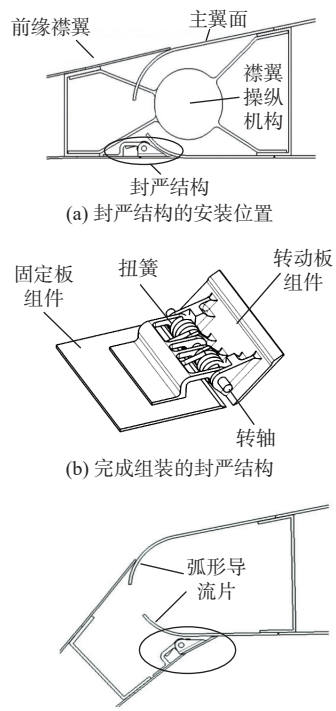


图 8 前缘襟翼缝隙封严结构[9]

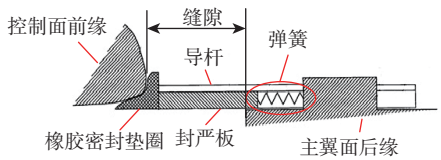
Fig. 8 A kind of gap sealing structure for leading edge of flap[9]

2.1.3 弹簧预加载式封严结构

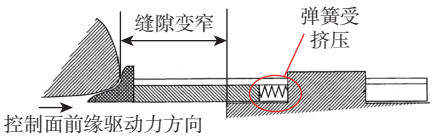
Wildman[10]提出了使用预加载弹簧来实现封严的结构,如图 9 (a) 所示,结构包含橡胶密封垫圈、

导杆、封严板 3 部分粘接成的整体和弹簧零件。控制面偏转时, 其前缘会给橡胶密封垫圈施加驱动力(见图 9(b)), 使导杆沿着固定轨道滑动。封严板的宽度始终大于缝隙宽度, 弹簧一直处于受压状态, 保持橡胶密封垫圈与控制面前缘贴合。

弹簧预加载式封严结构分组安装在机翼上, 相邻 2 组间通过卡槽相互搭接(见图 10(a))。现代客机机翼的弯曲幅度较大, 单组封严结构不能过宽, 整个机翼需要多组, 如图 10(b) 所示。机翼弯曲时, 搭接部位在卡槽内滑动(见图 10(c))。如果封严板使用全金属材质, 可能会导致机翼质量大幅增加, 严重影响控制面的偏转。



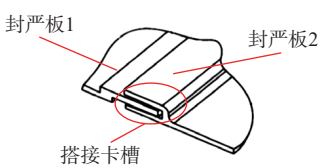
(a) 各组件图示说明



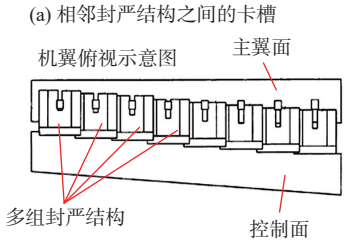
(b) 控制面前缘与主翼面后缘之间缝隙变窄

图 9　弹簧预加载式封严结构[10]

Fig. 9　Spring preloaded sealing structure[10]



(a) 相邻封严结构之间的卡槽



(b) 多组封严结构在机翼上的排布

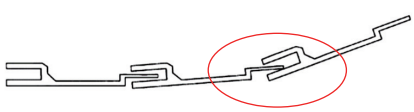

(c) 搭接部位在槽内滑动

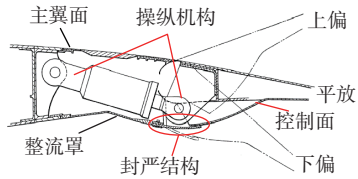
图 10　长机翼上的多组封严结构[10]

Fig. 10　Multiple sealing structures on long wings[10]

2.2　依靠材料的设计形式

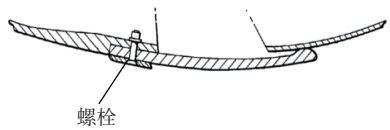
2.2.1　高弹性密封板式封严结构

Johnson[11] 设计了一种控制面整流罩封严结构, 如图 11(a) 所示。结构主体是一块纤维增强复合材料层合板, 安装之前具有较大的弯曲弧度 (见图 11(b))。图 11(c) 为安装后示意图, 封严板的一端使用螺栓

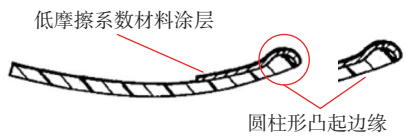


(a) 封严结构安装位置


(b) 未安装板材初始形貌



(c) 封严结构安装示意图放大图



(d) 防刮伤自由端细节设计

图 11　整流罩封严结构[11]

Fig. 11　A kind of fairing sealing structure[11]

固定在主翼面整流罩的外表面, 另一端通过板材的变形回弹紧密贴合在控制面外表面, 实现控制面与机翼之间的缝隙封严。图 11(d) 为自由端剖视图, 圆柱形凸起边缘可以减小自由端与可动翼面蒙皮的接触面积, 低摩擦系数材料涂层可以降低滑动摩擦力, 减缓蒙皮的磨损。

2.2.2　随动搭接式封严结构

王东等[12] 设计了一种搭接式的封严结构(见图 12)。该结构分为前封严板和后封严板, 用铆钉固支连接在主翼面和操纵面上。前后封严板的圆弧面互相贴合且过盈配合, 圆弧尺寸可适应操纵面偏转时的运动轨迹。这种设计减少了 2 个子部件之间的连接结构, 不会对操纵面转轴的运动造成干扰。但是, 前后封严板之间相对滑动距离较长, 贴合面极易磨损, 气密性降低。王东等[12] 用导电纳米材料对封严板进行了表面处理, 提高其表面粗糙度和耐磨性。该结构在圆弧面贴合处会出现鼓包(见图 12(c)), 可能导致气流在鼓包处提前转捩, 降低操纵效果。

2.3　2 种设计形式特点

依靠机构和依靠材料的设计形式各有优缺点, 总结如下:

1) 依靠机构的设计。结构较为复杂, 零部件数量较多, 部件一般分为传动机构和封严板 2 部分。可动翼面偏转过程中, 使用机械传动带动封严板迅速回弹来实现缝隙封严, 对材料的弹性变形性能及

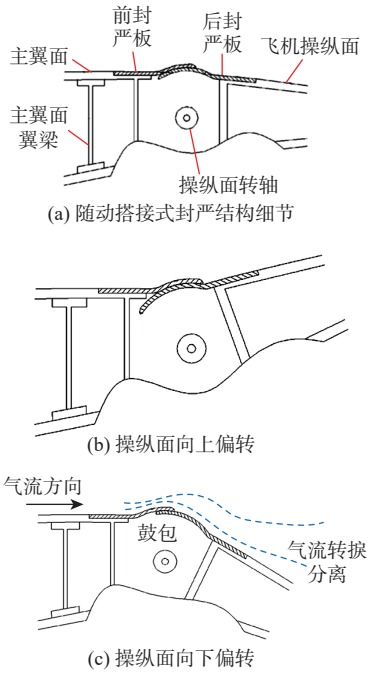


图 12 随动搭接式封严结构及偏转姿态[12]

Fig. 12 Follow-up overlapping sealing structure and deflection attitudes[12]

疲劳性能要求较低，抵抗气动载荷的能力较强，不会出现密封不严、板面振动等现象。但此类封严结构会导致机翼质量大幅增加，且拆装困难，不利于后期检修维护。如果传动机构布置不合理，会干涉其他操纵机构的运动。

2）依靠材料的设计。结构较为简单，零部件数量较少，部件一般仅为一块高弹性的板材，一端固支连接在主翼面上，另一端为自由端。高弹性板材在安装时便施加一个预变形，以保证在可动翼面的整个偏转角范围内，板材的回弹变形都可以使自由端紧密贴合在可动翼面表面。此类结构安装较为方便，不会给机翼增加较多质量，不会影响机翼内操纵机构的运动，但对材料的弹性大变形及疲劳性能要求较高，且易出现老化损坏，需要经常更换。

3 封严结构设计技术现状

3.1 选材制造技术

封严结构设计选材一般遵循经济性、可靠性、减重等原则，制造工艺因选材而异。

如果材料不发生大变形，可使用较为常见的不锈钢、铝合金等金属材料[9]。成骏等[13]对某型飞机封严板的制造工艺进行改进，发明了一种加工连接件安装孔的气动压窝器，孔周围没有出现大面积的起皱变形，表面质量大幅提升。对于发生材料大变形的结构，一般选用纤维增强复合材料层合板或橡胶。Blades[14]使用聚氨酯作为树脂基体。Johnson[11]使用液态橡胶材料浸润二维编织碳纤维，生成单个铺层，再将多个铺层铺叠加压固化。许腾飞等[15]设计了一种夹层封严结构，使用碳布和石英布铺层作为外层铺层，高弹性橡胶作为内部夹层。

需要在热带海域服役的舰载机[16]的封严结构对材料的湿热疲劳性能要求较高。新兴的竹纤维增强复合材料具有原料来源广、制备简单、耐湿热[17]等优点。竹纤维能提高复合材料断裂伸长率，甚至能提高树脂的结晶速率和结晶度[18]，可纳入选材范围。

NASA 格林研究中心[19]研发出了以耐高温材料编织物和陶瓷层叠薄片为基础的密封结构，这些耐高温结构在航天飞机和 X-38 空天往返飞行器上经受了考验。

3.2 自由端耐磨设计

如果自由端较为尖锐，易将可动翼面接触位置的蒙皮磨损划伤，设计时需重点考虑。

Lasar Aero Styling & Repair 公司的副翼封严结构表面涂覆了聚酯薄膜，可以减轻摩擦损伤，同时也可以降低滑动摩擦力，不会使副翼偏转受到阻滞。

Blades[14,20]在封严板自由端粘接了一块硅树脂橡胶（见图 13(a)），不同偏转状态下，橡胶变形后的轮廓不同，以贴合与其接触的可动翼面表面。考虑到胶粘的橡胶易脱落，Blades[14,20]设计了一种中空结构的自由端，内部填充硅树脂橡胶，如图 13(b) 所示，但硬质复合材料外壳使接触位置的弹性变差，一定程度上影响了气密性。

Northrop Grumman[21]公司提出可在蒙皮表面接触区域安装耐磨铁磁条（见图 14），既可以减少摩擦损伤，又可以给金属封严板提供一定的磁吸力，提高封严板的回弹速度。

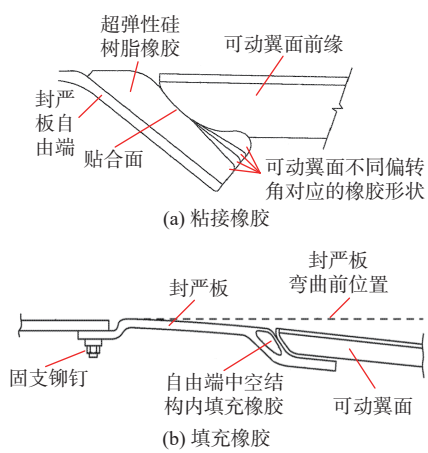


图 13 粘接或填充在封严板自由端的硅树脂橡胶[14,20]

Fig. 13 Silicone rubber bonded or filled at free end of sealing plate[14,20]

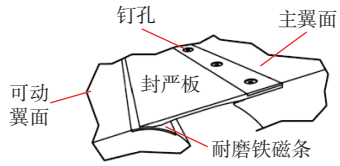


图 14　封严板自由端与可动翼面接触位置的耐磨铁磁条[21]

Fig. 14　Wear resistant ferromagnetic strip at contact position between free end of sealing plate and movable wing surface[21]

3.3　封严结构验证方法

封严结构的密封效果需要使用不同方法进行验证。

许腾飞等[15]通过静力试验测量了封严结构所用橡胶材料的实际力学性能, 用试验结果拟合出了该种橡胶材料的力学本构模型, 并在 ABAQUS 仿真软件中将模型用于橡胶夹层结构, 仿真结果显示, 该种橡胶材料具有良好的回弹特性, 满足封严板的设计要求。

除了仿真计算和力学试验, 风洞试验也是评估密封效果的重要方法, 空客公司设计增升装置时就使用了多个风洞进行试验验证。雷诺数对封严结构附近的气流流动, 包括附着边界层转捩、黏性尾迹的相互干扰、机翼尾部复杂流动等都有较大影响[22], 试验一般在高雷诺数风洞中进行。目前, 常见的高雷诺数风洞分为大气边界层风洞、压力风洞、低温风洞 3 种。大气边界层风洞通过提高马赫数来提高雷诺数, 适用于大尺寸全机模型的高雷诺数试验, 能对机翼尾部大片区域的气流特性进行监测, 观察封严结构对机翼表面气流的影响[23]。压力风洞靠增压来提高雷诺数, 但较大的压力会改变可动翼面与主翼面之间的缝道参数, 引起封严结构气动性能的变化[24]。低温风洞是近年来最常见的高雷诺数风洞, 不需要提高压力和马赫数, 适用于小型模型的气动验证, 虽然建造成本较高, 但收益很大[25]。

现阶段, 使用单一方法难以完成验证, 需要仿真计算、力学试验、风洞试验三者相结合。仿真主要进行前期气动载荷计算和后期力学结构优化; 力学试验可以对材料和结构的力学性能进行检测, 初步判断能否承受相应气动载荷; 风洞试验可对前 2 个方法的准确性进行验证, 每种设计方案都应经过风洞试验来最终定型。合理利用这 3 种方法可以节约研发成本, 缩短研发周期。

4　结 束 语

飞机在机翼安装封严结构后, 巡航速率、爬升速率、起降特性、操纵特性等都有所提升。本文罗列的封严结构产品均只适用于小型通用飞机, 大型飞机上的应用并不多见。对比分析几种封严结构的设计特点, 发现很难同时兼顾结构简单和材料易得 2 个方面。金属、橡胶、纤维增强树脂基复合材料是封严结构较为常见的材料, 就目前的技术而言, 在简化结构设计的同时, 材料的许多关键性能仍有待提升。

封严结构的设计建议如下:

1) 在实现缝隙密封的同时, 应保证从主翼面到可动翼面具有连续光滑的气动外形, 尽量减少台阶、鼓包的出现。

2) 依靠机构的设计方法, 应尽量减少零件数量, 减轻结构质量, 同时简化安装工艺, 便于后期检修。安装时, 尽量避免对蒙皮等原始结构造成损伤, 对使用的铆钉或沉头螺栓, 应严格把控拧紧力矩等工艺参数。

3) 依靠材料的设计方法, 材料应具有良好的弹性变形能力和优良的疲劳性能, 在经受气动载荷和可动翼面驱动力循环作用后, 仍能保持良好的回弹特性, 结构产生的永久变形应尽量小。同时, 应当考虑材料性能在湿热等特殊环境中的变化。

4) 将封严结构的整体刚度限制在一定范围内。刚度太小, 无法承受气动载荷, 会出现回弹滞后的现象, 影响密封效果; 刚度太大, 会给可动翼面施加较大的反作用力, 影响操纵效果。同时, 应考虑大板面的气动弹性效应, 如果出现振动, 要尽量避免发散导致的结构快速破坏。

5) 自由端需要进行特殊的表面处理, 使其具有良好的耐磨损性能, 可减少对翼面蒙皮的摩擦损伤。

6) 设计需要仿真计算、力学试验、风洞试验 3 种方法结合。

7) 严苛的受力状态和服役环境使封严结构成为机翼上最易损坏的结构之一, 需要经常检修, 可能需要进行场外临时更换。维护人员的工具包中应配备快速拆装设备, 如果是分块式设计, 应携带至少一组封严结构零部件以备更换使用。

参考文献（References）

[1] 李丽雅. 大型飞机增升装置技术发展综述[J]. 航空科学技术, 2015, 26(5): 1-10.
LI L Y. Review of high-lift device technology development on large aircrafts[J]. Aeronautical Science & Technology, 2015, 26(5): 1-10 (in Chinese).

[2] 金伟, 杨智春, 孟德虹, 等. 先进战斗机全动 V 尾抖振动强度设计与验证[J]. 航空学报, 2020, 41(6): 523473.
JIN W, YANG Z C, MENG D H, et al. Strength desigh and test of advanced fighter all-moving twin V-tail buffet[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(6): 523473(in Chinese).

[3] 聂春生, 阎君, 曹占伟, 等. 头部外形对升力体转捩影响的试验研

究[J]. 导弹与航天运载技术, 2022(4): 104-108.
NIE C S, YAN J, CAO Z W, et al. Study on the influence of head shape on transition of lifting body[J]. Missiles and Space Vehicles, 2022(4): 104-108(in Chinese).

[4] HERNANDEZ J, FOULIARD Q P, VO K, et al. Characterization of corrosive defects through pulsed eddy current thermography for aircraft panels: AIAA 2021-0433[R]. Reston: AIAA, 2021.

[5] 桑建华, 张宗斌, 王烁. 低RCS飞行器表面弱散射源研究[J]. 航空工程进展, 2012, 3(3): 257-262.
SANG J H, ZHANG Z B, WANG S. Research on the radar cross section of weak scatterers on stealth vehicle[J]. Advances in Aeronautical Science and Engineering, 2012, 3(3): 257-262(in Chinese).

[6] 王立研, 王菁华, 杨炳尉. 高超声速飞行器控制面动密封技术[J]. 宇航材料工艺, 2016, 46(3): 1-6.
WANG L Y, WANG J H, YANG B W. Dynamic seal technology for control surface of hypersonic vehicles[J]. Aerospace Materials & Technology, 2016, 46(3): 1-6(in Chinese).

[7] DECKER G R, ROBEDEAU S A. Slotted flaperon seal mechanism for aircraft devices: US11046421[P]. 2021-06-29.

[8] 潘立新, 孔斌, 何娅梅, 等. 一种飞行器舵面前缘随动封严结构: CN204433037U[P]. 2015-07-01.
PAN L X, KONG B, HE Y M, et al. A kind of follow-up sealing structure for the leading edge of aircraft rudder surface: CN 204433037U[P]. 2015-07-01(in Chinese).

[9] 胡利, 禹建军, 刘衍腾, 等. 一种前缘襟翼缝隙封严机构: CN108609160A[P]. 2021-05-07.
HU L, YU J J, LIU Y T, et al. A kind of gap sealing mechanism for the leading edge of flap: CN108609160A[P]. 2021-05-07(in Chinese).

[10] WILDMAN E. Flight surface seal: US8292236[P]. 2012-10-23.

[11] JOHNSON B D. Fairing arrangements for aircraft: US7051982[P]. 2006-05-30.

[12] 王东, 胡珺, 陈世春, 等. 一种飞机操纵面封严结构: CN207725605U[P]. 2018-08-14.
WANG D, HU J, CHEN S C, et al. Airplane control face structure of obturaging: CN207725605U[P]. 2018-08-14(in Chinese).

[13] 成骏, 阚小如, 陶义建. 基于Dynaform的飞机封严板压窝成形工艺分析与优化设计[J]. 锻压技术, 2021, 46(1): 37-42.
CHENG J, KAN X R, TAO Y J. Analysis and optimal design on press dimple forming process for aircraft seal plate based on Dynaform[J]. Forging & Stamping Technology, 2021, 46(1): 37-42(in Chinese).

[14] BLADES P. Seal for aircraft wing: US10480653[P]. 2019-11-19.

[15] 许腾飞, 王新峰, 郭树祥. 高弹性橡胶夹层结构封严板分析方法[J]. 南京航空航天大学学报, 2020, 52(3): 394-400.
XU T F, WANG X F, GUO S X. Analytical method of sealed plate with hyperelastic rubber sandwich structure[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(3): 394-400(in Chinese).

[16] 徐丽, 陈跃良, 武书阁, 等. 舰载机舰面停放环境及腐蚀情况研究[J]. 飞机设计, 2016, 36(6): 54-57.
XU L, CHEN Y L, WU S G, et al. The study about the deck park environment of carrier-based aircraft and corrosive state[J]. Aircraft Design, 2016, 36(6): 54-57(in Chinese).

[17] 徐海宇. PLA/竹纤维复合材料的制备及其性能研究[D]. 杭州: 浙江农林大学, 2021.
XU H Y. Preparation and properties of PLA/bamboo fiber composites[D]. Hangzhou: Zhejiang A & F University, 2021(in Chinese).

[18] 梁珊, 李杨, 吴建军, 等. 竹纤维/玻璃纤维混杂增强聚丙烯复合材料[J]. 塑料, 2012, 41(5): 86-88.
LIANG S, LI Y, WU J J, et al. Properties of bamboo fiber/glass fiber/hybrid reinforced polypropylene composite material[J]. Plastics, 2012, 41(5): 86-88(in Chinese).

[19] STEINETZ B M. Seal technology for hypersonic vehicle and propulsion: an overview[C]//Proceedings of the Short Course on Hypersonics Structures and Materials. Hampton: NTRS, 2008.

[20] BLADES P. Seal plate for an aerodynamic surface: US11299254[P]. 2022-04-12.

[21] NEAL M A, SMITH C R. Dynamic conformal aerodynamic seal (CAS) for aircraft control surfaces: US10017239[P]. 2018-07-10.

[22] POLL D I A. Transition in the infinite swept attachment line boundary layer[J]. Aeronautical Quarterly, 1979, 30(4): 607-629.

[23] HAFTMANN B, DEBBELER F J, GIELEN H. Takeoff drag prediction for airbus A300-600 and A310 compared with flight test results[J]. Journal of Aircraft, 1988, 25(12): 1088-1096.

[24] SPAID F W. High Reynolds number, multielement airfoil flowfield-measurements[J]. Journal of Aircraft, 2000, 37(3): 499-507.

[25] 赖欢, 祝长江, 陈万华, 等. 大型低温风洞结构设计关键技术分析[J]. 实验流体力学, 2022, 36(1): 19-26.
LAI H, ZHU C J, CHEN W H, et al. Key technology for mechanical design in large-scale cryogenic wind tunnel[J]. Journal of Experiments in Fluid Mechanics, 2022, 36(1): 19-26(in Chinese).

Study progress of gap sealing structure for aircraft movable wing

CHENG Xiaoquan[1,*], CAI Moquan[1], WANG Songwei[2]

(1. School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China;
2. The Military Representative Office of the Air Force Equipment Department in Shijiazhuang, Shijiazhuang 050081, China)

Abstract: The gap between the main wing surface and the movable wing surface might have a certain impact on the flight performance of the aircraft. The application of the gap-sealing structure will improve the surface smoothness of the wing while achieving effects such as increasing lift, reducing drag, and optimizing operation. This article compares various sealing structure design forms and divides them into two categories: relying on materials and relying on mechanisms. The benefits and drawbacks of each category are examined.The sealing structure products of several aviation parts companies are compared. Analysis is done on intricate designs like wear-resistant free ends. Design suggestions are provided at last, such as rapid disassembly and repair, limitation for the range of the structural stiffness, and a combination of simulation and experiment while designing.

Keywords: aircraft sealing structure; gap seal; flight performance improvement; large elastic deformation; fatigue property

Received: 2023-06-19; **Accepted**: 2023-09-08; **Published Online**: 2023-09-15 16:24
URL: link.cnki.net/urlid/11.2625.V.20230915.0957.001

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0413

基于功率损失特性分析的特种车辆传动性能研究进展

高钦和[1,*]，高蕾[1]，刘志浩[1]，王冬[1]，马栋[1]，章一博[2]
(1. 火箭军工程大学 导弹工程学院，西安 710025；2. 长安大学 公路学院，西安 710054)

摘 要：车辆传动系统是发动机与车轮负载之间的动力传递装置，作为动力系统、行走系统、制动系统的连接枢纽，保证特种设备在作战任务中的机动性与安全性。基于此，从传动系统结构组成特性、动力传递特性、能量流特性与整车动力特性4个方面论证了传动效率是评估传动系统性能的重要技术指标；在分析功率损失与传动效率关系的基础上，从产生机理与结构特性角度搭建了传动系统功率损失全局模型框架，为传动系统性能表征提供了理论依据；从传动效率的理论数值研究、建模仿真研究和试验论证研究3个方面论述了车辆传动系统功率损失的研究难点与重点，指出大型机械设备在复杂运行工况下的多介质、多参数耦合仿真分析与试验研究可为实现车辆传动部件的性能优化与健康管理、传动系统的性能评估与匹配设计提供指导思路；指出多因素下的综合传动效率分析，以及基于智能算法的功率损失特性、效率特性、性能退化特性耦合研究，对大型机械设备性能监测与评估具有实际工程意义。

关 键 词：特种车辆；传动系统；功率损失特性；传动效率；传动性能监测与评估
中图分类号：U463.21; TJ812; TH132.4
文献标志码：A **文章编号**：1001-5965(2025)06-1824-19

特种车辆作为大型设备承载、运输和发射的主要平台，其动力传动性能是实现快速机动作战任务的前提与基础。因此，特种车辆对行驶系统动力性指标[1]提出了更高的标准，即具备更高的机动速度、更短的加速时间和更大的爬坡度。研究表明，输入传动系统的能量仅约为动力源能量的20%，同时，由于系统内部阻力，动力在传动系统内部逐级传递时仍会产生约5.6%的能量损失[2]。特种车辆传动系统主要采用传统机械传动方式和结构布局，传动系统作为中间关键环节是整车驱动力的保证，传动效率将直接影响驱动力输出的有效性和可靠性。

相比普通两轴车辆，特种车辆轴数多、轴距大、轮距宽，使得传动系统布局复杂。由于累积效应，其在较多的动力传递层级和较长的传递路径中将产生更多传动环节的功率损失；多轴传动系统的子系统结构相对复杂，功率损失类型相应增多，同时，复杂行车环境增加了功率损失产生的不确定性。这均导致特种车辆传动系统性能分析与能量效率表征难度增大。对于传动系统功率损失与传动效率的关系，传动系统外部使用条件一定时，内部功率损失越小，传动效率越高。功率损失特性与各传动机械部件运行状态和车辆行驶工况密切相关，即由外部使用条件和内部技术状态共同决定，是传动系统动力传递状态和部件性能演变趋势的综合反映。因此，传动效率是表征传动系统综合性能的重要指标。研究传动系统功率损失组成特性和功率流传递特性，并分析系统内部能量损耗与整

收稿日期：2023-06-28；**录用日期**：2023-10-13；**网络出版时间**：2023-10-26 10:26
网络出版地址：link.cnki.net/urlid/11.2625.V.20231025.1705.004
基金项目：国家自然科学基金(51905541)；陕西省自然科学基础研究计划(2020JQ487)；陕西省高校科协青年人才托举计划(20190412)
*通信作者. E-mail：lycoris_lei@163.com

引用格式：高钦和，高蕾，刘志浩，等. 基于功率损失特性分析的特种车辆传动性能研究进展[J]. 北京航空航天大学学报，2025，51(6)：1824-1842. GAO Q H，GAO L，LIU Z H，et al. Research progress on transmission performance of special vehicle based on power loss characteristics analysis[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1824-1842 (in Chinese).

体能量传输多层次之间的理论关系, 对实现基于能效特性的复杂机械传动系统状态监测与性能评估具有重要的实际意义。

本文结合目前对机械传动系统传动效率的研究深度和广度, 围绕多轴驱动特种车辆传动系统, 在传动系统特性分析的基础上, 对传动系统功率损失的组成特性、效率特性及不同层级传动效率研究方法等方面进行阐述和梳理, 以期为特种车辆动力性能优化、快速机动能力评估、传动系统状态监测与寿命预测研究提供参考依据。

1　特种车辆传动系统特性分析

通过分析特种车辆传动系统结构组成, 确定传动系统的动力传递链路, 研究传动系统能量流特性及与整车动力特性之间的关联性, 是传动效率特性建模和传动性能分析的基础。

1.1　传动系统的结构组成特性

从层次结构和组成成分角度, 特种车辆传动系统层级划分如图 1 所示。特种车辆是机动发射作战单元中的子系统, 传动系统是特种车辆底盘系统的子系统, 也是发动机与驱动车轮之间的动力传动装置, 其作为行驶系统、制动系统、转向系统的连接枢纽, 保证了车辆在战训任务中的机动性与安全性。

传动系统主要由变速器、分动器、驱动桥、万向传动轴组成, 其中, 驱动桥包括主减速器、差速器、半轴和轮边减速器[3]。同时, 传动系统中的关键部件均采用机械传动方式, 以齿轮啮合传动为主, 轴承为支撑部件, 万向节和传动轴为连接部件, 因此, 不同结构、功能的传动模块均由齿轮、轴承、传动轴和万向节 4 类典型机械旋转部件构成, 也是传动系统功率损失组成分析和传动效率研究的主要对象。

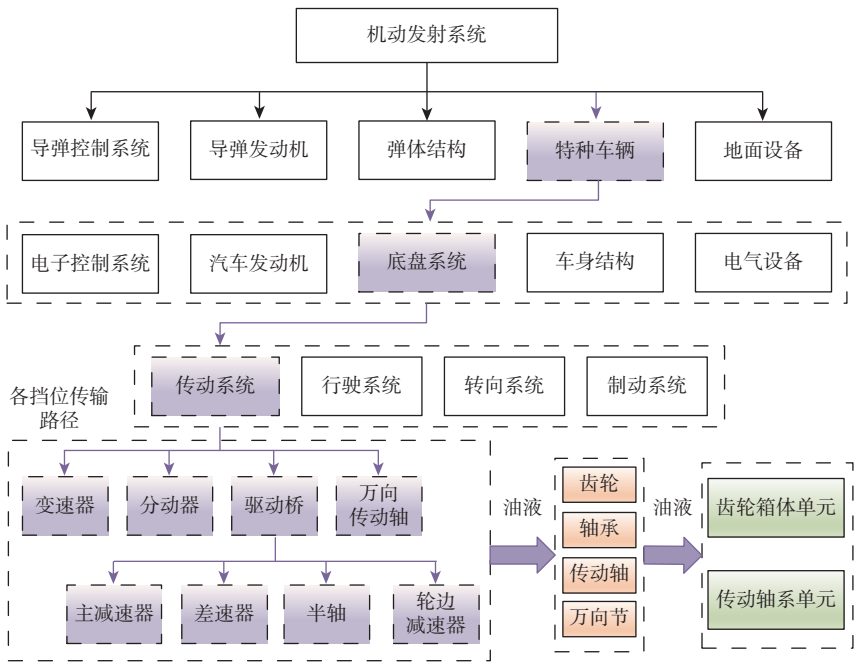


图 1　特种车辆传动系统层级划分

Fig. 1　Hierarchical division of special vehicle transmission system

1.2　传动系统的动力传递特性

特种车辆传动系统的动力传输结构如图 2 所示。动力传动的作用起点为变速器, 作用终点为轮边减速器, 主要通过变速器将发动机的动力按照一定传动比传递至驱动轮, 利用多级啮合齿轮实现转速和转矩的变化。同时, 由于该底盘布局, 传动系统需配置分动箱, 动力先由传动轴传递至分动箱, 再由分动箱分配驱动力至各个驱动轴。传动轴的传动作用借助万向传动装置实现, 解决了动力传递过程、适应转向和车辆运行时产生的上下跳动角度变化问题[4]。各驱动桥将经由万向传动装置的动力通过减速增扭的方式分配至同轴两端驱动轮, 最终产生驱动力矩。因此, 动力传递过程中存在多传动环节耦合作用, 导致传动效率衰减的累积效应更加突出。

1.3　传动系统的能量流特性

传动系统动力传动过程均伴随着能量传输, 系统中的每个传动环节都存在能量输入、存贮、释放、损耗和输出的动态过程, 且能量效率水平会直接影响驱动力的输出量值。其中, 机械传动系统能量传输具备 3 个主要特性[5]: ①传动系统能量流动状态是其能量特征变换与动力传动运行状态的综合反映; ②多轴驱动车辆行驶过程中能量损耗程度

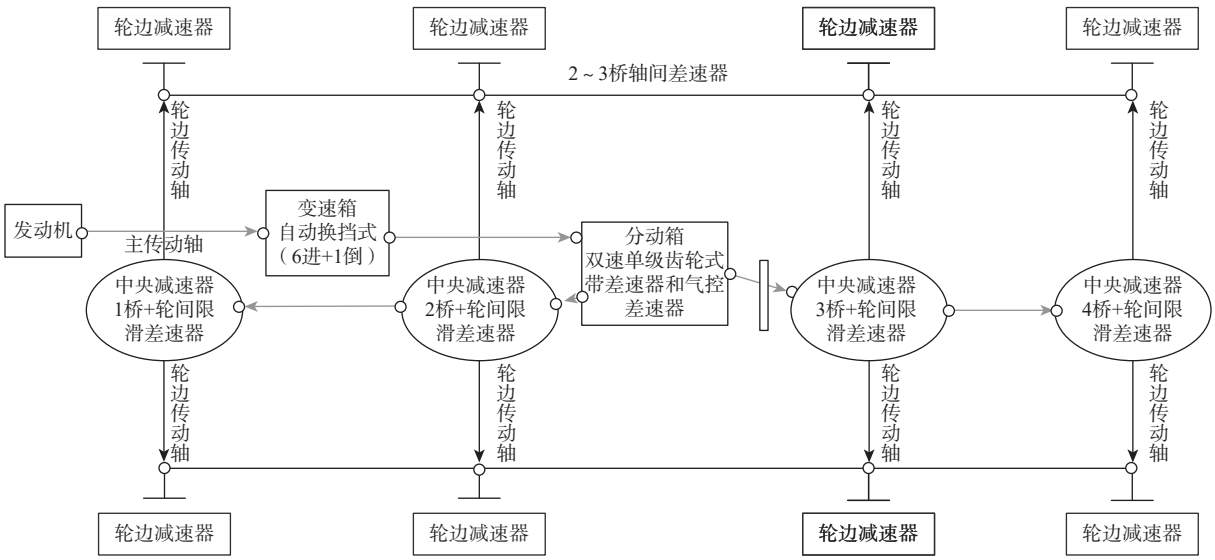


图2 特种车辆动力传递示意图

Fig. 2 Power transmission of special vehicles

较高，其与机械部件的摩擦学状态、运行工况和车辆行驶工况密切相关；③传动部件技术状态发生变化（如磨损、破损），通常会引起某一部分的能量状态发生变化，通过能量状态可直观反映传动部件内部异常。因此，能量流特性对车辆整体动力性能和传动部件结构性能的影响关系可表示为图3[5]。

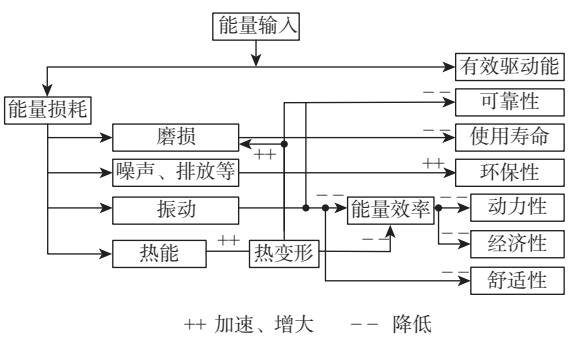


图3 特种车辆传动系统能量流特性[5]

Fig. 3 Energy flow characteristics of special vehicle transmission system[5]

目前，已有大量学者针对变速箱、驱动桥等不同类型齿轮箱结构单元的摩擦学能量损耗展开理论、仿真、试验研究[6-8]，分析了能量损耗与传动副动态磨损状态、润滑油液性能之间的关系[9-12]，以及能量损耗对部件承载能力、散热特性的影响机理[13]。但少有建立车辆传动系统能量特性与性能特性监测分析的实际联系，其在风电系统、机械制造系统等复杂机械系统中的应用为此提供了很好的研究思路与导向。例如，重庆大学团队通过建立传动系统能量传输数学模型实现了机床加工制造系统的能效预测，并构建了功率数据驱动的实时智能诊断系统[14]，对其机械加工过程开展了状态监测技术研究[15-16]。华南理工大学团队基于功率信号检测技术，从能量损耗的角度开展发动机、齿轮箱故障模式识别研究[17-19]。Gu等[19]提出了基于输出功率数据分析的风电机组运行效率在线监测方法。Ding等[20]利用改进的热弹性流体动力润滑（thermal elastohydrodynamic lubrication，TEHL）方法提出了一种新的能量损失和力学效率预测模型，为航发动力系统能量产生、转换、传递和利用提供重要的几何和物理评估途径。

因此，基于能量流理论的能量传输特性研究，能够从整体能量效率的角度评估机械传动系统的健康状态，可通过监测功率信号识别传动部件的异常状态，还可通过摩擦学能量损耗进行物理机理研究，建立能量损耗特性与传动部件内部结构损失退化特性之间的表征关系。

1.4 整车动力特性

结合传动系统的结构组成特性、动力传递特性和能量流特性可知，传动系统的性能直接关系到驱动力矩的输出，是车辆在各种特殊、极端工况下机动能力的保证，主要体现在车辆最高行驶速度、加速能力、极限爬坡能力3个主要方面。目前，广泛采用的表征车辆动力特性和传动系统特性的指标参数主要为各档最大驱动力、动力因数、最大爬坡度、最高车速、加速度、高档加速时间、传动效率等[21]。其中，华琛等[22]在车辆机动性评估方法研究中指出，车辆机动性指数与车辆的接地压系数、发动机系数、传动系数等有关。张金豹等[23]在车辆传动装置传动性能评估中建立了基于中位秩加权的综合传动效率指标。Dong等[24]以最小功率损耗为目标函数，采用动态规划方法评价多挡位变速器的能量利用性能。Lei等[25]从车辆层面的能耗特性、动力特性、平顺性建立了动力总成的综合性能

优化方法。Han 等[26]从齿轮传动误差角度, 基于偏差齿接触分析和负载齿接触分析对齿轮传动性能进行评价。同时, 车辆传动系统性能分析主要集中在动力性、燃油经济性、扭振动态特性[27]。通过传动系统参数匹配优化, 降低扭振特性, 增加传动系统的耐久性和可靠性, 并将传动效率作为提高传动系统动力性和燃油经济性的综合性能优化指标。

由此可见, 传动效率是评估车辆传动系统性能的重要技术指标, 包含丰富的运动变换和动力传递状态信息, 直接影响到整车的运动学、动力学性能和传动系统关键部件的可靠性, 也是特种车辆快速机动、灵活作战能力的保证。

2　传动系统效率特性基础分析

传动系统效率特性取决于系统输入功率和功率损失, 从功率流角度分析时, 主要不同点在于各部件机械结构、子系统间的功率传递路径。因此, 从该角度分析此类复杂机械系统性能更具适应性, 可将传动效率特性分析等价于对传动系统各子系统的关键部件功率损失组成成分、形成机理和主要影响因素的分析, 如机械传动部件的磨损、裂纹、润滑系统故障均会使系统功率损失增大。由此, 可进一步明确传动系统功率损失与系统输入输出工况、性能演变状态的关系, 实现传动系统的性能预测与评估, 指导传动系统优化设计等。

2.1　传动效率理论分析

依据现有汽车传动系统效率测试试验标准中对传动效率的定义[28], 并结合文献中对工程机械传动部件[29]、减速器[30]、驱动桥[31]等齿轮箱结构传动效率的理论分析, 设定车辆处于某一挡位、车速和动力传输路径下的传动系统效率η, 其中, 发动机输出功率P_{E}经传动系统传递至驱动轮的过程中, 为克服传动系统中的各项阻力, 需消耗一部分功率P_{L}, 即传动系统输入功率一定时, 传动系统子系统各部件的功率损失决定其传动效率。

$$\eta=\frac{P_{\mathrm{E}}-P_{\mathrm{L}}}{P_{\mathrm{E}}}=1-\frac{P_{\mathrm{L}}}{P_{\mathrm{E}}} \tag{1}$$

式中: η为传动效率; P_{E}为发动机输出功率; P_{L}为传动系统总功率损失。

对于传动系统效率试验测试, 可在系统输入、输出端分别安装转速转矩传感器, 在某一稳态运行工况下, 通过式(2)计算分析整个传动系统的传动效率:

$$\eta=\frac{P_{\mathrm{in}}}{P_{\mathrm{out}}}=\frac{T_{\mathrm{in}}n_{\mathrm{in}}}{T_{\mathrm{out}}n_{\mathrm{out}}} \tag{2}$$

式中: P_{in}、P_{out}分别为传动系统输入、输出功率; T_{in}、T_{out}分别为传动系统输入、输出端转矩; n_{in}、n_{out}分别为传动系统输入、输出端转速。

2.2　传动系统功率损失成分分析

传动系统中的机械传动部件均由不同类型的齿轮、轴承、万向节和传动轴类元件组成, 润滑系统则保证元件间的正常运转, 不同类型的元件产生的功率损失类型不同, 不同结构的传动部件产生功率损失的部位也不同。传动系统功率损失主要由子系统内部件间的负载摩擦作用、黏滞作用、惯性作用和热量传递产生。

1) 整车传动系统功率损失。莫易敏研究团队对微车传动系统效率进行了数值建模和试验研究[32-33], 以传动系统关键部件为分析单元, 以功率流传输路径为链接, 根据元件内部、元件之间的作用机理, 将传动系统功率损失分为机械摩擦阻力功率损失和搅动阻力功率损失两大类, 具体功率损失组成如图 4 所示。其中, 机械摩擦阻力功率损失在车辆正常运行工况下为主要损失, 若传动部件处于部分负载或较高的转速时, 搅动阻力功率损失占比则增大[8]。

田蜜[33]在传动系统参数优化数值分析中认为, 半轴的功率损失来自油封和支撑半轴的轴承。王熙[34]在传动系统功率损失分析中考虑了传动轴功率损失, 并认为轴类元件的功率损失主要由两端连接处的万向节轴颈摩擦产生。何平[5]对重型车传动系统能量特性的研究中认为, 传动系统能量损耗主要由 3 部分组成, 包括湿式换挡离合器的带排损失、摩擦副相对运动引起的机械摩擦阻力损失、旋转部件搅油摩擦阻力引起的搅油损失。与小型车辆不同之处在于: 重型车辆传动系统功率损失中占比较大的是前 2 种类型。对于使用液力机械式自动变速器的传动系统[3], 功率损失还应考虑液力变矩器的功率损失, 包括液力损失、机械损失和容积损失。

2) 传动系统主要传动部件功率损失。传动系统总功率损失由各个子系统传动部件累积传递作用产生, 目前, 对传动系统关键部件功率损失的研究主要集中在变速器总成、驱动桥总成或驱动桥中的主减速器。此类部件的功率损失组成可总结为齿轮箱结构单元功率损失组成。齿轮箱功率损失可分为空载功率损失和负载功率损失两大类[35], 在 Fernandes 等[36-37]建立的齿轮箱功率损失模型中(见图 5), 功率损失可划分为: ①啮合齿轮的负载功率损失P_{VZP}; ②啮合齿轮的非负载功率损失, 即搅油功率损失P_{VZ0}; ③轴承的负载功率损失与非负载功率损失总和P_{VL}; ④油封摩擦功率损失P_{VD}; ⑤空气阻力功率损失P_{VX}。

综上所述, 对传动系统功率损失成分分析需依据变速箱某一挡位下的动力传递路径, 结合传动系

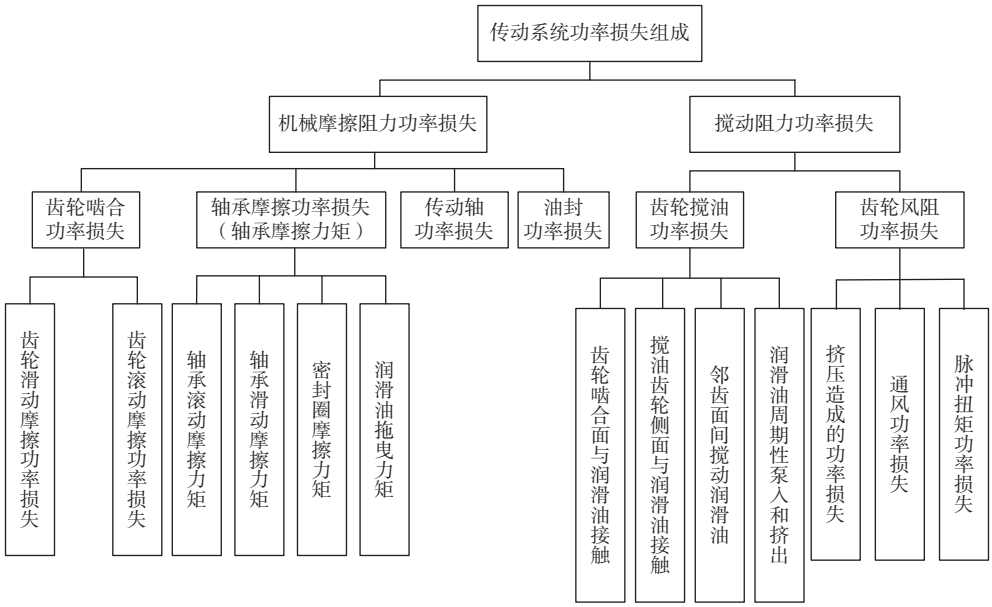


图4 传动系统功率损失组成成分

Fig. 4 Transmission system power loss composition

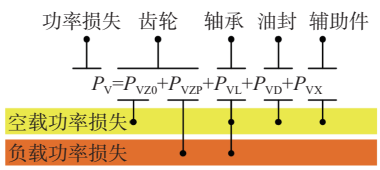


图5 齿轮箱功率损失组成[36-37]

Fig. 5 Gearbox power loss composition[36-37]

统的结构层次和齿轮、轴承、万向节、传动轴4类传动元件的功率损失类型，建立特种车辆传动系统功率损失全局模型框架，如图6所示。传动系统总的功率损失可表示为

$$P_L = P_{LT} + P_{LTC} + P_{LD} = \sum P_h + \sum P_g + \sum P_b + \sum P_{ch} + \sum P_{wind} + \sum P_{seal} + \sum P_{ui} \tag{3}$$

式中：P_{LT}为变速器总成功率损失；P_{LTC}为分动器总成功率损失；P_{LD}为驱动桥总成功率损失；$\sum P_h$为液力变矩器功率损失；$\sum P_g$为总齿轮啮合功率损失；$\sum P_b$为总轴承摩擦功率损失；$\sum P_{ch}$为总搅油功率损失；$\sum P_{wind}$为总风阻功率损失；$\sum P_{seal}$为总油封功率损失；$\sum P_{ui}$为总万向传动装置功率损失。

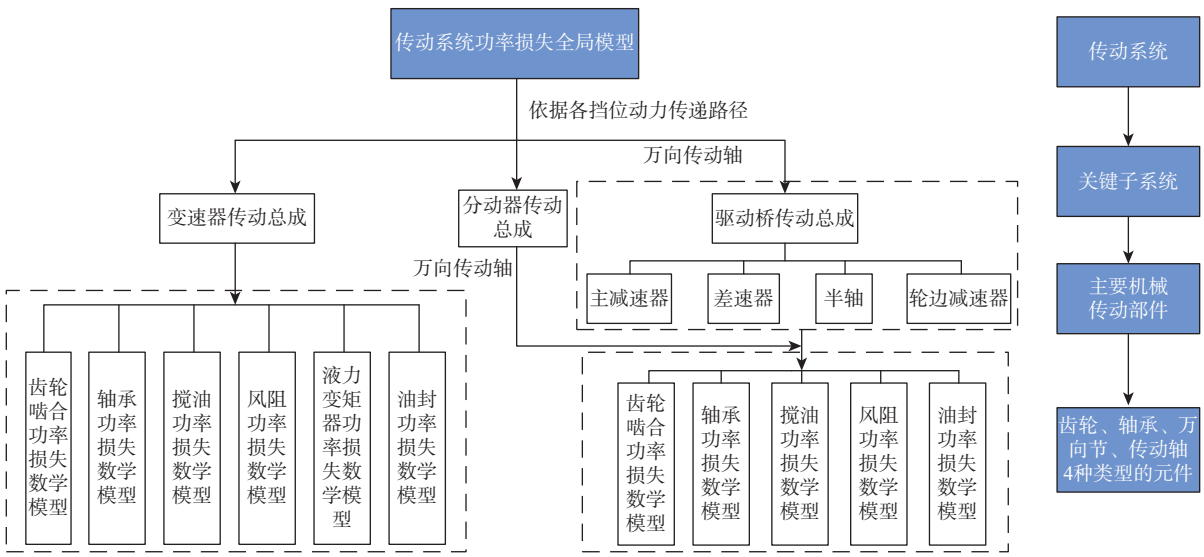


图6 传动系统功率损失全局模型框架

Fig. 6 Global power loss model framework of transmission system

3 传动系统效率的理论数值研究

传动系统功率损失的理论研究难点在于：多轴车底盘传动回路分支较多，各驱动回路的传动效率对车辆动力学特性造成的影响不同；传动部件内部结构复杂，元件类型不一，齿轮、轴承、传动轴之间

相互作用, 总功率损失不是单一元件的简单累加; 传动部件内的润滑系统流体力学特性复杂, 存在油气混合现象; 汽车传动部件工作时处于运动状态, 且输入转速和扭矩工况随实时车速、路况动态变化; 传动部件的温度场变化较大, 与润滑系统的流场存在耦合特性。

3.1　变速器总成传动效率理论数值研究

图 7[38] 为某型号变速器不同挡位动力传递示意图。变速器功率流路径基于挡位切换变化, 较其他传动部件更为特殊, 因此, 对变速器总成功率损失理论数值研究较多。张广庆等[29] 通过求解液压传动系统流量连续方程与变速器输入输出扭矩平衡方程, 建立了无级变速器的传动效率理论模型。王成[39] 采用啮合线积分法建立了齿轮副啮合效率模型, 并依据分流功率理论和虚功率理论构建了行星轮系的传动效率模型。柴蓉等[40] 结合离合器动力传递计算模型推导了带排功率损失计算方法, 分析了离合器对液力变速器传动效率的影响。冯能莲等[41] 在活齿无级变速器传动效率理论分析中研究了齿侧滑动摩擦功率损失、冲击功率损失、搅油功率损失和油封功率损失。李优华等[42] 建立的低速重载齿轮箱传动效率理论分析模型中包括齿轮啮合功率损失、轴承摩擦功率损失、搅油损失和风阻损失 4 种类型。王熙等[43] 针对车辆变速器建立了 5 类主要功率损失类型的理论分析模型, 并通过变速器台架试验验证了不同输入转矩和转速下模型的准确性。美国通用公司基于弹流润滑理论建立了某六挡变速器传动效率计算模型, 认为齿轮啮合功率损失和轴承功率损失为主要损失类型[44]。

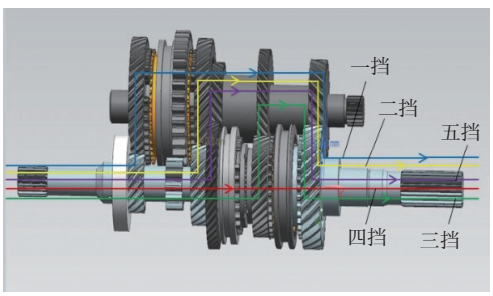


图 7　变速器不同挡位动力传递示意图[38]

Fig. 7　Schematic diagram of power flow in transmission gear position[38]

上述传动部件功率损失理论分析模型中, 搅油、风阻类型的损失多是基于试验方法拟合的经验公式或量纲分析法推导的计算模型, 而齿轮、轴承类型的损失是基于结构动力学原理、摩擦学原理推导的理论效率模型。此外, 也有学者将智能神经网络算法运用于传动效率预测。陈柯序等[45] 融入了反向传播(back propagation, BP)神经网络数学模型, 通过变速器台架试验获取数据集, 以润滑油温度、转速、转矩为模型输入, 以变速器效率为模型输出。Autiero 等[46] 以有限的工况点为训练集建立了基于神经网络的功率损耗评估降阶模型, 高效预测了齿轮箱整体的效率和温度。Karabacak 等[47] 利用神经网络开展了不同工况下蜗杆齿轮箱功率损失数建模, 并评估了健康齿轮箱和故障齿轮箱的效率特性。Fotso 等[48] 通过风电机组齿轮箱试验数据修正了 SKF(Svenska Kullager-Fabriken)轴承功率损失计算模型, 并基于 BP 神经网络对轴承功率损失进行了有效预测。

3.2　驱动桥总成传动效率理论数值研究

驱动桥总成功率损失理论数值研究方法与变速器基本一致, 区别在于: 驱动桥的作用主要是减速增扭, 运行工况多为低转速重负载; 主减速器中采用的螺旋锥齿轮齿面接触承载分析相对圆柱齿轮复杂。准双曲面齿轮几何学和运动学由 Barday 等[49] 识别和量化了卡车驱动桥的主要功率损失, 包括齿轮滑动摩擦损失、滚动轴承损失和油液损失。Kolivand 等[50] 结合计算效率接触模型和混合弹流润滑模型预测螺旋锥齿轮和准双曲面齿轮的功率损失, 量化了齿轮设计参数和工作参数对功率损失的影响规律。Xu 等[51-52] 建立了以非牛顿热弹流润滑理论为基础的平行轴齿轮副摩擦模型, 利用多元线性回归法研究了操作工况、表面光洁度和润滑剂性能对机械功率损失的影响。

国内学者多将驱动桥作为整体进行综合分析[53], 采用理论推导结合试验设计论证的方法研究驱动桥功率损失特性, 以实现提高驱动桥传动效率的目的。贾富淳[54] 面向驱动传动装置中的斜齿轮提出了基于切片法和线接触弹流润滑理论的摩擦功率损失计算方法。朱建杏[55] 基于膜厚比判定规则分别建立了边界润滑、弹流润滑和混合润滑的摩擦系数数学模型, 为减速器齿轮摩擦功率损失分析提供了计算方法。王钦等[56] 结合传动系统等效啮合模型与齿轮接触加载分析方法, 建立了系统变形情况下的驱动桥齿轮啮合效率理论模型。杨航[57] 建立了减速器传动效率理论模型, 并采用减速器总成传动效率试验验证了理论模型的准确性。

3.3　齿轮传动机构传动效率理论数值研究

由于齿轮传动机构元件类型相似, 不同型号、类型的齿轮传动机构功率损失理论研究方法基本相同。目前, 有较多学者对行星齿轮机构功率损失展开了大量研究。Fernandes 等[36] 对行星齿轮箱展开理论与试验分析, 建立的齿轮箱功率损失理论模型主要包含齿轮啮合损失模型、轴承摩擦功率损失模

型、搅油损失模型和油封损失模型。Grabovic 等[58]在锥齿轮传动研究中综合考虑了噪声、振动、声振粗糙度(noise vibration harshness, NVH)和摩擦功率损失,以弹流润滑理论为摩擦功率损失分析基础,以负载传输误差表征 NVH 水平,基于多目标优化实现齿轮最优设计。Yue 等[59]重点考虑机械传动过程中动态特性与齿轮摩擦之间的耦合效应,基于六自由度齿轮动力学模型分析了齿轮表面加工质量和载荷对功率损失的影响。

同时,不同类型齿轮箱的传动效率试验分析是理论数值研究的重要基础。陈芳超等[60]建立了车辆齿轮传动系统的齿轮副搅油损失流体力学分析模型,并开展了齿轮副搅油损失试验验证。罗冬源[61]利用多工况实验数据,基于量纲理论法和多元线性回归法建立了齿轮箱功率损失多工况预测模型。范桢[62]运用数值拟合法与量纲分析法推导了齿轮啮合功率损失分析模型。Guo 等[63]基于相似准则、尺寸协调原理和流体力学 π 定理建立了齿轮箱功率损失预测模型,并通过试验修正了相关影响参数系数。Dai 等[64-65]提出引入风差贡献因子和气油比演化模型的理论分析方法来预测单个正齿轮部分浸入油液时的搅动阻力损失。

综上,传动系统效率理论数值研究集中在机械摩擦阻力功率损失和搅动阻力功率损失 2 个方面,主要以摩擦理论、润滑理论、流体理论、齿轮动力学理论等为基础分别建立不同传动部件的功率损失理论分析模型,并基于能量流传输理论综合各部件、各类型功率损失,构建传动系统功率损失理论分析模型。此外,由于传动部件理论数值研究具有较强的针对性和应用性,需通过不同工况的台架试验修正各项影响参数以提高理论模型精度。

4 传动系统效率的建模仿真研究

由于传动系统组成部件的复杂性,理论分析模型的精度需要进行针对性试验修正,通用性有待提升,因此,关于齿轮传动系统功率损失的建模仿真研究发展较快。目前,对车辆传动系统效率建模仿真研究主要集中于齿轮啮合承载特性有限元仿真分析、传动部件润滑系统流态仿真分析、传动部件温度-功率损失耦合仿真分析、传动部件效率虚拟样机仿真分析 4 个方面。

4.1 齿轮啮合承载特性有限元仿真研究

齿轮啮合传动为动态变化过程,图 8[8]和图 9[8]分别为某型号直齿轮的载荷分布状态和齿轮承载接触区域分布示意图。齿面承载接触分析(loaded tooth contact analysis, LTCA)是研究齿轮传动效率

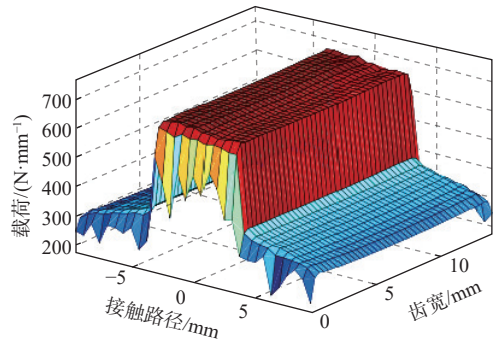


图 8 某型号直齿轮法向载荷分布状态[8]

Fig. 8 Normal load distribution of a certain type of spur gear[8]

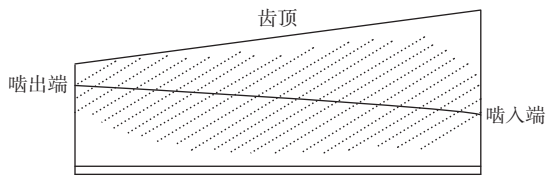


图 9 齿轮承载接触区域分布示意图[8]

Fig. 9 Schematic diagram of bearing contact area of gear[8]

动态特性的重要基础,齿面接触区域变化、齿面接触应力分布、齿间载荷分布、齿轮接触弹性变形等均对齿轮动态啮合功率损失有关键影响。

齿轮传动部件中,不同类型的啮合齿轮接触承载特性不同。王羽达等[66]利用有限元法(finite element method, FEM)分析了斜齿轮啮合对的应力分布状态,能够直接反映齿轮在不同负载工况下的接触特性,如图 10 所示。蒋进科[9,13]、贾超[67]等利用齿面承载接触分析方法,从抗磨损、动态冲击、传递误差等多个方面对齿轮啮合性能进行优化分析,建立了齿轮几何特性与力学特性的映射关系。刘德伟[68]利用 ANSYS 得到变速器斜齿轮静态接触压力分布数值,是对斜齿轮载荷分布理论的有效验证。Ravivarman 等[69]基于 ANSYS 有限元接触模型测量齿轮传动接触载荷,使用接触负荷分担

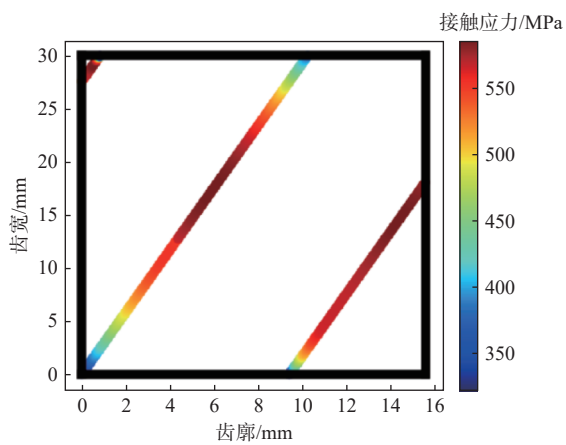


图 10 斜齿轮承载接触应力云图[66]

Fig. 10 Load contact stress nephogram of helical gear[66]

模型更准确地预测了功率损耗。疏奇等[70]通过参数化方式对螺旋锥齿轮进行有限元建模,得到了啮合齿轮的承载接触区域处啮合刚度、齿面、齿间载荷分布、传动误差等变化规律。陈文炜等[71]结合斜齿轮传动承载接触有限元分析与弹流润滑(elastohydrodynamic lubrication, EHL)摩擦系数计算模型,对行星轮系统动态啮合效率进行了计算分析。

因此,通过分析啮合冲击力、传动误差、齿轮磨损状态与承载特性的关系,进一步明确齿轮啮合效率与动力特性之间的耦合关系,对改善齿轮传动的稳定性和耐久性及齿轮动态啮合功率损失研究更具实际意义。

4.2　传动部件润滑系统流态仿真研究

润滑系统状态直接影响元件之间的摩擦作用,且齿轮转动过程中引起的搅油损失特性与润滑油液性质密切相关,可通过增加不同类型添加剂的方式改善特种车辆传动部件在高速重载、高温等极端运行条件下的磨损、疲劳问题。同时,受黏温特性、黏压特性影响的润滑油流态直接关系传动部件的温度场和压力场分布。针对润滑系统流态与齿轮传动部件功率损失的关联分析,目前多采用计算流体力学(computational fluid dynamic, CFD)、移动粒子半隐式法(moving particle semi-implicit, MPS)、光滑粒子流体动力学(smoothed particle hydrodynamics, SPH)法等物理仿真建模分析方法。

在润滑油液 CFD 仿真研究方面,主要应用研究对象有单直齿轮、斜齿轮,单级啮合齿轮,汽车变速器、减速器、驱动桥,飞机螺旋锥齿轮箱、行星轮系等[72-73],图 11[6]为某行星轮系传动部件搅油流态。CFD 建模方法主要基于流体体积法(volume of fluid, VOF)、有限体积法(finite volume method, FVM)和动网格技术,采用 Navier-Stokes 方程、湍流方程和压力隐式算子分裂法(pressure implicit with splitting of operators, PISO)等,建立齿轮箱搅油过程[74]和润滑系统[75]内流域仿真模型。通过分析不同润滑方式搅油流态,如飞溅润滑[76-77]、浸油润滑[78],得到转速转矩、浸油深度、润滑油黏度等因素对搅油特性的影响规律。在不同的搅油流态中,齿轮啮合接触压力、齿轮旋转阻力矩均不同,由此为齿轮啮合效率分析、传动部件搅油损失研究和润滑系统优化提供计算依据。此外,基于该理论还可分析搅动阻力损失中的风阻功率损失,其与搅油阻力损失类似,在齿轮高速旋转状态下占比较大[79-81]。

MPS 法是一种基于拉格朗日粒子的无网格方法,适用于求解不可压缩流体,模拟其自由界面运

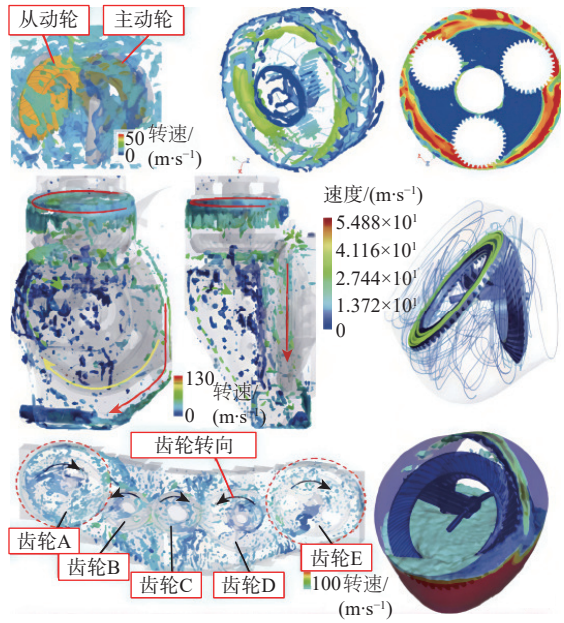


图 11　齿轮传动部件搅油流态 CFD 仿真[6]

Fig. 11　CFD simulation of stirred flow state of gear transmission parts[6]

动和多相流。SPH 法是一种拉格朗日无网格粒子计算方法,其对搅油损失的应用研究与 MPS 法类似[82]。刘娇[83]、Guo 等[84]、王交龙[85]结合齿轮副理论计算模型与 MPS 法建立减速器搅油损失仿真模型,得到了不同搅油流态下的搅油力矩值,分析了齿轮结构类型参数、润滑油特性参数、加油量、减速器倾斜角度、箱体结构变化对搅油损失的影响规律。刘恒龙[86]、谢迟新[87]等利用 MPS 法预测了齿轮箱飞溅润滑情况下的油液流场分布特性和搅油损失值,与 SPH 法相比精确度更高。

对比 CFD 法和 MPS 法,两者均用于分析齿轮箱体内部搅油阻力矩特性。其中,MPS 法仅针对油液流态进行分析,但运算速度较快,在复杂的车辆传动部件中,搅油过程存在油气混合现象,CFD 法可以分析油气两相,因此,目前采用 CFD 法研究搅油损失的较多。

4.3　传动部件的温度-功率损失耦合仿真研究

封闭式齿轮箱的热功率通常建立在各类型功率损耗源与传递至环境热量之间的平衡状态基础上,如图 12 所示。齿轮、轴承、传动轴、万向节类元件高速旋转产热并传热,导致温度变化,进而影响润滑油液性能,齿轮箱总功率损失受润滑系统变化的影响,箱体温度发生变化并反作用于传动部件。其中,功率损失类型主要为 2.2 节中归纳的 5 种,散热机制为通过齿轮箱壳体进行热传导、热对流和热辐射。传动部件的温度-功率损失耦合仿真分析可分为 2 类: 基于热网络模型的功率损失预测和基于热流场有限元仿真的功率损失预测。

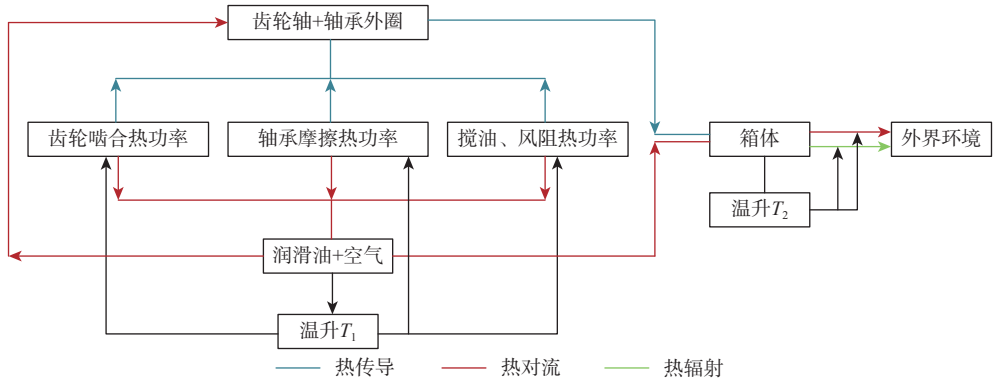


图 12 齿轮传动部件温度-功率损失耦合机理

Fig. 12 Coupling mechanism of temperature-power loss of gear transmission components

热网络法是一种集总参数法，可将传动装置部件转化为热源、热阻和热节点等元件，对于复杂齿轮传动系统是一种全局性的热功率特性分析方法[88]。Martins 等[89]早期通过大量试验对齿轮箱传动效率与热特性进行耦合分析，指出平衡状态下的油浴温度与传动效率是相关的。Changenet 等[90]基于热平衡理论建立了某手动变速器热网络模型，包括传动系统的功率损失、导热热阻和对流换热热阻模型。Koffel 等[91]建立高速旋转齿轮与轴承的热网络数值仿真模型，分析表明，温度分布预测对齿轮箱搅油流体阻力损失十分重要。随后，de Gevigney 等[92]基于 FZG 齿轮箱试验平台建立热网络平衡模型，并应用于车辆变速器功率损失预测。Navet 等[93]进一步结合 FZG 试验数据和热网络模型提出了一种润滑油缺乏时齿轮功率损失预测方法。Liu 等[94]采用热网络法建立了包括效率模型和热平衡模型的双速自动变速器集成模型，可预测不同条件下的功率损耗、体积温度和温度分布。Ning 等[95]重点利用热网络法分析了汽车驱动桥的温度特性，结合温度特性优化给出了提升驱动桥运行稳定性的优化策略。此外，还有学者结合热网络模型和机器学习算法，利用温度特性实现了对不同类型齿轮箱结构的故障检测[96-97]。

热流场的有限元仿真在齿轮或轴承元件中应用较多，能够建立准确的结构模型，但复杂传动部件建模时网格划分较为复杂，对计算机资源要求较高，需对模型进行一定的简化处理。范桢[62]利用 CFD 法建立了变速器润滑数值仿真模型，基于温度场特性、阻力矩变化特性等确定了最优润滑方式。辛相锦[98]利用 ANSYS 对变速器齿轮温度场进行了有限元仿真分析，并指出不同挡位的转速、转矩对齿轮热功率损失影响程度不同。毕旺洋[99]以传热学理论为基础，建立了汽车自动变速器齿轮有限元温度仿真分析模型，将热功率损失计算模型结果作为输入摩擦热，得到了齿轮稳态温度场分布规律。廖建明[100]利用有限元法对变速器温度-功率损失进行耦合分析，以变速器功率损失计算模型为热流场仿真分析提供初始热边界，以热流场仿真结果实时修正功率损失计算模型，得到稳定热流场状态下的功率损失。图 13 为稳定状态下的变速器温度分布特性[100]，相比初始边界条件固定的功率损失理论计算模型精度更高，更适用于模拟实际工程问题，也为优化分析车辆传动部件传动效率提供了新的研究方法。

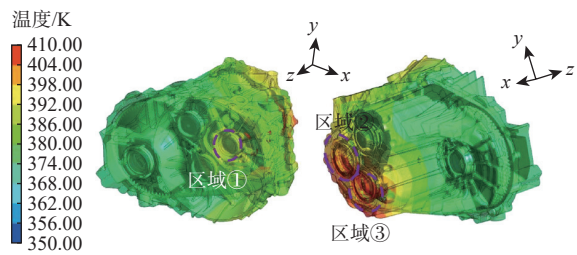


图 13 变速器稳态温度分布云图[100]

Fig. 13 Transmission steady state temperature distribution cloud[100]

4.4 传动部件效率虚拟样机仿真研究

传动部件传动效率仿真研究采用较多的是 RomaxDesigner 软件，其是机械传动系统开发领域用于齿轮传动部件、轴承建模分析的专业软件。对变速器，液力变矩器，驱动桥总成中的减速器、差速器，行星齿轮传动机构均能进行传动效率建模仿真分析。王维[101]、包英豪[102]等基于 RomaxDesigner 建立了机械式变速器的虚拟样机模型，如图 14 所示，包括传动轴、轴承、各挡位齿轮、同步器等关键元件的建模，通过添加功率流、设置循环仿真工况，分析了润滑油温度、挡位、输入转速、输入转矩与变速器综合传动效率之间的关系。吕希砚等[103]运用图论理论基本回路法建立双行星排式动力耦合系统的运动学关系，并基于此进行 RomaxDesigner 传动效率建模分析。张猛[104]利用 RomaxDesigner 建立斜齿轮传动模型，分析了齿轮修形前后的传

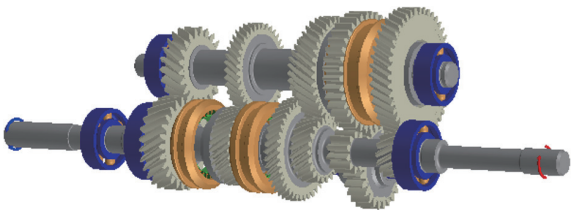


图 14　某型号变速器 RomaxDesigner 仿真模型[102]

Fig. 14　RomaxDesigner simulation model of a certain type of transmission[102]

动效率特性。郑光泽等[105]结合 ParticleWorks 与 RomaxDesigner 软件构建了减速器传动效率分析模型, 分析了齿轮啮合、轴承和油封摩擦的功率损失特性。

同时, Changenet 等[90]利用 ANSYS Workbench 平台建立了齿轮风阻功率损失仿真模型。Salawu 等[106]综合解析法、有限元法和 Comsol Multiphysics 软件建立了齿轮箱传动效率仿真模型。Modelica 开源仿真平台在汽车部件和整车建模中应用较为广泛, 国外学者利用 Modelica 虚拟仿真技术对齿轮箱传动负载功率损失进行了研究[107], 如图 15 所示[108], 主要将车辆的减速器、变速器模型与虚拟试验台架集成, 通过预先调整输入转矩和输出转速工作点仿真分析其阻力损失随负载的变化情况。王扬武等[108]基于 AVL-Cruise 车辆动力性能专业仿真平台建立了基于双离合变速器的整车仿真模型, 依据能量守恒理论研究了变速器在 NEDC 工况下的综合传动效率。

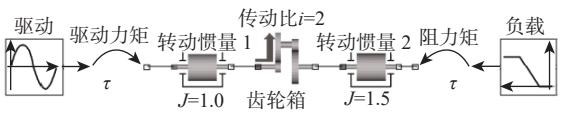


图 15　行星齿轮减速器 Modelica 负载损耗模型[108]

Fig. 15　Planetary gear reducer Modelica load loss model[108]

通过对传动系统功率损失机理建模, 仿真模拟指定范围参数对功率损失的影响特性, 为车辆传动效率与传动系统性能分析搭建了桥梁。对于进一步的整车传动系统效率评估, 则需要结合实际行驶循环工况进行综合效率计算分析。

5　传动系统效率的试验论证研究

汽车传动系统效率试验论证研究主要分为 3 个方面: ①传动系统总成; ②关键传动部件, 如变速器、驱动桥、主减速器等; ③齿轮、轴承类元件。试验类型主要为整车道路试验和台架试验 2 种。无论是理论数值研究还是建模仿真研究, 均需要进行试验论证, 包括试验系统的设计与搭建、试验方案和数据处理方法的论证与分析。检测不同负载和转速下的传动效率, 分析对传动系统效率影响最大的环节, 将影响因素重要度作为传动系统性能优化研究基础。

相比整车道路试验, 台架试验研究针对性更强、控制性高、试验周期短、可重复性强, 因此, 目前针对传动系统效率试验论证研究多采用台架试验。传动系统试验台架分为开式和闭式 2 种类型, 主要有国内的武汉理工大学、吉林大学、重庆大学、重庆理工大学、中北大学等高校团队与国外的密西根科技公司设计开发的车辆传动系统、传动部件的试验台架。同济大学设计的高精度行星减速器综合性能试验平台在对传动效率、功率研究的基础上, 还可对传动误差、寿命、温升、振动、噪声等指标进行测量与计算, 更有利于待测件的性能状态分析[109]。

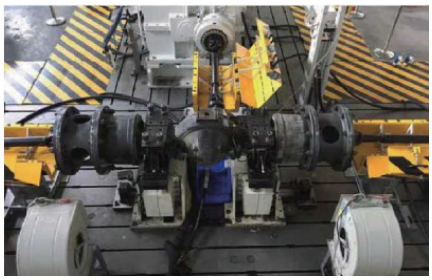
5.1　传动系统总成试验研究

在整车传动效率试验测试研究方面, 王旺平[110]、戴明灿[111]基于中国现有的传动效率试验标准设计了微车传动系统效率测试试验台架和试验方案, 验证了基于功率损失分析的传动效率理论计算模型的可行性。通过底盘测功机进行整车动力性能测试也可检测底盘损耗功率, 是机械底盘传动系统性能评价依据[112]。传动效率试验设计依据包括《汽车工程手册: 试验篇》[113]、《汽车试验学》[114]、《汽车传动轴总成技术条件及台架试验方法》[115]、《汽车机械式变速器总成台架试验方法第 4 部分: 重型》[116]、《汽车半轴技术条件和台架试验方法》[117]、《汽车驱动桥台架试验方法》[118]。其中, 王旺平[110]对试验台架的测量误差与精度做了详细的计算分析, 得到传动系统总效率的测试精度为 0.25%。戴明灿[111]利用正交试验设计理论开发出一套符合国家和行业相关标准的测试方案, 旨在量化研究传动效率影响因素及各因素之间的耦合关系。

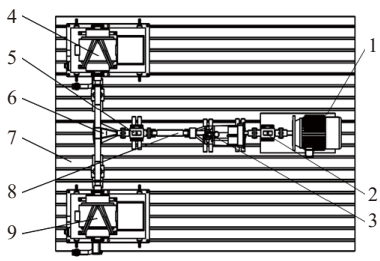
图 16[34,119]为闭式传动系统试验台架, 适用于车辆传动系统总成的功率测定, 主要由驱动设备、待测试传动系统、加载设备和传感器采集设备组成。其中, 拖动测功机和拖动变频器用于模拟发动机驱动设备, 为传动系统输入动力条件, 即转速和负载转矩, 并通过调节变速器挡位模拟车辆不同车速行驶工况。结合布置于变速器输入端和驱动桥半轴输出端的转速、转矩传感器测得相应试验数据, 由此得到不同运行工况下的传动系统效率及功率损失值。此外, 试验过程中应注意控制润滑油温在 (80±5) ℃, 保证润滑系统正常工作。

5.2　传动部件试验研究

车辆传动部件的试验研究相比整车传动系统试验研究较多, 可以利用传动部件试验台架深入分析不同型号部件的传动规律, 为优化内部结构参



(a) 试验台架实物图



1. 拖动测功机; 2, 5. 扭矩转速传感器; 3. 变速器; 4. 加载测功机; 6. 驱动桥; 7. 基础底座; 8. 传动轴; 9. 加载测功机。

(b) 试验台架机械结构

图 16 闭式传动系统试验台架[34,119]

Fig. 16 Closed-loop transmission system test bench[34,119]

数、运行控制策略、性能预测提供数据支撑。试验设备主要为机械式变速器试验台架、驱动桥试验台架、主减速器试验台架和半轴试验台架。主要是液力机械式变速器(hydraulic automatic transmission, AT)、金属带式无级变速器(metallic belt continuously variable transmission, CVT)、机械式变速器 3 种类型变速器的台架试验研究。由于变速器不同挡位下的动力传递路径不同,需分别分析各挡位下变速器传动效率的影响因素和变化规律,并作为理论计算模型的验证或修正依据。董柳杰等[120] 对新能源汽车齿轮箱进行了传动效率台架测试,验证了齿轮修形方法对优化传动效率的可行性。莫易敏等[121] 利用变速器试验台架分析指出,适当增大游隙,通过减小轴承摩擦功率损失可提升变速器传动效率。范文健等[122] 进一步试验分析指出,轴承精度的提高对低挡位变速器传动效率优化效果更佳,轴承密封结构产生的摩擦力矩占轴承总摩擦力矩最大,具有优化空间。

驱动桥传动效率的试验研究方法与变速器类似。占锐等[123] 结合中国重型汽车实际工况设计了重型汽车驱动桥传动效率试验方法。Cruz 等[124] 通过设计试验策略研究后桥齿轮传动空载功率损失,分析了后驱动桥主要部件功率损失的贡献度,确定了导致后桥传动效率低下的原因。王迪等[125] 对飞机传动系统主减速器提出了传动效率直接测试法和间接测试法 2 种试验方法,直接测试法类似于在车辆传动系统试验台架中的测试方法,而间接测试法基于热平衡理论,认为在保温绝热效果良好的条件下,减速器 99% 以上的机械功率损失转化为润滑油的温升,这为试验研究传动部件机械传动效率变化规律提供了新的思路方法。

5.3 齿轮、轴承元件试验研究

在齿轮对啮合效率、齿轮搅油损失、轴承摩擦功率损失试验研究方面,国外学者研究居多。Dindar 等[126] 提出了一种试验方法来分离齿轮箱功率损失的各个组成部分,并建立了基于转速和扭矩的功率损失函数。如图 17 所示,Fernandes 等[127] 利用 FZG 齿轮箱试验台对直齿轮啮合功率损失进行循环测试,重点分析了 5 种不同类型润滑油对齿轮摩擦功率损失的影响。Hammami 等[128] 在 FZG 齿轮箱试验台上测量扭矩损失并进行数值模拟,得到了不同基础油类型、油黏度和添加剂配比对齿轮扭矩损失的影响规律,结果表明,油液物理化学性质对机械摩擦功率损失影响较大,需关注润滑系统健康状态对传动部件运行状态的影响。

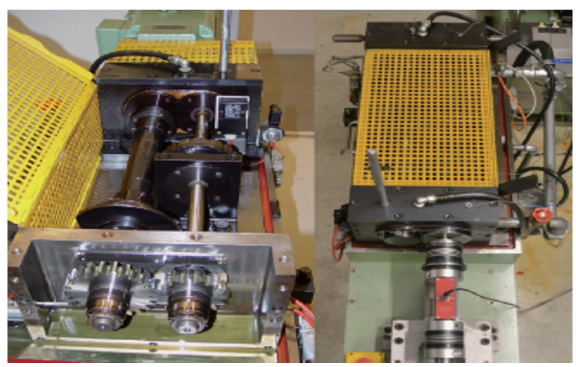


图 17 FZG 齿轮箱试验台[127]

Fig. 17 FZG gearbox test bed[127]

齿轮箱在高速旋转时会产生较大的搅油阻力矩,同样,轴承在润滑油液中也会产生阻力损失。Peterson 等[11] 设计研制了四轴承试验台,用于测量轴承在不同加载工况、不同润滑剂时的摩擦力矩,如图 18 所示,并通过分析完全浸没润滑轴承和最低高度润滑轴承之间的差值确定了流体阻力引起的轴承摩擦力矩。Giannetti 等[129] 试验验证了不同润滑和摩擦条件的滚动轴承与齿轮产生的风阻损失和流体动力损失预测模型,表明功率损失优化研究需考虑润滑系统的作用。Dindar 等[130] 通过试验测量球轴承机械功率损失,评估了大范围径向载荷与轴向载荷作用下球轴承摩擦动力学模型预测机械功率损失的准确性。Yilmaz 等[131] 通过齿轮箱功率损失测试试验分析了浸渍、喷射、最小润滑方式对轴承、齿轮功率损耗和温度行为的影响特性。

国内重庆理工大学团队搭建了大型齿轮箱功率损失多参数试验台架,如图 19 所示[61],能够测量

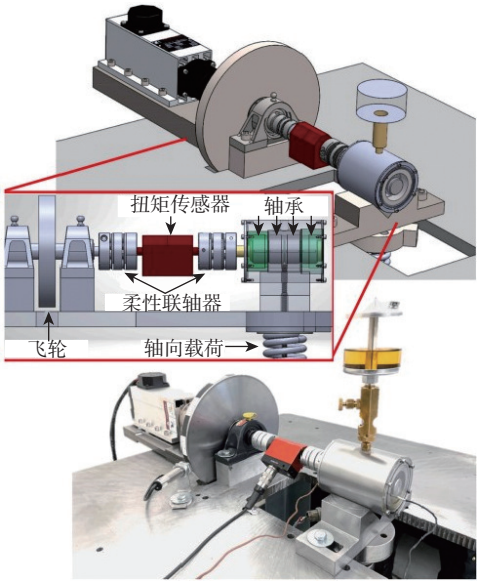
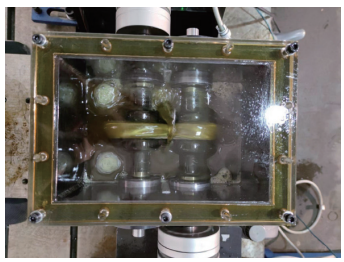
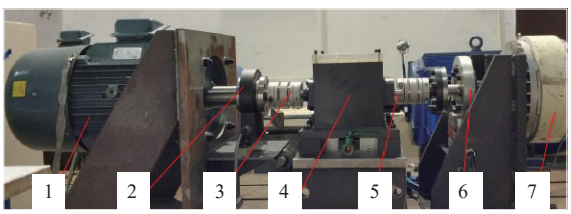


图 18　轴承摩擦力矩损失试验台[11]

Fig. 18　Bearing friction torque loss test bench[11]



(a) 齿轮箱实物图



1. 电机; 2, 6. 转速转矩传感器; 3, 5. 联轴器; 4. 齿轮箱; 7. 负载。

(b) 试验台架

图 19　大型齿轮箱功率损失多参数试验台架[61]

Fig. 19　Large gearbox power loss multi-parameter test bench[61]

不同类型结构参数、工况参数下齿轮箱空载损失与负载损失, 通过制定详细的试验方案, 设计不同参数的试验对象, 全面详细分析了不同转速、模数、传动比、润滑油、螺旋角、浸油深度、齿宽、齿高倍数、变位系数对齿轮箱功率损失的参数影响规律, 为齿轮箱功率损失试验分析提供了系统全面的参考方法。该团队为汽车传动系统减速器、变速器等齿轮箱结构单元的传动效率特性分析提供了较为完备的试验数据支撑。

6　结 束 语

本文从传动系统功率损失特性分析角度, 结合特种车辆在执行作战单元机动任务时对动力性能的要求, 论述了传动系统效率研究对提升传动系统性能的重要性。结合多轴车底盘系统的动力传动机理阐述了传动系统效率建模的思路, 分别从理论数值研究、建模仿真分析、试验论证研究 3 个方面归纳总结了传动系统、传动部件、齿轮轴承元件 3 个层次的传动效率研究方法, 结论如下:

1) 特种车辆传动系统的结构组成特性、动力传递特性、能量流特性与整车动力特性指标密切相关, 可作为建立特种车辆传动性能评估指标体系的分析基础。

2) 多轴传动系统具有布局复杂、传动部件数量多、功率损失类型多的特点, 需结合挡位信息、行驶速度、功率流传递路径等对各驱动轴分别分析, 某一驱动轴功率流传递链路失效即视为整车传动系统功能缺失。多驱动轴功率流传递链路有效数量与整车传动性能、动力性能可靠性之间的关系有待深入研究。

3) 动力传递路径是分析传动系统元件、部件、系统的关联网, 对于部件级的传动效率特性分析可以基于给定的转速、转矩工况, 但对于系统性的传动效率分析, 需以车辆实际运行工况和各传动部件挡位分布情况为依据。

4) 传动系统的功率损失类型具体可分为 5 种, 分别为齿轮啮合功率损失、轴承摩擦功率损失、搅油功率损失、风阻功率损失和油封功率损失。影响功率损失的关键因素可以分为两大类: ①外部施加工况; ②传动部件内部因素。在功率损失影响规律分析时, 多以单一因素线性变化为前提, 但实际内外部影响因素存在耦合关系, 如齿轮转速影响润滑状态从而改变齿面接触摩擦系数, 长时间运转下齿轮表面粗糙度也会受到影响, 因此, 分析耦合因素的不确定性、线性或非线性变化对功率损失的影响规律更具有实际意义。

5) 传动效率的研究方法主要有理论数值研究、建模仿真研究及试验论证研究, 其中, 对传动部件传动效率的理论数值研究与台架试验论证研究较多。试验台架以变速器总成、驱动桥总成、传动轴总成为对象, 涉及车辆底盘的整车道路试验主要与传动系统综合性能评价、NVH 性能优化、传动系统动态特性分析相关。多将传动系统效率同燃油经济性、整车动力性联合分析; 对于单一类型的功率损失, 如齿轮啮合功率损失、轴承摩擦功率损失、齿轮搅油损失, 采用建模仿真研究与齿轮箱试验论证研究方法的居多; 而基于机理分析的理论研究, 对本质上优化传动效率和指导结构设计具有重要意义。

车辆在长期使用过程中不可避免出现故障、老化等问题，可将典型故障特征如磨损、裂纹与传动效率特性联合分析，结合当前深度学习等智能算法的快速发展为建立识别传动系统功能异常或失效的监测系统提供新的研究思路。具体的研究点如下：

1）传动部件因摩擦作用产生大量热量并发生热变形、齿面胶合、齿面点蚀等问题，可深入开展基于温度信号的齿轮故障-温度场-功率损失耦合分析。

2）传动部件随车辆行驶过程中的动力学特性变化（如振动冲击、姿态变换等）均会影响其承载受力和力矩传递能力，可深入开展基于振动信号的动力学特性-功率损失耦合分析。

3）润滑系统失效、油液品质下降会加速部件的磨损与老化，润滑状态也会直接影响齿轮的传动效率，可深入开展基于油液状态的润滑接触-功率损失分析。

参考文献（References）

[1] 余志生. 汽车理论[M]. 5版. 北京: 机械工业出版社, 2009: 35-36.
YU Z S. Automobile theory[M]. 5th ed. Beijing: China Machine Press, 2009: 35-36(in Chinese).

[2] 张鹏程. 液力机械式自动变速器传动效率与整车燃油经济性研究[D]. 太原: 中北大学, 2017: 2-3.
ZHANG P C. Study on transmission efficiency of hydromechanical automatic transmission and fuel economy of vehicle[D]. Taiyuan: North University of China, 2017: 2-3(in Chinese).

[3] 崔志琴, 王晓华, 景银萍, 等. 装甲车辆底盘构造与原理[M]. 北京: 北京理工大学出版社, 2022: 20-21.
CUI Z Q, WANG X H, JING Y P, et al. Structure and principle of armored vehicle's chassis[M]. Beijing: Beijing Insititute of Technology Press, 2022: 20-21(in Chinese).

[4] 彭莫, 周良生, 岳惊涛, 等. 多轴汽车[M]. 北京: 机械工业出版社, 2014: 25-26.
PENG M, ZHOU L S, YUE J T, et al. Multi-axlecar[M]. Beijing: China Machine Press, 2014: 25-26(in Chinese).

[5] 何平. 重型汽车传动系统关键部件能量特性及节能评价研究[D]. 合肥: 合肥工业大学, 2015: 13-14.
HE P. Study on energy characteristics and energy-saving evaluation of key components of heavy-duty vehicle transmission system[D]. Hefei: Hefei University of Technology, 2015: 13-14(in Chinese).

[6] 王斌, 宁斌, 陈辛波, 等. 齿轮传动搅油功率损失的研究进展[J]. 机械工程学报, 2020, 56(23): 1-20.
WANG B, NING B, CHEN X B, et al. Research progress in churning losses of gear transmission[J]. Journal of Mechanical Engineering, 2020, 56(23): 1-20(in Chinese).

[7] 张林, 周建星, 章翔峰, 等. 风电齿轮传动系统弹流润滑特性研究[J]. 太阳能学报, 2022, 43(7): 385-394.
ZHANG L, ZHOU J X, ZHANG X F, et al. Research on elastohydrodynamic lubrication characteristics of gear transmission system in wind turbine[J]. Acta Energiae Solaris Sinica, 2022, 43(7): 385-394(in Chinese).

[8] LIU M Y, HAN X G, YAN C N, et al. Tribological characteristics evaluation of a helical gear pair based on the tribo-dynamic model[J]. Lubrication Science, 2023, 35(8): 616-635.

[9] 蒋进科, 刘钊, 刘红梅. Ease-off修形准双曲面齿轮齿面动态抗磨设计与分析[J]. 机械工程学报, 2021, 57(19): 155-164.
JIANG J K, LIU Z, LIU H M. Dynamic anti-wear design and analysis for hypoid gears with Ease-off flank modification[J]. Journal of Mechanical Engineering, 2021, 57(19): 155-164(in Chinese).

[10] ZHAO J, SHENG W, LI Z, et al. Study on the lubrication characteristics of spur gear pairs with low sliding ratio under mixed elastohydrodynamic lubrication[J]. Journal of Tribology, 2022, 144(7): 071604.

[11] PETERSON W, RUSSELL T, SADEGHI F, et al. Experimental and analytical investigation of fluid drag losses in rolling element bearings[J]. Tribology International, 2021, 161: 107106.

[12] FARRENKOPF F, SCHWARZ A, LOHNER T, et al. Analysis of a low-loss gear geometry using a thermal elastohydrodynamic simulation including mixed lubrication[J]. Lubricants, 2022, 10(9): 200.

[13] 蒋进科, 方宗德, 刘红梅. 行星传动多体齿轮承载接触特性分析[J]. 机械工程学报, 2019, 55(15): 174-182.
JIANG J K, FANG Z D, LIU H M. Loadedtooth contact characteristic analysis of multi-gear for planetary transmission[J]. Journal of Mechanical Engineering, 2019, 55(15): 174-182(in Chinese).

[14] LIANG Y C, WANG S, LI W D, et al. Data-driven anomaly diagnosis for machining processes[J]. Engineering, 2019, 5(4): 646-652.

[15] 王秋莲, 周啸宇, 黎敏, 等. 基于递归分析和机器学习的小批量机械加工过程状态监测[J]. 计算机集成制造系统, 2024, 30(12): 4339-4351.
WANG Q L, ZHOU X Y, LI M, et al. State monitoring of machining process in small batch production mode based on recurrence analysis and machine learning[J]. Computer Integrated Manufacturing Systems, 2024, 30(12): 4339-4351(in Chinese).

[16] WANG Q L, YANG H. Sensor-based recurrence analysis of energy efficiency in machining processes[J]. IEEE Access, 2020, 8: 18326-18336.

[17] 肖海兵. 基于能量耗损的机械设备故障诊断理论与方法研究[D]. 广州: 华南理工大学, 2013: 16-17.
XIAO H B. Research on theory and method of mechanical equipment fault diagnosis based on energy consumption[D]. Guangzhou: South China University of Technology, 2013: 16-17(in Chinese).

[18] 谢小鹏, 肖海兵, 冯伟, 等. 基于局部线性嵌入的能量耗损故障模式识别[J]. 华南理工大学学报 (自然科学版), 2012, 40(12): 1-6.
XIE X P, XIAO H B, FENG W, et al. Fault pattern recognition of energy loss based on locally linear embedding[J]. Journal of South China University of Technology (Natural Science Edition), 2012, 40(12): 1-6(in Chinese).

[19] GU Y J, XING Y. Online monitoring of wind turbine operation efficiency and optimization based on benchmark values[J]. IEEE Access, 2019, 7: 132193-132204.

[20] DING H, LI H P, CHEN S Y, et al. Energy loss and mechanical efficiency forecasting model for aero-engine bevel gear power

transmission[J]. International Journal of Mechanical Sciences, 2022, 231: 107569.

[21] 周良生, 曲学春, 周冬生, 等. 汽车动力传动系及动力性能计算[M]. 北京: 机械工业出版社, 2020: 15-16.
ZHOU L S, QU X C, ZHOU D S, et al. Automobile power train and dynamic performance calculation[M]. Beijing: China Machine Press, 2020: 15-16(in Chinese).

[22] 华琛, 牛润新, 余彪. 地面车辆机动性评估方法与应用[J]. 吉林大学学报 (工学版), 2022, 52(6): 1229-1244.
HUA C, NIU R X, YU B. Methods and applications of ground vehicle mobility evaluation[J]. Journal of Jilin University (Engineering and Technology Edition), 2022, 52(6): 1229-1244(in Chinese).

[23] 张金豹, 邹天刚, 高秀才, 等. 基于中位秩加权的车辆传动装置直驶模式传递效率性能评估[J]. 兵工学报, 2022, 43(增刊 1): 54-59.
ZHANG J B, ZOU T G, GAO X C, et al. Straight driving efficiency assessment of vehicle transmission device based on the-median-rank weighting[J]. Acta Armamentarii, 2022, 43(Sup 1): 54-59(in Chinese).

[24] DONG P, ZHAO J W, XU X Y, et al. Performance comparison of series-parallel hybrid transmissions with multiple gears and modes based on efficiency model[J]. Energy Conversion and Management, 2022, 274: 116442.

[25] LEI F, BAI Y C, ZHU W H, et al. A novel approach for electric powertrain optimization considering vehicle power performance, energy consumption and ride comfort[J]. Energy, 2019, 167: 1040-1050.

[26] HAN X H, HE C, ZHENG F Y, et al. Transmission performance evaluation of non-circular spur bevel gear based on a novel isochronal measurement method[J]. Mechanism and Machine Theory, 2023, 185: 105329.

[27] 吴兆雨. 汽车传动系统振动分析及参数优化[D]. 武汉: 武汉理工大学, 2020: 10-12.
WU Z Y. Vibration analysis and parameter optimization of automobile transmission system[D]. Wuhan: Wuhan University of Technology, 2020: 10-12(in Chinese).

[28] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 汽车动力性台架试验方法和评价指标: GB/T 18276—2017[S]. 北京: 中国标准出版社, 2017.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Test-bed methods and evaluating index of dynamic property of motor vehicles: GB/T 18276—2017[S]. Beijing: Standards Press of China, 2017(in Chinese).

[29] 张广庆, 汪开鑫, 肖茂华, 等. 基于液压机械传动扭矩比的 HMCVT 稳态传动效率研究[J]. 农业机械学报, 2021, 52(增刊 1): 533-541.
ZHANG G Q, WANG K X, XIAO M H, et al. Research on steady-state transmission efficiency of HMCVT based on torque ratio of hydro-mechanical transmission[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(Sup 1): 533-541(in Chinese).

[30] 陈文彬, 李晓阳, 童邦安, 等. 谐波减速器的传动效率确信可靠性建模与分析[J]. 振动工程学报, 2022, 35(1): 237-245.
CHEN W B, LI X Y, TONG B A, et al. Belief reliability modeling and analysis for transmission efficiency of harmonic gear[J]. Journal of Vibration Engineering, 2022, 35(1): 237-245(in Chinese).

[31] 章德平, 莫易敏, 高勇, 等. 微型汽车驱动桥传动效率的试验研究[J]. 广西大学学报 (自然科学版), 2019, 44(3): 613-620.
ZHANG D P, MO Y M, GAO Y, et al. Experimental study on transmission efficiency of mini-car's drive axle[J]. Journal of Guangxi University (Natural Science Edition), 2019, 44(3): 613-620 (in Chinese).

[32] 王旺平. 微型汽车传动系统效率建模及试验研究[D]. 武汉: 武汉理工大学, 2015: 12-13.
WANG W P. Efficiency modeling and experimental study of mini-car transmission system[D]. Wuhan: Wuhan University of Technology, 2015: 12-13(in Chinese).

[33] 田蜜. 基于减少功率损失的微型汽车传动系统参数优化[D]. 武汉: 武汉理工大学, 2012: 24.
TIAN M. Parameter optimization of mini-car transmission system based on reducing power loss[D]. Wuhan: Wuhan University of Technology, 2012: 24(in Chinese).

[34] 王熙. 基于传动系统效率的汽车燃油经济性研究[D]. 重庆: 重庆大学, 2010: 2-3.
WANG X. Research on automobile fuel economy based on transmission system efficiency[D]. Chongqing: Chongqing University, 2010: 2-3(in Chinese).

[35] HONG I, ANESHANSLEY E, CHAUDHURY K, et al. Stochastic microcontact model for the prediction of gear mechanical power loss[J]. Tribology International, 2023, 183: 108413.

[36] FERNANDES C M C G, MARQUES P M T, MARTINS R C, et al. Gearbox power loss. Part II: Friction losses in gears[J]. Tribology International, 2015, 88: 309-316.

[37] FERNANDES C M C G, MARQUES P M T, MARTINS R C, et al. Gearbox power loss. Part III: Application to a parallel axis and a planetary gearbox[J]. Tribology International, 2015, 88: 317-326.

[38] 李庚. 某微型汽车机械式变速器传动效率分析与研究[D]. 武汉: 武汉理工大学, 2017: 37-40.
LI G. Analysis and research on transmission efficiency of mechanical transmission of a mini-car[D]. Wuhan: Wuhan University of Technology, 2017: 37-40(in Chinese).

[39] 王成. 2K-H 型封闭式周转轮系功率流与传动效率计算[J]. 机械工程学报, 2023, 59(13): 59-67.
WANG C. Calculation of power flow and transmission efficiency of 2K-H enclosed revolving gear train[J]. Journal of Mechanical Engineering, 2023, 59(13): 59-67(in Chinese).

[40] 柴蓉, 张明柱, 王建华, 等. 离合器带排转矩对液压机械无级变速器效率的影响分析[J]. 机械传动, 2022, 46(4): 63-67.
CHAI R, ZHANG M Z, WANG J H, et al. Analysis of the influence of clutchdrag torque on the efficiency of hydro-mechanical continuously variable transmission[J]. Journal of Mechanical Transmission, 2022, 46(4): 63-67(in Chinese).

[41] 冯能莲, 米磊. 活齿无级变速器传动效率分析与试验研究[J]. 北京工业大学学报, 2016, 42(10): 1547-1551.
FENG N L, MI L. Analysis and experimental investigations of transmission efficiency of movable-teeth CVT[J]. Journal of Beijing University of Technology, 2016, 42(10): 1547-1551(in Chinese).

[42] 李优华, 李权才, 刘忠明. 低速重载齿轮箱传动效率分析与测试

[J]. 机械传动, 2017, 41(7): 111-115.
LI Y H, LI Q C, LIU Z M. Analysis and test of the transmission efficiency of low speed and heavy duty gearbox[J]. Journal of Mechanical Transmission, 2017, 41(7): 111-115(in Chinese).

[43] 王熙, 秦大同, 胡明辉, 等. 汽车变速器传动效率理论建模与实验测试[J]. 机械传动, 2010, 34(8): 21-24.
WANG X, QIN D T, HU M H, et al. Modeling and testing of the vehicle transmission efficiency[J]. Journal of Mechanical Transmission, 2010, 34(8): 21-24(in Chinese).

[44] ROSANDER P, BEDNAREK G, SEETHARAMAN S, et al. Development of an efficiency model for manual transmissions[J]. ATZ Worldwide, 2008, 110(4): 36-43.

[45] 陈柯序, 李海波, 赵小娟, 等. 基于BP神经网络的纯电动汽车动力传动系统效率建模及分析[J]. 机械设计与研究, 2021, 37(5): 180-185.
CHEN K X, LI H B, ZHAO X J, et al. Modeling and analysis of the drivetrain efficiency for pure electric vehicles based on BP neural network[J]. Machine Design & Research, 2021, 37(5): 180-185(in Chinese).

[46] AUTIERO M, CIRELLI M, PAOLI G, et al. A data-driven approach to estimate the power loss and thermal behaviour of cylindrical gearboxes under transient operating conditions[J]. Lubricants, 2023, 11(7): 303.

[47] KARABACAK Y E, BAŞ H S. Experimental investigation of efficiency of worm gears and modeling of power loss through artificial neural networks[J]. Measurement, 2022, 202: 111756.

[48] FOTSO H R F, KAZÉ C V A, KENMOÉG D. Real-time rolling bearing power loss in wind turbine gearbox modeling and prediction based on calculations and artificial neural network[J]. Tribology International, 2021, 163: 107171.

[49] BARDAY D, FOSSIER C, CHANGENET C, et al. Investigations on drive axle thermal behaviour: power loss and heat-transfer estimations[J]. SAE International Journal of Engines, 2018, 11(1): 55-66.

[50] KOLIVAND M, LI S, KAHRAMAN A. Prediction of mechanical gear mesh efficiency of hypoid gear pairs[J]. Mechanism and Machine Theory, 2010, 45(11): 1568-1582.

[51] XU H, KAHRAMAN A, ANDERSON N E, et al. Prediction of mechanical efficiency of parallel-axis gear pairs[J]. Journal of Mechanical Design, 2007, 129(1): 58-68.

[52] XU H, KAHRAMAN A. Prediction of friction-related power losses of hypoid gear pairs[J]. Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics, 2007, 221(3): 387-400.

[53] WANG W P, MO Y M. Test design on minicar transmission system efficiency using dimensional analysis[J]. Applied Mechanics and Materials, 2012, 192: 222-226.

[54] 贾富淳. 面向效率的高速电驱动系统传动装置润滑特性分析与优化[D]. 长春: 吉林大学, 2021: 17.
JIA F C. Analysis and optimization of lubrication characteristics of transmission device in efficiency-oriented high-speed electric drive system[D]. Changchun: Jilin University, 2021: 17(in Chinese).

[55] 朱建杏. 船用减速器功率损失研究[D]. 上海: 上海海洋大学, 2021: 9-14.
ZHU J X. Study on power loss of marine reducer[D]. Shanghai: Shanghai Ocean University, 2021: 9-14(in Chinese).

[56] 王钦, 贺迪, 桂良进, 等. 考虑系统变形的电驱动桥齿轮啮合效率研究[J]. 机械工程学报, 2023, 59(3): 66-75.
WANG Q, HE D, GUI L J, et al. Study of gear meshing efficiency of electric drive axle with considering system deformation[J]. Journal of Mechanical Engineering, 2023, 59(3): 66-75(in Chinese).

[57] 杨航. 考虑减速器NVH性能的传动效率分析优化[D]. 重庆: 重庆理工大学, 2020: 9-16.
YANG H. Analysis and optimization of transmission efficiency considering NVH performance of reducer[D]. Chongqing: Chongqing University of Technology, 2020: 9-16(in Chinese).

[58] GRABOVIC E, ARTONI A, GABICCINI M, et al. Exploration of trade-offs between NVH and efficiency in bevel gear design[J]. Forschung ImIngenieurwesen, 2023, 87(3): 933-947.

[59] YUE K, KANG Z, ZHANG M J, et al. Study on gear meshing power loss calculation considering the coupling effect of friction and dynamic characteristics[J]. Tribology International, 2023, 183: 108378.

[60] 郭栋, 陈芳超, 刘骄, 等. 齿轮副高速搅油阻力矩理论计算与试验研究[J]. 机械工程学报, 2021, 57(1): 49-60.
GUO D, CHEN F C, LIU J, et al. Theoretical and experimental study of oil churning resistance torque of high-speed gear pair[J]. Journal of Mechanical Engineering, 2021, 57(1): 49-60(in Chinese).

[61] 罗冬源. 齿轮箱功率损失的参数影响规律分析及实验研究[D]. 重庆: 重庆理工大学, 2022: 47.
LUO D Y.Analysis and experimental study on parameter influence law of gearbox power loss[D].Chongqing: Chongqing University of Technology,2022: 47(in Chinese).

[62] 范桢. 电动汽车变速箱功率损失及热流场分析与验证[D]. 福州: 福建工程学院, 2022: 9.
FAN Z. Analysis and verification of power loss and thermal flow field of electric vehicle gearbox[D]. Fuzhou: Fujian Institute of Engineering, 2022: 9(in Chinese).

[63] GUO D, WEN G B, WANG Y W, et al. A theoretical and experimental study on the power loss of gearbox based on dimensionless analysis[J]. Journal of Tribology, 2023, 145(10): 101701.

[64] DAI Y, MA F, ZHU X, et al. Development of an analytical model to estimate the churning power losses of a spiral bevel gear[J]. Tribology International, 2020, 151: 106536.

[65] DAI Y, ZHANG Y Y, ZHU X. Generalized analytical model for evaluating the gear power losses transition changing from windage to churning behavior[J]. Tribology International, 2023, 185: 108572.

[66] 王羽达, 吕宏展, 侯志勇, 等. 渐开线直齿轮啮合接触计算与分析[J]. 机械传动, 2021, 45(12): 41-47.
WANG Y D, LÜ H Z, HOU Z Y, et al. Calculation and analysis of meshing contact of involute spur gear[J]. Journal of Mechanical Transmission, 2021, 45(12): 41-47(in Chinese).

[67] 贾超, 方宗德. 高速齿轮传递误差和啮入冲击的激励模拟及齿面优化修形[J]. 振动与冲击, 2019, 38(23): 103-109.
JIA C, FANG Z D. Simulation for transmission error and mesh-in impact excitation of high speed gears and their tooth surface optimal modification[J]. Journal of Vibration and Shock, 2019, 38(23): 103-109(in Chinese).

[68] 刘德伟. 某轻型车用变速器传动效率分析计算[D]. 重庆: 重庆大学, 2017: 57.
LIU D W. Analysis and calculation of transmission efficiency of a light vehicle transmission[D]. Chongqing: Chongqing University, 2017: 57(in Chinese).

[69] RAVIVARMAN R, SEKAR R P. Load share model based gear loss factor prediction in high contact ratio spur gear drive[J]. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2023, 237(1): 103-118.

[70] 陆凤霞, 疏奇, 戈文昌, 等. 考虑安装误差的弧齿锥齿轮接触分析初始点计算方法[J]. 机械制造与自动化, 2020, 49(6): 106-109.
LU F X, SHU Q, GE W C, et al. Approach to determination of initial point for tooth contact analysis of spiral bevel gears in consideration of installation errors[J]. Machine Building & Automation, 2020, 49(6): 106-109(in Chinese).

[71] 陆凤霞, 陈文炜, 刘伟平, 等. 斜齿轮轮齿接触有限元分析的新方法研究[J]. 机械传动, 2018, 42(7): 38-43.
LU F X, CHEN W W, LIU W P, et al. Study on the new method of finite element analysis of helical gear contact[J]. Journal of Mechanical Transmission, 2018, 42(7): 38-43(in Chinese).

[72] HU X Z, WANG A, LI P P, et al. Influence of dynamic attitudes on oil supply for bearings and churning power losses in a splash lubricated spiral bevel gearbox[J]. Tribology International, 2021, 159: 106951.

[73] HU X Z, JIANG Y Y, LUO C, et al. Churning power losses of a gearbox with spiral bevel geared transmission[J]. Tribology International, 2019, 129: 398-406.

[74] MASTRONE M N, HARTONO E A, CHERNORAY V, et al. Oil distribution and churning losses of gearboxes: experimental and numerical analysis[J]. Tribology International, 2020, 151: 106496.

[75] 夏兴隆. 纯电动汽车变速器内部流场分析及其润滑优化[D]. 长春: 吉林大学, 2019: 42-45.
XIA X L. Internal flow field analysis and lubrication optimization of pure electric vehicle transmission[D]. Changchun: Jilin University, 2019: 42-45(in Chinese).

[76] DAI Y, CHEN X, YANG D, et al. Impulse power loss in orthogonal face gear-numerical and experimental results[J]. The European Physical Journal Plus, 2023, 138(1): 88.

[77] HU X Z, LI P P, WU M G. Influence of the dynamic motion of a splash-lubricated gearbox on churning power losses[J]. Energies, 2019, 12(17): 3225.

[78] LIU H, JURKSCHAT T, LOHNER T, et al. Detailed investigations on the oil flow in dip-lubricated gearboxes by the finite volume CFD method[J]. Lubricants, 2018, 6(2): 47.

[79] 陈涛, 周金阳, 高旭文, 等. 高速齿轮风阻功率损失的参数化仿真计算[J]. 热能动力工程, 2022, 37(增刊 1): 31-36.
CHEN T, ZHOU J Y, GAO X W, et al. Parametric simulation calculation of wind resistance power loss of high-speed gears[J]. Journal of Engineering for Thermal Energy and Power, 2022, 37(Sup 1): 31-36(in Chinese).

[80] ZHU X, DAI Y, MA F Y. On the estimation of the windage power losses of spiral bevel gears: an analytical model and CFD investigation[J]. Simulation Modelling Practice and Theory, 2021, 110: 102334.

[81] DAI Y, XU L, ZHU X, et al. Application of an unstructured overset method for predicting the gear windage power losses[J]. Engineering Applications of Computational Fluid Mechanics, 2021, 15(1): 130-141.

[82] LIU H, ARFAOUI G, STANIC M, et al. Numerical modelling of oil distribution and churning gear power losses of gearboxes by smoothed particle hydrodynamics[J]. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2019, 233(1): 74-86.

[83] 刘骄. 齿轮副搅油损失仿真分析及试验研究[D]. 重庆: 重庆理工大学, 2019: 12-19.
LIU J. Simulation analysis and experimental study on oil stirring loss of gear pair[D]. Chongqing: Chongqing University of Technology, 2019: 12-19(in Chinese).

[84] GUO D, CHEN F C, LIU J, et al. Numerical modeling of churning power loss of gear system based on moving particle method[J]. Tribology Transactions, 2020, 63(1): 182-193.

[85] 王交龙. 减速箱齿轮搅油损失仿真与实验研究[D]. 镇江: 江苏大学, 2021: 30-34.
WANG J L. Simulation and experimental study on oil stirring loss ofreducer gear[D]. Zhenjiang: Jiangsu University, 2021: 30-34(in Chinese).

[86] 刘桓龙, 谢迟新, 李大法, 等. 齿轮箱飞溅润滑流场分布和搅油力矩损失[J]. 浙江大学学报 (工学版), 2021, 55(5): 875-886.
LIU H L, XIE C X, LI D F, et al. Flow field distribution of splash lubrication of gearbox and churning gear torque loss[J]. Journal of Zhejiang University (Engineering Science), 2021, 55(5): 875-886 (in Chinese).

[87] 谢迟新, 刘桓龙, 贾瑞河, 等. 基于 MPS 方法的二级齿轮箱飞溅润滑特性研究[J]. 中国机械工程, 2021, 32(15): 1827-1835.
XIE C X, LIU H L, JIA R H, et al. Research on splash lubrication characteristics of two-stage gearboxes based on MPS method[J]. China Mechanical Engineering, 2021, 32(15): 1827-1835(in Chinese).

[88] LU Y B, LU X N, YE G, et al. Thermal failure analysis of gear transmission system[J]. Journal of Failure Analysis and Prevention, 2024, 24(2): 520-537.

[89] MARTINS R C, CARDOSO N F R, BOCK H, et al. Power loss performance of high pressure nitrided steel gears[J]. Tribology International, 2009, 42(11-12): 1807-1815.

[90] CHANGENET C, OVIEDO-MARLOT X, VELEX P. Power loss predictions in geared transmissions using thermal networks-applications to a six-speed manual gearbox[J]. Journal of Mechanical Design, 2006, 128(3): 618-625.

[91] KOFFEL G, VILLE F, CHANGENET C, et al. Investigations on the power losses and thermal effects in gear transmissions[J]. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2009, 223(3): 469-479.

[92] DE GEVIGNEY J D, CHANGENET C, VILLE F, et al. Thermal modelling of a back-to-back gearbox test machine: application to the FZG test rig[J]. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2012, 226(6): 501-515.

[93] NAVET P, CHANGENET C, VILLE F, et al. Thermal modeling of the FZG test rig: application to starved lubrication conditions[J]. Tribology Transactions, 2020, 63(6): 1135-1146.

[94] LIU Y G, PENG J Y, WANG B, et al. Bulk temperature prediction of a two-speed automatic transmission for electric vehicles using thermal network method and experimental validation[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2019, 233(10): 2585-2598.

[95] NING X F, CHEN M Z, ZHOU Z J, et al. Thermal analysis of automobile drive axles by the thermal network method[J]. World Electric Vehicle Journal, 2022, 13(5): 75.

[96] CORLEY B, KOUKOURA S, CARROLL J, et al. Combination of thermal modelling and machine learning approaches for fault detection in wind turbine gearboxes[J]. Energies, 2021, 14(5): 1375.

[97] CORLEY B, CARROLL J, MCDONALD A. Fault detection of wind turbine gearbox using thermal network modelling and SCADA data[J]. Journal of Physics: Conference Series, 2020, 1618(2): 022042.

[98] 辛相锦. 拖拉机动力换挡变速器润滑系统流动与传热仿真研究[D]. 长春: 吉林大学, 2021: 26-35.
XIN X J. Simulation study on flow and heat transfer in lubrication system of tractor power shift transmission[D]. Changchun: Jilin University, 2021: 26-35(in Chinese).

[99] 毕旺洋. 汽车自动变速器斜齿轮啮合热功率损失及温度场研究[D]. 天津: 河北工业大学, 2021: 41-48.
BI W Y. Study on thermal power loss and temperature field of helical gear meshing in automobile automatic transmission[D]. Tianjin: Hebei University of Technology, 2021: 41-48(in Chinese).

[100] 廖建明. 汽车变速器功率损失及热流场仿真研究[D]. 重庆: 重庆大学, 2019: 69-75.
LIAO J M. Simulation study on power loss and heat flow field of automobile transmission[D]. Chongqing: Chongqing University, 2019: 69-75(in Chinese).

[101] 王维, 回春, 汪洋, 等. 基于Romax软件的变速器传动效率仿真分析与试验研究[J]. 机械传动, 2017, 41(10): 178-184
ANG W, HUI C, WANG Y, et al. Simulation analysis and experimental research of the transmission efficiency of mechanical transmission based on Romax software[J]. Journal of Mechanical Transmission, 2017,41(10): 178-184(in Chinese).

[102] 包英豪,蔡文奇,彭显昌,等.电动车用两档变速器结构优化及效率分析[C]//2018 中国汽车工程学会年会.北京:机械工业出版社,2018: 876-881.
BAO Y H, CAI W Q, PENG X C, et al. Structural optimization and efficiency analysis of two-speed transmissions for electric vehicles[C]//2018 Proceedings of the Annual Meeting of the Chinese Society of Automotive Engineering. Beijing: Mechanical Industry Press, 2018: 876-881(in Chinese).

[103] 米林, 吕希砚, 谭伟. 双行星排式混合动力耦合机构效率特性分析[J]. 重庆理工大学学报 (自然科学), 2021, 35(4): 23-31.
MI L, LYU X Y, TAN W. Analysis of efficiency characteristics of double planet row hybrid coupling mechanism[J]. Journal of Chongqing University of Technology (Natural Science), 2021, 35(4): 23-31(in Chinese).

[104] 张猛. 高铁齿轮箱轮齿修形及传动效率研究[D]. 太原: 太原理工大学, 2020: 43-54.
ZHANG M. Study on gear tooth modification and transmission efficiency of high-speed rail gearbox[D]. Taiyuan: Taiyuan University of Technology, 2020: 43-54(in Chinese).

[105] 郑光泽, 杨航, 郝涛, 等. 齿轮搅油功率损失与减速器传动效率分析[J]. 机械传动, 2020,44(12): 49-54.
ZHENG G Z, YANG H, HAO T, et al. Analysis of gear oil churning power loss and transmission efficiency of reducer[J]. Journal of Mechanical Transmission, 2020, 44(12): 49-54(in Chinese).

[106] SALAWU E Y, AJAYI O O, OKENIYI J O, et al. On the prediction of power loss in helical gearbox via simulation approach[J]. Cleaner Engineering and Technology, 2021, 3: 100128.

[107] SCHLEGEL C, HÖSL A, DIEL S. Detailed loss modelling of vehicle gearboxes[C]//Proceedings of the 7th International Modelica Conference. Linköping: Linköping University Electronic Press, 2009: 434-443.

[108] 王扬武, 崔东伟, 张书光, 等. 关于汽车变速器NEDC工况综合效率的仿真计算[C]//第十七届河南省汽车工程科技学术研讨会. 郑州: 河南省汽车工程学会, 2020: 145-147.
WANG Y W, CUI D W, ZHANG S G, et al. Simulation calculation on the comprehensive efficiency of NEDC working condition of automotive transmission[C]//Proceedings of the 17th Henan Province Automotive Engineering Science and Technology Symposium. Zhengzhou: Henan Automotive Engineering Society, 2020: 145-147(in Chinese).

[109] 郑钰馨, 奚鹰, 袁浪, 等. RV减速器动力性能综合检测试验平台设计[J]. 中国工程机械学报, 2017, 15(6): 536-541.
ZHENG Y X, XI Y, YUAN L, et al. The design of testbed for RV reducer's dynamic characteristics comprehensive testing[J]. Chinese Journal of Construction Machinery, 2017, 15(6): 536-541(in Chinese).

[110] WANG W P. Efficiency modeling and experimental verification of minicar transmission system[J]. Applied Mechanics and Materials, 2013, 288: 30-36.

[111] 戴明灿. 微型汽车传动系统效率试验设计与研究[D]. 武汉: 武汉理工大学, 2012: 7-17.
DAI M C. Design and research on efficiency test of mini-car transmission system[D]. Wuhan: Wuhan University of Technology, 2012: 7-17(in Chinese).

[112] 赵玮, 王强, 李艳峰, 等. 基于底盘测功机的轮式工程机械底盘动力性能测试[J]. 中国工程机械学报, 2011, 9(3): 347-350.
ZHAO W, WANG Q, LI Y F, et al. Dynamical property testing on wheel construction machinery chasses using chassis dynamometer[J]. Chinese Journal of Construction Machinery, 2011, 9(3): 347-350(in Chinese).

[113] 《汽车工程手册》委员会. 汽车工程手册: 试验篇[M]. 北京: 人民交通出版社, 2001: 40-44.
Automotive Engineering Manual Committee.Automotive engineering manual: test article[M]. Beijing: People's Communications Publishing House, 2001: 40-44(in Chinese).

[114] 郭应时, 袁伟. 汽车试验学[M]. 3版. 北京: 人民交通出版社, 2022: 50-52.
GUO Y S, YUAN W. Automobile experimental science[M]. 3rd ed. Beijing: China Communications Press, 2022: 50-52(in Chinese).

[115] 中华人民共和国工业和信息化部. 汽车传动轴总成技术条件及台架试验方法: QC/T 29082—2019[S]. 北京: 北京科学技术出版社, 2019.
Ministry of Industry and Information Technology of the People's Republic of China. Technical conditions and bench test methods of

automobile transmission shaft assembly: QC/T 29082—2019[S]. Beijing: Beijing Science & Technology Press, 2019(in Chinese).

[116] 中华人民共和国工业和信息化部. 汽车机械式变速器总成台架试验方法 第 4 部分: 重型: QC/T 568.4—2011[S]. 北京: 全国汽车标准化技术委员会, 2011.
Ministry of Industry and Information Technology of the People's Republic of China. Automotive mechanical transmission assemblies-bench test methods-Part 4: Heavy duty: QC/T 568.4—2011 [S]. Beijing: National Technical Committee on Automobile Standardization, 2011(in Chinese).

[117] 中华人民共和国工业和信息化部. 汽车半轴技术条件和台架试验方法: QC/T 293—2019[S]. 北京: 北京科学技术出版社, 2019.
Ministry of Industry and Information Technology of the People's Republic of China. Technical conditions and bench test methods for automobile half-shafts: QC/T 293—2019[S]. Beijing: Beijing Science & Technology Press, 2019(in Chinese).

[118] 中华人民共和国机械工业部. 汽车驱动桥台架试验方法: QC/T 533—1999[S]. 北京: 全国汽车标准化技术委员会, 1999.
Ministry of Machinery Industry of the People's Republic of China. Benchtest method for automobile drive axle: QC/T 533—1999[S]. Beijing: National Technical Committee on Automobile Standardization, 1999(in Chinese).

[119] 姚哲皓, 刘金, 刘华军, 等. 商用车驱动桥传动效率分析与计算[C]//2021 中国汽车工程学会年会. 北京: 机械工业出版社, 2021: 23-28.
YAO Z H, LIU J, LIU H J, etal. Analysis and calculation of drive axle efficiency of commercial vehicles[C]//Proceedings of the Annual Meeting of the Chinese Society of Automotive Engineering 2021. Beijing: China Machine Press, 2021: 23-28(in Chinese).

[120] 董柳杰, 赵航, 杨钰洁, 等. 新能源汽车齿轮箱齿轮修形设计及效率分析[J]. 机床与液压, 2022, 50(17): 77-81.
DONG L J, ZHAO H, YANG Y J, et al. Gear modification design and efficiency analysis for new energy automobile gearbox[J]. Machine Tool & Hydraulics, 2022, 50(17): 77-81(in Chinese).

[121] 莫易敏, 徐伟, 胡杰, 等. 轴承游隙对变速器传动效率的影响研究[J]. 机械传动, 2018,42(10): 53-57.
MO Y M, XU W, HU J, et al. Research of the influence of bearing clearance on the transmission efficiency of gearbox[J]. Journal of Mechanical Transmission, 2018,42(10): 53-57(in Chinese).

[122] 范文健, 孔祥旭, 毛万鑫, 等. 轴承对 MT 变速器传动效率的影响及优化研究[J]. 机械传动, 2022, 46(7): 113-120.
FAN W J, KONG X K, MAO W X, et al. Study on the influence of bearings on transmission efficiency of manual transmission and optimization[J]. Journal of Mechanical Transmission, 2022, 46(7): 113-120(in Chinese).

[123] 占锐, 程华国, 李俊, 等. 重型汽车驱动桥传动效率试验台及方法研究[J]. 机械传动, 2017, 41(3): 197-202.
ZHAN R, CHENG H G, LI J, et al. Research of the method and test bench of drive axle transmission efficiency for heavy vehicle[J]. Journal of Mechanical Transmission, 2017, 41(3): 197-202 (in Chinese).

[124] CRUZ J A O, MARQUES P M T, SEABRA J H O, et al. No-load power loss of a rear axle gear transmission: measurement and validation[J]. Journal of Tribology, 2022, 144(9): 091202.

[125] 王迪, 夏海纯, 倪德, 等. 直升机主减速器传动效率试验方法[J]. 航空动力学报, 2020, 35(12): 2673-2680.
WANG D, XIA H C, NI D, et al. Experiment methods of transmission efficiency for helicopter main gearbox[J]. Journal of Aerospace Power, 2020, 35(12): 2673-2680(in Chinese).

[126] DINDAR A, CHAUDHURY K, HONG I, et al. An experimental methodology to determine components of power losses of a gearbox[J]. Journal of Tribology, 2021, 143(11): 111203.

[127] FERNANDES C, MARTINS R, SEABRA J, et al. FZG gearboxes lubricated with different formulations of polyalphaolefin wind turbine gear oils[C]//Proceedings of the International Gear Conference. Abington: Woodhead Publishing Limited, 2014: 699-708.

[128] HAMMAMI M, FERNANDES C M C G, MARTINS R, et al. Torque loss in FZG-A10 gears lubricated with axle oils[J]. Tribology International, 2019, 131: 112-127.

[129] GIANNETTI G, MELI E, RINDI A. Modelling and experimental study of power losses in toothed wheels[J]. Sensors, 2023, 23(12): 5541.

[130] DINDAR A, CHIMANPURE A S, KAHRAMAN A. Mechanical power losses of ball bearings: model and experimental validation [J]. Journal of Tribology, 2022, 144(5): 051603.

[131] YILMAZ M, ÖNÜT A, LOHNER T, et al. Gear and bearing power losses: from dip to minimum quantity lubrication[J]. Industrial Lubrication and Tribology, 2022, 74(9): 985-994.

Research progress on transmission performance of special vehicle based on power loss characteristics analysis

GAO Qinhe[1,*], GAO Lei[1], LIU Zhihao[1], WANG Dong[1], MA Dong[1], ZHANG Yibo[2]

(1. School of Missile Engineering, Rocket Force University of Engineering, Xi'an 710025, China;
2. School of Highway, Chang'an University, Xi'an 710054, China)

Abstract: A vehicle transmission system is a power transmission device between the engine and wheel load. As the connection hub of the power system, walking system and braking system, it ensures the mobility and safety of special equipment in combat tasks. Firstly, it was demonstrated that transmission efficiency was an important technical index to evaluate the transmission system performance from four aspects: structure characteristics, power transmission characteristics, energy flow characteristics and vehicle dynamic characteristics. Second, the transmission system's global power loss model framework was developed from the standpoint of generation mechanism and structural features based on the relationship analysis between power loss and transmission efficiency. This gave the transmission system performance characterization a theoretical foundation. Then, the research difficulties and emphases of vehicle transmission system power loss were summarized from three aspects: theoretical numerical study, simulation study and experimental demonstration study. It was noted that the experimental study of large mechanical equipment under complex operating conditions and the simulation analysis of multi-media and multi-parameter couplings can offer guidance for the matching design of the transmission system, performance evaluation, and health management of vehicle transmission components. Finally, it is considered that the comprehensive transmission efficiency analysis under multi-factors and the coupling research of power loss characteristics, efficiency characteristics and performance degradation characteristics based on intelligent algorithms have practical engineering significance for the performance monitoring and evaluation of large machinery equipment.

Keywords: special vehicle; transmission system; power loss characteristic; transmission efficiency; transmission performance monitoring and evaluation

Received: 2023-06-28; **Accepted**: 2023-10-13; **Published Online**: 2023-10-26 10:26

URL: link.cnki.net/urlid/11.2625.V.20231025.1705.004

Foundation items: National Natural Science Foundation of China (51905541); Natural Science Basic Research Program of Shaanxi (2020JQ487);Young Talent Fund of University Association for Science and Technology in Shaanxi, China (20190412)

* **Corresponding author**. E-mail: lycoris_lei@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0314

基于迭代最优阻抗的人机物理交互控制方法

刘微容*，魏子丰，晋振兵，孟家豪，王星琨，张浩琛
(兰州理工大学 电气工程与信息工程学院，兰州 730050)

摘 要：为提高人机物理交互的准确性和柔顺性，实现最优交互性能，针对基于迭代学习的阻抗控制方法需要多次重复同一任务的问题，借鉴迭代最优控制无需系统矩阵信息即可优化代价函数确定系统最优控制输入的机制，提出了基于迭代最优阻抗的人机物理交互控制方法。方法采用双环控制结构。面向任务的外环设计了迭代最优阻抗控制器 (IOIC)，将求取最优阻抗参数的问题描述成线性二次型调节器问题，利用迭代最优控制，求取最优反馈增益，使包括轨迹跟踪误差和交互力在内的代价函数最小化；同时引入软辅助函数，避免参数突变可能带来的机器人抖动问题。面向机器人的内环设计了非奇异终端滑模轨迹跟踪控制器 (NTSMTC)，使机器人实际轨迹跟踪外环输出的阻抗轨迹，通过饱和函数消减控制律的抖振。仿真结果证明：所提方法在人机协作任务中，仅利用一次任务初始阶段的交互信息即可求得最优阻抗参数，使任务过程中的轨迹跟踪误差和交互人所消耗的力最小化。

关 键 词：人机物理交互；阻抗控制；轨迹跟踪控制；最优控制；滑模控制
中图分类号：TP242
文献标志码：A **文章编号**：1001-5965(2025)06-1843-09

近年来，工业机器人已经成功应用于执行各类高精度重复性任务[1]。然而，仍有一些任务由于工序复杂、交互环境结构性差，导致机器人无法单独完成，需要人对机器人运动进行指导和调整[2]。在这种情况下，机器人和人不可避免地相互接触，形成人机耦合的动力学系统来协同完成任务[3]。因此，人机物理交互系统受到越来越多的研究人员关注。

为保证交互过程的安全、高效，机器人与人接触时，不仅需要控制交互位置，还需要控制交互力。阻抗控制因其能够调节交互位置和力之间的关系[4]，在人机物理交互系统中受到了广泛的研究与应用。但是，在执行任务的过程中，由于交互人的随机性，难以用固定的阻抗参数满足所有的交互需求，特别是面对具有变化和不确定性的交互对象[5]。

通过模拟人类特有的、可根据交互对象不同而主动调节肢体阻抗的属性，研究人员提出了多种变阻抗控制方法[6-9]，通过实时改变虚拟阻抗参数，提高人机交互系统的性能和安全性。Liu 等[10] 提出了一种基于速度的无传感器变阻抗控制方法，利用末端执行器速度和交互力在线调整阻抗参数，使末端执行器速度更符合人的行为，以提高人机物理交互过程中的顺应性。Grafakos 等[11] 提出了一种变导纳控制器，通过测量人体肌电信号，实时调节阻抗控制器的阻尼参数，以提高交互性能。然而，上述方法并不能在阻抗参数的变化过程中保证交互系统的稳定性，因此，基于稳定性的变阻抗方法受到广泛关注。Dong 等[12] 针对模型不确定的机器人提出了基于不确定性和扰动估计器 (uncertainty and disturbance estimator, UDE) 的变阻抗方法，利用

收稿日期：2023-06-05；**录用日期**：2023-10-16；**网络出版时间**：2023-11-24 14:33
网络出版地址：link.cnki.net/urlid/11.2625.V.20231123.1034.002
基金项目：国家自然科学基金 (62261032)；中央引导地方科技发展资金项目 (25ZYJA026)；甘肃省重点人才项目；甘肃省高等学校创新基金 (2022A-021)
* **通信作者**. E-mail：liuwr@lut.edu.cn

引用格式：*刘微容，魏子丰，晋振兵，等. 基于迭代最优阻抗的人机物理交互控制方法 [J]. 北京航空航天大学学报，2025，51（6）：1843-1851. LIU W R，WEI Z F，JIN Z B，et al. Human-robot physical interaction control method based on iterative optimal impedance [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（6）：1843-1851（in Chinese）.*

UDE 逼近不确定的被控模型，同时为阻抗参数的变化提供了稳定性条件，且该稳定性条件可在执行任务前离线验证。虽然基于稳定性的变阻抗控制方法可以保证交互过程的稳定性，但基于此设计的变阻抗策略往往比较保守，柔顺交互性能受到了较大的限制。总体而言，各类变阻抗控制方法虽然能够通过实时改变阻抗参数实现机器人与人的高效交互，提高了交互性能，但不能保证实现最优的交互性能[13]。

受到人类自主学习时通过积累先前失败经验，反复迭代练习从而提高对此技能掌控能力的启发，基于迭代学习的阻抗控制方法被提出。在每次执行任务过程中，利用先验信息，迭代更新阻抗参数，直至最优。Yang 等[14] 受到点对点运动的人机交互实验启发，基于迭代学习方法，设计包括前馈控制律和阻抗增益在内的阻抗参数迭代更新规则。Yang 等[15] 针对人体动力学时变难以建模的问题，利用期望轨迹和人体阻抗的空间周期性特征，提出一种空间重复阻抗学习控制策略，用于提高人机交互性能。然而，基于迭代学习的阻抗控制方法需要在相同的持续时间和初始条件下多次重复执行同一任务，在实际应用中，难以满足此严格条件[16]。

为实现最优交互性能，基于最优的阻抗控制同样受到研究人员的广泛关注。因为阻抗控制既需要控制力，又需要控制位置，所以，最优阻抗的思想是：权衡力和位置跟踪二者之间的关系[17]，定义包括力和位置跟踪误差在内的代价函数或奖励函数描述实际交互性能，通过最小化代价函数或最大化奖励函数求取最优的阻抗参数。Johansson 等[18] 针对环境模型已知的情况，利用线性二次型调节器（linear quadratic regulator, LQR），通过求解代数 Riccati 方程，确定最优阻抗参数。在此基础上，Matinfar 等[19] 考虑机器人与环境系统的动态特性，提出一类新的代价函数，不同于上述方法，采用在线求解 LQR 的方法求取最优阻抗参数。但是，上述方法均假设环境模型已知，因为 LQR 方法无法对参数未知的线性模型进行求取。而自适应动态规划（adaptive dynamic programming, ADP）和强化学习（reinforcement learning, RL）方法的提出，为解决上述问题提供了思路。Huang 等[20] 将机器人和环境的交互系统视为线性未知模型，利用策略迭代的方法，在线求取最优阻抗参数，实现了机器人与人的最优交互，使轨迹跟踪误差和交互力最小化。Guo 等[21] 利用相同的思路，设计了一种基于 Q-学习算法的自适应阻抗控制器，实现阻抗自适应，保证在人肢体模型动态完全未知的情况下，阻抗参数收敛到最优值，但在每次阻抗参数更新后，仍需重新收集数据进行计算，并未从根本上解决需要重复执行任务的问题。

基于上述讨论，本文提出了基于迭代最优阻抗的人机物理交互控制方法，仅利用执行 1 次任务时初始阶段的部分信息，通过迭代求取最优反馈增益，获得最优阻抗模型，明显减少了数据收集次数，节省了数据收集与计算的时间；在实际应用中，提高了系统运行效率和响应速度。同时，引入软辅助函数消减参数突变问题。利用非奇异终端滑模控制使机器人实际轨迹跟踪阻抗轨迹，并通过饱和函数消减控制力矩的抖振问题，实现了确保轨迹跟踪精度的同时最小化交互力。

1 控制系统方案

基于迭代最优阻抗的人机物理交互控制方法的结构如图 1 所示。该方法采用双环控制结构，由阻抗控制外环和轨迹跟踪内环构成。

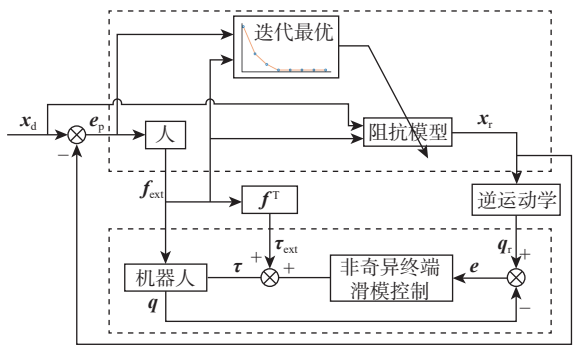


图 1 基于迭代最优阻抗的人机物理交互控制方法结构

Fig. 1 Structure of human-robot physical interaction control method based on iterative optimal impedance

外环面向任务，利用期望轨迹跟踪误差 $\boldsymbol{e}_{\mathrm{p}}$、人与机器人的交互力 $\boldsymbol{f}_{\mathrm{ext}}$ 作为迭代最优阻抗控制器（iterative optimal impedance controller, IOIC）的输入，求取最优阻抗参数。根据期望轨迹 $\boldsymbol{x}_{\mathrm{d}}$ 和交互力 $\boldsymbol{f}_{\mathrm{ext}}$ 得到阻抗轨迹 $\boldsymbol{x}_{\mathrm{r}}$，使得机器人的运动可以确保人以最小的工作量协助机器人跟踪期望轨迹，从而实现最优交互性能。

内环面向机器人，利用非奇异终端滑模轨迹跟踪控制器（nonsingular terminal sliding mode trajectory tracking controller, NTSMTC），使机器人的运动响应与所设计阻抗模型一致。

2 外环：迭代最优阻抗控制器

机器人与人交互和与环境交互不同。在交互力方面，环境只能根据机器人的运动被动改变交互力，而人能根据经验和感官对机器人当前所执行的

任务进行指导, 主动改变交互力的大小。因此, 与人交互的机器人需要调节自身阻抗参数, 以适应不同人员的技能水平差异, 与人共同完成指定任务。本节设计了 IOIC 以保证期望轨迹跟踪性能的同时减少人在任务过程中消耗的力。

2.1　人机物理交互系统

为构建轨迹跟踪误差和交互力之间的关系, 建立如下机器人阻抗模型:

$$\boldsymbol{M}_{\mathrm{d}}\ddot{\boldsymbol{e}}_{\mathrm{p}}+\boldsymbol{B}_{\mathrm{d}}\dot{\boldsymbol{e}}_{\mathrm{p}}+\boldsymbol{K}_{\mathrm{d}}\boldsymbol{e}_{\mathrm{p}}=-\boldsymbol{f}_{\mathrm{ext}} \tag{1}$$

式中: $\boldsymbol{M}_{\mathrm{d}}$、$\boldsymbol{B}_{\mathrm{d}}$和$\boldsymbol{K}_{\mathrm{d}}$分别为期望质量、阻尼和刚度参数; $\boldsymbol{f}_{\mathrm{ext}}$为交互力; $\boldsymbol{e}_{\mathrm{p}}$为期望轨迹跟踪误差。

人机协作型任务中, 人的动力学模型可建立为[22]

$$\boldsymbol{K}_{\mathrm{r}}\dot{\boldsymbol{f}}_{\mathrm{ext}}+\boldsymbol{K}_{\mathrm{f}}\boldsymbol{f}_{\mathrm{ext}}=\boldsymbol{K}_{\mathrm{e}}\boldsymbol{e}_{\mathrm{p}} \tag{2}$$

式中: $\boldsymbol{K}_{\mathrm{r}}$、$\boldsymbol{K}_{\mathrm{f}}$和$\boldsymbol{K}_{\mathrm{e}}$为未知增益, 表示人对所执行任务的熟练度。

从式(2)可以看出, 人对机器人施加的力是基于对任务完成度的判断。

为构建人机物理交互系统的状态空间方程, 基于上述模型和控制目标, 选取系统的状态$\boldsymbol{X}$为

$$\boldsymbol{X}=[\boldsymbol{e}_{\mathrm{p}}\quad\dot{\boldsymbol{e}}_{\mathrm{p}}\quad\boldsymbol{f}_{\mathrm{ext}}]^{\mathrm{T}} \tag{3}$$

则人机物理交互系统的状态方程可写为

$$\dot{\boldsymbol{X}}=\begin{bmatrix}\boldsymbol{0}&\boldsymbol{I}&\boldsymbol{0}\\ \boldsymbol{0}&\boldsymbol{0}&\boldsymbol{0}\\ \boldsymbol{K}_{\mathrm{r}}^{-1}\boldsymbol{K}_{\mathrm{e}}&\boldsymbol{0}&-\boldsymbol{K}_{\mathrm{r}}^{-1}\boldsymbol{K}_{\mathrm{f}}\end{bmatrix}\boldsymbol{X}+\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}\boldsymbol{u}_{\mathrm{e}}=\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}\boldsymbol{u}_{\mathrm{e}} \tag{4}$$

式中: $\boldsymbol{I}$为单位矩阵; $\boldsymbol{A}\in\mathbf{R}^{p\times p}$为状态矩阵; $\boldsymbol{B}\in\mathbf{R}^{p\times r}$为输入矩阵; $\boldsymbol{u}_{\mathrm{e}}\in\mathbf{R}^{r}$为交互系统的控制输入, 表示为

$$\boldsymbol{u}_{\mathrm{e}}=\ddot{\boldsymbol{e}}_{\mathrm{p}}=-\boldsymbol{M}_{\mathrm{d}}^{-1}\boldsymbol{B}_{\mathrm{d}}\dot{\boldsymbol{e}}_{\mathrm{p}}-\boldsymbol{M}_{\mathrm{d}}^{-1}\boldsymbol{K}_{\mathrm{d}}\boldsymbol{e}_{\mathrm{p}}-\boldsymbol{M}_{\mathrm{d}}^{-1}\boldsymbol{f}_{\mathrm{ext}}=-\boldsymbol{K}\boldsymbol{X} \tag{5}$$

其中: $\boldsymbol{K}=[\boldsymbol{M}_{\mathrm{d}}^{-1}\boldsymbol{K}_{\mathrm{d}}\quad\boldsymbol{M}_{\mathrm{d}}^{-1}\boldsymbol{B}\quad\boldsymbol{M}_{\mathrm{d}}^{-1}]$为反馈增益。

因此, 对于式(5)描述的人机物理交互系统, 求取最优阻抗参数$\boldsymbol{M}_{\mathrm{d}}^{*}$、$\boldsymbol{B}_{\mathrm{d}}^{*}$和$\boldsymbol{K}_{\mathrm{d}}^{*}$, 即求取最优反馈增益$\boldsymbol{K}^{*}$。

该问题可看作是一个 LQR 问题, 但系统中的状态矩阵$\boldsymbol{A}$包含了人体模型中的未知参数$\boldsymbol{K}_{\mathrm{r}}$、$\boldsymbol{K}_{\mathrm{f}}$和$\boldsymbol{K}_{\mathrm{e}}$, 考虑到 LQR 方法无法求解含有未知参数的线性模型, 借鉴 Jiang 等[23]所提迭代最优控制方法, 可以无需系统模型, 仅利用部分状态信息和输入信息, 即可优化代价函数, 确定系统最优控制输入的机制, 设计 IOIC。

2.2　迭代最优阻抗控制器设计

为实现最小化期望轨迹跟踪误差和交互力的控制目标, 构建代价函数L为

$$L=\int_{t_0}^{\infty}\left([\boldsymbol{e}_{\mathrm{p}}\quad\dot{\boldsymbol{e}}_{\mathrm{p}}\quad\boldsymbol{f}_{\mathrm{ext}}]\boldsymbol{Q}\begin{bmatrix}\boldsymbol{e}_{\mathrm{p}}\\ \dot{\boldsymbol{e}}_{\mathrm{p}}\\ \boldsymbol{f}_{\mathrm{ext}}\end{bmatrix}+\boldsymbol{u}_{\mathrm{e}}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{u}_{\mathrm{e}}\right)\mathrm{d}t=\int_{t_0}^{\infty}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{X}+\boldsymbol{u}_{\mathrm{e}}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{u}_{\mathrm{e}})\mathrm{d}t \tag{6}$$

式中: $\boldsymbol{Q}=\boldsymbol{Q}^{\mathrm{T}}>\boldsymbol{0}$为期望轨迹跟踪误差和交互力的权重; $\boldsymbol{R}=\boldsymbol{R}^{\mathrm{T}}>\boldsymbol{0}$为与期望轨迹跟踪误差和交互力呈线性依赖关系的控制输入的权重。

通过设计最优的控制输入$\boldsymbol{u}_{\mathrm{e}}^{*}$, 使代价函数$L$最小, 即可同时使轨迹跟踪误差和交互力最小, 实现最优交互性能。

Jiang 等[23]所提迭代最优控制方法通过收集一段时间内的状态和输入信息, 并经过采样, 构建成足够多的数据矩阵, 实现仅利用在该时间段内收集的数据, 迭代计算得出近似最优反馈增益矩阵。

因此, 定义如式 (7) 所示的 3 个算子, 用于在人机物理交互初始阶段$[0,t]$时间内, 收集状态$\boldsymbol{X}$和控制输入$\boldsymbol{u}_{\mathrm{e}}$的信息。

$$\begin{cases}\boldsymbol{\kappa}_{XX}=\left[\boldsymbol{X}^{\mathrm{T}}\otimes\boldsymbol{X}^{\mathrm{T}}\big|_{0}^{t_1},\boldsymbol{X}^{\mathrm{T}}\otimes\boldsymbol{X}^{\mathrm{T}}\big|_{t_1}^{t_2},\cdots,\boldsymbol{X}^{\mathrm{T}}\otimes\boldsymbol{X}^{\mathrm{T}}\big|_{t_{h-1}}^{t}\right]\\ \boldsymbol{\delta}_{XX}=\left[\int_{0}^{t_1}\boldsymbol{X}^{\mathrm{T}}\otimes\boldsymbol{X}^{\mathrm{T}}\mathrm{d}t,\int_{t_1}^{t_2}\boldsymbol{X}^{\mathrm{T}}\otimes\boldsymbol{X}^{\mathrm{T}}\mathrm{d}t,\cdots,\int_{t_{h-1}}^{t}\boldsymbol{X}^{\mathrm{T}}\otimes\boldsymbol{X}^{\mathrm{T}}\mathrm{d}t\right]\\ \boldsymbol{\delta}_{Xu}=\left[\int_{0}^{t_1}\boldsymbol{X}^{\mathrm{T}}\otimes\boldsymbol{u}_{\mathrm{e}}^{\mathrm{T}}\mathrm{d}t,\int_{t_1}^{t_2}\boldsymbol{X}^{\mathrm{T}}\otimes\boldsymbol{u}_{\mathrm{e}}^{\mathrm{T}}\mathrm{d}t,\cdots,\int_{t_{h-1}}^{t}\boldsymbol{X}^{\mathrm{T}}\otimes\boldsymbol{u}_{\mathrm{e}}^{\mathrm{T}}\mathrm{d}t\right]\end{cases} \tag{7}$$

式中: $\boldsymbol{\kappa}_{XX}$、$\boldsymbol{\delta}_{XX}$、$\boldsymbol{\delta}_{Xu}$为数据储存矩阵; “$\otimes$”表示克罗内克乘积; h为$[0,t_0]$时间段的采样个数。

在此数据收集阶段, 为保证收集到的数据丰富且非奇异, 设定输入为

$$\boldsymbol{u}_{\mathrm{e}}=-\boldsymbol{K}_0\boldsymbol{X}+\boldsymbol{v}_{\mathrm{n}} \tag{8}$$

式中: $\boldsymbol{K}_0$为使系统 (4) 稳定的初始反馈增益; $\boldsymbol{v}_{\mathrm{n}}$为满足持续激励的噪声函数。

计算$\boldsymbol{\kappa}_{XX}$、$\boldsymbol{\delta}_{XX}$和$\boldsymbol{\delta}_{Xu}$, 直到满足$\mathrm{rank}([\boldsymbol{\delta}_{XX},\boldsymbol{\delta}_{Xu}])=p(p+1)/2+pr$。

求解式 (9) 得到$\boldsymbol{P}_k$和$\boldsymbol{K}_{k+1}$:

$$\begin{bmatrix}\mathrm{vec}(\boldsymbol{P}_k)\\ \mathrm{vec}(\boldsymbol{K}_{k+1})\end{bmatrix}=\left(\boldsymbol{\Gamma}_k^{\mathrm{T}}\boldsymbol{\Gamma}_k\right)^{-1}\boldsymbol{\Gamma}_k^{\mathrm{T}}\boldsymbol{\Theta}_k \tag{9}$$

式中: $\mathrm{vec}(\boldsymbol{P}_k)$和$\mathrm{vec}(\boldsymbol{K}_{k+1})$表示待求解项, $\mathrm{vec}(\cdot)$表示矩阵向量化, $\boldsymbol{P}_k$为李雅普诺夫方程的对称正定解; $\boldsymbol{\Gamma}_k=[\boldsymbol{\kappa}_{XX}\quad-2\boldsymbol{\delta}_{XX}(\boldsymbol{I}\otimes\boldsymbol{K}_k^{\mathrm{T}}\boldsymbol{R})-2\boldsymbol{\delta}_{Xu}(\boldsymbol{I}\otimes\boldsymbol{R})]$为可在线测量的状态信息; $\boldsymbol{\Theta}_k=\boldsymbol{\delta}_{XX}\mathrm{vec}(\boldsymbol{Q}_k)$为输入信息, $\boldsymbol{Q}_k=\boldsymbol{Q}+\boldsymbol{K}_k^{\mathrm{T}}\boldsymbol{R}\boldsymbol{K}_k$。

通过不断重复迭代计算式 (9), 直到$\|\boldsymbol{P}_k-\boldsymbol{P}_{k-1}\|\leqslant\varepsilon$, ε为很小的正实数, 即可得到最优反馈增益的近似值$\boldsymbol{K}_{k+1}$, 从而计算得到最优阻抗参数。

特别注意的是, 当迭代结束得到最优反馈增益$\boldsymbol{K}_{k+1}$后, 系统的反馈增益会从初始的$\boldsymbol{K}_0$突变至

$\boldsymbol{K}_{k+1}$。这种突变会对系统控制输入$\boldsymbol{u}_{\mathrm{e}}$产生影响，使得加速度发生突变，可能导致机器人在运动过程中发生抖动，这种抖动在机器人的控制过程中是不允许存在的。

因此，为解决由于参数突变带来的问题，设计如下软辅助函数：

$$\boldsymbol{K}' = \frac{\boldsymbol{K}_{k+1}+\boldsymbol{K}_0}{2} - \frac{\boldsymbol{K}_{k+1}-\boldsymbol{K}_0}{2}\cos\left(\frac{\pi t}{T_{\mathrm{s}}}\right) \tag{10}$$

式中：$T_{\mathrm{s}} = T_1 - T_2$为软辅助函数的周期。

软辅助函数的变化曲线如图2所示。可以看到，为避免参数突变，设计的软辅助函数利用三角函数平滑连续的曲线特性，使初始反馈增益在时间$[T_1, T_2]$间隔内平滑、单调地变化到近似最优反馈增益$\boldsymbol{K}_{k+1}$。

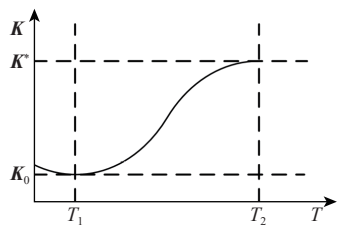


图2 软辅助函数曲线

Fig. 2 Soft auxiliary function curves

进一步，根据上述设计，本文所提IOIC的控制流程如下。

输入：初始反馈增益$\boldsymbol{K}_0$，权重$\boldsymbol{Q}$和$\boldsymbol{R}$，阈值ε，探测噪声$\boldsymbol{v}_{\mathrm{n}}$。

1. 根据选定的初始反馈增益$\boldsymbol{K}_0$和探测噪声$\boldsymbol{v}_{\mathrm{n}}$，设置机器人的初始阻抗增益，同时设置初始的系统输入$\boldsymbol{u}_{\mathrm{e}}$。当人与机器人开始交互后的$[0,t]$时间内，收集系统状态$\boldsymbol{e}_{\mathrm{p}}$、$\dot{\boldsymbol{e}}_{\mathrm{p}}$和$\boldsymbol{f}_{\mathrm{ext}}$信息，以及系统输入$\boldsymbol{u}_{\mathrm{e}}$信息。利用收集到的信息建立矩阵$\boldsymbol{\kappa}_{XX}$、$\boldsymbol{\delta}_{XX}$和$\boldsymbol{\delta}_{Xu}$，直到$\boldsymbol{\Gamma}_k$列满秩。

2. 利用式(9)，求解$\boldsymbol{P}_k$和$\boldsymbol{K}_{k+1}$。

3. 令$k \to k+1$，重复步骤2，直到$\|\boldsymbol{P}_k - \boldsymbol{P}_{k-1}\| \leqslant \varepsilon$为止。

4. 取$\boldsymbol{K}_{k+1}$为近似最优控制策略。

5. 在时间$[t, t+T_{\mathrm{s}}]$内，利用式(10)，求取由软辅助函数过渡的反馈增益$\boldsymbol{K}'$，使得反馈增益平滑地变化至最优。

输出：近似最优反馈增益$\boldsymbol{K}_{k+1}$，求取最优的阻抗增益$\boldsymbol{M}_{\mathrm{d}}^*$、$\boldsymbol{B}_{\mathrm{d}}^*$和$\boldsymbol{K}_{\mathrm{d}}^*$。

3 内环：非奇异终端滑模轨迹跟踪控制器

在第2节已完成外环交互控制设计的基础上，还需要进一步完成内环机器人的轨迹跟踪控制设计。因此，本节设计NTSMTC以保证机器人能跟踪外环输出的阻抗轨迹，从而使机器人的实际运动表现出外环阻抗控制器预设的阻抗行为。

3.1 非奇异终端滑模轨迹跟踪控制器设计

机器人和人的交互任务发生在笛卡儿空间，而机器人的控制发生在关节空间中，因此，机器人在关节空间的动力学方程可写为

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{G}(\boldsymbol{q}) = \boldsymbol{\tau} + \boldsymbol{\tau}_{\mathrm{ext}} + \boldsymbol{\tau}_{\mathrm{d}} \tag{11}$$

式中：$\boldsymbol{q}$、$\dot{\boldsymbol{q}}$、$\ddot{\boldsymbol{q}} \in \mathbf{R}^n$分别为机器人的实际关节角、关节角速度和关节角加速度；$\boldsymbol{M}(\boldsymbol{q}) \in \mathbf{R}^{n\times n}$为正定对称的惯性矩阵；$\boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}}) \in \mathbf{R}^{n\times n}$为哥氏离心力矩阵；$\boldsymbol{G}(\boldsymbol{q}) \in \mathbf{R}^n$为重力向量；$\boldsymbol{\tau} \in \mathbf{R}^n$为控制输入力矩；$\boldsymbol{\tau}_{\mathrm{ext}} = \boldsymbol{J}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{ext}} \in \mathbf{R}^n$为操作人员与机器人在关节空间中的交互力；$\boldsymbol{\tau}_{\mathrm{d}} \in \mathbf{R}^n$为外部干扰力矩向量，且假设$\|\boldsymbol{\tau}_{\mathrm{d}}\| \leqslant \eta$。

通过逆运动学将外环的阻抗轨迹$\boldsymbol{x}_{\mathrm{r}}$转换为关节空间中的阻抗轨迹$\boldsymbol{q}_{\mathrm{r}}$，设计滑模控制律使得机器人的关节位置$\boldsymbol{q}$跟踪$\boldsymbol{q}_{\mathrm{r}}$，再通过正运动学实现机器人的实际位置$\boldsymbol{x}$跟踪$\boldsymbol{x}_{\mathrm{r}}$。

为实现轨迹跟踪控制，定义阻抗轨迹位置跟踪误差为$\boldsymbol{e} = \boldsymbol{q}_{\mathrm{r}} - \boldsymbol{q}$，速度跟踪误差可定义为$\dot{\boldsymbol{e}} = \dot{\boldsymbol{q}}_{\mathrm{r}} - \dot{\boldsymbol{q}}$。根据机器人的动力学方程(11)，可以得到

$$\ddot{\boldsymbol{e}} = \boldsymbol{M}_0^{-1}(\boldsymbol{q})(\boldsymbol{\tau} - \boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})\boldsymbol{q} - \boldsymbol{G}_0(\boldsymbol{q}) + \boldsymbol{\tau}_{\mathrm{ext}} + \boldsymbol{\tau}_{\mathrm{d}}) - \ddot{\boldsymbol{q}}_{\mathrm{r}} \tag{12}$$

根据所定义的位置跟踪误差及速度跟踪误差，可设计如下非奇异终端滑模面：

$$\boldsymbol{s} = \dot{\boldsymbol{e}} + \boldsymbol{\lambda}_1\boldsymbol{e} + \boldsymbol{\lambda}_2\phi(\boldsymbol{e}) \tag{13}$$

式中：$\boldsymbol{\lambda}_j = \mathrm{diag}(\lambda_{j1},\cdots,\lambda_{jn}), j = 1,2$，为正定的对角矩阵；$\phi(\boldsymbol{e}) = [\phi(e(1)),\cdots,\phi(e(n))]^{\mathrm{T}}$为关于跟踪误差的函数，具体可以表示为

$$\phi(e(i)) = \begin{cases} e(i)^{b/a} & R_i = 0 \text{或} R_i \neq 0, |\boldsymbol{e}(i)| \geqslant \mu \\ \sigma_1 e(i) + \sigma_2 e(i)^2 \mathrm{sign}(e(i)) & R_i \neq 0, |\boldsymbol{e}(i)| < \mu \end{cases} \tag{14}$$

式中：$R_i = \dot{e}(i) + \lambda_{1i}e(i) + \lambda_{2i}e(i)^{b/a} (i = 1,\cdots,n)$；$\mu$为足够小的正常数；$a$和$b$为正奇数，且$1/2 < b/a < 1$；$\sigma_1 = (2-b/a)\mu^{b/a-1}$，$\sigma_2 = (b/a-1)\mu^{b/a-2}$；sign(·)表示符号函数。

为设计控制力矩，对滑模面(13)进行求导，可得

$$\dot{\boldsymbol{s}} = \ddot{\boldsymbol{e}} + \boldsymbol{\lambda}_1\dot{\boldsymbol{e}} + \boldsymbol{\lambda}_2\dot{\phi}(\boldsymbol{e},\dot{\boldsymbol{e}}) \tag{15}$$

将式(12)代入到式(15)得

$$\dot{\boldsymbol{s}} = \boldsymbol{M}_0^{-1}(\boldsymbol{q})(\boldsymbol{\tau} - \boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})\boldsymbol{q} - \boldsymbol{G}_0(\boldsymbol{q}) + \boldsymbol{\tau}_{\mathrm{ext}} + \boldsymbol{\tau}_{\mathrm{d}}) - \ddot{\boldsymbol{q}}_{\mathrm{r}} + \boldsymbol{\lambda}_1\dot{\boldsymbol{e}} + \boldsymbol{\lambda}_2\dot{\phi}(\boldsymbol{e},\dot{\boldsymbol{e}}) \tag{16}$$

因此，设计控制力矩为

$$\boldsymbol{\tau}=\boldsymbol{\tau}_{\mathrm{eq}}+\boldsymbol{\tau}_{\mathrm{sw}}=\boldsymbol{M}(\ddot{\boldsymbol{q}}_{\mathrm{r}}-\boldsymbol{\lambda}_1\dot{\boldsymbol{e}}-\boldsymbol{\lambda}_2\dot{\phi}(\boldsymbol{e},\dot{\boldsymbol{e}}))+\boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})\boldsymbol{q}+\boldsymbol{G}_0(\boldsymbol{q})-\boldsymbol{\tau}_{\mathrm{ext}}-\boldsymbol{M}(\eta\mathrm{sign}(\boldsymbol{s})+\boldsymbol{\beta}\boldsymbol{s}) \quad (17)$$

式中: $\boldsymbol{\tau}_{\mathrm{eq}}$ 为等效控制力矩; $\boldsymbol{\tau}_{\mathrm{sw}}$ 为切换控制力矩; $\boldsymbol{\beta}$为控制增益。

3.2　稳定性证明

定理 1　考虑非线性系统 (11), 在滑模面 (13) 和控制律 (17) 的作用下, 机器人系统阻抗轨迹跟踪误差能在有限时间内收敛至零。

证明　定义如下李雅普诺夫函数:

$$V=\frac{1}{2}\boldsymbol{s}^{\mathrm{T}}\boldsymbol{s} \quad (18)$$

对式 (18) 求导得

$$\dot{V}(\boldsymbol{s})=\boldsymbol{s}[\ddot{\boldsymbol{e}}+\boldsymbol{\lambda}_1\dot{\boldsymbol{e}}+\boldsymbol{\lambda}_2\dot{\phi}(\boldsymbol{e},\dot{\boldsymbol{e}})]=\boldsymbol{s}[\boldsymbol{M}_0^{-1}(\boldsymbol{q})(\boldsymbol{\tau}-\boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})\boldsymbol{q}-\boldsymbol{G}_0(\boldsymbol{q})+\boldsymbol{\tau}_{\mathrm{ext}}+\boldsymbol{\tau}_{\mathrm{d}})-\ddot{\boldsymbol{q}}_{\mathrm{r}}+\boldsymbol{\lambda}_1\dot{\boldsymbol{e}}+\boldsymbol{\lambda}_2\dot{\phi}(\boldsymbol{e},\dot{\boldsymbol{e}})] \quad (19)$$

将控制律 (17) 代入式 (19) 得

$$\dot{V}=\boldsymbol{s}^{\mathrm{T}}(-\eta\mathrm{sign}(\boldsymbol{s})-\boldsymbol{\beta}\boldsymbol{s}+\boldsymbol{\tau}_{\mathrm{d}})\leqslant\boldsymbol{s}^{\mathrm{T}}(-\eta\mathrm{sign}(\boldsymbol{s})+\boldsymbol{\tau}_{\mathrm{d}})\leqslant-\rho\|\boldsymbol{s}\|\leqslant-\rho V^{1/2} \quad (20)$$

式中: $\rho=\eta-\|\boldsymbol{\tau}_{\mathrm{d}}\|>0$。

V初始时刻状态记为$V(0)$, 设t_{r}为收敛时间, 则$V(t_{\mathrm{r}})=0$, 对式 (20) 从 0 时刻到t_{r}时刻积分得

$$\int_0^{t_{\mathrm{r}}}\mathrm{d}t\leqslant\int_{V(0)}^{V(t_{\mathrm{r}})}\frac{-\mathrm{d}V}{\rho V^{1/2}} \quad (21)$$

可以计算出收敛时间为

$$t_{\mathrm{r}}\leqslant\frac{1}{\rho}V^{1/2}(0) \quad (22)$$

假设t_{r}时刻$V(t_{\mathrm{r}})\neq0$, $\boldsymbol{e}$沿着非奇异终端滑模收敛到零的时间为$t_{\mathrm{r}}+t_{\mathrm{s}}$, 为方便计算, 令$t_{\mathrm{r}}=0$。

当$R_i=0$或$R_i\neq0, |\boldsymbol{e}(i)|\geqslant\mu$时, 令

$$\boldsymbol{s}=\dot{\boldsymbol{e}}+\boldsymbol{\lambda}_1\boldsymbol{e}+\boldsymbol{\lambda}_2\boldsymbol{e}^{b/a}=\boldsymbol{0} \quad (23)$$

令$\boldsymbol{y}=\boldsymbol{e}^{1-b/a}$, 把式(23)写成一阶线性微分方程标准形式为

$$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t}+\frac{a-b}{a}\boldsymbol{\lambda}_1\boldsymbol{y}=-\frac{a-b}{a}\boldsymbol{\lambda}_2 \quad (24)$$

通过求解一阶微分方程得

$$t_{\mathrm{s}}=\frac{a}{\boldsymbol{\lambda}_1(a-b)}\ln\left(\frac{\boldsymbol{\lambda}_1V(0)^{a-b/a}+\boldsymbol{\lambda}_2}{\boldsymbol{\lambda}_2}\right) \quad (25)$$

当$|\boldsymbol{e}(i)|\geqslant\mu$时, $\boldsymbol{e}$将沿着非奇异终端滑模面有限时间内收敛到零; 当$|\boldsymbol{e}(i)|<\mu$时, $\boldsymbol{e}$将在一般滑模面上做滑模运动。通过选取一个充分小的μ, 使$|\boldsymbol{e}(i)|\geqslant\mu$总能成立, 从而可以实现$\boldsymbol{e}$在非奇异终端滑模面上有限时间收敛到零, 即$\boldsymbol{e}$的有限时间稳定性能够得到保证。

由式 (22) 和式 (25) 可知, 通过设计的滑模面和控制律, 机器人能在有限时间内跟踪阻抗轨迹。

注 1　为减少控制律的抖振, 可以利用饱和函数sat(·)代替符号函数sign(·)。饱和函数可表示为

$$\mathrm{sat}(z)=\begin{cases}1 & z\geqslant\theta\\ z/\theta & -\theta<z<\theta\\ -1 & z\leqslant-\theta\end{cases} \quad (26)$$

4　仿真实验

本节搭建了二连杆机器人与人协同跟踪目标轨迹的场景, 机器人在操作人员外力的作用下完成跟踪期望轨迹的任务, 人机交互场景示意图如图 3 所示。为验证本文所提基于迭代最优阻抗的人机物理交互控制方法的有效性, 分别对外环阻抗控制器和内环轨迹跟踪控制器进行验证。外环控制目标为求取最优阻抗参数, 使阻抗轨迹跟踪期望轨迹, 同时最小化操作人员工作量。内环控制目标为机器人实际轨迹跟踪阻抗轨迹。在 MATLAB/Simulink 中进行仿真, 采用龙格库塔求解器, 仿真步长设置为 0.01 s, 仿真时长为 15 s。

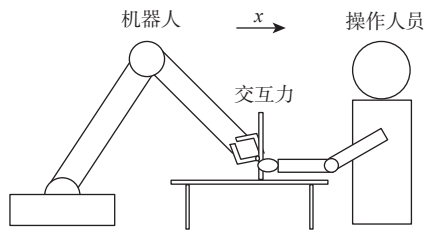


图 3　人机物理交互场景示意图

Fig. 3　Human-robot physical interaction scenario

4.1　控制外环仿真

为验证外环 IOIC 的有效性, 设置期望轨迹为沿着 x 方向上的点对点运动, 具体轨迹设置为幅值 0.4、周期 10 s 的方波信号。机器人的阻抗模型 (1) 中, 初始阻抗参数分别设置为 $\boldsymbol{M}_0$、$\boldsymbol{B}_0$和$\boldsymbol{K}_0$。人的运动学模型 (2) 中, 增益分别设置为$K_{\mathrm{r}}=0.5$、$K_{\mathrm{f}}=5$和$K_{\mathrm{e}}=20$, 这些增益对于 IOIC 来说是未知的。代价函数 (6) 中的权重$\boldsymbol{Q}$和R分别设置为

$$\begin{cases}\boldsymbol{Q}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\\ R=0.01\end{cases} \quad (27)$$

实验开始的前 0.5 s 是数据收集阶段, 设置此阶段控制输入 (8) 中的探测噪声v_{n}为

$$v_{\mathrm{n}}=\sum_{j=1}^{10}(0.05/j)\sin(100t) \quad (28)$$

1) IOIC 性能验证。为证明本文所提外环 IOIC 的性能, 与定阻抗控制 (fix impedance control, FIC) 和利用 LQR 法直接求取的最优阻抗控制 (optimal impedance control, OIC) 进行对比。其中, FIC 通常

需要在设计阶段手动选择适当的阻抗参数。在未知的环境和任务变化中，由于环境多变、任务多变，难以得到普适的阻抗参数，不具备自适应性；OIC在已知环境下被视为一种理想的控制策略，虽然性能最优，但难以在实际中取得应用。实验结果如图4～图6所示。

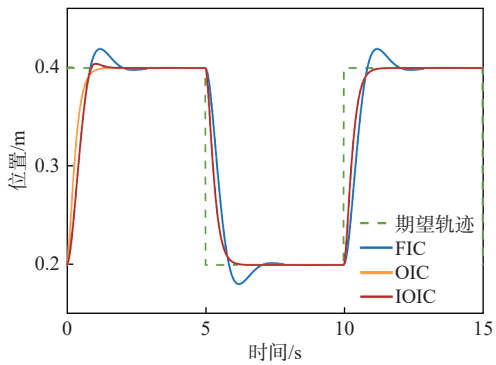


图4 期望轨迹和阻抗轨迹

Fig. 4 Desired trajectory and impedance trajectory

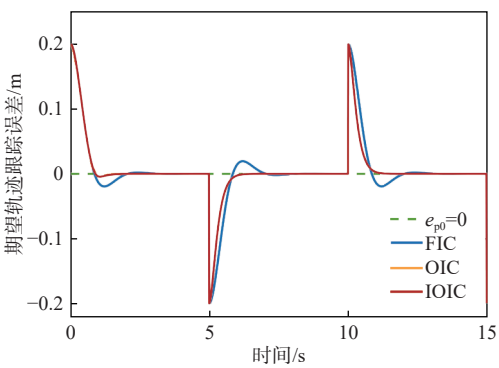


图5 期望轨迹跟踪误差

Fig. 5 Desired trajectory tracking error

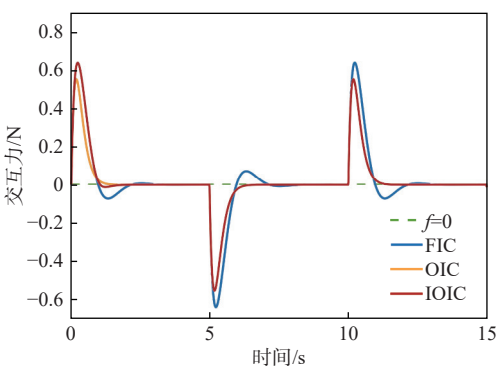


图6 机器人与人的实际交互力

Fig. 6 Actual interaction force between robots and humans

图4展示了IOIC、FIC和OIC在所搭建人机物理交互仿真实验中得到的阻抗轨迹结果。可以看到，0～5 s阶段，IOIC的阻抗轨迹在1 s后快速跟踪上期望轨迹，且在5～10 s和10～15 s这2个阶段中与OIC的最优阻抗轨迹基本重合，可以无超调地跟踪期望轨迹。这是因为IOIC能利用0～0.5 s内收集到的数据迭代求解出近似最优阻抗参数，从而实现与OIC几乎相同的性能。而FIC由于阻抗参数设置不合适，且无法改变，导致经过超调振荡后才能跟踪期望轨迹。由此可得，IOIC能够无需重复实验，仅利用任务初始阶段的交互信息，实现最优交互性能。

为验证IOIC能实现最小化期望轨迹跟踪误差和交互力的控制目标，在图5和图6分别展示了IOIC、FIC和OIC的期望轨迹跟踪误差e_{p0}和交互力f变化曲线。由图5可以看到，IOIC的期望轨迹跟踪误差在1 s后快速收敛至0，且在5～10 s和10～15 s这2个阶段中，误差曲线与OIC基本重合。由图6可以看到，0～5 s阶段，IOIC可以使交互力快速收敛至0，而FIC由于跟踪误差的存在，导致交互力经过几次振荡后才逐渐收敛至0。且通过5～10 s和10～15 s这2个阶段可以看出，IOIC的交互力峰值明显低于FIC，且能快速无超调地收敛至0。因此，结合图5和图6的仿真实验结果可以得出结论，IOIC可以实现最小化期望轨迹跟踪误差和交互力的控制目标。

2）软辅助函数有效性验证。为验证所提软辅助函数的作用，图7展示了人机物理交互系统的控制输入，即轨迹跟踪误差加速度。从图7中红色虚线和蓝色虚线可以看到，当不添加软辅助函数时，在0.5 s时刻，由于阻抗参数从初始值变为迭代得到的最优值，导致跟踪误差加速度有一个突变，取不同的阻抗参数时，其幅值突变大小不同，当所选的初始阻抗参数与迭代得到的最优阻抗参数差距过大时，会出现显著的突变。因此，可以得出结论：突变的大小与初始阻抗参数的选择有关。这种突变在实际应用中可能会造成机器人系统的抖动。如图7中红色实线和蓝色实线所示，当添加软辅助函数

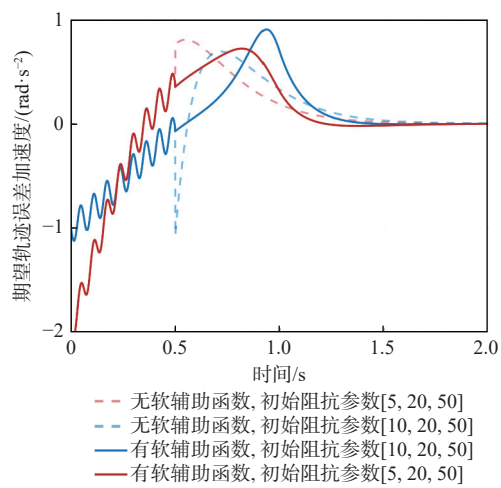


图7 期望轨迹跟踪误差加速度

Fig. 7 Desired trajectory tracking error acceleration

后, 可以看到, 控制输入曲线变化较为平缓和连续。由此验证了所提软辅助函数的有效性。同时, 图 7 中, 前 0.5 s 内有一定的抖振, 这是由于在系统输入中加入了探测噪声, 使收集到的数据丰富且非奇异。

3) IOIC 数值验证。为进一步展示本文所提 IOIC 能得到最优阻抗参数, 将 IOIC 与 OIC 结果进行比较。根据选择操作人员的模型参数, 可以得到系统 (4) 的状态矩阵 $\boldsymbol{A}$ 和输入矩阵 $\boldsymbol{B}$ 分别为

$$\begin{cases} \boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 40 & 0 & -10 \end{bmatrix} \\ \boldsymbol{B} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \end{cases} \tag{29}$$

OIC 利用 LQR 法直接求解 Riccati 方程, 可以得到最优的 $\boldsymbol{P}^*$ 和 $\boldsymbol{K}^*$ 分别为

$$\begin{cases} \boldsymbol{P}^* = \begin{bmatrix} 4.795\,8 & 0.383\,0 & 0.170\,8 \\ 0.383\,0 & 0.132\,9 & 0.007\,3 \\ 0.170\,8 & 0.007\,3 & 0.049\,7 \end{bmatrix} \\ \boldsymbol{K}^* = \begin{bmatrix} 38.296\,9 & 13.288\,9 & 0.733\,5 \end{bmatrix} \end{cases} \tag{30}$$

基于 IOIC 的控制流程, 根据选取的初始阻抗参数, 设置初始的反馈增益为

$$\boldsymbol{K}_0 = [10 \quad 4 \quad 0.2] \tag{31}$$

根据所设置的代价函数权重, 利用前 0.5 s 收集到的数据进行迭代。图 8 展示了矩阵 $\boldsymbol{P}$ 和反馈增益 $\boldsymbol{K}$ 的迭代过程。可以看到, 经过 7 次迭代, 迭代值与最优值的欧氏距离逐渐趋于 0, 即迭代反馈增益逐渐趋于最优反馈增益值。

最终得到的反馈增益为

$$\boldsymbol{K}_7 = [38.296\,6 \quad 13.288\,9 \quad 0.733\,6] \tag{32}$$

可以得到基于迭代求解的最优阻抗参数为

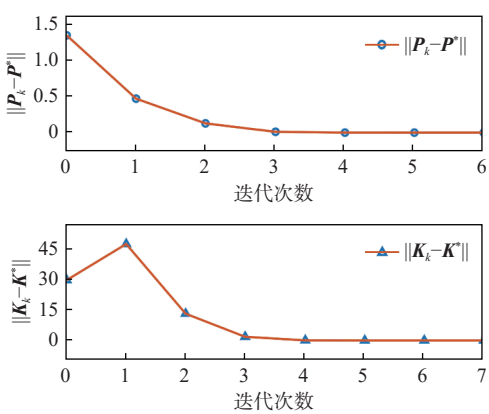


图 8　$\boldsymbol{P}$ 和 $\boldsymbol{K}$ 矩阵迭代过程

Fig. 8　$\boldsymbol{P}$ and $\boldsymbol{K}$ matrix iteration

$$\begin{cases} M^* = 1.363\,2 \\ B^* = 18.115\,0 \\ K^* = 52.204\,8 \end{cases} \tag{33}$$

通过比较式 (30) 和式 (32) 可以看出, IOIC 求解得到的反馈增益和 OIC 直接求解 Riccati 方程得到的最优反馈增益数值基本相同, 从数值上验证了 IOIC 的有效性。

4.2　控制内环仿真

为验证所提内环 NTSMTC 的有效性, 设置二连杆机器人的物理参数如表 1 所示。

表 1　二连杆机器人物理参数

Table 1　Physical parameters of two-link robot

参数	数值
连杆1长度 l_1/m	1.0
连杆2长度 l_2/m	1.0
连杆1质量 m_1/kg	1.5
连杆2质量 m_2/kg	1.5
连杆1惯性张量 I_1/(kg·m^2)	5.0
连杆2惯性张量 I_2/(kg·m^2)	5.0
重力常数 g/(m·s^{-2})	9.81

设置机器人运动学模型 (11) 中的转矩扰动向量为

$$\boldsymbol{\tau}_{\mathrm{d}} = \begin{bmatrix} \sin t + 0.2\sin(200\pi t) \\ \cos t + 0.2\cos(200\pi t) \end{bmatrix} \tag{34}$$

滑模面中的参数设置为: $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2 = \mathrm{diag}(3,3)$。式(13)中的参数分别设置为: $a = 5$, $b = 3$, $\mu = 0.005$。控制器中的参数分别设置为: $\boldsymbol{\beta} = \mathrm{diag}(0.5,1)$, $\eta = 1.5$。

仿真结果如图 9～图 11 所示。

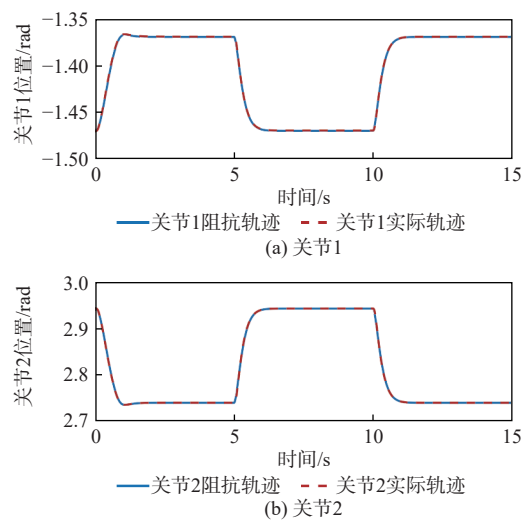


图 9　关节空间中的阻抗轨迹和实际轨迹

Fig. 9　Impedance trajectory and actual trajectory in joint space

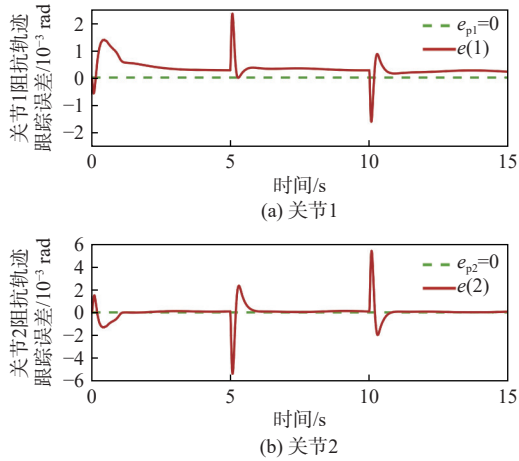


图 10 关节空间中的阻抗轨迹跟踪误差

Fig. 10 Impedance trajectory tracking error in joint space

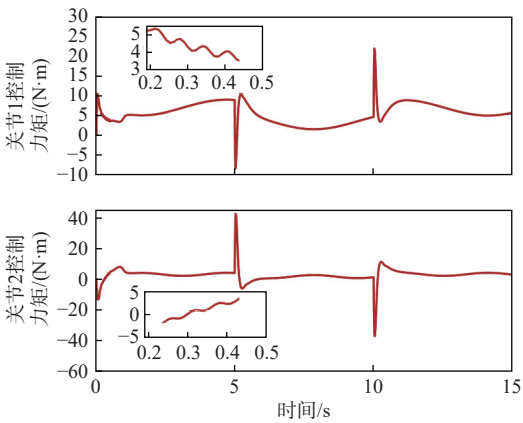


图 11 关节控制力矩

Fig. 11 Joint control torque

图 9 展示了关节空间中的阻抗轨迹和实际轨迹跟踪曲线。图 10 展示了机器人 2 个关节的轨迹跟踪误差，可以看出，关节 1 的轨迹跟踪误差始终在 3×10^{-3} rad范围内，关节 2 的轨迹跟踪误差始终在 6×10^{-3} rad范围内，证明了所提轨迹跟踪方法的有效性。图 11 展示了 2 个关节的控制力矩曲线，可以看到，关节 1 的控制力矩最大不超过 25 N·m，关节 2 的控制力矩最大不超过 42 N·m，在 5 s 和 10 s 时，关节力矩发生较大的变化，这是因为此刻阻抗轨迹突然变化所导致。0.5 s 前的抖振是由于加入的探测噪声所引起。整体来看，控制力矩平稳无明显抖振，验证了所提内环轨迹跟踪方法控制力矩的平稳性。

5 结 论

1）本文提出的基于迭代最优阻抗的人机物理交互控制方法，在人机协作型任务中，无需多次重复实验，能够实现机器人与人的最优交互性能，确保机器人轨迹跟踪精度的同时减少了人在执行任务过程中消耗的力。

2）外环 IOIC 无需交互人的模型信息，利用可在线测量的轨迹跟踪误差和交互力信息，求取近似最优阻抗参数，与 OIC 求取的最优阻抗参数基本一致。

3）内环 NTSMTC 在交互力和内部干扰的条件下具有很好的轨迹跟踪能力，控制力矩整体平滑且无明显抖振，并通过李雅普诺夫能量函数对控制律进行收敛性证明。

参考文献（References）

[1] 李琳, 肖佳栋, 张铁, 等. 基于自适应迭代的机器人曲面恒力跟踪[J]. 北京航空航天大学学报, 2019, 45(4): 641-649.
LI L, XIAO J D, ZHANG T, et al. Constant-force curved-surface-tracking with robotic manipulator based on adaptive iterative algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(4): 641-649(in Chinese).

[2] ZAHEDI F, CHANG D, LEE H. User-adaptive variable damping control using Bayesian optimization to enhance physical human-robot interaction[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 2724-2731.

[3] LIU X, GE S S, ZHAO F, et al. A dynamic behavior control framework for physical human-robot interaction[J]. Journal of Intelligent & Robotic Systems, 2020, 101(1): 14.

[4] HOGAN N. Impedance control: an approach to manipulation[J]. Journal of Dynamic Systems, Measurement, and Control, 1985, 107(1): 1-24.

[5] YU X B, LI B, HE W, et al. Adaptive-constrained impedance control for human-robot co-transportation[J]. IEEE Transactions on Cybernetics, 2022, 52(12): 13237-13249.

[6] SHARIFI M, ZAKERIMANESH A, MEHR J K, et al. Impedance variation and learning strategies in human-robot interaction[J]. IEEE Transactions on Cybernetics, 2022, 52(7): 6462-6475.

[7] ABU-DAKKA F J, SAVERIANO M. Variable impedance control and learning-a review[J]. Frontiers in Robotics and AI, 2020, 7: 590681.

[8] SUN T R, PENG L, CHENG L, et al. Stability-guaranteed variable impedance control of robots based on approximate dynamic inversion[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(7): 4193-4200.

[9] JIN Z H, QIN D D, LIU A D, et al. Model predictive variable impedance control of manipulators for adaptive precision-compliance tradeoff[J]. IEEE/ASME Transactions on Mechatronics, 2023, 28(2): 1174-1186.

[10] LIU X, WU J G, WANG L. Sensorless variable impedance compliant control for human-robot interaction[J]. AIP Advances, 2022, 12(11): 115121.

[11] GRAFAKOS S, DIMEAS F, ASPRAGATHOS N. Variable admittance control in pHRI using EMG-based arm muscles co-activation[C]//Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. Piscataway: IEEE Press, 2016: 1900-1905.

[12] DONG Y T, REN B B. UDE-based variable impedance control of uncertain robot systems[J]. IEEE Transactions on Systems, Man,

and Cybernetics: Systems, 2019, 49(12): 2487-2498.

[13] 梅雪松, 刘星, 赵飞, 等. 协作机器人外力感知与交互控制研究现状及展望[J]. 航空制造技术, 2020, 63(9): 22-32.
MEI X S, LIU X, ZHAO F, et al. Research status and prospects of external force perception and interaction control methods for collaborative robots[J]. Aeronautical Manufacturing Technology, 2020, 63(9): 22-32(in Chinese).

[14] YANG C G, GANESH G, HADDADIN S, et al. Human-like adaptation of force and impedance in stable and unstable interactions[J]. IEEE Transactions on Robotics, 2011, 27(5): 918-930.

[15] YANG J T, SUN T R, CHENG L, et al. Spatial repetitive impedance learning control for robot-assisted rehabilitation[J]. IEEE/ASME Transactions on Mechatronics, 2023, 28(3): 1280-1290.

[16] RAHIMI H N, HOWARD I, CUI L. Neural impedance adaption for assistive human-robot interaction[J]. Neurocomputing, 2018, 290: 50-59.

[17] XU Z H, LI X X, LI S, et al. Dynamic neural networks based adaptive optimal impedance control for redundant manipulators under physical constraints[J]. Neurocomputing, 2022, 471: 149-160.

[18] JOHANSSON R, SPONG M W. Quadratic optimization of impedance control[C]//Proceedings of the IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 1994: 1: 616-621.

[19] MATINFAR M, HASHTRUDI-ZAAD K. Optimization-based robot compliance control: geometric and linear quadratic approaches[J]. International Journal of Robotics Research, 2005, 24(8): 645-656.

[20] HUANG H H, YANG C G, CHEN C L P. Optimal robot-environment interaction under broad fuzzy neural adaptive control [J]. IEEE Transactions on Cybernetics, 2021, 51(7): 3824-3835.

[21] GUO Y D, TIAN Y, WANG H P. Learning-based adaptive optimal impedance control to enhance physical human-robot interaction performance[J]. International Journal of Control, Automation and Systems, 2022, 20(9): 3053-3062.

[22] SUZUKI S, FURUTA K. Adaptive impedance control to enhance human skill on a haptic interface system[J]. Journal of Control Science and Engineering, 2012, 2012: 365067.

[23] JIANG Y, JIANG Z P. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics [J]. Automatica, 2012, 48(10): 2699-2704.

Human-robot physical interaction control method based on iterative optimal impedance

LIU Weirong*, WEI Zifeng, JIN Zhenbing, MENG Jiahao, WANG Xingkun, ZHANG Haochen

(College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: In order to improve the accuracy and compliance of human-robot physical interaction and achieve optimal interaction performance, a human-robot physical interaction control method based on iterative optimal impedance was proposed to solve the problem that iterative learning-based impedance control method needs to repeat the same task many times. The proposed method draws on the mechanism by which iterative optimal control can optimize cost function to determine optimal control input to the system without information of the system matrix. A double-loop control structure was used for the proposed control method. An iterative optimal impedance controller (IOIC) was designed for a task-oriented outer loop. The problem of determining optimal impedance parameters was described as a linear quadratic regulator problem, which utilized iterative optimal control to find optimal feedback gain and minimize cost function including tracking error and interaction force. Robot jitter caused by parameter mutations was avoided by introducing soft auxiliary functions. A nonsingular terminal sliding mode trajectory tracking controller (NTSMTC) was used in the inner loop of the robot to make the actual trajectory of the robot track impedance trajectory output by the outer loop, and the chattering of control law was eliminated by saturation function. Simulation results prove that the proposed method can obtain optimal impedance parameters only by using interactive information in the initial stage of the task once in a human-robot collaborative task, so as to minimize the trajectory tracking error and the force consumed by the human during the task.

Keywords: human-robot physical interaction; impedance control; trajectory tracking control; optimal control; sliding mode control

Received: 2023-06-05; **Accepted**: 2023-10-16; **Published Online**: 2023-11-24 14:33

URL: link.cnki.net/urlid/11.2625.V.20231123.1034.002

Foundation items: National Nature Science Foundation of China (62261032); Central Government Guiding Funds for Local Science and Technology Development Program (25ZYJA026); Key Talent Project of Gansu Province; Gansu Education Science and Technology Innovation Fund (2022A-021)

* **Corresponding author**. E-mail: liuwr@lut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0322

边缘引导的双分支网络SAR图像相干斑抑制方法

朱磊*，姚同钰，车晨洁，姚丽娜，张博，潘杨
(西安工程大学 电子信息学院，西安 710048)

摘 要：为进一步提升深度学习方法对合成孔径雷达(SAR)图像相干斑的抑制与边缘保持性能，提出了一种边缘引导的双分支网络相干斑抑制方法。构建了一种由边缘信息提取模块与双分支抑斑网络2部分构成的新型抑斑网络模型。采用密集级联方式构建边缘信息提取模块，增强模型的边缘感知能力；利用基于通道注意力的残差抑斑子网络(CARNet)、基于混合注意力的增强抑斑子网络(MAENet)及基于多分支并行的多尺度特征融合模块(MPMFFB)共同形成双分支抑斑网络，实现在相干斑抑制的同时更好地保护边缘细节。实验结果表明：与SAR-Transformer、HTNet等先进方法相比，所提方法具有更好的相干斑抑制与边缘保持性能；对仿真SAR图像，峰值信噪比、结构相似性、边缘保持指数分别平均提升0.96 dB、2.60%、0.60%；对真实SAR图像，等效视数提升14.12%以上，边缘保持指数平均提升4.52%。

关 键 词：图像去噪；合成孔径雷达图像；相干斑抑制；双分支网络；多尺度特征融合
中图分类号：TN911.73; TP751
文献标志码：A **文章编号**：1001-5965(2025)06-1852-11

合成孔径雷达(synthetic aperture radar, SAR)作为一种新型遥感技术，在地质勘测、灾难监测等领域具有重大的应用价值。然而，由于SAR系统的相干特性，使其在成像过程中不可避免地产生一种称为相干斑的乘性噪声，相干斑的存在会极大降低后续分割、识别等SAR图像应用的精确度[1]。因此，在解译前对SAR图像进行抑斑处理，对提升SAR图像的质量与后续解译处理效果至关重要。

根据滤波方式不同，现有的相干斑抑制方法可以分为传统相干斑抑制方法和基于深度学习的相干斑抑制方法。传统相干斑抑制方法出现较早，主要以空域滤波方法[2-3]、变换域滤波方法[4-5]及各向异性扩散滤波方法[6-7]等为代表。随着深度学习在图像恢复领域的不断深入，利用深度学习进行相干斑抑制的方法也相继出现，其按照学习形式可以分为两大类，即监督学习方法和自监督学习方法。监督学习方法利用大量的清洁-含噪图像对来训练网络模型，通过学习相干斑噪声的统计特性，进一步减小真实值与估计值之间的误差。2017年，Chierchia等[8]提出了经典的SAR图像抑斑卷积神经网络(SAR image despeckling through convolutional neural network, SAR-CNN)。随后，Wang等[9]将比值残差思想应用于网络训练过程中，提出了端到端的图像抑斑卷积神经网络(image despeckling convolutional neural network, ID-CNN)。Vitale等[10]结合SAR图像的空间和统计特性，提出了多目标网络(multi-objective network, MONet)。Passah等[11]基于ResNet的跳跃连接方式，提出了一种用于SAR图像抑斑的深层卷积神经网络，并采用混合损失函数来加快模型收敛速度。2022年，Perera等[12]结合Transformer思想，允许模型学习不同图像区域之间

收稿日期：2023-06-06；**录用日期**：2023-09-14；**网络出版时间**：2023-09-28 16:51
网络出版地址：link.cnki.net/urlid/11.2625.V.20230928.1126.001
基金项目：国家自然科学基金(61971339)；陕西省重点研发计划(2019GY-113)；陕西省自然科学基础研究计划(2019JQ-361)
*通信作者. E-mail：zhulei791014@163.com

引用格式：朱磊，姚同钰，车晨洁，等. 边缘引导的双分支网络SAR图像相干斑抑制方法[J]. 北京航空航天大学学报，2025，51(6)：1852-1862. ZHU L，YAO T Y，CHE C J，et al. SAR image coherence speckle suppression method based on edge-guided dual-branch network[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1852-1862 (in Chinese).

的全局依赖关系, 提出了一种基于 Transformer 的抑斑新模型。2023 年, Cheng 等[13] 针对 SAR 图像中相干斑噪声的重尾特性, 设计了一种具有混合截断损失的双流卷积神经网络 (two-stream convolutional neural network with hybrid truncation loss, HTNet) 进行相干斑抑制。基于监督学习的抑斑方法作为目前 SAR 图像相干斑抑制领域的热门研究方向, 其可以通过非线性映射的方式提取图像深层特征, 克服了传统抑斑方法存在的局限性, 但在同步提升相干斑抑制与边缘细节保持性能方面仍具有提升空间。自监督学习方法则通过从大量无标签样本中挖掘自身监督信息, 从而进行类似于监督学习的过程。Lin 等[14] 利用自监督思想, 使用深度卷积神经网络隐含的图像先验制定学习去噪的正则化, 并通过先验估计和参数更新的循环执行实现 SAR 图像相干斑抑制。自监督学习方法不需要数据标签, 极大节省了人力资源, 但由于训练过程不稳定等问题导致其相关算法在 SAR 图像相干斑抑制领域还不够成熟。

针对 SAR 图像相干斑的抑制与边缘保持性能难以同步提升的问题, 本文以 SAR 图像乘性噪声分布为基础, 采用监督学习思想, 提出了一种边缘引导的双分支网络 SAR 图像相干斑抑制方法。主要贡献如下: ①构建了一种 SAR 图像相干斑抑制模型, 该模型主体部分采用双分支网络结构, 同时在网络输入端融入图像的边缘信息作为先验指导, 以同步提升同质区平滑滤波与边缘区细节保护性能。②提出了一种基于密集级联的边缘信息提取模块 (dense cascade based edge information extraction block, DCEEB), 用于准确获取 SAR 图像的边缘纹理信息, 引导模型学习更多的边缘细节信息。③采用双分支抑斑网络与注意力机制结合的策略进行相干斑抑制, 同时提出了一种基于多分支并行的多尺度特征融合模块 (multi-branch parallel based multi-scale feature fusion block, MPMFFB), 用于综合 2 个分支不同精细程度的图像特征, 在提升图像特征利用率的同时, 有利于抑斑后图像恢复出更丰富的纹理细节。

1　边缘引导的双分支抑斑网络

本文提出的抑斑网络结构如图 1 所示。该网络包含 DCEEB 模块和由基于通道注意力的残差抑斑子网络 (channel attention-based residual despeckling network, CARNet)、基于混合注意力的增强抑斑子网络 (mixed attention-based enhanced despeckling network, MAENet)、MPMFFB 模块组成的双分支抑斑网络 2 个部分。

对于输入的含噪图像, 首先, 经过 DCEEB 模块对图像的边缘信息进行提取。然后, 将含噪图像及其边缘图像一同送入网络中进行学习。CARNet 子网络采用 UNet 架构对相干斑进行抑制。MAENet 子网络主要用于弥补 CARNet 子网络对非局部特征提取能力不强的问题, 对相干斑进行进一步抑制, 实现更为精细的滤波。MPMFFB 模块则将 CARNet 子网络的多尺度特征融入到 MAENet 子网络中。最后, 运用集成学习的方法融合不同分支网络的图像特征, 并采用 1×1 卷积层进行维度压缩, 输出抑斑后图像。

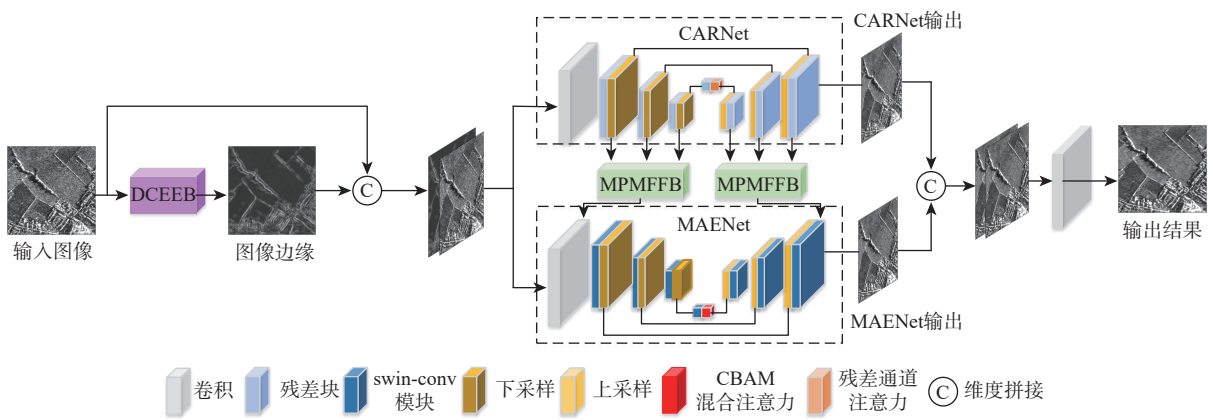


图 1　边缘引导的双分支抑斑网络总体结构

Fig. 1　Overall structure of edge-guided dual-branch speckle suppression network

1.1　基于密集级联的边缘信息提取模块

为改善 SAR 图像抑斑过程中存在的边缘纹理丢失问题, 本文基于 SAR 图像相干斑的乘性噪声特性, 提出 DCEEB 模块对图像的边缘信息进行提取。该模块利用均值比[15] 获取训练集图像的边缘强度信息, 将其送入 DCEEB 模块中进行训练, 从而获得 SAR 图像的精细边缘。对于 DCEEB 模块所获取的图像边缘信息, 将其与输入图像进行融合后

送入后续网络，以进一步增强边缘先验知识，达到在平滑滤波的同时实现弱边缘信息保护的目的。

如图2所示，对输入图像$\boldsymbol{F}_{\text{in}}$进行浅层特征提取，过程可表示为

$$\boldsymbol{F}_{-1}=H_{-1}(\boldsymbol{F}_{\text{in}}) \tag{1}$$

式中：$\boldsymbol{F}_{-1}$为提取到的浅层特征图；$H_{-1}(\cdot)$包括3×3卷积、批量标准化 (batch normalization, BN) 和ReLU激活。

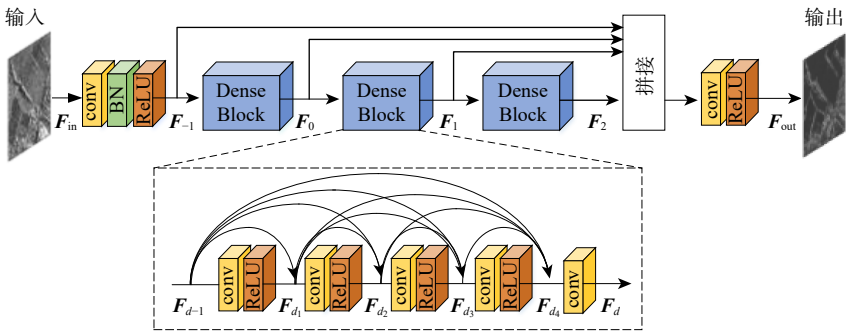


图2 DCEEB模块网络结构

Fig. 2 Network structure of DCEEB

然而，浅层特征所包含的语义信息并不够丰富，因此，采用3个Dense Block级联来提取更深层次的弱边缘强度信息。Dense Block利用其密集连接操作，将所有层的特征信息紧密结合，该过程可以表述为

$$\boldsymbol{F}_d=f_{3\times3}\left(\left[\boldsymbol{F}_{d-1},\boldsymbol{F}_{d_1},\boldsymbol{F}_{d_2},\boldsymbol{F}_{d_3},\boldsymbol{F}_{d_4}\right]\right) \tag{2}$$

式中：$\boldsymbol{F}_d$为第d个Dense Block的输出特征；$f_{3\times3}$为3×3卷积操作；$\boldsymbol{F}_{d_1},\cdots,\boldsymbol{F}_{d_4}$分别为第$d$个Dense Block第1个卷积层到第4个卷积层的输出特征图；$\left[\boldsymbol{F}_{d-1},\boldsymbol{F}_{d_1},\cdots,\boldsymbol{F}_{d_4}\right]$表示将前一个Dense Block的输出特征图与当前Dense Block各卷积层的输出特征图进行通道上的合并。

将不同层次的图像特征结合后，利用3×3的卷积操作进行通道压缩，得到边缘特征图，即

$$\boldsymbol{F}_{\text{out}}=\sigma(f_{3\times3}[\boldsymbol{F}_{-1},\boldsymbol{F}_0,\boldsymbol{F}_1,\boldsymbol{F}_2]) \tag{3}$$

式中：$\boldsymbol{F}_{\text{out}}$为DCEEB模块的输出；$\sigma$为激活函数ReLU；$[\boldsymbol{F}_{-1},\boldsymbol{F}_0,\boldsymbol{F}_1,\boldsymbol{F}_2]$表示将各层次所提取的特征进行拼接。

1.2 双分支抑斑网络

1.2.1 基于通道注意力的残差抑斑子网络

CARNet子网络主要由编码器和解码器2部分构成，其具体参数如表1所示。

CARNet子网络采用线性叠加的思想，通过多个残差块resblock级联进行特征提取，使网络更容易学习，同时避免网络在训练过程中出现梯度消失的现象。编码器端利用下采样操作获得不同尺度的特征，并在最小尺度上添加残差通道注意力RCABlock[16]，使网络对不同的特征信息赋予不同的权重，提升网络对重要细节特征的提取能力，使网络专注于更有用的信道。解码器端则通过上采样操作对图像信息进行重建。

表1 CARNet参数

Table 1 CARNet parameters

编码器/解码器	层级	网络层结构
编码器	L1	{conv3×3, 64}×1
	L2	{resblock, 64}×4, {strideconv2×2, 128}×1
	L3	{resblock, 128}×4, {strideconv2×2, 256}×1
	L4	{resblock, 256}×4, {strideconv2×2, 512}×1
	L5	{resblock, 512}×4, {RCABlock, 512}×4
解码器	L6	{convtranspose2×2, 256}×1, {resblock, 256}×4
	L7	{convtranspose2×2, 128}×1, {resblock, 128}×4
	L8	{convtranspose2×2, 64}×1, {resblock, 64}×4

1.2.2 基于混合注意力的增强抑斑子网络

在CARNet子网络中，由于卷积操作的局部感知特性及编码器和解码器之间的长跳跃连接方式，会使网络丢失一些在图像重建过程中发挥重要作用的上下文信息，本文提出MAENet子网络来提高模型的非局部特征提取能力，以弥补CARNet子网络的不足。在进行上/下采样前，MAENet子网络采用Zhang等[17]提出的Swin-Conv模块，该模块通过卷积、分裂和级联等操作，将Swin Transformer的非局部建模能力与带残差卷积RConv的局部建模能力结合，提升上下文特征信息的表达能力，克服了单一卷积操作存在的局限性。与CARNet子网络相似，MAENet子网络在最小尺度上采用CBAM混合注意力机制[18]，从空间域和通道域2个维度对获取到的图像特征进行矫正，在没有引入过多冗余计算量的同时达到自适应特征优化的目的。MAENet子网络的具体参数如表2所示。

表 2　MAENet 参数

Table 2　MAENet parameters

编码器/解码器	层级	网络层结构
编码器	L1	{conv3×3, 64}×1
	L2	{swin-conv block, 64}×2, {strideconv2×2, 128}×1
	L3	{swin-conv block, 128}×2, {strideconv2×2, 256}×1
	L4	{swin-conv block, 256}×2, {strideconv2×2, 512}×1
	L5	{swin-conv block, 512}×2, {CBAM, 512}×1
解码器	L6	{convtranspose2×2, 256}×1, {swin-conv block, 256}×2
	L7	{convtranspose2×2, 128}×1, {swin-conv block, 128}×2
	L8	{convtranspose2×2, 64}×1, {swin-conv block, 64}×2

1.2.3　基于多分支并行的多尺度特征融合模块

如图 3 所示, 受 Inception[19] 的启发, MPMFFB 模块采用多分支并行方式将 CARNet 子网络的多尺度信息融入到 MAENet 子网络。不同尺度特征所能提供的感受野大小不同, 将多尺度特征信息进行融合这一操作有利于图像恢复, 并且可以避免图像伪影, 进一步充分利用所提取的图像特征。

MPMFFB 模块主要由上采样层和1×1卷积层构成。1×1卷积层用于将 CARNet 子网络的不同尺度特征压缩至相同特征图数量, 以便于所有特征图添加到 MAENet 子网络时具有相同的权重。上采样层用于将特征图上采样到与将要添加到的 MAENet 子网络位置的相同特征图尺度, 本文使用 pixelshuffle 方式进行上采样操作, 其上采样的比例因子等于 MAENet 子网络中特定位置的特征图大小与 CARNet 子网络中相应特征图大小的比值。具体来说, 如式 (4) 所示, 对于输入的 3 个特征图 $\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$, 分别通过上采样模块和1×1大小的卷积操作得到 3 个特征图组 $\boldsymbol{Y}_1$、$\boldsymbol{Y}_2$、$\boldsymbol{Y}_3$, 该过程可以表述为

$$\begin{cases} \boldsymbol{Y}_1 = \mathrm{conv}_{1\times1}\left(\mathrm{upsample}\left(\boldsymbol{X}_1\right)\right) \\ \boldsymbol{Y}_2 = \mathrm{conv}_{1\times1}\left(\mathrm{upsample}\left(\boldsymbol{X}_2\right)\right) \\ \boldsymbol{Y}_3 = \mathrm{conv}_{1\times1}\left(\mathrm{upsample}\left(\mathrm{conv}_{1\times1}\left(\boldsymbol{X}_3\right)\right)\right) \end{cases} \tag{4}$$

将 $\boldsymbol{Y}_1$、$\boldsymbol{Y}_2$、$\boldsymbol{Y}_3$这 3 个特征图组进行元素相加, 获得 MPMFFB 模块最终的输出$\boldsymbol{Z}$, 用数学公式描述为

$$\boldsymbol{Z} = \boldsymbol{Y}_1 \oplus \boldsymbol{Y}_2 \oplus \boldsymbol{Y}_3 \tag{5}$$

式中: "⊕" 表示元素相加。

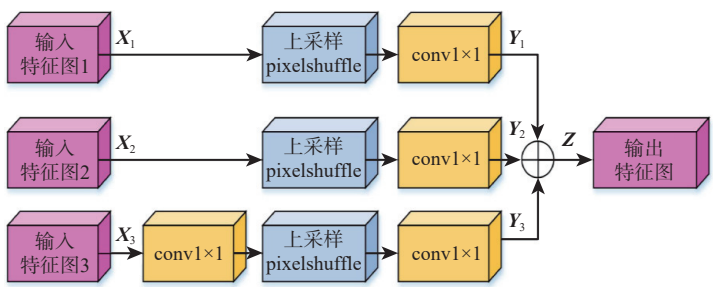


图 3　MPMFFB 模块网络结构

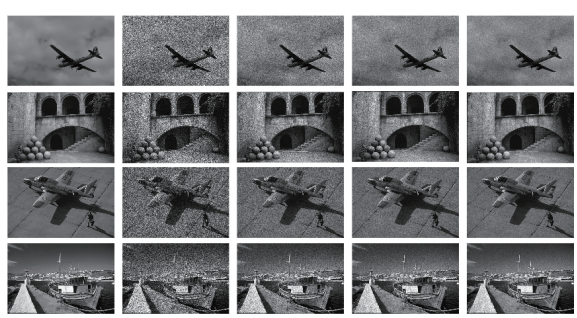
Fig. 3　Network structure of MPMFFB

2　实验结果与分析

2.1　实验数据及环境设置

由于 SAR 图像缺乏真实无噪图像, 本文实验采用仿真 SAR 图像来训练网络。实验选用 BSDS500 数据集来训练网络, 该数据集为伯克利大学提供的公开数据集, 包含 500 张自然图像。在网络的训练阶段, 随机选择 BSDS500 数据集的 400 张图像作为训练集, 其余 100 张作为验证集。仿真 SAR 图像利用视数为 L、噪声强度服从 Gamma 分布的相干斑噪声模型[20] 进行合成, L 分别设置为 1,4,8,16, 用于生成 4 种不同视数的噪声样本, 与上述无噪声样本构成训练图像对。图 4 为训练数据集部分样本示例, 图 4(a) 为原始图像, 图 4(b)～图 4(e) 分别为原始图像经过相干斑噪声模型构建的视数为 L=1,4,8,16 的仿真 SAR 图像。在测试阶段, 本文采用 Set12、Classic5 和 Kodak24 这 3 个公共数据集对仿真 SAR 图像进行测试, 同时使用 Amsterdam 地区真实 SAR 图像进行测试。

实验采用 64 位 Ubuntu20.04 操作系统,



(a) 原始图像　(b) L=1　(c) L=4　(d) L=8　(e) L=16

图 4　数据集样本示例

Fig. 4　Dataset samples

Pytorch1.11.0 框架，硬件方面采用 12th Gen Intel(R) Core(IM) i7-12700KF CPU 及 NVIDIA RTX 3060 GPU 对网络模型进行训练。模型训练的损失函数采用 L_1 损失，网络参数更新采用 Adam 优化器，设置初始学习率为 10^{-5}，网络每次训练选取的样本数为 16，总迭代次数为 5 000 次。

2.2 对比实验

为验证本文方法的相干斑抑制与边缘保持性能，实验采用 SAR-BM3D[21]、ID-CNN[9]、SAR-CNN[8]、MONet[10]、SAR-DCNN[11]、SAR-Transformer[12] 和 HTNet[13] 7 种代表性方法作为对比方法，分别进行仿真 SAR 图像和真实 SAR 图像 2 组实验。

2.2.1 仿真 SAR 图像实验

为验证边缘引导的双分支网络 SAR 图像相干斑抑制方法的有效性，先对仿真 SAR 图像进行实验。对于仿真 SAR 图像实验，采用峰值信噪比 (peak signal to noise ratio, PSNR)、结构相似性 (structural similarity index measure, SSIM) 及边缘保持指数 (edge preserve index, EPI) 这 3 种图像质量评价指标来评估抑斑性能。

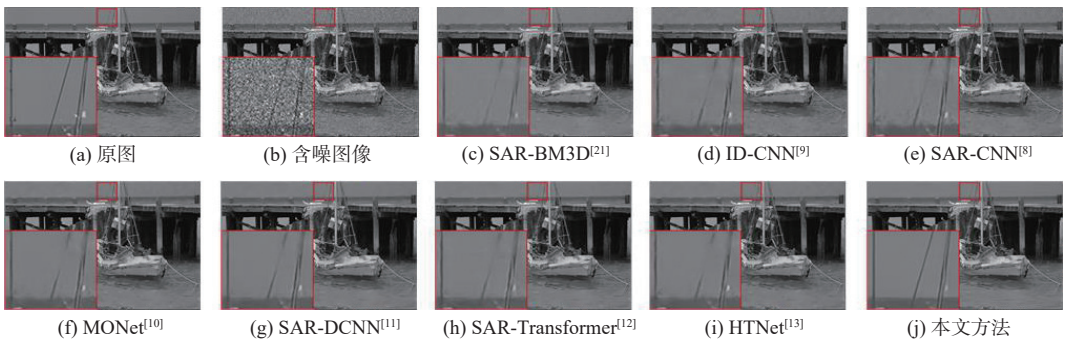


图 5 对 Kodak24 数据集 kodim11 仿真 SAR 图像抑斑的视觉效果对比 (L=4)

Fig. 5 Visual effect comparison of speckle suppression for kodim11 simulated SAR images in Kodak24 dataset (L = 4)

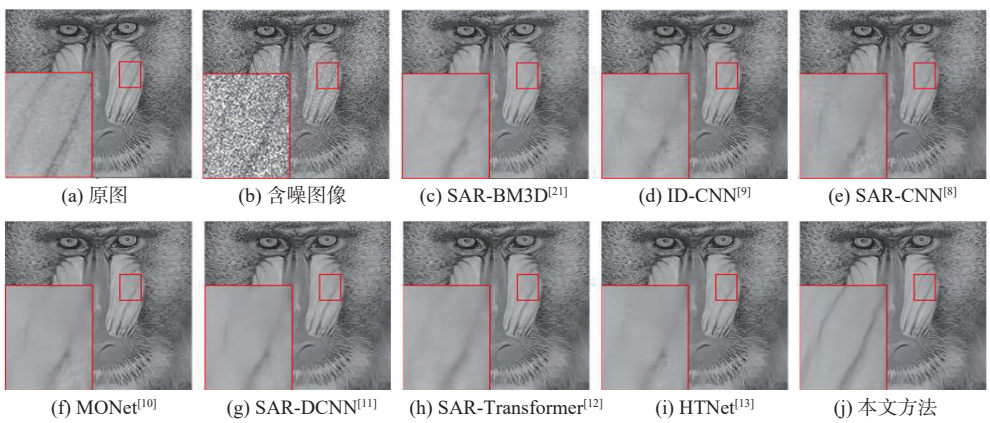


图 6 对 Classic5 数据集 baboon 仿真 SAR 图像抑斑的视觉效果对比 (L=4)

Fig. 6 Visual effect comparison of speckle suppression for baboon simulated SAR images in Classic5 dataset (L = 4)

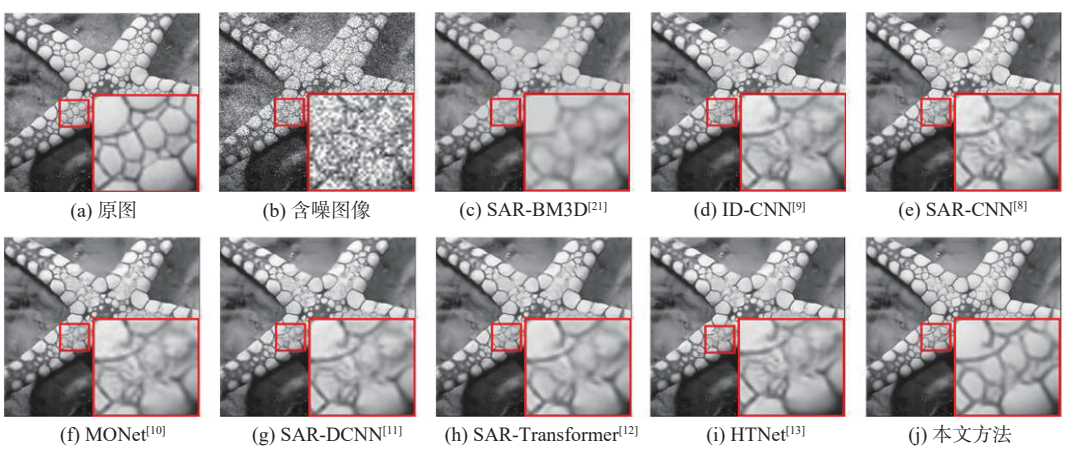


图 7 对 Set12 数据集 04 仿真 SAR 图像抑斑的视觉效果对比 (L=4)

Fig. 7 Visual effect comparison of speckle suppression for 04 simulated SAR images in Set12 dataset (L = 4)

图 5～图 7 为测试集中随机挑选的 3 个不同类别图像在 L=4 时不同方法的抑斑视觉效果对比。从图 5 和图 6 中可以看出, 其他方法均存在抑斑不足或纹理丢失的情况, 本文方法在实现一定去噪性能的基础上保留的细节信息更为丰富, 如对图 5 中“倾斜的绳子”和图 6 中“猩猩的面部褶皱”等边缘细节信息的重建效果更佳。图 7 中, SAR-BM3D 虽取得了一定的去噪性能, 但“海星触角”处的纹理模糊现象较为严重, 视觉成像不清晰; ID-CNN 和 SAR-CNN 相较于 SAR-BM3D 能够恢复出更多的图像细节, 在局部细节图中可以看到部分细节纹理, 但重建出的细节信息较为杂乱, 并不能很好地表征图像纹理; MONet、SAR-DCNN 和 HTNet 的抑斑效果有一定提升, 但在一些结构信息比较丰富的区域存在模糊现象; SAR-Transformer 的去噪效果较好, 但在细节放大后仍存在部分差异; 与其他 7 种方法相比, 本文方法能够在实现精细滤波的同时, 保留更多的边缘细节信息, “海星触角”处的纹理特征更为清晰, 获得最优的抑斑性能。

表 3 展示了 L=1,4,8,16 这 4 种不同视数的情况下, 不同方法在 3 个仿真 SAR 图像数据集上分别获取的 PSNR、SSIM 和 EPI 参数指标。可知, 相比于其

表 3　对仿真 SAR 图像抑斑的参数指标对比

Table 3　Parameter index comparison for speckle suppression of simulated SAR images

L	方法	PSNR/dB			SSIM			EPI		
		Set12	Classic5	Kodak24	Set12	Classic5	Kodak24	Set12	Classic5	Kodak24
1	SAR-BM3D[21]	24.01	24.57	23.90	0.767 0	0.776 7	0.791 4	0.957 9	0.942 6	0.948 5
	ID-CNN[9]	25.72	20.18	20.66	0.767 1	0.642 1	0.642 2	0.965 1	0.869 6	0.879 6
	SAR-CNN[8]	26.10	26.04	26.64	0.791 2	0.787 6	0.805 5	0.968 2	0.946 7	0.959 4
	MONet[10]	26.14	26.34	26.93	0.793 7	0.799 4	0.816 8	0.968 4	0.949 1	0.961 7
	SAR-DCNN[11]	26.25	26.45	27.03	0.796 6	0.805 2	0.822 0	0.969 2	0.950 1	0.962 5
	SAR-Transformer[12]	26.19	26.44	26.98	0.799 9	0.807 2	0.823 4	0.968 8	0.949 9	0.962 0
	HTNet[13]	26.22	26.53	26.91	0.808 3	0.821 4	0.836 7	0.969 1	0.949 4	0.961 3
	本文方法	**27.17**	**27.57**	**28.00**	**0.835 6**	**0.863 1**	**0.871 3**	**0.975 3**	**0.961 7**	**0.970 1**
4	SAR-BM3D[21]	27.85	28.40	28.01	0.856 7	0.884 5	0.888 9	0.980 7	0.972 2	0.974 9
	ID-CNN[9]	28.96	28.99	29.50	0.862 8	0.883 1	0.890 8	0.983 7	0.972 6	0.979 0
	SAR-CNN[8]	28.83	28.84	29.36	0.862 7	0.881 5	0.890 2	0.983 4	0.972 3	0.978 7
	MONet[10]	29.20	29.29	29.73	0.872 5	0.891 8	0.899 2	0.983 1	0.974 1	0.979 9
	SAR-DCNN[11]	29.31	29.39	29.82	0.874 9	0.895 0	0.902 2	0.984 9	0.974 6	0.980 3
	SAR-Transformer[12]	29.10	29.20	29.61	0.873 3	0.894 9	0.900 5	0.984 2	0.973 7	0.979 4
	HTNet[13]	29.13	29.18	29.65	0.870 7	0.888 6	0.897 2	0.984 3	0.973 4	0.979 6
	本文方法	**29.53**	**29.69**	**30.02**	**0.883 9**	**0.905 0**	**0.909 2**	**0.985 7**	**0.976 1**	**0.981 0**
8	SAR-BM3D[21]	29.52	29.98	29.82	0.888 3	0.916 8	0.922 2	0.986 4	0.979 7	0.982 4
	ID-CNN[9]	30.51	30.20	30.66	0.894 8	0.910 0	0.915 9	0.988 6	0.979 6	0.984 0
	SAR-CNN[8]	30.69	30.81	31.24	0.901 3	0.922 6	0.929 3	0.989 1	0.982 0	0.985 9
	MONet[10]	30.70	30.76	31.22	0.899 6	0.921 2	0.928 1	0.989 1	0.981 9	0.985 9
	SAR-DCNN[11]	30.83	30.88	31.32	0.903 1	0.924 1	0.930 1	0.989 4	0.982 2	0.986 1
	SAR-Transformer[12]	30.52	30.54	31.02	0.899 9	0.921 8	0.927 1	0.988 6	0.981 1	0.985 2
	HTNet[13]	30.67	30.71	31.16	0.900 2	0.920 6	0.927 4	0.989 0	0.981 6	0.985 7
	本文方法	**31.04**	**31.12**	**31.50**	**0.909 1**	**0.930 4**	**0.934 8**	**0.989 9**	**0.983 1**	**0.986 6**
16	SAR-BM3D[21]	31.27	31.58	31.68	0.917 3	0.942 9	0.948 0	0.990 7	0.985 8	0.988 2
	ID-CNN[9]	32.21	31.38	31.89	0.922 2	0.934 2	0.938 3	0.992 3	0.985 0	0.988 1
	SAR-CNN[8]	32.14	32.11	32.70	0.922 5	0.943 1	0.949 0	0.992 3	0.987 3	0.990 1
	MONet[10]	32.32	32.30	32.91	0.925 7	0.945 3	0.951 7	0.992 6	0.987 7	0.990 5
	SAR-DCNN[11]	32.30	32.39	33.02	0.926 9	0.946 8	0.952 7	0.992 7	0.987 9	0.990 6
	SAR-Transformer[12]	32.08	31.99	32.60	0.923 1	0.944 3	0.949 0	0.992 1	0.986 9	0.989 8
	HTNet[13]	32.32	32.28	32.88	0.925 1	0.945 1	0.951 1	0.992 5	0.987 6	0.990 4
	本文方法	**32.65**	**32.60**	**33.18**	**0.931 1**	**0.950 6**	**0.955 8**	**0.993 1**	**0.988 3**	**0.990 9**

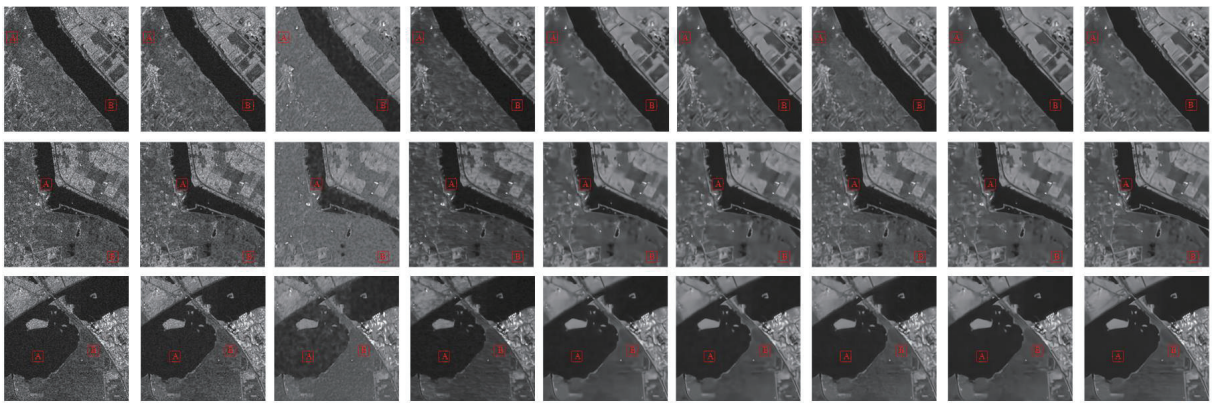
注: 加粗数值为最优值。

他 7 种方法，本文方法的 PSNR 增益分别为 1.96 dB、1.93 dB、0.71 dB、0.52 dB、0.42 dB、0.65 dB、0.54 dB，同时，SSIM 和 EPI 参数指标也最优。总的来说，本文方法在 4 种视数的定量评价指标上均优于其他方法。

2.2.2 真实 SAR 图像实验

为进一步验证本文方法的泛化能力，实验采用 Sentinel-1A 在 IW 扫描模式下拍摄的 Amsterdam 地区的真实 SAR 图像进行测试。先将图像裁剪为 256×256，再利用预先训练好的网络模型对真实 SAR 图像进行相干斑抑制，并采用 EPI 和等效视数(equivalent number of looks, ENL)2 个数据指标对其抑斑性能进行评估。图 8 为 Amsterdam 地区 3 个区域不同方法的抑斑视觉效果对比，其中，3 个区域分别命名为区域 1、2、3。本文随机选择明/暗 2 个同质区 A 和 B(图中 2 个红色方框处) 进行 ENL 定量评估。

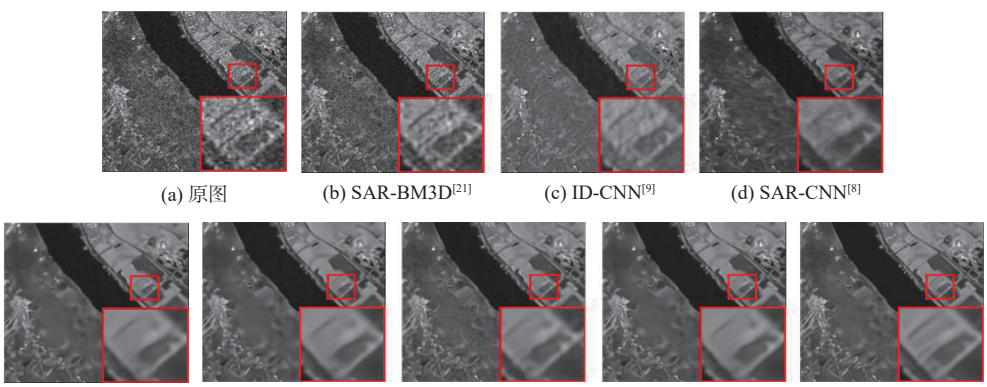
本文方法及其对比方法的局部细节对比如图 9～图 11 所示。可以看出，SAR-BM3D 的抑斑效果较



(a) 原图 (b) SAR-BM3D[21] (c) ID-CNN[9] (d) SAR-CNN[8] (e) MONet[10] (f) SAR-DCNN[11] (g) SAR-Transformer[12] (h) HTNet[13] (i) 本文方法

图 8 对 Amsterdam 地区真实 SAR 图像 3 个区域抑斑的视觉效果对比

Fig. 8 Visual effect comparison of speckle suppression in three regions for real SAR images in Amsterdam area

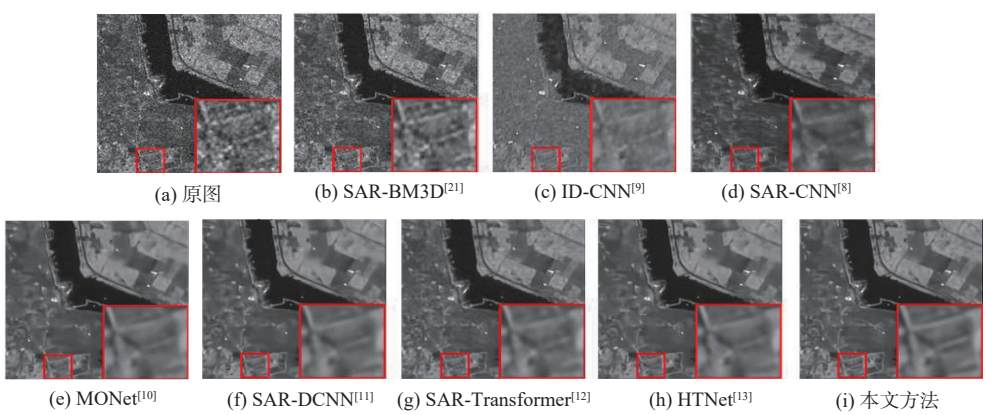


(a) 原图 (b) SAR-BM3D[21] (c) ID-CNN[9] (d) SAR-CNN[8]

(e) MONet[10] (f) SAR-DCNN[11] (g) SAR-Transformer[12] (h) HTNet[13] (i) 本文方法

图 9 对 Amsterdam 地区真实 SAR 图像区域 1 抑斑的局部细节对比

Fig. 9 Local detail comparison of speckle suppression in region 1 for real SAR images in Amsterdam area

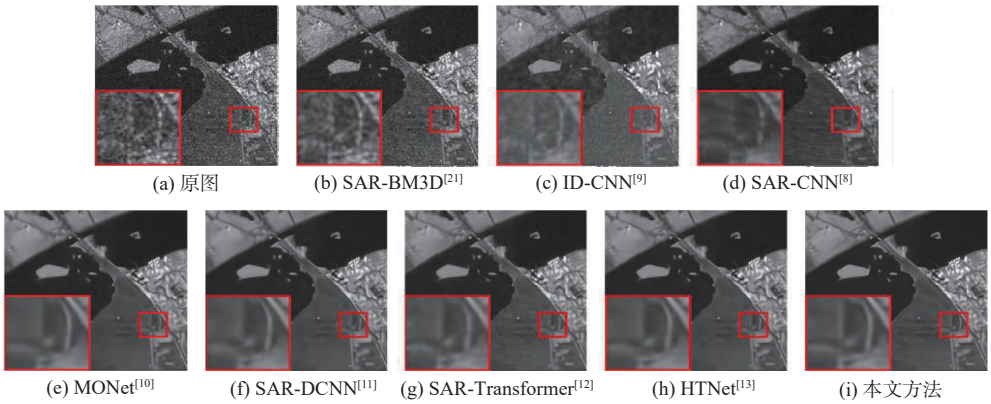


(a) 原图 (b) SAR-BM3D[21] (c) ID-CNN[9] (d) SAR-CNN[8]

(e) MONet[10] (f) SAR-DCNN[11] (g) SAR-Transformer[12] (h) HTNet[13] (i) 本文方法

图 10 对 Amsterdam 地区真实 SAR 图像区域 2 抑斑的局部细节对比

Fig. 10 Local detail comparison of speckle suppression in region 2 for real SAR images in Amsterdam area



(a) 原图　(b) SAR-BM3D[21]　(c) ID-CNN[9]　(d) SAR-CNN[8]

(e) MONet[10]　(f) SAR-DCNN[11]　(g) SAR-Transformer[12]　(h) HTNet[13]　(i) 本文方法

图 11　对 Amsterdam 地区真实 SAR 图像区域 3 抑斑的局部细节对比

Fig. 11　Local detail comparison of speckle suppression in region 3 for real SAR images in Amsterdam area

差, 图像抑斑后仍存在较多相干斑噪声, 有明显的抑斑不彻底问题。ID-CNN 和 SAR-CNN 的抑斑效果虽较 SAR-BM3D 有所提升, 但由于 ID-CNN 网络结构较浅导致的细节特征提取不充分及 SAR-CNN 利用 log 和 exp 同态变换导致的部分细节信息丢失的问题, 所获得的抑斑后图像结构信息丢失较多, 视觉成像较为模糊。MONet、SAR-DCNN、SAR-Transformer 和 HTNet 的去噪效果较好, 但仍然存在边缘区过于平滑或细节纹理丢失的问题。而本文方法能够较好地解决这一问题, 可以在实现 SAR 图像良好去噪效果的同时, 保留更多的边缘细节信息。

表 4 为图 8 中 Amsterdam 地区 3 个不同区域中区域 A 和区域 B 的 ENL 及 EPI 参数指标对比结果。可以看出, 本文方法在 2 个区域内的 ENL 指标, 除与 HTNet 方法区域 1 的区域 B 持平之外, 其余均高于其他方法, 说明本文方法的同质区更光滑, 去噪性能更好; 与其他方法相比, EPI 指标也位于最高值, 说明本文方法在边缘保持方面也具有良好的性能。

表 4　对真实 SAR 图像抑斑的参数指标对比

Table 4　Parameter index comparison for speckle suppression of real SAR images

区域	方法	ENL		EPI
		区域A	区域B	
1	SAR-BM3D[21]	67.10	4.73	0.866 3
	ID-CNN[9]	61.91	5.58	0.886 4
	SAR-CNN[8]	82.19	5.69	0.790 3
	MONet[10]	88.53	5.47	0.842 7
	SAR-DCNN[11]	124.49	5.65	0.843 7
	SAR-Transformer[12]	110.82	5.68	0.841 4
	HTNet[13]	115.75	**5.72**	0.864 2
	本文方法	**146.81**	**5.72**	**0.895 9**
2	SAR-BM3D[21]	71.29	48.44	0.861 0
	ID-CNN[9]	23.91	200.09	0.812 7
	SAR-CNN[8]	66.73	114.66	0.806 1
	MONet[10]	60.92	220.34	0.841 5
	SAR-DCNN[11]	104.03	215.62	0.836 4
	SAR-Transformer[12]	66.13	153.32	0.827 1
	HTNet[13]	97.05	207.31	0.863 0
	本文方法	**127.09**	**224.45**	**0.875 2**
3	SAR-BM3D[21]	115.69	76.93	0.883 8
	ID-CNN[9]	75.60	146.93	0.860 5
	SAR-CNN[8]	191.25	104.97	0.839 1
	MONet[10]	470.98	147.98	0.856 0
	SAR-DCNN[11]	484.92	162.35	0.868 9
	SAR-Transformer[12]	411.91	119.64	0.840 4
	HTNet[13]	581.17	150.81	0.880 2
	本文方法	**822.13**	**164.44**	**0.908 9**

注: 加粗数值为最优值。

2.3　消融实验

2.3.1　不同模块对网络性能的影响

为验证 DCEEB 模块、注意力机制 (Attention) 及 MPMFFB 模块的有效性, 利用控制变量法来衡量不同模块对本文模型的贡献。实验以去除 DCEEB、Attention 和 MPMFFB 这 3 个模块后的双分支网络为基准网络, 当视数 L=1 时, 进行包括 3 个模块单独作用、两两组合及 3 个模块共同作用共 8 组实验, 并以 PSNR、SSIM 和 EPI 为参数指标在 Set12 数据集上进行测试。

不同模块设置的抑斑结果如表 5 所示。以 PSNR 指标为例, 单独采用 DCEEB、Attention 和 MPMFFB 优化策略时, 其指标分别为 26.88 dB、26.39 dB 和 26.40 dB, 相比基准网络的 PSNR 指标 26.38 dB, 增益分别为 0.50 dB、0.01 dB、0.02 dB, 说明 3 种优化策略均能有效提升抑斑性能。当同时采用 3 个优化策略时, 本文模型的 PSNR 值高达 27.17 dB, 相比基准网络提升 0.79 dB, 表明 3 个模块共同作用时能提取更为精细的特征, 像素点的差异

表5 不同模块设置实验结果

Table 5 Experimental results for different block settings

基准网络	DCEEB	Attention	MPMFFB	PSNR/dB	SSIM	EPI
√				26.38	0.812 6	0.970 3
√	√			26.88	0.839 0	0.973 6
√	√	√		27.07	0.833 9	0.974 9
√		√		26.39	0.811 9	0.970 3
√		√	√	26.50	0.815 2	0.971 1
√			√	26.40	0.811 1	0.970 3
√	√		√	27.00	0.833 4	0.974 4
√	√	√	√	27.17	0.835 6	0.975 3

更小。在表5中，除Attention及MPMFFB单独作用时，EPI指标保持不变，SSIM指标分别下降0.000 7、0.001 5外，其余数据指标均呈上升趋势，另外，当Attention和MPMFFB这2个模块同时作用于网络时，相比于基准网络，PSNR、SSIM和EPI指标分别提升0.12 dB、0.002 6、0.000 8。结合8组实验的综合情况，2个及以上模块组合作用比各模块单独作用对模型性能的提升更大，将3个模块共同作用于抑斑网络的性能最优。

2.3.2 不同网络结构对网络性能的影响

1）单/双分支网络模型。为验证本文双分支网络的有效性，采用单分支网络模型和双分支网络模型分别进行抑斑处理，并分析其对抑斑性能的影响。在视数L=4时，本文设置3组消融实验进行分析：①仅保留MAENet子网络；②仅保留CARNet子网络；③结合CARNet子网络和MAENet子网络的双分支抑斑网络。不同网络结构的抑斑结果如表6所示。可以看出，采用MAENet和CARNet单分支网络模型所获得的参数指标要逊色于双分支网络模型，说明双分支网络模型的抑斑性能更好。本文方法在PSNR、SSIM、EPI指标上分别达到了29.53 dB、0.883 9和0.985 7。

表6 单/双分支网络模型实验结果

Table 6 Experimental results of single/dual-branch network model

MAENet	CARNet	PSNR/dB	SSIM	EPI
√		29.37	0.882 0	0.985 2
	√	29.45	0.882 3	0.985 4
√	√	29.53	0.883 9	0.985 7

2）不同双分支网络结构设置。为进一步探索不同双分支网络结构对网络性能的影响，将MAENet子网络和CARNet子网络在双分支网络中的位置进行不同组合。以MPMFFB模块为分界线，将MPMFFB模块获取输入数据的分支命名为分支1，则MPMFFB模块的输出特征融入到分支2。当视数L=1时，在Set12数据集上进行包括双分支均采用MAENet子网络进行构建，双分支均采用CARNet子网络进行构建，分支1采用MAENet子网络、分支2采用CARNet子网络及分支1采用CARNet子网络、分支2采用MAENet子网络共4组实验。

表7为不同双分支网络结构设置的实验结果。可以看出，2个分支采用不同网络结构所获取到的数据指标优于均采用相同网络结构；相比于分支1采用CARNet、分支2采用MAENet的策略，将MAENet子网络的输出经过MPMFFB模块再输入到CARNet子网络的情况下，PSNR、SSIM和EPI分别下降0.07 dB、0.000 5和0.000 6。综上，本文所使用的双分支网络结构，即分支1采用CARNet、分支2采用MAENet，2个分支相互促进，获得最佳的抑斑效果。

表7 不同双分支网络结构设置实验结果

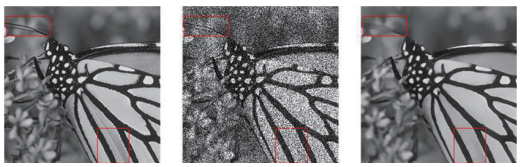
Table 7 Experimental results of different dual-branch network structure settings

分支1	分支2	PSNR/dB	SSIM	EPI
MAENet	MAENet	26.79	0.826 4	0.972 7
CARNet	CARNet	26.98	0.833 6	0.974 0
MAENet	CARNet	27.10	0.835 6	0.974 7
CARNet	MAENet	27.17	0.836 1	0.975 3

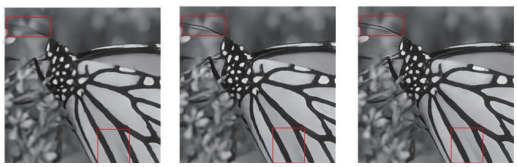
3）不同注意力机制设置。为验证不同注意力机制组合对网络性能的影响，当视数L=1时，在Set12数据集上分别进行Model 1、Model 2、Model 3、Model 4这4组实验。其中，Model 1为CARNet和MAENet子网络中均添加RCABlock，Model 2为CARNet和MAENet子网络中均添加CBAM，Model 3为CARNet子网络中添加CBAM，而MAENet子网络中添加RCABlock，Model 4为本文所使用的注意力组合，即CARNet子网络中添加RCABlock，而MAENet子网络中添加CBAM。

图12为4组不同注意力机制设置的抑斑效

果。可以看出, Model 1 和 Model 2 在昆虫“触角”处均存在模糊现象, Model 3 与 Model 1、Model 2 相比, 模糊现象有一定提升, 但在昆虫“翅膀”的条纹处出现了一定的失真, Model 4 重建出的图像质量更佳。



(a) 原始图像　(b) 含噪图像　(c) Model 1



(d) Model 2　(e) Model 3　(f) Model 4

图 12　不同注意力机制设置的抑斑视觉结果

Fig. 12　Visual results of speckle suppression with different attention mechanism settings

表 8 为不同注意力机制组合所获得的数据指标结果。可以看出, 本文采用的 Model 4, 即将 CARNet 与 RCABlock、MAENet 与 CBAM 进行组合, 可获得最佳的客观参数指标结果, 最大限度地提升模型的抑斑性能。

表 8　不同注意力机制设置数据指标结果

Table 8　Data index results of different attention mechanism settings

模型	PSNR/dB	SSIM	EPI
Model 1	27.07	0.833 8	0.974 5
Model 2	27.05	0.834 6	0.974 7
Model 3	27.13	0.835 2	0.975 1
Model 4	27.17	0.835 6	0.975 3

3　结　论

针对现有深度学习 SAR 图像相干斑抑制方法难以同时提升相干斑抑制与边缘细节保持性能的问题, 本文提出了一种边缘引导的双分支网络 SAR 图像相干斑抑制方法。

1) 采用双分支网络进行 SAR 图像相干斑抑制, 2 个分支相互成就, 克服了单一分支网络存在的局限性。相比于单分支网络 MAENet 和 CARNet, 本文的双分支网络模型在 L=4 时 Set12 数据集上的 PSNR 分别提升 0.16 dB、0.08 dB。

2) 将原始图像及其边缘强度图一同送入网络中进行训练, 最大限度地增强边缘信息的表达能力, 有效改善了 SAR 图像噪声抑制过程中的边缘纹理丢失问题。加入图像边缘信息后, L=1 时, Set12 数据集的 PSNR、SSIM、EPI 指数分别提升 0.50 dB、0.026 4、0.003 3。

3) 引入注意力机制, 使模型能够在提取到更多细节特征的同时增强其特征表达能力。当视数 L=1 时, Set12 数据集的 PSNR、SSIM、EPI 指数分别达到 26.39 dB、0.811 9、0.970 3。

4) 在双分支网络的基础上, 提出基于多分支并行的多尺度特征融合模块, 增强 2 个分支间的信息交流。以 PSNR 指标为例, 加入多尺度特征融合模块后, 当 L=1 时, Set12 数据集的 PSNR 提升 0.02 dB。

通过仿真 SAR 图像及真实 SAR 图像实验表明, 本文方法无论是在定量指标还是视觉成像效果方面均表现出优异的成绩, 但该网络模型只适用于特定类型的 SAR 图像降噪, 在后续的研究中, 将继续优化网络模型, 提升模型的学习能力, 以适用于更多类型的 SAR 图像进行相干斑抑制。

参考文献 (References)

[1] SINGH P, DIWAKAR M, SHANKAR A, et al. A review on SAR image and its despeckling[J]. Archives of Computational Methods in Engineering, 2021, 28(7): 4633-4653.

[2] 李敬曼, 朱磊, 张博, 等. 一种带约束搜索窗的非局部平均相干斑抑制算法[J]. 西安交通大学学报, 2020, 54(10): 54-62.
LI J M, ZHU L, ZHANG B, et al. Non-local means suppression algorithm of speckle with restrained search window[J]. Journal of Xi'an Jiaotong University, 2020, 54(10): 54-62(in Chinese).

[3] PAN Y, MENG Y H, ZHU L. SAR image despeckling method based on improved Frost filtering[J]. Signal, Image and Video Processing, 2021, 15(4): 843-850.

[4] 吴天琦, 肖文, 李仁剑, 等. 基于时域迭代小波变换的单分子定位图像背景去噪[J]. 中国激光, 2021, 48(13): 1307001.
WU T Q, XIAO W, LI R J, et al. Single-molecule localization image background denoising based on time-domain iterative wavelet transform[J]. Chinese Journal of Lasers, 2021, 48(13): 1307001 (in Chinese).

[5] CHU T Y, TAN Y M, LIU Q, et al. Novel fusion method for SAR and optical images based on non-subsampled shearlet transform[J]. International Journal of Remote Sensing, 2020, 41(12): 4590-4604.

[6] YU Y J, ACTON S T. Speckle reducing anisotropic diffusion[J]. IEEE Transactions on Image Processing, 2002, 11(11): 1260-1270.

[7] AJA-FERNÁNDEZ S, ALBEROLA-LÓPEZ C. On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering[J]. IEEE Transactions on Image Processing, 2006, 15(9): 2694-2701.

[8] CHIERCHIA G, COZZOLINO D, POGGI G, et al. SAR image despeckling through convolutional neural networks[C]//Proceedings of the IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2017: 5438-5441.

[9] WANG P Y, ZHANG H, PATEL V M. SAR image despeckling using a convolutional neural network[J]. IEEE Signal Processing Letters, 2017, 24(12): 1763-1767.

[10] VITALE S, FERRAIOLI G, PASCAZIO V. Multi-objective CNN-based algorithm for SAR despeckling[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(11): 9336-9349.

[11] PASSAH A, AMITAB K, KANDAR D. SAR image despeckling using deep CNN[J]. IET Image Processing, 2021, 15(6): 1285-1297.

[12] PERERA M V, BANDARA W G C, VALANARASU J M J, et al. Transformer-based SAR image despeckling[C]//Proceedings of the IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2022: 751-754.

[13] CHENG L, GUO Z C, LI Y, et al. Two-stream multiplicative heavy-tail noise despeckling network with truncation loss[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5213817.

[14] LIN H X, ZHUANG Y H, HUANG Y, et al. Self-supervised SAR despeckling powered by implicit deep denoiser prior[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 4514705.

[15] 朱磊, 李敬曼, 潘杨, 等. 自适应调节滤波强度的SAR图像非局部平均抑斑算法[J]. 电子与信息学报, 2021, 43(5): 1258-1266.
ZHU L, LI J M, PAN Y, et al. SAR image despeckling algorithm using non-local means with adaptive filtering strength[J]. Journal of Electronics & Information Technology, 2021, 43(5): 1258-1266(in Chinese).

[16] ZHANG Y L, LI K P, LI K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 294-310.

[17] ZHANG K, LI Y W, LIANG J Y, et al. Practical blind image denoising via Swin-Conv-UNet and data synthesis[EB/OL]. (2022-03-24) [2023-05-31]. https://arxiv.org/abs/2203.13278v4.

[18] 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.

[19] 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.

[20] LEE J S, HOPPEL K. Noise modeling and estimation of remotely-sensed images[C]//Proceedings of the 12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 1989: 1005-1008.

[21] PARRILLI S, PODERICO M, ANGELINO C V, et al. A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage[J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 50(2): 606-616.

SAR image coherence speckle suppression method based on edge-guided dual-branch network

ZHU Lei*, YAO Tongyu, CHE Chenjie, YAO Lina, ZHANG Bo, PAN Yang

(School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: In order to further improve the coherence speckle suppression and edge preservation performance of synthetic aperture radar (SAR) images by deep learning methods, a coherence speckle suppression method based on edge-guided dual-branch network was proposed. The method constructed a new speckle suppression network model, which consisted of an edge information extraction block and a dual-branch speckle suppression network. Firstly, a dense cascade strategy was used to build the edge information extraction block to enhance the edge perception capability of the model. Secondly, the channel attention-based residual despeckling network (CARNet), the mixed attention-based enhanced despeckling network (MAENet), and the multi-branch parallel based multi-scale feature fusion block (MPMFFB) were used to form a dual-branch speckle suppression network, so as to better preserve edge details while suppressing coherence speckles. The experimental results show that the proposed method has better coherence speckle suppression and edge preservation performance compared with recent state-of-the-art methods such as SAR-Transformer and HTNet. For the simulated SAR images, the peak signal to noise ratio, structural similarity index measure, and edge preserve index are improved by 0.96 dB, 2.60%, and 0.60% on average, respectively. For the real SAR images, the equivalent number of looks is improved by more than 14.12%, and the edge preserve index is improved by 4.52% on average.

Keywords: image denoising; synthetic aperture radar image; coherence speckle suppression; dual-branch network; multi-scale feature fusion

Received: 2023-06-06; **Accepted**: 2023-09-14; **Published Online**: 2023-09-28 16:51

URL: link.cnki.net/urlid/11.2625.V.20230928.1126.001

Foundation items: National Natural Science Foundation of China (61971339); Key Research and Development Program of Shaanxi (2019GY-113); Natural Science Basic Research Program of Shaanxi (2019JQ-361)

* **Corresponding author**. E-mail: zhulei791014@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0336

融合聚类特征的改进MRF轨面缺陷分割方法

闵永智[1,*]，刘洋[1,2]

(1. 兰州交通大学 自动化与电气工程学院，兰州 730070；2. 包头铁道职业技术学院 机械工程系，包头 014060)

摘 要：针对轨面缺陷样本数量少、种类多的特点,以及在真实场景下存在迁移学习效果不稳定、阈值分割易受环境因素影响的问题，提出一种零样本的改进马尔可夫随机场轨面缺陷分割方法。对采集的数据使用Gabor函数进行处理，突出缺陷特征，降低数据维度，得到降维特征图；对处理后的特征图进行Kmeans聚类，缩减数据分布，降低反光和阴影的影响，并将聚类结果作为预分类矩阵；通过降维特征图和预分类矩阵构建改进马尔可夫随机场(MRF)双层图模型，并进行推理；根据模型推理出的预分类矩阵特征值来分析缺陷部分的局部几何结构；标记出缺陷区域，并完成缺陷分割。使用自采样数据集进行对比实验和消融实验，结果表明：所提方法在自采样数据集上的像素准确率、平均像素准确率、加权交并比、均交并比分别达到93.6%、80.7%、89.4%、68.2%，超过对比检测方法精度。

关 键 词：图像处理；轨面缺陷；马尔可夫随机场；概率图模型；实例分割
中图分类号：TP391.4
文献标志码：A **文章编号**：1001-5965(2025)06-1863-10

铁路是中国交通系统的重要组成部分之一,随着中国铁路既有线提速及重载和高速铁路的发展,其运行环境复杂多变且易受各类客观因素影响,在列车运行过程中往往会有伤损的出现,该问题不仅会影响车辆的舒适度和平稳性,严重时还会危及行车安全。因此,能否及时检测出钢轨缺陷已经成为当前铁路运输部门的一项重要任务。随着深度学习在诸多领域的广泛使用,以及机器视觉技术的蓬勃发展,使用图像采集和图像处理技术完成轨面缺陷的自动检测已成为新的研究方向。深度学习和机器视觉的各种方法也已经广泛应用于轨面缺陷检测,其中,阈值分割和深度学习是当前轨面缺陷分割的2种主流方法[1-10]。

轨面缺陷阈值分割方法主要有背景差分法、最大熵值法等[1-2]。王国伟等[3]使用数据集建立图像背景模型,通过模型背景差分降低光照和反射的影响,结合遗传算法与最大熵值法计算最佳分割阈值,通过最佳分割阈值对轨面缺陷进行分割。程天栋等[4]使用Hough变换提取轨面区域,结合模糊理论与超熵理论计算图像的模糊超熵值,使用计算后的最优分割阈值作为适应度函数,通过遗传算法确定最优分割阈值,对轨面缺陷进行分割。曹义亲等[5]确定缝面缺陷占比较少的特征,根据全局缺陷比例进行阈值截断,增加差分图对比度,根据此特征进行最大熵分割。张辉等[6]采用马尔可夫模型对模糊规则前提部分进行推理,通过图像局部空间约束,迭代出新的自适应模糊集,对轨面缺陷实现自动分割。基于上述研究,轨面缺陷阈值分割方法的效果往往受轨面自身的光照条件影响,模型会对数据中的缺陷与轨面反光、阴影两者之间的关系造成误判,需要在常规的缺陷分割模型基础上考虑轨面的全局信息以降低光照对分割的影响。

收稿日期：2023-06-09；**录用日期**：2023-06-30；**网络出版时间**：2023-07-21 17:15
网络出版地址：link.cnki.net/urlid/11.2625.V.20230721.1442.002
基金项目：国家自然科学基金(62066024)；2023甘肃省优秀研究生“创新之星”项目(2023CXZX-615)
*通信作者. E-mail：minyongzhi@mail.lzjtu.cn

引用格式：*闵永智，刘洋. 融合聚类特征的改进MRF轨面缺陷分割方法[J]. 北京航空航天大学学报，2025，51(6)：1863-1872.*
MIN Y Z，LIU Y. Improved MRF rail surface defect segmentation method based on fusion of clustering features[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1863-1872 (in Chinese).

基于深度学习的轨面缺陷分割方法也是当前的重要研究方向。周雯等[7]使用特征金字塔网络提取缺陷在多个尺度上的融合特征,采用可变感受野的可变形卷积网络(deformable convolutional network, DCN)适应缺陷形态的多样性,实现对缺陷的精准分割。刘俊博等[8]提出一种少样本学习检测方法,通过改进分类损失函数和多层级特征融合的轨面缺陷检测网络,降低样本数量不平衡对模型分类性能的影响,实现少样本训练的轨面缺陷检测。赵晨阳等[9]利用注意力模块引导特征丛生成,提高缺陷检测速度,降低模型部署成本,对像素级标签进行数据处理,使标签的不同区域获得不同的关注,降低标签依赖性,使得检测模型鲁棒性提高。罗晖等[10]使用Retinex与Cascade R-CNN检测缺陷,采用交并比(intersection over union, IoU)平衡采样、完整交并比损失(complete intersection over union loss, CIoU)函数、感兴趣区域对齐,解决训练样本IoU分布与困难样本IoU分布不平衡、感兴趣区域池化中取整量化的问题,消除因样本不足导致的过拟合现象。基于上述研究,由于轨面缺陷具有缺陷样本少、种类繁多的特点,且在不同的线路上,设备和客观因素导致采集结果特征、色域、对比度等相差较大,在已有的缺陷数据集基础上,训练后的深度学习模型不能直接迁移到另外线路的检测设备上,直接使用其他设备的数据集会在实际检测与分割的过程中出现效果不理想、分割结果错误等问题,常规的深度学习模型检测并不能很好地解决这一问题[8,11-13]。

综上,阈值分割方法使用不同算法进行阈值划分,从阈值的角度降低反光、阴影的影响,区分缺陷与背景,但在整体数据集下,计算合适的阈值这一过程较为复杂;深度学习方法在仅有的缺陷数据集上使用加入注意力、感受野等方法,提取背景与缺陷的特征,以达到区分缺陷、背景与反光阴影的效果,本质上是加深模型在仅有样本数据上的拟合程度。相比于以上2种方法,概率图模型可以通过系统刻画输入变量与目标变量的关系,从数据中自动学习模型特征,并按照一定的原则在模型中进行推理。其中,马尔可夫随机场(Markov random field, MRF)是当前应用较广的无向概率图模型,不同于深度学习方法需要依赖大量的训练样本完成模型训练,该方法根据输入数据自行计算迭代推理且完成分割任务,已经在诸多分割任务场景下得到广泛应用[14-19],可以作为轨面缺陷分割的一种方法。

因此,本文提出一种融合Kmeans聚类特征的改进MRF轨面缺陷分割方法。不同于张辉等[6]使用的MRF缺陷分割方法,首先,使用不同方向的Gabor函数对输入图像进行处理,提升缺陷与钢轨部分的像素矩阵对比度,并降低输入数据的维度。其次,使用Kmeans算法[20]对处理后的像素灰度矩阵进行聚类,并生成对应的标签矩阵作为MRF的预分类矩阵。然后,将预分类矩阵结果作为标记层的迭代初始值,将Gabor卷积处理后的结果作为图像层输入,组成MRF双层图模型,使用条件迭代法进行最大后验概率迭代推理,得到分类后的像素矩阵,并将标记为缺陷的部分进行分割,展示最终结果。最后,对本文方法的改进部分进行消融实验分析,并利用自采样钢轨缺陷图像,与常见缺陷分割方法进行对比分析,以验证本文方法改进的有效性。

1 融合聚类特征的改进MRF模型

图像分割,本质上是一种多标签图像标记问题。MRF在给定拓扑结构下会在其节点嵌入独立性,可以刻画相邻像素之间的交互,其状态只与领域节点的状态有关,将其预期的相互依赖关系编码,整体图像上的局部平滑性和区域联通性可以通过局部相邻点间的依赖关系来系统执行[13]。

传统的MRF是将超像素分割作为模型建立的第一步[6,13],根据超像素分割结果与图像数据进行推理迭代,在实际的轨检设备检测时,缺陷的大小、形状、种类是随机且复杂的,且各类钢轨的自身环境、光照条件等各不相同,虽然传统的MRF可以刻画相邻像素关系,但这些客观因素增加了MRF模型推理的难度,并且会影响模型分割的准确性,同时由于超像素分割的组数阈值不好判定,加之轨面缺陷图像包含采集过程中的阴影和反光等不确定因素,使用超像素分割往往会扩大不确定因素对分割结果的影响,掩盖缺陷部分的特征,对分割结果造成极大的影响[14-15]。

相比于传统模型框架,本文改进MRF模型具有考虑轨面的全局信息以降低光照对分割的影响和根据数据自身的像素矩阵分布来判断该数据中缺陷部分的特征等特点。

1.1 Gabor卷积处理

对于钢轨缺陷图像来讲,钢轨的自身环境、光照条件等外界环境因素会影响模型的识别效果,因此,本文改进的MRF方法使用Gabor函数进行卷积预处理,以降低部分钢轨轨面反光及阴影带来的影响,同时处理后的数据特征突出且数据维度降低。Gabor函数的频率和方向类似于人的视觉,通常被用于分析某个点或区域是否在特定方向存在

特定频率的内容。Gabor 函数[13] 表达式为

$$G=\exp\left(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}\right)\exp\left(\mathrm{i}\left(2\pi\frac{x'}{\lambda}+\varphi\right)\right) \tag{1}$$

式中: $x'=x\cos\theta+y\sin\theta$, $y'=-x\sin\theta+y\cos\theta$, x、y表示像素坐标位置, θ为方向; λ为波长; σ为高斯因子的标准差; γ为空间纵横比; φ为相位偏移。

当给定输入图像 $\boldsymbol{M}\in\mathbf{R}^{h,w}$时($h$、$w$ 分别为输入图像像素值的长、宽尺寸), 将图像$\boldsymbol{M}$与 Gabor 滤波器作如下卷积运算, 计算过程如下:

$$\boldsymbol{F}_{\alpha\beta}=\boldsymbol{M}\otimes G \qquad \boldsymbol{F}_{\alpha\beta}\in\mathbf{R}^{\alpha,\beta} \tag{2}$$

式中: $\boldsymbol{F}_{\alpha\beta}$为卷积得到的特征图, α、β分别为卷积特征图像素值长、宽尺寸。

Gabor 卷积得到的特征图进一步突出了轨面缺陷与钢轨背景在像素级别的差异, 同时将输入图像的数据维度降低, 简化后续计算过程, 提升整体模型的性能。

1.2　融合聚类特征

Kmeans 聚类是在考虑数据全局信息的基础上对数据进行聚类[20], 对输入数据完成分类并输出分类矩阵, 且由于像素矩阵中数值范围为 0 ～255, 整体数值分布范围较大, 进行初步预分类在一定程度上会提升检测效果, 降低算法时间复杂度。因此, 使用 Kmeans 聚类算法, 对像素矩阵进行初步聚类分析, 并将聚类结果生成 label 标签矩阵作为 MRF 的预分类矩阵, 在预分类矩阵的基础上进行推理迭代, 不仅可以保留原始的空间信息, 还可以简化计算推理迭代的过程。

Kmeans 通过计算点与点之间的距离相似度来确定最佳类别归属[20]。对于 Gabor 卷积处理的特征图$\boldsymbol{F}_{\alpha\beta}$, 指定聚类中心的个数为$k$, 并通过将样本分离到$k$个簇中来最小化目标函数。Kmeans 算法的目标函数为

$$f_{\mathrm{SSE}}=\sum_{i=1}^{k}\sum_{s\in\boldsymbol{C}_i}(\boldsymbol{C}_i-\boldsymbol{s})^2 \tag{3}$$

对目标函数求偏导数, 如式 (4) 所示, 化简结果如式 (5) 所示。

$$\frac{\partial}{\partial\boldsymbol{C}_k}f_{\mathrm{SSE}}=\frac{\partial}{\partial\boldsymbol{C}_k}\sum_{i=1}^{k}\sum_{s\in\boldsymbol{C}_i}(\boldsymbol{C}_i-\boldsymbol{s})^2=\boldsymbol{0} \tag{4}$$

$$\boldsymbol{C}_i=\frac{1}{m_i}\sum_{s\in\boldsymbol{C}_i}\boldsymbol{s} \tag{5}$$

式中: m_i为簇中点的数量; $\boldsymbol{C}_i$为该组点的均值向量; $\boldsymbol{s}$为输入图像在各个特征维度上的值。

通过更新簇类的均值向量进行模型迭代, 当均值向量不再改变时, 模型构建完成, 使得数据与其最接近的聚类中心$\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_k$之间的距离尽可能小。Kmeans 预分类示例如图 1 所示。针对输入的特征图, Kmeans 按照设定的分类阈值进行像素与点之间的距离相似度计算, 并得出分类结果。

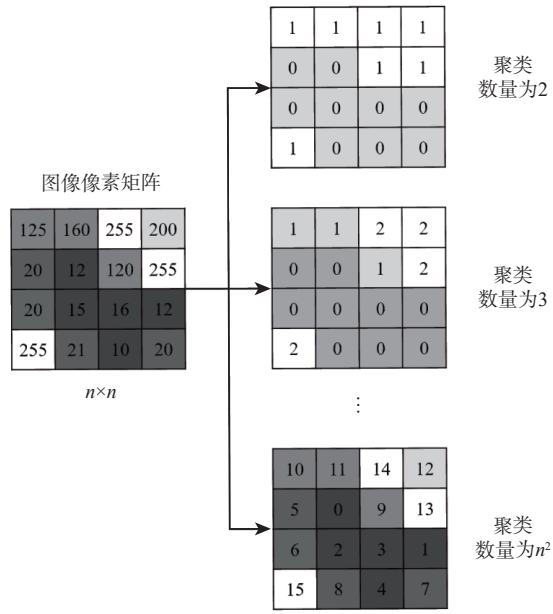


图 1　像素矩阵 Kmeans 预分类示例

Fig. 1　Pixel matrix K-means pre-classification example

1.3　改进 MRF 模型架构

给定输入向量$\boldsymbol{X}=[X_n]_{n=1}^{N}, n\in\nu=\{1,2,\cdots,N\}$, 其中, X_n为钢轨图像中某一像素点或一组图像像素强度或特征, ν为图像域, 即二维图像网络; 给定输出向量$\boldsymbol{Y}=[Y_n]_{n=1}^{N}$, 其中, Y_n为X_n的标签, 假设K个类别, 即$Y_n\in\tau=\{l_1,l_2,\cdots,l_K\}$, τ为图像像素可取范围内的对象标签值。

如图 2 所示, 顶部为标记层, 其中每个节点Y_n表示每个像素或一组像素的输出标签, 底层为图像层, 其中每个节点X_n表示一个或一组图像像素的特征或强度。将 Gabor 处理及 Kmeans 聚类结果代入图 2 所示的双层 MRF 架构, 图 2 中的X_n表示 Gabor 卷积结果得到的特征图, 即式 (2) 得到的$\boldsymbol{F}_{\alpha\beta}$, Y_n表示特征图经 Kmean 预分类后对应的类别。

由此, 将图像分割问题转化为最大后验 (maximum a posteriori estimation, MAP) 概率状态推理问

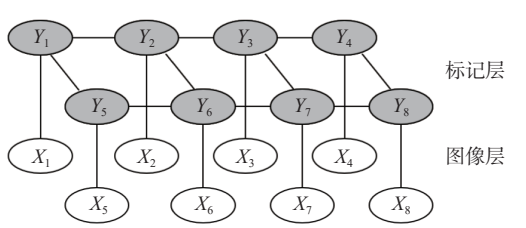


图 2　图像分割双层图模型

Fig. 2　Image segmentation two-layer graph model

题[13,19]，如式(6)所示，通过计算标记层的最大后验概率对图像层类别结果进行迭代推理。

$$y^* = \arg\max P_G(Y_n|X_n) \tag{6}$$

式中：P_G为X_n和Y_n的联合概率分布。

1.4 定义模型

1.4.1 局部领域

领域定义标记节点之间的连接，MRF进行图像标记，随机场遵循二维点阵图，其每个节点可对应一个像素矩阵。考虑到二阶领域系统能观测更多的领域信息，选用二阶领域系统作为局部领域的定义范围。一阶、二阶局部领域示意图如图3所示，局部领域表示每个节点与领域节点是否建立观测关系。

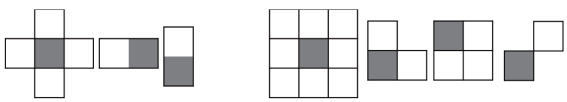


图3 一阶、二阶局部领域示意图

Fig. 3 One-order and two-order local domains

1.4.2 参数学习

标记观测MRF的势函数包含标记节点的成对势函数和标记节点与之对应图像节点直接的一元势函数，其中，一元势函数表示标记于其图像特征之间的关系[13,16]，成对势函数则表示局部数据的平滑性。在给定标记观测值的条件下，一元能量函数等于负似然对数，如下：

$$E(s_{\alpha,\beta}|z_{\alpha,\beta}) = -\ln p(s_{\alpha,\beta}|z_{\alpha,\beta}) \tag{7}$$

对于一元连续图像特征，标记为l的一元势函数为

$$\phi(s_{\alpha,\beta}|z_{\alpha,\beta}=l) = \frac{1}{\sqrt{2\pi}\sigma_l}\exp\left(-\frac{(s_{\alpha,\beta}-\mu_l)^2}{2\sigma_l^2}\right) \tag{8}$$

一元能量函数$E(s_{\alpha,\beta}|z_{\alpha,\beta}=l)$可以用二次项表示，如下：

$$E(s_{\alpha,\beta}|z_{\alpha,\beta}=l) = \frac{(s_{\alpha,\beta}-\mu_l)^2}{2\sigma_l^2} + \frac{1}{2}\ln\sigma_l^2 \tag{9}$$

式中：$s_{\alpha,\beta}$为图像层特征或强度；$z_{\alpha,\beta}$为标记层特征或强度；σ_l为方差；μ_l为均值。

1.5 模型推理方法

给定输入的标记层后，MRF模型在此基础上进行迭代计算，并在图像层进行MAP推理。MAP推理方法包括精确方法和近似方法，考虑到精确推理计算成本大，本文采用占用算力相对较少的近似推理方法，即条件迭代模式(iterated conditional modes, ICM)方法[21-22]。改进后MRF模型建立及推理过程如图4所示。

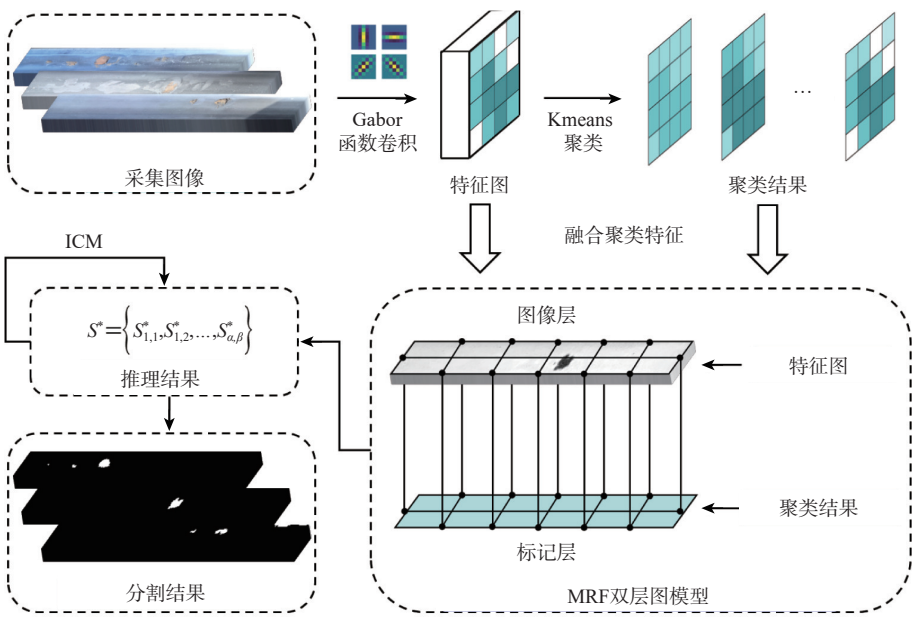


图4 改进MRF模型建立及推理过程

Fig. 4 Improve MRF model building and reasoning process

初始化所有节点，基于条件链式法则[11,13]的条件概率分布如下：

$$p(S|z) \approx \prod_{i=1}^{\alpha\times\beta} p(s_i|s_{-i},z) = \prod_{i=1}^{\alpha\times\beta} p(s_i|N_{s_i},z) \tag{10}$$

式中：N_{s_i}为对应节点。

图像层节点的条件概率近似因式分解$s_{\alpha,\beta}$的条件概率，使得可以对所有$s_{\alpha,\beta}$进行独立MAP估计[11]，如下：

$$s^*_{\alpha,\beta} = \arg\max p(s_{\alpha,\beta}|N_{s_{\alpha,\beta}},z) \tag{11}$$

可得图像层节点的MAP估计为

$$S^* = \left\{ s^*_{1,1}, s^*_{1,2}, \cdots, s^*_{\alpha,\beta} \right\} \tag{12}$$

根据其他节点的当前值更新每个节点值, 一直迭代到S^*收敛, 将最终迭代分类的标签记为缺陷的部分作为最终分类结果。

2　自采样数据集

2.1　轨检系统

本文的改进 MRF 算法基于笔者课题组自主搭建的轨检系统在采集样本过程中遇到的实际需求提出, 如图 5 所示, 该设备为课题组自主搭建的轨检系统实物, 系统包含图像采集模块、外部触发模块、照明模块, 可满足巡检、数据采集等常规任务。

设备使用线阵相机完成图像采集, 利用编码器与外部触发模块, 保证图像采集结果在设备运动过程中不失真, 同时设备加装了辅助光源和遮光罩以避免采集过程过度曝光的问题。



图 5　轨检系统

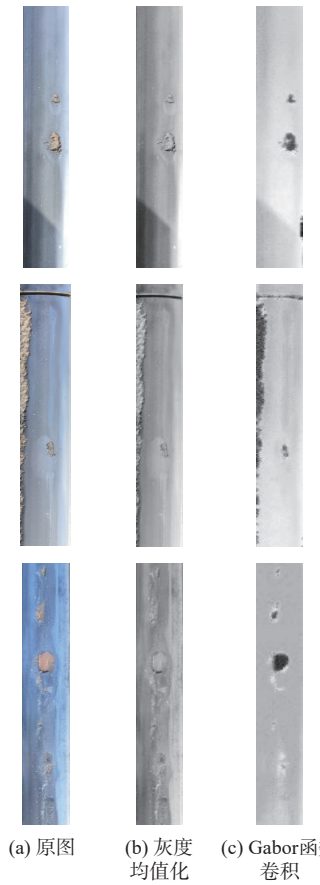
Fig. 5　Rail inspection system

2.2　测试数据集

本文使用的测试数据集为 2.1 节所述轨检系统实际巡检过程中自采样缺陷数据。缺陷测试数据为在不同路段采集的分辨率为 445×4 608 的轨面数据, 缺陷类型包含边缘磨损、接缝/断轨、剥落掉块等常见的 3 类轨面缺陷, 且各图像中的缺陷类型不唯一, 多数是多种类型缺陷的组合图像, 并且数据集采样的光照条件不同, 包含阴影、反光等常见影响因素。

由于轨检设备通常运行的天窗期多为夜晚, 且轨检采集设备成像结果多为灰度图, 因此, 将数据集转为灰度图后进行后续操作。测试数据集预处理及对比结果如图 6 所示。

通过对比结果可知, 将 Gabor 函数作为卷积核对数据进行卷积处理, 可在一定程度上克服阴影和反光等因素, 同时可以提高缺陷部分与钢轨部分的对比度, 使缺陷特征更加明显, 在消融实验部分将通过数据比对验证这一结论。同时, 本文将 Gabor 卷积处理后的结果作为后续对比实验中的测试数据集。



(a) 原图　(b) 灰度均值化　(c) Gabor函数卷积

图 6　测试数据集预处理结果

Fig. 6　Test dataset preprocessing results

3　实验结果与分析

3.1　对比实验结果分析

选取当前 3 种常用分割方法, 即 MRF[13]、模糊 C-均值聚类(fuzzy C-means, FCM)算法[6]、高斯混合模型(Gaussian mixture model, GMM)[6], 以及经少样本训练后的深度学习模型 U-Net[23]、DeepLabv3[24], 将其与本文方法做对比, 从自建轨面缺陷数据集中随机选取测试样本, 所有方法设置训练类别为二分类, 由于各方法参数不同, 分别设置多次实验取最优结果。为确定 Kmeans 聚类的簇数, 在测试样本上分别设置不同簇数进行测试, 结果表明, 簇数设置过大会导致模型运行时间增加, 簇数设置偏小又会丢失图像部分信息, 导致模型分割结果错误。因此, 结合上述测试结果, 将 Kmeans 聚类的簇数设置为训练类别的 2 倍, 即簇数设置为 4, 该设置不会导致模型运行时间大幅增加, 且不会导致图像丢失主要信息。

从测试样本中选取 4 张测试图像, 将分割结果进行展示, 如图 7 所示。将原图与对比方法的分割结果进行比对, 可更直观地看出分割效果差异。使用以下常见的实例分割模型评价指标:

1) 像素准确率 (pixel accuracy, PA)。表示标记

正确的像素占总像素的比例[13]，使用 K_{PA} 表示，如下：

$$K_{\mathrm{PA}}=\frac{\sum_{i=0}^{k'}p_{ii}}{\sum_{i=0}^{k'}\sum_{j=0}^{k'}p_{ij}} \tag{13}$$

2）平均像素准确率 (mean pixel accuracy, MPA)。



(a) 原图 (b) MRF[13] (c) FCM[6] (d) GMM[6] (e) U-Net[23] (f) DeepLabv3[24] (g) 本文方法

图 7 不同方法分割结果

Fig. 7 Segmentation results of different algorithms

计算每个类内被正确分类的像素数的比例后求所有类的平均[13]，如下：

$$K_{\mathrm{MPA}}=\frac{1}{k'+1}\sum_{i=0}^{k'}\left(\frac{p_{ii}}{\sum_{j=0}^{k'}p_{ij}}\right) \tag{14}$$

3）均交并比 (mean intersection over union, MIoU)。表示模型对每类预测的结果和真实值的交集与并集的比值求和再平均[13]，如下：

$$K_{\mathrm{MIoU}}=\frac{1}{k'+1}\sum_{i=0}^{k'}\left(\frac{p_{ii}}{\sum_{j=0}^{k'}p_{ij}+\sum_{j=0}^{k'}(p_{ji}-p_{ii})}\right) \tag{15}$$

4）加权交并比 (frequency weighted intersection over union, FWIoU)。根据每类出现频率作为该类权重[13]，如下：

$$K_{\mathrm{FWIoU}}=\frac{1}{\sum_{i=0}^{k'}\sum_{j=0}^{k'}p_{ij}}\sum_{i=0}^{k'}\left(\frac{p_{ii}}{\sum_{j=0}^{k'}p_{ij}+\sum_{j=0}^{k'}(p_{ji}-p_{ii})}\right) \tag{16}$$

式中：k'=2；p_{ij} 表示属于 i 类被预测为 j 类的数量；p_{ii} 表示属于 i 类被预测为 i 类的数量[13]。

5）使用模型每秒传输帧数 (frames per second, FPS) 作为运算时间的评价标准，如下：

$$K_{\mathrm{FPS}}=\frac{1}{t_{\mathrm{Latemcy}}} \tag{17}$$

式中：t_{Latency} 为网络预测一张图像所用的时间。

对不同方法的检测结果进行统计分析，对比实验结果如表 1 所示。数据显示，本文方法的 K_{PA}、K_{MPA}、K_{FWIoU}、K_{MIoU} 分别达到 93.6%、80.7%、89.4%、68.2%，对比其他方法，本文方法在以上 4 种评价指标下均为最佳。图 8 为不同方法分割结果在 4 类评价指标下的堆积条形图。方法的效果越好，单项评价指标越接近于 1，则图对应的整体长度越长。

表 1 对比实验结果分析

Table 1 Analysis of comparison experimental results

方法	K_{PA}/%	K_{MPA}/%	K_{FWIoU}/%	K_{MIoU}/%	K_{FPS}
MRF[13]	47.6	42.8	42.3	26.1	0.073
FCM[6]	79.6	72.4	74.6	55.0	**11.360**
GMM[6]	88.5	78.7	83.1	66.9	0.946
U-Net[23]	80.2	55.3	82.1	50.1	3.597
DeepLabv3[24]	90.8	59.8	83.0	55.1	3.257
本文方法	**93.6**	**80.7**	**89.4**	**68.2**	3.174

注：表中加粗部分表示评价指标达到最佳。

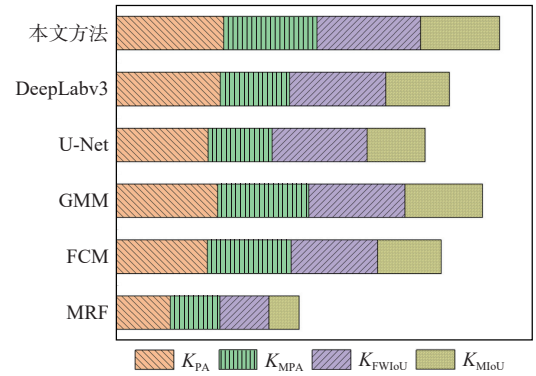


图 8　对比实验评价指标对比效果

Fig. 8　Comparative effects of comparison experiment evaluation index

未改进的 MRF 算法在整体分割效果上相对较差, 且模型计算时间较长, 4 类分割准确率的评价指标均不到改进后本文方法的 1/2; FCM 算法整体优于未改进的 MRF 算法; GMM 方法整体优于 FCM 算法; 经少样本训练后的 U-Net 与 DeepLabv3 方法整体优于前 3 种对比方法。

本文方法的 4 项分割准确率指标和实际分割效果均优于其他对比方法, 虽然 K_{FPS} 值不是最优, 但与当前应用较多的深度学习 U-Net 与 DeepLabv3 方法相近。根据综合评判, 本文方法满足实际工程应用需求。

3.2　消融实验结果分析

通过图 8 和表 1 的数据和结论可知, 本文方法的评价指标和实际分割效果均优于其他对比方法。为验证改进措施的有效性, 对本文方法的改进部分进行消融实验分析。方法 1 为未改进的 MRF 算法, 使用 ICM 作为 MAP 的推理方法, 方法 2 为融合 Kmeans 聚类特征作为双层图模型标记层的 MRF 算法, 方法 3 为加入 Gabor 卷积处理作为双层图模型图像层的 MRF 算法。实验结果如表 2 所示。

表 2　消融实验结果分析

Table 2　Analysis of ablation experiment results

方法	Gabor处理	融合聚类特征	ICM	K_{PA}/%	K_{MPA}/%	K_{FWIoU}/%	K_{MIoU}/%	K_{FPS}
MRF				47.6	42.8	42.3	26.1	0.073
方法1			√	45.2	40.2	39.4	25.7	0.612
方法2		√		53.8	49.7	72.8	40.7	0.469
方法3	√			86.4	**81.1**	81.6	61.5	2.096
本文方法	√	√	√	**93.6**	80.7	**89.4**	**68.2**	**3.174**

注: 表中加粗部分表示评价指标达到最佳。

表 2 的统计数据显示, 本文方法的 K_{PA}、K_{MPA}、K_{FWIoU}、K_{MIoU} 分别为 93.6%、80.7%、89.4%、68.2%, 对比其他方法, 本文方法在 K_{PA}、K_{FWIoU}、K_{MIoU} 这 3 种评价指标下均为最佳, K_{MPA} 指标略低于方法 3, 整体优于其他测试方法, 且每部分改进均在未改进 MRF 模型的基础上有一定的提升, 验证了本文方法改进的实际意义。

对比未改进 MRF 与方法 1 可知, 使用 ICM 作为 MAP 的推理方法, 使得 K_{FPS} 有成倍的提升, 可在一定程度上提升模型的推理速度, 但由于 ICM 是一种近似推理方法, 导致模型推理效果及实际分割准确率略有下降。

对比未改进 MRF 与方法 2 可知, 未改进的 MRF 模型缺陷分割的结果易受钢轨自身像素矩阵影响, 当钢轨自身像素矩阵数据分布略大时, MRF 无法正确识别缺陷部分, 只将钢轨像素矩阵中边缘信息明显的部分作为局部最优解的结果进行区分, 导致分割错误。而使用 Kmeans 聚类则可以一定程度上避免上述问题, 一方面, Kmeans 聚类在全局的角度对钢轨像素矩阵进行聚类, 使分割结果更加精确, 减少部分环境因素带来的误判, 可以对钢轨部分进行更好的识别; 另一方面, Kmeans 聚类结果可以简化后续 MRF 的推理迭代计算过程, 提升计算效率。

对比未改进 MRF 与方法 3 可知, 在整体方法中加入 Gabor 卷积处理, 可以极大地提升缺陷与轨面背景部分的对比效果, 使用 Gabor 卷积处理后进行缺陷分割, 4 类评价指标均高于原来的 1 倍以上, 并且加入 Gabor 卷积处理可以在很大程度上降低模型运算时间。消融实验评价指标对比效果如图 9 所示。

3.3　实验结果分析

本文通过设置对比实验与消融实验验证了融合聚类特征的 MRF 算法的有效性及改进模型各部分对整体模型的实际作用, 通过数据及分割效果图直观地展现出模型的分割能力, 但仍存在部分钢轨分割错误的情况。

图 10 为本文方法错误分割的 2 类示例。第 1 类为钢轨轨面出现较深的阴影, 此类图像由于阴影与轨面、缺陷色域差距过大, 导致在 Gabor 卷积的过程中进一步放大阴影的特征及像素级对比度,

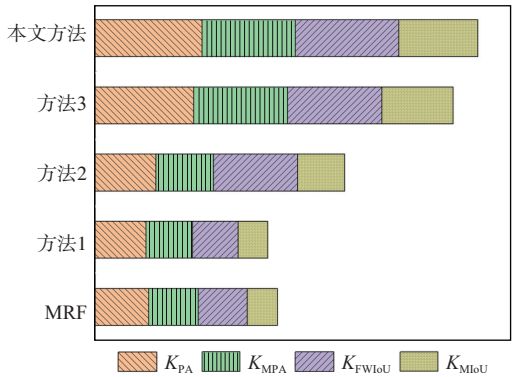


图 9 消融实验评价指标对比效果

Fig. 9 Comparative effects of ablation experiment evaluation index

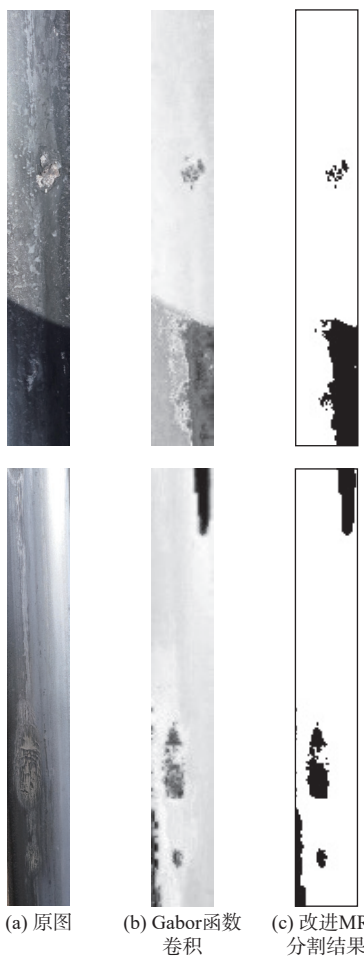


图 10 错误分割示例

Fig. 10 Incorrect segmentation

进而在后续的模型推理过程中将阴影与真实缺陷标注为一类；第 2 类为钢轨轨面部分出现较亮的反光光斑，其产生原理同第 1 类，主要原因是在 Gabor 卷积过程中加大了像素级对比度，导致结果将光斑也划分为轨面缺陷。

4 结 论

1）通过实验比较，本文方法的像素准确率、平均像素准确率、加权交并比、均交并比分别达到 93.6%、80.7%、89.4%、68.2%，每秒传输帧率达到 3.174，改进 MRF 算法在测试中具有较好的性能表现，面对不同种类缺陷具有较好的泛化能力。相比于传统的 MRF 模型，本文改进部分可以提高采集图像缺陷与钢轨背景部分的对比度，提高模型分割结果的准确性，保证模型对目标做出正确的判断及划分，避免出现背景与目标错分的问题。

2）文本方法属于无监督、零样本分割方法，不需要样本训练即可进行缺陷分割，可以规避只有少量样本的情况下训练模型效果不确定的问题。

3）实验部分使用的自采样数据集是轨检系统对真实钢轨缺陷的采集结果，且本文方法的实际检测效果和评价指标均优于其他对比方法，因此，本文方法对于铁路缺陷检测工作有着广泛的应用价值和较强的实际意义。

针对本文实验结果依然存在对反光、阴影等部位出现误判的问题，将在后续工作中考虑尝试将域自适应技术与改进 MRF 算法结合，针对当前轨面的环境进行迁移建模，进一步缩减因轨面反光或阴影带来的影响，而后再使用 MRF 模型对轨面缺陷进行分割。

参考文献（References）

[1] 黄鹏, 郑淇, 梁超. 图像分割方法综述[J]. 武汉大学学报 (理学版), 2020, 66(6): 519-531.
HUANG P, ZHENG Q, LIANG C. Overview of image segmentation methods[J]. Journal of Wuhan University (Natural Science Edition), 2020, 66(6): 519-531(in Chinese).

[2] CHOPRA M, PURWAR A. Recent studies on segmentation techniques for food recognition: a survey[J]. Archives of Computational Methods in Engineering, 2022, 29(2): 865-878.

[3] 王国伟, 陈光武, 魏宗寿. 基于背景差分与最大熵的轨面缺陷分割[J]. 计算机系统应用, 2022, 31(10): 184-190.
WANG G W, CHEN G W, WEI Z S. Rail surface defect segmentation based on background difference and maximum entropy[J]. Computer Systems and Applications, 2022, 31(10): 184-190(in Chinese).

[4] 程天栋, 闵永智, 马宏锋. 基于模糊超熵与遗传算法的轨面缺陷分割算法[J]. 兰州交通大学学报, 2016, 35(6): 46-50.
CHENG T D, MIN Y Z, MA H F. Rail surface defect segmentation algorithm based on fuzzy excess entropy and genetic algorithm[J]. Journal of Lanzhou Jiaotong University, 2016, 35(6): 46-50(in Chinese).

[5] 曹义亲, 刘龙标. 基于缺陷比例限制的背景差分钢轨表面缺陷检测方法[J]. 计算机应用, 2020, 40(10): 3066-3074.
CAO Y Q, LIU L B. Rail surface defect detection method based on background differential with defect proportion limitation[J]. Journal of Computer Applications, 2020, 40(10): 3066-3074(in Chinese).

[6] 张辉, 李平, WU Q M J, 等. 基于改进马尔可夫随机场的钢轨缺陷

分割[J]. 计算机工程与设计, 2020, 41(4): 1052-1061.
ZHANG H, LI P, WU Q M J, et al. Rail defect segmentation based on improved Markov random field[J]. Computer Engineering and Design, 2020, 41(4): 1052-1061(in Chinese).

[7] 周雯, 史天运, 李平, 等. 基于卷积神经网络的动车组行车安全图像缺陷检测与分割[J]. 铁道学报, 2019, 41(10): 76-83.
ZHOU W, SHI T Y, LI P, et al. Defects detection and segmentation of operation safety image of EMU based on convolutional neural network[J]. Journal of the China Railway Society, 2019, 41(10): 76-83 (in Chinese).

[8] 刘俊博, 杜馨瑜, 王胜春, 等. 基于少样本学习的钢轨表面缺陷检测方法[J]. 铁道学报, 2022, 44(7): 72-79.
LIU J B, DU X Y, WANG S C, et al. Rail surface defect detection method based on few-shot learning[J]. Journal of the China Railway Society, 2022, 44(7): 72-79(in Chinese).

[9] 赵晨阳, 张辉, 廖德, 等. 基于注意力机制与混合监督学习的钢轨表面缺陷检测模型[J]. 计算机科学, 2022, 49(增刊 2): 488-493.
ZHAO C Y, ZHANG H, LIAO D, et al. Rail surface defect detection model based on attention mechanism and mixed supervised learning[J]. Computer Science, 2022, 49(Sup 2): 488-493(in Chinese).

[10] 罗晖, 李健, 贾晨. 基于图像增强与改进 Cascade R-CNN 的钢轨表面缺陷检测[J]. 激光与光电子学进展, 2021, 58(22): 324-335.
LUO H, LI J, JIA C. Rail surface defect detection based on image enhancement and improved Cascade R-CNN[J]. Laser & Optoelectronics Progress, 2021, 58(22): 324-335(in Chinese).

[11] DURR O, SICK B, MURINA E. 概率深度学习: 使用 Python、Kears 和 TensorFlow Probability[M]. 崔亚奇, 唐田田, 但波, 译. 北京: 清华大学出版社, 2022: 3-28, 249-288.
DURR O, SICK B, MURINA E. Probability deep learning: using Python, Kears and TensorFlow Probability[M]. CUI Y Q, TANG T T, DAN B, translated. Beijing: Tsinghua University Press, 2022: 3-28, 249-288(in Chinese).

[12] TROMBINI M, SOLARNA D, MOSER G, et al. A goal-driven unsupervised image segmentation method combining graph-based processing and Markov random fields[J]. Pattern Recognition, 2023, 134: 109082.

[13] JI Q. 概率图模型及计算机视觉应用[M]. 郭涛, 译. 北京: 机械工业出版社, 2021: 15-38, 107-131.
JI Q. Probability graph model and computer vision application[M]. GUO T, translated. Beijing: Mechanical Industry Press, 2021: 15-38, 107-131(in Chinese).

[14] 卢印举, 马芳, 戴曙光, 等. 融合多尺度特征的马尔可夫随机场路面裂缝分割算法[J]. 计算机辅助设计与图形学学报, 2022, 34(5): 711-721.
LU Y J, MA F, DAI S G, et al. Markov random field road crack image segmentation algorithm integrating multi-scale features[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(5): 711-721(in Chinese).

[15] ZHANG P, CAI X Y, SHI J, et al. Feature extraction of fused residual network and single target-assisted vessel image recognition of MRF grayscale information[J]. Mobile Information Systems, 2022, 2022: 1041934.

[16] 吕宁, 肖剑, 高健, 等. 基于改进 MRF 的冲压件轮廓缺陷图像分割算法[J]. 锻压技术, 2022, 47(4): 101-109.
LYU N, XIAO J, GAO J, et al. Image segmentation algorithm on contour defects for stamping part based on improved MRF[J]. Forging & Stamping Technology, 2022, 47(4): 101-109(in Chinese).

[17] ZHU D C, LI J M, LI H, et al. A less-constrained sclera recognition method based on stem-and-leaf branches network[J]. Pattern Recognition Letters, 2021, 145: 43-49.

[18] ZITOUNI M, AFIF M, ZRIBI M. Tweedie hidden Markov random field and the expectation-method of moments and maximisation algorithm for brain MR image segmentation[J]. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2023, 11(1): 67-79.

[19] AZIZ K A A, SARIPAN M I, SAAD F F A, et al. A Markov random field approach for CT image lung classification using image processing[J]. Radiation Physics and Chemistry, 2022, 200: 110440.

[20] 崔鹏, 杨海峰, 蔡江辉, 等. 多尺度局部聚类的 Kmeans-DETR 目标检测方法[J]. 小型微型计算机系统, 2024, 45(5): 1136-1142.
CUI P, YANG H F, CAI J H, et al. Multi-scale local clustering method for object detection based on Kmeans-DETR[J]. Journal of Chinese Computer Systems, 2024, 45(5): 1136-1142(in Chinese).

[21] RAJPOOT V, DUBEY R, KHAN S S, et al. Orchard boumans algorithm and MRF approach based on full threshold segmentation for dental X-ray images[J]. Traitement Du Signal, 2022, 39(2): 737-744.

[22] FILALI H, KALTI K. Image segmentation using MRF model optimized by a hybrid ACO-ICM algorithm[J]. Soft Computing, 2021, 25(15): 10181-10204.

[23] RONNEBERGER O, FISCHER P, BROX T . U-Net: convolutional networks for biomedical image segmentation[EB/OL]. (2015-05-18) [2023-06-01]. https://arxiv.org/abs/1505.04597.

[24] LI X L, LI Y Y, AI J Q, et al. Semantic segmentation of UAV remote sensing images based on edge feature fusing and multi-level upsampling integrated with DeepLabv3[J]. PLoS One, 2023, 18(1): e0279097.

Improved MRF rail surface defect segmentation method based on fusion of clustering features

MIN Yongzhi[1,*], LIU Yang[1,2]

(1. School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. Department of Mechanical Egineering, Baotou Railway Vocational & Technical College, Baotou 014060, China)

Abstract: In view of the characteristics of the small number and many types of rail defect samples, unstable transfer learning effect in real scenes, and threshold segmentation susceptible to environmental factors, an improved Markov defect segmentation method with zero samples was proposed. Firstly, the collected data was processed using Gabor functions to highlight defect features and reduce data dimensionality, resulting in a reduced dimensionality feature map. To mitigate the effects of glare and shadows, the feature map was subjected to Kmeans clustering to reduce data distribution. The clustering results were then used as a pre-classification matrix. Based on the reduced dimensionality feature map and the pre-classification matrix, a two-layer graph model of the improved Markov random field (MRF) was constructed for inference. The model analyzed the local geometric structure of the defect area by using the eigenvalues of the classification matrix inferred by the model. Finally, the defect regions were labelled, and the defect segmentation was completed. In the experimental part, a self-sampled dataset was used to draw the final conclusion through comparative and ablation experiments. The experimental results show that the proposed method achieves pixel accuracy, mean pixel accuracy, weighted intersection over union, and mean intersection over union of 93.6%, 80.7%, 89.4%, and 68.2%, respectively, on the self-sampled dataset, surpassing the accuracy of other comparative detection algorithms.

Keywords: image processing; rail surface defect; Markov random field; probabilistic graphical model; instance segmentation

Received: 2023-06-09; **Accepted**: 2023-06-30; **Published Online**: 2023-07-21 17:15

URL: link.cnki.net/urlid/11.2625.V.20230721.1442.002

Foundation items: National Natural Science Foundation of China (62066024); 2023 Gansu Province Excellent Graduate Student "Innovation Star" Project (2023CXZX-615)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0340

基于飞联网运行的航空器多速度差跟驰模型

王莉莉*，赵云飞，郭微萌
(中国民航大学 民航飞联网重点实验室，天津 300300)

摘　　要：为提高空中交通流的稳定性，基于飞联网运行特性研究了考虑偏移的航空器多速度差跟驰模型。为定量描述航空器间的偏移对跟驰行为的影响,引入偏移阻碍作用,建立偏移与前导航空器速度的关系,将跟驰模型扩展到三维模式；考虑飞联网环境下的多航空器信息交互模式，构建航空器多速度差跟驰模型，并应用稳定性分析方法，推导所提模型稳定性判别条件，计算稳态通行能力；在对模型进行参数标定的基础上，以考虑3架前导航空器的多速度差跟驰模型为例，设计数值仿真实验。结果表明：阻碍作用随偏移量的增大而减小，在相同偏移量情况下，重型机的阻碍作用最大，轻型机最小；所提模型相比传统模型具备更优的稳定域，且考虑前导航空器数量越多、权重系数越大，所提模型的稳定性越好；相同取值条件下，所提模型的燃油消耗系数均低于传统模型，当敏感系数取1 s^{-1}时，燃油消耗系数降低27.12%。数值仿真表明航空器多速度差跟驰模型有利于提高空中交通流的稳定性，降低燃油消耗。

关　键　词：飞联网；多速度差；跟驰模型；稳定性分析；通行能力；燃油消耗
中图分类号：V355.1; TB553
文献标志码：A　　**文章编号**：1001-5965(2025)06-1873-09

随着空中交通流的快速增长，未来空中交通管理将从“航班控制”转向“交通流管理”[1]。对空中交通流仿真分析，提高空中交通流稳定性，是实现空域资源高效利用的关键途径。

飞联网是以互联互通、全息感知为核心的自主间隔技术，其通过综合航空器、外部环境信息，实现空域资源综合利用和综合交通协同。飞联网运行环境下，航空器将从传统单机隔离运行，向有人、无人、军民航混合运行及自由飞行演进；航空器将从传统多传感器的被动观测转向主动全息感知；空中交通流的微观特性会发生明显变化。

空中交通流的研究主要集中在空中交通流建模与预测。建模方面，赵嶷飞等[2]将多架航空器形成的一维或二维的航空器流纳入空域复杂性研究。此外，还有航班排队模型[3]、空中交通流机头时距模型[4]及元胞计算模型[5]等交通流建模方法。预测方面，Shafienya等[6]将蒙特卡罗与卷积神经网络和门控循环单元结合对4D航迹进行预测。Gui等[7]将航空大数据映射到航线上，利用长短期记忆(long short-term memory，LSTM)神经网络对不同城市之间的空中交通流量进行计数和预测。Zeng等[8]提出深度图嵌入长短时记忆网络模型，结合机场网络空间相关性预测航班延误。

关于空中交通流微观仿真的研究相对较少，且多是基于跟驰理论对间隔、容量等进行分析。黎新华等[9]在航空器跟驰稳定的基础上对终端区容量进行评估。王莉莉等[10]在空中高速路环境下基于跟驰理论构建了航空器的流量模型。张洪海等[11]基于跟驰和排队理论建立了多机场终端区交通流模型。之后，张洪海等[12]采用模糊逻辑方法提出了航空器动态期望间隔控制策略。上述研究多是利用传统的地面跟驰模型对空中交通流进行分析，并

收稿日期：2023-06-12；**录用日期**：2023-09-22；**网络出版时间**：2023-10-11 13:17
网络出版地址：link.cnki.net/urlid/11.2625.V.20231011.1038.001
基金项目：国家自然科学基金委员会与中国民用航空局联合基金(U1633124)
*通信作者. E-mail：llwang317@163.com

引用格式：王莉莉，赵云飞，郭微萌. 基于飞联网运行的航空器多速度差跟驰模型[J]. 北京航空航天大学学报，2025，51(6)：1873-1881. WANG L L，ZHAO Y F，GUO W M. Aircraft multi-velocity difference car-following model based on flight networking operation[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1873-1881 (in Chinese).

没有基于空中交通的运行环境和运行特点对空中交通流进行建模研究。此外，上述研究中的跟驰模型也仅考虑前后2架航空器的相互作用，并未针对航空器队列进行分析，与未来民航的“交通流管理”和多信息交互发展趋势不符。

因此，本文基于飞联网运行环境，并结合地面交通的相关理论，考虑航空器偏移的影响，构建了符合飞联网航空器运行特性的空中交通流多速度差跟驰模型，并对所建模型进行稳定性分析和参数标定。通过计算跟驰稳定情况下航路通行能力，证明所建模型可以表征航空器间的微观作用，提高空中交通流的稳定性，促进空中交通安全高效绿色运行。

1 飞联网环境下空中交通流多速度差跟驰模型

1.1 航空器三维偏移分析建模

航空器在航行时，为避免与前方航空器碰撞，飞行员和管制员要控制航速和距离，这会对后方航空器产生阻碍作用。阻碍作用受航空器间相对偏移的影响，航空器正前方的航空器对其产生的阻碍作用最大，随着前方航空器侧向偏移距离的增加，阻碍作用减小。当前方航空器的偏移超过一定程度时(航空器换道)，这种阻碍作用就会很小甚至消失，以致很难影响到目标航空器的运动。因此，在进行微观仿真时，为考虑三维偏移的影响，综合航空器高度、宽度及尾流影响，将其尾部影响抽象成椭圆锥面，如图1所示。

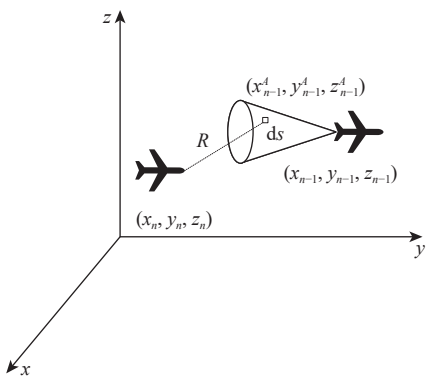


图1 航空器偏移示意图

Fig. 1 Schematic of aircraft offset

将前机对跟驰航空器的阻碍作用所产生的加速度表示为

$$a_n = a_2(a_3 v_{n-1}(t) - v_n(t)) \iint_{\Sigma} \frac{\mathrm{d}s}{R^w} \tag{1}$$

式中：$v_n(t)$为第n架航空器速度；a_2为偏移阻碍作用系数；a_3为前机速度系数；w为积分参数；$\iint_{\Sigma} \frac{\mathrm{d}s}{R^w}$为第$n$架航空器对前方航空器尾部椭圆锥面$A$的第一类曲面积分形式，表征了前方航空器对于目标航空器的阻碍作用。

为简化曲面积分的计算，将椭圆锥面A表示成参数方程，如下：

$$\begin{cases} y(u,\theta) = y^A_{n-1} + u \\ x(u,\theta) = x^A_{n-1} + tu\cos\theta \\ z(u,\theta) = z^A_{n-1} + u\sin\theta \\ \theta \in (0, 2\pi) \\ u \in (-k, 0) \end{cases} \tag{2}$$

式中：$x^A_{n-1} = x_{n-1}$；$y^A_{n-1} = y_{n-1} - L_{\mathrm{cra}}$；$z^A_{n-1} = z_{n-1}$；$L_{\mathrm{cra}}$为航空器长度；$(x_{n-1}, y_{n-1}, z_{n-1})$为第$n-1$架航空器机头中心的坐标；$t$为根据航空器类型确定的参数；$k$为根据航空器的宽度和尾流等级确定的参数。

令$w=2$，并采用参数情形下第一类曲面积分的计算规则[13]：如有参数方程$\begin{cases} x = f(u_a, v_a) \\ y = g(u_a, v_a) \\ z = h(u_a, v_a) \end{cases}$，则$\iint_{\Sigma} f(x, y, z)\mathrm{d}s = \iint_{D} f(x(u_a, v_a),\ y(u_a, v_a),\ z(u_a, v_a)) \cdot \sqrt{EG - F^2}\ \mathrm{d}u_a \mathrm{d}v_a$。其中：

$$\begin{cases} E = \left(\dfrac{\partial f(u_a, v_a)}{\partial u_a}\right)^2 + \left(\dfrac{\partial g(u_a, v_a)}{\partial u_a}\right)^2 + \left(\dfrac{\partial h(u_a, v_a)}{\partial u_a}\right)^2 \\ G = \left(\dfrac{\partial f(u_a, v_a)}{\partial v_a}\right)^2 + \left(\dfrac{\partial g(u_a, v_a)}{\partial v_a}\right)^2 + \left(\dfrac{\partial h(u_a, v_a)}{\partial v_a}\right)^2 \\ F = \dfrac{\partial f(u_a, v_a)}{\partial u_a} \cdot \dfrac{\partial f(u_a, v_a)}{\partial v_a} + \dfrac{\partial g(u_a, v_a)}{\partial u_a} \cdot \dfrac{\partial g(u_a, v_a)}{\partial v_a} + \\ \qquad \dfrac{\partial h(u_a, v_a)}{\partial u_a} \cdot \dfrac{\partial h(u_a, v_a)}{\partial v_a} \end{cases} \tag{3}$$

则积分式可以化简为

$$\iint_{\Sigma} \frac{\mathrm{d}s}{R^w} = \iint_{D} \frac{\sqrt{EG - F^2}\mathrm{d}u\mathrm{d}\theta}{a + o + c} \tag{4}$$

式中：(x_n, y_n, z_n)为第n架航空器机头中心的坐标；$a = (x_n - x^A_{n-1} - tu\cos\theta)^2$；$o = (z_n - z^A_{n-1} - u\sin\theta)^2$；$c = (y_n - y^A_{n-1} - u)^2$。

令

$$T = \iint_{D} \frac{\sqrt{de - f^2}\mathrm{d}u\mathrm{d}\theta}{g + h + i} \tag{5}$$

式中：$d = 1 + t^2\cos^2\theta + \sin^2\theta$；$e = t^2u^2\sin^2\theta + u^2\cos^2\theta$；$f = \sin^2\theta\cos^2\theta(u^2 + t^4u^2 - 2t^2u^2)$；$g = (x_n - x^A_{n-1} - tu\cos\theta)^2$；$h = (z_n - z^A_{n-1} - u\sin\theta)^2$；$i = (y_n - y^A_{n-1} - u)^2$。

T可以采用标准积分公式计算得到。由此，目标航空器n的偏移阻碍作用加速度可以表示为

$a_2(a_3v_{n-1}(t)-v_n(t))T$。

1.2　飞联网多前机速度差跟驰模型

1.2.1　飞联网航空器运行属性分析

飞联网的推广与实现存在阶段性与侧重点, 但其对航空器跟驰行为最直接的影响可概括为: 航空器之间可以信息交互与联动, 航空器能够更为准确、直接地感知多架航空器信息, 且感知能力加强, 范围增大, 跟驰航空器受到前方航空器信息与考虑范围的影响较之前更为明显。

航空器队列中, 各航空器的飞行速度都将受到相邻航空器运行状态的制约, 这种状态称为非自由运行状态, 在该状态下, 航空器队列具有制约性、延迟性和传递性 3 个特性。

1) 制约性。前机航速制约着后机航速和间隔。飞联网环境下, 跟驰航空器仍有紧随要求, 但航空器的安全间隔减小, 速度增大, 制约性减弱。

2) 延迟性。飞联网运行环境下, 系统的感知取代了传统交通环境中人为感知与判断, 总反应时间呈减小的趋势, 使延迟性减小。

3) 传递性。飞联网环境下, 航空器状态的传递并不是像传统交通一样, 依赖于航空器一架接一架地改变运动状态, 而是更广泛地通过信息系统交互传递, 传递的范围和速度将大幅提高。

1.2.2　飞联网运行环境假设

1) 仅考虑相邻 2 架航空器间偏移的影响。

2) 仅考虑飞联网航空器, 忽略混合运行。

3) 仅考虑中型航空器队列, 忽略机型差异。

4) 反应时间由系统延迟和人为延迟 (管制员和飞行员) 的总耗时构成。

1.3　飞联网环境下空中交通流多速度差跟驰模型构建

地面交通跟驰模型的研究多是基于车辆的速度信息, Bando 等[14] 根据交通流特征分析提出了优化速度 (optimal velocity, OV) 模型。

$$\dot{v}_n(t)=a_0(V(\Delta x_n(t))-v_n(t)) \tag{6}$$

式中: a_0为模型系数; $V(\cdot)$为速度关于间距的优化速度函数; 速度差$\Delta x_n(t)=x_{n-1}(t)-x_n(t)$。

该模型解决了无限加速问题, 但会产生过高的加速度或不切实际的减速度。针对该问题, Jiang 等[15] 根据正负速度差对跟驰模型的影响提出了全速度差 (full velocity difference, FVD) 模型。全速度差模型是后续电子节气门控制[16] 等跟驰模型的基础, 公式如下:

$$\dot{v}_n(t)=a(V(\Delta x_n(t))-v_n(t))+\lambda\Delta v_n(t) \tag{7}$$

式中: a 为模型系数; λ为速度差的反应系数。

针对空中交通跟驰特点和空中交通运行情况, 利用多架航空器之间的速度差信息和相邻航空器间偏移信息, 构建飞联网运行环境下的空中交通流多速度差跟驰模型, 如下:

$$\ddot{x}_n(t)=a_1(V(\Delta x_n(t))-v_n(t))+a_2(a_3v_{n-1}(t)-v_n(t))T+\sum_{j=1}^{m}\lambda_j\Delta v_{n-j+1,n}(t) \tag{8}$$

式中: a_1为敏感系数; $\Delta v_n(t)=v_{n-1}(t)-v_n(t)$; a_2为偏移阻碍作用系数; a_3为前机速度系数; $\Delta v_{n-j,n}(t)=v_{n-j}(t)-v_n(t)$; m表示前方有 m 架航空器, 且满足$m\ll n$, 当$m=1$时, $\sum_{l=1}^{m}\lambda_l=\lambda$, 该模型退化为全速度差模型, 当$m=1$且$\sum_{l=1}^{m}\lambda_l=\lambda=0$时, 该模型退化为优化速度模型。

2　稳定性分析

应用稳定性分析方法对本文所提跟驰模型进行稳定性分析, 用$n-1$表示航空器n的前机。假设给定初始状态为稳定态, 航空器的机头间距均为b, 对应的航速为$V(b)$。此时, 稳态空中交通流的航空器位置解为

$$x_n^0(t)=bn+V(b)t \tag{9}$$

对均匀流加一扰动$y_n(t)$, 可得

$$x_n(t)=x_n^0(t)+y_n(t) \tag{10}$$

由式 (10) 可得

$$y_n(t)=x_n(t)-x_n^0(t) \tag{11}$$

对式 (11) 分别求一阶导数和二阶导数, 即$\dot{y}_n(t)=\dot{x}_n(t)-\dot{x}_n^0(t)$, $\ddot{y}_n(t)=\ddot{x}_n(t)$。其中, $\dot{x}_n(t)=v_n(t)$; $\dot{x}_n^0(t)=V(b)$。

将上述结果代入式 (8) 可得

$$\ddot{y}_n(t)=a_1(V(\Delta x_n(t)-\Delta v_n(t))+a_2(a_3v_{n-1}(t)-v_n(t))T+\sum_{j=1}^{m}\lambda_j\Delta v_{n-j+1}(t)=a_1(V(b+\Delta y_n(t))-V(b)-\dot{y}_n(t))+a_2(a_3\dot{y}_{n-1}(t)-\dot{y}_n(t))T+\sum_{j=1}^{m}\lambda_j\Delta\dot{y}_{n-j+1}(t) \tag{12}$$

由一阶 Taylor 展开式可得

$$\ddot{y}_n(t)=a_1(V'(b)\Delta y_n(t)-\dot{y}_n(t))+a_2(a_3\dot{y}_{n-1}(t)-\dot{y}_n(t))T+\sum_{j=1}^{m}\lambda_j\Delta\dot{y}_{n-j+1}(t) \tag{13}$$

式中: $V'(b)=\left.\dfrac{\mathrm{d}V(x)}{\mathrm{d}x}\right|_{x=b}$; $\Delta x_n(t)=b+y_{n-1}(t)-y_n(t)=b+\Delta y_n(t)$; $\Delta v_n(t)=\dot{y}_{n-1}(t)-\dot{y}_n(t)=\Delta\dot{y}_n(t)$。

令 $y_n(t)=\mathrm{e}^{(N-n)\zeta i+zt}$, 将 $y_n(t)$按照傅里叶级数展

开,则有

$$\begin{cases}\dot{y}_n(t)=z\mathrm{e}^{(N-n)\zeta i+zt}\\ \ddot{y}_n(t)=z^2\mathrm{e}^{(N-n)\zeta i+zt}\\ \Delta y_n(t)=\mathrm{e}^{(N-n)\zeta i+zt}(\mathrm{e}^{\zeta i}-1)\\ \Delta\dot{y}_n(t)=z\mathrm{e}^{(N-n)\zeta i+zt}(\mathrm{e}^{\zeta i}-1)\end{cases} \tag{14}$$

将式(14)代入式(13)可得

$$z^2+z\left[a_1-\sum_{j=1}^{m}\lambda_j(\mathrm{e}^{j\zeta i}-\mathrm{e}^{(j-1)\zeta i})\right]=a_1V'(b)(\mathrm{e}^{\zeta i}-1)+a_2Tz(a_3\mathrm{e}^{\zeta i}-1) \tag{15}$$

令$z=\phi+\omega\mathrm{i}$,代入式(15),可得

$$\begin{aligned}&\phi^2+2\phi\omega\mathrm{i}-\omega^2+\phi\left[a_1-\sum_{j=1}^{m}\lambda_j(\mathrm{e}^{j\zeta\mathrm{i}}-\mathrm{e}^{(j-1)\zeta\mathrm{i}})\right]+\\&\omega\sigma_s-a_1V'(b)(\cos\zeta-1)-\\&a_2a_3T\phi\cos\zeta+a_2T\phi+a_2a_3T\omega\sin\zeta=\\&a_1V'(b)\mathrm{i}\sin\zeta+\mathrm{i}\omega^2\sin\zeta-\mathrm{i}\omega(a_1-\sigma_c)+\\&\mathrm{i}a_2a_3T\phi\sin\zeta+\mathrm{i}a_2a_3T\omega\cos\zeta-\mathrm{i}a_2T\omega\end{aligned} \tag{16}$$

式中:$\sigma_s=\sum_{j=1}^{m}\lambda_j\Big(\cos(j\zeta)-\cos((j-1)\zeta)\Big)$;$\sigma_c=\sum_{j=1}^{m}\lambda_j\cdot\Big(\sin(j\zeta)-\sin((j-1)\zeta)\Big)$。

令$\phi=0$,可得

$$\begin{aligned}&-\omega^2+\omega\sigma_s-a_1V'(b)(\cos\zeta-1)+a_2a_3T\omega\sin\zeta=\\&a_1V'(b)\mathrm{i}\sin\zeta+\mathrm{i}\omega^2\sin\zeta-\mathrm{i}\omega(a_1-\sigma_c)+\\&\mathrm{i}a_2T\omega(\cos\zeta-1)\end{aligned} \tag{17}$$

令式(17)实部、虚部分别为零,即

$$0=-\omega^2+\omega\sigma_s-a_1V'(b)(\cos\zeta-1)+a_2a_3T\omega\sin\zeta \tag{18}$$

$$\begin{aligned}0=&a_1V'(b)\sin\zeta+\omega^2\sin\zeta-\omega(a_1-\sigma_c)+\\&a_2a_3T\omega\cos\zeta-a_2T\omega\end{aligned} \tag{19}$$

直接求解式(19)中关于ω结果较为困难,但因为最终需要$\zeta\to0$(见式(22)),所以,可忽略式(19)中ω^2的系数$\sin\zeta$,则式(19)退化为

$$0=a_1V'(b)\sin\zeta-\omega(a_1-\sigma_c)+a_2a_3T\omega\cos\zeta-a_2T\omega \tag{20}$$

解得

$$\omega=\frac{a_1V'(b)\sin\zeta}{a_1-\sigma_c-a_2T(a_3\cos\zeta-1)} \tag{21}$$

代入式(18),可解得

$$V'(b)=[a_1-\sigma_c-a_2T(a_3\cos\zeta-1)]\sin\zeta+\frac{(1-\cos\zeta)[a_1-\sigma_c-a_2T(a_3\cos\zeta-1)]^2}{a_1\sin^2\zeta} \tag{22}$$

在$\zeta\to0$处,应用洛必达法则,可得临界稳定曲线:

$$V'(b)=\frac{a_1}{2}+\sum_{j=1}^{m}\lambda_j \tag{23}$$

稳定域为

$$a_1\geqslant2\left(V'(b)-\sum_{j=1}^{m}\lambda_j\right) \tag{24}$$

3 跟驰模型参数校准

3.1 校准方法

为具有优化速度函数的跟驰模型寻找最优参数,相当于对一个非线性优化问题数值求解,参考Kesting等[17]应用轨迹数据标定参数实验,利用航迹数据对优化速度函数进行重新标定,其中,优化速度函数如下:

$$V=\frac{1}{2}v_0\left[\tanh\left(\frac{h_c-l_{\min}-L_{\mathrm{cra}}}{b_x}-\beta\right)-\tanh(-\beta)\right] \tag{25}$$

式中:v_0为自由流速度,即自由交通条件下的理想速度;h_c为航空器机头间距;L_{cra}为航空器长度;$l_{\min}$为最小尾流距离;β为形式因子;b_x为参数。

3.1.1 模拟设置

航迹数据包含航空器的位置、速度、航向等信息,将其与跟驰模型的航迹数据进行比较,其中,前导航空器用作外部控制输入。用经验给定初始化速度差和距离差$v^{\mathrm{sim}}(t=0)=v^{\mathrm{data}}(0)$和$S^{\mathrm{sim}}(t=0)=S^{\mathrm{data}}(0)$。通过模拟航迹$x^{\mathrm{sim}}(t)$和前导航空器的给定位置$x_{\mathrm{lead}}^{\mathrm{data}}(t)$间的差来给出与前导航空器的间隙:

$$S^{\mathrm{sim}}(t)=x_{\mathrm{lead}}^{\mathrm{data}}(t)-x^{\mathrm{sim}}(t) \tag{26}$$

3.1.2 目标函数

校准过程旨在最小化实际航空器测量数据与跟驰模型模拟数据间的差异。使用间隙S作为误差测量,考虑以下3种不同的误差度量。

1)相对误差。

$$f_{\mathrm{rel}}(S^{\mathrm{sim}})=\sqrt{\left\langle\left(\frac{S^{\mathrm{sim}}-S^{\mathrm{data}}}{S^{\mathrm{data}}}\right)^2\right\rangle} \tag{27}$$

式中:〈 〉表示持续时间ΔP的时间序列的时间平均值,即

$$\langle x\rangle=\frac{1}{\Delta P}\int_0^{\Delta P}x(t)\mathrm{d}t \tag{28}$$

2)绝对误差。

$$f_{\mathrm{abs}}(S^{\mathrm{sim}})=\sqrt{\frac{\langle(S^{\mathrm{sim}}-S^{\mathrm{data}})\rangle^2}{\langle S^{\mathrm{data}}\rangle^2}} \tag{29}$$

3)由于绝对误差高估了高速范围内机头时距的误差,相对误差高估了低速范围内车头时距的误差,综合考虑这2种误差测量,定义混合误差测度:

$$f_{\mathrm{mix}}(S^{\mathrm{sim}})=\sqrt{\frac{1}{\langle|S^{\mathrm{data}}|\rangle}\left\langle\frac{(S^{\mathrm{sim}}-S^{\mathrm{data}})^2}{|S^{\mathrm{data}}|}\right\rangle} \tag{30}$$

3.2　参数约束

为将优化的参数空间限制为合理的参数值, 最小值和最大值应满足以下约束: 所需速度v_0的允许参数区间为 [60,300] m/s, 最小尾流距离$l_{\min}$取值为 7 400 m, 形式因子β的允许参数区间为 [0.10,10], 参数b_x的取值区间为 [10,1 000]。

3.3　遗传算法优化参数

使用遗传算法迭代寻优非线性优化问题的近似解, 实现过程如下:

1) 个体为跟驰模型参数集, 群体共 N 个集合。

2) 在每一代中, 通过目标函数 (27)、(29) 或 (30) 之一来确定群体中每个个体的适合度。

3) 基于轮盘赌法从当前群体中随机选择成对的 2 个个体, 并重新组合以产生新的个体。

4) 采用精英策略保留最佳个体, 其余所有个体的“基因”, 即它们的模型参数, 由给定概率控制的突变进行随机变化。

5) 终止标准分为 2 步。首先, 要达到给定数量的世代。然后, 在达到给定数量的世代的基础上, 根据最佳世代分数终止进化。

3.4　校准结果

通过应用所描述的优化方法, 确定跟驰模型与经验数据的最佳拟合, 所考虑的 3 个目标函数的校准结果如表 1 所示。

表 1　模型参数取值

Table 1　Parameter values of model

目标函数	v_0/(m·s^{-1})	β	b_x
$f_{\text{rel}}(S^{\text{sim}})$	157.8	1.02	136
$f_{\text{abs}}(S^{\text{sim}})$	196.4	2.56	204
$f_{\text{mix}}(S^{\text{sim}})$	250	1.13	160

4　仿真实验

4.1　稳定性仿真分析

令 $m=3$, 即考虑 3 架前导航空器的速度信息。与地面交通类似, 对航空器加速度起决定性作用的是式(8)的前半部分[18], 因此, 反应系数 $\lambda\in[0,1]$。相比地面交通, 航空器跟驰距离明显要长, 且随距离增加反应系数逐渐减小, 因此, 航空器速度差的反应系数λ取值小于地面交通, 设置为: $\lambda_1=0.1$, $\lambda_2=0.07$, $\lambda_3=0.04$。根据稳定性判别式 (24) 可求得选取的常规速度差模型, 以及 3 架前导航空器模型的稳定临界曲线。

当$m=0$时, 模型稳定条件为

$$a_1\geqslant 2V'(b) \tag{31}$$

当$m=1$时, 模型的稳定条件为

$$a_1\geqslant 2(V'(b)-\lambda_1) \tag{32}$$

当$m=2$时, 模型的稳定条件为

$$a_1\geqslant 2(V'(b)-\lambda_1-\lambda_2) \tag{33}$$

当$m=3$时, 模型的稳定条件为

$$a_1\geqslant 2(V'(b)-\lambda_1-\lambda_2-\lambda_3) \tag{34}$$

从不等式 (31)～(34) 可以看出, 模型的敏感系数a_1临界值依次减小, 说明考虑的前方航空器越多, 其稳定区域越广泛。图 2 给出了不同m取值的临界稳定曲线。可以看出, $m=3$时的稳定区域面积明显大于其他取值。由式 (24) 可知, m和λ_j越大, 模型稳定域越大, 因此, 考虑多速度差可以提高跟驰稳定性。

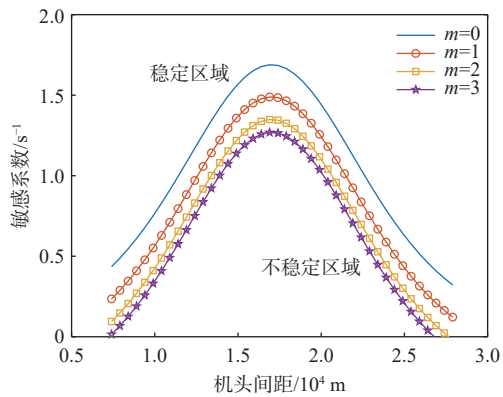


图 2　跟驰稳定域与敏感系数曲线

Fig. 2　Car-following stability domain and sensitivity coefficient curves

4.2　阻碍作用仿真

分别对轻型机、中型机和重型机在不同方向发生偏移的情形进行仿真, 结果对比如图 3 所示。

从图 3 可以看出, 无论是水平方向还是垂直方向, 阻碍作用均随偏移量的增大而减小, 当偏移增加到一定程度时, 阻碍作用可以忽略不记, 与实际运行符合; 分别对比 3 种机型, 在相同偏移情况下, 由于尾流间隔及机型不同, 重型机的阻碍作用最大, 轻型机最小。

4.3　通行能力

由式 (8)、式 (25) 可得飞联网环境下空中交通流多速度差跟驰模型下飞联网航空器的平衡态机头间距表达式, 如下:

$$h_c=\left(\operatorname{artanh}\left(\left(\frac{2v}{v_0}\right)+\tanh(-\beta)\right)+\beta\right)b_x+l_{\min}+L_{\text{cra}} \tag{35}$$

根据航空器航路密度与航空器机头间距的倒数关系, 可求出飞联网环境下航空器的交通流基本图模型:

$$k_c=\frac{1\,000}{h_c} \tag{36}$$

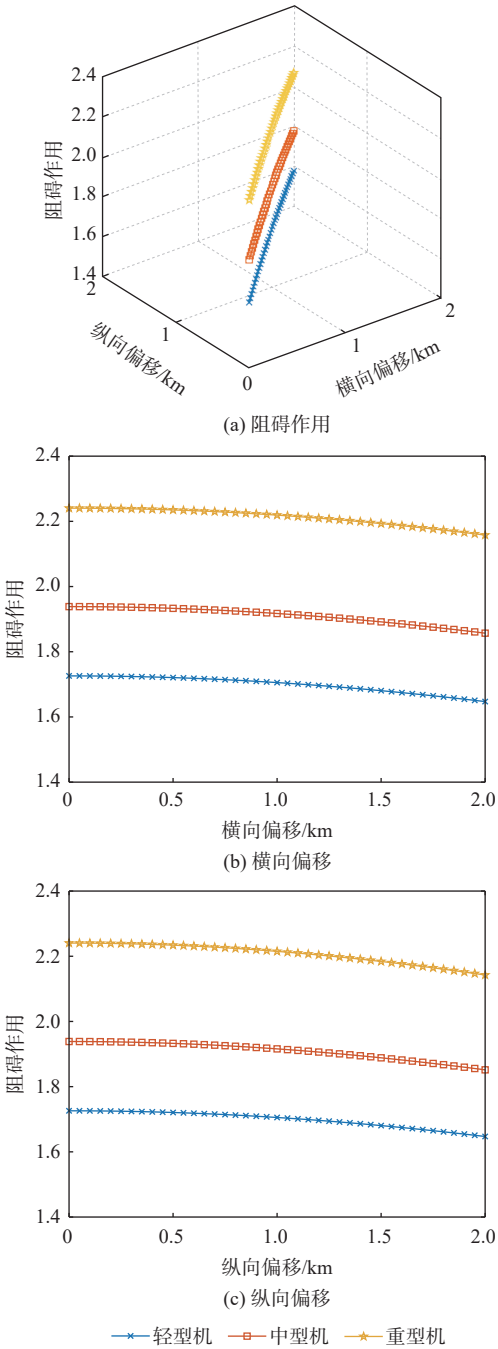


图3 航空器阻碍作用对比

Fig. 3 Comparison of aircraft hindrance effects

$$q_c = 3.6k_c v \tag{37}$$

式中：k_c为空中交通流密度；q_c为空中交通流流量。

由此可求得飞联网环境中空中交通流多速度差跟驰模型下的航空器巡航阶段的航路稳定状态时通行能力为$k_c = 65$架次/h。

4.4 航路环境仿真

为说明本文提出的多速度差跟驰模型有助于稳定性的提高，进行航路上的数值模拟，其中，偏移阻碍作用系数$a_2 = 0.46$，前机速度系数$a_3 = 1.03$、$\lambda_1 = 0.1$，$\lambda_2 = 0.07$，$\lambda_3 = 0.04$，航空器队列数量取20架，并与常规速度差跟驰模型进行对比。

取$a_1 = 1.026\ \mathrm{s}^{-1}$，由不等式(31)～(34)可知，常规速度差跟驰模型是不稳定的；$m = 1,2$时的多速度差跟驰模型也是不稳定的，$m = 3$时的空中交通流多速度差跟驰模型是稳定的。在有初始扰动的情况下，传统速度差跟驰模型和多速度差跟驰模型($m = 2,3$)中所有航空器的速度分布如图4所示。

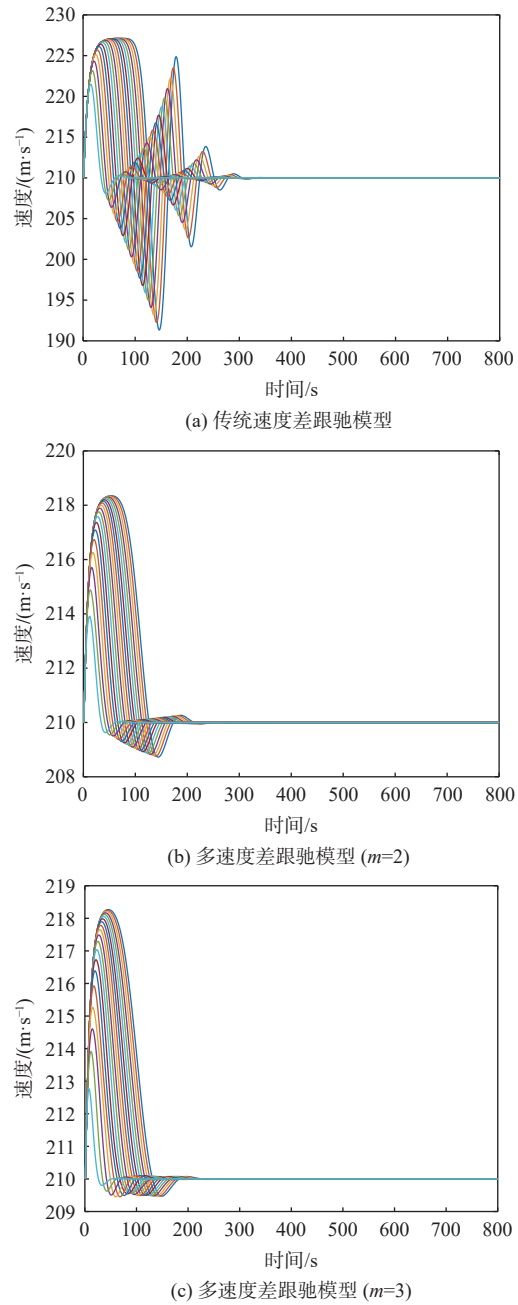


图4 航空器速度分布

Fig. 4 Aircraft velocity distribution

由图4可知，在加入初始扰动后，航空器速度都在初始速度210 m/s两侧出现波动，但多速度差跟驰模型的速度波动幅度明显小于传统速度差跟驰模型，且$m = 3$时的速度波动幅度小于$m = 2$时。传统速度差跟驰模型中，航空器速度波动幅度随时间的推移而减小，在0～350 s时段内，速度经过了

6 次幅度逐渐减小的波动, 空中交通流在此过程中是失稳的, 在 $t=400$ s 之后, 传统速度差跟驰模型的航空器速度不再发生明显变化。而多速度差跟驰模型 ($m=2,3$), 在 0～150 s 时段内, 航空器速度波动较为明显, 随着时间的推移, $t=200$ s 时航空器速度趋于平稳, 航空器的速度在 210 m/s 处的波动已经微乎其微, 系统已达到稳定状态, 交通流在演化过程中形成了稳态交通流。相比传统速度差跟驰模型, 多速度差跟驰模型无论是速度波动幅度还是恢复稳定所需时间, 均优于传统速度差跟驰模型, 这与前述线性稳定性分析结果一致。

此外, 与多速度差跟驰模型 ($m=2$) 相比, 多速度差跟驰模型 ($m=3$) 航空器的速度波动幅度减小程度不是很明显, 原因在于: 第 3 前机的敏感系数较小; 越靠前的航空器对跟驰航空器的影响越小, 跟驰航空器受第 1 前机的影响最大, 这与实际空中交通情况相符。

进一步验证所提多速度差跟驰模型空中交通流的演化规律及航空器因外部环境产生偏移或进行加减速时的变化情况, 仍采用上文初始条件进行数值模拟, 此时, 多速度差 ($m=3$) 跟驰模型是稳定的, 模拟结果如图 5 和图 6 所示。

从图 5 中可以看到, 前导航空器由于外部原因

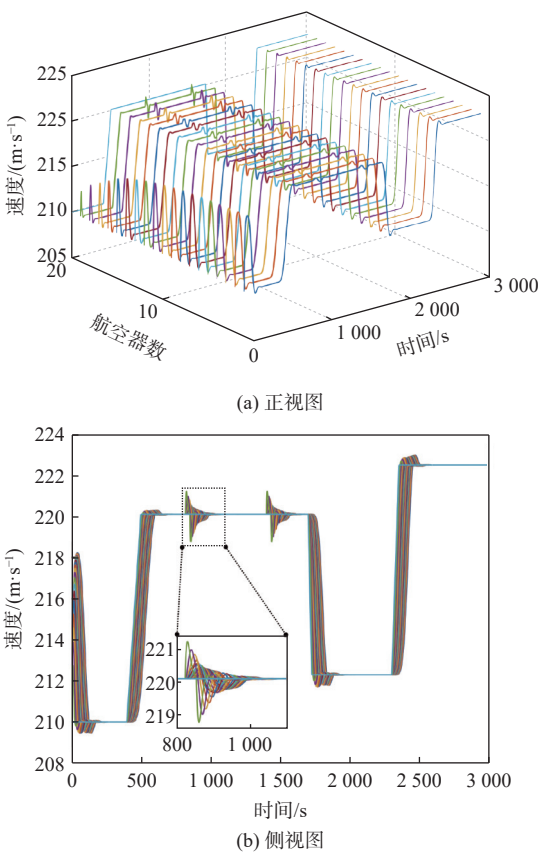


图 5　多速度差跟驰模型航空器速度分布

Fig. 5　Aircraft velocity distribution of multi-velocity difference car-following model

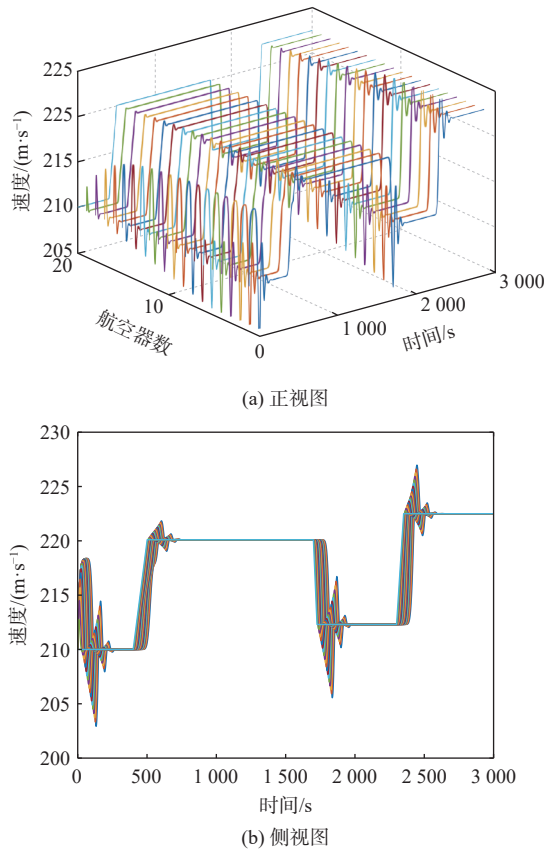


图 6　传统速度差跟驰模型航空器速度分布

Fig. 6　Aircraft velocity distribution of traditional velocity difference car-following model

引起的速度波动沿航空器队列向后传播, 且速度变化存在延迟, 速度波动幅度随着时间的增大而减小, 每架航空器的速度在平均速度上下起伏, 经过若干次波动后才平稳下来。体现出前文提到的飞联网航路运行状态下的非自由运行航空器队列的制约性、延迟性和传递性。此外, 所提多速度差跟驰模型无论是速度波动的幅度还是速度波动衰减的速率, 都优于传统的速度差跟驰模型。这是由于多速度差跟驰模型的稳定性比传统速度差跟驰模型强, 在多速度差跟驰模型中加入的扰动很快就消散了, 传统速度差跟驰模型稳定性差造成速度波动大, 衰减较慢。

通过图 5 还可以看出, 在 $t=820$ s 和 $t=1\,420$ s 时, 前导航空器产生偏移, 后方航空器因阻碍作用的减小而进行加速度逐渐减小的加速运动, 当前导航空器恢复原位置时, 后方航空器因阻碍作用增大而进行加速度逐渐减小的减速运动, 证明所建模型可以反映航空器间偏移的影响。

从资源角度看, 空中交通流速度的波动幅度与航空燃油消耗是正相关关系。空中交通流速度波动幅度减小及稳定性提高可以降低燃油的额外消耗。航空器性能数据库 BADA 是开展航空器性能分析和油耗计算等研究的重要工具。基于 BADA,

采用式(38)来衡量航空器的燃油消耗情况[19]。

$$P=\sum_{i=1}^{n}\int_{0}^{T_{\text{change},i}}[W_{r,i}(t)(1+A_{\text{cra},i}(t))]\mathrm{d}t \tag{38}$$

式中：n为队列航空器数；$T_{\text{change},i}$为航空器i速度变化的时间；$W_{r,i}(t)$为在航空器i速度曲线上t时刻的速度所对应发动机推力；$A_{\text{cra},i}$为航空器i在时刻t的加速度；P为燃油消耗指数，与油耗呈正比关系，但并不表示具体的燃油消耗量。

燃油消耗指数如图7所示。可以看出，敏感系数较小时，传统速度差跟驰模型的燃油消耗指数远大于多速度差跟驰模型，且$m=3$时的多速度差跟驰模型最小，在$a_1=1\ \text{s}^{-1}$时相比传统模型，所提模型的燃油指数降低了27.12%。随敏感系数增大，不同m取值的4种模型燃油消耗指数逐渐降低，且模型间的差距逐渐减小，即缩短系统延迟和人为延迟，能够更有效地减少燃油消耗，提高交通流的稳定性，民航飞联网技术可以有效提高系统敏感系数，因此，基于飞联网的多速度差跟驰模型对于绿色民航具有现实意义。

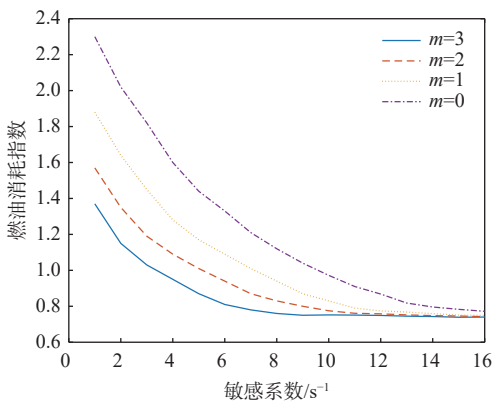


图7 燃油消耗指数

Fig. 7 Fuel consumption index

5 结 论

1）通过引入偏移阻碍作用的概念，建立偏移与前导航空器速度的关系，实现了偏移对跟驰影响的定量描述，并验证了阻碍作用随偏移量的增大而减小。

2）所构建的飞联网运行环境下的航空器多速度差模型无论是速度波动幅度还是速度波动衰减速率都优于传统的速度差模型，可以有效提高空中交通流队列的稳定性。

3）所构建的飞联网环境下的航空器多速度差模型，通过减少速度波动，提升航空器队列稳定性，从而达到降低航空器队列的燃油消耗的目的，符合未来绿色民航的发展需求。

未来应进一步考虑航空队列不同机型混合运行的场景，以及将后方航空器的速度信息纳入到跟驰模型中，进一步提高空中交通流的稳定性。

参考文献（References）

[1] 王飞, 韩翔宇. 基于分形插值的空中交通流量短期预测[J]. 航空学报, 2022, 43(9): 325585.
WANG F, HAN X Y. Short-term prediction of air traffic flow based on fractal interpolation[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(9): 325585(in Chinese).

[2] 赵嶷飞, 王梦琦. 空中交通工程学理论内涵与关键科学技术[J]. 航空学报, 2022, 43(12): 026537.
ZHAO Y F, WANG M Q. Important theories and critical scientific technology of air traffic engineering[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(12): 026537(in Chinese).

[3] 马园园, 胡明华, 张洪海, 等. 多机场终端区进场航班协同排序方法[J]. 航空学报, 2015, 36(7): 2279-2290.
MA Y Y, HU M H, ZHANG H H, et al. Optimized method for collaborative arrival sequencing and scheduling in metroplex terminal area[J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(7): 2279-2290(in Chinese).

[4] 王超, 郭宗震, 李冬. 基于三参数Burr分布的机头时距分布特性研究[J]. 计算机仿真, 2020, 37(5): 35-39.
WANG C, GUO Z Z, LI D. Research on the time-headway distribution characteristics of nose based on three-parameter Burr distribution[J]. Computer Simulation, 2020, 37(5): 35-39(in Chinese).

[5] ZHANG H H, XU Y, YANG L, et al. Macroscopic model and simulation analysis of air traffic flow in airport terminal area[J]. Discrete Dynamics in Nature and Society, 2014, 2014(1): 741654.

[6] SHAFIENYA H, REGAN A C. 4D flight trajectory prediction using a hybrid deep learning prediction method based on ADS-B technology: a case study of Hartsfield-Jackson Atlanta International Airport (ATL)[J]. Transportation Research Part C: Emerging Technologies, 2022, 144: 103878.

[7] GUI G, ZHOU Z Q, WANG J, et al. Machine learning aided air traffic flow analysis based on aviation big data[J]. IEEE Transactions on Vehicular Technology, 2020, 69(5): 4817-4826.

[8] ZENG W L, LI J, QUAN Z B, et al. A deep graph-embedded LSTM neural network approach for airport delay prediction[J]. Journal of Advanced Transportation, 2021, 2021: 6638130.

[9] 黎新华, 张兆宁, 王莉莉. 基于跟驰稳定的终端区容量评估方法[J]. 系统工程理论与实践, 2009, 29(2): 173-179.
LI X H, ZHANG Z N, WANG L L. Terminal area capacity evaluation method based on following stability[J]. Systems Engineering-Theory & Practice, 2009, 29(2): 173-179(in Chinese).

[10] 王莉莉, 张新瑜, 张兆宁. 空中高速路交通流的跟驰现象及流量模型[J]. 西南交通大学学报, 2012, 47(1): 158-162.
WANG L L, ZHANG X Y, ZHANG Z N. Following phenomenon and air freeway flow model[J]. Journal of Southwest Jiaotong University, 2012, 47(1): 158-162(in Chinese).

[11] 张洪海, 胡勇, 杨磊, 等. 多机场终端区微观交通流建模与仿真分析[J]. 西南交通大学学报, 2015, 50(2): 368-374.
ZHANG H H, HU Y, YANG L, et al. Modeling and simulation analysis of microscopic traffic flow in multi-airport terminal airspace[J]. Journal of Southwest Jiaotong University, 2015, 50(2):

368-374(in Chinese).

[12] 张洪海, 杨磊, 别翌荟, 等. 终端区进场交通流广义跟驰行为与复杂相变分析[J]. 航空学报, 2015, 36(3): 949-961.
ZHANG H H, YANG L, BIE Y H, et al. Analysis on generalized following behavior and complex phasetransition law of approaching traffic flow in terminal airspace[J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(3): 949-961(in Chinese).

[13] 任一雄, 徐怀. 参数情形下第一类曲面积分的计算[J]. 上饶师范学院学报, 2017, 37(6): 16-19.
REN Y X, XU H. Calculation for the first type surface integral in parameter equation[J]. Journal of Shangrao Normal University, 2017, 37(6): 16-19(in Chinese).

[14] BANDO M, HASEBE K, NAKAYAMA A, et al. Dynamical model of traffic congestion and numerical simulation[J]. Physical Review E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, 1995, 51(2): 1035-1042.

[15] JIANG R, WU Q, ZHU Z. Full velocity difference model for a car-following theory[J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2001, 64: 017101.

[16] 秦严严, 王昊, 冉斌. 考虑多前车反馈的智能网联车辆跟驰模型[J]. 交通运输系统工程与信息, 2018, 18(3): 48-54.
QIN Y Y, WANG H, RAN B. Car-following model of connected and autonomous vehicles considering multiple feedbacks[J]. Journal of Transportation Systems Engineering and Information Technology, 2018, 18(3): 48-54(in Chinese).

[17] KESTING A, TREIBER M. Calibrating car-following models by using trajectory data[J]. Transportation Research Record: Journal of the Transportation Research Board, 2008, 2088(1): 148-156.

[18] 吴兵, 王文璇, 李林波, 等. 多前车影响的智能网联车辆纵向控制模型[J]. 交通运输工程学报, 2020, 20(2): 184-194.
WU B, WANG W X, LI L B, et al. Longitudinal control model for connected autonomous vehicles influenced by multiple preceding vehicles[J]. Journal of Traffic and Transportation Engineering, 2020, 20(2): 184-194(in Chinese).

[19] 王湛, 熊佳俊, 陈浩. 航空器场面滑行速度与油耗研究[J]. 武汉理工大学学报 (交通科学与工程版), 2017, 41(3): 391-394.
WANG Z, XIONG J J, CHEN H. A research on aircraft taxiing speed and fuel consumption[J]. Journal of Wuhan University of Technology(Transportation Science & Engineering), 2017, 41(3): 391-394(in Chinese).

Aircraft multi-velocity difference car-following model based on flight networking operation

WANG Lili*, ZHAO Yunfei, GUO Weimeng

(Key Laboratory of Internet of Aircrafts, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In order to improve the stability of air traffic flow, this study investigated an aircraft multi-velocity difference car-following model considering offsets based on the characteristics of flight networking operations. Firstly, to quantitatively describe the influence of offsets between aircraft on the car-following behavior, the concept of offset hindrance was introduced, and the relationship between offset and the velocity of the leading aircraft was established, extending the car-following model to a three-dimensional mode. Secondly, by considering the multi-aircraft information interaction mode in the flight networking environment, a multi-velocity difference aircraft car-following model was constructed, and stability analysis methods were applied to derive the stability discrimination conditions and calculate the steady-state traffic capacity of the proposed model. Finally, based on the parameter calibration of the model, numerical simulation experiments were designed by using the multi-velocity difference car-following model considering three leading aircraft. The results show that the hindrance effect decreases as the offset increases, and for the same offset value, heavy aircraft have the highest hindrance effect while light aircraft have the lowest hindrance effect. The proposed model has a better stable region compared to traditional models, and the stability of the proposed model improves with an increase in the number of leading aircraft and the weight coefficient. Under the same parameter values, the fuel consumption coefficient of the proposed model is lower than that of the traditional car-following model, and when the sensitivity coefficient is set to 1 s^{-1}, the fuel consumption coefficient decreases by 27.12%. Numerical simulations demonstrate that the aircraft multi-velocity difference model contributes to improving the stability of air traffic flow and reducing fuel consumption.

Keywords: flight networking; multi-velocity difference; car-following model; stability analysis; traffic capacity; fuel consumption

Received: 2023-06-12; **Accepted**: 2023-09-22; **Published Online**: 2023-10-11 13:17

URL: link.cnki.net/urlid/11.2625.V.20231011.1038.001

Foundation item: Jointly Fund of National Natural Science Foundation of China and Civil Aviation Administration of China (U1633124)

* **Corresponding author**. E-mail: llwang317@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0404

基于双重随机扰动的人工大猩猩部队优化算法及工程应用

杜晓昕[1,2,*]，郝田茹[1]，王波[1,2]，王振飞[1]，张剑飞[1,2]，金梅[1,2]

(1. 齐齐哈尔大学 计算机与控制工程学院，齐齐哈尔 161006；
2. 齐齐哈尔大学 黑龙江省大数据网络安全检测分析重点实验室，齐齐哈尔 161006)

摘 要：针对人工大猩猩部队优化算法(GTO)存在易陷入局部最优、收敛速度慢、寻优精度低等问题，提出了基于双重随机扰动策略的人工大猩猩部队优化算法(DGTO)。引入Halton序列初始化种群，增加种群的多样性；在算法寻优阶段使用多维随机数策略，并在探索阶段提出自适应位置搜索机制，提高算法的收敛速度；提出双重随机扰动策略，解决大猩猩的群居效应，增强算法跳出局部最优的能力；采用逐维更新策略更新个体位置，提升算法的收敛精度。通过10个基准测试函数寻优结果及Wilcoxon秩和检验对比可知，改进算法在寻优精度、收敛速度上有较大提升。同时，通过工程优化问题的实验对比分析，进一步验证了改进算法在处理现实工程问题上的优越性。

关 键 词：人工大猩猩部队优化算法；Halton序列；自适应位置搜索；双重随机扰动策略；逐维更新

中图分类号：TP301.6

文献标志码：A **文章编号**：1001-5965(2025)06-1882-15

群智能优化算法具有参数少、简单易实现、不需要梯度信息等优点，通过模拟自然界中动植物的求偶、觅食等行为，可以很好地在合理的时间且高度复杂的约束条件下找到这些问题的最优解。经典的群智能优化算法有粒子群(particle swarm optimization, PSO)算法[1]、人工蜂群(artificial bee colony, ABC)算法[2]、蚁群(ant colony optimization, ACO)算法[3]等。随着技术的不断发展和更新，群智能优化算法在旅行商[4]、支持向量机优化[5]、路线规划[6]、电力系统控制[7]等问题中有着卓越表现。

人工大猩猩部队优化算法(artificial gorilla troops optimizer, GTO)[8]是由Abdollahzadeh等于2021年提出的一种新型元启发式优化算法，通过对自然界中大猩猩个体觅食及争夺配偶行为的模拟而构造的一种有效的优化方案，虽然该算法具有原理简单易实现、所需调节参数少等优点，但依然存在易陷入局部最优、收敛速度慢等不足。自该算法提出以来，国内外不少学者针对其不足之处做出了改进。例如，Xiao等[9]提出了基于镜像对立和自适应-β爬山算法的大猩猩优化算法，采用凸透镜成像反向学习扩大搜索范围，避免陷入局部最优，同时引入自适应-β爬山算法与GTO结合以提高求解精度；Liang等[10]提出了一种基于反向学习和并行策略的GTO算法，通过反向学习拓展算法的探索空间，提高算法的全局搜索能力，并行策略将种群分为多个群体进行探索，增加了种群的多样性；Wu等[11]利用基于二次插值的天牛须搜索增强银背大猩猩位置的多样性，引入教与学优化算法的教师阶段，

收稿日期：2023-06-25；录用日期：2023-09-11；网络出版时间：2023-10-20 11:49
网络出版地址：link.cnki.net/urlid/11.2625.V.20231020.0903.001
基金项目：黑龙江省省属高等学校基本科研业务费自然科学类青年创新人才项目(145209206)
*通信作者. E-mail：xiaoxindu@qqhru.edu.cn

引用格式：杜晓昕，郝田茹，王波，等. 基于双重随机扰动的人工大猩猩部队优化算法及工程应用[J]. 北京航空航天大学学报，2025，51(6)：1882-1896. DU X X，HAO T R，WANG B，et al. Artificial gorilla troops optimizer based on double random disturbance and its application of engineering problem[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1882-1896 (in Chinese).

以 50% 的概率更新跟随银背的行为, 并由准反射学习生成银背大猩猩的准反射位置; Alsolai 等[12] 提出增强型大猩猩, 利用 circle 混沌映射增加大猩猩的种群多样性, 并将其应用于无人机辅助智能车载网络中的分簇协议中; Mostafa 等[13] 引入精英反向学习增强种群的多样性, 利用柯西逆累积分布算子和切线飞行的融合提高种群的开发能力和收敛速度。尽管上述研究对 GTO 算法的寻优精度和速度有了一定的提升, 但该算法提出时间较短, 改进方法还不够完善, 仍然存在以下不足: ①迭代更新前的初始种群依赖初始条件较多, 健壮性不足; ②种群个体迭代更新策略较为单一, 且使用方式较为机械, 没有合理地平衡全局搜索与局部开发的关系; ③搜索过程中易掉入局部最优陷阱, 导致算法收敛精度不高; ④对于种群个体好坏的评价指标多样性差, 仅根据适应度函数的好坏便否决了种群个体的质量; ⑤种群个体位置更新方式不够细化。因此, 对于 GTO 算法的改进还需要更深入的研究。

本文提出基于双重随机扰动策略的人工大猩猩部队优化算法 (artificial gorilla troops optimizer based on double random disturbance strategy, DGTO) 。首先, 通过 Halton 序列生成均匀分布的初始种群, 增加初始种群的多样性; 其次, 在算法寻优阶段使用多维随机数策略, 并提出自适应位置搜索机制替代向已知位置搜索机制, 提高了算法的收敛速度; 然后, 提出一种双重随机扰动策略, 解决了大猩猩的群居效应, 避免算法过早收敛; 最后, 采用逐维更新策略更新个体位置, 提升了算法的收敛精度。通过对 10 个基准测试函数、Wilcoxon 秩和检验进行分析, 验证了本文算法的优越性, 并通过工程应用问题的实验结果分析, 验证了本文算法较其他算法更具有可行性。

1　人工大猩猩部队优化算法

GTO 算法是根据大猩猩的集体生活方式和社交方式提出的一种群智能优化算法。大猩猩生活在一个称为“部队”的群体中, 由成年雄性或银背大猩猩群体和几只成年雌性大猩猩及其后代组成, 银背大猩猩是群体的核心, 它做出所有决定, 调解战斗, 确定群体行动, 指导大猩猩寻找食物来源, 并为群体的安全福祉负责。在 GTO 算法中, 有 5 种不同的算子用于模拟大猩猩的群体行为, 主要分为 2 个阶段: 探索阶段和开发阶段。探索阶段采用 3 种不同的机制, 即迁移到未知位置、迁移到其他大猩猩位置及迁移到已知位置; 开发阶段采用跟随银背大猩猩和竞争成年雌性大猩猩 2 种社会行为。

大猩猩在探索阶段的位置更新方式如下:

$$G(t+1)=\begin{cases}(u_{\mathrm{b}}-l_{\mathrm{b}})r_1+l_{\mathrm{b}} & r<s\\ (r_2-C)X_{\mathrm{r}}(t)+LH & r\geqslant 0.5\\ X(t)-L[L(X(t)-G_{\mathrm{r}}(t))+r_3(X(t)-G_{\mathrm{r}}(t))] & r<0.5\end{cases}\tag{1}$$

式中: $G(t+1)$ 为下一次迭代中的大猩猩候选位置; $X(t)$ 为大猩猩的当前位置; r_1、r_2、r_3 和 r 表示取值为 (0,1) 的随机值; s 为优化操作前给定值的参数, 一般取 0.03, 其决定了大猩猩迁移到未知位置的概率; u_{b} 和 l_{b} 分别为变量的上限和下限; X_{r} 和 G_{r} 分别为从当前位置和候选位置中随机选择的大猩猩个体位置; C、L、H 分别表示为

$$C=(\cos(2r_4)+1)\left(1-\frac{I_t}{I_{t,\max}}\right)\tag{2}$$

$$L=Cl\tag{3}$$

$$H=ZX(t)\tag{4}$$

其中: I_t 为当前迭代值; $I_{t,\max}$ 为执行优化操作的迭代总值; r_4 表示取值为 (0,1) 的随机值; l 表示取值为 (−1,1) 的随机值; Z 表示问题维度中取值为 $[-C,C]$ 的随机值。

大猩猩在开发阶段选择跟随银背大猩猩机制, 其行为表示为

$$G(t+1)=LM(X(t)-X_{\mathrm{silverback}})+X(t)\tag{5}$$

$$M=\left(\left|\frac{1}{N}\sum_{i=1}^{N}G_i(t)\right|^{g}\right)^{\frac{1}{g}}\tag{6}$$

$$g=2^L\tag{7}$$

式中: $X_{\mathrm{silverback}}$ 为银背大猩猩 (最佳解) 的位置; $G_i(t)$ 为第 i 个候选大猩猩在第 t 次迭代时的位置; N 为大猩猩的总数。

大猩猩在开发阶段选择竞争成年雌性大猩猩机制, 其行为表示为

$$G(t+1)=X_{\mathrm{silverback}}-(X_{\mathrm{silverback}}Q-X(t)Q)A\tag{8}$$

$$Q=2r_5-1\tag{9}$$

$$A=\beta E\tag{10}$$

$$E=\begin{cases}N_1 & r\geqslant 0.5\\ N_2 & r<0.5\end{cases}\tag{11}$$

式中: Q 用于模拟冲击力; A 为冲突中暴力程度的系数; r_5 表示取值为 (0,1) 的随机值; β 为优化操作前给定的参数值; E 用于模拟暴力对解的维度的影响, 如果 $r\geqslant 0.5$, E 等于正态分布和问题维度中的随

机值 N_1，反之，E 等于正态分布中的随机值 N_2。

2 改进的人工大猩猩部队优化算法

在标准 GTO 算法中，种群初始化采用随机分布的方式，这种分布方式使得种群分布极不均匀，进而导致算法收敛速度缓慢；参数 C 随机生成时未考虑问题维度，使得到的种群多样性不高，同时，探索阶段的向已知位置迁移机制也未能很好地驱使种群迁移到更好的位置；种群跟随银背大猩猩和普通成年大猩猩进行位置更新，当种群个体陷入局部极值时，算法会出现停滞搜索现象；在更新个体位置时，算法中的种群个体在进行位置更新时，并未考虑到每一维度的位置信息，而是选择整体更新的方式，导致算法收敛精度变低。

本文针对标准 GTO 算法的不足之处，提出相应的策略对其进行改善。

2.1 Halton 序列初始化种群

种群初始化是群智能优化算法不可或缺的阶段，传统的种群初始化方法依赖伪随机数生成，该方式实现简单且随机性强，但种群个体不一定均匀地分布在搜索空间中[14]，从而使种群个体错过相当大的一部分种群空间。基于混沌技术生成初始化种群的方式也被广泛应用，已有研究表明，利用混沌序列进行种群初始化能取得比伪随机数更好的效果，但该方式整体稳定而局部不稳定，且对初始条件及其参数设置非常敏感。此外，多步骤技术初始化种群也较为流行，通常是在后续步骤中细化先前生成的种群，其中，最流行的是基于反向学习的技术。首先，产生原始群体的点；然后，使用反向学习技术对生成点中不符合条件的点进行反向学习，这种方式虽然可以取得良好的效果，但会受到第一步生成的种群点质量的影响，且需要消耗一部分计算预算来评估适应度函数。基于此，本文引入较为先进的准随机序列方式，即 Halton 序列来初始化种群。Halton 序列是一种低差异序列，以质数为基数的确定性方法构造，使得种群更加均匀地分布在整个解空间，增加了种群的多样性，从而加快了算法的收敛速度，并提高了算法的收敛精度。

假设搜索空间为二维，Halton 序列的实现过程为：选取 2 个质数作为基数，分别对应 2 个维度，每个维度根据基数分别对 (0,1) 不断地进行切分、迂回取值，从而形成不重复且均匀的点，其切分过程的数学模型为

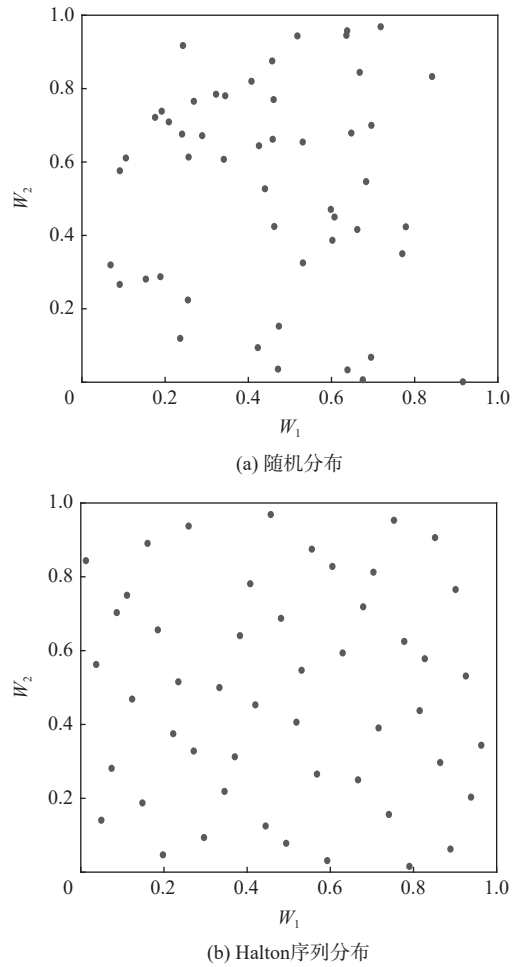
$$n=\sum_{i=0}^{m_0}a_i p_1^i=a_0+a_1p_1^1+\cdots+a_{m_0}p_1^{m_0} \tag{12}$$

$$\phi_{p_1}(n)=a_0p_1^{-1}+a_1p_1^{-2}+\cdots+a_{m_0}p_1^{-m_0-1} \tag{13}$$

$$H(n)=\{\phi_{p_{1,1}}(n),\phi_{p_{1,2}}(n)\} \tag{14}$$

式中：n 为 Halton 序列的序数；p_1 为 Halton 序列的基数，取值为大于或等于 2 的质数；m_0 为非负整数，用于确定求和的项数上限；$a_i\in\{0,1,2,\cdots,p_1-1\}$为常变量；$\phi_{p_1}(n)$为定义的序列函数；$H(n)$ 为二维均匀 Halton 序列。

在二维空间下，假设种群规模为 50，个体分布范围的上下界分别为 1 和 0，随机分布和 Halton 序列分布初始化种群空间个体分布对比如图 1 所示，其中，Halton 序列的基数分别为 2 和 3，W_1 和 W_2 为维度。由图 1 对比可知，通过 Halton 序列初始化种群空间的个体分布虽然降低了随机性，但分布更加均匀，增加了算法的多样性。



(a) 随机分布

(b) Halton序列分布

图 1 随机分布与 Halton 序列分布初始化种群空间

Fig. 1 Random distribution and Halton distribution initialize population space

2.2 自适应位置搜索机制

在标准 GTO 算法中，大猩猩探索阶段寻优时使用了 3 种不同的机制，即迁移到未知位置、迁移到其他大猩猩位置和向已知位置迁移。然而，原始的向已知位置迁移机制过多考虑大猩猩个体自身搜索的经验，没有自适应地平衡好大猩猩个体勘探

与开发之间的关系。针对群智能优化算法中普遍存在的缺乏自适应调整能力的问题, 出现了多种基于欧氏距离、适应度函数、迭代次数的自适应步长方法。文献 [15] 使用基于欧氏距离的自适应步长控制麻雀加入者和发现者各维度之间的距离, 当加入者和发现者最优位置较远时, 步长较大, 提高算法收敛速度, 反之, 减小步长, 提高算法的局部搜索能力; 文献 [16] 将果蝇种群每次寻优得到的适应度函数最好的全局最优解融入到寻优迭代的过程中, 避免果蝇个体盲目搜索; 文献 [17] 使用基于迭代次数的自适应步长方法优化步长因子, 使其随迭代次数增加不断减小。为使 GTO 算法能够自适应平衡全局搜索与局部开发之间的关系, 本文通过分别调整个体自身位置对迭代影响的自适应因子及其他大猩猩个体对本个体搜索影响的自适应因子, 提出一种自适应位置搜索机制替代该搜索机制, 具体表达式为

$$G(t+1)=TX(t)-r_6L(X(t)-G_{\mathrm{r}}(t)) \tag{15}$$

$$T=\left(\frac{1}{1+\mathrm{e}^{K}}\right)^{4} \tag{16}$$

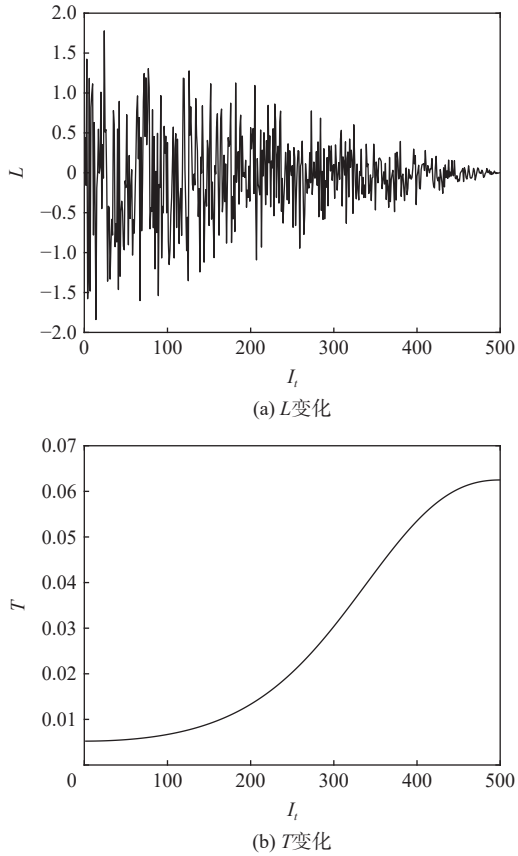
式中: L 的取值受 C 影响, 为使其考虑问题维度, 对 C 采用了多维随机数策略; r_6 表示取值为 (0,1) 的随机值; $K=\left(1-\dfrac{I_t}{I_{t,\max}}\right)^{2I_t/I_{t,\max}}$。

L 和 T 随迭代次数增加的变化曲线如图 2 所示。可知, 大猩猩个体自身位置的权重 T 取值从 0.005 不断上升到 0.063, 而为大猩猩个体受其他已知位置的影响部分 $X(t)-G_{\mathrm{r}}(t)$ 赋予权重 L, 其取值从 [−2,2] 不断向 0 靠近。在前 200 次迭代中, T 增长平缓, L 取值较大, 个体更注重其他位置的探索; 在 200～400 次迭代中, T 增长迅速, L 持续减小, 该阶段个体逐渐由探索其他位置转为开发自身邻域位置; 在最后 100 次迭代中, T 趋于平稳, 而 L 也逐渐稳定到 0 附近, 此时个体更注重自身邻域的搜索。

相比原始迁移到已知位置机制, 本文提出的自适应位置搜索机制更符合个体前期精于探索、后期精于开发的搜索模式, 进一步提高了算法的收敛速度。

2.3　双重随机扰动策略

大猩猩在社会生活中倾向于集体生活, 它们作为一个群体寻找食物, 并生活在银背大猩猩的领导下, 听从银背大猩猩的指挥决策。虽然大猩猩的集体生活方式通过个体之间的信息交流使得优质食物源更容易被个体搜寻到, 但也会因此产生潜在的群居效应, 即种群在找到自身认为是最优食物源后把它作为全局最好的食物源, 从而停止对更好食物源的探索。为减小群居效应对种群寻优产生的影响, 本文受元学习[18] 策略的启发, 通过对种群个体适应度值及适应度变化率的评估, 提出一种双重随机扰动策略帮助种群在完成位置迭代更新后跳出局部最优值。元学习策略即“学会学习的学习”, 即使使用少量样本数据, 也能快速学会新的概念或技能。元学习的样本是基于任务的, 其是从任务中学习特征表示, 从而实现在新任务上的泛化。然而, 各学习方法并不是单独存在的, 与其他学习方法存在潜在的联系。表 1 展示了元学习与其他学习方法之间的潜在关联。在双重随机扰动策略中, 寻找全局最优解过程中跳出局部最优解是种群个体的任务, 个体适应度值是否比适应度平均值低,



(a) L变化

(b) T变化

图 2　L 和 T 随迭代次数增加的变化曲线

Fig. 2　L and T change curves with increase of iterations

表 1　元学习与其他学习方法的潜在关联

Table 1　Potential correlation between meta-learning and other learning methods

学习方法	与元学习关联
对比学习	对比学习的核心是将正样本和负样本在特征空间进行对比, 使得模型能够学习到样本的重要内在特征。而元学习在学习特征表示时, 有时也会通过正样本和负样本比对的方式实现
迁移学习	迁移学习的核心是找到已有知识和新知识之间的相似性, 通过这种相似性的迁移达到迁移学习的目的。元学习基于任务展开学习, 然而生物体及其一生, 学习的永远不止一个任务, 因而迁移学习可以将元学习在某一任务的学习方式迁移到相似任务中

以及个体适应度变化率是否比优质个体适应度变化率大，即为种群个体的特征表示，个体根据自身特征选择不同的对策，双重随机扰动得以实现。由表1可知，双重随机扰动策略在元学习过程中，个体适应度值与适应度平均值的比较及个体适应度变化率与优质个体适应度变化率的比较结合了对比学习的核心思想，同时也提升了评判个体质量的多样性；而个体在双重随机扰动过程中学习到的如何跳出局部最优值的能力，也可以通过迁移学习迁移到更多相似的问题中。因此，该策略有良好的泛化性。双重随机扰动策略如图3所示，详细步骤如下：

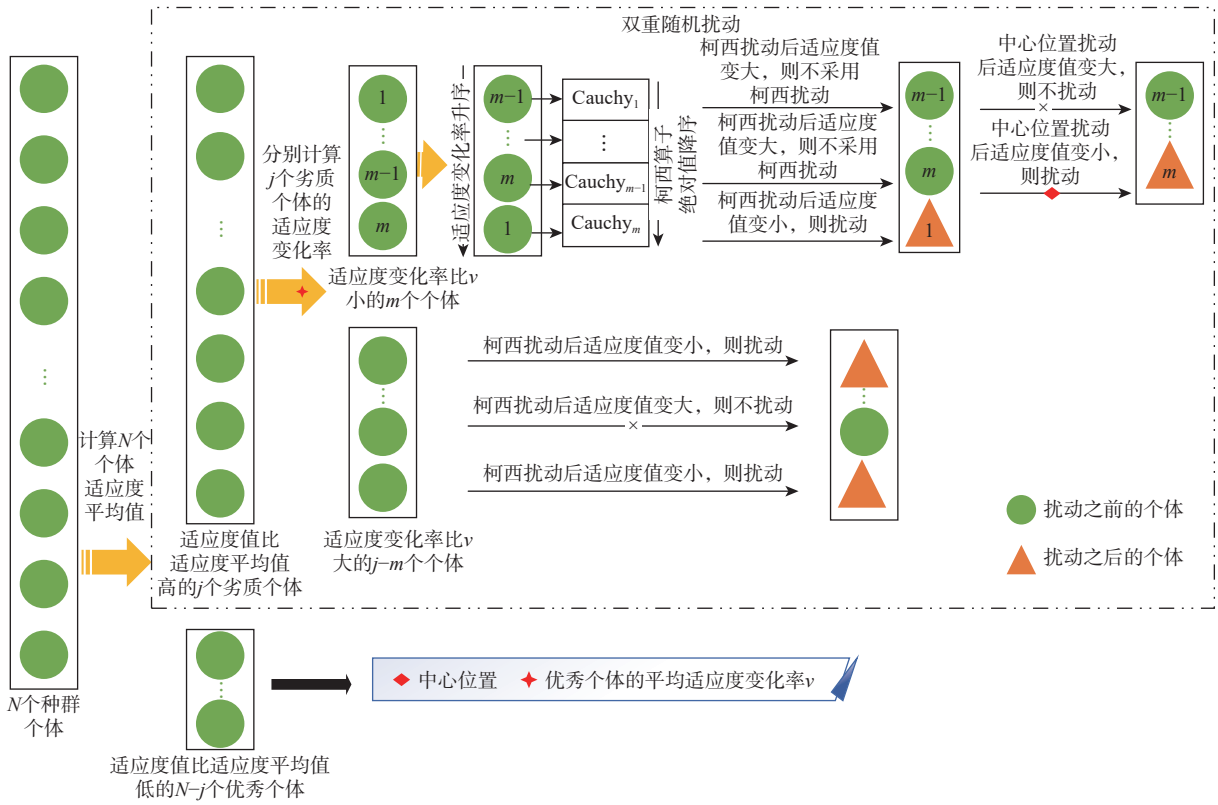


图3 双重随机扰动策略示意图

Fig. 3 Schematic diagram of double random disturbance strategy

步骤1 根据式(17)计算每次迭代后 N 个大猩猩个体的适应度平均值：

$$f_{\mathrm{v}}=\frac{\sum_{i=1}^{N} f_i}{N} \tag{17}$$

式中：f_i 为个体的适应度值。

步骤2 根据适应度平均值将种群个体分为 j 个比适应度平均值高的劣质个体和 $N-j$ 个比适应度平均值低的优质个体。

步骤3 根据式(18)分别计算 j 个劣质个体和 $N-j$ 个优质个体的适应度变化率 v_1，并求取优质个体的平均适应度变化率 v。

$$v_1=\frac{\Delta f}{\Delta I_t} \tag{18}$$

式中：Δf 为适应度的变化大小；ΔI_t 为迭代次数的变化大小。

步骤4 双重随机扰动策略。

1）第一重是柯西扰动[19]。一维标准柯西分布的概率密度函数表达式为

$$f(x)=\frac{1}{\pi}\left(\frac{1}{x^2+1}\right) \qquad -\infty<x<\infty \tag{19}$$

一维标准柯西分布与一维标准高斯分布概率密度曲线如图4所示。可知，柯西分布较高斯分布

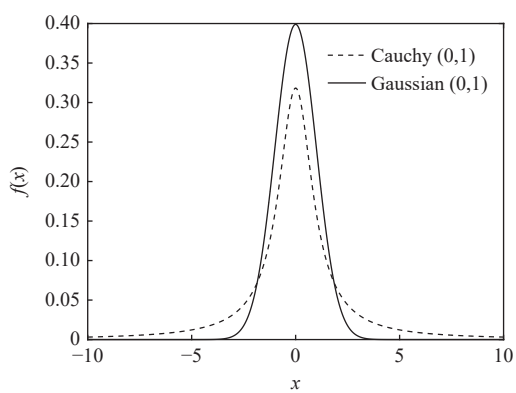


图4 一维标准柯西分布与高斯分布概率密度曲线

Fig. 4 Probability density curves of one-dimensional standard Cauchy distribution and Gaussian distribution

在原点处的峰值更小, 使得个体变异后使用较少的时间搜索相邻区域, 同时, 柯西分布相较高斯分布具有长尾特性, 使其有一定概率取到距离原点较远的随机数, 增加了全局搜索的可能性。这意味着经过柯西变异后的个体具有跳出局部最优值的能力, 其扰动方式如下:

$$G = X - K \cdot \text{Cauchy}(0,1) \tag{20}$$

式中: K 用来决定柯西算子扰动的作用大小; Cauchy(0,1) 为标准柯西分布。柯西分布随机变量生成函数为 $\eta = \tan((\xi - 0.5)\pi)$, ξ 为 [0,1] 上服从均匀分布的随机变量。

2) 第二重是中心位置扰动策略。先计算 N–j 个优质个体的中心位置, 再使用该位置对劣质个体进行扰动。中心位置 O 的计算如图 5 所示。图中: W 为个体位置的维度值。中心位置代表优质个体的平均水平, 该扰动策略促使陷入局部最优值的个体能从当前位置向优质个体的中心位置移动, 扰动方式如下:

$$G = X - r\,(X - O) \tag{21}$$

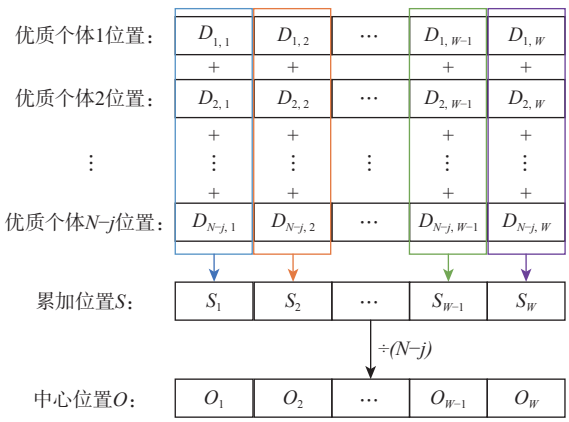


图 5　中心位置计算示意图

Fig. 5　Schematic diagram of central location calculation

步骤 5　对 j 个劣质个体进行分类。根据步骤 3 计算的适应度变化率将 j 个劣质个体分为 2 类: ①适应度变化率比优质个体平均适应度变化率低的 m 个个体; ②适应度变化率比优质个体平均适应度变化率高的 j–m 个个体。

步骤 6　对分类后的 j 个劣质个体进行双重随机扰动。第 1 类个体代表适应度达不到平均水平, 寻优速度也比优质个体平均寻优速度慢。对这类劣质个体进行第一重柯西扰动, 将第 1 类 j 个劣质个体按照适应度变化率进行排序, 适应度变化率小的个体较其他个体有更大的几率陷入局部最优值, 因此, 更需要大的步长对其进行扰动。同时, 生成 j 个柯西算子, 对其按照绝对值大小进行排序, 柯西算子绝对值越大, 代表扰动的步长越大。将适应度变化率较小的个体与绝对值较大的柯西算子(即大步长)一一对应, 利用式 (20) 对其进行扰动, 若扰动后适应度值变小, 则保留扰动后的位置, 否则, 根据式 (21) 启用第二重中心位置扰动。同样, 若扰动后适应度值变小, 则保留扰动后的位置, 否则, 不扰动。第 2 类个体代表适应度不及平均水平, 但寻优速度比优质个体的平均寻优速度快, 因此, 仅对其采用第一重柯西扰动, 若扰动后适应度值变小, 则保留扰动后的位置, 否则, 不扰动。

在个体完成位置更新后, 对个体进行双重随机扰动, 帮助个体减小陷入潜在局部最优的风险, 促使个体更好地向全局最优解的位置探索。

2.4　逐维更新

标准 GTO 算法在更新个体位置比较适应度时, 将种群个体的不同维度视为一个整体进行评估[20], 但某一维度上较好的结果可能会被忽略, 从而降低了算法的收敛速度和求解精度。因此, 本文引入一种逐维更新的策略, 用于更新大猩猩的位置, 该策略使得每一维的信息都得到关注。在大猩猩位置更新过程中, 每更新一个维度上的值, 就与其他维度的值组成一个完整的解, 再评价这个新的解, 直到每一维的值都被组合比较。

标准 GTO 算法位置更新方式与逐维更新位置方式如图 6 和图 7 所示。图中: D_{Xi}、D_{Gi} 分别为大猩猩在当前位置 X、候选位置 G 第 i 维度的位置信息, $f(X)$、$f(G)$ 表示适应度值。

通过引入逐维更新策略, 个体每一维的位置都得到关注, 促使每一维好的位置信息对寻优发挥一定的帮助, 也有效防止了每一维不好的位置信息对寻优结果产生坏的影响, 从而提高算法的寻优精度。

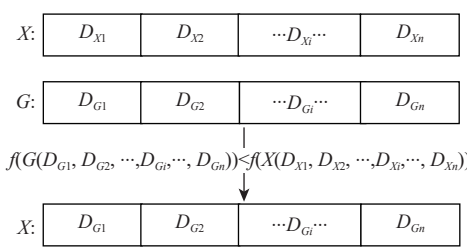


图 6　整体更新位置方式

Fig. 6　Entire update position way

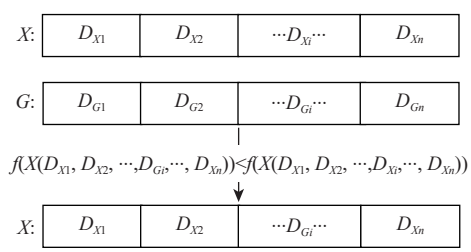


图 7　逐维更新位置方式

Fig. 7　Dimension-by-dimension update position way

2.5 算法实现流程

本文提出的DGTO算法步骤如下，其流程如图8所示。

步骤1 初始化相关参数：种群规模N、维度W、最大迭代次数$I_{t,\max}$、种群搜索范围u_b和l_b、探索阶段切换概率s、开发阶段切换概率w、参数β。

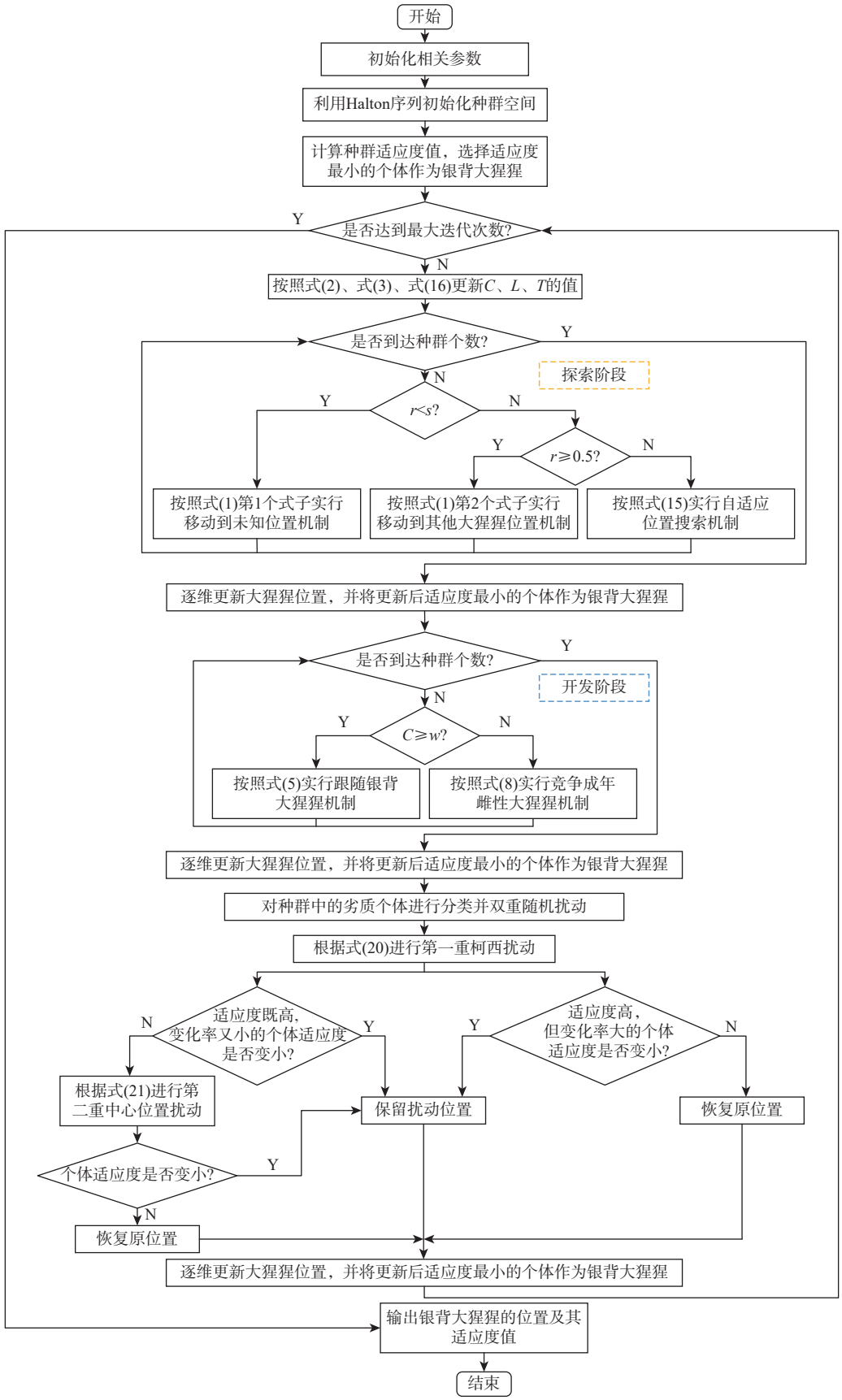


图8 本文算法流程

Fig. 8 Flowchart of the proposed algorithm

步骤 2　利用 Halton 序列初始化种群 X。

步骤 3　计算每个大猩猩个体 X 的适应度值, 并选择适应度最小的个体作为银背大猩猩 $X_{\text{silverback}}$。

步骤 4　按照式 (2)、式 (3)、式 (16) 计算参数 C、L、T 的值。

步骤 5　按照式 (1) 第 1 个式子、第 2 个式子及式 (15)、探索阶段切换概率 s 更新大猩猩候选位置 G, 并对更新后的候选位置 G 进行边界检查, 根据逐维更新策略更新大猩猩的位置 X, 选择适应度最小的大猩猩个体作为银背大猩猩 $X_{\text{silverback}}$。

步骤 6　按照式 (5)、式 (8) 及开发阶段切换概率 w 更新大猩猩候选位置 G, 并对更新后的位置 G 进行边界检查, 根据逐维更新策略更新大猩猩的位置 X, 选择适应度更小的大猩猩个体作为银背大猩猩 $X_{\text{silverback}}$。

步骤 7　按照双重随机扰动策略及式 (20)、式 (21) 对有可能陷入局部最优的个体进行位置扰动, 再根据逐维更新策略更新大猩猩的位置 X, 选择适应度更小的大猩猩个体作为银背大猩猩 $X_{\text{silverback}}$。

步骤 8　判断是否满足迭代终止条件, 满足, 则输出银背大猩猩位置 $X_{\text{silverback}}$ 及其适应度值, 否则, 进入步骤 4 继续执行。

3　实验仿真与结果分析

3.1　参数设置说明

基于 AMD Ryzen 75800H with Radeon Graphics CPU、3.20 GHz 主频、16 GB 内存及 Windows 11 (64 位) 的操作系统对 DGTO 算法进行仿真实验。编程软件为 MATLAB R2019a。选取标准 GTO 算法、灰狼优化算法 (grey wolf optimizer, GWO)[21]、差分进化 (differential evolution, DE) 算法[22]、鲸鱼优化算法 (whale optimization algorithm, WOA)[23] 与 DGTO 算法作对比, 基本参数统一设置为: 种群规模为 40, 最大迭代次数为 500, 维度为 10。各算法内部参数如表 2 所示。表中: F 为缩放因子; R_C 为交叉概率; b 为螺旋形状控制参数。

表 2　参数设置

Table 2　Parameters setting

算法	主要参数
GTO	s=0.03,w=0.8,β=3
GWO	
DE	F=0.5,R_C=0.3
WOA	b=1
DGTO	s=0.03,w=0.8,β=3

3.2　测试函数

为测试 DGTO 算法的寻优性能, 从基准测试函数中选取 10 个进行寻优测试, 其中, f_1～f_6 为单峰测试函数, f_7、f_8 为多峰测试函数, f_9、f_{10} 为固定维多峰测试函数。基准测试函数相关信息如表 3 所示。

表 3　基准测试函数

Table 3　Benchmark test functions

编号	函数名	定义域	维度	最优值
f_1	Sphere	[−100,100]	10/30	0
f_2	Schwefel' problem 2.22	[−10,10]	10/30	0
f_3	Schwefel' problem 1.2	[−100,100]	10/30	0
f_4	Schwefel' problem 2.21	[−100,100]	10/30	0
f_5	Generalized Rosenbrock's function	[−30,30]	10/30	0
f_6	Step function	[−100,100]	10/30	0
f_7	Generalized Schwefel's problem 2.26	[−500,500]	10/30	−418.98×维度
f_8	Generalized penalized function 2	[−50,50]	10/30	0
f_9	Kowalik's function	[−5,5]	4	0.000 3
f_{10}	Hatman's function 2	[0,1]	6	−3.32

3.3　DGTO 算法与其他改进 GTO 算法的性能对比及收敛性分析

为验证 DGTO 算法改进的有效性, 将其与标准 GTO 算法及基于镜像对立和自适应-β爬山算法改进的 GTO 算法 (improved gorilla troops optimizer based on lens opposition-based learning and adaptive β-hill climbing, IGTO)、用于全局优化问题的改进 GTO 算法 (modified gorilla troops optimizer for global optimization problem, MGTO) 在表 3 所示的基准测试函数上进行测试对比, IGTO 和 MGTO 的数据分别来自文献 [9] 和文献 [11]。为保证实验的可比性, 实验统一设置为: 种群规模为 30, 维度为 30, 最大迭代次数为 500。各算法分别运行 30 次取平均值和标准差, 对比结果如表 4 所示, 加粗字体表示

表4 DGTO算法与其他改进GTO算法寻优结果对比

Table 4 Comparison of optimization results of DGTO and other improved GTO algorithms

参数	算法	平均值	标准差
f_1	GTO	0	0
	IGTO	0	0
	MGTO	0	0
	DGTO	**0**	**0**
f_2	GTO	6.67×10^{-195}	0
	IGTO	0	0
	MGTO	0	0
	DGTO	**0**	**0**
f_3	GTO	0	0
	IGTO	0	0
	MGTO	0	0
	DGTO	**0**	**0**
f_4	GTO	1.01×10^{-192}	0
	IGTO	0	0
	MGTO	0	0
	DGTO	**0**	**0**
f_5	GTO	3.17×10^{0}	8.22×10^{0}
	IGTO	7.46×10^{-5}	8.57×10^{-5}
	MGTO	2.54×10^{-5}	3.94×10^{-5}
	DGTO	$\mathbf{7.42\times10^{-6}}$	$\mathbf{3.39\times10^{-5}}$
f_6	GTO	2.56×10^{-7}	4.19×10^{-7}
	IGTO	1.01×10^{-7}	1.15×10^{-7}
	MGTO	3.37×10^{-14}	3.76×10^{-14}
	DGTO	$\mathbf{2.03\times10^{-32}}$	$\mathbf{1.66\times10^{-32}}$
f_7	GTO	-1.26×10^{4}	3.05×10^{-5}
	IGTO	-1.26×10^{4}	3.02×10^{-5}
	MGTO	-1.26×10^{4}	1.03×10^{-6}
	DGTO	$\mathbf{-1.26\times10^{4}}$	$\mathbf{4.64\times10^{-12}}$
f_8	GTO	3.30×10^{-3}	5.12×10^{-3}
	IGTO	1.14×10^{-7}	3.34×10^{-7}
	MGTO	1.00×10^{-7}	1.43×10^{-7}
	DGTO	$\mathbf{7.24\times10^{-32}}$	$\mathbf{2.30\times10^{-31}}$
f_9	GTO	4.17×10^{-4}	3.01×10^{-4}
	IGTO	3.07×10^{-4}	4.06×10^{-19}
	MGTO	3.07×10^{-4}	3.29×10^{-19}
	DGTO	$\mathbf{3.07\times10^{-4}}$	$\mathbf{1.74\times10^{-19}}$
f_{10}	GTO	-3.29×10^{0}	5.39×10^{-2}
	IGTO	-3.31×10^{0}	3.63×10^{-2}
	MGTO	-3.30×10^{0}	4.51×10^{-2}
	DGTO	$\mathbf{-3.32\times10^{0}}$	**0**

在某一函数下各算法的最优结果。图9为各算法的部分收敛曲线。

由表4和图9可知，在10个基准测试函数中，DGTO算法较其他算法具有显著优势，其平均值、标准差大多具有量级优势。对于函数f_1～f_4，MGTO、IGTO算法可以找到理论最优值，稳定性良好，而DGTO算法通过不同的方式改进也达到了相同的效果，说明了DGTO算法的有效性，为算法改进提供了新的思路；对于函数f_5和f_{10}，DGTO算法的寻优结果最好，也最为稳定，f_5的收敛速度与其余算法相比较为明显，f_{10}甚至达到了标准差为0的稳定性；对于函数f_6，GTO、IGTO算法的收敛精度基本相同，可推断其陷入局部最优值而无法自行跳出，而MGTO算法收敛精度略高，但DGTO算法在双重随机扰动策略的帮助下，具备了更强的跳出局部最优值的能力，达到了更高的收敛精度及更强的稳定性，从收敛曲线也可看出，在收敛次数约为50次时，DGTO算法已经达到了其余算法的精度值，在其余算法逐渐收敛后，DGTO算法依旧探索更好的结果；对于函数f_8，IGTO、MGTO算法虽比GTO算法平均值更低，但DGTO算法还是达到了约10^{-32}的精度和标准差，差距较为明显；对于函数f_7，各算法虽然都可寻到最优值附近，但DGTO算法的性能更为稳定；对于函数f_9，改进后的GTO算法比标准GTO算法更靠近理论最优值，而DGTO算法的收敛曲线下降速率更大，说明其找到理论最优值所需的迭代次数更少，且更稳定。综上所述，DGTO算法的优势较为明显。

3.4 DGTO算法与其他算法的性能对比

为验证DGTO算法对于基准测试函数的寻优性能，选取标准GTO算法、GWO算法、DE算法、WOA算法进行寻优对比。为确保实验结果具有可比性，各算法实验统一设置为：种群规模为40，维度为10，最大迭代次数为500，各算法相关参数由表2给出，利用表3的10个基准测试函数对各算法进行寻优性能对比，各算法分别运行50次取最优值、最差值、平均值和标准差，对比结果如表5所示，加粗字体表示在某一函数下各算法的最优结果。

由表5可知，在求解函数f_1～f_{10}时，DGTO算法寻优结果及寻优稳定性均优于其他算法。在多个函数中，DGTO算法的最差值都优于其余算法的最优值。对于单峰函数f_1～f_4，DGTO算法每次都可以直接收敛到理论最优值0；对于单峰函数f_5，DGTO算法较其他算法更接近理论最优值；对于单峰函数f_6，DGTO算法最优可以收敛到理论最优值0；对于多峰函数f_7和固定维多峰函数f_9、f_{10}，DGTO算法更接近理论最优值−4 189.8、0.000 3和−3.32；对于多峰函数f_8，DGTO算法可以收敛到理论最优值附近。综上所述，DGTO算法较其他算法在对基准测试函数进行求解时具有明显优势。

3.5 DGTO算法与其他算法的收敛性对比

为验证DGTO算法的收敛性，图10给出了DGTO算法、标准GTO算法、GWO算法、DE算

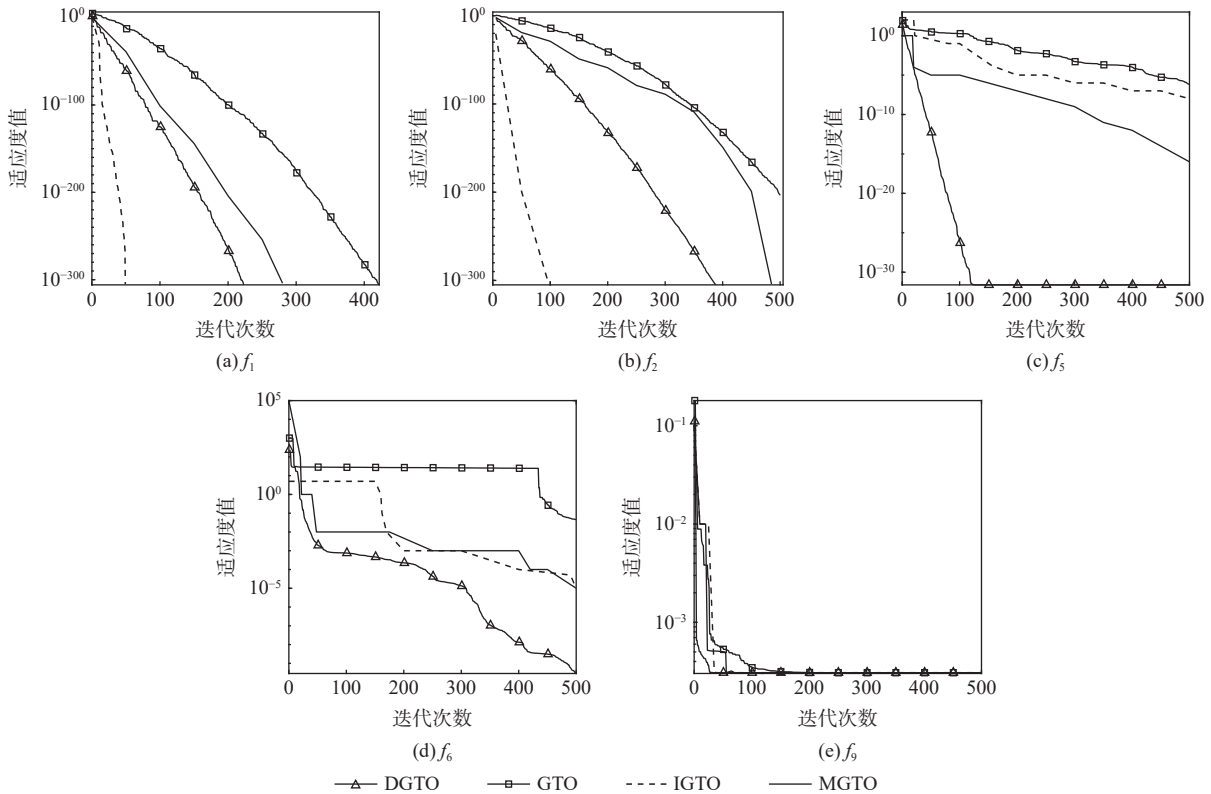


图 9　DGTO 算法与其他改进 GTO 算法的部分收敛曲线

Fig. 9　Partial convergence plots of DGTO and other improved GTO algorithms

法、WOA 算法在 10 个基本测试函数上的收敛曲线。各算法实验仍统一设置为: 种群规模为 40, 维度为 10, 最大迭代次数为 500。

由图 10 可知, 对于单峰函数 f_1～f_6 和多峰函数 f_7～f_{10}, 横向观察收敛曲线, DGTO 算法收敛曲线在整个寻优过程中均位于其余 4 个算法左方, 说明在初始迭代时, Halton 序列增加了种群的多样性, 随后自适应位置搜索机制合理地平衡了开发和探索之间的关系, 使得 DGTO 算法较其余 4 个算法有更好的收敛速度; 纵向观察收敛曲线, DGTO 收敛曲线在整个寻优过程中均位于其余 4 个算法下方, 说明双重随机扰动策略帮助算法跳出局部最优值, 最终逐维更新每一维度的信息进行比较更新, 使得 DGTO 算法较其他 4 个算法有更好的寻优精度。对于单峰函数 f_1、f_3, DGTO 算法分别仅需 208 次、252 次迭代便寻到理论最优值, 约为 GTO 算法寻优速度的 2 倍, 而 GWO、DE、WOA 算法经过 500 次迭代仍处于较低的寻优精度。对于函数 f_2、f_4, f_6, DGTO 算法分别在 380 次、398 次、102 次迭代找到了理论最优值, 而其余 4 个算法仍与理论最优值差距较大。对于单峰函数 f_5, DGTO 算法时而平缓时而下降, 说明算法在陷入局部最优时, 双重随机扰动策略可以帮助其跳出局部最优, 继而向全局最优解靠近。综上所述, DGTO 算法在对基准测试函数寻优时具有更快的收敛速度和更高的收敛精度, 说明了算法的有效性。

3.6　DGTO 算法与其他算法的平均绝对值误差比较

为评估 DGTO 算法与其他算法在求解 10 个基准测试函数时的寻优结果与理论最优值的差异情况, 使用平均绝对值误差[24] 对其进行排序比较。平均绝对值误差可以避免误差相互抵消的问题, 因而可以准确反映实际预测误差的大小。各算法实验统一设置为: 种群规模为 40, 维度为 10, 最大迭代次数为 500, 独立运行 50 次。平均绝对值误差的数学模型如式 (22) 所示, 各算法对 10 个基准函数的平均绝对值误差排序结果如表 6 所示。

$$Y=\frac{1}{n_1}\sum_{i=1}^{n_1}|\mu_i-\sigma_i| \tag{22}$$

式中: n_1 为选用基准函数的个数; μ_i 为算法运行 50 次寻得最优值的平均值; σ_i 为每个基准测试的理论最优值。

由表 6 可知, DGTO 算法的平均绝对值误差最小, 说明其寻优结果更接近函数的理论最优值, 收敛精度较其他算法更高, 进一步证明了算法的有效性, 较其他算法更有优势。

3.7　Wilcoxon 秩和检验

为进一步验证 DGTO 算法的有效性, 对其与其

表5 各算法寻优结果对比

Table 5 Comparison of optimization results of each algorithm

函数	算法	最优值	最差值	平均值	标准差
f_1	DGTO	**0**	**0**	**0**	**0**
	GTO	0	0	0	0
	GWO	2.62×10^{-69}	6.96×10^{-63}	3.93×10^{-64}	1.28×10^{-63}
	DE	6.09×10^{-20}	1.86×10^{-18}	5.05×10^{-19}	4.44×10^{-19}
	WOA	2.04×10^{-94}	5.80×10^{-81}	1.23×10^{-82}	8.19×10^{-82}
f_2	DGTO	**0**	**0**	**0**	**0**
	GTO	7.39×10^{-217}	3.67×10^{-197}	7.69×10^{-199}	0
	GWO	3.25×10^{-39}	8.12×10^{-36}	5.90×10^{-37}	1.30×10^{-36}
	DE	1.91×10^{-12}	2.93×10^{-11}	9.68×10^{-12}	4.55×10^{-12}
	WOA	1.10×10^{-61}	5.26×10^{-52}	1.12×10^{-53}	7.43×10^{-53}
f_3	DGTO	**0**	**0**	**0**	**0**
	GTO	0	0	0	0
	GWO	3.82×10^{-37}	3.02×10^{-27}	8.81×10^{-29}	4.33×10^{-28}
	DE	2.63×10^{0}	2.74×10^{1}	8.17×10^{0}	5.41×10^{0}
	WOA	2.73×10^{-8}	6.89×10^{2}	1.16×10^{2}	1.46×10^{2}
f_4	DGTO	**0**	**0**	**0**	**0**
	GTO	2.01×10^{-218}	4.03×10^{-200}	1.18×10^{-201}	0
	GWO	9.71×10^{-24}	2.86×10^{-19}	1.79×10^{-20}	4.48×10^{-20}
	DE	2.60×10^{-5}	1.31×10^{-4}	6.72×10^{-5}	2.38×10^{-5}
	WOA	2.21×10^{-5}	2.20×10^{1}	1.35×10^{0}	4.37×10^{0}
f_5	DGTO	**7.46×10^{-30}**	**2.25×10^{-11}**	**6.34×10^{-13}**	**3.28×10^{-12}**
	GTO	4.91×10^{-15}	1.41×10^{0}	1.07×10^{-1}	3.06×10^{-1}
	GWO	5.32×10^{0}	9.54×10^{0}	6.70×10^{0}	7.69×10^{-1}
	DE	1.74×10^{0}	1.75×10^{1}	7.58×10^{0}	2.78×10^{0}
	WOA	3.49×10^{0}	8.95×10^{0}	6.58×10^{0}	6.38×10^{-1}
f_6	DGTO	**0**	**1.23×10^{-32}**	**5.55×10^{-34}**	**1.94×10^{-33}**
	GTO	9.24×10^{-32}	9.98×10^{-23}	2.18×10^{-24}	1.14×10^{-23}
	GWO	1.11×10^{-6}	6.33×10^{-6}	2.97×10^{-6}	1.13×10^{-6}
	DE	5.02×10^{-20}	2.93×10^{-18}	4.00×10^{-19}	4.42×10^{-19}
	WOA	2.62×10^{-5}	1.17×10^{-3}	3.12×10^{-4}	2.70×10^{-4}
f_7	DGTO	**-4.19×10^{3}**	**-4.19×10^{3}**	**-4.19×10^{3}**	**2.13×10^{-12}**
	GTO	-4.19×10^{3}	-4.19×10^{3}	-4.19×10^{3}	5.29×10^{-12}
	GWO	-3.56×10^{3}	-2.08×10^{3}	-2.84×10^{3}	3.65×10^{2}
	DE	-7.94×10^{7}	-8.23×10^{14}	-2.09×10^{13}	1.20×10^{14}
	WOA	-4.19×10^{3}	-2.49×10^{3}	-3.46×10^{3}	5.61×10^{2}
f_8	DGTO	**1.35×10^{-32}**	**1.35×10^{-32}**	**1.35×10^{-32}**	**1.11×10^{-47}**
	GTO	7.26×10^{-30}	7.46×10^{-2}	5.34×10^{-3}	1.25×10^{-2}
	GWO	1.15×10^{-6}	1.04×10^{-1}	1.00×10^{-2}	3.04×10^{-2}
	DE	4.37×10^{-21}	4.46×10^{-19}	8.94×10^{-20}	8.94×10^{-20}
	WOA	2.57×10^{-4}	1.10×10^{-1}	1.44×10^{-2}	2.57×10^{-2}
f_9	DGTO	**3.07×10^{-4}**	**3.07×10^{-4}**	**3.07×10^{-4}**	**1.74×10^{-19}**
	GTO	3.07×10^{-4}	1.22×10^{-3}	4.17×10^{-4}	3.01×10^{-4}
	GWO	3.07×10^{-4}	2.04×10^{-2}	2.78×10^{-3}	6.56×10^{-3}
	DE	5.16×10^{-4}	1.30×10^{-3}	8.97×10^{-4}	1.48×10^{-4}
	WOA	3.08×10^{-4}	2.17×10^{-3}	6.93×10^{-4}	4.89×10^{-4}
f_{10}	DGTO	**-3.32×10^{0}**	**-3.32×10^{0}**	**-3.32×10^{0}**	**0**
	GTO	-3.32×10^{0}	-3.20×10^{0}	-3.29×10^{0}	5.39×10^{-2}
	GWO	-3.32×10^{0}	-3.09×10^{0}	-3.27×10^{0}	7.57×10^{-2}
	DE	-3.32×10^{0}	-3.20×10^{0}	-3.31×10^{0}	3.17×10^{-2}
	WOA	-3.32×10^{0}	-2.43×10^{0}	-3.22×10^{0}	1.45×10^{-1}

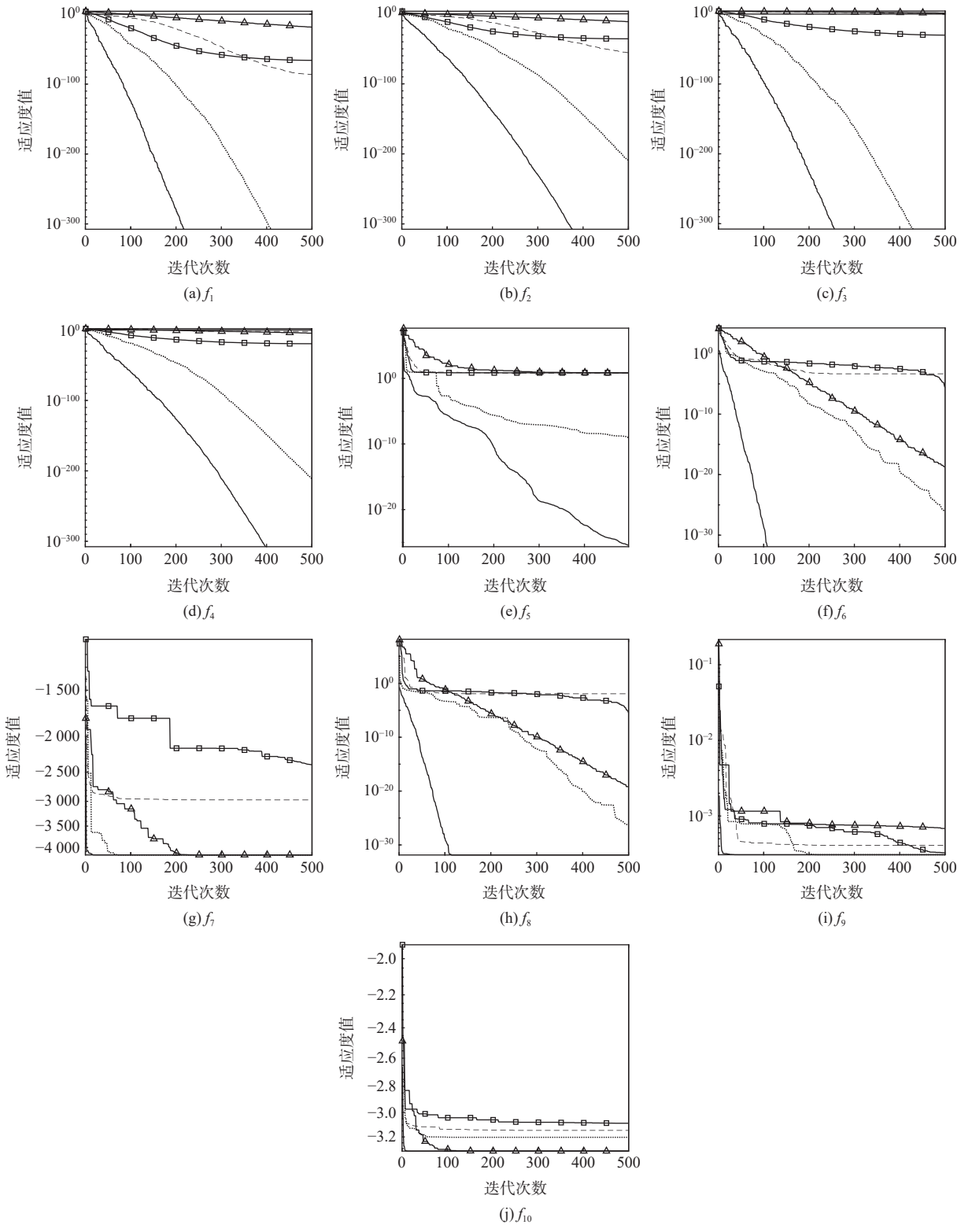


图 10　各算法在函数 f_1~f_{10} 的收敛曲线

Fig. 10　Convergence curves of each algorithm in functions f_1–f_{10}

他算法进行性能差异分析。对于 2 个独立的样本，常用的假设检验方法有参数检验和非参数检验。常见的参数假设检验方式有 t 检验和 z 检验，参数检验方法对数据要求较为严格，通常是在总体分布形式已知的情况下进行检验。对于非参数检验方式，符号检验和 Wilcoxon 秩和检验较为常见，这些方式在 2 个独立样本假设检验时无需知道样本的分布情况。而 Wilcoxon 秩和检验是在传统的符号检验方式上发展起来的，不仅可以直观显示 2 个独立样本是否有差异性，还可以通过 p 值判断出差异

性的程度，比传统的单独用正负号的检验更加有效。在对DGTO算法与其他算法进行性能差别分析前，样本分布情况未知，因此，在显著性水平为5%下，采用Wilcoxon秩和检验对DGTO算法与其余4个算法50次运行结果进行分析是否有显著的性能差异，并计算其p值。当p值小于5×10^{-2}时，否定对比算法与DGTO算法位置分布相似的假设[25]，即对比算法与DGTO算法的寻优性能具有明显差异。Wilcoxon秩和检验结果如表7所示。

表6 不同算法的平均绝对值误差

Table 6 Mean absolute error of different algorithms

算法	平均绝对值误差
DGTO	3.09×10^{-3}
GTO	1.73×10^{-2}
GWO	1.36×10^{2}
DE	2.09×10^{12}
WOA	8.54×10^{1}

由表7可知，DGTO算法与其他算法相比，大多数p值都小于5×10^{-2}，仅在函数f_1和f_3中，DGTO算法与GTO算法之间没有显著差异。因此，总体来说，与其他算法相比，DGTO算法的寻优性能被认为具有显著的优势。

表7 Wilcoxon秩和检验结果

Table 7 Wilcoxon rank sum test results

函数	p值			
	GTO	GWO	DE	WOA
f_1	NAN	3.31×10^{-20}	3.31×10^{-20}	3.31×10^{-20}
f_2	3.31×10^{-20}	3.31×10^{-20}	3.31×10^{-20}	3.31×10^{-20}
f_3	NAN	3.31×10^{-20}	3.31×10^{-20}	3.31×10^{-20}
f_4	3.31×10^{-20}	3.31×10^{-20}	3.31×10^{-20}	3.31×10^{-20}
f_5	7.97×10^{-18}	7.07×10^{-18}	7.07×10^{-18}	7.07×10^{-18}
f_6	6.35×10^{-19}	6.35×10^{-19}	6.35×10^{-19}	6.35×10^{-19}
f_7	4.32×10^{-12}	3.43×10^{-19}	3.43×10^{-19}	3.43×10^{-19}
f_8	1.30×10^{-18}	1.30×10^{-18}	1.30×10^{-18}	1.30×10^{-18}
f_9	2.79×10^{-16}	5.92×10^{-18}	5.92×10^{-18}	5.92×10^{-18}
f_{10}	2.71×10^{-7}	4.73×10^{-20}	6.69×10^{-11}	4.73×10^{-20}

注：参数p小于5×10^{-2}表示DGTO与对比算法有显著差异，NAN表示无法进行显著性判断。

4 工程优化设计问题

为验证DGTO算法改进的有效性，将其用于优化实际的拉伸/压缩弹簧设计问题。

4.1 拉伸/压缩弹簧设计问题

拉伸/压缩弹簧设计问题的目标是最小化弹簧拉伸/压缩的质量，其问题模型如图11所示。该优化设计问题受剪应力、振动频率和最小挠度约束条件的限制，包含3个决策变量，分别为弹簧线圈直径$d(x_1)$、平均线圈直径$D_a(x_2)$和有效绕圈数$P(x_3)$。

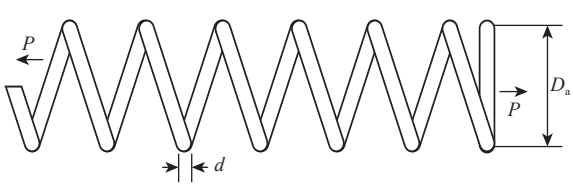


图11 弹簧设计模型

Fig. 11 Mode of spring design

拉伸/压缩弹簧设计问题的数学模型描述如下所示。

目标函数：

$$\min\ f(x)=(x_3+2)x_1^2x_2 \tag{23}$$

约束条件：

$$\begin{cases} g_1(x)=1-\dfrac{x_2^3x_3}{71\,785x_1^4}\leqslant 0\\ g_2(x)=\dfrac{4x_2^2-x_1x_2}{12\,566(x_1^3x_2-x_1^4)}+\dfrac{1}{5\,108x_1^2}-1\leqslant 0\\ g_3(x)=1-\dfrac{140.45x_1}{x_2^2x_3}\leqslant 0\\ g_4(x)=\dfrac{x_1+x_2}{1.5}-1\leqslant 0 \end{cases} \tag{24}$$

式中：$0.05\leqslant x_1\leqslant 2;0.25\leqslant x_2\leqslant 1.3;2\leqslant x_3\leqslant 15$。

4.2 实验结果及分析

为评估DGTO算法的性能，将其与标准GTO算法、GWO算法、DE算法、WOA算法、斑点鬣狗优化算法(spotted hyena optimizer, SHO)[26]、多元宇宙优化(multi-verse optimizer, MVO)算法[27]、正弦余弦算法(sine cosine algorithm, SCA)[28]和引力搜索算法(gravitational search algorithm, GSA)[29]进行实验对比，其中，SHO算法、MVO算法、SCA算法和GSA算法的实验数据来自文献[30]。各算法独立运行50次取最优值。各算法对拉伸/压缩弹簧设计问题的优化结果如表8所示。表中：加粗数据表示最优解。

表 8　弹簧设计问题各算法最优解

Table 8　Optimal solution of each algorithm for spring design problem

算法	d	D_a	P	最优解
DGTO	**0.051 72**	**0.357 50**	**11.242 94**	**0.012 665 252**
GTO	0.051 62	0.355 04	11.387 75	0.012 665 321
GWO	0.051 20	0.345 07	12.008 99	0.012 672 677
DE	0.051 61	0.354 79	11.403 14	0.012 665 876
WOA	0.051 48	0.351 82	11.582 31	0.012 666 117
SHO[30]	0.051 14	0.343 75	12.095 50	0.012 674 000
MVO[30]	0.050 00	0.315 96	14.226 23	0.012 816 930
SCA[30]	0.050 78	0.334 78	12.722 69	0.012 709 667
GSA[30]	0.050 00	0.317 31	14.228 67	0.012 873 881

由表 8 可知, DGTO 算法获得的使目标函数最小化的最优解为 $[x_1,x_2,x_3]$=[0.051 72,0.357 50,11.242 94], 最优值为 0.012 665 252。通过 DGTO 算法优化后的拉伸/压缩弹簧设计问题最优解优于其他算法最优解, 由此可说明 DGTO 算法在解决实际工程优化设计问题时具有一定的有效性, 进一步体现了算法的鲁棒性。

5　结　论

为改进标准 GTO 算法的寻优性能, 本文提出了一种基于双重随机扰动策略的人工大猩猩部队优化算法。

1) 在初始化阶段, 引入 Halton 序列对种群进行初始化, 增加了种群的多样性, 使得种群分布更加均匀。

2) 在探索阶段, 提出自适应位置搜索机制, 该机制有效提高了种群的收敛速度及收敛精度。

3) 为帮助算法跳出局部最优值, 提出双重随机扰动策略对种群位置进行扰动, 一重为柯西扰动, 另一重为中心位置扰动。

4) 在种群个体位置更新阶段, 采用逐维更新的方式对种群位置进行更新, 充分利用了种群个体每一维度的信息, 提高了算法的收敛精度。

5) 通过 10 个基准测试函数, 验证了本文算法改进的有效性, 并通过实际工程优化设计问题(即拉伸/压缩弹簧设计问题)证明了算法的鲁棒性, 为解决复杂的工程优化设计问题提供了新的思路。

在今后的工作中, 将会更加深入研究新的改进策略, 进一步提高标准 GTO 算法的性能。

参考文献（References）

[1] EBERHART R, KENNEDY J. A new optimizer using particle swarm theory[C]//Proceedings of the 6th International Symposium on Micro Machine and Human Science. Piscataway: IEEE Press, 2002: 39-43.

[2] CHEN D, ZHANG S, YANG Y, et al. Optimization of character image matching based on artificial bee colony algorithm[J]. Journal of Physics: Conference Series, 2021, 2035(1): 012034.

[3] MORIN M, ABI-ZEID I, QUIMPER C G. Ant colony optimization for path planning in search and rescue operations[J]. European Journal of Operational Research, 2023, 305(1): 53-63.

[4] AL-IBRAHIM A M H. Solving travelling salesman problem (TSP) by hybrid genetic algorithm (HGA)[J]. International Journal of Advanced Computer Science and Applications, 2020, 11(6): 376-384.

[5] LI J H, LEI Y S, YANG S H. Mid-long term load forecasting model based on support vector machine optimized by improved sparrow search algorithm[J]. Energy Reports, 2022, 8: 491-497.

[6] YAN Z P, ZHANG J Z, ZENG J, et al. Three-dimensional path planning for autonomous underwater vehicles based on a whale optimization algorithm[J]. Ocean Engineering, 2022, 250: 111070.

[7] GUHA D, ROY P K, BANERJEE S. Load frequency control of interconnected power system using grey wolf optimization[J]. Swarm and Evolutionary Computation, 2016, 27: 97-115.

[8] ABDOLLAHZADEH B, GHAREHCHOPOGH F S, MIRJALILI S. Artificial gorilla troops optimizer: a new nature-inspired metaheuristic algorithm for global optimization problems[J]. International Journal of Intelligent Systems, 2021, 36(10): 5887-5958.

[9] XIAO Y N, SUN X, GUO Y L, et al. An improved gorilla troops optimizer based on lens opposition-based learning and adaptive β-hill climbing for global optimization[J]. Computer Modeling in Engineering & Sciences, 2022, 131(2): 815-850.

[10] LIANG Q W, CHU S C, YANG Q Y, et al. Multi-group gorilla troops optimizer with multi-strategies for 3D node localization of wireless sensor networks[J]. Sensors, 2022, 22(11): 4275.

[11] WU T Y, WU D, JIA H M, et al. A modified gorilla troops optimizer for global optimization problem[J]. Applied Sciences, 2022, 12(19): 10144.

[12] ALSOLAI H, ALZAHRANI J S, MARAY M, et al. Enhanced artificial gorilla troops optimizer based clustering protocol for UAV-assisted intelligent vehicular network[J]. Drones, 2022, 6(11): 358.

[13] MOSTAFA R R, GAHEEN M A, ABD ELAZIZ M, et al. An improved gorilla troops optimizer for global optimization problems and feature selection[J]. Knowledge-Based Systems, 2023, 269: 110462.

[14] BANGYAL W H, TAYYAB H, BATOOL H, et al. An improved particle swarm optimization algorithm with Chi-square mutation strategy[J]. International Journal of Advanced Computer Science and Applications, 2019, 10(3): 481-491.

[15] 宋立钦, 陈文杰, 陈伟海, 等. 基于混合策略的麻雀搜索算法改进及应用[J]. 北京航空航天大学学报, 2023, 49(8): 2187-2199.
SONG L Q, CHEN W J, CHEN W H, et al. Improvement and application of hybrid strategy-based sparrow search algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(8): 2187-2199(in Chinese).

[16] 周理, 朱红求. 基于自适应步长果蝇算法的爬行机器人足端轨迹规划[J]. 机械设计与研究, 2021, 37(3): 60-63.
ZHOU L, ZHU H Q. Foot trajectory planning of creeping robot based on adaptive step fruit fly optimization algorithm[J]. Machine Design & Research, 2021, 37(3): 60-63(in Chinese).

[17] 宋阿妮, 包贤哲, 权轶. 基于混沌自适应萤火虫算法的 UAVs 分配策略[J]. 计算机应用与软件, 2022, 39(2): 300-306.
SONG A N, BAO X Z, QUAN Y. Uavs scheduling strategy based on chaotic adaptive firefly algorithm[J]. Computer Applications and

Software, 2022, 39(2): 300-306(in Chinese).

[18] 李凡长, 刘洋, 吴鹏翔, 等. 元学习研究综述[J]. 计算机学报, 2021, 44(2): 422-446.
LI F C, LIU Y, WU P X, et al. A survey on recent advances in meta-learning[J]. Chinese Journal of Computers, 2021, 44(2): 422-446 (in Chinese).

[19] LI K W, LI S H, HUANG Z C, et al. Grey wolf optimization algorithm based on Cauchy-Gaussian mutation and improved search strategy[J]. Scientific Reports, 2022, 12: 18961.

[20] 刘薇, 赵剑锟, 刘义保, 等. 基于改进型灰狼算法的γ能谱解析应用研究[J]. 核技术, 2021, 44(4): 31-36.
LIU W, ZHAO J K, LIU Y B, et al. Application research of γ energy spectrum analysis based on improved grey wolf algorithm[J]. Nuclear Techniques, 2021, 44(4): 31-36(in Chinese).

[21] KOHLI M, ARORA S. Chaotic grey wolf optimization algorithm for constrained optimization problems[J]. Journal of Computational Design and Engineering, 2018, 5(4): 458-472.

[22] STORN R, PRICE K. Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces[J]. Journal of Global Optimization, 1997, 11(4): 341-359.

[23] MIRJALILI S, LEWIS A. The whale optimization algorithm[J]. Advances in Engineering Software, 2016, 95: 51-67.

[24] NABIL E. A modified flower pollination algorithm for global optimization[J]. Expert Systems with Applications, 2016, 57: 192-203.

[25] 张新明, 王霞, 康强. 改进的灰狼优化算法及其高维函数和FCM优化[J]. 控制与决策, 2019, 34(10): 2073-2084.
ZHANG X M, WANG X, KANG Q. Improved grey wolf optimizer and its application to high-dimensional function and FCM optimization[J]. Control and Decision, 2019, 34(10): 2073-2084(in Chinese).

[26] LUO Q F, LI J, ZHOU Y Q, et al. Using spotted hyena optimizer for training feedforward neural networks[J]. Cognitive Systems Research, 2021, 65: 1-16.

[27] ABUALIGAH L. Multi-verse optimizer algorithm: a comprehensive survey of its results, variants, and applications[J]. Neural Computing and Applications, 2020, 32(16): 12381-12401.

[28] MIRJALILI S. SCA: a sine cosine algorithm for solving optimization problems[J]. Knowledge-Based Systems, 2016, 96: 120-133.

[29] YAZDANI S, NEZAMABADI-POUR H, KAMYAB S. A gravitational search algorithm for multimodal optimization[J]. Swarm and Evolutionary Computation, 2014, 14: 1-14.

[30] KAUR S, AWASTHI L K, SANGAL A L, et al. Tunicate swarm algorithm: a new bio-inspired based metaheuristic paradigm for global optimization[J]. Engineering Applications of Artificial Intelligence, 2020, 90: 103541.

Artificial gorilla troops optimizer based on double random disturbance and its application of engineering problem

DU Xiaoxin[1, 2, *], HAO Tianru[1], WANG Bo[1, 2], WANG Zhenfei[1], ZHANG Jianfei[1, 2], JIN Mei[1, 2]

(1. College of Computer and Control Engineering, Qiqihar University, Qiqihar 161006, China;
2. Heilongjiang Key Laboratory of Big Data Network Security Detection and Analysis, Qiqihar University, Qiqihar 161006, China)

Abstract: Traditional artificial gorilla troops optimizer (GTO) has the drawbacks of easily falling into local optimum, slow convergence speed, and low optimization accuracy. Aiming at these problems, an artificial gorilla troops optimizer based on a double random disturbance strategy (DGTO) was proposed. Firstly, the Halton sequence was introduced to initialize the population to increase the diversity of the population. Secondly, the method's convergence speed was increased by using the multi-dimensional random number technique during the algorithm optimization stage and proposing an adaptive position exploration mechanism. Thirdly, a double random disturbance strategy was proposed, which solved the group effect of gorillas and enhanced the ability of the algorithm to jump out of the local optimum. Finally, the individual position was updated by a dimension-by-dimension update strategy, which improved the convergence accuracy of the algorithm. It is evident that the enhanced technique has a greater improvement in optimization accuracy and convergence speed when comparing the Wilcoxon rank sum test results with the optimization results of ten benchmark test functions. In addition, through the experimental comparative analysis of one practical engineering optimization problem, the superiority of the proposed algorithm in dealing with practical engineering problems is further verified.

Keywords: artificial gorilla troops optimizer; Halton sequence; adaptive position exploration; double random disturbance strategy; dimension-by-dimension update

Received: 2023-06-25; **Accepted**: 2023-09-11; **Published Online**: 2023-10-20 11:49

URL: link.cnki.net/urlid/11.2625.V.20231020.0903.001

Foundation item: Heilongjiang Provincial Higher Education Institutions Basic Scientific Research Business Funds Natural Science Young Innovative Talents Program (145209206)

* **Corresponding author**. E-mail: xiaoxindu@qqhru.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0518

基于高增益扰动观测器的磁悬浮球系统滑模控制

林俊亭*，陈信州
(兰州交通大学 自动化与电气工程学院，兰州 730070)

摘 要：针对磁悬浮球系统存在建模误差与未知扰动导致控制性能下降的问题，设计了一种基于高增益扰动观测器(HGDO)的自适应非奇异终端滑模控制(ANTSMC)方法。建立磁悬浮球系统模型，并在平衡点处将模型线性化；为削弱滑模控制器的抖振并保证系统误差在有限时间收敛，设计了自适应非奇异终端滑模控制器，同时，采用高增益扰动观测器对系统中存在的总扰动进行估计，并通过理论验证了所设计高增益扰动观测器可以快速收敛到实际扰动值的可调邻域，基于扰动估计值设计带扰动补偿的自适应非奇异终端滑模控制律，并证明了在该控制律下系统是全局一致最终有界的；通过仿真验证了所设计方法在不同目标轨迹下的有效性。仿真与定量分析表明：与带广义比例积分观测器的控制器相比，带高增益扰动观测器的控制器对总扰动观测值的积分时间乘方误差(ITSE)降低了75%，积分时间绝对误差(ITAE)降低了60%，从而提高了在相同控制方法下系统的鲁棒性。

关 键 词：磁悬浮球系统；滑模控制；高增益扰动观测器；扰动补偿；轨迹跟踪
中图分类号：TP273.3
文献标志码：A **文章编号**：1001-5965(2025)06-1897-10

磁悬浮技术已在城市轨道交通[1]、人工肾脏泵[2]、同步电动机[3]等各个领域中发挥着越来越重要的作用。以噪声小、无摩擦阻力为特点的磁浮列车是磁悬浮技术在轨道交通领域中使用的代表，其由走行机构、车体、牵引系统、悬浮控制系统、制动系统等组成[4]。其中，悬浮控制系统是磁悬浮技术在轨道交通领域中使用的关键，但现行的磁悬浮系统的模型却不统一，大多数模型为基于假设铁圈磁阻不存在的情况下所构建的二阶模型[5-6]，少部分为三阶电流模型[7]。磁悬浮系统是一种典型的非线性系统，通常是开环不确定的，如何设计一个综合性能良好的悬浮控制系统是磁悬浮技术的主要问题。研究人员针对该问题提出许多相关的控制方法，如PID控制[8]、非线性模型预测控制[9]、鲁棒控制[10]、滑模控制(sliding mode control, SMC)[11]、分数阶控制[12]等，不同的控制方法对于被控对象的性能提升各不相同。

滑模控制具有无超调、鲁棒性强、动态响应快的优点，能够提高磁悬浮系统的鲁棒性与承载能力[13-14]。目前，基于滑模控制的磁悬浮系统悬浮控制策略研究是该领域的热点之一，文献[15]为提高磁悬浮系统的控制性能，采用变边界指数趋近律的滑模控制方法，并通过仿真验证了磁悬浮系统在存在未知扰动的干扰下能够实现对目标位置的追踪，但该方法存在收敛时间长、抗干扰能力弱、不能实现系统全局稳定的问题。文献[16]采用自适应快速终端滑模控制，实现对磁悬浮系统位置追踪的同时提高控制系统的收敛时间，但磁悬浮系统中不仅存在外部扰动的影响，还存在由参数不匹配带来的扰动，而该方法对总扰动的处理能力弱，进而影响

收稿日期：2024-07-05；**录用日期**：2024-09-06；**网络出版时间**：2024-09-26 14:47
网络出版地址：link.cnki.net/urlid/11.2625.V.20240926.1158.001
基金项目：国家自然科学基金(52162050)
*通信作者. E-mail：linjt@lzjtu.edu.cn

引用格式：林俊亭，陈信州. 基于高增益扰动观测器的磁悬浮球系统滑模控制[J]. 北京航空航天大学学报，2025，51(6)：1897-1906.
LIN J T，CHEN X Z. Sliding mode control of magnetic levitation ball systems based on a high-gain disturbance observer[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1897-1906 (in Chinese).

了系统的控制效果。并且传统的滑模控制存在明显的抖振问题，而抖振问题与控制律中的符号函数有关，为削弱控制中的抖振现象，可将符号函数替换为其他形式的函数，但在一定程度上削弱系统的鲁棒性，无法保证磁悬浮系统在总扰动下的控制精度。

为提高滑模控制器的控制性能，近年来，研究人员又提出通过扩张状态观测器 (extended state observer, ESO)[17]、广义比例积分观测器 (generalized proportional integral observer, GPIO)[18]、扰动观测器[19]、等效输入干扰[20]等方式对磁悬浮系统中所存在的未知扰动进行观测，基于干扰估计值修改控制器输出方程，从而进行扰动补偿。但 ESO 和 GPIO 需要调节的参数多且结构复杂。文献 [21] 中针对四旋翼无人驾驶飞行器易受各种未建模或难以测量的扰动影响，设计了一种结构简单，参数易调节，收敛速度快的高增益扰动观测器 (high gain disturbance observer, HGDO)对系统中所受到的未知扰动进行估计，并将干扰估计与传统的滑模控制复合，提高了无人机的抗干扰能力。

基于上述分析，考虑系统中存在的建模误差与未知扰动对磁悬浮系统控制性能的影响，本文提出一种基于 HGDO 的自适应非奇异终端滑模控制 (adaptive nonsingular terminal sliding mode control, ANTSMC) 方法，对磁悬浮系统进行悬浮高度控制。首先，采用 ANTSMC 对悬浮系统进行位置追踪并减小控制器输出抖振；其次，设计对扰动观测效果更好，结构更简单的 HGDO 输出实时扰动估计值，将扰动估计值通过前馈补偿的方式引入控制律中，以提高系统鲁棒性；最后，通过仿真对比，验证了在相同参数、不同目标轨迹追踪下本文方法在削弱滑模控制抖振的同时，提高了系统的鲁棒性。

1 磁悬浮球数学模型

为对磁悬浮球系统的悬浮高度进行控制，文献 [22] 通过解耦的方式将悬浮系统简化为单点磁悬浮系统的结构，并以单点磁悬浮系统作为磁悬浮系统悬浮高度控制的控制对象。因此，本节将建立磁悬浮球的单点磁悬浮数学模型。

磁悬浮球系统的结构如图 1 所示，由控制器、功率放大器、位置传感器、电磁铁、铁球等构成。通过位置传感器检测铁球悬浮位置，控制器根据实际位置与参考位置的误差确定输出的控制电压，通过功率放大器将控制电压转换为控制电流，进而实现铁球稳定悬浮。

由于磁悬浮系统的非线性，很难对其建立一个

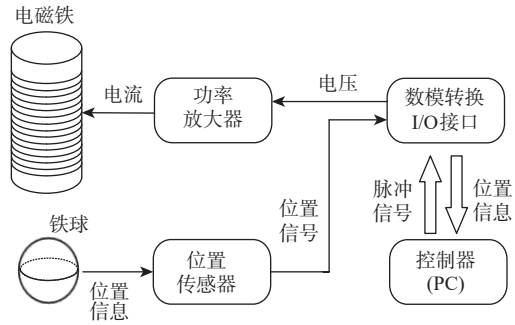


图 1 磁悬浮球系统结构

Fig. 1 Structure of magnetic levitation ball system

精确的模型，根据文献 [22-23]，对磁悬浮球系统做出如下假设：忽略漏磁影响，不考虑各个元器件之间传输的时滞，不考虑信号传输过程中的噪声，仅考虑悬浮球垂向动力学。则悬浮球系统的模型满足如下表达式：

$$\begin{cases} m\dfrac{\mathrm{d}^2x(t)}{\mathrm{d}t^2}=F(i,x)+mg+d(t) \\ F(i,x)=K(i/x)^2 \\ K=-\mu_0SN^2/4 \\ u(t)=Ri+L\dfrac{\mathrm{d}i}{\mathrm{d}t} \\ mg+F(i_0,x_0)=0 \end{cases} \tag{1}$$

式中：K为简化表达式的系数项；x为小球位置；i为线圈电流；$F(i,x)$为电磁力；$F(i_0,x_0)$为平衡点处的电磁力；m为小球质量；g为重力加速度；$d(t)$为系统中的外部扰动；$u(t)$为功率放大器的输入电压；R为线圈电阻；L为线圈的自感；μ_0为真空磁导率；N为线圈匝数；t为时间；S为磁导率面积。

通过文献 [18] 中将电磁力$F(i,x)$在平衡点(i_0,x_0)进行泰勒展开的方法，分析悬浮球系统的垂直方向动力学方程可得到系统的状态空间方程：

$$\begin{cases} \dot{x}_1=x_2 \\ \dot{x}_2=b_0u+a_0x_1+f \\ y=x_1 \end{cases} \tag{2}$$

式中：x_1为小球的位置，$x_1=x$；x_2为小球的速度，$x_2=\dot{x}$；u为磁悬浮球系统的控制律；a_0为系统中参数a的理想值，$a_0=-2Ki_0^2/mx_0^3$；b_0为系统中参数b的理想值，$b_0=2Ki_0/(mRx_0^2)$；f为由外部扰动与内部扰动所构成的总扰动，其表达式为

$$f=-\frac{2Ki_0}{mx_0^2}\cdot\frac{L\mathrm{d}i}{R\mathrm{d}t}+d(t)/m+F_\mathrm{h}/m+(a-a_0)x_1+(b-b_0)u \tag{3}$$

式中：F_h为平衡点处电磁力泰勒展开后的高次项。

由于磁悬浮球系统是一个强非线性系统，各个参数的实际值可能与理想值之间存在一定的误差，并且磁悬浮球系统存在未知的外部扰动的影响，难

以对该系统建立一个精确的模型, 很难针对该系统建立一个综合能力强的控制器, 因此, 本文将会针对该问题, 逐步设计一个基于 HGDO 的自适应非奇异终端滑模控制器。

2　控制器设计

本节将对磁悬浮球系统设计基于 HGDO 的 ANTSMC, 控制器的具体工作流程如图 2 所示, 其中, x_r 为磁悬浮球系统的参考位置, e_1 为磁悬浮球系统的位置误差, e_2 为磁悬浮球系统的速度误差, $\hat{f}$ 为扰动的观测值。

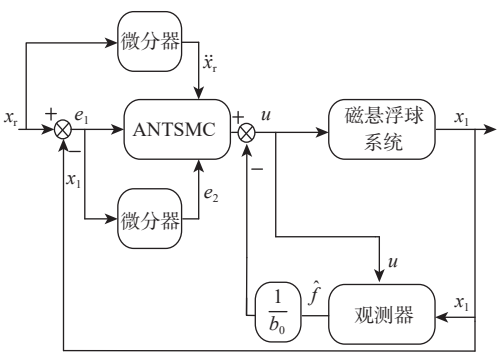


图 2　带观测器的 ANTSMC

Fig. 2　ANTSMC with observer

2.1　自适应滑模控制器的设计

针对磁悬浮球系统的位置追踪问题, 本节设计了非奇异终端滑模控制 (nonsingular terminal sliding mode control, NTSMC), 且通过 Lyapunov 证明该控制的稳定性。

令 x_r 为磁悬浮球系统的参考位置, 存在误差 $e_1 = x_r - x_1$, $e_2 = \dot{x}_r - \dot{x}_1$, 为构建非奇异终端滑模控制器, 对滑模面设计如下:

$$s = e_1 + e_2^{p/q}/\beta \tag{4}$$

式中: p 和 q 都为正奇数, 且满足 $1 < p/q < 2$; β 为正实数。设计 Lyapunov 函数为 $V_1 = s^2/2$, 对其求导可得

$$\dot{V}_1 = s\left\{\dot{e}_1 + \rho(e_2)\left[\ddot{x}_r - (b_0 u + a_0 x_1 + f)\right]\right\} \tag{5}$$

式中: $\rho(e_2) = pe_2^{p/q-1}/\beta q$。基于等速趋近律 $\dot{s} = -k\mathrm{sgn}(s)$ 与反演法可得到非奇异终端滑模面的控制律方程 u 为

$$u = \left(\ddot{x}_r - a_0 x_1 - f + \beta\frac{q}{p}e_2^{2-p/q} + \frac{k\mathrm{sgn}(s)}{\rho(e_2)}\right)\Big/ b_0 \tag{6}$$

式中: k 为趋近律的增益, 但式 (6) 中包含的总扰动 f 在 NTSMC 下是无法得到的, 同时方程中包含的 $\rho(e_2)$ 项在 $e_2 = 0$ 时, 会造成控制律的奇异问题, 因此, 将控制律设计为

$$u = \left(\ddot{x}_r - a_0 x_1 + \beta\frac{q}{p}e_2^{2-p/q} + k\mathrm{sgn}(s)\right)\Big/ b_0 \tag{7}$$

将设计的式 (7) 代入到式 (5) 中得到

$$\dot{V}_1 = -\rho(e_2)(k|s| + fs) \leqslant -\rho(e_2)|s|(k - |f|) \tag{8}$$

为使 $\dot{V}_1 \leqslant 0$, 需要 $k > |f|$, 并且总扰动 f 是有界的扰动。通过文献 [24] 可知对于 NTSMC 在满足 $k > |f|$、$e_2 \neq 0$ 的情况下, 小球的位置误差将在有限时间 t_r 内到达滑模面 $s = 0$, 并满足 $t_r \leqslant |s(0)|/(\rho(e_2)\cdot(k - |f|))$, t_r 为系统到达滑膜面 $s = 0$ 所需的时间。根据式 (4) 与 $s = 0$ 可得

$$e_1^{-q/p}e_2 = -\beta^{q/p} \tag{9}$$

设从 $e_1 \neq 0$ 到 $e_1 = 0$ 需要时间 t_s, 则由式 (9) 可得

$$t_s = \frac{p}{\beta^{q/p}(p-q)}e_1^{1-q/p}(t_r) \tag{10}$$

对式 (10) 进行分析可知, 系统收敛速度与 p、q 参数的大小有关, $p/q \to 2$ 则系统会牺牲计算速度以加快收敛速度, 使系统的状态误差在有限时间内收敛到 0。并且由于控制律 (7) 中不存在负幂, 从而所设计的控制器为非奇异终端滑模控制器。但在 NTSMC 下, 需要保证控制律中的 $k > |f|$, 但过大的 k 会导致明显的抖振现象, 过小会导致系统的抗干扰能力弱, 所以为提高控制器的综合性能。对控制器中的 k 设计如下自适应律:

$$\dot{k} = \begin{cases} \bar{k}|s|\mathrm{sgn}(|s| - \theta) & k > k_m \\ \mu & k \leqslant k_m \end{cases} \tag{11}$$

式中: k_m、θ、μ 和 $\bar{k}$ 均为正实数。通过自适应律可将控制律 (7) 进一步修改为

$$u = \left(\ddot{x}_r - a_0 x_1 + \beta\frac{q}{p}e_2^{2-p/q} + k(t)\mathrm{sgn}(s)\right)\Big/ b_0 \tag{12}$$

式中: $k(t)$ 为引入自适应律变化后的开关增益, 并假设存在 $k(t) > |f|$。

利用控制律 (12) 对系统 (2) 进行控制, 由于开关增益修改为时变参数, 因此, 构建的 Lyapunov 方程为 $V_2 = s^2/2 + \tilde{k}^2/2\gamma$, 其中, $\tilde{k} = k(t) - k^*$, k^* 为正实数, 且 $k^* > k(t)$, $k^* > f$, 对 V_2 求导可得

$$\begin{aligned}\dot{V}_2 &\leqslant -|s|\rho(e_2)(k^* - |f|) + \tilde{k}\dot{k}/\gamma \leqslant \\ &-|s|\rho(e_2)(k^* - |f|) + \rho(e_2)(k^* - k(t))|s| + \tilde{k}\dot{k}/\gamma \leqslant \\ &-\tilde{k}\left(\rho(e_2)|s| - \dot{k}/\gamma\right)\end{aligned} \tag{13}$$

假设式 (13) 在 $t > 0$ 的任意时刻 $k > k_m$。

1) 当 $|s| > \theta$ 时, 由式 (11) 可知 $\dot{k} = \bar{k}|s|$, 为满足 $-\tilde{k}\left(\rho(e_2)|s| - \dot{k}/\gamma\right) < 0$, 则有

$$\gamma < \bar{k}|s|/(\rho(e_2)|s|) \tag{14}$$

由式 (14) 可知, 在此情况下可以通过选择合适

的参数γ使$\dot{V}_2<0$，从而系统收敛。而γ并不是一个控制参数，仅在建立Lyapunov方程中出现，因此，可以判断系统在此情况下是能够收敛的。

2）当$|s|<\theta$，有$\dot{k}=-\bar{k}|s|$，则由式(13)易得$\dot{V}_2<0$。

本节根据所设计的自适应律，使ANTSMC可以在不损失系统抗扰动能力下提高系统的动态性能，但滑模控制难以避免的抖振问题仍然存在，因此，基于文献[15]中通过边界层的设置，使边界层外使用滑模控制，在边界层内使用连续控制，将符号函数替换为饱和函数。

通过饱和函数替代符号函数后，由于边界层的设计导致了系统的鲁棒性下降。对比控制律(6)和控制律(12)可知，ANTSMC的控制律中由于总扰动未知，并未包含对扰动的补偿。因此，参考如图2所示控制流程，通过观测器对未知扰动进行估计以估计值进行前馈补偿的方法，提高ANTSMC的鲁棒性。

2.2 观测器的设计

针对磁悬浮系统存在未知扰动的问题，设计了ESO、GPIO、HGDO对系统中所存在的扰动进行观测。将系统(2)中所存在的总扰动f看作一个新的状态变量x_3，总扰动f的变化率为x_4。将ESO设计为

$$\begin{cases}\dfrac{\mathrm{d}}{\mathrm{d}t}\hat{x}_1=\hat{x}_2+\dfrac{\alpha_1}{\varepsilon}(x_1-\hat{x}_1)\\ \dfrac{\mathrm{d}}{\mathrm{d}t}\hat{x}_2=a_0x_1+b_0u+\hat{x}_3+\dfrac{\alpha_2}{\varepsilon^2}(x_1-\hat{x}_1)\\ \dfrac{\mathrm{d}}{\mathrm{d}t}\hat{x}_3=\dfrac{\alpha_3}{\varepsilon^3}(x_1-\hat{x}_1)\end{cases}\tag{15}$$

式中：$\hat{x}_1$、$\hat{x}_2$、$\hat{x}_3$分别为观测器对x_1、x_2、x_3的观测值；α_1、α_2、α_3、ε为正实数。

为分析ESO的稳定性，设计误差矩阵：

$$\boldsymbol{\eta}=\begin{bmatrix}\eta_1 & \eta_2 & \eta_3\end{bmatrix}^{\mathrm{T}}\tag{16}$$

式中：$\eta_1=(x_1-\hat{x}_1)/\varepsilon^2$；$\eta_2=(x_2-\hat{x}_2)/\varepsilon$；$\eta_3=x_3-\hat{x}_3$。将式(15)与式(16)结合可以得到

$$\varepsilon\dot{\boldsymbol{\eta}}=\begin{bmatrix}-\alpha_1 & 1 & 0\\ -\alpha_2 & 0 & 1\\ -\alpha_3 & 0 & 0\end{bmatrix}\boldsymbol{\eta}+\varepsilon\begin{bmatrix}0\\0\\1\end{bmatrix}\dot{x}_3=\boldsymbol{A}\boldsymbol{\eta}+\varepsilon\boldsymbol{B}\dot{x}_3\tag{17}$$

式中：$\boldsymbol{A}$为观测误差的系数矩阵，当该矩阵的特征值均为负实部时矩阵为赫尔维茨矩阵；$\boldsymbol{B}$为扰动变化率的系数矩阵。

根据式(17)可知ESO的特征方程为$\lambda^3+\alpha_1\lambda^2+\alpha_2\lambda+\alpha_3=0$，所以为使ESO估计误差趋于0，减少观测器中调节的参数，并且使观测器误差矩阵的特征值全部分布在复平面的左半平面，则式(17)的特征方程可修改为

$$\lambda^3+\alpha_1\lambda^2+\alpha_2\lambda+\alpha_3=(\lambda+\lambda_{\mathrm{h}})^3\tag{18}$$

通过式(18)，可将观测器特征方程的特征根设计为$\lambda_{1,2,3}=-\lambda_{\mathrm{h}}$，则需要$\alpha_1=3\lambda_{\mathrm{h}}$，$\alpha_2=3\lambda_{\mathrm{h}}^2$，$\alpha_3=\lambda_{\mathrm{h}}^3$。于是通过对ESO特征根的设定，减少了观测器中需要条件的参数数目，并优化了参数设计过程，保证了观测器的稳定观测。

为建立各观测器之间的联系，基于文献[25]将GPIO修改设计为

$$\begin{cases}\dfrac{\mathrm{d}}{\mathrm{d}t}\hat{x}_1=\hat{x}_2+\dfrac{\alpha_1}{\varepsilon}(x_1-\hat{x}_1)\\ \dfrac{\mathrm{d}}{\mathrm{d}t}\hat{x}_2=a_0x_1+b_0u+\hat{x}_3+\dfrac{\alpha_2}{\varepsilon^2}(x_1-\hat{x}_1)\\ \dfrac{\mathrm{d}}{\mathrm{d}t}\hat{x}_3=\dfrac{\alpha_3}{\varepsilon^3}(x_1-\hat{x}_1)+\hat{x}_4\\ \dfrac{\mathrm{d}}{\mathrm{d}t}\hat{x}_4=\dfrac{\alpha_4}{\varepsilon^4}(x_1-\hat{x}_1)\end{cases}\tag{19}$$

式中：$\hat{x}_1$、$\hat{x}_2$、$\hat{x}_3$、$\hat{x}_4$分别为观测器对x_1、x_2、x_3、x_4的观测值；α_1、α_2、α_3、α_4为正实数。

为分析GPIO的稳定性，设计误差矩阵：

$$\boldsymbol{\eta}=\begin{bmatrix}\eta_1 & \eta_2 & \eta_3 & \eta_4\end{bmatrix}^{\mathrm{T}}\tag{20}$$

式中：$\eta_1=(x_1-\hat{x}_1)/\varepsilon^2$；$\eta_2=(x_2-\hat{x}_2)/\varepsilon$；$\eta_3=x_3-\hat{x}_3$；$\eta_4=\varepsilon(x_4-\hat{x}_4)$。

参考式(16)～式(18)过程，则可将GPIO中的参数设计为$\alpha_1=4\lambda_{\mathrm{h}}$，$\alpha_2=6\lambda_{\mathrm{h}}^2$，$\alpha_3=4\lambda_{\mathrm{h}}^3$，$\alpha_4=\lambda_{\mathrm{h}}^4$。

为建立各观测器之间的联系，基于文献[21]将HGDO的方程修改设计为

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{f}=\alpha_1(f-\hat{f})/\varepsilon\tag{21}$$

式中：α_1、ε为正实数构成观测器的增益；f为系统所受的总扰动；$\hat{f}$为观测器所观测到的总扰动。

HGDO的误差矩阵为$\boldsymbol{\eta}=[\eta_1]$，$\eta_1=f-\hat{f}$，参考式(16)～式(18)的过程，也有$\alpha_1=\lambda_{\mathrm{h}}$。

根据式(21)与式(2)可以得到磁悬浮球系统的HGDO方程为

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{f}=\alpha_1(\dot{x}_2-b_0u-a_0x_1-\hat{f})/\varepsilon\tag{22}$$

式(22)中存在$\dot{x}_2$即系统状态的的导数项，该项的存在会放大噪声的影响。可通过构建辅助变量的方式修改观测器方程，设计的辅助变量为

$$\gamma=\hat{f}-\alpha_1x_2/\varepsilon\tag{23}$$

对式(23)求导，并代入式(22)可得

$$\dot{\gamma}=\alpha_1(-b_0u-a_0x_1-\gamma-\alpha_1x_2/\varepsilon)/\varepsilon\tag{24}$$

为说明HGDO可以快速收敛到实际扰动值的可调邻域，设定扰动观测误差为$\tilde{f}=f-\hat{f}$，对其求导可得

$$\dot{\tilde{f}} = \dot{f} - (\dot{\gamma} + \alpha_1 \dot{x}_2/\varepsilon) \tag{25}$$

将式 (24) 与式 (2) 代入式 (25) 有

$$\dot{\tilde{f}} = \dot{f} - \frac{\alpha_1}{\varepsilon}\tilde{f} \tag{26}$$

对式 (26) 等号两边乘 $e^{\alpha_1 t/\varepsilon}$, 并在 [0,t] 上积分可得

$$\tilde{f}(t) = \tilde{f}(0)e^{-\frac{\alpha_1}{\varepsilon}t} + \int_0^t \dot{f}(\phi)e^{-\frac{1}{\varepsilon}(\alpha_1 t - \alpha_1 \phi)}\mathrm{d}\phi \tag{27}$$

式中: 对式 (27) 两边取绝对值, 利用 Holder's inequality[26] $\left|\int_0^t f(x)\mathrm{d}x\right| \leqslant \int_0^t |f(x)|\mathrm{d}x$ 可得

$$\left|\tilde{f}(t)\right| \leqslant \left|\tilde{f}(0)e^{-\frac{\alpha_1}{\varepsilon}t}\right| + \int_0^t \left|\dot{f}(\phi)e^{-\frac{1}{\varepsilon}(\alpha_1 t - \alpha_1 \phi)}\right|\mathrm{d}\phi \tag{28}$$

对式 (28) 进行 [0,t] 上积分可得

$$\int_0^t \left|\tilde{f}(\phi)\right|\mathrm{d}\phi \leqslant \int_0^t \left|\tilde{f}(0)e^{-\frac{\alpha_1}{\varepsilon}\phi}\right|\mathrm{d}\phi + \int_0^t \left(\int_0^\beta \left|\dot{f}(\phi)e^{-\frac{\alpha_1}{\varepsilon}(\beta-\phi)}\right|\mathrm{d}\phi\right)\mathrm{d}\beta \tag{29}$$

式 (29) 不等号右侧的第 1 项式的边界为 $\left|\tilde{f}(0)\right|\varepsilon/\alpha_1$, 第 2 项等价于 $\left|h(t) * \dot{f}(t)\right|$, 其中, $h(t) = e^{-\alpha_1 t/\varepsilon}$, "$*$" 表示卷积运算符。

根据杨氏卷积定理与 L_1 范数的定义[21] 有 $\left|h(t) * \dot{f}(t)\right| \leqslant |h(t)|\left|\dot{f}(t)\right|$, $|h(t)| \leqslant \varepsilon/\alpha_1$ 则

$$\int_0^t \left|\tilde{f}(\beta)\right|\mathrm{d}\beta \leqslant \left|\tilde{f}(0)\right|\varepsilon/\alpha_1 + \left|\dot{f}(t)\right|\varepsilon/\alpha_1 \tag{30}$$

式 (30) 不等号左侧可表示为 $\tilde{f}$ 的 L_1 范数, 从而有

$$\left|\tilde{f}(t)\right| \leqslant \left(\left|\tilde{f}(0)\right|\varepsilon\right)/\alpha_1 + \left(\varepsilon\left|\dot{f}(t)\right|\right)/\alpha_1 \tag{31}$$

式 (31) 表明扰动估计误差有界, 且受参数 ε 与 α_1 影响, 说明了 HGDO 可以快速收敛到实际扰动值的可调邻域, 即 α_1 越大则观测误差越小, ε 越小观测误差就越小。

本节设计结构更加简单的 HGDO 对系统进行扰动估计, 并通过参数 λ_h 将 ESO、GPIO、HGDO 互不相同的增益参数相互联系。使得对参数 λ_h 进行调节即可使观测器稳定, 简化了参数调节过程, 使各观测器能在相同参数 λ_h 下对总扰动的观测效果进行比较。

2.3　带扰动补偿的控制器设计

针对磁悬浮球系统滑模控制中, 降低滑模控制器输出抖振会导致系统鲁棒性下降的问题, 设计了带扰动补偿的自适应非奇异终端滑模控制, 并通过 Lyapnuov 验证了该控制的稳定性。为提高系统的鲁棒性, 通过如图 2 中前馈补偿的方式, 将观测器所得的扰动观测值与磁悬浮球系统中的控制律进行结合。

在避免控制律出现奇异问题的前提下, 基于式 (6) 与式 (12), 为将观测器所得到的扰动观测值引入到控制律方程中, 可通过用扰动观测值代替控制律 (6) 中扰动实际值的方法, 构建了带扰动补偿的自适应非奇异终端滑模控制律:

$$u = \left(\ddot{x}_r - a_0 x_1 - \hat{f} + \beta\frac{q}{p}e_2^{2-p/q} + k(t)\mathrm{sgn}(s)\right)\Big/ b_0 \tag{32}$$

式中: $\hat{f}$ 为扰动的观测值。可将 ESO、GPIO、HGDO 的观测结果代入其中。

根据控制律 (32) 构建的 Lyapunov 方程为 $V_3 = s^2/2 + \varepsilon\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\eta} + \tilde{k}^2/2\gamma$, 对 V_3 求导并结合式 (17) 可得

$$\dot{V}_3 = s\left\{\dot{e}_1 + \rho(e_2)\left[\ddot{x}_r - (b_0 u + a_0 x_1 + f)\right]\right\} + \boldsymbol{\eta}^{\mathrm{T}}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}\right)\boldsymbol{\eta} + 2\varepsilon\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{B}\dot{f} + \tilde{k}\dot{k}/\gamma \tag{33}$$

式中: $\boldsymbol{P}$ 为正定矩阵, 将控制律 (32) 代入式 (33) 可得:

$$\dot{V}_3 \leqslant |s|\rho(e_2)\left(k(t) - \left|\tilde{f}\right|\right) + \boldsymbol{\eta}^{\mathrm{T}}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}\right)\boldsymbol{\eta} + 2\varepsilon\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{B}\dot{f} + \tilde{k}\dot{k}/\gamma \tag{34}$$

参考式 (13) 有

$$\dot{V}_3 \leqslant -\tilde{k}\left(\rho(e_2)|s| - \dot{k}/\gamma\right) + \boldsymbol{\eta}^{\mathrm{T}}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}\right)\boldsymbol{\eta} + 2\varepsilon\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{B}\dot{f} \tag{35}$$

通过杨氏不等式[27] 修改式 (35) 可得

$$\dot{V}_3 \leqslant -\tilde{k}\left(\rho(e_2)|s| - \dot{k}/\gamma\right) + \boldsymbol{\eta}^{\mathrm{T}}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}\right)\boldsymbol{\eta} + \frac{1}{2\phi}\left(\varepsilon\dot{f}\right)^2 + 2\phi\left(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{B}\right)^2 \tag{36}$$

式中: ϕ 为正实数。利用 Riccati equation[28] 可知:

$$\bar{\boldsymbol{P}} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} + 2\phi\boldsymbol{P}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{0} \tag{37}$$

将式 (37) 代入式 (36) 即可得到

$$\dot{V}_3 \leqslant -\tilde{k}\left(\rho(e_2)|s| - \dot{k}/\gamma\right) - \boldsymbol{\eta}^{\mathrm{T}}\bar{\boldsymbol{P}}\boldsymbol{\eta} + \frac{1}{2\phi}\left(\varepsilon\dot{f}\right)^2 \leqslant -\lambda_{\min}\left(\bar{\boldsymbol{P}}\right)\left\|\boldsymbol{\eta}\right\|_2^2 + \frac{1}{2\phi}\left(\varepsilon\dot{f}\right)^2 - \tilde{k}\left(\rho(e_2)|s| - \dot{k}/\gamma\right) \tag{38}$$

式中: $\lambda_{\min}\left(\bar{\boldsymbol{P}}\right)$ 为 $\bar{\boldsymbol{P}}$ 的最小特征值; $\left\|\boldsymbol{\eta}\right\|_2^2 = \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta}$。

在控制过程中存在以下情况。

1) $|s| > \theta$, 此时 $\dot{k} = \bar{k}|s|$, 为满足 $\dot{V}_3 < 0$, $\left(\varepsilon\dot{f}\right)^2/2\phi + \left|-\tilde{k}\right|\left(\rho(e_2)|s| - \dot{k}/\gamma\right) < 0$ 则有

$$\gamma < \bar{k}|s| \Big/ \left[\rho(e_2)|s| + \frac{\left(\varepsilon\dot{f}\right)^2}{2\phi(k^* - k)}\right] \tag{39}$$

对式 (14) 分析相同, 故在此情况磁悬浮系统收敛。

2) 当 $|s| < \theta$, 则有 $\dot{k} = -\bar{k}|s|$, 为使 $\dot{V}_3 < 0$, 系统收敛则有

$$\left\|\boldsymbol{\eta}\right\|_2^2 \geqslant \frac{\left(\varepsilon\dot{f}\right)^2}{2\phi\lambda_{\min}\left(\bar{\boldsymbol{P}}\right)} + \frac{\left|-\tilde{k}\right|\left(\rho(e_2)|s| - \dot{k}/\gamma\right)}{\lambda_{\min}\left(\bar{\boldsymbol{P}}\right)} \cdot \frac{L}{\lambda_{\min}\left(\bar{\boldsymbol{P}}\right)} \tag{40}$$

式中：L为$\left(\varepsilon\dot{f}\right)^2\Big/2\phi+\left|-\tilde{k}\right|\left(\rho(e_2)|s|-\dot{k}/\gamma\right)$的上界。

式(40)基于文献[29]中的全局最终一致有界结论，得出系统是全局一致最终有界的，系统的误差满足式(40)时，V_3就会减小，由V_3的定义可知系统的误差减小。通过Lyapunov稳定性证明了带扰动补偿控制器的悬浮系统的稳定性。

本节基于观测器的扰动观测值与控制键式(12)和控制律式(32)，该控制律通过扰动补偿的方式削弱扰动对系统的影响，解决了饱和函数替代符号函数的滑模控制律输出抖振降低但磁悬浮球悬浮控制系统鲁棒性下降的问题。

3 仿真实验与结果分析

本文在MATLAB/Simulink中构造如图2所示磁悬浮控制模型，磁悬浮球系统部分的参数如表1所示。为比较各个控制器的性能，验证基于HGDO的自适应终端滑模控制器有效性，将扰动设置为$f=2\sin(2t+\pi/2)+0.1f_{\mathrm{B}}$，其中，$f_{\mathrm{B}}$设置为噪声强度0.05，采样时间0.03的白噪声，以模拟系统中具有未知不确定性的扰动。

表1 磁悬浮球系统仿真参数

Table 1 Parameters of magnetic levitation ball system in simulation

参数	数值
质量m/kg	0.17
线圈电阻R/Ω	13.577
线圈匝数N/T	1 057
真空磁导率μ/(H·m^{-1})	$4\pi\times10^{-7}$
磁导截面积S/m^2	$9\pi\times10^{-4}$
平衡位置x_{r}/m	0.04
平衡位置电流i_0/A	0.633

同时为对不同控制器进行公平比较，各控制器的参数如表2所示。并且，为定量描述各控制器的性能，采用文献[30]中积分时间乘方误差(integral time squared error, ITSE)与积分时间绝对误差(integral time absolute error, ITAE)的方法：

$$\begin{cases} e_{\mathrm{ITSE}}=\int_0^t te_{\mathrm{g}}^2\mathrm{d}t \\ e_{\mathrm{ITAE}}=\int_0^t t\left|e_{\mathrm{g}}\right|\mathrm{d}t \end{cases} \tag{41}$$

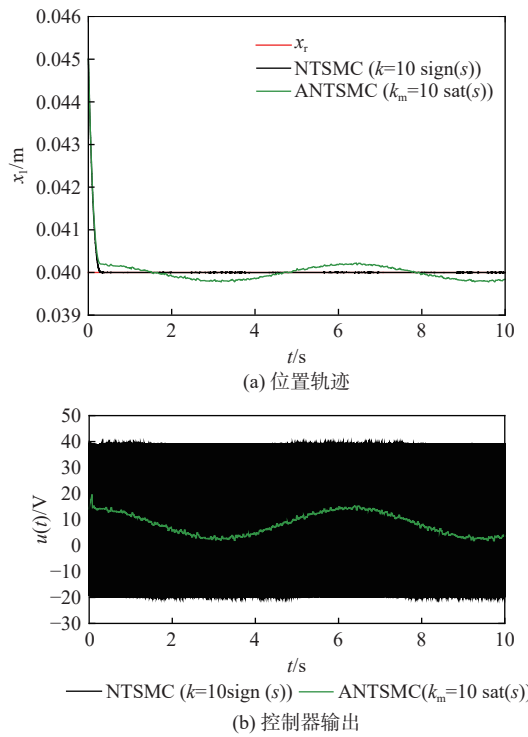
表2 控制器仿真参数

Table 2 Parameters of controller in simulation

控制方法	p	q	β	$\bar{k}$	k_{m}	λ_{h}	θ	μ
ANTSMC	5	3	1	100	10		0.03	30
ANTSMC+ESO	5	3	1	100	10	4	0.03	30
ANTSMC+GPIO	5	3	1	100	10	4	0.03	30
ANTSMC+HGDO	5	3	1	100	10	4	0.03	30

式中：e_{g}为观测器对扰动f的观测误差。

如图3所示，带饱和函数的ANTSMC可以明显的减小控制输出的抖振现象，但影响了控制器的抗干扰能力，说明了磁悬浮球系统在ANTSMC的控制下进行目标轨迹追踪时，不能兼顾控制器的鲁棒性与控制器的输出抖振。因此，利用观测器得到磁悬浮球系统中未知总扰动的估计值，通过前馈补偿的方式将估计值引入控制律中以解决该问题。



(a) 位置轨迹

(b) 控制器输出

图3 无扰动补偿的平稳轨迹追踪

Fig. 3 Smooth trajectory tracking without disturbance compensation

由于观测器的初始状态与实际值之间存在一定误差，会导致观测器输出存在峰值现象，造成观测器的收敛效果差，影响观测器的收敛效果。所以将观测器中的参数ε设计为

$$\frac{1}{\varepsilon}=\begin{cases}100t^3 & 0\leqslant t\leqslant 1\\ 100 & t>1\end{cases} \tag{42}$$

表3是对图4的定量分析，并在表3中对HGDO的定量分析结果进行加粗，从表3的定量分析可知，HGDO对系统总扰动的ITSE值与ESO和GPIO相比分别降低了84%和75%。HGDO对系统总扰动的ITAE值与ESO和GPIO相比分别降低了67%和60%。

根据各观测器的仿真结果与定量分析发现，各观测器的仿真结果与定量分析发现在时变ε、相同参数λ_{h}下HGDO对未知扰动f的观测效果，比ESO、GPIO具有更高的收敛速度与估计精度。

图5～图7为仿真结果，说明了在平稳轨迹，正

表 3　时变ε下各个观测器的性能

Table 3　Performance of different observers in time variationε

观测器类型	ITSE	ITAE
ESO	0.302 60	2.082 0
GPIO	0.209 70	1.708 0
HGDO	**0.049 25**	**0.690 6**

弦轨迹、方波轨迹下, 即使是系统存在不确定性与干扰, 通过扰动补偿的控制器会提高系统的抗干扰性能, 并进行的有效追踪, 削弱控制器输出抖振的同时, 提高了控制器的鲁棒性。由表 3 可知, HGDO 对未知扰动 f 的观测效果更好, 为控制器提供了更加精确的扰动值, 因此, 提高了带扰动补偿控制器的鲁棒性。

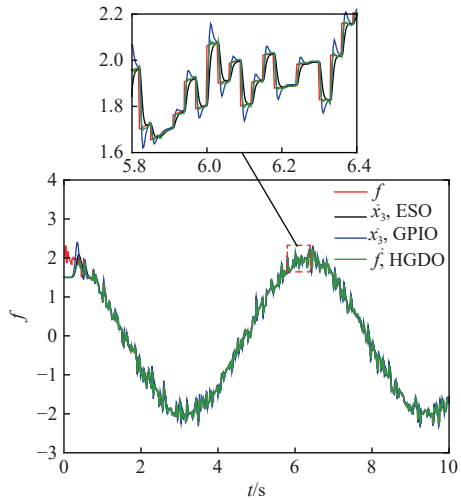
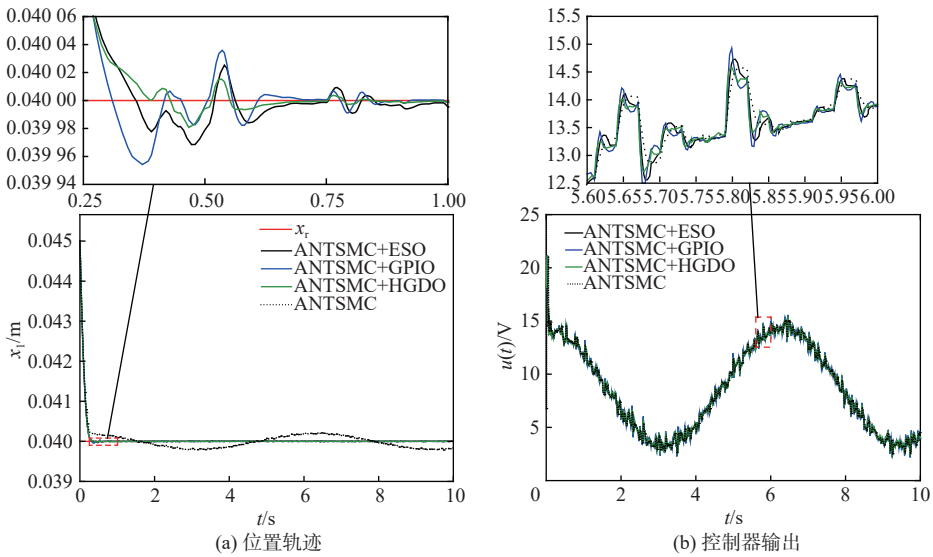


图 4　时变ε下的总扰动估计

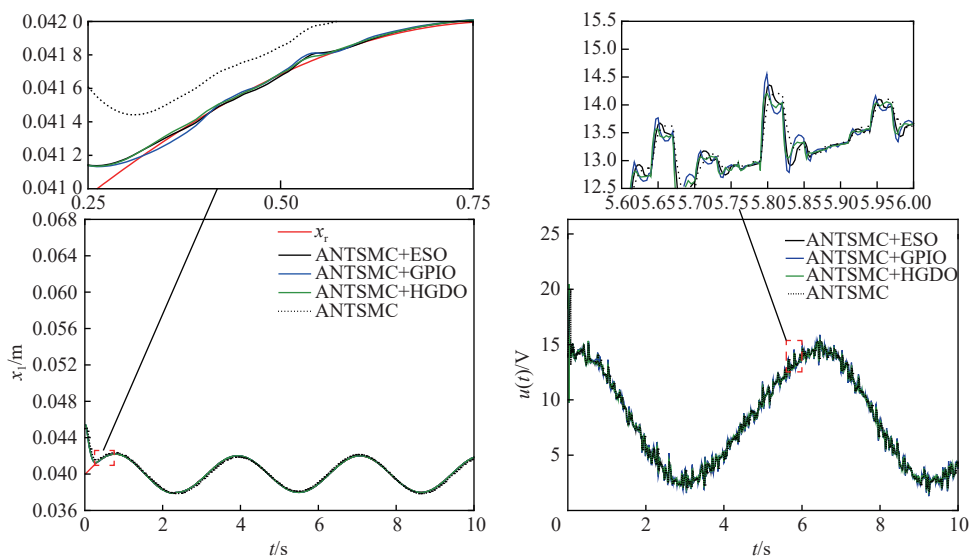
Fig. 4　Performance of different observers in time variation ε



(a) 位置轨迹　　(b) 控制器输出

图 5　带扰动补偿的平稳轨迹追踪

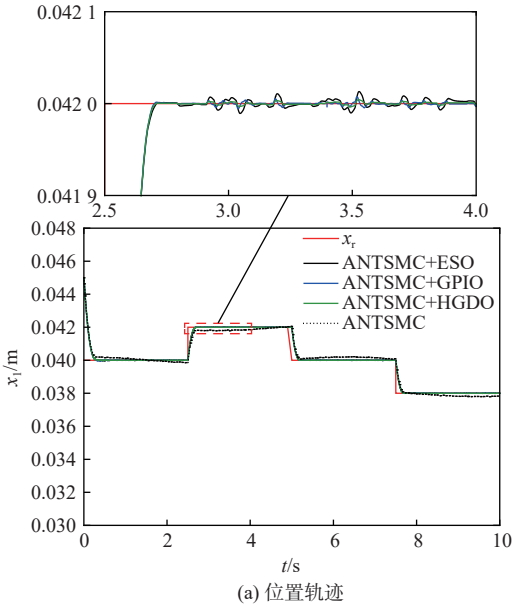
Fig. 5　Smooth trajectory tracking with disturbance compensation



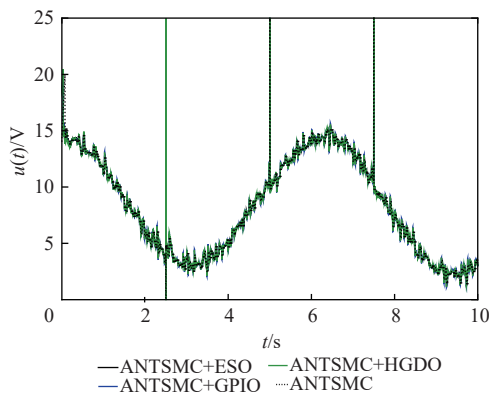
(a) 位置轨迹　　(b) 控制器输出

图 6　带扰动补偿的正弦轨迹追踪

Fig. 6　Sinusoidal trajectory tracking with disturbance compensation



(a) 位置轨迹



(b) 控制器输出

图 7 带扰动补偿的方波轨迹追踪

Fig. 7 Square trajectory tracking with disturbance compensation

4 结 论

本文研究了磁悬浮球系统存在未知扰动导致控制性能下降的问题，为提高控制器的综合动态性能，提出基于 HGDO 的 ANTSMC 方法。

1）考虑系统的外部扰动和参数不确定性构建磁悬浮球系统的数学模型。针对该模型设计了自适应非奇异终端滑模控制律以减小控制器输出的抖振现象，并证明了在该控制律下系统的稳定性。

2）针对时变总扰动，设计结构更简单的 HGDO 对其进行估计，为验证 HGDO 的观测结果，通过误差矩阵的特征值将 HGDO、ESO、GPIO 联系起来。

3）通过仿真结果比较分析了 HGDO+ANTSMC、ANTSMC、ANTSMC+ESO、ANTSMC+GPIO，说明本文的控制方案在不同目标轨迹下，降低了控制器抖振的同时，保证了系统的抗时变扰动能力与动态性能。

参考文献（References）

[1] BOLDEA I, TUTELEA L N, XU W, et al. Linear electric machines, drives, and MAGLEVs: an overview[J]. IEEE Transactions on Industrial Electronics, 2018, 65(9): 7504-7515.

[2] 金俊杰, 王岩峰, 徐程程, 等. 人工肾脏泵用磁悬浮轴承设计与磁力特性分析[J]. 西南交通大学学报, 2024, 59(4): 795-803.
JIN J J, WANG Y F, XU C C, et al. Design and magnetic force characteristic analysis of magnetic levitation bearing for artificial kidney pumps[J]. Journal of Southwest Jiaotong University, 2024, 59(4): 795-803(in Chinese).

[3] 雷城, 蓝益鹏, 孙云鹏. 直线同步电动机磁悬浮系统的模糊分数阶滑模控制[J]. 电机与控制学报, 2022, 26(3): 94-100.
LEI C, LAN Y P, SUN Y P. Fuzzy fractional sliding mode control of magnetic levitation system of linear synchronous motors[J]. Electric Machines and Control, 2022, 26(3): 94-100(in Chinese).

[4] 马卫华, 胡俊雄, 李铁, 等. EMS 型中低速磁浮列车悬浮架技术研究综述[J]. 西南交通大学学报, 2023, 58(4): 720-733.
MA W H, HU J X, LI T, et al. Technologies research review of electro-magnetic suspension medium-low-speed maglev train levitation frame[J]. Journal of Southwest Jiaotong University, 2023, 58(4): 720-733(in Chinese).

[5] WANG J X, YU L. Adaptive resonant-EIDO-based optimized position precision control for magnetic levitation system[J]. IEEE Transactions on Industrial Electronics, 2023, 70(5): 5013-5023.

[6] XU Y S, ZHAO Z G, YIN S, et al. Real-time performance optimization of electromagnetic levitation systems and the experimental validation[J]. IEEE Transactions on Industrial Electronics, 2023, 70(3): 3035-3044.

[7] SUN Y G, XU J Q, QIANG H Y, et al. Adaptive neural-fuzzy robust position control scheme for maglev train systems with experimental verification[J]. IEEE Transactions on Industrial Electronics, 2019, 66(11): 8589-8599.

[8] KHAN M K A A, MANZOOR S, MARAIS H, et al. PID Controller design for a Magnetic Levitation system[C]//Proceedings of the IEEE 4th International Symposium in Robotics and Manufacturing Automation. Piscataway: IEEE Press, 2018: 1-5.

[9] NI F, ZHENG Q H, XU J Q, et al. Nonlinear control of a magnetic levitation system based on coordinate transformations[J]. IEEE Access, 2019, 7: 164444-164452.

[10] ZHANG Y, XIAN B, MA S G. Continuous robust tracking control for magnetic levitation system with unidirectional input constraint [J]. IEEE Transactions on Industrial Electronics, 2015, 62(9): 5971-5980.

[11] 沈莹雅, 张建生, 张远, 等. 主动磁悬浮轴承的鲁棒反演滑模控制[J]. 机械制造与自动化, 2016, 45(1): 194-198.
SHEN Y Y, ZHANG J S, ZHANG Y, et al. Robust backstepping sliding-mode control of active magnetic bearing[J]. Machine Building & Automation, 2016, 45(1): 194-198(in Chinese).

[12] FOLEA S, MURESAN C I, DE KEYSER R, et al. Theoretical analysis and experimental validation of a simplified fractional order controller for a magnetic levitation system[J]. IEEE Transactions on

Control Systems Technology, 2016, 24(2): 756-763.

[13] ROY P, ROY B K. Sliding mode control versus fractional-order sliding mode control: applied to a magnetic levitation system[J]. Journal of Control, Automation and Electrical Systems, 2020, 31(3): 597-606.

[14] PANDEY S, DOURLA V, DWIVEDI P, et al. Introduction and realization of four fractional-order sliding mode controllers for nonlinear open-loop unstable system: a magnetic levitation study case[J]. Nonlinear Dynamics, 2019, 98(1): 601-621.

[15] SUN Y G, LI W L, QIANG H Y. The design and realization of magnetic suspension controller of low-speed Maglev train[C]//Proceedings of the IEEE/SICE International Symposium on System Integration. Piscataway: IEEE Press, 2016: 1-6.

[16] BOONSATIT N, PUKDEBOON C. Adaptive fast terminal sliding mode control of magnetic levitation system[J]. Journal of Control, Automation and Electrical Systems, 2016, 27(4): 359-367.

[17] CHEN Q, YANG Q, LIU M X, et al. Decentralized control design and implementation for magnetic levitation systems with extended state observer[J]. Control Theory and Technology, 2023, 21(3): 397-413.

[18] WANG J X, ZHAO L, YU L. Adaptive terminal sliding mode control for magnetic levitation systems with enhanced disturbance compensation[J]. IEEE Transactions on Industrial Electronics, 2020, 68(1): 756-766.

[19] NERKAR S S, LONDHE P S, PATRE B M. Design of super twisting disturbance observer based control for autonomous underwater vehicle[J]. International Journal of Dynamics and Control, 2022, 10: 306-322.

[20] HU K T, XU J Q, AI Y H, et al. Super-twisting-observer-based equivalent-input-disturbance compensation suspension control of medium-low speed maglev[J]. IEEE Transactions on Vehicular Technology, 2024, 73(6): 9009-9014.

[21] IZADI M, FAIEGHI R. High-gain disturbance observer for robust trajectory tracking of quadrotors[J]. Control Engineering Practice, 2024, 145: 105854.

[22] SUN Y G, XU J Q, QIANG H Y, et al. Adaptive sliding mode control of maglev system based on RBF neural network minimum parameter learning method[J]. Measurement, 2019, 141: 217-226.

[23] 曹泽华, 杨杰, 胡海林. 单点磁悬浮系统自抗扰自适应控制[J]. 哈尔滨工业大学学报, 2024, 56(8): 86-93.
CAO Z H, YANG J, HU H L. Active disturbance rejection and adaptive control of magnetic suspension system[J]. Journal of Harbin Institute of Technology, 2024, 56(8): 86-93(in Chinese).

[24] FENG Y, YU X H, MAN Z H. Non-singular terminal sliding mode control of rigid manipulators[J]. Automatica, 2002, 38(12): 2159-2167.

[25] WANG J X, RONG J Y, YU L. Dynamic prescribed performance sliding mode control for DC–DC buck converter system with mismatched time-varying disturbances[J]. ISA Transactions, 2022, 129: 546-557.

[26] WON D, KIM W, SHIN D, et al. High-gain disturbance observer-based backstepping control with output tracking error constraint for electro-hydraulic systems[J]. IEEE Transactions on Control Systems Technology, 2015, 23(2): 787-795.

[27] DIBENEDETTO E. Real analysis[M]. Boston: Birkhäuser Boston, 2002.

[28] GRIGORIADIS D E. A unified algebraic approach to control design[M]. Boca Raton: CRC Press, 1997.

[29] KHALIL H K. Nonlinear systems[M]. 3rd ed. Upper Saddle River: PrenticeHall, 2002.

[30] WANG J X, RONG J Y, YANG J. Adaptive fixed-time position precision control for magnetic levitation systems[J]. IEEE Transactions on Automation Science and Engineering, 2023, 20(1): 458-469.

Sliding mode control of magnetic levitation ball systems based on a high-gain disturbance observer

LIN Junting*, CHEN Xinzhou

(School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: Using a high-gain disturbance observer(HGDO), a self-adaptive nonsingular terminal sliding mode control (ANTSMC) approach is devised to solve the issue of control performance degradation in magnetic levitation ball systems caused by modeling errors and unknown disturbances. First, a model of the maglev ball system is developed and the model is linearized at the equilibrium point. Then, to weaken the chattering of the sliding mode controller and guarantee the finite-time convergence of tracking errors, an adaptive nonsingular terminal sliding mode controller is designed. To estimate the lumped disturbance, a high-gain disturbance observer is employed. Theoretical convergence findings demonstrate that the suggested high-gain disturbance observer may rapidly converge to an adjustable neighborhood of real disturbance values. Adaptive nonsingular terminal sliding mode control law with disturbance compensation is designed based on the disturbance estimation, the system is proven globally uniformly and ultimately bounded under the control law. According to simulation and quantitative analysis, the system's robustness under the same control method is improved when the controller with a high-gain disturbance observer lowers the integral time squared error(ITSE)value by 75% and the integral time absolute error(ITAE) value by 60% for the total disturbance observation error when compared to the controller with a generalized proportional-integral observer.

Keywords: magnetic levitation ball system; sliding mode control; high-gain disturbance observer; disturbance compensation; trajectory tracking

Received: 2024-07-05; **Accepted**: 2024-09-06; **Published Online**: 2024-09-26 14:47

URL: link.cnki.net/urlid/11.2625.V.20240926.1158.001

Foundation item: National Natural Science Foundation of China (52162050)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0402

基于QAR数据与关联规则的着陆阶段超限风险评估

汪磊*，张楠，高杉
(中国民航大学 安全科学与工程学院，天津 300300)

摘　　要：飞行品质监控(FOQA)中对超限事件进行监测是识别飞行潜在风险的重要手段之一，但该方法对各超限事件之间的关联性研究不足，无法量化存在相互影响关系的超限风险。为此，提出一种基于快速存取记录器(QAR)数据与关联规则挖掘的着陆阶段超限风险评估方法。对9 799个航段样本在进近着陆阶段发生各超限事件的频次进行统计，结合专家意见筛选出152～15 m下降率大、着陆坡度大、着陆速度大、着陆俯仰角小、15 m至接地距离远、着陆垂直载荷大6个着陆阶段风险评估指标；基于超限项目间关联性计算风险指标发生的可能性权重，依据风险指标的参数分布计算航班指标偏移量以评价超限风险的严重性，构建基于云模型的着陆超限风险评估模型。提取QAR数据应用该模型进行验证，结果显示风险评估结果具有区分度。所建模型为着陆阶段的超限风险评估提供了一种有效的方法，有利于飞行安全监管和飞行运行优化。

关　键　词：QAR数据；关联规则；着陆超限；PageRank算法；风险评估
中图分类号：V328.5
文献标志码：A　　**文章编号**：1001-5965(2025)06-1907-09

着陆阶段一直以来都是民航飞行的关键阶段。空客公司2023年的事故统计报告显示，着陆阶段仅占整体飞行平均时长的1%，但近20年内(2002—2022年)，超过50%的飞机机体损毁事故都发生在着陆期间[1]，因此，如何提升该阶段的飞行安全水平是目前研究的重点之一。同时，近年来随着飞机系统可靠性的持续提升及民航安全管理体系的不断完善，飞机的事故率逐年降低。波音公司2022年的事故统计报告同样显示，1959年至今，航空事故率整体呈下降趋势，并自1980年后长期维持在一个较低的水平[2]。因此，传统的基于事故及不安全事件报告的事后分析方法已无法满足日益增长的安全需求。中国民用航空局自1997年起逐渐开展飞行品质监控(flight operation quality assurance, FOQA)工作，在《飞行品质监控(FOQA)实施与管理》咨询通告[3]中提出，通过收集分析快速存取记录器(quick access recorder, QAR)数据中的超出监控项目阈值的超限事件，识别飞行中的安全隐患[4]。因此，研究着陆阶段的超限风险对飞行运行中的风险管控具有重要的现实意义。

目前国内外学者针对飞行数据的研究主要集中于两方面，一方面是基于飞行数据的航班运行异常检测和不安全事件预测分析研究，如Ayra等[5]建立了冲出跑道概率预测模型，以评估起飞、着陆等不同场景下冲出跑道的风险概率；孙瑞山等[6]以飞行参数为输入，基于高斯混合模型提出一种检测进近着陆阶段飞机运行异常的方法；祁明亮等[7]以着陆阶段的超限事件为研究对象，对样本进行约束划分空间建立数学规划模型，并对发生该类事件的高风险操作区域进行分析。另一方面是通过QAR数据对飞行阶段或不安全事件进行风险评估，如Hadjimichael等[8]结合实时飞行参数及天气信息等

收稿日期：2023-06-21；**录用日期**：2023-09-08；**网络出版时间**：2024-06-18 14:56
网络出版地址：link.cnki.net/urlid/11.2625.V.20240617.1513.002
基金项目：中国民用航空局安全能力建设资金项目(KJZ49420210076)
* **通信作者**. E-mail：wanglei0564@hotmail.com

引用格式：汪磊，张楠，高杉. 基于QAR数据与关联规则的着陆阶段超限风险评估[J]. 北京航空航天大学学报，2025，51(6)：1907-1915.
WANG L，ZHANG N，GAO S. Risk assessment of landing phase exceedance based on QAR data and association rule[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1907-1915 (in Chinese).

对风险因素进行辨识，建立了可控飞行撞地风险评估模型；汪磊等[9]提出一种基于QAR数据的重着陆风险评价模型，依据重着陆风险分布函数对风险值进行计算；陈农田等[10]建立了基于熵权可变模糊识别的深度学习模型来评估高高原机场进近着陆阶段的风险。综合来看，国内外应用飞行数据的研究虽然较多，但大多集中于特定不安全事件的分析或评估，从系统角度对飞行阶段进行风险评估的研究很少。同时现有的风险评估方法[11-12]鲜少考虑飞行过程的连续性所导致事件间的相互作用，如进近时的拉平操作会对着陆平飘距离及着陆垂直载荷产生很大影响[13]，下降速度过大会对飞机姿态产生影响等[14]。

关联规则算法是数据挖掘领域常用的无监督机器学习算法，主要用于挖掘海量数据中属性同时出现的规律与模式。该算法在交通安全领域的事故致因分析研究中被广泛应用，如Montella[15]通过关联规则算法挖掘道路环形交叉口的事故致因，确定了与环境相关的影响因素；Li等[16]通过挖掘燃气爆炸事故报告中的关联规则，得到了燃气爆炸事故的潜在规律性特征；Lan等[17]采用Éclat算法从船舶全损事故数据集中挖掘与事故类型和事故严重程度相关的关联规则，发现船龄超过20年是导致船舶全损的主要原因；时统宇等[18]采用先验频繁模式算法挖掘运行要素和延误事件之间的内在关联，发现了增大航班延误概率的影响因素；袁乐平等[19]通过关联规则算法挖掘事故/征候报告中与不安全事件严重程度有关的因素，得到了造成事故/征候的主要原因；谢嘉仪[14]通过关联规则挖掘出不稳定进近事件的触发诱因要素，并采用贝叶斯网络描述其相互关系，但并未根据其关联性进行更深入的研究。

综上，国内外学者对着陆阶段的风险评估及关联规则算法在民航领域的应用方面均取得了相应的成果，但少有人结合超限事件间关联性对该阶段的超限风险进行评估，因此，本文基于QAR数据和关联规则提出一种着陆阶段超限风险评估方法，对着陆阶段的超限风险进行量化评估，为航空公司的FOQA及风险管理提供决策支持。

1 超限风险评价指标选取

FOQA项目和监控标准是基于FOQA指导并结合各个公司的运行要求而制定的。因此，不同公司、不同机型的FOQA项目标准有所不同。为提高着陆超限风险评估的效率与准确率，本文基于某航空公司的B737-800机型的FOQA项目与标准，对其2014—2016年间进近着陆阶段的超限事件发生频次进行统计，具体结果如表1所示。

根据该公司进近着陆阶段超限事件的发生频次并结合飞行专家的意见，本文拟选取发生频次超过100次的着陆阶段超限事件（飞行高度为305 m（1 000 ft）以下监控项目）纳入评价指标体系，其中，E_1：152～15 m（500～50 ft）下降率大；E_2：着陆坡度大；E_3：着陆速度大；E_4：着陆俯仰角小；E_5：15 m（50 ft）至接地距离远；E_6：着陆垂直载荷大。编制超限事件在着陆阶段的重要性程度专家问卷，向2名资深飞行员、1名飞行管理人员及1名飞行教员发放，对每项超限事件进行1～5分的重要性程度打分，1～5分分别对应从“重要”到“非常重要”的5个

表1 进近着陆阶段超限事件统计

Table 1 Statistics of exceedance incidents in landing phase

超限事件	轻度超限阈值	轻度超限次数	严重超限阈值	严重超限次数	合计次数
A_1：GPWS警告		0	探测到	17	17
A_2：610～305 m(2 000～1 000 ft)下降率大	>7.620 m/s	138	>9.144 m/s	26	164
A_3：305～152 m(1 000～500 ft)下降率大	>6.604 m/s	18	>7.620 m/s	3	21
E_1：152～15 m(500～50 ft)下降率大	>5.588 m/s	414	>6.604 m/s	8	422
A_4：152～61 m(500～200 ft)进近坡度大	>15°	1	>20°	0	1
A_5：61～15 m(200～50 ft)进近坡度大	>8°	16	>10°	2	18
E_2：着陆坡度大	>4°	696	>6°	23	719
A_6：低空大速度	>118.22 m/s	89	>128.5 m/s	22	111
A_7：152～15 m(500～50 ft)进近速度大	>(V_{ref}+12.85) m/s	31	>(V_{ref}+15.42) m/s	5	36
E_3：着陆速度大	>(V_{ref}+7.71) m/s	167	>(V_{ref}+10.28) m/s	24	191
A_8：放起落架晚	<457.5 m	5	<6.604 m/s	0	5
A_9：着陆襟翼到位晚	<366 m	12	<5.08 m/s	2	14
A_{10}：着陆俯仰角大	>7.4°	9	>7.3°	0	9
E_4：着陆俯仰角小	<1°	132	<0.5°	32	164
A_{11}：接地点近	<224 m	9	<2.52 m/s	0	9
E_5：15 m(50 ft)至接地距离远	>762.5 m	5 029	>15.24 m/s	403	5 432
E_6：着陆垂直载荷大	>1.68g	213	>1.89g	13	226

注：g为重力加速度；V_{ref}为着陆参考速度。

重要性等级。最后, 采用 Kendall's W 检验对 4 名飞行专家对着陆阶段超限事件重要程度的一致性进行分析。结果表明, 4 名飞行专家对 E_1～E_6 的重要性评价分值均大于 4, 且一致性较好 (W^a=0.721, p=0.013), 其中, W^a为表明一致性的 Kendall 协同系数, 显著性p值小于 0.05, 表明一致性具有统计学意义。因此, 最终确定 E_1～E_6 作为风险指标集中的超限风险评价项目。

2　基于关联规则的超限风险评估模型构建

本文所采用的风险概念是指在非确定性情况下, 可能发生某种不利后果的概率与严重程度的综合[20]。基于关联规则的着陆超限风险评估建模过程如图 1 所示。首先, 通过 Apriori 算法挖掘超限监控项目间的关联规则, 根据所得关联规则并结合 PageRank 算法进行计算, 得出各指标的发生可能性权值; 然后, 提取大样本空间下各指标的测量值, 检验其分布并计算, 量化超限监控项目的偏移严重程度; 最后, 采用云模型建立着陆超限风险评估模型, 划分出风险等级。

2.1　风险指标可能性权重计算

2.1.1　关联规则算法—Apriori

关联规则概念由 Agrawal 等[21] 于 1993 年提出, 核心是通过寻找目标数据集中的频繁项集来发掘事物之间的潜在关联规则。关联规则反应出当 U 事务发生时, V 事务也随之发生的规律。本文选取支持度 S_{support}、置信度 C_{on} 和提升度 L_{ift} 作为关联规则是否成立的测度指标, 具体计算式如下:

$$S_{\text{support}}(U \to V) = P(UV) \tag{1}$$

$$C_{\text{on}}(U \to V) = \frac{P(UV)}{P(U)} \tag{2}$$

$$L_{\text{ift}}(U \to V) = \frac{P(UV)}{P(U)P(V)} \tag{3}$$

式中: 支持度 S_{support} 为事务集中中同时包含 U 和 V 的事务数与全部事务数的比值, 即事务 U 和事务 V 同时出现的概率; 置信度 C_{on} 为事务集中同时包含 U 和 V 的事务数与包含 U 的事务数的比值, 即在事件 U 发生的基础上事件 V 发生的概率; 提升度 L_{ift} 为同时包含 U 和 V 的事务数与包含 U 的事务数和包含 V 的事务数乘积的比值, 其值大于 1 且越高表明事务之间正相关性越高。

Apriori 算法是解决关联规则挖掘问题最经典、最具影响力的算法之一[22], 本文使用 Apriori 算法对超限事件之间的相关性进行挖掘。先采集某航空公司某类机型在一段时间内发生超限事件航班的超限类目数据, 由于超限监控项目类目较多, 为提高数据挖掘效率, 研究采用基于约束的挖掘方法进行数据预处理[23]。对同时存在 2 个及 2 个以上超限事件发生的航段数据, 以航班为单位进行统计, 得到 m 个航班的超限事务数据集 $T=\{T_1,T_2,\cdots,T_m\}$, 设置最小支持度 $S_{\text{support, min}}$ 及最小置信度 $C_{\text{on, min}}$。对超限事务数据集 T 进行第 1 次扫描, 生成候选 1(即项集 C_1), 搜索出所有支持度大于等于 $S_{\text{support, min}}$ 的超限事件项集, 组成频繁 1(即项集 L_1), 由 L_1 连接组成候选 2(即项集 C_2); 再次扫描超限事务数据集 T, 搜索出 C_2 中所有支持度大于等于 $S_{\text{support, min}}$ 的项集, 组成频繁 2(即项集 L_2), 连接 L_2 得到候选 3(即项集 C_3); 同上, 第 n 次扫描数据库, 搜索出 C_n 中所有支持度大于等于 $S_{\text{support, min}}$ 的项集组成频繁 n(即项集 L_n), 连接得候选 $(n+1)$ (即项集 C_{n+1}), 直至没有新的候选集产生。从频繁项集 $L=\{L_1,L_2,\cdots,L_n\}$中计算出所有的关联规则, 通过式(2)计算其置信度, 得出所有置信度大于等于 $C_{\text{on, min}}$ 的关联规则, 最终通过

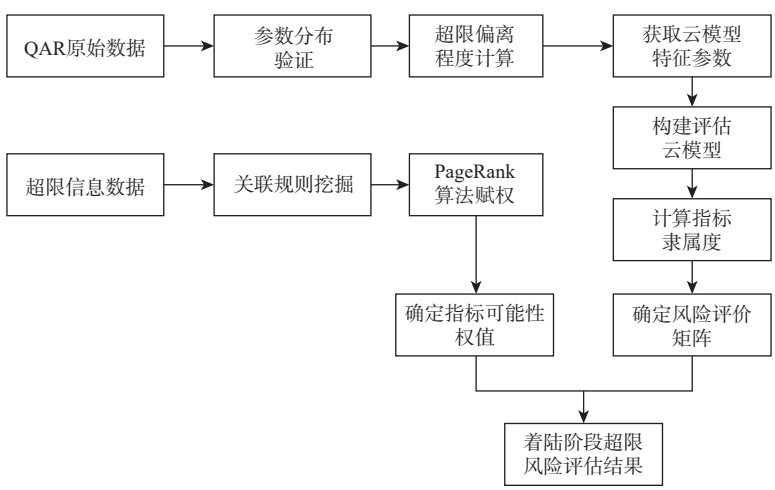


图 1　基于关联规则的着陆超限风险评估建模过程

Fig. 1　Modelling process for landing exceedance risk assessment based on association rule

式(3)筛选出提升度大于1的强关联规则，建立超限监控项目关联，Apriori算法流程如图2所示。

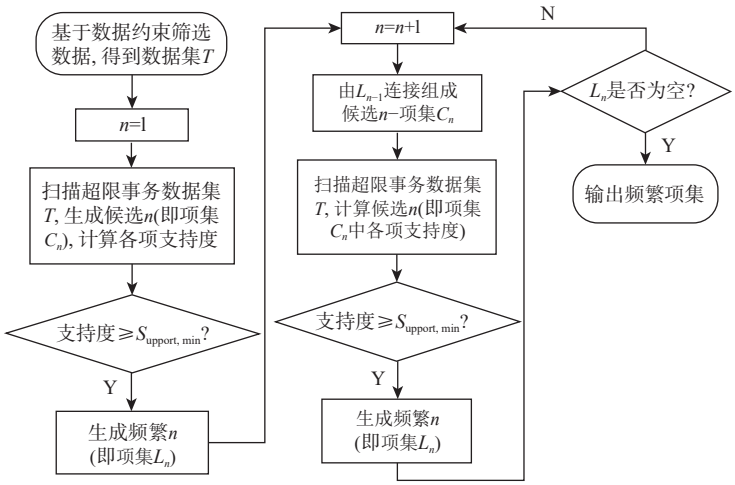


图2 Apriori算法流程

Fig. 2 Flow of Apriori algorithm

2.1.2 PageRank算法

PageRank是由Brin和Page提出的一种基于链接数量对网页进行重要度赋权的算法[24]，该算法核心思想是如果一个网页被较多其他网页所链接，则说明该网页有较高的重要程度，同时，被高重要程度网页所连接的网页，其重要程度也会相应提高。若在一个体系中存在独立网页不与其他网页相连，则可以将该独立网页视为自引用进行权重计算。本文通过Apriori算法挖掘得到超限监控项目间关联规则，以超限事件为节点，关联性为有向边，采用PageRank算法计算超限事件发生的可能性值。具体算法应用步骤如下。

步骤1 设置各节点初始P值

P值即一个网页被访问的概率，本文中将初始P值设定为各超限项目的发生次数占比，超限项目j的发生概率值为

$$P_j=\frac{C_j}{C_{\text{all}}} \tag{4}$$

式中：C_{all}为所有超限项目发生的总次数。对所得关联规则中各节点的P值进行归一化处理，如下：

$$Z=\{P_1,P_2,\cdots,P_k\} \tag{5}$$

$$P'_k=a\frac{P_k-\min(P)}{\max(P)-\min(P)}+b \tag{6}$$

式中：Z为各节点发生概率序列；P'_k为标准化后的k节点初始P值；$\max(P)$为所有节点中的最大P值，$\min(P)$为所有节点中的最小P值；a为归一化取值范围系数；b为归一化结果最小取值。

步骤2 对各节点P值进行更新得到各节点可能性权值，计算式为

$$P_{\text{E}}(k)=\frac{1-d}{N}+d\sum_{p\in M(k)}\frac{P'_p}{L(p)} \tag{7}$$

式中：k为超限项目节点；N为总节点数；d为任意节点发生可能性传递的概率，一般取值0.85；$M(k)$为指向k节点的节点集合；$L(p)$为节点p的出链数目。通过式(7)不断迭代计算各节点可能性权值至收敛，再对风险指标集中各指标的收敛值进行归一化处理即得各风险指标可能性权重。

2.2 严重度计算方法

当下航空公司主要参照FOQA制定超限标准，并且其阈值较FOQA标准偏低。对QAR数据进行分析可以发现，一些航段的监控数据未达到超限标准，但是已在标准附近浮动，这表明，这些航段发生超限事件的风险较高。同时QAR数据统计显示，在较大的样本空间里(s>100)，大部分参数在相同监控节点的取值都近似呈现正态分布或偏态分布[9]，因此，本文拟采用样本数据偏离指标均值的程度来表征超限风险的严重度。

检验各风险指标的参数值分布。对采集到的某机型某历史周期内m个航班的QAR数据进行处理，提取出各风险指标的数值。以某指标为例，计算其均值μ和方差S^2：

$$\mu=\frac{1}{m}\sum_{i=1}^{m}X_i\qquad i\in[1,m] \tag{8}$$

$$S^2=\frac{1}{m}\sum_{i=1}^{m}(X_i-\overline{X})\qquad i\in[1,m] \tag{9}$$

式中：X_i表示该指标单个样本值。

假设该数据样本服从正态分布$N(\mu,\sigma^2)$，对样本进行Anderson-Darling正态检验，若结果表明样本分布符合正态分布，则对样本方差进行偏差修

正, 得到总体方差的无偏估计:

$$\sigma^2=\frac{m}{m-1}S^2=\frac{1}{m-1}\sum_{i=1}^{m}(X_i-\overline{X})\qquad i\in[1,m]\quad(10)$$

当指标参数值越大, 风险越高时, 采用如下指标严重度计算式:

$$\begin{cases}S_{\mathrm{E}}(t)=(\alpha_t-\mu)/\sigma & \alpha_t>\mu\\ S_{\mathrm{E}}(t)=0 & \alpha_t<\mu\end{cases}\quad(11)$$

当指标参数值越小, 风险越高时, 则指标严重度计算式为

$$\begin{cases}S_{\mathrm{E}}(t)=(\mu-\alpha_t)/\sigma & \alpha_t<\mu\\ S_{\mathrm{E}}(t)=0 & \alpha_t>\mu\end{cases}\quad(12)$$

若检验结果表明样本分布为偏态分布, 则当分布为正偏态时, 指标严重度计算式为

$$\begin{cases}S_{\mathrm{E}}(t)=\dfrac{\alpha_t-Q_2}{Q_3-Q_1} & \alpha_t>Q_2\\ S_{\mathrm{E}}(t)=0 & \alpha_t<Q_2\end{cases}\quad(13)$$

当分布为负偏态时, 指标严重度计算式为

$$\begin{cases}S_{\mathrm{E}}(t)=\dfrac{Q_2-\alpha_t}{Q_3-Q_1} & \alpha_t<Q_2\\ S_{E}(t)=0 & \alpha_t>Q_2\end{cases}\quad(14)$$

式中: 严重度 $S_{\mathrm{E}}(t)\geqslant 0$; α_t为该趟航班风险指标 E_t 的取值; Q_1 为呈偏态分布指标的 25% 分位数; Q_2 为中位数; Q_3 为 75% 分位数。

2.3　超限风险评估模型构建

云模型理论是处理不确定性决策问题时描述系统模糊性和随机性, 实现系统评价定性与定量之间的相互转化映射的理论[25]。由于超限事件相关参数数据具有模糊性、随机性特点, 为使得评价结果更加客观合理, 研究采用云模型[26]对超限风险进行评估。具体计算如下。

1) 确定云模型参数。设定云模型期望值为该风险等级的中心值。同时, 由于指标等级区间存在上边界值 C_{L}、下边界值 C_{R} 的约束, 且边界值是相邻不同等级间的过渡值, 属于 2 种不同的风险状态, 因此, 其隶属于 2 个风险状态等级的不确定度相同。云模型参数计算式为

$$\begin{cases}E_{\mathrm{x}}=(C_{\mathrm{L}}+C_{\mathrm{R}})/2\\ E_{\mathrm{n}}=(C_{\mathrm{R}}-C_{\mathrm{L}})/2.355\\ H_{\mathrm{e}}=0.001\end{cases}\quad(15)$$

式中: E_{x} 为期望, 表示论域的中心值; E_{n} 为熵, 表示论域空间可被度量的范围; H_{e} 为超熵, 表示熵的不确定性。

2) 确定风险评价矩阵。根据咨询通告 AC-121/135-FS-2012-45R1[3], 结合过往研究文献将着陆阶段超限风险划分为“无”、“轻微”、“低”、“中”和“高” 5 个等级, 通过云模型将超限风险评估指标量化, 分别计算已选取指标对应的云模型特征参数, 由于正态分布的数据分布特性表明, 大部分数据集中在期望值附近的标准差范围内, 并且约 99.7% 的数据分布在期望值附近 3 个标准差范围内, 因此, 等级区间划分如表 2 所示。

表 2　云模型特征参数计算

Table 2　Calculation of cloud model parameter characteristics

特征参数	等级区间				
	$[0, x_1]$	$[x_1, x_2]$	$[x_2, x_3]$	$[x_3, x_4]$	$[x_4, 3]$
E_{x}	0	$\dfrac{x_1+x_2}{2}$	$\dfrac{x_2+x_3}{2}$	$\dfrac{x_3+x_4}{2}$	3
E_{n}	$\dfrac{x_1-0}{2.355}$	$\dfrac{x_2-x_1}{2.355}$	$\dfrac{x_3-x_2}{2.355}$	$\dfrac{x_4-x_3}{2.355}$	$\dfrac{3-x_4}{2.355}$
H_{e}	0.001	0.001	0.001	0.001	0.001

将指标严重度值代入式(16)计算各指标的风险等级隶属度 $\boldsymbol{r}^*=[\boldsymbol{r}_1,\boldsymbol{r}_2,\cdots,\boldsymbol{r}_A]$。

$$\begin{cases}E_{\mathrm{n}}^*=\mathrm{NORM}(E_{\mathrm{n}},H_{\mathrm{e}})\\ r=\exp\left[-(x^*-E_{\mathrm{x}})^2/2\left(E_{\mathrm{n}}^*\right)^2\right]\end{cases}\quad(16)$$

式中: E_{n}^*表示以 E_{n}为期望值; H_{e}为标准差的正态随机数; x^*为具体参数值。

对各指标风险等级隶属度向量进行组合, 得到风险评价矩阵 $\boldsymbol{R}$, 其中, A 为风险评估指标个数。

$$\boldsymbol{R}=\begin{bmatrix}\boldsymbol{r}_1\\ \boldsymbol{r}_2\\ \vdots\\ \boldsymbol{r}_A\end{bmatrix}=\begin{bmatrix}r_{11} & r_{12} & \cdots & r_{15}\\ r_{21} & r_{22} & \cdots & r_{25}\\ \vdots & \vdots & & \vdots\\ r_{A1} & r_{A2} & \cdots & r_{A5}\end{bmatrix}\quad(17)$$

最后, 结合 3.1 节所计算风险指标发生可能性指标权重向量, 与风险评价矩阵 $\boldsymbol{R}$ 相乘得出单次航班着陆超限风险隶属度结果 $\boldsymbol{W}=[w_1, w_2, w_3, w_4, w_5]$, 根据评估结果中的 $w_{\max}$ 所对应风险等级作为最终的风险等级。

3　模型验证与应用

本文使用国内某航空公司 B737-800 机型 2014—2016 年全航段超限事件信息, 数据包括航班日期、航班号、超限项目、超限值、超限监控项目阈值等。通过飞行品质分析软件导出所需关键飞行参数, 最终获得 9 799 个有效航段。基于数据约束, 对同时存在 2 个及 2 个以上超限事件发生的航段数据通过 Apriori 算法进行统计挖掘。由于超限事件项目较多且发生概率较小, 本文将支持度阈值设定为 0.01, 置信度设定为 0.2, 表 3 为进近着陆阶段强关联规则, $A_7\rightarrow E_3$表示这 2 项超限事件间存在关联性, 前项的发生可能会增大后项发生的概率。

依据式(4)～式(7)对指标可能性进行赋权, 采

表3 超限监控项目间关联规则

Table 3 Association rules among exceedance events

强关联规则	支持度	置信度	提升度
$A_7 \rightarrow E_3$	0.014	0.44	3.49
$A_5 \rightarrow E_1$	0.010	0.56	3.05
$A_1 \rightarrow E_1$	0.010	0.50	2.74
$A_5 \rightarrow E_2$	0.014	0.78	2.63
$E_3 \rightarrow E_4$	0.031	0.25	1.85
$E_4 \rightarrow E_3$	0.031	0.23	1.85
$A_2 \rightarrow E_1$	0.012	0.24	1.32
$E_4 \rightarrow E_5$	0.078	0.58	1.01

用PageRank算法确定各指标权重，最终归一化后的超限风险评价指标可能性权值如表4所示。

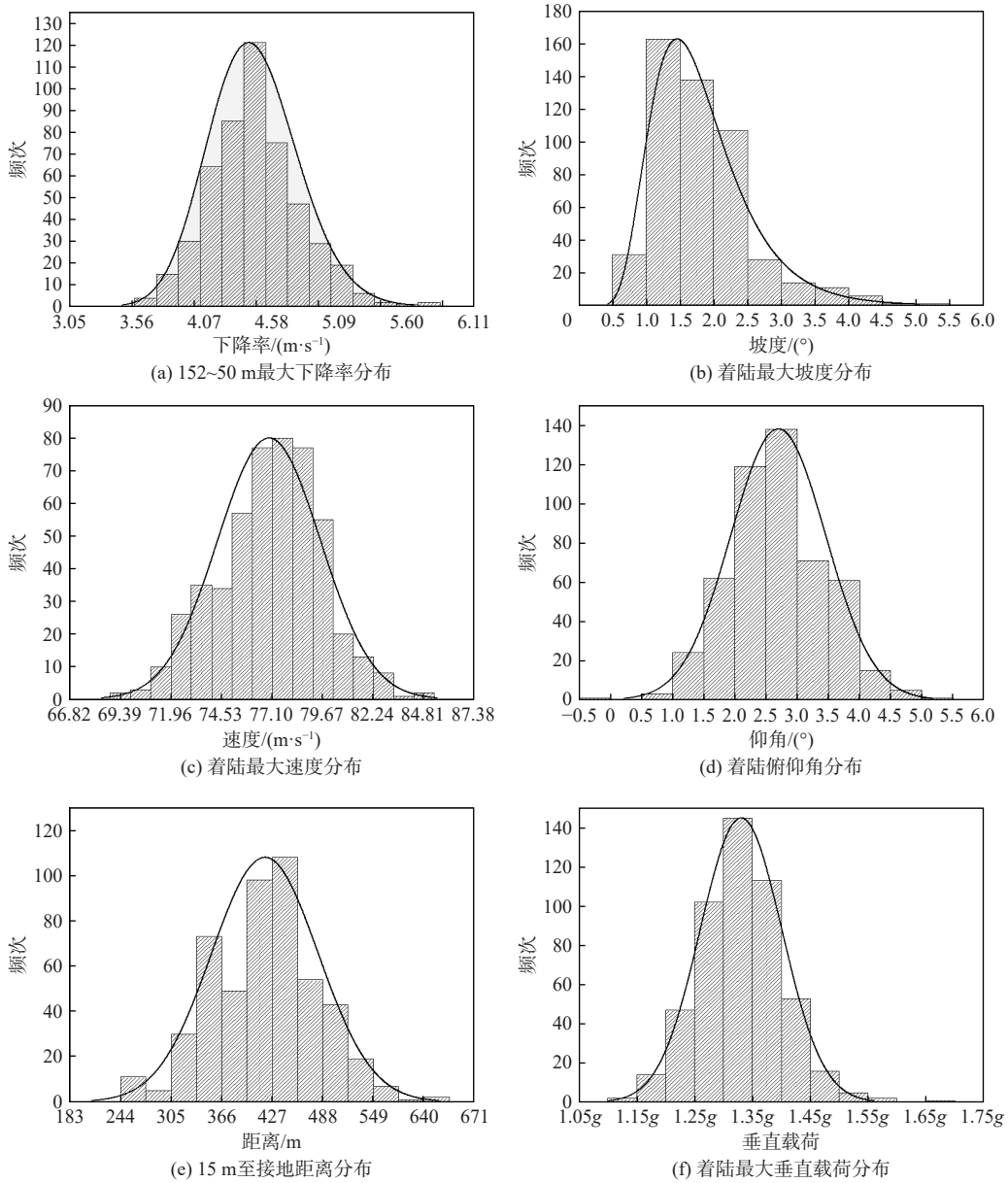
为验证所选指标分布，从某航司B737-800机队采集该机型500套QAR数据，通过python编程处理得到所选风险指标对应参数数据，其分布如图3所示。

经检验，152～15 m(500～50 ft)下降率最大值、着陆阶段速度最大值、着陆俯仰角、15 m(50 ft)至接地距离及着陆垂直载荷均符合正态分布，通过式(8)

表4 风险指标可能性权重

Table 4 Weight of risk indicator probability

指标	超限事件名称	可能性权值
E_1	152～15 m(500～50 ft)下降率大	0.131 0
E_2	着陆坡度大	0.048 0
E_3	着陆速度大	0.114 3
E_4	着陆俯仰角小	0.131 0
E_5	15 m(50 ft)至接地距离远	0.329 4
E_6	着陆垂直载荷大	0.246 3



(a) 152~50 m最大下降率分布 (b) 着陆最大坡度分布

(c) 着陆最大速度分布 (d) 着陆俯仰角分布

(e) 15 m至接地距离分布 (f) 着陆最大垂直载荷分布

图3 指标分布

Fig. 3 Distribution of indicators

和式(10)计算其期望和标准差，然后，依据式(11)和式(12)计算 $S_E(i)$ 值，具体结果如表 5 所示。

着陆阶段最大坡度样本经检验符合正偏态分布，计算得其 Q_1=1.23, Q_2=1.67, Q_3=2.11,依据式(13)可得

$$S_E(2)=\frac{\alpha_2-1.67}{2.11-1.23}=\frac{\alpha_2-1.67}{0.88}$$

以此方法对 A 航司 10 个航段的着陆超限风险进行评价，具体评估结果如表 6 所示。

表 5　风险指标严重度计算

Table 5　Calculation of risk indicator severity

风险指标	样本参数名称	期望	标准差	严重度计算公式
E_1	152～15 m(500～50 ft)下降率最大值	896.58	72.51	$S_E(1)=(\alpha_1-896.58)/72.51$
E_3	着陆阶段速度最大值	149.70	5.11	$S_E(3)=(\alpha_3-149.70)/5.11$
E_4	着陆最小俯仰角	2.70	0.76	$S_E(4)=(2.70-\alpha_4)/0.76$
E_5	15 m(50 ft)至接地距离	1 371.45	211.62	$S_E(5)=(\alpha_5-1\,371.45)/211.62$
E_6	着陆垂直载荷最大值	1.330 7	0.07	$S_E(6)=(\alpha_6-1.330\,7)/0.07$

表 6　样本风险等级评估结果

Table 6　Results of samples risk level assessment

样本序号	风险等级隶属度w_i					风险等级
	无	轻微	低	中	高	
1	0.839 6	0.102 8	0.022 1	0.026 0	1.868×10^{-12}	无风险
2	0.499 7	0.503 8	0.040 0	1.382×10^{-5}	1.490×10^{-42}	轻微风险
3	0.622 7	0.132 5	0.297 2	0.007 9	1.226×10^{-21}	无风险
4	0.700 4	0.095 6	0.238 4	0.008 2	9.208×10^{-23}	无风险
5	0.444 1	0.213 6	0.379 6	0.014 2	5.881×10^{-20}	无风险
6	0.293 3	0.372 6	0.390 8	0.024 6	7.196×10^{-20}	低风险
7	0.858 7	0.153 9	0.039 7	6.527×10^{-5}	1.307×0^{-35}	无风险
8	0.670 6	0.059 7	0.328 8	0.023 7	5.462×10^{-20}	无风险
9	0.380 5	0.442 0	0.141 2	0.036 9	1.311×10^{-14}	轻微风险
10	0.773 5	0.134 6	0.000 1	7.926×10^{-10}	2.425×10^{-66}	无风险

经计算，A 航司 10 个航段的超限风险评价结果大多为“无风险”等级，样本 2 评价结果为“轻微风险”，其 15 m(50 ft)至接地距离、着陆最小仰角及着陆最大坡度值均存在轻微偏移；样本 6 风险等级为“低风险”，其 15 m(50 ft)至接地距离值在正态分布中偏移较大，该指标是导致风险等级为“低风险的主要原因；样本 9 为“轻微风险”等级的原因是其 15 m(50 ft)至接地距离、152～15 m(500～50 ft)最大下降率、着陆最大坡度、着陆最大速度指标同时存在轻微的偏移。航空公司应对频繁出现“低”及以上等级航段的飞行员加以关注，当存在风险等级达到“高”时采取相应措施以降低风险。

本文方法未来可推广到其他飞行阶段的超限风险评估及飞行员个体的超限风险评估研究中。通过筛选出相关飞行阶段的关联规则，得到超限风险传递关系；然后，计算风险指标的可能性权值与偏移严重程度；最终，通过云模型进行评估得到航段在该飞行阶段的超限风险等级。进一步推广至飞行员个体的超限风险评估中，可对全飞行阶段的关联规则进行整理筛选，进行单次航班的超限风险评估，然后，对一段时间内该飞行员的多趟航班评价结果求均值，即可完成对飞行员个体的超限风险评估。

4　结　论

1）本文通过关联规则挖掘超限事件之间的相关性从而计算指标可能性权值，使其在风险评估模型中考虑着陆阶段的整体可能性而不是单个超限事件发生的可能性。

2）基于各指标在大样本空间下的分布，本文采用样本基于标准值的偏移量来评估严重性风险，消除传统超限标准的非普适性。

3）本文构建了有效的着陆超限风险定量评估模型，结合飞行数据分布，能够进一步识别潜在的超限风险，有利于干预和纠正。

本文方法今后可推广到其他飞行阶段的超限风险评估及飞行员个体超限风险等级评价。

参考文献（References）

[1] AIRBUS. A statistical analysis of commercial aviation accidents 1958-2022: X00D17008863[R]. Blagnac Cedex: AIRBUS, 2023: 27.

[2] Boeing Commercial Airplanes. Statistical summary of commercial jet airplane accidents worldwide operations 1959-2021: 23072954 [R]. Seattle: Boeing Commercial Airplanes, 2022: 10.

[3] 中国民用航空局. 飞行品质监控 (FOQA) 实施与管理: AC-121/135-FS-2012-45R1[R]. 北京: 中国民用航空局飞行标准司, 2015: 21-25.
Civil Aviation Administration of China. Flight quality assurance (FOQA) implementation and management: AC-121/135-FS-2012-45R1[R]. Beijing: Flight Standards Department of Civil Aviation Administration of China, 2015: 21-25(in Chinese).

[4] WANG L, GAO S, HONG R Y, et al. Effects of age and flight exposure on flight safety performance: evidence from a large cross-sectional pilot sample[J]. Safety Science, 2023, 165: 106199.

[5] AYRA E S, RÍOS INSUA D, CANO J. Bayesian network for managing runway overruns in aviation safety[J]. Journal of Aerospace Information Systems, 2019, 16(12): 546-558.

[6] 孙瑞山, 陈雄, 梁妍. 基于高斯混合模型的飞机进近着陆阶段运行异常检测[J]. 安全与环境学报, 2022, 22(3): 1371-1376.
SUN R S, CHEN X, LIANG Y. Abnormal operation detection for approach and landing phase based on Gaussian mixture model[J]. Journal of Safety and Environment, 2022, 22(3): 1371-1376(in Chinese).

[7] 祁明亮, 邵雪焱, 池宏. QAR 超限事件飞行操作风险诊断方法[J]. 北京航空航天大学学报, 2011, 37(10): 1207-1210.
QI M L, SHAO X Y, CHI H. Flight operations risk diagnosis method on quick-access-record exceedance[J]. Journal of Beijing University of Aeronautics and Astronautics, 2011, 37(10): 1207-1210 (in Chinese).

[8] HADJIMICHAEL M, OSBORNE D M, ROSS D, et al. The flight operationsriskassessmentsystem[J].SAEInternational,1999:150-156.

[9] 汪磊, 孙瑞山, 吴昌旭, 等. 基于飞行 QAR 数据的重着陆风险定量评价模型[J]. 中国安全科学学报, 2014, 24(2): 88-92.
WANG L, SUN R S, WU C X, et al. A flight QAR data based model for hard landing risk quantitative evaluation[J]. China Safety Science Journal, 2014, 24(2): 88-92(in Chinese).

[10] 陈农田, 满永政, 李俊辉. 基于 QAR 数据的民机高高原进近着陆风险评估方法[J]. 北京航空航天大学学报, 2024, 50(1): 77-85.
CHEN N T, MAN Y Z, LI J H. Risk assessment method for civil aircraft approach and landing at high plateau based on QAR data [J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(1): 77-85(in Chinese).

[11] 汪磊, 郭世广, 任勇. 基于飞行数据正态云的着陆操作风险评价方法[J]. 安全与环境学报, 2019, 19(5): 1555-1561.
WANG L, GUO S G, REN Y. Landing operation risk evaluation based on the normal cloud of the flight data[J]. Journal of Safety and Environment, 2019, 19(5): 1555-1561(in Chinese).

[12] 汪磊, 高杉, 张静怡, 等. 民航飞行员超限行为评价方法研究[J]. 安全与环境学报, 2021, 21(2): 695-700.
WANG L, GAO S, ZHANG J Y, et al. Of the airline transport pilots based on the QAR data Evaluation of the exceedance behaviors[J]. Journal of Safety and Environment, 2021, 21(2): 695-700 (in Chinese).

[13] WANG L, REN Y, WU C X. Effects of flare operation on landing safety: a study based on ANOVA of real flight data[J]. Safety Science, 2018, 102: 14-25.

[14] 谢嘉仪. 基于 QAR 飞行大数据的不稳定进近风险分析与预警技术研究[D]. 武汉: 武汉大学, 2021: 14-17.
XIE J Y. Research on unstable approach risk analysis and early warning technology based on QAR flight big data[D]. Wuhan: Wuhan University, 2021: 14-17(in Chinese).

[15] MONTELLA A. Identifying crash contributory factors at urban round abouts and using association rules to explore their relationships to different crash types[J]. Accident Analysis & Prevention, 2011, 43(4): 1451-1463.

[16] LI L, GUO H M, CHENG L H, et al. Research on causes of coal mine gas explosion accidents based on association rule[J]. Journal of Loss Prevention in the Process Industries, 2022, 80: 104879.

[17] LAN H, MA X X, MA L H, et al. Pattern investigation of total loss maritime accidents based on association rule mining[J]. Reliability Engineering & System Safety, 2023, 229: 108893.

[18] 时统宇, 王岩韬, 李鸿坤. 基于先验频繁模式算法的航班延误关联规则挖掘[J]. 飞行力学, 2023, 41(2): 87-94.
SHI T Y, WANG Y T, LI H K. Flight delay association rules mining based on Apriori algorithm[J]. Flight Dynamics, 2023, 41(2): 87-94(in Chinese).

[19] 袁乐平, 张文东, 赵力梵, 等. 基于 TEM 模型的民航不安全事件关联规则挖掘与分析[J]. 中国安全生产科学技术, 2022, 18(10): 31-36.
YUAN L P, ZHANG W D, ZHAO L F, et al. Mining and analysis of association rules in civil aviation unsafe incidents based on TEM model[J]. Journal of Safety Science and Technology, 2022, 18(10): 31-36(in Chinese).

[20] 王明涛. 证券投资风险本质属性探讨及其计量模型研究[J]. 河南金融管理干部学院学报, 2003, 21(4): 61-63.
WANG M T. An inquiry into essential attributes of securities investment risks and its quantitative pattern research[J]. Journal of Henan College of Financial Management Cadres, 2003, 21(4): 61-63 (in Chinese).

[21] AGRAWAL R, IMIELIŃSKI T, SWAMI A. Mining association rules between sets of items in large databases[J]. ACM SIGMOD Record, 1993, 22(2): 207-216.

[22] 毕建欣, 张岐山. 关联规则挖掘算法综述[J]. 中国工程科学, 2005, 7(4): 88-94.
BI J X, ZHANG Q S. Survey of the algorithms on association rule mining[J]. Strategic Study of CAE, 2005, 7(4): 88-94(in Chinese).

[23] 蔡伟杰, 张晓辉, 朱建秋, 等. 关联规则挖掘综述[J]. 计算机工程, 2001, 27(5): 31-33.
CAI W J, ZHANG X H, ZHU J Q, et al. Survey of association rule generation[J]. Computer Engineering, 2001, 27(5): 31-33(in Chinese).

[24] BRIN S, PAGE L. The anatomy of a large-scale hypertextual web search engine[J]. Computer networks and ISDN systems, 1998, 30(1-7): 107-117.

[25] 李德毅, 杜鹢. 不确定性人工智能[M]. 2 版. 北京: 国防工业出版社, 2014: 42-43.
LI D Y, DU Y. Artificial intelligence with uncertainty[M]. 2nd ed.

Beijing: National Defense Industry Press, 2014: 42-43(in Chinese).
[26] 叶琼, 李绍稳, 张友华, 等. 云模型及应用综述[J]. 计算机工程与设计, 2011, 32(12): 4198-4201.
YE Q, LI S W, ZHANG Y H, et al. Cloud model and application overview[J]. Computer Engineering and Design, 2011, 32(12): 4198-4201(in Chinese).

Risk assessment of landing phase exceedance based on QAR data and association rule

WANG Lei*, ZHANG Nan, GAO Shan

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

Abstract: The monitoring and analysis of exceedance events within the flight operation quality assurance (FOQA) program is one of the crucial ways to identify potential risks in flights. However, existing methods lack comprehensive consideration for the interrelationships between various exceedance events, thereby failing to systemically quantify the risks associated with several interdependent events. This study introduces a risk assessment method for landing exceedance events by integrating association rule mining and quick access recorder (QAR) data. Firstly, based on the frequency of exceedance events and the findings of an expert survey conducted during the approach and landing phases of 9 799 flight segments, six exceedance events were found to be indicators of risk assessment: high normal acceleration at landing, low pitch at landing, long landing, excessive bank angle below 50 m, high speed below 50 m, and high descent rate (152～15 m). Secondly, we calculated the possibility and severity of the six indicators based on their correlations and deviations, respectively. Subsequently, a risk assessment model for landing was constructed using the cloud model. Lastly, the real QAR data were used to verify the model. The results can differentiate risk levels associated with specific exceedance events. In order to improve flight safety supervision and operational efficiency, this model offers a useful method for evaluating the exceedance risk during the landing phase.

Keywords: QAR data; association rule; landing exceedance; PageRank algorithm; risk assessment

Received: 2023-06-21; **Accepted**: 2023-09-08; **Published Online**: 2024-06-18 14:56
URL: link.cnki.net/urlid/11.2625.V.20240617.1513.002
Foundation item: Safety Capacity Building Fund Project of Civil Aviation Administration of China (KJZ49420210076)
* **Corresponding author**. E-mail: wanglei0564@hotmail.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0790

基于内聚力模型的透波罩黏接强度性能研究

侯保江[1,2,*]，王姣[2]，邢誉峰[1]，张飞[2]，李焱喜[2]

(1. 北京航空航天大学 航空科学与工程学院，北京 100191；2. 北京航天长征飞行器研究所，北京 100076)

摘 要：透波罩与飞行器本体往往通过胶黏+辅助连接的方式进行连接，在实际工程应用中，通常将线弹性本构作为胶黏剂的本构模型进行分析，该方法难以对相关黏接结构的力学行为进行准确模拟。基于此，采取常用的双线性内聚力本构模型，通过胶黏剂实际性能的测试结果对本构模型进行校正，采用平板模拟实际黏接结构的性能试验进一步验证复杂受力环境下模型的有效性，并将验证后的本构模型用于分析实际整机级透波罩的黏接强度，并与实际试验结果进行对比，验证双线性内聚力本构模型的计算精度，为类似结构黏接强度的准确预测提供有效途径。

关 键 词：内聚力；双线性；透波罩；黏接强度；数值仿真

中图分类号：V221+.3; TB553

文献标志码：A **文章编号**：1001-5965(2025)06-1916-10

透波罩是飞行器上的重要结构部件，位于飞行器最前端，形状多为半球形和锥形，其具有防热、透波、承载和抗蚀等多种功能，在风雪、沙尘和太阳强辐射等恶劣环境条件下也能确保飞行器的通信系统、遥测系统、制导系统和引爆系统正常工作[1]。由于透波罩材料本身的特性，其无法实现与飞行器其他舱段的直接机械连接，往往通过金属连接环转接。透波罩与连接环之间通过高温胶黏剂+非金属销钉的方式进行连接[2]。在实际工程实践中，采用较为简单的方式评估透波罩的黏接强度，即在Abaqus有限元平台中将胶进行实体建模，胶与2种黏接介质采用“Tie”约束，并用线弹性本构进行计算，但强度计算结果与实际承载能力有较大差距。随着飞行器技术的发展，飞行器在大气层内飞行的时间越来越长，黏接部位的温度和载荷环境越来越严酷，使得该结构在高温、高载条件下承载能力的设计愈加困难，急需对黏接强度开展精确的理论预示，充分挖潜承载性能，为实际飞行器中透波罩的黏接部位的结构强度工程化设计提供有力的手段和途径。

对胶黏剂失效的研究最早可以追溯到对复合材料层间破坏现象，特别是无堆叠方向增强(如针刺缝合、Z-pin等)的复合材料层间剥离现象的研究。对失效过程的研究往往分成2部分，即损伤起始点的判断和损伤发展过程的分析。对损伤起始点的判断最传统的方式是基于应力或应变与特征长度等进行组合形成的特征量的判断[3-4]。

在预测损伤发展方面，比较常用的方法有2种，即线弹性断裂力学(linear elastic fracture mechcanics, LEFM)方法[5-6]与内聚力模型(cohesive zone model, CZM)方法。LEFM方法中最常用的也最成熟的是虚拟裂纹弥合技术(virtual crack closure technique, VCCT)[7-10]。在有限元计算中，VCCT较节省计算量，但该技术存在2个缺陷：①为保证计算精度，需要在裂纹尖端划分十分精细的网格；②需要事先确定初始裂纹，在特定情况下，结构初始裂纹位置的确定十分困难，而错误地选定初始裂纹位置可能导致计算结果与实际情况之间存在较大偏差。

CZM方法具有2个显著优点：①该模型为分层

收稿日期：2023-12-04；录用日期：2024-03-11；网络出版时间：2024-04-15 11:35
网络出版地址：link.cnki.net/urlid/11.2625.V.20240412.1238.002
*通信作者. E-mail：153327058@qq.com

引用格式：侯保江，王姣，邢誉峰，等. 基于内聚力模型的透波罩黏接强度性能研究[J]. 北京航空航天大学学报，2025，51(6)：1916-1925.
HOU B J，WANG J，XING Y F，et al. Performance study on bonding strength of radome based on cohesive zone model[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1916-1925 (in Chinese).

的萌生与扩展提供了一种统一的描述, 克服了断裂力学方法不能用于预测新裂纹萌生的困难; ②当表征断裂过程区的尺度参数接近于材料或结构的特征几何长度 (如裂纹长度) 时, 线弹性断裂力学的方法不再有效, 而黏聚区模型则不受此限制。此外, 使用界面元概念很容易实现基于 CZM 的有限元列式, 并易于嵌入商用有限元软件中。CZM 方法将材料分成连续体及连续体之间的黏聚层, 层间失效由被内聚力单元连接的 2 层节点间的相对位移来表征, 因此, 损伤的萌生与分层的扩展仅由内聚力单元描述即可。

本文采取常用的双线性内聚力本构模型[11-16], 通过胶黏剂实际性能测试结果修正本构模型, 并采用实际黏接结构的平板模型的性能试验结果进一步修正本构模型, 将修正后的本构模型用于分析实际整机级透波罩的黏接强度, 并与实际试验结果进行对比, 二者的一致性证明了修正后本构模型的实际计算精度。

1　内聚力模型

1.1　牵引力-位移曲线

CZM 方法在有限元分析中的应用主要通过内聚力单元实现。内聚力单元一般分为 2 种: 连续面间内聚力单元与点间内聚力单元。其中, 连续面间的内聚力单元可以存在有限厚度与零厚度 2 种, 而点间内聚力单元基本等价于非线性弹簧单元。以上单元形式中, 由于零厚度单元不影响整体结构相对尺寸, 也不会使仿真与实际时间存在几何差异, 在界面剥离尤其是胶层失效分析中具有显著优势。

内聚力单元的材料本构主要参数由牵引力-位移曲线、裂纹萌生判据和混合模式裂纹拓展模型共同确定。

牵引力-位移曲线主要表征材料在单一模式裂纹下的损伤模式, 也可以看作内聚力单元的单向本构。常见的内聚力模型有多项式型[17]、指数型[18-19]、三线型(梯形)[20] 与双线型[21-23]。通过试验对比发现, 三线型内聚力模拟的延展性黏接结构精度较高, 但双线型内聚力模型也能达到精度要求, 且在有限元软件中易于实现和收敛较快, 因此, 目前广泛采用的内聚力模型为双线型模型。双线性模型的牵引力 T 与位移 δ 的关系如图 1 所示。

临界断裂能的表达式为

$$G_{\mathrm{c}}=\int_0^{\delta_{\mathrm{c}}} T(\delta)\mathrm{d}\delta \tag{1}$$

其牵引力-位移关系如下:

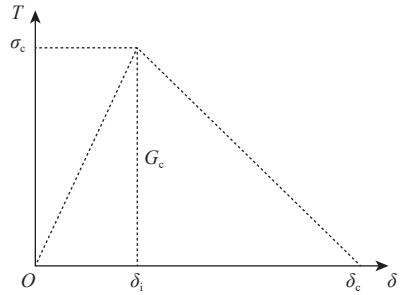


图 1　双线型模型牵引力-位移曲线

Fig. 1　Traction-displacement curve of bilinear model

$$T(\delta)=\begin{cases}\dfrac{\delta}{\delta_{\mathrm{i}}}\sigma_{\mathrm{c}} & \delta\leqslant\delta_{\mathrm{i}}\\ \dfrac{\delta_{\mathrm{c}}-\delta}{\delta_{\mathrm{c}}-\delta_{\mathrm{i}}}\sigma_{\mathrm{c}} & \delta>\delta_{\mathrm{i}}\end{cases} \tag{2}$$

式中: T 为牵引力; δ 为裂纹尖端分离位移; 峰值 σ_{c} 为材料的内聚力强度(下文称黏聚强度); δ_{i} 为峰值 σ_{c} 对应的位移; δ_{c} 为内聚力为 0 时对应的位移。图 1 中, G_{c} 为曲线下方的面积, 即临界能量释放率。

由图 1 与式 (1) 和式 (2) 可知, 内聚力本构中最关键的参数主要有初始刚度、黏聚强度和临界能量释放率(断裂韧性), 分别为牵引力-位移曲线的上升段斜率、峰值和曲线下面积, 这 3 个参数可以完全确定双线型模型牵引力-位移曲线的形状。

通常情况下, 胶黏剂的初始刚度值由胶黏界面的材料刚度与尺寸确定, 但由于胶黏层厚度趋近于零, 导致胶黏层的初始刚度趋近于无限大, 处理方式一般为选取一个较大的数值作为初始刚度, 使之不至于过大导致计算不收敛, 也不至于过小导致软化过程过短或没有软化过程[24]。已经有多项研究表明, 初始刚度的变化会对黏接强度的模拟精度产生较大的影响[25-27], 特别是在存在混合模式的界面失效情况下[28-30]。

临界能量释放率可由标准的断裂韧性试验测得, 而黏聚强度一般可通过拉剪、压剪、拉离试验获取。胶接结构的主要失效形式与传统断裂力学问题一致, 可以概括为Ⅰ型、Ⅱ型、Ⅲ型裂纹及三者的组合。在有限元计算中, Ⅱ型与Ⅲ型 2 种状态可以统一为一种剪切状态, 只用剪切面上不同方向表征面内各向异性。单一型断裂的双线型本构关系可用图 2 表示[31]。其中, k 为界面的初始刚度, k_{p} 为线性退化刚度, δ_{i}^{0} 和 $\delta_{\mathrm{i}}^{\mathrm{f}}$ 分别为界面发生初始损伤和完全破坏时的相对位移, d 为与界面相对位移相关的损伤变量, N、S_1、S_2 分别为法向黏聚强度、纵向剪切黏聚强度、横向剪切黏聚强度, G_{IC}、G_{IIC} 和 G_{IIIC} 分别为纯拉伸型分层(Ⅰ型)、纵向剪切型分层(Ⅱ型)和横向剪切型分层(Ⅲ型)的临界能量释放率(即材料临界断裂能)。

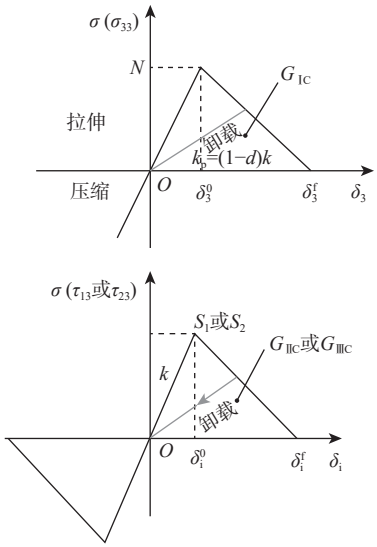


图2 单一型断裂双线性本构关系[31]

Fig. 2 Bilinear constitutive relation of single-mode fracture[31]

1.2 混合模式裂纹拓展

复杂结构中，常见的裂纹形式是混合型裂纹，胶接界面在实际受力中也往往产生混合型断裂。以能量准则作为判断混合模式失效的标准假定，当能量释放率G超过临界能量释放率G_c时，界面分层开始扩展。最广泛使用的预示混合载荷模式下界面分层扩展的准则是根据能量释放率间的线性或二次方关系建立的幂指数型准则[32]，其形式如下：

$$\left(\frac{G_{\mathrm{I}}}{G_{\mathrm{IC}}}\right)^{\alpha}+\left(\frac{G_{\mathrm{shear}}}{G_{\mathrm{shearC}}}\right)^{\alpha}=1 \tag{3}$$

式中：$G_{\mathrm{shear}}=G_{\mathrm{II}}+G_{\mathrm{III}}$；$G_{\mathrm{shearC}}=G_{\mathrm{IIC}}+G_{\mathrm{IIIC}}$；$G_{\mathrm{I}}$、$G_{\mathrm{II}}$、$G_{\mathrm{III}}$分别为Ⅰ型、Ⅱ型、Ⅲ型裂纹扩展时的能量释放率；$\alpha$为材料常数。

试验表明，指数型准则对热塑性PEEK材料具有较好的适应性，但该准则不适用于环氧类胶黏剂。

针对环氧类胶黏剂，Benzeggagh和Kenane于1996年提出对其适应性较好的B-K准则[33]，该准则表达式如下：

$$G_{\mathrm{mC}}=G_{\mathrm{IC}}+(G_{\mathrm{IIC}}-G_{\mathrm{IC}})\left(\frac{G_{\mathrm{shear}}}{G_{\mathrm{I}}+G_{\mathrm{shear}}}\right)^{\eta} \tag{4}$$

式中：G_{mC}为混合型断裂的等效临界能量释放率；η为与材料相关的常数。

基于此，可以进一步定义损伤演化变量D为

$$D=\frac{G_{\mathrm{I}}+G_{\mathrm{shear}}}{G_{\mathrm{mC}}} \tag{5}$$

当D=1时，胶接界面破坏，进而导致裂纹扩展。

1.3 初始失效判据

在单一型断裂的本构关系中，失效过程开始发生的判据可以直接用界面临界应力σ_{33}^{c}、τ_{13}^{c}和τ_{23}^{c}表示，即应力超过临界应力时，开始发生失效。混合模式的裂纹中，必须考虑不同种模式裂纹之间的耦合。较常用的是Ye在1988年提出的二次名义应力准则[34]，其表达式如下：

$$\left(\frac{\langle\sigma_{33}\rangle}{\sigma_{33}^{\mathrm{c}}}\right)^{2}+\left(\frac{\tau_{13}}{\tau_{13}^{\mathrm{c}}}\right)^{2}+\left(\frac{\tau_{23}}{\tau_{23}^{\mathrm{c}}}\right)^{2}=1 \tag{6}$$

式中：σ_{33}、τ_{13}、τ_{23}分别为各方向独自承载时的名义应力；〈〉为Macaulay括号，含义为$\langle x\rangle=\max(x,0)$。

2006年，Turon等提出了一种结合B-K准则的初始失效判据，但该判据与二次应力准则偏差不大[13]。

2 模型验证及参数确定

2.1 胶黏剂材料性能

基于工程研制需求，针对胶黏剂的高温、高强度目标，攻关阶段共研制了3个系列：PFE系列（酚醛改性）、EA系列（酚醛改性）、CYH系列（硅改性）环氧基体胶黏剂。最终，综合考虑性能和工艺性，选定CYH系列环氧树脂为胶黏剂基体。

对新研耐高温、高强度环氧胶黏剂全面进行性能表征发现，室温拉剪强度均在13 MPa以上，−40 ℃拉剪强度达15 MPa以上，250 ℃拉剪强度为8.08～8.74 MPa，−40 ℃拉离强度在49～52 MPa之间，室温拉离强度均在56 MPa以上，250 ℃拉离强度达12.82 MPa，满足型号使用要求。

2.2 模型验证

为验证内聚力模型的有效性，开展了胶接界面黏接性能的有限元仿真预示及试验研究。有限元模型中，选择二次名义应力准则（式(6)）作为Cohesive单元的初始损伤准则，损伤演化规律采用基于能量的B-K准则（式(4)）。

基于试片级试验结果，给出了Cohesive单元的模型参数，如表1所示，其中，$G_{\mathrm{IIIC}}=G_{\mathrm{IIC}}$。

表1 Cohesive单元的模型参数

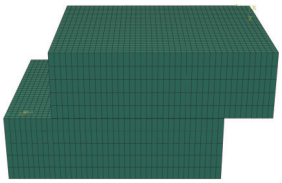
Table 1 Model parameters of Cohesive elements

温度/℃	弹性模量/GPa	G_{IC}/(J·m^{-2})	G_{IIC}/(J·m^{-2})
室温	3.2	78.42	148.99
170	2.6	51.53	91.40
250	1.1	16.39	30.17

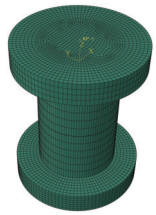
基于室温拉剪、压剪和拉离试验，在Abaqus中分别建立相应的有限元模型，分别进行计算与试验的对比分析，如图3和表2所示。其中，试片材料为铝合金，试片尺寸及加载形式参见相关标准[35-37]。基于胶黏剂性能试验及修正结果，得到适用于环氧



(a) 拉剪模型



(b) 压剪模型



(c) 拉离模型

图 3　胶层性能测试有限元模型

Fig. 3　Finite element model of performance test for adhesive layer

表 2　胶层力学性能参数

Table 2　Mechanical parameters of adhesive layer

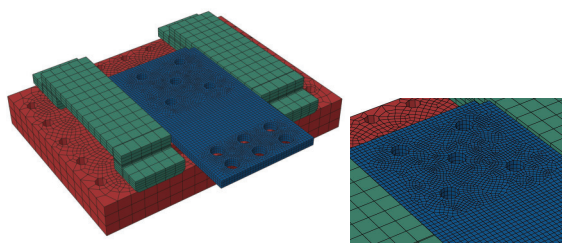
加载方式	强度测试值/MPa							仿真结果/MPa	误差/%
	1#	2#	3#	4#	5#	6#	平均值		
拉剪	15.3	17	20.5	15	14.8	14.4	16.17	17.1	5.75
压剪	74.3	80.9	78.3	66.7	70.2	75.7	74.35	71.08	−4.4
拉离	66.9	51.4	56.4	47.9	36.9	68.5	56.42	53.37	−5.41

类胶黏剂的 B-K 准则常数 η 为 1.39。

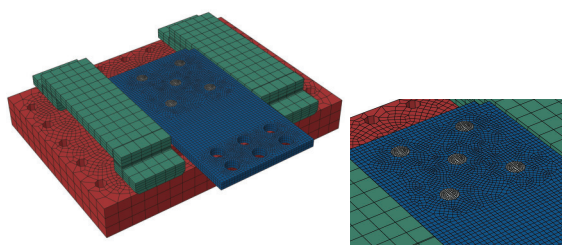
胶层强度实际测试结果与模拟结果对比如表 2 所示, 其中, 1#～6#对应 6 组试验数据。可知, 胶层性能的模拟结果和试验结果吻合较好, 对单向受力状态下胶层性能拟合的相对误差不超过 6%, 验证了本构模型的有效性。

为进一步验证复杂受力环境下模型的有效性, 建立平板连接结构有限元模型进行仿真分析, 并开展平板连接件承载试验进行验证。

图 4 为胶接和胶接+非金属销钉连接 2 种连接状态试验件的有限元模型, 除胶层使用黏聚力单元 COH3D8 外, 其余结构均划分为 8 节点六面体线性减缩积分单元 C3D8R, 线性减缩积分单元计算精度较高, 且用于复合材料结构时便于其方向定义。钢钢板与石英平板分别与胶层之间建立 Tie 约束, 约束工装与石英平板试验件之间建立接触关系, 销钉与钢钢平板之间建立 Tie 约束、与石英平板之间建立接触关系。



(a) 胶接连接状态有限元模型

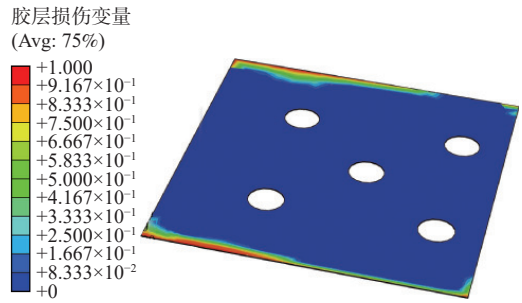


(b) 胶接+销钉连接状态有限元模型

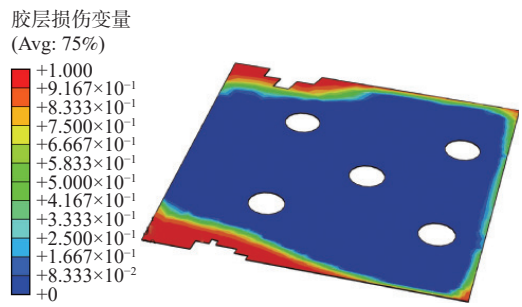
图 4　平板级连接有限元模型

Fig. 4　Finite element model at level of surface plate

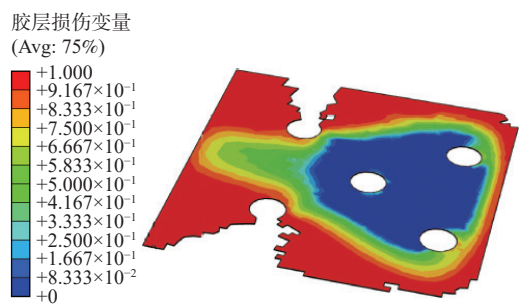
根据室温环境下的仿真结果, 图 5(a)～图 5(c)给出了胶接连接状态的连接失效演化过程。其



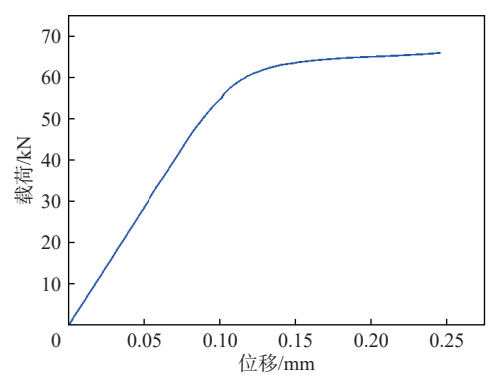
(a) t=0.6



(b) t=0.8



(c) t=0.95

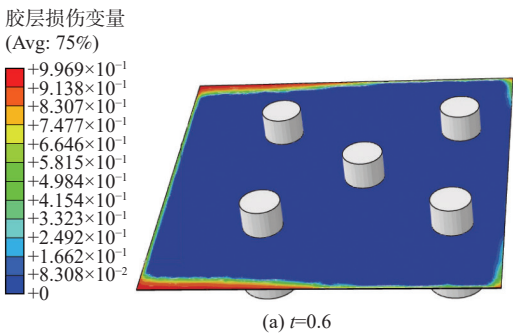


(d) 胶接状态载荷-位移曲线

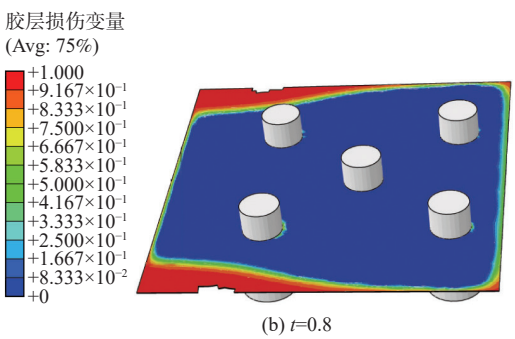
图 5　胶接状态预测结果

Fig. 5　Predicted results of adhesive bonding state

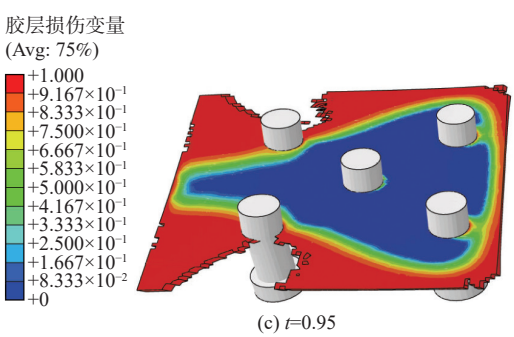
中，胶层损伤变量为0表示黏接界面未发生损伤，胶层损伤变量为1表示完全损伤。图6(a)～图6(c)展示了胶接+销钉连接状态的连接失效演化过程，极限载荷下连接面失效形式与试验件破坏形式一致。图5(d)和图6(d)分别为2种状态的载荷-位移曲线，结构破坏时，2种状态的极限载荷分别为65.8 kN、76.1 kN。图7、图8分别给出了室温环境下试验得到的2种连接状态下的载荷曲线。其中，“设定”指试验控制曲线，“实际”指实际响应



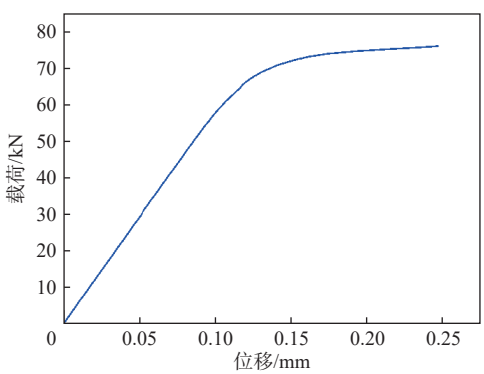
(a) t=0.6



(b) t=0.8



(c) t=0.95



(d) 胶接+销钉状态载荷-位移曲线

图6 胶接+销钉状态预测结果

Fig. 6 Predicted results of adhesive bonding + pin state

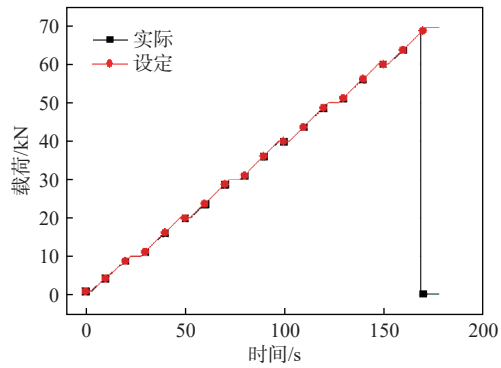


图7 胶接状态试验载荷曲线

Fig. 7 Curves of test load of adhesive bonding state

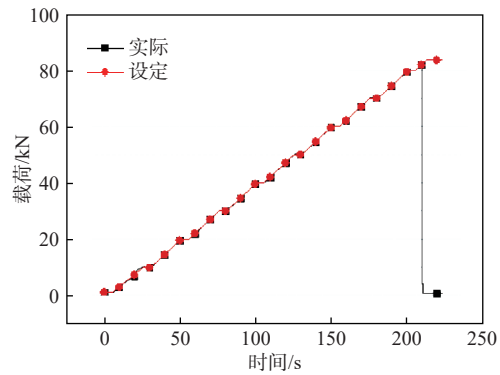


图8 胶接+销钉状态试验载荷曲线

Fig. 8 Curves of test load of adhesive bonding + pin state

曲线。由图得到极限载荷分别为68 kN、82 kN。计算与试验结果的相对误差分别为3.2%、7.2%，一致性较好。

为进一步验证考虑温度效应时分析方法的有效性，对170 ℃、250 ℃工况分别进行了仿真分析，计算与试验结果比较参见表3，最大相对误差为7.2%，计算精度能够满足工程使用要求。

以上结果验证了本文所建内聚力模型的有效性，为进一步研究系统级透波罩承载性能提供了支撑。

表3 极限载荷试验结果与计算结果对比

Table 3 Comparison between experimental and calculated results under ultimate load

连接状态	温度条件	载荷/kN		误差/%
		计算值	试验值	
胶接	室4温	65.8	68	−3.2
胶接+销钉	室温	76.1	82	−7.2
胶接	170 ℃	28.6	30.4	−5.9
胶接+销钉	170 ℃	47.7	50	−4.6
胶接	250 ℃	1.06	1	6
胶接+销钉	250 ℃	22.5	23.3	−3.4

3　系统级透波罩承载性能

某型号透波罩为面对称结构, 透波罩与连接环之间通过高温胶黏剂+非金属销钉的方式进行连接, 如图 9 所示。为适应长时间飞行环境, 其高温胶黏剂采用第 2 节所述的新研耐 250 ℃ 高温的高强度环氧胶黏剂。本文以该型透波罩为研究对象, 基于内聚力本构模型, 进行系统级透波罩连接区极限承载能力评估及仿真预示, 并开展对应的试验研究, 将仿真结果与试验结果进行对比。

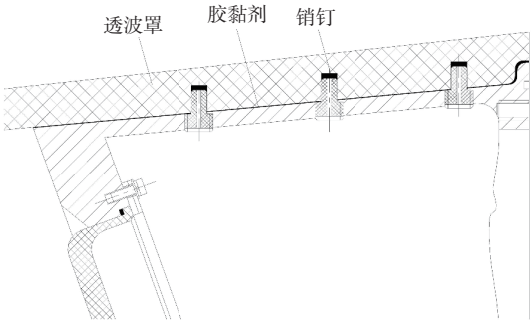


图 9　透波罩连接结构示意图

Fig. 9　Connection structure for radome

建立透波罩 1/2 有限元模型, 如图 10 所示。透波罩与锕钢环之间通过胶接+非金属销钉的方式进行连接, 透波罩和锕钢环分别与胶层之间建立 Tie 约束; 销钉与锕钢环为紧配合状态, 通过 Tie 约束连接; 销钉与透波罩之间建立接触关系。采用常规计算方法时, 胶层与其他结构使用 C3D8R 单元, 本构为简单的线弹性本构。基于内聚力模型时, 胶层使用内聚力单元 COH3D8, 其余结构均使用 C3D8R 单元划分网格。

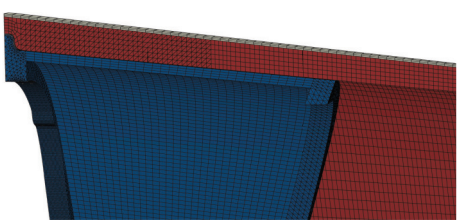


图 10　透波罩有限元模型局部结构

Fig. 10　Local structure of finite element model for radome

3.1　静力载荷工况

对静力载荷工况下透波罩承载能力进行分析, 预测结构达到承载极限时胶层和透波罩本体发生破坏的先后顺序, 评估结构设计的合理性。

图 11 为相应的载荷-位移曲线。根据仿真结果可知, 载荷加载至 369% 最大使用载荷时, 透波罩本体结构发生破坏, 连接区黏接界面完好, 胶层及非金属销钉结构均未发生破坏。由图 11 可以看出, 0%～348% 最大使用载荷时, 位移随载荷线性变化; 348%～369% 最大使用载荷时, 位移随载荷变化曲线曲率发生明显变化, 呈逐渐增大趋势; 369% 最大使用载荷时, 位移发生突变, 整体结构无法继续承载。最大使用载荷时, 结构变形达 3.68 mm, 与试验结果得到的 3.93 mm 误差约为 6.8%, 一致性较好。

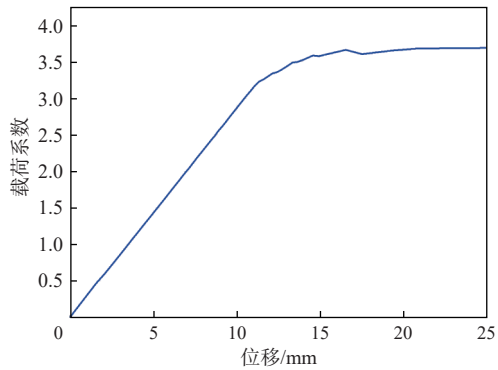


图 11　静力载荷工况下变形情况

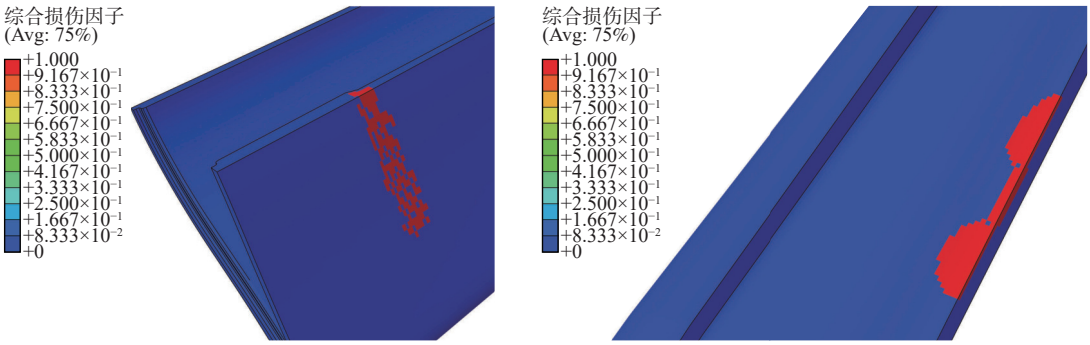
Fig. 11　Deformation under static load conditions

图 12(a)为极限载荷下罩体结构破坏形式, 可以发现, Ⅰ象限连接区小端刚度突变区域外表面和大面积受压区域内表面同时发生宏观破坏。图 12(b)、图 12(c)分别为Ⅰ象限连接区小端和大面积受压区域母向及环向损伤系数分量, Ⅰ象限连接区小端为母向拉伸破坏, 大面积受压区域为母向及环向同时破坏。其中, 综合损伤因子、母向损伤系数、环向损伤系数均为复合材料损伤变量, 其值为 0 表示材料未发生损伤, 为 1 表示完全损伤。

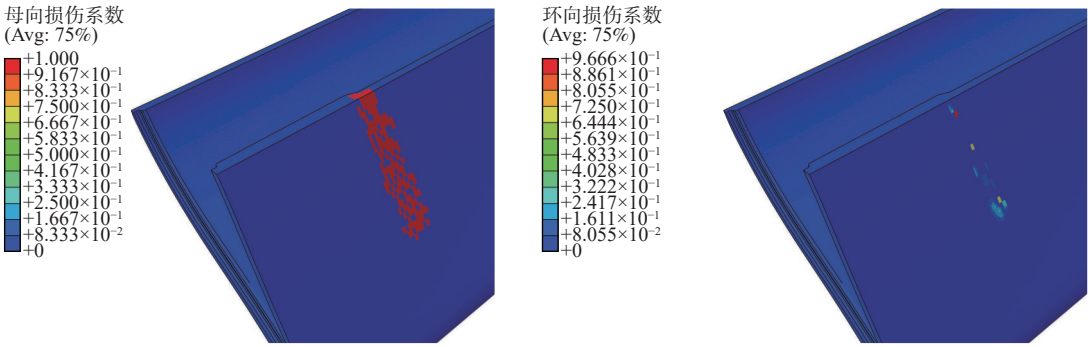
3.2　力热耦合工况

力热耦合工况时胶黏剂性能降低, 且由于金属与非金属结构线膨胀系数的差异引起热应力等, 导致透波罩连接区承载性能明显下降。仿真分析时, 先对透波罩施加热载荷模拟飞行热环境, 待热载荷加载完毕后立刻施加载荷。为便于对比分析, 试验加载程序与仿真分析一致。

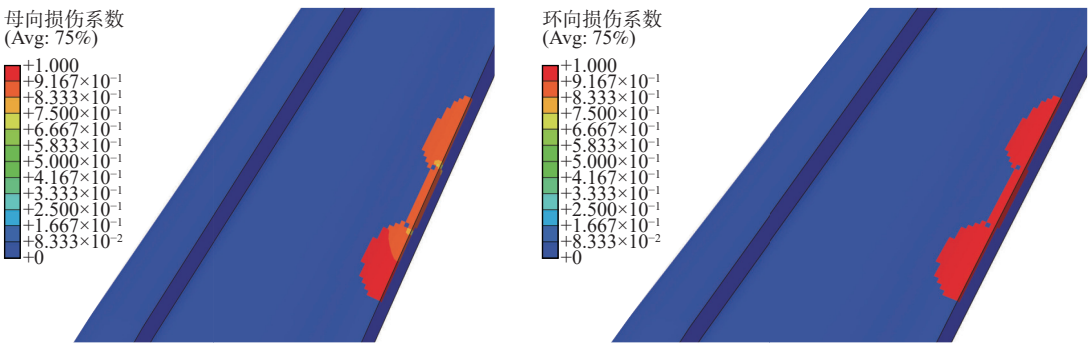
图 13 为仿真得到的力热耦合工况下胶层失效及力学性能。根据仿真结果, 载荷加载至 265% 最大使用载荷时, 透波罩连接区胶层大面积脱胶且销钉断裂。图 13(a)为黏接胶层失效情况, 从载荷-位移曲线可以看出, 该时刻位移发生突变, 载荷-位移曲线出现拐点, 连接区失效后透波罩继续承载, 但连接区罩体局部应力和钢环应力重新分布, 加载至 303% 最大使用载荷后透波罩两侧Ⅱ、Ⅳ象限环向拉断(见图 14), 结构不能继续承载。最大使用载荷



(a) 极限载荷下透波罩连接区小端及大面积破坏形式



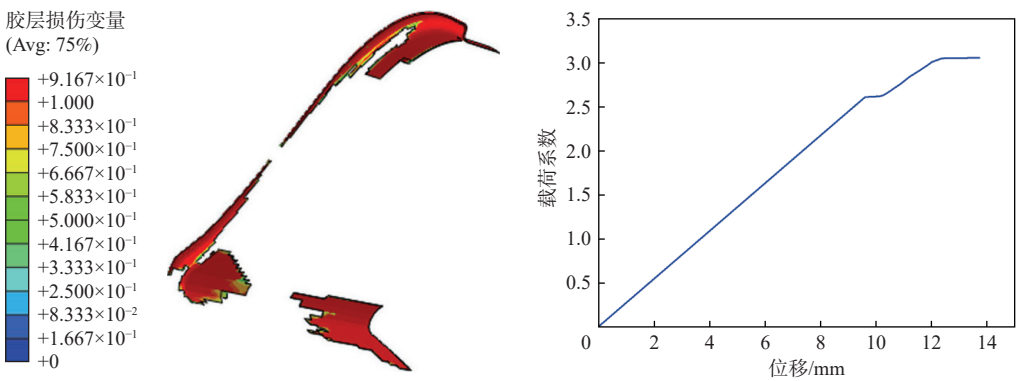
(b) 透波罩连接区小端母向及环向损伤系数分量



(c) 透波罩大面积受压区域母向及环向损伤系数分量

图 12 静力载荷工况下罩体损伤分布

Fig. 12 Damage distribution of radome under static load



(a) 黏接胶层失效情况 (b) 力热耦合工况下载荷-位移曲线

图 13 力热耦合工况下胶层失效力学性能

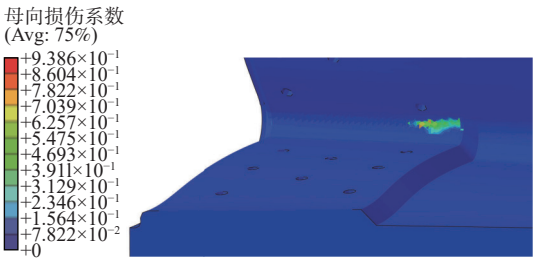
Fig. 13 Adhesive failure and mechanical property under thermo-mechanical coupling

时结构变形达 3.72 mm, 与试验结果得到的 3.99 mm 误差约为 7.26%, 一致性较好。

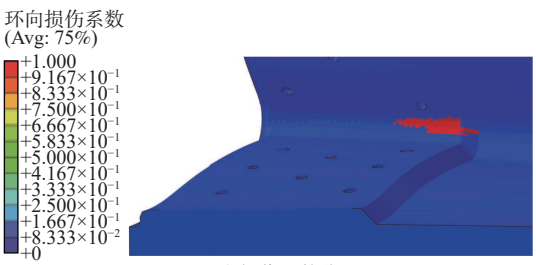
因此, 透波罩破坏模式为: 265% 使用载荷时, 透波罩连接区失效→连接区罩体局部应力和钢环应力重新分布后透波罩继续承载→303% 使用载荷时透波罩整体破坏。



(a) 透波罩罩体损伤情况



(b) 母向损伤系数分量



(c) 环向损伤系数分量

图 14　力热耦合工况下罩体损伤分布

Fig. 14　Damage distribution of radome under thermo-mechanical coupling

4　结　论

1) 将内聚力模型应用于胶黏剂的界面失效分析, 开展拉剪、压剪、拉离试片和平板级黏接结构失效分析与试验验证, 确定了适用于环氧类胶黏剂的 B-K 准则常数, 验证了内聚力模型的有效性。

2) 基于内聚力本构模型, 实现了透波罩整体承载能力的精确预测, 为解决系统级透波罩结构承载问题提供了理论支持和有效途径。

参考文献（References）

[1] 彭望泽. 防空导弹天线罩[M]. 北京: 宇航出版社, 1993: 2-3.
PENG W Z. Air defense missile radome[M]. Beijing: Aerospace Publishing House, 1993: 2-3(in Chinese).

[2] 李富志. 我国胶粘剂应用的发展趋势[J]. 粘接, 2011, 32(7): 22-25.
LI F Z. The development trend of adhesive application in China[J]. Adhesion, 2011, 32(7): 22-25(in Chinese).

[3] CAMANHO P P, MATTHEWS F L. Delamination onset prediction in mechanically fastened joints in composite laminates[J]. Journal of Composite Materials, 1999, 33(10): 906-927.

[4] DAVILA C G, JOHNSON E R. Analysis of delamination initiation in postbuckled dropped-ply laminates[J]. AIAA Journal, 1993, 31(4): 721-727.

[5] SHENG L. Quasi-impact damage initiation and growth of thick-section and toughened composite materials[J]. International Journal of Solids and Structures, 1994, 31(22): 3079-3098.

[6] ZOU Z, REID S R, LI S, et al. Modelling interlaminar and intralaminar damage in filament-wound pipes under quasi-static indentation[J]. Journal of Composite Materials, 2002, 36(4): 477-499.

[7] RYBICKI E F, KANNINEN M F. A finite element calculation of stress intensity factors by a modified crack closure integral[J]. Engineering Fracture Mechanics, 1977, 9(4): 931-938.

[8] RAJU I S. Calculation of strain-energy release rates with higher order and singular finite elements[J]. Engineering Fracture Mechanics, 1987, 28(3): 251-274.

[9] ZOU Z, REID S R, SODEN P D, et al. Mode separation of energy release rate for delamination in composite laminates using sub-laminates[J]. International Journal of Solids and Structures, 2001, 38(15): 2597-2613.

[10] KRUEGER R. The virtual crack closure technique: history, approach and applications: NASA/CR-2002-211628[R]. Washington, D. C.: NASA, 2002.

[11] TURON A, DÁVILA C G, CAMANHO P P, et al. An engineering solution for mesh size effects in the simulation of delamination using cohesive zone models[J]. Engineering Fracture Mechanics, 2007, 74(10): 1665-1682.

[12] 陈兴. 基于内聚力模型的不锈钢钎焊接头裂纹扩展研究[D]. 上海: 华东理工大学, 2014: 8-13.
CHEN X. Study on crack propagation of stainless steel brazed joint based on cohesion model[D]. Shanghai: East China University of Science and Technology, 2014: 8-13(in Chinese).

[13] TURON A, CAMANHO P P, COSTA J, et al. A damage model for the simulation of delamination in advanced composites under variable-mode loading[J]. Mechanics of Materials, 2006, 38(11): 1072-1089.

[14] DAVILA C G, CAMANHO P P, TURON A. Cohesive elements for shells: NASA/TP-2007-214869[R]. Washington, D. C.: NASA, 2007.

[15] 周储伟, 杨卫, 方岱宁. 内聚力界面单元与复合材料的界面损伤分析[J]. 力学学报, 1999, 31(3): 372-377.
ZHOU C W, YANG W, FANG D N. Cohesive interface element and interfacial damage analysis of composites[J]. Chinese Journal of Theoretical and Applied Mechanics, 1999, 31(3): 372-377(in Chinese).

[16] MI U, CRISFIELD M A, DAVIES G A O. Progressive delamination using interface elements[J]. Journal of Composite Materials, 1998, 32(14): 1246-1272.

[17] NEEDLEMAN A. A continuum model for void nucleation by inclusion debonding[J]. Journal of Applied Mechanics, 1987, 54(3): 525.

[18] NEEDLEMAN A. An analysis of tensile decohesion along an interface[J]. Journal of the Mechanics and Physics of Solids, 1990, 38(3): 289-324.

[19] XU X P, NEEDLEMAN A. Numerical simulations of fast crack growth in brittle solids[J]. Journal of the Mechanics and Physics of Solids, 1994, 42(9): 1397-1434.

[20] 徐建新, 宋依良, 郭巧荣. 一种考虑脆性断裂的三线性内聚力模型[J]. 航空科学技术, 2023, 34(3): 97-103.
XU J X, SONG Y L, GUO Q R. A trilinear cohesive model considering brittle fracture[J]. Aeronautical Science & Technology, 2023,

34(3): 97-103(in Chinese).

[21] 饶玉文. 复合材料分层内聚力模型参数反演研究[D]. 南京: 南京航空航天大学, 2020: 10-12.
RAO Y W. Study on parameter inversion of layered cohesion model of composite materials[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2020: 10-12(in Chinese).

[22] CAMANHO P P, DAVILA C G, DE MOURA M F. Numerical simulation of mixed-mode progressive delamination in composite materials[J]. Journal of Composite Materials, 2003, 37(16): 1415-1438.

[23] 何振鹏, 邓殿凯, 刘国峰, 等. 基于内聚力模型的复合材料裂纹扩展研究[J]. 复合材料科学与工程, 2022(1): 5-12.
HE Z P, DENG D K, LIU G F, et al. Research on crack propagation of composite materials based on cohesive zone model[J]. Composites Science and Engineering, 2022(1): 5-12(in Chinese).

[24] SCHELLEKENS J C J, DE BORST R. A non-linear finite element approach for the analysis of mode-I free edge delamination in composites[J]. International Journal of Solids and Structures, 1993, 30(9): 1239-1253.

[25] YUAN H, LI X. Effects of the cohesive law on ductile crack propagation simulation by using cohesive zone models[J]. Engineering Fracture Mechanics, 2014, 126: 1-11.

[26] ZHAO L B, GONG Y, ZHANG J Y, et al. Simulation of delamination growth in multidirectional laminates under mode Ⅰ and mixed mode Ⅰ/Ⅱ loadings using cohesive elements[J]. Composite Structures, 2014, 116: 509-522.

[27] GUO X, ZHANG W J, ZHU L L, et al. Mesh dependence of transverse cracking in laminated metals with nanograined interface layers[J]. Engineering Fracture Mechanics, 2013, 105: 211-220.

[28] TURON A, CAMANHO P P, COSTA J, et al. Accurate simulation of delamination growth under mixed-mode loading using cohesive elements: definition of interlaminar strengths and elastic stiffness[J]. Composite Structures, 2010, 92(8): 1857-1864.

[29] SARRADO C, TURON A, RENART J, et al. Assessment of energy dissipation during mixed-mode delamination growth using cohesive zone models[J]. Composites Part A: Applied Science and Manufacturing, 2012, 43(11): 2128-2136.

[30] TURON A, GONZÁLEZ E V, SARRADO C, et al. Accurate simulation of delamination under mixed-mode loading using a cohesive model with a mode-dependent penalty stiffness[J]. Composite Structures, 2018, 184: 506-511.

[31] 叶强. 层合复合材料的粘聚区模型及其应用研究[D]. 南京: 南京航空航天大学, 2012: 14-16.
YE Q. Study on cohesion zone model of laminated composites and its application[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012: 14-16(in Chinese).

[32] WU E M, REUTER R C. Crack extension in fiberglass reinforced plastics: 1967-0570[R]. Urbana: University of Illinois, 1965.

[33] BENZEGGAGH M L, KENANE M. Measurement of mixed-mode delamination fracture toughness of unidirectional glass/epoxy composites with mixed-mode bending apparatus[J]. Composites Science and Technology, 1996, 56(4): 439-449.

[34] YE L. Role of matrix resin in delamination onset and growth in composite laminates[J]. Composites Science and Technology, 1988, 33(4): 257-277.

[35] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 胶粘剂 拉伸剪切强度的测定 (刚性材料对刚性材料): GB/T 7124—2008[S]. 北京: 中国标准出版社, 2008.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, National Standardization Administration. Adhesives-Determination of tensile lap-shear strength of rigid-to-rigid bonded assemblies: GB/T 7124—2008[S]. Beijing: Standards Press of China, 2008(in Chinese).

[36] 中国航天工业总公司. 胶粘剂压缩剪切强度试验方法: QJ 1634A—96[S]. 北京: 中国标准出版社, 1996.
China Aerospace Industry Corporation. Test method for compressive shear strength of adhesives: QJ 1634A—96[S]. Beijing: Standards Press of China, 1996(in Chinese).

[37] 国防科学技术工业委员会. 胶粘剂高温拉伸强度试验方法 (金属对金属): GJB 445—1988[S]. 北京: 中国标准出版社, 1988.
Commission of Science, Technology and Industry for National Defense. Test method for tensile of adhesives at high temperature (metal to metal):GJB 445—1988[S]. Beijing: Standards Press of China, 1988(in Chinese).

Performance study on bonding strength of radome based on cohesive zone model

HOU Baojiang[1, 2, *], WANG Jiao[2], XING Yufeng[1], ZHANG Fei[2], LI Yanxi[2]

(1. School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China;
2. Beijing Institute of Space Long March Vehicle, Beijing 100076, China)

Abstract: The radome and aircraft are often connected by adhesive bonding and auxiliary connection. With the development of aircraft technologies, higher requirements are raised by severe thermo-mechanical environments for the properties of bonding structure during flight. In practical engineering applications, the linear elastic constitutive model is usually used as the constitutive model of adhesives for analysis. The calculation results often greatly deviate from the actual bearing capacity, and it is difficult to accurately simulate the mechanical behavior of the related bonding structures. The frequently used bilinear cohesive constitutive model was employed in this paper. First, the constitutive model was calibrated by the test results of the actual performance of the adhesive. Then, the effectiveness of the model under a complex stress environment was further verified by the performance test of the actual bonding structure at the level of the surface plate. The verified constitutive model was used to analyze the bonding strength for aircraft-level radome, and the results were compared with the actual test results. The calculation accuracy of the model was verified, which provided an effective way for accurately predicting the bonding strength of similar structures in the future.

Keywords: cohesion; bilinear; radome; bonding strength; numerical simulation

Received: 2023-12-04; **Accepted**: 2024-03-11; **Published Online**: 2024-04-15 11:35
URL: link.cnki.net/urlid/11.2625.V.20240412.1238.002
* **Corresponding author**. E-mail: 153327058@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0405

考虑多车型协同的机坪保障车辆双阶段调度

姜雨*，张诗淼，陈名扬，张文静，吴薇薇
(南京航空航天大学 民航学院，南京 211106)

摘　　要：大型机场机坪保障车辆一直面临较大运行压力，而多种机坪保障服务的配合约束和动态航班信息对车辆调度提出更高挑战。考虑连续工作、容量限制的连续工作和往返工作3种不同模式车辆运行约束差异，以车辆数量最少和行驶总距离最小为目标，建立多车型协同的机坪保障车辆调度模型。根据机场实际运行情况，对该模型进行双阶段求解，在静态阶段设计一种融合局部搜索的非支配排序遗传算法Ⅱ(LS-NSGA Ⅱ)，在动态阶段设计一种类邻域搜索算法求解多车型协同调度问题。数据仿真结果表明：所建模型静态调度阶段结果相对于先到先服务(FCFS)，车辆数量和行驶总距离分别降低了18.9%和8.9%；动态调度阶段结果能保持原调度计划车辆数量，行驶距离调整量相较于大规模邻域搜索算法降低了25.8%。研究结果可以为大型机场机坪保障车辆调度管理和决策提供一定的指导。

关 键 词：机坪保障车辆；双阶段调度；路径规划；多车型协同；改进NSGA Ⅱ
中图分类号：V351.11
文献标志码：A　　**文章编号**：1001-5965(2025)06-1926-09

机坪保障车辆作为保障航班服务质量和时效性的重要支持设备，现阶段普遍采用基于先到先服务(first-come-first-served, FCFS)的人工指派调度[1-3]，而这种调度策略已经无法满足日趋繁忙的大型枢纽机场，亟待寻求科学高效的机坪保障车辆调度策略。由于航空器机坪服务所需保障车辆类型众多、服务种类多样、车辆服务时间约束迥异、车载资源受限，因此，若能根据不同车型工作方式设置独特约束条件，再将所有车辆看作一个整体进行统一调度，则能进一步提高保障车辆的服务效率。

国内外众多学者对大型机场机坪保障车辆调度策略已经进行了大量研究，绝大多数学者的主要研究对象为某类单一车型，并考虑单车型的独特运行模式[4-5]，如加油车[6-7]、摆渡车[8-9]、行李车[10]、除冰车[11]等。

关于多车型协同的机坪保障车辆调度研究发展迅速。多车型协同是指将2种或2种以上的车型作为一个整体进行协同配置，以实现更大范围的高效率机坪保障车辆调度。Zhao等[12]以摆渡车和牵引车为研究对象，以车辆数量最少和任务均衡量最大为目标建立双目标混合整数规划(mixed integer programming, MIP)模型。Zhu等[13]以加油车和摆渡车为研究对象，设计了考虑两者时间约束的多目标混合整数规划模型，并使用CPLEX求解器进行求解。

在机场的实际运行过程中，机坪保障车辆调度可以分为2个阶段：静态调度阶段和动态调度阶段。上述研究[12-13]主要针对静态阶段，而针对动态阶段的车辆调度策略研究逐渐成为近年来的研究热点。在机坪保障车辆动态调度问题中很多相关信息是不确定且随机的，如服务航班的进离港时间发生变化[14-15]或航班更改停靠机坪[16]等，因此，在

收稿日期：2023-06-21；录用日期：2023-08-24；网络出版时间：2023-09-05 15:00
网络出版地址：link.cnki.net/urlid/11.2625.V.20230905.1321.004
基金项目：国家自然科学基金(U1933118)
*通信作者. E-mail：jiangyu07@nuaa.edu.cn

引用格式：姜雨，张诗淼，陈名扬，等. 考虑多车型协同的机坪保障车辆双阶段调度[J]. 北京航空航天大学学报，2025，51(6)：1926-1934.
JIANG Y，ZHANG S M，CHEN M Y，et al. Dual-phase scheduling of apron support vehicles considering multi-vehicle coordination[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1926-1934 (in Chinese).

调度保障车辆时须根据实时信息对车辆路径进行及时调整。大型机场机坪保障车辆动态调度问题的本质是带时间窗的动态车辆路径问题(dynamic vehicle routing problem with time windows, DVRPTW), 因此, 可以参考这一领域内的相关研究[17-19]。

综上, 目前已有的研究中多车型协同主要体现在车型种类, 鲜有从车辆工作方式的角度出发进行研究; 在动态调度研究中, 大都是将动态问题直接转化为静态模型并使用传统 DVRPTW 的求解算法和思路, 未考虑机坪保障车辆实时调度场景的适用性。因此, 本文基于对机坪保障车辆工作方式的分析, 选取代表性车型进行协同调度研究, 并设计双阶段算法进行求解。

1　问题描述

所有类型的机坪保障车辆都必须按照民用机场规定的运行模式进行调度和服务。本文分析了机坪保障车辆的工作方式并将其分为 3 类, 如图 1 所示。

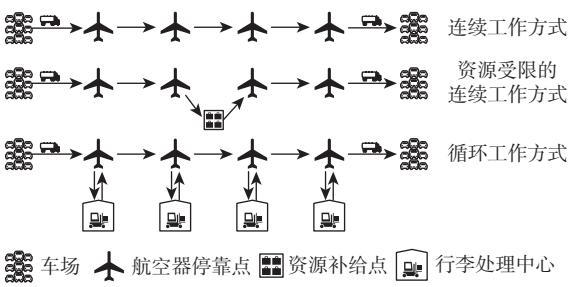


图 1　机坪保障车辆工作方式

Fig. 1　Apron ground support vehicles working mode

连续工作方式指机坪保障车辆从车场出发后前往服务点进行地面保障服务, 结束任务后继续前往下一服务点直至完成所有任务后返回车场; 资源受限的连续工作方式指当剩余车载资源无法满足下一个服务点的需求, 必须返回补给点补充资源后再执行后续任务; 循环工作方式是指机坪保障车辆在进行服务需要往返于多个服务点, 是最复杂的车辆工作方式。

本文以传送带车、清水车和行李牵引车 3 种典型车辆为研究对象, 主要功能是完成进离港航班的行李服务和清水服务, 且传送带车和行李牵引车需要相互协同配合工作, 从而完成机坪保障服务。

机坪保障车辆从航班i停靠点到航班j停靠点转场时的行驶距离与车辆工作方式有关。本文使用希腊字母α、β和γ分别表示行李牵引车、传送带车和清水车。调度有向图G中, d_{ij}表示航班i停靠点到航班j停靠点的实际距离, s_{ij}^{q}表示q类型车辆从航班i停靠点到航班j停靠点的转场距离, a、b和c分别表示清水补给点、行李中心停靠点和车场。

行李牵引车为循环工作方式, 其转场距离与航班性质有关, 具体如表 1 所示。传送带车为连续工作方式车辆, 其转场距离$s_{ij}^{\beta}=d_{ij}$。清水车为资源受限的连续工作方式, 当资源不够需要前往补给点时, 其转场距离$s_{ij}^{\gamma}=d_{ia}+d_{aj}$, 其余情况$s_{ij}^{\gamma}=d_{ij}$。

表 1　行李牵引车转场距离

Table 1　Luggage tractor transfer distance

前序航班i性质	后序航班j性质	车辆转场方向	转场距离s_{ij}^{α}
进港	进港	行李中心→停机位	d_{bj}
进港	离港	行李中心→行李中心	0
离港	进港	停机位→停机位	d_{ij}
离港	离港	停机位→行李中心	d_{ib}

2　机坪保障车辆协同调度模型

2.1　模型假设

1) 航班机型及车辆基本信息完备无误, 且不会发生变更。

2) 机坪保障车辆一旦开始服务, 不会因任何情况中断直至该项任务结束。

3) 同类型的机坪保障车辆性能相同, 车辆操作人员的工作能力和效率相同。

4) 车辆均从车场出发, 完成所有任务后再次返回车场, 且行驶速度不变。

2.2　单车型调度模型

机坪保障车辆单车型调度模型目标函数如下:

$$Z_1^q=\min\sum_{k\in N^q}y_k^q \tag{1}$$

$$Z_2^q=\min\sum_{k\in N^q}\sum_{i,j\in H^q}s_{ij}^q x_{ijk}^q \tag{2}$$

式中: N^q为q车型车辆集合; H^q为q车型需要服务的航班集合。式 (1) 表示q车型车辆数目最少; 式 (2) 表示q车型行驶距离最小。

式 (3)～式 (9) 为通用约束。式 (3) 表示每个航班只接受一次q类型车辆的服务; 式 (4)～式 (6) 表示车辆路的流量限制, 即车辆最多同时服务一个航班; 式 (7) 表示使用的机坪保障车辆不能超过可用车辆数量; 式 (8) 表示车辆k到达航空器j停靠点的时间不晚于该航空器服务开始时间; 式 (9) 表示q车型开始服务航班i的时间必须在该类型服务开始时间窗内。

$$\sum_{k\in N^q}z_{ik}^q=1\qquad\forall i\in H^q \tag{3}$$

$$\sum_{i\in H^q} x_{cik}^q \leqslant 1 \qquad \forall k \in N^q \tag{4}$$

$$\sum_{j\in H^q} x_{ijk}^q - \sum_{j\in H^q} x_{jik}^q = 0 \qquad \forall k \in N^q, \forall i \in H^q \tag{5}$$

$$\sum_{i\in H^q} x_{ick}^q \leqslant 1 \qquad \forall k \in N^q \tag{6}$$

$$\sum_{k\in N^q} y_k^q \leqslant U^q \tag{7}$$

$$E_{ik}^q + \frac{s_{ij}^q}{v} - (1-x_{ijk}^q)M \leqslant S_{jk}^q \qquad \forall k \in N^q,\ \forall i,j \in H^q \tag{8}$$

$$A_i^q z_{ik}^q \leqslant S_{ik}^q z_{ik}^q \leqslant B_i^q z_{ik}^q \qquad \forall k \in N^q,\ \forall i \in H^q \tag{9}$$

式中：z_{ik}^q为决策变量；U^q为q车型可用车辆数量；E_{ik}^q为q车型车辆k结束服务航班i的时间；S_{jk}^q为q车型车辆k开始服务航班j的时间；v为机坪保障车辆行驶速度；M为无穷大正数；A_i^q为q车型服务航班i的最早开始时间；B_i^q为q车型服务航班i的最晚开始时间。

式(10)～式(11)为清水车特定约束。式(10)表示清水车从车场出发时为空载状态；式(11)表示清水车在运行过程中剩余清水量不小于0也不超过最大容积。

$$c_{ck}^\gamma y_k^\gamma = 0 \qquad \forall k \in N^\gamma \tag{10}$$

$$0 \leqslant c_{ik}^\gamma z_{ik}^\gamma \leqslant P \qquad \forall i \in H^\gamma,\ \forall k \in N^\gamma \tag{11}$$

式中：c_{ck}^γ为清水车k服务完航班i后的剩余清水容量；P为清水车最大容量。

x_{ijk}^q为0-1决策变量，当q车型的车辆k依次服务航班i、j时其值为1，否则为0；y_k^q为0-1决策变量，当q车型的车辆k被启用时其值为1，否则为0；z_{ik}^q为0-1决策变量，当q车型的车辆k服务航班i时其值为1，否则为0。

2.3 多车型协同调度模型

在机场地面保障实际运行环节中，需要不同车型相互配合完成服务。本文主要考虑传送带车和行李牵引车之间的协同配合。

机坪保障车辆多车型协同调度的目标函数如下：

$$Z_1 = \min \sum_{k\in N^q} \sum_{q\in Q} y_k^q \tag{12}$$

$$Z_2 = \min \sum_{k\in H^q} \sum_{i,j\in H^q} \sum_{q\in Q} s_{ij}^q x_{ijk}^q \tag{13}$$

式中：Q为机坪保障车辆类型集合。

式(12)表示所有机坪保障车辆数量最少；式(13)表示所有机坪保障车辆行驶总距离最小。最小化车辆数量是为探索完成保障所必须的资源支持数量，最小化行驶总距离是为减少车辆运行成本的同时尽可能避免车辆在飞行区大面积活动。

在行李装卸服务中，行李牵引车和传送带车必须同时抵达服务地点才可以工作。因此，协同调度在满足单车型调度约束式(3)～式(11)外，还需要满足如下约束条件：

$$T_{ik}^\alpha \geqslant T_{il}^\beta \qquad \forall i \in H, \forall k \in N^\alpha, \forall l \in N^\beta \tag{14}$$

$$S_{ik}^\alpha = S_{il}^\beta = \begin{cases} T_{ik}^\alpha & T_{ik}^\alpha > A_i^\alpha \\ A_i^\alpha & T_{ik}^\alpha \leqslant A_i^\alpha \end{cases} \qquad \forall i \in H, \forall k \in N^\alpha, \forall l \in N^\beta \tag{15}$$

$$E_{ik}^\alpha = E_{il}^\beta \qquad \forall i \in H, \forall k \in N^\alpha, \forall l \in N^\beta \tag{16}$$

式中：T_{ik}^q为q车型车辆k到达航班i停靠点的时间。式(14)～式(16)为车辆协同约束。式(14)表示传送带车必须在行李牵引车前抵达；式(15)表示只有传送到车和行李牵引车同时到达服务点才可以开始行李服务；式(16)表示服行李牵引车结束服务后传送带车方可结束服务。

3 双阶段调度求解方法

机坪保障车辆调度分为双阶段，即静态调度和动态调度。静态调度是指根据待执行的航班计划表求解多车型协同配合的机坪保障车辆调度方案。动态调度是根据临时发生的航班信息变化对静态调度方案进行实时调整，快速高效地得到新调度方案。本文设计的多车型协同的机坪保障车辆双阶段求解流程如图2所示。

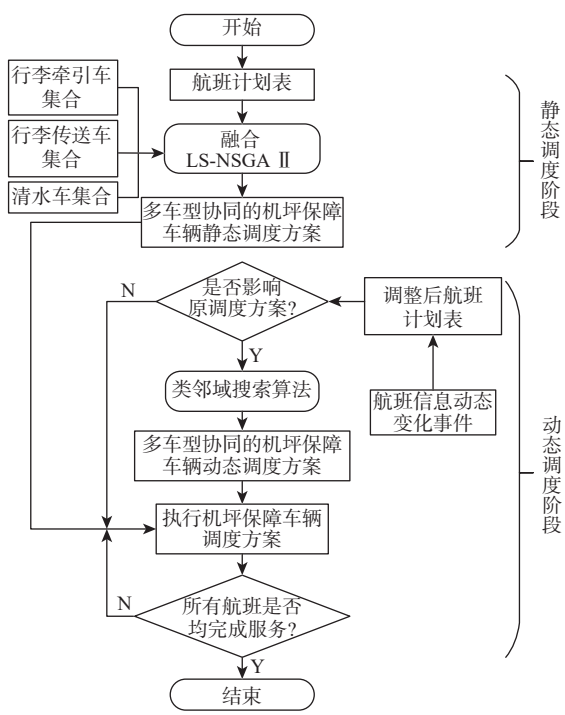


图2 多车型协同的机坪保障车辆双阶段调度算法流程

Fig. 2 Two-stage scheduling algorithm process of multi-vehicle cooperative apron support vehicle

针对多车型协同的机坪保障车辆静态调度, 本文采用融合局部搜索的非支配排序遗传算法Ⅱ(local search non-dominated sorting genetic algorithms Ⅱ, LS-NSGA Ⅱ)。首先, 采用节约里程算法构建初始解以保证初始解质量; 其次, 使用 LS-NSGA Ⅱ对初始解进行迭代计算, 得到 Pareto 最优解集; 最后, 使用多属性决策中有序加权平均算子得到满足协同约束的多车型调度问题的最优解。

针对多车型协同的机坪保障车辆动态调度, 本文参考邻域搜索算法中的破坏和修复理念设计类邻域搜索(similar neighborhood search, SNS)算法。首先, 判断原有的调度计划是否满足变化的航班计划表; 其次, 筛选出原方案中需要调整服务车辆的航班; 最后, 将需要调整的航班按照新时间窗和协同约束插入到车辆路径中, 以达到快速响应的目的。

3.1　静态调度阶段算法设计

本文在传统 NSGA Ⅱ基础的交叉变异操作后增加局部搜索操作, 在保证算法具备全局搜索能力的同时提高局部搜索能力, 能更快贴近最优解, 提高求解质量和效率。

局部搜索操作的核心是破坏操作和修复操作。通过破坏算子筛选出种群中需要调整的航班, 再根据修复算子将这些航班重新插入车辆路径中, 得到新调度方案。根据机坪保障车辆调度模型的特点, 引入随机移除和最差值移除 2 种破坏算子, 以及最小插入成本、最大遗憾值 2 种修复算子, 且根据新个体的质量选择破坏算子和修复算子。LS-NSGA Ⅱ主函数伪代码如下:

输入　迭代前种群C, 迭代次数gen, 交叉概率P_c, 变异概率P_m, 航班信息集合s, 停机位间距离矩阵$\boldsymbol{d}$, 清水车容量限制P。

输出　迭代后种群C'。

$F \leftarrow C$ //生成父代初始种群F
对F进行非支配排序和拥挤度计算
$t \leftarrow 1$
while $t \leqslant$ gen
　$Q \leftarrow \varnothing$
　for $i \leftarrow 1$ to length[F]
　　使用二元锦标赛方法对F进行选择操作, 选出个体f
　　$Q \leftarrow Q \cup f$
　end for
　for $j \leftarrow 1$ to length[Q]
　　$g \leftarrow Q(j,:)$ //提取第j个个体
　　对g进行交叉、变异、局部搜索操作, 生成新个体g'
　　$Q(j,:) \leftarrow g'$
　end for
　$W \leftarrow Q \cup F$
　对W进行非支配排序计算
　$F \leftarrow$对W进行精英选择策略, 得到新父代种群
　$t \leftarrow t+1$
end while
$C' \leftarrow F$
return C'

3.2　动态调度阶段算法设计

本文设计了一种类邻域搜索算法, SNS 算法的整体思想如图 3 所示。航班信息发生变化后, 将需要调整保障车辆的航班移动至待调度航班集合H^q_{adjust}中, 完成破坏操作; 再将待调度航班插入至新路径中, 完成修复操作。为满足动态调度的时效性, 若调整结果没有违反任何单车型调度和协同调度约束条件, 则完成动态调度, 不再进行迭代寻找最优解。

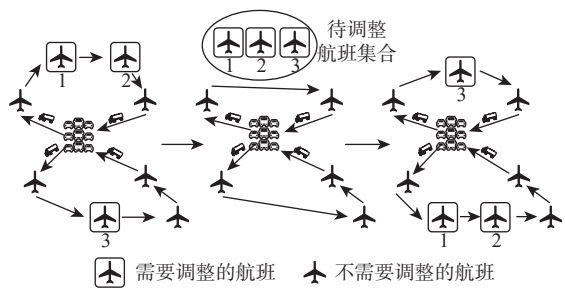


图 3　SNS 算法示意图

Fig. 3　Schematic diagram of SNS algorithm

本文假设在开始动态调整时, 已经完成或正在进行的调度计划不会发生变化。SNS 算法主函数伪代码如下:

输入　原始调度方案X, 原始航班信息集合s, 动态航班信息集合s', 停机位间距离矩阵$\boldsymbol{d}$, 清水车容量限制P。

输出　动态调度方案X'。

$H^{\alpha}_{\text{adjust}}; H^{\beta}_{\text{adjust}}; H^{\gamma}_{\text{adjust}} \leftarrow s, s', X$
　//得到待重新调整的航班集合
$[V^{\alpha}; V^{\beta}; V^{\gamma}] \leftarrow X$ //各车型原始调度路径
while $H^{\alpha}_{\text{adjust}} \neq \varnothing \&\& H^{\beta}_{\text{adjust}} \neq \varnothing \&\& H^{\gamma}_{\text{adjust}} \neq \varnothing$
　while $H^{\alpha}_{\text{adjust}} \neq \varnothing$
　　$V^{\alpha} \leftarrow \text{repair}(H^{\alpha}_{\text{adjust}})$//行李牵引车动态调度
　end while
　while $H^{\gamma}_{\text{adjust}} \neq \varnothing$
　　$V^{\gamma} \leftarrow \text{repair}(H^{\gamma}_{\text{adjust}})$//清水车动态调度
　end while
　while $H^{\beta}_{\text{adjust}} \neq \varnothing$

12. $V^{\beta} \leftarrow \text{repair}(H^{\beta}_{\text{adjust}})$//清水车动态调度
13. end while
14. $X' \leftarrow [V^{\alpha}; V^{\beta}; V^{\gamma}]$
15. if X'中有违反任意多车型协同调度模型约束的航班
16. $H^{q}_{\text{adjust}} \leftarrow$违反约束条件的航班
17. end if
18. end while
19. return X'

4 实例分析

4.1 实验数据

本文以中国某大型机场T2航站楼实际运行数据为例，选取国庆假期某日高峰时刻进离港航班数据，验证多车型协同机坪保障车辆调度模型和双阶段求解算法的可靠性。

4.1.1 航班与车辆信息

本文共选取145个航班，其中，进港航班67个，离港航班78个，始发航班11个，过站航班共有67对。航班计划表要素如表2所示。

表2 航班计划表（示例）

Table 2 Flight schedule (example)

航班编号	机型	机位	计划进/离港时刻	航班性质	过站时长/min
1	A32B	268	10:00	进港	65
2	ERJ190	434	10:00	离港	65

车辆信息主要包括车辆可用数量、车辆所需数量和工作时长，如表3所示。

表3 车辆所需数量和服务时长

Table 3 Number of vehicles required and minutes of service

机型	车辆所需数量/台			服务时长/min				
	行李牵引车	传送带车	清水车	行李牵引车		传送带车		清水车
				进港	离港	进港	离港	
C	1	1	1	15	25	15	25	5
D	2	1	1	15	25	15	25	7
E	2	1	1	15	25	15	25	7
F	3	2	1	20	30	20	30	9

4.1.2 算法参数

本文设计的算法均在MATLAB平台上实现，LS-NSGA Ⅱ相关参数设置如表4所示。

表4 LS-NSGA Ⅱ参数设置

Table 4 LS-NSGA Ⅱ parameters settings

参数	数值
LS-ANSGA Ⅱ迭代次数	300
LS-ANSGA Ⅱ种群大小	150
交叉概率	0.7
变异概率	0.05
目标函数权重	0.5

4.1.3 动态调度信息

本文设计的动态情景为得知未来2 h内发生的航班变动信息，表5为不同类型航班信息变化数量。变化出现频率及变更时长符合该机场历史数据，因此，本文设计的动态情景能在一定程度上还原实际运行中的动态变化。

表5 动态变化的航班数量

Table 5 Number of dynamic change flights

航班性质	航班数量		
	时间变更	停机位变更	航班取消
进港	33	2	1
离港	14	2	1

4.2 静态调度结果分析

4.2.1 静态调度算法对比

使用LS-NSGA Ⅱ对静态阶段求解得到Pareto最优解集，并与传统NSGA Ⅱ结果进行比较如表6所示，计算结果散点图如图4所示。其中，NSGA Ⅱ相关参数的设置与LS-NSGA Ⅱ一致。

通过有序加权平均算子得到多车型协同的整体最佳调度方案，最优结果在表6中加粗显示。同时与FCFS进行对比，对比结果如表7所示。

由图4可知，LS-NSGA Ⅱ求得的Pareto最优解集整体优于FCFS算法求得的唯一解，并通过更少的迭代次数获得比NSGA Ⅱ更好的优化结果。

由表7可知，LS-NSGA Ⅱ相较于FCFS算法需要的车辆总数减少7辆，降幅18.9%；行驶总距离减少18 406 m，降幅8.9%，这是由于FCFS策略只按照时间序列进行车辆调度，没有全局优化理念。相较于传统的NSGA Ⅱ，本文设计的LS-NSGA Ⅱ使用更少时间求解出相同车辆数目，并将行驶总距离减少了3 852 m，这是由于增加了局部搜索操作，在保证原有全局搜索能力的基础上加速局部搜索效率，

表 6　LS-NSGA Ⅱ、NSGA Ⅱ和 FCFS 结果对比

Table 6　LS-NSGA II, NSGA II and FCFS results comparison

车辆类型	Z_1^i/台			Z_2^i/m		
	行李牵引车	传送带车	清水车	行李牵引车	传送带车	清水车
LS-NSGA Ⅱ	14	10	3	128 456	48 467	22 336
	15	**11**	**4**	**121 648**	**45 679**	**22 194**
	16	12	5	120 358	44 974	22 074
	17	13		119 365	44 075	
	18	14		118 453	43 754	
NSGA Ⅱ	14	10	3	129 375	49 447	23 257
	15	**11**	**4**	**123 956**	**46 346**	**23 071**
	16	12	5	121 365	45 355	22 887
	17	13		120 745	44 698	
	18	14		118 753	44 256	
FCFS	18	14	5	131 965	50 486	25 476

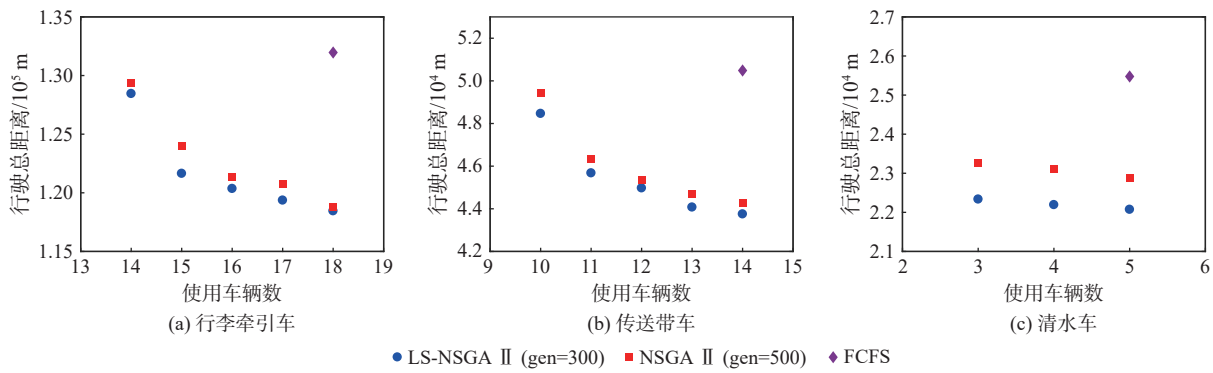


图 4　LS-NSGA Ⅱ、NSGA Ⅱ和 FCFS 求解结果散点图

Fig. 4　LS-NSGA Ⅱ, NSGA Ⅱ and FCFS solution result scatter plot

加速了求解速度并提高结果质量。

综上, 本文所建立的静态阶段算法在质量和效率上均优于对比的基准算法。

4.2.2　协同调度与顺序调度对比

表 8 为顺序调度和协同调度结果对比。顺序调度是指按照单车型调度模型分别求解 3 种车型调度结果, 其中, 传送带车需根据行李牵引车的调度结果进行求解以满足行李装卸服务的协同约束条件。

表 7　最优解对比

Table 7　Comparison of optimal solutions

算法	Z_1/台	Z_2/m
LS-NSGA Ⅱ	30	189 521
NSGA Ⅱ	30	193 373
FCFS	37	207 927

由表 8 可知, 相较于顺序调度, 本文建立的协同调度模型车辆使用数减少了 2 辆, 降幅 6.3%, 车辆行驶距离减少了 2 352 m。其中, 行李牵引车和传送带车的数量均减少 1 辆; 行李牵引车、传送带车和清水车的行驶距离分别减少了 867 m、1 354 m 和 131 m。这是由于顺序调度中传送带车在行李牵引车之后求解, 为迎合牵引车的时间窗协同完成行李装卸服务, 才造成传送带车数量上涨。而清水车

表 8　顺序调度与协同调度结果对比

Table 8　Comparison of sequential scheduling and collaborative scheduling

调度模式	Z_1/台	Z_2/m	Z_1^i/台			Z_2^i/m		
			行李牵引车	传送带车	清水车	行李牵引车	传送带车	清水车
顺序调度	32	191 873	16	12	4	122 515	47 033	22 325
协同调度	30	189 521	15	11	4	121 648	45 679	22 194

没有协同约束条件，因此，其数量与顺序调度一样，车辆行驶距离仅有小幅降低。

综上，本文建立的多车型协同调度模型能对机坪保障车辆进行更系统化的求解，并对2个目标函数的优化效果均优于独立调度。

4.3 动态调度结果分析

4.3.1 动态调度算法对比

对本文4.1.3节设置的动态情景求解，并与文献[20]中的自适应大邻域搜索(adaptive large neighborhood search, ALNS)算法进行对比，结果如表9所示。2种算法的车辆数涨幅相同，但与静态调度行驶距离的差值SNS算法比ALNS算法少25.8%。并且SNS算法能够使用更短的时间获得与ALNS算法所得相近的实时调度方案，结果表明，本文设计的SNS算法能根据航班动态信息快速高效获得实时调度方案，且调整速度满足机场实际运行需求。

表9 SNS和ALNS算法结果对比

Table 9 SNS and ALNS algorithms results comparison

动态调度算法	Z_1/台	与静态调度差值/台	Z_2/m	与静态调度差值/m	计算时间/s
SNS	32	+2	217 111	+27 590	6.53
ALNS[20]	32	+2	228 146	+38 625	142.72

4.3.2 动态调度与静态调度对比

使用SNS算法进行动态求解后结果与静态调度结果对比如表10所示。3种车型行驶总距离均有所增长，平均每辆行李牵引车增加约429 m，平均每辆传送带车增加582 m，平均每辆清水车增加372 m。2种车型使用数量均增加1辆，其中，清水车数量与原方案相同。

为分析各车型机坪保障车辆路径变化情况，表11统计了不同信息变更情景下的航班数量。清水车调度中调整的航班数量不多，这是因为该车型主要服务离港航班，通常在发生延误前就已经完成清水操作，无需因延误调整调度计划，只有在停机位变更和航班取消时才必须做出实时调整。行李牵引车和传送带车的时间窗与计划进港/离港时间关系密切，几乎所有的航班信息变更情况都需要这2类车调整调度结果，因此，需要调整的航班数量较多，并且会影响到一部分正常航班。

为研究车辆服务过程中在场面的流动情况，本文将航站楼分为7个区域。由于优化目标包含最小化车辆行驶距离，会使调度结果中车辆呈现分区服务的状态，因为这样可以减少车辆在场面大面积和长距离流通，从而达到降低行驶距离的目的。由于行李牵引车使用数量最多且需要调整的航班最多，因此，以行李牵引车为例分析动态调度与静态调度车辆服务范围对比，结果如表12所示。其中，车辆16是在动态阶段的新增车辆，因此，该车无静态调度范围。

综上，本文提出的SNS算法对连续工作方式和资源受限的连续工作方式的机坪保障车辆仍能保持原有的服务范围；对于循环工作方式的车辆虽然相较于静态方案有所改变，但仍能与原方案服务范围保持较高的相似度。

表10 动态调度与静态调度结果对比

Table 10 Comparison of dynamic scheduling and static scheduling

调度模式	Z_1/台	Z_2/m	Z_1^i/台			Z_2^i/m		
			行李牵引车	传送带车	清水车	行李牵引车	传送带车	清水车
动态调度	32	217 111	16	12	4	136 614	56 816	23 681
静态调度	30	189 521	15	11	4	121 648	45 679	22 194

表11 动态调度中不同情景下的航班数量

Table 11 Number of flights under different scenarios in dynamic scheduling

机坪保障车辆类型	航班信息改变		航班信息未改变	
	保障车辆改变	保障车辆未改变	保障车辆改变	保障车辆未改变
行李牵引车	37	16	24	68
传送带车	28	25	15	77
清水车	3	14	0	61

表 12　行李牵引车双阶段调度结果对比

Table 12　Comparison of two-stage scheduling results of luggage tractor

车辆序号	航站楼分区						
	1	2	3	4	5	6	7
1				√	√		
	○			○			
2				√	√		
	○			○	○		
3		√					
	○	○		○			
4	√	√	√				
					○		
5	√	√	√				
	○	○		○			
6	√	√	√				
	○	○	○	○	○	○	○
7	√	√	√	√	√	√	√
	○	○	○	○	○	○	○
8		√	√				
		○	○				
9	√	√					
	○	○	○				
10	√	√	√	√	√	√	√
	○	○	○				
11			√				
		○	○				
12			√	√		√	
			○	○		○	
13		√	√				
		○	○				
14		√	√				
		○					
15		√	√				
		○	○				
16							
	○	○	○				

注: √表示静态调度, ○表示动态调度。

5　结　论

分析机坪保障车辆工作方式、车辆转场距离、车辆特定运行规则、协同运行规则等车辆约束条件, 在此基础上建立考虑多车型的机坪保障车辆协同调度模型, 根据机场实际运行特点设计静态调度和动态调度双阶段求解算法。以中国某大型机场实际运行数据为案例, 验证了本文模型和算法的有效性。本文主要结论如下。

1) 本文建立的多车型协同调度模型与各车型独立调度相比, 车辆数量降低了 6.3%, 且每种车辆的行驶距离均有所下降。

2) 静态调度阶段在传统的 NSGA Ⅱ 基础上加入局部搜索操作, 在保证算法全局搜索能力的同时加速局部搜索效率。与传统的先到先服务策略相比, 能减少 18.9% 的车辆数量和 8.9% 的行驶总距离。

3) 动态阶段的 SNS 算法能满足机场实际动态变化情景, 在计算时间和与原计划差值上均优于对比算法, 并且动态调度结果能在车辆服务范围上与原计划保持高相似度。

机坪保障车辆系统不是独立运行的, 停机位分配方案可能会使车辆调度计划无法满足服务需求。未来研究可以针对机坪保障车辆资源和停机位资源进行联合调度研究, 分析联合调度的耦合因素和求解策略。对推动智慧机场发展, 提高机场运营管理水平具有重要意义。

参考文献 (References)

[1] KUHN K, LOTH S. Airport service vehicle scheduling[J]. Air Traffic Control Quarterly, 2010, 18(1): 63-83.

[2] NORIN A, GRANBERG T A, YUAN D, et al. Airport logistics–a case study of the turn——around process[J]. Journal of Air Transport Management, 2012, 20: 31-34.

[3] NORIN A, YUAN D, GRANBERG T A, et al. Scheduling de-icing vehicles within airport logistics: a heuristic algorithm and performance evaluation[J]. Journal of the Operational Research Society, 2012, 63(8): 1116-1125.

[4] 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.

[5] 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).

[6] GAMAYANTI N, SAHAL M, WIBISONO A. Optimization of vehicle routing problem with tight time windows, short travel time and re-used vehicles (vrptsr) for aircraft refueling in airport using ant colony optimization algorithm[J]. Journal on Advanced Research in Electrical Engineering, 2018, 2(1): 43-46.

[7] 衡红军, 戚馨桐. 考虑任务均衡的加油车动态调度问题[J]. 计算机工程与科学, 2020, 42(5): 923-930.
HENG H J, QI X T. Dynamic refueling vehicle scheduling considering task balance[J]. Computer Engineering & Science, 2020, 42(5): 923-930(in Chinese).

[8] ZHAO P X, HAN X, WAN D. Evaluation of the airport ferry vehicle scheduling based on network maximum flow model[J]. Omega, 2021, 99: 102178.

[9] HAN X, ZHAO P X, KONG D X. Two-stage optimization of airport ferry service delay considering flight uncertainty[J]. European Journal of Operational Research, 2023, 307(3): 1103-1116.

[10] GUO W A, XU P, ZHAO Z, et al. Scheduling for airport baggage transport vehicles based on diversity enhancement genetic algorithm[J]. Natural Computing, 2020, 19(4): 663-672.

[11] 袁成斌. 基于启发式算法的机场除冰资源调度策略研究[D]. 武汉: 华中科技大学, 2020: 51-63.
YUAN C B. Research on airport deicing resource scheduling strategy based on heuristic algorithm[D]. Wuhan: Huazhong University of Science and Technology, 2020: 51-63(in Chinese).

[12] ZHAO P X, GAO W Q, HAN X, et al. Bi-objective collaborative scheduling optimization of airport ferry vehicle and tractor[J]. International Journal of Simulation Modelling, 2019, 18(2): 355-365.

[13] ZHU S R, SUN H J, GUO X. Cooperative scheduling optimization for ground-handling vehicles by considering flights' uncertainty[J]. Computers & Industrial Engineering, 2022, 169: 108092.

[14] GUIMARANS D, PADRÓN S. A stochastic approach for planning airport ground support resources[J]. International Transactions in Operational Research, 2022, 29(6): 3316-3345.

[15] 吴枕. 大型机场地面保障车辆协同与动态调度研究[D]. 北京: 北京交通大学, 2021: 27-38.
WU Z. Research on coordination and dynamic scheduling of ground support vehicles in large airports[D]. Beijing: Beijing Jiaotong University, 2021: 27-38(in Chinese) .

[16] 蔡畅. 枢纽机场摆渡车动态调度方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2021: 31-44.
CAI C. Research on dynamic dispatching method of hub airport car ferry[D]. Harbin: Harbin Institute of Technology, 2021: 31-44(in Chinese) .

[17] XUE G Q, WANG Y, GUAN X Y, et al. A combined GA-TS algorithm for two-echelon dynamic vehicle routing with proactive satellite stations[J]. Computers & Industrial Engineering, 2022, 164: 107899.

[18] 王庞伟, 邓辉, 于洪斌, 等. 车路协同系统下区域路径实时决策方法[J]. 北京航空航天大学学报, 2019, 45(7): 1349-1360.
WANG P W, DENG H, YU H B, et al. Real-time regional path decision method in cooperative vehicle infrastructure system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(7): 1349-1360(in Chinese).

[19] 范厚明, 张跃光, 田攀俊, 等. 时变路网下异型车辆动态配置与路径优化[J]. 系统工程理论与实践, 2022, 42(2): 455-470.
FAN H M, ZHANG Y G, TIAN P J, et al. Dynamic vehicle routing problem of heterogeneous fleets with time-dependent networks[J]. Systems Engineering-Theory & Practice, 2022, 42(2): 455-470(in Chinese).

[20] CHEN S F, CHEN R, WANG G G, et al. An adaptive large neighborhood search heuristic for dynamic vehicle routing problems[J]. Computers & Electrical Engineering, 2018, 67: 596-607.

Dual-phase scheduling of apron support vehicles considering multi-vehicle coordination

JIANG Yu*, ZHANG Shimiao, CHEN Mingyang, ZHANG Wenjing, WU Weiwei

(College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: The coordination restrictions of several apron support services and dynamic flight information have resulted in increased operational demand on large airport apron support vehicles and more difficulties with vehicle scheduling. Considering the difference of vehicle operation constraints in three different modes of continuous operation, capacity-limited continuous operation and round-trip operation, a multi-vehicle cooperative apron support vehicle scheduling model is established with the goal of minimizing the number of vehicles and the total driving distance. The model is solved in two stages according to the actual operation of the airport. In the static stage, a local search non-dominated corting genetic algorithms Ⅱ (LS-NSGA II) algorithm integrating local search is designed. In the dynamic stage, a similar neighborhood search algorithm is designed. The static results show that the number of vehicles and the total driving distance are reduced by 18.9% and 8.9% respectively compared with the first-come-first-served. In comparison to the big neighborhood search technique, the dynamic results can keep the number of cars in the static scheduling plan while reducing the adjusted driving distance by 25.8%. The research results can provide a certain guiding significance for the scheduling management and decision-making of large airport apron support vehicles.

Keywords: apron support vehicles; two-stage scheduling; path planning; multi-vehicle collaboration; improved NSGA II

Received: 2023-06-21; **Accepted**: 2023-08-24; **Published Online**: 2023-09-05 15:00

URL: link.cnki.net/urlid/11.2625.V.20230905.1321.004

Foundation item: National Natural Science Foundation of China (U1933118)

* **Corresponding author**. E-mail: jiangyu07@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0341

基于部件特性图优化的民航发动机性能退化建模

郭庆[1,*]，黄启廉[1]，陈金亮[2]

(1. 中国民航大学 航空工程学院，天津 300300；2. 山东航空股份有限公司烟台分公司，烟台 264000)

摘 要：为从单元体层级给出民航发动机气路性能退化的理论依据，以CFM56-3发动机为研究对象，在使用特性图缩放法获取部件特性方程的基础之上，优化了风扇通用特性图缩放基准点的选取过程，提出特性图的曲面拟合方法，构建出稳态工况下符合特定转速条件的发动机部件级基准性能模型。通过引入故障因子生成故障系数矩阵，计算发动机监控参数随部件效率下降的偏离量，并与美国通用电气公司培训手册的指印图资料作对比，验证了融合风扇特性图缩放基准点优化及特性图曲面拟合方法的发动机稳态性能模型在气路性能退化分析中具有较好的精度和使用前景。

关 键 词：民航发动机；气路分析；部件级建模；缩放基准点优化；曲面拟合

中图分类号：V263.6

文献标志码：A **文章编号**：1001-5965(2025)06-1935-11

安全是民航发展的基石，民航发动机作为飞机的“心脏”，工作环境较为恶劣，为保证飞行安全的同时降低发动机的运维成本，对发动机进行状态监控与故障诊断，确保发动机在全寿命周期内安全可靠地工作至关重要[1-2]。

气路分析作为民航发动机状态监控与故障诊断的重要组成部分，目前主要有基于模型和基于数据驱动2种方法[3-4]，且基于模型的诊断方法根据对模型的处理过程不同又可分为基于线性模型的诊断方法和基于非线性模型的诊断方法。其中，基于线性模型的诊断方法起源较早，20世纪70年代初期，Urban[5]提出气路分析法，并对小偏差方法和故障方程开展了一系列研究[6]。范作民等[7-8]在其基础之上深入研究故障因子理论，并提出故障主因子模型，有效克服了多重共线性的影响。基于非线性模型的诊断方法的提出有效解决了发动机非线性系统建模困难的问题，Stamatis等[9]通过在部件特性图中引入特性修正因子，实现了非线性发动机模型的自适应建模，提高了实际发动机的性能模拟精度。

基于数据驱动的诊断方法一般需要大量的发动机运维数据作为训练样本或用于构建知识库，并且该方法缺乏与发动机机理层面的结合，对故障诊断结果的解释通常缺乏说服力。基于模型的诊断方法对数据的需求量较小，并且可从发动机原理的角度进行建模仿真[10-11]。气路分析法是一种基于线性模型的诊断方法，在采用该方法进行气路性能退化分析时，由于发动机在特定转速控制条件下部件特性参数之间的关系较难获得，因此，对发动机基准性能模型的构建往往较为困难。此外，为有效结合发动机的实际工作环境，在对风扇单元体进行研究时，需要综合考虑民航发动机风扇叶片叶尖处的超音速气流对发动机建模的影响。

针对上述2个问题，为结合发动机机理层面的基础理论推导构建发动机部件级模型，本文首先建立起风扇叶片进口气流轴向马赫数和低压转子转速间的相关关系，优化风扇通用特性图缩放基准点的坐标选取，提高了特性图缩放法的精度，使缩放

收稿日期：2023-06-12；**录用日期**：2023-08-11；**网络出版时间**：2023-08-22 14:40
网络出版地址：link.cnki.net/urlid/11.2625.V.20230822.0851.001
基金项目：2022年天津市研究生科研创新项目(2022SKY156)
*通信作者. E-mail：qguocauc@sina.com

引用格式：郭庆，黄启廉，陈金亮. 基于部件特性图优化的民航发动机性能退化建模[J]. 北京航空航天大学学报，2025，51(6)：1935-1945. GUO Q，HUANG Q L，CHEN J L. Performance degradation modelling of civil aviation engines based on component characteristic map optimization[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1935-1945 (in Chinese).

后的风扇特性图更符合工程实际；其次，本文在发动机数学模型中引入了由特性图缩放法得到的特定型号发动机部件特性图，并通过特性图的三维曲面拟合求解出特定转速控制条件下的部件特性方程，有效构建起反映转速特征的发动机基准性能模型，为发动机气路性能退化研究提供了相关的理论和模型支持。

1 发动机稳态性能建模方法

1.1 发动机基准性能模型的构建

部件级性能建模是采用气路分析法进行故障诊断的基础，是联系部件性能和整机性能的桥梁。本文构建的发动机基准性能模型即发动机部件级性能模型，主要包括发动机的外界条件、控制条件、部件匹配方程和部件特性方程4部分。其中，外界条件和控制条件取决于发动机的运行环境和具体型号，一般可从发动机的试车数据中获取；部件匹配方程表征稳态条件下压气机和涡轮的共同工作，主要由流量连续、功率平衡及能量平衡等方程组成，一般由发动机原理及空气动力学的相关理论推导得到；部件特性方程反映了压气机和涡轮自身工作特性，一般从部件特性图中得到，该部分为本文的研究重点。

部件特性方程由发动机在某特定工作转速下的部件特性线拟合获取，并随着控制条件的变化进行动态调整。部件特性线为部件特性图中的各条等换算转速线，主要包含了发动机循环参考点（cycle reference point, CRP）和各个非设计点的相对换算转速、增压比、效率和流量。为方便从特性图中获取发动机不同工作点的部件特性参数，在流量阻塞线和喘振边界线之间对各部件特性线进行线性插值，将不同转速线上的插值点平滑连接，作若干条与等换算转速线相交的曲线，并用辅助坐标为0～1之间的β线来定义[12]。故每个工作点的部件特性参数可由β线和相对换算转速N_C确定。

发动机部件特性图属于原始设备制造厂商的保密资料，一般无法直接获取[13]。且同一型号的不同发动机间会由于制造和装配的偏差而导致部件特性图间呈现个体差异。为获得特定型号发动机的部件特性图，筛选增压比相近的成熟型号发动机的部件特性图，视作与特定型号发动机部件特性图具有相似特征的通用特性图，采用特性图缩放法[14]对其进行缩放。具体方法如下：首先，根据特定型号发动机的试车数据随机选取高换算转速下的某个运行工作点作为循环参考点，并在通用特性图上选取对应转速下$\beta=0.5$的坐标点作为缩放基准点（scaling reference point, SRP）；接着，由式(1)计算部件特性缩放因子；最后，采用缩放因子对通用特性图的特性线形状及位置沿β线进行缩放，实现缩放基准点与循环参考点部件性能参数的匹配，绘制出特定发动机的部件特性图。该方法的实现过程如图1所示，其中，风扇通用特性图中缩放基准点的选取位置需要进行优化，N_{SRP}为缩放基准点的相对换算转速，N_{CRP}为循环参考点的相对换算转速。

$$\begin{cases} F_W=\dfrac{W_{CRP}}{W_{SRP}} \\ F_\pi=\dfrac{\pi_{CRP}-1}{\pi_{SRP}-1} \\ F_\eta=\dfrac{\eta_{CRP}}{\eta_{SRP}} \end{cases} \tag{1}$$

式中：F为部件特性缩放因子；W为流量；π为压比；η为效率；下标CPR表示循环参考点；下标SRP表示缩放基准点。

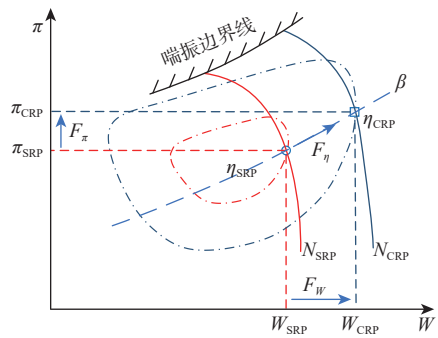


图1 特性图缩放法

Fig. 1 Characteristic map scaling method

工程上多采用特性曲线的插值法来获得部件特性参数之间的关系，张冬阳[15]提出了压气机特性曲线的两步一元多项式拟合，解决了传统插值方法中存在的精度低和工作量大的问题。然而这种特性图的平面曲线拟合方法中所包含的部件特性信息有限，无法直接在部件特性方程中反映出转子转速与部件特性间的关系。欧阳辉和朱之丽[16]从发动机单元体层级出发，构建了小型涡喷发动机的稳态性能计算模型，并提出了改型设计方案。为将部件级性能建模方法应用于民航涡扇发动机上，对风扇通用特性图缩放基准点的选取位置进行优化，进一步提高特性图缩放法的精度，并利用部件特性曲线的形状和位置随转子转速渐进变化的特点，对缩放后的发动机部件特性图进行三维曲面拟合，代入转速控制条件对拟合曲面方程进行降维处理，得到一组流量、压比和效率之间的部件特性参数隐函数关系，实现任意转速条件下部件特性线的提取。

由部件特性方程与部件匹配方程即可构建特定外界条件和控制条件下的发动机基准性能模型，求解模型得到发动机各性能参数或监控参数的基

准值或健康值, 用于构建发动机的故障模型。

1.2　发动机故障模型的构建

发动机的实际故障通常表现在部件结构尺寸的变化上, 此变化导致部件性能产生变化, 并由征兆量的变化直接体现[17]。反映部件性能变化的部件特性参数主要为流经该部件的气体流量和部件效率。故障因子是能够表征发动机各部件自身性能是否处于故障状态的变量, 主要包括各部件的流量故障因子和效率故障因子, 若故障因子为 0, 则表示该部件特性参数正常。

假设发动机故障处于小偏差范围内, 为研究部件性能变化与征兆量之间的映射关系, 在部件特性方程中分别引入各部件特性故障因子, 采用隐函数平移法则对部件特性线进行平移, 得到以故障因子为自变量, 以征兆量或部件特性参数为响应的发动机故障模型。该线性模型中的故障方程可以用矩阵形式表示, 如下:

$$\delta \boldsymbol{Y} = \boldsymbol{A}\delta \tilde{\boldsymbol{X}} \tag{2}$$

式中: $\delta \boldsymbol{Y} = (\boldsymbol{Y}' - \boldsymbol{Y}_0)/\boldsymbol{Y}_0$为 $m\times 1$的响应偏离向量, $\boldsymbol{Y}'$为故障响应向量, $\boldsymbol{Y}_0$为健康响应向量, m为响应参数数量, δ为相对偏差量; $\delta \tilde{\boldsymbol{X}} = \Delta \tilde{\boldsymbol{X}} / \boldsymbol{X}_0$为 $n\times 1$的故障因子相对偏差向量, $\Delta \tilde{\boldsymbol{X}}$为部件特性故障因子, $\boldsymbol{X}_0$为部件特性参数健康向量, n为故障因子数量; $\boldsymbol{A} = [a_{ij}]$为 $m\times n$的故障系数矩阵, a_{ij}为某部件特性故障因子变化一个单位且其他故障因子均为 0 时某响应的变化量。

根据故障系数的定义, 通过数值线性化方法, 利用有限差分代替微分分别求解每个故障系数, 如下:

$$a_{ij} = \frac{(\boldsymbol{Y}' - \boldsymbol{Y}_0)\boldsymbol{X}_0}{\boldsymbol{Y}_0 \Delta \tilde{\boldsymbol{X}}} \tag{3}$$

在中国民用航空发动机维修领域中, 发动机制造商为更加直观地展示部件性能参数与发动机监控参数之间的相关关系, 选取了排气温度(exhaust gas temperature, EGT)、燃油流量 W_{F}、高压转子换算转速 $N_{2\mathrm{C}}$这 3 个征兆量[17]绘制 CFM56-3 系列发动机的指印图, 通过对比指印图中征兆量间的比例关系与实际监控数据中征兆量间的比例关系, 进行部件级故障隔离与部件性能退化分析。

2　CFM56-3 发动机建模与方法验证

本文采用某航空公司提供的 CFM56-3 发动机试车符合性报告(correlation test report, CTR)构建该型发动机部件级性能模型, 结合建模所得的发动机基准状况计算监控参数EGT、W_{F}及$N_{2\mathrm{C}}$随各部件效率下降1%的偏离量, 并通过美国通用电气公司故障诊断培训手册提供的指印图资料进一步验证基准模型的精准度。技术路线图如图 2 所示。

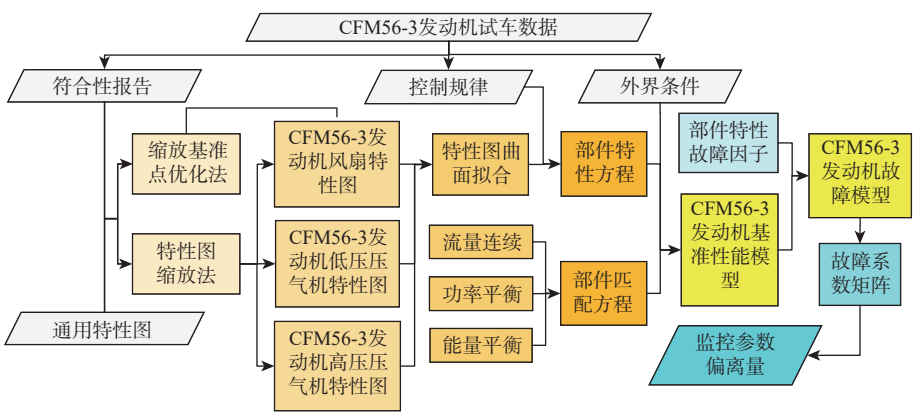


图 2　技术路线图

Fig. 2　Technology roadmap

2.1　CFM56-3 发动机部件特性图缩放基准点优化

为建立 CFM56-3 发动机的基准性能模型, 先采用特性图缩放法对通用特性图进行缩放, 生成该型发动机的特定特性图。其中, 考虑到民航发动机风扇叶片叶尖处可能为超音速气流, 根据气流速度三角形相似, 在高换算转速区, 风扇进口气流的轴向马赫数和低压转子相对换算转速成正比。又由于风扇特性图高换算转速线远离喘振边界端的形状一般较为陡峭[18], 故同一条转速线该区域的空气流量及对应的轴向马赫数相近, 为确定气流轴向马赫数 Ma_x, 在通用特性图相对换算转速线 $N_{\mathrm{C}} = 1.0$和 $N_{\mathrm{C}} = 1.1$笔直段上任意选取 2 个工作点 A 和 B, 如图 3 所示, 其中, $W_{2\mathrm{C}}$ 为相对换算转速下的相对换算流量。工作点 A 和 B 之间的关系如下:

$$\frac{Ma_{x,B}}{Ma_{x,A}} = \frac{N_{\mathrm{C},B}}{N_{\mathrm{C},A}} \tag{4}$$

流量计算式为

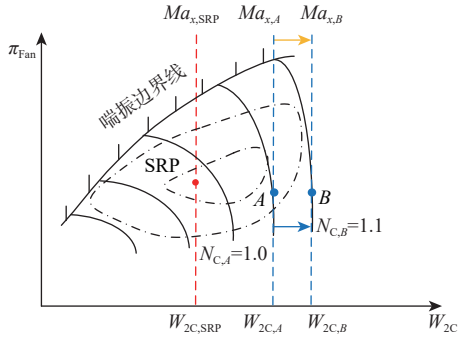


图3 通用特性图缩放基准点选取

Fig. 3 Selection of general characteristic map scaling reference point

$$\begin{cases} W = Z\dfrac{P^*}{\sqrt{T^*}}A_{\mathrm{r}}q(Ma_x) \\ q(Ma_x) = Ma_x\left(\dfrac{2}{\gamma+1}+\dfrac{\gamma-1}{\gamma+1}Ma_x^2\right)^{-\frac{\gamma+1}{2(\gamma-1)}} \end{cases} \tag{5}$$

式中：$Z=0.040\,4$, s·$\sqrt{\mathrm{k}}$/m；P^*为总压；T^*为总温；A_{r}为流场截面积；$q(Ma_x)$为密流函数；γ为绝热指数。

将式(5)推导得到的方程通过指数拟合法拟合出工作点A和B与缩放基准点SRP之间轴向马赫数Ma_x与空气流量W_{SRP}的映射关系：

$$\begin{cases} Ma_{x,A} = 6.0\times10^{13}\mathrm{e}^{-44.9W_{\mathrm{SRP}}}+3.1\mathrm{e}^{-1.8W_{\mathrm{SRP}}} \\ Ma_{x,B} = 1.2\times10^{14}\mathrm{e}^{-44.0W_{\mathrm{SRP}}}+3.0\mathrm{e}^{-1.7W_{\mathrm{SRP}}} \end{cases} \tag{6}$$

选取与CFM56-3发动机风扇单元体目标特性图上CRP增压比相近的部件特性图作为通用特性图[19]。根据Kurzke和Halliwell[18]的结论，通用特性图上的SRP和CFM56-3发动机风扇特性图上的CRP之间的气流轴向马赫数Ma_x相等。故由CTR数据及式(5)可初步估算出CRP的$Ma_x=0.5$。结合式(4)和式(6)可得一个仅包含参数W_{SRP}的适应度函数$f(W_{\mathrm{SRP}})$，如式(7)所示。采用标准粒子群算法[20-21]寻优，适应度进化曲线如图4所示，最优解$W_{\mathrm{SRP}}=0.81$。由此确定SRP在通用特性图中的横坐标。又缩放基准点需尽量选在特性图中效率较高且具有一定的喘振裕度的区域，故最终选取$\beta=0.4$、$N_{\mathrm{C}}=0.8$的工作点作为通用特性图中的SRP。

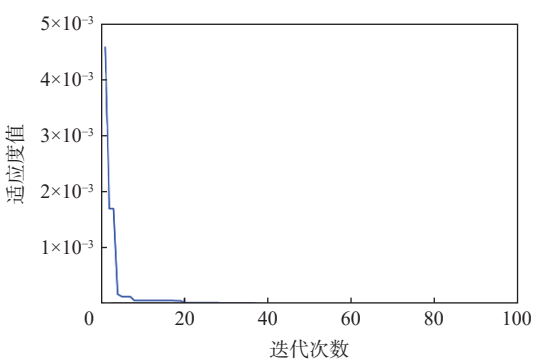


图4 适应度进化曲线

Fig. 4 Fitness evolution curve

$$f(W_{\mathrm{SRP}}) = \left|\frac{Ma_{x,B}}{Ma_{x,A}}-1.1\right| \tag{7}$$

根据所选取缩放基准点，由式(1)计算得到风扇单元体流量、压比和效率缩放因子分别为$F_W=387.40$、$F_\pi=1.70$和$F_\eta=0.92$，利用缩放因子对通用特性图进行缩放，所得的CFM56-3发动机风扇特性如图5所示，其中，$W_{2\mathrm{Rstd}}$为国际标准大气修正流量，N_{C1}为低压转子相对换算转速。同理可得CFM56-3发动机低压压气机和高压压气机部件特性。

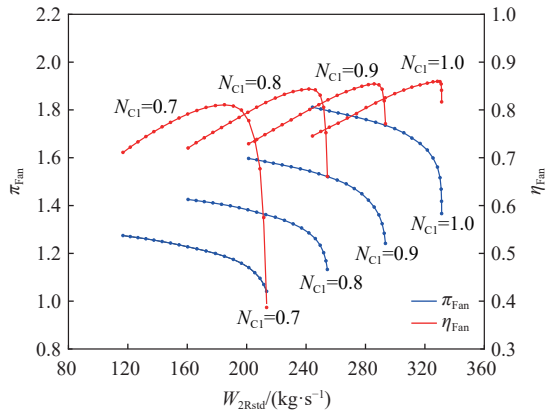


图5 CFM56-3发动机风扇特性

Fig. 5 Fan characteristic of CFM56-3 engine

2.2 CFM56-3发动机基准性能模型构建

在避免欠拟合现象及简化模型计算的前提下，对缩放后的部件特性图进行多项式曲面拟合，得到CFM56-3发动机各部件特性曲面。其中，风扇单元体的流量-增压比-转速特性曲面如图6所示，其中，圆点表示特征图不同转速线上工作点的位置。

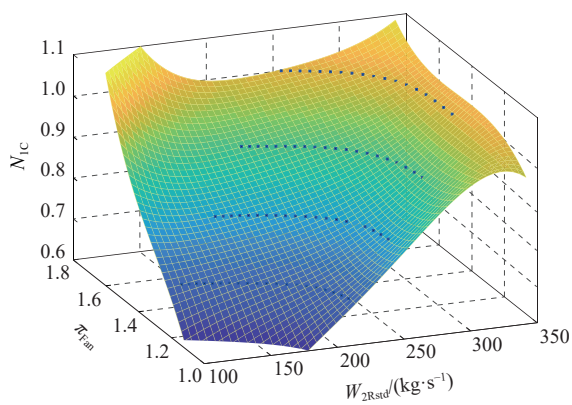


图6 风扇部件特性曲面

Fig. 6 Fan component characteristic surface

由于CFM56-3发动机的控制条件是低压转子转速$N_{1\mathrm{C}}$等于常数，故通过在部件特性曲面中设定参数$N_{1\mathrm{C}}$即可拟合出特定转速下的部件特性方程。为建立一个简化的封闭发动机数学模型，实现未知数个数与方程个数的匹配，对模型做出如下假设：

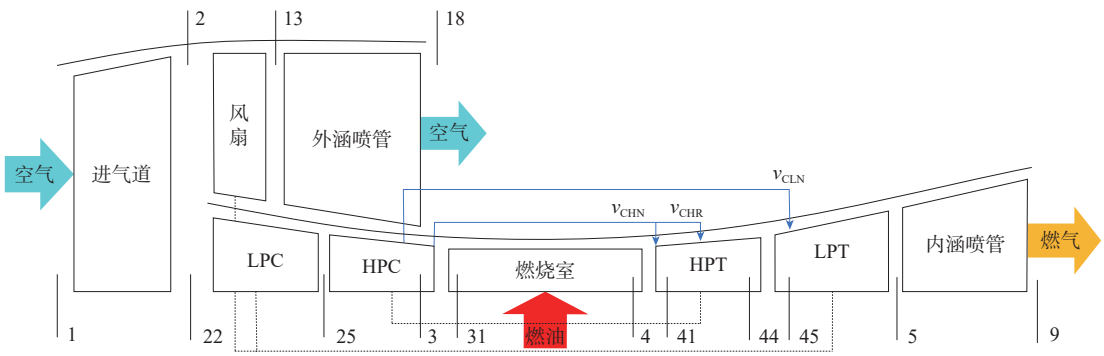
1) 将发动机中实际空气的三维非定常黏性流动简化为理想气体的一维定常无黏性流动。

2)将内涵风扇部分作为低压压气机的一部分,故外涵风扇(后文统称为风扇)和低压压气机进口气流总温和总压相同。

3) 仅简单考虑由高压压气机五级引向低压涡轮导向器的冷却空气、由高压压气机末级引向高压涡轮导向器及高压轴的冷却空气,忽略其他引气量。

4) 发动机在中高转速下运转,高、低压涡轮导向器最小截面及尾喷管出口截面处于临界或超临界状态。

结合假设 2,本文建立的 CFM56-3 发动机气路简化模型如图 7 所示。模型的输入参数及该型发动机的构型数据[18,22-23]如表 1 所示。其中,建模选取的外界条件与指印图相同,并将飞行高度 $H=10.668$ km,飞行马赫数 $Ma=0.76$结合进气道总压恢复系数σ_{I}进一步换算为风扇进口空气的总温 T_2^*和总压P_2^*。



LPC: 低压压气机, HPC: 高压压气机, HPT: 高压涡轮, LPT: 低压涡轮, 各数字: 发动机的位置, 1: 进气道入口, 2: 风扇入口, 13: 风扇出口, 18: 外涵喷管出口, 22: LPC 入口, 25: HPC 入口, 3: HPC 五级出口, 31: HPC 出口, 4: 燃烧室出口, 41: HPT 导向器出口, 44: HPT 出口, 43: LPT 导向器出口, 5: LPT 出口, 9: 内涵喷管出口。

图 7　CFM56-3 发动机气路简化模型

Fig. 7　Simplified gas path model of CFM56-3 engine

表 1　参数信息

Table 1　Parameters information

参数	数值
涵道比R_{BPR}	5.024 5
气体常数R/(J·(kg·K)$^{-1}$)	287.0
燃油热值H_{U}/10^7 (J·kg^{-1})	4.277
低压转子转速$N_{1\mathrm{C}}$	0.930 0
燃烧效率η_{B}	0.996 0
风扇进口空气总温T_2^*/K	244.14
风扇进口空气总压P_2^*/10^4 Pa	3.461 5
国际标准大气总温T_{ISA}^*/K	288.15
国际标准大气总压P_{ISA}^*/10^5 Pa	1.013 25
高压涡轮轴机械效率η_{MH}	0.99
低压涡轮轴机械效率η_{ML}	1.00
进气道总压恢复系数σ_{I}	0.99
中介机匣总压恢复系数σ_{24}	0.98
HPT转子轴承冷却空气比例ν_{CHR}	0.05
HPT进口导向叶片冷却空气比例ν_{CHN}	0.06
LPT进口导向叶片冷却空气比例ν_{CLN}	0.02

根据表 1 中的已知参数,利用发动机各单元体间的流量连续、功率平衡和能量平衡关系推导部件匹配方程,如式(8)～式(29)所示。其中,由假设 3 可得流经燃烧室的高温燃气、冷却空气及掺混后燃气间的能量平衡方程,如式(17)、式(20)和式(22)所示。式(25)～式(27)为实际空气流量与标准大气条件下修正空气流量间的换算方程。

$$R_{\mathrm{BPR}}=\frac{W_{13}}{W_{22}} \tag{8}$$

$$W_2=W_{13}+W_{22} \tag{9}$$

$$T_{13}^*-T_2^*=T_2^*\frac{\pi_{\mathrm{Fan}}^{m_{y1}}-1}{\eta_{\mathrm{Fan}}} \tag{10}$$

$$T_{24}^*-T_{22}^*=T_{22}^*\frac{\pi_{\mathrm{LPC}}^{m_{y2}}-1}{\eta_{\mathrm{LPC}}} \tag{11}$$

$$T_3^*-T_{25}^*=T_{25}^*\frac{\pi_{\mathrm{HPC}}^{m_{y3}}-1}{\eta_{\mathrm{HPC}}} \tag{12}$$

$$W_{25}(1-\nu_{\mathrm{CLN}})=W_3 \tag{13}$$

$$W_3(1-\nu_{\mathrm{CHN}}-\nu_{\mathrm{CHR}})=W_{31} \tag{14}$$

$$W_{31}+W_F=W_4 \tag{15}$$

$$\frac{1}{m_{y5}}RT_4^*W_4-\frac{1}{m_{y4}}RT_3^*W_{31}=W_{\mathrm{F}}H_{\mathrm{U}}\eta_{\mathrm{B}} \tag{16}$$

$$T_{41}^*W_{41}=T_4^*W_4+T_3^*\nu_{\mathrm{CHN}}W_{25} \tag{17}$$

$$W_4+\nu_{\mathrm{CHN}}W_{25}=W_{41} \tag{18}$$

$$T_{41}^*-T_{43}^*=\eta_{\mathrm{HPT}}T_{41}^*\left(1-\frac{1}{\pi_{\mathrm{HPT}}^{m_{y5}}}\right) \tag{19}$$

$$T_{44}^{*}W_{44}=T_{43}^{*}W_{41}+T_{3}^{*}\nu_{\mathrm{CHR}}W_{25} \tag{20}$$

$$W_{41}+\nu_{\mathrm{CHR}}W_{25}=W_{44} \tag{21}$$

$$T_{45}^{*}W_{45}=T_{44}^{*}W_{44}+\frac{1}{3}\left(2T_{3}^{*}+T_{25}^{*}\right)\nu_{\mathrm{CLN}}W_{25} \tag{22}$$

$$W_{44}+\nu_{\mathrm{CLN}}W_{25}=W_{45} \tag{23}$$

$$T_{45}^{*}-T_{5}^{*}=\eta_{\mathrm{LPT}}T_{45}^{*}\left(1-\frac{1}{\pi_{\mathrm{LPT}}^{m_{\gamma 6}}}\right) \tag{24}$$

$$W_{2}=W_{2\mathrm{Rstd}}\sqrt{\frac{T_{\mathrm{ISA}}^{*}}{T_{2}^{*}}}\cdot\frac{P_{2}^{*}}{P_{\mathrm{ISA}}^{*}} \tag{25}$$

$$W_{22}=W_{22\mathrm{Rstd}}\sqrt{\frac{T_{\mathrm{ISA}}^{*}}{T_{2}^{*}}}\cdot\frac{P_{2}^{*}}{P_{\mathrm{ISA}}^{*}} \tag{26}$$

$$W_{22}=W_{25\mathrm{Rstd}}\sqrt{\frac{T_{\mathrm{ISA}}^{*}}{T_{25}^{*}}}\cdot\frac{P_{2}^{*}\pi_{\mathrm{LPC}}\sigma_{24}}{P_{\mathrm{ISA}}^{*}} \tag{27}$$

$$\frac{1}{m_{\gamma 3}}\left(T_{3}^{*}-T_{25}^{*}\right)W_{25}=\frac{1}{m_{\gamma 5}}\left(T_{41}^{*}-T_{43}^{*}\right)W_{41}\eta_{\mathrm{MH}} \tag{28}$$

$$\frac{1}{m_{\gamma 1}}\left(T_{13}^{*}-T_{2}^{*}\right)W_{13}+\frac{1}{m_{\gamma 2}}\left(T_{24}^{*}-T_{22}^{*}\right)W_{22}=\frac{1}{m_{\gamma 6}}\left(T_{45}^{*}-T_{5}^{*}\right)W_{45}\eta_{\mathrm{ML}} \tag{29}$$

式中:T^{*}为空气或燃气总温,下标的数字表示发动机的不同站位;W为空气或燃气流量;W_{Rstd}为国际标准大气修正流量;W_{F}为燃油流量;ν为压比;η为效率;下标Fan表示风扇,下标ISA表示国际标准大气;m_{γ}为绝热系数,且$m_{\gamma 1}=0.286\,2$,$m_{\gamma 2}=0.285\,8$,$m_{\gamma 3}=0.279\,6$,$m_{\gamma 4}=0.270\,1$,$m_{\gamma 5}=0.239\,0$,$m_{\gamma 6}=0.253\,0$。

部件特性方程如式(30)～式(40)所示。其中,式(30)～式(38)由特性曲面拟合得到;式(39)和式(40)为涡轮效率特性方程,由假设4,涡轮效率特性变化对压气机和涡轮共同工作的敏感度较低,可取为一常数[8,24]。

$$\begin{bmatrix}1 & W_{2\mathrm{Rstd}} & W_{2\mathrm{Rstd}}^{2}\end{bmatrix}\begin{bmatrix}-1.88\times10^{-1} & 5.18\times10^{-1} & 1.90\times10^{-1}\\ 2.49\times10^{-3} & -3.00\times10^{-3} & 0\\ 5.90\times10^{-6} & 0 & 0\end{bmatrix}\begin{bmatrix}1\\ \pi_{\mathrm{Fan}}\\ \pi_{\mathrm{Fan}}^{2}\end{bmatrix}=N_{1\Omega} \tag{30}$$

$$\begin{bmatrix}1 & \eta_{\mathrm{Fan}} & \eta_{\mathrm{Fan}}^{2} & \eta_{\mathrm{Fan}}^{3}\end{bmatrix}\begin{bmatrix}7.19 & -16.93 & 9.98 & -4.42\times10^{-1}\\ 6.77 & 7.69 & -10.33 & 0\\ -19.14 & 14.16 & 0 & 0\\ 1.40 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}1\\ \pi_{\mathrm{Fan}}\\ \pi_{\mathrm{Fan}}^{2}\\ \pi_{\mathrm{Fan}}^{3}\end{bmatrix}=N_{1\Omega} \tag{31}$$

$$\begin{bmatrix}1 & W_{2\mathrm{Rstd}} & W_{2\mathrm{Rstd}}^{2}\end{bmatrix}\begin{bmatrix}2.73\times10^{-1} & 2.37 & -2.99\\ -2.87\times10^{-3} & 7.50\times10^{-3} & 0\\ -2.42\times10^{-6} & 0 & 0\end{bmatrix}\begin{bmatrix}1\\ \eta_{\mathrm{Fan}}\\ \eta_{\mathrm{Fan}}^{2}\end{bmatrix}=N_{1\Omega} \tag{32}$$

$$\begin{bmatrix}1 & W_{22\mathrm{Rstd}} & W_{22\mathrm{Rstd}}^{2} & W_{22\mathrm{Rstd}}^{3}\end{bmatrix}\begin{bmatrix}-8.22\times10^{-1} & 1.07 & -4.31\times10^{-2} & 3.21\times10^{-2}\\ 3.14\times10^{-2} & -2.63\times10^{-2} & -1.32\times10^{-3} & 0\\ -3.37\times10^{-5} & 1.71\times10^{-4} & 0 & 0\\ 7.81\times10^{-7} & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}1\\ \pi_{\mathrm{LPC}}\\ \pi_{\mathrm{LPC}}^{2}\\ \pi_{\mathrm{LPC}}^{3}\end{bmatrix}=N_{1\Omega} \tag{33}$$

$$\begin{bmatrix}1 & \eta_{\mathrm{LPC}} & \eta_{\mathrm{LPC}}^{2} & \eta_{\mathrm{LPC}}^{3}\end{bmatrix}\begin{bmatrix}-41.76 & 11.82 & -1.73 & -5.02\times10^{-2}\\ 125.40 & -17.98 & 2.04 & 0\\ -130.80 & 6.06 & 0 & 0\\ 47.59 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}1\\ \pi_{\mathrm{LPC}}\\ \pi_{\mathrm{LPC}}^{2}\\ \pi_{\mathrm{LPC}}^{3}\end{bmatrix}=N_{1\Omega} \tag{34}$$

$$\begin{bmatrix}1 & W_{22\mathrm{Rstd}} & W_{22\mathrm{Rstd}}^{2}\end{bmatrix}\begin{bmatrix}-2.64 & 4.77 & -1.62\\ 4.03\times10^{-2} & -2.94\times10^{-2} & 0\\ -1.19\times10^{-5} & 0 & 0\end{bmatrix}\begin{bmatrix}1\\ \eta_{\mathrm{LPC}}\\ \eta_{\mathrm{LPC}}^{2}\end{bmatrix}=N_{1\Omega} \tag{35}$$

$$\begin{bmatrix}1 & W_{25\mathrm{Rstd}} & W_{25\mathrm{Rstd}}^{2} & W_{25\mathrm{Rstd}}^{3}\end{bmatrix}\begin{bmatrix}2.02\times10^{-1} & 1.30\times10^{-1} & 9.95\times10^{-3} & 1.79\times10^{-4}\\ 2.30\times10^{-2} & -1.67\times10^{-2} & -5.80\times10^{-4} & 0\\ 1.87\times10^{-3} & 5.28\times10^{-4} & 0 & 0\\ -7.23\times10^{-5} & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}1\\ \pi_{\mathrm{HPC}}\\ \pi_{\mathrm{HPC}}^{2}\\ \pi_{\mathrm{HPC}}^{3}\end{bmatrix}=N_{2\Omega} \tag{36}$$

$$\begin{bmatrix}1 & \eta_{\mathrm{HPC}} & \eta_{\mathrm{HPC}}^{2} & \eta_{\mathrm{HPC}}^{3}\end{bmatrix}\begin{bmatrix}60.50 & 1.88 & 3.77\times10^{-2} & 6.98\times10^{-4}\\ -249.80 & -5.58 & -7.25\times10^{-2} & 0\\ 347.70 & 4.35 & 0 & 0\\ -162.10 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}1\\ \pi_{\mathrm{HPC}}\\ \pi_{\mathrm{HPC}}^{2}\\ \pi_{\mathrm{HPC}}^{3}\end{bmatrix}=N_{2\Omega} \tag{37}$$

$$\begin{bmatrix}1 & W_{25\mathrm{Rstd}} & W_{25\mathrm{Rstd}}^2 & W_{25\mathrm{Rstd}}^3\end{bmatrix}\begin{bmatrix}9.46 & -45.11 & 83.08 & -54.54\\ 3.80\times10^{-4} & -8.05\times10^{-1} & 1.51 & 0\\ 1.17\times10^{-2} & -3.43\times10^{-2} & 0 & 0\\ 2.59\times10^{-4} & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}1\\ \eta_{\mathrm{HPC}}\\ \eta_{\mathrm{HPC}}^2\\ \eta_{\mathrm{HPC}}^3\end{bmatrix}=N_{2\Omega} \tag{38}$$

$$\eta_{\mathrm{HPT}}=0.821\,5 \tag{39}$$

$$\eta_{\mathrm{LPT}}=0.886\,0 \tag{40}$$

式中：N_Ω为系数，$N_{1\Omega}=N_{1\mathrm{C}}(288.15/T_2^*)^{0.5}$，$N_{2\Omega}=N_{2\mathrm{C}}(369.92/T_{25}^*)^{0.5}$。

式(8)～式(40)组成了 CFM56-3 发动机在外界条件 $H=10.668\,\mathrm{km}$(35 000ft)、$Ma=0.76$，控制条件 $N_{1\mathrm{C}}=0.93$时的稳态性能模型，采用 MATLAB 软件的 Fsolve 函数求解该非线性方程组，得到各参数基准健康值，如表 2 所示。其中，CFM56-3 发动机的 EGT传感器位于低压涡轮第 2 级静叶[25]，其表达式为

$$T_{\mathrm{EGT}}^*=0.976\left[T_{45}^*-0.217\left(T_{45}^*-T_5^*\right)\right] \tag{41}$$

2.3　CFM56-3 发动机故障系数矩阵生成

在 CFM56-3 发动机基准性能模型的部件特性方程中引入效率故障因子和流量故障因子，通过特性曲面平移建立故障模型，采用数值线性化方法计算故障系数矩阵。

此处以引入高压压气机效率故障因子为例，替换基准性能模型中的式(37)和式(38)，得到的高压压气机效率故障特性如式(42)和式(43)所示。

仍采用 Fsolve 函数求解故障模型，得到与基准性能模型相同外界条件和控制条件下高压压气机效率故障因子对应的故障响应值。同理可得其他部件特性故障因子对应的故障响应值。根据表 2 中提供的基准健康响应值，由式(2)计算各部件效率和流量分别下降 1% 时各响应的偏离量，所得故障系数如表 3 所示，由这些故障系数a_{ij}组成矩阵$\boldsymbol{A}$。

$$\begin{bmatrix}1 & \eta'_{\mathrm{HPC}}-\Delta\tilde{\eta}_{\mathrm{HPC}} & \left(\eta'_{\mathrm{HPC}}-\Delta\tilde{\eta}_{\mathrm{HPC}}\right)^2 & \left(\eta'_{\mathrm{HPC}}-\Delta\tilde{\eta}_{\mathrm{HPC}}\right)^3\end{bmatrix}\cdot\begin{bmatrix}60.5 & 1.881 & 3.772\times10^{-2} & 6.976\times10^{-4}\\ -249.8 & -5.581 & -7.253\times10^{-2} & 0\\ 347.7 & 4.350 & 0 & 0\\ -162.1 & 0 & 0 & 0\end{bmatrix}\cdot\begin{bmatrix}1\\ \pi_{\mathrm{HPC}}\\ \pi_{\mathrm{HPC}}^2\\ \pi_{\mathrm{HPC}}^3\end{bmatrix}=N_{2\Omega} \tag{42}$$

表 2　CFM56-3 发动机参数基准值

Table 2　Parameter reference values of CFM56-3 engine

参数	基准值	参数	基准值
T_{13}^*/K	286.820 0	W_{F}/(kg·s^{-1})	0.380 0
T_{25}^*/K	317.240 0	W_4/(kg·s^{-1})	17.507 6
T_3^*/K	663.350 0	W_{41}/(kg·s^{-1})	18.693 4
T_4^*/K	1 345.718 8	W_{44}/(kg·s^{-1})	19.680 1
T_{41}^*/K	1 305.858 0	W_{45}/(kg·s^{-1})	20.075 2
T_{43}^*/K	991.862 8	π_{Fan}	1.660 0
T_{44}^*/K	976.215 3	π_{LPC}	2.248 0
T_{45}^*/K	967.715 9	π_{HPC}	10.702 0
T_5^*/K	717.994 7	π_{HPT}	4.256 6
W_{13}/(kg·s^{-1})	99.182 0	π_{LPT}	3.899 3
W_2/(kg·s^{-1})	118.921 0	η_{Fan}	0.893 1
$W_{2\mathrm{Rstd}}$/(kg·s^{-1})	320.420 0	η_{LPC}	0.870 0
W_{22}/(kg·s^{-1})	19.739 6	η_{HPC}	0.861 6
$W_{22\mathrm{Rstd}}$/(kg·s^{-1})	53.186 0	η_{HPT}	0.821 5
$W_{25\mathrm{Rstd}}$/(kg·s^{-1})	27.487 0	η_{LPT}	0.886 0
W_3/(kg·s^{-1})	19.345 0	$N_{2\mathrm{C}}$	0.909 9
W_{31}/(kg·s^{-1})	17.174 7	T_{EGT}^*/K	891.601 7

$$\begin{bmatrix}1 & W_{25\mathrm{Rstd}} & W_{25\mathrm{Rstd}}^2 & W_{25\mathrm{Rstd}}^3\end{bmatrix}\cdot\begin{bmatrix}9.456 & -45.11 & 83.08 & -54.54\\ 3.795\times10^{-4} & -8.052\times10^{-1} & 1.512 & 0\\ 1.165\times10^{-2} & -3.433\times10^{-2} & 0 & 0\\ 2.592\times10^{-4} & 0 & 0 & 0\end{bmatrix}\cdot\begin{bmatrix}1\\ \eta'_{\mathrm{HPC}}-\Delta\tilde{\eta}_{\mathrm{HPC}}\\ \left(\eta'_{\mathrm{HPC}}-\Delta\tilde{\eta}_{\mathrm{HPC}}\right)^2\\ \left(\eta'_{\mathrm{HPC}}-\Delta\tilde{\eta}_{\mathrm{HPC}}\right)^3\end{bmatrix}=N_{2\Omega} \tag{43}$$

式中：η'_{HPC}为 HPC 效率故障值；$\Delta\tilde{\eta}_{\mathrm{HPC}}$为 HPC 效率故障因子；$\Delta\tilde{\eta}_{\mathrm{HPC}}=-1\%\eta_{\mathrm{HPC}}$。

2.4　CFM56-3 发动机监控参数偏离量计算与指印图验证

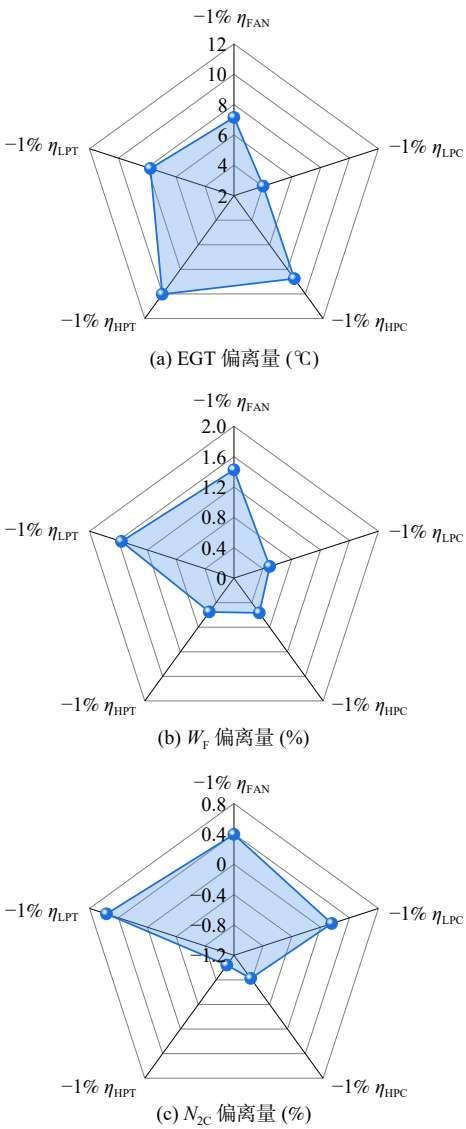
根据表 3 中各部件效率下降 1% 时EGT、W_{F}及$N_{2\mathrm{C}}$这 3 个监控参数的故障系数值，采用基准健康值进行相对化处理，计算该组参数在发生部件性能退化时相对基准性能模型的偏离量，如图 8 所示，其中，−1% 表示效率下降 1%。由图 8 可以直观得到 3 个监控参数对各部件性能退化的敏感程度。

为验证由发动机基准性能模型及故障系数矩阵计算得到的各监控参数偏离量的实验精度及应用可行性，采用由美国通用电气公司故障诊断培训手册提供的指印图资料进行对照。考虑到发动机

表3 不同监控参数偏离量和故障因子下的故障系数

Table 3 Fault coefficients under different monitoring parameter deviations and fault factors

监控参数偏移	故障系数a_{ij}/%							
	$\delta\tilde{W}_{2\mathrm{Rstd}}$	$\delta\tilde{\eta}_{\mathrm{Fan}}$	$\delta\tilde{W}_{22\mathrm{Rstd}}$	$\delta\tilde{\eta}_{\mathrm{LPC}}$	$\delta\tilde{W}_{25\mathrm{Rstd}}$	$\delta\tilde{\eta}_{\mathrm{HPC}}$	$\delta\tilde{\eta}_{\mathrm{HPT}}$	$\delta\tilde{\eta}_{\mathrm{LPT}}$
δT^*_{13}	−0.32	0.15	0.02	0	−0.01	0.02	0.03	−0.01
δT^*_{25}	0.21	−0.06	−0.39	0.23	−0.03	0.15	0.19	−0.09
δT^*_{3}	−0.63	0.21	−0.06	0.25	0.07	0.24	−0.37	0.31
δT^*_{4}	−0.99	0.71	0.19	0.40	−0.11	0.80	0.64	0.76
δT^*_{41}	−0.98	0.69	0.18	0.39	−0.11	0.79	0.61	0.74
δT^*_{43}	−0.93	0.80	0.19	0.45	−0.20	0.99	1.17	0.81
δT^*_{44}	−0.92	0.78	0.18	0.44	−0.19	0.97	1.12	0.79
δT^*_{45}	−0.92	0.78	0.18	0.44	−0.19	0.97	1.11	0.79
δT^*_{5}	−0.88	0.93	0.20	0.51	−0.22	1.07	1.20	1.29
$\delta\pi_{\mathrm{Fan}}$	−0.96	0	0.06	0	0	0.06	0.12	0
$\delta\pi_{\mathrm{LPC}}$	1.16	−0.40	−1.47	0.04	−0.13	0.80	1.02	−0.58
$\delta\pi_{\mathrm{HPC}}$	−2.79	0.92	1.00	0.07	0.28	−1.28	−1.77	1.32
$\delta\pi_{\mathrm{HPT}}$	−0.40	−0.52	−0.02	−0.24	0.15	−0.31	−0.36	−0.27
$\delta\pi_{\mathrm{LPT}}$	−0.31	−0.66	−0.06	−0.30	0.12	−0.30	−0.17	−0.55
δW_{F}	−2.30	1.43	−0.20	0.50	−0.14	0.57	0.55	1.56
$\delta N_{2\mathrm{C}}$	−1.19	0.39	0.27	0.16	0.79	−0.82	−1.04	0.56
T_{EGT}	−0.91	0.80	0.18	0.45	−0.20	0.98	1.13	0.87



(a) EGT 偏离量 (℃)

(b) W_{F} 偏离量 (%)

(c) $N_{2\mathrm{C}}$ 偏离量 (%)

图8 监控参数偏离量实验结果

Fig. 8 Experimental results of monitoring parameter deviation

所处的外界环境和工况条件对故障诊断结果的影响，在CFM56-3发动机指印图的实际工程应用中，根据待诊断发动机与指印图中EGT、W_{F}及$N_{2\mathrm{C}}$这3个监控参数偏离方向初步定位发生故障的单元体是属于高压系统或低压系统，因此，在验证过程中应重点关注实验与指印图所得监控参数偏离方向的一致性。由图8可知，在实验结果中只有HPC和HPT的效率下降会导致监控参数$N_{2\mathrm{C}}$降低，其余部件的效率下降都会使3个监控参数值增加，此偏离方向与指印图一致。

根据工程经验，当待诊断发动机的EGT、W_{F}及$N_{2\mathrm{C}}$这3个监控参数的偏离方向与指印图中所示的偏离方向一致时，常根据EGT、W_{F}这2个监控参数偏离量之间的比例关系与指印图中的对应情况，将待诊断发动机的故障隔离至单元体。因此，为进一步验证实验结果的工程实用性，分别计算实验结果与指印图中各部件效率下降导致监控参数偏离量的比例关系如表4所示，由表可知，实验与指印图中EGT、W_{F}这2个监控参数偏离量之间的比例关系基本相近，但在HPC和HPT发生性能退化时存在一定的实验误差。这是由于高压系统工作环境复杂多变，并且在实际故障诊断中涉及到具体的引气量，而本文中所采用的引气量源于表1中该型发动机的参考值。在实际应用过程中，高压系统是否发生性能衰退可以由EGT、W_{F}及$N_{2\mathrm{C}}$这3个监控参数的偏离方向进行直接判断，因此，这种误差是可接受的。由表4可知，HPC与HPT效率下降导致的EGT、W_{F}及$N_{2\mathrm{C}}$偏离量比例关系都基本相同，因此，若要具体区分是HPC还是HPT发生故障，应结

合表 3 中其他征兆量的变化或孔探等无损检测措施进行更为精确的故障定位。

表 4　监控参数偏离量比例关系

Table 4　Proportional relationship of monitoring parameters deviation

监控参数偏离量比例关系	$\delta T_{EGT}:\delta W_F:\delta N_{2C}$	
	实验	指印图
$-1\%\ \eta_{PAN}$	1:0.20:0.05	1:0.22:0.12
$-1\%\ \eta_{LPC}$	1:0.12:0.04	1:0.10:0.03
$-1\%\ \eta_{HPC}$	1:0.06:−0.09	1:0.10:−0.13
$-1\%\ \eta_{HPT}$	1:0.06:−0.10	1:0.10:−0.13
$-1\%\ \eta_{LPT}$	1:0.20:0.07	1:0.21:0.10

此外, 表 4 中还展示了包含 N_{2C} 在内的 3 个监控参数偏离量的比例关系供参考, 但由于 N_{2C} 受外界条件和控制规律的影响较大, 而指印图中提供的是 CFM56 系列发动机不同具体型号下大量运行故障数据样本整合得到的监控参数偏离量的平均值, 因此, N_{2C} 的偏离量在具体发动机的实际故障诊断中应用较少, 仅起辅助参考作用。

综上所述, 虽然实验结果中 N_{2C} 偏离量具体大小与指印图之间存在一定波动, 但由于实验结果与指印图中各监控参数的偏离方向一致、EGT和 W_F 偏离量的比例关系相近, 因此, 间接验证了基于风扇特性图缩放基准点优化和特性曲面拟合的发动机基准性能模型具有较好的精度, 并可用于气路性能退化分析和故障诊断。

3　结　论

本文以 CFM56-3 发动机为例, 提出一套从发动机机理层面充分利用部件特性图信息进行稳态性能建模, 并将其作为基准状态进行气路性能退化研究的方法, 所得结论如下。

1) 基于粒子群优化算法的缩放基准点的选取可使风扇通用特性图的缩放更符合发动机的实际工作环境, 提高特性图缩放法的精度。

2) 通过部件特性图的曲面拟合方法可将特性图中的转速特性引入部件特性方程, 使所建立的部件级基准性能模型能反映发动机的转速控制条件。

3) 采用构建的基准性能模型生成故障系数矩阵, 计算各监控参数EGT、W_F 及 N_{2C} 随部件效率下降 1% 的偏离量, 并结合指印图进行对比验证, 表明了本文部件级基准性能模型可用于民航发动机的气路性能退化分析, 并具有一定的精度和理论依据。

在后续研究中, 可通过多组不同工况点的试车数据实现特性图缩放过程的自适应优化, 并可改进发动机基准性能模型非线性方程组, 进一步提高建模精度。

参考文献 (References)

[1] 中国民用航空局. 2021 年民航行业发展统计公报[EB/OL]. (2022-05-18)[2023-05-30]. http://www.caac.gov.cn/XXGK/XXGK/TJSJ/202205/t20220518_213297.html.
Civil Aviation Administration of China. Statistical bulletin on the development of civil aviation industry in 2021[EB/OL]. (2022-05-18)[2023-05-30]. http://www.caac.gov.cn/XXGK/XXGK/TJSJ/202205/t20220518_213297.html(in Chinese).

[2] 林京, 张博瑶, 张大义, 等. 航空燃气涡轮发动机故障诊断研究现状与展望[J]. 航空学报, 2022, 43(8): 626565.
LIN J, ZHANG B Y, ZHANG D Y, et al. Research status and prospect of fault diagnosis for gas turbine aeroengine[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(8): 626565(in Chinese).

[3] 李少尘, 陈敏, 胡金涛, 等. 航空燃气涡轮发动机气路故障诊断进展[J]. 航空发动机, 2022, 48(2): 33-49.
LI S C, CHEN M, HU J T, et al. A review of research progress on aircraft gas turbine engines gas path fault diagnosis[J]. Aeroengine, 2022, 48(2): 33-49(in Chinese).

[4] SAJTEEV A M. Gas turbine diagnostics based on gas path analysis[D]. Netherlands: Delft University of Technology, 2015: 19-29.

[5] URBAN L A. Gas path analysis applied to turbine engine condition monitoring[J]. Journal of Aircraft, 1973, 10(7): 400-406.

[6] URBAN L A, VOLPONI A J. Mathematical methods of relative engine performance diagnostics[C]//Proceedings of the Aerospace Technology Conference and Exposition. Warrendale: SAE International, 1992: 922048.

[7] 范作民, 孙春林, 林兆福. 发动机故障诊断的主因子模型[J]. 航空学报, 1993, 14(12): 588-595.
FAN Z M, SUN C L, LIN Z F. Primary factor model for jet engine fault diagnosis[J]. Acta Aeronautica et Astronautica Sinica, 1993, 14(12): 588-595(in Chinese).

[8] 范作民, 孙春林, 白杰. 航空发动机故障诊断导论[M]. 北京: 科学出版社, 2004: 2-45.
FAN Z M, SUN C L, BAI J. Introduction to aeroengine fault diagnosis[M]. Beijing: Science Press, 2004: 2-45(in Chinese).

[9] STAMATIS A, MATHIOUDAKIS K, PAPAILIOU K D. Adaptive simulation of gas turbine performance[J]. Journal of Engineering for Gas Turbines and Power, 1990, 112(2): 168-175.

[10] 黄金泉, 王启航, 鲁峰. 航空发动机气路故障诊断研究现状与展望[J]. 南京航空航天大学学报, 2020, 52(4): 507-522.
HUANG J Q, WANG Q H, LU F. Research status and prospect of gas path fault diagnosis for aeroengine[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(4): 507-522(in Chinese).

[11] 曹明, 黄金泉, 周健, 等. 民用航空发动机故障诊断与健康管理现状、挑战与机遇Ⅰ: 气路、机械和 FADEC 系统故障诊断与预测[J]. 航空学报, 2022, 43(9): 625573.
CAO M, HUANG J Q, ZHOU J, et al. Current status, challenges and opportunities of civil aero-engine diagnostics & health management Ⅰ: diagnosis and prognosis of engine gas path, mechanical and FADEC[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(9): 625573(in Chinese).

[12] 骆广琦, 桑增产, 王如根, 等. 航空燃气涡轮发动机数值仿真[M]. 北京: 国防工业出版社, 2007: 49-51.
LUO G Q, SANG Z C, WANG R G, et al. Numerical methods for aviation gas turbine engine simulation[M]. Beijing: National Defense Industry Press, 2007: 49-51(in Chinese).

[13] 李书明, 王泽芃, 孙爽, 等. 基于贝兹曲线的压气机特性图生成方法[J]. 科学技术与工程, 2021, 21(33): 14428-14433.
LI S M, WANG Z P, SUN S, et al. Compressor engine component maps generating method based on Bézier curve[J]. Science Technology and Engineering, 2021, 21(33): 14428-14433(in Chinese).

[14] 黄开明, LI Y G, 张伟, 等. 某型涡轴发动机性能衰减与部件退化评估[J]. 航空动力学报, 2015, 30(11): 2673-2679.
HUANG K M, LI Y G, ZHANG W, et al. Performance degradation and components deterioration degree estimation for a turboshaft engine[J]. Journal of Aerospace Power, 2015, 30(11): 2673-2679(in Chinese).

[15] 张冬阳. 压气机特性的系数拟合法[J]. 燃气轮机技术, 1993, 6(3): 27-32.
ZHANG D Y. Coefficient fitting method for compressor characteristics[J]. Gas Turbine Technology, 1993, 6(3): 27-32(in Chinese).

[16] 欧阳辉, 朱之丽. 某涡喷发动机数值建模与改型设计[J]. 北京航空航天大学学报, 2008, 34(3): 311-314.
OUYANG H, ZHU Z L. Modeling and retrofit of certain turbojet[J]. Journal of Beijing University of Aeronautics and Astronautics, 2008, 34(3): 311-314(in Chinese).

[17] VERBIST M. Gas path analysis for enhanced aero-engine condition monitoring and maintenance[D]. Netherlands: Delft University of Technology, 2017: 15-38.

[18] KURZKE J, HALLIWELL I. Propulsion and power: an exploration of gas turbine performance modeling[M]. Berlin: Springer, 2018: 491-495.

[19] CUMPSTY N. Jet propulsion: a simple guide to the aerodynamic and thermodynamic design and performance of jet engines[M]. Edinburgh: Cambridge University Press, 1997.

[20] POLI R, KENNEDY J, BLACKWELL T. Particle swarm optimization[J]. Swarm Intelligence, 2007, 1(1): 33-57.

[21] KENNEDY J. Particle swarm optimization[C]//Proceedings of the Encyclopedia of Machine Learning. Berlin: Springer, 2011: 760-766.

[22] RIDAURA J A R. Correlation analysis between HPC blade chord and compressor efficiency for the CFM56-3[D]. Lisbon: University of Lisbon, 2013.

[23] 楚武利, 刘前智, 胡春波. 航空叶片机原理[M]. 西安: 西北工业大学出版社, 2009: 9.
CHU W L, LIU Q Z, HU C B. Principle of aviation blade machine[M]. Xi'an: Northwestern Polytechnical University Press, 2009: 9(in Chinese).

[24] TSOUTSANIS E, LI Y-G, PILIDIS P, et al. Non-linear model calibration for off-design performance prediction of gas turbines with experimental data[J]. The Aeronautical Journal, 2017, 121(1245): 1758-1777.

[25] MARTINS D A R. Off-design performance prediction of the CFM56-3 aircraft engine[D]. Lisbon: University of Lisbon, 2015.

Performance degradation modelling of civil aviation engines based on component characteristic map optimization

GUO Qing[1,*], HUANG Qilian[1], CHEN Jinliang[2]

(1. School of Aeronautical Engineering, Civil Aviation University of China, Tianjin 300300, China;
2. Shandong Airlines Co., Ltd. Yantai Branch, Yantai 264000, China)

Abstract: In order to provide a theoretical basis for the gas path performance degradation of civil aviation engines at the module level, the CFM56-3 engine was taken as the research object. Firstly, the characteristic map scaling method was used to obtain the component characteristic equations, and the selection process of the general fan characteristic map scaling reference point was optimized. A characteristic map surface fitting method was proposed to construct an engine component-level benchmark performance model that conformed to specific speed conditions under steady-state operating mode. Then, by introducing fault factors to generate a fault coefficient matrix, the deviation of engine monitoring parameters with the decrease in component efficiency was calculated and compared with the fingerprint diagram data in the General Electric Company training manual. The results show that the engine steady-state performance model integrating fan characteristic map scaling reference point optimization and characteristic map surface fitting methods has good accuracy and application prospects in gas performance degradation analysis.

Keywords: civil aviation engine; gas path analysis; component-level modelling; scaling reference point optimization; surface fitting

Received: 2023-06-12; **Accepted**: 2023-08-11; **Published Online**: 2023-08-22 14:40
URL: link.cnki.net/urlid/11.2625.V.20230822.0851.001
Foundation item: 2022 Tianjin Postgraduate Research Innovation Project (2022SKY156)
* **Corresponding author**. E-mail: qguocauc@sina.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0323

基于几何蒸馏和特征自适应的磁共振成像重建

朵琳*，任勇，许渤雨，杨新
(昆明理工大学 信息工程与自动化学院，昆明 650500)

摘 要：虽然现有基于深度学习的压缩感知磁共振成像(CS-MRI)方法已经取得了较好的效果，但这些方法的可解释性仍然面临挑战，并且从理论分析到网络设计的过渡并不够自然。为解决上述问题，提出深度双域几何蒸馏特征自适应网络(DDGD-FANet)。该深度展开网络将磁共振成像重建优化问题迭代展开成3个子部分：数据一致性模块、双域几何蒸馏模块和自适应网络模块，不仅可以补偿重建图像丢失的上下文信息，恢复更多的纹理细节，还可以去除全局伪影，进一步提高重建效果。在公开数据集使用3种不同的采样模式进行实验，结果表明：DDGD-FANet在3种采样模式下均取得了更高的峰值信噪比和结构相似性指数，在笛卡儿10%压缩感知(CS)比率下，峰值信噪比较迭代收缩阈值算法(ISTA-Net+)、快速ISTA(FISTA)-Net和DGDN模型分别提高了5.01 dB、4.81 dB和3.34 dB。

关 键 词：磁共振成像；图像重建；深度学习；几何蒸馏；自适应网络；注意力机制
中图分类号：TP391; R319
文献标志码：A **文章编号**：1001-5965(2025)06-1946-09

磁共振成像(magnetic resonance imaging, MRI)作为一种先进的影像学技术，可精确生成用于疾病诊断和医学分析的体内组织图像。MRI数据从频域(k空间)采集得到，采集时间通常与k空间相位编码步长成正比，一个典型MR图像序列的扫描时间通常至少需要30 min，具体取决于被扫描的身体部位。这比大多数其他成像方式要长得多，长时间的扫描过程会导致患者不适或将运动伪影引入MR图像并降低诊断的可访问性。因此，加快MRI扫描速度对于提高MR图像质量至关重要且意义重大[1]。

为克服上述问题，基于压缩感知的磁共振成像(compressed sensing-magnetic resonance imaging, CS-MRI)[2]重建技术应运而生，CS-MRI重构旨在通过数学算法和信号处理方法从欠采样的k空间数据中恢复清晰的MR图像。

传统的CS-MRI方法将采样和压缩结合在一起，如果所获取的数据存在某个稀疏变换域，则压缩感知需要的测量比理论上必要的测量要少[3]。因此，采用压缩传感对k空间数据进行欠采样，可以减少成像过程中的扫描时间[4]。经典的CS-MRI方法，如傅里叶变换[5]、小波变换[6]和离散余弦变换[7]已被成功应用于MRI重建。由于具有简单有效的优点，总变分(total variation, TV)[8-9]正则化也被广泛应用于MRI重建中，尽管其在重建中引入了阶梯伪影。一些复杂的非局部稀疏方法使用局部相似块来利用非局部自相似性，可以捕获更多纹理先验并显著提高重建性能，如基于块的定向小波(patch-based directional wavelets, PBDW)[10]，基于补丁的非局部算子(patch-based nonlocal operator, PANO)[11]。许多研究将压缩感知(compressed sensing, CS)重构问题表述为稀疏正则化优化问题，然后，使用不同的迭代算法来求解，如迭代收缩阈值算法(iterative

收稿日期：2023-06-07；**录用日期**：2023-08-18；**网络出版时间**：2023-09-12 16:19
网络出版地址：link.cnki.net/urlid/11.2625.V.20230911.1850.002
基金项目：国家自然科学基金(61962032)
*通信作者. E-mail：duolin2003@126.com

引用格式：朵琳，任勇，许渤雨，等. 基于几何蒸馏和特征自适应的磁共振成像重建[J]. 北京航空航天大学学报，2025，51(6)：1946-1954. DUO L，REN Y，XU B Y，et al. MRI reconstruction based on geometry distillation and feature adaptation[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1946-1954 (in Chinese).

shrinkage-thresholding algorithm, ISTA)[12] 和交替方向乘子方法 (alternating direction method of multipliers, ADMM)[13] 等。上述方法均基于强理论和良好预定义的图像先验, 具有可解释性强的优点。然而, 他们通常需要昂贵的计算, 并且正则化器和模型的选择也是一个挑战。因此, 经典 CS-MRI 方法的重建结果通常不令人满意。

近年来, 随着硬件、软件及相关算法的发展, 深度学习 (deep learning, DL)[14] 在众多领域取得了巨大成功。基于此, 人们开始将深度学习应用于 MRI 中[15]。Wang 等[16] 将卷积神经网络 (convolutional neural networks, CNN) 用于 MRI, 设计了一种离线 CNN 来学习 k 空间欠采样图像和全采样图像之间的映射关系, 有效恢复了精细结构和细节。深度去混叠生成对抗网络 (de-aliasing generative adversarial networks, DAGAN) 将对抗性损失与内容损失结合起来, 以更好地保留 CS-MRI 中的纹理和边缘, 并且还结合了频域信息以增强图像和频域的相似性[17]。这些现有的基于深度学习的 CS-MRI 方法以数据驱动的方式将网络作为黑匣子训练, 并且也没有任何先验模型, 虽然重建效果明显优于传统方法, 但可解释性一直是一个悬而未决的问题。

深度展开网络继承了传统 CS-MRI 方法和基于深度学习 CS-MRI 方法的优点, 具有足够强的解释性和良好的性能[18-20]。作为开创性的工作, Yang 等[21] 提出一种新颖的深层 ADMM-Net 框架来监督图像重建网络中的数据流图。Zhang 和 Ghanem [22] 通过将 ISTA 映射到深度 CNN 框架来学习近端映射, 提出 ISTA-Net。为将快速迭代收缩阈值算法 (fast iterative shrinkage-thresholding algorithm, FISTA) 嵌入到深度网络中, FISTA-Net[23] 将多个梯度下降、近端映射和加速模块堆叠在一起, 该网络继承了 FISTA 的良好可解释性, 同时也在 CNN 的支持下具有强大的学习能力, 可以去除噪声和恢复细节。Schlemper 等[24] 提出具有数据一致性层 (data consistency, DC) 的深度级联卷积网络, 其保留了原始的欠采样数据, 这一思想在深度展开网络中被广泛使用。Fan 等[25] 将 CS-MRI 优化问题展开为线性逼近和特征几何补偿 2 个子问题, 并把几何补偿子模块通过泰勒展开式展开, 提出深度几何蒸馏网络, 有效恢复了图像纹理细节, 但随着网络的加深, 几何特征丢失严重, 并且也没有考虑全局伪影问题。

此外, 有研究表明, 在频域使用卷积算子可以实现非局部感受野, 弥补图像域卷积处理局部感受野的局限性[26]。在此基础上, Singh 等[27] 联合频域和图像空间重建, 有效去除了图像的局部和全局伪影。另一方面, 目前基于深度学习的 MRI 重建方法通常平等的对待通道和空间特征, 有研究表明着重关注通道和空间中的重要特征可以提升网络性能[28]。受上述问题的启发, 本文联合双域与几何蒸馏, 设计了一个可以关注通道和空间重要信息的特征自适应网络来对 MR 图像进行重建, 将该架构命名为: 深度双域几何蒸馏特征自适应网络 (deep dual-domain geometry distillation feature adaptive network, DDGD-FANet)。实验结果表明, 本文网络在不同采样模式不同采样比率下, 均取得了更好的重建效果。

1　基于 DDGD-FANet 的 MRI 重建

1.1　问题表述

在 MRI 中, MR 设备采集 k 空间数据, 从欠采样 k 空间数据重建原始图像可以描述为

$$\boldsymbol{y}=\boldsymbol{\Phi x} \tag{1}$$

式中: $\boldsymbol{x}$为原始图像信号; $\boldsymbol{y}$为欠采样的 k 空间数据; 线性映射矩阵 $\boldsymbol{\Phi}=\boldsymbol{PS}$, $\boldsymbol{P}$为采样算子, $\boldsymbol{S}$为傅里叶变换。一般来说, 欠采样 k 空间数据是通过亚奈奎斯特采样获得的, 意味着矩阵 $\boldsymbol{\Phi}$是高度病态且稀疏的。所以, 想要通过直接求解该逆问题来重构图像十分困难, 通常需要结合一些先验信息的正则化来重建图像[29]。因此, 图像重构可以表示为

$$\boldsymbol{x}=\arg\min_{\boldsymbol{x}}\{\mathcal{E}(\boldsymbol{x})+\gamma\mathcal{R}(\boldsymbol{x})\} \tag{2}$$

其中, $\mathcal{E}(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{\Phi x}-\boldsymbol{y}\|_2^2$为数据一致性项; γ为正则化参数, 其选择方式对方法的重构性能有着重要影响; $\mathcal{R}(\boldsymbol{x})=\|\boldsymbol{\Psi x}\|_1$表示先验解的正则化器, 其可以通过强制执行先验信息来提高重构图像质量。由于式(2)常通过迭代算法求解, 因此, 迭代正则化已经发展为一种极具潜力的正则化方法。传统的展开迭代方法往往依靠手工制作的正则化器或惩罚函数, 要找到合适的先验表示并不容易[30]。使用网络模块来替代相应算法的正则化器或惩罚函数, 然后, 对展开的深度神经网络(deep neural networks, DNN)进行端到端训练, 以数据驱动的方式实现先验学习, 可以达到更好的重建效果和更快的重建速度。因此, 将深度学习与优化模型相结合的磁共振图像重建已成为目前的研究重点。

1.2　DDGD-FANet 架构

DDGD-FANet 结构如图 1 所示, 随着网络的加深, 特征域的纹理信息会丢失, 双域几何蒸馏 (double domain geometry distillation, DDGD)模块将 CNN 架构与非线性几何算子的泰勒展开相结合, 可以有效补偿几何特征的局部和全局纹理信息。为更好地融合来自图像域和频域的特征, 本文提

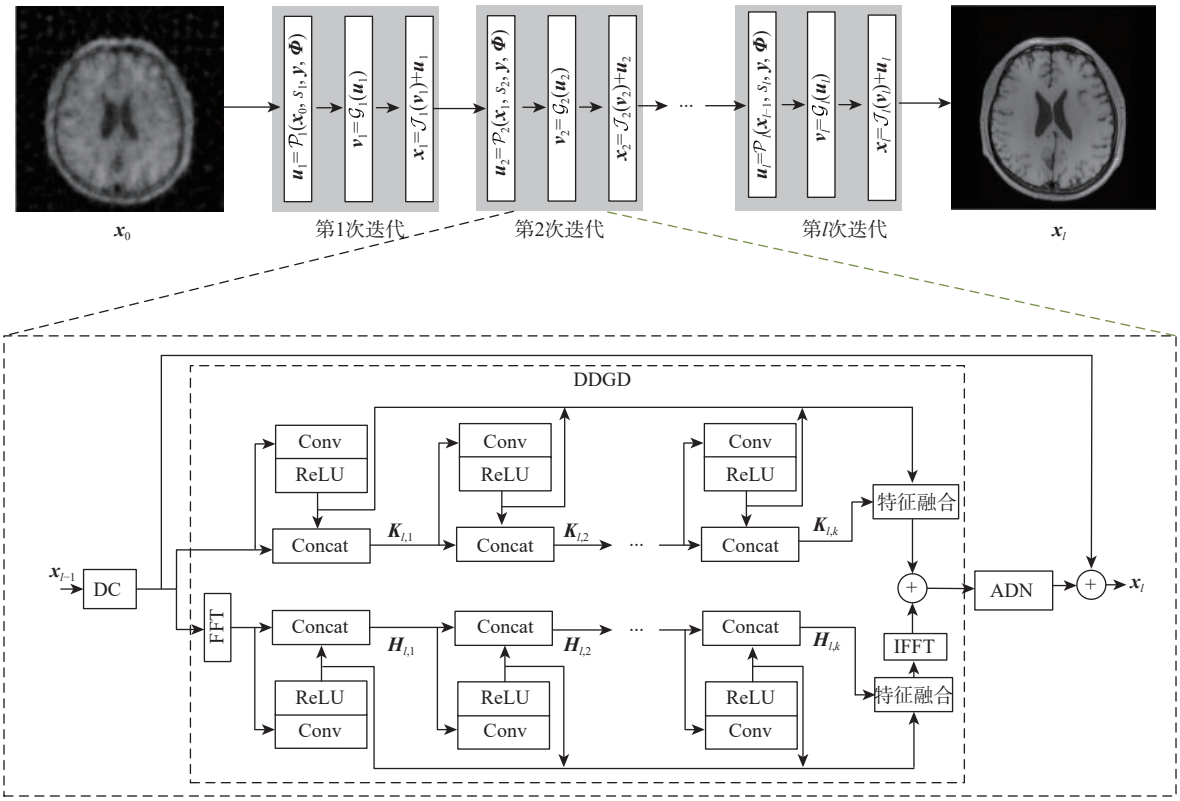


图1 DDGD-FANet结构

Fig. 1 Structure of DDGD-FANet

出一种新颖且高效的自适应网络 (adaptive network, ADN), 该网络由注意力模块 (attention module, AM) 和特征自适应融合 (feature adaptive fusion, FAF) 模块组成, 可以考虑不同通道和空间之间的差异, 关注更重要的特征, 自适应的融合特征, 进一步去除伪影, 恢复图像结构。对于第l次迭代, 第k层蒸馏层, 设$\mathcal{E}_l(\boldsymbol{x})$为$\mathcal{E}(\boldsymbol{x})$在$\boldsymbol{x}_{l-1}$处的泰勒二阶展开, 其可以描述为

$$\mathcal{E}(\boldsymbol{x}) \approx \mathcal{E}_l(\boldsymbol{x}) := \frac{1}{2s_l}\left\|\boldsymbol{x}-[\boldsymbol{x}_{l-1}-s_l\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi}\boldsymbol{x}_{l-1}-\boldsymbol{y})]\right\|_2^2+d \tag{3}$$

式中: $1/s_l$为 Lipschitz 常数; d为常数。

具体来说, 式 (2) 的解可以通过以下 3 步迭代完成:

$$\boldsymbol{u}_l=\mathcal{P}_l(\boldsymbol{x}_{l-1}, s_l, \boldsymbol{y}, \boldsymbol{\Phi}) := \arg\min_{\boldsymbol{x}} \mathcal{E}_l(\boldsymbol{x}) \tag{4}$$

$$\boldsymbol{v}_l=\mathcal{G}_l(\boldsymbol{u}_l) := \arg\min_{\boldsymbol{x}}\left\{\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{u}_l\|_2^2+\lambda_l\mathcal{R}(\boldsymbol{x})\right\} \tag{5}$$

$$\boldsymbol{x}_l=\mathcal{J}_l(\boldsymbol{v}_l)+\boldsymbol{u}_l \tag{6}$$

式中: $\lambda_l=s_l\gamma$。

1.2.1 数据一致性层

如图 1 所示, 在每次迭代开始时, 引入数据一致性层。利用已知的物理原理和数据特性, 对重构图像进行约束, 强制每次迭代的输出图像的 k 空间数据取一个与原始欠采样数据接近的值, 以提高其准确性和质量。在数学上, 将式 (4) 重写为

$$\boldsymbol{u}_l=\mathcal{P}_l(\boldsymbol{x}_{l-1}, s_l, \boldsymbol{y}, \boldsymbol{\Phi}) := \boldsymbol{x}_{l-1}-s_l\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi}\boldsymbol{x}_{l-1}-\boldsymbol{y}) \tag{7}$$

$$s_l=\ln(1+\exp(a_1 l+a_2)) \qquad l=1,2,\cdots,L \tag{8}$$

式中: s_l为步长; a_1和a_2为可学习步长参数。

1.2.2 双域几何蒸馏模块

传统上, 通过傅里叶逆变换将频率空间信号转换为图像空间进行重建, 每个单独的频率空间测量对图像空间中的所有输出像素都有贡献[27]。因此, 获取的频率空间数据的局部变化会对输出图像产生全局影响。为产生准确的图像重建, 除执行精细的图像空间处理外, 还必须校正这些全局伪影。受此启发, 本文联合图像空间与频率空间设计了一个双域几何蒸馏模块, 其结构如图 1 中内层虚线所示, 双域几何蒸馏模块呈差分对称结构, 图像空间卷积处理有助于对图像域中的局部伪影进行补充校正, 而频率空间中的卷积运算有望对图像空间全局伪影进行校正, 以此来补偿丢失的纹理信息。通过从数据一致性模块的输出中提取成分, 可以获得来自不同蒸馏层的特征图, 最后, 将这些特征聚合并细化提纯, 以得到更丰富、更高效的几何信息。值得注意的是, 由于图像空间和频率空间的特征差异, 图像域与频域几何蒸馏模块并不共享参数。

如上所述, 数据一致性模块的输出存在严重的伪影, 由 CNN 组成的几何蒸馏网络可以逐层蒸馏图像特征, 通过学习来补偿更多的纹理细节。设 Concat(.,.)表示以下拼接运算符:

$$\mathrm{Concat}(\boldsymbol{f}_1,\boldsymbol{f}_2)=(\boldsymbol{f}_1,\boldsymbol{f}_2)\in\mathbf{R}^{n\times m\times b}$$

式中: $\boldsymbol{f}_1\in\mathbf{R}^{n\times m\times b_1}$; $\boldsymbol{f}_2\in\mathbf{R}^{n\times m\times b_2}$; $b=b_1+b_2$。设 $\boldsymbol{K}_{l,i+1}$ 和 $\boldsymbol{H}_{l,i+1}$ 分别为第 l 次迭代第 $i+1$ 层的图像域和频域特征图; $\boldsymbol{M}_l$ 和 $\boldsymbol{N}_l$ 分别为第 l 次迭代的图像域和频域几何蒸馏输出, 则双域几何蒸馏模块可定义为

$$\begin{cases}\boldsymbol{K}_{l,0}=\boldsymbol{u}_l\\ \boldsymbol{K}_{l,1}=(\boldsymbol{K}_{l,0},\mathcal{A}_l^0(\boldsymbol{K}_{l,0}))\\ \boldsymbol{K}_{l,2}=(\boldsymbol{K}_{l,1},\mathcal{A}_l^1(\boldsymbol{K}_{l,1}))\end{cases}\tag{9}$$

$$\boldsymbol{K}_{l,i+1}=(\boldsymbol{K}_{l,i},\mathcal{A}_l^i(\boldsymbol{K}_{l,i}))\qquad k-1\geqslant i\geqslant 0\tag{10}$$

$$\boldsymbol{M}_l=(\mathcal{A}_l^0(\boldsymbol{K}_{l,0}),\mathcal{A}_l^1(\boldsymbol{K}_{l,1}),\cdots,\mathcal{A}_l^{k-1}(\boldsymbol{K}_{l,k-1}),\boldsymbol{K}_{l,k})\tag{11}$$

$$\begin{cases}\boldsymbol{H}_{l,0}=\mathrm{FFT}(\boldsymbol{u}_l)\\ \boldsymbol{H}_{l,1}=(\boldsymbol{H}_{l,0},\mathcal{B}_l^0(\boldsymbol{H}_{l,0}))\\ \boldsymbol{H}_{l,2}=(\boldsymbol{H}_{l,1},\mathcal{B}_l^1(\boldsymbol{H}_{l,1}))\end{cases}\tag{12}$$

$$\boldsymbol{H}_{l,i+1}=(\boldsymbol{H}_{l,i},\mathcal{B}_l^i(\boldsymbol{H}_{l,i}))\qquad k-1\geqslant i\geqslant 0\tag{13}$$

$$\boldsymbol{N}_l=\mathrm{IFFT}((\mathcal{B}_l^0(\boldsymbol{H}_{l,0}),\mathcal{B}_l^1(\boldsymbol{H}_{l,1}),\cdots,\mathcal{B}_l^{k-1}(\boldsymbol{H}_{l,k-1}),\boldsymbol{H}_{l,k}))\tag{14}$$

$$\boldsymbol{v}_l=\boldsymbol{M}_l+\boldsymbol{N}_l\tag{15}$$

其中: FFT(·)为快速傅里叶变换; IFFT(·)为快速傅里叶反变换; $\mathcal{A}_l^i$ 和 $\mathcal{B}_l^i$ 为具有卷积 w 和具有激活函数 ReLU 的块, 可以定义为

$$\mathcal{A}_l^i(.)\left(\mathcal{B}_l^i(.)\right):=\mathrm{ReLU}(w(.)):\mathbf{R}^{n\times m\times(1+ip)}\to\mathbf{R}^{n\times m\times p}\tag{16}$$

通过由许多卷积和 ReLU 激活函数组成的组合块, 网络可以学习到式 (5) 的近似最优解。

1.2.3　自适应网络

1.2.2 节所述的双域几何蒸馏模块中, 模块的输出数据来自不同特征域的不同蒸馏层, 每个通道和空间对网络的贡献也有差异, 此时再平等的对待这些数据会丢失特征间的差异性。因此, 本文引入通道注意力和空间注意力, 如图 2(a) 的注意力模块所示, 利用特征的通道关系和空间关系生成通道和空间注意力图, 沿通道和空间 2 个维度关注重要特征并且抑制不必要的特征。此外, 为有效地帮助信息在网络中流动, 如图 2(b) 所示, 本文设计了一个由 CNN 和 ReLU 函数组成的特征自适应融合模块, 自适应融合特征, 同时减少通道融合中产生的数据抖动, 更好的恢复图像的纹理细节。

对于第 l 次迭代, 设注意力模块的入特征图 $\boldsymbol{F}_l\in\mathbf{R}^{C\times H\times W}$, 本文使用平均池化AvgPool和最大池化MaxPool操作聚合特征图的通道和空间信息, 得到 2 个不同的特征上下文描述, 然后, 通过卷积层生成通道注意力图 $\boldsymbol{W}_l^C\in\mathbf{R}^{C\times1\times1}$ 和空间注意力图 $\boldsymbol{W}_l^S\in\mathbf{R}^{1\times H\times W}$, 将其分别与输入特征图 $\boldsymbol{F}_l$ 和通道注意力输出 $\boldsymbol{F}_l^C\in\mathbf{R}^{C\times H\times W}$ 相乘, 最终得到空间注意力输出 $\boldsymbol{F}_l^S\in\mathbf{R}^{C\times H\times W}$。特征自适应融合模块级联在注意力模块之后, 由 4 个级联的卷积层组成, 通过在卷积层中引入非线性激活函数 ReLU 来提高网络的表达能力。需要注意的是, 最后一个卷积层不使用激活函数, 因为这样会截断部分网络输出图像数据, 降低图像重构质量, 减缓网络收敛速度, 相关的消融实验将在后面叙述。设 $\mathcal{D}_l$ 为第 l 次迭代自适应网络的输出, 则自适应网络可以描述为

$$\boldsymbol{W}_l^C(\boldsymbol{F}_l)=\sigma(f_l^{1\times1}(\mathrm{AvgPool}(\boldsymbol{F}_l))+f_l^{1\times1}(\mathrm{MaxPool}(\boldsymbol{F}_l)))\tag{17}$$

$$\boldsymbol{F}_l^C=\boldsymbol{W}_l^C(\boldsymbol{F}_l)\otimes\boldsymbol{F}_l\tag{18}$$

$$\boldsymbol{W}_l^S(\boldsymbol{F}_l^C)=\sigma(f_l^{7\times7}(\mathrm{AvgPool}(\boldsymbol{F}_l^C),\mathrm{MaxPool}(\boldsymbol{F}_l^C)))\tag{19}$$

$$\boldsymbol{F}_l^S=\boldsymbol{W}_l^S(\boldsymbol{F}_l^C)\otimes\boldsymbol{F}_l^C\tag{20}$$

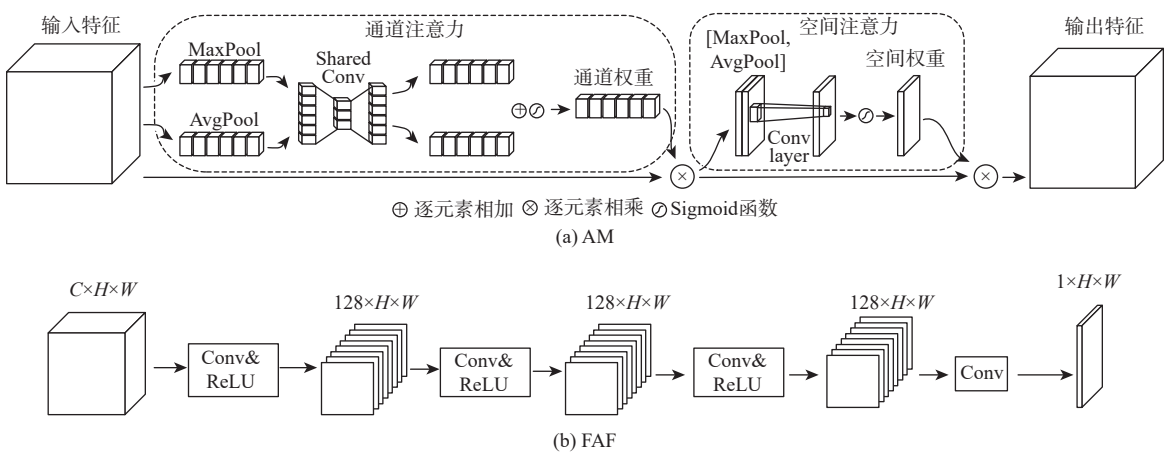


图 2　自适应网络结构

Fig. 2　Structure of adaptive network

$$\mathcal{D}_l = \mathcal{J}_l(\boldsymbol{v}_l) = f_l^{3\times3}\mathcal{C}_l^1\mathcal{C}_l^2\mathcal{J}_l(\boldsymbol{F}_l^{\mathrm{S}}) \tag{21}$$

式中：σ表示 Sigmoid 函数；$f_l^{1\times1}$、$f_l^{7\times7}$和$f_l^{3\times3}$分别为卷积核大小为 1×1、7×7 和 3×3 的卷积算子；$\mathcal{J}_l$、$\mathcal{C}_l^1$和$\mathcal{C}_l^2$为具有卷积和非线性激活函数 ReLU 的块。

1.3 损失函数

损失函数通常用来最小化网络输出重构图像$\hat{\boldsymbol{x}}$与真实图像$\boldsymbol{x}$之间的距离以寻找真实目标图像，常用的距离度量包括l_1范数和l_2范数。然而，最近有相关研究表明，像素级别的l_2范数难以捕捉感知高频几何细节[31]。因此，本文采用l_1范数来约束优化过程，结合深监督技术，在网络的中间迭代阶段添加辅助约束来监督主干网络，使优化过程更加灵活。对于给定的小批量训练样本$\{(\boldsymbol{y}^j,\boldsymbol{x}^j)\}_{j=1}^{n_{\mathrm{S}}}$，损失函数可以定义为

$$\mathcal{L} = \frac{1}{n_{\mathrm{S}}m_{\mathrm{S}}}\sum_{j=1}^{n_{\mathrm{S}}}\left(\left\|\boldsymbol{x}_{\bar{l}}^j-\boldsymbol{x}^j\right\|_1+\left\|\hat{\boldsymbol{x}}^j-\boldsymbol{x}^j\right\|_1\right) \tag{22}$$

式中：n_{S}为小批量大小；m_{S}为$\boldsymbol{x}^j$的像素数；$\bar{l}=(L+1)/2$，L为网络迭代次数。

2 实 验

2.1 数据集与评价指标

实验在 MR 大脑数据集上评估不同方法的性能，该数据集来自不同的 MRI 设备，由健康和阿尔兹海默患者的 T1 加权二维图像组成，其中，800 张图片用于训练，50 张图片用于测试[25]。为评估图像重建质量，实验使用峰值信噪比 (peak signal-to-noise ratio, PSNR) 和结构相似性指数 (structural similarity index measure, SSIM) 作为评价指标[1]。在测试过程中，需先对数据进行归一化处理，设全采样数据为$\boldsymbol{x}$，重建图像为$\hat{\boldsymbol{x}}$，则 PSNR 和 SSIM 的计算方式如下：

PSNR 定义为

$$P_{\mathrm{SNR}} = 20\lg\left(1/\sqrt{M}\right) \tag{23}$$

式中：M为$\boldsymbol{x}$与$\hat{\boldsymbol{x}}$之间的均方误差。

SSIM 定义为

$$S_{\mathrm{SIM}} = \frac{(2u_xu_{\hat{x}}+c_1)(2\sigma_{x\hat{x}}+c_2)}{(u_x^2+u_{\hat{x}}^2+c_1)(\sigma_x^2+\sigma_{\hat{x}}^2+c_2)} \tag{24}$$

式中：u_x和$u_{\hat{x}}$分别为$\boldsymbol{x}$和$\hat{\boldsymbol{x}}$的均值；σ_x^2和$\sigma_{\hat{x}}^2$分别为$\boldsymbol{x}$和$\hat{\boldsymbol{x}}$的方差；$\sigma_{x\hat{x}}$为$\boldsymbol{x}$和$\hat{\boldsymbol{x}}$的协方差；c_1和c_2为常数，默认使用$c_1=0.01$，$c_2=0.03$[1]。

2.2 实验环境与参数设置

实验在配置为 Inter Xeon CPU E5-2696, 16 GB 内存，NVIDIA RTX3090(24 GB 显存)，pytorch1.7.1，64 位 Windows10 操作系统的服务器上完成，采用 3 种欠采样方式进行采样，分别是笛卡儿欠采样、伪径向欠采样和二维随机欠采样，每种欠采样方式分别使用 10%、20% 和 30% 的 CS 比率来评估模型的重构性能。结合硬件资源考虑，实验中将 batch_size 设置为 1，网络迭代次数L=13，蒸馏层数k=8，特征图通道数p=32，使用 Adam 优化器优化模型，学习率设置为 0.000 1，epoch 为 100，除特别说明外，网络中使用的卷积核大小均为 3×3。

2.3 实验结果与分析

实验将本文框架与先进的方法进行比较，包括传统方法(Zero-filling[32]、PBDW[10]、PANO[11])和深度展开方法(DC-CNN[24]、ISTA-Net[22]、ISTANet+[22]、FISTA-Net[23]、DGDN[25])，为公平起见，DGDN 的迭代阶段数设置为 13。表 1 为大脑数据集上本文方法与对比方法在笛卡儿采样模式下不同采样率重建图像的评估结果，以测试集中所有重建图像与对应全采样图像之间的平均 PSNR 和 SSIM 为评价指标。其中，加黑数值为最大值，由表可知，相比于其他方法，本文方法在所有 CS 比率下都具有更高的

表 1 不同 CS 比率笛卡儿采样模式下大脑数据集重建效果对比

Table 1 Comparison of reconstruction effects in different CS ratios of Cartesian sampling pattern in brain dataset

掩码类型	方法	PSNR/dB			SSIM		
		CS比率为10%	CS比率为20%	CS比率为30%	CS比率为10%	CS比率为20%	CS比率为30%
笛卡儿	Zero-filling[32]	23.84	26.40	31.14	0.574 3	0.661 1	0.786 5
	PBDW[10]	26.82	32.41	35.05	0.735 5	0.866 0	0.911 6
	PANO[11]	28.98	34.62	36.75	0.789 7	0.893 3	0.922 2
	DC-CNN[24]	30.04	35.75	38.48	0.823 5	0.920 5	0.949 3
	ISTA-Net[22]	30.28	36.58	39.17	0.825 6	0.933 4	0.956 8
	ISTA-Net+[22]	30.86	37.03	39.63	0.845 5	0.935 5	0.959 7
	FISTA-Net[23]	31.06	37.29	39.86	0.844 4	0.939 6	0.960 7
	DGDN[25]	32.53	38.31	40.61	0.886 3	0.949 9	0.965 6
	DDGD-FANet	**35.87**	**39.55**	**41.41**	**0.932 0**	**0.959 3**	**0.970 1**

PSNR 和 SSIM, 尤其是在 10% 的低 CS 比率下 PSNR 达到了 35.87 dB, SSIM 值为 0.9320, PSNR 较 DGDN 实现了 3.34 dB 的提升, 在 3 种 CS 比率下 PSNR 平均提高了 1.79 dB。

为验证本文方法的泛化性, 实验测试了各种方法在伪径向采样和二维随机采样模式下不同 CS 比率的 PSNR 值。如表 2 所示, 表中, 加黑数值为最大值。实验结果表明, 本文方法在 2 种采样模式下均实现了最好的重建结果。具体来说, 相对于 DGDN 模型, 在伪径向 10%CS 比率下, PSNR 实现了 1.17 dB 的提升, 3 种 CS 比率下 PSNR 平均提高了 0.57 dB。在二维随机 10%CS 比率下, PSNR 实现了 0.4 dB 的提升, 3 种 CS 比率下 PSNR 平均提高了 0.26 dB。由此可见, 本文方法始终如一的保持着更好的重建性能。

图 3 比较了所有方法在 3 种采样模式下 CS 比率为 10% 的重建视觉效果。结果表明, 传统方法可以去除一些噪声, 但纹理信息重建效果比较差, 尽管 FISTA-Net 和 DGDN 效果明显优于传统方法, 但对局部信息的恢复并不理想。由于 DDGD 模块和 ADN 模块的有效性, 本文方法的重建图像更接近真实图像, 重建误差更小, 实现了更好的视觉质量和更高的 PSNR, 在恢复局部几何特征的同时, 还可以去除全局伪影。

表 2　不同 CS 比率伪径向和二维随机采样模式下大脑数据集建效果对比

Table 2　Comparison of reconstruction effects in different CS ratios of pseudo-radial and 2D random sampling pattern in brain dataset

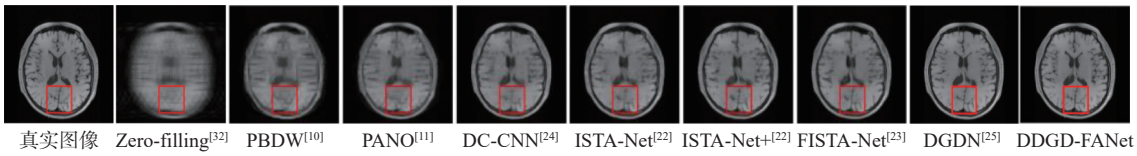
掩码类型	方法	PSNR/dB		
		CS比率为10%	CS比率为20%	CS比率为30%
伪径向	Zero-filling[32]	26.64	30.32	32.93
	PBDW[10]	32.48	36.53	38.95
	PANO[11]	33.61	37.01	39.29
	DC-CNN[24]	34.32	38.43	40.75
	ISTA-Net[22]	34.70	38.74	41.01
	ISTA-Net+[22]	34.83	38.75	41.00
	FISTA-Net[23]	35.10	39.05	41.19
	DGDN[25]	35.91	39.33	41.34
	DDGD-FANet	**37.08**	**39.67**	**41.54**
二维随机	Zero-filling[32]	27.71	29.36	31.13
	PBDW[10]	35.68	38.49	40.83
	PANO[11]	36.60	39.57	42.03
	DC-CNN[24]	37.65	40.81	43.40
	ISTA-Net[22]	38.22	41.40	43.97
	ISTA-Net+[22]	38.41	41.66	44.20
	FISTA-Net[23]	38.36	41.75	44.29
	DGDN[25]	38.85	41.91	44.38
	DDGD-FANet	**39.25**	**42.13**	**44.53**

2.4　消融实验

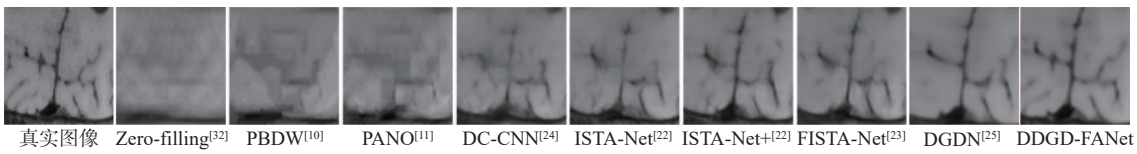
实验以 DGDN 作为基准模型, 在大脑数据集笛卡儿 10%CS 比率下进行了一组消融实验来验证 DDGD 和 ADN 对重建性能的有效性。DGDN 模型只在图像域中使用几何蒸馏模块, 可以很好的恢复图像的局部伪影, 但却很难处理全局伪影。如表 3 所示, DDGN 模块和 ADN 分别带来了 2.02 dB 和 1.05 dB 的 PSNR 提升。由此可见, 本文的 DDGD 模块可以有效地去除这些全局伪影, 达到更好的重建效果, ADN 可以消除数据抖动, 自适应融合特征, 进一步提升重建质量。

接下来, 实验验证了在 FAF 模块最后一个卷积层使用 ReLU 函数对网络性能的影响。如图 4 所示, 结果表明, 在 FAF 模块末尾使用非线性变换会严重影响图像重构质量和收敛速度。从量化数值上讲, 在笛卡儿 10%CS 比率下, PSNR 降低了 0.7 dB, 收敛速度减少了 25%。这是由于 ReLU 激活函数截断了迭代输出, 致使重构图像数据缺失, 从而影响网络性能。

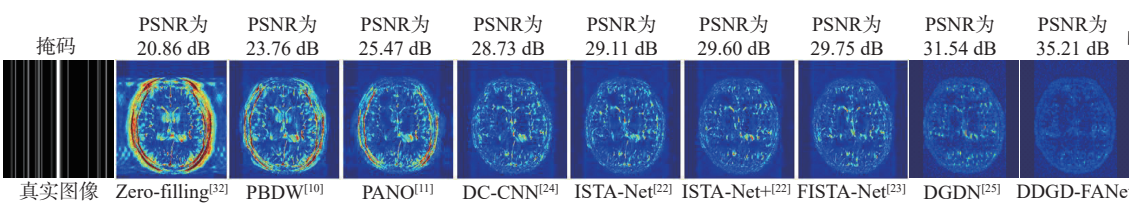
最后, 为验证数据一致性层的有效性, 在笛卡儿 10%CS 比率下进行实验。其结果如表 4 所示, 表中的 DDGD-FANet_NODC 表示没有数据一致性



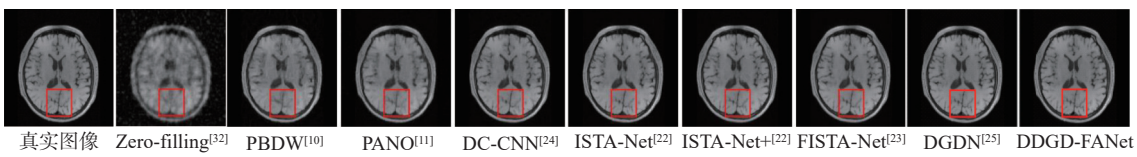
(a) 笛卡儿10%CS比率下重建后效果



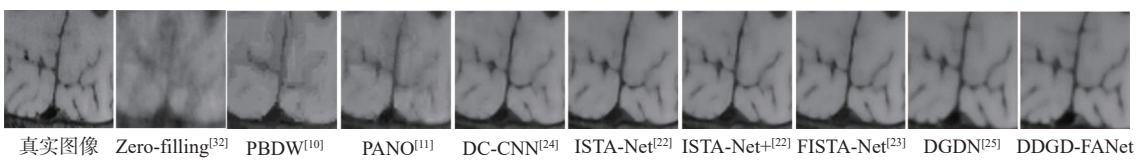
(b) 笛卡儿10%CS比率下重建后局部放大



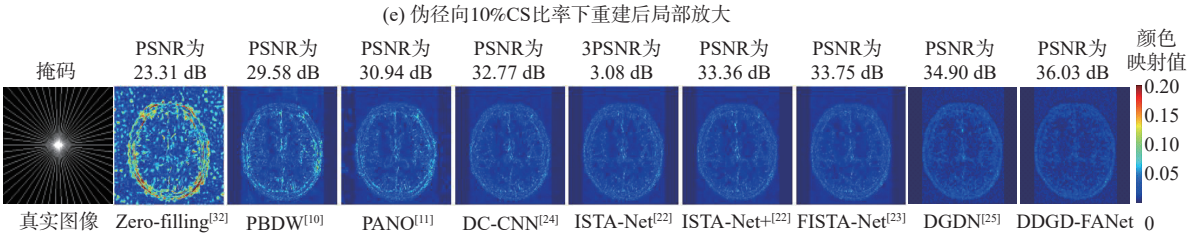
(c) 笛卡儿10%CS比率下重建后误差



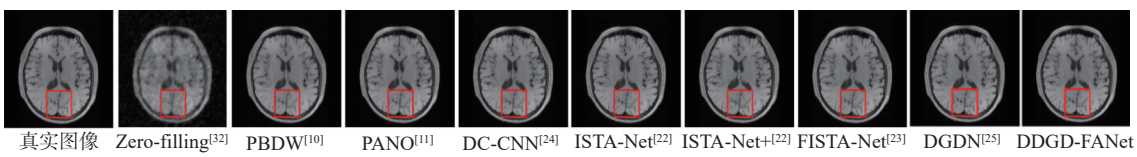
(d) 伪径向10%CS比率下重建后效果



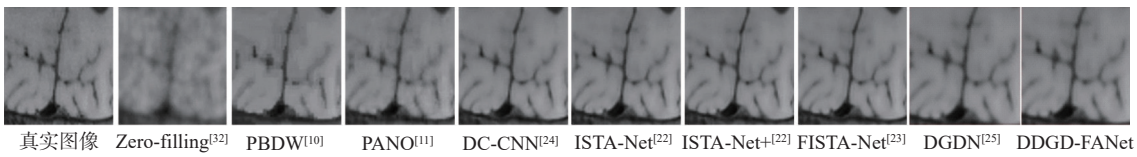
(e) 伪径向10%CS比率下重建后局部放大



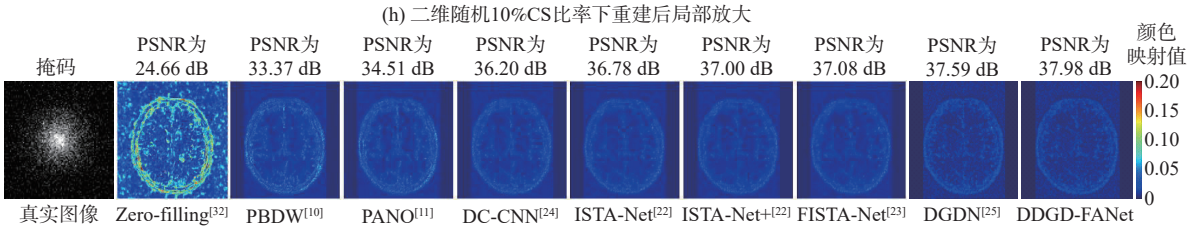
(f) 伪径向10%CS比率下重建后误差



(g) 二维随机10%CS比率下重建后效果



(h) 二维随机10%CS比率下重建后局部放大



(i) 二维随机10%CS比率下重建后误差

图 3 大脑数据集上使用不同采样模式的重建效果对比

Fig. 3 Comparison of reconstruction effects using different sampling patterns in brain dataset

表 3　本文模型消融实验结果

Table 3　Results of the proposed model in ablation experiment

模型	ADN		PSNR/dB	SSIM
	AM	FAF		
DGDN[25]			32.53	0.886 3
	√	√	33.58	0.902 1
DGDN[25]+DDGD			34.55	0.916 5
	√		34.70	0.917 6
		√	35.67	0.930 2
	√	√	35.87	0.932 0

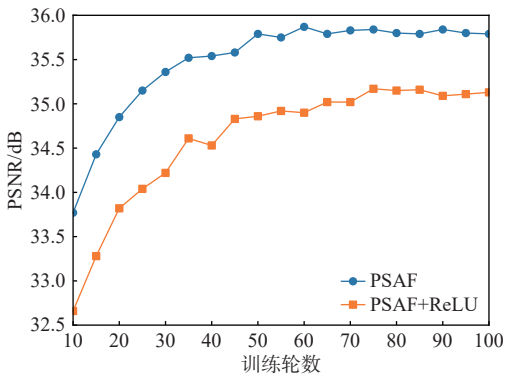


图 4　FAF 末尾使用激活函数对网络性能影响

Fig. 4　Effect of activation functions at end of FAF on network performance

层的 DDGD-FANet 模型, 结果表明数据一致性层是提升 MR 图像重建质量的重要方法。

表 4　数据一致性层消融实验结果

Table 4　Ablation experiment results of data consistency layer

模型	PSNR/dB	SSIM
DDGD-FANet_NODC	34.73	0.921 5
DDGD-FANet	35.87	0.932 0

3　结　论

1) 本文方法可以实现较为优异的重建性能, 在大脑数据集上, 笛卡儿 10%CS 比率情况下, PSNR 值达到了 35.87 dB, SSIM 指标为 0.932 0, 较对比方法有着不错的提升。

2) 由重建视觉效果对比可以得出, 双域几何蒸馏模块可以还原更多的纹理细节, 有着更高的重建质量, 保证了重构精度。

3) 本文方法不在特征自适应模块后使用非线性变换策略, 不但提升了重构性能, 还加快了模型的收敛速度。

在今后的工作中, 将细化双域几何蒸馏网络, 以减少更多的特征冗余。同时, 在迭代优化步骤中引入其他先验信息来更好的指导网络学习。

参考文献（References）

[1] BIAN W Y. Optimization-based deep learning methods for magnetic resonance imaging reconstruction and synthesis[D]. Gainesville: University of Florida, 2022.

[2] LUSTIG M, DONOHO D L, SANTOS J M, et al. Compressed sensing MRI[J]. IEEE Signal Processing Magazine, 2008, 25(2): 72-82.

[3] DONOHO D L. Compressed sensing[J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289-1306.

[4] QU X B, GUO D, CHEN Z, et al. Compressed sensing MRI based on nonsubsampled contourlet transform[C]//Proceedings of the IEEE International Symposium on IT in Medicine and Education. Piscataway: IEEE Press, 2008: 693-696.

[5] STALIDIS G, MAGLAVERAS N, DIMITRIADIS A, et al. Parametric 4D modeling of myocardial surface motion applied to MRI data[C]//Proceedings of the Computers in Cardiology. Piscataway: IEEE Press, 1997: 319-322.

[6] SONAWANE S S, AGNIHOTRI S P. Compressed MR imaging using wavelet transform[C]//Proceedings of the International Conference on Micro-Electronics and Telecommunication Engineering. Piscataway: IEEE Press, 2016: 259-264.

[7] LINGALA S G, JACOB M. Blind compressive sensing dynamic MRI[J]. IEEE Transactions on Medical Imaging, 2013, 32(6): 1132-1145.

[8] LIANG D, WANG H F, YING L. SENSE reconstruction with nonlocal TV regularization[C]//Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Piscataway: IEEE Press, 2009: 1032-1035.

[9] OUYANG Y, CHEN Y, WU Y. A spatial regularization framework of orientation diffusion functions using total variation and wavelet [C]//Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Piscataway: IEEE Press, 2011: 272-275.

[10] QU X B, GUO D, NING B D, et al. Undersampled MRI reconstruction with patch-based directional wavelets[J]. Magnetic Resonance Imaging, 2012, 30(7): 964-977.

[11] QU X B, HOU Y K, LAM F, et al. Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator[J]. Medical Image Analysis, 2014, 18(6): 843-856.

[12] BECK A, TEBOULLE M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1): 183-202.

[13] BOYD S, PARIKH N, CHU E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. Foundations and Trends in Machine Learning, 2010, 3(1): 1-122.

[14] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[15] WANG T H, LEI Y, FU Y B, et al. Medical imaging synthesis using deep learning and its clinical applications: a review[EB/OL]. (2020-04-21)[2023-02-01]. https://arxiv.org/abs/2004.10322v1.

[16] WANG S S, SU Z H, YING L, et al. Accelerating magnetic resonance imaging via deep learning[C]//Proceedings of the IEEE 13th International Symposium on Biomedical Imaging. Piscataway: IEEE Press, 2016: 514-517.

[17] YANG G, YU S M, DONG H, et al. DAGAN: deep de-aliasing

generative adversarial networks for fast compressed sensing MRI reconstruction[J]. IEEE Transactions on Medical Imaging, 2018, 37(6): 1310-1321.

[18] YANG Y, SUN J, LI H B, et al. ADMM-CSNet: a deep learning approach for image compressive sensing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(3): 521-538.

[19] WANG S S, XIAO T H, LIU Q G, et al. Deep learning for fast MR imaging: a review for learning reconstruction from incomplete k-space data[J]. Biomedical Signal Processing and Control, 2021, 68: 102579.

[20] AGHABIGLOU A, EKSIOGLU E M. Deep unfolding architecture for MRI reconstruction enhanced by adaptive noise maps[J]. Biomedical Signal Processing and Control, 2022, 78: 104016.

[21] YANG Y, SUN J, LI H B, et al. Deep ADMM-net for compressive sensing MRI[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: ACM, 2016: 10-18.

[22] ZHANG J, GHANEM B. ISTA-net: interpretable optimization-inspired deep network for image compressive sensing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 1828-1837.

[23] XIANG J X, DONG Y G, YANG Y J. FISTA-net: learning a fast iterative shrinkage thresholding network for inverse problems in imaging[J]. IEEE Transactions on Medical Imaging, 2021, 40(5): 1329-1339.

[24] SCHLEMPER J, CABALLERO J, HAJNAL J V, et al. A deep cascade of convolutional neural networks for dynamic MR image reconstruction[J]. IEEE Transactions on Medical Imaging, 2018, 37(2): 491-503.

[25] FAN X H, YANG Y, ZHANG J P. Deep geometric distillation network for compressive sensing MRI[C]//Proceedings of the IEEE EMBS International Conference on Biomedical and Health Informatics. Piscataway: IEEE Press, 2021: 1-4.

[26] LU C, JIANG B R, MU Y D. Fast fourier convolution[C]//Proceedings of the Neural Information Processing Systems. Vancouver: Curran Associates Inc, 2020: 4479-4488.

[27] SINGH N M, IGLESIAS J E, ADALSTEINSSON E, et al. Joint frequency and image space learning for MRI reconstruction and analysis[J]. The Journal of Machine Learning for Biomedical Imaging, 2022, 2022: 018.

[28] 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.

[29] CHENG J, WANG H F, ZHU Y J, et al. Model-based deep MR imaging: the roadmap of generalizing compressed sensing model using deep learning[EB/OL]. (2029-09-05)[2023-02-01]. https://arxiv.org/abs/1906.08143.

[30] CUI Z X, ZHU Q Y, CHENG J, et al. Deep unfolding as iterative regularization for imaging inverse problems[EB/OL]. (2022-11-24)[2023-02-01]. https://arxiv.org/abs/2211.13452v1.

[31] JIANG K, WANG Z Y, YI P, et al. Deep distillation recursive network for remote sensing imagery super-resolution[J]. Remote Sensing, 2018, 10(11): 1700.

[32] BERNSTEIN M A, FAIN S B, RIEDERER S J. Effect of windowing and zero-filled reconstruction of MRI data on spatial resolution and acquisition strategy[J]. Journal of Magnetic Resonance Imaging, 2001, 14(3): 270-280.

MRI reconstruction based on geometry distillation and feature adaptation

DUO Lin*, REN Yong, XU Boyu, YANG Xin

(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Although the existing compressed sensing-magnetic resonance imaging (CS-MRI) methods based on deep learning have achieved good results, the interpretability of these methods still faces challenges, and the transition from theoretical analysis to network design is not natural enough. In order to solve the above problems, this paper proposed a deep dual-domain geometry distillation feature adaptive network (DDGD-FANet). The deep unfolding network iteratively expanded the MRI reconstruction optimization problem into three sub-modules: data consistency module, dual-domain geometry distillation module, and adaptive network module. It could compensate for the lost context information of the reconstructed image, restore more texture details, remove global artifacts, and further improve the reconstruction effect. Three different sampling modes were used in the public dataset. The results show that DDGD-FANet achieves a higher peak signal-to-noise ratio and structural similarity index in all three sampling modes. At the Cartesian 10% compressed sensing(CS)ratio, the peak signal-to-noise ratio is increased by 5.01 dB, 4.81 dB, and 3.34 dB, respectively, higher than that of iterative shrinkage-thresholding algorithm (ISTA)-Net +, fast ISTA (FISTA)-Net, and DGDN models.

Keywords: magnetic resonance imaging; image reconstruction; deep learning; geometry distillation; adaptive network; attention mechanism

Received: 2023-06-07; **Accepted**: 2023-08-18; **Published Online**: 2023-09-12 16:19

URL: link.cnki.net/urlid/11.2625.V.20230911.1850.002

Foundation item: National Natural Science Foundation of China (61962032)

* **Corresponding author**. E-mail: duolin2003@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0334

基于多注意力与双模板更新的视觉跟踪算法

马素刚[1,2,*]，孙思维[1]，侯志强[1]，余旺盛[3]，蒲磊[4]

(1. 西安邮电大学 计算机学院，西安 710121；2. 西安邮电大学 陕西省网络数据分析与智能处理重点实验室，西安 710121；
3. 空军工程大学 信息与导航学院，西安 710077；4. 火箭军工程大学 作战保障学院，西安 710025)

摘 要：针对全卷积孪生网络(SiamFC)跟踪器在复杂场景下表征能力不足且缺乏在线更新问题，提出一种基于多注意力与双模板更新的视觉跟踪算法。使用VGG16网络替换AlexNet，用SoftPool代替最大池化层，构建特征提取网络；在骨干网络后添加多注意力模块(MAM)，增强网络对目标特征的提取能力；设计双模板进行特征融合和响应图融合，使用平均峰值相关能量(APCE)判断是否更新动态模板，有效提高跟踪鲁棒性。在GOT-10k数据集上对所提算法进行训练，并分别在OTB2015、VOT2018和UAV123数据集上进行测试，实验结果表明：相较于基准SiamFC算法，所提算法在OTB2015和UAV123数据集上，跟踪成功率分别提高了0.085和0.037，精确度分别提升了0.118和0.058；在VOT2018数据集上，跟踪准确率、鲁棒性和期望平均重叠率(EAO)分别提升了0.030、0.295和0.139。所提算法在复杂场景下取得了较高的跟踪准确度，并且运行速度达到33.9帧/s，满足实时跟踪要求。

关 键 词：视觉目标跟踪；孪生网络；SoftPool；注意力机制；特征融合
中图分类号：TP391.4
文献标志码：A **文章编号**：1001-5965(2025)06-1955-10

目标跟踪是计算机视觉领域研究的一个热点问题，在视频监控[1]、智能交通[2]、医学诊断[3]、人机交互[4]等许多场景有着广泛的应用前景。该任务的目标是对于任一视频序列，在初始帧中给定被跟踪的目标，在后续每帧中对目标运动状态进行预测并标定目标的位置和大小。由于在跟踪过程中算法只能在初始帧中对目标进行学习，且跟踪过程的环境复杂多变，在实际应用中难度变大。因此，如何在目标遮挡、变形、背景干扰等因素的影响下，实现精确、高速地跟踪，仍然具有很大的挑战性。

近年来，基于深度学习的目标跟踪算法成为主流。2015年，Danelljan等[5]结合浅层卷积特征与空间正则化判别相关滤波器(spatially regularized discriminative correlation filters, SRDCF)算法[6]提出深度空间正则化判别相关滤波器(deep spatially regularized discriminative correlation filters, DeepSRDCF)算法，但其速度慢，跟踪实时性较差。Bertinetto等[7]提出的全卷积孪生网络(fully-convolutional Siamese networks, SiamFC)算法，其核心是将模板和搜索区域输入AlexNet网络提取特征，网络共享权重，通过模板匹配定位目标，用互相关操作计算相似度得到目标位置。虽然SiamFC算法获得了不错的跟踪精确度和速度，但存在如下问题：①用层数较少的AlexNet网络提取目标特征，该浅层网络在复杂环境下的跟踪效果较差。②为保证速度，不进行模板更新，导致模型的鲁棒性不佳。

针对这些问题，出现了很多改进算法。在提高

收稿日期：2023-06-09；**录用日期**：2023-07-21；**网络出版时间**：2023-08-25 14:30
网络出版地址：link.cnki.net/urlid/11.2625.V.20230824.1905.002
基金项目：国家自然科学基金(62072370)；陕西省自然科学基金(2023-JC-YB-598)；西安市科技计划(22GXFW0125)
*通信作者. E-mail：msg@xupt.edu.cn

引用格式：马素刚，孙思维，侯志强，等. 基于多注意力与双模板更新的视觉跟踪算法[J]. 北京航空航天大学学报，2025，51(6)：1955-1964. MA S G，SUN S W，HOU Z Q，et al. Visual tracking algorithm based on multi-attention and dual-template update[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1955-1964 (in Chinese).

特征表示能力方面，Zhang和Peng[8]提出内部裁剪残差(cropping-inside-residual, CIR)模块以取代残差网络(residual network, ResNet)中的基本模块，实现更广更深网络的应用。Li等[9]提出孪生区域候选网络(Siamese region proposal network, SiamRPN)算法，通过对目标的尺度估计和目标定位，获得更精确的目标边界框。Danelljan等[10]提出通过重叠最大化实现精确跟踪(accurate tracking by overlap maximization, ATOM)算法，将目标跟踪分为目标估计和目标分类2部分，分别用于精确的边界框输出和目标置信度判别。基于ATOM网络框架，Danelljan等又提出使用18层深度残差网络的判别模型预测(discriminative model prediction with an eighteen-layer ResNet network, DiMP18)算法[11]，通过具有判别能力强的损失函数和快速迭代方法优化网络，在算法精度上表现突出。Wang等[12]提出一种结合通道注意力与空间注意力机制的残差注意力模块，提升了模型的特征选择能力。在模板更新方面，使用方向梯度直方图特征和颜色特征的高效卷积算子(efficient convolution operators with HOG and CN, ECO-HC)算法[13]采用稀疏更新策略，每隔5帧做一次优化更新模型参数。Zhang等[14]设计了一个自适应的更新模块UpdateNet，采用残差拟合的模式更新初始模板。

本文在SiamFC算法框架基础上，提出一种基于多注意力与双模板更新的视觉跟踪算法。主要创新点如下：

1）在改进的16层视觉几何组网络(visual geometry group network with 16 layers, VGG16)[15]骨干网络中引入软池化SoftPool[16]，保留更多相关信息的特征，缓解网络过拟合风险。

2）在骨干网络后添加多注意力模块(multi-attention module, MAM)，使跟踪器关注重要的通道和空间信息，增强网络的特征表达能力。

3）设计双模板进行特征融合和响应图融合，使用平均峰值相关能量(average peak correlation energy, APCE)判断是否更新动态模板，提高模型鲁棒性。

4）将本文算法在GOT-10k[17]数据集上训练，在OTB2015[18]、VOT2018[19]和UAV123[20]数据集上进行大量实验测试算法性能。

1 本文算法

本文算法的总体框架如图1所示。卷积神经网络(convolutional neural network, CNN)为使用SoftPool改进的VGG16网络，用来提取模板与搜索区域的深度特征。引入MAM对提取特征的通道和空间信息进行增强。利用双模板更新策略，选取初始图像中的目标作为初始模板，采用APCE判断是否更新动态模板，设计了双模板，进行特征融合和响应图融合，提高模板的抗干扰能力，有效提高了跟踪鲁棒性。

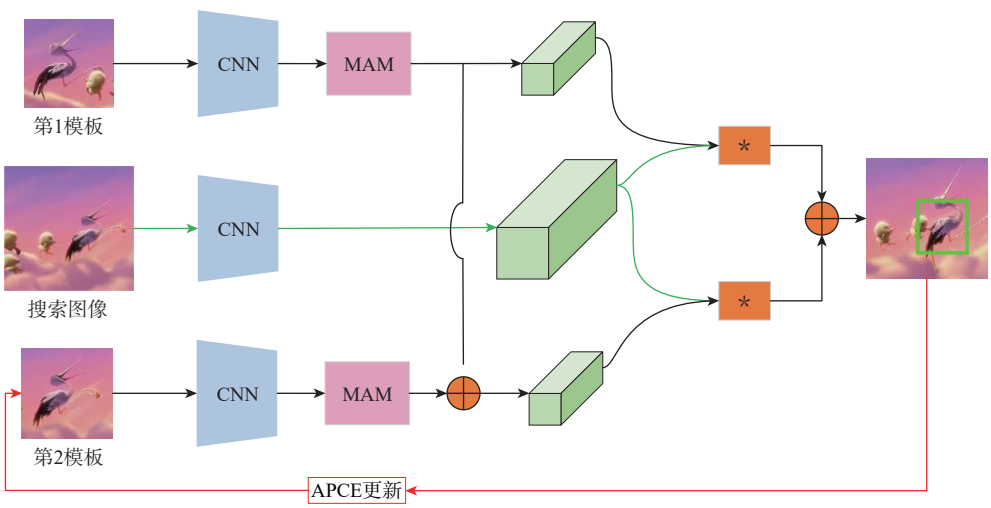


图1 本文算法总框架

Fig. 1 Overall framework of the proposed algorithm

1.1 骨干网络

池化操作的本质是对特征图下采样，降低网络训练的难度。用VGG16中的最大池化进行下采样时可能会导致信息丢失。针对这个问题，本文选择软池化SoftPool替换最大池化。SoftPool是一种高效、快速的池化方式，在特征图的池化区域R内通过Softmax方式实现池化。在池化过程中，SoftPool为R内的每个元素a_i应用一个权重w_i，该权重为元素a_i的自然指数相对于R内所有元素的自然指数之和的比值。SoftPool输出值$\tilde{a}$为池化区域R内所有

元素的加权求和, 如下:

$$\tilde{a}=\sum_{i\subset R}w_i a_i=\sum_{i\subset R}\left(\mathrm{e}^{a_i}a_i/\sum_{j\subset R}\mathrm{e}^{a_j}\right) \tag{1}$$

式中: $|R|=k^2$, k为池化区域的宽或高。

骨干网络结构如图 2 所示。

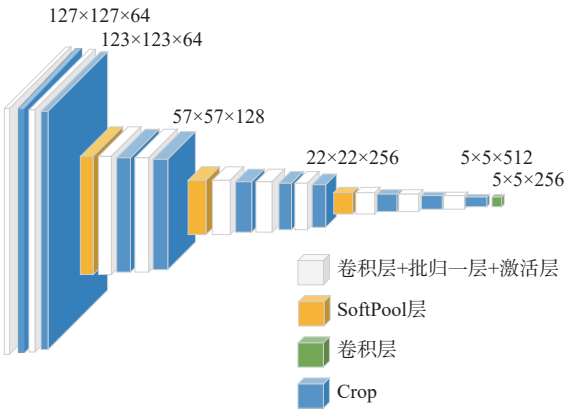


图 2　骨干网络

Fig. 2　Backbone network

1.2　多注意力模块

注意力机制来源于人类的视觉注意力机制, 将有限的注意力集中在真正需要关注的信息上, 过滤掉无效信息。通道注意力对于提升模型性能具有显著效果, 经典的 Self-attention[21] 通过建模通道关系来重新衡量每个通道的重要性, 但忽略了位置信息, 而位置信息对于视觉中目标的空间结构有重要作用。卷积块注意力模块(convolutional block attention module, CBAM)[22] 通过在通道上进行全局池化引入位置信息, 将通道注意力与空间注意力联合起来, 但 CBAM 只能捕获局部的信息, 而无法获取长范围依赖的信息。

本文设计了一种 MAM 模块, 如图 3 所示, 由门控通道变换单元(gated channel transformation, GCT)[23] 和坐标注意力(coordinate attention, CA)[24] 2 个子模块组成。输入特征分别经过 GCT 和 CA 注意力, 然后进行加权融合, 再与原始特征加权得到最终的输出特征。

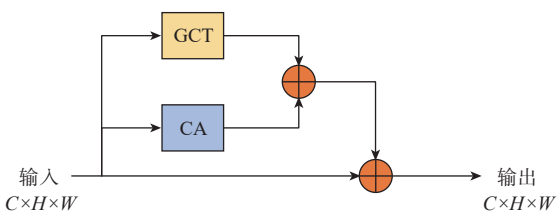


图 3　MAM

Fig. 3　MAM

1.2.1　门控通道变换单元

门控通道变换单元是一种轻量级、可解释的单元, 可以促进特征通道间的合作或竞争关系。

该模块通过自适应学习每个特征通道并给其分配不同的权重, 学习每个特征通道的重要程度, 突出有用特征并削弱无用特征, 从而促进特征间的关系。

定义输入信号 $X=\{\boldsymbol{x}_1,\boldsymbol{x}_2\cdots,\boldsymbol{x}_C\}$, 其中, $\boldsymbol{x}_c=\left[\boldsymbol{x}_c^{i,j}\right]_{H\times W}\in R^{H\times W}$, 通道索引 $c\in\{1,2,\cdots,C\}$, C为通道数, H和W为高和宽。$\boldsymbol{x}_c$为通道c对应的输入特征。GCT 结构分为以下 3 部分:

1) 全局上下文嵌入。大的感受野意味着可能蕴含更为全局、语义层次更高的特征, 所以嵌入全局上下文信息并在归一化之前控制每个通道的权重。嵌入权重 $\boldsymbol{\alpha}=\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_C\}$, 此模块如下:

$$\boldsymbol{s}_c=\boldsymbol{\alpha}_c\|\boldsymbol{x}_c\|_2=\boldsymbol{\alpha}_c\left\{\left[\sum_{i=1}^{H}\sum_{j=1}^{W}\left(x_c^{i,j}\right)^2\right]+\xi\right\}^{\frac{1}{2}} \tag{2}$$

式中: $\boldsymbol{s}_c$为激励函数; ξ为较小常数, 避免零点的推导问题; $\boldsymbol{\alpha}_c$为可训练参数, 控制每个通道的权重。

2) 通道归一化。定义 $\boldsymbol{S}=\{\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_C\}$, 通过少量的计算建立通道之间的竞争关系。此模块如下:

$$\hat{\boldsymbol{s}}_c=\frac{\sqrt{C}\boldsymbol{s}_c}{\|\boldsymbol{S}\|_2}=\frac{\sqrt{C}\boldsymbol{s}_c}{\left[\left(\sum_{c=1}^{C}\boldsymbol{s}_c^2\right)+\xi\right]^{\frac{1}{2}}} \tag{3}$$

式中: $\sqrt{C}$为尺度因子, 作用是规范化$\hat{\boldsymbol{s}}_c$, 防止当通道数C较大时$\hat{\boldsymbol{s}}_c$的比例过小。

3) 门限自适应。可以适应原有特征, 使得 GCT 模块促进通道间竞争或合作关系。此模块如下:

$$\hat{\boldsymbol{x}}_c=\boldsymbol{x}_c\left[1+\tanh\left(\gamma_c\hat{\boldsymbol{s}}_c+\beta_c\right)\right] \tag{4}$$

式中: $\hat{\boldsymbol{x}}_c$为通道c的输出信号; γ_c、β_c为特征权值, 是一组可训练参数, 为控制激活通道特征。当一个通道的特征权值γ_c大于 0 时, GCT 会促进该通道与其他通道竞争; 相反, GCT 模块便会促进该通道与其他通道进行合作。

1.2.2　坐标注意力

坐标注意力机制提出分解通道注意力为 2 个并行的一维特征编码, 高效地整合空间坐标信息到生成的注意力图像中。

对于给定的输入$\boldsymbol{x}$, 分别使用尺寸为$H\times1$和$1\times W$的池化层沿着水平坐标和垂直坐标对每个通道进行编码。高度为h的第c通道的一维特征编码 $\boldsymbol{z}_c^h$如下:

$$\boldsymbol{z}_c^h(h)=\frac{1}{W}\sum_{0\leqslant i<W}x_{\mathrm{c}}(h,i) \tag{5}$$

同样地, 宽度为w的第c通道的一维特征编码 $\boldsymbol{z}_c^w$如下:

$$z_c^w(w)=\frac{1}{H}\sum_{0\leqslant j<H}\boldsymbol{x}_c(j,w) \tag{6}$$

将式(6)生成的2个一维特征编码沿着空间维度进行拼接，再经过1×1卷积操作和激活函数得到

$$\boldsymbol{f}=\text{h_swish}(F_1([\boldsymbol{z}^h,\boldsymbol{z}^w])) \tag{7}$$

式中：[,]为沿空间维数的拼接操作。

之后沿着空间维度将$\boldsymbol{f}$分解为2个单独的向量$\boldsymbol{f}^h$和$\boldsymbol{f}^w$，利用2个1×1卷积和sigmoid激活函数得到2个注意力向量$\boldsymbol{g}^h$和$\boldsymbol{g}^w$，如下：

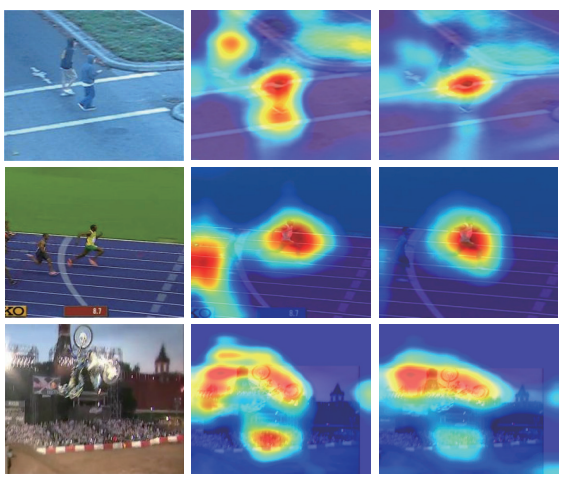
$$\boldsymbol{g}^h=\text{sigmoid}(F_h(\boldsymbol{f}^h)) \tag{8}$$

$$\boldsymbol{g}^w=\text{sigmoid}(F_w(\boldsymbol{f}^w)) \tag{9}$$

最后，将2个注意力向量变换到$C\times H\times W$的维度后与经过残差操作的输入特征图$\boldsymbol{x}_c$进行点乘操作，得到输出特征$\boldsymbol{y}$如下：

$$\boldsymbol{y}=\boldsymbol{x}\cdot\boldsymbol{g}^h\cdot\boldsymbol{g}^w \tag{10}$$

图4为使用MAM前后的注意力可视化图，其中，红色区域为网络模型的重点关注区域，使用MAM后网络将注意力集中在目标信息上，过滤掉了更多干扰信息。



(a) 搜索图像 (b) 使用MAM前 (c) 使用MAM后

图4 注意力可视化

Fig. 4 Visualization of attention

1.3 双模板更新策略

单一固定的初始模板无法很好地应对目标的各种复杂变化，进而导致跟踪漂移。为缓解这一问题，本文采用双模板的更新策略进行跟踪，除使用初始帧中标注的目标作为初始模板外，增加一个动态模板分支，使用跟踪置信度APCE对响应图进行评价，判断动态模板是否更新。当APCE值突然减少时，可能发生目标遮挡或丢失等情况，容易发生模型漂移。只有当APCE值及响应图峰值$F_{\max}$分别以一定比例μ大于历史APCE均值和历史峰值均值时更新动态模板，具体计算式为

$$\begin{cases}V_{\text{APCE}}>\mu\text{mean}\left(\sum V_{\text{APCE}}\right)\\ F_{\max}>\mu\text{mean}\left(\sum F_{\max}\right)\end{cases} \tag{11}$$

式中：权重$\mu\in[0,1]$。μ越小，模板更新越频繁，动态模板的可靠性可能降低，跟踪速度也可能降低。

假设初始模板为z_{r}，动态模板为z_{t}，搜索区域为x，骨干网络提取特征操作为φ。为充分发挥双模板的优势，利用特征融合和响应图融合，在响应图融合前先融合初始模板分支和动态模板分支的特征，再融合初始模板的响应图和融合后特征生成的响应图，从而得到更加准确的目标位置。

1）特征融合(Update-A)。按比例融合初始模板特征$\varphi(z_{\text{r}})$和动态模板特征$\varphi(z_{\text{t}})$，得到最终的融合后的模板特征$\varphi(z_{\text{r}},z_{\text{t}})$，如下：

$$\varphi(z_{\text{r}},z_{\text{t}})=\alpha\varphi(z_{\text{r}})+(1-\alpha)\varphi(z_{\text{t}}) \tag{12}$$

2）响应图融合(Update-B)。利用融合后的模板特征和搜索区域特征计算结果响应图$f(z_{\text{t}},x)$。对初始模板结果响应图$f(z_{\text{r}},x)$与融合模板特征结果响应图$f(z_{\text{t}},x)$采用加权求和，得到最终的结果响应图$f(z_{\text{r}},z_{\text{t}},x)$。最终的响应图$F(z_{\text{r}},z_{\text{t}},x)$计算式为

$$\begin{cases}f(z_{\text{r}},x)=\varphi(z_{\text{r}})\varphi(x)\\ f(z_{\text{t}},x)=\varphi(z_{\text{r}},z_{\text{t}})\varphi(x)\\ F(z_{\text{r}},z_{\text{t}},x)=\beta f(z_{\text{r}},x)+(1-\beta)f(z_{\text{t}},x)\end{cases} \tag{13}$$

式中：$\alpha\in[0,1]$为特征融合时的权重；$\beta\in[0,1]$为响应图融合时的权重。

对上述融合方法进行实验测试表明，使用简单的响应图融合方式可获得性能提升，在添加特征融合后再融合响应图性能可以进一步得到提升。

1.4 本文算法流程

为保证跟踪的实时性，本文算法采用了三尺度$[1.037\,5^{-1},1.037\,5^{0},1.037\,5^{1}]$搜索的尺度估计策略。具体算法流程如下：

输入 图像序列：$I=\{I_1,I_2,\cdots,I_n\}$，初始目标位置：$P_1=(x_1,y_1,w_1,h_1)$。

输出 第i帧的目标位置：$P_i=(x_i,y_i,w_i,h_i)$ (i=2,3,$\cdots$,n)。

for i =1, 2, $\cdots$, n, do:

if i=1:

1.裁剪出当前序列中的感兴趣区域，缩放到127×127大小。

2.通过主干网络提取深度特征，并通过模板分支的MAM进行特征增强。将初始目标位置P_1作为初始模板，记为z_{r}，动态模板记为z_{t}。

else:

1.根据前一帧的跟踪结果P_{i-1}获取当前帧的搜索区域x。

2.计算初始模板的特征$\varphi(z_r)$及生成的响应图$f(z_r, x)$。

3.使用双模板更新策略。如果$V_{APCE} > \mu \text{mean}\left(\sum V_{APCE}\right)$ and $F_{max} > \mu \text{mean}\left(\sum F_{max}\right)$, 则更新动态模板$z_t$; 计算动态模板$z_t$的特征$\varphi(z_t)$及生成的响应图$f(z_t, x)$。

4.计算最终响应图, 根据响应图对目标位置进行计算, 输出当前帧的跟踪结果P_i。

Return P_i

end if

end for

2　实　验

本文算法在 Ubuntu18.04.6 系统环境, 基于深度学习框架 Pytorch1.8.1, 用 Python 编程实现, 并使用 Pysot 工具箱对算法结果进行评估。硬件环境为: Intel 4 216 CPU, 128 GB 内存, 2.10 GHz 和 NVIDIA GTX1080Ti GPU。

在训练阶段, 使用 GOT-10k 作为训练数据集, 该数据集包含 66 GB 图片序列, 超过 10 000 条视频和 563 个目标类别。测试阶段采用 OTB2015、VOT2018 和 UAV123 基准数据集, 与其他一些先进跟踪器的性能作对比。

2.1　OTB2015 数据集实验分析

OTB2015 数据集由 100 个具有挑战性的视频序列组成, 是一个用于跟踪算法性能评估的代表性数据集。跟踪算法的评估有 2 个指标: 精确度 (Precision) 和成功率 (Success Rate)。精确度曲线代表预测框中心和真实目标框中心小于规定阈值的帧数占总帧数的百分比, 通常阈值为 20 像素。成功率代表算法输出预测框与真实目标框的交并比大于规定阈值的帧数占总帧数的百分比。

通过大量实验, 实验结果如表 1 所示, 其中, 粗体为本文算法在 OTB2015 数据集的参数取值, μ为 0.85, α、β权重分别设置为 0.5、0.5。

表 1　不同取值在 OTB2015 数据集上的实验结果

Table 1　Experimental results of different values on OTB2015 dataset

μ	成功率	精确度
0.6	0.657	0.884
0.65	0.664	0.892
0.7	0.651	0.875
0.75	0.658	0.882
0.8	0.659	0.884
0.85	**0.667**	**0.899**
0.9	0.661	0.885
0.95	0.657	0.880
1.0	0.655	0.881

2.1.1　定性分析

为分析本文算法的性能, 图 5 展示了本文算法与干扰感知的孪生区域建议网络 (distractor-aware Siamese region proposal network, DaSiamRPN)[25]、SRDCF、SiamFC、ATOM 和更深更宽的孪生网络

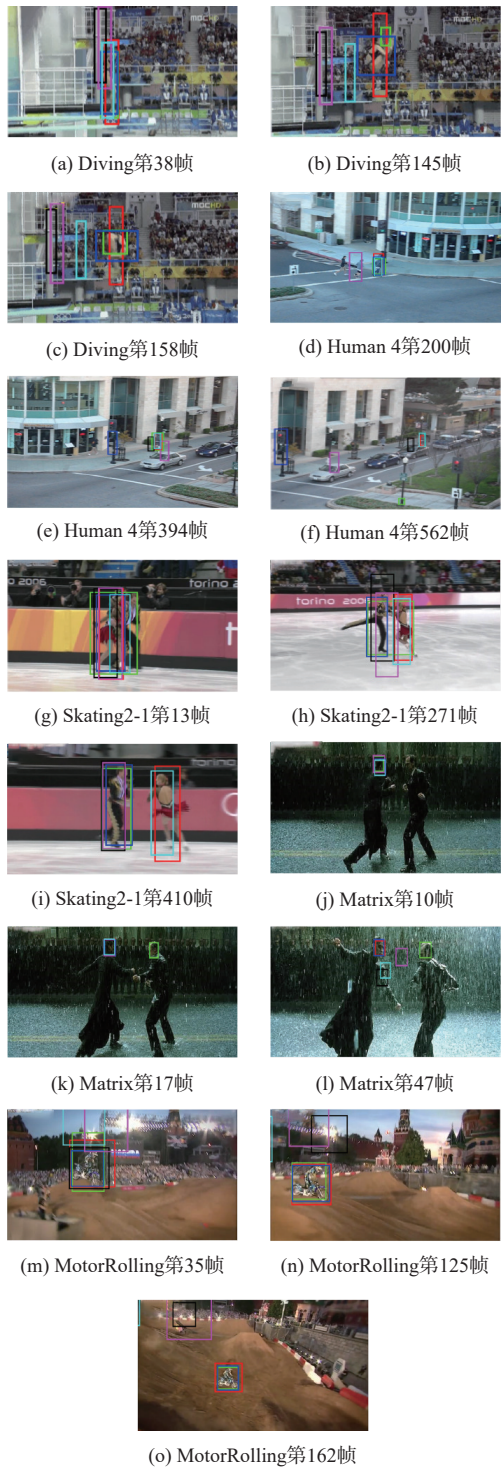


图 5　5 组视频序列的跟踪效果对比

Fig. 5　Comparison of tracking effect of five groups of video sequences

(deeper and wider siamese networks, SiamDW)[8]跟踪算法在5组视频序列(Diving、Human4、Skating2-1、Matrix和MotorRolling)的定性对比结果。

1)目标旋转。在视频序列Diving、Skating2-1和Matrix中,图5(b)、图5(h)和图5(l)出现目标平面内旋转现象,本文算法均能较好地完成跟踪。由此可见,本文算法构建提取深度特征网络有助于稳定跟踪。

2)目标遮挡。目标遮挡是跟踪中常见的问题。图5(f)中因为目标被遮挡,ATOM、DaSiamRPN、SiamFC和SiamDW算法均出现目标丢失,而本文算法可以跟踪到目标。由此可见,本文算法通过双模板更新策略,相比其他算法可以有效避免由遮挡引起的跟踪漂移问题。

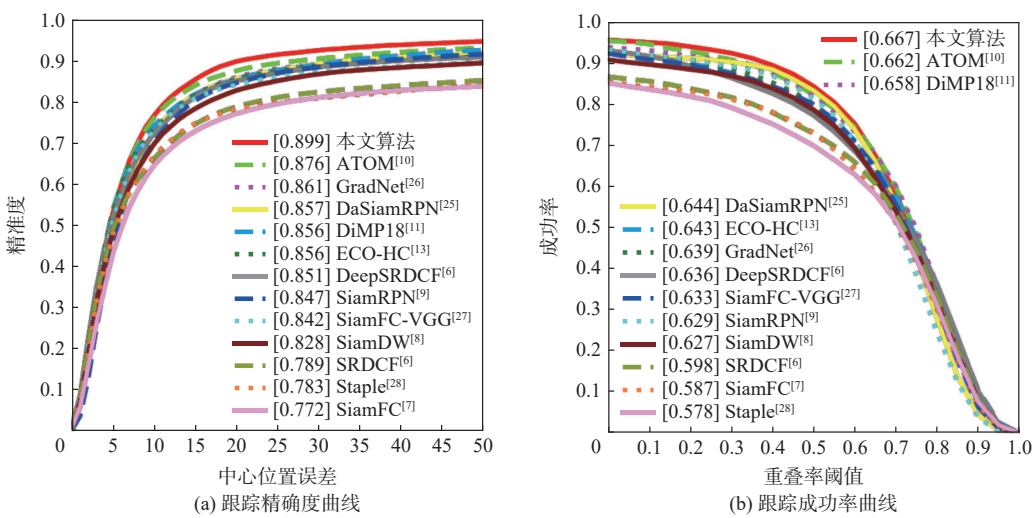
3)运动模糊。目标位置快速改变,导致图像模糊。图5(i)中,除本文算法和SRDCF算法,其他全部跟踪失败。图5(m)中,SiamF和SRDCF算法丢失目标,图5(n)中,SiamDW算法丢失目标,而本文算法能够很好地应对目标的运动模糊,较好地完成跟踪。由此可见,本文算法通过改进的VGG16网络结构和MAM提取特征,包含大量位置信息,并通过双模板更新补充信息,这使得本文算法适用于快速移动和运动模糊的场景。

4)背景干扰。图5(l)中,只有本文算法和SRDCF算法跟踪结果依然良好,其他均丢失目标。SRDCF算法通过在目标函数中加入显式的空间正则化项来抑制背景干扰。而本文算法有较高的特征提取能力,能更准确地区分相似背景,定位重点区域的确切位置,实现稳定跟踪。

2.1.2 定量分析

为进一步验证本文算法的性能,在OTB2015数据集上进行评估,选取了一些最新的基于孪生网络的跟踪算法,以及其他主流的基于判别相关滤波器的跟踪算法进行比较。实验中,将本文算法与SiamDW、DaSiamRPN、ECO-HC、梯度引导网络(gradient-guided network, GradNet)[26]、SRDCF、SiamRPN、SiamFC-VGG[27]、ATOM、Staple[28]、Dimp-18、DeepSRDCF、SiamFC这12种有代表性的跟踪算法进行比较。图6给出了各算法在OTB2015数据集上的精确度和成功率曲线。可以看出,本文算法的跟踪成功率和精确度分别为0.667和0.899,与基准SiamFC算法相比,分别提高了0.085和0.118。相比于算法ATOM,本文算法在跟踪精确度方面仍具有优势。

本文提供了在OTB2015数据集上基于11种属性的性能分析,将本文算法与对比算法分别进行成功率和精确度比较,如表2和表3所示,表中,加粗数值表示对应属性下的最优结果,下划线数值为次优结果。可以看出,本文算法在背景干扰、光照变化、运动模糊、平面外旋转属性下,跟踪成功率均处于最优。



(a) 跟踪精确度曲线 (b) 跟踪成功率曲线

图6 不同算法在OTB2015数据集上的精确度和成功率曲线

Fig. 6 Curves of precision rate and success rate of different algorithms on OTB2015 dataset

本文算法在OTB2015数据集的平均跟踪速度为33.9帧/s,基准算法SiamFC的跟踪速度为70帧/s。相较于基准算法,跟踪速度有所下降,但仍能满足实时跟踪的需要,同时跟踪性能得到明显提升。

2.2 消融实验

为验证本文算法各组成部分的有效性,在OTB2015数据集上对各个模块进行有效性测试。消融实验结果如表4所示,以SiamFC算法为基准,替换网络为VGG16网络,添加SoftPool;在添加SoftPool模块的基础上加入MAM,MAM中2个模块具有互补性,两者相互配合更加有效地结合空间与通道信息;再添加2种融合方式Update-A和Update-

表 2　不同属性在 OTB2015 数据集上的成功率

Table 2　Success rate of different attributes on OTB2015 dataset

算法	背景干扰	形变	快速运动	光照变化	平面内旋转	低分辨率	运动模糊	遮挡	平面外旋转	出视野	尺度变化
本文	**0.663**	0.624	0.664	**0.690**	0.655	0.678	**0.691**	0.636	**0.647**	0.621	0.644
ATOM[10]	0.619	**0.631**	0.642	0.655	0.635	**0.705**	0.659	**0.637**	0.629	0.609	0.671
DiMP18[11]	0.616	0.617	**0.674**	0.654	0.650	0.611	0.679	0.636	0.635	**0.628**	**0.675**
DaSiamRPN[25]	0.625	0.599	0.655	0.670	**0.669**	0.594	0.651	0.577	0.636	0.605	0.652
ECO-HC[13]	0.636	0.595	0.634	0.640	0.582	0.533	0.627	0.629	0.608	0.594	0.611
GradNet[26]	0.611	0.572	0.625	0.643	0.627	0.673	0.647	0.617	0.629	0.585	0.614
DeepSRDCF[6]	0.627	0.566	0.628	0.621	0.589	0.564	0.643	0.602	0.607	0.555	0.606
SiamFC-VGG[27]	0.591	0.603	0.600	0.631	0.623	0.689	0.630	0.576	0.614	0.533	0.619
SiamRPN[9]	0.591	0.617	0.600	0.649	0.628	0.642	0.623	0.586	0.625	0.544	0.615
SiamDW[8]	0.574	0.560	0.630	0.622	0.606	0.598	0.654	0.602	0.612	0.592	0.614
SRDCF[6]	0.583	0.544	0.597	0.613	0.544	0.514	0.541	0.559	0.550	0.474	0.561
SiamFC[7]	0.527	0.512	0.571	0.572	0.559	0.621	0.555	0.550	0.561	0.511	0.557
Staple[28]	0.560	0.551	0.540	0.592	0.548	0.394	0.594	0.543	0.533	0.460	0.521

表 3　不同属性在 OTB2015 数据集上的精确度

Table 3　Precision of different attributes on OTB2015 dataset

算法	背景干扰	形变	快速运动	光照变化	平面内旋转	低分辨率	运动模糊	遮挡	平面外旋转	出视野	尺度变化
本文	**0.895**	**0.870**	**0.890**	**0.914**	0.912	0.983	**0.912**	**0.859**	**0.903**	**0.845**	**0.884**
ATOM[10]	0.806	0.856	0.828	0.866	0.869	0.993	0.855	0.835	0.851	0.808	0.877
DiMP18[11]	0.801	0.823	0.857	0.849	0.866	0.856	0.850	0.835	0.843	0.820	0.872
DaSiamRPN[25]	0.843	0.814	0.858	0.895	**0.913**	0.922	0.856	0.764	0.867	0.787	0.868
ECO-HC[13]	0.850	0.806	0.829	0.820	0.800	0.888	0.802	0.848	0.834	0.818	0.822
GradNet[26]	0.822	0.795	0.838	0.844	0.860	**0.999**	0.855	0.838	0.872	0.789	0.841
DeepSRDCF[6]	0.841	0.783	0.814	0.791	0.818	0.847	0.823	0.825	0.835	0.781	0.819
SiamFC-VGG[27]	0.761	0.828	0.777	0.817	0.845	0.997	0.817	0.764	0.833	0.706	0.834
SiamRPN[9]	0.799	0.825	0.789	0.859	0.854	0.978	0.816	0.780	0.851	0.726	0.838
SiamDW[8]	0.762	0.763	0.808	0.794	0.824	0.901	0.841	0.798	0.829	0.781	0.819
SRDCF[6]	0.775	0.734	0.768	0.792	0.745	0.760	0.765	0.734	0.741	0.594	0.745
SiamFC[7]	0.692	0.691	0.744	0.736	0.743	0.900	0.707	0.723	0.758	0.673	0.736
Staple[28]	0.749	0.752	0.708	0.783	0.768	0.690	0.698	0.726	0.737	0.664	0.726

表 4　本文算法在 OTB2015 数据集上的消融实验

Table 4　Ablation experiment of the proposed algorithm on OTB2015 dataset

SiamFC[7]	VGG16[15]	SoftPool[16]	CA[24]	GCT[23]	MAM	Update-A	Update-B	成功率	精确度
√								0.582	0.781
√	√							0.632	0.848
√	√	√						0.643	0.855
√	√	√	√					0.648	0.866
√	√	√		√				0.655	0.875
√	√	√			√			0.660	0.888
√	√	√			√		√	0.663	0.892
√	√	√			√	√	√	0.667	0.899

B。所有组件均不同程度地提高了跟踪器的跟踪性能。

2.3　VOT2018 数据集实验分析

VOT2018 数据集包含 60 个经过精确标注的短时跟踪视频序列, 包括目标遮挡、尺度变化、运动变化、光照变化和相机运动 5 种挑战属性。VOT2018 数据集主要通过 3 个指标分析算法性能, 分别是准确率 (accuracy, A)、鲁棒性 (robustness, R) 及期望平均重叠率 (expected average overlap, EAO)。

准确率是有效帧上跟踪预测框与真实框之间重叠率的平均值，鲁棒性是跟踪失败次数的具体衡量，EAO是通过计算跟踪器在大量的具有相同视觉特性的短期序列上所期望得到的平均重叠率。

在VOT2018数据集上确定了μ为0.9，α、β权重分别设置为0.45、0.5。在VOT2018数据集上与其他7个算法进行比较，如表5所示，实验结果表明，本文算法相较于基准SiamFC算法，跟踪准确率、鲁棒性和EAO分别提升了0.030、0.295、0.139，取得了较高的跟踪性能。

表5 不同算法在VOT2018数据集上的性能比较

Table 5 Comparisons of performance among different algorithms on VOT2018 dataset

算法	准确率	鲁棒性	EAO
ECO-HC[13]	0.484	0.276	0.280
SiamRPN[9]	0.490	0.460	0.244
CCOT[29]	0.494	0.318	0.267
SiamFC[7]	0.503	0.585	0.188
DSiam[30]	0.512	0.646	0.196
UpdateNet[14]	0.518	0.454	0.244
SiamFC-VGG[27]	0.531	0.318	0.286
本文算法	0.533	0.290	0.327

2.4 UAV123数据集实验分析

UAV123数据集是一个无人机捕获的数据集，包括123个低空空中视频，每个视频平均915帧。在UAV123数据集上确定了μ为0.85，α、β权重分别设置为0.5、0.55。如表6所示，本文算法的成功率和精确度分别为0.542、0.764，相较于基准算法分别提高了0.037和0.058。

表6 不同算法在UAV123数据集上的性能比较

Table 6 Comparisons of performance among different algorithms on UAV123 dataset

算法	成功率	精确度
MCCT[31]	0.453	0.656
SRDCF[6]	0.464	0,676
ARCF[32]	0.465	0.669
AutoTrack[33]	0.467	0.686
STRCF[34]	0.478	0.678
ECO-HC[13]	0.493	0.707
CCOT[29]	0.502	0.729
SiamFC[7]	0.505	0.706
SiamRPN[9]	0.535	0.751
本文算法	0.542	0.764

3 结 论

1）本文提出一种基于多注意力与双模板更新的视觉跟踪算法，在OTB2015、VOT2018和UAV123数据集上，该算法取得了良好的跟踪性能。

2）通过在VGG16骨干网络中引入SoftPool和MAM，获取丰富的通道和空间信息，增强了网络模型的特征表达能力。

3）采用双模板更新策略，增强模板的抗干扰能力，提高算法在目标遮挡、旋转等复杂场景下的跟踪性能。

在未来工作中，将进一步探索特征融合的方法，不断提升算法跟踪性能。

参考文献（References）

[1] DONG E Z, ZHANG Y, DU S Z. An automatic object detection and tracking method based on video surveillance[C]//Proceedings of the IEEE International Conference on Mechatronics and Automation. Piscataway: IEEE Press, 2020: 1140-1144.

[2] ZHAI L, WANG C P, HOU Y H, et al. MPC-based integrated control of trajectory tracking and handling stability for intelligent driving vehicle driven by four hub motor[J]. IEEE Transactions on Vehicular Technology, 2022, 71(3): 2668-2680.

[3] MANGAL N K, TIWARI A K. Kinect v2 tracked body joint smoothing for kinematic analysis in musculoskeletal disorders [C]//Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society. Piscataway: IEEE Press, 2020: 5769-5772.

[4] DEWANGAN D K, SAHU S P. Real time object tracking for intelligent vehicle[C]//Proceedings of the 1st International Conference on Power, Control and Computing Technologies. Piscataway: IEEE Press, 2020: 134-138.

[5] 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, 2015: 621-629.

[6] 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, 2015: 4310-4318.

[7] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional Siamese networks for object tracking[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2016: 850-865.

[8] ZHANG Z P, PENG H W. Deeper and wider Siamese networks for real-time visual tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4586-4595.

[9] 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.

[10] DANELLJAN M, BHAT G, KHAN F S, et al. ATOM: accurate tracking by overlap maximization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4655-4664.

[11] BHAT G, DANELLJAN M, VAN GOOL L, et al. Learning discriminative model prediction for tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2020: 6181-6190.

[12] WANG Q, TENG Z, XING J L, et al. Learning attentions: residual attentional Siamese network for high performance online visual tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4854-4863.

[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] ZHANG L C, GONZALEZ-GARCIA A, VAN DE WEIJER J, et al. Learning the model update for Siamese trackers[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 4009-4018.

[15] ZHANG C Y, WANG H, WEN J W, et al. Deeper Siamese network with stronger feature representation for visual tracking[J]. IEEE Access, 2020, 8: 119094-119104.

[16] STERGIOU A, POPPE R, KALLIATAKIS G. Refining activation downsampling with SoftPool[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 10337-10346.

[17] HUANG L H, ZHAO X, HUANG K Q. GOT-10k: a large high-diversity benchmark for generic object tracking in the wild[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5): 1562-1577.

[18] WU Y, LIM J, YANG M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.

[19] KRISTAN M, LEONARDIS A, MATAS J, et al. The sixth visual object tracking VOT2018 challenge results[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2019: 3-53.

[20] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for UAV tracking[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2016: 445-461.

[21] 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.

[22] 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.

[23] YANG Z X, ZHU L C, WU Y, et al. Gated channel transformation for visual recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11794-11803.

[24] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 13708-13717.

[25] ZHU Z, WANG Q, LI B, et al. Distractor-aware Siamese networks for visual object tracking[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 103-119.

[26] LI P X, CHEN B Y, OUYANG W L, et al. GradNet: gradient-guided network for visual object tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 6161-6170.

[27] LI Y H, ZHANG X F, CHEN D M. SiamVGG: visual tracking using deeper Siamese networks[EB/OL]. (2022-07-04)[2023-02-01]. https://arxiv.org/abs/1902.02804v4.

[28] 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.

[29] DANELLJAN M, ROBINSON A, SHAHBAZ KHAN F, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2016: 472-488.

[30] GUO Q, FENG W, ZHOU C, et al. Learning dynamic Siamese network for visual object tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 1781-1789.

[31] WANG N, ZHOU W G, TIAN Q, et al. Multi-cue correlation filters for robust visual tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4844-4853.

[32] HUANG Z Y, FU C H, LI Y M, et al. Learning aberrance repressed correlation filters for real-time UAV tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 2891-2900.

[33] LI Y M, FU C H, DING F Q, et al. AutoTrack: towards high-performance visual tracking for UAV with automatic spatio-temporal regularization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11920-11929.

[34] LI F, TIAN C, ZUO W M, et al. Learning spatial-temporal regularized correlation filters for visual tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4904-4913.

Visual tracking algorithm based on multi-attention and dual-template update

MA Sugang[1, 2, *], SUN Siwei[1], HOU Zhiqiang[1], YU Wangsheng[3], PU Lei[4]

(1. School of Computer Science & Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China;
2. Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an 710121, China;
3. School of Information and Navigation, Air Force Engineering University, Xi'an 710077, China;
4. School of Operational Support, Rocket Force Engineering University, Xi'an 710025, China)

Abstract: To address the problem of insufficient representational capability and lack of online update of the fully-convolutional Siamese network (SiamFC) tracker in complex scenes, this paper proposed a visual tracking algorithm based on multi-attention and dual-template update. First, the feature extraction network was constructed by replacing AlexNet with the VGG16 network, and SoftPool was used to replace the maximum pooling layer. Secondly, the multi-attention module (MAM) was added after the backbone network to enhance the network's ability to extract object features. Finally, a dual template was designed for feature fusion and response map fusion, and average peak-to-correlation energy (APCE) was used to determine whether to update the dynamic templates, which effectively improved the tracking robustness. The proposed algorithm was trained on the GOT-10k dataset and tested on the OTB2015, VOT2018, and UAV123 datasets. The experimental results show that, compared with the benchmark algorithm SiamFC, the tracking success rate of the proposed algorithm is increased by 0.085 and 0.037, and the accuracy is increased by 0.118 and 0.058 on the OTB2015 and UAV123 datasets, respectively. On the VOT2018 dataset, the tracking accuracy, robustness, and expected average overlap (EAO) are improved respectively by 0.030, 0.295 and 0.139. The proposed algorithm achieves high tracking accuracy in complex scenes, and the tracking speed reaches 33.9 frame/s, which meets the real-time tracking requirements.

Keywords: visual object tracking; Siamese network; SoftPool; attention mechanism; feature fusion

Received: 2023-06-09; **Accepted**: 2023-07-21; **Published Online**: 2023-08-25 14:30

URL: link.cnki.net/urlid/11.2625.V.20230824.1905.002

Foundation items: National Natural Science Foundation of China (62072370); Natural Science Foundation of Shaanxi Province (2023-JC-YB-598); Science and Technology Project of Xi'an (22GXFW0125)

* **Corresponding author**. E-mail: msg@xupt.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0335

EHL下考虑冲击力的局部缺陷角接触球轴承动力学建模

雷春丽[1]，薛伟[1,*]，樊高峰[2]，宋瑞哲[1]，刘凯[1]

(1. 兰州理工大学 机电工程学院，兰州 730050；2. 云南文山铝业有限公司，文山 663000)

摘 要：为详细地分析具有局部缺陷的角接触球轴承(ACBBs)的运行状态，针对传统模型未考虑轴承润滑及缺陷引起的冲击力问题，建立ACBBs在弹流润滑(EHL)条件下考虑冲击力的局部缺陷ACBBs动力学模型。建立考虑自旋的缺陷轴承弹流润滑模型，计算出缺陷轴承的油膜厚度及油膜刚度，再提出ACBBs局部缺陷时变位移激励模型和时变刚度模型，在此基础上建立与缺陷尺寸及轴承转速有关的瞬时冲击力函数；基于Hertz接触理论和冲击力函数，提出外圈具有局部缺陷ACBBs动力学计算方法；研究具有故障的角接触球轴承振动特性，并通过实验验证，分析了不同参数对轴承动力学响应的影响规律。计算结果表明：轴承在润滑条件下，冲击力对轴承的振动影响明显减弱；随着缺陷尺寸和载荷的增大，轴承故障特征的频率均不变，但其幅值增大；随着轴承转速的增大，轴承故障特征频率及其幅值均增大，为滚动轴承的故障诊断和维护提供了有力依据。

关 键 词：角接触球轴承；局部缺陷；瞬时冲击力；弹流润滑；时变位移；动力学建模

中图分类号：TH133.3

文献标志码：A **文章编号**：1001-5965(2025)06-1965-13

轴承作为机械设备的旋转支承核心部件，其动态特性直接影响着轴承-转子系统乃至整机装备的运行状态和服役寿命[1]。由于角接触球轴承可以承受变化的轴向力和径向力，在长期运行的情况下容易产生裂纹、凹坑、胶合等局部缺陷，使得轴承各部件之间的接触变形和受力关系变得更为复杂[2]。轴承滚动体经过缺陷区域接触时，会产生冲击激励，引起系统谐振，导致振动增大，缩短轴承的使用寿命，而弹流润滑作用能够有效地减少轴承各部间之间的磨损，使其运行得更加平稳、可靠，在弹流润滑作用下对含有局部缺陷的角接触球轴承进行动力学建模，研究轴承在特定工况和故障下所产生的时变位移激励，并对其动响应进行分析具有重要的现实意义。

目前，对于轴承弹流润滑的研究及其故障动力学建模，国内外学者做了大量的工作。在弹流润滑方面，杨沛然[3]将弹流润滑理论和数值分析结合，给出滚动轴承在多种接触情况下弹流数值解计算式；王鹏等[4]通过数值求解研究表面凸起和凹坑缺陷对点接触等温稳态弹流润滑油膜厚度和压力的影响，并讨论了缺陷位置和尺寸的影响；雷春丽等[5]通过建立考虑自旋运动的弹流润滑的角接触球轴承模型，分析了在不同接触压力、滚道表面粗糙度下自旋运动对角接触球轴承弹流润滑和油膜刚度的影响；路遵友等[6]运用热弹性流体动压润滑理论，考虑了润滑油膜温升变化引起的圆柱滚子轴承中滚子和内圈接触表面的径向热弹性变形和表面粗糙度的影响，提出一种计入热弹性变形和粗糙度影响的圆柱滚子轴承线接触热弹性流体动压润滑分析方法。Wijnant等[7]提出弹性流体动力学简化

收稿日期：2023-06-09；**录用日期**：2023-07-21；**网络出版时间**：2023-08-24 15:36
网络出版地址：link.cnki.net/urlid/11.2625.V.20230823.1459.003
基金项目：国家自然科学基金(51465035)
*通信作者. E-mail：2542908749@qq.com

引用格式：雷春丽，薛伟，樊高峰，等. EHL下考虑冲击力的局部缺陷角接触球轴承动力学建模[J]. 北京航空航天大学学报，2025，51(6)：1965-1977. LEI C L，XUE W，FAN G F，et al. Dynamic modelling of angular contact ball bearings with local defects under EHL considering impact force[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1965-1977 (in Chinese).

模型，该模型适合研究不同机械系统的弹性流体动力学。Yan等[8]基于等温弹流润滑理论，建立了含局部缺陷滚动轴承的五自由度动力学模型，研究了滚动轴承在不同转速和载荷下油膜刚度、综合刚度和振动特性的变化规律；在轴承动力学建模方面，曹宏瑞等[9]利用冲击脉冲描述滚动体在经过轴承局部损伤时产生的激励力，建立了含有损伤产生激励力的轴承-转子系统动力学模型；罗茂林等[10]基于Hertz接触理论，分析了滚动体与内圈剥落区之间时变接触刚度特性，建立了时变位移激励的角接触球轴承局部故障动力学模型；剡昌锋等[11]建立了滚动轴承在弹流润滑及滑动作用下局部缺陷轴承位移激励的二自由度动力学模型，分析了滚动轴承局部缺陷动态响应特性；李志农等[12]探索了滚动体故障经过内、外圈时变形量的释放变化规律，给出了渐变释放的变形量的计算方法，建立了含滚动体点蚀故障的滚动轴承二自由度动力学渐变模型；黄文涛等[13]在传统动力学模型基础上考虑滚动体退出缺陷区域时的撞击力，搭建了三自由度滚动轴承故障振动模型，模拟了内滚道缺陷的振动响应；崔玲丽等[14]为解决轴承损伤尺寸难以诊断的问题，建立滚动轴承非线性振动的动力学模型，通过解析法求解轴承4种不同大小的故障振动信号响应，对轴承故障的双冲击现象进行了定量分析，胡爱军等[15]提出时变接触变形与时变刚度耦合的滚动轴承故障非线性六自由度动力学模型，结果显示不同缺陷尺寸下，滚动体进入和离开缺陷过程中产生的冲击响应明显不同；Cui等[16]提出一个滚动轴承五自由度振动模型，对比分析了不同外圈缺陷尺寸下的仿真信号与实验信号；随后，Cui等[17]考虑了缺陷的不同位置和尺寸，建立了一种用于轴承外圈故障的定量估计模型；Zhang等[18]建立了考虑轴承径向游隙和时变位移激励的二自由度滚动轴承内圈局部缺陷故障动力学模型，有效地模拟了故障信号特征参数随缺陷宽度、径向载荷和转速的变化情况；Wen等[19]建立了完整的角接触球轴承多自由度动力学模型，分析了不同尺寸和位置角度的外滚道缺陷对轴承动态行为的影响；Niu等[20]建立了滚道表面局部缺陷的滚动球轴承动力学模型，对其振动响应进行了时域和频域仿真分析。

上述研究对故障轴承动力学建模提供了坚实的理论基础，但其未考虑角接触球轴承在弹流润滑作用下局部缺陷与冲击力之间的关系及冲击力对轴承振动响应的影响。为使模型更符合轴承工作的实际环境，本文以含有外圈局部缺陷的角接触球轴承为研究对象，采用等温点接触弹流润滑理论，结合角接触球轴承的自旋特点，建立了考虑自旋的故障轴承弹流润滑模型，进而求得轴承的油膜厚度与油膜刚度，通过联立滚动体与内外圈的接触变形、滚动体经过缺陷时的时变位移和接触刚度，得到轴承的时变位移模型和时变刚度模型；依据滚动体经过缺陷区域时变位移激励函数，推导出滚动体经过缺陷区域时的瞬时冲击力激励函数；基于Hertz接触理论，建立在弹流润滑作用下考虑冲击力的角接触球轴承动力学模型，研究润滑作用及不同因素对角接触球轴承振动响应的影响规律。

1 考虑自旋的局部缺陷角接触球轴承弹流润滑建模

1.1 角接触球轴承自旋运动模型

角接触球轴承中的运动学关系如图1所示，由图示关系可知，接触椭圆法线不垂直于滚动体自转轴线，从而在接触法线方向产生相应的角速度分量，使得滚动体与内、外圈之间沿接触法线方向存在相对旋转运动，其称为自旋运动，ω_{si}、ω_{so}分别为滚动体在内、外圈上的自旋角速度。

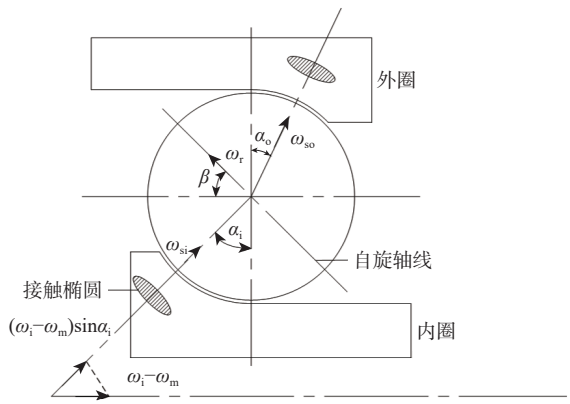


图1 角接触球轴承中的自旋运动

Fig. 1 Spin motion in angular contact ball bearings

由几何关系可得自旋角速度的具体表达为

$$\begin{cases}\omega_{si}=\omega_r\sin(\alpha_i-\beta)-(\omega_i-\omega_m)\sin\alpha_i\\ \omega_{so}=\omega_r\sin(\alpha_o-\beta)-\omega_m\sin\alpha_o\end{cases} \tag{1}$$

式中：ω_i为轴承内圈角速度；ω_m为滚动体公转角速度；ω_r为滚动体自转角速度；α_i、α_o为轴承内、外圈接触角；β为姿态角。

1.2 含局部缺陷角接触球轴承弹流润滑模型

结合角接触球轴承自旋运动的典型特点，将滚动体与内、外圈接触的自旋弹流润滑副等效为无限大平面与椭球的接触。椭球在无限大平面上以滚动体和内、外圈滚道的平均速度u_b运动，如图2所示。同时，平面以ω_s的自旋角速度旋转，接触区任意点速度分量分别为v_R、u_R。图2中任意点处v_R

和 u_R 的表达式分别为

$$\begin{cases} u_R(x,y)=\omega_s y/2+u_b \\ v_R(x,y)=-\omega_s x/2 \end{cases} \tag{2}$$

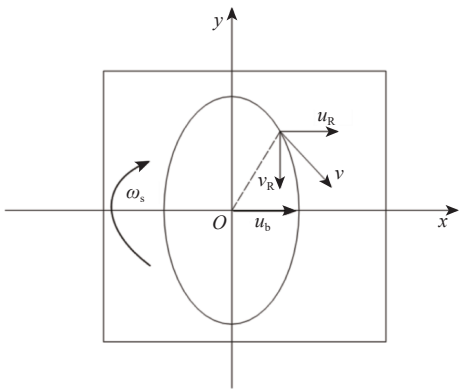


图 2　自旋运动等效模型

Fig. 2　Equivalent model of spin motion

根据本文角接触球轴承点接触自旋运动模型，选用等温点接触弹流润滑 Reyonds 方程，其基本形式如下：

$$\frac{\partial}{\partial x}\left(\frac{\rho h^3}{\eta}\cdot\frac{\partial p}{\partial x}\right)+\frac{\partial}{\partial y}\left(\frac{\rho h^3}{\eta}\cdot\frac{\partial p}{\partial y}\right)= 2\frac{\partial}{\partial x}(u_R\rho h)+12\frac{\partial}{\partial x}(v_R\rho h) \tag{3}$$

式中：ρ为润滑油密度；p 为润滑油压力分布；y 为润滑油的动力黏度。

其边界条件为：

入口区：$P(X_0, Y)=0$；

出口区：$P(X_e, Y)=0$，$\dfrac{\partial P(X_e,Y)}{\partial X}=0$；

端部：$P|_{Y=\pm 1}=0$。

本文假设缺陷在轴承外圈上是光滑过渡的，将轴承表面局部缺陷视为单点深凹坑粗糙度函数，其特征记为 $S(x,y)$，函数表达式为

$$S(x,y)=\begin{cases} -\dfrac{A}{2}\left[1+\cos(\pi\sqrt{x^2+y^2}/R_d)\right] \\ x^2+y^2\leqslant {R_d}^2 \end{cases} \tag{4}$$

式中：A 和 R_d 分别为局部缺陷的深度和宽度。

考虑凹坑的油膜厚度计算式为

$$h(x,y)=h_0+\frac{x^2}{2R_x}+\frac{y^2}{2R_y}+ \frac{2}{\pi E}\iint\frac{p(s,t)\,\mathrm{d}s\mathrm{d}t}{\sqrt{(x-s)^2+(y-t)^2}}-S(x,y) \tag{5}$$

式中：h_0 为滚动体与滚道的初始间隙；R_x、R_y 分别为等效椭球体沿 x 轴、y 轴方向的当量曲率半径；E 为材料的弹性模量值；s 和 t 对应 x 轴和 y 轴的附加坐标。

黏-压方程[21] 为

$$\eta=\eta_0\exp\left\{(\ln\eta_0+9.67)\left[(1+p/p_0)^z-1\right]\right\} \tag{6}$$

$$z=\frac{\alpha}{5.1\times10^{-9}(\ln\eta_0+9.67)} \tag{7}$$

式中：η_0 为润滑油大气压下的动力黏度；α 为 Barus 黏压系数。

密-压方程[22] 为

$$\rho=\rho_0\left(1+\frac{0.6p_0}{1+1.7p_0}\right) \tag{8}$$

式中：p_0 为标准大气压强；ρ_0 为标准大气压下润滑油密度。

当润滑油将两接触体分离后，润滑油受外载荷影响，外载荷平衡方程[23] 为

$$\int_{x_0}^{x_e}\int_{y_0}^{y_e}p\,\mathrm{d}x\mathrm{d}y=w \tag{9}$$

式中：w 为单位长度上的法向外载荷。

式 (5)～式 (9) 的参数取值范围如表 1 所示。

表 1　弹流润滑参数

Table 1　EHL parameter values

动力黏度 η/(Pa·s)	润滑油密度 ρ_0/(kg·m^{-3})	黏压系数 α/(m^2·N^{-1})	环境温度/℃	标准大气压 p_0/Pa
0.05	0.91	1.2×10^{-8}	25	1.01×10^5

1.3　油膜刚度模型

根据刚度定义，轴承油膜刚度的基本表达式如下：

$$K_f=\lim_{\delta\to0}\frac{\Delta w}{\delta}=-\frac{\mathrm{d}w}{\mathrm{d}h_{min}} \tag{10}$$

根据式 (10) 和最小油膜厚度 h_{min} 推导可得

$$K_f=-\frac{\mathrm{d}w}{\mathrm{d}h_{min}}=0.162\,19ER_x^3/\left[0.107\,7\times10^3R_x\,h_{min}- 0.958\,8G^{0.800\,1}U^{0.72}k^{1.616}\left(1+21.459\mathrm{e}^{-0.3k}\right)\cdot \left(1-1.310\,1\times10^{10}\ \Omega_s^{1.024}\right)\right] \tag{11}$$

式中：U 为速度参数；G 为材料参数；k 为椭圆率；Ω_S 自旋速度参数；K_f 为考虑自旋运动时的油膜刚度[24]。式(11)各参数取值范围如表 2 所示，表中，

表 2　各参数取值范围

Table 2　Range of values for each parameter

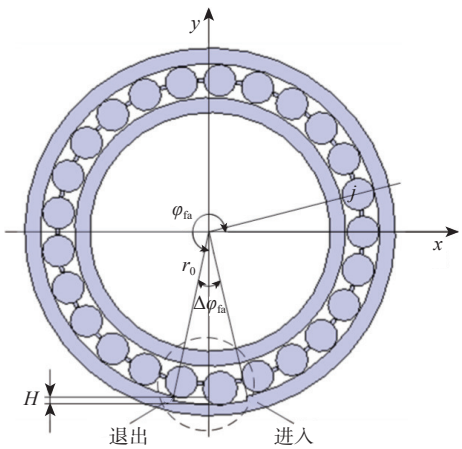
参数	取值
k	1～9.5
u_b/(m·s^{-1})	1～40
w/N	200～10^4
G	2 500～5 000
W	3.23×10^{-4}～1.94×10^{-3}
U	1.68×10^{-11}～6.74×10^{-10}
Ω_s	0～4.41×10^{-11}

u_b 为滚动体和内外圈滚道的平均速度；W 为载荷参数。

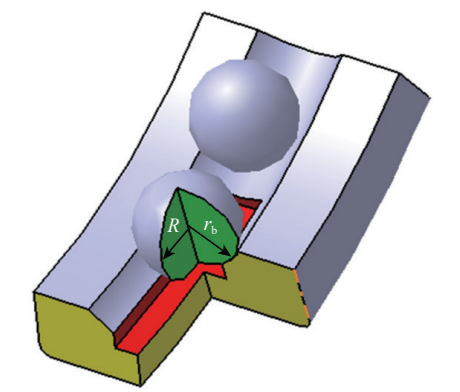
2 时变位移模型

2.1 矩形缺陷时变位移

轴承外圈局部缺陷轮廓等效为矩形形状，如图 3 所示。图 3(a) 中 φ_{fa} 为缺陷中心所在的方位角，$\Delta\varphi_{fa}$ 为缺陷的环向跨度角，H 为缺陷深度，r_0 为轴承外滚道半径；图 3(b) 中 R 为滚动体半径，r_b 为外沟道曲率半径。



(a) 整体模型



(b) 残缺部分

图 3 轴承局部缺陷轮廓描述

Fig. 3 Outline description of local defects in bearings

在实际工况中，滚动体进入和退出缺陷区域时，滚动体与滚道的接触缺陷深度是变化的，为更加准确地描述这个过程，需要计算出滚动体与滚道的接触缺陷深度。

当保持架位于初始位置时，第 j 个滚动体的方位角为

$$\varphi_j=\omega_c t+\frac{2\pi(j-1)}{N}\qquad j=1,2,\cdots,N \tag{12}$$

式中：ω_c 为保持架转速；N 为滚动体个数。

图 3 (a) 中滚动体与缺陷接触深度可表示为

$$d(\varphi)=\begin{cases}\min\{H,R-\sqrt{R^2-0.25r_o^2(\varphi_j-\varphi_{fa}+0.5\Delta\varphi_{fa})^2}\} & \varphi_{fa}-\dfrac{1}{2}\Delta\varphi_{fa}<\varphi_j<\varphi_{fa}\\ \min\{H,R-\sqrt{R^2-0.25r_o^2(\varphi_{fa}-\varphi_j+0.5\Delta\varphi_{fa})^2}\} & \varphi_{fa}<\varphi_j<\varphi_{fa}+\dfrac{1}{2}\Delta\varphi_{fa}\\ 0 & \varphi_{fa}+\dfrac{1}{2}\Delta\varphi_{fa}<\varphi_j<\varphi_{fa}-\dfrac{1}{2}\Delta\varphi_{fa}\end{cases} \tag{13}$$

设矩形缺陷时变位移为 H_d，由此可得矩形缺陷的时变位移表达式为

$$H_d=d(\varphi) \tag{14}$$

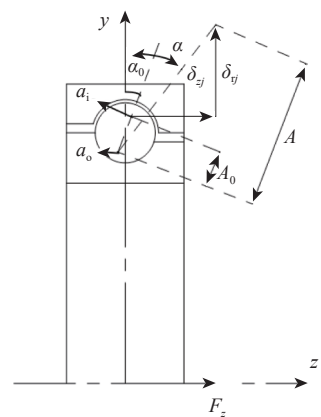
2.2 油膜变化时变位移

轴承在受到径向和轴向载荷时，轴承会产生轴向和径向接触变形，如图 4 所示，通过图中的几何关系可得接触变形为

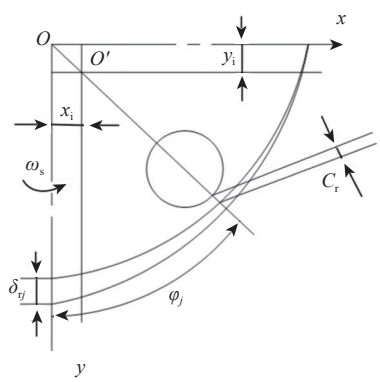
$$\delta_j=A-A_0 \tag{15}$$

$$A=\left[(A_0\cos\alpha_0+\delta_{rj})^2+(A_0\sin\alpha_0+\delta_{zj})^2\right]^{0.5} \tag{16}$$

式中：A 和 A_0 分别为内滚道槽曲率中心 a_i 和外滚道槽曲率中心 a_o 之间的加载和卸载相对距离；δ_{rj} 为轴承径向接触变形；δ_{zj} 为轴承轴向接触变形。



(a) 整体接触变形



(b) 径向接触变形

图 4 轴承的接触变形

Fig. 4 Contact deformation of bearings

如图 4 (b) 所示，若在 φ_j 角度时第 j 滚动体进入滚道的矩形缺陷，则轴承的径向接触变形 δ_{rj} 和轴向接触变形 δ_{zj} 分别可表示为

$$\delta_{rj}=(x_1-x_2)\cos\varphi_j+(y_1-y_2)\sin\varphi_j-C_r-h(x,y)-H_d\cos\alpha_0 \tag{17}$$

$$\delta_{zj}=C_d+z_1-H_d\sin\alpha_0 \tag{18}$$

式中：C_r 和 C_d 分别为轴承的径向游隙和轴向游隙；z_1 为轴承的轴向初始位移。

综合式 (15)～式 (18)，可计算出 δ_j 为

$$\delta_j=\{[A_0\cos\alpha_0+(x_1-x_2)\cos\varphi_j+(y_1-y_2)\sin\varphi_j-0.5C_r-h(x,y)-H_d\cos\alpha_0]^2+[A_0\sin\alpha_0+C_d+z_1-H_d\sin\alpha_0]^2\}^{1/2}-A_0 \tag{19}$$

3　时变刚度模型

通过 Hertz 接触理论可得滚动体与滚道的接触力为

$$F=K\delta_j^n \tag{20}$$

式中：K 为滚动体与内、外圈等效接触刚度；δ_j 表示接触变形；n 为载荷-变形指数，角接触球轴承取 n=1.5，当轴承运行时，滚动体与滚道会产生的接触刚度[25] 为

$$\begin{cases}K_i=4\sqrt{2}\left(\dfrac{E}{1-\mu^2}\right)\left(\dfrac{1}{\delta_i}\right)^{3/2}\Big/3\sqrt{\sum\rho_{in}}\\ K_o=4\sqrt{2}\left(\dfrac{E}{1-\mu^2}\right)\left(\dfrac{1}{\delta_o}\right)^{3/2}\Big/3\sqrt{\sum\rho_{ou}}\end{cases} \tag{21}$$

式中：K_i、K_o 分别为滚动体与内、外滚道之间的接触刚度；$\sum\rho_{in}$ 和 δ_i 分别为内滚道的曲率和和接触位移；$\sum\rho_{ou}$ 和 δ_o 分别为外滚道的曲率和接触位移；μ 为材料的泊松比值。

角接触球轴承的刚度是分析轴承在运行过程中振动特性的基础，在润滑油充足的情况下，角接触球轴承中的内、外滚道被润滑油膜支撑，油膜在外力的作用下出现油膜刚度。根据 1.2 节建立的弹流润滑模型，可以求出滚动体和滚道之间的油膜刚度。角接触球轴承的内、外圈刚度由两部分组成，一部分是赫兹接触刚度 K_i/K_o，另一部分是润滑油膜刚度 K_{fi}/K_{fo}，将 2 部分刚度串联得到滚动体和内、外滚道之间的综合刚度 K_{ii}/K_{oo}：

$$\begin{cases}K_{ii}=(K_i^{-1}+K_{fi}^{-1})^{-1}\\ K_{oo}=(K_o^{-1}+K_{fo}^{-1})^{-1}\end{cases} \tag{22}$$

根据轴承综合刚度计算式，第 j 个滚动体与内、外圈在径向和轴向上的综合刚度分量 K_{rij}、K_{roj}、K_{aij}、K_{aoj} 分别为

$$\begin{cases}K_{rij}=K_{ii}\cos^2\alpha_{ij}\\ K_{roj}=K_{oo}\cos^2\alpha_{oj}\\ K_{aij}=K_{ii}\sin^2\alpha_{ij}\\ K_{aoj}=K_{oo}\sin^2\alpha_{oj}\end{cases} \tag{23}$$

轴承的综合径向和轴向刚度分别为

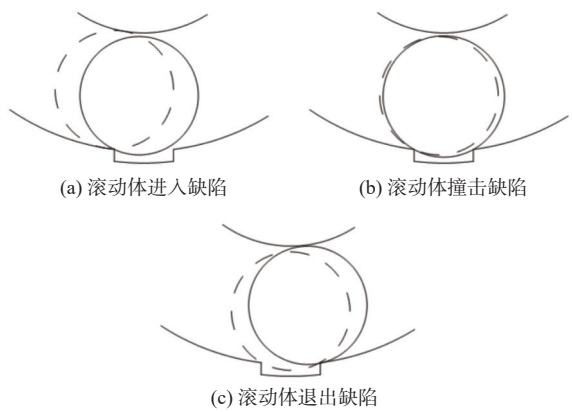
$$K_r=\sum_{j=1}^{j=Z}\frac{K_{rij}K_{roj}}{K_{rij}+K_{roj}}\cos^2\left[\frac{2\pi}{Z}(j-1)\right] \tag{24}$$

$$K_a=\sum_{j=1}^{j=Z}\frac{K_{aij}K_{aoj}}{K_{aij}+K_{aoj}} \tag{25}$$

4　考虑瞬时冲击力的角接触球轴承动力学模型

4.1　瞬时冲击力计算方法

滚动体在经过缺陷区域时会存在一种瞬时的冲击过程，这个冲击过程可以划分为 3 个阶段，具体描述过程如图 5 所示。



(a) 滚动体进入缺陷　(b) 滚动体撞击缺陷　(c) 滚动体退出缺陷

图 5　滚动体经过缺陷的冲击过程

Fig. 5　Impact process of rolling element passing through defects

冲击过程的第 2 阶段如图 6 所示，此时滚动体中心与缺陷中心在垂直的位置重合，滚动体与缺陷边沿部分发生碰撞，产生附加的冲击力作用。本文基于第 2 节矩形缺陷时变位移函数，考虑故障尺寸及轴承转速对冲击力的影响，从而提出一种改进的瞬时冲击力激励函数[26]，详细推导过程如下。滚动体在进入缺陷区域后存在切向和径向的速度分别为 v_h 和 v_n，v_n 则用矩形缺陷时变位移激励函数的一阶导数来表示：

$$v_n=\frac{0.25r_o^2[(\varphi_j-\varphi_{fa})+0.5\Delta\varphi_{fa}]\omega_c}{\sqrt{R^2-0.25r_o^2[(\varphi_j-\varphi_{fa})]^2}} \tag{26}$$

本文的冲击力使用冲量定理进行计算，在径向方向上对滚动体运用冲量定理：

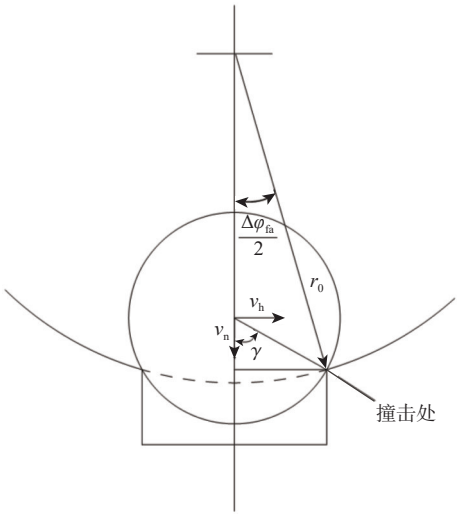


图 6 滚动体与故障边缘碰撞简化图

Fig. 6 Simplified diagram of collision between rolling element and fault edge

$$F_{imp}\Delta t\cos\gamma = m_b v_n \tag{27}$$

式中：F_{imp} 为冲击力；Δt 为冲击力作用的时间；m_b 为滚动体质量；γ 为冲击力方向与径向的夹角。

冲击力作用的时间为滚动体从接触缺陷后边沿直至离开缺陷区域所经历的时间：

$$\Delta t = \frac{\Delta\varphi_{fa}}{2\omega_c} \tag{28}$$

由几何关系可知：

$$\cos\gamma = \sqrt{R^2-(L/2)^2}\Big/R \tag{29}$$

由式 (27) 可得瞬时冲击力表达式为

$$F_{imp} = \frac{m_b v_n}{\Delta t\cos\gamma} \tag{30}$$

瞬时冲击力可化简为

$$F_{imp} = \frac{0.5r_o^2(\varphi_j-\varphi_{fa}+0.5\Delta\varphi_{fa})\omega_c^2}{\sqrt{R^2-0.25r_o^2(\varphi_j-\varphi_{fa}+0.5\Delta\varphi_{fa})^2}}\cdot \frac{m_b R}{\Delta\varphi_{fa}\sqrt{R^2-(L/2)^2}} \tag{31}$$

4.2 动力学模型的建立

为简化研究，本文对研究对象做出如下假设：

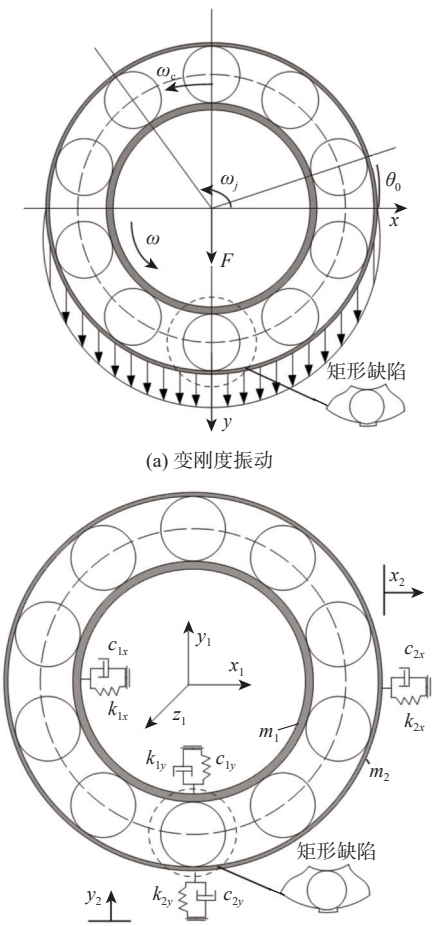
1）轴承的故障为外滚道单一局部剥落，剥落形式为与滚道等宽的矩形凹槽，且凹槽具有一定深度；

2）滚动体均匀分布在滚道中且互不接触，忽略滚动体奇偶数目对轴承内部载荷分布的影响；

3）滚动体与内、外滚道接触满足 Hertz 接触理论；

4）轴承的外圈与轴承座、内圈与转轴皆为刚性连接。

本文研究针对轴承运转过程中产生的早期缺陷，故障尺寸不足以导致轴承失效；在上述条件下，将角接触球轴承系统简化为弹簧-阻尼模型，如图 7 所示。图 7(a)中 F 为轴承所受的外力分量，φ_j 为第 j 个滚动体与内外圈的角位置，ω 和 ω_c 分别为轴承转轴的转速和保持架的转速；图 7(b)中 m_1、m_2、x_1、x_2、y_1、y_2 分别为轴承内圈和外圈的质量、径向水平绝对位移及径向垂直绝对位移；z_1 为内圈的轴向绝对位移；k_{1x}、k_{2x}、k_{1y}、k_{2y}、c_{1x}、c_{2x}、c_{1y}、c_{2y} 分别为轴承内圈和外圈在径向水平方向、垂直方向的刚度和阻尼。设 k_z 和 c_z 分别为轴承轴向刚度和阻尼；f_x、f_y 和 f_z 分别为轴承非线性支承反力在径向水平、垂直方向和轴向的分量；F_x、F_y 和 F_z 分别为作用在转子上的径向水平力、垂直力和轴向力在轴承处的外力分量。



(a) 变刚度振动

(b) 动力学模型

图 7 角接触球轴承动力学模型

Fig. 7 Dynamic model of angular contact ball bearing

结合式 (27)，提出考虑冲击力的角接触球轴承五自由度简化动力学模型：

$$\begin{cases} m_1\ddot{x}_1 + c_{1x}\dot{x}_1 + k_{1x}x_1 + f_x + F_{imp}\sin\gamma = F_x \\ m_1\ddot{y}_1 + c_{1y}\dot{y}_1 + k_{1y}y_1 + f_y + F_{imp}\cos\gamma = F_y \\ m_1\ddot{z}_1 + c_z\dot{z}_1 + k_z z_1 + f_z = F_z \\ m_2\ddot{x}_2 + c_{2x}\dot{x}_2 + k_{2x}x_2 - f_x = 0 \\ m_2\ddot{y}_2 + c_{2y}\dot{y}_2 + k_{2y}y_2 - f_y = 0 \end{cases} \tag{32}$$

5　计算结果分析

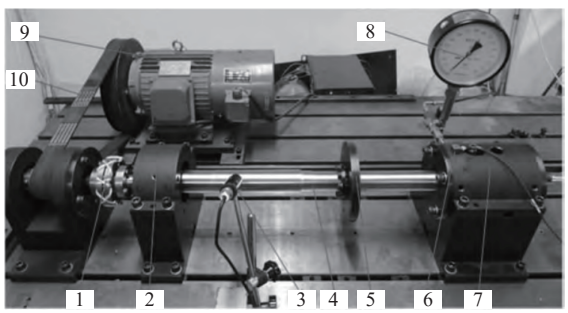
5.1　模型验证

为证明本文建立的角接触球轴承动力学模型的正确性, 采用文献 [27] 的实验结果来验证。文献 [27] 中角接触球轴承的型号为 HRB7206AC, 轴承参数如表 3 所示。文献 [27] 为模拟缺陷对轴承的影响, 采用电火花加工方法在角接触轴承的滚道表面上人为加工一个微小的凹坑, 凹坑的宽度为 0.3 mm, 轴承的预紧力为 600 N, 轴的转速为 1 200 r/min。图 8 为文献 [27] 的实验台。

表 3　HRB7206AC 轴承基本参数[27]

Table 3　Basic parameters of HRB7206AC bearing[27]

参数	数值
内圈直径d_i/mm	30
外圈直径d_o/mm	62
滚动体直径d_b/mm	9.525
滚动体个数Z/个	12
轴承宽度D/mm	16
初始接触角α_0/(°)	25
弹性模量E/Pa	2.1×10^{11}
泊松比	0.3



1. 柔性联轴器; 2. 左侧轴承座; 3. 转速传感器; 4. 主轴; 5. 质量盘; 6. 加速度传感器; 7. 右侧轴承座; 8. 预紧力显示用液压表; 9. 驱动电机; 10. 传动带。

图 8　实验台[27]

Fig. 8　Experiment platform[27]

为获得该轴承的动态特性, 将表 1 中轴承各参数代入本文模型中, 采用 4 阶 Runge-kutta 法对式(32)进行求解。选取步长Δt=1×10^{-4} s, x 和 y 方向位移初始值分别为 x_0=10^{-6} mm 和 y_0=10^{-6} mm。计算可得轴承动态特性频谱图如图 9 所示。

计算结果与文献 [27] 中轴承动态特性的时域图、频谱图进行对比, 该轴承的外圈故障特征频率 f_o=97.48 Hz, 本文模型得到的外圈缺陷轴承故障特征频率 f_o=95.21 Hz, 误差为 2.3%。采用本文模型得到的振动响应频谱图外圈故障特征频率 f_o 及其倍

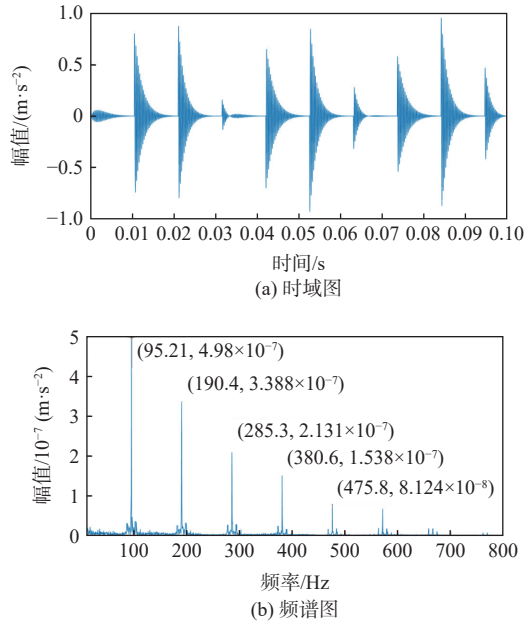


图 9　本文模型得到的文献 [27] 缺陷轴承外圈故障振动响应

Fig. 9　Outer ring fault vibration response of [27] defective bearing in literature obtained by the proposed model

频清晰可见, 与文献 [27] 结果吻合。由此验证了本文外圈缺陷轴承动力学模型的准确性。

5.2　弹流润滑作用下不同工况参数对角接触球轴承振动特征的影响

假设轴承的外圈矩形缺陷深度 H=0.5 mm, 缺陷跨度角$\Delta\varphi_{fa}$ 分别为 0.6°、0.8°和 1°, 根据本文矩形缺陷的时变位移计算公式, 可得滚动体经过矩形缺陷的时变位移如图 10 所示, 为直观表示流动体陷入矩形缺陷时的位移, 纵轴数值由大到小排列。可以看出滚动体经过缺陷时, 位移变化是一条曲线, 证明滚动体不是直接接触底部, 滚动体的运动位移是一个变化的过程, 而且随着缺陷跨度角$\Delta\varphi_{fa}$ 的增大而增大。

图 11 为瞬时冲击力与缺陷角度和转速之间的

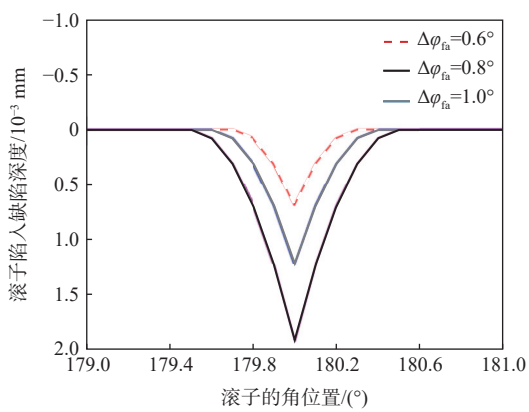
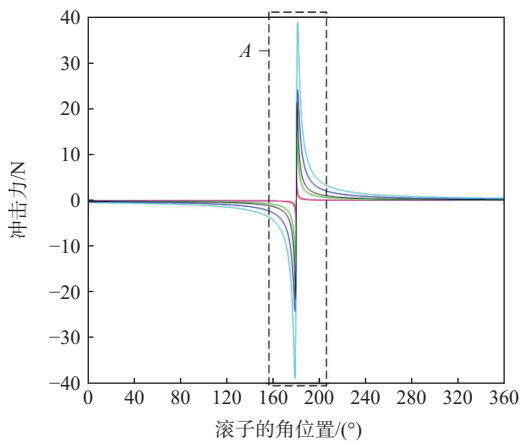


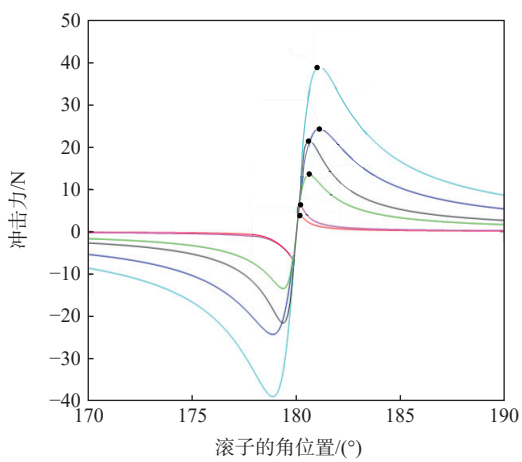
图 10　滚动体陷入矩形缺陷时的位移

Fig. 10　Displacement of rolling element when trapped in a rectangular defect

关系。由图 11 (a) 可知，当滚动体进入外圈缺陷时，缺陷区域前边缘的接触变形迅速变小，滚动体与外圈之间的接触力也迅速减小；当滚动体运动至外圈缺陷区域中央位置时，缺陷处的接触变形减小至最小，同时滚动体与外圈之间的接触力减小至最小值；随后，滚动体在保持架的带动下以一定的速度与外圈缺陷区域后边缘发生剧烈的碰撞，对外圈产生了冲击作用，诱发冲击力，这时的冲击力从其最小值快速增大到最大值；接着，滚动体在保持架的带动下退出缺陷区域，滚动体的冲击力逐渐减小，直到滚动体退出缺陷区时恢复到其正常的接触状况。



(a) 瞬时冲击力



(b) *A*处瞬时冲击力放大图

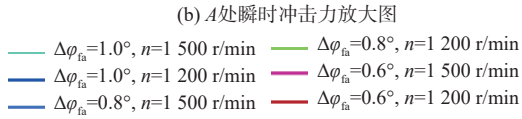


图 11 瞬时冲击力与缺陷角度和转速之间的关系

Fig. 11 Relationship among instantaneous impact force, defect angle, and rotational speed

由图 11 (b) 和表 4 可知，当转速一定时，随着缺陷尺寸的增大，冲击力增大；当缺陷尺寸一定时，随着转速的增大，冲击力也增大。此外，比较冲击力随缺陷跨度角和轴承转速的变化幅度，可以看出轴承转速对冲击力的影响大于缺陷跨度角。

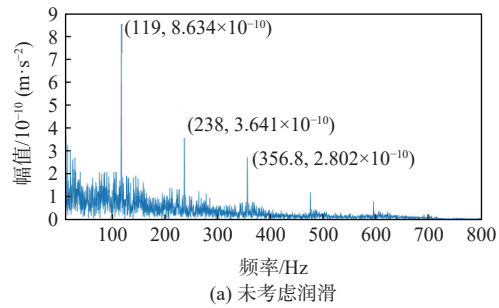
表 4 转速、缺陷角度对冲击力的影响

Table 4 Effect of rotational speed and defect angle on impact force

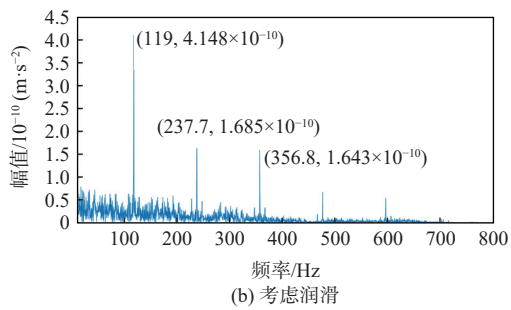
转速n/(r·min^{-1})	缺陷角度$\Delta\varphi_{fa}$/(°)	冲击力F_{imp}/N
1 200	0.6	3.917
1 200	0.8	6.284
1 200	1	13.44
1 500	0.6	21.56
1 500	0.8	24.25
1 500	1	38.67

5.2.1 润滑和缺陷跨度角的影响

本节分析弹流润滑作用和缺陷跨度角对故障轴承振动响应的影响规律，设置轴的转速 n=1 500 r/min，载荷 F_x=1 000 N、F_y=0、F_z=600 N，外圈缺陷跨度角$\Delta\varphi_{fa}$分别为 0.6°、0.8°和 1°。计算结果如图 12～图 14 所示。



(a) 未考虑润滑



(b) 考虑润滑

图 12 $\Delta\varphi_{fa}$=0.6°时外圈的振动响应

Fig. 12 Vibration response of outer ring when $\Delta\varphi_{fa}$ = 0.6°

由图 12 可知，当缺陷跨度角$\Delta\varphi_{fa}$=0.6°时，轴承在考虑润滑作用下，外圈故障振动响应的主频率不受影响，但其振动响应的幅值明显小于未考虑润滑时的幅值；图 13 和图 14 分别为$\Delta\varphi_{fa}$=0.8°、$\Delta\varphi_{fa}$=1°时外圈的振动响应，可得出与图 12 相同的规律。由图 12～图 14 可知，随着轴承缺陷跨度角的增大，外圈故障振动响应频率值不变，但其幅值增大；同时外圈故障振动响应的幅值在考虑润滑作用时均较小，对比变小的幅度可以得出，轴承外圈缺陷较大时润滑减振作用尤其明显。因为幅值的大小反映了振动能量的大小，幅值小说明轴承运转时内部激励小，因此，当轴承在充足的润滑作用下，轴承的刚度随之增大，振动响应幅值随之减小，进而使轴承

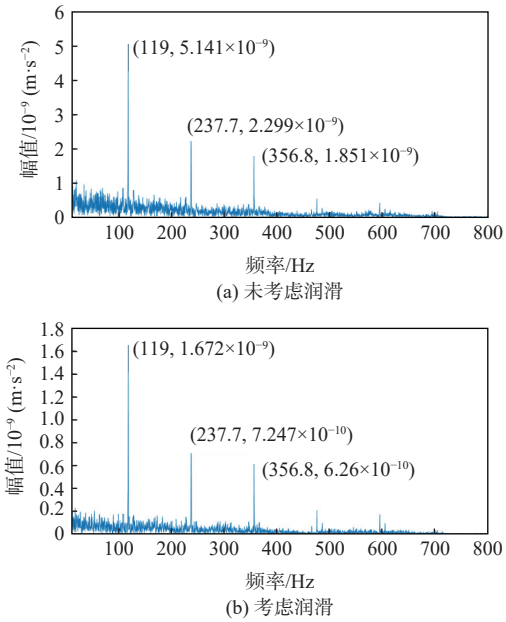


图 13　$\Delta\varphi_{fa}$=0.8°时外圈的振动响应

Fig. 13　Vibration response of outer ring when $\Delta\varphi_{fa}$=0.8°

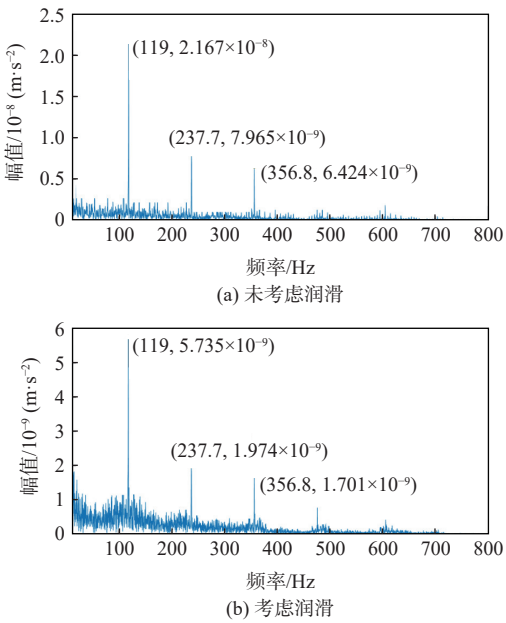


图 14　$\Delta\varphi_{fa}$=1°时外圈的振动响应

Fig. 14　Vibration response of outer ring when $\Delta\varphi_{fa}$ = 1°

内部的振动激励力减小，减缓轴承缺陷变大，延长轴承的使用寿命。

为进一步研究角接触球轴承在弹流润滑作用下不同缺陷跨度角对故障轴承振动响应的影响，设置其转速 n=1 500 r/min，载荷 F_x=1 000 N、F_y=0、F_z=600 N，缺陷跨度角 $\Delta\varphi_{fa}$ 分别为 0.6°、0.5°、0.4°、0.3°、0.2°、0.1°时进行分析，图 15 为滚动体经过不同外圈缺陷跨度角时的瞬时冲击力；将不同缺陷跨度角下轴承振动响应的主频率幅值进行统计，所得的幅值趋势如图 16 所示。

由图 15 可知，随着外圈缺陷跨度角的增大，滚

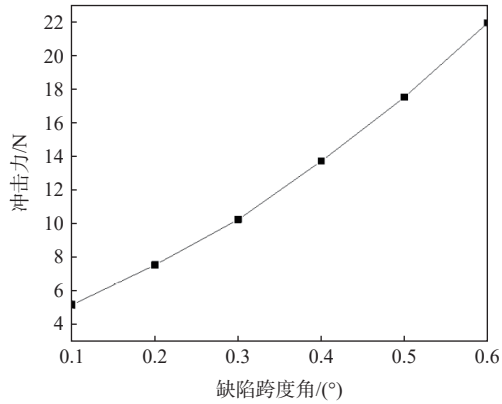


图 15　瞬时冲击力与缺陷跨度角之间的关系

Fig. 15　Relationship between instantaneous impact force and defect span angle

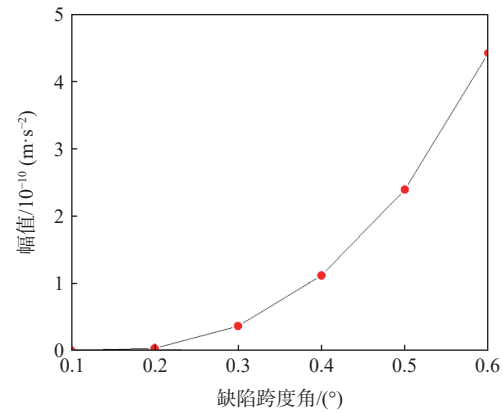


图 16　不同缺陷跨度角的振动幅值变化趋势

Fig. 16　Variation trend of vibration amplitude for different defect span angles

动体经过轴承外圈缺陷时的瞬时冲击力增大。由图 16 可知，当外圈缺陷跨度角$\Delta\varphi_{fa}$ < 0.2°时，其幅值接近于 0，随着外圈缺陷跨度角的增大，轴承的振动响应频率幅值增加。

接着，更详细地研究了外圈缺陷跨度角较小时轴承振动响应的变化规律。当外圈缺陷跨度角 $\Delta\varphi_{fa}$ 分别为 0.1°、0.2°、0.3°时，轴承的振动响应如图 17 所示。由图 17(a)可知，外圈缺陷跨度角 $\Delta\varphi_{fa}$=0.1°时，外圈的振动响应主频率及其幅值无法明显的辨识；由图 17(b)可知，外圈缺陷跨度角 $\Delta\varphi_{fa}$=0.2°时，外圈的振动响应主频率及其幅值虽然可以辨识，但其倍频周围的边频幅值很大；依据图 17(c)可知，外圈缺陷跨度角$\Delta\varphi_{fa}$=0.3°时，外圈的振动响应主频率及其幅值可以清晰辨识，其倍频周围的边频幅值很小。由于幅值大小反映了轴承外圈缺陷引起的振动能量的大小，幅值小说明轴承运转时内部振动激励力小，因此，当外圈缺陷跨度角$\Delta\varphi_{fa}$ < 0.2°时，缺陷引起的轴承内部振动激励力很小，其对轴承振动响应的影响小于轴承的

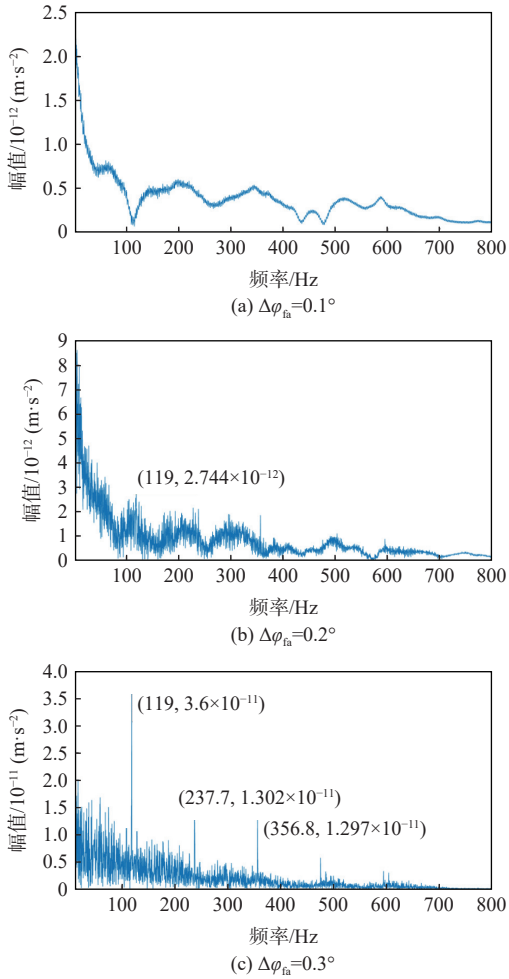


图 17 不同缺陷跨度角时外圈的振动响应

Fig. 17 Vibration response of outer ring at different defect span angles

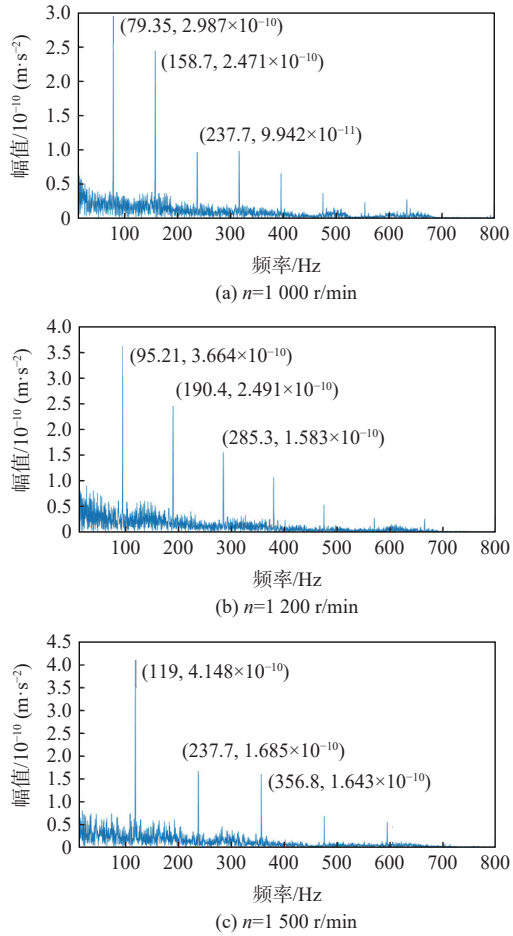


图 18 $\Delta\varphi_{fa}$=0.6°, 不同转速对轴承振动幅值的影响

Fig. 18 Effect of different rotational speeds on bearing vibration amplitude when $\Delta\varphi_{fa}$=0.6°

其他因素(如轴承离心力), 这使得缺陷引起的轴承振动响应不能准确识别, 而当外圈缺陷跨度角 $\Delta\varphi_{fa}>0.2°$时, 外圈缺陷对轴承振动响应有明显影响, 且外圈缺陷跨度角越大, 外圈缺陷对轴承振动的影响越显著, 这就容易识别出轴承出现了故障及其故障类型。

5.2.2 缺陷跨度角和转速的影响

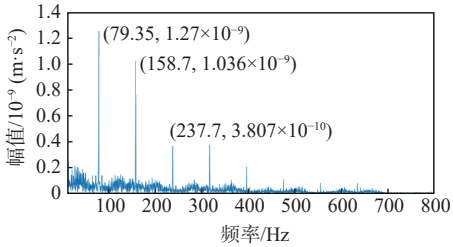
为研究角接触球轴承在弹流润滑作用下不同缺陷尺寸和转速对轴承振动响应的影响, 对轴承在不同转速、同一缺陷尺寸和同一转速、不同缺陷尺寸状态下的轴承振动响应进行研究, 结果如图 18～图 20 所示。

由图 18 可知, 当外圈缺陷跨度角一定时, 轴承外圈故障振动响应的主频率和幅值随着转速的增加而增大。由图 19 和图 20 可知, 当转速一定时, 轴承外圈故障振动响应的主频率不变, 幅值随着外圈缺陷跨度角的增大而增大。因为轴承转速与外圈故障振动响应的主频率成正比, 转速越大外圈故障振动响应的主频率越大, 同时轴承外圈缺陷跨度角和转速的增加都将引起轴承内部振动激励力的增加, 使得外圈故障振动响应的幅值变大, 破坏轴承的正常运转。

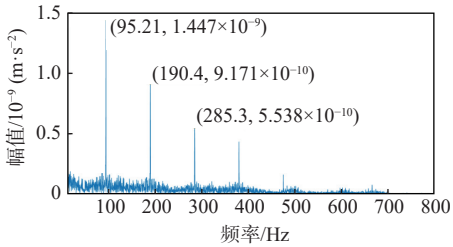
5.2.3 载荷的影响

为获得角接触球轴承在弹流润滑作用下不同方向载荷对轴承振动响应的影响规律, 对轴承在不同方向、不同大小载荷下外圈故障振动响应进行研究。图 21 为轴承外圈缺陷缺陷角 $\Delta\varphi_{fa}$=1°, 转速 n=2 000 r/min, 受不同方向、不同大小载荷时外圈故障振动响应的频谱图。

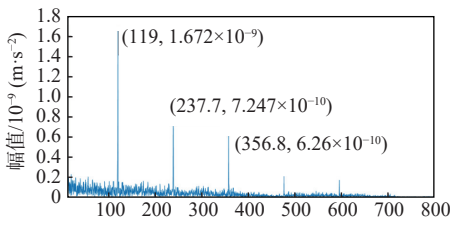
当轴承外圈缺陷尺寸和转速一定, 即径向载荷 F_x 增大、F_y 不变, 轴向载荷 F_z 不变时, 轴承外圈故障振动响应的主频率不变, 幅值随径向载荷 F_x 的增大而增大; 而对比图 21 (b) 和图 21 (c), 发现径向载荷 F_x、F_y 不变, 轴向载荷 F_z 变大时同样可得上述规律。因为轴承所受载荷增加, 将导致轴承各部件间的间隙和缺陷处接触变形变大, 从而使得外圈故障振动响应的幅值增加; 比较图 21 (a) 和图 21 (b), 在轴向载荷 F_z 不变, 径向载荷 F_x、F_y 大小相等, 但其单独作用于轴承时, 可得外圈故障振动响应的主频



(a) n=1 000 r/min



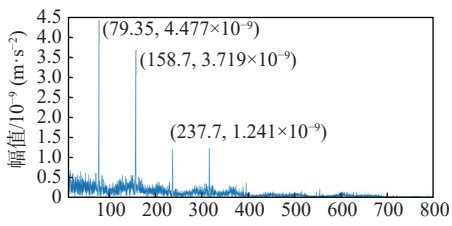
(b) n=1 200 r/min



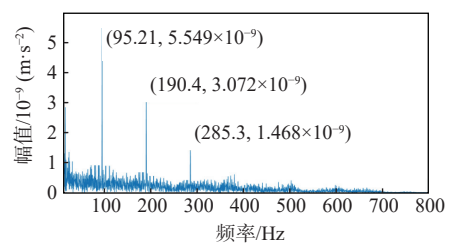
(c) n=1 500 r/min

图 19　$\Delta\varphi_{fa}$=0.8°, 不同转速对轴承振动幅值的影响

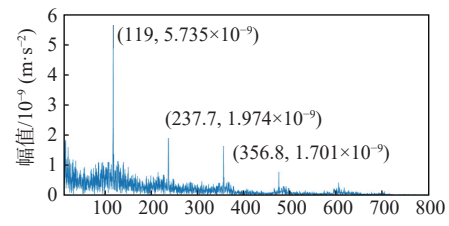
Fig. 19　Effect of different rotational speeds on bearing vibration amplitude when $\Delta\varphi_{fa}$=0.8°



(a) n=1 000 r/min



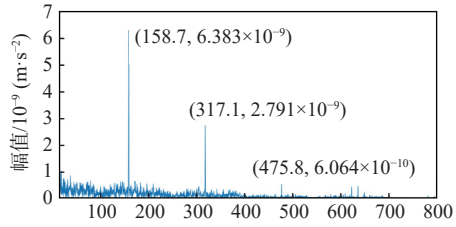
(b) n=1 200 r/min



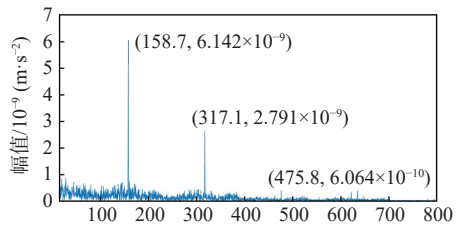
(c) n=1 500 r/min

图 20　$\Delta\varphi_{fa}$=1°, 不同转速对轴承振动幅值的影响

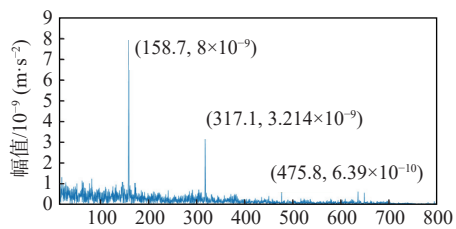
Fig. 20　Effect of different rotational speeds on bearing vibration amplitude when $\Delta\varphi_{fa}$=1°



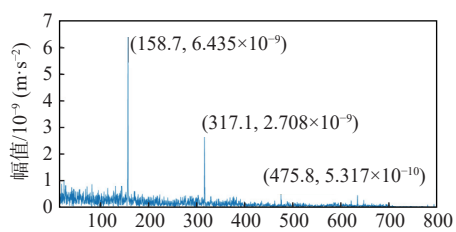
(a) F_x=0 N, F_y=1 000 N, F_z=600 N



(b) F_x=1 000 N, F_y=0 N, F_z=600 N



(c) F_x=1 000 N, F_y=0 N, F_z=1 200 N



(d) F_x=1 500 N, F_y=0 N, F_z=600 N

图 21　载荷对轴承外圈故障振动响应的影响

Fig. 21　Effect of load on vibration response of outer ring fault of bearing

率相同，轴承在径向载荷 F_y 作用下的幅值更大，这可判断出，在轴承 x 方向的径向载荷对外圈故障振动响应影响小于 y 方向，对于故障轴承的维修提供了有效的理论依据。

6　结　论

本文通过建立角接触球轴承的弹流润滑模型和轴承经过局部缺陷的瞬时冲击力模型，得出轴承在弹流润滑条件下转速和缺陷跨度角与冲击力之间的关系。在此基础上，提出在弹流润滑条件下考虑冲击力的局部缺陷角接触球轴承动力学模型，研究了外圈具有缺陷的轴承的振动特性，并对不同参数下轴承的动态响应规律进行分析，得到如下结论：

1）当轴承转速和缺陷跨度角增大时，轴承经过

缺陷区域产生瞬时冲击力随之增大；当轴承考虑弹流润滑作用时，其故障特征频率及倍频处的幅值小于未考虑弹流润滑作用时的幅值，这证明弹流润滑作用可以大大降低瞬时冲击力对轴承外圈故障振动响应的影响。

2）通过分析不同参数下的振动响应可知，缺陷跨度角的增大并不会影响轴承故障特征频率及其倍频的大小，但其幅值会增大，轴承振动幅度加剧；轴承故障特征频率值与载荷方向和大小的变化无关，载荷方向和大小的变化会影响轴承故障特征频率及倍频处的幅值；轴承转速的增大会使轴承故障特征频率和倍频数值增大，同时幅值会随转速的增大而增大，从而加剧轴承振动，降低轴承寿命。

参考文献（References）

[1] 李天夫, 屈伸, 董晨, 等. 数控机床主轴轴承失效分析[J]. 金属热处理, 2022, 47(4): 258-262.
LI T F, QU S, DONG C, et al. Failure analysis on spindle bearing of numerical control machine tool[J]. Heat Treatment of Metals, 2022, 47(4): 258-262(in Chinese).

[2] 卓兰. 滚动轴承的失效形式与故障诊断[J]. 赤峰学院学报 (自然科学版), 2012, 28(12): 115-116.
ZHUO L. Failure forms and fault diagnosis of rolling bearings[J]. Journal of Chifeng University (Natural Science Edition), 2012, 28(12): 115-116(in Chinese).

[3] 杨沛然. 流体润滑数值分析[M]. 北京: 国防工业出版社, 1998.
YANG P R. Numerical analysis of fluid lubrication[M]. Beijing: National Defense Industry Press, 1998 (in Chinese).

[4] 王鹏, 刘晓玲, 王召岩, 等. 表面缺陷对点接触等温稳态弹流润滑的影响[J]. 润滑与密封, 2014, 39(5): 28-34.
WANG P, LIU X L, WANG Z Y, et al. Influence of surface defect on the steady state isothermal EHL of point contacts[J]. Lubrication Engineering, 2014, 39(5): 28-34(in Chinese).

[5] 雷春丽, 巩宝儒, 贾希斌, 等. 自旋对角接触球轴承弹流润滑与油膜刚度的影响[J]. 润滑与密封, 2020, 45(12): 19-25.
LEI C L, GONG B R, JIA X B, et al. Influence of spinning on elasto-hydrodynamic lubrication and film stiffness of angular contact ball bearings[J]. Lubrication Engineering, 2020, 45(12): 19-25(in Chinese).

[6] 路遵友, 吕延军, 李莎, 等. 考虑热弹性变形和表面粗糙度的圆柱滚子轴承热弹流润滑分析[J]. 机械工程学报, 2018, 54(13): 159-169.
LU Z Y, LYU Y J, LI S, et al. Thermal elastohydrodynamic lubrication analysis of cylindrical roller bearing considering thermal elastic deformation and surface roughness[J]. Journal of Mechanical Engineering, 2018, 54(13): 159-169(in Chinese).

[7] WIJNANT Y H, WENSING J A, NIJEN G C. The influence of lubrication on the dynamic behaviour of ball bearings[J]. Journal of Sound and Vibration, 1999, 222(4): 579-596.

[8] YAN P F, YAN C F, WANG K, et al. 5-DOF dynamic modeling of rolling bearing with local defect considering comprehensive stiffness under isothermal elastohydrodynamic lubrication[J]. Shock and Vibration, 2020, 2020(1): 9310278.

[9] 曹宏瑞, 李亚敏, 成玮, 等. 局部损伤滚动轴承建模与转子系统振动仿真[J]. 振动 测试与诊断, 2014, 34(3): 549-552.
CAO H R, LI Y M, CHENG W, et al. Rolling bearing modeling with localized defects and vibration response simulati on of rotor-bearing system[J]. Journal of Vibration, Measurement & Diagnosis, 2014, 34(3): 549-552(in Chinese).

[10] 罗茂林, 郭瑜, 伍星. 基于时变接触刚度的球轴承双冲击现象动力学建模[J]. 振动工程学报, 2018, 31(5): 875-882.
LUO M L, GUO Y, WU X. Dynamic modeling for double impulses behavior of a spalled ball bearing based on time-varying contact stiffness[J]. Journal of Vibration Engineering, 2018, 31(5): 875-882(in Chinese).

[11] 剡昌锋, 康建雄, 苑浩, 等. 考虑弹流润滑及滑动作用下滚动轴承系统局部缺陷位移激励动力学建模[J]. 振动与冲击, 2018, 37(5): 56-64.
YAN C F, KANG J X, YUAN H, et al. Dynamic modeling for local defect displacement excitation in rolling bearing systems under elasto-hydrodynamic lubrication and slip[J]. Journal of Vibration and Shock, 2018, 37(5): 56-64(in Chinese).

[12] 李志农, 周泽文, 胡莺庆, 等. 考虑冲击力的球轴承复合缺陷动力学建模[J]. 沈阳工业大学学报, 2022, 44(2): 173-179.
LI Z N, ZHOU Z W, HU N Q, et al. Dynamic modeling of composite defects of ball bearing with consideration of impact force[J]. Journal of Shenyang University of Technology, 2022, 44(2): 173-179 (in Chinese).

[13] 黄文涛, 董振振, 孔繁朝. 引入撞击力的滚动轴承内圈故障振动模型[J]. 振动与冲击, 2016, 35(17): 121-126.
HUANG W T, DONG Z Z, KONG F C. Vibration model of rolling element bearings with inner race faults considering impact force[J]. Journal of Vibration and Shock, 2016, 35(17): 121-126(in Chinese).

[14] 崔玲丽, 张宇, 巩向阳, 等. 基于振动响应机理的轴承故障定量诊断及量化分析[J]. 北京工业大学学报, 2015, 41(11): 1681-1687.
CUI L L, ZHANG Y, GONG X Y, et al. Vibration mechanism based quantitative diagnosis and quantization analysis of rolling bearing fault[J]. Journal of Beijing University of Technology, 2015, 41(11): 1681-1687(in Chinese).

[15] 胡爱军, 吉新星, 向玲, 等. 滚动轴承非线性时变参数动力学模型与故障机理研究[J]. 机械工程学报, 2022, 58(19): 139-147.
HU A J, JI X X, XIANG L, et al. Nonlinear time-varying parameter dynamic model of rolling bearing and failure mechanism research[J]. Journal of Mechanical Engineering, 2022, 58(19): 139-147 (in Chinese).

[16] CUI L L, ZHANG Y, ZHANG F B, et al. Vibration response mechanism of faulty outer race rolling element bearings for quantitative analysis[J]. Journal of Sound and Vibration, 2016, 364: 67-76.

[17] CUI L L, HUANG J F, ZHANG F B. Quantitative and localization diagnosis of a defective ball bearing based on vertical-horizontal synchronization signal analysis[J]. IEEE Transactions on Industrial Electronics, 2017, 64(11): 8695-8706.

[18] ZHANG F L, ZHANG Y W, GUAN J Y, et al. Fault dynamic modeling and characteristic parameter simulation of rolling bearing with inner ring local defects[J]. Shock and Vibration, 2021, 2021(1): 5077366.

[19] WEN C W, MENG X H, FANG C C, et al. Dynamic behaviors of angular contact ball bearing with a localized surface defect considering the influence of cage and oil lubrication[J]. Mechanism and

Machine Theory, 2021, 162: 104352.

[20] NIU L K, CAO H R, HE Z J, et al. Dynamic modeling and vibration response simulation for high speed rolling ball bearings with localized surface defects in raceways[J]. Journal of Manufacturing Science and Engineering, 2014, 136(4): 041015.

[21] 黄平. 润滑数值计算方法[M]. 北京: 高等教育出版社, 2012.
HUANG P. Lubrication numerical calculation methods[M]. Beijing: Higher Education Press, 2012 (in Chinese).

[22] ROELANDS C J A, WINER W O, WRIGHT W A. Correlational aspects of the viscosity-temperature-pressure relationship of lubricating oils(dr in dissertation at technical university of delft, 1966)[J]. Journal of Lubrication Technology, 1971, 93(1): 209-210.

[23] 黄平. 弹性流体动压润滑数值计算方法 [M]. 北京: 清华大学出版社, 2013.
HUANG P. Numerical calculation method for elastohydrodynamic lubrication [M]. Beijing: Tsinghua University Press, 2013(in Chinese).

[24] 吴明星, 吴维, 胡纪滨, 等. 考虑自旋的高速角接触球轴承油膜刚度计算[J]. 振动与冲击, 2014, 33(10): 38-42.
WU M X, WU W, HU J B, et al. Oil film stiffness calculation of high speed angular contact ball bearings considering spinning[J]. Journal of Vibration and Shock, 2014, 33(10): 38-42(in Chinese).

[25] HARRIS T A, KOTZALAS M N. Essential concepts of bearing technology[M]. Boca Raton: CRC Press, 2006.

[26] 李昊泽, 贺雅, 冯坤, 等. 考虑时变激励的滚动轴承局部故障动力学建模[J]. 航空学报, 2022, 43(8): 625176.
LI H Z, HE Y, FENG K, et al. Dynamic modeling of rolling bearing local fault considering time-varying excitation[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(8): 625176(in Chinese).

[27] 周智, 朱永生, 闫柯, 等. 考虑缺陷因素的角接触球轴承七自由度动力学模型[J]. 计算机集成制造系统, 2014, 20(11): 2836-2842.
ZHOU Z, ZHU Y S, YAN K, et al. Seven degrees of freedom dynamic model of angular contact ball bearing with defects[J]. Computer Integrated Manufacturing Systems, 2014, 20(11): 2836-2842(in Chinese).

Dynamic modelling of angular contact ball bearings with local defects under EHL considering impact force

LEI Chunli[1]，XUE Wei[1,*]，FAN Gaofeng[2]，SONG Ruizhe[1]，LIU Kai[1]

(1. School of Mechanical and Electrical Engineering，Lanzhou University of Technology，Lanzhou 730050，China；
2. Yunnan Wenshan Aluminum Co.，Ltd.，Wenshan 663000，China)

Abstract：In order to precisely analyze the operating status of angular contact ball bearings (ACBBs) with local defects and response to the impact force caused by bearing lubrication and defects in the traditional model, a dynamic model of ACBBs with local defects considering impact force under elastohydrodynamic lubrication (EHL) conditions was proposed. Firstly, an EHL model for defective bearings considering spin was established, and the oil film thickness and oil film stiffness of the defective bearings were calculated. Time-varying displacement excitation model and time-varying stiffness model for local defects in ACBBs were presented. On this basis, an instantaneous impact force function related to defect size and bearing speed was established. Secondly, based on Hertz contact theory and impact force function, a dynamic calculation method for ACBBs with local defects on the outer ring was proposed. Finally, the vibration characteristics of ball bearings with faults were investigated, and the influence of different parameters on the dynamic response of the bearings was analyzed through experiments. The calculation results show that under lubrication conditions, the impact force on the vibration of the bearing is significantly weakened. With the increase in defect size and load, the frequency of bearing fault characteristics remains unchanged, but its amplitude increases. As the bearing speed increases, the characteristic frequency and amplitude of bearing faults increase, providing a powerful basis for fault diagnosis and maintenance of rolling bearings.

Keywords：angular contact ball bearings; local defect; instantaneous impact force; elastohydrodynamic lubrication; time-varying displacement; dynamic modellin

Received：2023-06-09；**Accepted**：2023-07-21；**Published Online**：2023-08-24 15：36
URL：link.cnki.net/urlid/11.2625.V.20230823.1459.003
Foundation item：National Natural Science Foundation of China (51465035)
* **Corresponding author**. E-mail：2542908749@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0342

基于K互近邻与核密度估计的DPC算法

周玉*，夏浩，刘虹瑜，白磊
(华北水利水电大学 电气工程学院，郑州 450045)

摘 要：快速搜索和发现密度峰值聚类(DPC)算法是一种基于密度的聚类算法。该算法不需要迭代和过多的设定参数，但由于计算局部密度时没有考虑数据的局部结构，导致无法识别簇密度小的聚类中心。针对此问题，提出基于K互近邻(KN)和核密度估计(KDE)的DPC(KKDPC)算法。通过K近邻和核密度估计方法得到数据点的K互近邻数量和局部核密度；将K互近邻数量与局部核密度进行加和获得新的局部密度；根据数据点的局部密度得到相对距离，并通过构建决策图选取聚类中心及分配非中心点。利用人工数据集和真实数据集进行实验，并与DPC、基于密度的噪声空间聚类应用(DBSCAN)、K-means、模糊C均值聚类算法(FCM)、基于K近邻的DPC(DPC-KNN)、近邻优化DPC(DPC-NNO)、基于模糊加权共享邻居的DPC(DPC-FWSN)算法进行对比。通过计算调整互信息(AMI)、调整兰德指数(ARI)、归一化互信息(NMI)来验证KKDPC算法的性能。实验结果表明：KKDPC算法能更加准确地识别聚类中心，有效地提高聚类精度。

关 键 词：聚类算法；密度峰值；K近邻；K互近邻；核密度估计
中图分类号：TP311
文献标志码：A **文章编号**：1001-5965(2025)06-1978-13

聚类是数据挖掘的主要方法之一，也是无监督学习的重要组成部分。根据对象的相似性度量将对象划分为子集或聚类，使得聚类内的对象具有高度相似性，并且属于不同聚类的对象具有高度不相似性[1]。聚类分析能从复杂的数据中挖掘出潜在规律并提取有价值的信息，其广泛应用于生物检测[2-3]、图像处理[4-5]、入侵检测[6-7]等。在过去的几十年里，研究者已经开发出不同类型的聚类算法，经典的聚类算法包括基于划分的K-means[8]算法和K-medoids[9]算法，基于层次的基于代表点的聚类(clustering using representatives, CURE)[10]算法和利用层次方法的平衡迭代规约聚类(balanced iterative reducing and clustering using hierarchies, BIRCH)[11]算法，基于密度的噪声空间聚类应用(density-based spatial clustering of applications with noise, DBSCAN)[12]算法和通过点排序识别聚类结构(ordering points to identify the clustering structure, OPTICS)[13]算法，基于网格的WaveCluster[14]算法和统计信息网格(statistical information grid, STING)[15]聚类算法。

快速搜索和发现密度峰值(clustering by fast search and find of density peaks, DPC)算法是2014年提出的一种新的基于密度的聚类算法[16]。相比传统聚类算法，DPC算法优点在于能够快速发现聚类中心且聚类过程不需要迭代、能够适应任意形状簇、适用于大规模数据等。尽管DPC算法与其他算法相比具有很多明显的优势，但算法本身也存在如下缺点：①截断距离 d_c 为人工设定参数，在面对未知数据集时，截断距离 d_c 难以确定；②由于计算局部密度时没有考虑数据的局部结构，所以当类簇之间具有不同的密度时，DPC算法无法识别稀疏簇的聚类中心。

针对上述问题，有很多学者提出改进的DPC算法。为考虑数据点的局部特征，Du等[17]将K近

收稿日期：2023-06-12；**录用日期**：2023-10-13；**网络出版时间**：2023-10-20 10:53
网络出版地址：link.cnki.net/urlid/11.2625.V.20231019.1645.003
基金项目：国家自然科学基金(U1504622, 31671580)；河南省青年骨干教师培养计划(2018GGJS079)
*通信作者. E-mail：zhouyu_beijing@126.com

引用格式：周玉，夏浩，刘虹瑜，等. 基于K互近邻与核密度估计的DPC算法[J]. 北京航空航天大学学报，2025，51(6)：1978-1990.
ZHOU Y，XIA H，LIU H Y，et al. DPC algorithm based on K-reciprocal neighbors and kernel density estimation[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1978-1990 (in Chinese).

邻算法引入到局部密度的计算中, 使得数据点均有稳定的局部密度, 同时解决截断距离 d_c 难以确定的问题。同样基于 K 近邻算法, Jiang 等[18] 使用基尼系数找到最优截止距离, 然后, 用 K 近邻找到中心点, 在一定程度上能够解决密度差异大的数据集, 但该方法处理复杂数据能力较弱。Yuan 等[19] 提出自适应合并策略的 K 近邻密度峰值聚类算法, 利用 K 最近邻居策略计算局部密度和分配非中心点, 并通过自适应分段聚类合并策略, 提高聚类算法处理复杂数据能力。针对 DPC 算法分配非中心点策略单一的问题, Xie 等[20] 提出基于模糊加权 K 近邻的密度峰值检测和数据点分配的鲁棒聚类。该算法根据 K 个最近邻居计算局部密度并使用双分配策略提高了 DPC 算法的准确性。Liu 等[21] 提出基于共享最近邻的密度峰值算法, 根据共享最近邻方法计算数据点的局部密度, 并在考虑邻域信息的情况下提出新的相对距离, 优化非中心点的分配策略。然而, 当数据集中簇密度差距过大时, 上述算法依然无法对密度稀疏簇中的聚类中心进行有效地选取。针对密度分布不均的问题, Chen 等[22] 将 K 近邻与逆近邻算法引入到局部密度的计算中, 提高稀疏样本的局部密度, 使 DPC 算法更准确地找到聚类中心。Zhao 等[23] 通过 K-最近邻和模糊邻域得到最近邻模糊核函数, 然后, 利用最近邻模糊核函数重新定义局部密度, 平衡样本密度在密集和稀疏区域的贡献。

针对 DPC 算法难以识别稀疏簇的聚类中心问题, 本文提出基于 K 互近邻和核密度估计 (kernel density estimation, KDE) 的 DPC (KKDPC) 算法。首先, 根据数据点的 K 近邻得到 K 互近邻个数, 并将 K 互近邻个数进行量化; 其次, 利用 KDE 算法得到数据点的局部核密度; 接着, 将 K 互近邻个数与局部核密度进行相加获得新的局部密度, 用新的局部密度代替传统 DPC 算法的局部密度; 最后, 根据局部密度与相对距离构建决策图选取聚类中心并进行聚类。无论聚类之间的密度差异有多大, 该算法都能够平衡密度, 从而更加准确地识别聚类中心。

本文的主要创新点如下:

1) 使用 K 互近邻算法平衡不同类簇之间的密度差异, 在决策图中突出簇密度小的聚类中心, 使得能够更加准确地选取聚类中心。

2) 提出用高斯核函数和 K 近邻的算法对数据点邻域进行 KDE, 用来保留数据点的邻域特点。在数据点的邻域内拥有着相同的 K 互近邻数量情况下, 此算法能够突出邻域内密度最高的数据点, 并保证在新的局部密度中, 数据点间仍存在着差异, 避免非中心点在分配过程中存在错乱的情况。

1　DPC 算法

1.1　DPC 算法介绍

DPC 算法是通过构建关于密度与距离的决策图来确定聚类中心。关于聚类中心的解释: 聚类中心的密度最大且被密度小于其数据点所包围, 而不同聚类中心间的距离较远。根据数据点归类于密度大且距离最近的簇的原则进行聚类。

DPC 算法中定义的 2 个参数:

定义 1　局部密度:

$$\rho_i = \sum_i \chi(d_{ij} - d_c) \tag{1}$$

当 $x<0$ 时, $\chi(x)=1$, 否则, $\chi(x)=0$。d_{ij} 为数据点 i 与 j 之间的欧式距离, d_c 为截断距离。d_c 为超参数, 通常被定义为: d_c=p%, 表示所有数据点的平均邻居数为数据总量的 p%。避免数据点局部密度出现相同的情况, 采用高斯函数来代替式 (1), 如下:

$$\rho_i = \sum_{i\neq j} \mathrm{e}^{-\left(\frac{d_{ij}}{d_c}\right)^2} \tag{2}$$

定义 2　相对距离:

$$\delta_i = \min_{j:\rho_j>\rho_i} d_{ij} \tag{3}$$

计算数据点与密度大于其最近数据点的距离。当数据点 i 的密度为数据集中最大时, 式 (3) 不成立。此时, 其相对距离为距离 i 最远点的距离, 即 $\delta_i=\max_j d_{ij}$。利用式 (2) 和式 (3) 获取数据点的局部密度和相对距离构建决策图来获取初始聚类中心。计算局部密度和相对距离的乘积来对聚类中心进行选取, 如下:

$$\gamma = \rho\delta \tag{4}$$

图 1 为一组二维数据, 图中包含数据规模不同的 2 个簇。通过 DPC 算法构建局部密度-相对距离决策图如图 2 所示。由于红色点聚类形成簇的数据规模大, 通过截断距离计算得到的聚类中心的局部密度最大, 且为整个数据集中局部密度最大的数据点, 位于决策图的最右上方。其余大部分红色

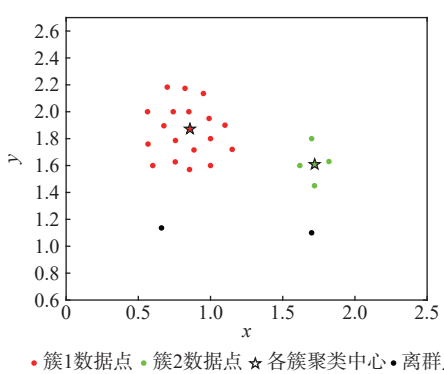


图 1　二维数据

Fig. 1　Two-dimensional data

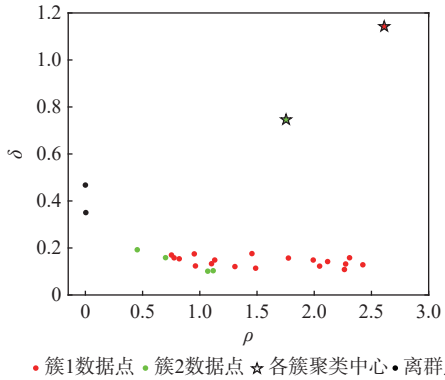


图2 决策图

Fig. 2 Decision graph

数据点的局部密度虽大但相对距离小，则位于决策图的右下方。绿色数据点对应的簇，聚类中心的局部密度虽不是很大，但相对距离大，在决策图中位于右上方。离群点的局部密度最小，则位于决策图的左下方。通过上述可知，能从决策图直观地选取聚类中心，接着，根据局部密度和相对距离的值将剩余数据点分配到密度更高且距离最近数据点所属类中。

1.2 DPC 算法分析

DPC 算法优点在于不需要迭代过程和人为设定过多参数，可以在没有先验知识的前提下，通过构建决策图对数据集的聚类中心进行选择，但聚类算法本身存在一些不足。本节将通过人工合成数据集 Data 和数据集 Jain 对 DPC 算法中存在的不足进行解释。DPC 算法中截断距离 d_c 的计算方法是：人为设定参数，使数据点的邻居平均数量约为数据集中总数据量的 1%～2%。使用 DPC 算法构建决策图并分别对图 3 和图 4 的数据集进行聚类。为解释数据点的聚类情况，分别绘制了每个数据点的邻居平均数量约为数据集中总数据量的 2%、5%、10%、20%、50% 和 100% 的决策图，并对应于不同 d_c 值，如图 5 和图 6 所示。

由图 5 可知，由红色数据点聚类形成的簇的聚类中心容易被识别出来，因为红色数据点分布最密集。由绿色数据点聚类形成的簇，其聚类中心在图 5(c) 中参数 p=10 之后更加明显，易于识别，而在此之前，会被视为离群点。而蓝色数据点分布最为稀疏，所以由蓝色数据点聚类形成的簇的聚类中心无法被识别，且极容易被视为离群点。

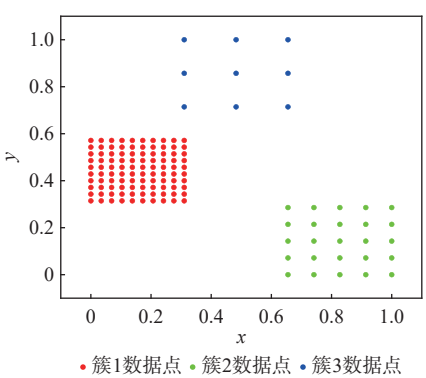


图3 Data 数据集

Fig. 3 Data dataset

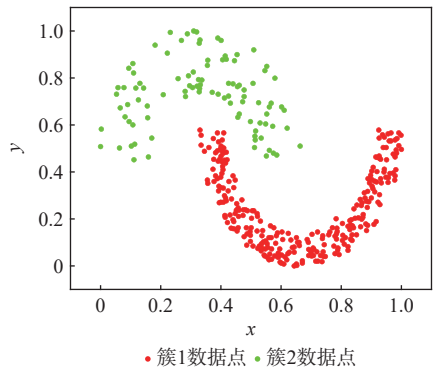


图4 Jain 数据集

Fig. 4 Jain dataset

由图 6 可知，红色数据点数量多且分布密集，而绿色数据点数量少且分布稀疏，导致绿色数据点的局部密度小且相对距离小。在决策图中，绿色数据点分布于决策图的左下方，而红色数据点分布在决策图的右下方。无论截断距离如何设定，都无法在决策图中选取到绿色数据点部分的聚类中心。分别根据图 5(c) 及图 6(a) 选取聚类中心，聚类结果如图 7 和图 8 所示。

2 KKDPC 算法

面对 DPC 结果受截断距离影响大，簇密度小的聚类中心很难被选取的问题。本文提出 KKDPC 算法，首先，通过 K 互近邻方法得到数据点的 K 互近邻数量，K 互近邻数量越多，该数据点越有可能为聚类中心；其次，使用 KDE 算法计算数据点的局部核密度；最后，将数据点的 K 互近邻数量与局部核密度相加得到新的局部密度，用新的局部密度代替原 DPC 算法中的局部密度。

2.1 K 互近邻与核密度估计

首先，将 K 近邻的思想引入到局部密度的计算中。给定数据集 $X=\{x_1,x_2,\cdots,x_N\}$，$x_i,x_j\in X$，$d(\cdot,\cdot)$为欧氏距离函数，$N_{Nk}(x_i)$为距离 x_i第 k 近邻的数据点，则 K 近邻的定义如下：

$$k_{NN}(x_i)=\{x_j\in X|d(x_i,x_j)\leqslant d(x_i,N_{Nk}(x_i))\} \tag{5}$$

通过 K 近邻可以得知数据点的邻域信息，若 $x_j\in k_{NN}(x_i)$且 $x_i\in k_{NN}(x_j)$，则 x_i与 x_j为 K 互近邻。若数据点的互近邻个数越多，则该数据点越有可能为聚类中心。K 互近邻优点在于：只考虑数据点与近邻之间的分布紧密程度，不考虑类簇之间的密度差

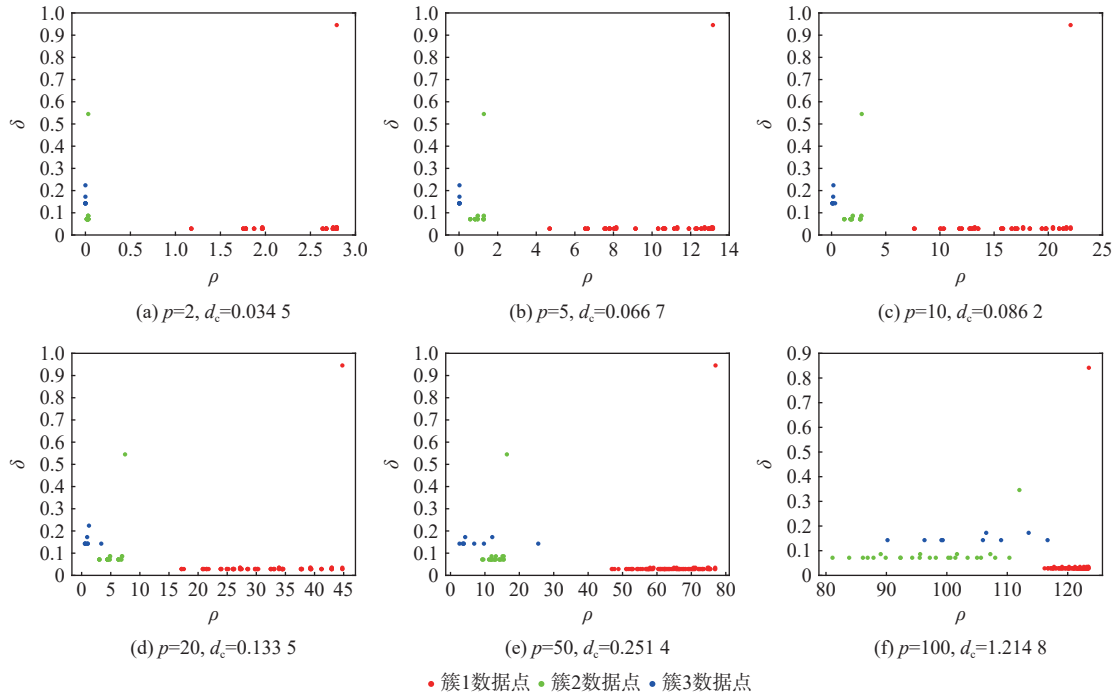


图 5　Data 数据集在不同 d_c 值下的决策图

Fig. 5　Decision graphs of Data dataset under different values of d_c

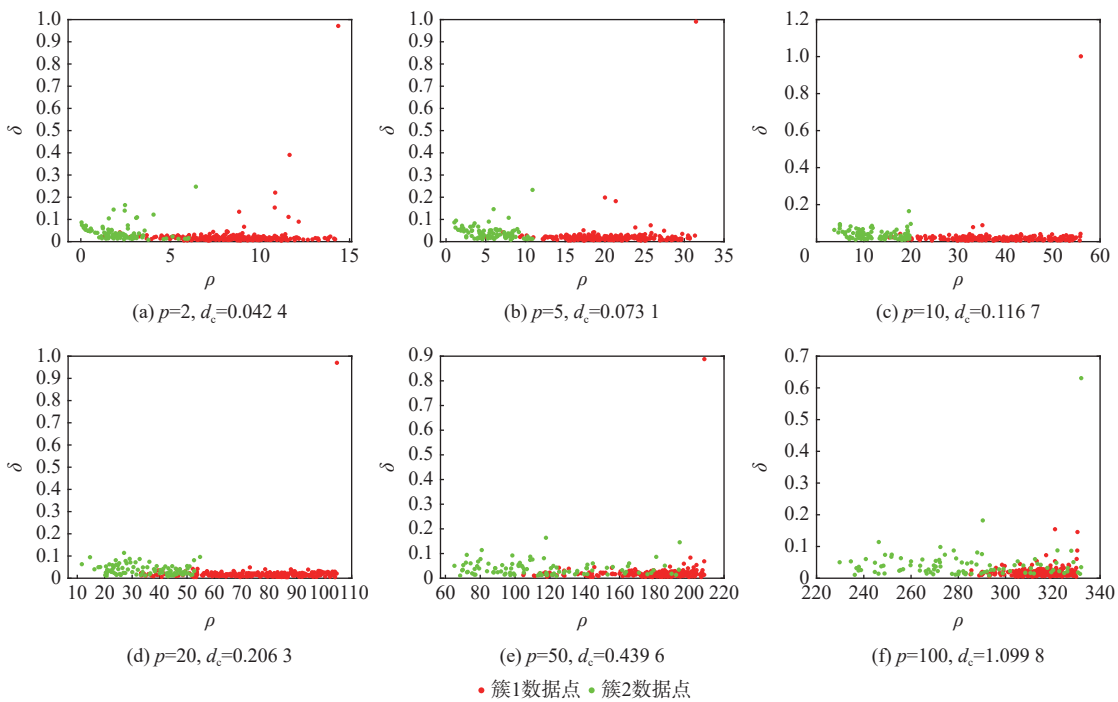


图 6　Jain 数据集在不同 d_c 值下的决策图

Fig. 6　Decision graphs of Jain dataset under different values of d_c

异，可以使密度小的簇的聚类中心也能有较大的 K 互近邻个数。可根据式 (6) 和式 (7) 得到数据点的互近邻个数。

$$K_N(x_i,x_j)=\begin{cases}1 & x_j\in k_{NN}(x_i)且x_i\in k_{NN}(x_j)\\ 0 & 其他\end{cases} \tag{6}$$

$$n_{KN}(x_i)=\sum_{x_j\in k_{NN}(x_i)}K_N(x_i,x_j) \tag{7}$$

其中，若 $K_N(x_i,x_j)=1$，表示 x_i 和 x_j 互为近邻，反之亦然。$n_{KN}(x_i)$表示 x_i 的互近邻个数，如果数据点 x_i的第 k 近邻居不止一个，可能会导致 $n_{KN}(x_i)$值大于 k。本文算法针对 K 互近邻个数加个限定条件：$n_{KN}(x_i)\leqslant k$。

接着，根据高斯核函数和 KDE 算法，计算数据点局部核密度。其目的：①在保留数据点本身的分布的特点；②当数据点邻域均有相同的 n_{KN} 时，突出

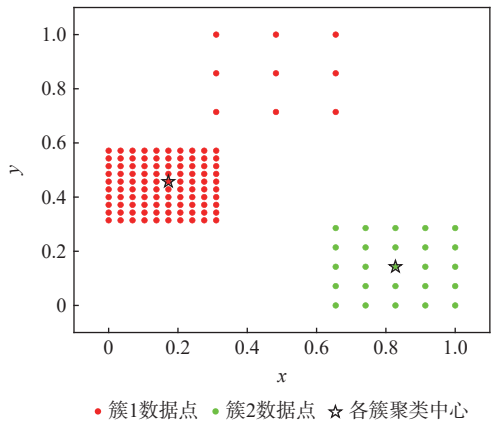


图 7 Data 数据集聚类结果

Fig. 7 Clustering results of Data dataset

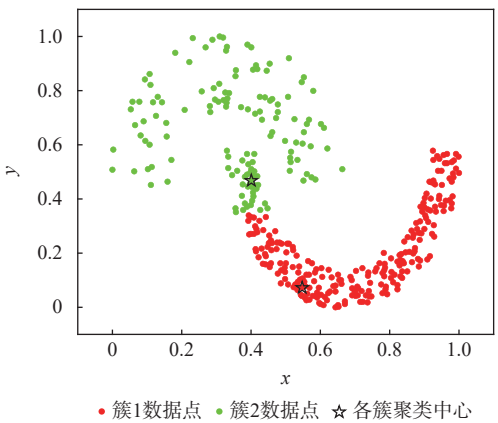


图 8 Jain 数据集聚类结果

Fig. 8 Clustering results of Jain dataset

局部核密度最高的数据点。高斯核函数和 KDE 的计算式为

$$K_{\text{Gaussian}}(\boldsymbol{x})=\frac{1}{(2\pi)^d}\exp\left(-\frac{\|\boldsymbol{x}\|^2}{2}\right) \tag{8}$$

$$K_{\text{DE}}(x_i)=\sum_{x_j\in k_{\text{NN}}(x_i)}\frac{1}{(2\pi)^d}\exp\left(-\frac{d_k(x_i,x_j)^2}{2}\right) \tag{9}$$

式中：$k_{\text{NN}}(x_i)$为数据点x_i的 k 个最近邻居；$\|\boldsymbol{x}\|$为 $\boldsymbol{x}$ 的范数；$d_k(x_i,x_j)$为数据点x_i与第 k 近邻的数据点x_j间的距离；d 为数据维度。

最后，将数据点的 K 互近邻数量与数据点的局部核密度相加得到数据点的局部密度，用新的局部密度代替传统 DPC 算法中根据截断距离计算的局部密度：

$$\rho(x_i)=n_{\text{KN}}(x_i)+K_{\text{DE}}(x_i) \tag{10}$$

同样，在非中心分配过程中，遵循传统 DPC 算法的分配策略。由式 (3) 得到数据点的相对距离，并根据新的局部密度与相对距离构建新的决策图，相比传统 DPC 算法能够更加容易地选取聚类中心。

2.2 KKDPC 算法步骤

输入 数据集 X、近邻个数 k。

输出 聚类中心 center、数据集的类标签 label(X)。

1. 计算数据点间的欧氏距离 dist(x_i,x_j)，并构建距离矩阵 $\boldsymbol{D}$；

2. 根据距离矩阵 $\boldsymbol{D}$ 和式 (5) 获取数据点的 k 近邻居$k_{\text{NN}}(x_i)$；

3. 根据式 (6) 和式 (7) 获取数据点的 K 互近邻数量$n_{\text{KN}}(x_i)$；

4. 根据式 (8) 和式 (9) 计算数据点的局部核密度$K_{\text{DE}}(x_i)$；

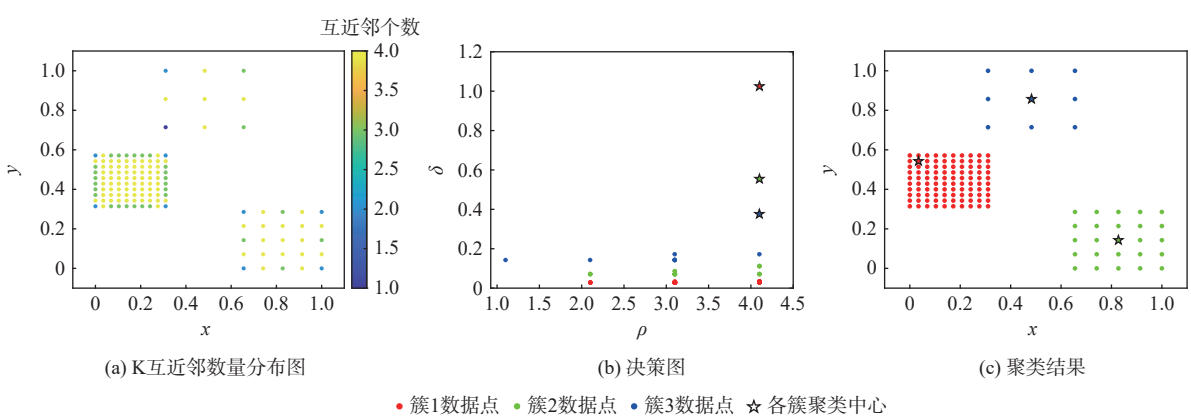
5. 根据式 (10) 获取数据点的局部密度ρ，并根据式 (3) 获取相对距离δ；

6. 根据局部密度ρ和相对距离δ构建决策图选取聚类中心 center，并赋予类标签 label(center)；

7. 将非中心数据点依次分配到局部密度更高且距离最近数据点所属类中，得到数据集的类标签 label(X)。

2.3 KKDPC 算法分析

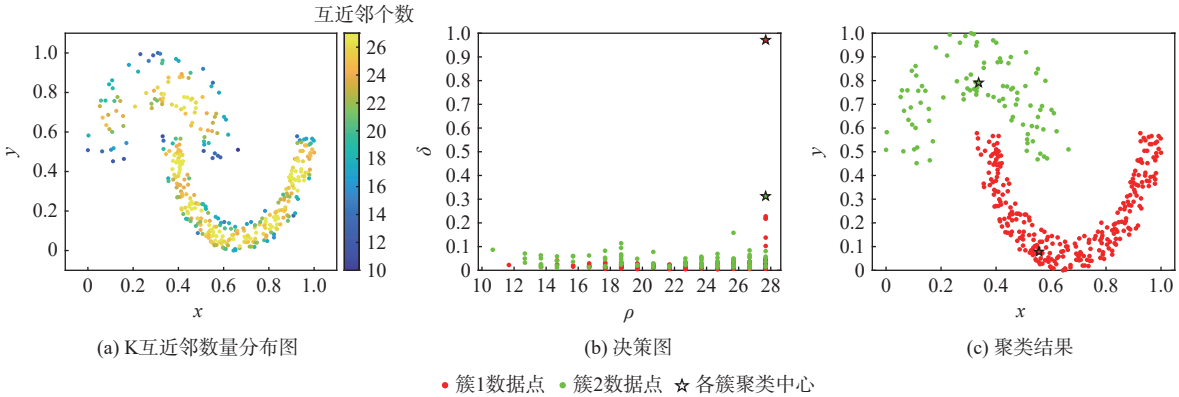
为验证 KKDPC 算法的有效性，同样使用 Data 和 Jain 数据集进行实验分析，如图 9 和图 10 所示。首先，根据实验参数 k=4 及 k=27 计算数据点



(a) K互近邻数量分布图 (b) 决策图 (c) 聚类结果

图 9 Data 数据集聚类中心分析

Fig. 9 Analysis of cluster centers in Data dataset



(a) K互近邻数量分布图　　(b) 决策图　　(c) 聚类结果

图 10　Jain 数据集聚类中心分析

Fig. 10　Analysis of cluster centers in Jain dataset

的 K 互近邻数量，图 9(a) 和图 10(a) 直观展示出数据点的 K 互近邻数量。由图可知，颜色越浅（黄），数据点的 n_{KN} 值越大。无论簇密度大小如何，每个簇中的部分数据点都会存在较高的 n_{KN} 值。接着，计算数据点的局部核密度 KDE，根据式 (10) 得到数据点的局部密度，并通过构建决策图选取聚类中心。在图 9(b) 和图 10(b) 中，五角星表示被选取的聚类中心。最后，将非中心数据点依次分配到局部密度更高且距离最近数据点所属类中，聚类结果及聚类中心位置如图 9(c) 和图 10(c) 所示。

由图 9 和图 10 可知：KKDPC 算法可以有效平衡不同簇之间的密度差异，并且通过优化聚类中心的选取，进而提高聚类精度。

关于 KKDPC 计算复杂度的分析如下：

1）计算数据点间的距离所需要的时间复杂度为 $O(N^2)$；

2）计算数据点间的 K 近邻数量、局部核密度、局部密度、相对距离的时间复杂度均为 $O(N)$；

3）非中心点分配策略的时间复杂度为 $O(N)$；

4）KKDPC 算法的计算复杂度近似为 $O(N^2)$。

3　实验与分析

为更加全面验证 KKDPC 算法的性能，通过人工数据集和真实数据集进行实验，并与 K-means[8]、模糊 C 均值聚类（fuzzy C-means clustering, FCM）[24]、DPC[16]、DBSCAN[12]、基于 K 近邻的 DPC(DPC based on K nearest neighbors, DPC-KNN)[17]、近邻优化 DPC(DPC with nearest neighbor optimization, DPC-NNO)[22]、基于模糊加权共享邻居的 DPC(DPC based on fuzzy and weighted shared neighbor, DPC-FWSN)[23] 这 7 种算法进行比较。为评估聚类算法的性能，采用了 3 种常用的聚类评估指标：调整互信息 (adjusted mutual information, AMI) [25]、调整兰德指数 (adjusted rand index, ARI) [25]、归一化互信息 (normalized mutual information, NMI) [26]，各项指标值最优值为 1，越接近 1 则算法对数据集的聚类效果越好。算法采用 MATLAB R2020a 编写，实验环境为：AMDR7 3.2 GHz CPU, 8.00 GB 内存, Windows11 操作系统。

3.1　人工数据集

首先，通过人工数据集验证算法的性能，人工数据集包含任意形状的簇、簇与簇之间相互连接、类簇之间密度差异大等特点，如表 1 所示。

表 1　人工数据集

Table 1　Artificial datasets

数据集	数据量	维度	类簇个数
A3[27]	7 500	2	50
Aggregation[28]	788	2	7
Flame[29]	219	2	2
Jain[30]	373	2	2
Line blobs[31]	266	2	3
Spiral[32]	312	2	3

在聚类之前，为消除不同维度之间的量纲大小对实验产生的影响，故对数据进行 [0,1] 的归一化处理。为充分验证并对比实验结果，K-means 与 FCM 算法均运行 50 次选取最优的结果；DPC 算法的 d_c 值的设定标准为：使数据点的邻居平均数量约为数据集中总数据量的 1%～100%，步长为 1%；DBSCAN 聚类算法中 2 个参数设置半径和密度阈值取值范围分别为 (0,1] 和 [1,50]，步长分别为 0.01 和 1；KKDPC 算法、DPC-KNN、DPC-NNO 和 DPC-FWSN 算法中 k 值的选取范围为 [1,100]，步长为 1。表 2 为各个聚类算法在人工数据集中的聚类结果，最优结果用粗体表示。图 11～图 16 为不同聚类算法在人工数据集上的聚类结果，图中，不同颜色的点被分为不同的类簇，五角星表示类簇的聚类中心，黑色点为 DBSCAN 算法确定的噪声点。

表 2 人工数据集实验结果

Table 2 Experimental results on artificial datasets

算法	AMI					
	A3	Aggregation	Flame	Jain	Line blobs	Spiral
KKDPC	0.977 5	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**
DPC[16]	0.976 7	0.992 2	**1.000 0**	0.709 2	0.779 9	**1.000 0**
DBSCAN[12]	0.867 4	0.984 1	0.938 6	**1.000 0**	**1.000 0**	**1.000 0**
K-means[8]	0.963 9	0.835 7	0.550 6	0.491 6	0.590 5	−0.004 7
FCM[24]	0.937 3	0.805 0	0.589 7	0.515 2	0.586 6	−0.004 7
DPC-KNN[17]	0.977 4	0.992 2	**1.000 0**	0.631 9	0.779 4	**1.000 0**
DPC-NNO[22]	**0.979 4**	0.990 5	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**
DPC-FWSN[23]	0.977 3	0.995 5	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**

算法	NMI					
	A3	Aggregation	Flame	Jain	Line blobs	Spiral
KKDPC	0.965 9	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**
DPC[16]	0.962 9	0.995 6	**1.000 0**	0.822 4	0.721 0	**1.000 0**
DBSCAN[12]	0.685 9	0.991 1	0.986 6	**1.000 0**	**1.000 0**	**1.000 0**
K-means[8]	0.921 4	0.765 3	0.565 8	0.576 7	0.500 8	−0.005 0
FCM[24]	0.859 3	0.731 8	0.622 4	0.585 3	0.494 2	−0.005 0
DPC-KNN[17]	0.965 6	0.995 6	**1.000 0**	0.757 7	0.717 9	**1.000 0**
DPC-NNO[22]	**0.966 5**	0.994 9	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**
DPC-FWSN[23]	0.963 9	0.997 8	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**

算法	NMI					
	A3	Aggregation	Flame	Jain	Line blobs	Spiral
KKDPC	0.978 6	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**
DPC[16]	0.977 8	0.992 4	**1.000 0**	0.742 9	0.788 5	**1.000 0**
DBSCAN[12]	0.920 4	0.986 0	0.969 0	**1.000 0**	**1.000 0**	**1.000 0**
K-means[8]	0.968 6	0.883 6	0.565 2	0.528 8	0.595 8	0.001 1
FCM[24]	0.945 7	0.840 1	0.604 8	0.554 7	0.591 5	0.000 4
DPC-KNN[17]	0.978 7	0.992 4	**1.000 0**	0.676 2	0.787 1	**1.000 0**
DPC-NNO[22]	**0.980 3**	0.991 5	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**
DPC-FWSN[23]	0.978 4	0.995 8	**1.000 0**	**1.000 0**	**1.000 0**	**1.000 0**

由人工数据集的聚类结果可知，A3 数据集为凸形簇数据集，且含有多个簇数，KKDPC、DPC、DPC-KNN、DPC-NNO 和 DPC-FWSN 算法有良好聚类性能，而 K-means、FCM 聚类中心选取具有随机性，聚类结果不稳定，DBSCAN 聚类算法对参数选取过于敏感，会聚类产生很多异常点。在 Aggregation 数据集中，KKDPC 精度能达到 100%，优于其他聚类算法。在 Flame 和 Spiral 数据集中，由于簇形状的原因，导致 K-means、FCM 聚类结果较差。在 Jain 数据集中，由于存在类簇之间密度不均等的情况，DPC、DPC-KNN 算法均无法选取到密度较低簇中的聚类中心，导致聚类结果差。而因为本身形状为非凸簇，K-means 和 FCM 结果也不好。在 Line blobs 数据集中，DPC、DPC-KNN 算法错误地把红色的部分数据点归类于蓝色或绿色数据点所对应的簇，导致聚类结果变差。

3.2 UCI 数据集

本节将使用 18 个 UCI 数据集[33] 进行实验，并与 DPC、DPC-KNN、DBSCAN、K-means、FCM、DPC-NNO 和 DPC-FWSN 这 7 种聚类算法进行对比。同样使用 AMI、ARI 和 NMI 这 3 个指标验证聚类算法性能。表 3 为实验所需要的 UCI 数据集，表 4 为各个聚类算法的实验结果，最优结果用粗体表示。

由表 4 可知，在 UCI 数据集中 KKDPC 算法结果总体优于其他聚类算法，其中，8 个数据集的聚类指标最优。相比 DPC-KNN、DPC-NNO 和 DPC-FWSN 这 3 个改进 DPC 算法，KKDPC 算法在 11 个数据集中聚类指标最优。而且 KKDPC 在 Ecoli、Ionosphere、WDBC 和 MFCCs 数据中的 ARI、AMI 和 NMI 这 3 个指标提升尤为明显，原因在于 UCI

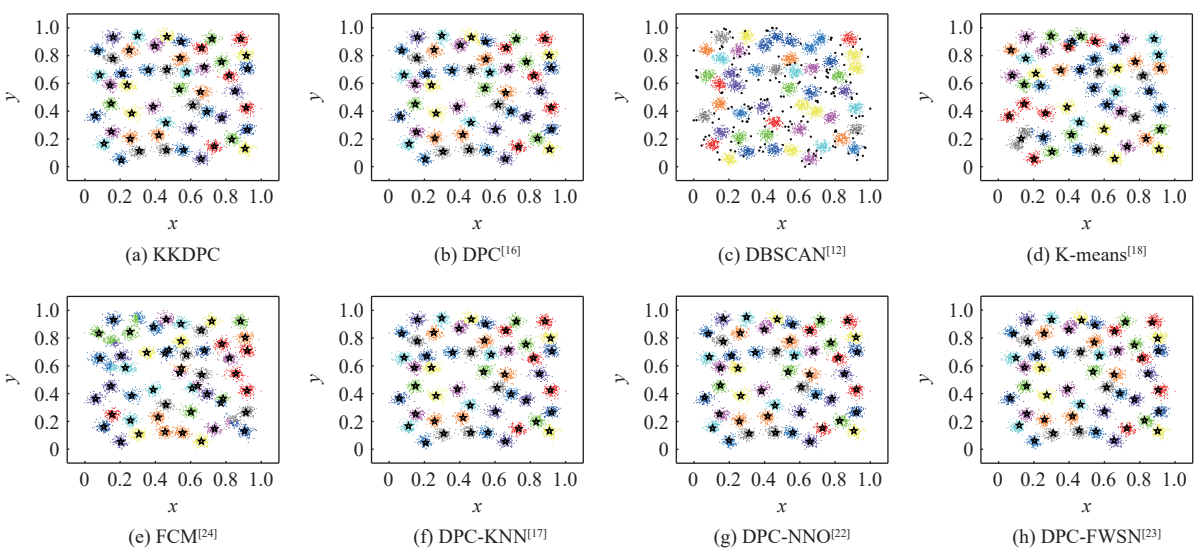
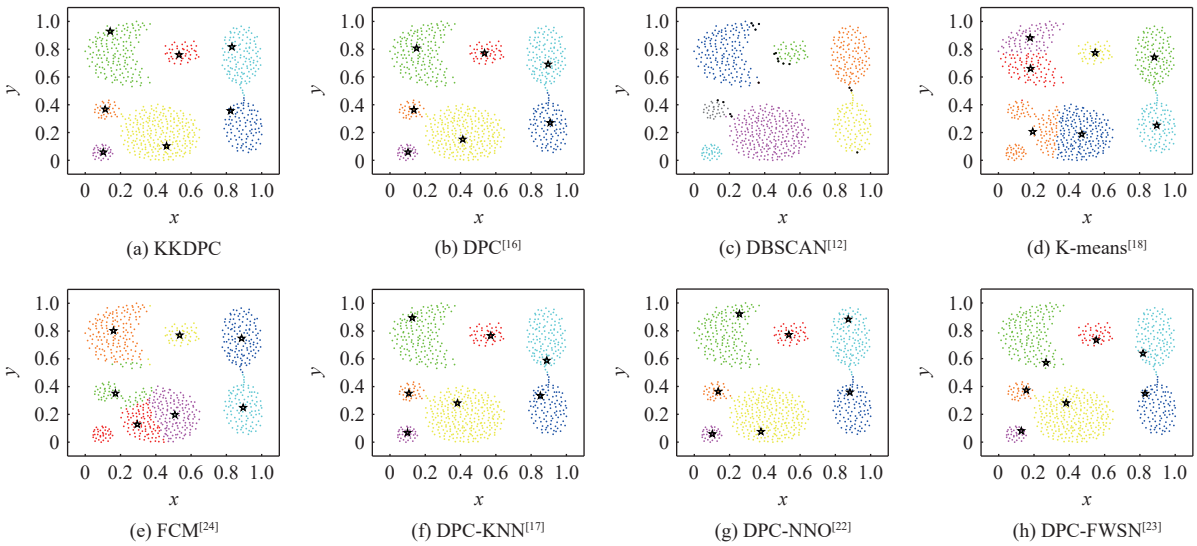


图 11 A3 数据集[27] 聚类结果

Fig. 11 A3 dataset[27] clustering results

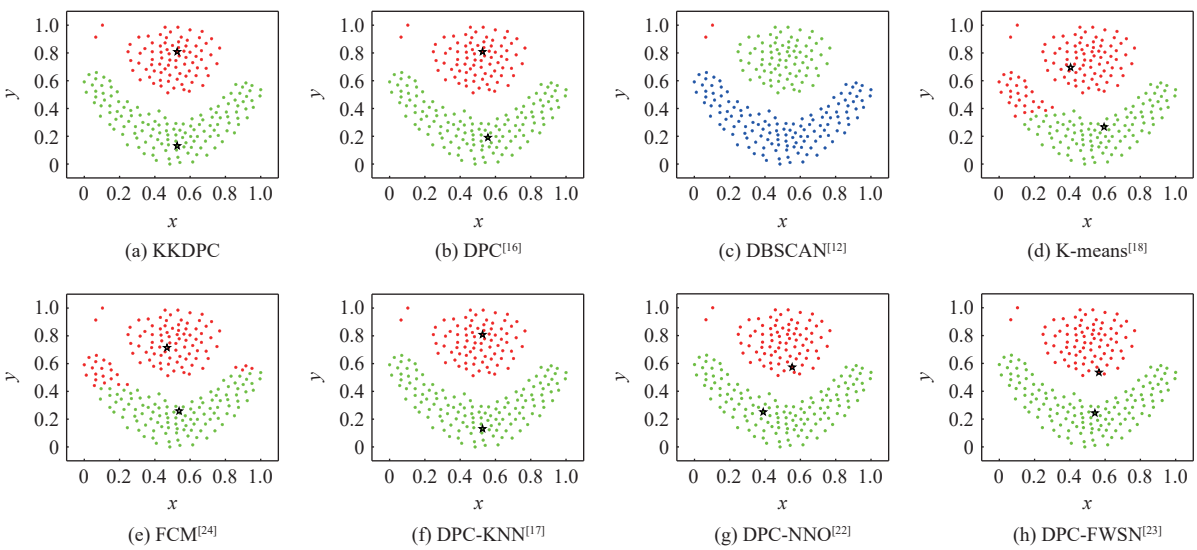


(a) KKDPC　(b) DPC[16]　(c) DBSCAN[12]　(d) K-means[18]

(e) FCM[24]　(f) DPC-KNN[17]　(g) DPC-NNO[22]　(h) DPC-FWSN[23]

图 12　Aggregation 数据集[28] 聚类结果

Fig. 12　Aggregation dataset[28] clustering results

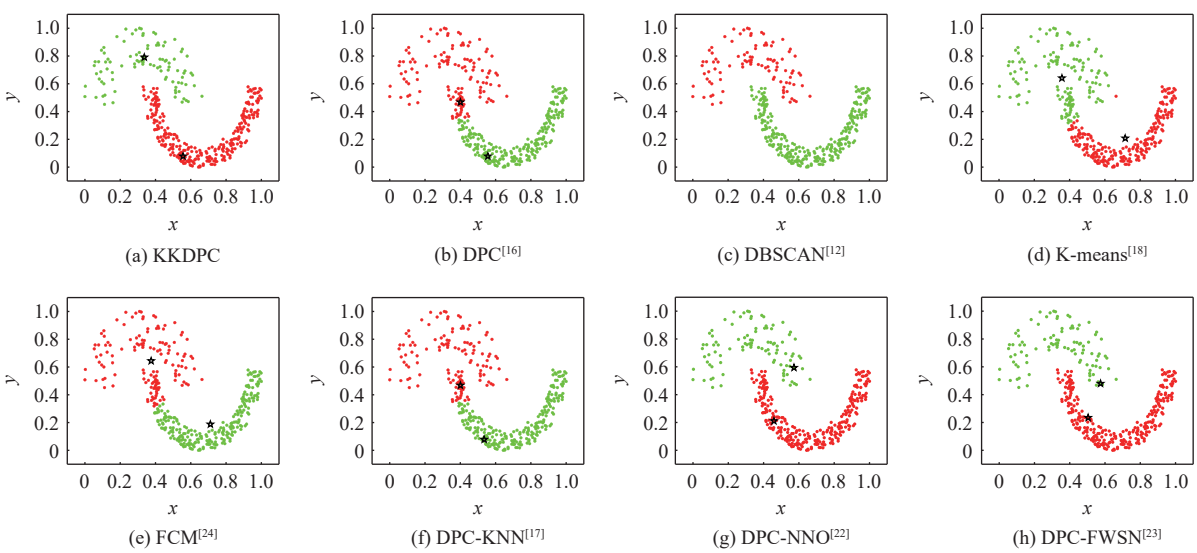


(a) KKDPC　(b) DPC[16]　(c) DBSCAN[12]　(d) K-means[18]

(e) FCM[24]　(f) DPC-KNN[17]　(g) DPC-NNO[22]　(h) DPC-FWSN[23]

图 13　Flame 数据集[29] 聚类结果

Fig. 13　Flame dataset[29] clustering results



(a) KKDPC　(b) DPC[16]　(c) DBSCAN[12]　(d) K-means[18]

(e) FCM[24]　(f) DPC-KNN[17]　(g) DPC-NNO[22]　(h) DPC-FWSN[23]

图 14　Jain 数据集[30] 聚类结果

Fig. 14　Jain dataset[30] clustering results

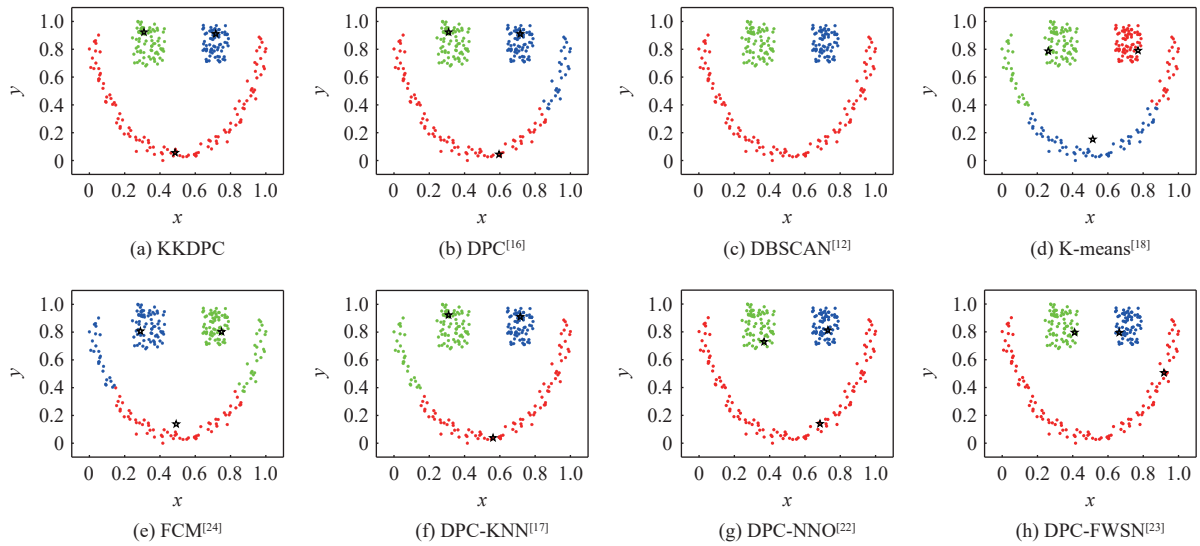


图 15 Line blobs 数据集[31] 聚类结果

Fig. 15 Line blobs dataset[31] clustering results

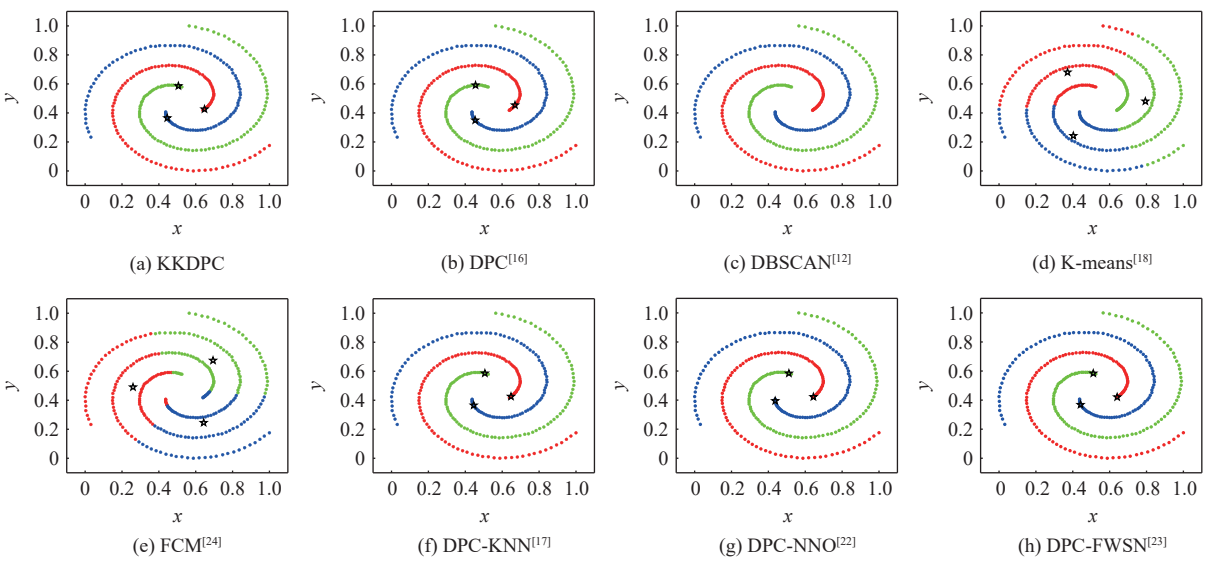


图 16 Spiral 数据集[32] 聚类结果

Fig. 16 Spiral dataset[32] clustering results

数据集的类簇密度不平衡较为常见，而 KKDPC 处理类簇密度不平衡的能力更好。在 Ionosphere、Iris、Seeds、Wine 和 Pendigits 数据集中，DPC-NNO 和 DPC-FWSN 算法优于其他聚类算法，说明优化非中心点的非配策略确实能够在一些数据集中表现出更加出色的聚类结果。

3.3 Olivetti Face 数据集

本节通过 Olivetti Face 数据集[34] 将 KKDPC 算法与 DPC 算法进行对比实验，图 17 为 Olivetti Face 数据集的前 100 组数据。其中，白色数据点为聚类中心，灰色不属于任何簇，相同颜色属于同一类。由图 17 可知，KKDPC 算法准确识别出 9 个聚类中心，而 DPC 算法识别出 8 个聚类中心。相比 DPC 算法，KKDPC 算法有着更高的聚类精度及更好的泛化能力。

表 3 UCI 数据集

Table 3 UCI datasets

UCI数据集	数据量	维度	类簇个数
Blood	748	4	2
Breast	277	9	2
Diabetes	768	8	2
Ecoli	336	7	8
Heart	270	13	2
Ionosphere	351	34	2
Iris	150	4	3
Pima	768	8	2
Seeds	210	7	3
Thyroid	215	5	3
Vehicle	846	18	4
WDBC	569	30	2
Wine	178	13	3
Zoo	101	16	7
Banknote	1 372	4	2
Waveform	5 000	21	3
MFCCs	7 195	22	10
Pendigits	10 992	16	10

表 4　UCI 数据集实验结果

Table 4　Experimental results on UCI datasets

UCI数据集	AMI							
	KKDPC	DPC[16]	DBSCAN[12]	K-means[8]	FCM[24]	DPC-KNN[17]	DPC-NNO[22]	DPC-FWSN[23]
Blood	0.059 1	**0.088 5**	0.008 4	−0.000 8	−0.000 7	0.010 1	0.043 7	0.050 3
Breast	**0.074 1**	−0.001 4	0.001 7	0.074 1	−0.001 4	0.067 2	0.063 4	0.069 6
Diabetes	0.045 0	0.034 1	**0.077 6**	0.051 2	0.062 5	0.001 7	0.044 0	0.002 2
Ecoli	**0.679 8**	0.489 0	0.530 5	0.585 5	0.465 3	0.614 6	0.676 2	0.647 7
Heart	**0.313 3**	0.229 7	0.085 0	0.283 2	0.255 2	0.245 3	0.279 3	0.263 2
Ionosphere	0.321 3	0.087 6	0.212 3	0.129 4	0.124 6	0.150 4	0.366 7	**0.381 1**
Iris	0.862 3	0.862 3	0.604 4	0.733 1	0.737 2	0.862 3	**0.883 1**	**0.883 1**
Pima	0.045 0	0.034 1	**0.077 6**	0.051 2	0.077 1	0.001 7	0.044 0	0.002 2
Seeds	0.723 3	0.717 2	0.541 4	0.661 5	0.678 5	0.706 3	0.773 8	**0.785 5**
Thyroid	0.513 3	0.354 4	0.550 1	0.516 8	0.589 7	0.194 5	**0.668 5**	0.412 3
Vehicle	**0.236 5**	0.164 7	0.163 6	0.158 4	0.094 9	0.173 6	0.193 7	0.181 4
WDBC	**0.779 2**	0.026 7	0.374 9	0.611 0	0.607 6	0.506 9	0.735 5	0.707 6
Wine	0.778 3	0.706 5	0.548 4	0.873 5	0.828 1	0.739 1	**0.888 6**	**0.888 6**
Zoo	**0.899 4**	0.889 6	0.820 6	0.858 1	0.740 9	0.747 1	0.848 7	0.726 2
Banknote	0.755 3	0.931 7	0.762 4	0.016 8	0.032 9	0.923 5	**0.937 7**	0.935 9
Waveform	**0.367 9**	0.240 5	0.012 7	0.364 2	0.330 2	0.356 8	0.362 9	0.355 0
MFCCs	**0.796 1**	0.639 0	0.726 6	0.754 1	0.461 6	0.745 1	0.764 1	0.723 3
Pendigits	0.733 2	0.732 1	0.517 5	0.702 9	0.618 1	0.769 6	**0.808 2**	0.782 1

UCI数据集	ARI							
	KKDPC	DPC[16]	DBSCAN[12]	K-means[8]	FCM[24]	DPC-KNN[17]	DPC-NNO[22]	DPC-FWSN[23]
Blood	0.111 0	**0.196 9**	0.032 8	−0.006 2	−0.007 2	0.031 1	0.026 2	0.065 2
Breast	**0.162 2**	−0.003 1	0.015 9	0.157 0	−0.003 1	0.152 8	0.128 4	0.162 2
Diabetes	0.107 2	0.078 0	**0.148 1**	0.104 0	0.106 9	0.014 3	0.099 5	0.013 1
Ecoli	**0.760 1**	0.561 8	0.649 1	0.537 3	0.368 9	0.734 7	0.734 6	0.741 0
Heart	**0.403 6**	0.305 6	0.018 0	0.366 6	0.331 4	0.322 7	0.366 3	0.348 5
Ionosphere	0.435 0	0.131 7	0.222 7	0.177 6	0.172 7	0.235 7	0.484 0	**0.491 5**
Iris	0.885 7	0.885 7	0.639 7	0.761 3	0.728 7	0.885 7	**0.903 8**	**0.903 8**
Pima	0.107 2	0.078 0	**0.148 1**	0.104 0	0.117 6	0.014 3	0.099 5	0.013 1
Seeds	0.764 1	0.734 1	0.544 9	0.693 4	0.714 9	0.741 9	0.824 6	**0.836 9**
Thyroid	0.599 9	0.425 8	**0.793 2**	0.638 2	0.692 7	0.273 1	0.771 2	0.455 0
Vehicle	**0.157 8**	0.127 3	0.068 2	0.109 2	0.074 5	0.109 3	0.155 2	0.136 9
WDBC	**0.870 2**	0.004 8	0.494 1	0.730 2	0.730 5	0.577 3	0.830 6	0.811 3
Wine	0.786 9	0.672 4	0.529 2	0.899 2	0.849 8	0.726 9	**0.914 9**	**0.914 9**
Zoo	**0.957 0**	0.957 0	0.807 4	0.894 2	0.644 4	0.622 4	0.926 0	0.770 6
Banknote	0.837 9	0.962 4	0.826 6	0.022 3	0.045 2	0.956 7	**0.962 4**	0.965 3
Waveform	**0.316 1**	0.226 7	0.003 6	0.253 6	0.243 6	0.305 7	0.297 7	0.303 2
MFCCs	**0.905 3**	0.729 2	0.855 4	0.895 6	0.208 0	0.841 1	0.884 2	0.813 3
Pendigits	0.579 0	0.635 0	0.598 0	0.623 3	0.529 1	0.623 8	**0.700 9**	0.6645

UCI数据集	NMI							
	KKDPC	DPC[16]	DBSCAN[12]	K-means[8]	FCM[24]	DPC-KNN[17]	DPC-NNO[22]	DPC-FWSN[23]
Blood	0.067 0	**0.092 7**	0.026 1	0.000 3	0.000 4	0.035 0	0.048 3	0.057 5
Breast	**0.107 4**	0.001 3	0.017 2	0.078 4	0.001 3	0.071 0	0.107 4	0.096 2
Diabetes	0.062 3	0.035 6	**0.080 6**	0.052 7	0.064 7	0.017 1	0.059 8	0.015 4
Ecoli	**0.739 4**	0.576 1	0.588 9	0.665 3	0.562 0	0.707 7	0.690 3	0.709 1
Heart	**0.316 1**	0.238 1	0.151 3	0.286 2	0.258 3	0.255 0	0.286 8	0.269 6
Ionosphere	0.351 6	0.091 6	0.272 3	0.134 9	0.129 9	0.152 9	0.401 8	**0.415 3**
Iris	0.864 2	0.864 2	0.676 2	0.741 9	0.743 3	0.864 2	**0.885 1**	**0.885 1**
Pima	0.062 3	0.035 6	**0.080 6**	0.052 7	0.080 2	0.017 1	0.059 8	0.015 4
Seeds	0.726 6	0.723 8	0.609 0	0.665 4	0.682 2	0.710 3	0.775 9	**0.787 5**
Thyroid	0.583 3	0.482 5	0.648 7	0.602 7	0.665 7	0.333 3	**0.691 1**	0.502 1
Vehicle	**0.266 8**	0.196 0	0.193 5	0.203 4	0.098 6	0.203 3	0.211 4	0.214 1
WDBC	**0.781 1**	0.056 2	0.377 2	0.623 2	0.615 2	0.547 5	0.747 1	0.718 2
Wine	0.783 8	0.710 4	0.591 8	0.878 2	0.833 6	0.743 5	**0.892 6**	**0.892 6**
Zoo	**0.916 9**	0.911 9	0.856 8	0.889 3	0.815 8	0.804 8	0.899 8	0.795 0
Banknote	0.755 3	0.931 7	0.762 4	0.017 4	0.032 9	0.923 5	**0.931 7**	0.935 9
Waveform	**0.367 9**	0.240 5	0.344 9	0.364 2	0.330 2	0.356 8	0.362 9	0.355 0
MFCCs	**0.796 1**	0.639 0	0.726 6	0.754 1	0.461 6	0.745 1	0.764 1	0.723 3
Pendigits	0.773 9	0.754 4	0.699 0	0.706 5	0.639 3	0.796 7	**0.824 9**	0.812 6



(a) DPC[16]算法聚类结果



(b) KKDPC算法聚类结果

图 17 Olivetti Face 数据集[34] 聚类结果

Fig. 17 Olivetti Face dataset[34]clustering results

4 结 论

针对 DPC 算法未能充分考虑数据点的局部特征，导致其在识别密度稀疏簇的聚类中心方面存在不足，本文提出 KKDPC 算法。

1）本文引入 K 近邻思想，通过 K 互近邻方法突出密度稀疏簇的聚类中心；采用核密度估计方法计算数据点的局部核密度，以保留其局部特征。与传统 DPC 算法相比，KKDPC 算法在参数设定上更加简便，并且对数据集中簇的形状和密度具有更好的适应能力。

2）在人工数据集和 UCI 数据集中，KKDPC 算法能够有效处理密度分布不均的数据集，并且在复杂 UCI 数据集上的表现优于其他对比算法。

在未来研究中，可以通过优化非中心分配策略进一步提升聚类精度。

参考文献（References）

[1] ZUO W D, HOU X M. An improved probability propagation algorithm for density peak clustering based on natural nearest neighborhood[J]. Array, 2022, 15: 100232.

[2] KHATER I M, NABI I R, HAMARNEH G. A review of super-resolution single-molecule localization microscopy cluster analysis and quantification methods[J]. Patterns, 2020, 1(3): 100038.

[3] GRIFFIÉ J, BURN G L, WILLIAMSON D J, et al. Dynamic Bayesian cluster analysis of live-cell single molecule localization microscopy datasets[J]. Small Methods, 2018, 2(9): 1800008.

[4] FANG U, LI J X, LU X Q, et al. Robust image clustering via context-aware contrastive graph learning[J]. Pattern Recognition, 2023, 138: 109340.

[5] GUAN J Y, LI S, HE X X, et al. Peak-graph-based fast density peak clustering for image segmentation[J]. IEEE Signal Processing Letters, 2021, 28: 897-901.

[6] HU N, TIAN Z H, LU H, et al. A multiple-kernel clustering based intrusion detection scheme for 5G and IoT networks[J]. International Journal of Machine Learning and Cybernetics, 2021, 12(11): 3129-3144.

[7] 张喜梅, 解滨, 徐童童, 等. 基于反向 K 近邻和密度峰值初始化的加权 Kmeans 聚类入侵检测算法[J]. 南京理工大学学报, 2023, 47(1): 56-65.
ZHANG X M, XIE B, XU T T, et al. Intrusion detection algorithm based on weighted Kmeans clustering with reverse K-nearest neighbor and density peak initialization[J]. Journal of Nanjing University of Science and Technology, 2023, 47(1): 56-65(in Chinese).

[8] MACQUEEN J. Some methods for classification and analysis of multivariate observations[J]. Berkeley Symposium on Mathematical Statistics and Probability, 1967, 1: 281-297.

[9] PARK H S, JUN C H. A simple and fast algorithm for K-medoids clustering[J]. Expert Systems with Applications, 2009, 36(2): 3336-3341.

[10] GUHA S, RASTOGI R, SHIM K. Cure: an efficient clustering algorithm for large databases[J]. Information Systems, 2001, 26(1): 35-58.

[11] ZHANG T, RAMAKRISHNAN R, LIVNY M. BIRCH: an efficient data clustering method for very large databases[J]. ACM SIGMOD Record, 1996, 25(2): 103-114.

[12] ESTER M, KRIEGEL H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. Palo Alto: AAAI Press, 1996: 226-231.

[13] ANKERST M, BREUNIG M M, KRIEGEL H P, et al. OPTICS: ordering points to identify the clustering structure[J]. ACM SIGMOD Record, 1999, 28(2): 49-60.

[14] SHEIKHOLESLAMI G, CHATTERJEE S, ZHANG A D. WaveCluster: a wavelet-based clustering approach for spatial data in very large databases[J]. The VLDB Journal, 2000, 8(3): 289-304.

[15] WANG W, YANG J, MUNTZ R R. STING: a statistical information grid approach to spatial data mining[C]//Proceedings of the 23rd International Conference on Very Large Data Bases. New York: ACM, 1997: 186-195.

[16] RODRIGUEZ A, LAIO A. Clustering by fast search and find of density peaks[J]. Science, 2014, 344(6191): 1492-1496.

[17] DU M J, DING S F, JIA H J. Study on density peaks clustering based on K-nearest neighbors and principal component analysis[J]. Knowledge-Based Systems, 2016, 99: 135-145.

[18] JIANG D, ZANG W K, SUN R, et al. Adaptive density peaks clustering based on K-nearest neighbor and gini coefficient[J]. IEEE Access, 2020, 8: 113900-113917.

[19] YUAN X N, YU H, LIANG J, et al. A novel density peaks clustering algorithm based on K nearest neighbors with adaptive merging strategy[J]. International Journal of Machine Learning and Cybernetics, 2021, 12(10): 2825-2841.

[20] XIE J Y, GAO H C, XIE W X, et al. Robust clustering by detecting density peaks and assigning points based on fuzzy weighted K-nearest neighbors[J]. Information Sciences, 2016, 354: 19-40.

[21] LIU R, WANG H, YU X M. Shared-nearest-neighbor-based cluster-

ing by fast search and find of density peaks[J]. Information Sciences, 2018, 450: 200-226.

[22] 陈蔚昌, 赵嘉, 肖人彬, 等. 面向密度分布不均数据的近邻优化密度峰值聚类算法[J]. 控制与决策, 2024, 39(3): 919-928.
CHEN W C, ZHAO J, XIAO R B, et al. Density peaks clustering algorithm with nearest neighbor optimization for data with uneven density distribution[J]. Control and Decision, 2024, 39(3): 919-928 (in Chinese).

[23] ZHAO J, WANG G, PAN J S, et al. Density peaks clustering algorithm based on fuzzy and weighted shared neighbor for uneven density datasets[J]. Pattern Recognition, 2023, 139: 109406.

[24] BEZDEK J C. Pattern recognition with fuzzy objective function algorithms[M]. New York : Plenum Press, 1981.

[25] DING S F, LI C, XU X, et al. A sampling-based density peaks clustering algorithm for large-scale data[J]. Pattern Recognition, 2023, 136: 109238.

[26] CHENG D D, ZHANG S L, HUANG J L. Dense members of local cores-based density peaks clustering algorithm[J]. Knowledge-Based Systems, 2020, 193: 105454.

[27] KARKKAINEN I, FRANTI P. Dynamic local search algorithm for the clustering problem[M]. Joensuu: University of Joensuu, 2002.

[28] GIONIS A, MANNILA H, TSAPARAS P. Clustering aggregation[J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1): 4.

[29] FU L M, MEDICO E. FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data[J]. BMC Bioinformatics, 2007, 8(1): 3.

[30] JAIN A K, LAW M H C. Data clustering: a user's dilemma[M]// Pattern Recognition and Machine Intelligence. Berlin: Springer, 2005: 1-10.

[31] XU X, DING S F, WANG L J, et al. A robust density peaks clustering algorithm with density-sensitive similarity[J]. Knowledge-Based Systems, 2020, 200: 106028.

[32] CHANG H, YEUNG D Y. Robust path-based spectral clustering[J]. Pattern Recognition, 2008, 41(1): 191-203.

[33] BACHE K, LICHMAN M. UCI machine learning repository [EB/OL]. (2020-12-04)[2023-02-01]. http://archive.ics.uci.edu/ml.

[34] SAMARIA F S, HARTER A C. Parameterisation of a stochastic model for human face identification[C]// Proceedings of The IEEE Workshop on Applications of Computer Vision. Piscataway: IEEE Press, 1994: 138-142.

DPC algorithm based on K-reciprocal neighbors and kernel density estimation

ZHOU Yu*，XIA Hao，LIU Hongyu，BAI Lei

(School of Electrical Engineering，North China University of Water Resources and Electric Power，Zhengzhou 450045，China)

Abstract：Clustering by fast search and find of density peaks (DPC) algorithm is a density-based clustering algorithm that does not require iteration or too many parameter settings. However, it fails to identify cluster centers with low cluster density because the local structure of data is not considered when computing local density. To solve this problem, a DPC algorithm based on K-reciprocal neighbors (KN) and kernel density estimation (KDE), called KKDPC was proposed. The number of KN and local kernel density of data points were obtained using the *k*-nearest neighbor and KDE methods. The number of KNs and local kernel density were weighted to obtain the new local density. The relative distance of data points was obtained based on their local density, and cluster centers and non-center points were selected based on the decision graph. Experiments were conducted on artificial and real datasets and compared with seven clustering algorithms including DPC, density-based spatial clustering of applications with noise (DBSCAN), K-means, fuzzy C-means clustering (FCM), DPC based on K nearest neighbors (DPC-KNN) algorithm, DPC with nearest neighbor optimization (DPC-NNO) algorithm, and DPC-FWSN algorithm. The performance of the KKDPC algorithm was verified by calculating the adjusted mutual information (AMI), adjusted Rand index (ARI), and normalized mutual information (NMI). The experimental results show that the proposed KKDPC algorithm can accurately identify cluster centers and improve clustering accuracy effectively.

Keywords：clustering algorithm；density peak；K-nearest neighbor；K-reciprocal neighbor；kernel density estimation

Received：2023-06-12；**Accepted**：2023-10-13；**Published Online**：2023-10-20 10：53

URL：link.cnki.net/urlid/11.2625.V.20231019.1645.003

Foundation items：National Natural Science Foundation of China (U1504622,31671580); Henan Province Young Backbone Teachers Training Program (2018GGJS079)

* **Corresponding author**. E-mail：zhouyu_beijing@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0326

冰晶颗粒在低压压气机中的融化特性

贾惟[1,2,*]，张峰[1]，杨博文[3]，郑淼[3]

(1. 中国民航大学 安全科学与工程学院，天津 300300；2. 中国空气动力研究与发展中心 结冰与防除冰重点实验室，绵阳 621000；
3. 中国民航大学 中欧航空工程师学院，天津 300300)

摘 要：冰晶在压气机内部发生融化，容易造成叶片结冰，深入分析冰晶融化特性对研究压气机冰晶结冰规律具有重要意义。基于压气机一维气动特性，提出一种冰晶融化率的快速计算方法，并进行验证。以某大涵道比涡扇发动机的低压压气机为研究对象，分析冰晶粒径和环境温度对等粒径球形冰晶融化特性的影响。结果表明：冰晶粒径增加导致冰晶融化的起始位置有向低压压气机后面级移动的趋势，冰晶粒径减小或环境温度增加导致冰晶融化速率增大。在此基础上突破等粒径球形冰晶假设，分别考虑冰晶形状和粒径分布，并分析非球形冰晶的球形度和球形冰晶的粒径分布对冰晶融化特性的影响。结果表明：球形度在一定范围内对冰晶融化率有影响，当球形度在0.710～0.958之间时，随着球形度的增大，冰晶融化率减小；相比球形度，非球形冰晶的等效直径对冰晶融化率的影响更为显著；当冰晶的平均体积直径(MVD)不变时，粒径分布的不同导致压气机内同一位置的冰晶融化率偏差最大可达0.205；对于MVD相同而变异系数不同的粒径分布，其融化特性曲线前半部分由小粒径组分冰晶的融化占主导，变异系数较大的分布对应的融化率相对较高，后半部分由大粒径组分冰晶的融化占主导，变异系数较小的分布对应的融化率相对较高。

关 键 词：压气机；融化特性；冰晶融化率；球形度；朗缪尔分布
中图分类号：V2
文献标志码：A **文章编号**：1001-5965(2025)06-1991-13

2013年，一架装有4台GEnx发动机的飞机途经成都附近上空约12 km高度飞行时，由于发动机吸入冰晶造成核心机结冰，引起2号发动机喘振停车，1号发动机转速显著下降且无法恢复到原值，同时3台发动机的高压压气机中有多排叶片被冰块打坏[1]。这是一起非常典型的高空吸入冰晶造成发动机推力损失、压气机结构损伤的事件。据统计，1990—2014年之间发生了62起发动机冰晶结冰事件，其中，大约90%发生在北纬38°和南纬38°之间，中国东南沿海地区是发动机冰晶结冰发生率相对较高的地区[2]。

研究表明，冰晶结冰与传统的过冷水结冰明显不同[3]。冰晶结冰发生的平均温度为−36 ℃，平均高度约为11 km，处于对流云层上部，一般对应飞机的巡航状态，而过冷水结冰很少发生在7 km以上的高空。传统的过冷水结冰发生在发动机整流帽罩、进口支板和出口导叶等位置，这些区域的结冰可提前预防。冰晶结冰主要发生在压气机内部，不容易探测，也无法提前预防，即使飞行员发现压气机结冰，也很难采取有效措施，因此，其危害远大于过冷水结冰。冰晶结冰可导致压气机有效流通面积减小、内部流场紊乱，进而引起发动机转速降低、推力下降，甚至是发动机喘振。脱落的积冰还可能造成压气机叶片断裂、燃烧室熄火等一系列危

收稿日期：2023-06-08；**录用日期**：2023-09-13；**网络出版时间**：2024-04-03 20:08
网络出版地址：link.cnki.net/urlid/11.2625.V.20240403.1509.008
基金项目：国家自然科学基金(12102457)；中国空气动力研究与发展中心结冰与防除冰重点实验室开放课题(IADL20210303)
*通信作者. E-mail：caucjiawei@163.com

引用格式：贾惟，张峰，杨博文，等. 冰晶颗粒在低压压气机中的融化特性[J]. 北京航空航天大学学报，2025，51(6)：1991-2003.
JIA W，ZHANG F，YANG B W，et al. Ice crystal particles melting characteristics in a low-pressure compressor[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：1991-2003 (in Chinese).

害性后果，严重影响飞行安全[4-5]。因此，评估压气机冰晶结冰风险是适航符合性验证的一项重要工作。2011年，美国和欧盟系统开展了高空冰晶结冰研究工作[6-7]。

Mason等[8]总结并分析了46起发动机推力损失事件后指出，冰晶进入压气机后，由于周围空气温度高于冰点，部分冰晶融化，在叶片表面形成水膜。随后冰晶撞击到叶片表面时，由于水膜的存在，部分冰晶“黏附”在表面并发生热量交换。由于不断有冰晶通过撞击黏附在叶片表面，与周围水膜及叶片发生热量交换，温度不断降低。当叶片表面温度低于冰点时，水开始凝固成冰[9]。Hauk等[10]使用空客公司的结冰研究设备开展冰晶撞击刚性平面的实验，冰晶经颗粒喷射模块加速至设定值后撞击到目标铝板上，由高速摄像机和压力传感器阵列记录结果，实验显示真实冰晶撞击后会出现破碎现象，产生粒径更小的不规则冰晶颗粒。Veres等[11]以Honeywell涡扇发动机(LF01)为研究对象开展整机实验，模拟发生推力损失的环境工况，研究压气机的结冰风险参数和结冰区域。他们发现，当冰晶平均体积直径(median volume diameter, MVD)在1～9.5 μm范围内时，第2级静子表面发生积冰黏附。Ashlie[12]使用Honeywell经认证的实验发动机为研究对象开展了整机实验，发现设备入口处不同的冰晶粒径会导致不同的积冰特性，冰晶积冰量随着MVD的增大而减小，当MVD较大时则不会发生黏附积冰现象。Villedieu等[13]开发了ONERA 2D工具包，考虑了真实冰晶球形度对冰晶融化相变和轨迹的影响，并用美国国家航空航天局(National Aeronautics and Space Administration, NASA)的实验结果对计算结果进行了校验。Iuliano等[14]提出冰晶粒子球形度模型，该模型以圆柱体、方柱体和球体等冰晶主要外形特征为研究对象，使用球形度作为分析冰晶不规则程度的重要参数，以球形度为核心构建不规则冰晶融化相变的模化方法，并被Norde[15]采用。Al-Khakik等[16]采用与过冷大水滴具有类似粒径分布的冰晶进行了实验，并对不同条件下的冰形进行了研究。

在国内，姜飞飞等模拟了低压压气机中冰晶的运动轨迹和相变过程[17]。谭燕基于Messinger模型对某楔形体冰晶结冰进行了数值模拟[18]。卜雪琴等基于改进的Messinger模型对NACA0012冰晶结冰情况进行了模拟[19]。郭向东等发展了基于欧拉法的冰晶运动和传热传质耦合计算方法，模拟了大型结冰风洞构型内冰晶运动和传热过程，从沉降收缩、动量平衡和热平衡3个方面研究了颗粒形状和体积密度的影响[20]。黄平等对结冰及混合相结冰问题进行了综述并指出，冰晶运动融化相变、冰晶黏附、侵蚀和结冰过程是冰晶结冰数值模拟的关键问题[21]。马乙楗等基于NNWICE开发了计算程序，系统研究了翼面结冰过程中的冰晶相变和黏附特性[22]。

综上可以看出，对于冰晶结冰问题，国内的研究大多是在外流环境中针对经典翼型开展数值模拟工作，对压气机内流环境中冰晶融化相变规律的研究相对较少。为能够定量分析压气机冰晶融化特性，本文提出一种压气机冰晶融化特性快速分析方法。在此基础上，分析了冰晶粒径、环境温度、冰晶球形度及冰晶粒径分布对冰晶融化特性的影响。

1 数学模型

为了能够对压气机内冰晶融化特性进行快速的定量分析，基于压气机一维气动性能，建立了冰晶融化率数学模型。

1.1 压气机气动模型

冰晶的融化过程与压气机的气动参数紧密相关，如气流的压力和温度等。为快速计算压气机的气动性能，采用一维平均线模型，忽略气流在径向和周向的变化，用平均半径处基元叶栅的气动参数表征整个叶片排的性能[23]。研究表明，只要选用的损失模型和落后角模型合适，就可以获得与实验吻合较好的结果[24]。图1为压气机转子进、出口的速度三角形。其中，下标1表示转子进口，下标2表示转子出口，α为绝对气流角，β为相对气流角。

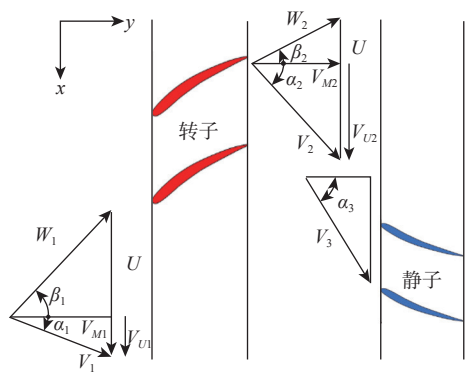


图1 速度三角形

Fig. 1 Velocity triangle

转子进口相对气流角β_1和转子进口相对速度W_1可以分别表示为

$$\beta_1 = \arctan\left(\frac{U - V_{U1}}{V_{M1}}\right) \tag{1}$$

$$W_1 = \frac{V_{M1}}{\cos\beta_1} \tag{2}$$

式中: U 为转子的周向速度; V_{U1} 和 V_{M1} 分别为转子进口绝对速度的周向分量和轴向分量, m/s。

转子出口速度的轴向分量 V_{M2} 和周向分量 V_{U2} 可以分别表示为

$$V_{M2}=\frac{\dot{m}T_{S2}R}{P_{S2}A_2} \tag{3}$$

$$V_{U2}=U-V_{M2}\tan(\beta_{B2}+\delta) \tag{4}$$

式中: A_2 为转子出口面积, m^2; $\dot{m}$为空气的质量流量, kg/s; T_{S2} 为转子出口的静温, K; P_{S2} 为转子出口的静压, Pa; β_{B2} 为转子叶片出口构造角。δ 为叶片出口的落后角, 其表达式为[25]

$$\delta=\frac{\left(0.092f_{\max}^2-0.002\beta_{B2}+0.18\right)K_d}{\sqrt{\sigma}(\theta-0.002K_d)} \tag{5}$$

式中: $f_{\max}$ 为叶型中线最大弯度点相对坐标; σ 为叶片稠度; θ 为叶型弯角。K_d 为修正系数, 其表达式为

$$K_d=K_xK_cK_b \tag{6}$$

式中: K_x 为叶型修正系数, 针对本文选取的叶型, K_x 取 1.1; K_c 为速度修正系数, 考虑了叶片进口和出口轴向速度的变化对落后角的影响; K_b 为角度修正系数, 考虑出口气流角对落后角的影响。针对不同的进口轮毂比 D_{ratio}, 速度修正系数的表达式为

$$K_c=\begin{cases}1 & D_{ratio}\leqslant 0.4\\ \dfrac{1.3(D_{ratio}-0.4)}{V_{M1}r_{1T}/V_{M2}r_{2T}-0.545}-2.76D_{ratio}+2.145 & 0.4<D_{ratio}\leqslant 0.75\\ \dfrac{0.455}{V_{M1}r_{1T}/V_{M2}r_{2T}-0.545} & D_{ratio}>0.75\end{cases} \tag{7}$$

式中: r_T 和 r_H 分别为叶尖和叶根半径。当出口气流角小于 60°时, 角度修正系数表达式为

$$K_b=1+0.001\,5(60-\beta_2)+0.000\,3(60-\beta_2)^2 \tag{8}$$

由于只考虑压气机平均半径处的流动, 因此, 忽略二次流和叶尖泄漏等损失, 主要考虑叶型损失。叶型损失系数 Ω 表达式如下[26]:

$$\Omega=\frac{2\sigma}{\cos\beta_2}\left(\frac{\cos\beta_1}{\cos\beta_2}\right)^2\frac{\xi}{C}\cdot\left\{\left(\frac{2H}{3H-1}\right)\Big/\left[1-\frac{\xi}{C}\cdot\frac{\sigma H}{\cos\beta_2}\right]^3\right\} \tag{9}$$

式中: C 为弦长; ξ 和 H 分别为叶片尾迹的动量厚度和形状因子。叶型损失系数由马赫数、雷诺数和流道收敛比修正得到。

1.2　冰晶融化率模型

冰晶通过吸热使温度达到冰点后发生融化。对每个冰晶颗粒, 考虑质量守恒定律和 Fick 扩散定律, 可以得到

$$\frac{dm}{dt}=Sh\pi D\rho_aD_v(\omega_\infty-\omega_s) \tag{10}$$

式中: Sh 为 Sherwood 数; D_v 为水蒸汽在空气中的扩散系数, m^2/s; ρ_a 为空气密度, kg/m^3; D 为冰晶直径, m; ω 为质量分数, 是相对湿度和水蒸气分压的函数, ω_∞为自由来流的水蒸气质量分数, ω_s为冰晶表面的水蒸气质量分数。

对于固定直径的球形冰晶, 可得

$$\frac{dm}{dt}=\frac{d}{dt}\left(\rho_i\frac{\pi D^3}{6}\right)=\frac{\pi}{2}\rho_iD^2\frac{dD}{dt} \tag{11}$$

式中: ρ_i 为冰晶的密度, kg/m^3。联立式(10)和式(11)可得

$$D\frac{dD}{dt}=-\frac{2Sh\rho_aD_v}{\rho_i}(\omega_s-\omega_\infty) \tag{12}$$

对式(12)积分可得

$$d^2=d_0^2-\frac{4Sh\rho_aD_v}{\rho_iV_\infty}(\omega_s-\omega_\infty)(x-x_0) \tag{13}$$

式中: V_∞为气流的速度, m/s; $(x-x_0)$为当前站点和前一站点之间的距离, m。

考虑冰晶焓的增加主要来源于空气与冰晶之间的换热及固态冰升华所吸收的热量。根据能量守恒关系可得

$$\frac{\partial H_i}{\partial t}=\dot{q}_{conv}+\dot{q}_{sub} \tag{14}$$

式中: H_i 为冰晶的焓; $\dot{q}_{conv}$为空气与冰晶之间交换的热量; $\dot{q}_{sub}$为冰晶升华从空气中吸收的热量。

考虑球形冰晶可得

$$\rho_ic_{pi}\frac{\pi D^3}{6}\cdot\frac{dT_i}{dt}=h\pi D^2(T_a-T_i)+\pi D^2\rho_ah_mL_{sub}(\omega_\infty-\omega_s) \tag{15}$$

式中: L_{sub} 为升华潜热, J/kg; h 为对流换热系数; h_m 为传质系数; c_{pi} 为冰晶的定压比热容, $(J\cdot kg^{-1})/K$; T_a 为空气的温度, K; T_i 为冰晶的温度, K。

热响应时间是描述冰晶对空气温度场变化响应能力的物理参数[27]。根据空气与冰晶之间的换热关系可得

$$mc_{pi}\frac{dT_i}{dt}=Nu\pi k_aD(T_a-T_i) \tag{16}$$

式中: Nu 为 Nusselt 数; k_a 为空气的导热系数, $(W\cdot m^{-1})/K$。

考虑球形冰晶可得

$$\frac{dT_i}{dt}=Nu\frac{6k_a}{\rho_ic_{pi}D^2}(T_a-T_i) \tag{17}$$

定义热响应时间 τ_T 为

$$\tau_T=\frac{\rho_ic_{pi}D^2}{6k_a} \tag{18}$$

当冰晶温度未达到冰点时，从空气中吸收的热量仅用于使冰晶温度上升，此时可以认为冰晶的直径保持不变[15]。考虑 Pr 和 Sc 的定义，可以得到冰晶融化前的温度 T_i 随轴向运动距离 x 的关系式：

$$\frac{\mathrm{d}T_i}{\mathrm{d}x}=\frac{Nu}{\tau_T V_\infty}(T_a-T_i)+\frac{L_{sub}Sh}{\tau_T c_{pa}V_\infty}\cdot\frac{Pr}{Sc}(\omega_\infty-\omega_s) \tag{19}$$

对于球形冰晶，Nu 和 Sh 可以分别表示为[28]

$$Nu=2+0.6Re^{1/2}Pr^{1/3} \tag{20}$$

$$Sh=2+0.6Re^{1/2}Sc^{1/3} \tag{21}$$

当温度达到冰点后冰晶开始融化，生成均匀覆盖冰晶表面的水膜。在冰晶完成融化前冰晶温度保持不变，冰晶（冰水混合物）吸收的热量仅用于使冰晶融化，式（19）可以改写为

$$\frac{\mathrm{d}\psi}{\mathrm{d}x}=\frac{c_{pi}}{L_f}\cdot\frac{Nu}{\tau_T V_\infty}(T_a-T_m)+\frac{c_{pi}}{L_f}\cdot\frac{L_{sub}Sh}{\tau_T c_{pa}V_\infty}\cdot\frac{Pr}{Sc}(\omega_\infty-\omega_m) \tag{22}$$

式中：ψ 为冰晶融化率，是一个表示冰晶融化程度的无量纲参数；L_f 为融化潜热，J/kg。

2 求解方法与验证

2.1 计算站设置

轴流压气机通常为多级结构，一级压气机包括一排转子、一排静子和2个轴向间隙。对压气机沿流向设置计算站点如图2所示，包括转子进口1站、转子出口1站、静子进口1站和静子出口1站，一级压气机共设置了4个计算站。

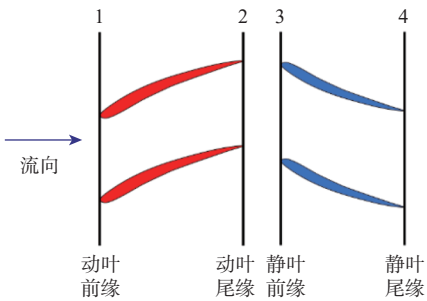


图2 计算站

Fig. 2 Calculation stations

通过假设下一站点的静温，结合进口的总温和总压，可以得到马赫数，根据静温、马赫数及叶片几何可以得到出口速度三角形。采用空间推进的方法，沿轴向依次对每个站点进行计算，直到完成一级的计算。

2.2 求解方法

给定压气机几何和进口条件，沿轴向求解计算站的气动参数。冰晶通过压气机时不断从周围的空气中吸收热量，温度升高导致发生冰晶升华、融化和蒸发等相变的现象。空气与冰晶之间的热量传递可以表示为焓的变化，冰晶焓的增加等于空气焓的减少。

考虑转子通道中冰晶与空气的能量关系，对叶轮机欧拉方程进行改写，可得

$$c_{pa}(T-T_0)=(UV_U-U_0V_{U0})-\Delta H_i \tag{23}$$

静子通道和轴向间隙中冰晶与空气的能量关系可以表示为

$$c_{pa}(T-T_0)=\Delta H_i \tag{24}$$

式中：下标0表示前一站点，无下标表示当前站点；ΔH_i 为冰晶焓的变化量，考虑了融化前冰晶的升华和融化中的液态水的蒸发，完全融化后该项等于0。假设站点静温，通过计算压气机性能得到各站点的压力和速度等气动参数。根据冰晶粒径和温度计算冰晶的融化情况，获得气流损失的热量和冰晶融化所吸收的能量。结合给定的冰晶特征参数，同时考虑冰晶与空气的能量交换，最终得到计算站的气动参数和冰晶融化率。按照各个计算站的轴向顺序，气动参数和冰晶融化率的迭代计算会依次遍历各个计算站，最终得到各个计算站的冰晶融化率。图3为数值求解的流程，图中，I_{WC} 为单位体积空气中所含有的冰晶质量。

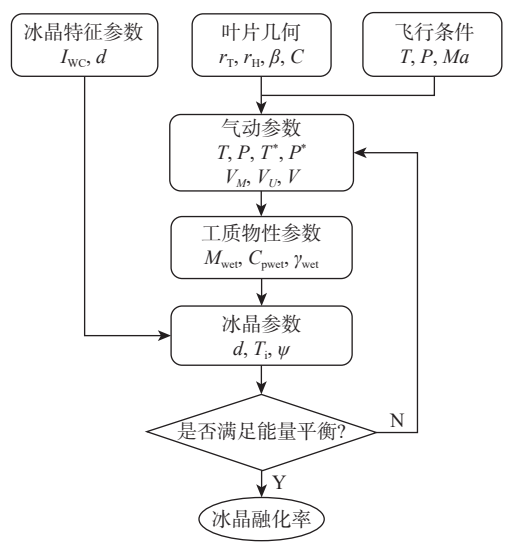


图3 数值求解流程

Fig. 3 Flow of numerical solution

2.3 方法验证

采用C#语言编写计算程序。选取NASA某大涵道比涡扇发动机的低压压气机为对象，对计算程序进行验证，叶片几何详见参考文献[29]。考虑典型的巡航状态：飞行高度 H=10 668 m、环境温度 $T=I_{SA}$+20 K、冰晶粒径 d=5 μm、冰晶含量 I_{WC}=2 g/m^3。其中，I_{SA} 表示国际标准大气条件下的环境温度。利用本文发展的计算程序，可在3 min左右的时间内完成一个工况的计算。

对于多相流问题, 一般用微粒浓度(particulate loading, PL)表示相与相之间的相互作用, 定义如下:

$$P_{\mathrm{L}}=\frac{\alpha_{\mathrm{d}}\rho_{\mathrm{d}}}{\alpha_{\mathrm{c}}\rho_{\mathrm{c}}} \tag{25}$$

式中: 下标 d 表示分散相(即本文中的冰晶), 下标 c 表示载体相(即本文中的空气); 容积分数 α 表示某个相所占有的体积与总体积之比。对于 P_{L}<0.1 的两相流动, 可以认为其是单向耦合的。考虑本文所涉及的工况, P_{L} 值的量级大约为

$$P_{\mathrm{L}}=\frac{\alpha_{\mathrm{d}}\rho_{\mathrm{d}}}{\alpha_{\mathrm{c}}\rho_{\mathrm{c}}}\simeq\frac{0.002}{1\times1}=2\times10^{-3} \tag{26}$$

P_{L} 值远小于临界值 0.1。由此可知, 压气机内的气体与冰晶之间是单向耦合的, 即压气机内的空气流场影响冰晶的运动。实际上, 因为冰晶的密度相对较大, 其运动轨迹会向外径方向偏离。但本文是基于压气机一维气动特性发展的冰晶融化率快速计算方法, 因此, 暂不考虑因冰晶轨迹偏离产生的影响。

图 4 为某低压压气机的冰晶融化率曲线, 并与文献 [30] 结果进行比较。可以看出, 本文得到的冰晶融化率的变化趋势与文献 [30] 结果完全一致, 特别是冰晶融化的起始位置和终止位置与文献 [30] 结果完全重合。同时, 每个计算站的冰晶融化率与文献 [30] 结果基本吻合, 最大相对误差不超过 5%。因此, 本文方法对研究压气机内冰晶融化问题是适用的, 冰晶融化率的计算结果是准确可信的。

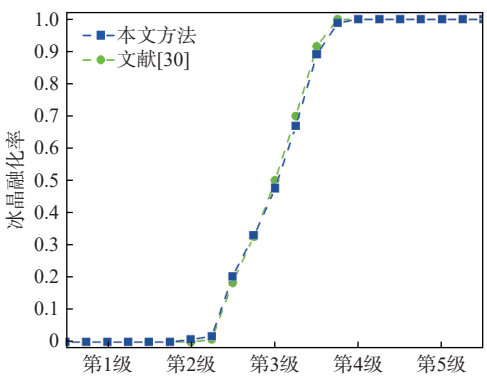


图 4　某低压压气机冰晶融化率

Fig. 4　Ice crystal melting ratio in a low-pressure compressor

3　等粒径球形冰晶的融化特性

3.1　研究对象

选取某大涵道比涡扇发动机的低压压气机为研究对象, 包括 1 级风扇核心(R1 和 S1)和 3 级增压级(R2～S4), 如图 5 所示。其中, 红色表示动叶, 蓝色表示静叶。对于风扇转子叶片, 仅考虑内涵部分。由于增压级最后一级的静子叶片(S5)相对较长, 因此, 在该静子叶片中设置了 3 个计算站。

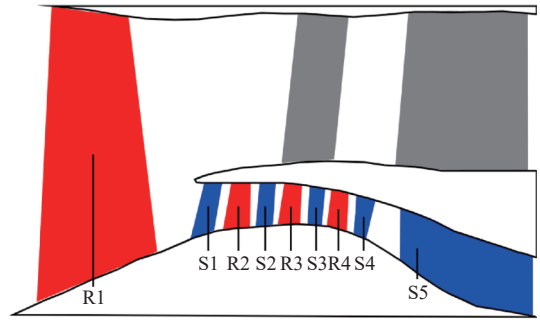


图 5　某大涵道比涡扇发动机低压压气机

Fig. 5　Low-pressure compressor of turbofan engine with large bypass ratio

Mason 等的统计数据表明, 压气机冰晶结冰主要发生在 4 600 m 以上的巡航和下降阶段[8]。FAR33 部附录 D[31] 中强对流天气下冰晶及混合相结冰包线覆盖 1 500～14 000 m 的高度范围。因此, 本节重点考虑巡航阶段和下降阶段, 分别选择巡航阶段的 10 700 m 和下降阶段的 4 700 m 这 2 个典型高度分别进行分析。

3.2　粒径和环境温度对融化特性的影响

已有研究表明, 导致压气机发生结冰的冰晶粒径一般在几微米到几十微米。本节选择冰晶直径为 3, 5, 8, 11, 25 μm 共 5 种工况进行计算。Mason 等指出, 发生冰晶结冰的环境温度一般在 (I_{SA}+10) K～(I_{SA}+20) K 之间[8], 本节考虑环境温度为 (I_{SA}+10) K 和 (I_{SA}+20) K。此外, 考虑不同的冰晶含量并分别进行计算, 本文只给出 I_{WC}=2 g/m^3 的计算结果。巡航阶段的计算工况如表 1 所示。

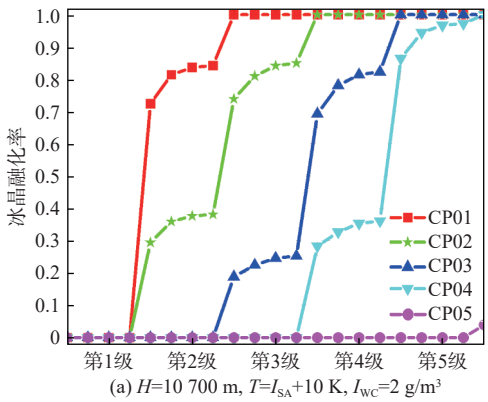
表 1　巡航阶段计算工况

Table 1　Cases for cruising phase

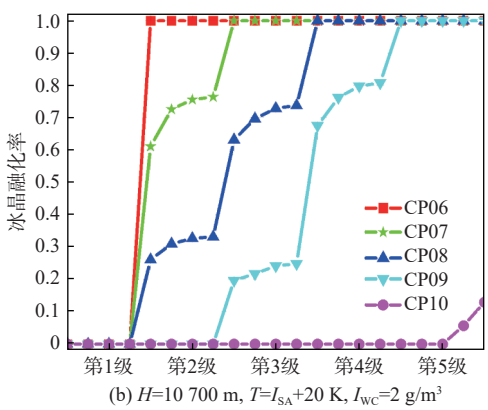
工况	H /m	T /K	d /μm	I_{WC}/(g·m^{-3})
CP01	10 700	I_{SA} +10	3	2
CP02	10 700	I_{SA} +10	5	2
CP03	10 700	I_{SA} +10	8	2
CP04	10 700	I_{SA} +10	11	2
CP05	10 700	I_{SA} +10	25	2
CP06	10 700	I_{SA} +20	3	2
CP07	10 700	I_{SA} +20	5	2
CP08	10 700	I_{SA} +20	8	2
CP09	10 700	I_{SA} +20	11	2
CP10	10 700	I_{SA} +20	25	2

图 6 为巡航阶段不同粒径冰晶的融化特性曲线。可以看出, 随着冰晶粒径的增大, 冰晶颗粒融化的起始位置向后面级移动, 这是由于大粒径冰晶需要从空气中吸收更多的热量才能升温到冰点。对比图 6(a) 和图 6(b) 可以看出, 在同一飞行高度下, 当环境温度升高, 冰晶颗粒融化的起始位置向

前面级移动，同时冰晶融化速率也会加快。由图 6 可知，环境温度上升 10 K 后，粒径为 8 μm 冰晶的融化起始位置从第 2 级提前到第 1 级，粒径为 11 μm 冰晶的融化起始位置从第 3 级提前到第 2 级。这是因为环境温度的升高导致空气与冰晶之间的温差增大、换热量增加。当粒径增大到 25 μm 后可以看到，冰晶在低压压气机内的融化率未超过 0.15。因此可以推测，随着粒径的进一步增大，冰晶颗粒很难在低压压气机内发生融化。



(a) H=10 700 m, T=I_{SA}+10 K, I_{WC}=2 g/m³



(b) H=10 700 m, T=I_{SA}+20 K, I_{WC}=2 g/m³

图 6 巡航阶段不同粒径冰晶融化特性

Fig. 6 Ice crystal melting characteristics for different particle sizes at cruising phase

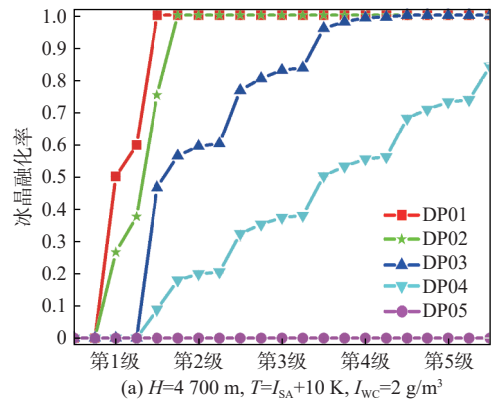
参考巡航阶段，下降阶段考虑冰晶直径为 3,5,8,11,25 μm，环境温度为 I_{SA} +10 K 和 I_{SA} +20 K，冰晶含量 I_{WC} =2 g/m³，具体计算工况如表 2 所示。

图 7 为下降阶段不同粒径冰晶的融化特性曲线。可以看出，在下降阶段，冰晶颗粒融化的起始位置随冰晶粒径的增大向后面级移动，但变化的幅度不如巡航阶段明显。同时，冰晶颗粒的融化速率随着冰晶粒径的增大而降低，这与巡航阶段的变化趋势基本一致。这是因为冰晶粒径的增大会使得冰晶与空气总的换热量降低，粒径增大会延长冰晶升温到冰点的时间，所以冰晶融化速率减小。对比图 7(a) 和图 7(b) 可知，环境温度的增加对冰晶融化的起始位置基本没有影响，但是会显著增加冰晶的融化速率。

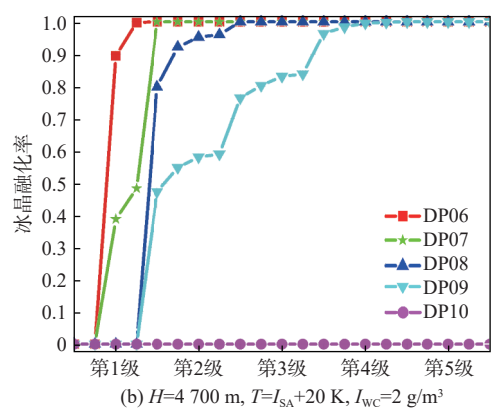
表 2 下降阶段计算工况

Table 2 Cases for descent phase

工况	H/m	T/K	d/μm	I_{WC}/(g·m⁻³)
DP01	4 700	I_{SA} +10	3	2
DP02	4 700	I_{SA} +10	5	2
DP03	4 700	I_{SA} +10	8	2
DP04	4 700	I_{SA} +10	11	2
DP05	4 700	I_{SA} +10	25	2
DP06	4 700	I_{SA} +20	3	2
DP07	4 700	I_{SA} +20	5	2
DP08	4 700	I_{SA} +20	8	2
DP09	4 700	I_{SA} +20	11	2
DP10	4 700	I_{SA} +20	25	2



(a) H=4 700 m, T=I_{SA}+10 K, I_{WC}=2 g/m³



(b) H=4 700 m, T=I_{SA}+20 K, I_{WC}=2 g/m³

图 7 下降阶段不同粒径冰晶融化特性

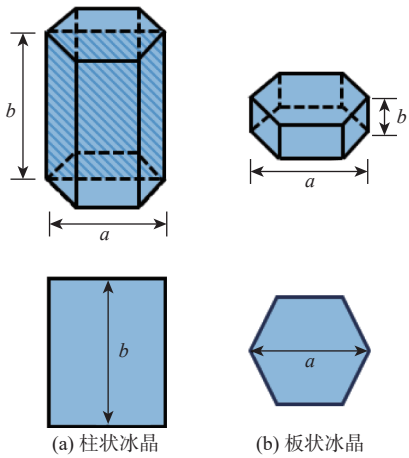
Fig. 7 Ice crystal melting characteristics for different particle sizes at descent phase

4 考虑真实冰晶特征的融化特性

4.1 球形度的影响

在真实的大气环境中，大多数冰晶不会以球形存在。即使是球形冰晶，在通过转子叶片通道时受到转子叶片的“切割”作用后也会发生破碎，生成形状不规则的小粒径冰晶。参考 Iuliano 等[14] 的工作，以球形度为主要参数表征冰晶的不规则程度。统计数据表明，冰晶颗粒的生长机理是以水蒸气凝华为主，以凝华形式增长的冰颗粒通常为单晶结

构, 冰晶形状以柱状和板状为主[32]。因此, 本节考虑的不规则冰晶主要包括柱状和板状 2 种, 如图 8 所示。定义 a 为冰晶颗粒的径向长度, b 为冰晶颗粒的轴向长度。



(a) 柱状冰晶　　(b) 板状冰晶

图 8　柱状和板状冰晶示意图

Fig. 8　Cylindrical and plate shape ice crystals

冰晶颗粒纵横比 E 定义如下:

$$E = b/a \tag{27}$$

定义 d_p 为等效直径, 即与该不规则冰晶的体积相等的球形冰晶的直径, 表达式为

$$d_p = a\sqrt[3]{3E/2} \tag{28}$$

分别定义球形度 Φ 和横向球形度 $\Phi_\perp$ 如下:

$$\Phi = \frac{d_p}{a}\left(\frac{1}{2}+E\right)^{-1} \tag{29}$$

$$\Phi_\perp = \min\left(\left(\frac{d_p}{a}\right)^2, \frac{\pi}{4E}\left(\frac{d_p}{a}\right)^2\right) \tag{30}$$

球形度是描述冰晶的真实形状与球形的接近程度。根据定义可知, 球形度 Φ 越接近于 1, 冰晶越接近球形。不规则的外形使得非球形冰晶与空气的换热过程显著区别于球形冰晶。对于非球形冰晶, 可以使用球形度和横向球形度对 Nusselt 数和 Sherwood 数进行修正, 分别表示为

$$Nu = 1.76 + 0.55\Phi Pr^{\frac{1}{3}}\sqrt{Re}\Phi_\perp^{0.075} + 0.001\,4Pr^{\frac{1}{3}}Re^{\frac{2}{3}}(\Phi/\Phi_\perp)^{7.2} \tag{31}$$

$$Sh = 1.76 + 0.55\Phi Sc^{\frac{1}{3}}\sqrt{Re}\Phi_\perp^{0.075} + 0.001\,4Sc^{\frac{1}{3}}Re^{\frac{2}{3}}(\Phi/\Phi_\perp)^{7.2} \tag{32}$$

本节选择飞行高度为 7 600 m、环境温度为 I_{SA} +10 K、冰晶含量 I_{WC} =2 g/m^3 作为计算条件, 考虑 3 种不同的等效直径。表 3 为不同球形度冰晶的几何参数, 包括等效直径、径向长度、轴向长度和纵横比。其中, NS5-1、NS8-1 和 NS11-1 分别是对应等效直径下的球形冰晶。

表 3　不同球形度计算工况

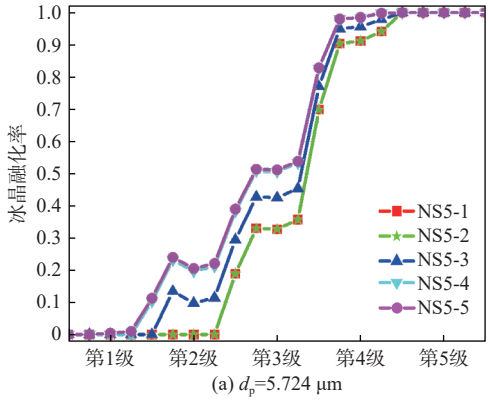
Table 3　Cases for different sphericities

工况	a/μm	b/μm	E	d_p/μm	Φ	$\Phi_\perp$
NS5-1	5.724	5.724	1	5.724	1	
NS5-2	7.5	2.22	0.296	5.724	0.958	0.582
NS5-3	10	1.25	0.125	5.724	0.915	0.327
NS5-4	15	0.555	0.037	5.724	0.710	0.145
NS5-5	20	0.312 5	0.015	5.724	0.555	0.081
NS8-1	8.013	8.013	1	8.013	1	
NS8-2	10.5	3.11	0.296	8.013	0.958	0.582
NS8-3	14	1.75	0.125	8.013	0.915	0.327
NS8-4	21	0.777	0.037	8.013	0.710	0.145
NS8-5	28	0.437 5	0.015	8.013	0.555	0.081
NS11-1	11.447	11.447	1	11.447	1	
NS11-2	15	4.44	0.296	11.447	0.958	0.582
NS11-3	20	2.5	0.125	11.447	0.915	0.327
NS11-4	30	1.11	0.037	11.447	0.710	0.145
NS11-5	40	0.625	0.015	11.447	0.555	0.081

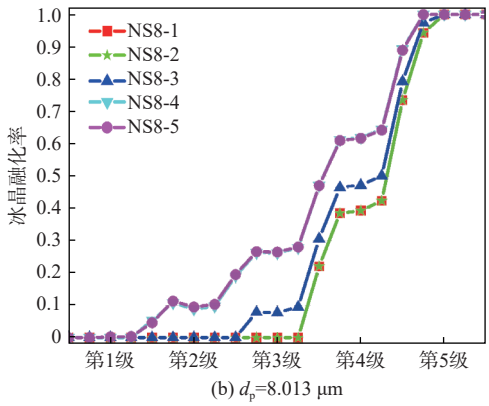
图 9 为不同球形度下冰晶的融化特性曲线。可以看出, 随着球形度和横向球形度的增加, 冰晶颗粒融化的起始位置向后面级移动。以 d_p=5.7 μm 为例, 球形度较小(NS5-4 和 NS5-5)时冰晶颗粒在第 1 级出口开始融化, 球形度略有增大(NS5-3)后冰晶颗粒在第 2 级进口开始融化, 球形度较大(NS5-1 和 NS5-2)时冰晶颗粒在第 2 级出口开始融化。这是因为球形度较小的冰晶表面积相对较大, 吸热速度较快, 相同时间内从空气中吸收的热量较多, 因此, 冰晶会更早的到达冰点。进一步分析可以看出, 同一位置不同球形度冰晶的融化率的变化速率基本一致, 表现为融化率曲线基本平行, 这是因为冰晶与空气的温差是影响冰晶开始融化后融化率曲线变化的主要因素。

对比图 9(a)、图 9(b) 和图 9(c) 可知, 当 Φ=0.958 时, 冰晶融化率曲线与等粒径球形冰晶的融化率曲线几乎重合。因此, 球形度继续增大对冰晶融化率几乎没有影响。而当球形度降低到一定程度之后, 冰晶的表面积较大, 继续降低球形度对表面积的变化率影响不大, 对冰晶与空气换热量的影响有限。对比 Φ=0.710 和 Φ=0.555 这 2 种工况可知, 冰晶融化率曲线几乎重合。因此, 球形度在一定范围内对冰晶融化率有影响。在本文所考虑的工况中, 当 Φ>0.958 和 Φ<0.710 时, 球形度对冰晶融化率的影响并不明显; 当 0.710<Φ<0.958 时, 随着球形度的增大, 冰晶融化率减小。

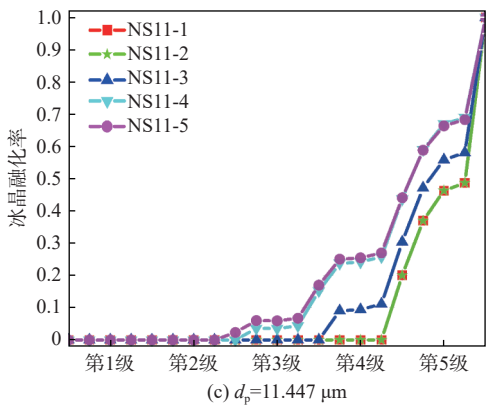
进一步, 给定径向长度并改变轴向长度得到了不同的纵横比, 通过纵横比由小到大的变化模拟板



(a) d_p=5.724 μm



(b) d_p=8.013 μm



(c) d_p=11.447 μm

图 9 不同球形度冰晶融化特性

Fig. 9 Ice crystal melting characteristics for different sphericities

状冰晶向柱状冰晶的变化过程，具体参数如表 4 所示。可以看出，随着纵横比的增加，冰晶等效直径增大、球形度减小。纵横比越接近 1，横向球形度越大。随着纵横比的增加，横向球形度经历了先增大再减小的变化过程。

图 10 为不同纵横比条件下冰晶的融化特性曲线。可以看出，随着纵横比的增加，冰晶的融化的起始位置后移。一方面，冰晶轴向距离的增大使得等效直径增加；另一方面，冰晶轴向距离的增大使得球形度降低，冰晶的表面积增大，换热量增加。这 2 种因素对冰晶融化起始位置的影响趋势是相反的。在冰晶纵横比变化的过程中，这 2 种因素相互制约，但是等效直径的变化起主导作用，所以冰晶融化的起始位置向后移动。因此，相比球形度，

表 4 不同纵横比计算工况

Table 4 Cases for different aspect ratios

工况	a/μm	b/μm	E	d_p/μm	Φ	$\Phi_\perp$
NS1	3	0.5	0.17	3.347 1	0.956	0.397
NS2	3	1.0	0.33	4.542 8	0.909	0.630
NS3	3	3.0	1.0	5.723 5	0.763	1.029
NS4	3	4.5	1.5	6.551 8	0.655	0.899
NS5	3	6.0	2.0	7.211 2	0.577	0.816
NS6	3	9.0	3.0	8.254 8	0.472	0.713
NS7	3	18.0	6.0	10.400 4	0.320	0.566

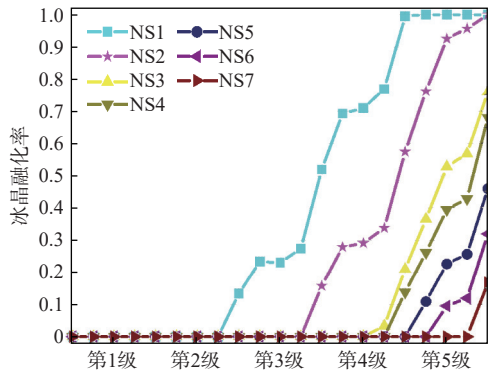


图 10 不同纵横比冰晶融化特性

Fig. 10 Ice crystal melting characteristics for different aspect ratios

非球形冰晶的等效直径对冰晶融化率的影响更显著。

4.2 粒径分布的影响

在真实的大气环境中，冰晶并非呈现出单一的粒径。即使是等粒径冰晶，当通过转子叶片后也会破碎成大小不一的冰晶颗粒。对于包含多种粒径的冰晶群进入压气机后，每种粒径的冰晶都会和压气机中的空气进行单独的热量交换，各部分冰晶交换的热量相加得到总的吸热量，以总吸热量表征冰晶与空气的热量交换。多粒径冰晶群融化相变过程如图 11 所示。

参考过冷水滴常用的 Langmuir 分布[16,33]，构建冰晶的平均体积直径 M_{VD}=7 μm 时冰晶粒径的 Langmuir-B 分布（B 分布）、Langmuir-C 分布（C 分布）、Langmuir-D 分布（D 分布）和 Langmuir-E 分布（E 分布）。每种分布由 7 种不同粒径的冰晶组成，每种粒径的占比是固定的，具体粒径分布如表 5 所

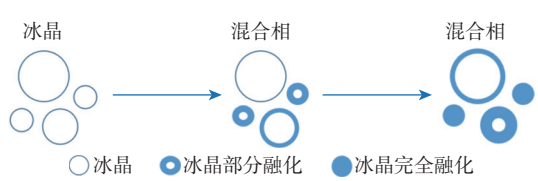


图 11 多粒径冰晶群融化过程示意图

Fig. 11 Melting process of ice crystals with multiple particle sizes

示。以 B 分布为例, 表中第 1 行粒径为 3.92 μm 的冰晶在整个 B 分布的占比为 5%, 对应编码为 B7-1。

表 5　M_{VD}=7 μm 的 Langmuir 分布

Table 5　Langmuir distributions for M_{VD} = 7 μm

占比/%	粒径/μm			
	B分布	C分布	D分布	E分布
5	3.92	2.94	2.17	1.61
10	5.04	4.27	3.64	3.08
20	5.88	5.39	4.97	4.55
30	7.00	7.00	7.00	7.00
20	8.19	8.82	9.59	10.36
10	9.24	10.57	12.18	14
5	10.43	12.67	15.54	18.97

图 12 为 M_{VD}=7 μm 时, 不同粒径分布条件下冰晶的融化特性曲线。以 B 分布为例, 最小粒径冰晶 B7-1 和最大粒径冰晶 B7-7 占比相同。对比图 12(a) 中 B7-1 和 B7-7 的融化特性曲线可以看出, B7-1 对应的冰晶在第 2 级进口处开始融化, 在第 4 级进口处完全融化。B7-7 对应的冰晶在第 4 级进口处开始融化, 在第 5 级出口处完全融化。对于其他分布, 亦可发现类似的规律。在一个粒径符合 Langmuir 分布的多粒径冰晶群中, 不同粒径的冰晶融化速率不同, 小粒径冰晶相比大粒径更快升温到冰点, 融化速度也更快。

对比 B 分布和 E 分布, 最小粒径从 3.92 μm 降低至 1.61 μm, 最大粒径从 10.43 μm 增加到 18.97 μm。对粒径大于 MVD 的组分(简称大粒径组分), 其对应粒径逐渐增大, 而对粒径小于 MVD 的组分(简称小粒径组分), 其对应粒径逐渐减小。因此, 可以认为, E 分布中冰晶粒径比 B 分布更加分散。用变异系数(标准差/均值)表征不同粒径分布的分散程度, 如表 6 所示。变异系数相比方差, 可以消除平均值的影响。可以看出, 随着粒径分散程度的增加, 变异系数增大。

对比 B7-7、C7-7、D7-7 和 E7-7 的融化特性曲线可知, 粒径最小的 B7-7 在第 4 级进口处开始融化, 在低压压气机出口处全部融化, 而粒径最大的 E7-7 在第 5 级中后部才开始融化, 在低压压气机出口处融化率不超过 0.15。因此, MVD 不变的情况下, 粒径分布越分散, 大粒径组分的冰晶粒径更大, 冰晶融化的起始位置后移, 同时小粒径组分的冰晶粒径更小, 冰晶融化的起始位置前移。

由图 12 可知, 变异系数较大的 D 分布和 E 分布相比变异系数较小的 B 分布和 C 分布, 由于小粒径组分的冰晶粒径更小且大粒径组分的冰晶粒径更大, 故小粒径组分的冰晶融化更快, 而大粒径组分的冰晶融化更慢。可以推断, 由于小粒径冰晶吸热升温较快, 融化起始点靠前, 因此, 总融化率曲线前段的变化由小粒径冰晶主导; 大粒径冰晶吸热升

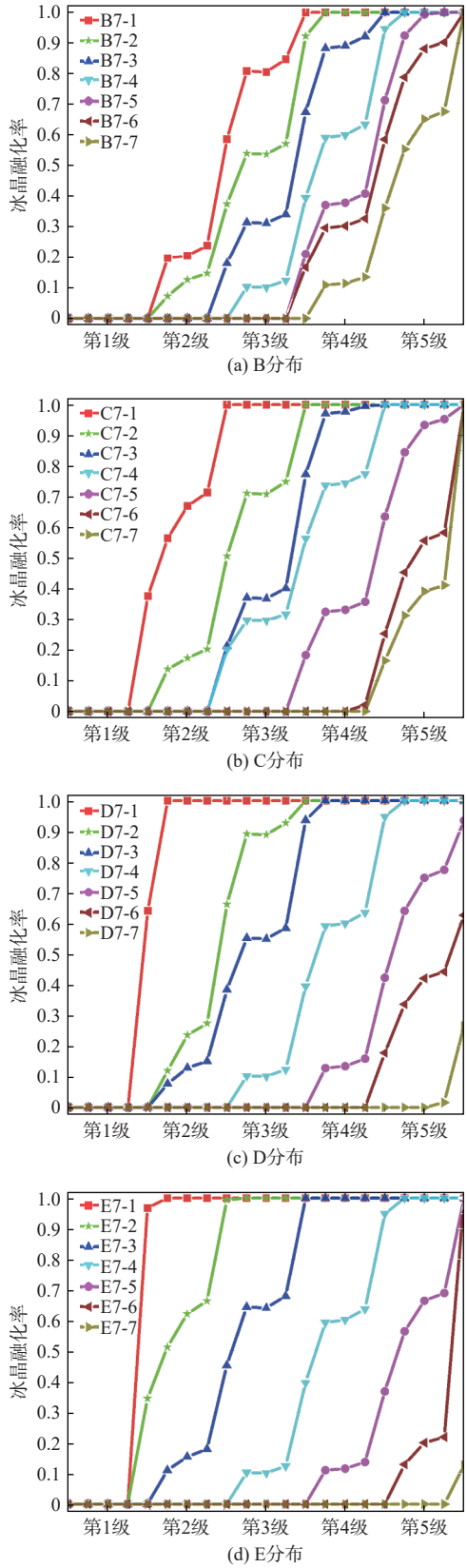


图 12　M_{VD}=7 μm 时, 不同组分冰晶融化特性

Fig. 12　Ice crystal melting characteristics for different components when M_{VD} = 7 μm

表6 Langmuir分布的变异系数

Table 6 Coefficient of variation for Langmuir distribution

B分布	C分布	D分布	E分布
0.22	0.33	0.46	0.61

温较慢，融化起始点靠后，总融化率曲线后段的变化由大粒径冰晶主导。

图13为不同冰晶粒径分布的融化特性曲线。可以看出，在前面3级，变异系数较大的E分布的总融化率相对较大，而变异系数较小的B分布的总融化率相对较小；在后面2级，变异系数较大的E分布的总融化率相对较小，而变异系数较小的B分布的总融化率相对较大。总融化率曲线的变化印证了上述推断。当冰晶融化率超过0.1后，低压压气机可能存在冰晶结冰的风险[33]。由图13可知，E分布的融化率在第2级进口已达到0.1，而B分布的融化率在第3级进口才达到0.1。不同粒径分布下，压气机内的冰晶融化率相差较大。因此，考虑冰晶粒径的分布对开展压气机冰晶结冰风险评估是十分重要的。

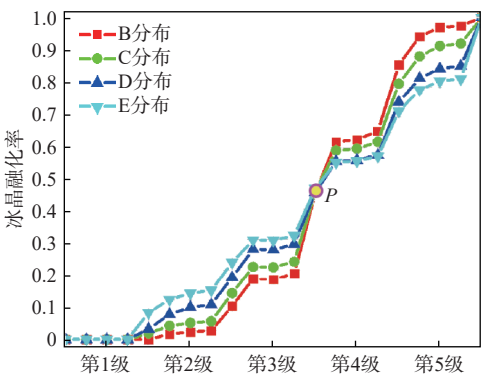


图13 M_{VD}=7 μm时，不同粒径分布的融化特性

Fig. 13 Melting characteristics for different particle size distributions when M_{VD} = 7 μm

选择M_{VD}=5 μm、M_{VD}=10 μm和M_{VD}=15 μm的冰晶群为研究对象，分别考虑4种不同的粒径分布，具体参数如表7～表9所示。以M_{VD}=10 μm的B分布为例，其粒径组分如图14所示。可知，Langmuir分布中粒径组分呈现对称分布。因此，当M_{VD}相同时，不同分布的总融化率曲线总存在一个交叉点(见图13中的P点)。

图15为M_{VD}=5 μm、M_{VD}=10 μm和M_{VD}=15 μm时不同粒径分布下冰晶的融化特性曲线。可以看出，融化刚开始时E分布的冰晶融化率增长较快，而B分布的冰晶融化率增长相对较慢。在交叉点之后，融化率的增长呈现相反的趋势，即E分布的冰晶融化率增长较慢，而B分布的冰晶融化率增长较快，这与M_{VD}=7 μm时的规律完全一致。此外，

表7 M_{VD}=5 μm的Langmuir分布

Table 7 Langmuir distributions for M_{VD} = 5 μm

占比/%	粒径/μm			
	B分布	C分布	D分布	E分布
5	2.8	2.1	1.55	1.15
10	3.6	3.05	2.6	2.2
20	4.2	3.85	3.55	3.25
30	5	5	5	5
20	5.85	6.3	6.85	7.4
10	6.6	7.55	8.7	10
5	7.45	9.05	11.1	13.55

表8 M_{VD}=10 μm的Langmuir分布

Table 8 Langmuir distributions for M_{VD} = 10 μm

占比/%	粒径/μm			
	B分布	C分布	D分布	E分布
5	5.6	4.2	3.1	2.3
10	7.2	6.1	5.2	4.4
20	8.4	7.7	7.1	6.5
30	10	10	10	10
20	11.7	12.6	13.7	14.8
10	13.2	15.1	17.4	20
5	14.9	18.1	22.2	27.1

表9 M_{VD}=15 μm的Langmuir分布

Table 9 Langmuir distributions for M_{VD} = 15 μm

占比/%	粒径/μm			
	B分布	C分布	D分布	E分布
5	8.4	6.3	4.65	3.45
10	10.8	9.15	7.8	6.6
20	12.6	11.55	10.65	9.75
30	15	15	15	15
20	17.55	18.9	20.55	22.2
10	19.8	22.65	26.1	30
5	22.35	27.15	33.3	40.65

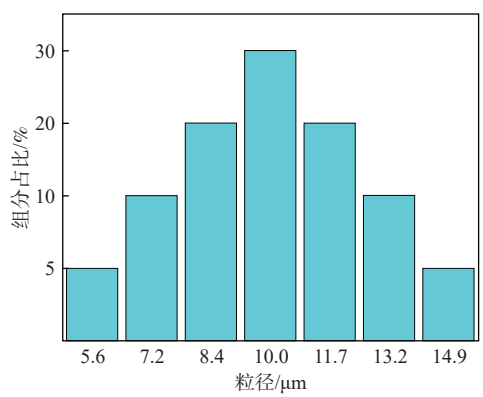


图14 M_{VD}=10 μm时，B分布的粒径组分

Fig. 14 Proportion of components for B distribution when M_{VD} = 10 μm

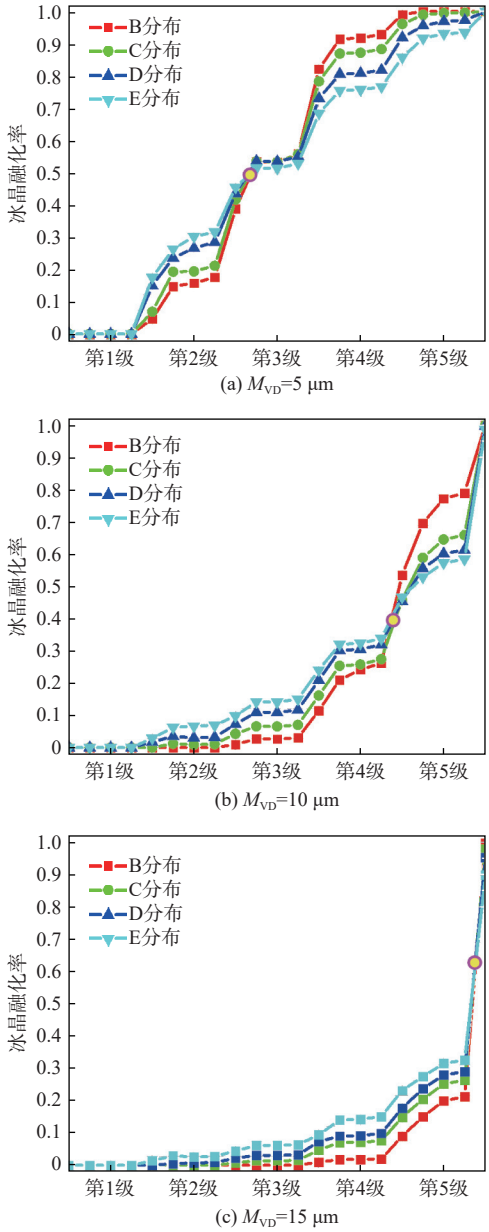


图 15　不同粒径分布的冰晶融化特性

Fig. 15　Ice crystal melting characteristics for different particle size distributions

随着 M_{VD} 的增加, 不同分布对应的交叉点逐渐后移。当 M_{VD} =5 μm 时, 交叉点位于第 3 级进口, 交叉点前不同分布的冰晶融化率最大相差 0.144, 交叉点后不同分布的冰晶融化率最大相差 0.163; 当 M_{VD} =10 μm 时, 交叉点位于第 4 级出口, 交叉点前不同分布的冰晶融化率最大相差 0.126, 交叉点后不同分布的冰晶融化率最大相差 0.205; 当 M_{VD} = 15 μm 时, 交叉点接近低压压气机出口, 交叉点前不同分布的冰晶融化率最大相差 0.142。在交叉点之前, 以小粒径冰晶融化、大粒径冰晶吸热为主, 因此, 融化率曲线的变化是由冰晶群中小粒径组分冰晶的融化主导的。在交叉点之后, 小粒径冰晶已经完全融化或接近完全融化, 大粒径冰晶开始融化, 融化率曲线的变化是由冰晶群中大粒径组分冰晶的融化主导的。M_{VD} 越大, 冰晶群中小粒径组分对应的冰晶粒径更大, 融化的起始点后移、融化率增长变缓, 导致交叉点向后移动。

5　结　论

本文提出一种压气机冰晶融化率快速计算方法, 以典型大涵道比涡扇发动机的低压压气机为研究对象, 分析了冰晶粒径、环境温度、冰晶球形度和冰晶粒径分布对低压压气机内冰晶融化特性的影响规律, 主要结论如下:

1) 对于等粒径球形冰晶, 在相同的环境温度下, 冰晶粒径增加导致冰晶融化的起始位置有向压气机后面级移动的趋势; 冰晶粒径减小或环境温度增大导致冰晶融化速率增加。

2) 对于非球形冰晶, 球形度在一定范围内对冰晶融化率有影响。当 Φ>0.958 和 Φ<0.710 时, 球形度对融化率的影响并不明显; 当 0.710<Φ<0.958 时, 随着球形度的增大, 冰晶融化率减小。当等效直径和球形度同时发生变化时, 等效直径对冰晶融化率的影响更为显著。

3) 粒径分布对冰晶的融化特性的影响显著。MVD 不变时, 粒径分布不同导致压气机内同一位置的冰晶融化率偏差最大可达 0.205。对于 MVD 相等而变异系数不同的对称分布, 其融化特性曲线存在一个交叉点。交叉点前小粒径组分冰晶的融化占主导, 变异系数较大的分布对应的融化率较高; 交叉点后大粒径组分冰晶的融化占主导, 变异系数较小的分布对应的融化率较高。

参考文献 (References)

[1] HRADECKY S. Incidents and news in aviation [EB/OL]. (2013-07-31) [2023-02-01]. https://avherald.com/h?article=46679ba1&opt=0.

[2] BRAVIN M, STRAPP J W, MASON J. An investigation into location and convective lifecycle trends in an ice crystal icing engine event database[C]//Proceedings of the SAE International Conference on Icing of Aircraft, Engines, and Structures. Warrendale: SAE International, 2015.

[3] 沈浩, 韩冰冰, 张丽芬. 航空发动机中冰晶结冰的研究进展[J]. 实验流体力学, 2020, 34(6): 1-7.
SHEN H, HAN B B, ZHANG L F. Research progress of the ice crystal icing in aero-engine[J]. Journal of Experiments in Fluid Mechanics, 2020, 34(6): 1-7(in Chinese).

[4] 袁庆浩, 樊江, 白广忱. 航空发动机内部冰晶结冰研究综述[J]. 推进技术, 2018, 39(12): 2641-2650.
YUAN Q H, FAN J, BAI G C. Review of ice crystal icing in aero-engines[J]. Journal of Propulsion Technology, 2018, 39(12): 2641-2650(in Chinese).

[5] ADDY Jr H E, VERES J P. An overview of NASA engine ice-crys-

tal icing research[C]//Proceedings of the SAE International Conference on Aircraft and Engine Icing and Ground Deicing. Warrendale: SAE International, 2011.

[6] DEZITTER F, GRANDIN A, BRENGUIER J L, et al. HAIC-high altitude ice crystals[C]//Proceedings of the 5th AIAA Atmospheric and Space Environments Conference. Reston: AIAA, 2013.

[7] NAGORSKI M, KOCH C, STAUDACHER S. Boundary conditions for compressor cascade ice crystal icing testing[J]. Journal of Turbomachinery, 2021, 143(10): 101013.

[8] MASON J, STRAPP W, CHOW P. The ice particle threat to engines in flight[C]//Proceedings of the 44th AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2006.

[9] MACLEOD J, FULEKI D. Ice crystal accretion test rig development for a compressor transition duct[C]//Proceedings of the AIAA Atmospheric and Space Environments Conference. Reston: AIAA, 2010.

[10] HAUK T, BONACCURSO E, ROISMAN I V, et al. Ice crystal impact onto a dry solid wall. Particle fragmentation[J]. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2015, 471(2181): 20150399 .

[11] VERES J P, JORGENSON P, JONES S M. Modeling of highly instrumented honeywell turbofan engine tested with ice crystal ingestion in the NASA propulsion system laboratory[C]//Proceedings of the 8th AIAA Atmospheric and Space Environments Conference. Reston: AIAA, 2016.

[12] ASHLIE F. Ice crystal icing investigation on a Honeywell uncertified research engine in an altititude simulation icing facility[J]. Journal of Turbomachinery, 2021, 143(10): 1-22.

[13] VILLEDIEU P, TRONTIN P, CHAUVIN R. Glaciated and mixed phase ice accretion modeling using ONERA 2D icing suite[C]//Proceedings of the 6th AIAA Atmospheric and Space Environments Conference. Reston: AIAA, 2014.

[14] IULIANO E, MONTREUIL E, NORDE E, et al. Modelling of non-spherical particle evolution for ice crystals simulation with an eulerian approach [C]//Proceedings of the SAE International Conference on Icing of Aircraft, Engines, and Structures. Warrendale: SAE International, 2015.

[15] NORDE E. Eulerian method for ice crystal icing in turbofan engines [D]. Enschede: University of Twente, 2017.

[16] AL-KHALIL K, IRANI E, MILLER D. Mixed phase icing simulation and testing at the cox icing wind tunnel[C]//Proceedings of the 41st Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2003.

[17] 姜飞飞, 董威, 郑梅, 等. 冰晶在涡扇发动机内相变换热特性[J]. 航空动力学报, 2019, 34(3): 567-575.
JIANG F F, DONG W, ZHENG M, et al. Phase change heat transfer characteristic of ice crystal ingested into turbofan engine[J]. Journal of Aerospace Power, 2019, 34(3): 567-575(in Chinese).

[18] 谭燕. 基于欧拉方法的2维翼型冰晶结冰数值计算[J]. 航空发动机, 2020, 46(4): 30-35.
TAN Y. Numerical calculation of 2D airfoil ice crystal icing based on Euler method[J]. Aeroengine, 2020, 46(4): 30-35(in Chinese).

[19] 卜雪琴, 李皓, 黄平, 等. 二维机翼混合相结冰数值模拟[J]. 航空学报, 2020, 41(12): 124085.
BU X Q, LI H, HUANG P, et al. Numerical simulation of mixed phase icing on two-dimensional airfoil[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(12): 124085(in Chinese).

[20] 郭向东, 胡站伟, 丁亮, 等. 大型结冰风洞中冰晶热/力平衡特性数值研究[J]. 航空动力学报, 2022, 37(3): 478-491.
GUO X D, HU Z W, DING L, et al. Numerical investigation of thermal and mechanical equilibrium characteristics of ice crystal in large icing wind tunnel[J]. Journal of Aerospace Power, 2022, 37(3): 478-491(in Chinese).

[21] 黄平, 卜雪琴, 刘一鸣, 等. 混合相/冰晶条件下的结冰研究综述[J]. 航空学报, 2022, 43(5): 025178.
HUANG P, BU X Q, LIU Y M, et al. Mixed phase/glaciated ice accretion: Review[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(5): 025178(in Chinese).

[22] 马乙棣, 柴得林, 王强, 等. 翼面结冰过程中的冰晶运动相变与黏附特性[J]. 航空学报, 2023, 44(1): 35-46.
MA Y J, CHAI D L, WANG Q, et al. Phase change and adhesion characteristics of ice crystal movements in wing icing[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(1): 35-46(in Chinese).

[23] 王进, 周玲, 季路成. 轴流压气机一维特性计算方法简介及展望[J]. 实验流体力学, 2021, 35(2): 1-12.
WANG J, ZHOU L, JI L C. Brief introduction and prospect of calculation methods for one-dimensional characteristics of axial flow compressor[J]. Journal of Experiments in Fluid Mechanics, 2021, 35(2): 1-12(in Chinese).

[24] WHITE N M, TOURLIDAKIS A, ELDER R L. Axial compressor performance modelling with a quasi-one-dimensional approach[J]. Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, 2002, 216(2): 181-193.

[25] 张军. 多级轴流压气机方案设计与特性计算研究[D]. 北京: 北京理工大学, 2016.
ZHANG J. Study on scheme design and characteristic calculation of multistage axial compressor[D]. Beijing: Beijing Institute of Technology, 2016(in Chinese).

[26] 张长生, 王小莉. 轴流压气机设计点流动损失的预估和分析[J]. 航空动力学报, 1993, 8(1): 29-33.
ZHANG C S, WANG X L. Estimation of flow loss at design point in an axial-flow compressor[J]. Journal of Aerospace Power, 1993, 8(1): 29-33(in Chinese).

[27] CROWE C T, SCHWARZKOPF J D, SOMMERFELD M, et al. Multiphase flows with droplets and particles[M]. Boca Raton: CRC Press, 2011.

[28] WRIGHT W, JORGENSON P, VERES J. Mixed phase modeling in GlennICE with application to engine icing[C]//Proceedings of the AIAA Atmospheric and Space Environments Conference. Reston: AIAA, 2010.

[29] OLIVER M. Ice crystal icing engine testing in the NASA Glenn research center's propulsion systems laboratory: altitude investigation[J]. SAE International Journal of Aerospace, 2015, 8(1): 33-37.

[30] VERES J, JORGENSON P, WRIGHT W. A model to assess the risk of ice accretion due to ice crystal ingestion in a turbofan engine and its effects on performance[C]//Proceedings of the 4th AIAA Atmospheric and Space Environments Conference. Reston: AIAA, 2012.

[31] Federal Aviation Administration. Appendix D to Part 33-mixed phase and ice crystal icing envelope (deep convective clouds): 79 FR 65538 [S]. Washington, D.C.: Federal Aviation Administration, 2014: 743–745.

[32] PRUPPACHER H R, KLETT J D. Microphysics of clouds and precipitation[M]. Berlin: Springer, 1978.

[33] VERES J P, JORGENSON P C E, JONES S M, et al. Modeling of a turbofan engine with ice crystal ingestion in the NASA propulsion system laboratory[C]//Proceedings of the Turbomachinery Technical Conference and Exposition. New York: ASME, 2017.

Ice crystal particles melting characteristics in a low-pressure compressor

JIA Wei[1, 2, *], ZHANG Feng[1], YANG Bowen[3], ZHENG Miao[3]

(1. College of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China;
2. Key Laboratory of Icing and Anti/De-icing, China Aerodynamics Research and Development Center, Mianyang 621000, China;
3. Sino-European Institute of Aviation Engineering, Civil Aviation University of China, Tianjin 300300, China)

Abstract: The melting of ice crystals in the compressor leads to ice formation on the blades. It is of great significance to analyze the ice crystal melting characteristics to study the ice crystal icing in the compressor. Based on the one-dimensional aerodynamic characteristics of the compressor, a rapid method for calculating the ice crystal melting ratio was developed and validated. A low-pressure compressor of a turbofan engine with a large bypass ratio was selected to study the influence of ice crystal size and ambient temperature on the ice crystal melting characteristics at the same particle size. The results show that the initial position of ice crystal melting tends to move to the rear stage of the low-pressure compressor with the increase in ice crystal size, and the decrease in ice crystal size or the increase in ambient temperature leads to the increase in ice crystal melting ratio. On this basis, the hypothesis of spherical ice crystals with equal particle size is broken through, and the influence of sphericity of non-spherical ice crystals and the particle size distribution of spherical ice crystal on the melting characteristics were analyzed by considering ice crystal shape and particle size distribution. The results demonstrate that the sphericity has an effect on the ice crystal melting ratio within a certain range. When the sphericity is between 0.710 and 0.958, the ice crystal melting ratio decreases with the increase in sphericity. Compared with the sphericity, the equivalent diameter of non-spherical ice crystals has a more significant effect on the ice crystal melting ratio. When the mean volume diameter (MVD) of the ice crystal is constant, the difference in particle size distribution leads to the maximum deviation of 20.5% in the ice crystal melting ratio at the same position in the compressor. For the particle size distributions with the same MVD but different coefficients of variation, in the front half of the ice crystal melting characteristic curve, the melting of small-sized ice crystals is dominant, and the particle size distribution with a larger coefficient of variation has a relatively higher melting ratio. However, in the latter part, the melting of large-sized ice crystals is dominant, and the particle size distribution with a smaller coefficient of variation has a relatively higher melting ratio.

Keywords: compressor; melting characteristic; ice crystal melting ratio; sphericity; Langmuir distribution

Received: 2023-06-08; **Accepted**: 2023-09-13; **Published Online**: 2024-04-03 20:08

URL: link.cnki.net/urlid/11.2625.V.20240403.1509.008

Foundation items: National Natural Science Foundation of China (12102457); Open Fund for Key Laboratory of Icing and Anti/De-icing of China Aerodynamics Research and Development Center (IADL20210303)

* **Corresponding author**. E-mail: caucjiawei@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0339

飞行器大气数据系统构型与解算演变机理

熊亮[1,2,*]，张睿[1]，许斌[1]，黄巧平[2]
(1. 西北工业大学 自动化学院，西安 710129；2. 成都凯天电子股份有限公司 大气数据系统研究室，成都 610091)

摘 要：大气数据传感系统为飞行器飞行控制及投放初始状态给定与轨迹控制提供必要的静压、总压、迎角、侧滑角等参量，飞行器跨代发展诱发了大气数据传感系统参数解算基本要素获取方式由直接型演变为解耦提取。对大气数据传感系统构型形态和发展历程进行总结、提炼，系统分析了大气数据传感系统构型演变原因及过程、大气参数解算模型原理、故障大气参数诊断与检测流程、系统重构模型，总结了未来飞行器大气数据传感系统构型与参数解算设计重点研究方向。

关 键 词：大气数据系统；大气数据系统构型；大气参数解算；容错大气数据系统；大气数据系统重构
中图分类号：V247.1+6
文献标志码：A **文章编号**：1001-5965(2025)06-2004-10

飞行器大气数据传感系统提供的飞行大气参数表征了飞行器运动与来流空气相互关系，实现静压、总压、迎角和侧滑角的测量及进行气压高度、速度等参数计算。飞行器大气数据传感系统构型、参数解算、修正模型与气动外形息息相关[1-2]，具有与载机平台唯一匹配的特点。

源于喷气动力技术突破诞生的一代战斗机及源于跨声速面积率、大后掠/三角翼布局、薄翼型、动力涡喷发动机等一系列技术突破的二代机[3]，飞行大气参数仅作为飞行员对飞机操控的引导；得益于电传飞控、综合化航电技术、加力涡扇发动机技术及隐身技术发展诞生的三代机、四代机[3]，飞行安全控制及精准投放对飞行大气参数的依赖性逐渐增强，机上的飞控系统、发动机控制系统、火控系统、导航系统、环控系统及空管系统等均需要大气数据传感系统提供飞行参数[4]。技术进步与需求牵引推动了飞行器跨代发展，飞行器的作战需求诱发了大气数据传感系统构型及参数解算模型适应性演变。

1 系统构型

国内外的二代、三代、四代战斗机和民用飞机主要采用了基于飞行气压测量的大气数据传感技术，随着平台跨代发展需求牵引，推动了大气数据传感技术呈现出了集中式、分布式和嵌入式3种物理形态。基于激光雷达原理的大气数据测量技术从被动式探测转向主动适应性探测，具有不受载机气动外形影响、无管路迟滞、响应速度快的显著特点，为解决低速甚至负速度下大气参数测量精度下降问题和最优飞行校准基准方案带来了新思路，已引发国内外多家公司和科研机构开展原理研究和验证[5-18]。

集中式大气数据传感系统对静压、总压、迎角和侧滑角4个基本要素直接测量，基于标准大气数据方程进行飞行大气参数计算，压力受感器与压力转换器之间气压传递管路的低通滤波特性[19-21]明显，使得动态响应特性存在滞后。结合风洞试验、地面滑跑及飞行数据进行受感器源位置误差修正

收稿日期：2023-06-12；录用日期：2023-09-11；网络出版时间：2023-10-10 19:32
网络出版地址：link.cnki.net/urlid/11.2625.V.20231009.1254.002
*通信作者. E-mail：caic02253@126.com

引用格式：熊亮，张睿，许斌，等. 飞行器大气数据系统构型与解算演变机理[J]. 北京航空航天大学学报，2025，51(6)：2004-2013.
XIONG L，ZHANG R，XU B，et al. Configuration and calculation evolution mechanism of air data system for aircraft[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2004-2013 (in Chinese).

能够得到较真实的飞行大气参数, 可直接参与飞行控制, 保证飞行安全及作为分布式和嵌入式大气数据系统的校准基准, 在二代机、三代机得到广泛应用。如美国 F-4、F-15, 苏联安-124、安-225, 俄罗斯米格-21、苏-27, 中国歼-7、歼-10 战斗机和 A320 民用飞机; 以及作为四代机的校准基准, 如美国 F-22、F-35[22], 俄罗斯 Su-35、Su-57 等四代战斗机。

分布式大气数据传感系统采用了集压力受感与参数解算一体化综合设计, 依托气压与飞行姿态的拟合关系获得迎角、侧滑角, 气动隐身综合设计使得低雷达散射性能显著, 具有功能集成度高、管路迟滞误差小、受感器数量少、动态响应特性好的特点, 已在四代机及先进运输机得到广泛应用。如美国 F-22、F-35[22], 俄罗斯 Su-35、Su-57 等四代战斗机, 巴西 KC-390, 欧盟 A400M 等军用运输机, 中国歼-10B、歼-20 战斗机和 A-350、A-380、B-787 等先进民用飞机。

嵌入式大气数据传感系统根据飞行器机身表面多处局部流场信息进行飞行大气参数的解算。局部流场压力信息与飞行姿态之间的相关性增强、耦合度高, 需通过解耦方式获得静压、总压、迎角、侧滑角 4 个基本要素, 凭其对整机气动及隐身影响小、耐高温能力强、数据异常诊断能力强和系统重构性能好的优点已在飞翼气动布局飞机及高超音速类飞机得到广泛应用。如美国的 X-15、F-14、航天飞机、F-18、X-31、X-33、X-38、X-34、X-43A、日本 HYFLEX 及德国 SHEFEX Ⅱ[23-38] 等高超音速飞行器和新一代 B-2、X-47B、RQ-170 等高隐身性能的飞翼气动布局飞机。

2　系统构型演变原因

受作战需求牵引, 飞行器面临低气动阻力、高动态响应、高安全性和可靠性、低可探测性等需求, 飞行器大气数据传感系统从集中式演变到分布式再到嵌入式, 随后诞生了基于激光雷达原理的主动探测方式。具体体现在以下 4 个方面。

1) 减小对平台气动阻力的影响。以机头空速管直接感受静压和飞行总压为典型特征的集中式大气数据传感系统, 逐步过渡到通过机身压力受感器进行总压、静压和与迎角、侧滑角等姿态相关的压力测量, 再到打破理想的总压、静压测量局限, 参数解算所需压力直接来源于飞行器机身局部稳定流场信息。凸出机身的受感器逐渐被与机身蒙皮共形的受感器替代, 适应了减小飞行器气动阻力的总体设计需要。

2) 高动态响应特性。飞行控制和空气动力学关键技术突破[39-40], 实现了从侦察与弹药投放功能单一的一代战斗机逐步跃升到具有隐身、超音速巡航、超机动和超视距打击的典型 4S 特征的四代机[3], 飞行速度覆盖了几十公里每小时到 20 马赫范围[41], 飞行空间从航空空间拓展到了临近空间和航天空间[42]。飞行姿态的实时性测量是安全飞行控制的根本, 推动了大气数据传感器系统管路迟滞问题、测量稳定性问题的解决, 实现了压力受感与参数解算一体化综合设计、压力受感器埋入式设计和基于主动探测方式的高精度低速飞行状态测量思路的提出。

3) 高安全性、可靠性。飞行器作战任务的实施和预期作战效能的实现以安全飞行为保障, 飞行大气参数作为飞行安全控制和精准导航的重要信息源, 对大气数据传感系统的安全性、数据可靠性提出更高要求, 系统冗余、异常状态识别与判断、系统重构能力增强。

4) 低可探测性。飞行器对敌近距侦察、长期监视和毁灭性突袭等作战需要, 探测技术的进一步发展及隐身技术的研究与突破共同带来了飞行器超低可探测性需求, 迫使飞行器机体外露凸出物数量大量减少, 机体表面更加光滑连续[43-44], 推动了翼身融合气动设计技术[45-46] 及扁平翼身融合体飞行控制技术[47-50] 获得快速突破和试验验证, 打破了大气数据探测依托特定局部流场信息的局限, 突破了大气数据传感器与翼身融合, 实现了依托静压、总压、迎角、侧滑角耦合特征明显的多处局部流场压力信息进行飞行大气参数的解算, 嵌入式大气数据传感系统是飞翼气动布局飞机大气数据探测的最优方案, 避免了对整机隐身性能的破坏。

3　构型演变过程及机理

大气数据传感系统中压力受感方式制约了参数测量范围、雷达散射性能、参数解算模型的复杂度、校准与飞行验证需求, 大气数据传感系统构型的演变与静压、总压、迎角、侧滑角的获取方式密切相关。不同大气数据系统构型在物理形态上虽然存在较大的差异, 但是依据静压、总压、迎角、侧滑角 4 个基本要素解算大气参数和依托温度传感器获得总温或静温的基本思想没有发生变化, 静压、总压、迎角、侧滑角的获取原理和方式差异是导致形态差异的根本原因。3 种系统构型的静压、总压、迎角、侧滑角获取原理演变过程如图 1 所示。

集静压受感器、总压受感器、迎角风标传感器、侧滑角风标传感器为一体的机头直杆空速管, 可直接感受静压、总压、迎角、侧滑角的变化量, 随

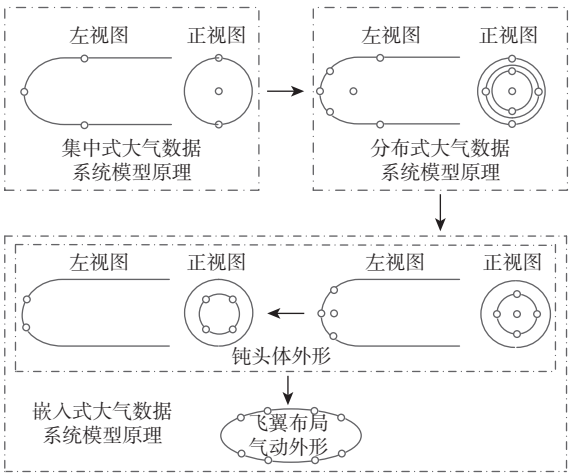


图1 参数解算基本要素获取方式演变流程

Fig. 1 Flow of evolution of basic element acquisition method for parameter calculation

着雷达探测技术的发展和超距离探测能力的需要，为避免对机载雷达信号幅度和频率调制的影响[51-52]，静压受感器、总压受感器、迎角风标传感器、侧滑角风标传感器安装位置从机头移向机身，集中式大气数据传感系统呈现出另一种状态，增强了静压、总压、迎角、侧滑角测量源位置误差[1-2]，其中，静压源位置误差$\Delta P/Q_c$分布如图2所示，需要进行源位置误差修正。

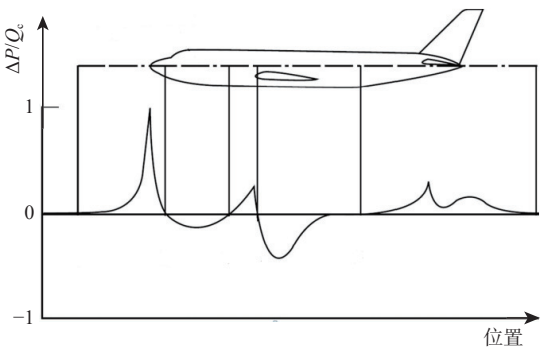


图2 源位置误差分布

Fig. 2 Distribution of source position error

随着电传飞控的产生和应用，航电系统朝着综合化方向发展，集中式大气数据传感系统气压传递管路低通滤波特性导致压力传递迟滞[19-21]，不能满足电传飞控的动态响应特性需求，静压受感器、总压受感器及压力转换实现一体化，压力管路迟滞减小，分布式大气数据系统的典型特征得到彰显。与集中式大气数据传感系统相比，静压、总压、迎角、侧滑角依然可以通过受感器直接获得，可以直接采用飞行大气参数解算基础模型进行飞行大气参数的计算，采用基于飞行姿态多参量的飞行大气参数误差补偿修正方法消除测量值与飞行姿态耦合带来的误差。

伴随探测技术的进一步发展及隐身技术的研究与突破，航电综合技术出现跨越式发展，机体外露物数量大量减少，基于气压随迎角、侧滑角的变化规律实现迎角、侧滑角解算技术的突破，具有活动部件的迎角风标传感器、侧滑角风标传感器被综合替代，完全依托静压、总压、迎角压差感受的一体化探头实现参数解算所需压力的探测，基于飞行姿态多参量的误差补偿修正在提升数据测量精度方面更加重要，系统数据处理能力提升的同时也简化了系统架构。形成了分布式大气数据系统构型气动隐身一体化、高动态响应典型特征。

随着飞机隐身技术的深入发展，翼身融合技术及扁平翼身融合体飞行控制技术[47-50]、高超声速飞行控制技术[53-55]得到突破，分布式大气数据传感系统不仅无法耐受高超声速飞行时气动热环境，也成了影响隐身性能的重要因素，为满足总体隐身及气动性能的需要，集静压、总压、迎角压差感受的一体化探头与机身融合，将机身或钝头体头部作为探头的本体，根据翼身融合体特征，完全依托与静压、总压、迎角、侧滑角特征明显的机身局部流场信息进行飞行大气参数的解算，诞生了嵌入式大气数据传感系统。嵌入式大气数据传感系统成为钝头体气动外形及飞翼气动布局飞行器大气数据探测的最佳解决方案。

4 参数解算模型的演变

4.1 基于基础数学的大气参数解算模型

大气数据传感系统中迎角、侧滑角通过气压与飞行姿态之间的变化规律换算得到，气压高度、指示空速等参数解算的基本原理均是基于静压、总压、温度3个基本参数按通过标准大气数据方程获得[56-57]，如图3所示。

4.2 基于数据拟合的大气参数解算模型

分布式大气数据传感系统以取消迎角、侧滑角风标传感器为显著特征，关键在于可由相对飞行器横向平面对称的受感器气动形面特性、机身左右压力差分布规律确定局部迎角、侧滑角与压力的拟合关系，直接测量得到局部总压、静压，实现了大气数据传感系统完全依托感受气压的静压、总压、迎角、侧滑角4个基本参数的获取，参数解算模型如图4所示。

4.3 基于交叉解耦的参数解算模型

钝头体外形高超声速飞行器大气数据系统测压孔气压与分布几何关系特征鲜明，选取3个典型几何特征的测压数据组合实现参数解算；飞翼气动布局大气数据系统测量压力耦合动压分量比重明

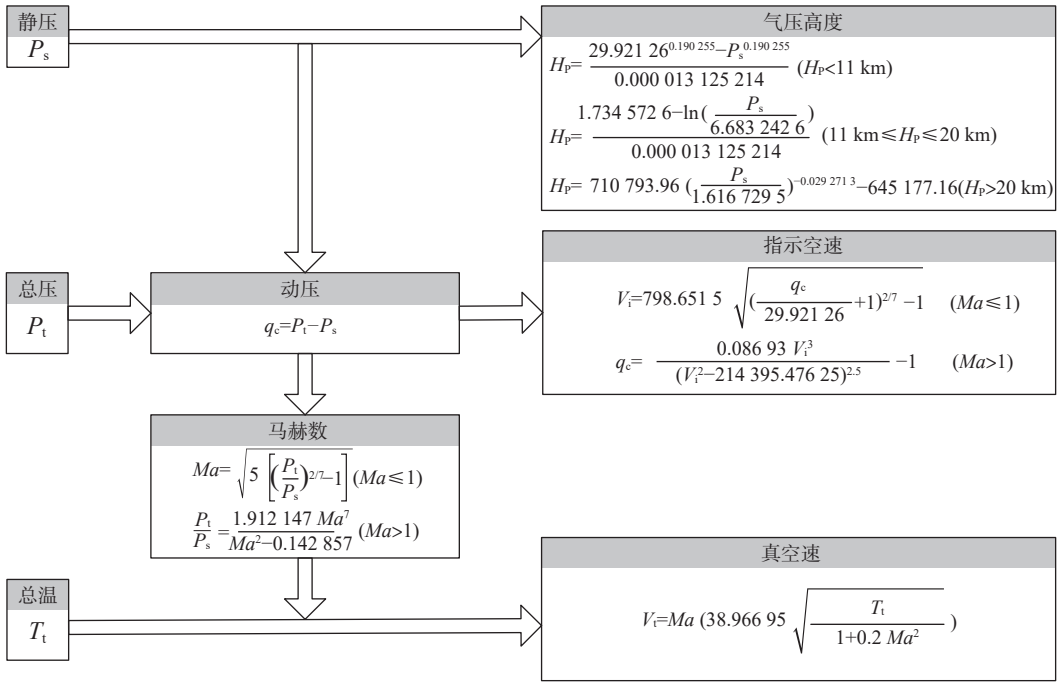


图 3　基于基础数学的大气参数解算模型

Fig. 3　Air parameter calculation model based on basic mathematics

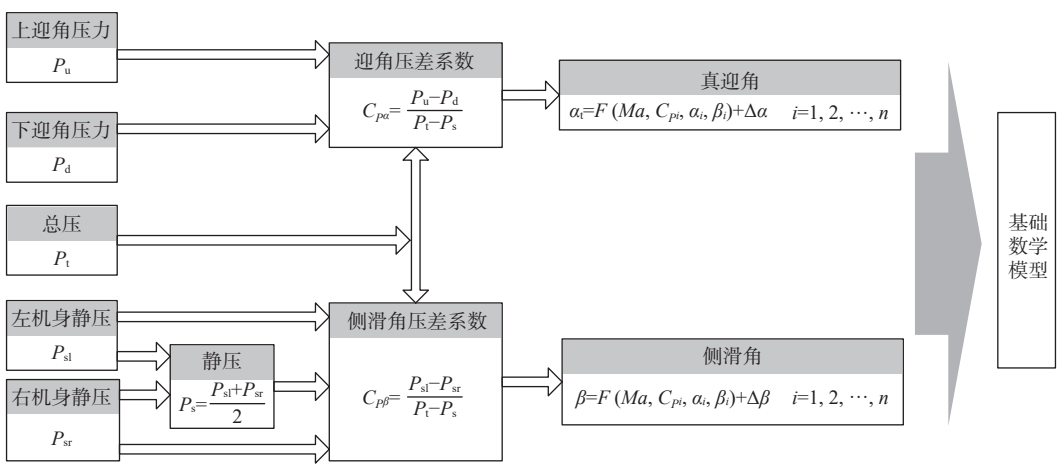


图 4　基于数据拟合的大气参数解算模型

Fig. 4　Air parameter calculation model based on data fitting

显增加, 需要通过消参降耦逐步进行参数求解。

4.3.1　钝头体气动外形

对于钝头体嵌入式大气数据传感系统, 采用美国国家航空航天局(National Aeronautics and Space Administration, NASA)开发的三点法相继获得迎角、侧滑角和迭代求得动压、静压[58-61]。

根据位势流理论和修正的牛顿流理论建立钝头体外形嵌入式大气系统的空气动力学模型, 第i个测压孔表面压力可表示为与入射角、迎角、侧滑角相关的关系式, 即

$$P(\theta_i)=Q_c\left[\cos^2(\theta_i)+\varepsilon\sin^2(\theta_i)\right]+P_s \tag{1}$$

式中: ε为形压系数, 与压缩效应、气动外形、系统影响等因素有关; θ_i为来流入射角, 即该点的曲面法线与来流速度矢量的夹角, 表达式为

$$\cos\theta_i=\cos\alpha\cos\beta\cos\lambda_i+\sin\beta\sin\phi_i\sin\lambda_i+\sin\alpha\cos\beta\cos\phi_i\sin\lambda_i \tag{2}$$

式中: ϕ_i为i点的圆周角; λ_i为i点的圆锥角。选取3个测压点, 通过组合和化简消元得到α和β的解析值, 然后, 把α和β代入到式(1)中, 可以计算出入射角θ_i的值。矩阵形式:

$$\begin{bmatrix}P_1\\P_2\\\vdots\\P_n\end{bmatrix}=\begin{bmatrix}\cos^2\theta_1+\varepsilon\sin^2\theta_1 & 1\\\cos^2\theta_2+\varepsilon\sin^2\theta_2 & 1\\\vdots & \vdots\\\cos^2\theta_n+\varepsilon\sin^2\theta_n & 1\end{bmatrix}\begin{bmatrix}Q_c\\P_s\end{bmatrix} \tag{3}$$

令$\Omega_i=\cos^2\theta_i+\varepsilon\sin^2\theta_i$, 使用加权线形回归方法

可解得

$$\begin{bmatrix}\hat{Q}_c\\ \hat{P}_s\end{bmatrix}=\begin{bmatrix}\sum_{i=1}^{n}q_i & -\sum_{i=1}^{n}q_i(\Omega_i)\\ -\sum_{i=1}^{n}q_i(\Omega_i) & -\sum_{i=1}^{n}q_i(\Omega_i)^2\end{bmatrix}\begin{bmatrix}\sum_{i=1}^{n}(\Omega_i q_i p_i)\\ \sum_{i=1}^{n}q_i p_i\end{bmatrix}\Bigg/\left(\left[\sum_{i=1}^{n}q_i\right]\left[\sum_{i=1}^{n}q_i(\Omega_i)^2\right]-\left[\sum_{i=1}^{n}q_i(\Omega_i)^2\right]\right) \tag{4}$$

式中：q_i为取值为0或1的权值，q_i取0时，将第i个压力测量点的数值从算法中排除，从而满足容错机制。式(4)是闭合形式解，ε为Q_c和P_s的函数，且由于式(3)其实是非线性的，需要进行反复迭代。令ε^0假设为−0.05(超音速时)或−0.5(亚音速时)以初始化迭代，$\varepsilon^{(j)}$的求解则采用最小二乘法求解获得的参数矩阵：

$$\begin{cases}\boldsymbol{M}^{(j)}=\begin{bmatrix}\cos^2\theta_1+\varepsilon^{(j)}\sin^2\theta_1 & 1\\ \cos^2\theta_2+\varepsilon^{(j)}\sin^2\theta_2 & 1\\ \vdots & \vdots\\ \cos^2\theta_n+\varepsilon^{(j)}\sin^2\theta_n & 1\end{bmatrix}\\ \boldsymbol{Q}=\begin{bmatrix}q_1 & \cdots & 0\\ \vdots & & \vdots\\ 0 & \cdots & q_n\end{bmatrix}\end{cases} \tag{5}$$

迭代表达式为

$$\begin{bmatrix}\hat{Q}_c\\ \hat{P}_s\end{bmatrix}^{j+1}=\left[\boldsymbol{M}^{(j)\mathrm{T}}\boldsymbol{Q}\boldsymbol{M}^{(j)}\right]^{-1}\boldsymbol{M}^{(j)\mathrm{T}}\boldsymbol{Q}\begin{bmatrix}P_1\\ P_2\\ \vdots\\ P_n\end{bmatrix} \tag{6}$$

式中：下标j表示第j次迭代的结果；$\boldsymbol{M}^{(j)}$在计算中，采用的是上一次迭代中的Q_c和P_s的值决定的马赫数的值来计算ε；$\boldsymbol{Q}$为权值矩阵，根据P_i数据是否有效，其对角线元素q_1-q_n取值为1或0。

求出Q_c和P_s之后，在亚音速条件下，可以直接根据等熵流的规律来计算。当比热容比$\gamma=1.4$时，有

$$\frac{Q_c}{P_s}=\left(1.0+0.2Ma^2\right)^{3.5}-1 \tag{7}$$

在超音速条件下根据绝热的正激波关系有

$$\frac{Q_c}{P_s}=166.92Ma^2\left(\frac{Ma^2}{7Ma^2-1}\right)^{2.5}-1 \tag{8}$$

$$\begin{cases}Ma=\sqrt{\left(\sum_{i=0}^{9}r_i Z^i\right)\Big/Z}\\ Z=0.839\,371\left(\dfrac{P_s}{Q_c+P_s}\right)\end{cases} \tag{9}$$

4.3.2 飞翼布局气动外形

对于飞翼气动布局嵌入式大气数据传感系统，根据流场数据通过解耦的方式逐步获得飞行大气参数计算所需的静压、总压、迎角、侧滑角4个基本要素。

各嵌入式压力传感器处实测局部压力值表示为

$$P_i=P_s+C_{P_i}(Ma,\alpha,\beta)Q_c \tag{10}$$

式中：$C_{P_i}(Ma,\alpha,\beta)$为各测压点处的压力系数值，选取3个测压点处的实测压力值，P_i、P_j、P_k作差并求比值可得

$$K=\frac{P_i-P_j}{P_k-P_i}=\frac{C_{P_i}-C_{P_j}}{C_{P_k}-C_{P_i}} \tag{11}$$

选定3个压力后，K值为定值，则有

$$C_{P_i}-C_{P_j}=K\left(C_{P_k}-C_{P_i}\right) \tag{12}$$

在给定Ma、β初值情况下，$C_{P_i}-C_{P_j}$和$C_{P_k}-C_{P_i}$为迎角α的函数，则有

$$K_1\alpha+\alpha_{01}=K(K_2\alpha+\alpha_{02}) \tag{13}$$

式(13)仅α一个未知变量，可由此求解α。再选取2个压力值P_m、P_n，联立2个压力表达式得

$$\begin{cases}P_m=P_s+C_{P_m}Q_c\\ P_n=P_s+C_{P_n}Q_c\end{cases} \tag{14}$$

解二元一次方程组可得动、静压表达式为

$$Q_c=\frac{P_m-P_n}{C_{P_m}-C_{P_n}} \tag{15}$$

$$P_s=P_m-C_{P_m}\frac{P_m-P_n}{C_{P_m}-C_{P_n}} \tag{16}$$

式中：C_{P_m}、$C_{P_m}-C_{P_n}$由式(13)求解的α代入拟合的直线得到。

由动、静压求和得到总压：

$$P_t=Q_c+P_s \tag{17}$$

根据标准大气方程解算得到Ma，再由α和Ma确定的机身左右压力系数差随β变化的曲线求得β。

4.4 故障大气参数诊断与检测

早期大气数据传感系统输出的飞行大气参数仅作为飞行操控引导，压力数据异常或不准确均不会带来机毁人亡的灾难。常规气动布局飞行器中使用的分布式大气数据传感系统采用独立多余度配置，飞行大气参数失效率低于10^{-7}，参数耦合度极低，常规气动布局的高稳定性优点使得飞行控制通道表决选择阈值门限较宽。飞翼气动布局飞行器飞行控制系统各通道表决选择阈值门限相对较窄，各参数之间的强耦合使得任一局部流场异常导致通道数据分离的概率增大，甚至造成灾难性事故发生[62]。而在装备全天候、跨区域长航时作战备战要

求下不可避免遇到复杂气象环境造成气压通路进水、结冰、异物堵塞, 使得压力数据无效、不准确, 参数解算异常、误差增大, 需要通过故障诊断与检测[63]过程识别出无效压力数据, 对通道健康状态进行监测, 故障诊断能力得到提升。

文献 [64-65] 基于各测压点残差标准化后的平方和服从于卡方分布规律, 将卡方检测应用于嵌入式大气数据系统故障检测中。文献 [65] 基于 3 点组合冗余性, 利用奇偶方程和粗糙集理论, 实现了嵌入式大气数据系统故障数据高效检测与识别。鉴于嵌入式大气数据系统通常在同区域流场或各余度通道均会布局多个测压点, 可采用同区域数据分布规律和设置阈值, 以及采用预置压力遍历检查方法对大气参数解算过程中使用的异常压力进行检测更具有普适性, 流程如图 5 和图 6 所示。

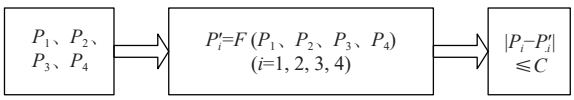


图 5　基于同区域压力分布规律故障诊断流程

Fig. 5　Fault diagnosis process based on pressure distribution law in same region

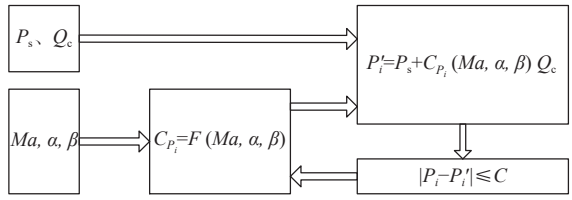


图 6　基于预置压力故障诊断流程

Fig. 6　Fault diagnosis process based on preset pressure

4.5　基于多源数据融合的大气参数重构模型

根据系统健康监测情况, 采用系统内其他通道压力信息或综合利用惯导信息、飞行器动力学模型进行大气数据解析实现大气参数的间接测量, 对故障通道进行重构, 保证通道的完整性。

基于系统内压力的重构就是在出现压力数据异常或不可用时, 借助其他压力数据, 使用最小二乘法计算的重构压力替换异常压力。鉴于惯性导航系统不受机体表面气动流场分布影响, 测量参数实时反映了机体姿态的变化, 且不受外部电磁干扰的影响, 低动压大迎角飞行时采用大气/惯导融合参数估计方法, 可有效解决高机动飞行参数波动和准确性问题[66-70]。而估计参数的动态误差大小受大气系统、惯导系统工作完好性的影响, 且大气数据系统的容错能力有限, 考虑到飞机的六自由度数学模型建立了飞机的飞行状态参数与飞行控制量之间的非线性函数关系, 反映了飞机所受合外力、合外力矩及发动机推力对飞行状态参数的影响[69], 采用惯导/飞控系统融合实现以飞机动力学模型为状态方程和惯导系统输出为观测量的大气参数计算, 通过卡尔曼滤波算法[71-72]构建独立的大气数据系统补充通道和大气数据系统的故障容错的辅助判据, 提高系统的安全性, 演变流程如图 7 所示。

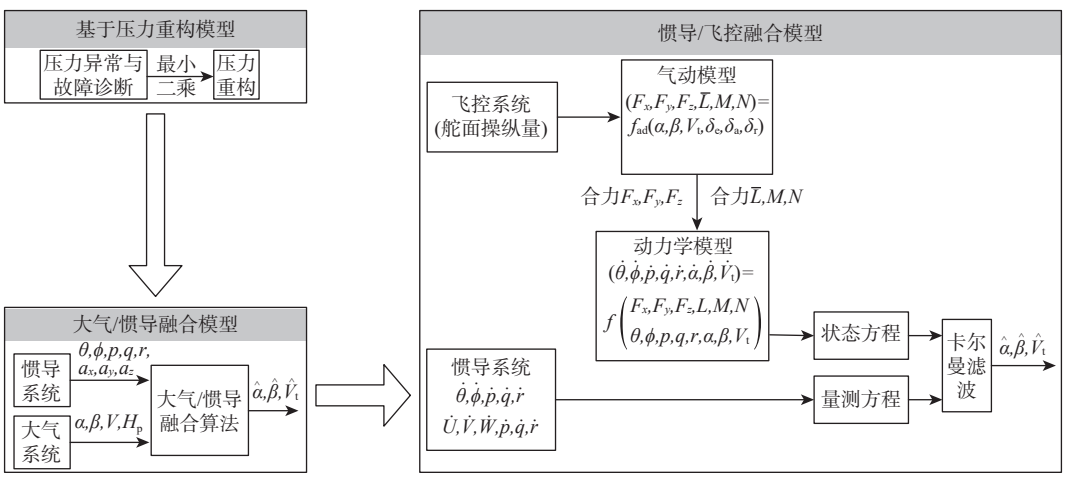


图 7　系统重构模型演变流程

Fig. 7　Evolution of system reconstruction model

5　结 束 语

为适应飞行器低气动阻力、高动态响应、高安全性和可靠性、低可探测性等需求, 大气数据传感系统构型和解算数学模型实现了适应性演变, 但伴随飞行控制操纵纵向、横向稳定性及鲁棒性问题得到解决, 以及未来飞行器具有宽隐身、强感知、远航程、超敏捷、云网络、闪杀伤、积木式组合架构、软件化定义功能等典型能力特征[73-75], 飞行器大气数据系统设计发展方向预测如下:

1) 基于多源信息融合的大气参数解算技术。针对未来飞行器扁平飞翼气动布局特征更加明显,

局部稳定流场压力信息与飞行器姿态之间的变化规律一致性被破坏，需要研究基于与飞行姿态相关联的信息进行飞行大气参数解算，如舵面伺服传感器、进气道流量、发动机转速、惯导信息等。

2）基于多参量传感器网络的飞行大气数据测量技术研究。基于集成温度、应变、流场等多参量传感器的蒙皮和多参量测量微系统网络数据无线传输，建立基于机体表面数十至数百处多参量微系统网络数据及传感器数据融合飞行大气参数解算，增强飞行器在复杂环境中的自适应前相控制、超态势感知能力，提升全频谱超低可探测性性能。

6 结 论

1）本文系统分析了飞行器大气数据系统构型和解算模型演变过程和机理，总结出飞行器大气数据传感系统构型形态的变更归根结底是为适应飞行器新的作战需求。飞行器大气数据系统构型在发展过程中逐步打破了参数解算需直接获得静压、总压、迎角、侧滑角的局限，流场压力与飞行姿态高度耦合性增强了依托对飞行压力与飞行姿态之间的对应关系的挖掘逐次获得静压、总压、迎角、侧滑角的求解，参数解算难度增加。

2）基于大气数据传感系统内压力冗余、大气/惯导、惯导/动力学模型实现数据异常识别与故障诊断、系统重构，提升了大气数据传感系统飞行大气参数测量的准确性、可靠性、容错能力和保证通道的完整性。为适应未来飞行器发展趋势，大气数据传感系统设计应重点发展基于多源信息融合的大气参数解算技术和实现基于多参量传感器网络的飞行大气数据测量技术。

参考文献（References）

[1] HAERING JR E A. Airdata measurement and calibration[EB/OL]. (1995-12-01)[2023-06-04]. https://ntrs.nasa.gor/api/citations/19990065780/downloads/19990063780.pdf.

[2] UTC Aerospace Systems. Air data probes[EB/OL]. (2017-01-01)[2023-10-01]. https://prd-sc102-cdn.rtx.com/-/media/ca/product-assets/files/commercial/flight-deck/surveillance/air-data-systems/air-data-probes.pdf.

[3] 杨伟. 关于未来战斗机发展的若干讨论[J]. 航空学报, 2020, 41(6): 524377.
YANG W. Development of future fighters[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(6): 524377(in Chinese).

[4] 史彦斌, 高宪军, 王远达. 航空电子系统导论[M]. 北京: 国防工业出版社, 2019: 254-258.
SHI Y B, GAO X J, WANG Y D. Introduction to avionics systems[M]. Beijing: National Defense Industry Press, 2019: 254-258 (in Chinese).

[5] 岳亚洲, 李彬, 雷宏杰. 激光大气运动参数测量技术研究进展及展望 (特邀)[J]. 光子学报, 2022, 51(4): 0428001.
YUE Y Z, LI B, LEI H J. Advances and prospects of laser measurement technology for air motion parameters(invited)[J]. Acta Photonica Sinica, 2022, 51(4): 0428001(in Chinese).

[6] CALDWELL L M, TANG S Y, O'BRIEN M. Optical air data systems and methods: US7760339[P]. 2010-07-20.

[7] CALDWELL L M, ACOTT P E, O'BRIEN M. Optical air data systems and methods: US11300584[P]. 2022-04-12.

[8] CALDWELL L M, O'BRIEN M J, WEIMER C S, et al. Optical air data systems and methods: US6894768[P]. 2005-05-17.

[9] CALDWELL L M, TANG S Y, ACOTT P E, et al. Optical air data systems and methods: US8072584[P]. 2011-12-06.

[10] 孙友师. 光学大气数据测量系统的发展研究[C]//航空试验测试技术峰会论文集. 北京: 中国航空学会 2010: 6-9.
SUN Y S. Progress of optical air data measurement systems [C]//Proceedings of the Aviation Test Technology Summit. Beijing: Chinese Society of Aeronautics and Astronautics, 2010, 29: 6-9(in Chinese).

[11] BOGUE R. Recent flight-test results of optical airdata techniques[C]//Proceedings of the 6th AIAA Biennial Flight Test Conference. Reston: AIAA, 1992.

[12] MAMIDIPUDI P, DAKIN E A, DAKIN D C, et al. LandSafe precision flight instrumentation system: the DVE solution[C]//Proceedings of the Airborne Intelligence, Surveillance, Reconnaissance Systems and Applications. Bellinham: SPIE, 2012.

[13] 秋路, 屈飞舟, 惠辉辉. 机载激光测速技术在大气数据校准领域的应用研究[J]. 航空科学技术, 2019, 30(2): 32-36.
QIU L, QU F Z, HUI H H. Research on applications of airborne laser anemometry in air data calibration[J]. Aeronautical Science & Technology, 2019, 30(2): 32-36(in Chinese).

[14] BOGUE R K, JENTINK H W. Optical air flow measurements in flight: NASA/TP-2004-210735 [R]. Washington, D. C. : NASA, 2018.

[15] VERBEEK M J, JENTINK H W. Optical air data system flight testing: NLR-TP-2012-068[R]. Munich : NLR, 2012.

[16] 梁应剑, 梅运桥, 程丽媛, 等. 基于米散射的光学大气数据系统研究[J]. 测控技术, 2015, 34(1): 32-34.
LIANG Y J, MEI Y Q, CHENG L Y, et al. Research on optical air data system based on Mie scattering[J]. Measurement & Control Technology, 2015, 34(1): 32-34(in Chinese).

[17] 王晓维, 梁应剑, 李翔, 等. 基于光学多普勒频移的低空速测量方法研究[J]. 激光技术, 2016, 40(5): 629-632.
WANG X W, LIANG Y J, LI X, et al. Research of low-airspeed measurement based on optical Doppler frequency shift[J]. Laser Technology, 2016, 40(5): 629-632(in Chinese).

[18] 龙彦志, 梁应剑, 黄巧平, 等. 基于多普勒频移的光学大气测速系统设计[J]. 北京航空航天大学学报, 2018, 44(12): 2521-2527.
LONG Y Z, LIANG Y J, HUANG Q P, et al. Design of optical airspeed measurement system based on Doppler shift[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(12): 2521-2527(in Chinese).

[19] 宋秀毅, 陆宇平. 嵌入式大气数据传感系统压力传感器设计研究[J]. 计测技术, 2007, 27(5): 8-10.
SONG X Y, LU Y P. Research on design of pressure sensor of em-

bedded airdata sensing system[J]. Metrology & Measurement Technology, 2007, 27(5): 8-10(in Chinese).

[20] 沈国清. 嵌入式大气数据传感系统误差分析及其消除方法研究[D]. 南京: 南京航空航天大学, 2012: 15-23.
SHEN G Q. Research on error analysis and elimination method of embedded air data sensing system[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012: 15-23(in Chinese).

[21] WHITMORE S, MOES T. The effects of pressure sensor acoustics on airdata derived from a high-angle-of-attack flush airdata sensing system[C]//Proceedings of the 29th Aerospace Sciences Meeting. Reston: AIAA, 1991.

[22] 陈怦. F-35”闪电”Ⅱ战斗机飞行试验全纪录[M]. 北京: 航空工业出版社, 2019: 59-67.
CHENG P. Full flight test record of F-35 lightning fighter jet[M]. Beijing: Aviation Industry Press, 2019: 59-67(in Chinese).

[23] 丁智坚, 周欢, 吴东升, 等. 嵌入式大气数据测量系统技术研究进展[J]. 宇航学报, 2019, 40(3): 247-257.
DING Z J, ZHOU H, WU D S, et al. Review of flush air data sensing system[J]. Journal of Astronautics, 2019, 40(3): 247-257(in Chinese).

[24] CARY J, KEENER E R. Flight evaluation of the X-15 ball-nose flow-direction sensor as an air-data system: NASA TN D-2923 [R]. Washington, D. C. : NASA, 1965.

[25] ROW P, FISCHEL J. Operational flight-test experience with the X-15 airplane[C]//Proceedings of the Space Flight Testing Conference. Reston: AIAA, 1963.

[26] WOESTE T. Shuttle entry air data system - an experimental investigation of calibration for ascent flight[C]//Proceedings of the 30th Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 1992.

[27] SIEMERS P M, WOLF H, FLANAGAN P F, et al. Shuttle entry air data system concepts applied to space shuttle orbiter flight pressure data to determine air data-STS 1-4[C]//Proceedings of the 21st Aerospace Sciences Meeting. Reston: AIAA, 1983.

[28] HENRY M, WOLF H, SIEMERS P. An evaluation of shuttle entry air data system (seads) flight pressures-comparisons with wind tunnel and theoretical predictions[C]//Proceedings of the 15th Aerodynamic Testing Conference. Reston: AIAA, 1988.

[29] WHITMORE S, MOES T, LARSON T. Preliminary results from a subsonic high-angle-of-attack flush airdata sensing (HI-FADS) system-design, calibration, algorithm development, and flight test evaluation[C]//Proceedings of the 28th Aerospace Sciences Meeting. Reston: AIAA, 1990.

[30] WHITMORE S A. Development of a pneumatic high-angle-of-attack flush airdata sensing (HI-FADS) system[J]. Control and Dynamic Systems, 1992, 52: 453-511.

[31] WEISS S. Comparing three algorithms for modeling flush air data systems[C]//Proceedings of the 40th AIAA Aerospace Sciences Meeting & Exhibit. Reston: AIAA, 2002.

[32] WHITMORE S, COBLEIGH B, HAERING E Jr. Design and calibration of the X-33 flush airdata sensing (FADS) system[C]//Proceedings of the 36th AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 1998.

[33] COBLEIGH B, WHITMORE S, HAERING E Jr, et al. Flush airdata sensing (FADS) system calibration procedures and results for blunt forebodies[C]//Proceedings of the 36th 9th International Space Planes and Hypersonic Systems and Technologies Conference. Reston: AIAA, 1999.

[34] ELLSWORTH J, WHITMORE S. Reentry air data for a sub-orbital spacecraft based on X-34 design[C]//Proceedings of the 45th AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2007.

[35] WHITMORE S A, ELLSWORTH J C. Simulation of a flush airdata system for transatmospheric vehicles[J]. Journal of Spacecraft and Rockets, 2008, 45(4): 716-732.

[36] BAUMANN E, PAHLE J W, DAVIS M C, et al. X-43A flush airdata sensing system flight-test results[J]. Journal of Spacecraft and Rockets, 2010, 47(1): 48-61.

[37] TAKAKI R, TAKIZAWA M, TAKAKI R, et al. ADS measurement of HYFLEX (hypersonic flight experiment) [C]//Proceedings of the 35th Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 1997.

[38] THEIL S, SCHLOTTERER M, CONRADT M, et al. Hybrid navigation system for the SHEFEX-2 mission[C]//Proceedings of the 36th AIAA Guidance, Navigation and Control Conference and Exhibit. Reston: AIAA, 2008.

[39] 丁玉临, 韩忠华, 乔建领, 等. 超声速民机总体气动布局设计关键技术研究进展[J]. 航空学报, 2023, 44(2): 626310.
DING Y L, HAN Z H, QIAO J L, et al. Research progress in key technologies for conceptual-aerodynamic configuration design of supersonic transport aircraft[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(2): 626310(in Chinese).

[40] 蔡琰. 国外射流飞行控制技术发展及前景分析[J]. 航空科学技术, 2020, 31(1): 85-86.
CAI Y. Development and prospect analysis of jet flight control technology abroad[J]. Aeronautical Science & Technology, 2020, 31(1): 85-86(in Chinese).

[41] 顾诵芬, 史超礼. 世界航空发展史[M]. 郑州: 河南科学技术出版社, 1998: 138.
GU S F, SHI C L. The history of the world aviation development[M]. Zhengzhou: Henan Science and Technology Press, 1998: 138(in Chinese).

[42] 秦伟伟, 刘刚, 赵欣, 等. 临近空间高超声速飞行器控制系统基本原理[M]. 北京: 北京航空航天大学出版社, 2019: 24-26.
QIN W W, LIU G, ZHAO X, et al. Basic principle of control system for near-space hypersonic vehicle[M]. Beijing: Beijing University of Aeronautics & Astronautics Press, 2019: 24-26(in Chinese).

[43] 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).

[44] 魏小龙, 韩欣珉, 李益文, 等. 动态可调等离子体隐身技术[M]. 北京: 科学出版社, 2021.
WEI X L, HAN X M, LI Y W, et al. Dynamic adjustable plasma stealth technology[M]. Beijing: Science Press, 2021(in Chinese).

[45] 蒋瑾, 钟伯文, 符松. 翼身融合布局飞机总体参数对气动性能的影响[J]. 航空学报, 2016, 37(1): 278-289.
JIANG J, ZHONG B W, FU S. Influence of overall configuration parameters on aerodynamic characteristics of a blended-wing-body aircraft[J]. Acta Aeronautica et Astronautica Sinica, 2016, 37(1): 278-289(in Chinese).

[46] 李晓勇, 张淼, 鲁素芬, 等. 翼身融合飞机的空气动力学研究进展[J]. 飞机设计, 2007, 27(2): 1-9.

LI X Y, ZHANG M, LU S F, et al. Advances in aerodynamics for blended-wing-body aircraft[J]. Aircraft Design, 2007, 27(2): 1-9(in Chinese).

[47] 葛志闪. 飞翼布局无人机控制律设计[D]. 西安: 西北工业大学, 2007: 1-83.
GE Z S. Design of control law for UAV with flying wing layout[D]. Xi'an: Northwestern Polytechnical University, 2007: 1-83(in Chinese).

[48] 李永丰, 吕永玺, 史静平, 等. 深度确定性策略梯度和预测相结合的无人机空战决策研究[J]. 西北工业大学学报, 2023, 41(1): 56-64.
LI Y F, LYU Y X, SHI J P, et al. UAV's air combat decision-making based on deep deterministic policy gradient and prediction[J]. Journal of Northwestern Polytechnical University, 2023, 41(1): 56-64 (in Chinese).

[49] 张羽白, 肖成方, 邹俊俊, 等. 飞翼布局无人机控制律设计[J]. 测控技术, 2021, 40(1): 123-131.
ZHANG Y B, XIAO C F, ZOU J J, et al. Design of flight control law for a flying-wing UAV[J]. Measurement & Control Technology, 2021, 40(1): 123-131(in Chinese).

[50] 范力元, 张浩哲, 徐钊, 等. 基于安全飞行走廊的无人机密集障碍规避算法[J]. 西北工业大学学报, 2022, 40(6): 1288-1296.
FAN L Y, ZHANG H Z, XU Z, et al. A dense obstacle avoidance algorithm for UAVs based on safe flight corridor[J]. Journal of Northwestern Polytechnical University, 2022, 40(6): 1288-1296(in Chinese).

[51] 刘宝泉. 机载预警雷达构型探测性能影响分析[J]. 科技创新与应用, 2019, 9(12): 39-40.
LIU B Q. Analysis on the influence of airborne early warning radar configuration detection performance[J]. Technology Innovation and Application, 2019, 9(12): 39-40(in Chinese).

[52] 颜仙荣, 李其虎. 旋转螺旋桨对机载雷达探测性能影响研究[J]. 中国电子科学研究院学报, 2020, 15(2): 141-146.
YAN X R, LI Q H. Study on effect of the rotating propeller for airborne radar detection performance[J]. Journal of China Academy of Electronics and Information Technology, 2020, 15(2): 141-146(in Chinese).

[53] 甄子洋, 朱平, 江驹, 等. 基于自适应控制的近空间高超声速飞行器研究进展[J]. 宇航学报, 2018, 39(4): 355-367.
ZHEN Z Y, ZHU P, JIANG J, et al. Research progress of adaptive control for hypersonic vehicle in near space[J]. Journal of Astronautics, 2018, 39(4): 355-367(in Chinese).

[54] 吴宏鑫, 孟斌. 高超声速飞行器控制研究综述[J]. 力学进展, 2009, 39(6): 756-765.
WU H X, MENG B. Review on the control of hypersonic flight vehicles[J]. Advances in Mechanics, 2009, 39(6): 756-765(in Chinese).

[55] 孙长银, 穆朝絜, 余瑶. 近空间高超声速飞行器控制的几个科学问题研究[J]. 自动化学报, 2013, 39(11): 1901-1913.
SUN C Y, MU C X, YU Y. Some control problems for near space hypersonic vehicles[J]. Acta Automatica Sinica, 2013, 39(11): 1901-1913(in Chinese).

[56] 中华人民共和国航空工业部. 飞行大气参数: HB 6127—1986[S]. 中国: 航空工业部, 1987: 1-498.
Ministry of Aviation Industry of the People's Republic of China. Flight atmosphere parameter: HB 6127—1986[S] . China: Ministry of aviation Industry, 1987: 1-498(in Chinese).

[57] 肖建德. 大气数据计算机系统[M]. 北京: 国防工业出版社, 1992: 35-45.
XIAO J D. Air data computer system[M]. Beijing: National Defense Industry Press, 1992: 35-45(in Chinese).

[58] 陈康, 符文星, 闫杰. 不同布局下高超声速飞行器 FADS 求解精度[J]. 固体火箭技术, 2014, 37(4): 453-457.
CHEN K, FU W X, YAN J. Solving accuracy for flush airdate sensing system of hypersonic aircraft under different configuration[J]. Journal of Rocket Technology, 2014, 37(4): 453-457(in Chinese).

[59] COBLEIGH B R, WHITMORE S A, HAERING E A, et al. Flush airdata sensing (FADS) system calibration procedures and results for blunt forebodies[C]//Proceedings of the 9th International Space Planes and Hypersonic Systems and Technologies Conference. Reston: AIAA, 1999.

[60] 方习高, 陆宇平. 嵌入式大气数据传感系统的求解算法研究[J]. 计算机测量与控制, 2008, 16(3): 398-400.
FANG X G, LU Y P. Research on algorithms of flush airdata sensing system[J]. Computer Measurement & Control, 2008, 16(3): 398-400(in Chinese).

[61] 方习高. 嵌入式大气数据传感系统的技术及应用研究[D]. 南京: 南京航空航天大学, 2007: 12-15.
FANG X G. Research on technology and application of embedded air data sensing system[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2007: 12-15(in Chinese).

[62] 新浪军事. 美军公布 B2 坠机直接原因: 空速管受潮失灵(图)[EB/OL]. (2008-06-08)[2023-10-01]. http://mil.news.sina.com.cn/p/2008-06-08/1053504376.html.
Sina Military. The US military announced the immediate cause of the B-2 crash: the airspeed tube was damp and malfunctioned (with pictures)[EB/OL]. (2008-06-08)[2023-10-01]. http://mil.news.sina.com.cn/p/2008-06-08/1053504376.html(in Chinese).

[63] 张颖伟, QIN S J. 复杂工业过程的故障诊断[M]. 沈阳: 东北大学出版社, 2007.
ZHANG Y W, QIN S J. Fault diagnosis of complex industrial process[M]. Shenyang: Northeast University Press, 2007(in Chinese).

[64] WHTITMORE S A, MOES T R. Failure detection and fault management techniques for a pneumatic high-angle-of-attack flush airdata sensing (HI-FADS) System: NASA TM-4335 [R]. Washington, D. C.: NASA, 1992.

[65] 赵磊. 嵌入式大气数据传感系统故障检测与处理算法研究[D]. 南京: 南京航空航天大学, 2010: 33-43.
ZHAO L. Research on fault detection and processing algorithm of embedded air data sensing system[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2010: 33-43(in Chinese).

[66] 马航帅, 雷廷万, 李荣冰, 等. 大迎角下基于信息融合的迎角/侧滑角估计方法[J]. 电光与控制, 2012, 19(8): 1-5.
MA H S, LEI T W, LI R B, et al. Estimation method for angle-of-attack and sideslip angle based on information fusion at high angle of attack[J]. Electronics Optics & Control, 2012, 19(8): 1-5(in Chinese).

[67] LONG H, SONG S J. Method of estimating angle-of-attack and sideslip angel based on data fusion[C]//Proceedings of the 2nd International Conference on Intelligent Computation Technology and Automation. Piscataway: IEEE Press, 2009: 641-644.

[68] PERRY J, MOHAMED A, JOHNSON B, et al. Estimating angle of attack and sideslip under high dynamics on small UAVs[C]// Proceedings of the 21st International Technical Meeting of the Satellite Division of The Institute of Navigation. London: Institute of Navigation , 2008: 1165-1173.

[69] 马航帅. 基于虚拟大气的大迎角大气数据估计技术研究[D]. 南京: 南京航空航天大学, 2012: 1-59.
MA H S. Research on estimation technology of atmospheric data at high angle of attack based on virtual atmosphere[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012: 1-59(in Chinese).

[70] 蒋保睿, 肖地波, 张勇, 等. 融合气象观测数据的高精度大气数据估计算法研究[J]. 计算机测量与控制, 2022, 30(12): 276-283.
JIANG B R, XIAO D B, ZHANG Y, et al. High precision air data estimation algorithm with meteorological observation[J]. Computer Measurement & Control, 2022, 30(12): 276-283(in Chinese).

[71] 肖地波, 蒋保睿, 刘鹏. 基于 FADS/IMU 信息融合的大气数据测量方法研究[J]. 测控技术, 2022, 41(9): 50-55.
XIAO D B, JIANG B R, LIU P. Research on atmospheric data measurement method based on FADS/IMU information fusion[J]. Measurement & Control Technology, 2022, 41(9): 50-55(in Chinese).

[72] 蒋保睿, 刘鹏, 肖地波. 基于 UKF 的 FADS/INS 融合大气数据估计[J]. 宇航计测技术, 2022, 42(4): 31-36.
JIANG B R, LIU P, XIAO D B. Air data estimation of FADS/INS fusion based on UKF[J]. Journal of Astronautic Metrology and Measurement, 2022, 42(4): 31-36(in Chinese).

[73] 闫晓婧, 杨涛, 药红红. 国外第六代战斗机概念方案与关键技术[J]. 航空科学技术, 2018, 29(4): 18-26.
YAN X J, YANG T, YAO H H. Conceptual scheme and key technologies of sixth generation fighters abroad[J]. Aeronautical Science & Technology, 2018, 29(4): 18-26(in Chinese).

[74] 王锴, 丁宇, 何大龙. 第六代战斗机发展动向及能力分析[J]. 光电技术应用, 2019, 34(5): 1-6.
WANG K, DING Y, HE D L. Development trend and capability analysis of the sixth generation fighter[J]. Electro-Optic Technology Application, 2019, 34(5): 1-6(in Chinese).

[75] 赵保军, 陈士涛, 李大喜, 等. 国外六代机发展及作战概念分析[J]. 现代防御技术, 2022, 50(6): 19-25.
ZHAO B J, CHEN S T, LI D X, et al. Analysis of the sixth generation fighter development and operational concept[J]. Modern Defence Technology, 2022, 50(6): 19-25(in Chinese).

Configuration and calculation evolution mechanism of air data system for aircraft

XIONG Liang[1, 2, *], ZHANG Rui[1], XU Bin[1], HUANG Qiaoping[2]

(1. College of Automation, Northwestern Polytechnical University, Xi'an 710129, China;
2. Department of Air Data System, Chengdu Caic Electronics Co., Ltd., Chengdu 610091, China)

Abstract: The air data sensing system provides the necessary parameters for aircraft flight control, delivery and trajectory control, such as static pressure, total pressure, angle of attack, and sideslip angle. The inter-generation development of aircraft has induced the evolution of the basic element acquisition method for parameter calculation of air data sensing systems from direct type to decoupled extraction. This paper summarized and refined the configuration and development history of the air data sensing system and systematically analyzed the causes and processes of configuration evolution for the air data sensing system, the principle of air parameter calculation model, the diagnosis and detection process of fault air parameters, and system reconstruction model. It also summarized the key research directions for configuration and parameter solution design of air data sensing systems for aircraft in the future.

Keywords: air data system; air data system configuration; air parameter calculation; fault-tolerant air data system; air data system reconstruction

Received: 2023-06-12; **Accepted**: 2023-09-11; **Published Online**: 2023-10-10 19:32
URL: link.cnki.net/urlid/11.2625.V.20231009.1254.002

* **Corresponding author**. E-mail: caic02253@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0194

星载准绝对式光电编码器抗污方法

韩庆阳*，沈宏海，马天翔，梁超，王志冲
(中国科学院长春光学精密机械与物理研究所，长春 130033)

摘　　要：为提高星载准绝对式光电编码器的抗油污能力，提出多参考点绝对角位移识别和多读数头数据融合方法。介绍星载准绝对式光电编码器的组成和处理电路设计；提出多参考点绝对角位移的识别方法，防止因油污造成的绝对角位移误识别，并采用多读数头数据融合的方法，判断并消除错误的读数头测量数据；通过角位移曲线和角速度曲线对比实验，验证所提方法的有效性，并使用自准直平行光管和23面体检测精度。实验结果表明：所提方法能够有效的解决圆光栅被污染带引起的角位移错误问题，确保角位移正确可靠；经检验角位移误差最大值为4.6"，最小值为−0.6"，峰峰值为5.2"，均方差为4.3"。满足星载准绝对式光电编码器抗污能力和精度要求，并已成功在工程中应用，效果良好。所提方法对提高星载准绝对式光电编码器抗油污能力，有实用价值。

关 键 词：星载；准绝对式光电编码器；抗污能力；多参考点；多读书头
中图分类号：V221+.3; TB553
文献标志码：A　　**文章编号**：1001-5965(2025)06-2014-08

随着星载激光通信技术的发展，对激光通信终端二维转台的技术要求日益提高[1-2]。光电编码器为转台提供高精度角位移，进而完成终端设备的捕获、跟踪和通信功能[3]，因此，对光电编码器的要求随之提高。光电编码器是一种集光、机、电于一体的数字化精密角位移测量装置[4-5]。具有精度高、分辨力高、环境适应力强和易于集成等优点，常用于工业和其他科技领域的实时测量控制系统中[6-8]。从编码方式上分，其可分为增量式、准绝对式和绝对式[9]。准绝对式即多参考点光电编码器，相对单参考点增量式光电编码器，准绝对式光电编码器绝对角位移识别效率高[10]。同时又兼具单参考点增量式光电编码器编码方式简单、结构简单、反应速度快和工作稳定等优点[11]，故在星载激光通信终端有广泛应用。但由于准绝对式光电编码器的核心部件，即圆光栅在使用过程中容易被污染，导致测量的角位移错误，这对星载准绝对式光电编码器的影响是灾难性的。因此，亟需提高星载准绝对式光电编码器的抗污能力，保证角位移的正确。

导致星载准绝对式光电编码器因圆光栅污染造成角位移测量错误的主要原因是使用时对圆光栅的污染引起的虚假参考点和莫尔条纹信号跳动。通常的做法是将圆光栅密封和读数头备份。由于不可能完全密封，且即使完全密封，也存在因使用等原因造成的密封泄露。一旦密封泄露，使用时环境中的污渍极易污染圆光栅，导致测量得到的角位移错误。读数头备份是星载设备的常用手段，但难点在于判断主备测量的角位移是否正常。于海等[12]提出基于图像补偿的抗污方法，有效的提高了图像式光电编码器的抗污能力，这种方法并不适用于准绝对式光电编码器。文献[13-15]提出多读数头圆光栅自校准方法，有效的提高了圆光栅的测量精度，但无法解决圆光栅的污染问题。

星载准绝对式光电编码器圆光栅的污染通常有且只有一处。当该污染出现在圆光栅的参考点编码环或光栅环时，会导致测量的角位移错误。因

收稿日期：2024-04-07；**录用日期**：2024-07-30；**网络出版时间**：2024-08-02 10:14
网络出版地址：link.cnki.net/urlid/11.2625.V.20240801.1544.001
基金项目：国家自然科学基金(52075520)
*通信作者. E-mail：lanbohesky@126.com

引用格式：韩庆阳，沈宏海，马天翔，等. 星载准绝对式光电编码器抗污方法[J]. 北京航空航天大学学报，2025，51(6)：2014-2021.
HAN Q Y，SHEN H H，MA T X，et al. Anti-stain method of satellite-borne absolute-to-be optical encoder[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2014-2021 (in Chinese).

此, 本文提出从绝对角位移识别方法及多读数头测量融合两方面, 提高星载准绝对式光电编码器的抗污能力, 保证角位移正确可靠。首先, 介绍星载准绝对式光电编码器的组成和信号处理电路; 其次, 设计准绝对式光电编码器的圆光栅图案, 提出多参考点绝对位置识别法和 3 组测量读数头数据融合方法, 消除油污对圆光栅污染的影响; 最后, 设计开发了圆光栅直径为 112 mm, 光栅刻线数为 16 384 的准绝对式角位移测量装置, 进行抗油污实验和精度检测。结果表明: 本文方法有效的消除油污污染对角度测量的影响, 经检测误差最大值为 4.6", 最小值为−0.6", 峰峰值为 5.2", 均方差为 4.3"; 满足系统对星载准绝对式光电编码器角位移测量的可靠性和精度要求。这对提高星载准绝对式光电编码器抗油污能力, 具有实用价值, 并已成功应用。

1　星载准绝对式光电编码器的组成

星载准绝对式光电编码器是基于圆光栅和指示光栅干涉产生的莫尔条纹实现角位移精密测量的。其由圆光栅、指示光栅、机械轴系和处理电路等部分组成。当轴系转动时圆光栅随之转动, 圆光栅和指示光栅发生相对运动。此时光源发出的光经过圆光栅和指示光栅后产生莫尔条纹, 光电接收管接收莫尔条纹并将其转化成电流信号, 送给后续处理电路进行处理, 即可得到轴系转动的角位移。其组成如图 1 所示。

后续处理电路完成星载准绝对式光电编码器的信号处理和角位移解算, 主要由: 整形电路、放大电路、模数转换 (analog-to-digital converter, AD) 采集电路、现场可编程门阵列 (field programnable gate array, FPGA) 和通信电路构成; 其电路原理框如图 2 所示。

星载准绝对式光电编码器的处理电路元器

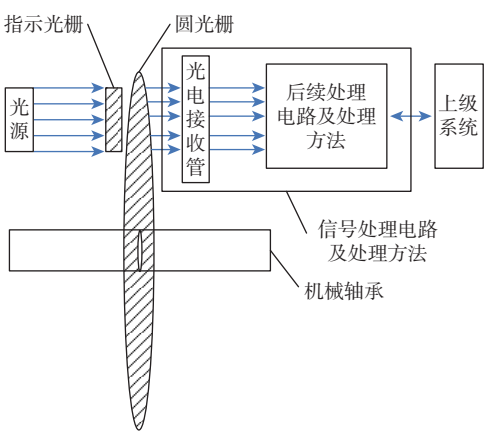


图 1　星载准绝对式光电编码器的组成

Fig. 1　Composition of satellite-borne absolute-to-be optical encoder

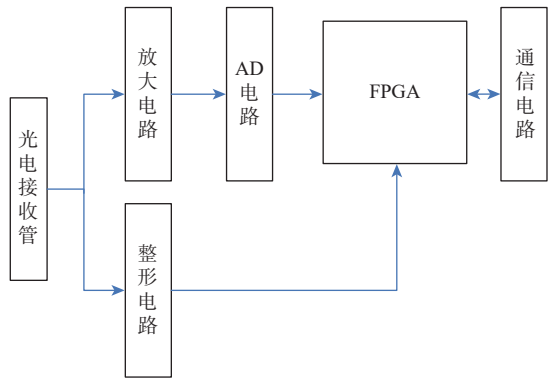


图 2　星载准绝对式光电编码器处理电路原理框

Fig. 2　Block of satellite-borne absolute-to-be optical encoder

件选用要充分考虑航天环境使用要求。因此, 核心器件需满足宇航级器件要求, 保证电路稳定可靠。

2　抗污方法

星载准绝对式光电编码器的圆光栅按照一定的规律在一圈设置多个参考点。转过一个小角位移, 保证经过 2 个参考点即可确定绝对角位移。但这种识别方法易受使用环境污染影响产生虚假参考点, 一旦产生虚假参考点, 则按照两参考点识别得到的绝对角位移是错误的。因此, 提出多参考点绝对角位移识别方法消除虚假参考点。而光栅环的污染, 则通过设置 3 个测量读数头的方式, 采用多读数头数据融合和跳码识别的方法, 剔除角位移测量不正确的读数头数据, 保证编码器工作正常。

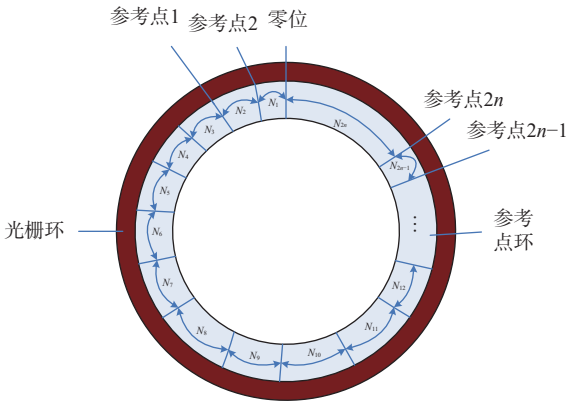
2.1　参考点抗污方法

2.1.1　多参考点圆光栅设计

多参考点准绝对式光电编码器的圆光栅在单参考点的基础上按照规律插入多个参考点, 只需要旋转小角位移, 就能确定绝对角位移。因此, 其绝对角位移识别效率比单参考点高。如图 3 所示, 是一种常见的编码方式。首先, 将圆光栅按照均分刻线数 m 插入参考点, 其个数为 n, 每 2 个参考点之间的刻线数为 m/n; 其次, 在每 2 个参考点之间插入 n 个参考点, 插入的方法如下:

$$\begin{cases} N_1 = \dfrac{m}{2n} - 1 \\ N_2 = \dfrac{m}{2n} + 1 \\ N_3 = \dfrac{m}{2n} - 2 \\ N_4 = \dfrac{m}{2n} + 2 \\ \quad\vdots \\ N_{2n-1} = \dfrac{m}{2n} - n \\ N_{2n} = \dfrac{m}{2n} + n \end{cases} \tag{1}$$

式中：N_1～N_{2n}从绝对角位移的零点逆时针依次排布，得到的圆光栅图案。其设计图和实物图如图3所示。



(a) 设计图



(b) 实物图

图3 星载准绝对式光电编码器圆光栅设计

Fig. 3 Design of satellite-borne absolute-to-be optical encoder's circle grating

2.1.2 多参考点绝对位置识别方法

由于多参考点一圈设置了多个参考点，因此，采用过3次或4次参考点决定一个绝对位置的方法。本文方法的流程如图4所示。

利用相邻2个不同参考点之间的线条数预判下一个参考点之间的线条数，进而确定绝对角位移。若$N_{2n}\geqslant N_i+N_{i+1}$，其中，$i$=2,3,⋯, 2$n$-1；当虚假参考点出现在$N_1$～$N_{2n}$任一区域，或任意几个区域时，只要剩余连续3个区域(即4个参考点)未被污染，即可得到绝对角位移n_5。若$N_{2n}<N_i+N_{i+1}$，其中，i=2, 3, ⋯, 2n-1，则只需经3个参考点即可得到绝对角位移，即如图4中遇到第3个参考点得到的绝对角位移n_4，无需第4参考点。

2.2 光栅环抗污染方法

采用三读数头测量数据融合的方法，克服光栅环的污染问题。具体如下：设3个测量读数头测量得到的角位移分别为A_1、A_2、A_3，第2测量读数头逆时针方向与第1测量读数头相差α，第3测量读数头逆时针方向与第1测量读数头相差β。融合后的角位移值A表达式为

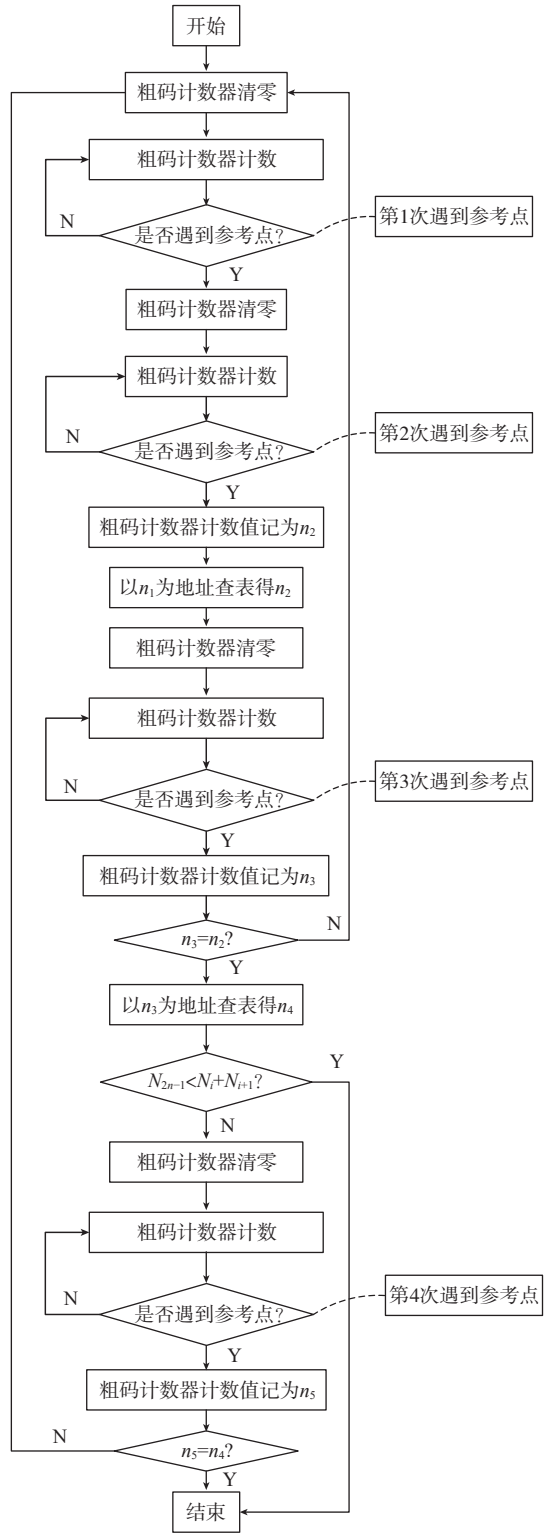


图4 星载准绝对式光电编码器绝对角位移识别方法流程

Fig. 4 Flow of the method for identify absolute angular displacements of on-board quasi-absolute photoelectric encoders

$$A=\begin{cases}\{\gamma_1A_1+\gamma_2[360°-(\alpha-A_2)]+\gamma_3[360°-(\beta-A_3)]\}/\\ \quad\gamma_1+\gamma_2+\gamma_3 & A_2<\alpha,A_3<\beta\\ \{\gamma_1A_1+\gamma_2(A_2-\alpha)+\gamma_3[360°-(\beta-A_3)]\}/\gamma_1+\gamma_2+\gamma_3 & A_2>\alpha,A_3<\beta\\ \{\gamma_1A_1+\gamma_2[360°-(\alpha-A_2)]+\gamma_3(A_3-\beta)\}/\gamma_1+\gamma_2+\gamma_3 & A_2<\alpha,A_3>\beta\\ [\gamma_1A_1+\gamma_2(A_2-\alpha)+\gamma_3(A_3-\beta)]/\gamma_1+\gamma_2+\gamma_3 & A_2>\alpha,A_3>\beta\\ 0° & A_2=\alpha,A_3=\beta\end{cases}\tag{2}$$

式中: $0°\leqslant\alpha\leqslant360°$, $0°\leqslant\beta\leqslant360°$; γ_1、γ_2 和 γ_3 至多有一个为 0, 当第 i ($1\leqslant i\leqslant3$, 且为整数) 个测量读数头经过被污染圆光栅区时, γ_i=0, 反之 γ_i=1。

本文方法将 3 个测量读数头均布在圆光栅上, 即在圆光栅上相隔 120°位置分别放置一个测量读数头, 即第 2 测量读数头逆时针方向与第 1 测量读数头相差 240°, 第 3 测量读数头逆时针方向与第 1 测量读数头相差 120°, 如图 5 所示。

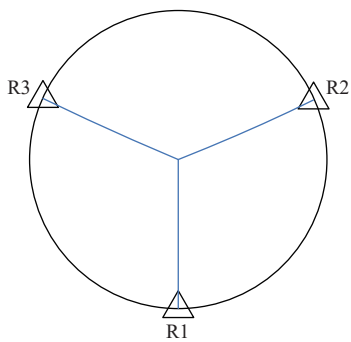


图 5　星载准绝对式光电编码器测量读数头分布

Fig. 5　Distribution of measurement readheads for on-board quasi-absolute optical encoders

若 3 个测量读数头均未经过被污染的圆光栅区最终的角位移 A 的计算式为

$$A=\begin{cases}\{A_1+[360°-(240°-A_2)]+[360°-(120°-A_3)]\}/3 & A_2<240°,A_3<120°\\ \{A_1+[360°-(240°-A_2)]+(A_3-120°)\}/3 & A_2<240°,A_3>120°\\ \{A_1+(A_2-240°)+[360°-(120°-A_3)]\}/3 & A_2>240°,A_3<120°\\ [A_1+(A_2-240°)+(A_3-120°)]/3 & A_2>240°,A_3>120°\\ 0° & A_2=240°,A_3=120°\end{cases}\tag{3}$$

式中: A_2、A_3 经过旋转到 A_1, 再取 3 个角位移的平均值作为光电编码器的最终结果 A。

设第 i 个读数头经过被污染圆光栅区域的判据如下: 连续 2 次角位移的角增量小于刻线数的最小分辨率, 则认为是正常的, 因为圆光栅被污染会引起莫尔条纹整形或细分错误, 角位移表现为粗码跳动一个分辨率:

$$\begin{cases}\Delta A_i=\left|A_{it_2}-A_{it_1}\right|<360°/l & A_i\text{正常}\\ \Delta A_i=\left|A_{it_2}-A_{it_1}\right|\geqslant360°/l & A_i\text{异常}\end{cases}\tag{4}$$

式中: i 为 1、2 或 3; t_1 和 t_2 为连续 2 个采样时刻; l 为圆光栅光栅环的刻线数, $360°/l$ 为星载准绝对式光电编码器的粗码分辨率。如果 A_i 异常, 则该读数头测量测量得到的连续 2 个采样周期测量角位移差 ΔA_i 就会出超出粗码分辨率, 需要舍弃。此时, 使用另外 2 个正常工作的测量读数头测量得到的角位移, 取二者的平均值作为 A。如假设 i=3 即第 3 测量读数头经过圆光栅污染区, 则 A_3 需要被舍弃。所以 A 为第 1 测量读数头和第 2 测量读数头的角位移融合:

$$A=\begin{cases}\{A_1+[360°-(240°-A_2)]\}/2 & A_2<240°\\ [A_1+(A_2-240°)]/2 & A_2>240°\\ 0° & A_2=240°\end{cases}\tag{5}$$

当 ΔA_i 取值恢复正常, 则读数头 i 需重新按照图 4 所示方法确定绝对角位移, 无误后在与其他两头按照式 (3) 合成最终的角位移 A。

3　实　验

3.1　抗污实验

为验证本文方法的有效性, 设计了准绝对式角位移测量装置, 其中, 圆光栅直径为 112 mm, 光栅刻线数为 16 384, 然后, 先均匀插入 64 个参考点, 再按照式 (1) 插入 64 个参考点。在圆光栅每相隔 120°均布 3 个读数头, 经 1 024 细分后, 实现 24 bit 角位移测量。分别采用不透明的航天用润滑油和透明的酒精作为污渍, 滴在准绝对式角位移测量装置的圆光栅上。

1) 航天用润滑油。在圆光栅上, 放置一处航天用润滑油作为油污, 该油污覆盖圆光栅的部分光栅线条和参考点。具体如下: 用滴管取一滴 (约 0.05 mL) 的宇航用润滑油, 滴在圆光栅上, 圆光栅的编码环和光栅环同时被油污污染, 直径尺寸约为 8.0 mm 如图 6 所示。

测量得到的莫尔条纹信号与无污渍的对比如图 7 所示。

由图 7 可知, 被污染的圆光栅产生的莫尔条纹信号有明显的漂移, 这种漂移导致角位移的测量出现跳数, 对星载准绝对式光电编码器的影响是灾难性的。

未经过本文方法的角位移曲线和经过本文方

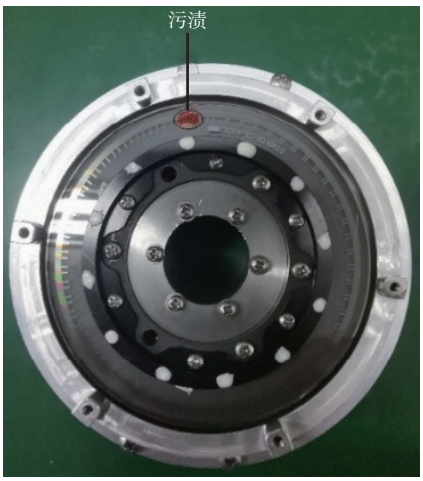
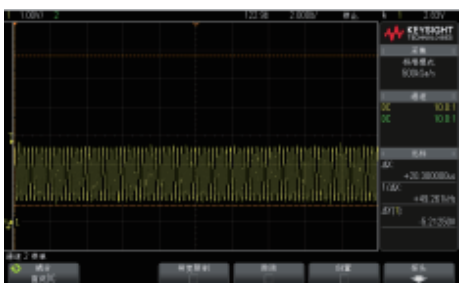


图 6 圆光栅被航天用润滑油污染

Fig. 6 Circle grating with aerospace lube stain



(a) 被航天用润滑油污染的莫尔条纹信号



(b) 未被航天用润滑油污染的莫尔条纹信号

图 7 莫尔条纹信号对比

Fig. 7 Comparison of Moore's stripe signals

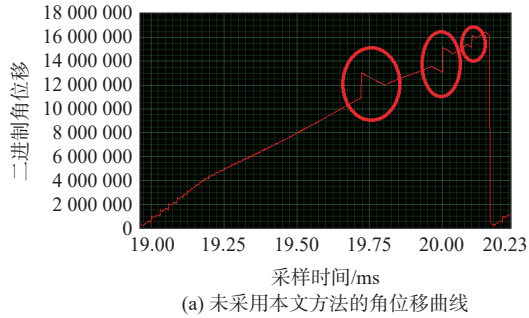
法的角位移曲线对比如图 8 所示。

由图 8 可知，未经本文方法角位移有明显的错误，而经过本文方法后角位移曲线正常。为进一步验证本文方法的有效性，将角速度曲线也测量出来，即：连续 2 个采样时刻的角位移差与采样时间间隔的比，此值应小于粗码的分辨率与采样时间的比：

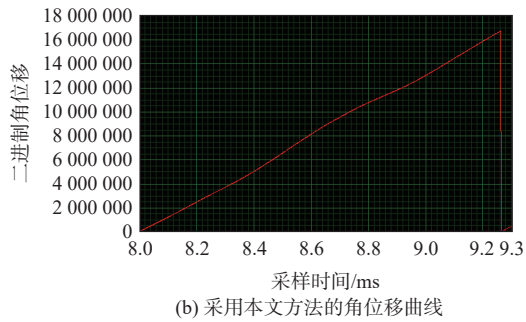
$$\omega < \frac{360°}{lt_{\text{sample}}} \tag{6}$$

式中：ω 为角速度；t_{sample} 为采样时间间隔；l 为圆光栅光栅环的刻线数。未经过本文方法的角速度曲线和经过本文方法的角速度曲线对比如图 9 所示。

由图 9 可知，未经本文方法的角速度曲线明显有跳动大于粗码分辨率，这种跳动是错误的。而经过本文方法后，角速度曲线恢复正常。因此，本文方法可有效的克服圆光栅污染对角位移测量引起的影响。



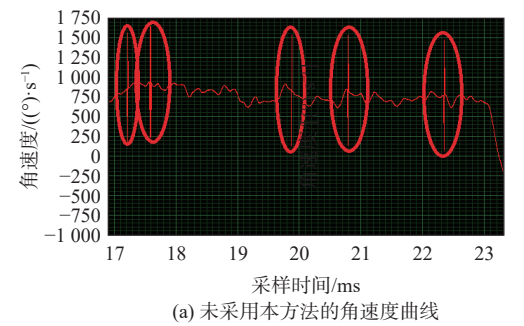
(a) 未采用本文方法的角位移曲线



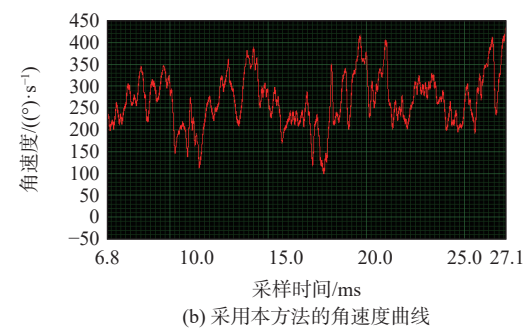
(b) 采用本文方法的角位移曲线

图 8 角位移结果对比

Fig. 8 Comparison of angle shift curve



(a) 未采用本方法的角速度曲线



(b) 采用本方法的角速度曲线

图 9 角速度结果对比

Fig. 9 Comparison of angular velocity curve

2）酒精。为对比油污透明度的影响，用滴管在圆光栅上滴入一滴酒精，该滴酒精尺寸约为 10 mm，其位置同时覆盖圆光栅的编码环和光栅环如图 10 所示。

分别测量未被酒精污染和被酒精污染的莫尔条纹信号，如图 11 所示。

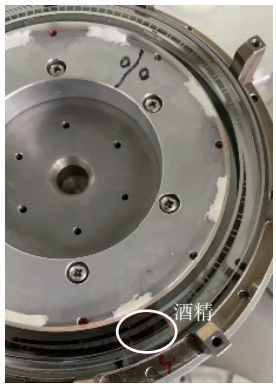
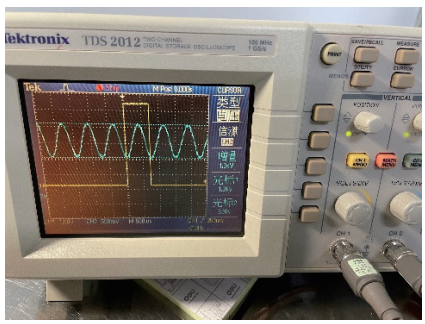
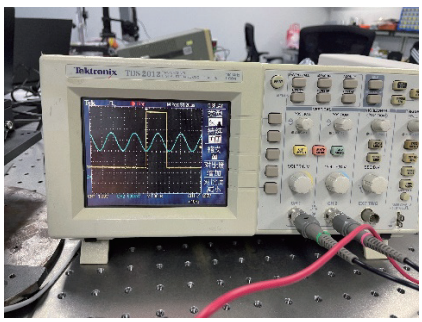


图 10　圆光栅被酒精油污染

Fig. 10　Circle grating with alcohol stain



(a) 未被酒精污染的莫尔条纹信号



(b) 被酒精污染的莫尔条纹信号

图 11　莫尔条纹对比

Fig. 11　Comparison of Moore's Stripes

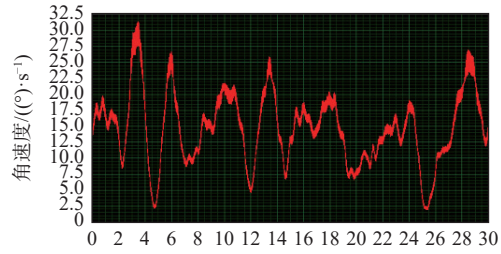
由图 11 可知, 蓝色的莫尔条纹未发生漂移, 黄色的参考环信号正常, 未消失。可见酒精由于透明, 未对准绝对式光电编码器的信号产生影响。

分别测量 2 种情况的角速度曲线, 如图 12 所示。

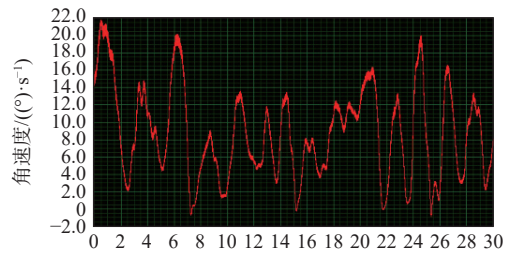
由图 12 可知, 被酒精污染相对未被酒精污染得到的角速度曲线无明显跳动, 由于酒精是透明的, 因此, 其污染圆光栅不会对测量的结果产生灾难性的误码影响。

3.2　精度检测

采用自准直平行光管结合 23 面体检测准绝对式光电编码器精度, 23 面体与星载准绝对式光电编码器的轴系通过检验工装连接, 准绝对式光电编码器和自准直平行光管的测量结果通过 PC 显示, 如图 13 所示。



(a) 被酒精污染的角速度曲线



(b) 被酒精污染的角速度曲线

图 12　角速度结果对比曲线

Fig. 12　Comparison of angular velocity curves

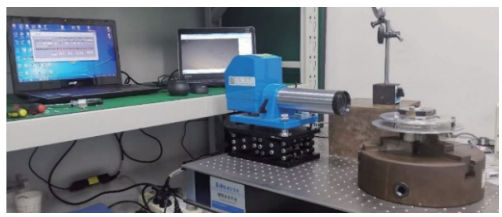


图 13　精度检测系统

Fig. 13　Accuracy detection system

23 面体相邻 2 个面与分别自准直平行光管垂直时编码器转过的角度为 360°/23。首先, 调整编码器识别绝对角位移, 并使绝对角位移显示为 0, 此时再调整自准平行直光管的光轴与 23 面体的第一面垂直, 即自准直平行光管的读数为 0, 当与第 i (1≤i≤23) 面垂直时, 理论上转过的角度为

$$\theta = (i-1)\times(360°/23) \tag{7}$$

此时准绝对式光电编码器显示的读数 θ' 与 θ 之差即为此位置处的误差。在实际测量中, 编码器的转动比较好操作, 故将编码器转到 (i−1)×(360°/23) 位置处, 此时自准直平行光管的误差即为编码器在此处的误差, 由于精度比较高, 故计算时只需将理论角度的角秒计算进去即可。最后, 最终误差还要将编码器的分辨率误差和 23 面体的修正误差考虑进去。检验结果如表 1 所示。

由表 1 可知, 星载准绝对式编码器: 角位移误差最大值为 4.6", 最小值为−0.6", 峰峰值为 5.2", 均方差为 4.3"。满足系统对角位移测量的精度要求。

表1 星载准绝对式光电编码器精度检测结果

Table 1 Accuracy test results of satellite-borne absolute-to-be optical encoder's

测量位置	$\theta'-\theta$ /(")	理论值/(")	修正值/(")	最终精度/(")
1	0	0	0	0
2	10.5	7.8	−0.5	3.2
3	18.0	15.7	−0.3	2.6
4	22.6	23.5	−0.3	−0.6
5	32.5	31.3	−0.2	1.4
6	40.0	39.1	−0.5	1.4
7	48.8	47.0	0.2	1.7
8	57.5	54.8	0.7	2.0
9	4.2	2.6	0.4	1.2
10	11.7	10.4	0.4	0.8
11	18.9	18.3	−0.2	0.8
12	24.6	26.1	−0.7	−0.8
13	35.3	33.9	−0.3	1.6
14	44.6	41.7	0.1	2.7
15	51.7	49.6	−0.9	3.1
16	61.0	57.4	−0.2	3.8
17	7.6	5.2	−0.8	3.2
18	14.9	13.0	−0.4	2.3
19	22.8	20.9	−0.7	2.7
20	32.2	28.7	−1.1	4.6
21	39.1	36.5	0.1	2.4
22	47.0	44.3	−0.2	2.9
23	52.3	52.2	0.3	−0.2

4 结 论

本文方法解决了因圆光栅被污染造成角位移错误。

1）通过角位移曲线和角速度曲线实验结果表明：当圆光栅被污染，莫尔条纹信号输出异常，角位移测量依旧正确。

2）经检验星载准绝对式光电编码器的误差最大值为4.6"，最小值为−0.6"，峰峰值为5.2"，均方差为4.3"。本文方法已在实际工程中应用，效果良好。

3）本文方法对提高星载准绝对式光电编码器的抗油污能力，防止油污造成灾难性的影响，保证测量的角位移正确可靠，有实用价值。

参考文献（References）

[1] 董全睿, 陈涛, 高世杰, 等. 星载激光通信技术研究进展[J]. 中国光学, 2019, 12(6): 1260-1270.
DONG Q R, CHEN T, GAO S J, et al. Progress of research on satellite-borne laser communication technology[J]. Chinese Optics, 2019, 12(6): 1260-1270(in Chinese).

[2] 秦涛, 郭骏立, 张美丽, 等. 星载二维转台结构设计及刚度分析[J]. 红外与激光工程, 2022, 51(5): 300-308.
QIN T, GUO J L, ZHANG M L, et al. Structure design and stiffness analysis of spaceborne two-dimensional turntable[J]. Infrared and Laser Engineering, 2022, 51(5): 300-308(in Chinese).

[3] 韩旭东, 徐新行, 刘长顺, 等. 用于星载激光通信终端的绝对式光电角度编码器[J]. 光学 精密工程, 2016, 24(10): 2424-2431.
HAN X D, XU X H, LIU C S, et al. Absolute optical angle encoder used for laser communication terminal on satellite platform[J]. Optics and Precision Engineering, 2016, 24(10): 2424-2431(in Chinese).

[4] 叶盛祥. 光电位移精密测量技术[M]. 成都: 四川科学技术出版社, 2003.
YE S X. Precise photoelectric displacement measurement technology [M]. Chengdu: Science and Technology Press, 2003 (in Chinese).

[5] 刘长顺, 王显军, 韩旭东, 等. 八矩阵超小型绝对式光电编码器[J]. 光学 精密工程, 2010, 18(2): 326-333.
LIU C S, WANG X J, HAN X D, et al. Ultra miniature absolute optical encoders based on eight-matrix coding[J]. Optics and Precision Engineering, 2010, 18(2): 326-333(in Chinese).

[6] 王亚洲, 于海, 易进, 等. 图像式角位移测量的光栅偏心度监测系统[J]. 光学 精密工程, 2020, 28(5): 1038-1045.
WANG Y Z, YU H, YI J, et al. Grating eccentricity monitoring system for image-based angular displacement measurement[J]. Optics and Precision Engineering, 2020, 28(5): 1038-1045(in Chinese).

[7] YU H, WAN Q H, ZHAO C H, et al. Error compensation for low-density circular gratings based on linear image-type angular displacement measurements[J]. IEEE Transactions on Industrial Electronics, 2022, 69(12): 13736-13743.

[8] 刘永坤, 丁红昌, 向阳, 等. 利用反射式圆光栅的振镜转角测量[J]. 中国光学, 2021, 14(3): 643-651.
LIU Y K, DING H C, XIANG Y, et al. Rotational angle measurement of galvanometer using reflective circular grating[J]. Chinese Optics, 2021, 14(3): 643-651(in Chinese).

[9] 韩庆阳, 陈赟, 张红胜, 等. 耐高温增量式光电编码器的研制[J]. 光学 精密工程, 2019, 27(7): 1458-1464.
HAN Q Y, CHEN Y, ZHANG H S, et al. Development of high-temperature resistant incremental encoder[J]. Optics and Precision Engineering, 2019, 27(7): 1458-1464(in Chinese).

[10] 裘祖荣, 周磊, 薛浩, 等. 精密减速器检验仪测角误差补偿[J]. 光学 精密工程, 2021, 29(11): 2662-2631.
QIU Z R, ZHOU L, XUE H, et al. Angle error compensation of precision reducer tester[J]. Optics and Precision Engineering, 2021, 29(11): 2662-2631(in Chinese).

[11] 于连栋, 鲍文慧, 赵会宁, 等. 新型圆光栅测角误差补偿方法及其应用[J]. 光学 精密工程, 2019, 27(8): 1719-1726.
YU L D, BAO W H, ZHAO H N, et al. Application and novel angle measurement error compensation method of circular gratings[J]. Optics and Precision Engineering, 2019, 27(8): 1719-1726(in Chinese).

[12] YU H, WAN Q H, ZHAO C H, et al. Anti-stain algorithm of angular displacement based on a single image sensor[J]. Applied Optics, 2020, 59(7): 1985-1990.

[13] 张文颖, 劳达宝, 周维虎, 等. 基于多头读数布局的圆光栅自校准方法研究[J]. 光学学报, 2018, 38(8): 0812001.
ZHANG W Y, LAO D B, ZHOU W H, et al. Self-calibration meth-

od based on multi-head reading layout[J]. Acta Optica Sinica, 2018, 38(8): 0812001(in Chinese).

[14] 杨雪, 梁煜, 张为, 等. 应用于反射式编码器的光电集成芯片设计[J]. 光学 精密工程, 2023, 31(8): 1136-1149.
YANG X, LIANG Y, ZHANG W, et al. Design of an optoelectronic integrated chip for reflective encoder application[J]. Optics and Precision Engineering, 2023, 31(8): 1136-1149(in Chinese).

[15] 殷景志, 余芝帅, 彭火山, 等. 小型伪随机绝对式光电编码器芯片[J]. 光学 精密工程, 2016, 24(10): 169-174.
YIN J Z, YU Z S, PENG H S, et al. Applied of pseudo-rondom code in miniature absolute encoder chip[J]. Optics and Precision Engineering, 2016, 24(10): 169-174(in Chinese).

Anti-stain method of satellite-borne absolute-to-be optical encoder

HAN Qingyang*, SHEN Honghai, MA Tianxiang, LIANG Chao, WANG Zhichong

(Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China)

Abstract: In order to improve the pollutant resistance of satellite-borne absolute-to-be optical encoder, a method based on multiple references and probes is proposed. First, the composition and circuit design of the satellite-borne absolute-to-be optical encoder was introduced. Second, the data from several probes was combined and references were utilized to differentiate absolute angle shift. This solved the problem which is caused by contaminated circle grating. At last, angular curve and velocity were employed to prove this. The accuracy is tested using autocollimation and a 23-surface polyhedron. The outcome demonstrates its effectiveness. The biggest error is 4.6", the smallest error is −0.6", the pink is 5.2", and the RSME is 4.3". This method could meet the requirements and has been successfully applied in engineering. The proposed method has practical value for improving the oil pollution resistance of spaceborne quasi-absolute optical encoders.

Keywords: satellite-borne; absolute-to-be optical encoder; pollutant-resistance; multiple reference; multiple probes

Received: 2024-04-07; **Accepted**: 2024-07-30; **Published Online**: 2024-08-02 10:14

URL: link.cnki.net/urlid/11.2625.V.20240801.1544.001

Foundation item: National Natural Science Fouddation of China (52075520)

* **Corresponding author**. E-mail: lanbohesky@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0408

基于深度图卷积网络的自监督群体发现模型

王栽胜，汪晓锋*，沈国栋，张增杰，全大英
(中国计量大学 信息工程学院，杭州 310018)

摘 要：为缓解图神经网络在群体发现任务中对先验知识的过度依赖并提高识别准确性，提出了一种基于自监督学习和深度图卷积网络(GCN)的群体发现模型。该模型充分利用少量标记节点的语义特征，并通过语义对齐机制获得未知节点的伪标签，从而引入一种自监督学习模块以缓解GCN模型训练过程中对大量先验标签的依赖性。同时，为通过获取网络全局信息以提高群体识别的准确性，通过堆叠多个自监督学习模块构建一种深度图自监督学习模型，并利用初始残差和恒等映射2种策略来克服深度模型带来的过平滑问题。在公开数据集上的实验表明：在给定少量先验标签和加深模型深度的情况下，所提模型与现有模型相比在群体识别精度上表现出了明显优势。

关 键 词：复杂网络；群体发现；图卷积网络；残差连接；自监督学习
中图分类号：TP39
文献标志码：A **文章编号**：1001-5965(2025)06-2022-11

目前，复杂网络已经成为描述和建模复杂系统的通用语言[1]，其将复杂系统中的个体及个体间的相互关系表示为由节点和连边组成的图结构。同时，复杂网络广泛应用于现实网络场景中，如社交网络、生物网络、引文网络等[2]。大量研究表明，复杂网络中的节点呈现出普遍的群体聚集特性，群体结构已成为复杂网络的一种重要中观结构特征[3]。群体结构通常与复杂系统中相应功能模块及群体属性有着紧密联系，因此，有效揭示复杂网络中的群体结构对于分析网络节点特征、结构属性、群体行为模式等问题有着重要意义，同时为研究网络动态演化机制、信息传播规律、群体行为分析等提供重要的理论支撑[4]。群体发现是网络科学中的重要研究问题之一，其相关技术已大量应用于社交网络分析、信息检索与推荐、公共安全事件管控与生物结构预测等领域[5]。近年来，随着信息技术和大数据的快速发展，网络应用场景变得复杂多样，除了复杂的拓扑结构外，还存在大量的与节点及网络相关的属性信息，如文本、图像、节点语义等，这给群体发现任务带来了极大挑战[6]。

群体发现旨在从未标记的网络中识别具有潜在相似属性或功能的节点集合，通常被当作一种节点聚类任务来研究。由于不同领域学者的研究视角不同，目前对群体结构没有统一的定义，从而导致群体发现方法多样但普适性较低。早期的群体发现工作也称社区发现，主要基于网络拓扑特性将整个网络按某种相似性或划分标准划分为不同子图结构，典型的方法有基于图划分和节点/边聚类等[7]。然而，针对当前多源异构的复杂网络，传统方法很难同时整合高维结构信息和内容属性信息，如矩阵分解[8]、贝叶斯模型[9]等。为解决该问题，近年来深度学习技术不断被引入到图数据挖掘中，形成了新兴的图深度学习研究领域[10]。基于图深度学习的群体发现方法通过对高维图数据进行分布式表示学习，并提取具有判别性的节点特征用于群体识别。相比传统方法，图深度学习技术在网络群体

收稿日期：2023-06-21；**录用日期**：2023-11-21；**网络出版时间**：2023-12-06 13:16
网络出版地址：link.cnki.net/urlid/11.2625.V.20231205.1639.003
基金项目：浙江省自然科学基金(LQ20F020021)；国家重点研发计划(2019YFB1707104)
* **通信作者**. E-mail：xfwang@cjlu.edu.cn

引用格式：王栽胜，汪晓锋，沈国栋，等. 基于深度图卷积网络的自监督群体发现模型[J]. 北京航空航天大学学报，2025，51(6)：2022-2032. WANG Z S，WANG X F，SHENG G D，et al. Self-supervised learning for community detection based on deep graph convolutional networks[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2022-2032 (in Chinese).

发现任务中有效融合了高维图数据和节点属性, 从而提高群体发现的效率及准确性。此外, 图深度学习模型相比传统模型能更有效地处理大规模、稀疏的复杂网络数据。

目前, 许多用于群体发现的图深度学习模型相继被提出[11], 包括图卷积网络 (graph convolutional network, GCN)[12]、图注意力网络 (graph attention network, GAT)[13] 和图生成对抗网络 (graph generative adversarial network, GraphGAN)[14] 等。其中, GCN 是图深度学习的一个重要分支, 在各种图挖掘任务中广泛应用。对于群体发现问题, GCN 通过图卷积操作来聚合邻域节点信息, 可同时融合网络拓扑结构和节点属性来学习复杂的节点特征以识别群体。近年来, GCN 及其变体在群体发现任务中表现出了明显优势[15-17], 但大多数方法仍是基于浅层的 GCN 模型, 即利用 2～3 层图卷积操作来聚合邻域信息, 从而忽略了高阶邻域信息对群体发现的影响。然而, 简单地叠加更多的卷积层会导致模型过平滑问题, 即随着卷积层增加, 网络节点的特征表示会趋于相似而变得难以区分, 使得模型性能急剧下降。为此, 近年来一些学者提出了改进模型以解决深度 GCN 带来的过平滑问题[18-20]。尽管这些改进模型一定程度上缓解了过平滑问题, 但随着模型进一步加深, 性能下降明显。

同时, 作为一种半监督学习模型, GCN 模型在训练过程中依赖大量先验的带标签节点, 而真实网络场景下大多数节点的群体标签是未知的, 这给半监督的群体发现任务带来了极大挑战。近年来, 自监督学习 (self-supervised learning, SSL) 由于可缓解模型对先验知识的依赖而受到大量关注, 并逐渐应用于图深度学习领域[21]。自监督学习旨在根据数据本身设计辅助任务以挖掘数据的表征特性作为监督信息, 来提升模型的特征提取能力。You 等[22] 利用预先计算的聚类索引作为自监督信息, 提出一种自监督节点聚类方法。为同时利用网络结构和属性信息, 无监督图表示学习的集群感知图神经网络 (cluster-aware graph neural network, CAGNN) 模型[23] 利用聚类标签作为自监督信号来实现无监督的网络表示学习。此外, 多阶段自监督学习 (multi-stage self-supervised learning, M3S) 模型 [24] 及自监督语义对齐图卷积网络 (self-supervised semantic alignment graph convolution network, Self-SAGCN)[25] 等典型的图自监督学习模型相继被提出以缓解 GCN 对先验标签的过度依赖, 但这些方法仍基于浅层的网络结构。

综上所述, 针对群体发现任务, 目前基于 GCN 的图神经网络模型大都采用半监督学习范式, 尽管相对于无监督模型能够取得较高的识别精度, 但过度依赖于先验标签。同时, 浅层的 GCN 结构进一步限制了模型捕获全局信息的能力。尽管自监督学习技术的引入不同程度上缓解了模型对大量标签的依赖, 但模型性能仍会随着模型的加深而不断下降。因此, 如何设计一种不依赖大量训练标签同时保持较高性能的 GCN 模型仍然是一个值得研究的问题。

因此, 本文提出一种基于深度 GCN 的自监督群体发现模型。该模型在图卷积操作的基础上引入了一种自监督学习模块, 以利用特征聚合从少量标记节点中提取语义信息, 并利用语义对齐机制来获得未知节点的伪标签, 采用伪标签训练模型, 从而降低模型对已知标签的依赖性。同时, 为利用高阶领域信息来提高群体识别的准确性, 通过堆叠多层自监督学习模块构建一种深度图自监督学习模型, 利用初始残差和恒等映射 2 种策略来缓解深度模型带来的过平滑问题。通过在真实网络数据集的实验, 验证了该模型在给定少量先验标签的情况下, 加深模型深度仍能取得较好的效果。

1　相关工作

1.1　群体发现

群体发现作为网络科学中的一个重要课题, 在过去 10 年中得到了广泛研究。早期的大量研究工作对网络群体结构提出了各种不同定义, 并相应地设计了各种群体发现方法。传统方法包括层次聚类[26]、模块度优化[27] 及谱聚类[28] 等, 主要利用网络拓扑特征来识别群体, 也称为社区发现。Louvain 方法[29] 通过贪婪地优化模块度以揭示社区结构的层次性, 随机块模型 (stochastic block model, SBM)[30] 通过控制连接边生成的概率来生成不相交的群体结构。此外, 带有扰动的谱聚类 (spectral clustering with perturbations, SCP)[31] 方法通过归一化拉普拉斯矩阵和正则化邻接矩阵来划分网络。然而, 传统方法特别是涉及到矩阵分解的方法存在计算复杂度高的问题, 无法应用于大规模网络。同时, 随着当前网络环境的复杂化, 除网络拓扑场景外, 大量与网络或节点相关的高维属性信息未得到充分利用。因此, 大多数传统方法已无法适用于大规模、多源异构的复杂网络。

近年来, 深度学习技术被引入到图表示学习中以克服上述挑战[5]。基于图深度学习的群体发现方法能同时利用网络结构特征和高维属性信息, 可更好地适应复杂场景中的群体发现任务。GCN[12] 作

为图深度学习中最重要的图神经网络模型之一，在群体发现中被广泛应用。Jin等[15]提出了一个基于半监督GCN的MRFasGCN模型，该模型通过将马尔可夫随机场(Markov random field, MRF)扩展为一个新的卷积层来表征隐藏群体，再与GCN模型相结合以构成端对端的群体发现模型。为实现无监督群体发现，Hu等[16]提出了一种图卷积梯形网络(graph convolutional ladder-shape network, GCLN)模型，先对称地构建收缩路径和扩展路径，再从收缩路径中捕获上下文特征，与在扩展路径中学习到的本地化信息融合在一起进行群体发现。同时，Jin等[17]提出了一种通过嵌入属性网络的联合GCN进行社区检测(community detection via joint graph convolutional network embedding in attribute network, JGE-CD)模型，该模型由基于GCN的双编码器、群体发现模块及拓扑重构模块组成，通过联合嵌入学习过程来实现群体发现任务。此外，一些工作将对比学习与GCN相结合来识别属性网络中的群体结构。Zhang等[32]通过节点和群体上的互信息(mutual information, MI)联合优化图表示和聚类，提出了一种基于互信息最大化图对比学习的群体发现模型(community deep graph infomax, CommDGI)。Zhao等[33]提出了一种图去偏对比学习模型以同时进行节点表示学习和聚类，该模型通过将节点表示与聚类结果进行对齐以实现联合优化，从而学习到更具有区分性的节点表示和聚类结果。

总体来说，半监督的GCN[12]被大量用于群体发现问题，但GCN在训练过程中需要大量带标签节点，同时随着网络层次的加深，不可避免地带来了过平滑问题。为缓解深度模型带来的过平滑问题，基于GCN的各种变体及改进模型相继被提出。Xu等[18]提出的JKNet (jumping knowledge network)模型，以半监督的方式将前一层的输出与密集跳跃连接相结合，以保留节点的局部特征。在此基础上，Chen等[20]提出了GCNII(graph convolutional network via initial residual and identity mapping)模型，通过初始残差和恒等映射2种方法，极大地缓解了深度卷积网络带来的过度平滑问题。Rong等[19]通过随机删除输入图中的一些边来减少过平滑的影响，提出了一种DropEdge模型。此外，为降低训练模型对标签的依赖性，Wang等[34]提出了一种基于GCN的属性网络社区检测无监督学习模型，通过局部标签采样模型定位群体的结构中心，并通过GCN对网络拓扑和节点属性进行编码，从而在没有任何先验信息的情况下推断节点的群体标签。无监督方法减少了对已知标签的依赖，同时也降低了模型的性能。

1.2 图自监督学习

自监督学习通过精心设计的辅助任务来挖掘数据自身的表征特性作为监督信息，提升模型的特征提取能力，已成为图数据挖掘的一个重要方向[35]。由于图域的复杂网络有着复杂的数据结构，图域的自监督学习可通过设计各种辅助任务来获取监督信息。在图自监督学习中，图神经网络[10]通常作为编码器，从下游任务的输入图中学习节点表示。图自监督学习的核心是如何设计有效的辅助任务。现有相关工作中的辅助任务可主要分为3类：掩码特征回归、图对比学习和辅助属性预测(auxiliary property prediction, APP)。本文工作主要涉及基于辅助属性预测的图自监督学习。

基于辅助属性预测的自监督学习方法的基本思想是：通过预测结构和属性构建各种辅助任务来提供自监督信息。Jin等[36]提出通过预测节点特定属性或全局结构信息，如计算未标记节点和预定义集群之间的距离，以构建局部结构感知的辅助任务。此外，另一部分基于分类的辅助属性预测方法通过构建伪标签来指导模型训练。Sun等[24]引入了基于DeepCluster的多阶段自训练机制，并利用嵌入空间对齐机制来获得高置信度的伪标签以扩展标签数据，从而提出了一种M3S模型以缓解GCN对大量先验标签的依赖。然而，由于过平滑问题，如大多数基于GCN的自监督学习模型(包括M3S)都利用了GCN的浅层架构，极大限制了节点标签信息的有效传播。因此，Yang等[25]提出了Self-SAGCN模型，该网络通过特征聚合和语义对齐来缓解多层引起的过平滑和过拟合问题，然而，其仅在4层结构时达到了最佳性能，随着更多卷积层的堆叠，性能下降明显。因此，如何缓解深度模型的过平滑问题和降低模型对已知标签依赖性仍然是一个挑战。为此，本文引入了一种用于群体发现的深度自监督学习模型，该模型有效弥补了GCN及自监督学习方法存在的不足。

2 基于深度自监督GCN的群体发现

2.1 基本定义

对于一个无向的属性网络，可表示为图$G=(V,E,\boldsymbol{X})$，其中，$V=\{v_i\}_{i=1,2,\cdots,N}$表示网络节点集合，$N$表示节点数量，$E$为节点连边集合，$\boldsymbol{X}\in\mathbf{R}^{N\times F}$为网络属性矩阵，其包含各节点的$F$维实值属性向量。$\boldsymbol{A}\in\mathbf{R}^{N\times N}$表示网络对应的邻接矩阵，当节点$v_i$和节点$v_j$之间存在连边时，则$A_{ij}=1$，否则$A_{ij}=0$。

$\boldsymbol{D} \in \mathrm{diag}(d_1, d_2, \cdots, d_i)$表示图$G$对应的度矩阵, 其中, $d_i = \sum_{j=1}^{n} A_{ij}$。

GCN 将传统的卷积神经网络 (convolutional neural network, CNN) 推广到图域, 目前主要有空域 GCN 和频域 GCN 这 2 种类型, 本文主要采用频域 GCN, 也即常规意义下的 GCN。GCN 通过在频域上分解图信号并对频域分量应用频谱滤波器来定义卷积操作。图卷积的逐层传播规则可以表示为

$$\boldsymbol{H}^{(l+1)} = \sigma\left(\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{H}^{(l)}\boldsymbol{W}^{(l)}\right) \tag{1}$$

式中: $\widetilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}_N$为添加了自环的无向图$G$的邻接矩阵, $\boldsymbol{I}_N$为单位矩阵; $\boldsymbol{H}(l)$为第 l 层的输出矩阵, $\boldsymbol{H}^{(0)} = \boldsymbol{X}$, $l = \{0, 1, \cdots, L-1\}$为卷积层层数; $\boldsymbol{W}^{(l)}$为特定层的可训练权重矩阵; $\sigma(\cdot)$表示激活函数, 如 ReLU。

通过重归一化技巧, 将$\tilde{\boldsymbol{P}}$表示为$\widetilde{\boldsymbol{P}} = \widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}$, 则式 (1) 可进一步简化为

$$\boldsymbol{H}^{(l+1)} = \sigma\left(\widetilde{\boldsymbol{P}}\boldsymbol{H}^{(l)}\boldsymbol{W}^{(l)}\right) \tag{2}$$

2.2　模型整体框架

本文基于自监督学习和深度 GCN(self-supervision learning based on deep graph convolutional network, SDGCN) 的群体发现模型的整体结构如图 1 所示, 其中, 输入图中的不同颜色表示不同类别的节点。一方面, 为充分利用节点的语义信息, 模型在传统 GCN 基础上增加了一个语义特征提取通道, 并以自监督学习的方式来降低模型对大量先验标签的依赖; 另一方面, 为有效缓解过平滑影响, 模型引入了初始残差和恒等映射来加深模型深度。此外, 为加快模型训练过程, 本文模型的输入和输出端分别使用了一个全连接层以降低节点特征维度。从模型构成来看, 该模型可分为自监督学习模块和深度 GCN 这 2 个层次。具体地, 自监督学习模块利用语义对齐机制来获取节点特征为未知节点提供判别信息以扩展训练标签, 以自监督学习模块作为基础层进行叠加来构建深度自监督学习 GCN。同时, 在每层输入都融合初始输入, 以确保模型在堆叠了很多层后仍使得每个节点表示中至少保留了一部分来自输入层的原始特征。此外, 恒等映射策略的引入有效缓解了深度模型带来的训练参数过多的问题。与 Self-SAGCN 相比, 本文模型将带有自监督学习机制的 GCN 进行了加深, 更有效缓解了过平滑问题。

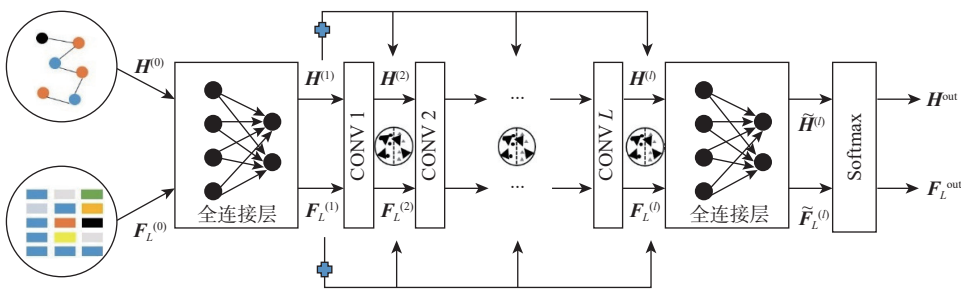


图 1　本文模型整体框架

Fig. 1　Overall framework of the proposed model

2.3　自监督学习模块

基于大量先验的训练数据, 传统 GCN 在半监督节点分类任务中取得了巨大的成功, 然而在群体发现任务中, 大多数节点的标签信息通常是难以获取的, 而人工标注成本是非常昂贵的。因此, 利用自监督学习技术设计辅助任务来挖掘无标签数据自身的表征特性作为监督信息以减少模型对先验知识的依赖成为了一种有效途径。对于群体发现来说, 属于同一群体结构的节点间的连接更加紧密, 同时更倾向于共享相同的语义特征及标签。因此, 在给定少量先验标记节点的情况下, 模型训练过程可利用 GCN 的局部特征聚合过程, 通过自监督学习的方式提取节点语义信息来分配伪标签并扩展训练集。因此, 本文基于 GCN 模型引入了一种自监督学习模块, 将属性图和先验标签信息分别作为 2 个不同的输入, 通过语义对齐机制提取监督信息, 该模块结构如图 2 所示, 其中, 输入图中的不同颜色表示不同类别的节点。

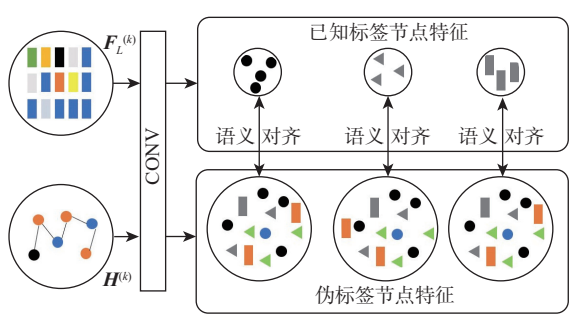


图 2　自监督学习模块结构框图

Fig. 2　Structure of the self-supervised learning module

给定一个无向图G,将其特征矩阵$\boldsymbol{H}^{(k)}$作为第一个输入,k为卷积层层数,且$\boldsymbol{H}^{(0)}=\boldsymbol{X}$。为从已知节点学习语义特征,构造一个输入$\boldsymbol{F}_E^{(k)}$,通过单位矩阵$\boldsymbol{I}$的特征聚合提取标记节点集$E$的语义信息,其中,$\boldsymbol{F}_E^{(0)}=\boldsymbol{X}_E$表示带标签的节点属性。因此,该输入经过卷积层的输出可表示为

$$\boldsymbol{H}^{(k+1)}=\sigma\left(\widetilde{\boldsymbol{P}}\boldsymbol{H}^{(k)}\boldsymbol{W}^{(k+1)}\right) \tag{3}$$

$$\boldsymbol{F}_E^{(k+1)}=\sigma\left(\boldsymbol{I}\boldsymbol{F}_E^{(k)}\boldsymbol{W}^{(k+1)}\right) \tag{4}$$

为在模型训练期间利用来自少数标记节点的先验信息并指导特征$\boldsymbol{H}^{(k)}$,本文进一步采用语义对齐来确保$\boldsymbol{H}^{(k)}$和$\boldsymbol{F}_E^{(k)}$中来自同一群体的节点特征在邻域映射。对于$\boldsymbol{H}^{(k)}$和$\boldsymbol{F}_E^{(k)}$中的带标签节点,直接应用相应的标签作为集群信息,对于$\boldsymbol{H}^{(k)}$中的未标记节点,根据当前网络参数为节点分配伪标签。将伪标签与从已知标签中提取的语义信息进行语义对齐。具体来说,模型在训练过程中,利用已知标签的语义信息建立与伪标签节点间的语义对应关系,从而确定具有相同类别特征的未标记节点。为减轻伪标签对节点划分的负面影响,利用标记节点和未标记节点的中心相似性来提供额外的监督信息。自监督学习模块的损失定义如下:

$$\mathcal{L}_{\mathrm{S}}=\sum_{j=1}^{c}d\left(C_j\left(\boldsymbol{F}_E^{(k)}\right),C_j\left(\boldsymbol{H}^{(k)}\right)\right) \tag{5}$$

式中:d为平方欧几里得距离函数;$C_j(\cdot)$表示第j类特征的质心。此外,$\boldsymbol{H}^{(k)}$中的伪标签在$\boldsymbol{H}^{\mathrm{out}}$中采用。

另外,由于利用伪标签构建的质心在半监督群体发现任务中可能缺乏稳定性,模型在每次迭代过程中,用$\boldsymbol{H}^{(k)}$和$\boldsymbol{F}_E^{(k)}$对应的特征来更新$C_j^t\left(\boldsymbol{H}^{(k)}\right)$和$C_j^t\left(\boldsymbol{F}_E^{(k)}\right)$的质心,并添加质心最后一次迭代来抑制不稳定性。质心更新过程定义如下:

$$C_j^t\left(\boldsymbol{H}^{(k)}\right)\leftarrow(1-\alpha)C_j^t\left(\boldsymbol{H}^{(k)}\right)+\alpha C_j^{(t-1)}\left(\boldsymbol{H}^{(k)}\right) \tag{6}$$

$$C_j^t\left(\boldsymbol{F}_E^{(k)}\right)\leftarrow(1-\alpha)C_j^t\left(\boldsymbol{F}_E^{(k)}\right)+\alpha C_j^{(t-1)}\left(\boldsymbol{F}_E^{(k)}\right) \tag{7}$$

式中:$\alpha\in[0,1)$用来平衡权重。

2.4 深度GCN

尽管自监督学习模块理论上可利用节点语义特征来降低模型对先验标签的依赖,但浅层结构仍难以捕获高阶邻域信息而取得理想的群体识别精度,因此,需要进一步加深模型结构。然而,与原始GCN模型类似,叠加更多的自监督学习模块同样会带来过平滑问题。受ResNet[37]模型的启发,本文尝试在模型中引入初始残差连接和恒等映射2种模型优化策略。

研究表明,残差连接只能缓解深度GCN模型的过度平滑问题,随着层数的增加,其性能仍然会下降。因此,受GCNII模型启发,本文将残差连接替换为初始残差连接,区别在于:初始残差连接模拟ResNet中的跳跃连接,其构建初始层到输出层的连接,而不是深度模型中常用的逐层残差连接。在使用初始残差连接情况下,即使模型堆叠了更多层,也能确保最终节点表示保留至少一部分来自输入层的信息。与传统GCN模型不同,SDGCN群体发现模型提供2个不同的输入,其中,使用特征矩阵$\boldsymbol{H}^{(l)}$作为第一个输入,为提取已知标签中的语义信息,通过特征聚合构造另一个输入$\boldsymbol{F}^{(l)}$。因此,将初始残差加入到GCN模型后,网络模型的卷积层定义如下:

$$\boldsymbol{H}^{(l+1)}=\sigma(((1-\alpha_l)\widetilde{\boldsymbol{P}}\boldsymbol{H}^{(l)}+\alpha_l\boldsymbol{H}^{(1)})\boldsymbol{W}^{(l)} \tag{8}$$

$$\boldsymbol{F}^{(l+1)}=\sigma(((1-\alpha_l)\boldsymbol{I}\boldsymbol{F}^{(l)}+\alpha_l\boldsymbol{F}^{(1)})\boldsymbol{W}^{(l)} \tag{9}$$

式中:α_l为超参数;$\widetilde{\boldsymbol{P}}=\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}$为重标准化的图卷积矩阵;$\boldsymbol{I}$为单位矩阵;$\boldsymbol{H}^{(1)}$为第一层的初始残差连接。因此,将初始层$\boldsymbol{H}^{(1)}$添加到卷积层$\widetilde{\boldsymbol{P}}\boldsymbol{H}^{(l)}$中,通过$\alpha_l$来控制初始残差的比重,其完整形式为$(1-\alpha_l)\widetilde{\boldsymbol{P}}\boldsymbol{H}^{(l)}+\alpha_l\boldsymbol{H}^{(1)}$。

此外,模型同时引入了恒等映射作为补充策略。虽然初始残差极大缓解了深度图神经网络中过度平滑的问题,但多层卷积层带来的训练参数过多的问题仍有待解决。Klicpera等[38]提出了一种基于GCN的近似个性化传播模型APPNP,极大简化了训练参数。受APPNP的启发,本文模型将ResNet的恒等映射应用到权重矩阵上,当超参数β_l足够小时,本质上可以模拟APPNP。因此,恒等映射可有效提高模型学习效率,同时配合初始参数策略可确保深层模型达到与浅层模型相近的性能。实验结果也证明了恒等映射可有效提高模型的性能。基于自监督学习模块的深度GCN的卷积层重新定义如下:

$$\boldsymbol{H}^{(l+1)}=\sigma(((1-\alpha_l)\widetilde{\boldsymbol{P}}\boldsymbol{H}^{(l)}+\alpha_l\boldsymbol{H}^{(1)})((1-\beta_l)\boldsymbol{I}_n+\beta_l\boldsymbol{W}^{(l)})) \tag{10}$$

$$\boldsymbol{F}^{(l+1)}=\sigma(((1-\alpha_l)\boldsymbol{I}\boldsymbol{F}^{(l)}+\alpha_l\boldsymbol{F}^{(1)})((1-\beta_l)\boldsymbol{I}_n+\beta_l\boldsymbol{W}^{(l)})) \tag{11}$$

式中:β_l为超参数,单位矩阵$\boldsymbol{I}_n$添加到权重矩阵$\boldsymbol{W}^{(l)}$中,通过β_l来控制恒等矩阵权重,其完整形式为$(1-\beta_l)\boldsymbol{I}_n+\beta_l\boldsymbol{W}^{(l)}$,体现了恒等映射的思想。

将深度GCN的输出映射为节点标签,并在最小化交叉熵损失的约束下进行模型训练,该交叉熵损失函数定义如下:

$$\mathcal{E}_{\mathrm{D}}=-\sum_{i\in E}\sum_{j=1}^{c}(Y_{ij}\ln\boldsymbol{H}_{ij}^{\mathrm{out}}+Y_{ij}\ln\boldsymbol{F}_{ij}^{\mathrm{out}}) \tag{12}$$

式中: E为标记节点的集合; $Y_i(i \in E)$表示Y中第i个标记节点的标记指示向量; $\boldsymbol{H}^{\text{out}} = \text{softmax}(\widetilde{\boldsymbol{H}}^{(l)}) \in \mathbf{R}^{n\times c}$, $\boldsymbol{F}^{\text{out}} = \text{softmax}(\widetilde{\boldsymbol{F}}^{(l)}) \in \mathbf{R}^{n\times c}$, c为群体的数量。

将 2 部分损失联合起来, 得到整个模型最终的损失函数:

$$\mathcal{L} = \mathcal{L}_{\text{D}} + \lambda \mathcal{L}_{\text{S}} \tag{13}$$

式中: λ为超参数, 用于调整自监督信息权重。

3　实验结果与分析

本节通过实验验证本文模型在给定少量已知节点情况下的性能表现, 以及在加深模型深度的情况下对过平滑问题的有效抑制。实验主要采用 3 个真实的引文数据集: Cora、Citeseer、PubMed。Cora 数据集由 2 708 篇机器学习领域相关的论文组成, 一篇论文至少引用或被引用于另外一篇论文, 每篇论文的特征使用 1 433 个单词表示。Citeseer 数据集由 3 327 篇机器学习领域相关的论文组成, 保留论文间引用和被引用的关系, 每篇论文的特征使用 3 703 个单词表示。PubMed 数据集由 19 717 篇生物医学领域相关的论文组成, 保留论文之间的引用和被引用关系, 每篇论文使用 500 个单词表示。论文之间的引用和被引用关系均视为无向连接。实验数据集具体信息如表 1 所示。

表 1　数据集基本信息

Table 1　Basic information of datasets

数据集	节点数	边数	社群数	特征维度
Cora	2 708	5 429	7	1 433
Citeseer	3 327	4 732	6	3 703
PubMed	19 717	44 338	3	500

本文采用 4 种不同类型的基线模型进行对比实验: ①多层感知机 (multilayer perceptron, MLP) 作为简单的神经网络模型; ②传统的图神经网络模型, 包括 GCN[12] 和 GAT[13]; ③深度图神经网络模型 (deep adaptive graph neural network, DAGNN)[39] 模型; ④基于 GCN 改进的自监督学习模型 Self-SAGCN[25]。另外, 在深度模型实验中, 本文模型 SDGCN 与另外 3 种半监督深度模型 (即 GCNdrop[16]、JKNet[18] 和 GCNII[20]) 进行了对比。

实验参数设置如下: 本文在实验中分别将卷积层设置为 4/16/32/64 层, 分别进行 10 轮训练, 每次训练 1 000 次, 将结果取平均值。模型使用 Adam 优化器进行实验, 训练迭代 1 000 次, 学习率设置为 0.05, 权重损失设置为 5×10^{-4}。对于卷积层, 其超参数α_l设置为 0.1。在实验中, $\lambda \leftarrow \dfrac{2}{1+\text{e}^{-10p}} - 1$用于在训练的初始阶段抑制噪声标签, 其中, p为迭代次数, $\lambda \in (0,1)$随着迭代次数逐渐增加。其余基线模型按照论文的参数设置。在实验中, 评价指标采用聚类精度 (cluster accuracy, AC) 来评估机器学习中各种算法的性能。给定真实标签t_i和预测标签p_i, 聚类精度可以定义为

$$A_{\text{C}} = \frac{\sum_{i=1}^{n} \delta(t_i, \text{map}(p_i))}{N} \tag{14}$$

式中: n为数据总的个数; δ表示指示函数; t_i 表示样本 i 对应的真实标签; map(·)表示最佳类标的重现分配, 以确保统计的正确。

实验环境如下: 操作系统为 Windows10 64 bit, CPU 为 Intel(R)Core i7-7700HQ, 频率为 3.4 GHz, GPU 为 NVIDIAGeForce1060, 存储器为 DDR4 16 GB, 编程语言及环境为 Python3.6+Pytorch1.7.0。

3.1　少量已知标签实验

为验证本文模型 SDGCN 在少量标签情况下在群体发现任务中的性能, 基于 3 个基准数据集, 在分别提供少量已知标签情况下, 将 SDGCN 与其他基线模型进行对比, 实验结果如表 2 所示。

表 2　在不同数量的已知标签条件下模型聚类精度比较

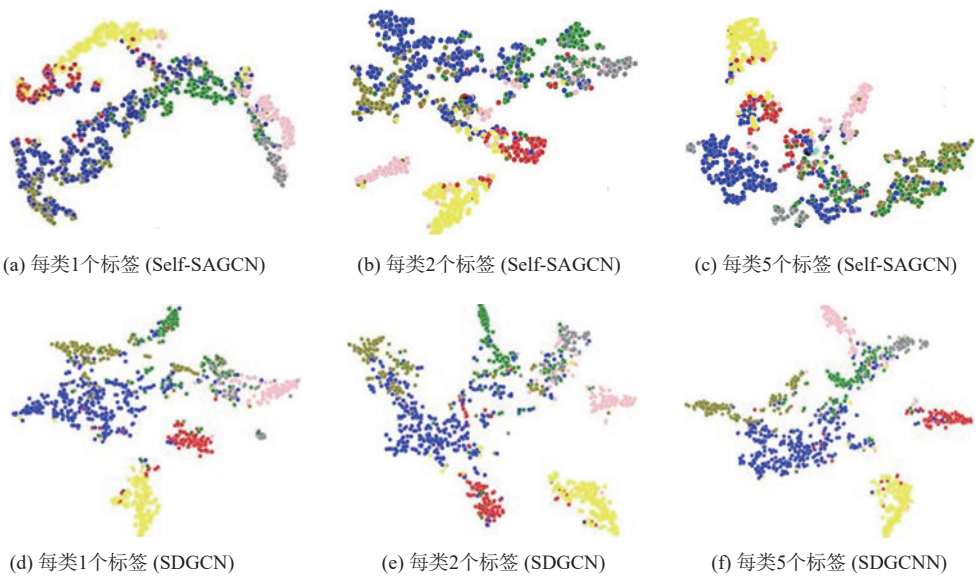
Table 2　Comparison of clustering accuracy of various models under different numbers of known labels　%

模型	每类1个标签			每类2个标签			每类5个标签		
	Cora	Citeseer	PubMed	Cora	Citeseer	PubMed	Cora	Citeseer	PubMed
MLP	42.5	27.9	49.5	48.9	33.6	55.9	59.9	43.2	65.1
GCN[12]	43.2	33.6	51.9	50.1	41.8	54.9	69.0	53.2	68.1
GAT[13]	44.6	33.9	52.1	59.6	44.1	58.6	70.6	54.3	68.8
DAGNN[39]	59.1	44.3	56.3	64.2	53.8	64.8	71.2	54.7	70.6
Self-SAGCN[25]	61.2	56.8	63.2	70.8	63.9	68.8	75.1	66.9	71.9
SDGCN	63.8	57.8	65.3	71.5	64.4	70.9	78.6	68.1	73.5

从表 2 中可以看出, 简单的神经元网络 MLP 模型相比传统的 GCN 模型性能较差。而深度半监督模型 DAGNN 的性能比浅层模型 GCN 和 GAT 性能表现更好, 验证了深度模型确实能够提升模型性

能。本文模型SDGCN和DAGNN相比，在每类仅提供1个已知标签情况下，在聚类精度上分别提升了4.7%、13.5%、9.0%；在每类提供2个已知标签情况下，在聚类精度上分别提升了7.3%、10.6%、6.1%。随着训练规模的增加，SDGCN模型仍然优于DAGNN模型，证明了本文模型的有效性。其原因在于：本文模型引入了自监督学习模块，大大降低了群体发现中对先验标签的依赖。即使与自监督学习模型Self-SAGCN相比，SDGCN模型在每类提供1个已知标签情况下，在聚类精度上仍有提升，分别提升了2.6%、1.0%、2.1%。这是由于本文模型使用了更深层次的模型架构，可有效获取高阶邻居信息来提升群体发现准确性。

图3为在不同数量标记节点条件下SDGCN模型和Self-SAGCN模型在Cora数据集上获得的节点特征分布t-SNE可视化结果。可以看出，在不同数量的已知标签下，SDGCN模型相比Self-SAGCN模型获得了更清晰的聚类特征。随着先验标签数量的增加，其结果变得更具区分性，这也从聚类精度的提高上得到了证实。



(a) 每类1个标签 (Self-SAGCN) (b) 每类2个标签 (Self-SAGCN) (c) 每类5个标签 (Self-SAGCN)

(d) 每类1个标签 (SDGCN) (e) 每类2个标签 (SDGCN) (f) 每类5个标签 (SDGCNN)

图3 不同已知标签数量条件下2种模型在Cora数据集上获得的节点特征t-SNE可视化

Fig. 3 t-SNE visualization of node features obtained by two model on Cora dataset under different numbers of known labels

3.2 深度模型实验

为进一步验证本文模型能够有效缓解深度架构所造成的过平滑影响，实验将对比SDGCN模型与其他深度模型在不同层数深度条件下的性能表现。为保证所有基线模型有一致的实验条件，实验在给定每类20个已知标签情况下，调节不同的模型深度来验证包括SDGCN在内的深度模型在本地环境中的性能。实验结果如表3所示。从实验结果可以发现，SDGCN模型能够有效缓解过度平滑的问题，在将自监督学习模块扩展到深度模型时，其聚类精度仍然优于其他模型，且SDGCN模型聚类精度随着层数的增加仍在不断上升。

与GCNII相比，SDGCN模型在大型数据集上的深度架构表现更好。例如，在Citeseer数据集上，SDGCN模型在聚类精度上分别提升了3.2%、1.4%、1.9%、0.6%。对于大型网络，每个类别仅提供20个已知标签时，半监督模型仍不足以将标签信息传遍到全局网络。SDGCN模型引入自监督学习模块来学习已知节点的语义信息，并进一步使用伪标签来不断扩展已知节点，因此,可以进一步提高性能。此外，与自监督学习模型Self-SAGCN相比，SDGCN模型通过堆叠更多的卷积层(自监督学习模块)，可更好地利用伪标签信息。同时，从表3中可以观察到，Self-SAGCN模型在16层后性能急剧下降，而SDGCN模型性能依然能够保持稳定且有上升趋势。其主要原因是：在深度架构中引入了初始残差和恒等映射，大大降低了过度平滑的影响，同时也提高了模型性能。由于Citeseer数据集的边集数量较少，在所有模型上的精度都低于其他数据集。

同时，从图4中对Cora数据集上的t-SNE可视化结果可以观察到，随着层数的叠加，不同类别节点的特征区分性愈加明显。对比GCNII在64层的t-SNE可视化图可以发现，SDGCN模型获得的节点特征更具有判别性。这是因为：模型中的自监督学习模块为节点特征提供了判别语义信息，且在初始残差和恒等映射的辅助下，更有效缓解了过度平滑问题。

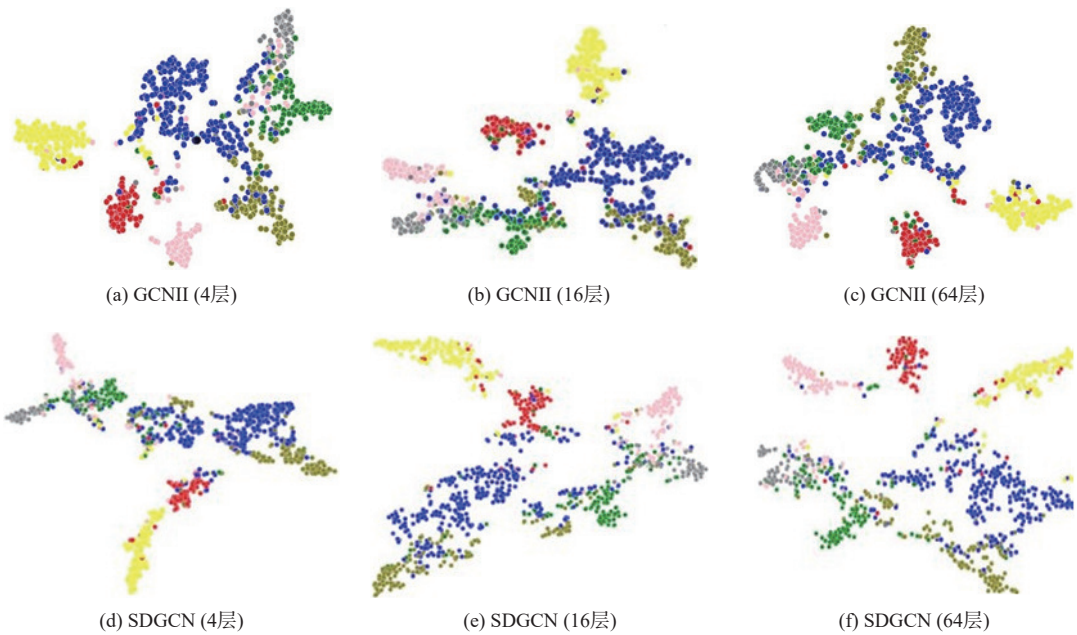
3.3　超参数分析

进一步研究了 SDGCN 模型的超参数敏感性。实验在 Cora 上测试不同损失权重下准确率的变化, 如图 5 (a) 所示。其表明 SDGCN 模型对参数 λ 在 [0.1,0.5] 范围内不敏感。在极少量已知标签情况下, 当 $\lambda = 0.3$时, 提供了自监督损失和图卷积损失之间的最佳权衡, 取得最优结果。在每类提供 5 个或 20 个已知标签情况下, $\lambda = 0.5$决定了自监督损失和图卷积损失之间达到最佳权衡。此外, 随着标记节点数量的增加, SDGCN 模型可以保持稳定的性

表 3　在不同模型深度条件下模型聚类精度对比

Table 3　Comparison of clustering accuracy of various models under different model depth

模型	层数为4			层数为16			层数为32			层数为64		
	Cora	Citeseer	PubMed	Cora	Citeseer	PubMed	Cora	Citeseer	PubMed	Cora	Citeseer	PubMed
GCN[12]	67.8	15.6	17.2	16.4	50.6	18.9	14.9	17.6	71.6	40.2	36.6	39.9
GCN(drop)	80.6	75.7	62.5	49.5	68.6	57.3	41.1	33.2	78.3	77.6	76.2	61.1
JKNet[18]	78.9	79.3	80.6	70.9	67.9	68.2	69.2	63.1	78.1	72.1	73.6	73.1
GCNII[20]	79.2	83.1	83.9	84.1	67.0	70.2	70.8	71.6	79.1	77.6	79.6	79.9
Self-SAGCN[25]	75.1	34.6	31.6	30.8	62.5	48.9	26.2	18.1	75.3	50.8	43.7	40.1
SDGCN	81.2	82.6	84.1	85.2	70.2	71.6	72.7	72.2	78.6	79.2	80.1	80.6

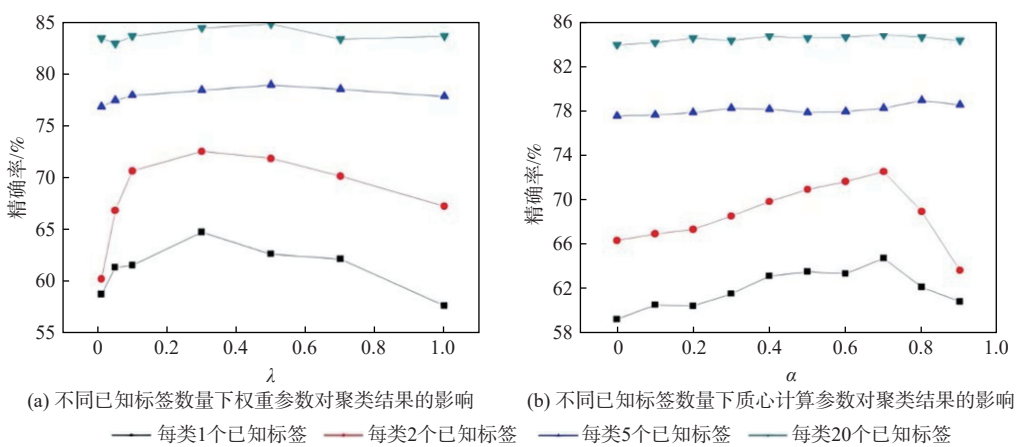


(a) GCNII (4层)　(b) GCNII (16层)　(c) GCNII (64层)

(d) SDGCN (4层)　(e) SDGCN (16层)　(f) SDGCN (64层)

图 4　不同模型深度下 2 种模型在 Cora 数据集上获得的节点特征 t-SNE 可视化

Fig. 4　t-SNE visualization of node features obtained by two models on Core dataset under different model depth



(a) 不同已知标签数量下权重参数对聚类结果的影响　(b) 不同已知标签数量下质心计算参数对聚类结果的影响

图 5　超参数对模型聚类准确率的影响

Fig. 5　The influence of hyper-parameters on model clustering accuracy

能。此外，图 5 (b) 表明，语义对齐中采用的中心相似度优化方法可以有效提高模型性能，主要是因为逐步更新机制抑制了伪标签的噪声信号。

4 结 论

为解决 GCN 过度依赖先验标签和深度模型所导致的过平滑问题，本文提出了一种基于深度 GCN 的自监督学习群体发现模型。

1）本文模型通过引入自监督学习模块，利用少量先验节点标签信息，通过语义对齐机制来推断未知节点的标签，有效降低了 GCN 对大量先验标签的依赖。

2）本文模型提出利用初始残差和恒等映射 2 种策略，有效缓解了深度 GCN 架构带来的过平滑问题，从而获得网络全局信息并提高模型性能。

3）实验结果表明，即使在给定少量先验标签条件下，本文模型在深度达到 64 层构架时，仍然能够保持接近 80% 的识别准确率。相较于其他模型，该模型在群体识别精度上有显著的提升。

4）为进一步提高模型的计算效率和性能，仍需要优化模型结构，如引入图注意力机制和多层次特征融合技术，以更精确地捕获全局信息。

参考文献（References）

[1] BEDRU H D, YU S, XIAO X R, et al. Big networks: a survey[J]. Computer Science Review, 2020, 37: 100247.

[2] DA MATA A S. Complex networks: a mini-review[J]. Brazilian Journal of Physics, 2020, 50(5): 658-672.

[3] 潘理, 吴鹏, 黄丹华. 在线社交网络群体发现研究进展[J]. 电子与信息学报, 2017, 39(9): 2097-2107.
PAN L, WU P, HUANG D H. Reviews on group detection in online social networks[J]. Journal of Electronics & Information Technology, 2017, 39(9): 2097-2107(in Chinese).

[4] CAI B, ZENG L N, WANG Y P, et al. Community detection method based on node density, degree centrality, and K-means clustering in complex network[J]. Entropy, 2019, 21(12): 1145.

[5] SU X, XUE S, LIU F Z, et al. A comprehensive survey on community detection with deep learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(4): 4682-4702.

[6] 齐金山, 梁循, 李志宇, 等. 大规模复杂信息网络表示学习: 概念、方法与挑战[J]. 计算机学报, 2018, 41(10): 2394-2420.
QI J S, LIANG X, LI Z Y, et al. Representation learning of large-scale complex information network: concepts, methods and challenges[J]. Chinese Journal of Computers, 2018, 41(10): 2394-2420(in Chinese).

[7] ZHOU Z, AMINI A A. Analysis of spectral clustering algorithms for community detection: the general bipartite setting[J]. The Journal of Machine Learning Research, 2019, 20(1): 1774-1820.

[8] ZHANG Y, YEUNG D Y. Overlapping community detection via bounded nonnegative matrix tri-factorization[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2012: 606-614.

[9] YANG B, ZHAO X H, LIU X Y. Bayesian approach to modeling and detecting communities in signed network[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Austin: AAAI, 2015: 1952-1958.

[10] WU Z H, PAN S R, CHEN F W, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.

[11] 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述[J]. 计算机学报, 2020, 43(5): 755-780.
XU B B, CEN K T, HUANG J J, et al. A survey on graph convolutional neural network[J]. Chinese Journal of Computers, 2020, 43(5): 755-780(in Chinese).

[12] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2017-02-22) [2023-05-25]. https://arxiv.org/abs/1609.02907.

[13] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2018-02-04) [2023-05-25]. https://arxiv.org/abs/1710.10903.

[14] WANG H W, WANG J, WANG J L, et al. GraphGAN: graph representation learning with generative adversarial nets[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans: AAAI, 2018: 2508-2515.

[15] JIN D, LIU Z Y, LI W H, et al. Graph convolutional networks meet Markov random fields: semi-supervised community detection in attribute networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu: AAAI, 2019: 152-159.

[16] HU R Q, PAN S R, LONG G D, et al. Going deep: graph convolutional ladder-shape networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. NewYork: AAAI, 2020: 2838-2845.

[17] JIN D, LI B Y, JIAO P F, et al. Community detection via joint graph convolutional network embedding in attribute network [C]//Proceedings of International Conference on Artificial Neural Networks. Berlin: Springer, 2019: 594-606.

[18] XU K, LI C, TIAN Y, et al. Representation learning on graphs with jumping knowledge networks[C]//Proceedings of the International conference on machine learning. Stockholm: PMLR, 2018: 5453-5462.

[19] RONG Y, HUANG W B, XU T Y, et al. DropEdge: towards deep graph convolutional networks on node classification[EB/OL]. (2010-03-12) [2023-05-26]. https://arxiv.org/abs/1907.10903.

[20] CHEN M, WEI Z, HUANG Z, et al. Simple and deep graph convolutional networks[C]//Proceedings of the International Conference on Machine Learning. Vienna: PMLR, 2020: 1725-1735.

[21] LIU X, ZHANG F J, HOU Z Y, et al. Self-supervised learning: generative or contrastive[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(1): 857-876.

[22] YOU Y N, CHEN T L, WANG Z Y, et al. When does self-supervision help graph convolutional networks?[EB/OL].(2020-06-18)[2023-05-26]. https://arxiv.org/abs/2006.09136.

[23] ZHU Y Q, XU Y C, YU F, et al. CAGNN: cluster-aware graph neural networks for unsupervised graph representation learning[EB/OL]. (2020-09-03) [2023-05-26]. https://arxiv.org/abs/2009.01674.

[24] SUN K, LIN Z C, ZHU Z X. Multi-stage self-supervised learning

for graph convolutional networks on graphs with few labeled nodes [C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020: 5892-5899.

[25] YANG X, DENG C, DANG Z Y, et al. Self-SAGCN: Self-supervised semantic alignment for graph convolution network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 16770-16779.

[26] JIA S W, GAO L, GAO Y, et al. Defining and identifying cograph communities in complex networks[J]. New Journal of Physics, 2015, 17(1): 013044.

[27] ZHANG P, MOORE C. Scalable detection of statistically significant communities and hierarchies, using message passing for modularity[J]. Proceedings of the National Academy of Sciences of the United States of America, 2014, 111(51): 18144-18149.

[28] FANUEL M, ALAÍZ C M, SUYKENS J A K. Magnetic eigenmaps for community detection in directed networks[J]. Physical Review E, 2017, 95(2-1): 022302.

[29] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 2008: 10008.

[30] HOLLAND P W, LASKEY K B, LEINHARDT S. Stochastic block models: first steps[J]. Social Networks, 1983, 5(2): 109-137.

[31] LIU J H, MA G X, JIANG F, et al. Community-preserving graph convolutions for structural and functional joint embedding of brain networks[C]//Proceedings of the IEEE International Conference on Big Data. Piscataway: IEEE Press, 2019: 1163-1168.

[32] ZHANG T Q, XIONG Y, ZHANG J W, et al. CommDGI: community detection oriented deep graph infomax[C]//Proceedings of the ACM International Conference on Information & Knowledge Management. New York: ACM, 2020: 1843-1852.

[33] ZHAO H, YANG X, WANG Z, et al. Graph debiased contrastive learning with joint representation clustering[C]//Proceedings of the International Joint Conference on Artificial Intelligence. Montreal: IJCAI, 2021: 3434-3440.

[34] WANG X F, LI J H, YANG L, et al. Unsupervised learning for community detection in attributed networks based on graph convolutional network[J]. Neurocomputing, 2021, 456: 147-155.

[35] QIAO F, HUANG C X, XU S, et al. Community detection model based on graph representation and self-supervised learning[C]//Proceedings of the Artificial Intelligence and Security. Berlin: Springer, 2021: 27-40.

[36] JIN W, DERR T, LIU H, et al. Self-supervised learning on graphs: deep insights and new direction[EB/OL]. (2020-06-17)[2023-05-20]. https://doi.org/10.48550/arXiv.2006.10141.

[37] 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.

[38] KLICPERA J, BOJCHEVSKI A, GÜNNEMANN S. Predict then propagate: graph neural networks meet personalized PageRank[EB/OL]. (2022-04-05) [2023-05-28]. https://arxiv.org/abs/1810.05997.

[39] LIU M, GAO H Y, JI S W. Towards deeper graph neural networks[C]//Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 338-348.

Self-supervised learning for community detection based on deep graph convolutional networks

WANG Zaisheng，WANG Xiaofeng*，SHEN Guodong，ZHANG Zengjie，QUAN Daying

(College of Information Engineering，China Jiliang University，Hangzhou 310018，China)

Abstract：To alleviate the excessive dependence of graph neural networks on prior knowledge in community discovery and improve recognition accuracy, a novel self-supervised learning model for community detection based on a deep graph convolutional network (GCN) is proposed. The model makes full use of the semantic features of a small number of nodes and obtains pseudo-labels of unknown nodes through a semantic alignment mechanism, and thus introduces a self-supervised module to alleviate the dependence on a large number of prior labels during the training of GCN. Furthermore, by stacking self-supervised modules, a deep graph self-supervised learning model is built to increase the accuracy of community detection by obtaining the global information of networks. Two strategies, identity mapping and initial residual, are employed to address the over-smoothing issues that the deep model introduces. According to experiments conducted on publicly available datasets, the suggested approach outperforms current models in terms of community recognition accuracy when a limited number of prior labels are used and the model depth is increased.

Keywords：complex network；community detection；graph convolutional network；residual connection；self-supervised learning

Received：2023-06-21；**Accepted**：2023-11-21；**Published Online**：2023-12-06 13：16

URL：link.cnki.net/urlid/11.2625.V.20231205.1639.003

Foundation items：Zhejiang Provincial Natural Science Foundation of China (LQ20F020021); National Key Research and Development Program of China (2019YFB1707104)

* **Corresponding author**. E-mail：xfwang@cjlu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0376

基于滑模控制律的机翼极限环主动抑制

王诗其[1,2]，张征[1]，宋晨[2,*]，杨超[2]，郑焱午[1]

(1. 中国商用飞机有限责任公司北京民用飞机技术研究中心，北京 102211；
2. 北京航空航天大学 航空科学与工程学院，北京 100191)

摘 要：间隙非线性机翼系统在一定速度范围内会发生极限环振荡，传统线性控制律对极限环抑制效果通常不显著且鲁棒性不足，因此，需设计相应的非线性控制律来提高系统的控制效果及鲁棒性。通过对间隙非线性二元翼段气动弹性系统进行线性二次调节 (LQR) 控制律设计，结果显示：极限环速度范围由 11.4～16.9 m/s 优化至 12.9～19.2 m/s。同时，对间隙非线性二元翼段气动弹性系统进行滑模控制律 (SMC) 设计，结果显示：极限环速度范围由 11.4～16.9 m/s 优化至 29.4～39.3 m/s。算例结果表明：针对间隙非线性系统，所设计 SMC 效果优于 LQR 控制律。

关 键 词：间隙非线性；极限环振荡；控制律；线性二次调节；滑模控制律
中图分类号：V215.3
文献标志码：A **文章编号**：1001-5965(2025)06-2033-08

受安装误差、加工工艺精度、使用过程磨损等因素影响，飞机机翼结构往往存在间隙非线性现象，结构非线性会使得飞机在一定速度范围内发生极限环振荡、混沌等非线性气动弹性现象[1-2]。与线性系统颤振相同的是，极限环振荡也是一种自激振荡现象，不仅会影响飞机飞行品质，而且长时间疲劳会导致飞机结构破坏，从而影响飞行安全[3-6]。

对于线性机翼系统，通过改变机翼结构、质量分布的方式，可在一定范围内提升系统的颤振速度，但该方法会伴随机翼结构质量的增加，影响飞机整体性能。颤振主动抑制(active flutter suppression, AFS)是当前解决颤振问题的热门研究方法，该方法不会对机翼结构质量产生明显变化。AFS 首次在美国的“主动气动弹性机翼(active aeroelastic wing, AAW)计划”中得到验证，该计划以 F/A-18A 战机为受控对象，通过控制飞机舵面运动，从而改变飞机气动力分布，最终达到提高临界颤振速度、避免结构增重、机动载荷减缓的效果[7-8]。杨超等采用 PID 控制、线性二次调节(linear quadratic regulator, LQR)控制等控制方法对线性机翼系统进行了 AFS 控制律设计，结果表明，以上控制律均能在一定程度上提高系统颤振速度[9-10]。通常情况下，AFS 采用的是线性控制律，对于间隙非线性系统，采用线性控制律效果不显著，且鲁棒性较差。李道春和向锦武针对间隙非线性二元翼段设计了一套 LQG 控制律，结果表明，闭环状态下系统极限环得到了抑制，但抑制效果次于线性系统[11-12]。宋晨等采用滑模变结构控制方法设计了线性二元翼段颤振控制律，结果显示，该方法能提升机翼的颤振速度，同时保证了系统的鲁棒性及抗干扰能力[13-15]。

本文以带控制面二元翼段为研究对象，在俯仰刚度添加间隙非线性因素，通过施加滑模控制律 (sliding mode control law, SMC) 达到极限环抑制效果。文中采用的气动力模型为准定常气动力模型，结构模型为具有俯仰、沉浮 2 自由度二元翼段，对比 SMC、LQR 控制律对间隙非线性二元翼段极限环抑制效果，结果显示，本文设计 SMC 对间隙非线性系统极限环抑制有效，且抑制效果优于 LQR 控制律。

收稿日期：2023-06-15；**录用日期**：2023-10-30；**网络出版时间**：2023-11-17 16:09
网络出版地址：link.cnki.net/urlid/11.2625.V.20231116.1736.004
基金项目：国家自然科学基金 (11402013)
*通信作者. E-mail：songchen@buaa.edu.cn

引用格式：王诗其，张征，宋晨，等. 基于滑模控制律的机翼极限环主动抑制 [J]. 北京航空航天大学学报，2025，51（6）：2033-2040.
WANG S Q，ZHANG Z，SONG C，et al. Limit cycle oscillation suppression of an airfoil based on sliding mode control law [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（6）：2033-2040 （in Chinese）.

1 模型描述

本文研究的二元翼段模型如图1所示[13]。采用准定常气动力理论，忽略控制面动力学响应，对应的气动弹性方程可描述为

$$\boldsymbol{M}\ddot{\boldsymbol{q}}+\boldsymbol{C}\dot{\boldsymbol{q}}+\boldsymbol{K}\boldsymbol{q}=-\boldsymbol{Q}_{\mathrm{G}}\boldsymbol{q}-\boldsymbol{Q}_{\mathrm{Z}}\dot{\boldsymbol{q}}+\rho V^2 b[\ c_{l\beta}\quad c_{m\beta}\ ]^{\mathrm{T}}\beta \tag{1}$$

式(1)等式左边为二元翼段受到的惯性力、阻尼力和弹性力，等式右边为气动力，式中各矩阵对应关系如下[16]：

$$\begin{cases}\boldsymbol{M}=\begin{bmatrix} m & mx_\alpha b \\ mx_\alpha b & I_\alpha \end{bmatrix}\\ \boldsymbol{C}=\begin{bmatrix} c_h & 0 \\ 0 & c_\alpha \end{bmatrix}\\ \boldsymbol{K}=\begin{bmatrix} k_h & 0 \\ 0 & k_\alpha(\alpha) \end{bmatrix}\\ \boldsymbol{q}=\begin{bmatrix} h & \alpha \end{bmatrix}^{\mathrm{T}}\\ \boldsymbol{Q}_{\mathrm{G}}=2\pi\rho Vb\begin{bmatrix} 0 & V \\ 0 & -\left(\frac{1}{2}+a\right)bV \end{bmatrix}\\ \boldsymbol{Q}_{\mathrm{Z}}=2\pi\rho Vb\begin{bmatrix} 1 & \left(\frac{1}{2}-a\right)b \\ -\left(\frac{1}{2}+a\right)b & a^2b^2 \end{bmatrix}\end{cases} \tag{2}$$

式中：h为沉浮方向位移；a为弹性轴到中点距离；m为质量；x_α为弹性轴到重心距离；c_α为俯仰阻尼系数；$c_{m\beta}$为控制面力矩系数；β为控制面偏转角；k_h为沉浮刚度系数；α为俯仰角；b为半弦长；I_α为惯性矩；c_h为沉浮阻尼系数；$c_{l\beta}$为控制面升力系数；ρ为来流大气密度；V为来流大气速度；$k_\alpha(\alpha)$为俯仰刚度系数。

俯仰力矩$M(\alpha)$随俯仰角α的变化关系如图2所示[11]。

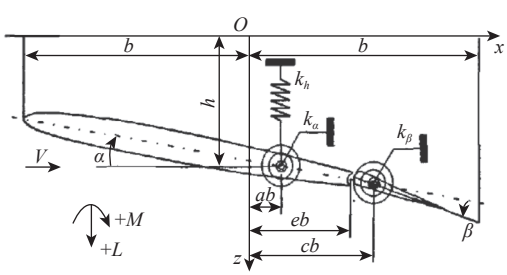


图1 二元翼段[13]

Fig. 1 Two-dimensional airfoil[13]

俯仰力矩$M(\alpha)$随俯仰角α变化关系为(k为斜率)

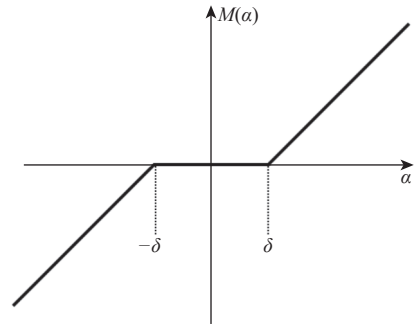


图2 位移-力关系曲线[11]

Fig. 2 Displacement-force curve[11]

$$M(\alpha)=\begin{cases} k\delta\left(\dfrac{\alpha}{\delta}-1\right) & \dfrac{\alpha}{\delta}\geqslant 1\\ 0 & -1<\dfrac{\alpha}{\delta}<1\\ k\delta\left(\dfrac{\alpha}{\delta}+1\right) & \dfrac{\alpha}{\delta}\leqslant -1 \end{cases} \tag{3}$$

式中：δ为二元机翼俯仰间隙角。

引入状态变量$\boldsymbol{x}=\begin{bmatrix} h & \alpha & \dot{h} & \dot{\alpha}\end{bmatrix}^{\mathrm{T}}$，则式(1)所示动力学方程可表示为状态空间方程的形式：

$$\dot{\boldsymbol{x}}=\boldsymbol{A}(\boldsymbol{x})\boldsymbol{x}+\boldsymbol{B}\beta \tag{4}$$

$$\begin{cases}\boldsymbol{A}(\boldsymbol{x})=\begin{bmatrix} \boldsymbol{0}_{2\times 2} & \boldsymbol{I}_{2\times 2} \\ -\boldsymbol{M}^{-1}(\boldsymbol{Q}_{\mathrm{G}}+\boldsymbol{K}) & -\boldsymbol{M}^{-1}(\boldsymbol{Q}_{\mathrm{Z}}+\boldsymbol{C}) \end{bmatrix}\\ \boldsymbol{B}=\rho V^2 b\begin{bmatrix} \boldsymbol{M}^{-1} & \boldsymbol{0}_{2\times 2} \\ \boldsymbol{0}_{2\times 2} & \boldsymbol{M}^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{0}_{2\times 1} \\ -c_{l\beta} \\ \left(\frac{1}{2}+a\right)bc_{l\beta}+2bc_{m\beta} \end{bmatrix}\end{cases} \tag{5}$$

2 控制律设计

2.1 LQR控制律

对于式(4)所示系统，考虑二次型指标最优控制问题，使得如下性能指标取最小值：

$$J=\int_0^{+\infty}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}+R\beta^2)\mathrm{d}t \tag{6}$$

式中：$\boldsymbol{Q}$为4×4正定对称矩阵；R>0。通过调节$\boldsymbol{Q}$和R的值可获得理想闭环控制效果。

对于非线性系统，直接求解最优控制律较为困难，但通过状态依赖Riccati方程的求解可获得控制律次优解。为确保状态依赖Riccati方程有解，需满足式(4)所示系统可控。对于本文所研究俯仰刚度间隙非线性系统，可将其分解成3段线性系统，要确保系统可控，则需保证系统在3个分段线性区域均可控，即

$$\det\begin{bmatrix}\boldsymbol{B} & \boldsymbol{A}(\boldsymbol{x})\boldsymbol{B} & \boldsymbol{A}^2(\boldsymbol{x})\boldsymbol{B} & \boldsymbol{A}^3(\boldsymbol{x})\boldsymbol{B}\end{bmatrix}\neq 0 \tag{7}$$

在确保式(4)所示系统可控的前提下求解以下状态依赖Riccati方程：

$$\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{P}+\boldsymbol{P}\boldsymbol{A}(\boldsymbol{x})-\boldsymbol{P}\boldsymbol{B}R^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{P}+\boldsymbol{Q}=\boldsymbol{0} \tag{8}$$

在求得式 (8) 中正定矩阵 $\boldsymbol{P}$ 后, 可得式 (4) 系统对应非线性控制律如下:

$$\beta(t) = -R^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x} \tag{9}$$

对应闭环系统控制框图如图 3 所示。

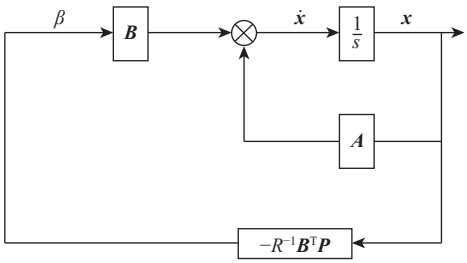


图 3　闭环系统框图(LQR)

Fig. 3　Block diagram of closed loop system (LQR)

2.2　SMC 控制律

2.2.1　SMC 机理

对如下任意动力学系统:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},t) \qquad \boldsymbol{x}\in\mathbf{R}^n,\boldsymbol{u}\in\mathbf{R}^m,t\in\mathbf{R} \tag{10}$$

式中: $\boldsymbol{x}$ 为状态变量; $\boldsymbol{u}$ 为控制律。在 n 维相空间里都能找到一组滑模面 $\boldsymbol{s}$, 通过调节滑模面 $\boldsymbol{s}$ 参数可完成控制律 $\boldsymbol{u}$ 的求解。

$$\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{x}) \qquad \boldsymbol{s}\in\mathbf{R}^m \tag{11}$$

$$u_i(\boldsymbol{x}) = \begin{cases} u_i^{+}(\boldsymbol{x}) & s_i(\boldsymbol{x}) > 0 \\ u_i^{-}(\boldsymbol{x}) & s_i(\boldsymbol{x}) < 0 \end{cases},\ i = 1,2,\cdots,m \tag{12}$$

在滑模面参数设计时, 需同时考虑滑模系统的存在性、可达性和稳定性, 即满足以下不等式:

$$\lim_{s_i\to 0} s_i\dot{s}_i < 0 \tag{13}$$

滑模面通常为系统状态变量的线性组合, 即

$$\boldsymbol{s}(\boldsymbol{x}) = \boldsymbol{c}\boldsymbol{x} \qquad \boldsymbol{c}\in\mathbf{R}^{m\times n} \tag{14}$$

为确保滑模系统的存在性、可达性和稳定性, 有

$$\boldsymbol{s}\boldsymbol{c}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},t) < 0 \tag{15}$$

针对式 (4) 对应二元翼段系统, 为满足式 (15) 不等式成立, 引入符号函数 sign(), 利用符号函数设计等速趋近率如下:

$$\dot{\boldsymbol{s}} = -\boldsymbol{\varepsilon}[\mathrm{sign}(\boldsymbol{s}_1),\ \mathrm{sign}(\boldsymbol{s}_2),\cdots,\mathrm{sign}(\boldsymbol{s}_m)]^{\mathrm{T}} \qquad \boldsymbol{\varepsilon} > 0 \tag{16}$$

式中: 符号函数 sign(·) 可表示为

$$\mathrm{sign}(x) = \begin{cases} 1 & x \geqslant 0 \\ -1 & x < 0 \end{cases} \tag{17}$$

将式 (4)、式 (14) 代入式 (16), 可得

$$\boldsymbol{c}(\boldsymbol{A}(\boldsymbol{x})\boldsymbol{x} + \boldsymbol{B}\beta) + \boldsymbol{\varepsilon}\mathrm{sign}(\boldsymbol{s}) = 0 \tag{18}$$

式中: β 为式 (10) 中对应控制律 $\boldsymbol{u}$。

将式 (18) 化简, 可得

$$\beta(t) = -(\boldsymbol{c}\boldsymbol{B})^{-1}[\boldsymbol{c}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}\mathrm{sign}(\boldsymbol{s})] \tag{19}$$

对应闭环系统控制框图如图 4 所示。

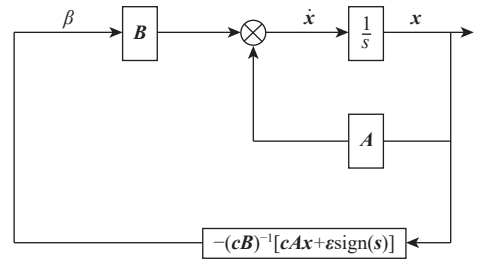


图 4　闭环系统框图(SMC)

Fig. 4　Block diagram of closed loop system (SMC)

2.2.2　SMC 设计

对于式 (4) 所示系统, 假设其滑模控制面参数为 $\boldsymbol{c}_s$=[c_{s1}　c_{s2}　c_{s3}　c_{s4}], 则滑模面函数可写为

$$\boldsymbol{s}(t) = \begin{bmatrix} c_{s1} & c_{s2} & c_{s3} & c_{s4} \end{bmatrix}\boldsymbol{x}(t) \tag{20}$$

根据矩阵正交化理论易得, 存在正交矩阵 $\boldsymbol{T}_{\mathrm{r}}$ 使得

$$\boldsymbol{T}_{\mathrm{r}}\boldsymbol{B} = \begin{bmatrix} 0 & 0 & 0 & d \end{bmatrix}^{\mathrm{T}} \qquad d\in\mathbf{R} \tag{21}$$

令 $\boldsymbol{z}(t)$=$\boldsymbol{T}_{\mathrm{r}}\boldsymbol{x}(t)$, 其中,

$$\boldsymbol{z}(t) = \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} \qquad z_1(t)\in\mathbf{R}^{3\times 1}, z_2(t)\in\mathbf{R}^{1\times 1} \tag{22}$$

则有

$$\dot{\boldsymbol{z}}(t) = \boldsymbol{T}_{\mathrm{r}}\boldsymbol{A}\boldsymbol{T}_{\mathrm{r}}^{\mathrm{T}}\boldsymbol{z}(t) + \boldsymbol{T}_{\mathrm{r}}\boldsymbol{B}\beta(t) \tag{23}$$

令 $\bar{\boldsymbol{A}} = \boldsymbol{T}_{\mathrm{r}}\bar{\boldsymbol{A}}\boldsymbol{T}_{\mathrm{r}}^{\mathrm{T}} = \begin{bmatrix} \bar{\boldsymbol{A}}_{11} & \bar{\boldsymbol{A}}_{12} \\ \bar{\boldsymbol{A}}_{21} & \bar{\boldsymbol{A}}_{22} \end{bmatrix}$, 其中, $\bar{\boldsymbol{A}}_{11}\in\mathbf{R}^{3\times 3}$, $\bar{\boldsymbol{A}}_{12}\in\mathbf{R}^{3\times 1}, \bar{\boldsymbol{A}}_{21}\in\mathbf{R}^{1\times 3}, \bar{\boldsymbol{A}}_{22}\in\mathbf{R}^{1\times 1}$。

因此, 式 (23) 可写成如下分解形式:

$$\begin{bmatrix} \dot{z}_1(t) \\ \dot{z}_2(t) \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{A}}_{11} & \bar{\boldsymbol{A}}_{12} \\ \bar{\boldsymbol{A}}_{21} & \bar{\boldsymbol{A}}_{22} \end{bmatrix}\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ b \end{bmatrix}\beta \tag{24}$$

又 $\boldsymbol{s}(t)$=$\boldsymbol{c}_s\boldsymbol{x}(t)$=$\boldsymbol{c}_s\boldsymbol{T}_{\mathrm{r}}^{\mathrm{T}}\boldsymbol{z}(t)$=$\boldsymbol{c}_t\boldsymbol{z}(t)$, 其中:

$$\boldsymbol{c}_t = \boldsymbol{c}_s\boldsymbol{T}_{\mathrm{r}}^{\mathrm{T}} = \begin{bmatrix} c_{t1} & c_{t2} \end{bmatrix} = c_{t2}\begin{bmatrix} \boldsymbol{\Omega} & 1 \end{bmatrix}$$
$$c_{t1}\in\mathbf{R}^{1\times 3}, c_{t2}\in\mathbf{R}^{1\times 1} \tag{25}$$

对于式 (4) 所示系统, 考虑滑模控制问题, 使得如下性能指标取最小值:

$$J = \int_0^{+\infty}(\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{Q}_1\boldsymbol{x}(t))\mathrm{d}t \tag{26}$$

式中: $\boldsymbol{Q}_1$ 为正定对称矩阵。

用 $\boldsymbol{z}(t)$ 代替式 (26) 中的 $\boldsymbol{x}(t)$, 得

$$J = \int_0^{+\infty}(\boldsymbol{z}^{\mathrm{T}}(t)\boldsymbol{T}_{\mathrm{r}}\boldsymbol{Q}_1\boldsymbol{T}_{\mathrm{r}}^{\mathrm{T}}\boldsymbol{z}(t))\mathrm{d}t = \int_0^{+\infty}(\boldsymbol{z}^{\mathrm{T}}(t)\bar{\boldsymbol{Q}}\boldsymbol{z}(t))\mathrm{d}t \tag{27}$$

式中:

$$\bar{\boldsymbol{Q}} = \boldsymbol{T}_{\mathrm{r}}\boldsymbol{Q}\boldsymbol{T}_{\mathrm{r}}^{\mathrm{T}} = \begin{bmatrix} \bar{\boldsymbol{Q}}_{11} & \bar{\boldsymbol{Q}}_{12} \\ \bar{\boldsymbol{Q}}_{21} & \bar{\boldsymbol{Q}}_{22} \end{bmatrix} \tag{28}$$

其中: $\bar{\boldsymbol{Q}}_{11}\in\mathbf{R}^{3\times 3}$; $\bar{\boldsymbol{Q}}_{12}\in\mathbf{R}^{3\times 1}$; $\bar{\boldsymbol{Q}}_{21}\in\mathbf{R}^{1\times 3}$; $\bar{\boldsymbol{Q}}_{22}\in\mathbf{R}^{1\times 1}$。

现引入中间变量 $\boldsymbol{v}(t)$:

$$\boldsymbol{v}(t) = z_2(t) + \bar{\boldsymbol{Q}}_{22}^{-1}\bar{\boldsymbol{Q}}_{12}^{\mathrm{T}}z_1(t) \tag{29}$$

则式 (27) 所示最小性能指标可变换成如下形式:

$$J=\int_0^{+\infty}(z_1^{\mathrm{T}}\hat{\boldsymbol{Q}}z_1+\bar{\boldsymbol{Q}}_{22}\boldsymbol{v}^2)\mathrm{d}t \tag{30}$$

$$\hat{\boldsymbol{Q}}=\bar{\boldsymbol{Q}}_{11}-\bar{\boldsymbol{Q}}_{12}\bar{\boldsymbol{Q}}_{22}^{-1}\bar{\boldsymbol{Q}}_{12}^{\mathrm{T}} \tag{31}$$

因此, 可将式 (4) 所示状态空间方程转化成如下形式:

$$\dot{z}_1(t)=\hat{\boldsymbol{A}}z_1(t)+\bar{\boldsymbol{A}}_{12}\boldsymbol{v}(t) \tag{32}$$

$$\hat{\boldsymbol{A}}=\bar{\boldsymbol{A}}_{11}-\bar{\boldsymbol{A}}_{12}\bar{\boldsymbol{Q}}_{22}^{-1}\bar{\boldsymbol{Q}}_{12}^{\mathrm{T}} \tag{33}$$

此时, 式 (32) 与式 (30) 即经典的 LQR 控制问题, 且$\boldsymbol{v}(t)$为式 (32) 方程对应次优控制律, 即

$$\boldsymbol{v}(t)=-(\bar{\boldsymbol{Q}}_{22}^{-1}\bar{\boldsymbol{A}}_{12}^{\mathrm{T}}\boldsymbol{P}_1)z_1(t) \tag{34}$$

式中: $\boldsymbol{P}_1$ 需满足如下 Riccati 方程:

$$\hat{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{P}_1+\boldsymbol{P}_1\hat{\boldsymbol{A}}-\boldsymbol{P}_1\bar{\boldsymbol{A}}_{12}\bar{\boldsymbol{Q}}_{22}^{-1}\bar{\boldsymbol{A}}_{12}^{\mathrm{T}}\boldsymbol{P}_1+\hat{\boldsymbol{Q}}=\boldsymbol{0} \tag{35}$$

又因为当系统状态到达滑模面时有 $\boldsymbol{s}(t)$=0, 则在滑模面上需满足:

$$c_{t2}\begin{bmatrix}\boldsymbol{\Omega} & 1\end{bmatrix}\begin{bmatrix}z_1(t)\\ z_2(t)\end{bmatrix}=0 \tag{36}$$

由式 (36), 可得

$$z_2(t)=-\boldsymbol{\Omega}z_1(t) \tag{37}$$

联立式 (29)、式 (34), 可得

$$z_2(t)+\bar{\boldsymbol{Q}}^{-1}\bar{\boldsymbol{Q}}_{12}^{\mathrm{T}}z_1(t)=-(\bar{\boldsymbol{Q}}_{22}^{-1}\bar{\boldsymbol{A}}_{12}^{\mathrm{T}}\boldsymbol{P}_1)z_1(t) \tag{38}$$

将式 (38) 化简后, 可得

$$z_2(t)=-\bar{\boldsymbol{Q}}^{-1}(\bar{\boldsymbol{Q}}_{12}^{\mathrm{T}}+\bar{\boldsymbol{A}}_{12}^{\mathrm{T}}\boldsymbol{P}_1)z_1(t) \tag{39}$$

联立式 (37)、式 (39), 可得

$$\boldsymbol{\Omega}=\bar{\boldsymbol{Q}}^{-1}(\bar{\boldsymbol{Q}}_{12}^{\mathrm{T}}+\bar{\boldsymbol{A}}_{12}^{\mathrm{T}}\boldsymbol{P}_1) \tag{40}$$

因此, 通过以上推导计算可求得 $\boldsymbol{\Omega}$, 从而可求得 $\boldsymbol{c}_t$(c_{t2} 可任选), $\boldsymbol{c}_s=\boldsymbol{c}_t\boldsymbol{T}_r$,最终求得 SMC。

3 结果分析

本文研究的二元翼段各参数值如表 1 所示[14]。

3.1 系统开环响应

在不考虑俯仰方向间隙的情况下, 本文研究对象为线性气动弹性系统。通过对式 (1) 所示动力学方程进行频域分析, 可得二元翼段颤振速度为 16.9 m/s, 且在 16.9 m/s 速度下, 式 (4) 对应系统开环响应曲线如图 5 所示。

在俯仰刚度加入间隙非线性因素, 间隙角 δ=1°, 通过时域分析, 可得此时系统在 11.4～16.9 m/s 速度范围内发生极限环振荡, 同时对 16 m/s 速度下系统进行时域仿真, 可得系统开环响应如图 6 所示。

3.2 系统闭环响应

3.2.1 LQR 控制

对线性系统施加 LQR 控制律, 此时闭环系统颤振速度为 24.9 m/s, 较开环颤振速度提高 47%, 且 16.9 m/s、24. 9m/s 速度下系统闭环响应如图 7、图 8 所示。

对间隙非线性系统施加 LQR 控制律, 此时闭环系统极限环速度范围为 12.9～19.2 m/s, 极限环速度下限相比开环系统提高 13%, 极限环速度上限相比开环系统提高 14%, 且 16 m/s、24.9 m/s 速度下系统闭环响应如图 9、图 10 所示, 对比图 8、图 10 可知, LQR 控制律对线性系统控制效果优于间隙非线性系统。

3.2.2 SMC 控制

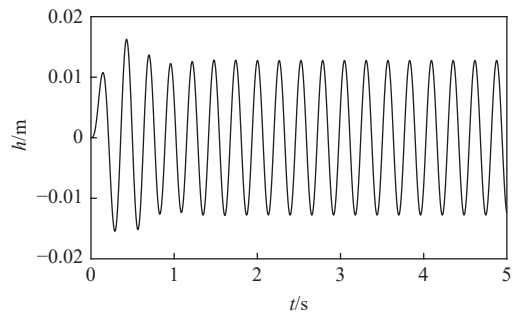
针对线性二元翼段气动弹性系统, 其开环颤振速度为 19.5 m/s, 宋晨等对以上系统设计了对应的 SMC, 在 SMC 作用下, 闭环系统在来流速度为 22.5 m/s (比开环颤振速度高 15.3%)时依然收敛[13]。

对本文线性系统施加 SMC, 此时闭环系统颤振速度为 39.3 m/s, 较开环颤振速度提高 1.33 倍, 不仅扩展了闭环滑模控制系统收敛速度包线, 还预测了闭

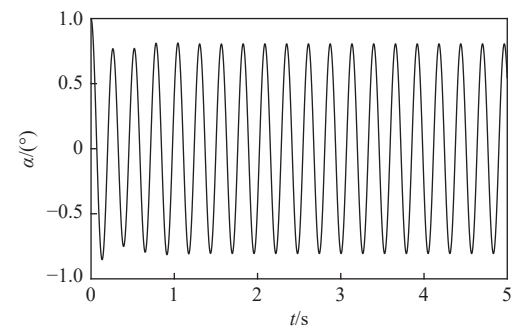
表 1 二元翼段参数[14]

Table 1 Parameters of two-dimensional airfoil[14]

参数	数值	参数	数值
a	−0.68	b/m	0.135
m/kg	12.387	I_α	0.065
x_α	0.326 7	c_h/(N·s·m^{-1})	27.43
c_α/(N·s)	0.036	$c_{l\beta}$	3.358
$c_{m\beta}$	−0.635	ρ/(kg·m^{-3})	1.225
k_h/(N·m^{-1})	2 884.4	k	28



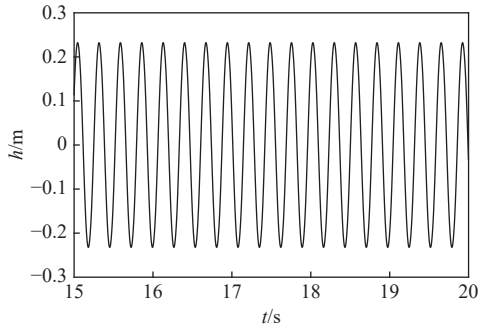
(a) 沉浮位移曲线



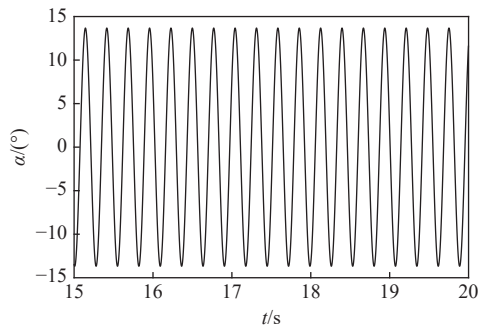
(b) 俯仰角曲线

图 5 线性系统开环响应(速度为 16.9 m/s)

Fig. 5 Open-loop response of linear system (the speed is 16.9 m/s)



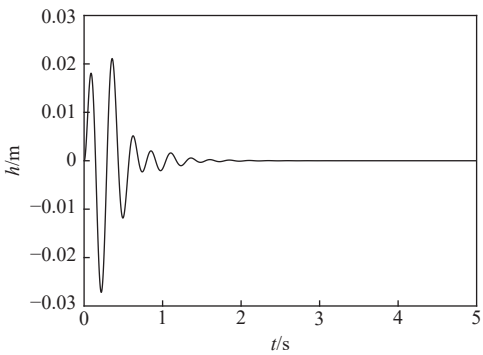
(a) 沉浮位移曲线



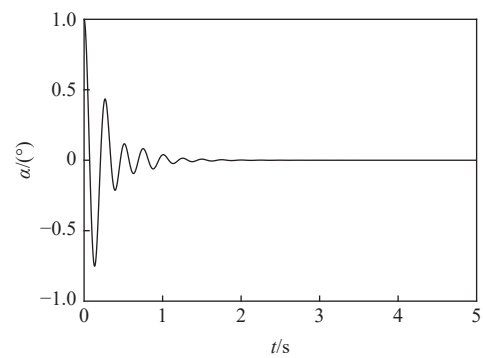
(b) 俯仰角曲线

图 6　非线性系统开环响应(速度为 16 m/s)

Fig. 6　Open-loop response of nonlinear system (the speed is 16 m/s)



(a) 沉浮位移曲线



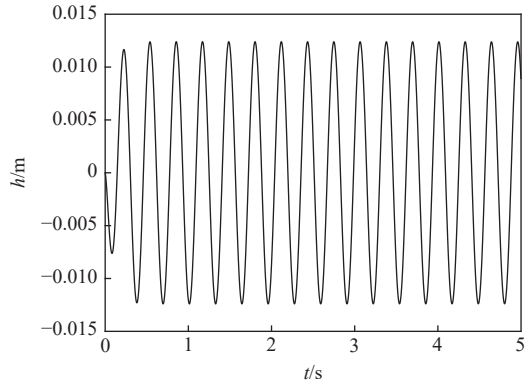
(b) 俯仰角曲线

图 7　线性系统闭环响应(LQR, 速度为 16.9 m/s)

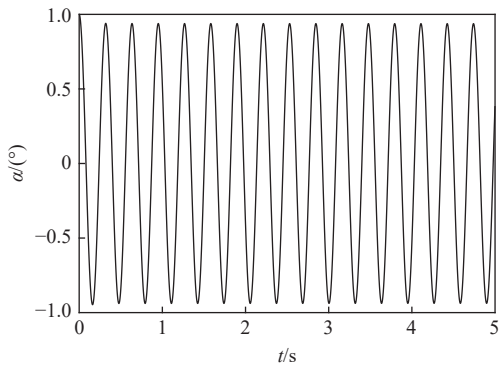
Fig. 7　Closed-loop response of linear system (LQR, the speed is 16.9 m/s)

环滑模控制系统颤振速度边界(39.3 m/s), 且 24.9 m/s、39.3 m/s 速度下系统闭环响应如图 11、图 12 所示。

对间隙非线性系统施加 SMC, 此时闭环系统极



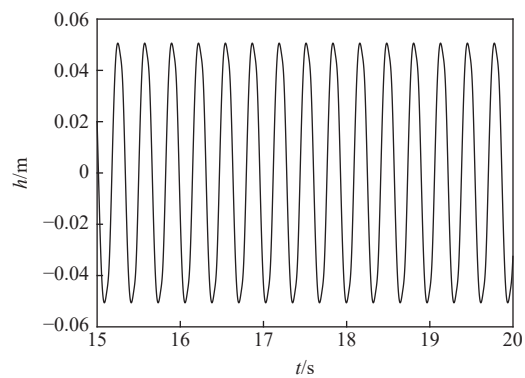
(a) 沉浮位移曲线



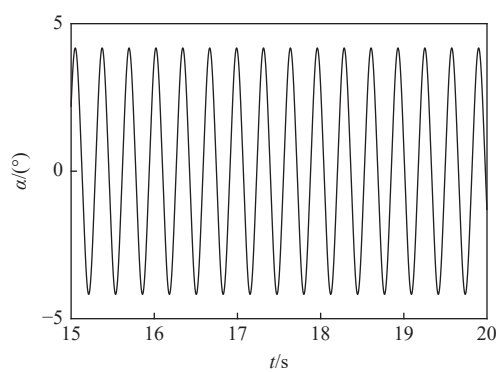
(b) 俯仰角曲线

图 8　线性系统闭环响应(LQR, 速度为 24.9 m/s)

Fig. 8　Closed-loop response of linear system (LQR, the speed is 24.9 m/s)



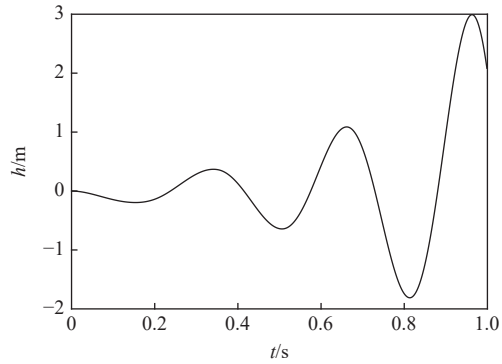
(a) 沉浮位移曲线



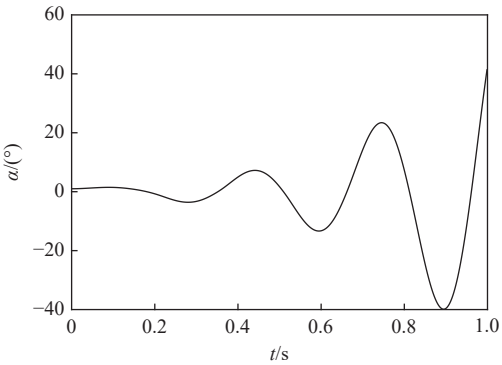
(b) 俯仰角曲线

图 9　非线性系统闭环响应(LQR, 速度为 16 m/s)

Fig. 9　Closed-loop response of nonlinear system (LQR, the speed is 16 m/s)



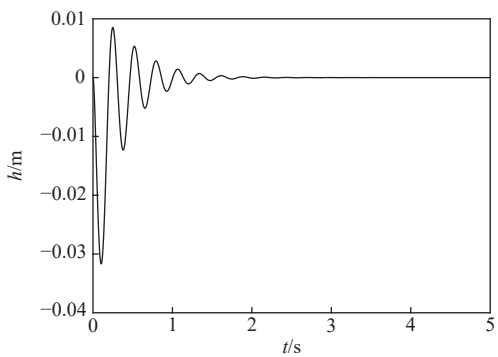
(a) 沉浮位移曲线



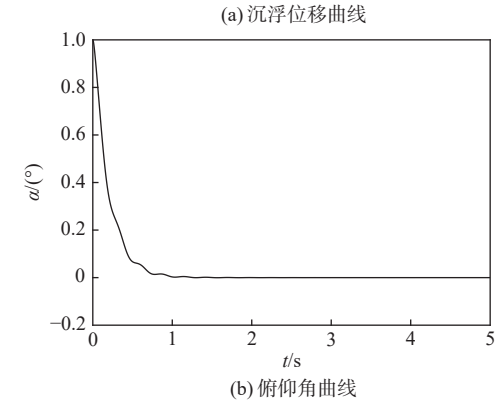
(b) 俯仰角曲线

图 10 非线性系统闭环响应(LQR, 速度为 24.9 m/s)

Fig. 10 Closed-loop response of nonlinear system (LQR, the speed is 24.9 m/s)



(a) 沉浮位移曲线

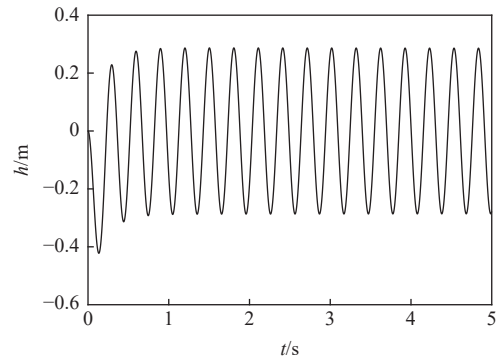


(b) 俯仰角曲线

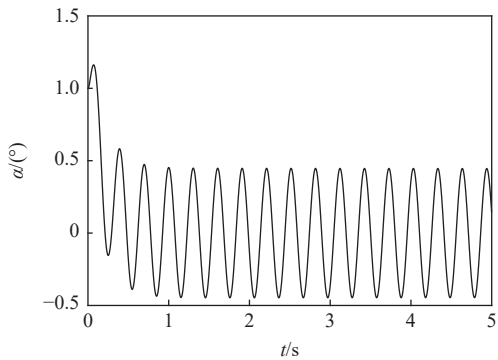
图 11 线性系统闭环响应(SMC, 速度为 24.9 m/s)

Fig. 11 Closed-loop response of linear system (SMC, the speed is 24.9 m/s)

限环速度范围为 29.4～39.3 m/s, 极限环速度下限相比开环系统提高 1.58 倍, 极限环速度上限相比开环系统提高 1.33 倍, 且 24.9 m/s、29.4 m/s速度下系统闭环响应如图 13、图 14 所示。



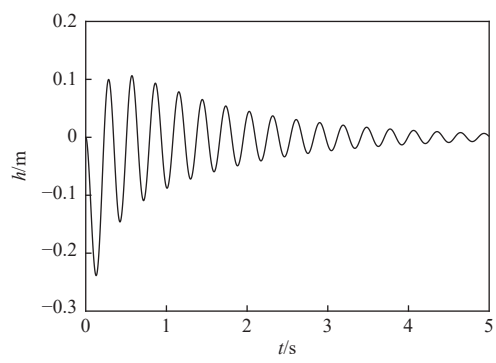
(a) 沉浮位移曲线



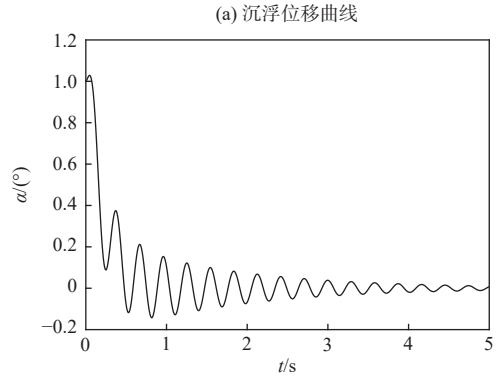
(b) 俯仰角曲线

图 12 线性系统闭环响应(SMC, 速度为 39.3 m/s)

Fig. 12 Closed-loop response of linear system (SMC, the speed is 39.3 m/s)



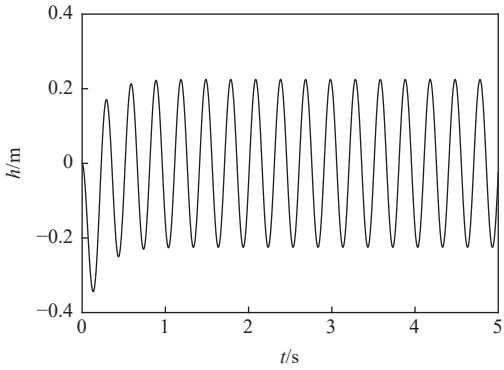
(a) 沉浮位移曲线



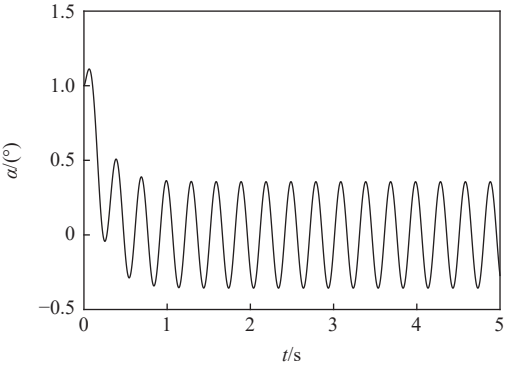
(b) 俯仰角曲线

图 13 非线性系统闭环响应(SMC, 速度为 24.9 m/s)

Fig. 13 Closed-loop response of nonlinear system (SMC, the speed is 24.9 m/s)



(a) 沉浮位移曲线



(b) 俯仰角曲线

图 14　非线性系统闭环响应(SMC, 速度为 29.4 m/s)

Fig. 14　Closed-loop response of nonlinear system (SMC, the speed is 29.4 m/s)

4　结　论

本文分别采用 LQR 控制律、SMC 作用于线性气动弹性系统和间隙非线性气动弹性系统, 根据 2 种控制律作用效果可得如下结论:

1) LQR 控制律、SMC 对线性气动弹性系统颤振抑制、间隙非线性气动弹性系统极限环抑制均有效。

2) LQR 控制律将线性气动弹性系统颤振速度由 16.9 m/s 提升至 24.9 m/s, 将间隙非线性气动弹性系统极限环速度范围由 11.4～16.9 m/s 提升至 12.9～19.2 m/s, 因此 LQR 控制律对线性系统的控制效果优于间隙非线性系统。

3) SMC 将间隙非线性系统极限环速度范围由 11.4～16.9 m/s 提升至 29.4～39.3 m/s, 因此, SMC 对间隙非线性气动弹性系统的控制效果优于 LQR 控制律。

参考文献 (References)

[1] 赵永辉, 胡海岩. 具有操纵面间隙非线性二维翼段的气动弹性分析[J]. 航空学报, 2003, 24(6): 521-525.
ZHAO Y H, HU H Y. Aeroelastic analysis of a two-dimensional airfoil with control surface freeplay nonlinearity[J]. Acta Aeronautica et Astronautica Sinica, 2003, 24(6): 521-525(in Chinese).

[2] DOWELL E H, COX D, CURTISS H C, et al. A modern course in aeroelasticity[M]. Dordrecht: Kluwer Academic Pub lisher, 2004.

[3] 胡海岩, 赵永辉, 黄锐. 飞机结构气动弹性分析与控制研究[J]. 力学学报, 2016, 48(1): 1-27.
HU H Y, ZHAO Y H, HUANG R. Studies on aeroelastic analysis and control of aircraft structures[J]. Chinese Journal of Theoretical and Applied Mechanics, 2016, 48(1): 1-27(in Chinese).

[4] 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).

[5] 李家旭, 田玮, 谷迎松. 考虑间隙非线性的控制舵非线性气动弹性分析[J]. 航空工程进展, 2020, 11(6): 827-835.
LI J X, TIAN W, GU Y S. Nonlinear aero-elastic analysis of control fin with free-play nonlinearity[J]. Advances in Aeronautical Science and Engineering, 2020, 11(6): 827-835(in Chinese).

[6] 李家旭, 田玮, 谷迎松. 间隙非线性气动弹性系统颤振及控制问题研究进展[J]. 强度与环境, 2021, 48(4): 18-25.
LI J X, TIAN W, GU Y S. Advance in the study on nonlinear flutter and control of aeroelastic system with freeplay nonlinearity[J]. Structure & Environment Engineering, 2021, 48(4): 18-25(in Chinese).

[7] 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.

[8] 陈桂彬, 邹丛青, 杨超. 气动弹性设计基础[M]. 北京: 北京航空航天大学出版社, 2004.
CHEN G B, ZOU C Q, YANG C. Aeroelastic design foundation [M]. Beijing: Beihang University Press, 2004(in Chinese).

[9] 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会议. 北京: 北京航空航天大学飞机所气动弹性研究室, 2009: 230-235.
SONG C, WU Z G, YANG C. Design active flutter suppression control law based on PID controller[C]//Proceedings of the 11th National Airelastic Symposimn. Beijing: BUAA Aeroelatictty Lab, 2009: 230-235(in Chinese).

[10] 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).

[11] 李道春, 向锦武. 间隙非线性气动弹性颤振控制[J]. 北京航空航天大学学报, 2007, 33(6): 640-643.
LI D C, XIANG J W. Flutter control of aeroelasticity with freeplay nonlinearity[J]. Journal of Beijing University of Aeronautics and Astronautics, 2007, 33(6): 640-643(in Chinese).

[12] 李道春, 向锦武. 控制面间隙对非线性二元机翼气动弹性响应的影响[J]. 航空学报, 2009, 30(8): 1385-1391.
LI D C, XIANG J W. Effect of control surface freeplay on nonlinear aeroelastic responses of an airfoil[J]. Acta Aeronautica et Astronautica Sinica, 2009, 30(8): 1385-1391(in Chinese).

[13] 宋晨, 吴志刚, 杨超. 二元机翼滑模变结构控制颤振主动抑制[J]. 北京航空航天大学学报, 2010, 36(12): 1400-1403.
SONG C, WU Z G, YANG C. Sliding mode variable structure control of flutter suppression for two-dimensional wing[J]. Journal of Beijing University of Aeronautics and Astronautics, 2010, 36(12):

1400-1403(in Chinese).

[14] 杨超, 宋晨, 吴志刚, 等. 滑模输出反馈控制在二元机翼颤振主动抑制中的应用[J]. 中国科学: 技术科学, 2010, 40(7): 745-756.
YANG C, SONG C, WU Z G, et al. Application of sliding mode output feedback control in active flutter suppression of two-dimensional wing[J]. Scientia Sinica (Technologica), 2010, 40(7): 745-756 (in Chinese).

[15] 宋晨, 王诗其, 杨超. 基于滑模观测器的机翼颤振主动抑制设计[J]. 北京航空航天大学学报, 2017, 43(6): 1098-1104.
SONG C, WANG S Q, YANG C. Active flutter suppression design of a wing based on sliding mode observer[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(6): 1098-1104(in Chinese).

[16] SINGH S N, YIM W. State feedback control of an aeroelastic system with structural nonlinearity[J]. Aerospace Science and Technology, 2003, 7(1): 23-31.

Limit cycle oscillation suppression of an airfoil based on sliding mode control law

WANG Shiqi[1, 2], ZHANG Zheng[1], SONG Chen[2, *], YANG Chao[2], ZHENG Yanwu[1]

(1. COMAC Beijing Aircraft Technology Research Institute, Beijing 102211, China;
2. School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China)

Abstract: Limit cycle oscillation happens in airfoil systems with freeplay nonlinearity within a specific velocity range. Because the traditional linear control law's limit cycle suppression effect is typically insignificant and its robustness is insufficient, a corresponding nonlinear control law must be designed in order to improve the system's control effect and robustness. According to this article, the limit cycle speed range is optimized between 11.4−16.9 m/s and 12.9−19.2 m/s. The linear quadratic regulartor (LQR) control law is created for the 2D airfoil aeroelastic system with freeplay nonlinearity. At the same time, the sliding mode control law (SMC) is designed for the 2D airfoil aeroelastic system with freeplay nonlinearity above, it shows that the speed range of the limit cycle is optimized from 11.4−16.9 m/s to 29.4−39.3 m/s. The result of the calculating example shows that the SMC law is better than that of LQR for nonlinear systems in this paper.

Keywords: freeplay nonlinearity; limit cycle oscillation; control law; linear quadratic regulator; sliding mode control law

Received: 2023-06-15; **Accepted**: 2023-10-30; **Published Online**: 2023-11-17 16:09
URL: link.cnki.net/urlid/11.2625.V.20231116.1736.004
Foundation item: National Natural Science Foundation of China (11402013)
* **Corresponding author**. E-mail: songchen@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0350

热隔膜成形中隔膜材料性能测试及本构模型选择

赵月青*，林德志，陈汇，汤家力，陈萍
(上海飞机制造有限公司 复合材料中心，上海 201324)

摘 要：隔膜本构模型的选择直接关系到热隔膜成形仿真的精度。为分析5种超弹性本构模型对不同隔膜材料的适用性，对尼龙隔膜和硅橡胶隔膜2种不同类型的隔膜材料进行不同温度单轴拉伸测试，通过应力-应变曲线、断裂延伸率和弹性模量分析2类材料的性能差异；采用5种超弹性本构模型分别对隔膜材料的单轴拉伸试验数据进行拟合，并将试验数据与隔膜材料的单轴拉伸和热隔膜成形过程仿真数据进行对比。结果表明：硅橡胶隔膜材料具有比尼龙隔膜更大的断裂延伸率，但其弹性模量远小于尼龙隔膜材料。Ogden、Polynomial (*N*=2) 和 Marlow 模型对尼龙隔膜和硅橡胶隔膜单轴拉伸数据均具有较高的拟合度，Yeoh 模型对尼龙隔膜单轴拉伸数据拟合度较低，对硅橡胶隔膜的拟合度较高。Mooney-Rivlin 模型无法准确反映大变形状态下超弹性材料的应力-应变关系，对2种隔膜材料的单轴拉伸数据拟合度都较低。以单轴拉伸数据建立的 Ogden 和 Polynomial 模型 (*N*=2) 模型在热隔膜成形模拟时存在过刚现象，而 Marlow 模型则具有很高的预测精度。

关 键 词：隔膜；热隔膜成形；超弹性本构模型；断裂延伸率；有限元仿真
中图分类号：V258; TB125
文献标志码：A **文章编号**：1001-5965(2025)06-2041-10

在热隔膜成形过程中，隔膜材料位于预浸料平面叠层的上表面(单隔膜成形)或上下表面(双隔膜成形)，在真空和热的作用下，依靠自身延展性及与预浸料之间的摩擦力促使预浸料平面叠层变形为模具表面形状。在成形过程中，隔膜材料发生大变形且保持拉伸状态，从而限制了预成形件褶皱、劈裂和薄斑等缺陷的产生[1]。隔膜材料是热隔膜成形的关键之一。

在热隔膜成形过程中，隔膜材料发生大变形，因此，要求隔膜材料具有很好的弹性性能。除此之外，还要求隔膜材料耐温性和刚性好，当预浸料叠层变形时，能对预浸料叠层产生较大的拉力或抵消压应力[2]。在热隔膜成形过程中，隔膜材料的刚度对成形质量具有很大的影响，隔膜厚度过薄容易导致成形件褶皱的产生，因此，在工业生产中通常使用较厚较硬的隔膜来缓解剪切引起的平面外屈曲[3]。Monaghan等[4]研究了隔膜刚度对碳纤维增强聚醚醚酮预浸料制件成形质量的影响，主要涉及3种聚合物隔膜材料(Upilex-R、Upilex-S 和 Kapton-H)及1种金属隔膜，结果表明，在成形速率相同的条件下，制件表面质量随隔膜刚度增加而有所改善，面外屈曲受到抑制，隔膜刚度越大，所需要的成形压力越大，铺层受到较大的压实力，纤维横向流动作用强烈，会造成制件较大的厚度改变。成形过程中，上下隔膜可以采用不同类型的隔膜材料来满足成形的需要。Mallon 等[5]采用双隔膜成形工艺成形半球状零件，上下隔膜的延伸率不同，上隔膜的最大延伸率出现在模具的边缘区域，而下隔膜最大延伸率出现在弯曲的区域。受隔膜的作用，表层预浸料易沿横向方向滑动，隔膜沿着纤维方向产生较大的拉伸应力，容易造成横向方向隔膜的破裂。在成形时，可通过降低隔膜与模具表面的摩擦缓解隔膜的破坏，如在

收稿日期：2023-06-13；录用日期：2023-07-22；网络出版时间：2023-08-31 15:44
网络出版地址：link.cnki.net/urlid/11.2625.V.20230830.1609.004
*通信作者. E-mail：yqz_666@163.com

引用格式：赵月青，林德志，陈汇，等. 热隔膜成形中隔膜材料性能测试及本构模型选择 [J]. 北京航空航天大学学报，2025，51(6)：2041-2050. ZHAO Y Q，LIN D Z，CHEN H，et al. Performance test and constitutive model selection of diaphragm materials in hot diaphragm forming [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2041-2050 (in Chinese).

模具表面擦涂脱模剂。Bersee和Beukers等[6]研究了温度、压强、成形速率对耐高温聚酰亚胺(PI)隔膜变形的影响,成形温度低于275 ℃时,上隔膜滑过下隔膜具有很好的变形性,温度在305 ℃以上时,上隔膜不能在下隔膜上滑动,温度在275～305 ℃之间时,是之前2种情况的过渡。此外,压力对隔膜变形性能的影响很小,成形速率对下隔膜的变形基本没有影响,而上隔膜则对成形速率表现出了一定的依赖性。Walczyk和Hardt[7]通过研究可重构的模具,将表面离散成独立的单元,并通过调整每个单元的高度获得所需要的模具表面形状,使用不同种类的隔膜材料来研究成形制件的质量,结果表明,聚酰亚胺和聚乙烯组合的材料成形的制件质量最好,且发现单隔膜即可实现成形和质量要求,没有必要采用双隔膜,隔膜的刚度及施加的成形压力对成形件的褶皱情况影响很大,且提高真空压力与提高隔膜刚度具有相同的效果。Elgamel[8]研究了压力对隔膜形变的影响,热塑性材料的成形压力一般为0.1～1.7 MPa,而热固性材料一般为0.1 MPa,隔膜材料要有更好的刚性才能减少热塑性材料成形过程中产生的褶皱。文献[6, 9]的研究结果表明,热塑性材料厚度的改变主要依赖于隔膜的高刚度,制备复杂制件需要高的成形压力。

数值模拟在优化成形工艺参数过程中起着重要的作用,一些文献报道了针对热隔膜成形过程的仿真分析方法。O'Brádaigh等[10]在数值模拟过程中将隔膜材料视为各向同性不可压缩牛顿流体。Sorrentino和Bellini[11]使用塑性材料G'Sell模型模拟隔膜材料,该模型基于Maxwell各向同性非线性黏弹性壳单元。Sjölander等[12]使用超弹性Mooney-Rivlin橡胶模型建立隔膜的模型,材料参数通过单轴拉伸测试获得并假设材料不可压缩。Chen等[13-14]采用Ogden超弹性橡胶模型描述硅橡胶的变形行为,材料参数通过单轴和双轴拉伸测试得到,将测试数据与Ogden分析模型及模拟数据进行对比,发现测试数据与Ogden分析模型和模拟数据拟合度很高。之后,该模型被Thompson等[15]用于模拟织物材料的热隔膜成形。杨小亮等[16]研究了DP1000隔膜材料在不同温度和拉伸速率下的力学性能,并采用Polynomial模型建立了超弹性本构模型。王立东等[17-18]基于应变能量密度方程并考虑了温度对隔膜性能的影响,建立了DP1000隔膜材料的超弹性本构模型,在ABAQUS中使用自定义子程序UHYPER实现热隔膜成形过程中隔膜材料模型的建立。Yu等[19]基于隔膜材料单轴拉伸试验的数据和ABAQUS软件自带Marlow模型推导隔膜材料双轴拉伸和平面剪切的应力-应变曲线,结果表明,由Marlow模拟和试验得到的2种应力-应变曲线具有很高的吻合度,证明了Marlow模型的可靠性。

研究人员采用不同本构模型模拟成形过程中隔膜的变形行为,但很少有文献关注隔膜材料模型对热隔膜成形模拟准确性的影响。本文对2种不同类型的隔膜材料进行力学性能测试,采用5种超弹性本构模型对隔膜的应力-应变曲线进行拟合,分析这5种超弹性本构模型的适用性。同时,采用有限元仿真方法对隔膜单轴拉伸和成形过程进行分析,验证超弹性本构模型的适用性。

1 试验材料及方法

1.1 材 料

隔膜材料选用美国Airtech公司一种高温尼龙隔膜和一种硅橡胶隔膜。其中,高温尼龙隔膜厚度为0.06 mm,硅橡胶隔膜厚度为0.5 mm。

1.2 试样制备

中国对于薄膜类塑料材料的拉伸性能测试通常采用GB/T 1040.3—2006标准[20]。由于受测试设备行程和制样条件限制,参考GB/T 1040.3—2006,对其中的1B型试样尺寸进行局部调整,具体的试样尺寸如图1所示,其中,圆弧半径为15 mm。制备完成的试样应无扭曲变形,相邻的切割边缘应互相垂直,表面和边缘应无划痕、孔洞、凹陷和毛刺等缺陷。隔膜材料厚度小,测试过程中变形量很大,无法通过在隔膜材料上贴应变片或加引伸计的方式来获得测试过程中试样的应变情况。为解决该问题,在哑铃型试样中间的均匀宽度区域做标记,如图1中虚线所示。

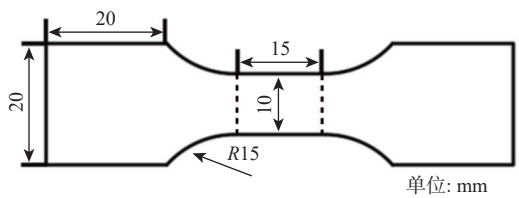


图1 试样尺寸

Fig. 1 Dimensions of specimens

1.3 试样测试

采用长春科新试验仪器有限公司的wdw-100力学试验机进行隔膜材料的拉伸性能测试,该试验机的拉力范围为0～750 N,行程为600 mm。

在25, 60, 70, 80, 90, 100 ℃,拉伸速率为20 mm/min条件下,对隔膜材料进行不同温度条件下的拉伸性能测试。测试过程中,先将隔膜材料装在专门设计的夹具上,再连同夹具一起安装在力学试验机上,将HYGD-40环境试验箱温度从室温逐渐增加到试

验温度，当试样达到试验温度后保温 10 min，以确保试样温度分布的均匀性。

为测试不同温度下的隔膜材料性能，使用无锡惠尔环境试验设备厂型号为 HYGD-40 的环境试验箱进行温度调节，温度控制精度为±0.5℃，如图 2 所示。该环境试验箱的温度控制范围为−120～200 ℃。为监测拉伸过程中的试样变形情况，采用索尼 Alpha 6000 APS-C 微单数码相机对拉伸过程的试样表面进行连续记录。

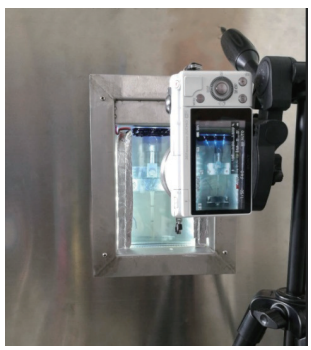


图 2　隔膜材料拉伸测试过程监测

Fig. 2　Tensile test process monitoring of diaphragm materials

测试过程中，采用相机记录拉伸过程试样的变形情况。对获得的视频采用视频播放软件 Potplayer 每隔 10 s 截取一张图片，再采用 ImageJ 软件量取试样拉伸过程中不同时刻标记段的长度，根据式 (1) 计算得到某一时刻的应变，建立时间和应变的关系曲线，结合力学试验机采集的测试过程中力和时间的数据，获得应力-应变的对应数据。

$$\varepsilon = \frac{l}{l_0} \tag{1}$$

$$\varepsilon_{\max} = \frac{l_{\max}}{l_0} \tag{2}$$

式中：ε为某一时刻的应变；l为标距段某一时刻长度；$l_{\max}$为标距段断裂时长度；l_0为初始标距段断裂时长度。

为提高测试过程的数据可靠性，每个测试条件至少都进行 3 次重复测试，并对测试结果取平均值进行分析。

2　基本理论

对于各向同性的超弹性材料，可以用应变能函数来描述[18]：

$$W = (\boldsymbol{C}) \tag{3}$$

式中：W为应变能函数；$\boldsymbol{C}$为右柯西-格林应变张量，$\boldsymbol{C} = \boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}$，$\boldsymbol{F}$为变形梯度，$\boldsymbol{F} = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$。

等温条件下，各向同性超弹性材料的应变能函数也可以用应变不变量表示，即

$$W = W(I_1, I_2, I_3) \tag{4}$$

各不变量I_i可以由右柯西-格林张量求得

$$I_1(\boldsymbol{C}) = \mathrm{tr}(\boldsymbol{C}) \tag{5}$$

$$I_2(\boldsymbol{C}) = \frac{1}{2}\left(\mathrm{tr}^2(\boldsymbol{C}) - \mathrm{tr}(\boldsymbol{C})^2\right) \tag{6}$$

$$I_3(\boldsymbol{C}) = \det(\boldsymbol{C}) \tag{7}$$

第二皮奥拉-基尔霍夫（Piola-Kirchhoff）应力张量$\boldsymbol{K}$可以由应变能函数 (4) 求导获得，即

$$\boldsymbol{K} = 2\frac{\partial W}{\partial \boldsymbol{C}} = 2\sum_{i=1}^{3}\left(\frac{\partial W}{\partial I_i}\cdot\frac{\partial I_i}{\partial \boldsymbol{C}}\right) = 2\left[W_1\boldsymbol{I} + W_2\left(I_1\boldsymbol{I} - \boldsymbol{C}^{\mathrm{T}}\right) + I_3W_3\boldsymbol{C}^{-\mathrm{T}}\right] \tag{8}$$

式中：$\boldsymbol{I}$为单位矩阵；W_i为应变能函数对相应的不变量函数的导数，表示为

$$W_i = \frac{\partial W}{\partial I_i} \tag{9}$$

I_1、I_2、I_3分别为

$$I_1 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 \tag{10}$$

$$I_2 = (\lambda_1\lambda_2)^2 + (\lambda_2\lambda_3)^2 + (\lambda_3\lambda_1)^2 \tag{11}$$

$$I_3 = (\lambda_1\lambda_2\lambda_3)^2 \tag{12}$$

式中：$\lambda = 1 + \varepsilon$，$\varepsilon$为隔膜材料单向拉伸的名义应变，下标 1～3 分别表示 3 个相互正交的方向。

假设隔膜材料为各向同性不可压缩材料，则

$$I_3 = (\lambda_1\lambda_2\lambda_3)^2 = 1 \tag{13}$$

$$W_3 = 0 \tag{14}$$

在单轴拉伸中，对于各向同性材料，有

$$\lambda_2 = \lambda_3 = \lambda^{-\frac{1}{2}} \tag{15}$$

因此，可以分别转化为

$$I_1 = \lambda^2 + \frac{2}{\lambda} \tag{16}$$

$$I_2 = 2\lambda + \frac{1}{\lambda^2} \tag{17}$$

隔膜的应力-应变关系可用应变能函数对λ_i的部分考虑来表示，并结合第二皮奥拉-基尔霍夫和右柯西-格林应变张量，得到应力张量$\boldsymbol{\sigma}_{ij}$与应变张量$\boldsymbol{\gamma}_{ij}$的关系式。

$$\boldsymbol{\sigma}_{ij} = \frac{\partial W}{\partial \boldsymbol{\gamma}_{ij}} = \frac{\partial W}{\partial I_1}\cdot\frac{\partial I_1}{\partial \boldsymbol{\gamma}_{ij}} + \frac{\partial W}{\partial I_2}\cdot\frac{\partial I_2}{\partial \boldsymbol{\gamma}_{ij}} + \frac{\partial W}{\partial I_3}\cdot\frac{\partial I_3}{\partial \boldsymbol{\gamma}_{ij}} \tag{18}$$

将式 (16)、式 (17) 代入式 (18) 中，可得

$$\sigma_{11} = \frac{\partial W}{\partial I_1}\cdot\frac{\partial I_1}{\partial \lambda} + \frac{\partial W}{\partial I_2}\cdot\frac{\partial I_2}{\partial \lambda} = 2\left(1 - \frac{1}{\lambda^3}\right)\left(\lambda\frac{\partial W}{\partial I_1} + \frac{\partial W}{\partial I_2}\right) \tag{19}$$

1）Polynomial 模型。该模型最早由 Rivlin[21] 提

出，其最一般的形式为

$$W=\sum_{i+j=1}^{N}C_{ij}(I_1-3)^i(I_2-3)^j+\sum_{i}^{N}\frac{1}{D_i}(J-1-V)^{2i} \tag{20}$$

式中：J为变形前和变形后的体积比，对于不可压缩的材料，体积保持不变，因此，$J=\lambda_1\lambda_2\lambda_3=1$；假设体积膨胀随温度变化为零，即 $V(T)=0$。

因此，式 (20) 可简化为

$$W=\sum_{i+j=1}^{N}C_{ij}(I_1-3)^i(I_2-3)^j \tag{21}$$

将式 (21) 代入式 (19)，可得

$$\sigma_{11}=2\left(1-\frac{1}{\lambda^3}\right)\left[\lambda\sum_{i+j=1}^{N}C_{ij}(I_1-3)^{i-1}(I_2-3)^j+\sum_{i+j=1}^{N}C_{ij}(I_1-3)^i(I_2-3)^{j-1}\right] \tag{22}$$

当 N=1 时，Rivlin 模型可以简化为 Mooney-Rivlin 模型：

$$W=C_{10}(I_1-3)+C_{01}(I_2-3) \tag{23}$$

同时，式 (22) 可转换为

$$\sigma_{11}=2\left(1-\frac{1}{\lambda^3}\right)(\lambda C_{10}+C_{01}) \tag{24}$$

当 N=2 时，Polynomial 模型为

$$W=C_{10}(I_1-3)+C_{01}(I_2-3)+C_{11}(I_1-3)(I_2-3)+C_{20}(I_1-3)^2+C_{02}(I_2-3)^2 \tag{25}$$

工程应力表示为

$$\sigma_{11}=2\left(1-\frac{1}{\lambda^3}\right)\{\lambda[C_{10}+C_{11}(I_2-3)+2C_{20}(I_1-3)]+C_{01}+C_{11}(I_1-3)+2C_{02}(I_2-3)\} \tag{26}$$

当 $\lambda\to1$ 时，式 (26) 对应变求导，可得

$$E=\lim_{\lambda\to1}\frac{\partial\sigma_{11}}{\partial\lambda}=6(C_{10}+C_{01}) \tag{27}$$

式中：E为初始弹性模量。式(27)可用于计算隔膜材料的初始弹性模量[17]。

2）Reduction Polynomial 模型。对于各向同性和不可压缩材料，简化多项式模型的应变能函数可写为

$$W=\sum_{i=1}^{N}C_{i0}(I_1-3)^i \tag{28}$$

将式 (28) 代入式 (19)，可得

$$\sigma_{11}=2\left(\lambda-\frac{1}{\lambda^2}\right)\sum_{i=1}^{N}C_{i0}(I_1-3)^{i-1} \tag{29}$$

当 N=3 时，即可得到 Yeoh 模型，其表达式如下：

$$W=C_{10}(I_1-3)+C_{20}(I_1-3)^2+C_{30}(I_1-3)^3 \tag{30}$$

将式 (30) 代入式 (19)，可得

$$\sigma_{11}=2\left(\lambda-\frac{1}{\lambda^2}\right)C_{10}+\left[C_{20}\left(\lambda^2+\frac{2}{\lambda}-3\right)+C_{30}\left(\lambda^2+\frac{2}{\lambda}-3\right)^2\right] \tag{31}$$

3）Ogden 模型。Ogden[22] 于 1972 年提出 Ogden 模型，具有各向同性和不可压缩性材料的应变能函数可以写为

$$W=\sum_{i=1}^{N}\frac{2\mu_i}{\alpha_i^2}\left(\lambda_1^{-\alpha_i}+\lambda_2^{-\alpha_i}+\lambda_3^{-\alpha_i}-3\right) \tag{32}$$

式中：α_i、μ_i 为模型参数，可由试验获得。

代入式 (19)，可得

$$\sigma_{11}=\sum_{i=1}^{3}\frac{2\mu_i}{\alpha_i}\left(\lambda^{\alpha_i-1}-\lambda^{-\frac{1}{2}\alpha_i-1}\right) \tag{33}$$

4）Marlow 模型。Marlow 模型为 ABAQUS 软件内置的超弹性本构模型，其能量密度函数为[23]

$$W=W_{\mathrm{dev}}(I_1)+W_{\mathrm{vol}}(J_{\mathrm{el}}) \tag{34}$$

式中：$W_{\mathrm{dev}}(I_1)$为偏量部分；$W_{\mathrm{vol}}(J_{\mathrm{el}})$为体积部分；$J_{\mathrm{el}}$为弹性体积比。

实际上，Marlow 模型没有特定的应变能密度函数，在超弹性本构关系的理论框架下，通过平滑、插值、外推等数值方法，直接确定了 Marlow 模型的应变能密度与应力应变的数值关系。

3 结果与分析

3.1 试验结果

由于常用的热固性单向带预浸料热隔膜成形温度在 60～100 ℃ 之间，选择在 25, 60, 70, 80, 90, 100 ℃，拉伸速率为 20 mm/min 条件下，对 2 种隔膜材料进行不同温度条件下的拉伸性能测试。尼龙隔膜和硅橡胶隔膜的应力-应变曲线分别如图 3 和图 4 所示。可以看出，随着温度的升高，隔膜材料的应力随之下降。对于超弹性膜材料，在拉伸开始时是各向同性材料，屈服之后，高分子链段沿着拉伸方向取向，此时应变快速增加，当高分子链完全取向后，力值快速增加。由图 4 可以看出，硅橡胶隔膜材料的应力水平远低于尼龙隔膜材料，在室温具有最高的应力水平，当应变为 6.5 时，应力值也仅为 4.5 MPa。当温度在 60～100 ℃ 内变化时，硅橡胶隔膜材料的变形能力相差很小，应力-应变曲线接近重合，说明硅橡胶隔膜材料在此温度区间具有较好的力学性能稳定性。

图 5 和图 6 分别为尼龙隔膜和硅橡胶隔膜在不同温度下的断裂延伸率。从图 5 可以看出，在测试温度范围内，尼龙隔膜断裂延伸率随着温度的升高

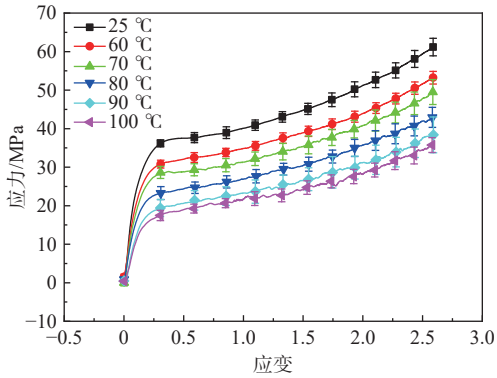


图 3　尼龙隔膜不同温度条件下的应力-应变曲线

Fig. 3　Stress-strain curves of nylon diaphragm at different temperatures

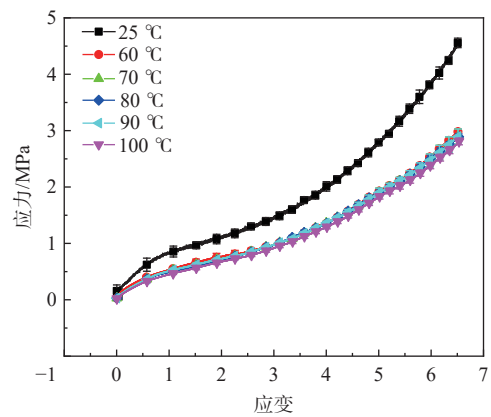


图 4　硅橡胶隔膜不同温度条件下的应力-应变曲线

Fig. 4　Stress-strain curves of silicone rubber diaphragm at different temperatures

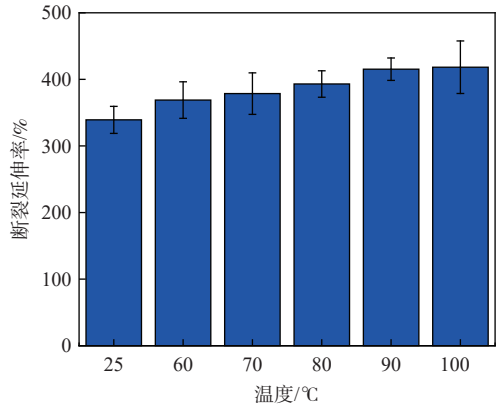


图 5　不同温度条件下尼龙隔膜的断裂延伸率

Fig. 5　Fracture elongation of nylon diaphragm at different temperatures

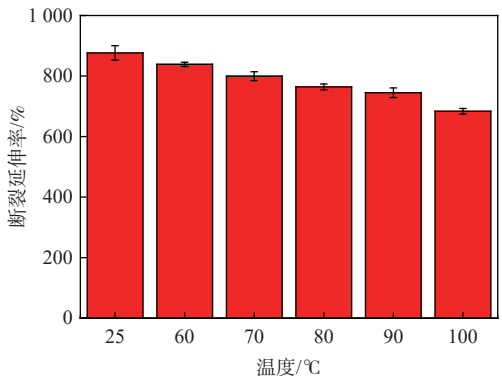


图 6　不同温度条件下硅橡胶隔膜的断裂延伸率

Fig. 6　Fracture elongation of silicone rubber diaphragm at different temperatures

而逐渐升高，断裂延伸率在 300%～450% 之间变化。硅橡胶隔膜相比尼龙隔膜材料具有较好的延伸性能，其断裂延伸率在 600%～900% 之间变化。与尼龙隔膜材料不同的是，硅橡胶隔膜材料的断裂延伸率随着温度的升高反而下降。可能是因为随着温度的升高，橡胶分子链运动变得越来越活跃，分子链的柔顺性增加，在外力的作用下，分子链的取向变得容易；同时温度升高后，伴随着橡胶体积发生膨胀，加大分子链间的自由空间，使得分子链的运动更加自由。这 2 种效果同时作用，一定程度上降低了分子链的物理交联点，减小了分子间作用力，宏观表现为断裂延伸率下降。

3.2　隔膜材料弹性模量计算

由图 3 和图 4 可以看出，隔膜材料在屈服前的变形阶段并非线性变化，采用传统的计算模量的方法测量初始变形的模量难以取值。将 2 种隔膜材料在不同温度测试获得的应力-应变曲线按照式 (26) 进行拟合，得到C_{10}、C_{01}、C_{11}、C_{20}、C_{02}，并结合式 (27) 进行计算，得到不同隔膜材料对应的初始弹性模量[17]，结果分别如表 1 和表 2 所示。从表 1 可以看出，随着温度的升高，隔膜材料的弹性模量降低。尼龙隔膜材料的弹性模量由室温的 345.65 MPa 降低到 100 ℃ 的 151.49 MPa，降低明显。而硅橡胶隔膜在室温时弹性模量最大，为 1.87 MPa，温度从 60 ℃ 升至 100 ℃，弹性模量从 1.04 MPa 降到 0.65 MPa。同种温度条件下，硅橡胶隔膜的弹性模量远小于尼龙隔膜材料。

3.3　模型对试验数据拟合结果

采用常用的 5 种超弹性本构模型（包括 Mooney-Rivlin 模型、Ogden 模型、Yeoh 模型、Polynomial 模型和 Marlow 模型）对隔膜材料在 80 ℃ 获得的应力-应变曲线进行拟合，其中，Marlow 模型的数据是通过 ABAQUS 软件的 Evaluate 功能对测试数据进行的模型评估，其他 4 种模型均通过自定义公式在 Origin 软件中拟合得到，2 种隔膜材料单轴拉伸测试数据的模型拟合参数分别如表 3 和表 4 所示。

由图 7 的拟合结果可以看出，对于尼龙隔膜材料，Ogden、Polynomial 和 Marlow 模型都能获得很高的拟合度，而 Mooney-Rivlin 和 Yeoh 模型拟合度

表 1 尼龙隔膜不同温度下的模型参数 (Polynomial, N=2)

Table 1 Model parameters of nylon diaphragm at different temperatures (Polynomial, N= 2)

参数/℃	C_{01}	C_{10}	C_{02}	C_{20}	C_{11}	E/MPa
25	194.61	−136.99	50.81	1.29	−7.81	345.65
60	176.74	−126.26	50.07	1.59	−9.40	302.89
70	160.36	−114.28	43.46	1.17	−7.09	276.45
80	131.94	−93.75	35.60	0.94	−5.72	229.14
90	95.45	−66.07	25.45	0.79	−4.44	176.28
100	76.43	−51.18	19.01	0.59	−3.14	151.49

表 2 硅橡胶隔膜不同温度下的模型参数 (Polynomial, N=2)

Table 2 Model parameters of silicone rubber diaphragm at different temperatures (Polynomial, N= 2)

参数/℃	C_{01}	C_{10}	C_{02}	C_{20}	C_{11}	E/MPa
25	−0.07	0.39	−0.10	1.81×10^{-3}	2.87×10^{-3}	1.87
60	−0.22	0.39	−0.12	2.27×10^{-4}	0.01	1.04
70	−0.28	0.43	−0.13	3.69×10^{-5}	0.01	0.90
80	−0.26	0.39	−0.12	-1.56×10^{-4}	0.01	0.80
90	−0.36	0.48	−0.14	-1.27×10^{-4}	0.01	0.72
100	−0.29	0.40	−0.12	-7.75×10^{-5}	0.01	0.65

表 3 尼龙隔膜不同模型的拟合参数

Table 3 Fitting parameters of different models for nylon diaphragm

模型	参数
Mooney-Rivlin	C_{10}=1.95, C_{01}=12.73
Ogden	u_1=15.70, u_2=15.65, u_3=0.15, α_1=−1.53, α_2=−1.53, α_3=−12.00
Yeoh	C_{10}=11.10, C_{20}=−1.51, C_{30}=0.11
Polynomial, N=2	C_{01}=131.94, C_{10}=−93.75, C_{02}=35.61, C_{20}=0.94, C_{11}=−5.72

表 4 硅橡胶隔膜不同模型的拟合参数

Table 4 Fitting parameters of different models for silicone rubber diaphragm

模型	参数
Mooney-Rivlin	C_{10}=0.21, C_{01}=−0.29
Ogden	u_1=0.29, u_2=0.05, u_3=0.09, α_1=0.09, α_2=0.09, α_3=−6.54
Yeoh	C_{10}=0.12, C_{20}=7.65×10^{-4}, C_{30}=1.30×10^{-5}
Polynomial, N=2	C_{01}=−0.26, C_{10}=0.39, C_{02}=−0.12, C_{20}=-1.57×10^{-4}, C_{11}=0.01

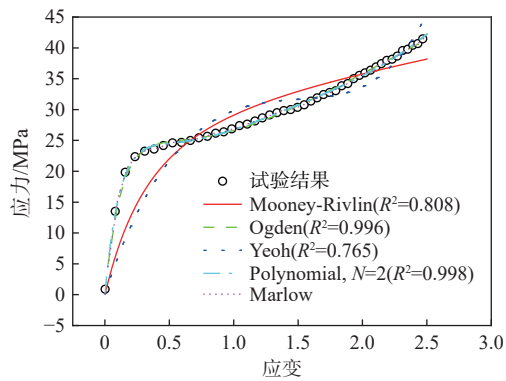


图 7 尼龙隔膜试验数据拟合结果

Fig. 7 Fitting results of experimental data of nylon diaphragm

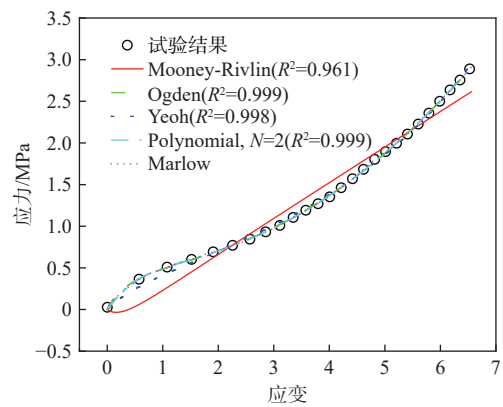


图 8 硅橡胶隔膜试验数据拟合结果

Fig. 8 Fitting results of experimental data of silicone rubber diaphragm

相对较低，这 2 种模型都没能很好地捕捉应力-应变曲线在大应变处的转折。但从图 8 可以看出，对于硅橡胶材料而言，采用 Ogden、Yeoh、Polynomial 和 Marlow 模型进行拟合都能获得较高的拟合度，拟合曲线几乎与测试数据完全重合，而 Mooney-Rivlin 模型的拟合度相对比较低。这是由于 Mooney-Rivlin 模型仅保留了有限的项数，无法准确反映大变形状态下超弹性材料的应力-应变关系，尤其是“S”型应力-应变曲线。Yeoh 模型则是舍弃了式 (21) 中的 I_2 项，并保留小于或等于三阶的项，对于具有大非线性特征的应力-应变曲线，会产生较大偏差。综上，对于本文选取的 2 种隔膜，Ogden、Polynomial 和 Marlow 模型对于单轴拉伸测试数据具有较高的

适用性。

3.4　模拟结果与试验结果对比

3.4.1　隔膜材料单轴拉伸模拟

由于橡胶隔膜材料的厚度较大，较难保证预成形体厚度的均匀性，目前，航空结构件预成形主要采用塑料类隔膜材料，在本文中只针对尼龙隔膜材料进行本构模型的准确性验证。在 ABAQUS 中根据图 1 所示的隔膜材料的尺寸建立隔膜材料的模型。由 3.3 节可知，对于尼龙隔膜材料，Ogden 模型、Polynomial 模型和 Marlow 模型具有较高的拟合度，因此，在本文中分别采用这 3 种模型建立隔膜材料的本构模型，输出隔膜材料的应力-应变曲线并与试验值进行对比。

在模拟过程中，由于隔膜材料很薄，隔膜材料采用 Membrane 单元建立模型，单元类型为 M3D4R，网格尺寸为 0.5 mm。隔膜两端为夹持端，在仿真过程中设置为刚体，在拉伸过程中不发生变形。隔膜材料一端固定，另外一端设置位移约束。

采用 3 种超弹性材料本构模型对隔膜材料的拉伸过程进行模拟，提取试样中央均匀变化区域的应变值。图 9 为尼龙隔膜材料拉伸试验与模拟结果的应力-应变曲线对比。可以看出，3 种超弹性本构模型的模拟结果与试验结果几乎完全重合。

3.4.2　隔膜材料热隔膜成形铺覆性模拟

利用自主搭建的热隔膜成形装置进行隔膜材料成形试验，如图 10(a) 所示。热隔膜成形装置包括紧固夹、固定框、隔膜材料、真空管、硅橡胶垫等。硅橡胶按照固定框的尺寸进行裁剪并置于隔膜上下面边缘，保证装置的气密性，在尼龙隔膜材料上画上边长为 25 mm 的小方格，方便观察隔膜材料的变形情况。试验过程中，C 型模具置于腔体中间并用黄胶固定位置，防止变形过程中位置发生移动。腔体底部铺上透气毡，底部四周需要铺上三角形硅橡胶，防止隔膜在此处架桥应力过大导致隔膜破裂，成形之后隔膜的变形如图 10(b)所示。

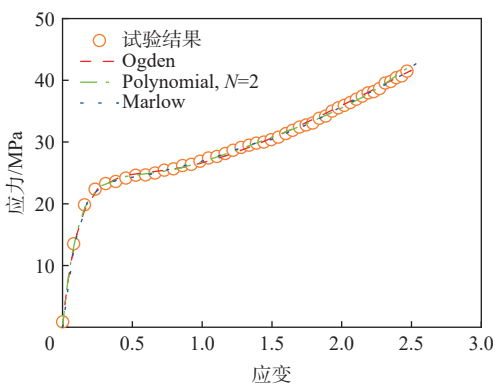
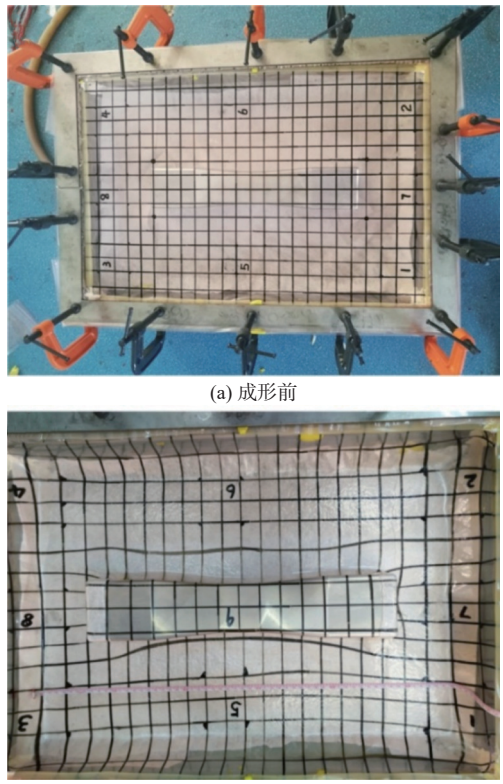


图 9　尼龙隔膜应力-应变曲线试验与模拟结果对比

Fig. 9　Comparison of experimental and simulation results of stress-strain curves for nylon diaphragm



(a) 成形前

(b) 成形后

图 10　隔膜成形试验

Fig. 10　Diaphragm forming test

在 ABAQUS 中建立隔膜成形的模型，如图 11 所示。隔膜采用 M3D4R 单元进行网格划分，C 型模具及腔体均采用刚性单元进行网格划分。模型中，C 型模具和成形腔体按照实际成形的尺寸建立，隔膜材料位于 C 型模具上表面，隔膜四周固定，上表面施加 0.1 MPa 的压力，在成形过程中，缓慢达到设定的压力。隔膜材料分别采用 Ogden, N=3 模型、Marlow 模型和 Polynomial, N=2 模型，模拟结果如图 12(a)所示。从图 12(b)可以看出，成形后模具底部的架桥宽度约为 25 mm，而采用 Marlow 模型获得的模拟结果显示的架桥宽度约为 25 mm，Ogden, N=3 模型和 Polynomial, N=2 模型约为 45 mm，说明仅仅采用单轴拉伸数据建立 Ogden 和 Polynomial 模型，在进行热隔膜成型模拟时，隔膜材料表现出

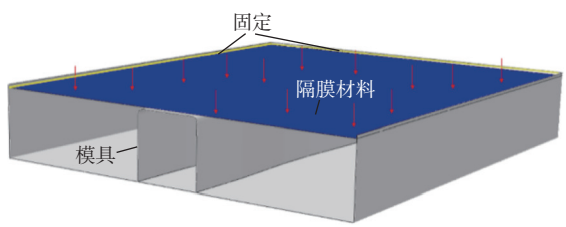
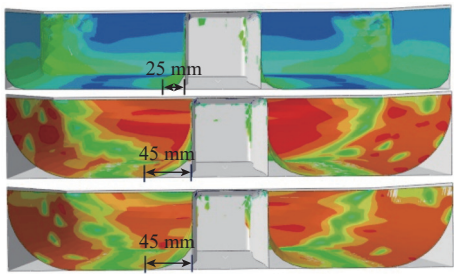
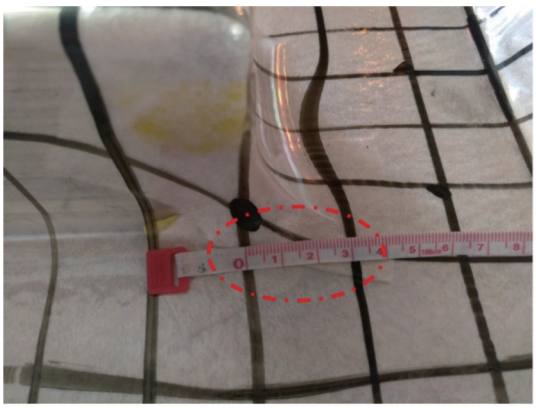


图 11　隔膜成形有限元模型

Fig. 11　Finite element model of diaphragm forming



(a) 3种本构模型的模拟结果



(b) 试验值

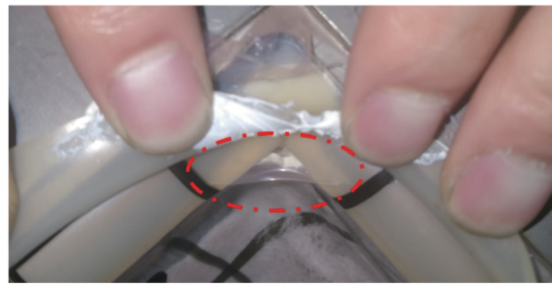
图 12 隔膜成形底部架桥宽度

Fig. 12 Bridge width at bottom of diaphragm forming

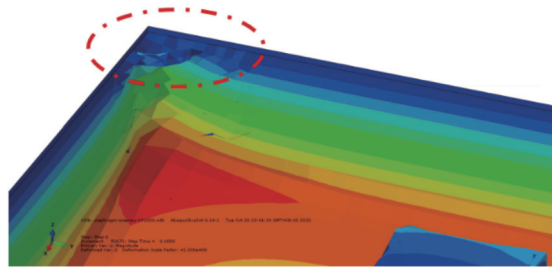
过刚行为。这可能是由于热隔膜成型时，隔膜材料受到多轴拉伸或剪切作用，由单轴拉伸测试数据获得的 Ogden 和 Polynomial 模型不能很好地预测复杂载荷下的应变，该现象与文献 [19, 24]中的结论一致。隔膜材料架桥过大可导致成形结构件缘条的底端不完全贴向模具面，造成翘曲。因此，在模拟过程中建议采用 Marlow 模型。如图 13(a)所示，在实际成形结束后，腔体的 4 个边角位置隔膜材料架空并产生褶皱，在模拟结果中，也观察到了同样的现象，进一步证明了采用 Marlow 模型作为隔膜材料本构模型的准确性。

3.4.3 预浸料热隔膜成形过程模拟

采用 ABAQUS 对 80 ℃ 环境下预成型体的热隔膜预成型过程进行仿真，进一步验证 Marlow 模型作为隔膜材料本构模型的准确性。本文采用双隔膜的成形方式。预浸料铺层为 $[45/90/-45/90]_{3s}$，初始尺寸为 300 mm×150 mm。预浸料采用 S4R 单元进行网格划分，预浸料位于 2 层隔膜材料之间。在整个双隔膜成型过程中分 2 步进行成型。首先，固定隔膜四周，在上下隔膜靠近预浸料一侧各施加 0.1 MPa 的负压，模拟预制体的封装；然后，在上隔膜上表面缓慢施加 0.1 MPa 的压力，隔膜与预浸料在压力的作用下与 C 型模具贴合，形成所需要的形状。试验件仿真与试验结果对比如图 14 所示。可以看出，在 C 型结构件腹板凹陷区域中间和缘条位



(a) 试验结果



(b) 模拟结果

图 13 隔膜成形边角变形

Fig. 13 Diaphragm forming edge deformation

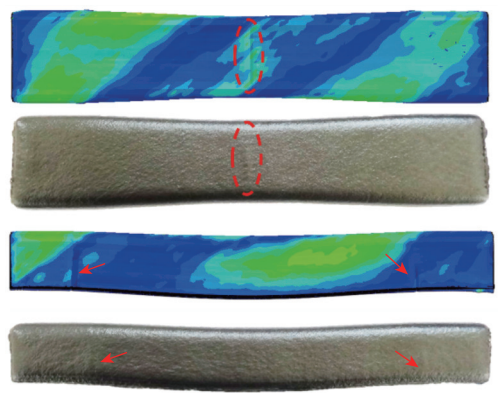


图 14 双隔膜成形试验件仿真与试验结果对比

Fig. 14 Comparison of simulation and experiment results of double-diaphragm forming specimen

置的过渡区域都出现褶皱，所建立的有限元仿真方法能有效预测预成型体的褶皱。

4 结 论

1）Ogden、Polynomial (N=2) 和 Marlow 这 3 种模型对于尼龙隔膜和硅橡胶隔膜材料单轴拉伸测试数据具有较高的拟合度，Yeoh 模型对尼龙隔膜单轴拉伸数据拟合度较低，对硅橡胶隔膜的拟合度较高。Mooney-Rivlin 模型对 2 种隔膜材料的单轴拉伸数据拟合度都较低。

2）采用基于单轴拉伸数据建立的 Ogden 和 Polynomial (N=2) 模型进行热隔膜成形模拟时，隔膜材料表现出过刚行为。Marlow 模型更适合作为尼龙隔膜材料的本构模型。

3）当隔膜材料测试数据有限时，应采用多种本

构模型对测试数据进行拟合，并结合成型过程仿真分析进行验证，最终确定合理的本构模型。

参考文献（References）

[1] TANG Z X, YE L. Characterisation of mechanical properties of thin polymer films using a bi-axial tension based on blow-up test[J]. Plastics Rubber and Composites, 2003, 32(10): 459-465.

[2] 吴志恩. 复合材料热隔膜成型[J]. 航空制造技术, 2009, 52(增刊 2): 113-116.
WU Z E. Hot drape forming of composites [J]. Aeronautical Manufacturing Technology, 2009, 52(Sup 2): 113-116(in Chinese).

[3] O'BRÁDAIGH C M, PIPES R B, MALLON P J. Issues in diaphragm forming of continuous fiber reinforced thermoplastic composites[J]. Polymer Composites, 1991, 12(4): 246-256.

[4] MONAGHAN M R, MALLON P J, O'BRADAIGH C M, et al. The effect of diaphragm stiffness on the quality of diaphragm formed thermoplastic composite components[J]. Journal of Thermoplastic Composite Materials, 1990, 3(3): 202-215.

[5] MALLON P J, O'BRÁDAIGH C M, PIPES R B. Polymeric diaphragm forming of complex-curvature thermoplastic composite parts [J]. Composites, 1989, 20(1): 48-56.

[6] BERSEE H E N, BEUKERS A. Diaphragm forming of continuous fibre reinforced thermoplastics: influence of temperature, pressure and forming velocity on the forming of Upilex-R® diaphragms[J]. Composites Part A: Applied Science and Manufacturing, 2002, 33(7): 949-958.

[7] WALCZYK D F, HARDT D E. Design and analysis of reconfigurable discrete dies for sheet metal forming[J]. Journal of Manufacturing Systems, 1998, 17(6): 436-454.

[8] ELGAMEL H E. Closed-form expressions for the relationships between stress, diaphragm deflection, and resistance change with pressure in silicon piezoresistive pressure sensors[J]. Sensors and Actuators A: Physical, 1995, 50(1-2): 17-22.

[9] GUTOWSKI T G, DILLON G, CHEY S, et al. Laminate wrinkling scaling laws for ideal composites[J]. Composites Manufacturing, 1995, 6(3-4): 123-134.

[10] O'BRÁDAIGH C M, MCGUINNESS G B, PIPES R B. Numerical analysis of stresses and deformations in composite materials sheet forming: central indentation of a circular sheet[J]. Composites Manufacturing, 1993, 4(2): 67-83.

[11] SORRENTINO L, BELLINI C. Potentiality of hot drape forming to produce complex shape parts in composite material[J]. The International Journal of Advanced Manufacturing Technology, 2016, 85(5): 945-954.

[12] SJÖLANDER J, HALLANDER P, ÅKERMO M. Forming induced wrinkling of composite laminates: a numerical study on wrinkling mechanisms[J]. Composites Part A: Applied Science and Manufacturing, 2016, 81: 41-51.

[13] CHEN S, MCGREGOR O P L, HARPER L T, et al. Defect formation during preforming of a bi-axial non-crimp fabric with a pillar stitch pattern[J]. Composites Part A: Applied Science and Manufacturing, 2016, 91: 156-167.

[14] CHEN S, MCGREGOR O P L, ENDRUWEIT A, et al. Double diaphragm forming simulation for complex composite structures[J]. Composites Part A: Applied Science and Manufacturing, 2017, 95: 346-358.

[15] THOMPSON A J, BELNOUE J P, HALLETT S R. Modelling defect formation in textiles during the double diaphragm forming process[J]. Composites Part B: Engineering, 2020, 202: 108357.

[16] 杨小亮, 彭雄奇, 王立冬, 等. 热隔膜成形中隔膜的超弹性本构模型[J]. 模具技术, 2016(6): 14-18.
YANG X L, PENG X Q, WANG L D, et al. Hyperelastic constitutive model of diaphragm in hot diaphragm forming[J]. Die and Mould Technology, 2016(6): 14-18(in Chinese).

[17] WANG L D, WANG J C, LIU M R, et al. Development and verification of a finite element model for double diaphragm preforming of unidirectional carbon fiber prepreg[J]. Composites Part A: Applied Science and Manufacturing, 2020, 135: 105924.

[18] 王立冬, 魏冉, 徐鹏, 等. 基于温度的隔膜超弹性本构模型[J]. 上海交通大学学报, 2017, 51(9): 1025-1030.
WANG L D, WEI R, XU P, et al. A temperature-dependent hyperelastic constitutive model for diaphragm[J]. Journal of Shanghai Jiao Tong University, 2017, 51(9): 1025-1030(in Chinese).

[19] YU F, CHEN S, HARPER L T, et al. Double diaphragm forming simulation using a global-to-local modelling strategy for detailed defect detection in large structures[J]. Composites Part A: Applied Science and Manufacturing, 2021, 147: 106457.

[20] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 塑料 拉伸性能测试 第 3 部分: 薄膜和薄片的试验条件: GB/T 1040.3—2006[S]. 北京: 中国标准出版社, 2006.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China and Standardization Administration of the People's Republic of China. Plastics- Determination of tensile properties- Part 3: Test conditions for films and sheets: GB/T1040.3-2006 [S]. Beijing: China Standard Press, 2006.

[21] RIVLIN R S. Large elastic deformations of isotropic materials IV. further developments of the general theory[J]. Philosophical Transactions of the Royal Society of London, Series A, Mathematical and Physical Sciences, 1948, 241(835):379-397.

[22] OGDEN R W. Large deformation isotropic elasticity-on the correlation of theory and experiment for incompressible rubberlike solids[J]. Proceedings of the Royal Society of London Series A, 1972, 326(1567): 565-584.

[23] Dassault Systemes. Abaqus simulia user assistance[EB/OL]. [2023-05-16]. https://help.3ds.com/2022/english/dssimulia_established/SIMACAEMATRefMap/simamat-c-hyperelastic.htm?contextscope=all#simamat-c-hyperelastic-marlow.

[24] 王永冠, 李心, 黄友剑. 橡胶计算中本构模型的选择[C] // "时代新材杯"全国橡胶制品技术研讨会. 北京: 中国化工学会, 中国汽车工程学会, 2007: 443-449.
WANG Y G, LI X, HUANG Y J. Selection of constitutive models in rubber calculation[C]//National Rubber Product Technology Seminar on "New Materials Cup of the Era". Beijing: China Chemical Society, China Automotive Engineering Society, 2007: 443-449(in Chinese).

Performance test and constitutive model selection of diaphragm materials in hot diaphragm forming

ZHAO Yueqing*，LIN Dezhi，CHEN Hui，TANG Jiali，CHEN Ping

(Composites Center，Shanghai Aircraft Manufacturing Co.，Ltd.，Shanghai 201324，China)

Abstract：The selection of the diaphragm constitutive model is directly related to the precision of hot diaphragm forming simulation. In order to analyze the applicability of five hyperelastic constitutive methods to different diaphragm materials, uniaxial tensile tests were carried out on two different types of diaphragm materials (nylon diaphragm and silicone rubber diaphragm) at different temperatures, and the performance differences of the two types of materials were analyzed by stress-strain curves, fracture elongation, and modulus. Five hyperelastic constitutive models were used to fit the data of diaphragm materials in the uniaxial tensile tests, and the data were compared with the simulation data of the uniaxial tension and hot diaphragm forming process of the diaphragm materials. The results indicate that the rubber diaphragm material has a greater fracture elongation than the nylon diaphragm, but its modulus is much smaller than the nylon diaphragm material. The Ogden, Polynomial (N = 2), and Marlow models have a high degree of fit for uniaxial tensile data of nylon and rubber diaphragms. However, the Yeoh model has a low degree of fit for uniaxial tensile data of nylon diaphragms but a high degree of fit for rubber diaphragms. Mooney-Rivlin model fails to accurately reflect the stress-strain relationship of hyperelastic material under large deformation, and the degree of fit of uniaxial tensile data of two diaphragm materials is relatively low. The Ogden and Polynomial (N = 2) models established based on uniaxial tensile data exhibit excessive stiffness during the simulation of hot diaphragm forming, while the Marlow model has high prediction accuracy.

Keywords：diaphragm；hot diaphragm forming；hyperelastic constitutive model；fracture elongation；finite element simulation

Received：2023-06-13；**Accepted**：2023-07-22；**Published Online**：2023-08-31 15：44

URL：link.cnki.net/urlid/11.2625.V.20230830.1609.004

* **Corresponding author**. E-mail：yqz_666@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0349

基于自适应积分电容的高动态像素结构

刘绥阳，郭仲杰*，余宁梅，许睿明
(西安理工大学 自动化与信息工程学院，西安 710048)

摘 要：由于红外图像传感技术具备不受环境影响、目标识别度好、抗干扰能力强等优点而受到广泛关注，但随着红外焦平面集成度的提升，光电系统的动态范围、噪声和满阱之间的制约关系尤为突出。为解决弱光下噪声和强光下满阱容量的矛盾，在5T红外像素电路中，利用反型MOS电容在特定电压区间内电容值和电压的关系，使红外图像传感器积分电容从6.5 fF到37.5 fF自动变化，提出一种基于自适应积分电容的高动态像素结构，并基于55 nm CMOS工艺技术，在12 288×12 288像素规模的红外图像传感器中研究其性能参数。结果表明：5.5 μm×5.5 μm的小尺寸像素具有1.31 Me^-的大满阱容量和可变的转换增益，噪声电子数小于0.43 e^-，动态范围超过130 dB。

关 键 词：CMOS红外图像传感器；自适应积分电容；高动态范围；大满阱容量；像素读出电路

中图分类号：TN47

文献标志码：A **文章编号**：1001-5965(2025)06-2051-09

CMOS图像传感器自20世纪中期以来已代替CCD成为主流，广泛流通于智能手机、监控摄像、自动驾驶、人工智能等各级消费市场[1-3]。但一般的线性CMOS红外图像传感器探测到的光范围小，不能兼具低照明和高照明环境下的实物信息辨认，具有较小的动态范围，影响红外图像传感器的输出图像质量。高动态图像传感器获取的图像能保留更多的细节信息，因此，其相关的研究至关重要。高动态红外图像传感器可通过改善系统和读出电路的方法实现，如搭建高动态图像数据获取系统和多次捕获读出电路，实现多个图像的组合和高动态成像[4-7]。但是，多次图像的叠加导致噪声不断叠加，而高动态红外像素在像素内部进行改善，单次成像避免了噪声叠加的缺点，并具有结构简单的优点。

CMOS红外图像传感器的核心在于像素，因此，研究人员针对动态范围较小的问题，提出各具优势的解决方法[8]。

文献[9]通过3D堆叠技术增大积分电容，提升满阱容量(full well capacity, FWC)和动态范围。文献[10]提出在每个像素中采用横向溢流电容器，其电容密度为24.3 fF/μm^2，约为平面MOS电容的5倍，可以实现非常高的满阱容量，可达24.3 Me^-，提升动态范围至130 dB[11]。Uchida等开发了一种光电二极管结构，将动态范围提升至97 dB[12]。但上述3种方法都具有工艺繁杂和成品率难以提升的缺点[13-14]。

文献[15]通过像素内集成相关双采样(correlated double sampling, CDS)技术来减小噪声，把动态范围从60 dB提升到74 dB，但内部集成的CDS将占用额外的面积。文献[16]提出一种像素内部电荷相减的方法，既提升昏暗信号的信噪比，又避免了明亮信号的饱和，提升动态范围至74 dB。但是，像素内部具有比较器和减法器占用面积太

收稿日期：2023-06-12；**录用日期**：2023-08-24；**网络出版时间**：2023-09-08 16:27
网络出版地址：link.cnki.net/urlid/11.2625.V.20230908.0902.001
基金项目：国家自然科学基金(62171367)；陕西省重点研发计划(2021GY-060)；陕西省创新能力支撑计划(2022TD-39)；西安理工大学校企协同基金(252062213)；西安理工大学博士创新基金(103-252072401)
* **通信作者**. E-mail：zjguo@xaut.edu.cn

引用格式：刘绥阳，郭仲杰，余宁梅，等. 基于自适应积分电容的高动态像素结构[J]. 北京航空航天大学学报，2025，51(6)：2051-2059. LIU S Y，GUO Z J，YU N M，et al. Structure of highly dynamic pixel based on adaptive integral capacitance[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2051-2059 (in Chinese).

多[17]。文献[18]利用线性像素在低照度下灵敏度较高、对数像素在高照度下灵敏度较高的特点，提出了一种线性对数有源像素结构，同时获得线性像素和对数像素的动态范围，实现了152 dB的高动态范围。但是，采用2套像素单元将占用更多的面积。总之，上述3种方法都会影响像素的填充因子[19]。

文献[20]提出通过调整衬底电压和摆幅来提升动态范围，但衬底电压的提升势必会带来更高的功耗。文献[21]提出在读出电路的每个单元中自适应调整积分时间，在中度光照下保证线性度和转换增益，但在接近满阱的情况下降低敏感度，这种方法可提升动态范围至100 dB。但是，这种调整积分时间的做法将给时序的安排、数字电路的规划带来不小的麻烦。

动态范围的提升是指图像传感器在弱光下对光变化十分敏感，而强光下不易达到满阱。因此，本文在分析红外图像传感器噪声与满阱间制约关系的基础上，基于动态范围的定义建立数学模型，提出自适应积分电容红外图像传感器的像素结构。同时，研究自适应积分电容容值与光强间的关系，分析自适应红外像素电路的线性度、转换增益、满阱容量与噪声。在一款12 288×12 288大面阵红外图像传感器中应用，提升动态范围。

1 噪声与满阱的制约关系分析

红外图像传感器的积分光电流 i_{ph} 和信噪比(signal-to-noise ratio, S/N)R_{SN} 的关系为[22]

$$R_{SN}(i_{ph})=\frac{(i_{ph}t_{int})^2}{qi_{ph}t_{int}+q^2\sigma^2_{Readout}} \qquad i_{ph}\leqslant\frac{qQ_{FWC}}{t_{int}} \tag{1}$$

式中：i_{ph} 为光电流；t_{int} 为积分时间，在研究过程中固定设置为47 μs，在具体工作时可进行调整；q 为电子电荷；Q_{FWC} 为满阱电荷及阱电容；$\sigma^2_{Readout}$ 为影响图像传感器的所有噪声，其类型包含与工艺偏差相关的固定模式噪声、读出过程中的复位噪声、与温度和沟道长度相关的1/f噪声和散粒噪声。

高动态红外图像传感器的噪声和满阱容量存在制约关系。图像传感器的噪声电子数 Q_{noise}(见式(2))，包括像素内部产生的所有噪声，占主要部分的是读出过程的复位噪声和散粒噪声。由于将可编程增益放大器(programmable gain amplifier, PGA)设置为增益可调的，在暗光条件下最大可进行4倍放大[23]。此时，后级读出电路的噪声折合到像素时，噪声电子数将小于一个电子，读出电路产生的噪声可忽略。

$$Q_{noise}=\frac{C_{int}V_{noise}}{q} \tag{2}$$

式中：C_{int} 为积分电容；V_{noise} 为噪声电压。

红外图像传感器的满阱容量可表示为[24]

$$Q_{FWC}=\frac{1}{q}\int C_{int}V\mathrm{d}V \tag{3}$$

式中：V 为红外图像传感器的摆幅。

红外图像传感器可进行光电转换的动态范围表示为[25]

$$D_R=20\lg\left(\frac{Q_{FWC}}{Q_{noise}}\right) \tag{4}$$

从式(2)和式(3)可见，大的积分电容 C_{int} 将给图像传感器带来大的噪声和满阱容量，对动态范围的提高形成了制约。如式(4)所示，提升红外图像传感器的动态范围需要同时提高满阱容量和减小噪声，而满阱容量的提升需要增大积分电容，噪声的下降需要减小积分电容。因此，电容值在强光下变大、弱光下变小的自适应积分电容将满足高动态红外图像传感器的需要。

2 自适应积分电容设计

红外图像传感器中承载光电流的积分电容是设计的关键，积分电容上的电压 V_{int} 会根据光电流 i_{ph} 的大小与积分时间 t_{int} 呈现一定的变化关系，其中，复位电压表示为 V_{RST} 表示为

$$V_{int}=V_{RST}-i_{ph}t_{int}C_{int} \tag{5}$$

提出自适应积分电容的思想正是基于这一特点，而MOS电容天然具有栅源电压 V_{gs} 与电容的对应关系。但是一般而言，MOS电容变化区间对应的 V_{gs} 变化区间较小。因此，为了加宽MOS电容随 V_{gs} 变化的区间，积分电容采用3种不同阈值电压的MOS电容并联，来减小电容随电压变化曲线的斜率，提升高动态红外图像传感器的线性度。基于商用CMOS工艺，选取了3种阈值电压 V_{th} 的MOS器件，其特性如表1所示。

表1 C_1、C_2 和 C_3 的阈值电压

Table 1 Threshold voltage of C_1, C_2, and C_3

MOS管类型	V_{th}/V
C_1	0.62
C_2	0.40
C_3	0.30

为了研究自适应电容作积分电容提升像素动态范围最佳方案，对宽和长均为3 μm、MOS管类型为 C_1、C_2 和 C_3 的MOS电容，在25 ℃和tt工艺角的条件下进行仿真，得到MOS电容值随 V_{gs} 变化的曲线，如图1所示。图1展示了 C_1、C_2、C_3 电容在强反型区和积累区具有类似的电容值，但阈值电压

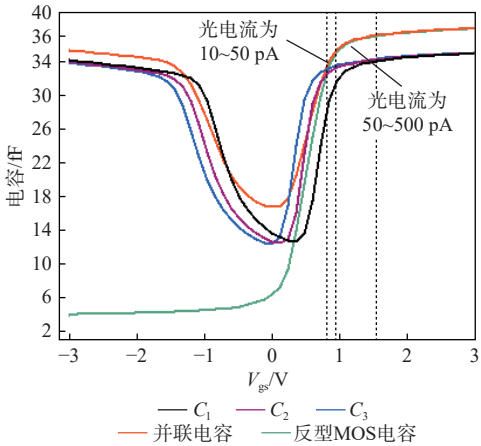


图 1　MOS 电容的容值随 V_{gs} 的变化而变化曲线

Fig. 1　The capacitance value of the MOS capacitor varies with changes in the V_{gs}

不同; 且反型区的电容值随 V_{gs} 变化的曲线均具有陡峭的斜率, 作红外像素的积分电容使红外图像传感器输出 CDS 信号的线性度较差。因此不能直接用作红外图像传感器的积分电容。

为了增大 MOS 电容的反型区电压范围, 得到一个电容值随 V_{gs} 变化更平缓的曲线。将面积均为 1 μm×3 μm 的 C_1、C_2、C_3 电容并联, 得到总体面积还为 9 μm^2 的积分电容。其电容与电压曲线如图 1 中橙色曲线所示, 并联电容不仅具有 0～1.5 V 的 V_{gs} 可用范围, 且其电容值随 V_{gs} 变化更平缓, 比单独采用一种类型的 MOS 电容具有更大的电容值。这意味着采用 C_1、C_2、C_3 并联作自适应积分电容, 可使红外像素在不同工作电压下均获得更大的满阱电荷。

根据反型 MOS 电容的设计原理, 将并联电容的衬底接地, 其电容电压曲线如图 1 中绿色曲线所示。其在 0 V 的 V_{gs} 下仅有 6.5 fF 的电容, 而在 1 V 的 V_{gs} 下具有 37.5 fF 的电容。

为了研究反型 MOS 电容作高动态红外图像传感器积分电容的实际效果, 基于国内 55 nm 1P4M CIS 工艺平台, 搭建基于自适应积分电容的 12 288×12 288 面阵规模红外图像传感器结构。

3　自适应积分电容红外图像传感器结构

自适应积分电容红外图像传感器结构如图 2 所示, 虚线框为自适应积分电容像素结构。

自适应积分电容红外图像传感器主要包括 12 288×12 288 面阵规模像素阵列、行驱动和读出电路 3 个部分。行驱动电路用于传输全局复位 RST 信号、电荷转移 TX 信号, 行复位 RST1 信号、行选 ROW 信号。像素单元为五管 (5T) 像素[26], 包含 RST

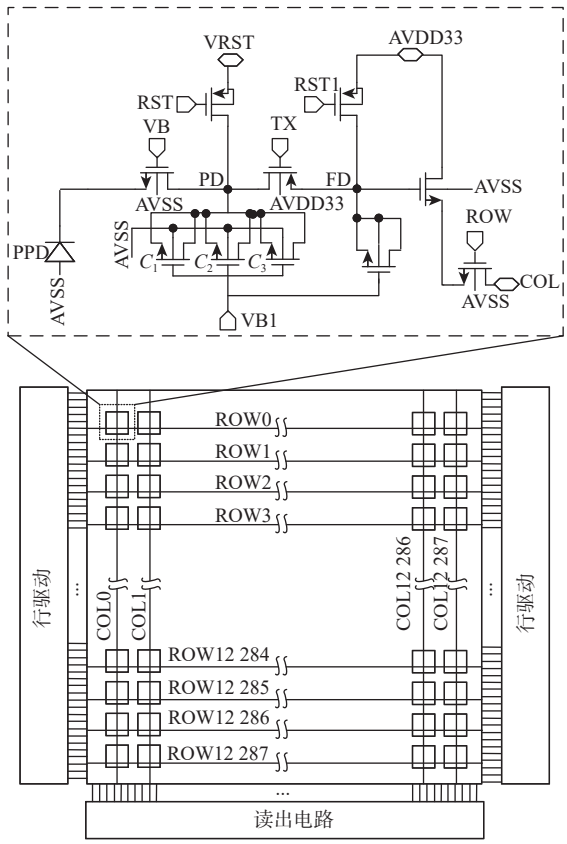


图 2　基于自适应积分电容的 CMOS 红外图像传感器的基本架构

Fig. 2　Basic architecture of a CMOS infrared image sensor based on adaptive integral capacitance

全局复位管、TX 电荷转移管、RST1 行复位管、ROW 行选控制管和源随器。除此之外, 还包含光电二极管 PPD、电压偏置管 VB 和自适应积分电容。定义积分电容节点为 PD 点, 转移电容节点也是浮动扩散节点 (floating diffusion, FD) 点。为了实现对像素单元的单独测试, 将复位电源和源跟随器电源分开处理。

像素采用反型 MOS 电容作自适应积分电容。由式(5)与式(6)得到式(7), 推导光电流与积分电容容值的关系(见式(8))。

$$V_{gs} = V_{VB1} - V_{PD} \tag{6}$$

$$V_{gs} = V_{VB1} - V_{RST} + i_{ph} t_{int} C_{int} \tag{7}$$

式中: V_{VB1} 为图 2 中 VB1 点的偏置电压; V_{RST} 为图 2 中 RST 点的复位电压, 均设置为 3.3 V。

则 V_{gs} 与 C_{int} 的关系为

$$V_{gs} = i_{ph} t_{int} C_{int} \tag{8}$$

如图 2 所示, 在像素积分之前, RST 打开, V_{RST} 将 MOS 电容的源漏端复位至 3.3 V。此时, V_{gs} 等于 0 V, 随着积分的进行, MOS 电容的源漏端电压逐渐下降, 对应图 1 横坐标电容 V_{gs} 的变化, 随着积

分电荷的增多，V_{gs}从0 V逐渐向x轴的正半轴增加。而V_{gs}的最大值为

$$V_{gs,max}=V_{VB1}-V_{PD,min}=V_{VB1}-(V_{ta}+V_{th})=(3.3-(0.2+0.7))V=2.4\ V \quad (9)$$

式中：V_{ta}为尾电流的电压；V_{th}为阈值电压，即可被源随器输出的最小电压。

式(8)中，在强光下，V_{gs}最大可增至2.4 V，而在理想的暗光下，V_{gs}最小可减至0 V。对应图1中的曲线，强光下的电容值是暗光下的5.4倍，使自适应积分电容红外像素在强光下具有更大的满阱容量，暗光下具有更低的噪声，红外图像传感器将实现更大的动态范围。

图3展示了前4行像素的行驱动时序。首先，在复位阶段，RST作为全局复位信号，先有效，复位积分电容节点PD；然后，RST10～RST13晶体管同时导通复位转移电容节点FD；其次，信号TX用于转移光生电子；最后，在读出阶段，t_1～t_4期间ROW0～ROW3相继保持高电平，表示像素信号逐行读出，像素的列总线电压下降，显示信号电压，RST10～RST13逐行有效，像素的列总线电压升高，显示复位电压，后级读出电路采用CDS技术对信号电压和复位电压进行采样。

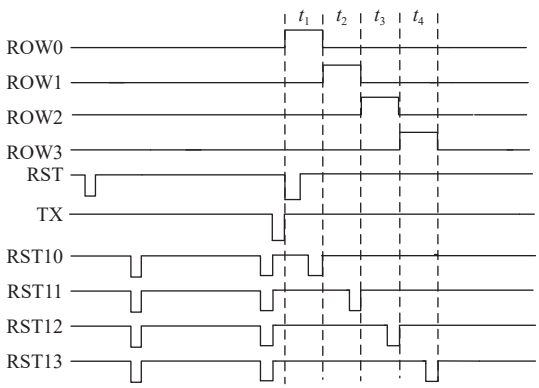


图3 像素的时序图

Fig. 3 Timing diagram of pixel

4 版图设计和仿真数据分析

为了研究基于自适应积分电容的高动态红外图像传感器的成像效果，基于国内55 nm 1P4M CIS工艺平台，进行红外图像传感器版图设计和仿真数据分析。

像素的版图设计如图4所示。为了紧凑布局，版图设计上采用2×2的像素单元作为像素面阵的最小重复单元，进一步共用N阱。像素中心距为5.5 μm。图4展示了积分电容采用不同类型的NMOS电容分布在版图的上下侧，中间设计PMOS复位管共用N阱，实现较大的空间利用率和像素电压摆幅。图中，白色正方形为与光二极管连接的孔，尺寸为3 μm×3 μm，用于倒焊互连铟柱。

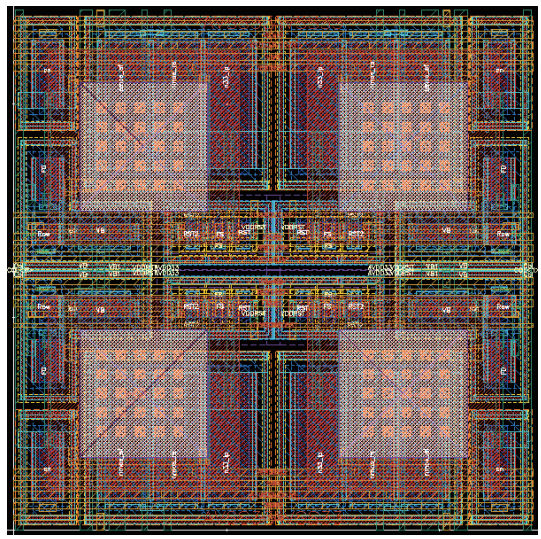


图4 2×2像素版图

Fig. 4 2 × 2 pixel layout

基于自适应积分电容的高动态红外图像传感器整体电路与版图设计如图5所示，展示了将光电流转换为电压信号的自适应积分电容像素阵列和用于信号处理的列级读出电路。图5中，列级读出电路包括：PGA根据应用场景实现电压信号的1、2、4倍放大，列缓冲实现信号的并转串输出，CDS消除低频噪声。同时，研究了自适应积分电容对红外探测器的噪声、FWC以及动态范围的影响。

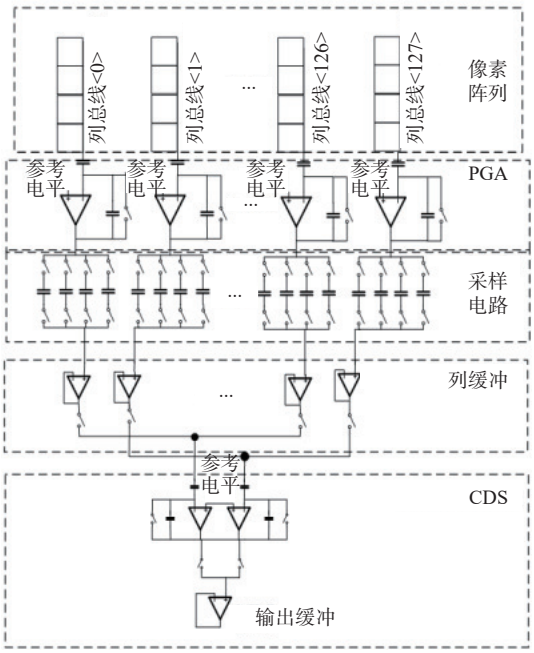
为研究自适应积分电容5T红外像素的输出信号质量，在强光和弱光的条件下分别分析红外像素的列线信号如图6所示。在25 ℃和tt工艺角的条件下进行仿真，在光电流为10～50 pA的弱光下，仿真出红外像素的列线信号如图6(a)所示，在光电流为50～500 pA的强光条件下，仿真出红外像素的列线信号如图6(b)所示。可见，随着光强的增大，自适应红外像素列线电压信号下降，CDS信号增大，像素功能正常。

线性度是红外图像传感器的重要性能指标。转换增益是CDS信号与光生电子的比值，用于衡量图像传感器的灵敏度[27]。转换增益G_C计算为

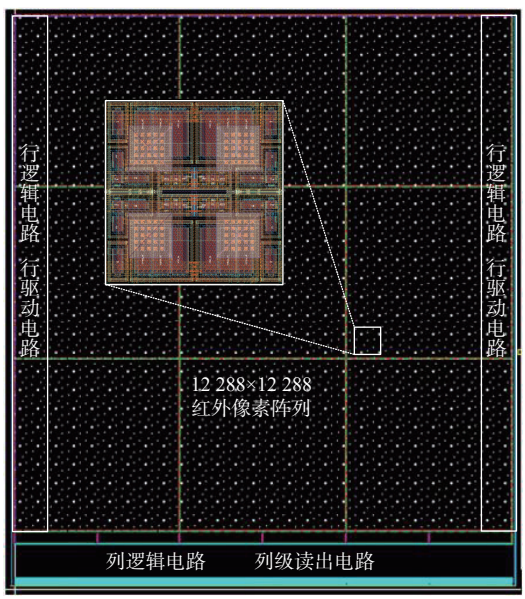
$$G_C=\frac{1\times q}{C_{int}} \quad (10)$$

因此，通过分析采用自适应积分电容的5T红外像素CDS信号和转换增益随光变化的曲线，研究自适应积分电容对图像传感器线性度和灵敏度的影响。

通过图7(a)所示的红外像素CDS信号和输入光电流的曲线，得出红外像素的列线CDS信号在



(a) 电路设计



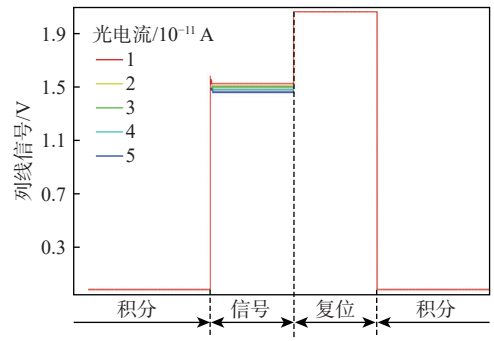
(b) 版图设计

图 5　自适应积分电容的红外图像传感器整体电路设计和版图设计

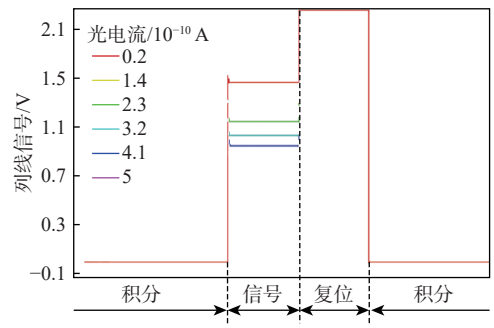
Fig. 5　Overall circuit design and layout design of infrared image sensor with adaptive integral capacitance

光电流为 10～50 pA 的弱光和光电流为 50～500 pA 的强光下, 均具有良好的线性, 其线性度分别为 96.7% 和 86.6%。

在弱光和强光下, 自适应积分电容像素的转换增益如图 7(b)所示, 在光电流为 10～50 pA 的弱光下, 转换增益从 183.1 μV/e⁻下降到 40.8 μV/e⁻。在光电流为 50～500 pA 的强光下, 转换增益从 40.8 μV/e⁻下降到 8.1 μV/e⁻。可见, 在不同光照环境下, 自适应积分电容使像素具有可调节的转换增益, 且自适应积分电容红外图像传感器具有对弱光感应更灵敏的优良性能。



(a) 10~50 pA光电流下的列线信号



(b) 50~500 pA光电流下的列线信号

图 6　采用自适应电容的 5T 红外像素在 10～50 pA 和 50～500 pA 光电流下的列线信号

Fig. 6　Line signal of a 5T infrared pixel using adaptive capacitance under photocurrent of 10−50 pA and 50−500 pA

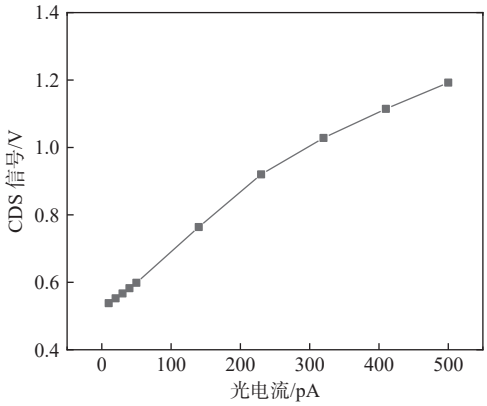
图像传感器的线性度是通过图 7(a)中 CDS 信号与光电流曲线的线性度来计算的。计算光电流导致电容上电荷发生变化, 最终转换为 CDS 电压的公式为

$$Q_{\text{in}} = C_{\text{int}} V_{\text{CDS}} = i_{\text{ph}} t_{\text{int}} \tag{11}$$

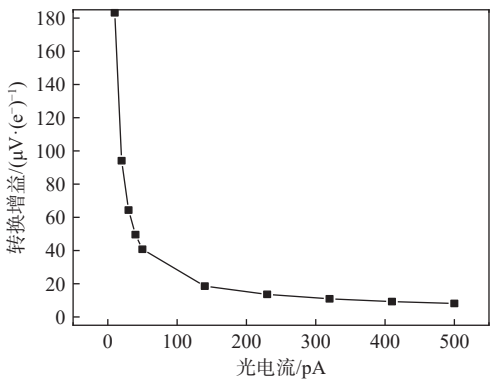
式中: Q_{in} 为积分电荷; V_{CDS} 为 CDS 信号, V_{CDS} 随 i_{ph} 变化曲线的斜率为 $t_{\text{int}}/C_{\text{int}}$。本文中, 积分时间为定值 47 μs, 则图像传感器的线性度只与 $1/C_{\text{int}}$ 的变化有关, $1/C_{\text{int}}$ 只要线性变化就可以保证图像传感器的线性度。

图 1 表示了光电流为 10～50 pA 和 50～500 pA 的积分电容变化。根据图 1, 光电流为 10～50 pA 的区域内, 积分电容的变化线性度好。由于像素固有的衬偏效应, 当光电流在 10～500 pA 范围内, V_{th} 从 1.1 V 向 1 V 变化。光电流为 50～500 pA 的区域内, 由于阈值电压的存在, 使积分电容变化存在一个拐点, $1/C_{\text{int}}$ 的变化具有非线性, 从而影响了图像传感器的线性度。

图 7(b)也同样可以说明这个问题。虽然光电流为 10～50 pA 的区域内转换增益的变化率高, 但是变化是线性的。而光电流为 50～500 pA 的区域内, 转换增益的变化率低, 但在 140 pA 处存在一个



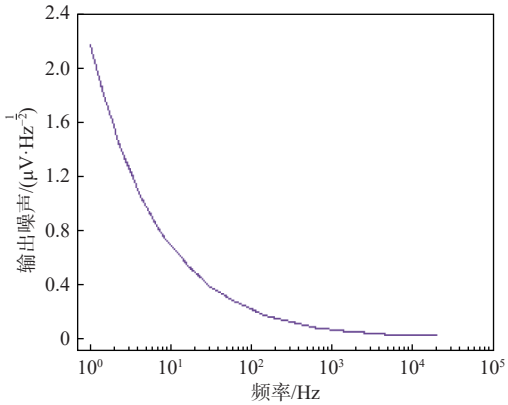
(a) CDS信号随光变化的曲线



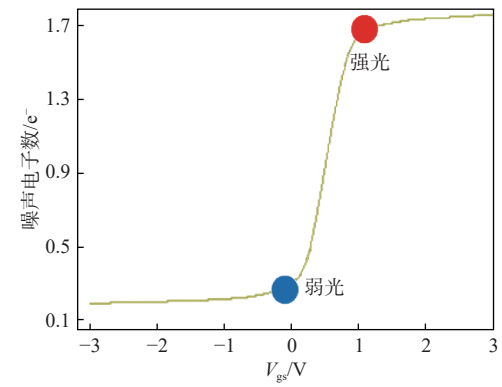
(b) 转换增益随光变化的曲线

图7 采用自适应电容的5T红外像素的CDS信号和转换增益随光电流变化的曲线

Fig. 7 Variation of CDS signal of 5T infrared pixel using adaptive capacitance and conversion gain with photocurrent



(a) 噪声曲线



(b) 噪声电子数

图8 采用自适应电容的5T红外像素的噪声和噪声电子数曲线

Fig. 8 Noise and noise electronic number of 5T infrared pixel using adaptive capacitance

拐点，使转换增益的变化产生了非线性，影响了图像传感器的线性度。该线性度问题与传统的像素结构一致，且广泛存在。由于篇幅问题，且本文的重点是解决高动态问题，该非线性问题并未展开分析，可采用文献[28]中的解决方案予以消除。

通过图8自适应积分电容像素的噪声分析，来研究自适应积分电容对红外图像传感器噪声性能的影响。像素的噪声频域曲线如图8(a)所示。自适应红外像素噪声主要为低频信号，以闪烁噪声为主，其噪声电压对频率的积分为7.45 μV。根据式(2)，由于采用自适应积分电容，不同栅源电压下的噪声电子数随着电容值的变化而变化，曲线如图8(b)所示。根据式(8)，图8(b)中蓝色表示电容V_{gs}为0 V的弱光条件下，噪声仅为0.3 e^-，红点表示电容V_{gs}为1 V的强光条件下，噪声为1.8 e^-。可见，噪声电子数随着光强的增加而增大，从而实现了在弱光下噪声较小、强光下噪声较大的优良性能。

为研究自适应积分电容红外图像传感器在强光和弱光下的电容变化情况，将电容值在6.5～37.5 fF变化的自适应积分电容红外像素与6.5 fF、37.5 fF的定值积分电容像素进行比较，分析其CDS信号随光电流变化的曲线，如图9所示。图中，黑色曲线展示了在5T红外像素中，使用电容值在6.5～37.5 fF变化的自适应电容作积分电容，其CDS信号随光电流变化的曲线。可见，在0～230 pA的弱光条件下，自适应电容的CDS信号随光电流变化

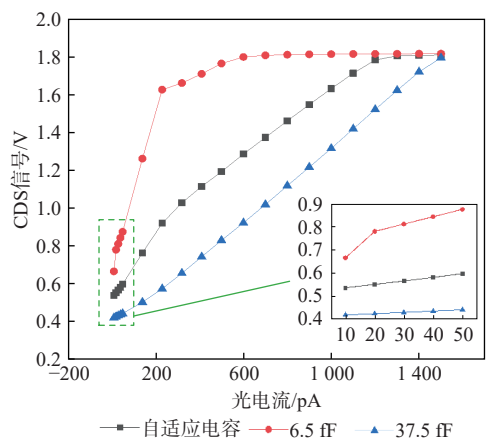


图9 采用自适应电容和6.5 fF、37.5 fF固定电容作5T红外像素的积分电容时CDS信号随光电流变化的曲线

Fig. 9 Variation of CDS signal using adaptive capacitance and fixed capacitance of 6.5 fF and 37.5 fF as integral capacitance of 5T infrared pixel with photocurrent

的曲线斜率更贴近 6.5 fF 的曲线; 而在 230～1 500 pA 的强光条件下, 自适应电容的 CDS 信号随光电流变化的曲线斜率更贴近 37.5 fF 的曲线。说明自适应积分电容在弱光下接近 6.5 fF, 强光下接近 37.5 fF。

为研究自适应积分电容红外图像传感器降低噪声和提升满阱的效果, 将电容值在 6.5～37.5 fF 变化的自适应积分电容红外像素与 6.5 fF、37.5 fF 的定值电容作积分电容的红外像素进行比较, 根据式(2)、式(3)计算出不同电容作积分电容时, 5T 红外像素的噪声和满阱容量如图 10 所示。图 10 展示了在 5T 红外像素中, 自适应电容作积分电容具有与 6.5 fF 小电容相比拟的更小的噪声电子数(0.3 e^-～1.8 e^-)和与 37.5 fF 大电容相比拟的更大的满阱容量(1.1 Me^-)。说明采用自适应电容作积分电容的 5T 红外像素具有更大的动态范围, 根据式(4)计算其值为 130 dB。根据式(4)可知, 动态范围由 Q_{FWC} 和 Q_{noise}2 个因素决定, 即满阱容量和噪声电子数。如图 8(b)所示, 采用可变积分电容的图像传感器, 其噪声电子为变量, 范围为 0.3 e^-～1.8 e^-, 根据式(4)计算出的动态范围为 130～116 dB。

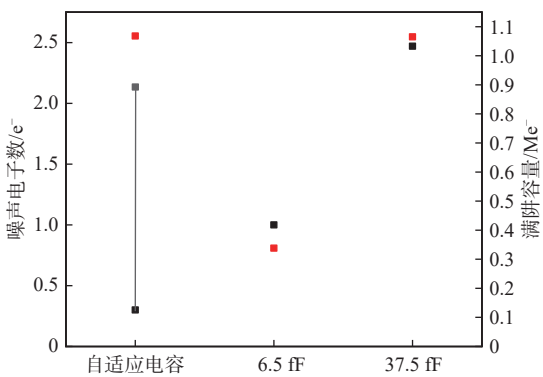


图 10　采用自适应电容和 6.5 fF、37.5 fF 固定电容作 5T 红外像素的积分电容的满阱容量和噪声

Fig. 10　Full well capacity and noise of using adaptive capacitance and fixed capacitance of 6.5 fF and 37.5 fF as integral capacitance of 5T infrared pixel

为了进一步研究自适应积分电容红外图像传感器的实际性能, 基于 55 nm CMOS 工艺平台进行后仿真, 其工艺仿真文件经过笔者课题组流片和修正, 与实际测试结果保持一致[29-31]。在 25 ℃ 和 tt 工艺角下, 得出符合实际的性能参数如表 2 所示。其各项参数在光电流为 10～50 pA 的弱光和光电流为 50～500 pA 的强光条件下均较优于前仿真, 并实现了前后仿真的一致。

由于采用工艺仿真文件经过实际流片的修正, 表 2 中的性能参数具有与其他高动态图像传感器研究直接比较的合理性, 如表 3 所示。自适应积分电容像素中心距仅为 5.5 μm, 远小于文献 [11] 的 16 μm、文献 [16] 的 22 μm、文献 [21] 的 15 μm, 极具优势。相比于文献 [12] 的增加减法器、文献 [16] 的增加计数器、文献 [21] 的优化光电二极管结构的高动态方法, 自适应积分电容设计简单, 具有更大的电容可用面积和 1.31 Me^-的满阱容量。文献 [18] 的动态范围虽然大, 但由于采用对数像素, 线性度很差。通过比较表 3 中噪声和动态范围, 自适应积分电容的高动态图像传感器在可比拟的 130 dB 的动态范围条件下具有最小的噪声电子数, 仅为 0.43 e^-。

表 2　5T 红外像素电路的后仿真性能参数

Table 2　Post-simulation performance parameters of 5T infrared pixel circuit

光电流/pA	线性度/%	转换增益/(μV·$(e^-)^{-1}$)	噪声电子数/e^-	动态范围/dB	满阱容量/Me^-
10～50	97.5	110～24.8	0.43～2.2	130～116	1.31
50～500	96.6	24.8～6.5			

表 3　图像传感器的性能对比

Table 3　Performance comparison of image sensors

方法	工艺条件/nm	像素尺寸/(μm×μm)	噪声电子数/e^-	满阱容量/Me^-	采用方法	动态范围/dB
[11]	180	16×16	7.6	24.3	横向溢流集成沟槽电容器	130
[12]		2.9×2.9	0.78	0.06	新的光电二极管结构	97
[16]	65	22×22	61	0.3	双曝光电荷相减法	74
[18]	180	10.5×8.5			线性-对数像素	152
[21]		15×15	6	0.6	自适应调整积分时间	100
本文	55	5.5×5.5	0.43～2.2	1.31	自适应积分电容	130～116

5　结　论

1) 在红外图像传感器向小尺寸、大满阱、低噪声发展的过程中, 高动态成像具有现实意义。而本文提出的自适应积分电容可广泛应用于各种类型的像素结构中, 实现高动态范围成像。

2) 基于国内 55 nm CMOS 工艺平台, 在 12 288×12 288 超大面阵规模红外图像传感器芯片中进行自适应积分电容提升动态范围的研究。

3) 研究结果表明, 在光电流为 10～50 pA 的弱光信号下, 实现积分电容接近 6.5 fF, 增益高达 39.2 μV/e^-, 噪声电子数低至 0.43 e^-; 当光电流为 50～500 pA 的强光信号下, 电容接近 37.5 fF; 增益减小到 2.3 μV/e^-, 满阱容量提升至 1.31 Me^-, 动态范围可达 130 dB。

本文方法解决了像素设计中噪声和满阱两个关键参数的共赢问题, 为高性能大面阵 CMOS 图像

传感器的设计提供了一定的理论指导。

参考文献（References）

[1] SEO M W, CHU M, JUNG H Y, et al. A 2.6 e-rms low-random-noise, 116.2 mW low-power 2-mp global shutter CMOS image sensor with pixel-level ADC and in-pixel memory[C]//Proceedings of the 2021 Symposium on VLSI Circuits. Piscataway: IEEE Press, 2021: 1-2.

[2] KIM M S, KANG D U, LEE D H, et al. Fast signal transfer in a large-area X-ray CMOS image sensor[J]. Journal of Instrumentation, 2014, 9(8): P08011.

[3] XU J T, LI W, NIE K M, et al. A method to reduce the effect on image quality caused by resistance of column bus[J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2018, 27(1): 173-181.

[4] KIM H J, HWANG S I, KWON J W, et al. A delta-readout scheme for low-power CMOS image sensors with multi-column-parallel SAR ADCs[J]. IEEE Journal of Solid-State Circuits, 2016, 51(10): 2262-2273.

[5] PARK H, YU C Z, KIM H, et al. Low power CMOS image sensors using two step single slope ADC with bandwidth-limited comparators & voltage range extended ramp generator for battery-limited application[J]. IEEE Sensors Journal, 2020, 20(6): 2831-2838.

[6] NIE K M, ZHA W B, SHI X L, et al. A single slope ADC with row-wise noise reduction technique for CMOS image sensor[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2020, 67(9): 2873-2882.

[7] PARK I, JO W, PARK C, et al. A 640 × 640 fully dynamic CMOS image sensor for always-on operation[J]. IEEE Journal of Solid-State Circuits, 2020, 55(4): 898-907.

[8] CHEUNG V, WESTLAND S, THOMSON M. Accurate estimation of the nonlinearity of input/output response for color cameras[J]. Color Research & Application, 2004, 29(6): 406-412.

[9] ZAITSU K, MATSUMOTO A, NISHIDA M, et al. A 2-layer transistor pixel stacked CMOS image sensor with oxide-based full trench isolation for large full well capacity and high quantum efficiency[C]//Proceedings of the 2022 IEEE Symposium on VLSI Technology and Circuits. Piscataway: IEEE Press, 2022: 286-287.

[10] KIM H, KIM Y H, MOON S, et al. A 0.64 μm 4-photodiode 1.28 μm 50 Mpixel CMOS image sensor with 0.98e- temporal noise and 20Ke- full-well capacity employing quarter-ring source-follower[C]//Proceedings of the 2023 IEEE International Solid-State Circuits Conference. Piscataway: IEEE Press, 2023: 1-3.

[11] MURATA M, KURODA R, FUJIHARA Y, et al. A high near-infrared sensitivity over 70-dB SNR CMOS image sensor with lateral overflow integration trench capacitor[J]. IEEE Transactions on Electron Devices, 2020, 67(4): 1653-1659.

[12] UCHIDA T, YAMASHITA K, MASAGAKI A, et al. A 2.9 μm pixel CMOS image sensor for security cameras with high FWC and 97 dB single-exposure dynamic range[C]//Proceedings of the 2021 IEEE International Electron Devices Meeting. Piscataway: IEEE Press, 2021: 30.3.1-30.3.4.

[13] GOTO M, HONDA Y, NANBA M, et al. Pixel-parallel three-layer stacked CMOS image sensors using double-sided hybrid bonding of SOI wafers[J]. IEEE Transactions on Electron Devices, 2023, 70(9): 4705-4711.

[14] SUN K Y, PAN L Y, WU D, et al. A gate-controlled lateral thyristor-based pixel with low-light-level extension for high dynamic range applications[J]. IEEE Sensors Journal, 2022, 22(24): 23856-23865.

[15] ZHUO Y, LU W G, YU S Z, et al. A low-noise CTIA-based pixel with CDS for SWIR focal plane arrays[C]//Proceedings of the 2021 6th International Conference on Integrated Circuits and Microsystems. Piscataway: IEEE Press, 2021: 258-262.

[16] QIAN X Y, YU H, CHEN S S, et al. A high dynamic range CMOS image sensor with dual-exposure charge subtraction scheme[J]. IEEE Sensors Journal, 2015, 15(2): 661-662.

[17] WOCIAL T, STEFANOV K D, MARTIN W E, et al. A method to achieve high dynamic range in a CMOS image sensor using interleaved row readout[J]. IEEE Sensors Journal, 2022, 22(22): 21619-21627.

[18] KUO K C. A novel linear-logarithmic active pixel CMOS image sensor with wide dynamic range[C]//Proceedings of the 2021 IEEE International Midwest Symposium on Circuits and Systems. Piscataway: IEEE Press, 2021: 1100-1103.

[19] WANG F, HAN L Q, THEUWISSEN A J P. Development and evaluation of a highly linear CMOS image sensor with a digitally assisted linearity calibration[J]. IEEE Journal of Solid-State Circuits, 2018, 53(10): 2970-2981.

[20] ZHANG Q, GUO Z J. An 8T global shutter pixel with extended output range for CMOS image sensor[C]//Proceedings of the 2019 IEEE International Conference on Electron Devices and Solid-State Circuits. Piscataway: IEEE Press, 2019: 1-3.

[21] KUZNETSOV P A, MOSCHEV I S. Photodetector of shortwave infrared range of format 640 × 512 elements with increased dynamic range[J]. Journal of Communications Technology and Electronics, 2019, 64(9): 1034-1037.

[22] KAVUSI S, GHOSH K, EL GAMAL A. Architectures for high dynamic range, high speed image sensor readout circuits[M]//DE MICNELI G, MIR S, REIS R. VLSI-SoC: research trends in VLSI and systems on chip. Berlin: Springer, 2008: 1-23.

[23] GUO Y M, GUO Z J, WANG Y L, et al. Research on dark current compensation method of analog front end based on synchronous adaptive matching[C]//Proceedings of the 2023 6th International Conference on Electronics Technology. Piscataway: IEEE Press, 2023: 31-35.

[24] WANG F, THEUWISSEN A J P. Pixel optimizations and digital calibration methods of a CMOS image sensor targeting high linearity[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2019, 66(3): 930-940.

[25] LOU S S, QU Y, ZHONG G Q, et al. An over 140 dB dynamic range CMOS image sensor combined DCG and logarithmic response[J]. IEEE Transactions on Electron Devices, 2023, 70(9): 4719-4724.

[26] GAO Z Y, YANG C J, XU J T, et al. A dynamic range enhanced readout technique with a two-step TDC for high speed linear CMOS image sensors[J]. Sensors, 2015, 15(11): 28224-28243.

[27] 杜雅娟, 高静, 高志远, 等. 基于自对准技术的4管像素寄生电容建模与分析[J]. 激光与光电子学进展, 2023, 60(19): 210-216.

DU Y J, GAO J, GAO Z Y, et al. Modeling and analysis of parasitic capacitance in 4-transistor pixels based on self-alignment technique[J]. Laser & Optoelectronics Progress, 2023, 60(19): 210-216(in Chinese).

[28] GUO Z, YU N, WU L. An improved global shutter pixel with extended output range and linearity of compensation for CMOS image sensor[J]. Chinese Journal of Electronics, 2021, 30(1): 102-108.

[29] GUO Z J, YU N M, WU L S. A synchronous driving approach based on adaptive delay phase-locked loop for stitching CMOS image sensor[J]. IEICE Electronics Express, 2020, 17(3): 20190642.

[30] GUO Z J, YU N M, WU L S. Research on column FPN and black level calibration in large array CMOS image sensor[J]. Chinese Journal of Electronics, 2021, 30(2): 268-274.

[31] GUO Z J, YU N M, WU L S. An improved segmented DAC for column readout circuit correction of large array CMOS image sensor[J]. IEICE Electronics Express, 2020, 17(10): 20200094.

Structure of highly dynamic pixel based on adaptive integral capacitance

LIU Suiyang，GUO Zhongjie*，YU Ningmei，XU Ruiming

(School of Automation and Information Engineering，Xi'an University of Technology，Xi'an 710048，China)

Abstract: Infrared image sensing technology has received widespread attention due to its advantages of not being affected by the environment, good target recognition, and strong anti-interference ability. However, with the improvement of the integration of the infrared focal plane, the constraints among the dynamic range, noise, and full well capacity of the photoelectric system are particularly prominent. Therefore, in order to solve the contradiction between noise in low light and full well capacity in strong light, in the 5T infrared pixel circuit, the relationship between the capacitance value and voltage of the inverse MOS capacitor in a specific voltage interval was used to automatically change the integral capacitance of the infrared image sensor from 6.5 fF to 37.5 fF, and a highly dynamic pixel structure based on adaptive integral capacitance was proposed. Based on 55 nm CMOS process technology, the performance parameters of an infrared sensor with a 12 288 × 12 288 pixel scale were studied. The research results show that a small-size pixel of 5.5 μm × 5.5 μm has a large full well capacity of 1.31 Me^-, and a variable conversion gain. The noise is less than 0.43 e^-, and the dynamic range is more than 130 dB.

Keywords: CMOS infrared image sensor; adaptive integral capacitance; highly dynamic range; large full well capacity; pixel readout circuit

Received：2023-06-12；**Accepted**：2023-08-24；**Published Online**：2023-09-08 16：27

URL：link.cnki.net/urlid/11.2625.V.20230908.0902.001

Foundation items：National Natural Science Foundation of China (62171367); Key Research and Development Program of Shaanxi (2021GY-060); Innovation Capability Support Program of Shaanxi (2022TD-39); Xi'an University of Technology School Enterprise Collaborative Foundation (252062213); Doctoral Innovation Fund of Xi'an University of Technology (103-252072401)

* **Corresponding author**. E-mail：zjguo@xaut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0375

针对高机动目标的深度强化学习智能拦截制导

张豪[1]，朱建文[1,2,*]，李小平[1]，包为民[1,3]

(1. 西安电子科技大学 空间科学与技术学院，西安 710126；2. 火箭军工程大学 导弹工程学院，西安 710025；
3. 中国航天科技集团有限公司，北京 100048)

摘 要：针对固定系数比例导引律在拦截高机动目标时脱靶量与能量损耗过大的问题，提出一种基于深度强化学习智能调参的自适应比例导引律。建立基于实时飞行状态的状态空间和包含横向、纵向导航比的动作空间，以及综合不同状态的奖励函数模型；在奖励函数的模型设计中,引入预测-校正方法，以提高动作评估的准确性；利用柔性演员-评论 (SAC) 算法，根据拦截器和目标的相对运动状态，训练获得综合考虑脱靶量和能量损耗的网络参数与制导参数决策系统。仿真结果表明：相较于传统比例导引律，所提制导策略可在保证低脱靶量的同时显著降低能量损耗，并且对未训练场景具有良好的适应性。

关 键 词：智能制导；机动目标；深度强化学习；比例导引；马尔可夫决策过程
中图分类号：V448; TJ765.3
文献标志码：A **文章编号**：1001-5965(2025)06-2060-10

高速机动目标的拦截制导问题一直是研究的热点和难点。比例导引律 (proportional navigation guidance, PNG) 以形式简单和易于实现的优点被广泛应用,其主要思想是通过零化视线角速率迫使拦截器和目标保持在碰撞三角形上。然而,随着飞行任务复杂化,制导性能指标已不仅是零脱靶量,还包括其他约束,如能量最优[1]。对于无机动目标,PNG 展现出优异的拦截性能,并在理论上已经证明导航比为 3 时能量最优[2]。而面对高速机动目标时,为了追求低脱靶量,常采用大导航比,导致拦截器对目标机动过于敏感,造成不必要的能量损耗。基于现代控制理论的制导律,如滑模制导[3]、最优制导[4]、微分对策制导[5] 和动态规划[6] 等,同样已用于解决该问题,但这些制导律往往需要剩余飞行时间或目标加速度等额外信息,尤其当考虑能量损耗时,制导指令会更加复杂。因此,研究具有能量最优的 PNG 是非常必要的。

作为比较成熟的制导律,已有许多文献对 PNG 及其变体进行了研究。文献 [7] 针对大气层外的机动目标,分析了现实真比例导引律对任意机动目标的捕获区域,并给出了拦截器不超过自身饱和过载下的比例系数范围。文献 [8] 提出了一种广义相对偏置比例制导律,以解决常系数比例法捕获域较小及碰撞角约束问题。然而,文献 [7-8] 均研究平面拦截问题,在实际应用时,需额外设计制导方案以保证两者共面,这将导致原制导律性能下降。因此,对三维 PNG 进行研究是非常重要的。文献 [9] 通过在旋转视线坐标系中建立相对运动方程,利用非线性输出调节方法研究了纯比例导引 (pure proportional navigation, PPNG) 在三维空间中的捕获能力。文献 [10] 讨论了三维 PPNG 中加速度方向对脱靶量的影响,并将其修正到最小化零控脱靶量的加速度方向,进一步增强 PPNG 的拦截性能。由于目标机动使拦截器和目标间相对运动变得异常复杂,加入其他约束将很难进行理论分析和求解。因此,对三维 PNG 的研究大多集中在最小化脱靶

收稿日期：2023-06-15；录用日期：2023-09-11；网络出版时间：2023-09-27 18:44
网络出版地址：link.cnki.net/urlid/11.2625.V.20230927.1317.001
*通信作者. E-mail：zhujianwen1117@163.com

引用格式：张豪，朱建文，李小平，等. 针对高机动目标的深度强化学习智能拦截制导 [J]. 北京航空航天大学学报，2025，51（6）：2060-2069. ZHANG H，ZHU J W，LI X P，et al. Deep reinforcement learning intelligent guidance for intercepting high maneuvering targets [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（6）：2060-2069（in Chinese）.

量上, 很少涉及能量损耗。

得益于人工智能技术的高速发展, 强化学习 (reinforcement learning, RL) 技术作为一种体现智能决策的算法, 得到了众多学者的认可。文献 [11-12] 将状态空间和动作空间离散化, 为实现零脱靶量提出了一种结合 Q-learning 和 PNG 的制导律, 但该制导律仅适用于低维空间, 且制导性能与离散程度相关。深度强化学习 (deep reinforcement learning, DRL) 结合了深度学习(deep learning, DL)的感知能力和 RL 的决策能力, 为解决连续高维空间的制导难题提供了新的可能。针对机动目标的拦截问题, 文献 [13] 基于信赖域策略优化 (trust region policy optimization, TRPO) 算法将飞行状态直接映射为制导指令, 并在训练时随机初始化拦截弹和目标状态, 以提高制导策略的泛化能力, 仿真表明, 该制导策略相比于 PNG 和增强比例导引律, 在训练场景和未学习场景下均具有更低的脱靶量。文献 [14] 将 TRPO 替换为近端策略优化 (proximal policy optimization, PPO) 算法, 并在目标函数中引入熵项, 进一步提高了拦截概率。文献 [15] 将加速度指令构造为零控脱靶量和待飞时间的函数, 利用深度确定性策略梯度 (deep deterministic policy gradient, DDPG) 算法确定制导增益, 与飞行状态直接映射为加速度指令相比, 可以实现更快的收敛。然而, 文献 [13-15] 均研究二维拦截问题, 将其扩展到三维空间时, 势必因为飞行器纵向和侧向之间的耦合关系及动作维度的增加而增大训练难度。文献 [16] 研究了三维空间中仅有角度信息的拦截问题, 通过在 PPO 中引入循环网络实现了比增强零控脱靶量制导律更低的脱靶量, 但其奖励函数具有一定的稀疏性, 导致训练效率降低。可见, 尽管 DRL 已经取得了较好的成果, 但将其应用于三维空间拦截问题时, 由于动作维度增加及稀疏的奖励函数, 仍存在训练效率低的问题。

针对上述问题, 本文提出了一种基于柔性演员-评论(soft actor-critic, SAC)的三维自适应 PNG。相较于 DDPG, SAC 具有超参数少、训练稳定的优势, 同时具有比 TRPO 和 PPO 更高的数据利用率, 因此更加适合复杂制导问题。为此, 将相对运动模型构建为马尔可夫决策过程(Markov decision processes, MDP), 通过 DRL 将飞行状态直接映射为导航比, 并引入预测-校正方法以提高训练效率, 同时为了提高制导策略的适应性, 目标机动方式在每回合随机初始化。仿真结果表明, 提出的制导策略可以实现高精度的拦截, 并且相比于传统 PNG 具有更低的能量损耗。

1　问题描述与相对运动建模

1.1　拦截相对运动建模

三维空间中, 拦截弹和目标的相对运动关系可由图 1 描述, 其中, v_{M}和v_{T}分别为拦截弹和目标的速度, R为两者距离, θ和σ分别为速度倾角和速度方位角, λ_{D}和λ_{T}分别为视线高低角和视线方位角。

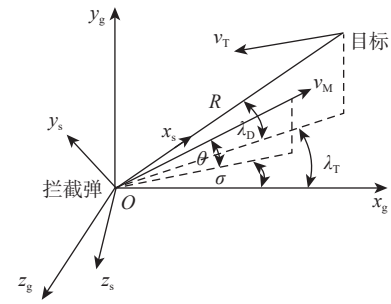


图 1　拦截弹-目标相对运动关系

Fig. 1　Interceptor-target relative motion relationship

由于末制导阶段拦截弹和目标相对距离近且飞行速度极快, 忽略两者速度损失, 假定v_{M}和v_{T}为常值。根据图 1 几何关系可得

$$\begin{cases} R=\sqrt{(x-x_{\mathrm{T}})^2+(y-y_{\mathrm{T}})^2+(z-z_{\mathrm{T}})^2} \\ \lambda_{\mathrm{D}}=\arcsin((y-y_{\mathrm{T}})/R) \\ \lambda_{\mathrm{T}}=\arctan((z-z_{\mathrm{T}})/(x-x_{\mathrm{T}})) \end{cases} \tag{1}$$

式中: (x,y,z)和$(x_{\mathrm{T}},y_{\mathrm{T}},z_{\mathrm{T}})$分别为拦截器和目标在地面系中的位置。

将式(1)两边对时间求导, 得相对运动方程为

$$\begin{cases} \dot{\lambda}_{\mathrm{D}}=\dfrac{\Delta\dot{y}R-\Delta y\dot{R}}{R\sqrt{R^2-\Delta y^2}} \\ \dot{\lambda}_{\mathrm{T}}=\dfrac{\Delta\dot{z}\Delta x-\Delta z\Delta\dot{x}}{\Delta x^2+\Delta z^2} \\ \dot{R}=\dfrac{\Delta x\Delta\dot{x}+\Delta y\Delta\dot{y}+\Delta z\Delta\dot{z}}{R} \end{cases} \tag{2}$$

式中: $\Delta x=x-x_{\mathrm{T}}$; $\Delta\dot{x}=\dot{x}-\dot{x}_{\mathrm{T}}$; Δy、$\Delta\dot{y}$、Δz和$\Delta\dot{z}$采用相同定义。

1.2　目标机动规律

目标在纵向和侧向分别采用正弦机动和余弦机动以增加突防能力:

$$\begin{cases} a_{\mathrm{D}}=-ng\sin(\omega t) & 0\leqslant t\leqslant t_{\mathrm{p}} \\ a_{\mathrm{T}}=-ng\cos(\omega t) & 0\leqslant t\leqslant t_{\mathrm{p}} \end{cases} \tag{3}$$

式中: n和ω分别为机动过载和机动频率; g 为重力加速度; $t_{\mathrm{p}}=2\pi/\omega$为机动周期。当飞行时间$t>t_{\mathrm{p}}$时, 目标不再进行机动, 即$a_{\mathrm{D}}=a_{\mathrm{T}}=0$。

1.3　比例导引律

三维制导律的经典设计方法是将三维运动解耦到两个互相垂直的平面, 即俯冲平面和转弯平

面，分别在2个平面内进行制导律求解，并在三维空间中进行综合。对PNG的改进，主要体现在比例系数和偏置项上，采用后者需要拦截器时刻计算偏置值，将非常考验弹载计算机的计算性能。因此，本文设计了如下自适应PNG：

$$\begin{cases}\dot{\gamma}_{\mathrm{D}}=K_{\mathrm{D}}\dot{\lambda}_{\mathrm{D}}\\ \dot{\gamma}_{\mathrm{T}}=K_{\mathrm{T}}\dot{\lambda}_{\mathrm{T}}\end{cases} \tag{4}$$

式中：$\dot{\gamma}_{\mathrm{D}}$和$\dot{\gamma}_{\mathrm{T}}$分别为速度方位角在俯冲平面和转弯平面内的角加速度；K_{D}和K_{T}分别为纵向和侧向导航比。采用DRL方法对其进行离线学习并在线修正。

2 基于SAC的制导系数智能调参

2.1 SAC智能调参建模

RL问题通常被构建为MDP，其通过状态、动作和奖励描述了智能体和环境的交互，通常被定义为一个五元组$<S,A,\boldsymbol{P},G,\gamma>$，其中，$S$为状态的集合，$A$为智能体动作的集合，$\boldsymbol{P}$为状态转移概率矩阵，$G$为奖励函数，$\gamma$为折扣率。在MDP中，智能体在时间$t$时刻观测到环境状态$\boldsymbol{s}_t\in S$，并根据策略$\Phi(\boldsymbol{a}_t|\boldsymbol{s}_t)$采取动作$\boldsymbol{a}_t\in A$作用于环境，环境将依据状态转移概率$P(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t,\boldsymbol{a}_t)\in\boldsymbol{P}$反馈给智能体下一状态$\boldsymbol{s}_{t+1}$和奖励值$r(\boldsymbol{s}_t,\boldsymbol{a}_t)\in G$。$\gamma\in[0,1]$决定了未来奖励在当前时刻的折扣程度，$\gamma\to1$表示智能体更加注重未来奖励，$\gamma\to0$则优先考虑当前奖励。

2.1.1 状态空间设计

拦截弹根据状态空间中的信息感知当前状态并作出决策。繁杂的状态空间会导致智能体无法提取有用的信息，同时引起计算效率的下降，而状态信息的不足则会缩小各状态之间的差异，使智能体无法辨别，作出错误决策。因此，根据相对运动方程，设计了如式(5)所示的归一化状态空间，包括拦截弹和目标间的相对信息、拦截弹自身状态及目标的位置信息。

$$S=\left\{\bar{R},\bar{\lambda}_{\mathrm{D}},\bar{\lambda}_{\mathrm{T}},\bar{\theta},\bar{\sigma}\right\} \tag{5}$$

其中：

$$\begin{cases}\bar{R}=\dfrac{R}{R_0}\\ \bar{\lambda}_{\mathrm{D}}=\dfrac{2\lambda_{\mathrm{D}}}{\pi},\bar{\lambda}_{\mathrm{T}}=\dfrac{\lambda_{\mathrm{T}}}{2\pi}\\ \bar{\theta}=\dfrac{2\theta}{\pi},\bar{\sigma}=\dfrac{\sigma}{2\pi}\end{cases} \tag{6}$$

式中：R_0和R分别为初始时刻和当前时刻拦截弹和目标间的距离。

2.1.2 动作空间设计

DRL的目标是获得不同状态下的最优动作。根据制导需求，显然应将动作选为导航比K_{D}和K_{T}，即

$$A=\{K_{\mathrm{D}},K_{\mathrm{T}}\} \tag{7}$$

式中：$K_{\mathrm{D}}\in[0,K_{\mathrm{DMAX}}]$和$K_{\mathrm{T}}\in[0,K_{\mathrm{TMAX}}]$均为连续变量，$K_{\mathrm{DMAX}}$和$K_{\mathrm{TMAX}}$分别为纵向和侧向的最大导航比。

2.1.3 奖励函数设计

合适的奖励函数是成功应用DRL算法的关键，其不仅决定了算法优化的目标，同时也影响着算法收敛程度和速度。在本问题中，制导指标为脱靶量和能量损耗，其中，能量损耗以拦截器的总过载和表征。对于机动目标，解析给出其与飞行状态之间的关系是不现实的，若只在终端设置奖励，则会产生奖励稀疏现象，导致训练效率下降。为此，引入预测-校正方法为飞行过程中的决策提供过程奖励。在t时刻，利用策略网络计算状态$\boldsymbol{s}_t$下的导航比K_{D}和K_{T}，并以该导航比进行积分预测，根据预测的终端脱靶量和总过载和计算当前决策的过程奖励，最终将该奖励反馈给智能体，对导航比进行校正。过程奖励和终端奖励可用统一的奖励函数计算，如式(8)所示，由脱靶量项奖励R_{d}、能量损耗项奖励R_{n}及偏置项奖励R_{b}组成。若当前状态$\boldsymbol{s}_{\mathrm{c}}$为终端状态$\boldsymbol{s}_{\mathrm{e}}$，即$\boldsymbol{s}_{\mathrm{c}}=\boldsymbol{s}_{\mathrm{e}}$，根据式(8)计算得终端奖励，否则为过程奖励。

$$G=R_{\mathrm{d}}+R_{\mathrm{n}}+R_{\mathrm{b}} \tag{8}$$

脱靶量项奖励R_{d}为

$$R_{\mathrm{d}}=\begin{cases}k_{\mathrm{dp}}D_{\mathrm{p}} & \boldsymbol{s}_{\mathrm{c}}\neq\boldsymbol{s}_{\mathrm{e}}\\ k_{\mathrm{de}}D_{\mathrm{e}} & \boldsymbol{s}_{\mathrm{c}}=\boldsymbol{s}_{\mathrm{e}}\end{cases} \tag{9}$$

能量损耗项奖励R_{n}为

$$R_{\mathrm{n}}=\begin{cases}k_{\mathrm{np}}N_{\mathrm{p}} & \boldsymbol{s}_{\mathrm{c}}\neq\boldsymbol{s}_{\mathrm{e}}\\ k_{\mathrm{ne}}N_{\mathrm{e}} & \boldsymbol{s}_{\mathrm{c}}=\boldsymbol{s}_{\mathrm{e}}\end{cases} \tag{10}$$

偏置项奖励R_{b}为

$$R_{\mathrm{b}}=\begin{cases}k_{\mathrm{b}} & D_{\mathrm{e}}\leqslant D_{\mathrm{lim}}\\ 0 & D_{\mathrm{e}}>D_{\mathrm{lim}}\end{cases} \tag{11}$$

式中：D和N为脱靶量和总过载和，下标p和e分别表示预测和真实终端值；k_{dp}、k_{de}、k_{np}、k_{ne}和k_{b}为权重常数；D_{lim}为判定拦截成功的最大脱靶量。

由于过程奖励基于预测值计算，无法反映真实的终端信息，R_{d}和R_{n}中权重$k_{\mathrm{dp}}<k_{\mathrm{de}}$，$k_{\mathrm{np}}<k_{\mathrm{ne}}$。另外，设置偏置项奖励$R_{\mathrm{b}}$，当$D_{\mathrm{e}}\leqslant D_{\mathrm{lim}}$，即拦截成功时，给予正奖励$k_{\mathrm{b}}$，促使策略朝着低脱靶量方向优化。通过多次训练测试，各参数的具体值如表1所示。

2.2 SAC算法学习流程

SAC是集合了异策略、最大熵和演员-评论家的RL算法。作为异策略算法中的一员，SAC创建

表 1　奖励函数中的权重

Table 1　Weights in the reward function

k_{dp}	k_{de}	k_{np}	k_{ne}	k_{b}
−0.1	−0.5	−1	−4	100

一个大小为$|\mathcal{D}|$的经验缓存池来提高数据利用率，并在网络更新时，从中随机抽取M条经验进行反向传播，打破了数据之间的关联性。作为最大熵算法，SAC 的目的是寻找最优策略Φ^*，使 RL 中的累计奖励和熵的期望同时最大化，即

$$\Phi^* = \arg\max_{\Phi} \sum_t E_{(\boldsymbol{s}_t,\boldsymbol{a}_t)\sim\rho_\Phi}\left[\gamma^t\left(r(\boldsymbol{s}_t,\boldsymbol{a}_t)+\tau H(\Phi(\cdot|\boldsymbol{s}_t))\right)\right] \tag{12}$$

式中：τ为温度系数，决定了熵在策略优化时的重要性；$\rho_\Phi(\boldsymbol{s}_t)$和$\rho_\Phi(\boldsymbol{s}_t,\boldsymbol{a}_t)$分别为状态和状态-动作对在策略$\Phi$下的分布；熵度量了动作的不确定性，可根据式(13)计算：

$$H(\Phi(\cdot|\boldsymbol{s}_t)) = E_{\boldsymbol{a}\sim\Phi(\cdot|\boldsymbol{s}_t)}\left[-\log\Phi(\boldsymbol{a}|\boldsymbol{s}_t)\right] \tag{13}$$

熵项的引入大大提高了智能体对环境扰动的鲁棒性及对环境的探索能力，可有效防止过早陷入局部极小值点。SAC 采用了演员-评论家架构，即同时存在评价网络和策略网络，网络更新频率为N，具体的更新过程如图 2 所示，分为策略评估和策略提升 2 个步骤。

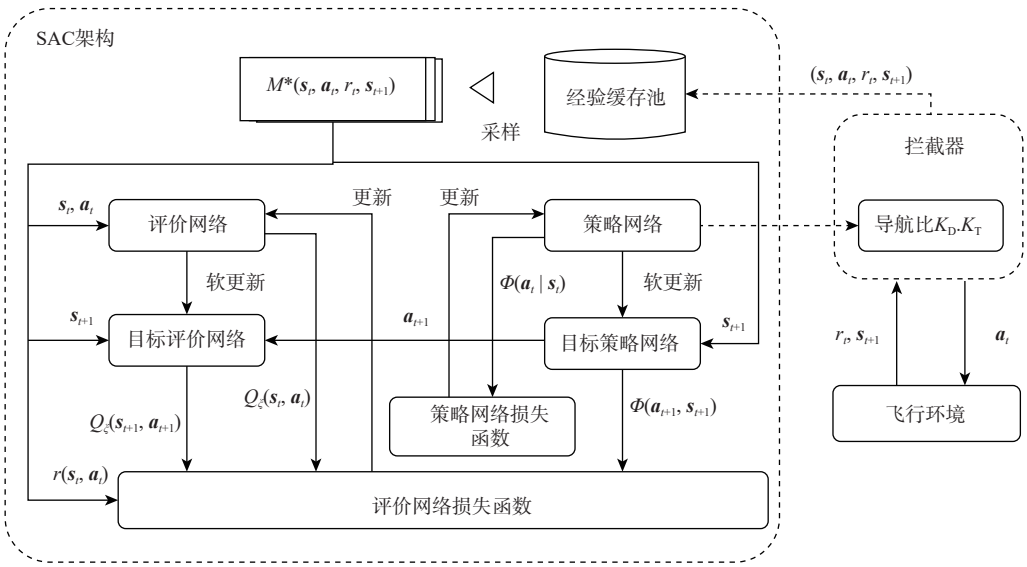


图 2　基于 SAC 的策略训练架构

Fig. 2　Strategy training architecture based on SAC

2.2.1　策略评估

策略评估是策略提升的基础，目的是获得真实的软动作价值函数。在 SAC 中，其被定义为奖励和熵和的期望，如下：

$$Q_\Phi(\boldsymbol{s},\boldsymbol{a}) = E_{(\boldsymbol{s}_t,\boldsymbol{a}_t)\sim\rho_\Phi}\left[\sum_{t=0}^{\infty}\gamma^t r(\boldsymbol{s}_t,\boldsymbol{a}_t)+\tau\sum_{t=1}^{\infty}\gamma^t\mathcal{H}(\Phi(\cdot|\boldsymbol{s}_t))\right] \tag{14}$$

评价网络通过时序差分算法最小化损失函数J_Q来提高对式 (14) 的拟合程度：

$$J_Q(\boldsymbol{\xi}) = E_{(\boldsymbol{s}_t,\boldsymbol{a}_t,\boldsymbol{s}_{t+1})\sim\mathcal{D}\boldsymbol{a}_{t+1}\sim\pi_{\boldsymbol{\psi}}}\left[\frac{1}{2}\left(Q_{\boldsymbol{\xi}}(\boldsymbol{s}_t,\boldsymbol{a}_t)-Q'_{\boldsymbol{\xi}}(\boldsymbol{s}_t,\boldsymbol{a}_t)\right)^2\right] \tag{15}$$

式中：$Q'_{\boldsymbol{\xi}}(\boldsymbol{s}_t,\boldsymbol{a}_t)=r(\boldsymbol{s}_t,\boldsymbol{a}_t)+\gamma(Q_{\boldsymbol{\xi}}(\boldsymbol{s}_{t+1},\boldsymbol{a}_{t+1})-\tau\ln(\Phi_{\boldsymbol{\psi}}(\boldsymbol{a}_{t+1}|\boldsymbol{s}_{t+1})))$，$\boldsymbol{\psi}$和$\boldsymbol{\xi}$分别为策略网络和评价网络参数。

评价网络可通过随机梯度下降算法对其进行更新：

$$\boldsymbol{\xi}' = \boldsymbol{\xi}-\alpha\nabla J_Q(\boldsymbol{\xi}) \tag{16}$$

式中：α为学习率。

2.2.2　策略提升

策略提升将对现有策略Φ进行优化。由软动作价值定义和式 (12) 可知，最优策略中动作的概率分布应与真实的软动作价值分布相同。因此，将软动作价值化为玻尔兹曼分布形式，通过最小化两者的相对熵进行策略提升：

$$\Phi_{\mathrm{new}} = \arg\min_{\Phi'\in\Pi} D_{\mathrm{KL}}\left(\Phi'(\cdot|\boldsymbol{s}_t)\left\|\frac{\exp\left(\tau^{-1}Q^{\Phi_{\mathrm{old}}}(\boldsymbol{s}_t,\cdot)\right)}{Z^{\Phi_{\mathrm{old}}}(\boldsymbol{s}_t)}\right.\right) \tag{17}$$

式中：Π为可选策略的集合；$Z^{\Phi_{\mathrm{old}}}(\boldsymbol{s}_t)$为偏分函数。

为了降低方差，同时方便梯度回传，网络更新采用了重参数化技术，即将策略写为$\boldsymbol{a}_t=f_{\boldsymbol{\psi}}(\boldsymbol{\varepsilon}_t;\boldsymbol{s}_t)$，$\boldsymbol{\varepsilon}_t$为输入噪声[17]。化简式 (17) 可得用于策略网络的损失函数为

$$J_{\Phi}(\boldsymbol{\psi})=E_{s_t\sim\mathcal{D},a_t\sim\Phi_{\psi}}[\ln\Phi_{\psi}\left(f_{\psi}(\boldsymbol{\varepsilon}_t;\boldsymbol{s}_t)\middle|\boldsymbol{s}_t\right)-\tau^{-1}Q_{\xi}\left(\boldsymbol{s}_t,f_{\psi}(\boldsymbol{\varepsilon}_t;\boldsymbol{s}_t)\right)+\ln Z(\boldsymbol{s}_t)] \tag{18}$$

温度系数τ控制了熵的重要程度，由于策略的熵在训练过程中不断变化，固定的值将很难使算法收敛。为此，本文采用了文献[18]中改进的SAC，其将τ构造成带约束的优化问题：

$$\begin{cases}\max\limits_{\Phi_0,\cdots,\Phi_{T_f}} E\left[\sum\limits_{t=0}^{T_f}r(\boldsymbol{s}_t,\boldsymbol{a}_t)\right]\\ \text{s.t. }\forall t,\ H(\Phi_t)\geqslant H_{\text{m}}\end{cases} \tag{19}$$

式中：H_{m}为期望的最小策略熵。

通过求解式(19)，得到温度系数的损失函数$J(\tau)$，同样采用梯度下降方法对参数τ进行更新。

$$J(\tau)=E_{\boldsymbol{a}_t\sim\Phi_t}[-\tau\ln\Phi_t(\boldsymbol{a}_t|\boldsymbol{s}_t)-\tau H_{\text{m}}] \tag{20}$$

为降低网络参数剧烈波动引起的训练不稳定性，引入了目标网络和软更新技术。评价网络和策略网络对应的目标网络分别由$\boldsymbol{\xi}_t$和$\boldsymbol{\psi}_t$参数化，其网络参数通过软更新技术更新：

$$\begin{cases}\boldsymbol{\xi}_t'=(1-\epsilon)\boldsymbol{\xi}_t+\epsilon\boldsymbol{\xi}\\ \boldsymbol{\psi}_t'=(1-\epsilon)\boldsymbol{\psi}_t+\epsilon\boldsymbol{\psi}\end{cases} \tag{21}$$

式中：ϵ为平滑常数。ϵ越小，目标网络参数变化越缓慢，训练过程越稳定。

3 仿真分析与验证

为评估基于SAC制导策略的性能，对其进行训练测试。在初始时刻，拦截弹和目标在地面坐标系中的位置分别为(0, 25, 0) km和(20, 30, 20) km。拦截弹速度为1 500 m/s，初始时刻速度方向指向目标，最大过载30 g，最大导航比K_{DMAX}和K_{TMAX}均为10。目标在地面坐标系中的初始速度为(−2 000, 0, −1 500) m/s。积分步长为0.1 s，当两者距离小于350 m时，改为0.000 1 s。判定拦截成功的最大脱靶量D_{lim}为1 m。

策略网络和评价网络均采用具有3个隐藏层的全连接网络，每层的节点数如表2所示。输入层和隐藏层均采用ReLU作为激活函数，因为其不含指数项，可以有效提高计算速度，输出层无激活函数。

表2 策略网络和评价网络结构

Table 2 Architecture of the policy and critic networks

网络层	节关数	
	策略网络	评价网络
输入层	7	9
隐藏层1	256	256
隐藏层2	256	256
隐藏层3	256	256
输出层	4	1

SAC采用了随机策略，即策略网络输出每个动作的高斯分布的参数：均值和方差。根据前述动作空间设定，策略网络需要输出纵向和侧向导航比，故其输出层大小为4。同时，为了将输出值限定在预设范围内，先利用双曲正切函数压缩到[−1, 1]，再对其进行扩展。

3.1 最优导航比策略训练

针对本文中的拦截问题，利用SAC算法对制导策略进行训练。为了提高策略适应性，每回合目标机动的幅值、周期均随机改变。由于机动幅值越大，拦截成功率越低，对该场景设置更多回合数。机动幅值从[6, 7, 8, 9]序列中随机选取，每数被选择的概率是前一数的1.1倍。机动周期则在[6, 6.5, 7, 7.5, 8]中等概率选取。采样点数为10，即每回合策略网络给出10次导航比。最大训练回合数设为1 000。超参数的具体数值总结在表3中。

表3 超参数设置

Table 3 Hyperparameters setting

超参数	数值
经验缓存池大小$\lvert\mathcal{D}\rvert$	5 000
采样经验大小M	128
折扣率γ	0.98
学习率α	0.000 5
期望熵H_0	−5
温度系数τ	1
平滑常数ϵ	0.005
网络更新频率N	60

总奖励及其每50步的平滑曲线如图3所示。图4、图5和图6展示了脱靶量、总过载和及温度系数的变化。在训练前200次，由于温度系数较大，智能体更加注重探索，且此时网络参数随机初始化产生，因此，总奖励会大幅度振荡。之后，随着温度系数逐渐降低，总奖励开始收敛。500次后，总奖励维持在较高水平，脱靶量在0.05 m和0.35 m之

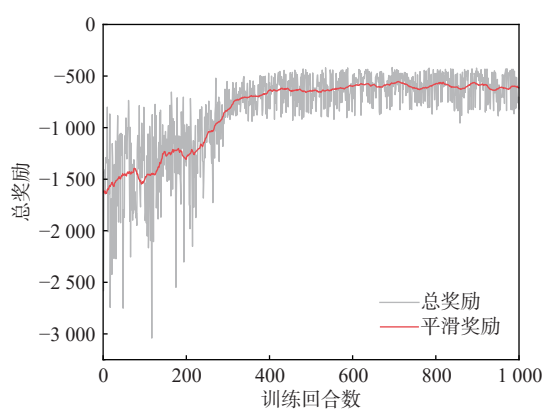


图3 总奖励曲线

Fig. 3 Curves of total reward

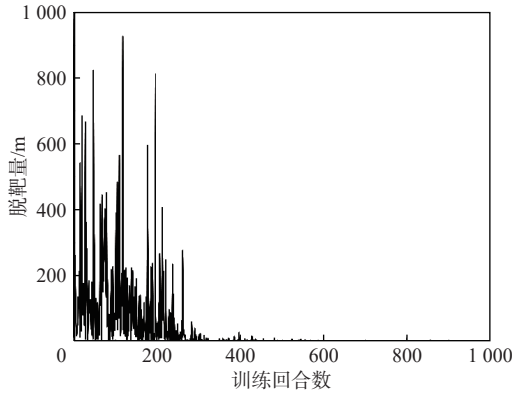


图 4　脱靶量曲线

Fig. 4　Curve of miss distance

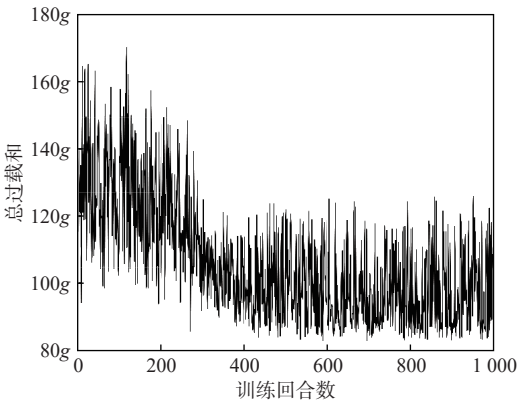


图 5　总过载和曲线

Fig. 5　Curve of total overload

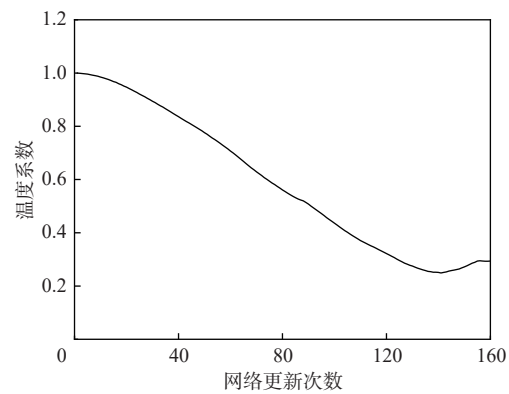


图 6　温度系数曲线

Fig. 6　Curve of temperature coefficient

间, 总过载和则因为目标机动形式改变而发生明显变化。

奖励稀疏性是影响训练效率的关键因素。图 7 对比了有无过程奖励下 4 000 次训练的终端奖励平滑曲线。可见, 当存在过程奖励时, 终端奖励从 500 次开始一直到训练后期均维持在最大值附近, 并且中间没有大幅度下降; 而无过程奖励时在 3 000 次左右才开始收敛, 并且两者收敛到相同值, 说明引入预测-校正方法不仅可以提高训练效率, 而且不会破坏策略的最优性。

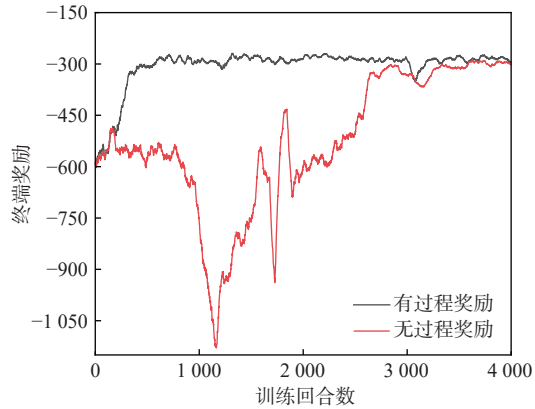


图 7　有无过程奖励下终端奖励曲线

Fig. 7　Terminal reward curves with or without process reward

3.2　制导性能对比分析

为了测试较提出制导策略 (G1) 的性能, 分别与导航比为 7, 8, 9, 10 的 PNG 和同样利用 SAC 算法训练的制导律 (G2) 进行对比分析, 其中 G2 仅考虑脱靶量, 即式 (8) 奖励函数中无能量损耗项R_n。

3.2.1　训练场景下的测试

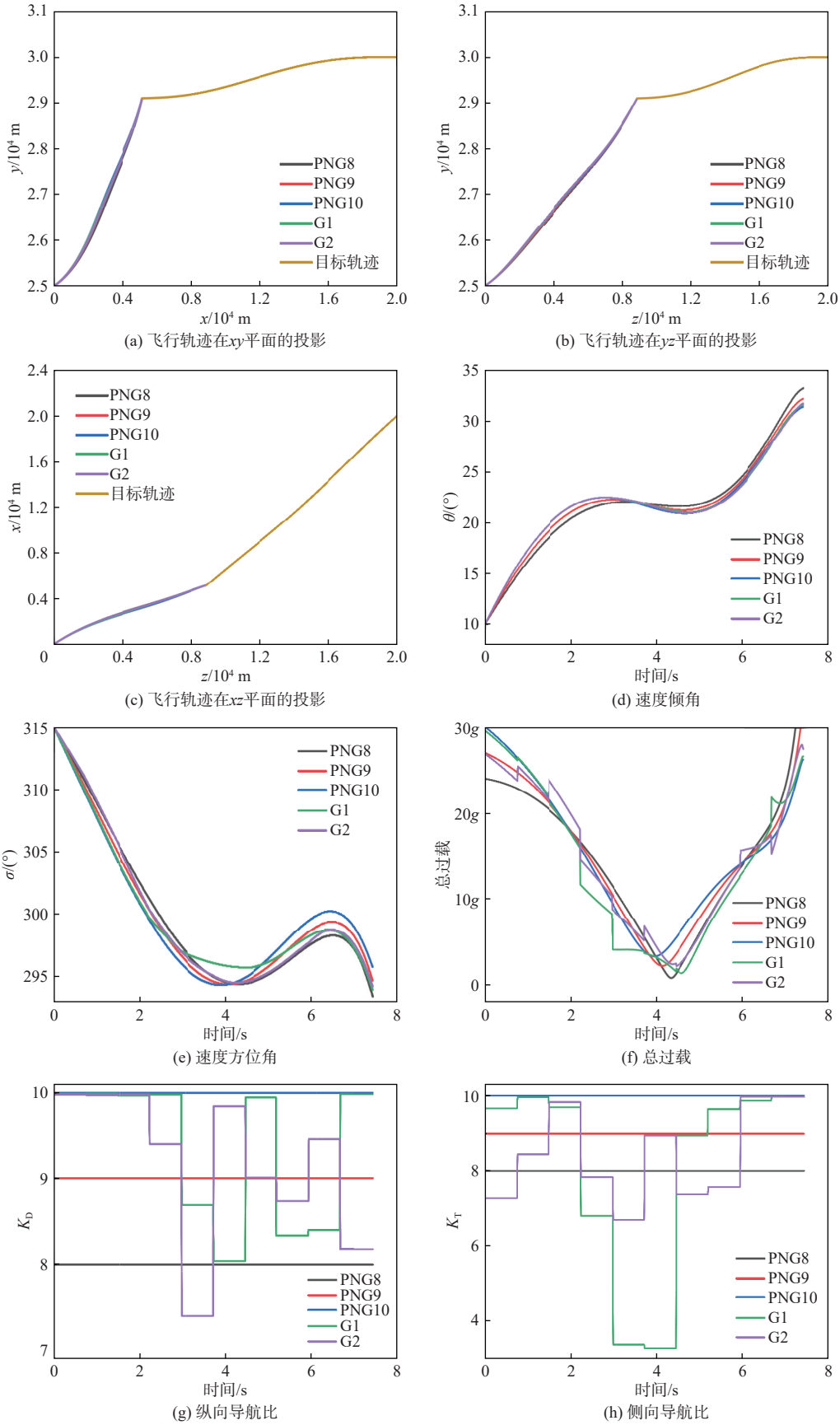
在训练过的场景下进行测试, 目标机动过载 $n=9$, 周期 $T_0=8$ s。拦截情况如表 4 所示, 其中, PNG7 表示拦截器纵向和侧向均采用导航比为 7 的 PNG, 其余类似。可见, 除导航比为 7 的 PNG 外, 其余制导律均可成功拦截目标, 并且导航比越大, 脱靶量和拦截时间越小, 但也会消耗更多能量。G1 中加入了对过载的惩罚, 实现低脱靶量的同时具有最低的能量损耗。

表 4　不同制导律下的终端参数

Table 4　Terminal parameters with different guidance laws

制导律	飞行时间/s	脱靶量/m	总过载和
PNG7	7.439	2.974	107.217 *g*
PNG8	7.435	0.780	108.323 *g*
PNG9	7.432	0.263	110.543 *g*
PNG10	7.430	0.113	112.474 *g*
G1	7.430	0.220	103.326 *g*
G2	7.432	0.248	109.267 *g*

图 8 和图 9 分别给出了拦截成功时各拦截器的飞行状态变化曲线和目标速度的变化曲线。由于目标在纵向做正弦机动, 其y_g向速度的绝对值在 $t=4$即半周期时达到最大, 之后降低, 反映在高度上即变化率先增大后减小, 而高度的变化也导致拦截器和目标间的纵向视线角速率变小, 根据式 (4), 拦截器的速度倾角θ将产生同样变化。速度方位角受目标纵向和侧向机动的双重影响, 在 4 s 后, 目标x_g向速度的绝对值先减小后增大, 而z_g向速度先增大后减小, 由 PNG 的物理意义可知, 拦截器的速度



(a) 飞行轨迹在xy平面的投影　(b) 飞行轨迹在yz平面的投影

(c) 飞行轨迹在xz平面的投影　(d) 速度倾角

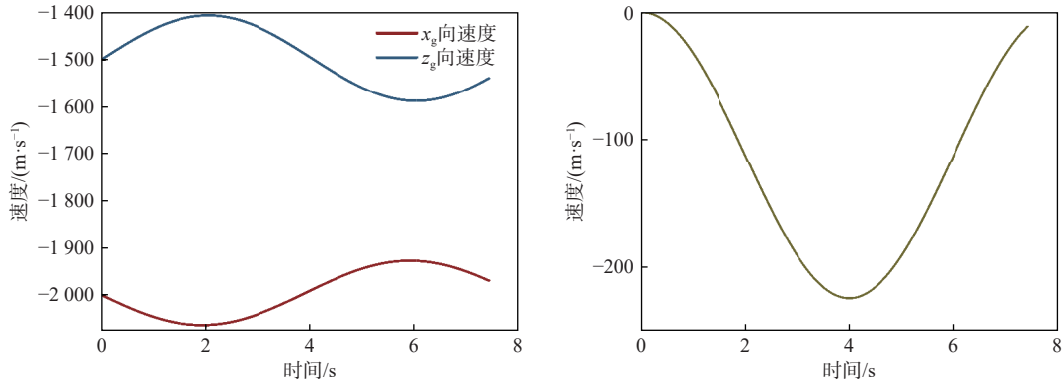
(e) 速度方位角　(f) 总过载

(g) 纵向导航比　(h) 侧向导航比

图 8　不同制导律下拦截器参数变化曲线

Fig. 8　Curves of interceptor parameters with different guidance laws

方位角曲线将先上升后下降。观察图 8(f) 可知，在目标加速度反向时，采用大导航比 PNG 的拦截器更加敏感，即具有最短的反应时间和最大的过载，可有效降低末端的拦截压力，防止过载饱和。结合



(a) 目标x_g和z_g向速度曲线　　(b) 目标y_g向速度曲线

图 9　目标速度曲线

Fig. 9　Target speed profile

图 8(g) 和图 8(h)，G1 会根据双方距离大小采用小导航比K_T以降低机动幅度，节省能量，而在末端为降低脱靶量则快速提高增大K_T。

3.2.2　未训练场景下的测试

本节将在 3 种未训练场景下测试制导策略的性能，以验证该策略的适应性。场景 1 对训练场景进行扩展，其中目标机动幅值和周期分别在 6～9 和 6～8 s 之间均匀选取。表 5 给出了上述各制导律在该场景下的 2 000 次蒙特卡罗仿真测试结果。可见，导航比为 7 的 PNG 脱靶量均值大于 1 m，意味着大多数情况下拦截失败。而本文提出的制导策略 G1 具有极低的脱靶量均值和方差，且和其余制导律相比具有最低的能量损耗。

表 5　未训练场景下蒙特卡罗结果统计

Table 5　Monte Carlo results statistics in untrained scenarios

制导律	脱靶量均值/m	脱靶量方差/m^2	总过载和均值
PNG7	1.793	2.504	105.983 g
PNG8	0.380	0.116	104.811 g
PNG9	0.192	0.012	105.242 g
PNG10	0.187	0.012	106.297 g
G1	0.189	0.012	98.532 g
G2	0.188	0.018	104.906 g

场景 2 在训练场景的基础上将正弦机动改为方波机动，机动方程如式 (22) 所示，当$t > T_0$时，$a_D = a_T = 0$。场景 3 则将拦截器最大过载降至 20g。

$$\begin{cases} a_D = -ng\mathrm{sgn}(\sin(\omega t)) & 0 \leqslant t \leqslant T_0 \\ a_T = -ng\mathrm{sgn}(\cos(\omega t)) & 0 \leqslant t \leqslant T_0 \end{cases} \tag{22}$$

图 10 和图 11 分别给出了目标方波机动和拦截器低过载时的终端参数。可见，在这两种场景下，G1 的最大脱靶量均在 PNG9 和 PNG10 之间，而能量损耗却是最小的。对场景 3 来说，降低拦截器最大过载导致末端没有充足的能力对目标机动作出反应，此时各制导律均有拦截失败的情况。

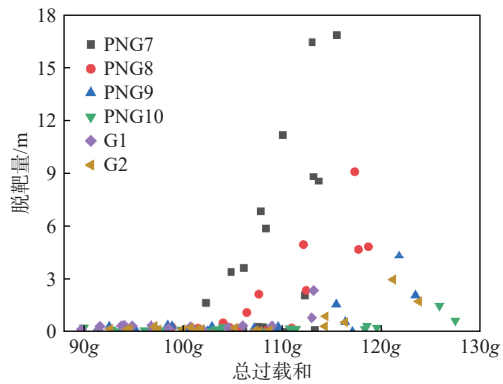


图 10　方波机动下终端参数

Fig. 10　Terminal parameters under square wave maneuvering

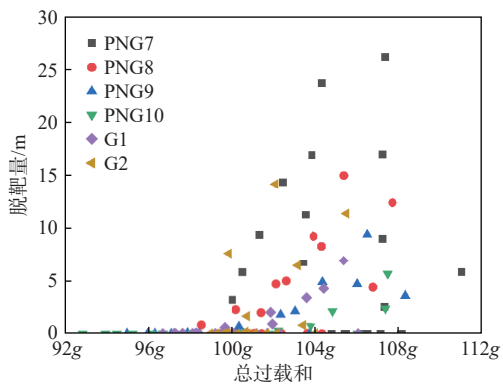


图 11　低过载下终端参数

Fig. 11　Terminal parameters under low overload

从训练场景和未训练场景中的仿真分析可见，采用本文基于 SAC 的 PNG，针对不同的机动方式、不同机动形式的目标及采用低过载拦截器均有较好的拦截性能，证明了所提制导策略的有效性和泛化性。

4　结　论

1）针对高机动目标拦截问题，同时考虑了脱靶量和能量损耗，提出了一种基于 DRL 的自适应

PNG。相较于传统PNG,在保证低脱靶量的同时,可降低能量损耗。

2)相较于平面拦截,三维空间中的拦截问题由于动作维度的增加和决策空间的复杂化导致严重的奖励稀疏现象,在构建奖励函数时引入预测-校正思想以提高训练效率。

3)仿真表明,在不同的目标机动形式、不同的机动周期和低可用过载等场景下,本文方法同样具有良好的性能,证明了该方法的适应性。

在后续的研究中,将考虑更加真实的拦截环境,同时进一步提高PNG性能。

参考文献(References)

[1] 纪毅, 王伟, 张宏岩, 等. 面向高机动目标拦截任务的空空导弹制导方法综述[J]. 航空兵器, 2022, 29(6): 15-25.
JI Y, WANG W, ZHANG H Y, et al. A survey on guidance method of air-to-air missiles facing high maneuvering targets[J]. Aero Weaponry, 2022, 29(6): 15-25(in Chinese).

[2] JEON I S, KARPENKO M, LEE J I. Connections between proportional navigation and terminal velocity maximization guidance[J]. Journal of Guidance, Control, and Dynamics, 2020, 43(2): 383-388.

[3] ZHANG B L, ZHOU D. Optimal predictive sliding-mode guidance law for intercepting near-space hypersonic maneuvering target[J]. Chinese Journal of Aeronautics, 2022, 35(4): 320-331.

[4] LI C D, WANG J, HE S M, et al. Collision-geometry-based generalized optimal impact angle guidance for various missile and target motions[J]. Aerospace Science and Technology, 2020, 106: 106204.

[5] YE D, SHI M M, SUN Z W. Satellite proximate interception vector guidance based on differential games[J]. Chinese Journal of Aeronautics, 2018, 31(6): 1352-1361.

[6] 郭建国, 胡冠杰, 郭宗易, 等. 天线罩误差下基于ADP的机动目标拦截制导策略[J]. 宇航学报, 2022, 43(7): 911-920.
GUO J G, HU G J, GUO Z Y, et al. ADP-based guidance strategy for maneuvering target interception under radome errors[J]. Journal of Astronautics, 2022, 43(7): 911-920(in Chinese).

[7] 白志会, 黎克波, 苏文山, 等. 现实真比例导引拦截任意机动目标捕获区域[J]. 航空学报, 2020, 41(8): 323947.
BAI Z H, LI K B, SU W S, et al. Capture region of RTPN guidance law against arbitrarily maneuvering targets[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(8): 323947(in Chinese).

[8] 王荣刚, 唐硕. 拦截高速运动目标广义相对偏置比例制导律[J]. 西北工业大学学报, 2019, 37(4): 682-690.
WANG R G, TANG S. Intercepting higher-speed targets using generalized relative biased proportional navigation[J]. Journal of Northwestern Polytechnical University, 2019, 37(4): 682-690(in Chinese).

[9] LI K B, SHIN H S, TSOURDOS A, et al. Capturability of 3D PPN against lower-speed maneuvering target for homing phase[J]. IEEE Transactions on Aerospace and Electronic Systems, 2020, 56(1): 711-722.

[10] SHIN H S, LI K B. An improvement in three-dimensional pure proportional navigation guidance[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(5): 3004-3014.

[11] 张秦浩, 敖百强, 张秦雪. Q-learning强化学习制导律[J]. 系统工程与电子技术, 2020, 42(2): 414-419.
ZHANG Q H, AO B Q, ZHANG Q X. Reinforcement learning guidance law of Q-learning[J]. Systems Engineering and Electronics, 2020, 42(2): 414-419(in Chinese).

[12] 李庆波, 李芳, 董瑞星, 等. 利用强化学习开展比例导引律的导航比设计[J]. 兵工学报, 2022, 43(12): 3040-3047.
LI Q B, LI F, DONG R X, et al. Navigation ratio design of proportional navigation law using reinforcement learning[J]. Acta Armamentarii, 2022, 43(12): 3040-3047(in Chinese).

[13] 邱潇颀, 高长生, 荆武兴. 拦截大气层内机动目标的深度强化学习制导律[J]. 宇航学报, 2022, 43(5): 685-695.
QIU X Q, GAO C S, JING W X. Deep reinforcement learning guidance law for intercepting endo-atmospheric maneuvering targets[J]. Journal of Astronautics, 2022, 43(5): 685-695(in Chinese).

[14] CHEN W X, GAO C S, JING W X. Proximal policy optimization guidance algorithm for intercepting near-space maneuvering targets[J]. Aerospace Science and Technology, 2023, 132: 108031.

[15] HE S M, SHIN H S, TSOURDOS A. Computational missile guidance: a deep reinforcement learning approach[J]. Journal of Aerospace Information Systems, 2021, 18(8): 571-582.

[16] GAUDET B, FURFARO R, LINARES R. Reinforcement learning for angle-only intercept guidance of maneuvering targets[J]. Aerospace Science and Technology, 2020, 99: 105746.

[17] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor[C]//Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018, 80: 1861-1870.

[18] HAARNOJA T, ZHOU A, HARTIKAINEN K, et al. Soft actor-critic algorithms and applications[EB/OL]. (2019-01-29)[2023-06-01]. https://arxiv.org/abs/1812.05905v2.

Deep reinforcement learning intelligent guidance for intercepting high maneuvering targets

ZHANG Hao[1]，ZHU Jianwen[1, 2, *]，LI Xiaoping[1]，BAO Weimin[1, 3]

(1. School of Aerospace Science and Technology，Xidian University，Xi'an 710126，China；
2. College of Missile Engineering，Rocket Force University of Engineering，Xi'an 710025，China；
3. China Aerospace Science and Technology Corporation，Beijing 100048，China)

Abstract：An adaptive proportional navigation law with intelligent parameter adjustment by deep reinforcement learning is presented to address the issue of excessive miss distances and energy loss in the interception of moving targets using fixed coefficient proportional navigation law. First, a state space based on real-time flight states, an action space containing lateral and vertical gains, and a reward function model integrating different states is established. Meanwhile, a prediction-correction method is introduced to improve the accuracy of action evaluation in the model design of the reward function. Secondly, the soft actor-critic (SAC) algorithm is employed to train a network parameter and guidance parameter decision system that takes into account the miss distances and energy loss according to the relative motion states of the interceptor and the target. In comparison to traditional proportional navigation guidance, the simulation results demonstrate that the guidance technique has strong adaptability and can greatly minimize energy loss while retaining low miss distances.

Keywords：intelligent guidance；maneuvering targets；deep reinforcement learning；proportional navigation；Markov decision process

Received：2023-06-15；**Accepted**：2023-09-11；**Published Online**：2023-09-27 18：44
URL：link.cnki.net/urlid/11.2625.V.20230927.1317.001
* **Corresponding author**. E-mail：zhujianwen1117@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0410

VSV 调节机构运动精度可靠度计算及误差分析

杨奕凤[1,2]，王艺[3,4,*]，刘傲宇[1,2]，李佳[1]，谢里阳[1,2]

(1. 东北大学 机械工程与自动化学院，沈阳 110819；2. 东北大学 航空动力装备振动及控制教育部重点实验室，沈阳 110819；
3. 中国航发沈阳发动机研究所，沈阳 110015；4. 北京航空航天大学 航空发动机研究院，北京 102206)

摘 要： 以某航空发动机可调静子叶片 (VSV) 调节机构为研究对象，建立了含运动副间隙的运动学模型。考虑到驱动误差、尺寸误差和装配间隙等因素的随机性，建立了考虑各静子叶片运动精度失效相关的 VSV 调节机构运动精度可靠性模型。引入 BP 神经网络 (BPNN) 代理模型，采用 Monte Carlo 方法计算了机构的可靠度，给出了可靠度随静子叶片数量和失效阈值的变化规律。通过与单元失效独立假设的系统模型和完全相关系统模型进行对比，验证了考虑失效相关性的机构运动精度可靠性模型的合理性。提出了一种估算代理模型误差造成的复杂系统可靠度误差的方法，并证明了该 VSV 调节机构的可靠度计算结果是准确可信的。

关 键 词： 可调静子叶片调节机构；失效相关性；BP 神经网络；可靠度；误差分析
中图分类号： V233.95; TB114
文献标志码： A **文章编号：** 1001-5965(2025)06-2070-11

可调静子叶片(variable stator vane, VSV)调节机构是现代航空发动机及燃气轮机中的重要运动机构，其功能是调节静子叶片角度，扩大压气机稳定工作范围，提高发动机喘振裕度[1-3]。VSV 调节机构是由多个简单构型运动机构高度集成的多级联调机构，其零件数量众多，各零件的连接关系多样，加之驱动误差、尺寸误差和装配间隙等因素的影响，在恶劣的服役条件下极易出现可靠性问题。现有针对 VSV 调节机构的研究大多是机构的运动学分析[4]、动力学分析[5-6] 和优化设计[7] 等，而关于机构可靠性分析的研究较少。因此，开展 VSV 调节机构的可靠性分析具有重要意义。

机构运动可靠性的定义为：机构在规定的使用条件下，规定的使用时间内，精确、及时、协调地完成规定运动功能的能力[8]。近年来，机构运动可靠性分析得到了广泛关注。李铭等[9] 利用最小次序统计量的概念建立了行星齿轮机构的可靠度计算模型，定量分析了偏载程度与行星齿轮系可靠度之间的关系。周长聪等[10] 考虑了摩擦因数、装配误差等因素的随机不确定性，对单失效模式和多失效模式串联情况下的飞机起落架收放机构的可靠性和灵敏度进行了分析。张政等[11] 考虑销轴半径、摩擦系数等输入变量含区间分布参数，建立了结合拒绝采样和主动学习 Kriging 代理模型的分析方法，对尾喷管调节机构的可靠性进行了分析。Li 等[12] 针对飞机舱门锁机构，提出了考虑磨损和多源不确定因素的基于性能裕度的可靠性分析方法。Zeng 等[13] 对具有混合不确定性的机械系统提出了基于拟 Monte Carlo 模拟方法的运动可靠性分析方法。Zhan 等[14] 分析了多种不确定性因素影响下平面 3-RRR 并联机械手的运动可靠性。王博文等[15] 讨论了折叠翼展开机构的可靠性分析中的共因失效现象。刘胜利等[16] 结合区间优化算法和 Monte Carlo 算法建立了多源不确定性下变胞机构全构态运动可靠性模型。

收稿日期： 2023-06-24；**录用日期：** 2023-08-29；**网络出版时间：** 2023-09-06 08:44
网络出版地址： link.cnki.net/urlid/11.2625.V.20230905.0952.002
基金项目： 国家科技重大专项 (J2019-Ⅳ-0002-0069)
* **通信作者**. E-mail：diyuanyinian@163.com

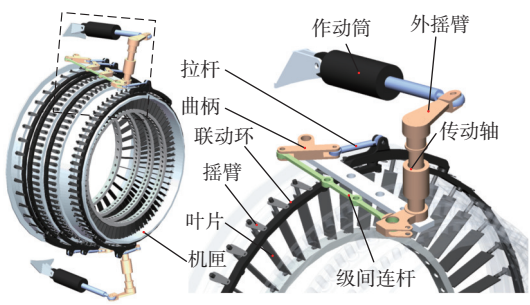
引用格式：杨奕凤，王艺，刘傲宇，等. VSV 调节机构运动精度可靠度计算及误差分析 [J]. 北京航空航天大学学报，2025，51（6）：2070-2080. YANG Y F，WANG Y，LIU A Y，et al. Calculation and error analysis of kinematic accuracy reliability of VSV adjustment mechanism [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（6）：2070-2080 （in Chinese）.

本文采用运动学和静力学理论建立含运动副间隙的 VSV 调节机构模型, 考虑多种影响因素的随机性, 根据全概率公式建立考虑各静子叶片运动精度失效之间相关性的机构运动精度可靠性模型。基于所构建的 BP 神经网络 (BP neural network, BPNN) 代理模型, 借助 Monte Carlo 技术计算调节机构运动精度可靠度。提出一种分析代理模型误差对可靠度计算误差影响的方法, 可用于基于代理模型的复杂系统可靠度计算精度的验证。

1　VSV 调节机构运动学模型

VSV 调节机构是由包含相同运动关系的多级联动静叶调节机构组成的复杂空间运动机构。如图 1(a) 所示, VSV 调节机构有 4 级静叶调节机构, 由 2 个液压作动筒驱动, 其中, 第 0 级静子叶片 42 个, 第 1 级静子叶片 84 个, 第 2 级静子叶片 82 个, 第 3 级静子叶片 72 个。在实际运动中, VSV 调节机构同时驱动 280 个静子叶片运动, 包含 2 300 个以上零件、超过 1 200 个运动副。

VSV 调节机构的工作原理如图 1(b) 所示。置于外机匣的液压作动筒推动外摇臂旋转, 带动传动轴旋转; 传动轴与级间连杆连接, 级间连杆带动各级曲柄同步旋转。在每级调节机构中, 曲柄通过拉杆带动联动环绕机匣中心轴旋转的同时沿机匣中心轴作轴向移动; 摇臂一端与联动环连接绕叶片轴旋转, 另一端与叶片轴固定连接, 带动静子叶片旋转相应角度, 起到整流作用。



(a) VSV调节机构组成　　(b) VSV调节机构工作原理

图 1　VSV 调节机构

Fig. 1　The variable stator vane adjustment mechanism

VSV 调节机构运动简图如图 2 所示, 其中, P、R、C 和 S 分别表示移动副、旋转副、圆柱副和球副。由于加工装配等因素, 机构运动副间隙不可避免。本文采用连续接触模型[17] 对摇臂-联动环运动副间隙进行分析。如图 3 所示, 假设销轴与轴套中心由无质量的刚性杆连接, 销轴中心始终在误差圆上运动, 其半径为间隙轴、孔的半径之差。则摇臂的有效长度 L_D 可以表示为

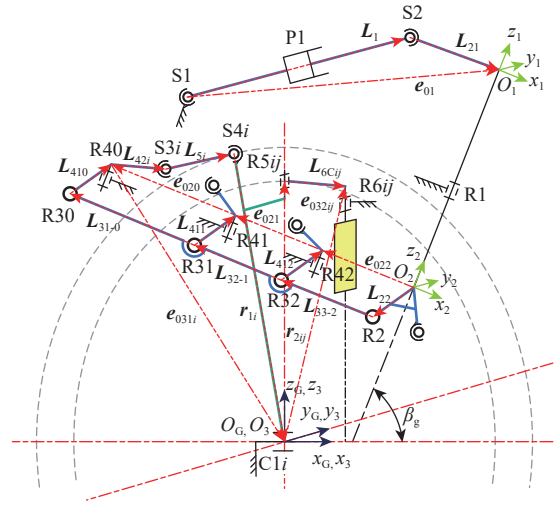


图 2　VSV 调节机构运动简图

Fig. 2　The kinematic sketch of VSV adjustment mechanism

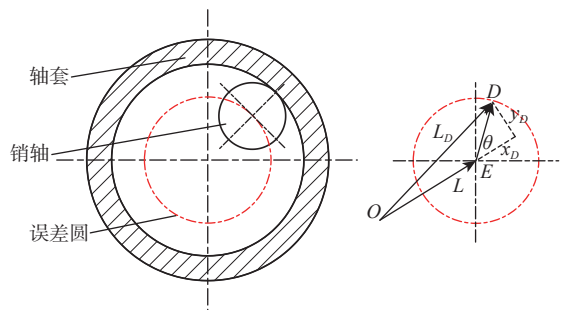


图 3　有效长度模型示意图

Fig. 3　Schematic diagram of the effective link model

$$L_D = \sqrt{(L + x_D)^2 + y_D^2} \tag{1}$$

式中: L 为摇臂的实际长度; x_D 和 y_D 为局部坐标系中销轴中心位置的坐标, 局部坐标系原点与 E 点重合, x 轴与 OE 平行。

VSV 调节机构的运动学模型为

$$\begin{cases} \boldsymbol{L}_1 + \boldsymbol{L}_{21} - \boldsymbol{e}_{01} = \boldsymbol{0} \\ \boldsymbol{L}_{22} + \boldsymbol{L}_{33\text{-}2} + \boldsymbol{L}_{412} - \boldsymbol{e}_{022} = \boldsymbol{0} \\ \boldsymbol{L}_{22} + (\boldsymbol{L}_{33\text{-}2} + \boldsymbol{L}_{32\text{-}1}) + \boldsymbol{L}_{411} - \boldsymbol{e}_{021} = \boldsymbol{0} \\ \boldsymbol{L}_{22} + (\boldsymbol{L}_{33\text{-}2} + \boldsymbol{L}_{32\text{-}1} + \boldsymbol{L}_{31\text{-}0}) + \boldsymbol{L}_{410} - \boldsymbol{e}_{020} = \boldsymbol{0} \\ \boldsymbol{L}_{42i} + \boldsymbol{L}_{5i} + \boldsymbol{r}_{1i} - \boldsymbol{e}_{031i} = \boldsymbol{0} \\ \boldsymbol{r}_{2ij} + \boldsymbol{L}_{6Cij} - \boldsymbol{e}_{032ij} = \boldsymbol{0} \end{cases} \tag{2}$$

式中: $\boldsymbol{L}_1$ 为作动筒向量; $\boldsymbol{L}_{21}$ 为外摇臂向量; $\boldsymbol{e}_{01}$ 为外摇臂安装位置相对于作动筒安装位置的向量; $\boldsymbol{L}_{22}$ 为第 3 级与级间连杆连接的曲柄向量, 且 $\boldsymbol{L}_{22}$ 相对于 $\boldsymbol{L}_{21}$ 固定; $\boldsymbol{L}_{412}$、$\boldsymbol{L}_{411}$、$\boldsymbol{L}_{410}$ 分别为第 2 级、第 1 级、第 0 级与级间连杆连接的曲柄向量, 且$-\boldsymbol{L}_{22}=\boldsymbol{L}_{412}=\boldsymbol{L}_{411}=\boldsymbol{L}_{410}$; $\boldsymbol{e}_{022}$、$\boldsymbol{e}_{021}$、$\boldsymbol{e}_{020}$ 分别为第 2 级、第 1 级、第 0 级曲柄安装位置相对于传动轴安装位置的向量, 且向量 $\boldsymbol{e}_{022}$、$\boldsymbol{e}_{021}$ 和 $\boldsymbol{e}_{020}$ 的方向相同; $\boldsymbol{L}_{33\text{-}2}$、$\boldsymbol{L}_{32\text{-}1}$ 和 $\boldsymbol{L}_{31\text{-}0}$ 分别为 3-2 级、2-1 级、1-0 级的级间连杆向量, 且 $\boldsymbol{L}_{33\text{-}2}$、$\boldsymbol{L}_{32\text{-}1}$、$\boldsymbol{L}_{31\text{-}0}$ 与 $\boldsymbol{e}_{022}$、$\boldsymbol{e}_{021}$、$\boldsymbol{e}_{020}$ 的方向相同; i 表示第 i 级

调节机构，i=0,1,2,3；j 表示某级调节机构中的第 j 个静子叶片；$\boldsymbol{L}_{42i}$ 为第 i 级与拉杆连接的曲柄向量，且向量 $\boldsymbol{L}_{42i}$ 相对于第 i 级与级间连杆连接的曲柄向量固定；$\boldsymbol{L}_{5i}$ 为第 i 级拉杆向量；$\boldsymbol{L}_{6Cij}$ 为第 i 级第 j 个摇臂有效长度向量，$\boldsymbol{L}_{6Cij}=\boldsymbol{L}_{6ij}+\boldsymbol{L}_{Cij}$，$\boldsymbol{L}_{6ij}$ 为第 i 级第 j 个摇臂向量，$\boldsymbol{L}_{Cij}$ 为第 i 级第 j 个摇臂-联动环间隙向量；$\boldsymbol{e}_{031i}$ 为第 i 级曲柄的定位向量；$\boldsymbol{r}_{1i}$ 为第 i 级联动环连接拉杆处的位置向量；$\boldsymbol{r}_{2ij}$ 为第 i 级联动环上第 j 个摇臂安装位置向量，且向量 $\boldsymbol{r}_{2ij}$ 相对于 $\boldsymbol{r}_{1i}$ 固定；$\boldsymbol{e}_{032ij}$ 为第 i 级第 j 个叶片的定位向量。

由于考虑运动副间隙，式 (2) 为欠定方程，需要引入附加方程进行求解。如图 2 所示，第 i 级曲柄-第 i 级拉杆-第 i 级联动环-第 i 级第 j 个摇臂-第 i 级第 j 个叶片组成闭环系统。因此，根据静力学分析可得

$$\begin{cases}\boldsymbol{F}_{6ij}\times\boldsymbol{L}_{6ij}-\boldsymbol{T}_{ij}=\boldsymbol{0}\\ \boldsymbol{F}_{6ij}\times\boldsymbol{r}_{2ij}+\boldsymbol{F}_{5i}\times\boldsymbol{r}_{1i}=\boldsymbol{0}\end{cases} \tag{3}$$

式中：$\boldsymbol{T}_{ij}$ 为第 i 级第 j 个叶片所受载荷力矩；$\boldsymbol{F}_{6ij}$ 为第 i 级联动环施加给第 i 级第 j 个摇臂的力；$\boldsymbol{F}_{5i}$ 为第 i 级联动环施加给第 i 级拉杆的力。

式 (2) 中各构件的尺寸和安装位置为已知条件，构件向量的方向为未知量。在式 (3) 中，$\boldsymbol{T}_{ij}$ 已知，$\boldsymbol{F}_{6ij}$ 和 $\boldsymbol{F}_{5i}$ 为未知量，其中，$\boldsymbol{F}_{6ij}$ 与 $\boldsymbol{L}_{Cij}$ 的方向相同，$\boldsymbol{F}_{5i}$ 与 $\boldsymbol{L}_{5i}$ 的方向相同，$\boldsymbol{F}_{5i}$ 和 $\boldsymbol{F}_{6ij}$ 在固定坐标系 $O_Gx_Gy_Gz_G$ 的 x_G 轴方向的分量大小相等方向相反。显然，式 (2) 与式 (3) 联立后，未知量与方程个数相同，系统有确定解。其中，$\boldsymbol{L}_{6ij}$ 在固定坐标系 $O_Gx_Gy_Gz_G$ 的 X_GY_G 平面上的投影与 x_G 轴的夹角为第 i 级第 j 个叶片转角。

在本文研究中，作动筒作匀速直线运动，且运动距离为 45 mm。通过上述方程可以得到各级静子叶片在理想状态下的转角变化规律，如图 4 所示。

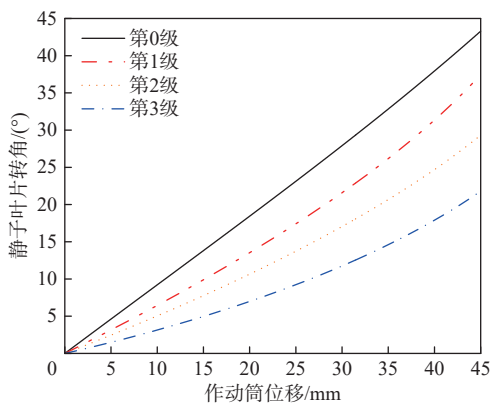


图 4 各级静子叶片转角随作动筒位移的变化

Fig. 4 The variation of stator vane rotation angles at each stage with actuator displacement

2 考虑失效相关性的 VSV 调节机构运动精度可靠性模型

由于 VSV 调节机构是混联机构，且各静子叶片处于同一驱动下，因此，机构系统存在失效相关现象。本文考虑各静子叶片运动精度的失效相关性建立 VSV 调节机构的运动精度可靠性模型。

以 VSV 调节机构的静子叶片转角误差超额为失效模式。以驱动误差、尺寸误差与运动副间隙等作为输入变量 $\boldsymbol{X}$，则输出响应静子叶片转角误差 Δ 为

$$\Delta=|Y(\boldsymbol{X})-Y^*| \tag{4}$$

式中：$Y(\boldsymbol{X})$ 为静子叶片转角；Y^* 为理想状态下的静子叶片转角。

只有静子叶片转角误差小于转角误差阈值才能实现调节功能，否则认为发生失效。因此，该调节机构运动精度失效对应的极限状态函数可表示为

$$g(\boldsymbol{X})=\Delta_0-\Delta \tag{5}$$

式中：Δ_0 为转角误差阈值。

VSV 调节机构的运动精度可靠度为

$$R=P(g(\boldsymbol{X})>0) \tag{6}$$

假设输入变量 x 的概率密度函数为 $h(x)$，输出响应的概率密度函数为 $f_\theta(\Delta,x)$，则基于全概率公式[18]，可靠度为

$$R=\int_a^b h(x)\left[\int_0^{\Delta_0} f_\theta(\Delta,x)\,\mathrm{d}\Delta\right]\mathrm{d}x \tag{7}$$

式中：b、a 分别为输入变量 x 的取值区间上、下限。

由于 VSV 调节机构有多个静子叶片，任意静子叶片的运动精度不足都会导致机构失效，因此可以视为串联系统。机构的各个静子叶片同时受到驱动误差、曲柄尺寸误差、联动环尺寸误差、曲柄 x 方向定位误差、外摇臂尺寸误差等因素的影响，因此，不能假设各静子叶片运动精度失效是相互独立的随机事件进行可靠性建模。而这些同时影响各静子叶片的因素又可以分为 2 类：①同时影响 4 级调节机构中各静子叶片转角的因素，如驱动误差等；②同时且仅影响某级调节机构中各个静子叶片转角的因素，如第 0 级联动环尺寸误差仅影响第 0 级的各静子叶片转角。显然，不考虑 4 级调节机构受共同影响因素作用时，各级静子叶片运动精度失效是相互独立的；更进一步，不考虑某级调节机构中各静子叶片受共同影响因素作用时，同级内各静子叶片运动精度失效也是

相互独立的。

根据上述分析, 基于式 (7) 所示的可靠度模型, 机构系统的运动精度可靠度为

$$R=\int_{a_Q}^{b_Q}h_Q\left(x_Q\right)\cdots\int_{a_q}^{b_q}h_q\left(x_q\right)\cdots\int_{a_1}^{b_1}h_1\left(x_1\right)\cdot\prod_{i=0}^{3}\left\{\int_{a_{Mi}}^{b_{Mi}}h_{Mi}\left(x_{Mi}\right)\cdots\int_{a_{mi}}^{b_{mi}}h_{mi}\left(x_{mi}\right)\cdots\int_{a_{1i}}^{b_{1i}}h_{1i}\left(x_{1i}\right)\cdot\left(\int_{a_{Lij}}^{b_{Lij}}h_{Lij}\left(x_{Lij}\right)\cdots\int_{a_{lij}}^{b_{lij}}h_{lij}\left(x_{lij}\right)\cdots\int_{a_{1ij}}^{b_{1ij}}h_{1ij}\left(x_{1ij}\right)\cdot\int_0^{\Delta_0}f_{\theta_{ij}}\left(\Delta_{ij},x_{1ij},\cdots,x_{lij},\cdots,x_{Lij},x_{1i},\cdots,x_{mi},\cdots,x_{Mi},x_1,\cdots,x_q,\cdots,x_Q\right)\mathrm{d}\Delta_{ij}\cdot\mathrm{d}x_{1ij}\cdots\mathrm{d}x_{lij}\cdots\mathrm{d}x_{Lij}\right)^{k_i}\mathrm{d}x_{1i}\cdots\mathrm{d}x_{mi}\cdots\mathrm{d}x_{Mi}\right\}\mathrm{d}x_1\cdots\mathrm{d}x_q\cdots\mathrm{d}x_Q \tag{8}$$

式中: $f_{\theta ij}(\Delta_{ij}, x_{1ij}, ..., x_{lij}, ..., x_{Lij}, x_{1i}, ..., x_{mi}, ..., x_{Mi}, x_1, ..., x_q, ..., x_Q)$ 为第 i 级第 j 个静子叶片转角误差的概率密度函数, 其中, 同级静子叶片的概率密度函数相同; k_i 为第 i 级静子叶片的个数; $h_{lij}(x_{lij})$ 为第 l 个仅影响第 i 级第 j 个静子叶片的输入变量的概率密度函数; $h_{mi}(x_{mi})$ 为第 m 个同时且仅影响第 i 级调节机构中各静子叶片的输入变量的概率密度函数; $h_q(x_q)$ 为第 q 个同时影响 4 级调节机构中各静子叶片的输入变量的概率密度函数。

式 (8) 表示 VSV 调节机构中任一静子叶片转角误差小于转角误差阈值的概率, 即机构系统的运动精度可靠度。

3　基于代理模型的 VSV 调节机构的运动精度可靠度计算

在第 2 节中建立了考虑失效相关性的 VSV 调节机构运动精度可靠性模型, 但是模型过于复杂, 难以直接计算可靠度。而采用机构的真实物理模型结合 Monte Carlo 方法计算可靠度则耗时过长。因此, 综合考虑计算的准确性与计算效率, 采用 BP 神经网络代理模型结合 Monte Carlo 方法对机构可靠度进行计算。

3.1　BP 神经网络代理模型

人工神经网络 (artificial neural network, ANN) 是一种对生物大脑神经网络的模拟和近似, 是大量神经元通过相互连接构成的自适应非线性动态网络系统[19]。本文采用目前较为成熟的 BP 神经网络[20] 进行建模, 其由输入层、隐藏层和输出层组成, 如图 5 所示。

分析驱动误差、尺寸误差和装配间隙等因素对静子叶片转角的影响, 其中, 影响较大的主要有以

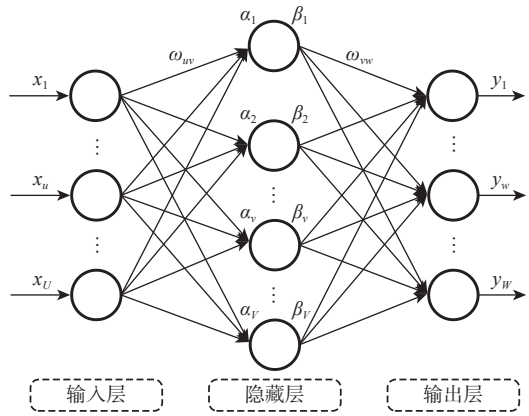


图 5　BP 神经网络的示意图

Fig. 5　Schematic depiction of an BPNN

下参量: 驱动误差、摇臂尺寸误差、各级曲柄尺寸误差、各级联动环尺寸误差、外摇臂尺寸误差、各级曲柄 x 方向定位误差、各级摇臂-联动环运动副间隙。在本文所研究的机构构件参数及设定转角误差阈值 Δ_0=1°的条件下, VSV 调节机构的第 0 级静子叶片在整个运动过程中误差都比其他级大, 且只有其存在精度失效问题, 如图 6 所示。因此, 本文的可靠性分析中仅考虑第 0 级调节机构。

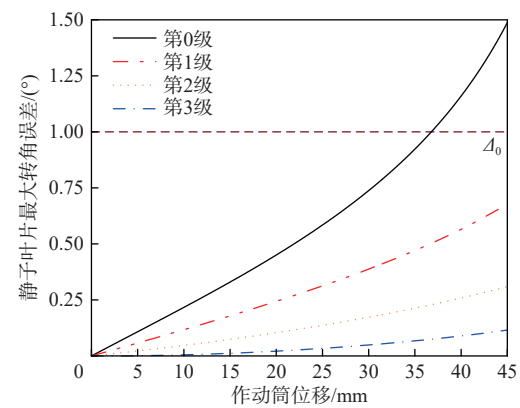


图 6　各级静子叶片最大转角误差随作动筒位移的变化

Fig. 6　The variation of the maximum rotation angle error of stator vanes at each stage with actuator displacement

将对第 0 级调节机构影响较大的第 0 级曲柄尺寸误差等 7 个参量设置为输入变量, 单个静子叶片的转角误差为输出响应, 构建 BP 神经网络以建立输入变量和输出响应之间的关系。这 7 个随机变量服从截尾正态分布, 基于 6σ 原则[12] 计算标准差, 则其概率分布参数如表 1 所示。BP 神经网络代理模型的训练结果如图 7 所示, 预测响应误差的范围为 $[-10^{-5}, 10^{-5}]$, 模型的拟合精度较高。

3.2　VSV 调节机构运动精度可靠度的计算流程

本节基于 BP 神经网络代理模型, 借助 Monte Carlo 技术, 对 VSV 调节机构运动精度可靠度进行

表 1 主要影响因素的分布

Table 1 Distribution of main influencing factors

影响因素	均值/mm	方差/mm^2	下限/mm	上限/mm
驱动误差	0	0.150^2	−0.45	0.45
曲柄尺寸误差	0	0.020^2	−0.06	0.06
联动环尺寸误差	0	0.033^2	−0.1	0.1
外摇臂尺寸误差	0	0.023^2	−0.07	0.07
曲柄x方向定位误差	0	0.333^2	−1	1
摇臂尺寸误差	0	0.050^2	−0.15	0.15
摇臂-联动环运动副间隙	0.09	0.010^2	0.06	0.12

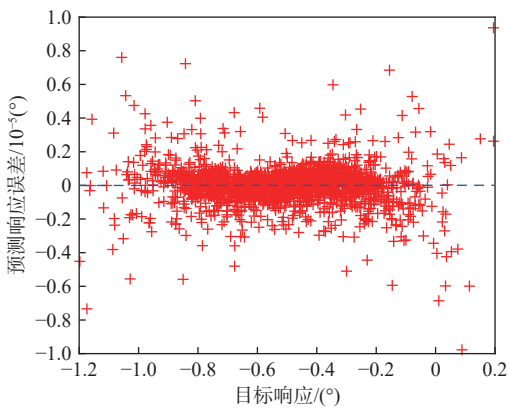


图 7 BP 神经网络代理模型训练结果

Fig. 7 The training results of the BPNN surrogate model

计算，流程如图 8 所示。

具体步骤包括以下 4 步：

1）生成样本池。根据表 1 所示的输入变量的分布生成样本池，该样本池由 N 组样本 $\boldsymbol{X}_n$ 组成 (N=500 000)，每组样本中有 42 个摇臂尺寸误差和摇臂-联动环运动副间隙，对应 42 个静子叶片。

2）计算功能函数值。调用 3.1 节中构建好的 BP 神经网络代理模型计算输出响应 $\boldsymbol{\Delta}_n$，即第 0 级各个静子叶片的转角误差；然后计算功能函数值 $g(\boldsymbol{X}_n)$。

3）计算失效概率及其变异系数。根据 $g(\boldsymbol{X}_n)$ 的符号判断样本 $\boldsymbol{X}_n$ 是否落入失效域内，然后根据 Monte Carlo 方法计算失效概率估计值及其变异系数。

4）计算可靠度。当变异系数小于 5% 时，认为失效概率的估计值可以接受，此时计算可靠度；否则，更新 Monte Carlo 样本池，返回第 1 步。

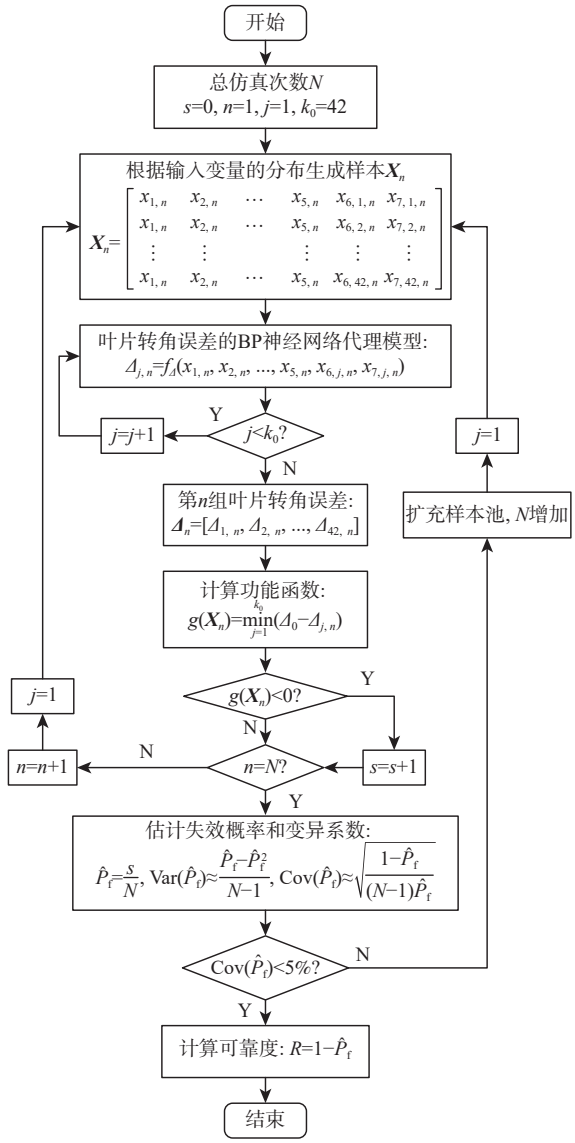


图 8 基于 BP 神经网络的 VSV 调节机构可靠度计算流程

Fig. 8 Reliability calculation process of VSV adjustment mechanism based on BPNN

4 可靠度的计算结果及误差分析

4.1 可靠度的计算结果

本文针对 VSV 调节机构运动精度失效问题，设定失效判据为：任意静子叶片的转角误差超过 1°。统计每组样本中最大的静子叶片转角误差，获得如图 9 所示的 VSV 调节机构转角误差频率分布直方图，由统计结果可知，机构的系统可靠度 R 为 92.25%。图 10 给出了第 0 级某个静子叶片转角误差的频率分布直方图，第 0 级其他静子叶片的转角误差统计结果与其相同。当各静子叶片运动精度失效完全相关时，可以使用单个静子叶片的可靠度表示机构的系统可靠度，根据 Monte Carlo 仿真结果，单个叶片的可靠度 R_θ 为 98.86%。当不考虑各静子叶片运动精度失效之间的相关性时，机构系统是单元失效相互独立且同分布的串联系统，则系统可靠度 $R=R_\theta^{k_0}$，通过计算可靠度 R 为 61.78%。

由上述分析可知，VSV 调节机构系统存在失效相关现象，验证了第 2 节中所构建的考虑失效相关性的机构运动精度可靠性模型的合理性。因此，在

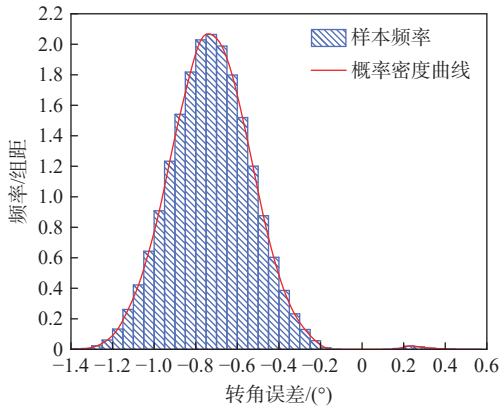


图 9　VSV 调节机构转角误差频率分布直方图

Fig. 9　The frequency distribution histogram of rotation angle error of VSV adjustment mechanism

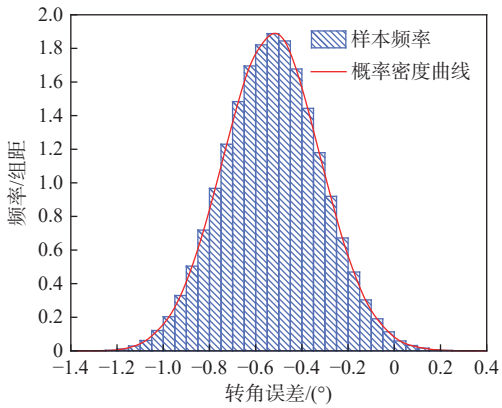


图 10　某个静子叶片转角误差频率分布直方图

Fig. 10　The frequency distribution histogram of a certain stator vane rotation angle error

研究 VSV 调节机构系统可靠性问题时, 不能直接使用完全相关或单元失效独立假设的系统模型的可靠度作为系统可靠度。

如图 11 所示, 在失效阈值 Δ_0=1°的条件下对比分析失效完全相关、失效相关和单元失效独立假设的系统模型可以看出, 静子叶片数量由 1 增至 42 个时, 失效完全相关系统模型的可靠度保持不变, 单元失效独立假设系统模型和考虑失效相关系统模型的可靠度随着静子叶片数量的增加而降低。当系统内叶片的数量不大于 2 时, 3 种模型计算得到的系统可靠度相差不多, 但当叶片数量大于 2 时, 失效完全相关系统模型的可靠度明显大于考虑失效相关性的系统模型可靠度, 单元失效独立假设系统模型的可靠度明显小于考虑失效相关性的系统可靠度, 且 3 种模型可靠度之间的差值随着静子叶片数量的增加而增加。由此可知, 在计算 VSV 调节机构的系统可靠度时, 若采用单元失效独立假设系统模型, 所得结果过于保守; 而采用失效完全相关系统模型, 则可能导致结果过于激进。特别是在静子叶片数量较多时, 应充分考虑失效相关性, 以实现更为准确的可靠度评估。

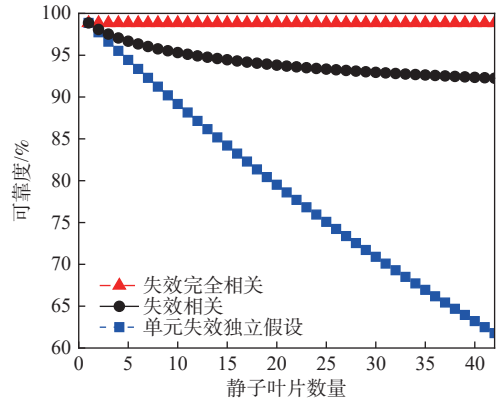


图 11　系统可靠度与静子叶片数量之间的关系

Fig. 11　The relationship between system reliability and the number of stator vanes

VSV 调节机构在失效完全相关、失效相关及单元失效独立假设 3 种系统模型下的可靠度随失效阈值的变化规律如图 12 所示。随着失效阈值 Δ_0 从 0°增加至 1.3°, 3 种模型的可靠度也逐渐从 0 增加至 100%, 表明机构规定转角误差阈值越大, 即对运动精度要求越低, 机构可靠性越高。其中, 失效完全相关系统模型的可靠度始终大于考虑失效相关的系统可靠度, 单元失效独立假设的系统可靠度始终小于考虑失效相关的系统可靠度。因此, 在不同失效阈值下, 前文中得到的针对 VSV 调节机构系统可靠度的结论依然成立。

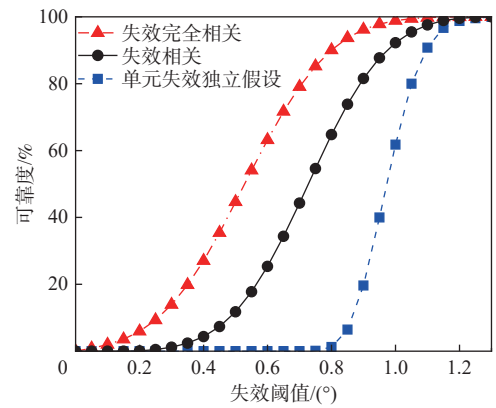


图 12　系统可靠度与失效判据之间的关系

Fig. 12　The relationship between system reliability and failure threshold

4.2　可靠度的计算误差分析

使用基于代理模型的 Monte Carlo 方法计算可靠度时, 计算结果与理论结果之间的误差主要来源于 2 个方面: ①使用 Monte Carlo 方法得到的可靠度估计值和可靠度真实值的误差; ②代理模型和理论模型之间的误差所引起的可靠度计算误差。在随机变量的样本分布确定的情况下, 使用 Monte Carlo 方法得到的可靠度估计值的精度与随机变量的样

本个数 N 有关，只要 N 足够大，就能得到高精度的可靠度估计值。因此，本节主要讨论使用代理模型造成的可靠度误差。

在可靠性分析中，关键是要得到性能响应量的统计规律，而响应量的统计规律是由输入变量的统计规律通过输入变量-性能响应量之间的函数关系传递得到的[21]。如图 13 所示，采用代理模型近似输入变量-性能响应量之间的函数关系（即理论模型）时，得到的响应与真实响应之间存在误差。而该误差会影响由代理模型得到的响应的统计规律，导致代理模型响应和理论模型响应的概率密度函数有一定的差异，如图 14 所示。其中，x 为输入变量，y 为输出响应，$f(y)$ 为响应的概率密度函数。

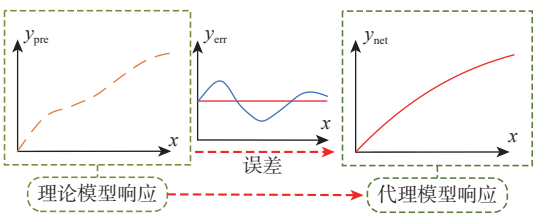


图 13 理论模型响应和代理模型响应的关系

Fig. 13 Relationship between theoretical and surrogate model response

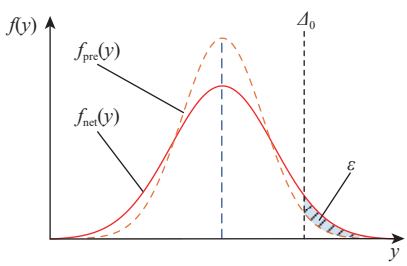


图 14 可靠度计算误差

Fig. 14 Calculation errors of reliability

此时，代理模型与理论模型之间的误差造成的可靠度的计算误差 ε 可以用两者响应的概率密度函数在失效域内差值的积分表示。其中，失效域定义为 $F=\{x: y(x)-\Delta_0>0\}$。

则可靠度误差 ε 为

$$\varepsilon = \left| \int_{\Delta_0}^{+\infty} \left(f_{\text{pre}}(y) - f_{\text{net}}(y) \right) \mathrm{d}y \right| \tag{9}$$

式 (9) 是目前文献中[22-25]常用的分析可靠度误差的方法，即采用基于理论模型的 Monte Carlo 方法作为参考解来检验其他新方法的计算精度。然而，在实际使用中，针对一些复杂的机构，难以在较为经济的时间内通过理论模型获得满足 Monte Carlo 仿真计算所需的大量样本[26]。同时，由于没有严格的方法来分析不确定性在代理模型中的传递[27-28]，很难甚至无法准确地量化代理模型误差所造成的可靠度误差，导致通过代理模型计算得到的可靠度不完全可信[29]。针对这种情况，提出一种基于已有信息估算代理模型误差造成的复杂机构系统可靠度误差的方法。

通过对比所构建的代理模型和理论模型的响应值，可以得到相应的误差，其中，绝对值最大的误差值记为 $\delta_{\max}$。由此认为，代理模型在 $[\Delta_0-\delta_{\max}, \Delta_0+\delta_{\max}]$ 区间内响应的误差可能会影响对该区间内样本是否失效的判断，导致可靠度误差，如图 15 所示。因此，代理模型响应的概率密度函数在此区间的积分可近似表示由代理模型误差导致的可靠度误差 ε：

$$\varepsilon \approx \varepsilon_{\text{net}} = \int_{\Delta_0-\delta_{\max}}^{\Delta_0+\delta_{\max}} f_{\text{net}}(y)\mathrm{d}y \tag{10}$$

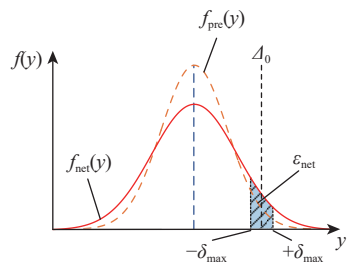


图 15 可靠度误差估算示意图

Fig. 15 Diagram of error estimation of reliability

基于代理模型，借助 Monte Carlo 技术计算可靠度时，可以认为在 $[\Delta_0-\delta_{\max}, \Delta_0+\delta_{\max}]$ 区间内的响应样本点存在的误差可能会影响到对该样本点是否失效的判断，导致可靠度误差。此时的 ε_{net} 可以由 $\varepsilon_{\text{sample}}$ 表示：

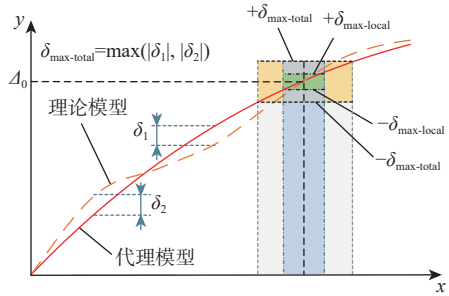
$$\varepsilon_{\text{net}} \approx \varepsilon_{\text{sample}} = n_{\text{error}}/N \tag{11}$$

式中：n_{error} 为处于 $[\Delta_0-\delta_{\max}, \Delta_0+\delta_{\max}]$ 区间内的样本点个数。

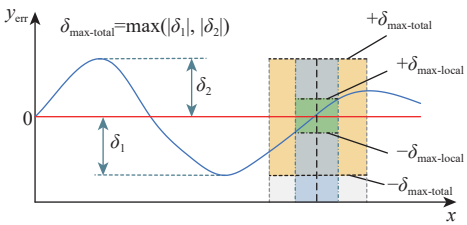
使用式 (11) 估算可靠度误差，关键是确定代理模型和理论模型的响应误差区间，即绝对值最大的误差值 $\delta_{\max}$。通常，可以根据用于训练代理模型的样本点的真实响应值和代理模型响应值计算得到误差区间 $[\Delta_0-\delta_{\text{max-total}}, \Delta_0+\delta_{\text{max-total}}]$。但此区间是代理模型在整个输入变量空间内的输出响应误差，这比实际会影响可靠度计算结果的失效阈值 Δ_0 附近的局部输出响应误差区间 $[\Delta_0-\delta_{\text{max-local}}, \Delta_0+\delta_{\text{max-local}}]$ 大，从而导致估算的可靠度误差偏大，如图 16 所示。因此，在估算可靠度误差时，需要根据已知的训练集样本信息确定代理模型在 Δ_0 附近的局部输出响应误差区间。

根据已知的训练集样本信息计算代理模型在 Δ_0 附近的输出响应误差区间，进而估算可靠度计算误差的流程如图 17 所示，具体步骤包括以下 4 步：

1）通过训练样本点的真实响应值和代理模型



(a) 输入变量-性能响应量之间的函数关系示意图



(b) 代理模型响应误差示意图

图 16　全局与局部输出响应误差区间对比

Fig. 16　Comparison of global and local output response error interval

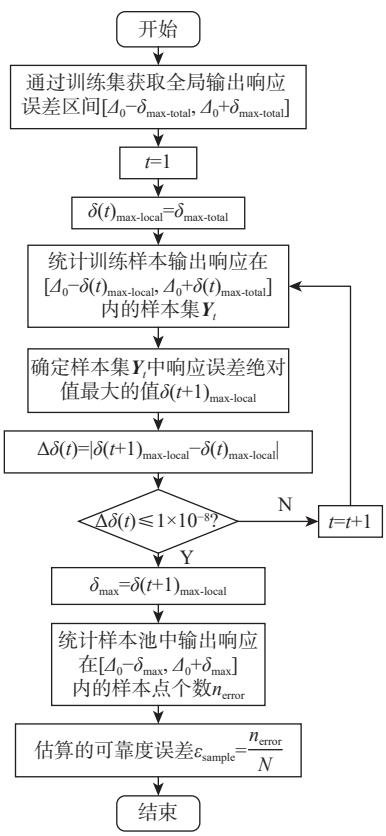


图 17　可靠度计算误差分析流程

Fig. 17　Reliability calculation error analysis process

响应值计算得到全局输出响应误差区间, 定义此区间作为初始误差区间。

2) 统计响应在 $[\Delta_0-\delta(t)_{\text{max-local}}, \Delta_0+\delta(t)_{\text{max-local}}]$ 区间内的训练集样本点的输出响应误差, 确定绝对值最大的误差值记为 $\delta(t+1)_{\text{max-local}}$, 获得新的误差区间 $[\Delta_0-\delta(t+1)_{\text{max-local}}, \Delta_0+\delta(t+1)_{\text{max-local}}]$。

3) 判断第 2 步中得到的误差区间是否满足设定条件, 若满足则跳至第 4 步; 否则, 返回第 2 步进行下一次迭代计算, 直至满足设定要求。记满足设定要求的区间为 $[\Delta_0-\delta_{\max}, \Delta_0+\delta_{\max}]$。

4) 统计用于可靠度计算的样本池内代理模型响应值在区间 $[\Delta_0-\delta_{\max}, \Delta_0+\delta_{\max}]$ 内的样本点个数 n_{error}, 使用式 (11) 计算可靠度误差。

如果需要获得更精确的误差估计, 可以扩充样本池和进行更多的迭代计算, 这需要根据实际情况进行取舍。由于本文方法认为样本池内响应值在 $[\Delta_0-\delta_{\max}, \Delta_0+\delta_{\max}]$ 内的样本点都受代理模型误差的影响导致失效判断错误, 在使用式 (11) 估算可靠度误差时将其都计入 n_{error}, 从而导致估算的误差结果会偏大。

4.3　基于代理模型的 VSV 调节机构运动精度可靠度误差分析

表 2 给出了失效阈值 Δ_0 为 1°时, 不同代理模型精度下 VSV 调节机构运动精度可靠度计算误差的变化规律。本文在讨论可靠度相关的数值时, 仅保留百分数的小数点后 2 位。其中, 基于式 (9) 采用 Monte Carlo 方法计算可靠度误差 ε。

表 2　不同代理模型精度下的可靠度误差

Table 2　Reliability error under different accuracy of the surrogate model

训练次数	$\delta_{\text{max-total}}$	n_{error}	$\varepsilon_{\text{sample}}$/%	ε/%
112	1×10^{-3}	685	0.14	0.12
366	5×10^{-4}	284	0.06	0.04
892	1×10^{-4}	90	0.02	0.01
3 585	5×10^{-5}	53	0.01	0.01
8 936	1×10^{-5}	23	0.00	0.00
18 626	5×10^{-6}	7	0.00	0.00

如表 2 所示, 当 BP 神经网络的训练次数增加, 即代理模型的精度提高时, 由代理模型误差导致的可靠度误差 ε 逐渐减小; 而相应的由训练样本点的真实响应值和代理模型响应值计算得到输出响应误差区间也在减小, 落入此区间内的样本个数 n_{error} 减少, 因而本文方法所得的可靠度误差估计值 $\varepsilon_{\text{sample}}$ 逐渐减小, 与 ε 呈现相同的趋势, 这在一定程度上表明了本文估算可靠度误差方法的合理性。

当代理模型精度较高时, 本文方法所估算的可靠度误差与实际可靠度误差数值非常接近, 精度较高。当 $\delta_{\text{max-total}}=1\times10^{-5}$ 时, 可靠度误差值百分数的小数点后两位都为 0, 因此, 本文基于 BP 神经网络代理模型计算得到的可靠度 R=92.25% 是准确可信的。还可以发现, 本文方法所估算的 $\varepsilon_{\text{sample}}$ 始终大

于或等于 ε，这也符合 4.2 节中的分析，即本文提出的估算代理模型误差造成的可靠度误差的方法偏保守，这在一定程度上说明了方法的泛用性。

表 3 讨论了在 $\delta_{\text{max-total}}=1\times10^{-5}$ 的情况下，不同失效阈值下可靠度误差的变化规律。随着失效阈值 Δ_0 的增大，可靠度 R 逐渐增大，可靠度误差 ε 逐渐减小；而样本池中落入误差区间的样本个数 n_{error} 也在不断减少，因此，本文方法所得的可靠度误差估计值 $\varepsilon_{\text{sample}}$ 也在逐渐减小。$\varepsilon_{\text{sample}}$ 与 ε 的变化趋势一致。这符合式 (10) 的原理，即随着失效阈值的增大，表示可靠度误差的积分面积减小，可靠度误差减小。为了减小低可靠度时的计算误差，可以适当提高代理模型的精度。同时，由表 3 可知，本文所提的可靠度误差估算方法偏保守，符合前文所述。

表 3 不同失效阈值下的可靠度误差

Table 3 Reliability error under different failure threshold

Δ_0/(°)	R/%	n_{error}	$\varepsilon_{\text{sample}}$/%	ε/%
0.75	54.61	292	0.06	0.05
0.80	64.76	203	0.04	0.04
0.85	73.84	161	0.03	0.03
0.90	81.55	103	0.02	0.02
0.95	87.71	60	0.01	0.01
1.00	92.25	23	0.00	0.00
1.05	95.47	11	0.00	0.00
1.10	97.58	0	0.00	0.00

通过本节对可靠度误差的分析，可以证明 3.1 节中所构建的 BP 神经网络代理模型精度满足可靠度计算要求，4.1 节中的可靠度计算结果是准确可信的。

5 结 论

1）采用运动学和静力学理论基于连续接触模型建立了含运动副间隙的 VSV 调节机构模型。在此基础上，进一步考虑驱动误差、尺寸误差和装配间隙等因素的随机不确定性，以及各静子叶片运动精度失效的相关性，基于全概率公式构建了 VSV 调节机构系统运动精度可靠性模型。

2）基于 BP 神经网络代理模型，借助 Monte Carlo 方法对 VSV 调节机构进行了可靠性分析。分析结果表明，机构系统可靠度为 92.25%，存在失效相关性。由于各静子叶片运动精度失效相关，采用单元失效独立假设的系统模型计算得到的机构系统可靠度过于保守，会增加机构在设计、制造、装配过程中的成本；而采用失效完全相关系统模型得到的机构系统可靠度过于激进，会低估由于机构运动精度失效而导致的发动机出现喘振的风险。

3）提出了一种估算代理模型误差所引起的可靠度误差的方法，证明了基于 BP 神经网络代理模型的可靠度计算结果是准确可信的。该方法具有较高的效率和精度。

4）本文提出的可靠度误差估计方法偏保守、适用性较强，可推广至基于其他代理模型的复杂系统可靠度分析中，即可在无法获得基于理论模型的 Monte Carlo 结果作为对照基准的情况下，验证基于代理模型的可靠度结果的准确性。

参考文献（References）

[1] 王云. 航空发动机原理[M]. 北京: 北京航空航天大学出版社, 2009: 85-99.
WANG Y. Aeroengine principle[M]. Beijing: Beihang University Press, 2009: 85-99(in Chinese).

[2] 曹志鹏, 刘波, 丁伟. 静叶角度调节对组合压气机性能优化机理[J]. 北京航空航天大学学报, 2007, 33(8): 878-881.
CAO Z P, LIU B, DING W. Stator setting angles adjustment on performance improvement of axial-centrifugal compressor[J]. Journal of Beijing University of Aeronautics and Astronautics, 2007, 33(8): 878-881(in Chinese).

[3] 隋岩峰, 邴连喜, 施磊, 等. 某发动机高压压气机进口可调静子叶片角度控制方法[J]. 航空动力学报, 2010, 25(11): 2589-2593.
SUI Y F, BING L X, SHI L, et al. Control method of compressor stator vane angle for an aero-engine[J]. Journal of Aerospace Power, 2010, 25(11): 2589-2593(in Chinese).

[4] 于嘉鹏, 孙加明, 纪福森, 等. 航空发动机静叶联调机构运动分析及优化[J]. 航空动力学报, 2019, 34(6): 1193-1200.
YU J P, SUN J M, JI F S, et al. Motion analysis and optimization of jointly adjusting mechanism of aero-engine stator vane[J]. Journal of Aerospace Power, 2019, 34(6): 1193-1200(in Chinese).

[5] 张哲, 王汉平, 孙浩然, 等. VSV 调节机构阻滞力和调节精度的归因分析[J]. 航空学报, 2020, 41(12): 380-390.
ZHANG Z, WANG H P, SUN H R, et al. Attribution analysis of blocking force and adjustment accuracy of adjusting mechanism of variable stator vane[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(12): 380-390(in Chinese).

[6] 高华, 翟敬宇, 张昊, 等. 考虑转动副特性的发动机静叶调节机构动态性能分析[J]. 动力学与控制学报, 2022, 20(1): 51-59.
GAO H, ZHAI J Y, ZHANG H, et al. Dynamic performance analysis of engine VSV adjustment mechanism considering the characteristics of revolute joint[J]. Journal of Dynamics and Control, 2022, 20(1): 51-59(in Chinese).

[7] 唐佑远, 郭为忠. 静叶调节机构尺度全局优化设计方法研究[J]. 机械工程学报, 2020, 56(11): 26-36.
TANG Y Y, GUO W Z. Global dimensional optimization for the design of adjusting mechanism of variable stator vanes[J]. Journal of Mechanical Engineering, 2020, 56(11): 26-36(in Chinese).

[8] 庞欢. 飞机典型机构运动功能可靠性建模方法及应用[D]. 西安: 西北工业大学, 2016: 13-20.
PANG H. Modeling method and application of motion function reliability of typical aircraft mechanisms[D]. Xi'an: Northwestern Polytechnical University, 2016: 13-20(in Chinese).

[9] 李铭, 谢里阳, 丁丽君. 行星机构的可靠性分析与计算[J]. 航空学报, 2017, 38(8): 421145.
LI M, XIE L Y, DING L J. Reliability analysis and calculation for planetary mechanism[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(8): 421145(in Chinese).

[10] 周长聪, 吉梦瑶, 张屹尚, 等. 多失效模式下起落架机构可靠性及灵敏度研究[J]. 西北工业大学学报, 2021, 39(1): 46-54.
ZHOU C C, JI M Y, ZHANG Y S, et al. Mechanism reliability and sensitivity analysis of landing gear under multiple failure modes[J]. Journal of Northwestern Polytechnical University, 2021, 39(1): 46-54(in Chinese).

[11] 张政, 王攀, 周瀚渊. 含区间分布参数的尾喷管调节机构可靠性分析[J]. 北京航空航天大学学报, 2023, 49(12): 3377-3385.
ZHANG Z, WANG P, ZHOU H Y. Reliability analysis of nozzle adjustment mechanism with interval distribution parameters[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(12): 3377-3385(in Chinese).

[12] LI X Y, CHEN W B, KANG R. Performance margin-based reliability analysis for aircraft lock mechanism considering multi-source uncertainties and wear[J]. Reliability Engineering & System Safety, 2021, 205: 107234.

[13] ZENG C D, QIU Z C, ZHANG F H, et al. Error modelling and motion reliability analysis of a multi-DOF redundant parallel mechanism with hybrid uncertainties[J]. Reliability Engineering & System Safety, 2023, 235: 109259.

[14] ZHAN Z H, ZHANG X M, JIAN Z C, et al. Error modelling and motion reliability analysis of a planar parallel manipulator with multiple uncertainties[J]. Mechanism and Machine Theory, 2018, 124: 55-72.

[15] 王博文, 谢里阳, 樊富友, 等. 考虑共因失效的折叠翼展开机构可靠度分析[J]. 机械工程学报, 2020, 56(5): 161-171.
WANG B W, XIE L Y, FAN F Y, et al. Reliability analysis of folding wing deployable mechanism considering common cause failure [J]. Journal of Mechanical Engineering, 2020, 56(5): 161-171(in Chinese).

[16] 刘胜利, 王兴东, 孔建益, 等. 多源不确定性下平面变胞机构运动可靠性分析[J]. 机械工程学报, 2021, 57(17): 64-75.
LIU S L, WANG X D, KONG J Y, et al. Kinematic reliability analysis of planar metamorphic mechanism with multi-source uncertainties[J]. Journal of Mechanical Engineering, 2021, 57(17): 64-75(in Chinese).

[17] 李媛媛. 考虑铰链间隙的太阳翼刚柔耦合多体动力学特性研究[D]. 哈尔滨: 哈尔滨工业大学, 2019: 6-8.
LI Y Y. Study on rigid-flexible coupling multi-body dynamic characteristics of solar wing considering hinge clearance[D]. Harbin: Harbin Institute of Technology, 2019: 6-8(in Chinese).

[18] 谢里阳, 王正, 周金宇, 等. 机械可靠性基本理论与方法[M]. 2 版. 北京: 科学出版社, 2012: 120-128.
XIE L Y, WANG Z, ZHOU J Y, et al. Basic theory and method of mechanical reliability[M]. 2nd ed. Beijing: Science Press, 2012: 120-128(in Chinese).

[19] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
ZHOU F Y, JIN L P, DONG J. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251(in Chinese).

[20] CHEN Y F, HOU F H, DONG S H, et al. Reliability evaluation of corroded pipeline under combined loadings based on back propagation neural network method[J]. Ocean Engineering, 2022, 262: 111910.

[21] 吕震宙, 宋述芳, 李璐祎, 等. 结构/机构可靠性设计基础[M]. 西安: 西北工业大学出版社, 2019: 1-4.
LV Z Z, SONG S F, LI L Y, et al. Fundamentals of structure/mechanism reliability design[M]. Xi'an: Northwestern Polytechnical University Press, 2019: 1-4(in Chinese).

[22] 常泽明, 李璐祎. 基于 NARX 和 Kriging 的时变可靠性分析双层代理模型[J]. 北京航空航天大学学报, 2023, 49(7): 1802-1812.
CHANG Z M, LI L Y. Double-loop surrogate model for time-dependent reliability analysis based on NARX and Kriging models[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(7): 1802-1812(in Chinese).

[23] 智鹏鹏, 汪忠来, 李永华, 等. 基于 RMQGS-APS-Kriging 的主动学习结构可靠性分析方法[J]. 机械工程学报, 2022, 58(16): 420-429.
ZHI P P, WANG Z L, LI Y H, et al. RMQGS-APS-Kriging-based active learning structural reliability analysis method[J]. Journal of Mechanical Engineering, 2022, 58(16): 420-429(in Chinese).

[24] ECHARD B, GAYTON N, LEMAIRE M. AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation[J]. Structural Safety, 2011, 33(2): 145-154.

[25] ZHANG X F, WANG L, SØRENSEN J D. REIF: a novel active-learning function toward adaptive Kriging surrogate models for structural reliability analysis[J]. Reliability Engineering & System Safety, 2019, 185: 440-454.

[26] SU G S, PENG L F, HU L H. A Gaussian process-based dynamic surrogate model for complex engineering structural reliability analysis[J]. Structural Safety, 2017, 68: 97-109.

[27] LI M Y, WANG Z Q. Surrogate model uncertainty quantification for reliability-based design optimization[J]. Reliability Engineering & System Safety, 2019, 192: 106432.

[28] LI L X, CHEN G H, FANG M X, et al. Reliability analysis of structures with multimodal distributions based on direct probability integral method[J]. Reliability Engineering & System Safety, 2021, 215: 107885.

[29] DUBOURG V, SUDRET B, DEHEEGER F. Metamodel-based importance sampling for structural reliability analysis[J]. Probabilistic Engineering Mechanics, 2013, 33: 47-57.

Calculation and error analysis of kinematic accuracy reliability of VSV adjustment mechanism

YANG Yifeng[1, 2], WANG Yi[3, 4, *], LIU Aoyu[1, 2], LI Jia[1], XIE Liyang[1, 2]

(1. School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China;
2. Key Laboratory of Vibration and Control of Aero-Propulsion Systems of Ministry of Education, Northeastern University, Shenyang 110819, China;
3. AECC Shenyang Engine Research Institute, Shenyang 110015, China;
4. Research Institute of Aero-Engine, Beihang University, Beijing 102206, China)

Abstract: A kinematic model considering joint clearance is developed for the variable stator vane (VSV) adjustment mechanism of an aero-engine. A reliability model of the mechanism's kinematic accuracy is constructed considering the failure correlation among stator vanes, as well as the stochastic nature of driving errors, dimensional tolerances, and assembly clearances. Based on the BP neural network (BPNN) surrogate model, the mechanism's reliability is evaluated through Monte Carlo simulation. The effects of the number of stator vanes and the failure threshold on the system's reliability are investigated. Compared with the system model of the unit failure independent assumption and the complete correlation, the mechanism system kinematic accuracy reliability model considering failure correlation is more reasonable. Lastly, a method for estimating the complex system's reliability error caused by surrogate model approximation is proposed, and the accuracy and credibility of the VSV adjustment mechanism's reliability assessment are confirmed.

Keywords: variable stator vane adjustment mechanism; failure correlation; BP neural network; reliability; error analysis

Received: 2023-06-24; **Accepted**: 2023-08-29; **Published Online**: 2023-09-06 08:44
URL: link.cnki.net/urlid/11.2625.V.20230905.0952.002
Foundation item: National Science and Technology Major Project (J2019-Ⅳ-0002-0069)
* **Corresponding author**. E-mail: diyuanyinian@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0797

多元件相关的列车牵引系统可靠性评估与寿命预测

田贵双，王少萍*，石健
(北京航空航天大学 自动化科学与电气工程学院，北京 100191)

摘　　要：牵引系统作为城市轨道交通列车的核心，其可靠性对保障列车的运行安全具有重要意义。针对牵引系统结构复杂且失效模式多样的问题，开展可靠性评估与寿命预测研究。着重围绕牵引系统中的关键元件：牵引电机和绝缘栅双极晶体管(IGBT)，构建牵引电机退磁故障和IGBT键合线失效的性能退化模型，采用融合失效机理的维纳过程描述2个元件的性能退化过程，并采用Copula函数描述两者的相关性。对于离线可靠性评估，采用贝叶斯马尔可夫链蒙特卡罗方法进行未知参数估计；对于在线剩余使用寿命预测，采用贝叶斯与期望最大相融合的算法更新模型中的未知参数。基于牵引系统的性能退化试验数据，验证所提模型和算法，结果表明：考虑牵引电机和IGBT 2个元件相关的可靠性模型能够精准实现可靠性评估，采用贝叶斯与期望最大相融合的参数更新算法可有效提升寿命预测精度。

关　键　词：列车牵引系统；维纳过程；Copula函数；可靠性；寿命预测
中图分类号：TP202+.1
文献标志码：A　　**文章编号**：1001-5965(2025)06-2081-10

城市轨道交通车辆牵引系统的主要作用是完成列车的牵引、制动和控制，对列车的安全可靠运行至关重要。牵引系统一旦发生故障，不仅会造成乘客出行的延误，甚至危及乘客安全。为了确保城市轨道交通牵引系统在运行过程中万无一失，就必须掌握其退化机理和剩余使用寿命，提高系统的可靠性、安全性和综合保障性。

据统计，制约牵引系统可靠性的关键元件是牵引电机和牵引逆变器[1]，一旦牵引电机或其逆变器发生性能退化，将影响整车动力性能，甚至影响行车安全。近年来，随着中国城市轨道交通的迅速发展，越来越多的学者关注其牵引系统的可靠性和寿命研究。马伊光[2]针对高转矩密度电机的失效机理和电机内部应力场对电机可靠性的影响，构建了多场耦合作用下的永磁体和绕组绝缘失效的可靠性模型。Guo等[3]提出了基于双环数字孪生架构的永磁同步电机全生命周期健康管理模型。唐春萍[4]以CRH2型动车组牵引电机为研究对象，根据异步电机驱动系统中发生不同类型复合故障时的数据特点，提出了复合故障诊断方案。对于牵引逆变器而言，绝缘栅双极晶体管(insulated gate bipolar transistor, IGBT)是其核心装置，具备开关速度快、驱动功率低、密度载流大等优势。然而，IGBT长期处于恶劣的外部工作环境，并承受着电热效应的内部持续作用，这使其成为牵引逆变器中最易发生故障的器件。因此，针对IGBT进行失效机理分析并建立模型以准确掌握寿命状态，有利于降低列车维护成本，提升列车运行可靠度。罗俊[5]针对IGBT的键合线断裂故障，从有限元仿真和实验测试2个方面进行了热形变场特性分析，并根据所获取的热

收稿日期：2023-12-09；**录用日期**：2023-12-30；**网络出版时间**：2024-01-10 14:13
网络出版地址：link.cnki.net/urlid/11.2625.V.20240110.1019.002
基金项目：北京市自然科学基金-丰台轨道交通前沿研究联合基金(L221008)
*通信作者. E-mail：shaopingwang@vip.sina.com

引用格式：*田贵双，王少萍，石健. 多元件相关的列车牵引系统可靠性评估与寿命预测[J]. 北京航空航天大学学报，2025，51(6)：2081-2090. TIAN G S，WANG S P，SHI J. Reliability assessment and lifetime prediction for train traction system considering multiple dependent components[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2081-2090 (in Chinese).*

形变场分布规律开展了基于热形变场信号的故障诊断应用。杨德浩[6]提出了一种IGBT的结温数值计算方法，以便引入优化算法来提升牵引逆变器的可靠性。在不同应力条件下，Yang等[7]利用断裂力学理论，提出了基于应变强度因子的键合线热疲劳寿命预测方法。目前，针对牵引电机和IGBT涌现出不少研究成果，但未考虑复杂工况下牵引电机和IGBT同时出现失效模式的情况，且退化过程中存在随机性和不确定性，这使得可靠性分析问题更具挑战性。

基于随机过程的元件性能退化建模是以随机过程和数理统计理论为基础，融合元件在退化过程中存在的随机不确定性，得到元件寿命的概率分布，并量化可靠性评估和寿命预测中的不确定性[8]。Yang等[9]分析了牵引电机的失效机理，将稳态电流变化的绝对值作为性能退化指标，采用维纳过程描述了电机的性能退化过程。朱琦[10]针对电机系统中母线电容建立了基于随机过程的性能退化模型。曾东[11]以维纳过程描述了IGBT饱和压降为性能退化指标的退化过程。王宇琦[12]通过5种随机过程模型对IGBT老化性能退化建模，并采用参数估计方法完成了分布模型的表征，进而模拟出IGBT老化过程。

上述研究工作大多聚焦于牵引电机或IGBT单一性能指标的退化建模，但在实际工程应用中，往往存在多个性能退化指标可以反映牵引系统的性能退化过程。在系统退化建模时考虑多个性能退化指标，有助于提高可靠性评估的精度。Copula函数凭借其简洁结构和较高计算效率，成为系统多失效模式耦合建模的有效工具。刘小平等[13]利用基于随机效应的维纳过程及Copula函数建立了齿轮泵容积效率和总效率相关的退化过程模型，并借助Bootstrap抽样方法实现了个体差异参数的再抽样。王少萍等[14]利用Copula函数揭示了飞机液压泵转子-配流盘摩擦副和轴尾密封副的摩擦磨损之间的相关关系。Zhang等[15]针对唇形密封的2种失效模式进行相关性分析，通过多项式时变Copula函数描述了两者相关关系，并根据赤池信息量准则选择出最佳Copula函数和最佳时变模型的阶次，以防止过拟合现象的发生。Chen等[16]将Copula函数与层次贝叶斯网络相融合，建立了考虑系统中多元件相关的系统可靠性模型。

为了准确描述牵引系统多个故障之间的相关性，本文从牵引系统2个关键元件—牵引电机和IGBT失效机理出发，建立牵引电机退磁故障和IGBT键合线失效的性能退化模型。首先，采用维纳过程描述2个健康因子的性能退化过程，并采用Gaussian Copula函数描述2个健康因子之间的相关性。然后，针对牵引系统可靠性模型中未知参数，通过贝叶斯与期望最大(expectation maximum, EM)相融合的算法进行实时更新，从而实现牵引系统剩余使用寿命的精准预测。最后，基于本文所提模型开展牵引系统性能退化试验，以验证其有效性。

1 牵引系统失效机理分析

1.1 牵引系统工作原理

列车牵引系统包括受电弓、高速断路器、牵引控制单元、牵引电机和IGBT等，具体结构如图1所示。在牵引电机驱动系统中，将单相交流电输入到牵引变流器(由脉冲整流器、中间直流环和牵引逆变器组成)，通过控制器对IGBT进行控制，得到电压和频率都可以调节的三相交流电，再输入到牵引电机从而控制齿轮与轮对的运转以牵引列车前行。牵引电机和IGBT是牵引系统中的关键元件，共同负责将电能转换为机械能以推动列车，一旦其发生故障将会致使整个列车系统运行性能下降。本文将针对牵引电机和IGBT的典型故障模式进行建模，分析其退化过程。

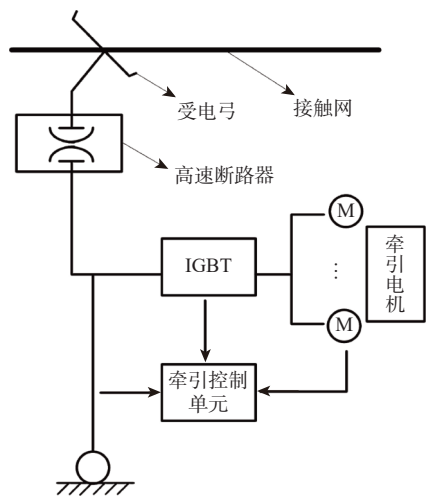


图1 牵引系统工作原理

Fig. 1 The working principle of traction system

1.2 牵引电机失效机理与性能退化建模

退磁故障作为牵引电机特有的一类故障，是关系电机能否安全可靠运行的最关键因素之一[2]。由于牵引电机长期处于高转矩运行状态，内部温度急剧攀升，但自身散热能力有限，使得电机永磁体磁感应强度减弱，与此同时退磁膝点随之升高。一旦发生退磁故障，将导致电机输出转矩减小，所产生的转矩脉动、振动和过热可能会引起匝间短路、偏心故障、轴承磨损等其他故障，而且退磁故障发生

初期往往难以直接检测, 因此, 开展退磁故障机理的分析和寿命预测研究显得尤为重要。

发生永磁体退磁故障后, 可以通过转子各个磁极上的永磁体磁链变化在 d-q 轴上的投影反映牵引电机退磁故障的情况。当牵引电机发生退磁故障时, 其电流变化率方程可以表示为

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}=\underbrace{\begin{bmatrix}-\dfrac{\rho}{L_d} & 0\\ 0 & -\dfrac{\rho}{L_q}\end{bmatrix}}_{\boldsymbol{A}}\boldsymbol{x}(t)+\underbrace{\begin{bmatrix}\dfrac{1}{L_d} & 0\\ 0 & \dfrac{1}{L_q}\end{bmatrix}}_{\boldsymbol{B}}\boldsymbol{u}(t)+\underbrace{\begin{bmatrix}\dfrac{\omega_{\mathrm{e}}L_q}{L_d}i_q+\dfrac{\omega_{\mathrm{e}}\psi_{gq}(t)}{L_d}\\ -\dfrac{\omega_{\mathrm{e}}L_d}{L_q}i_d-\dfrac{\omega_{\mathrm{e}}\psi_{gd}(t)}{L_q}\end{bmatrix}}_{\boldsymbol{d}_{\mathrm{f}}(t)} \tag{1}$$

式中: $\boldsymbol{x}(t)=\left[i_d(t),i_q(t)\right]^{\mathrm{T}}$, $i_d(t)$和$i_q(t)$为 d-q 轴上定子电流; $\boldsymbol{u}(t)=\left[u_d(t),u_q(t)\right]^{\mathrm{T}}$, $u_d(t)$和$u_q(t)$为 d-q 轴上定子电压; $\boldsymbol{A}$和$\boldsymbol{B}$为系统矩阵; $\boldsymbol{d}_{\mathrm{f}}(t)$为系统扰动项; ρ为定子相电阻; ω_{e}为转子角速度; L_d和L_q为 d-q 轴上定子绕组电感; $\psi_{gd}(t)$和$\psi_{gq}(t)$为永磁体磁链变化$\psi_{\mathrm{g}}(t)$在 d-q 轴上的投影。

通常, $\psi_{gd}(t)$和$\psi_{gq}(t)$难以获取, 本文采用扩张状态观测器方法, 得到永磁体磁链变化的估计信息。将状态方程 (1) 基于采样周期S离散化, 构造如下扩张状态观测器:

$$\begin{cases}\hat{\boldsymbol{d}}_{\mathrm{f}}(s)=\hat{\boldsymbol{d}}_{\mathrm{f}}(s-1)-\tau_1 S\boldsymbol{e}_x(s)\\ \hat{\boldsymbol{x}}(s+1)=\hat{\boldsymbol{x}}(s)+S\left(\boldsymbol{A}\boldsymbol{x}(s)+\boldsymbol{B}\boldsymbol{u}(s)\right)+\\ \qquad S\left(\hat{\boldsymbol{d}}_{\mathrm{f}}(s)-\tau_2\boldsymbol{e}_x(s)\right)\\ \boldsymbol{e}_x(s)=\hat{\boldsymbol{x}}(s)-\boldsymbol{x}(s)\end{cases} \tag{2}$$

式中: τ_1和τ_2为满足$0<\tau_1<\left(4+\tau_2^2\right)/(4S)$和$0<\tau_2<4$的参数, 则$\hat{\boldsymbol{d}}_{\mathrm{f}}(s)=\left[\hat{d}_{fd}(s),\hat{d}_{fq}(s)\right]^{\mathrm{T}}$能够收敛到$\boldsymbol{d}_{\mathrm{f}}(s)$。

由此, 牵引电机发生退磁故障后, 磁链在 d-q 轴的投影可以通过估计得到:

$$\begin{bmatrix}\hat{\psi}_{gd}(s)\\ \hat{\psi}_{gq}(s)\end{bmatrix}=\begin{bmatrix}-L_d i_d(s)\\ L_q i_q(s)\end{bmatrix}-\frac{1}{\omega_{\mathrm{e}}(k)}\begin{bmatrix}L_q\hat{d}_{fq}\\ L_d\hat{d}_{fd}\end{bmatrix} \tag{3}$$

基于退磁故障的影响, 定义电机性能退化指标如下:

$$Y_{\mathrm{M}}(l)=\frac{\sqrt{\hat{\psi}_{gd}^2(s)+\hat{\psi}_{gq}^2(s)}}{\psi_{\mathrm{f}}^*} \tag{4}$$

式中: ψ_{f}^*为正常情况下永磁体磁链值。

1.3　IGBT 键合线失效与性能退化建模

IGBT 是牵引逆变器实现电能转换的核心单元, 同时也是逆变器中失效率最高的元件。在长期运行过程中, IGBT 高频导通和关断会产生大量焦耳热, 且因各层封装材料的热膨胀系数不匹配, 在热应力反复冲击下易造成封装材料疲劳损伤, 最终导致失效, 威胁牵引逆变器运行。因热应力而失效的形式主要包含键合线失效和焊料层失效, 其中, 前者占 IGBT 总失效的 70% 左右。因此, 本文主要关注 IGBT 的键合线失效。本质上, 键合线的失效是由热应力引起, 且热应力多聚集于键合线底部, 这使得键合点的两侧出现裂纹, 并向中间扩展, 最终引发键合线完全脱落, 如图 2 所示。

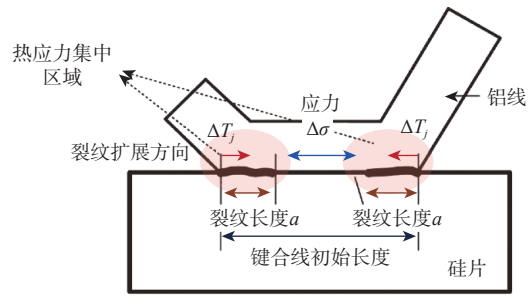


图 2　IGBT 键合线裂纹扩展示意图

Fig. 2　IGBT bonding wire crack propagation diagram

采用 Paris 方程描述裂纹扩展状态[17], 具体如下:

$$\frac{\mathrm{d}a}{\mathrm{d}N}=\vartheta(\Delta K)^{\xi} \tag{5}$$

式中: a为裂纹长度; N为应力循环次数; ϑ和ξ为与键合线材料相关的系数; ΔK为应力强度因子, 满足:

$$\Delta K=\beta\Delta\sigma\sqrt{\pi a} \tag{6}$$

其中: β为几何尺寸因子; $\Delta\sigma$为与结温变化ΔT相关的应力幅值, 即

$$\Delta\sigma=\kappa E\Delta\varepsilon=\kappa E(\alpha_{\mathrm{Al}}-\alpha_{\mathrm{Si}})\Delta T \tag{7}$$

式中: $\Delta\varepsilon$为应变; κ为常值系数; E为铝材料的弹性模量; α_{Al}和α_{Si}分别为铝材料和硅材料的热膨胀系数。

因此, 结合以上分析可以得到裂纹扩展状态方程为

$$\frac{\mathrm{d}a}{\mathrm{d}N}=\vartheta[\kappa\beta E(\alpha_{\mathrm{Al}}-\alpha_{\mathrm{Si}})\Delta T]^{\xi} \tag{8}$$

随着键合点的裂纹扩展, IGBT 外部电学特性的变化表现为集电极-发射极导通压降$V_{\mathrm{ce(on)}}$逐渐上升, 因此, 可将导通压降$V_{\mathrm{ce(on)}}$作为特征参数。在式 (8) 的基础上, $V_{\mathrm{ce(on)}}$的表达式如下:

$$V_{\mathrm{ce(on)}}=V_{\mathrm{ce(on)}}^0+f(a(N)) \tag{9}$$

式中: $V_{\mathrm{ce(on)}}^0$为伏安特性曲线于横轴的交点电压; $f(a(N))$由历史和仿真数据拟合确定, $a(N)$由式 (8) 得到。

定义 IGBT 的性能退化指标如下:

$$Y_{\mathrm{IGBT}}(N)=\frac{V_{\mathrm{ce(on)}}}{V^{*}_{\mathrm{ce(on)}}} \tag{10}$$

式中：$V^{*}_{\mathrm{ce(on)}}$为IGBT导通压降的失效阈值。

2 牵引系统二元相关可靠性建模与剩余使用寿命预测方法

IGBT需要根据列车的牵引需求，调整输出电能的特性，以满足牵引电机的工作要求。IGBT故障会影响牵引电机将电能转换为机械能，从而影响整个列车正常服役。牵引电机退磁故障将导致电机的输出转矩下降，在输入功率不变的情况下，电机的恒转矩阶段占整个牵引时间的比重上升，IGBT的功率损耗相比恒功率阶段而言较大，进而导致IGBT功率损耗上升，影响IGBT正常工作。牵引电机和逆变器之间的协同确保了电能的高效转换，促使牵引系统能够在不同速度和负载条件下实现最佳性能，因此，二者性能之间存在一定的相关性。本节分别针对牵引电机和IGBT进行性能退化建模，并基于此建立考虑二者相关的性能退化模型。

2.1 融合失效机理的维纳过程模型

随服役时间的推移，牵引电机与IGBT在退化过程中存在一定的随机性。本文采用维纳过程来描述二者的性能退化。维纳过程初始退化量为零，退化量服从正态分布：

$$Y_k(t)\sim N\left(\mu_k\Lambda_k(t;\gamma_k),\sigma_k^2\Lambda_k(t;\gamma_k)\right)\qquad(k=\mathrm{M},\mathrm{IGBT}) \tag{11}$$

式中：$\mu_k>0$为漂移参数，用于反映退化速率；$\sigma_k^2>0$为扩散参数，用于反映退化过程中时间随机性；$\Lambda_k(t;\gamma_k)$为单调递增的时间尺度函数，用于反映性能指标的退化轨迹，γ_k为时间尺度函数中的相关参数；k代表牵引电机和IGBT。基于牵引电机和IGBT的失效机理分析，确定2个元件性能指标的退化轨迹形式。

给定牵引电机和IGBT性能退化指标的失效阈值为d_k，则失效时间H_k定义为首次达到阈值的时间，即

$$H_k=\inf\{t:Y_k(t)>d_k\} \tag{12}$$

对于维纳过程，失效时间H_k服从逆高斯分布。因此，单一元件的可靠度函数$R_k(t)$表示为

$$R_k(t)=\Phi\left(\frac{d_k-\mu_k\Delta\Lambda_k(t;\gamma_k)}{\sigma_k\sqrt{\Delta\Lambda_k(t;\gamma_k)}}\right)-\exp\left(\frac{2\mu_k d_k}{\sigma_k^2}\right)\Phi\left(-\frac{d_k+\mu_k\Delta\Lambda_k(t;\gamma_k)}{\sigma_k\sqrt{\Delta\Lambda_k(t;\gamma_k)}}\right) \tag{13}$$

给定性能退化指标的性能退化数据为$Y_{k,1:j}=\{Y_k(t_1),Y_k(t_2),\cdots,Y_k(t_j)\}$，且在$t_j$时刻未失效，即$Y_k(t_j)<d_k$。剩余使用寿命$U_j$表示元件失效前可以保持正常工作的时间，其表达式为

$$U_j=\inf\{l_j:Y(t_j+l_j)\geqslant d_k,Y_{k,1:j},Y_k(t_j)<d_k\} \tag{14}$$

剩余使用寿命可以认为是随机过程$\{Y'_k(l_j),l_j>0\}$超过失效阈值$d_k-Y_k(t_j)$的时间，其中，$Y'_k(l_j)=Y_k(l_j+t_j)-Y_k(t_j)$。与式(13)类似，剩余使用寿命累积分布$F_{k,U_j}(l_j)$的表达式为

$$F_{k,U_j}(l_j)=\Phi\left(\frac{-(d_k-Y_k(t_j))+\mu_k\Delta\Lambda_k(l_j;\gamma_k)}{\sigma_k\sqrt{\Delta\Lambda_k(l_j;\gamma_k)}}\right)+\exp\left(\frac{2\mu_k(d_k-Y_k(t_j))}{\sigma_k^2}\right)\cdot\Phi\left(-\frac{(d_k-Y_k(t_j))+\mu_k\Delta\Lambda_k(l_j;\gamma_k)}{\sigma_k\sqrt{\Delta\Lambda_k(l_j;\gamma_k)}}\right) \tag{15}$$

由于加工工艺和设计误差等因素，元件间存在个体差异性。不同工况下牵引系统性能退化也有所不同，进而导致其性能退化过程中存在不确定性，且这些不确定性可以体现在退化量均值$E[Y_k(t)]=\mu_k\Lambda(t;\gamma_k)$上。本文假设退化速率$\mu_k$服从正态分布，即$\mu_k\sim N(\eta_{\mu,k},\sigma_{\mu,k}^2)$。此时，对于考虑个体间差异性的维纳过程的可靠度函数$R_k^{\mathrm{v}}(t)$和剩余使用寿命分布$F_{k,U_j}^{\mathrm{v}}(l_j)$，分别通过求可靠度函数$R_k(t)$和剩余使用寿命分布$F_{k,U_j}(l_j)$关于$p(\mu_k)$的期望得到，并根据文献[18-19]所给出的定理可以得到其解析表达式。

2.2 Copula函数

Copula函数可以作为连接函数，通过相应的函数结构连接多个随机变量的边缘分布，从而描述性能指标之间的相关性。本文采用Copula函数来描述牵引电机磁链和IGBT导通压降之间的相关性。根据Sklar定理[20]和二元Copula函数，2个退化指标的联合累积分布函数表示为

$$J(y_{\mathrm{M}},y_{\mathrm{IGBT}})=C(J_{\mathrm{M}}(y_{\mathrm{M}}),J_{\mathrm{IGBT}}(y_{\mathrm{IGBT}});\theta) \tag{16}$$

式中：$J_{\mathrm{M}}(y_{\mathrm{M}})$和$J_{\mathrm{IGBT}}(y_{\mathrm{IGBT}})$分别为2个性能指标的边缘分布函数；$\theta$为Copula函数中的相关系数，用于反映2个退化指标之间的相关性强度。

2个性能指标的联合概率密度函数表示为

$$f(y_{\mathrm{M}},y_{\mathrm{IGBT}})=c(J_{\mathrm{M}}(y_{\mathrm{M}}),J_{\mathrm{IGBT}}(y_{\mathrm{IGBT}});\theta)f(y_{\mathrm{M}})f(y_{\mathrm{IGBT}}) \tag{17}$$

式中：$f(y_{\mathrm{M}})$和$f(y_{\mathrm{IGBT}})$分别为2个随机变量的概率密度函数；$c(J_{\mathrm{M}}(y_{\mathrm{M}}),J_{\mathrm{IGBT}}(y_{\mathrm{IGBT}});\theta)$为对应Copula函数的概率密度函数，且有

$$c(J_{\mathrm{M}}(y_{\mathrm{M}}),J_{\mathrm{IGBT}}(y_{\mathrm{IGBT}});\theta)=\frac{\partial C(J_{\mathrm{M}}(y_{\mathrm{M}}),J_{\mathrm{IGBT}}(y_{\mathrm{IGBT}});\theta)}{\partial J_{\mathrm{M}}(y_{\mathrm{M}})\,\partial J_{\mathrm{IGBT}}(y_{\mathrm{IGBT}})} \tag{18}$$

Gaussian Copula 函数呈现出对称结构[20], 本文采用该函数来描述牵引电机和 IGBT 退化的相关性。

2.3 基于相关性函数的系统多性能指标相关的寿命预测

给定牵引电机磁链变化和 IGBT 集电极-发射极导通压降的失效阈值分别为d_{M}和d_{IGBT}, 且当任意一个性能指标达到失效阈值时即认为失效。考虑 2 个性能指标的可靠度函数表示为

$$R(t)=\Pr\{Y_{\mathrm{M}}(t)<d_{\mathrm{M}},Y_{\mathrm{IGBT}}(t)<d_{\mathrm{IGBT}}\} \tag{19}$$

牵引系统寿命表示为

$$H_{\mathrm{system}}=\min\{H_{\mathrm{M}},H_{\mathrm{IGBT}}\} \tag{20}$$

由 Copula 函数性质可知, 考虑 2 个性能指标相关的可靠度函数$R(t)$和剩余使用寿命分布$F_{U_j}(l_j)$如下:

$$R(t)=C(R_{\mathrm{M}}(t),R_{\mathrm{IGBT}}(t)) \tag{21}$$

$$F_{U_j}(l_j)=C\left(J_{\mathrm{M},U_j}(l_j),J_{\mathrm{IGBT},U_j}(l_j)\right) \tag{22}$$

根据全概率公式, 考虑元件个体间差异和 2 个性能退化指标相关的可靠度函数$R^{\mathrm{v}}(t)$及系统牵引系统剩余使用寿命分布$F^{\mathrm{v}}_{U_j}(l_j)$分别表示为

$$R^{\mathrm{v}}(t)=\iint R(t|\mu_{\mathrm{M}},\mu_{\mathrm{IGBT}})\,g(\mu_{\mathrm{M}})\,g(\mu_{\mathrm{IGBT}})\,\mathrm{d}\mu_{\mathrm{M}}\mathrm{d}\mu_{\mathrm{IGBT}} \tag{23}$$

$$F^{\mathrm{v}}_{U_j}(l_j)=\iint F_{U_j}(l_j|\mu_{\mathrm{M}},\mu_{\mathrm{IGBT}})\,g(\mu_{\mathrm{M}})\,g(\mu_{\mathrm{IGBT}})\,\mathrm{d}\mu_{\mathrm{M}}\mathrm{d}\mu_{\mathrm{IGBT}} \tag{24}$$

3 统计推断

3.1 离线参数估计

假设在试验中共有N个牵引电机和 IGBT, 以及在给定试验时间内每个试验样件有M个性能退化观测值。假设第$i(i=1,2,\cdots,N)$个元件第k个性能退化指标的第$j(j=1,2,\cdots,M)$个观测量为$Y_k(t_{ij})$, 则退化增量可以表示为$\Delta Y_k(t_{ij})=Y_k(t_{ij})-Y_k(t_{i,j-1})$。时间尺度函数参数$\gamma_k$由元件失效机理来确定。令$\boldsymbol{\Theta}=(\eta_{\mu,\mathrm{M}},\sigma^2_{\mu,\mathrm{M}},\sigma^2_{\mathrm{M}},\eta_{\mu,\mathrm{IGBT}},\sigma^2_{\mu,\mathrm{IGBT}},\sigma^2_{\mathrm{IGBT}},\theta)$表示可靠性模型中所有的未知参数, 其中, $\boldsymbol{\Theta}^{\mathrm{W}}_k=(\eta_{\mu k},\sigma^2_{\mu k},\sigma^2_k)$为参数估计时维纳过程中的未知参数。令$\boldsymbol{D}=(\boldsymbol{D}_{\mathrm{M}},\boldsymbol{D}_{\mathrm{IGBT}})$表示退化数据, 其中, $\boldsymbol{D}_{\mathrm{M}}$为牵引电机的退化数据, $\boldsymbol{D}_{\mathrm{IGBT}}$为 IGBT 的退化数据。由此, 构建考虑 2 个性能指标相关的可靠性模型, 其表达式如下:

$$\begin{cases}\Delta Y_k(t_{ij})\sim N\left(\mu_k\Delta\Lambda(t_{ij};\gamma_k),\sigma_k^2\Delta\Lambda(t_{ij};\gamma_k)\right)\\ \mu_k\sim N(\eta_{\mu k},\sigma^2_{\mu k})\\ f\left(\Delta Y_{\mathrm{M}}(t_{ij}),\Delta Y_{\mathrm{IGBT}}(t_{ij})|\boldsymbol{D},\mu_{\mathrm{M}},\mu_{\mathrm{IGBT}}\right)=\\ \quad\prod_{k=1}^{2}\prod_{i=1}^{N}\prod_{j=1}^{M}f_k\left(\Delta Y_k(t_{ij})\right)\cdot\\ \quad\prod_{i=1}^{N}\prod_{j=1}^{M}c\left(J_{\mathrm{M}}(\Delta Y_{\mathrm{M}}(t_{ij})),J_{\mathrm{IGBT}}(\Delta Y_{\mathrm{IGBT}}(t_{ij}))\right)\\ \quad i=1,2,\cdots,N;j=1,2,\cdots,M\end{cases} \tag{25}$$

完全数据$(\boldsymbol{D},\mu_{\mathrm{M}},\mu_{\mathrm{IGBT}})$的对数似然函数可以表示为

$$\begin{aligned}\ln\Gamma(\boldsymbol{\Theta})=&\sum_{i=1}^{N}\sum_{j=1}^{M}\ln c\left(J_{\mathrm{M}}(\Delta Y_{\mathrm{M}}(t_{ij})|\mu_{\mathrm{M},i}),\right.\\ &\left.J_{\mathrm{IGBT}}(\Delta Y_{\mathrm{IGBT}}(t_{ij})|\mu_{\mathrm{IGBT},i});\theta\right)+\\ &\sum_{k=1}^{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\ln\left[f_k(\Delta Y_k(t_{ij})|\mu_{k,i})\right]+\sum_{k=1}^{2}\sum_{i=1}^{N}\ln g(\mu_{k,i})\end{aligned} \tag{26}$$

采用贝叶斯马尔可夫链蒙特卡罗方法对模型中的未知参数进行估计, 即

$$\pi(\boldsymbol{\Theta}|\boldsymbol{D})\propto\ln\Gamma(\boldsymbol{\Theta})\cdot\pi(\boldsymbol{\Theta}) \tag{27}$$

式中: $\pi(\boldsymbol{\Theta})$和$\pi(\boldsymbol{\Theta}|\boldsymbol{D})$分别为未知参数$\boldsymbol{\Theta}$的先验概率和后验概率, 且本文参数的先验分布均选为无信息先验分布。基于元件的历史退化数据, 借助 OpenBUGS 完成未知参数估计, 并以此作为在线参数估计的初值。在本文提出的模型下, 基于联合后验概率密度函数的可靠度函数如下:

$$R^{\mathrm{v}}(t|\boldsymbol{D})=\int_{\boldsymbol{\Theta}}R^{\mathrm{v}}(t|\boldsymbol{\Theta})\pi(\boldsymbol{\Theta}|\boldsymbol{D})\,\mathrm{d}\boldsymbol{\Theta}=E_{\pi(\boldsymbol{\Theta}|\boldsymbol{D},T)}[R^{\mathrm{v}}(t|\boldsymbol{\Theta})] \tag{28}$$

3.2 在线参数更新

本节基于在线退化数据, 提出参数更新和剩余使用寿命计算方法。针对模型中未知参数$\boldsymbol{\Theta}=(\eta_{\mu,\mathrm{M}},\sigma^2_{\mu,\mathrm{M}},\sigma^2_{\mathrm{M}},\eta_{\mu,\mathrm{IGBT}},\sigma^2_{\mu,\mathrm{IGBT}},\sigma^2_{\mathrm{IGBT}},\theta)$, 采用具有计算效率优势的两阶段估计法完成未知参数的估计[20]。首先, 对边缘分布中的未知参数进行估计, 将参数估计由多维简化到一维; 然后, 估计 Copula 函数中的参数, 以反映边缘分布之间的相关关系。

3.2.1 第 1 阶段: 边缘分布参数更新

当获得第m个退化增量数据时, 标记此时在线单性能指标退化数据为$\boldsymbol{D}_{k,1:m}=[\Delta X_k(t_1),\Delta X_k(t_2),\cdots,\Delta X_k(t_m)]$。由贝叶斯定理可知:

$$p\left(\mu_k|\boldsymbol{D}_{k,1:m}\right)\propto p\left(\boldsymbol{D}_{k,1:m}|\mu_k\right)\cdot p(\mu_k) \tag{29}$$

式中: $p(\mu_k)$为先验概率, 其服从正态分布, 即$\mu_k\sim N(\eta_{\mu k},\sigma^2_{\mu k})$; $p\left(\boldsymbol{D}_{k,1:m}|\mu_k\right)$为似然函数; $p\left(\mu_k|\boldsymbol{D}_{k,1:m}\right)$

为后验概率。由指数族分布的共轭条件可知,后验概率$p\left(\mu_k|\boldsymbol{D}_{k,1:m}\right)$仍为正态分布。当获得新息$\boldsymbol{D}_{k,m}$时,漂移参数$\mu_k$的后验概率的均值和方差为

$$\begin{cases}\eta_{\mu k,m}^{\text{posterior}}=\dfrac{X_k(t_m)\sigma_{\mu k}^2+\eta_{\mu k}}{\Lambda(t_m;\gamma_k)\sigma_{\mu k}^2+\sigma_k^2}\\ \left(\sigma_{\mu k,m}^2\right)^{\text{posterior}}=\dfrac{\sigma_k^2\sigma_{\mu k}^2}{\Lambda(t_m;\gamma_k)\sigma_{\mu k}^2+\sigma_k^2}\end{cases}\tag{30}$$

式中:$\eta_{\mu k,m}^{\text{posterior}}$和$\left(\sigma_{\eta k,m}^2\right)^{\text{posterior}}$随着迭代次数的变化而变化,在降低不确定性的基础上不断更新漂移参数μ_k。

由于μ_k服从正态分布,无法使用似然函数通过极大似然估计直接得到。为解决此问题,本文采用EM算法,对存在隐变量的位置参数进行估计[21]。EM算法具有计算简单、可保证收敛性的特点,可以通过给定的数据集寻找不可观测数据的极大似然估计,使得算法的每一步都能够以封闭或直接的形式获得参数估计的更新值。

完全数据$\left(\boldsymbol{D}_{k,1:m},\mu_k\right)$的对数似然函数表示为

$$\begin{aligned}\ln p\left(\boldsymbol{D}_{k,1:m},\mu_k|\boldsymbol{\Theta}\right)&=\ln p\left(\boldsymbol{D}_{k,1:m}|\mu_k,\boldsymbol{\Theta}\right)+\ln p\left(\mu_k|\boldsymbol{\Theta}\right)\propto\\&-\frac{m}{2}\ln\sigma_k^2-\sum_{j=1}^{m}\frac{\left(\Delta X\left(t_j\right)-\mu_k\Delta\Lambda\left(t_j;\gamma_k\right)\right)^2}{2\sigma_k^2\Delta\Lambda\left(t_j;\gamma_k\right)}-\\&\frac{1}{2}\ln\sigma_{\mu k}^2-\frac{\left(\mu_k-\eta_{\mu k}\right)^2}{2\sigma_{\mu k}^2}\end{aligned}\tag{31}$$

当得到时刻t_m时,基于退化数据$\boldsymbol{D}_{k,1:m}$,在EM算法中已经进行到第g步获得参数的估计值$\hat{\boldsymbol{\theta}}_{k,m}^{\text{W}(g)}=\left[\eta_{\mu k}^{(g)},\sigma_{\mu k}^{2(g)},\sigma_k^{2(g)}\right]$的条件并得到后验值$\hat{\boldsymbol{\theta}}_{k,m}^{\text{Wposterior}(g)}$,则EM算法中求期望步骤中,即求$\ln p\left(\boldsymbol{D}_{k,1:m},\mu_k|\boldsymbol{\Theta}\right)$关于$p\left(\mu_k|\boldsymbol{D}_{k,1:m},\hat{\boldsymbol{\theta}}_{k,m}^{\text{Wposterior}(g)}\right)$的数学期望$Q\left(\boldsymbol{\theta}_k^{\text{W}}|\boldsymbol{D}_{k,1:m},\hat{\boldsymbol{\theta}}_{k,m}^{\text{Wposterior}(g)}\right)$,则有

$$\begin{aligned}Q\left(\boldsymbol{\theta}_k^{\text{W}}|\boldsymbol{D}_{k,1:m},\hat{\boldsymbol{\theta}}_{k,m}^{\text{Wposterior}(g)}\right)&=\frac{m}{2}\ln\sigma_k^2-\sum_{j=1}^{m}\frac{1}{2\sigma_k^2\Delta\Lambda\left(t_j;\gamma_k\right)}\cdot\\&\left\{\left[\Delta X\left(t_j\right)\right]^2-2\eta_{\mu k,m}^{(g)}\Delta\Lambda\left(t_j;\gamma_k\right)\Delta X\left(t_j\right)+\right.\\&\left.\left(\eta_{\mu k,m}^{2(g)}+\sigma_{\mu k,m}^{2(g)}\right)\left[\Delta\Lambda\left(t_j;\gamma_k\right)\right]^2\right\}-\frac{1}{2}\ln\sigma_{\mu k}^2-\\&\frac{1}{2\sigma_{\mu k}^2}\left(\eta_{\mu k,m}^{2(g)}+\sigma_{\mu k,m}^{2(g)}-2\eta_{\mu k,m}^{2(g)}\eta_{\mu k}+\eta_{\mu k}^2\right)\end{aligned}\tag{32}$$

令$\dfrac{\partial}{\partial\boldsymbol{\theta}_k^{\text{W}}}Q\left(\boldsymbol{\theta}_k^{\text{W}}|\boldsymbol{D}_{k,1:m},\hat{\boldsymbol{\theta}}_{k,m}^{\text{Wposterior}(g)}\right)=0$,则有

$$\eta_{\mu k,m}^{(g+1)}=\eta_{\mu k,m}^{\text{posterior}(g)}\tag{33}$$

$$\sigma_{\mu k,m}^{2(g+1)}=\left(\sigma_{\mu k,m}^2\right)^{\text{posterior}(g)}\tag{34}$$

$$\begin{aligned}\sigma_{k,m}^{2(g+1)}=\frac{1}{m}\sum_{j=1}^{m}&\left\{\left[\Delta X\left(t_j\right)\right]^2-2\eta_{\mu k,m}^{(g+1)}\Delta\Lambda\left(t_j;\gamma_k\right)\Delta X\left(t_j\right)+\right.\\&\left.\left(\eta_{\mu k,m}^{2(g+1)}+\sigma_{\mu k,m}^{2(g+1)}\right)\left[\Delta\Lambda\left(t_j;\gamma_k\right)\right]^2\right\}/\Delta\Lambda\left(t_j;\gamma_k\right)\end{aligned}\tag{35}$$

上述计算所获得的点是唯一最大值,即每次迭代均有新的表达式。当满足收敛阈值时,可以获得参数的估计值。至此,即可完成2个性能指标维纳过程中模型参数的更新。

3.2.2 第2阶段:相关系数更新

在完成边缘分布的参数估计后,可以得到每个性能指标的边缘分布,进而以此作为Copula函数的输入,对Gaussian Copula函数中相关系数θ进行更新。Gaussian Copula的似然函数表示为

$$\begin{aligned}&\ln\Gamma^c\left(\theta_m|\boldsymbol{D}_{1:m}\right)=\\&\sum_{j=1}^{m}\ln\left[c\left(J_{\text{M}}\left(\Delta Y_k\left(t_j\right)\right),J_{\text{IGBT}}\left(\Delta Y_k\left(t_j\right)\right);\theta_m\right)\right]\end{aligned}\tag{36}$$

基于似然函数,采用极大似然估计得到Gaussian Copula函数中相关系数θ_m的估计值,即

$$\hat{\theta}_m=\arg\max_{\theta}\ln\Gamma^c\left(\theta|\boldsymbol{D}_{1:m}\right)\tag{37}$$

至此,完成在线可靠性模型中所有未知参数的估计,进而经计算可得元件剩余使用寿命的分布。注意到,该算法中维纳过程模型的参数由显式表达,因此,可以显著提高计算速度。算法1总结了在线参数估计的算法流程。

算法1 多元性能指标相关的动态可靠性模型在线参数估计算法。

输入: 模型参数的初值$\eta_{\mu k,0}$、$\sigma_{\mu k,0}^2$、$\sigma_{k,0}^2$、θ_0(由基于退化数据与故障时间数据的离线估计得到);在线退化数据$\boldsymbol{D}_{1:m}$;迭代次数κ;EM算法迭代阈值ε。

输出: $\eta_{\mu k,m}$、$\sigma_{\eta k,m}^2$、$\sigma_{k,m}^2$、θ_m。

在$t=t_m$时刻:

步骤1 初始化$\left\{\eta_{\mu k}^{(0)},\sigma_{\mu k}^{2(0)},\sigma_k^{2(0)}\right\}$。

For $g=1:\kappa$ or $\left|\hat{\boldsymbol{\theta}}_k^{\text{IG}(g+1)}-\hat{\boldsymbol{\theta}}_k^{\text{IG}(g)}\right|\geqslant\varepsilon$

由贝叶斯式(30)计算后验值$\left\{\eta_{\mu k,m}^{\text{posterior}(g)},\left(\sigma_{\mu k,m}^2\right)^{\text{posterior}(g)}\right\}$;

由EM算法式(33)~式(35)更新$\left\{\eta_{\mu k,m}^{(g+1)},\sigma_{\mu k,m}^{2(g+1)},\sigma_{k,m}^{2(g+1)}\right\}$;

End for

步骤2 由极大似然估计式(37)获得$\hat{\theta}_m$。

步骤3 获得$t=t_m$时刻所有未知参数的估计值,即可计算剩余使用寿命。

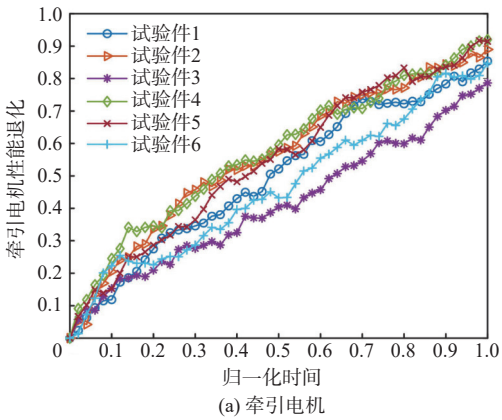
获得下一时刻退化数据$\boldsymbol{D}_{m+1}$,重复步骤1~步骤3。

4 实例分析

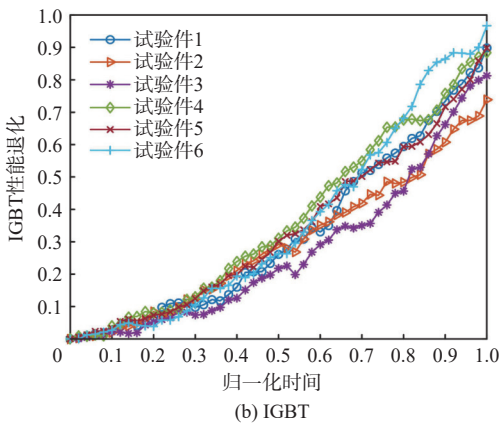
4.1 试验数据

本节通过牵引系统服役工况下的退化试验实例验证所提算法的有效性。牵引系统退化试验中共有6套试验件,通过性能退化试验采集电机定子的电流、电压和IGBT的导通压降,得到电机和

IGBT 性能指标随时间变化曲线, 如图 3 所示。将原始退化数据归一化处理到[0,1]之间, 将退化时间也归一化处理。采用维纳过程描述牵引电机磁链和 IGBT 导通压降退化过程, 采用 Gaussian Copula 函数描述 2 个性能指标之间的相关性。



(a) 牵引电机



(b) IGBT

图 3　牵引电机和 IGBT 性能退化曲线

Fig. 3　The degradation curves of the traction motor and IGBT

4.2　离线可靠性评估

根据留出法原则, 前 5 个牵引电机和 IGBT 的退化数据用于估计离线可靠性模型中的未知参数, 最后 1 个牵引电机和 IGBT 的数据用于预测剩余使用寿命。由图 3 可知, 牵引电机磁链退化逐渐上升并趋于放缓。因此, 设时间尺度函数为 $\Lambda_{\mathrm{M}}(t|\gamma_{\mathrm{M}})=\alpha\ln(1+t^{\gamma_{\mathrm{M}}})$, 并经曲线拟合得到 $\alpha=3.46$, $\gamma_{\mathrm{M}}=1.12$; IGBT 的导通电压退化曲线为凹曲线, 因此, 设时间尺度函数为 $\Lambda_{\mathrm{IGBT}}(t|\gamma_{\mathrm{IGBT}})=t^{\gamma_{\mathrm{IGBT}}}$, 并经曲线拟合得到 $\gamma_{\mathrm{IGBT}}=1.668$。

将本文所提出的可靠性评估方法与现有相关方法进行对比, 以验证方法的准确性。对比模型如下:

1)模型 1(model 1): 仅考虑牵引电机磁链性能退化, 时间尺度函数由失效机理确定。

2)模型 2(model 2): 仅考虑 IGBT 导通电压性能退化, 时间尺度函数由失效机理确定。

3)模型 3(model 3): 仅考虑牵引电机磁链性能退化, 退化轨迹采用线性函数 $\Lambda_{\mathrm{M}}(t)=t$。

4)模型 4(model 4): 同时考虑牵引电机磁链和 IGBT 导通电压 2 个性能退化相关, 电机时间尺度函数采用线性函数 $\Lambda_{\mathrm{M}}(t)=t$, IGBT 时间尺度函数由失效机理确定。

5)模型 5(model 5): 同时考虑牵引电机和 IGBT 2 个性能退化相关, 两者时间尺度函数均由失效机理确定, 即本文方法。

通过贝叶斯马尔可夫链蒙特卡罗方法得到上述 5 个模型中未知参数的后验概率分布, 并以退化收敛的后验概率均值为参数估计值。基于模型 5 的参数估计结果如表 1 所示。

表 1　模型 5 未知参数估计结果

Table 1　Unknown parameter estimation results for model 5

牵引电机/IGBT	η_μ	σ_μ^2	σ^2
牵引电机	1.124 9	0.458 2	0.217 9
IGBT	1.095 0	0.143 1	0.112 1

注: 相关系数 $\theta=0.854\,6$。

基于上述基于贝叶斯马尔可夫链蒙特卡罗方法的参数估计结果, 由式 (28) 可以得到不同模型下的可靠度随时间变化的曲线, 如图 4 所示。由此看出, 根据不同可靠性模型得到的可靠度随时间变化曲线存在差异。在服役初期, 牵引电机比 IGBT 的退化速率快, 牵引电机可靠度曲线比 IGBT 低; 在服役后期, IGBT 比牵引电机的退化速率快, IGBT 可靠度曲线比牵引电机低。因此, 考虑 2 个性能指标相关的可靠度曲线在服役初期与电机可靠度曲线接近, 在服役后期与 IGBT 可靠度曲线接近。当描述牵引电机性能退化时间尺度函数为线性函数时, 比电机性能退化时间尺度函数为由失效机理确定的差别较大, 且影响了考虑 2 个性能退化指标相关的可靠度曲线。

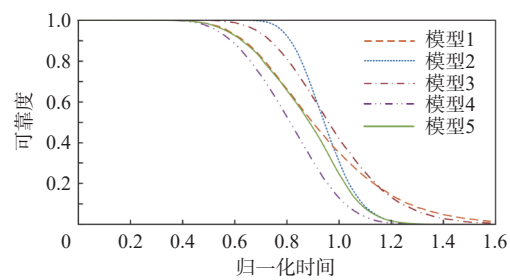


图 4　不同模型得到的牵引系统可靠度随时间变化曲线

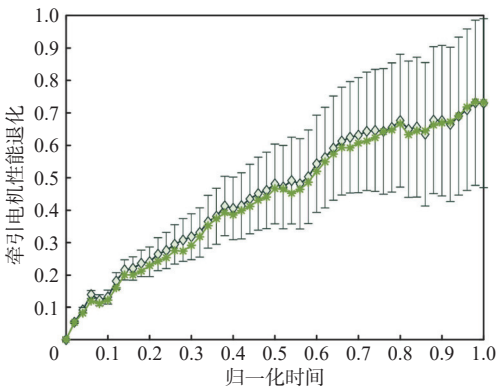
Fig. 4　The reliability curves of traction system based on different models

4.3　在线剩余使用寿命预测

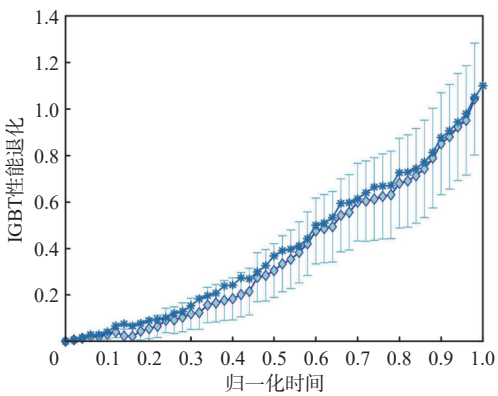
根据试验件 6 的退化数据进行剩余使用寿命的评估, 以验证在线统计推断算法的准确性。为此, 以 4.2 节中得到的模型参数估计值为初值, 采用

本文所提出的贝叶斯与EM相融合的在线参数估计算法，估计2个性能指标维纳过程中未知参数及Gaussian Copula的相关系数。

图5为2个性能指标的退化轨迹预测值及误差带。可以看到，在初期模型参数估计值并不准确，在服役初期预测结果也不准确。由于退化过程中存在不确定性和波动性，基于贝叶斯算法更新模型中的参数，与此同时采用EM算法进行迭代，因此，模型中的未知参数不断发生变化。注意到，随着退化数据的不断积累，各未知参数的估计值不断趋于稳定，退化轨迹预测的准确率逐渐上升。



(a) 牵引电机磁链实际退化轨迹与估计退化轨迹对比



(b) IGBT导通电压实际退化轨迹与估计退化轨迹对比

图5 实际退化轨迹与估计退化轨迹对比

Fig. 5 The comparison between the true degradation path and estimated degradation path

采用均方误差来描述估计值与实际值之间的误差，表2列出不同算法下退化量预测值与真实值的均方误差。由此可知，相较于仅使用EM算法，本文所提出的贝叶斯与EM相融合的算法能够得到更小的均方误差，因此，更有助于提升未知参数估计的准确性。

类似地，将本文所提出的在线可靠性评估方法与现有相关方法进行对比，以验证方法的准确性。对比模型如下：

1)模型1(model 1)：仅考虑牵引电机性能退化。

2)模型2(model 2)：仅考虑IGBT性能退化。

表2 不同算法得到的退化量估计值与实际值的均方误差

Table 2 The MSEs between the estimated degradation path and true degradation path based on different algorithms

算法	均方误差	
	牵引电机	IGBT
贝叶斯-EM	0.024 7	0.016 6
EM	0.046 82	0.034 11

3)模型3(model 3)：同时考虑牵引电机和IGBT性能退化相关，即本文方法。

基于以上3种模型，图6为剩余使用寿命的概率密度分布图。随着退化数据的累积，概率密度分布逐渐变窄，表明预测值逐渐精准，具有较高的预测精度。值得指出的是，考虑牵引电机和IGBT 2个元件得到的概率密度分布相较于仅考虑单一元件的概率密度分布更高，说明寿命预测精度更高。

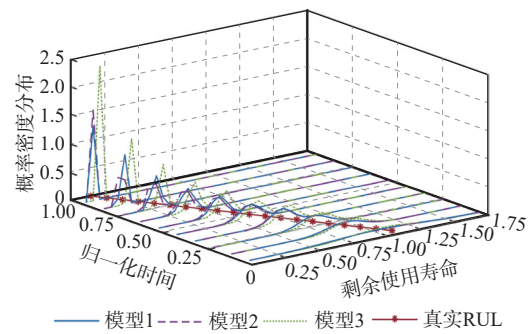


图6 估计剩余使用寿命与实际剩余使用寿命的概率密度分布对比

Fig. 6 The comparisons between the estimated PDF of RUL and true RUL

表3列出不同模型下剩余使用寿命估计值与实际值的均方误差。由此可知，模型3下剩余使用寿命估计值的均方误差小于模型1和模型2。因此，考虑牵引电机和IGBT 2个元件相关的寿命估计值比仅考虑一个元件更契合实际。

表3 不同预测模型得到的剩余使用寿命估计值与实际剩余使用寿命的均方误差

Table 3 The MSEs between the estimated RUL and true RUL based on different prediction models

模型	均方误差
模型1	0.089 74
模型2	0.076 12
模型3	0.036 72

5 结 论

针对列车牵引系统的寿命预测问题，提出了考虑牵引电机和IGBT 2个元件相关的可靠性评估与剩余使用寿命预测算法，通过性能退化试验，得到

如下结论:

1) 针对牵引电机退磁故障和 IGBT 键合线失效 2 种故障, 建立了两者失效机理与性能退化模型, 有效反映了 2 个元件的性能退化过程。

2) 构建了牵引电机和 IGBT 性能退化相关的可靠性模型, 不仅反映了两者之间的相关性, 且有助于提高可靠性评估和寿命预测的准确性。

3) 提出了基于贝叶斯与期望最大相融合的动态参数估计算法, 能够实现精准剩余使用寿命预测, 并通过试验验证了本文算法的有效性。

参考文献（References）

[1] 牛刚, 刘森铁. 永磁牵引系统退磁故障自主延寿控制[J]. 仪器仪表学报, 2017, 38(9): 2251-2257.
NIU G, LIU S Y. Autonomous life extending control for PMSM rail vehicles under demagnetizing fault[J]. Chinese Journal of Scientific Instrument, 2017, 38(9): 2251-2257(in Chinese).

[2] 马伊光. 高转矩密度永磁同步电机的可靠性研究[D]. 哈尔滨: 哈尔滨工业大学, 2019.
MA Y G. Research on reliability of high torque density permanent magnet synchronous motor[D]. Harbin: Harbin Institute of Technology, 2019(in Chinese).

[3] 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, 2024, 37(6): 244-261.

[4] 唐春萍. 高速列车牵引电机驱动系统复合故障诊断[D]. 南京: 南京航空航天大学, 2018.
TANG C P. Compound fault diagnosis of traction motor drive system of high-speed train[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2018(in Chinese).

[5] 罗俊. IGBT 模块键合引线断裂故障的热形变场特性分析及应用[D]. 成都: 电子科技大学, 2022.
LUO J. Analysis and application of thermal deformation field characteristics of IGBT module bonding wire fracture fault[D]. Chengdu: University of Electronic Science and Technology of China, 2022(in Chinese).

[6] 杨德浩. 考虑牵引逆变器 IGBT 模块可靠性的列车运行曲线优化方法研究[D]. 北京: 北京交通大学, 2022.
YANG D H. Research on optimization method of train operation curve considering reliability of IGBT module of traction inverter [D]. Beijing: Beijing Jiaotong University, 2022(in Chinese).

[7] YANG L, AGYAKWA P A, JOHNSON C M. Physics-of-failure lifetime prediction models for wire bond interconnects in power electronic modules[J]. IEEE Transactions on Device and Materials Reliability, 2013, 13(1): 9-17.

[8] RAMEZANI S B, CUMMINS L, KILLEN B, et al. Scalability, explainability and performance of data-driven algorithms in predicting the remaining useful life: a comprehensive review[J]. IEEE Access, 2023, 11: 41741-41769.

[9] YANG T H, LI S H, DUAN S Y, et al. Performance degradation model and reliability evaluation of brush DC motor for the intelligent on-off valve[J]. Journal of Electrical Engineering & Technology, 2023, 18(3): 1909-1918.

[10] 朱琦. 电机驱动系统母线电容状态监测与寿命预测方法研究[D]. 武汉: 华中科技大学, 2022.
ZHU Q. Research on monitoring and life prediction method of bus capacitance in motor drive system[D]. Wuhan: Huazhong University of Science and Technology, 2022(in Chinese).

[11] 曾东. 基于加速老化试验 IGBT 性能退化特征参量的可靠性评估[D]. 重庆: 重庆大学, 2019.
ZENG D. Reliability evaluation of IGBT performance degradation characteristic parameters based on accelerated aging test[D]. Chongqing: Chongqing University, 2019(in Chinese).

[12] 王宇琦. IGBT 疲劳老化失效剩余使用寿命预测的研究[D]. 西安: 西安电子科技大学, 2017.
WANG Y Q. Study on prediction of remaining service life of IGBT in fatigue aging failure[D]. Xi'an: Xidian University, 2017(in Chinese).

[13] 刘小平, 郭斌, 崔德军, 等. 基于二元维纳过程的小样本齿轮泵可靠寿命预测[J]. 中国机械工程, 2020, 31(11): 1315-1322.
LIU X P, GUO B, CUI D J, et al. Q-precentile life prediction based on bivariate Wiener process for gear pumps with small sample sizes[J]. China Mechanical Engineering, 2020, 31(11): 1315-1322(in Chinese).

[14] 王少萍, 陈仁同, 张超. 飞机液压泵二维性能退化的可靠性评估[J]. 北京航空航天大学学报, 2022, 48(9): 1613-1623.
WANG S P, CHEN R T, ZHANG C. Reliability estimation for aircraft hydraulic pump based on bivariate performance degradation analysis[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(9): 1613-1623(in Chinese).

[15] ZHANG C, CHEN R T, BAI G H, et al. Reliability estimation of rotary lip seal in aircraft utility system based on time-varying dependence degradation model and its experimental validation[J]. Chinese Journal of Aeronautics, 2020, 33(8): 2230-2241.

[16] CHEN R T, ZHANG C, WANG S P, et al. Importance measures for critical components in complex system based on Copula hierarchical Bayesian network[J]. Reliability Engineering & System Safety, 2023, 230: 108883.

[17] SI X S, ZHOU D H. A generalized result for degradation model-based reliability estimation[J]. IEEE Transactions on Automation Science and Engineering, 2014, 11(2): 632-637.

[18] SI X S, WANG W B, HU C H, et al. Remaining useful life estimation based on a nonlinear diffusion degradation process[J]. IEEE Transactions on Reliability, 2012, 61(1): 50-67.

[19] NELSEN R B. An introduction to Copulas[M]. Berlin: Springer, 2006.

[20] JOE H. Multivariate models and multivariate dependence concepts [M]. London: Chapman & Hall, 1997.

[21] NG S K, KRISHNAN T, MCLACHLAN G J. The EM algorithm [M]//GENTLE J E, HÄRDLE W K, MORI Y. Handbook of computational statistics. Berlin: Springer, 2011: 139-172.

Reliability assessment and lifetime prediction for train traction system considering multiple dependent components

TIAN Guishuang，WANG Shaoping*，SHI Jian

(School of Automation Science and Electrical Engineering，Beihang University，Beijing 100191，China)

Abstract：The traction system, serving as the power core of urban rail transit trains, plays a crucial role in ensuring the safe operation of the trains. Reliability assessment and lifespan prediction are investigated in order to tackle the difficulties brought about by the traction system's intricate structure and numerous failure types. The physics of failure model for motor demagnetization and insulated gate bipolar transistor (IGBT) are constructed. The degradation processes for those performance indicators are described by the Wiener process fusing failure mechanism while considering unit-to-unit variability. The Copula function is used to describe the dependent relationship between performance indicators. As for off-line parameter estimation, the Bayesian Markov chain Monte Carlo method estimates unknown parameters. As for online remaining useful life prediction, the algorithm combining Bayesian and expectation-maximization is implemented to update unknown parameters. The proposed model and algorithm are validated by the degradation data of the traction system. The results indicate that the reliability model considering the dependent relationship between the motor and IGBT improves the accuracy of reliability assessment. The remaining useful life prediction accuracy is improved by the parameter updating approach that combines Bayesian and expectation-maximization.

Keywords：train traction system; Wiener process; Copula function; reliability; lifetime prediction

Received：2023-12-09；**Accepted**：2023-12-30；**Published Online**：2024-01-10 14：13

URL：link.cnki.net/urlid/11.2625.V.20240110.1019.002

Foundation item：Beijing Municipal Natural Science Foundation-Fengtai Rail Transit Frontier Research Joint Foundation (L221008)

* **Corresponding author**. E-mail：shaopingwang@vip.sina.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0352

中心分级环形燃烧室液雾点熄火边界试验

王智辉[1,2,3]，张弛[1,2,3]，甘志超[1,2,3]，高安雯[1,2,3]，韩啸[1,2,3,*]，陶雯婕[4]

(1. 北京航空航天大学 航空发动机研究院，北京 102206；2. 北京航空航天大学 航空发动机气动热力国家级重点实验室，北京 102206；3. 北京航空航天大学 先进航空发动机协同创新中心，北京 102206；4. 中国航发商用航空发动机有限责任公司，上海 200241)

摘 要：针对中心分级环形燃烧室液雾点熄火问题，对环形模型燃烧室进行了试验研究。在由16个中心分级燃烧室头部组成的环形模型燃烧室中，在常温常压条件下，对航空煤油RP-3的液雾点熄火过程进行了试验研究。试验测试了该环形燃烧室的点火和熄火边界，获得了贫油点火和熄火边界曲线。在燃烧室压降0.5%～3%范围内，随着压降增大，点火边界油气比呈先降低后升高的趋势，熄火边界油气比随压降增大逐渐降低后基本保持恒定。研究结果表明：通过在燃烧室头部安装导流环能够有效降低环形燃烧室贫油点火边界。使用单反相机记录了环形燃烧室点火和熄火过程，不同压降下的火焰周向传播相似，火焰传播存在周向不对称性，头部熄火顺序则与燃油均匀性相关。

关 键 词：环形燃烧室；中心分级；液雾火焰；点火；熄火

中图分类号：V231.2

文献标志码：A **文章编号**：1001-5965(2025)06-2091-08

由于环形燃烧室具有结构紧凑质量轻、压力损失小等优点，目前先进航空发动机均采用环形燃烧室。燃烧室点熄火性能对航空发动机的起动、高空飞行安全边界有着重大影响，燃烧室的点火可靠性和稳定工作范围决定了发动机的稳定工作边界。环形燃烧室点火过程包含单头部点火及周向的火焰传播过程，对环形燃烧室点熄火过程的研究具有重要价值。

对于环形燃烧室周向点火过程，国内外学者开展了相关研究。巴黎理工大学EM2C实验室发展了包含16个单级径向旋流喷嘴的MICCA环形燃烧室，研究了预混的丙烷/空气点火联焰过程。Bourgouin等[1]通过高速摄像技术研究了不同来流速度工况下，预混丙烷火焰周向传播联焰过程的差异，研究发现火焰从火核到周向联焰所需时间与来流速度成反比，火焰面移动主要受湍流火焰速度、气体膨胀效应的影响。Philip等[2]采用大涡模拟(large eddy simulation, LES)方法对燃烧室内点火过程进行计算，与实验结果有较好的一致性。Prieur等[3]使用MICCA-Spray燃烧室对比了正庚烷、十二烷和丙烷燃料的点火特性差异，研究表明，3种燃料周向点火过程近似，但液体燃料周向点火时间会显著增加。Machover和Mastorakos[4-5]也发展了一种具有18个单级旋流器头部的环形燃烧室，对预混和非预混模式下的甲烷周向点火过程进行了研究，预混燃烧模式下，头部间距对点火特性影响很小，火焰周向传播速度主要受来流速度、当量比和旋流强度的影响。而对于非预混燃烧模式，减小头部间距、增大当量比都会加速火焰周向传播，同时受头部旋流影响，火焰向两侧的传播出现了明显的不对

收稿日期：2023-06-13；**录用日期**：2023-07-21；**网络出版时间**：2023-08-16 09:58
网络出版地址：link.cnki.net/urlid/11.2625.V.20230815.1716.001
基金项目：中央高校基本科研业务费专项资金；国家自然科学基金(52106128)
* **通信作者**. E-mail：han_xiao@buaa.edu.cn

引用格式：王智辉，张弛，甘志超，等. 中心分级环形燃烧室液雾点熄火边界试验[J]. 北京航空航天大学学报，2025，51(6)：2091-2098. WANG Z H，ZHANG C，GAN Z C，et al. Spray ignition and blow-out boundary of centrally staged annular combustor[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2091-2098 (in Chinese).

称性。浙江大学发展了TurboCombo实验平台，燃烧室也具有16个单级径向旋流喷嘴，对丙烷预混火焰的点火周向传播过程进行了研究。令狐昌鸿等[6]对比了先点火后通气和先通气后点火模式下的差异，发现先通气后点火模式具有更短的周向点火时间，点火过程更加稳定。叶沉然等[7-8]研究了周向速度分量和涡轮导叶对周向点火过程的影响，研究表明：大的周向速度分量会显著改变火焰周向传播特性，使得火焰仅单向传播，而涡轮导叶导致的燃烧室非定常流的变化使得周向联焰时间缩短。上述研究以气态燃料的单级旋流火焰为主，与真实航空发动机环形燃烧室存在明显差异。

在工程实践方面，现行民用航空发动机污染排放标准日益严格，先进民用航空发动机普遍采用低排放燃烧技术[9]。其中，采用贫预混燃烧的中心分级结构燃烧室成为主流，其中心为预燃级，采用扩散燃烧模式，用于保证小工况下的燃烧效率，同时为主燃级提供持续的点火源，保证燃烧稳定。主燃级采用预混燃烧模式，降低污染排放。对于中心分级燃烧室，由于主燃级点火时不供油，主燃级气流对预燃级油雾存在遮挡，且对初始火核存在强烈的耗散作用，而环形燃烧室主燃级气流又会阻碍火焰的周向传播，使得点火更为困难。目前针对单头部中心分级燃烧室点火已经开展了较多的研究，而对中心分级环形燃烧室液雾点火的研究还不够丰富。Kobayashi等[10]研究了预燃级旋流角度及唇口长度对中心分级燃烧室点火性能的影响，并提出了通过头部加帽罩从而提升燃烧室点火性能的方案。William[11]研究了Lean-Burn燃烧室点火火核的发展过程，只有当火核能够持续并发展迁移到喷嘴下游才能够成功点火。Wang等[12]通过高速摄像等技术研究了中心分级燃烧室主燃级旋流强度对点火过程的影响，研究发现主燃级流动影响了油气分布最终导致了点火性能差异。Matsuyama等[13]研究了中心分级燃烧室气动结构与油雾分布的关系及其对点火性能的影响。Fu等[14]通过实验和数值方法研究了某中心分级燃烧室点熄火边界及燃烧振荡和排放特性。Wang等[15]在高温高压条件下研究了一中心分级燃烧室火焰和排放特性，提出新的NO_x排放预估模型。并采用中心分级环形燃烧室研究了丙烷非预混火焰点火过程，对点火时间、火焰传播模式和火焰前沿动力学进行了分析。中国科学院工程热物理研究所Gao等[16]发展了一种中心分级环形液雾燃烧室，包含18个中心分级头部，预燃级由两级轴向旋流器和文氏管组成，主燃级为一级径向旋流器。研究发现，火焰周向传播模式主要取决于来流速度和当量比，流速的增加会显著降低火焰高度。随着流速和当量比的变化，将火焰传播分为3种不同的模式。Gao等[17]对周向传播过程中的火焰传播时间进行了研究，说明了来流速度和当量比对不同阶段火焰传播速度的影响，其中强调了雾化的重要影响。丁国玉等[18]研究了某型全环燃烧室点火特性，得到了燃烧室进口参数对成功点火油气比的影响，拟合了点火边界与流速之间的函数关系式。

总的来说，目前国内外均发展了实验室级别的环形模型燃烧室，但研究多集中于气态燃料和单级旋流结构，对于中心分级头部的液雾点火过程研究还不够充分，鲜有对于其熄火过程的研究。本文设计了一种中心分级环形燃烧室，研究了其点熄火边界及头部导流环对点火性能的影响，以及不同压降下点熄火过程的变化。

1 试验系统

1.1 环形模型燃烧室

本文采用的中心分级环形模型燃烧室是从实际工程燃烧室进行简化而来，其具体结构如图1和图2所示。其中，图1表征了单头部中心分级结构及其气动热力特征，中心为预燃级，外围为主燃级，预燃级和主燃级均为一级轴向旋流器结构，旋流方向均为俯视顺时针。预燃级采用离心喷嘴对RP-3燃油进行雾化，喷嘴流量数为1.0 kg/(h·MPa$^{0.5}$)。由于点熄火试验仅涉及预燃级火焰，故主燃级不设置油路及喷嘴。燃烧室单头部直径为30 mm，内外火焰筒半径差为60 mm。

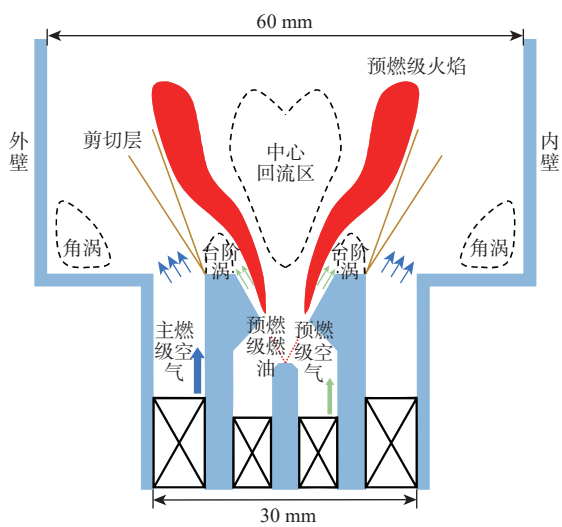


图1 单头部中心分级结构

Fig. 1 Single-injector centrally staged structure

图2为环形模型燃烧室实物，该燃烧室具有16个等间距分布的中心分级头部，需注意的是，图中燃烧室头部均安装了导流环，试验发现其能够优

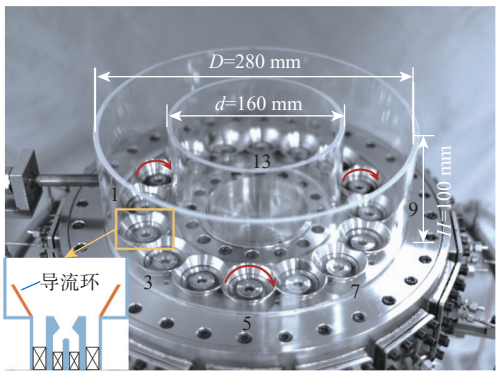


图 2　中心分级环形燃烧室(安装导流环)

Fig. 2　Centrally staged annular combustor (with flow deflector)

化中心分级燃烧室的点火性能。燃烧室内外火焰筒采用 100 mm 高的石英玻璃, 实现环形燃烧室周向点火过程的可视化。将头部按照顺序进行编号: 火器安装位置为 1 号头部, 其余按俯视逆时针分别编号为 2～16 号。

1.2　环形燃烧室试验系统

环形燃烧室试验系统如图 3 所示, 主要分为环形模型燃烧室、气路系统、油路系统、排气系统和测试系统。

气路系统采用空压机压缩空气, 文丘里流量计测量空气质量流量, 测量精度为 1%, 来流经空气流量计后进入机匣, 经匀流段整流后通过燃烧室头部进入火焰筒。油路系统使用高压气瓶进行挤压式供油, 通过调节阀调节燃油流量, 使用科里奥利流量计进行质量流量测量, 测量精度为 0.1%, 安装两位三通阀和旁路喷嘴, 用于点火时燃油从旁路喷嘴向燃烧室内的切换, 燃油经燃油总管后分别进入各头部, 经离心喷嘴雾化后参与燃烧。采用 K 型热电偶测试燃烧室进口温度, 采用差压变送器测量燃烧室压降。试验中采用高能电火花点火器, 点火能量 12 J, 点火频率 6～12 次/s, 点火器轴向位置距头部安装平面为 24.65 mm。采用安装 50 mm 定焦镜头的尼康 Z5 单反相机拍摄点熄火过程, 帧率为 60 帧, 相机布置于点火器对侧即 9 号头部进行斜角俯拍。

试验在常温常压条件下进行, 依次调节空气流量分别使得燃烧室压降为 0.5%～3%, 保持空气流量稳定, 通过调节燃油流量依次进行点熄火试验。

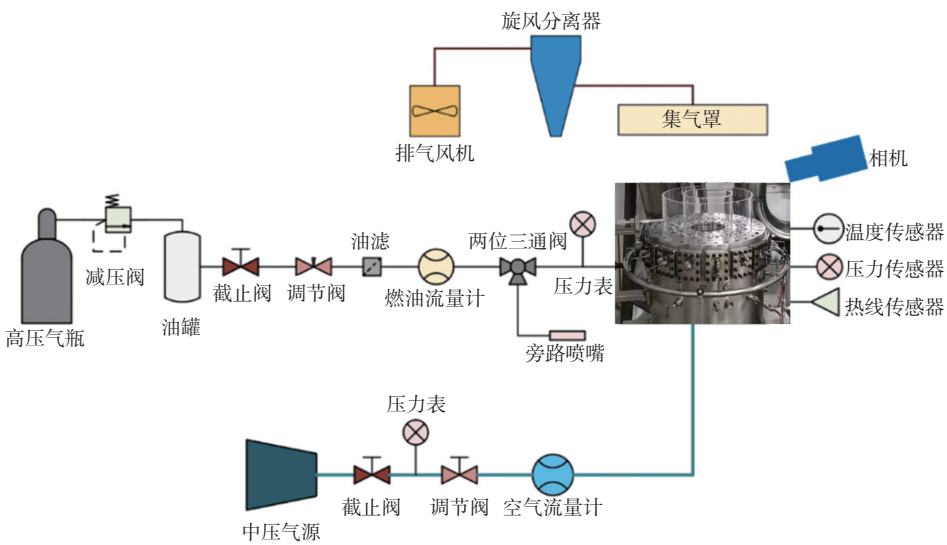


图 3　环形燃烧室试验系统

Fig. 3　Test system for annular combustor

2　试验件的标定

2.1　旋流器有效面积标定

为保证环形模型燃烧各头部气流量均匀性, 对环形燃烧室各头部分别进行有效面积标定工作。将各旋流器先后单独安装于测试工装, 在常温常压条件下测量旋流器前后压降, 进而得到各旋流器总有效流通面积, 测试结果如图 4 所示。可以看出, 13 个头部有效面积偏差在 1% 以内, 剩余 3 个头部有效面积偏差也保持在 2% 以内, 具有较好的一致性, 可保证头部进气的均匀性。

2.2　预燃级喷嘴的标定

环形燃烧室中, 预燃级燃油喷嘴为离心喷嘴, 喷嘴流量数为 1.0 kg/(h·MPa$^{0.5}$), 内部 3 条旋流槽宽度仅为 0.25 mm。由于喷嘴尺寸过小, 微小的加工偏差会对喷嘴性能造成较大的影响, 加工难度大, 难以保证喷嘴很好的均匀性。

将头部各喷嘴与燃油总管连接, 通过油路系统向燃油总管供油, 给定时间内, 用量筒收集各喷嘴喷射的燃油进行体积测量, 测量结果如图 5 所示。

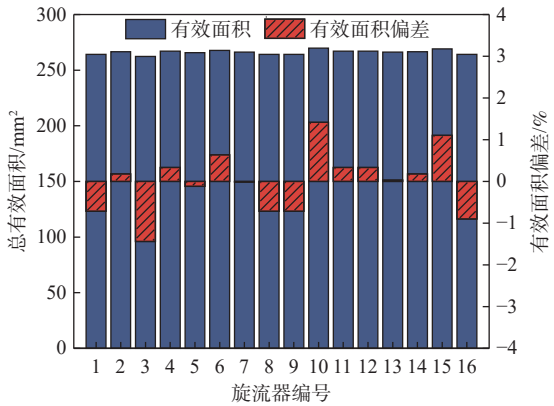


图4 旋流器有效面积标定结果

Fig. 4 Calibration results of effective area of swirlers

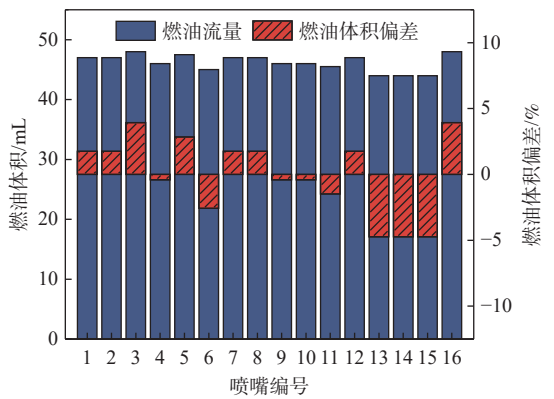


图5 喷嘴均匀性标定结果

Fig. 5 Calibration results of nozzle uniformity

可见，多数喷嘴燃油流量偏差在5%内，有4个喷嘴燃油体积偏差较大，但也在±10%范围内，基本符合测试要求。

3 试验结果与讨论

3.1 火焰形态

图6展示了环形燃烧室在3%压降下点火成功后的火焰照片，火焰为淡蓝色，呈“V”型独立依附于各喷嘴上方。各火焰总体形态如火焰高度等没有明显的差异，即当前空气燃油均匀性可支持后续试验。



图6 中心分级环形燃烧室火焰图像(无导流环)

Fig. 6 Flame image of centrally staged annular combustor (without flow deflector)

3.2 点火边界

点火试验在常温常压条件下进行，试验时调节空气质量流量至恰当的压降，调节燃油流量至设定初值，启动点火器点火，通过两位三通阀将燃油从旁路切换至燃烧室内，头部全部点燃即点火成功，逐渐减小燃油流量至能够点火成功的最小值，在该流量下重复3次，记为贫油点火边界，边界油气比如图7所示。

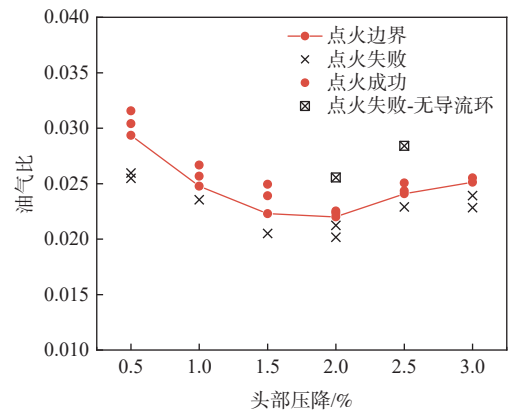


图7 环形燃烧室点火边界

Fig. 7 Ignition boundary of annular combustor

图7包含了头部不安装导流环与安装导流环的点火试验数据。在不安装导流环情况下点火困难，由于供油能力限制难以获取完整点火边界曲线，仅在2个压降下进行了点火试验，均无法点火成功。

安装导流环后，燃烧室油雾分布发生变化，能够成功点火，作用机理如图8所示，导流引导了主燃级气流进一步扩张，增大了射流角度，进而引导油雾输运，改变了燃烧室内流动结构和油雾分布，提高了点火器附近油气比，故而优化了中心分级环形燃烧室点火性能。这在单头部中心分级燃烧室中已有验证[19]，图9为单头部中心分级燃烧室安装导流环前后的油雾场和轴向速度分布，可见导流环起到了引导主燃级气流扩张和预燃级油雾锥扩张的作用，从而提升了中心分级燃烧室点火性能。本文进一步验证了导流环对环形燃烧室点火及火焰

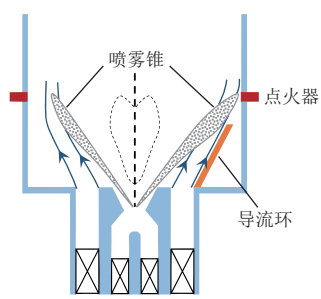


图8 导流环作用机理

Fig. 8 Mechanism of flow deflector

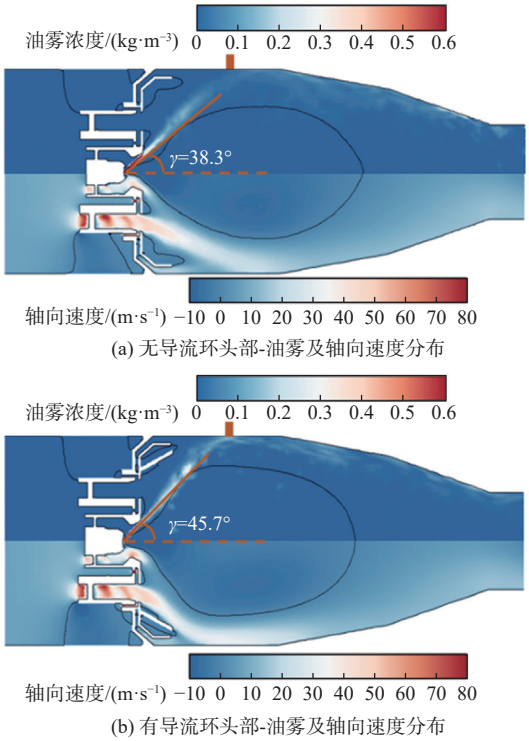


图 9　导流环作用数值计算结果[19]

Fig. 9　Numerical calculation results of flow deflector effect [19]

传播过程的促进作用。

在安装导流环情况下，燃烧室点火边界油气比随着燃烧室压降的增大呈先降低后升高的趋势，这与常规单头部中心分级燃烧室点火边界曲线相似[20]。本文预燃级采用离心喷嘴，燃油流量对雾化质量会有显著影响从而改变点火性能，而来流气动力对雾化效果影响相对较弱。在压降较小时，燃油流量小使得雾化压力小，而随着压降增大，燃油流量增大雾化压力增大，使得离心喷嘴雾化效果变好，燃油颗粒索太尔平均直径（Sauter mean diameter, SMD）减小，从而加速燃油的蒸发混合，降低了点火边界油气比[21]。当雾化压力达到一定值，喷嘴雾化 SMD 不会再有明显变化[22]，但随着压降的进一步增大，空气流量增大，回流区流速随之增大，不利于初始火核生成和火焰的稳定，贫油点火油气比随之增大[20]。

3.3　熄火边界

在点火试验完成后，同样在 0.5%～3% 压降下进行了环形燃烧室熄火试验。根据行业标准 HB 7485—2012[23] 中关于环形及多头部燃烧室熄火判据，1/3 头部熄火即认为熄火，故本文在熄火试验中，熄火头部达到 5 个即认为燃烧室熄火。

燃烧室点火成功后，逐渐降低燃油流量，直至熄火头部达到 5 个，记录此时的燃油流量，即得到当前压降下的熄火油气比，试验结果如图 10 所示。

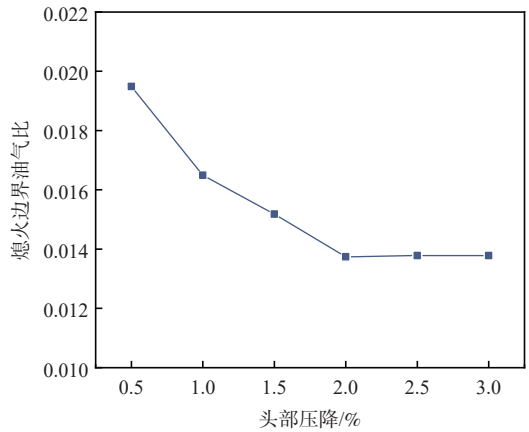


图 10　环形燃烧室熄火边界

Fig. 10　Blow-out boundary of annular combustor

在 0.5% 压降下，熄火边界油气比最高，随着压降的增加，熄火边界油气比逐渐降低，当压降达到 2% 时，熄火油气比基本恒定，不再随压降的升高而降低。在压降较小时，同样压力雾化起主导作用，导致熄火油气比较高，随着压降增大，雾化效果变好，熄火油气比下降。与点火边界相比，燃烧室熄火油气比明显低于点火边界油气比，这是液雾燃烧的典型特征。

将燃烧室熄火边界油气比按照主预燃级气量分配换算为预燃级熄火当量比如图 11 所示。试验中预燃级熄火当量比在 1.7～2.5 之间，远高于碳氢燃料的 0.5 当量比的可燃边界，这与常规单头部中心分级燃烧室预燃级熄火的表现一致[24]，是中心分级燃烧的特点，这一方面是由于主燃级冷气流对预燃级火焰的耗散，另一方面是由于熄火时燃油流量小雾化压力小，离心喷嘴还未达到正常工作状态，油雾颗粒大，混合蒸发水平差，而此时的气流速度高于稳定燃烧边界进而造成熄火。

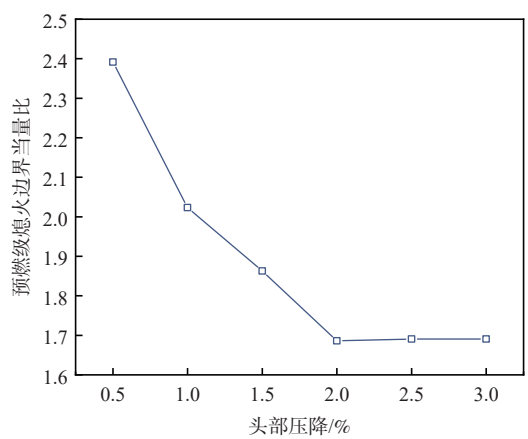


图 11　预燃级熄火当量比

Fig. 11　Blow-out equivalence ratio of pilot stage

3.4　点熄火过程

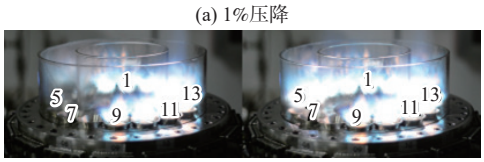
在点火及熄火试验中，使用单反相机拍摄完整

的点熄火过程，研究中心分级环形燃烧室液雾点熄火过程的特性。研究发现，在燃烧室点熄火过程中，在不同压降工况下，火焰周向传播过程存在一定相似性；不同压降下的熄火过程也存在一定相似性。

在点火过程中，火焰传播出现了明显的周向不对称性。图 12 为 1% 压降和 3% 压降下火焰周向传播的部分过程。在不同压降下，1 号头部被点火器成功点燃后，火焰沿顺时针方向迅速经 16 号头部传播至 9 号头部，而逆时针方向火焰传播明显较慢，两侧火焰最终在 7 号和 8 号头部汇合。这种周向不对称性在单级环形燃烧室[25-26]和中心分级环形燃烧室[27]中均有类似发现。



(a) 1%压降



(b) 3%压降

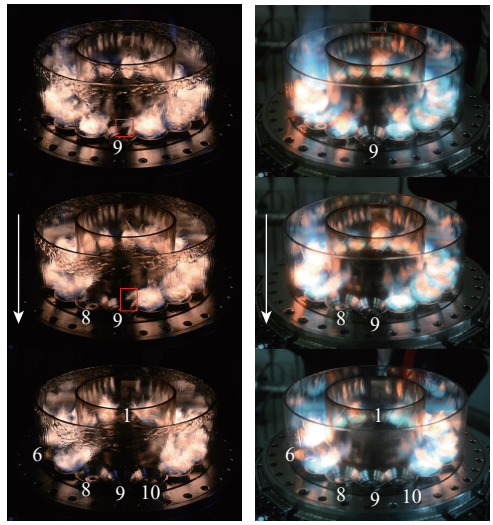
图 12 环形燃烧室火焰周向传播不对称性

Fig. 12 Circumferential propagation asymmetry of flame in annular combustor

这种不对称传播现象与旋流器导致的旋流方向有关，如图 2 所示，顺时针的旋流方向造成了火焰筒外壁面形成了顺时针速度分量，而内壁面则具有逆时针速度分量，Machover 和 Mastorakos[5] 对非预混的甲烷/空气火焰周向传播过程的研究中存在这种周向不对称传播现象，而在对甲烷预混火焰的研究中并没有出现明显的不对称性，故这种环形中的旋流流动和其导致的油雾输运差异是导致环形旋流燃烧室液雾火焰周向传播不对称的原因。

在熄火过程中，头部熄火顺序在不同压降下也具有一致性，燃烧室熄火过程如图 13 所示。图 13(a) 为 1% 压降工况，图 13(b) 为 2.5% 压降工况，图中标注了熄火头部的序号。

由图 13 可知，在不同压降下，头部熄火顺序保持一致，首先为 9 号头部熄火，之后 8 号头部熄火，最后熄火头部达到 5 个，判定为环形燃烧室熄火。这样的熄火顺序与各头部燃油均匀性有关。结合图 5 中喷嘴标定数据，可以看出，9 号头部喷嘴和 8 号头部喷嘴燃流量均明显偏小，导致在降低燃油流量过程中，喷嘴雾化质量恶化程度更加严重，因而率先熄火。1 号和 6 号喷嘴燃油流量也偏小，10 号喷嘴燃油流量虽略偏高，但可能受到 9 号头部



(a) 1%压降 (b) 2.5%压降

图 13 燃烧室熄火过程

Fig. 13 Blow-out process of annular combustor

熄火后的未燃冷空气的淬熄作用，导致熄火。而 4 号由于与其相邻喷嘴燃油流量偏高，释热更强，有助于火焰的稳定。由于头部燃油流量偏低的喷嘴会率先熄火，在喷嘴配置时尽可能提高各头部间燃油的均匀性应能够提升环形燃烧室熄火性能。

在熄火过程中，可以明显观察到未熄头部试图重新引燃熄火头部的现象，如图 13(a) 中 9 号头部上方框中所示，熄火的 9 号头部与 10 号头部之间形成了明显的火焰刷结构，但由于 9 号燃油雾化质量差无法重新引燃。这一现象说明环形燃烧室在熄火过程中，各头部之间互相作为点火/稳火源，存在强烈的相互稳火作用。

4 结 论

采用试验方法，在常温常压条件下对中心分级环形燃烧室进行了点熄火试验，主要结论如下：

1）中心分级环形燃烧室液雾点火边界随压降升高呈先降低后升高的趋势，头部安装导流环能够改变燃烧室内流动结构和油雾分布，从而提升点火性能。液雾熄火边界随压降的升高先降低后基本保持恒定，熄火边界明显低于点火边界。

2）中心分级环形燃烧室液雾火焰传播过程中，会出现明显的周向不对称现象，这与环形旋流流动和液雾燃烧特性相关。

3）中心分级环形燃烧室熄火过程中，头部熄火顺序与燃油均匀性高度相关，各头部之间存在互稳火作用。

本文研究有助于加深对航空发动机中心分级环形燃烧室液雾点熄火过程的认识，对提升航空发动环形燃烧室点熄火性能具有借鉴意义。

参考文献（References）

[1] BOURGOUIN J F, DUROX D, SCHULLER T, et al. Ignition dynamics of an annular combustor equipped with multiple swirling injectors[J]. Combustion and Flame, 2013, 160(8): 1398-1413.

[2] PHILIP M, BOILEAU M, VICQUELIN R, et al. Large eddy simulations of the ignition sequence of an annular multiple-injector combustor[J]. Proceedings of the Combustion Institute, 2015, 35(3): 3159-3166.

[3] PRIEUR K, DUROX D, BEAUNIER J, et al. Ignition dynamics in an annular combustor for liquid spray and premixed gaseous injection[J]. Proceedings of the Combustion Institute, 2017, 36(3): 3717-3724.

[4] MACHOVER E, MASTORAKOS E. Experimental investigation on spark ignition of annular premixed combustors[J]. Combustion and Flame, 2017, 178: 148-157.

[5] MACHOVER E, MASTORAKOS E. Spark ignition of annular non-premixed combustors[J]. Experimental Thermal and Fluid Science, 2016, 73: 64-70.

[6] 令狐昌鸿, 王高峰, 钟亮, 等. 环形旋流燃烧室模型点火过程的实验[J]. 航空动力学报, 2018, 33(7): 1767-1778.
LINGHU C H, WANG G F, ZHONG L, et al. Experiment on ignition process in annular swirling combustor model[J]. Journal of Aerospace Power, 2018, 33(7): 1767-1778(in Chinese).

[7] 叶沉然, 王高峰, 马承飚, 等. 斜喷环流环形燃烧室点火实验研究[J]. 工程热物理学报, 2018, 39(11): 2549-2558.
YE C R, WANG G F, MA C B, et al. Experimental investigations of ignition process in an annular combustor with circumferential flow via oblique injection[J]. Journal of Engineering Thermophysics, 2018, 39(11): 2549-2558(in Chinese).

[8] 叶沉然, 王高峰, 方元祺, 等. 涡轮导叶对环形燃烧室点火的影响[J]. 燃烧科学与技术, 2020, 26(1): 75-80.
YE C R, WANG G F, FANG Y Q, et al. Ignition dynamics in annular combustor with turbine guide vanes[J]. Journal of Combustion Science and Technology, 2020, 26(1): 75-80 (in Chinese).

[9] 张弛, 林宇震, 徐华胜, 等. 民用航空发动机低排放燃烧室技术发展现状及水平[J] 航空学报, 2014, 35(2): 332-350.
ZHANG C, LIN Y Z, XU H S, et al. Development status and level of low emissions combustor technologies for civil aero-engine[J]. Acta Aeronautica et Astronautica Sinica, 2014, 35(2): 332-350(in Chinese).

[10] KOBAYASHI M, OGATA H, ODA T, et al. Improvement on ignition performance for a lean staged low NO*x* combustor[C]//ASME 2011 Turbo Expo: Turbine Technical Conference and Exposition. New York: ASME, 2012: 997-1004.

[11] WILLIAM R R. Experimental investigations into high-altitude relight of a gas turbine[D]. Cambridge: University of Cambridge, 2008.

[12] WANG B, ZHANG C, LIN Y Z, et al. Influence of main swirler vane angle on the ignition performance of TeLESS-II combustor[J]. Journal of Engineering for Gas Turbines and Power, 2017, 139(1): 011501.

[13] MATSUYAMA R, KOBAYASHI M, OGATA H, et al. Development of a lean staged combustor for small aero-engines[C]//ASME Turbo Expo 2012: Turbine Technical Conference and Exposition. New York: ASME, 2013: 211-218.

[14] FU Z B, LIN Y Z, LI L, et al. Experimental and numerical studies of a lean-burn internally-staged combustor[J]. Chinese Journal of Aeronautics, 2014, 27(3): 488-496.

[15] WANG Z C, LIN Y Z, WANG J C, et al. Experimental study on NO_x emission correlation of fuel staged combustion in a LPP combustor at high pressure based on NO-chemiluminescence[J]. Chinese Journal of Aeronautics, 2020, 33(2): 550-560.

[16] GAO W, YANG J H, LIU F Q, et al. Experimental investigation on the flame propagation pattern of a staged partially premixed annular combustor[J]. Combustion and Flame, 2021, 230: 111445.

[17] GAO W, YANG J H, MU Y, et al. Experimental investigation on spark ignition of a staged partially premixed annular combustor[J]. Fuel, 2021, 302: 121062.

[18] 丁国玉, 马丹, 高雅, 等. 某型全环燃烧室点火特性试验[J]. 航空动力学报, 2023, 38(6): 1299-1305.
DING G Y, MA D, GAO Y, et al. Experiment on ignition performance of a full annular combustor[J]. Journal of Aerospace Power, 2023, 38(6): 1299-1305 (in Chinese).

[19] MA J L, HUI X, HAN X, et al. The effect and mechanism of the flow deflector on ignition performance of the centrally staged combustor[J]. Physics of Fluids, 2023, 35(2): 027113.

[20] 王延胜, 林宇震, 李林, 等. 中心分级燃烧室点火性能试验研究[J]. 推进技术, 2016, 37(1): 98-104.
WANG Y S, LIN Y Z, LI L, et al. Experimental investigation on ignition performance of internally-staged combustor[J]. Journal of Propulsion Technology, 2016, 37(1): 98-104 (in Chinese).

[21] 陈坚, 李建中, 袁丽, 等. 雾化特性对喷雾燃烧点火过程的影响[J]. 推进技术, 2017, 38(6): 1318-1326.
CHEN J, LI J Z, YUAN L, et al. Effects of spray characteristics on ignition process in spray combustion[J]. Journal of Propulsion Technology, 2017, 38(6): 1318-1326 (in Chinese).

[22] 张永良, 王宝瑞, 孔文俊, 等. 离心喷嘴实验与流场结构的数值模拟[J]. 工程热物理学报, 2013, 34(4): 760-764.
ZHANG Y L, WANG B R, KONG W J, et al. Experiment and numerical studies on the flow field of a pressure atomizer[J]. Journal of Engineering Thermophysics, 2013, 34(4): 760-764(in Chinese).

[23] 中国燃气涡轮研究院, 中国航空综合技术研究所, 沈阳发动机设计研究所. 航空燃气涡轮发动机燃烧室性能试验方法: HB 7485—2012[S] . 北京: 中国航空综合技术研究所, 2013.
AECC Sichuan Gas Turbine Establishment, AVIC China Aero-polytechnology Establishment, AVIC Shenyang Aircraft Design and Research Institute. Performance test method of combustor for aircraft gas turbine engine: HB 7485—2012[S]. Beijing: AVIC China Aero-polytechnology Establishment, 2013(in Chinese).

[24] 李海涛, 许全宏, 付镇柏, 等. 中心分级燃烧室预燃级贫油熄火性能试验[J]. 航空动力学报, 2014, 29(9): 2188-2194.
LI H T, XU Q H, FU Z B, et al. Experiment on lean blow-out performance of pilot stage in internally-staged combustor[J]. Journal of Aerospace Power, 2014, 29(9): 2188-2194(in Chinese).

[25] XIA Y F, LINGHU C H, ZHENG Y, et al. Experimental investigation of the flame front propagation characteristic during light-round ignition in an annular combustor[J]. Flow, Turbulence and Combustion, 2019, 103(1): 247-269.

[26] ZHAO D M, XIA Y F, GE H W, et al. Simulations of flame propagation during the ignition process in an annular multiple-injector combustor[J]. International Journal of Numerical Methods for Heat & Fluid Flow, 2019, 29(6): 1947-1964.

[27] WANG H, ZHONG L, BARAKAT E, et al. Experimental investigation on the ignition dynamics of an annular combustor with multiple centrally staged swirling burners[J]. Physics of Fluids, 2022, 34(7): 075103.

Spray ignition and blow-out boundary of centrally staged annular combustor

WANG Zhihui[1, 2, 3], ZHANG Chi[1, 2, 3], GAN Zhichao[1, 2, 3], GAO Anwen[1, 2, 3], HAN Xiao[1, 2, 3, *], TAO Wenjie[4]

(1. Research Institute of Aero-Engine, Beihang University, Beijing 102206, China;
2. National Key Laboratory of Science and Technology on Aero-Engine Aerothermodynamics, Beihang University, Beijing 102206, China;
3. Collaborative Innovation Center for Advanced Aero-Engine, Beihang University, Beijing 102206, China;
4. AECC Commercial Aircraft Engine Co., Ltd., Shanghai 200241, China)

Abstract: An experimental study of an annular model combustor was conducted to investigate spray ignition and blow-out problems in a centrally staged annular combustor. In the annular model combustor composed of 16 centrally staged injectors, the spray ignition and blow-out process of aviation kerosene RP-3 was experimentally studied under atmospheric temperature and pressure. Within the range of 0.5%–3% of the pressure drop in the combustor, with the increase in the pressure drop, the fuel-air ratio of the ignition boundary first decreases and then increases, while the fuel-air ratio of the blow-out boundary gradually decreases and then remains constant. The research results show that the lean oil ignition boundary of the annular combustor can be significantly improved by installing the flow deflector at the injectors. The ignition and blow-out process of the annular combustor is recorded with a digital single lens reflex. The circumferential propagation of flame is similar under different pressure drops, and the flame propagation is circumferential asymmetry. The blow-out sequence of the injector is related to the fuel uniformity.

Keywords: annular combustor; centrally staged; spray flame; ignition; blow-out

Received: 2023-06-13; **Accepted**: 2023-07-21; **Published Online**: 2023-08-16 09:58

URL: link.cnki.net/urlid/11.2625.V.20230815.1716.001

Foundation item: The Fundamental Research Funds for the Central Universities; National Natural Science Foundation of China (52106128)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0370

人-车-智共享道路空间行人安全行为建模方法

仵若珺，杨德真*，任羿，贾露露，李晓宾，王自力
(北京航空航天大学 可靠性与系统工程学院，北京 100191)

摘 要：在未来自动驾驶汽车将与传统交通参与者共享道路资源的发展背景下，对人-车-智共享道路空间中的行人安全行为建模，对于自动驾驶汽车的仿真开发和可靠性、安全性测试等工作至关重要。针对该需求，分析行人的安全运动行为，在此基础上改进传统的行人社会力模型，提出共享道路空间中行人与周围其他行人、传统车辆和自动驾驶汽车交互运动的动量模型；采用人-车交互运动的数据集和遗传算法对模型的安全参数进行标定，并搭建典型的共享道路空间场景以验证模型的有效性。仿真结果表明：所提模型能够模拟出行人真实的安全运动行为，用于人-车-智共享道路空间仿真等任务中的模拟效果更佳。

关 键 词：共享道路空间；行人；自动驾驶汽车；安全；动量
中图分类号：U491.2+26; X951
文献标志码：A **文章编号**：1001-5965(2025)06-2099-07

自动驾驶汽车目前在学术界和工业界已取得广泛关注，未来自动驾驶汽车将同传统车辆一起在混合道路上行驶，与其他交通参与者共享道路资源，因此，在自动驾驶汽车投入实际使用前通过仿真等方法确保其在各种道路情况下能够安全、稳定地行驶十分重要。尤其是在中国行人众多的道路上，车辆与行人碰撞引起的人员伤亡是较为突出的交通问题。为保障自动驾驶汽车与行人的交互安全，实现对自动驾驶汽车运行场景高还原度的仿真，研究并构建行人在人-车-智共享道路空间中的安全行为模型是一项关键技术。

经调研，已有研究将行人的行为模型主要分为数据驱动模型和专家模型[1]，其中，数据驱动模型均需要大量的数据训练才能模拟出行人的多样化行为，如长短时记忆网络模型[2-3]、生成对抗网络模型[4-5]等，但目前公开的数据集中关于行人与不同交通参与者的交互行为数据较少且单一，无法满足大规模的训练需求，且这些模型更多地用于短时预测行人未来轨迹，不能实现长时间的实时生成。专家模型可以通过元胞自动机[6-7]、几何速度障碍[8-9]或社会力[10-11]等形式建模。其中，元胞自动机模型和基于速度障碍的几何模型以生成最优无碰撞的轨迹为目标，交互形式固定，无法准确地再现行人现实情况下的复杂行为。基于社会力的模型使用物理力来模拟行人之间的运动和微观的相互作用，是行人模拟中广泛使用的技术。在该模型中，行人被视为一个质点，周围环境对行人的影响用力的形式表示，包括目标吸引力、周围其他物体的吸引力和排斥力，计算得到行人当前时刻受到的合力后可以采用牛顿第二定律计算其加速度，从而输出行人下一时刻的速度和位置。多位学者在最初的社会力模型上进行改进，将使用场景拓展到多种道路使用者[12]，并将决策模型[11]、博弈论[13]等方法融合其中，使得社会力模型能够模拟出多样化的行人行为。

但现有的社会力模型中没有考虑自动驾驶汽车对行人运动的影响，无法应用于未来传统车辆与

收稿日期：2023-06-15；**录用日期**：2023-08-04；**网络出版时间**：2023-09-05 11:52
网络出版地址：link.cnki.net/urlid/11.2625.V.20230905.0928.001
基金项目：可靠性与环境工程技术重点实验室基金 (6142004210108)
* **通信作者**. E-mail：09967@buaa.edu.cn

引用格式：仵若珺，杨德真，任羿，等. 人-车-智共享道路空间行人安全行为建模方法 [J]. 北京航空航天大学学报，2025，51(6)：2099-2105. WU R Y，YANG D Z，REN Y，et al. Modelling method for pedestrian safety behavior in shared road spaces of pedestrian-traditional vehicle-autonomous vehicle[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2099-2105 (in Chinese).

自动驾驶汽车混合行驶的城市道路。此外，社会力模型中周围环境的作用力影响了行人运动加速度的大小，即行人下一时刻的速度由当前时刻的速度和速度变化量加和得到。但在共享空间车辆与行人的交互过程中，行人运动变化较为迅速，下一时刻的速度只由当前环境决定，与之前时刻的速度无关。如在过马路时，行人如果预判到碰撞可能，会立即改变速度大小或方向来避免，且变化时间较短。因此，现有的行人社会力模型需要综合考虑自动驾驶汽车与行人运动行为的特点进行改进。

鉴于此，本文拟通过分析行人在人-车-智共享道路空间中交互运动的安全行为，以行人社会力模型为基础，建立行人运动的动量模型，并引入自动驾驶汽车对行人的影响，最后给出典型场景下的行人运动轨迹生成案例，为人-车-智共享道路空间的仿真提供技术支撑。

1 行人交互运动的安全行为分析

人-车-智共享道路空间中的道路参与者包括行人、传统车辆和自动驾驶汽车，在该环境中，所有参与者都具有平等的道路空间使用权利，无明确的交通规则限制，其运动行为主要受社交礼仪的约束，如在无交通信号灯的交叉路口区域，所有参与者通过互动、协商等方式确定道路使用优先级，从而做出下一步的行为决策，最终在不断的交互中完成行动目标[14]。

共享道路空间中的行人在保证安全的前提下为到达目的地会表现出多样化的行为，行动过程可概括为感知-决策-行动3个部分，如图1所示。

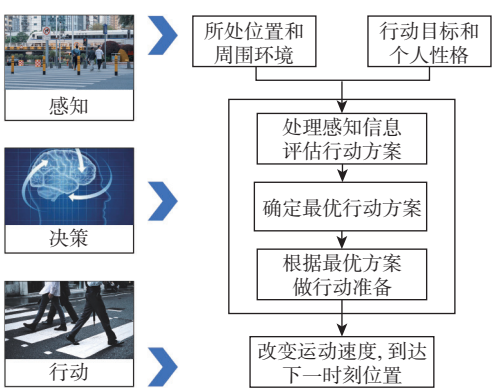


图1 行人交互运动的行为变化过程

Fig. 1 Process of pedestrian interaction behavior

本文在传统行人行为研究的基础上重点分析了行人与自动驾驶汽车间交互运动的安全行为。与传统车辆的交互不同，感知过程中行人与自动驾驶汽车交互时缺少了和驾驶员间的交流，如果车内没有驾驶员或驾驶员注意力不在周围路况上，行人会因为无法交流而表现出较为保守的行为；决策过程中，行人的行为与其对自动驾驶汽车的认知程度有关，如果行人对自动驾驶汽车有较为充分的了解且信任自动驾驶汽车会准确识别出前方行人并做出避撞决策，行人会更自信地通过，如果对自动驾驶汽车保持质疑或不认可的态度，行人会更谨慎地通过；行动过程中，如果行人未接触过自动驾驶汽车，没有交互经验，会根据行人性格的不同表现出不同的犹豫时间。

2 行人安全运动行为模型构建

根据分析得到的行人安全运动行为特点，本文对传统的行人社会力模型[15]进行改进，提出行人动量模型，将行人周围环境的影响用动量的形式体现，使其能够直接反映到行人的速度和位置变化中去。确定行人动量模型共包含4种要素，分别是目标驱动量、周围行人影响量、周围车辆影响量和周围自动驾驶汽车影响量，行人总动量等于4个动量的矢量和，即

$$\boldsymbol{P}_i = \boldsymbol{P}_i^{\mathrm{d}} + \boldsymbol{P}_i^{\alpha} + \boldsymbol{P}_i^{\beta} + \boldsymbol{P}_i^{\eta} \tag{1}$$

式中：$\boldsymbol{P}_i$为行人的总动量；$\boldsymbol{P}_i^{\mathrm{d}}$为目标驱动量；$\boldsymbol{P}_i^{\alpha}$为周围行人影响量；$\boldsymbol{P}_i^{\beta}$为周围车辆影响量；$\boldsymbol{P}_i^{\eta}$为周围自动驾驶汽车影响量。

人-车-智共享道路空间中的行人动量模型如图2所示，其中，行人1的总动量由目标位置吸引产生的动量$\boldsymbol{P}_1^{\mathrm{d}}$、周围行人$\alpha_1$和$\alpha_2$排斥产生的动量$\boldsymbol{P}_1^{\alpha_1}$和$\boldsymbol{P}_1^{\alpha_2}$、周围车辆$\beta_1$排斥产生的动量$\boldsymbol{P}_1^{\beta_1}$和周围自动驾驶汽车$\eta_1$排斥产生的动量$\boldsymbol{P}_1^{\eta_1}$组成。

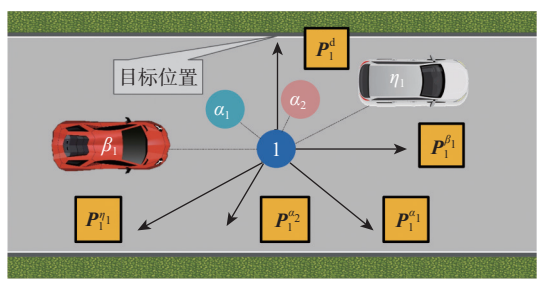


图2 共享道路空间中的行人动量模型

Fig. 2 Pedestrian momentum model in shared road spaces

2.1 目标驱动量

给定一个目标位置，行人的运动方向由当前位置指向目标位置，运动速度为行人的最佳速度，则行人的目标驱动量可以表示为

$$\boldsymbol{P}_i^{\mathrm{d}} = m_i \mu^{\mathrm{d}} \boldsymbol{v}_i^{\mathrm{d}} \tag{2}$$

$$\boldsymbol{v}_i^{\mathrm{d}} = \boldsymbol{v}_i \cdot \frac{\boldsymbol{s}_i^{\mathrm{d}} - \boldsymbol{s}_i^{t}}{\sqrt{\left|\boldsymbol{s}_i^{\mathrm{d}} - \boldsymbol{s}_i^{t}\right|^2}} \tag{3}$$

$$\mu^{\mathrm{d}}=\max\left\{\min\left[\frac{1}{P_2-P_1}\left(\left|\boldsymbol{P}_i^{\beta}+\boldsymbol{P}_i^{\eta}\right|-P_2\right),1\right],0\right\} \tag{4}$$

式中: m_i为行人质量; $\boldsymbol{v}_i^{\mathrm{d}}$为指向目标点的期望速度; $\boldsymbol{s}_i^{\mathrm{d}}$为目标位置; $\boldsymbol{s}_i^{t}$为行人当前位置; $\boldsymbol{v}_i$为行人步行最佳速度; μ^{d}为周围传统车辆影响量和自动驾驶汽车影响量之和$\boldsymbol{P}_i^{\beta}+\boldsymbol{P}_i^{\eta}$的函数, 能够根据$\boldsymbol{P}_i^{\beta}+\boldsymbol{P}_i^{\eta}$的大小调节目标驱动量的大小, 其取值范围为[0,1]; P_1和P_2为 2 个阈值。

2.2　周围行人影响量

周围行人影响量等于该行人周围其他行人对其影响量的总和。根据行人在人群中行走时会保持足够个人空间的社会规则, 周围行人影响量的方向为排斥方向, 即由周围其他行人指向该行人, 其计算式为

$$\boldsymbol{P}_i^{\alpha}=\sum_{j\in Q(i)}\boldsymbol{P}_i^{\alpha,j} \tag{5}$$

$$\boldsymbol{P}_i^{\alpha,j}=m_iU_{\alpha}\exp\left[\frac{d_{ij}^{\mathrm{s}}-d_{ij}}{\sigma_{\alpha}}\right]Q_{ij}\boldsymbol{n}_{ij} \tag{6}$$

$$Q_{ij}=\lambda+(1-\lambda)(1+\cos\varphi_{ij})/2 \tag{7}$$

式中: $\boldsymbol{P}_i^{\alpha,j}$为单个行人的影响量; U_{α}为行人交互影响的安全系数; d_{ij}为行人间的实际距离; d_{ij}^{s}为行人间能接受的最小安全距离; σ_{α}为行人间相互排斥的作用范围; $\boldsymbol{n}_{ij}$为行人影响的方向; Q_{ij}为各项异性参数, 能够根据周围行人的位置方向调整行人影响量的大小, 取值范围为[0,1], 反映了自我行人主要受视野范围内行人的影响, 且影响大小根据角度的增加而减小, 本文采用正弦方式计算各项异性参数, 其中, λ为调节各项异性特性的参数, φ_{ij}为行人位置之间的夹角。

2.3　周围车辆影响量

周围车辆影响量等于该行人周围所有车辆对其影响量的总和。首先, 定义车辆的安全轮廓如图 3 所示, 在车辆原始轮廓的基础上增加一个延伸长度l_{s}形成缓冲区, l_{s}为行人想要与车辆保持的最小安全距离。此外, 由于车辆前方的危险性更高, 在车辆前方增加额外的延伸长度$r_{\mathrm{s}}u_{\mathrm{veh}}$形成三角缓冲区, 其中, r_{s}为影响系数, u_{veh}为车辆的纵向速度, 即车速越快, 缓冲区域越大。

确定行人到车辆安全轮廓的最短距离为两者间距值d_{ic}, 计算车辆对行人的影响量:

$$\boldsymbol{P}_i^{\beta}=\sum_{c\in Q(i)}\boldsymbol{P}_i^{\beta,c} \tag{8}$$

$$\boldsymbol{P}_i^{\beta,c}=m_iV_{\beta}\exp\left[\frac{-d_{ic}}{\sigma_{\beta}}\right]Q_{ic}\boldsymbol{n}_{ic} \tag{9}$$

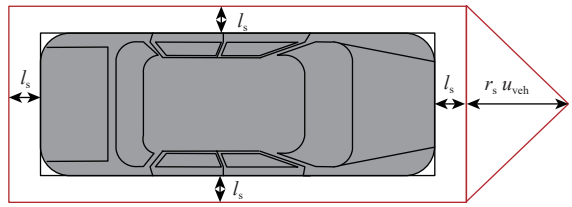


图 3　车辆安全轮廓

Fig. 3　Safety profile of vehicle

$$Q_{ic}=\lambda+(1-\lambda)(1+\cos\varphi_{ic})/2 \tag{10}$$

式中: $\boldsymbol{P}_i^{\beta,c}$为单个车辆的影响量; V_{β}为车辆对行人交互影响的安全系数; σ_{β}为车辆与行人间相互排斥的作用范围; $\boldsymbol{n}_{ic}$为车辆影响量的方向, 即从车辆指向行人; Q_{ic}为各项异性参数, 与周围行人影响量中的Q_{ij}相同, 取值范围为[0,1]; φ_{ic}为车辆和行人位置之间的夹角。

2.4　周围自动驾驶汽车影响量

考虑到行人对传统车辆和自动驾驶汽车反应行为的不同, 在周围车辆影响量的基础上加入影响系数D_{η}, 定义周围自动驾驶汽车影响量为

$$\boldsymbol{P}_i^{\eta}=\sum_{a\in Q(i)}\boldsymbol{P}_i^{\eta,a} \tag{11}$$

$$\boldsymbol{P}_i^{\eta,a}=D_{\eta}m_iV_{\beta}\exp\left[\frac{-d_{ia}}{\sigma_{\beta}}\right]Q_{ia}\boldsymbol{n}_{ia} \tag{12}$$

式中: $\boldsymbol{P}_i^{\eta,a}$为单个自动驾驶汽车的影响量。根据行人运动变化过程中的感知、决策、行动 3 个阶段, 定义影响系数D_{η}为

$$D_{\eta}=1+(\omega_1D_1+\omega_2D_2+\omega_3D_3) \tag{13}$$

式中: D_1、D_2和D_3分别为自动驾驶汽车对行人感知、决策和行动过程的影响因子; ω_1、ω_2、ω_3为对应影响因子的权重系数。当$D_{\eta}>1$时, 自动驾驶汽车对行人的排斥影响相对于传统车辆增加; 当$D_{\eta}<1$时, 自动驾驶汽车对行人的排斥影响相对于传统车辆减小; 当$D_{\eta}=1$时, 自动驾驶汽车对行人的排斥影响与传统车辆相同。

如表 1 所示, 影响系数D_{η}中感知影响因子的取值取决于自动驾驶汽车的外部标志和驾驶模式; 决策影响因子取决于行人对自动驾驶汽车安全性的信任程度和行人的心理特征; 行动影响因子取决于行人做出行动改变的犹豫时间。当影响因子的取值大于 0 时, 自动驾驶汽车对行人的额外影响增大, 行人远离汽车方向的动量增加; 当影响因子的取值小于 0 时, 自动驾驶汽车对行人的额外影响减小, 行人远离汽车方向的动量减小。

在行人的总动量计算完成后, 可以根据动量表达式计算行人的期望速度$\boldsymbol{w}_i$, 即

表1 影响因子在不同情况下的取值范围

Table 1 Range of impact factor in different situations

D_1取值范围	汽车外部标志	汽车驾驶模式
$D_1 \leqslant 0$	外部有显示决策行为的实时标志	有驾驶员且驾驶员与行人有眼神交流
$D_1 > 0$	外部无显示决策行为的实时标志	无驾驶员或有驾驶员但注意力不在周围路况
D_2取值范围	行人信任程度	行人心理特征
$D_2 \leqslant 0$	对自动驾驶汽车了解少，持怀疑态度	保守型性格
$D_2 > 0$	对自动驾驶汽车了解多，相信避撞技术	冲动型性格
D_3取值范围	行人犹豫时间	
$D_3 \leqslant 0$	行动能力强，犹豫时间短，能及时改变行为	
$D_3 > 0$	行动能力弱，犹豫时间长，不能及时改变行为	

$$\boldsymbol{w}_i = \boldsymbol{P}_i / m_i \tag{14}$$

为完善模型还需加入行人的最大速度限制，保证行人行走在安全合理的速度区间内，假设行人最大速度为$v^{\max}$，则可以计算最终的行人运动速度$\boldsymbol{v}_i$和位移$\boldsymbol{s}_i$为

$$\boldsymbol{s}_i = \boldsymbol{v}_i = \boldsymbol{w}_i g(v^{\max}/\|\boldsymbol{w}_i\|) \tag{15}$$

$$g(v^{\max}/\|\boldsymbol{w}_i\|) = \begin{cases} 1 & \|\boldsymbol{w}_i\| \leqslant v^{\max} \\ v^{\max}/\|\boldsymbol{w}_i\| & \text{其他} \end{cases} \tag{16}$$

3 模型参数标定

模型参数标定的目标是找到模型中各项参数的最佳值，使得这些值能够最大限度地减小模拟到的行人轨迹和真实行人轨迹间的误差。本文采用遗传算法对模型中的交互安全系数进行标定，标定数据采用Yang等[16]在美国俄亥俄州立大学控制与智能交通研究(control and intelligent transportation research, CITR)实验室附近停车场采集的数据集，CITR数据集中包含了4种不同的人-车交互场景下的26个轨迹片段，每个片段中的数据包含当前时刻的采样点、车辆和行人的标签及他们在不同时刻上的位置和速度。

遗传算法是一类模仿生物系统中自然选择的进化算法，能够用于寻找复杂系统的最优解，采用遗传算法进行模型参数标定的流程如图4所示。

需要标定的参数包括行人与行人之间交互的安全系数U_α和相互排斥的作用范围σ_α，以及车辆与行人之间交互的安全系数V_β和相互排斥的作用范围σ_β，他们统一构成了参数集Θ，即

$$\Theta = \{U_\alpha,\ \sigma_\alpha,\ V_\beta,\ \sigma_\beta\} \tag{17}$$

本文所采用的CITR数据集中共有26个场景，每个场景中有若干名行人，在参数标定过程中，需要对26个场景中的每个行人进行模拟以得到其模拟轨迹。假设数据集中行人在全部时刻的真实运动位置为$\left(x_i^{\mathrm{r}}(t), y_i^{\mathrm{r}}(t)\right)$，$t \in \{1,2,\cdots,T\}$，采用动量模型计算得到的行人模拟运动位置为$\left(x_i^{\mathrm{s}}(t), y_i^{\mathrm{s}}(t)\right)$，$t \in \{1,2,\cdots,T\}$，$T$为每个场景采样时间的总序列，可以得到该行人真实轨迹和模拟轨迹之间距离的平均值d_i为

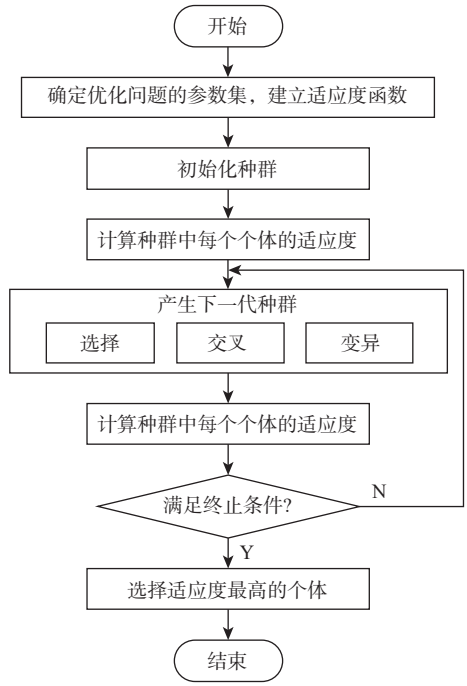


图4 遗传算法标定流程

Fig. 4 Calibration process of genetic algorithm

$$d_i = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\left(x_i^{\mathrm{r}}(t)-x_i^{\mathrm{s}}(t)\right)^2+\left(y_i^{\mathrm{r}}(t)-y_i^{\mathrm{s}}(t)\right)^2} \tag{18}$$

参数标定的目标是模型产生的行人轨迹与原数据集中行人轨迹之间的误差最小，则可以定义遗传算法的适应度函数为所有场景下全部行人模拟轨迹与真实轨迹之间距离平均值的倒数，如式(19)所示，适应度函数的值越大，参数集Θ的模拟效果越好，被遗传到下一代的概率就越大：

$$F(\Theta) = \frac{1}{\sum_{e=1}^{E}\left(\sum_{u=1}^{U} d_i \Big/ U\right) \Big/ E} \tag{19}$$

式中：U为数据集每个场景中行人的数量；E为数据集中场景的数量。

初始种群是随机产生的一组可行解组成的一组个体，根据本文待标定参数数量，可以确定每个个体所代表的染色体长度L为4，定义种群大小M为40。在生成初始种群之前需要确定待标定参数可行解的范围，以避免生成不切实际的值，从而加快算法的运行速度和效率。根据文献[17]中手动校准行人运动模型参数的结果，可以确定本文拟

标定参数的取值范围如下:

$$\begin{cases} 0.1 \leqslant U_\alpha \leqslant 15.1 \\ 0.01 \leqslant \sigma_\alpha \leqslant 0.21 \\ 0.1 \leqslant V_\beta \leqslant 23.0 \\ 0.05 \leqslant \sigma_\beta \leqslant 0.55 \end{cases} \tag{20}$$

设置遗传算法的最大进化代数$T=50$, 交叉概率为 0.9, 变异概率为 0.01, 计算得到每代个体中适应度的最高得分如图 5 所示。

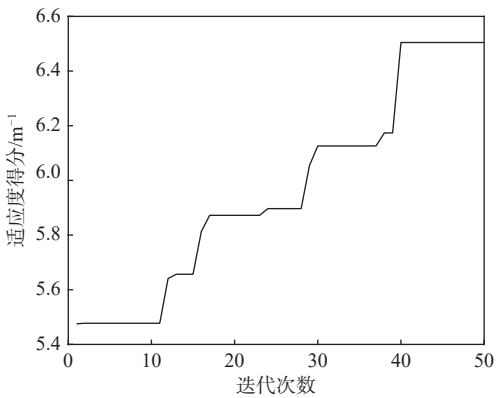


图 5　遗传算法标定结果

Fig. 5　Calibration result of genetic algorithm

已知遗传算法训练得到的适应度得分最终为 6.504 39 m^{-1}, 即采用本文模型产生的行人轨迹与真实行人轨迹之间的距离差为 d_i=1/6.504 39 m= 0.153 742 m, 其在误差允许范围内,此时对应的参数集为$\Theta=\{0.6,0.06,3.02,0.09\}$, 即模型中未标定的参数值分别为: $U_\alpha=0.6$, $\sigma_\alpha=0.06$, $V_\beta=3.02$, $\sigma_\beta=0.09$。

4　案例验证

本文搭建了典型的人-车-智共享道路空间场景进行案例验证, 如图 6 所示, 道路上共有 3 辆汽车和 8 名行人, 所有车辆均沿道路直线行驶, 其中, A 车为自动驾驶汽车, B 车、C 车为传统汽车, 8 名行人分别从道路两侧出发通过马路, 每名行人均由不同颜色的编号表示。

针对以上场景建立坐标系, 定义自西向东为 x 轴正方向, 自南向北为 y 轴正方向, 分别设置所有行人和车辆的初始位置及速度, 采用标定完成后的行人安全行为模型进行模拟, 得到行人完整的轨迹和速度值如图 7 所示。其中, 车辆和行人的起始位置用矩形框标注。

可以看出: 采用安全行为模型生成的行人运动方向和速度大小会根据周围其他交通参与者的影响变化。在可能与车辆发生碰撞之前, 行人①、行人⑧改变运动方向分别绕行至 C 车和 A 车前方并加速通过; 行人②、行人③加速向 C 车后方运动以

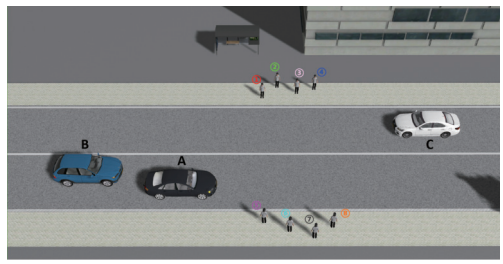
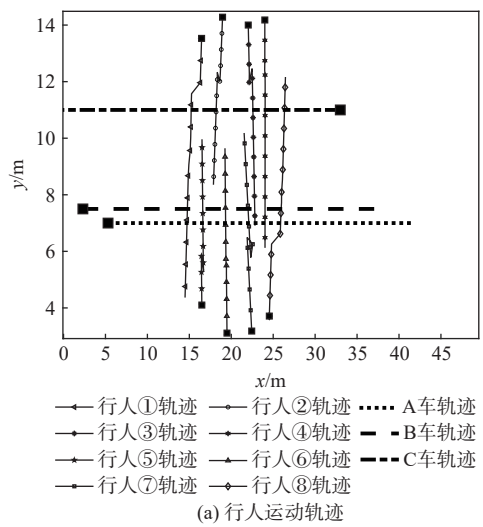
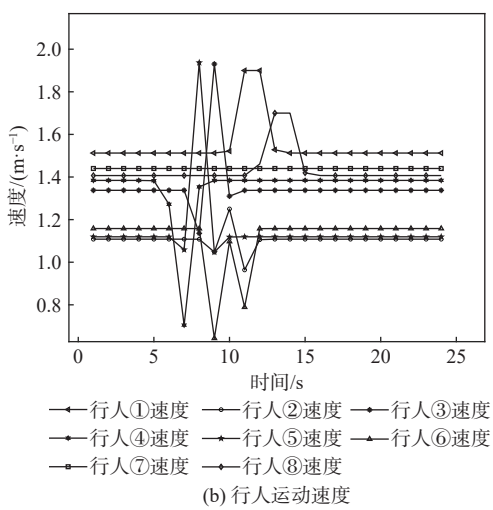


图 6　人-车-智共享道路空间场景

Fig. 6　Shared road space scene of pedestrian-traditional vehicle-autonomous vehicle



(a) 行人运动轨迹



(b) 行人运动速度

图 7　人-车-智交互运动结果

Fig. 7　Interaction result of pedestrian-traditional vehicle-autonomous vehicle

躲避车辆; 行人④减速等待 C 车先通过, 运动方向不变; 行人⑤加速向后方运动等待 A 车和 B 车通过后再前进; 行人⑥在遇到 A 车和 B 车之前分别进行了 2 次减速等待车辆先通过, 运动方向不变; 行人⑦改变运动方向绕行至 A 车和 B 车后避免碰撞, 运动速度不变。以上行为均符合现实生活中行人为保证个人安全的运动习惯。

为对比自动驾驶汽车对行人运动过程的不同影响，在以上模拟实验的基础上，改变A车的影响系数D_η使其分别等于0.5、1和2重新进行模拟，得到模拟结果后计算不同时刻下行人与A车的距离，得到距离的平均值和最小值如表2所示。可以看出：D_η取值越大，A车对行人的排斥作用越强，两者间的距离越大；D_η取值越小，A车对行人的排斥作用越弱，两者间的距离越小。

表2 不同影响系数模拟生成的不同距离

Table 2 Different distances simulated under different impact factors

D_η取值	距离平均值/m	距离最小值/m
0.5	11.281 7	1.272 9
1	11.295 6	1.335 5
2	11.344 7	1.361 8

为验证本文模型优于传统的社会力模型，在相同的场景下采用与动量模型相同的计算方法计算行人各时刻的社会力，根据牛顿第二定律模拟得到行人的运动轨迹如图8所示。

可以看出：在相同的模拟时间内，除受车辆影响较小的行人①、行人②、行人⑧穿过马路到达另一侧外，其余行人均向后运动以躲避车辆，但车辆排斥力的持续作用使得行人在车辆已经远离的情况下也无法及时以最佳速度前进，直到模拟结束时行人才将速度方向修正为指向目标点的方向，行动轨迹变化迟缓，与现实情况差别较大。

同时，本文对比了动量模型和社会力模型模拟结果中行人与3个车辆间的距离平均值和最小值，如表3所示。

从表3可以看出，动量模型模拟生成的行人与车辆间的距离平均值小于社会力模型，但距离最小值大于社会力模型，这表明动量模型在有限的道路空间内模拟得到的行人运动效率和行为安全度均高于传统的社会力模型。

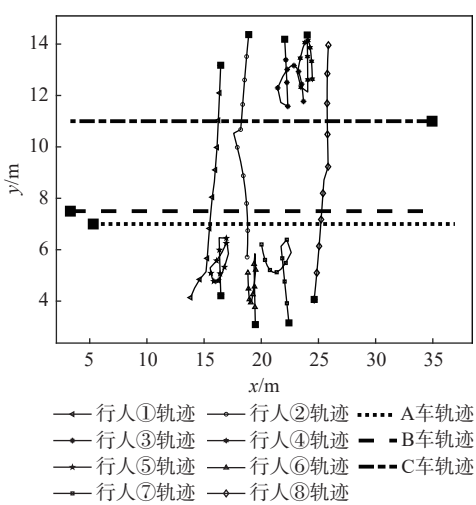


图8 社会力模型模拟结果

Fig. 8 Simulation result of social force model

表3 不同模型模拟生成的不同距离

Table 3 Different distances simulated under different models

模型类型	距离平均值/m	距离最小值/m
动量模型	11.295 6	1.335 5
社会力模型	12.004 9	1.229 1

5 结 论

1）通过分析行人在共享道路空间中的安全行为特点，综合考虑了目标位置、周围行人、周围车辆和周围自动驾驶汽车对行人的影响，搭建了安全行为的动量模型，该模型能够通过行人的周围环境信息实时计算出下一步的速度和位置，并迭代得到完整的行人运动轨迹。

2）对本文模型进行案例验证，结果表明：该模型能够模拟出行人与周围其他交通参与者交互运动时的多种行为，且不同的自动驾驶汽车影响系数模拟出的行人行为不同，模拟结果中行人的行为安全性较高且具有多样性。

3）对比本文模型与传统社会力模型的模拟效果，结果表明：本文模型的模拟效果优于传统的社会力模型，且模拟出的行人行为在有限空间内与车辆的碰撞可能更低，行为更安全。

综上所述，本文模型模拟效果较好，能够用于人-车-智共享道路空间的仿真环境搭建中，解决自动驾驶汽车可靠性测试过程中测试场景单一、固定且不能反映真实路况的问题。

参考文献（References）

[1] CHENG H, JOHORA F T, SESTER M, et al. Trajectory modelling in shared spaces: expert-based vs. deep learning approach? [M]// Multi-Agent-Based Simulation XXI. Berlin: Springer, 2021: 13-27.

[2] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: human trajectory prediction in crowded spaces[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 961-971.

[3] CHENG H, SESTER M. Modeling mixed traffic in shared space using LSTM with probability density mapping[C]//Proceedings of the 21st International Conference on Intelligent Transportation Systems. Piscataway: IEEE Press, 2018: 3898-3904.

[4] PANG S M, CAO J X, JIAN M Y, et al. BR-GAN: a pedestrian trajectory prediction model combined with behavior recognition[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(12): 24609-24620.

[5] HUANG L, ZHUANG J H, CHENG X M, et al. STI-GAN: multimodal pedestrian trajectory prediction using spatiotemporal inter-

actions and a generative adversarial network[J]. IEEE Access, 2021, 9: 50846-50856.

[6] FELICIANI C, CROCIANI L, GORRINI A, et al. A simulation model for non-signalized pedestrian crosswalks based on evidence from on field observation[J]. Intelligenza Artificiale: the International Journal of the AIxIA, 2017, 11(2): 117-138.

[7] WU W J, CHEN R C, JIA H F, et al. Game theory modeling for vehicle-pedestrian interactions and simulation based on cellular automata[J]. International Journal of Modern Physics C, 2019, 30(4): 1950025.

[8] MA Y X, MANOCHA D, WANG W P. AutoRVO: local navigation with dynamic constraints in dense heterogeneous traffic[EB/OL]. (2018-11-14)[2023-06-15]. https://arxiv.org/abs/1804.02915v2.

[9] CHARLTON J, GONZALEZ L R M, MADDOCK S, et al. Simulating crowds and autonomous vehicles[M]// Transactions on Computational Science XXXVII. Berlin: Springer, 2020: 129-143.

[10] ANVARI B, BELL M G H, SIVAKUMAR A, et al. Modelling shared space users via rule-based social force model[J]. Transportation Research Part C: Emerging Technologies, 2015, 51: 83-103.

[11] CHAO Q W, DENG Z G, JIN X G. Vehicle–pedestrian interaction for mixed traffic simulation[J]. Computer Animation and Virtual Worlds, 2015, 26(3-4): 405-412.

[12] RINKE N, SCHIERMEYER C, PASCUCCI F, et al. A multi-layer social force approach to model interactions in shared spaces using collision prediction[J]. Transportation Research Procedia, 2017, 25: 1249-1267.

[13] JOHORA F T, MÜLLER J P. Modeling interactions of multimodal road users in shared spaces[C]//Proceedings of the 21st International Conference on Intelligent Transportation Systems. Piscataway: IEEE Press, 2018: 3568-3574.

[14] SCHÖNAUER R, STUBENSCHROTT M, SCHROM-FEIERTAG H, et al. Social and spatial behaviour in shared spaces[C]//Proceedings of the 17th International Conference on Urban Planning and Regional Development in the Information Society GeoMultimedia. Piscataway: IEEE Press, 2012: 759-767.

[15] HELBING D, MOLNÁR P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51(5): 4282-4286.

[16] YANG D F, LI L H, REDMILL K, et al. Top-view trajectories: a pedestrian dataset of vehicle-crowd interaction from controlled experiments and crowded campus[C]//Proceedings of the IEEE Intelligent Vehicles Symposium. Piscataway: IEEE Press, 2019: 899-904.

[17] YANG D F, ÖZGÜNER Ü, REDMILL K. A social force based pedestrian motion model considering multi-pedestrian interaction with a vehicle[J]. ACM Transactions on Spatial Algorithms and Systems, 2020, 6(2): 1-27.

Modelling method for pedestrian safety behavior in shared road spaces of pedestrian-traditional vehicle-autonomous vehicle

WU Ruoyu，YANG Dezhen*，REN Yi，JIA Lulu，LI Xiaobin，WANG Zili

(School of Reliability and Systems Engineering，Beihang University，Beijing 100191，China)

Abstract: Under the development background that autonomous vehicles will share road resources with traditional traffic participants in the future, pedestrian safety behavior modelling in the shared road spaces of pedestrian-traditional vehicle-autonomous vehicle is crucial to the simulation development and reliability and safety testing of autonomous vehicles. To satisfy this demand, the safe movement behavior of pedestrians was analyzed firstly, and the traditional social force model of pedestrians was improved to build a momentum model of the interaction between pedestrians and other pedestrians, traditional vehicles, and autonomous vehicles in the shared road spaces. Then, a dataset of pedestrian-vehicle interaction motion and a genetic algorithm were adopted to calibrate the safety parameters of the model, and a typical shared road space scene was proposed to verify the effectiveness of the model. The simulation results indicate that the proposed model can simulate the safe movement behavior of pedestrians, and it has better effects in tasks such as shared road space simulation of pedestrian-traditional vehicle-autonomous vehicle.

Keywords: shared road space; pedestrian; autonomous vehicle; safety; momentum

Received：2023-06-15；**Accepted**：2023-08-04；**Published Online**：2023-09-05 11：52

URL：link.cnki.net/urlid/11.2625.V.20230905.0928.001

Foundation item：Science &Technology on Reliability & Enviromental Engineering Laboratory (6142004210108)

* **Corresponding author**. E-mail：09967@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0395

星载 GNSS 反射信号台风检测及风眼位置估计

甄佳欢[1]，朱云龙[1]，杨东凯[1]，张国栋[2]，王峰[1,3,*]

(1. 北京航空航天大学 电子信息工程学院，北京 100191；2. 中国空间技术研究院 通信与导航卫星总体部，北京 100094；
3. 中国气象科学研究院灾害天气国家重点实验室，北京 100081)

摘　　要：全球导航卫星系统(GNSS)反射信号主要利用时延-多普勒图(DDM)的奇异性对台风进行研究。但由于模糊函数的存在，会使得这一奇异特征被模糊从而影响检测性能。通过将DDM反卷积去除模糊函数的影响后再重构归一化散射系数(NBRCS)并将其映射到空间域，对台风事件检测算法及风眼位置估计算法进行研究。根据台风场内的DDM时间序列重构得到空间域NBRCS分布序列，定义台风敏感特征观测量，并基于置信区间距离半径的滑动窗口异常检测方法来检测台风事件；提出一种台风眼位置估计算法，通过模型匹配算法将重构所得空间域散射系数与模型数据集匹配以得到风眼位置。结果表明：台风检测算法的检测性能对信噪比(SNR)敏感，信噪比越高检测性能越好，当SNR为8时，台风检测性能接收机操作特性(ROC)曲线的曲线下面积(AUC)为0.80；在4种SNR下风眼位置估计的平均误差不超过0.3°，均方根误差不超过0.6°。

关　键　词：全球导航卫星系统反射测量技术；台风检测；台风中心；全球导航卫星系统；热带气旋

中图分类号：TN967.1

文献标志码：A　　**文章编号**：1001-5965(2025)06-2106-13

据国家气候中心2023汛期气候预测，2023年在西北太平洋和南海生成的台风个数为24～27，登陆中国的台风个数为6～8[1]。作为一种典型的海洋动力灾害，台风常伴有强降水、狂风、冰雹等灾害天气现象，还会诱发山体滑坡、泥石流、城市内涝等其他灾害，使沿海地区人民生命和财产安全受到威胁[2]。因此，准确检测台风并监测台风的中心位置，提高台风预警能力，具有迫切的现实需求。

传统的海面风场探测手段包括岸基台站、浮标及卫星遥感等。岸基台站和浮标测量精度高，但空间分布稀疏，不适用于监测发展迅速、影响范围广的台风天气[3-4]。目前，实际的台风业务化预报工作依赖于高性能计算机进行数值模式的运算，辅以人工预报员的经验，但是数值模式需要一系列的大气和海洋参数作为初始条件，代入空气运动方程中来求解天气状况。这种方法受模型自身物理框架、初始场的条件和台风的结构等影响，导致模型结果不稳定[5]。卫星遥感具有覆盖面积大、观测范围广、时间分辨率高等优点，能够全天候监测台风，已成为监测台风的主要手段[6]。在台风中心定位方面，目前国内外学者的研究主要基于卫星云图或微波散射计[7]。在实际业务中，预报员通过观测卫星云图序列中云系的发展来确定台风中心。这些方法的局限性在于需要提前判定台风模式，且只能应对某类特定的台风云系[8]。

全球导航卫星系统(global navigation satellite system, GNSS)是以人造地球卫星作为导航基准台

收稿日期：2023-06-19；**录用日期**：2023-09-15；**网络出版时间**：2023-10-16 14:06
网络出版地址：link.cnki.net/urlid/11.2625.V.20231012.1316.002
基金项目：灾害天气国家重点实验室开放课题（2024LASW-B20）；博士后创新人才支持计划(BX20200039)
*通信作者. E-mail：wangf.19@163.com

引用格式：甄佳欢，朱云龙，杨东凯，等. 星载GNSS反射信号台风检测及风眼位置估计[J]. 北京航空航天大学学报，2025，51(6)：2106-2118. ZHEN J H，ZHU Y L，YANG D K，et al. Detection of typhoons and estimation of eye position using satellite-based GNSS-reflectometry[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2106-2118 (in Chinese).

的星基无线电导航系统, 在全球范围内提供全天候的定位、测速和授时服务[9]。20 世纪 90 年代左右, GNSS 反射信号作为一种新型的遥感信号源进入大家的视野, 其波形、幅度、相位等相对于直射信号的变化可以反映地表物理参数的特征[10]。经过 30 余年的发展, GNSS 反射测量技术(GNSS-reflectometry, GNSS-R)已经成为国内外遥感探测和导航技术领域的研究热点之一[11-13]。

GNSS-R 在台风领域有很大的应用潜力。在台风检测方面: 2013 年, Camps 等首次从国际空间站的 GNSS-R 仪器中看到穿越飓风 Sandy 时延-多普勒图(delay-Doppler map, DDM)形状和峰值的演变, 提出通过检测 DDM 峰值(非镜面反射点)时延和多普勒频移的演变来确定飓风眼位置[14] 的方法; 2023 年, 北京航空航天大学王峰等基于台风风场结构的特殊性, 定义了星载全时延多普勒图(full DDM, FDDM)与台风事件有关的特征观测量, 构建了 FDDM 时序台风事件独立检测算法[15]。在台风位置估计方面: 2014 年, Park 等在之前工作的基础上, 通过国际空间站的 GNSS-R 仪器, 提出从包含飓风眼的模拟 DDM 中重建空间域的散射系数以确定飓风眼位置的方法[16]; Li 等基于 Zavorotny 和 Voronovich (Z-V)GNSS 散射模型模拟热带气旋海况下 DDM 图, 并利用空间积分法从模拟 DDM 计算散射系数, 结果表明可以从中看到热带气旋的位置[17-18]; 2016 年美国发射了 CYGNSS 卫星星座, 星载 GNSS-R 在海洋遥感方面进入了一个快速发展的时期。2017 年, Morris 和 Ruf 提出一种利用 CYGNSS 观测资料估算热带气旋综合动能的方法, 并利用模拟 CYGNSS 观测值表征了其性能[19]; 同年, Morris 和 Ruf 利用模拟 CYGNSS 观测值开发了一种估算热带气旋参数的方法, 证明了 CYGNSS 对热带气旋结构参数反演的能力[20]; 2019 年, Mayers 和 Ruf 提出一种估算热带气旋中心位置的方法, 通过文献[20] 的方法得到风场参数化模型后, 对模型中的假定风暴中心进行调整, 使其与观测值的残差最小, 结果与国家飓风中心相比不确定性降低了 4 倍以上[21]。

目前, 国内外学者对于 GNSS-R 在台风领域的研究主要依据台风场景 DDM 在时延-多普勒域的奇异性, 但模糊函数的存在使得这一奇异特征被模糊, 并且很难从 DDM 中确定台风眼的位置。针对 GNSS-R 在空间域对台风事件的应用研究较少的问题, 本文将 DDM 通过约束最小二乘滤波器消除 Woodward 模糊函数的影响后, 通过重构得到空间域归一化雷达散射截面(normalized bistatic radar cross section, NBRCS)时间序列, 基于此定义台风敏感特征观测量来检测台风事件。当检测到台风事件发生后, 将重构 NBRCS 与仿真 NBRCS 数据集进行模型匹配得到最优解, 从而估计台风眼位置。

1　GNSS-R 台风模型仿真及验证

GNSS-R 技术所使用的 L 波段信号对降雨不敏感, 能穿越云层和强降雨, 深入到台风内核的风场; 并且丰富的 GNSS 信号源使得该技术对大面积海况作多方位的持续监测成为可能。

1.1　星载 GNSS-R 几何模型

GNSS-R 常被称为被动式的双/多基遥感系统, 接收机同时接收导航卫星直射信号和反射信号, 并通过协同处理直射、反射信号来获取地球表面物理参数的信息。在对 GNSS-R 进行理论或仿真分析时, 需要构建合理的几何观测场景, 即观测坐标系[22-23], 如图 1 所示。原点 O 位于地球质心, z 轴指向镜面反射点切面的法向, yOz 平面与入射信号共面, 导航卫星与接收卫星均位于 yOz 平面且分居 y 轴正负半轴两侧。

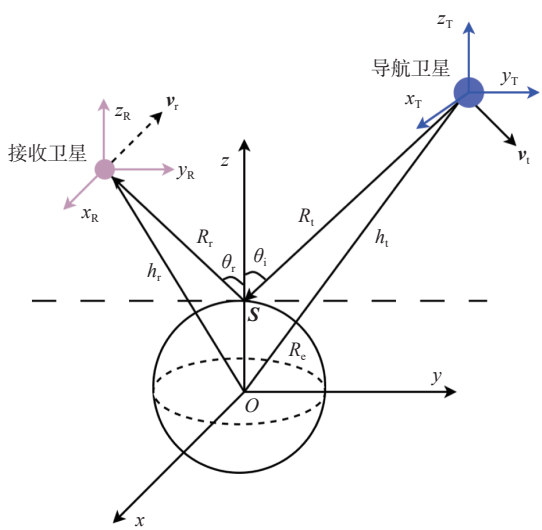


图 1　星载 GNSS-R 空间坐标系

Fig. 1　Spaceborne GNSS-R space coordinate system

导航卫星 $\boldsymbol{T}$、镜面反射点 $\boldsymbol{S}$ 和接收卫星 $\boldsymbol{R}$ 的位置可表示为

$$\boldsymbol{T}=(0,R_t\sin\theta_i,R_e+R_t\cos\theta_i) \tag{1}$$

$$\boldsymbol{S}=(0,0,R_e) \tag{2}$$

$$\boldsymbol{R}=(0,-R_r\sin\theta_r,R_e+R_r\cos\theta_r) \tag{3}$$

式中: θ_i 和 θ_r 分别为导航信号的入射角和散射角; R_e 为地球半径; R_t 和 R_r 分别为导航卫星和接收卫星距镜面反射点的距离; 可分别表示为

$$R_t = -R_e\cos\theta_i + \sqrt{(h_t+R_e)^2 - R_e^2\sin^2\theta_i} \tag{4}$$

$$R_r = -R_e\cos\theta_r + \sqrt{(h_r+R_e)^2 - R_e^2\sin^2\theta_r} \tag{5}$$

其中：h_t和h_r分别为导航卫星和接收卫星距地面的高度。

1.2 台风模型

研究台风风场特征的主要手段是仿真模拟，开展物理实验难度较大。参数化台风风场模型通过建立经验或半经验的风场模型描述台风风场，被广泛应用在台风结构分析、台风灾害研究等方面。Willoughby模型构建的台风二维风场的大小、分布和衰减率与实际台风结构较为吻合，准确度高，故本文选择Willoughby模型[24]。

Willoughby模型中，根据台风最大风速V_{max}和台风眼纬度ϕ可以计算得到最大风速半径R_{max}和参数n：

$$R_{max} = 46.4\exp(-0.015\,5V_{max} + 0.016\,9\phi) \tag{6}$$

$$n = 0.406\,7 + 0.014\,4V_{max} - 0.003\,8\phi \tag{7}$$

式中：R_{max}为最大风速半径；n为幂函数指数。

Willoughby模型将台风风场分为3个区域：风眼区、过渡区和外围区。风眼区即台风眼半径R_{in}内的区域，其风速V_{in}可表示为

$$V(r) = V_{in} = V_{max}\left(\frac{r}{R_{max}}\right)^n \qquad 0 \leqslant r < R_{in} \tag{8}$$

式中：r为到台风中心的距离。过渡区为台风眼与台风外围背景风场之间的区域，其风速V_{tr}可表示为

$$V(r) = V_{tr} = V_{in}(1-w) + V_{out}w \qquad R_{in} \leqslant r < R_{out} \tag{9}$$

式中：R_{out}为过渡区最大半径；w为权重因子，可表示为

$$w(\xi) = \begin{cases} 0 & \xi \leqslant 0 \\ 126\xi^5 - 420\xi^6 + 540\xi^7 - 315\xi^8 + 70\xi^9 & 0 < \xi < 1 \\ 1 & \xi \geqslant 1 \end{cases} \tag{10}$$

$$\xi = \frac{r - R_{in}}{R_{out} - R_{in}} \tag{11}$$

外围区为台风外围稳定的背景风场区域，其风速V_{out}可表示为

$$V_{out} = V_{max}(1-A)\exp\left(-\frac{r-R_{max}}{X_1}\right) + V_{max}A\exp\left(-\frac{r-R_{max}}{X_2}\right) \qquad r > R_{out} \tag{12}$$

式中：A为权重因子；X_1和X_2为指数函数衰减长度，由下式得出：

$$\begin{cases} A = 0.069\,6 + 0.004\,9V_{max} - 0.006\,4\phi \\ X_1 = 317.1 - 2.026V_{max} + 1.915\phi \\ X_2 = 25 \end{cases} \tag{13}$$

台风是运动的。引入台风中心移动的速度和方向参数对Willoughby模型改进优化，可以得到更接近真实场景的移行台风风场，表示为

$$V = \sqrt{(-V_a\cos\theta_r)^2 + \left(V_a\sin\theta_r + \frac{V_sR_{max}r}{R_{max}^2 + r^2}\right)^2} \tag{14}$$

式中：V_a和V分别为加入台风移动因素前后的风场；θ_r为相对于台风运动方向的方位角；V_s为台风移动速度。

1.3 双基电磁散射模型

导航卫星发射的电磁波信号向外辐射时遇到海面会发生反射、散射等现象。大量的理论分析和实验验证证明，除了低入射掠角外，海面反射的GNSS信号大部分是非相干的[25]。

NBRCS又称归一化散射系数，被用于描述海面对入射信号的散射强度，其定义为

$$\sigma_{qp} = \frac{4\pi R^2|\boldsymbol{E}_s|^2}{A|\boldsymbol{E}_i|^2} \tag{15}$$

式中：R为散射点与观测点的距离；$\boldsymbol{E}_s$和$\boldsymbol{E}_i$分别为散射和入射电场；A为信号照射区域面积；下标q和p分别表示入射信号和散射信号的极化方式。

海表面的电磁散射模型可以用来描述海水和海波对电磁波的反射、折射和散射等特性。本文采用双尺度模型(two scale model, TSM)。如图2所示，TSM模型将海面粗糙度划分为大尺度和小尺度，其中，大尺度粗糙度的散射系数由基尔霍夫近似-几何光学(Kirchhoff Approximation-Geometric Optics, KA-GO)模型计算，小尺度粗糙度的散射系数由SPM模型计算。TSM模型的散射系数表示为

$$\sigma_{pq,\mathrm{TSM}} = \sigma_{pq,\mathrm{KA\text{-}GO}} + \sigma_{pq,\mathrm{SPM}} \tag{16}$$

$\sigma_{pq,\mathrm{KA\text{-}GO}}$表示为[26]

$$\sigma_{pq,\mathrm{KA\text{-}GO}} = \pi|\Re_{pq}|^2\frac{|\boldsymbol{q}|^4}{\boldsymbol{q}_z^4}P\left(-\frac{\boldsymbol{q}_\perp}{\boldsymbol{q}_z}\right) \tag{17}$$

式中：$\Re_{pq}$为菲涅尔反射系数；P为海表面均方坡度

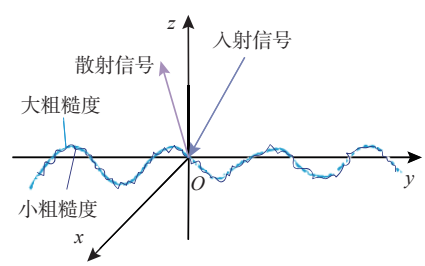


图2 海面TSM模型示意图

Fig. 2 Schematic diagram of TSM model at sea

的概率分布, 假设其为零均值的高斯分布; $\boldsymbol{q}$ 为散射向量; $\boldsymbol{q}_{\perp}$为散射矢量 $\boldsymbol{q}$ 的水平分量; $\boldsymbol{q}_z$为散射矢量 $\boldsymbol{q}$ 的垂直分量。

$\sigma_{pq,\mathrm{SPM}}$表示为[27]

$$\sigma_{pq,\mathrm{SPM}} = 8\left|k^2\cos\theta_{\mathrm{i}}\cos\theta_{\mathrm{s}}\Re\right|^2 S_{\mathrm{sea}}\left(2k\sin\theta_{\mathrm{i}},\varphi_{\mathrm{i},w},u_{10}\right) \tag{18}$$

式中: θ_{i}和θ_{s}分别为入射角和散射角; $S_{\mathrm{sea}}(\cdot)$为海浪谱; $\varphi_{\mathrm{i},w}$为入射信号相对于风向的方位角; u_{10}为距海面 10 m 处的风速; k 为电磁波波数。

1.4　信号模型

GNSS-R 接收机在处理反射信号时, 会产生不同时延和多普勒频率的本地复制信号与反射信号进行相关运算。将相关函数进行非相干累加可得反射信号的相关功率, 也称 DDM, 可表示为[26]

$$\langle|Y_{\mathrm{incoh}}(\tau,f)|^2\rangle = A^2T_{\mathrm{i}}^2\iint\frac{G_{\mathrm{r}}^2(\boldsymbol{r})}{4\pi R_{\mathrm{t}}^2(\boldsymbol{r})R_{\mathrm{r}}^2(\boldsymbol{r})}\times \Lambda^2\left[\delta\tau(\boldsymbol{r})\right]\times\left|S\left[\delta f(\boldsymbol{r})\right]\right|^2\sigma_0(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \tag{19}$$

相关功率是时延τ和多普勒频移f的二维函数。若将海面看作一个系统, GNSS-R 相关功率模型可以看作 Woodward 模糊函数与空间场景函数的卷积[28], 即

$$<|Y(\tau,f_{\mathrm{D}})|^2>=\chi^2(\tau,f)*\sum(\tau,f) \tag{20}$$

式中: $\chi(\tau,f)$为 Woodward 模糊函数, 约为 2 个已知独立函数的乘积:

$$\chi(\Delta\tau(\boldsymbol{r}),\Delta f(\boldsymbol{r}))\cong\Lambda(\Delta\tau(\boldsymbol{r}))S(\Delta f(\boldsymbol{r})) \tag{21}$$

$\sum(\tau,f_{\mathrm{D}})$为空间场景函数, 与散射表面、观测几何结构、天线方向图等相关:

$$\sum(\Delta\tau,\Delta f_{\mathrm{D}}) = T_{\mathrm{i}}^2A\iint\frac{D^2(\boldsymbol{r})\sigma^0(\boldsymbol{r})}{R_{\mathrm{t}}^2(\boldsymbol{r})R_{\mathrm{r}}^2(\boldsymbol{r})}\delta(\Delta\tau)\delta(\Delta f)\mathrm{d}^2\boldsymbol{r} \tag{22}$$

式中: $A=\dfrac{P_{\mathrm{T}}G_{\mathrm{T}}\lambda^2}{4\pi^3}$; T_{i}为相干积分时间; $\Delta\tau=\tau-\tau(\boldsymbol{r})$和$\Delta f=f-f(\boldsymbol{r})$分别为每个表面点相对于镜点的延迟和多普勒频移; D为天线辐射方向图。

实际接收机接收 GNSS 反射信号时会引入噪声, 导致生成的 DDM 中存在加性热噪声和乘性斑点噪声。本文使用的噪声模型为[29]

$$Y_{\mathrm{s}}(\tau,f)=\sqrt{|Y(\tau,f)|^2/2}Z_{\mathrm{s}}+\sqrt{N_0/2}Z_{\mathrm{n}} \tag{23}$$

式中: Z_{s}和Z_{n}分别为具有独立实部和虚部且服从方差为 1 的复高斯分布的随机变量。式 (23) 的前半部分代表乘性斑点噪声, 后半部分代表加性热噪声。

本文基于星载 GNSS-R 仿真平台[30] 分别仿真了台风海况和常规海况下的 DDM, 并分析了台风海况下的 DDM 分布特征。利用 CYGNSS CDR v1.0 中 2018 年 09 月 14 日 CY02 轨道的数据仿真了台风山竹(国际气候管理最佳路径档案风暴标识符 sid=2018250N12170)风场的星载 GNSS 反射信号 DDM, GNSS-R 仿真参数如表 1 所示, 台风“山竹”的参数如表 2 所示, 根据 Willoughby 得到的台风“山竹”的风场如图 3 所示。

表 1　台风海况 GNSS-R 仿真参数

Table 1　GNSS-R simulation parameters of tropical cyclone

参数	描述	取值范围
$\lvert\boldsymbol{T}\rvert$/m	GNSS卫星位置(CY02-DDMI2)	
$\lvert\boldsymbol{R}\rvert$/m	GNSS-R卫星位置(CY02-DDMI2)	
$\lvert\boldsymbol{v}_{\mathrm{t}}\rvert$/(m·s^{-1})	GNSS卫星速度(CY02-DDMI2)	
$\lvert\boldsymbol{v}_{\mathrm{r}}\rvert$/(m·s^{-1})	GNSS-R卫星速度(CY02-DDMI2)	
P_{t}/W	GNSS发射功率	26.8
G_{t}/dB	GNSS发射天线增益	12.1
G_{r}/dB	GNSS-R接收天线增益	1
u_{10}/(m·s^{-1})	台风风场风速	
S/km	散射区	100×100
S_{xy}/km	散射单元	1×1

表 2　台风“山竹”参数

Table 2　Parameters of the tropical cyclone “Mangkhut”

台风名称	最大风速/(m·s^{-1})	最大风速半径/km	台风眼位置
山竹	72.7	55.56	(124.353 7°, 17.151 1°)

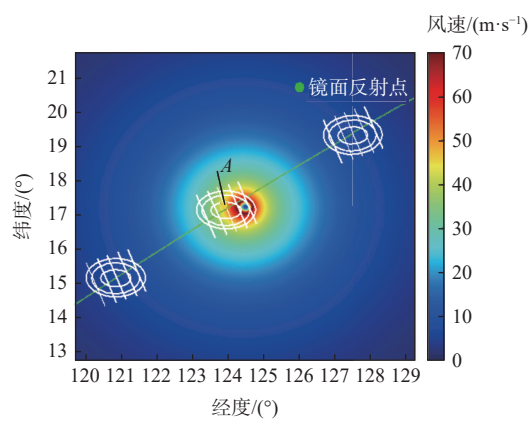


图 3　空间观测范围相对台风风场的分布

Fig. 3　Distribution of spatial observation range relative to tropical cyclone wind field

当 DDM 空间观测范围为 100 km×100 km 时, 其与台风风场的相对分布如图 3 所示。随着镜面反射点穿过台风风场, DDM 空间观测区域所覆盖的台风风场也发生变化。特别地, 在某段镜面反射点序列中, DDM 的空间观测区域会覆盖台风眼, 以镜面反射点 A(124.102 2°, 17.265 3°)为例: A 点距台风眼的距离为(−0.251 5°, 0.114 2°), 其 DDM 观测结果如图 4(a)和图 4(b)所示。

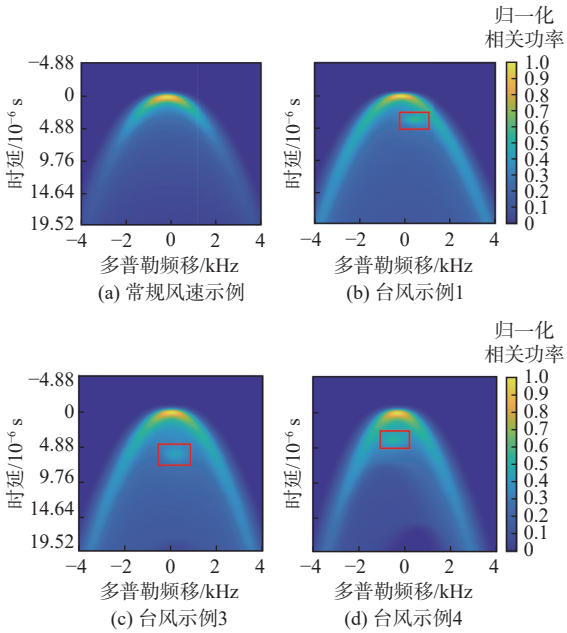


图 4 台风风场和常规风场 DDM 分布

Fig. 4 DDM distribution of tropical cyclone wind field and conventional wind field

图 4(a)和图 4(b)分别是镜面反射点 A 处的台风“山竹”台风风场和常规风场的 DDM，A 点处常规风场的风速为 5 m/s，风向为 0°。图 4(c)和图 4(d)分别是与图 4(a)不同的台风眼位置处的 DDM。由图 4 可以看出，当 DDM 观测区域包含台风眼时，与常规风场相比台风风场 DDM 的分布存在显著不同。具体表现在台风风场 DDM 在除镜面反射点外的某处有高亮，而常规风场则仅在镜面反射点处有峰值。

分析这一高亮区域出现的原因：台风眼处气压最低，风速最小，所以风眼区域内的散射系数高于周围，因而台风风场 DDM 出现除镜面反射点外的高亮，即散射区海面风速分布的不同导致散射系数分布的差异。DDM 的这一奇异性可用于台风检测和台风眼位置估计，但由于模糊函数的存在，使得风眼处相关功率的高亮与镜面反射点相比不够明显，即风眼在 DDM 中的这一奇异特征易被模糊，导致台风不易被检测到。若想使风眼更清晰，可以采取反卷积的方法去除模糊函数的影响。

2 散射系数重构

2.1 DDM 反卷积

由式 (20) 可知，DDM 可看作 Woodward 模糊函数与空间场景函数的卷积。通过反卷积可消除模糊函数的影响，得到散射功率在时延-多普勒域分布。约束最小二乘滤波器[31]是一种反卷积的常用方法：

$$\mathcal{F}[\widetilde{\sum}(\Delta\tau,\Delta f)] = K_{\mathrm{CLS}}\cdot\mathcal{F}[<|Y(\tau,f_{\mathrm{D}})|^2>] \tag{24}$$

式中：$\mathcal{F}$ 表示傅里叶变换；K_{CLS} 为 CLS 滤波器，定义如下：

$$K_{\mathrm{CLS}} = \frac{\mathcal{F}[\chi^2]^*}{|\mathcal{F}[\chi^2]|^2+\gamma p} \tag{25}$$

其中：参数 γ 控制噪声的放大程度和反卷积过程的分辨率；p 为平滑标准函数的傅里叶变换，本文指二阶拉普拉斯算子，如式 (26) 所示：

$$p = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} \tag{26}$$

本节以 CYGNSS 2018 年 09 月 14 日 CY02 轨道为例，基于星载 GNSS-R 仿真平台[30]做仿真，仿真参数如表 1 所示。CY02 轨道相对台风山竹风场的位置如图 3 所示。图 5 为在本节仿真情形下根据收发卫星位置和速度得到的等时延线和等多普勒线的分布情况。其中，等时延线为具有相同时延差的曲线，呈椭圆形；等多普勒线为具有相同多普勒频率差的曲线，呈双曲线形。相邻等时延线间隔为 1 码片，相邻等多普勒线间隔为 1 000 Hz。

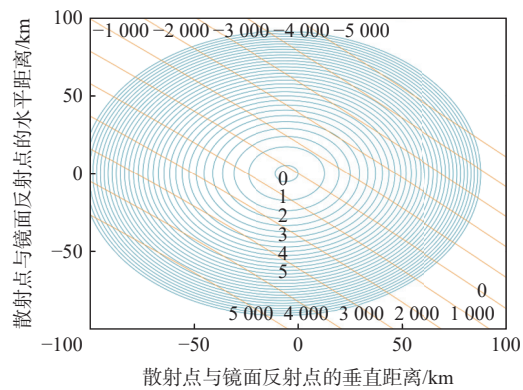


图 5 仿真场景下等时延线和等多普勒线的分布

Fig. 5 Distribution of iso-delay lines and iso-Doppler lines in simulation scenarios

图 6(a)为将图 4(a)的 DDM 通过约束最小二乘滤波器进行反卷积处理后得到的时延-多普勒域散射功率分布，以及不同信噪比下的 DDM 反卷积结果。由图 6 可以看出，散射功率分布中有近似圆形的高亮区域，并且其在时延-多普勒域的位置与图 4 高亮度值位置相同。此高亮代表台风眼处有相对较高的散射功率，这是由于风眼处风速明显低于附近区域，因此其散射系数明显高于附近区域。

由于反卷积运算是一个病态问题(数学上称为不适定)，即使模糊图像中携带很小的噪声，都会导致反卷积后结果与无噪时差异很大。由图 6 可知，随着信噪比降低，高亮区域逐渐扩散。当信噪比

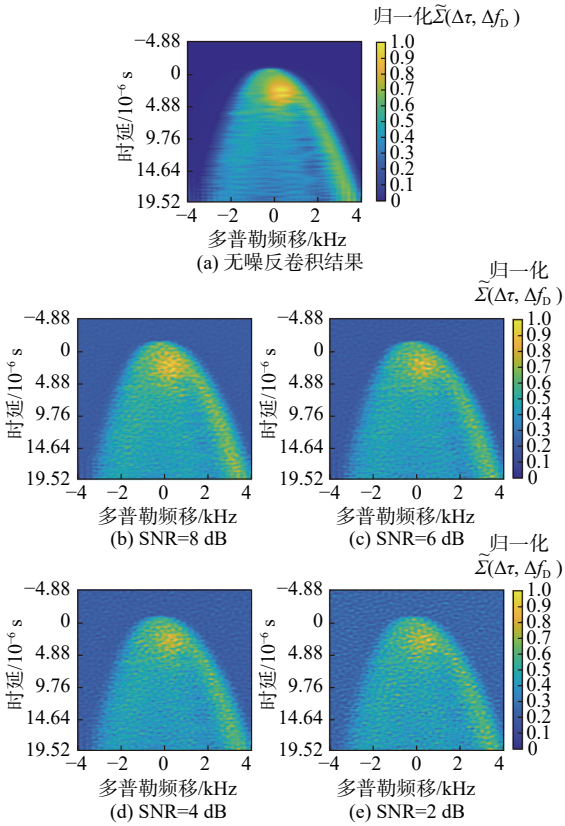


图 6　反卷积结果

Fig. 6　Result of deconvolution

SNR=2 时, 由图 6(e)中虽然仍能看到台风眼处的高亮区域, 但与图 6(b)相比, 高亮区域边缘扩散严重, 形状不规则。这说明反卷积结果对噪声十分敏感。因而本文所提出的台风检测和风眼位置估计算法要求接收机的信噪比较高。

2.2　空间模糊度消除

空间模糊问题是指海面上 2 个不同的物理区域对应于同一个时延-多普勒区域, 如图 7 所示。这意味着每个时延-多普勒区域的散射功率由 2 个独立的物理区域内的散射功率累加得到。在这种情况下, 对每个时延-多普勒域的散射系数, 难以确定 2 个独立的空间域中各自的散射系数。

图 8 为镜点 A 附近某镜点处 CYGNSS 的实测 DDM(A 点 CYGNSS DDM 无效), 该镜点距离 A 点的距离为(0.095 3°, 0.061 4°)。由图 8 可以看出, 实

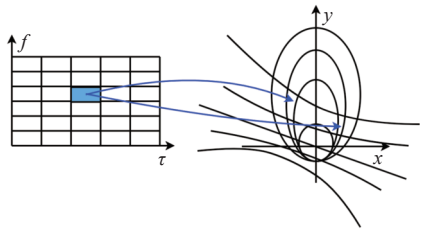


图 7　空间模糊问题

Fig. 7　Problem of spatial ambiguity

测数据中台风风场的“高亮”区域不明显, 这是由于空间模糊度的存在, 即每个时延-多普勒区域内的散射功率由 2 个独立的空间物理区域内的散射功率相加而成。

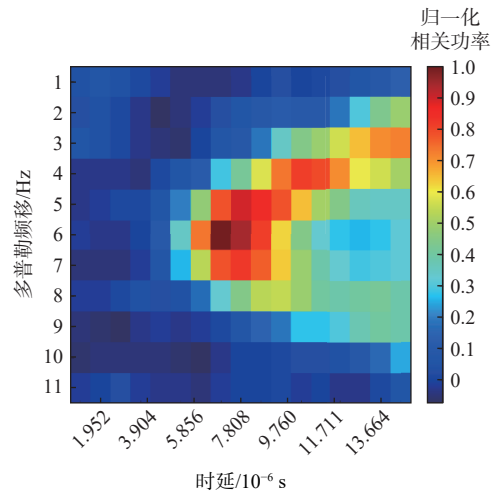


图 8　CYGNSS 实测 DDM

Fig. 8　Actually measured DDM of CYGNSS

为消除空间模糊度, 本文使用模拟的与 SAR 系统相似的观测几何结构[27,32], 即调整接收天线的指向, 使得观测区域只对应于每个时延-多普勒域所对应的 2 个空间区域中的一个, 如式 (27) 所示, 由此可以得到散射系数在空间域的分布情况:

$$\sigma^0(\boldsymbol{r}(\Delta\tau,\Delta f_{\mathrm{D}})) \to \sigma^0(\boldsymbol{r}) \tag{27}$$

由于 CYGNSS 接收天线尚不具备消除实测数据空间模糊度的能力, 故本文采用仿真的形式验证由 DDM 重构 NBRCS 的方法。由 2.1 节可得到去模糊后的 DDM, 即时延-多普勒域散射功率的分布。通过修正收发机几何结构、接收天线增益及散射区域面积等参数的影响, 得到时延-多普勒域散射系数的分布, 进一步建立时延-多普勒域与空间域的映射关系来求解空间域的散射系数分布:

$$\sigma^0(\boldsymbol{r}(\Delta\tau,\Delta f_{\mathrm{D}})) = \sum(\Delta\tau,\Delta f_{\mathrm{D}})\frac{1}{|J(\Delta\tau,\Delta f_{\mathrm{D}})|}\times \frac{R_{\mathrm{i}}^2(\boldsymbol{r}(\Delta\tau,\Delta f_{\mathrm{D}}))R_{\mathrm{r}}^2(\boldsymbol{r}(\Delta\tau,\Delta f_{\mathrm{D}}))}{T_{\mathrm{i}}^2AD^2(\boldsymbol{r}(\Delta\tau,\Delta f_{\mathrm{D}}))} \tag{28}$$

当镜面反射点轨迹穿过台风风场时, DDM 空间观测区域内的风场发生变化。图 9(a)为镜面反射点 A 处的观测风场。图中呈圆形、风速低的区域为台风眼。图 9(b)～图 9(e)为基于台风场景下的海面坡度概率密度函数(slope probability density function, SPDF)[31] 及 TSM 电磁散射模型来计算观测区域内的各网格点处的 NBRCS, 即散射系数真值分布。由图 9(b)可知, 与风场分布相反, 圆形台风眼区域的 NBRCS 高于附近值。图 9(c)～图 9(e)

分别是无噪、高信噪比和低信噪比条件下的空间域NBRCS映射结果。与图9(b)NBRCS真值对比可发现：无噪时在空间域中可以清晰地看到近似圆形的高亮区域，即台风眼；当SNR=8时，从图9(d)也可以观察到近似圆形的高亮区域，台风眼的表现比较明显；但当信噪比较低(SNR=2)时，虽然从图9(e)中仍能看到空间域的高亮，但与图9(c)和图9(d)相比，该高亮区域发生扩散且形状不规则。这说明当信噪比较低时，不仅反卷积结果受噪声影响较大，空间域NBRCS也会失真。需要指出的是，图9中的空白区域是多普勒频移±5 kHz以外区域，不作为研究区域讨论。

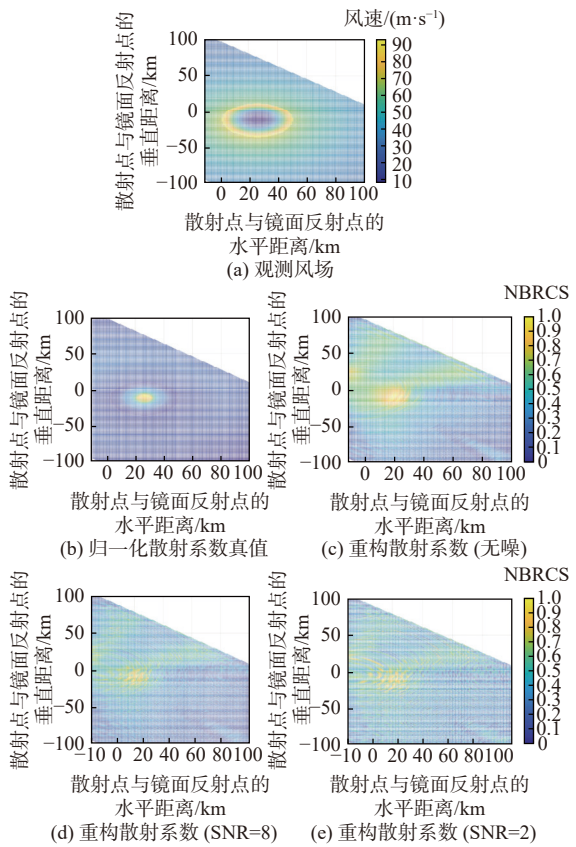


图9 空间域NBRCS映射结果

Fig. 9 Mapping results of spatial domain NBRCS

3 台风检测

针对星载GNSS-R台风事件检测问题，本节从由DDM重构的空间域散射系数出发，定义台风事件敏感特征观测量，研究台风事件检测算法。

3.1 台风事件检测算法

由图9可以看出，空间域中台风风场散射系数分布的显著特点是台风眼处的高亮，这一特点会影响散射系数分布的方差。为表征这一特点，定义空间域内的归一化散射系数方差D_σ为台风敏感特征观测量。假设空间域内的归一化散射系数为σ，空间域内有N个网格点，则台风敏感特征观测量可表示为

$$D_\sigma = \frac{1}{N-1}\sum_{j=1}^{N}\left(\sigma_j - \frac{1}{N}\sum_{i=1}^{N}\sigma_i\right)^2 \tag{29}$$

由于不同台风各个轨道的散射系数值域有差距，这些差距可能对衡量台风数据有影响，故将D_σ的值域进行规范化处理，如式(30)所示：

$$\hat{y}_i = \frac{y_i - \min(Y)}{\max(Y) - \min(Y)} \qquad i = 1,2,\cdots,N \tag{30}$$

式中：y_i为台风敏感特征观测量；Y为台风敏感特征观测序列。

将一条镜面反射点轨迹对应的DDM序列称为一个DDM时间序列。根据式(29)和式(30)处理穿越台风风场内的DDM事件序列，可得到对应的台风敏感特征观测量D_σ序列。

图10(a)为通过处理CYGNSS CDR v1.0中2021年08月18日CY02DDMI4、CY06DDMI1、CY07DDMI1和CY07 DDMI2 4条轨道的数据计算得到的台风Linda(sid=2021222N12262)的规范化台风敏感特征观测量随镜面反射点序列的变化；图10(b)为通过处理CYGNSS CDR v1.0中2019年10月08日CY01 DDMI2、CY02 DDMI2、CY03 DDMI2和CY05 DDMI1 4条轨道的数据计算得到台风海贝思(sid=2019278N16165)的规范化台风敏感特征观

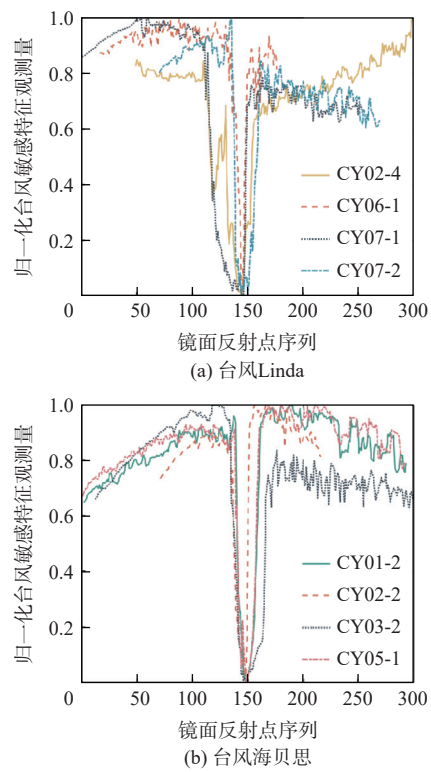


图10 特征观测量D_σ变化趋势

Fig. 10 Changing tendency of D_σ

测量随镜面反射点序列的变化。据 DDM 时间序列计算得到的 D_σ 序列, 称为 D_σ 时间序列。由图 10 可以看出, D_σ 时间序列的变化趋势为: 当镜面反射点穿越台风外围大风区时, D_σ 随之缓慢地变化; 在镜面反射点靠近台风眼时, D_σ 突然下降; 随着镜面反射点远离台风眼, D_σ 又恢复至原来缓慢变化的趋势。因此, 可以通过检测 D_σ 时间序列异常来检测台风事件。

本文选择基于置信区间距离半径的滑动窗口时序异常检测方法。根据时序 DDM 得到的长度为 n 的规范化 D_σ 时间序列可表示为

$$Y=\{\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_i,\cdots,\hat{y}_n\},\qquad i=1,2,\cdots,\ n \tag{31}$$

采用长度为 $l(l<<n)$ 的滑动窗口对该时间序列进行等长度的分割, 之后以步长为 r 连续滑动 $(n-l)/r$ 次窗口后, 得到 $(n-l)/r+1$ 个子序列, 记为

$$Y_{\text{cut}}=\{(\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_l)\cdots(\hat{y_{n-l}},\cdots,\hat{y}_n)\}\qquad i=1,2,\cdots,n \tag{32}$$

置信区间所展现的是真值以一定的概率落在测量结果周围的程度。采用子序列的斜率的置信区间距离半径作为其特征表示能够准确的描述子区间的结构特征[33], 其提取步骤如下。

以宽度 l=10、步长 r=1 的滑动窗口对特征观测量子序列进行分割, 设子序列集合为

$$Y_{\text{cut}}=\{Y_1,Y_2,\cdots,Y_j,\cdots,Y_{(n-l)+1}\} \tag{33}$$

计算子序列 Y_j 中相邻点的斜率为

$$xlv_i=\frac{Y_j(i)-Y_j(i-1)}{i-(i-1)}\qquad i=2,3,\cdots,l \tag{34}$$

进一步可得置信区间距离半径为

$$d_j^{\text{r}}=\left|\overline{C}_j^{\text{l}}-\underline{C}_j^{\text{l}}\right|\Big/2 \tag{35}$$

$\overline{C}_j^{\text{l}}$ 和 $\underline{C}_j^{\text{l}}$ 分别为置信上限和置信下限:

$$\begin{cases}\overline{C}_j^{\text{l}}=\bar{s}_j+\dfrac{\sigma_j}{\sqrt{l}}Z_{\alpha/2}\\ \underline{C}_j^{\text{l}}=\bar{s}_j-\dfrac{\sigma_j}{\sqrt{l}}Z_{\alpha/2}\end{cases} \tag{36}$$

式中: $\bar{s}_j$ 和 σ_j 分别为子序列的均值和方差; Z 为满足标准正态分布的随机变量; α 为置信水平, 取 C^{l} 为 95% 置信区间, 则 $\alpha=0.05$。

图 11 为根据式 (31)～式 (36) 处理 3.1 节中台风 Linda 和台风海贝思的规范化台风敏感特征观测量 D_σ 所得到的置信区间距离半径的变化趋势。由图 11 可知, 镜面反射点穿过台风外围风场时, D_σ 没有剧烈波动, 置信区间距离半径 d_j^{r} 也缓慢变化; 随着镜面反射点处逐渐靠近台风眼区域, 由图 10 可知 D_σ 发生异常, 而 d_j^{r} 也骤然上升; 当镜面反射点远离台风眼时, 与 D_σ 相同, d_j^{r} 也恢复至缓慢变化的状态。

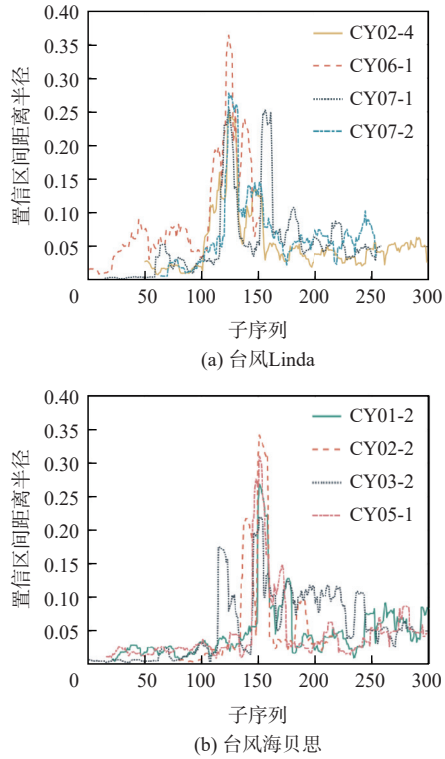


图 11　子序列置信区间距离半径变化

Fig. 11　Changing tendency of subsequence confidence interval distance radius

根据子序列置信区间半径这一特征可以建立异常检测模型。设子序列 Y_j 的置信区间距离半径为 d_j^{r}。通过将其与设定的阈值 γ 相比较:

$$d_j^{\text{r}}>\gamma \tag{37}$$

当 $d_j^{\text{r}}>\gamma$ 时, 表明此子序列含有异常值, 即可判定台风眼的出现, 反之则无。

3.2　台风检测仿真验证

如表 3 所示, 选取 2018—2021 年的 10 个典型的台风作为样本, 根据 Willoughby 模型生成仿真台风风场, 筛选穿过台风眼附近的 CYGNSS 轨道, 并按照 3.1 节的步骤得到若干组台风敏感特征观测量 D_σ 时间序列, 而后求得各 D_σ 时间序列对应的置信区间距离半径 d_j^{r} 时间序列, 通过检测 d_j^{r} 时间序列的异常来检测台风事件的发生。

距离判决阈值对台风事件的检测性能有很大的影响。本文利用接收机操作特性(receiver operating characteristic, ROC)曲线来评价不同阈值下台风事件检测算法的性能。

图 12(a)为不同阈值下台风事件检测算法检测性能的 ROC 曲线。曲线的横坐标代表虚警概率, 纵坐标代表检测概率, 这代表着 ROC 曲线越接近 (0, 1)点, 台风事件检测算法的效果越好; ROC 曲线下面积 (area under curve, AUC) 越大证明性能越

表3 台风样本

Table 3 Tropical cyclone samples

序号	台风编号	台风名称
1	2018250N12170	台风“山竹”
2	2021197N19135	台风“烟花”
3	2021266N11149	台风“Mindulle”
4	2021222N12262	台风“Linda”
5	2019278N16165	超强台风“海贝思”
6	2018263N12146	台风“潭美”
7	2021172N11147	台风“Champi”
8	2020299N11144	超强台风“天鹅”
9	2019268N10155	台风“米娜”
10	2018263N12146	台风“玛莉亚”

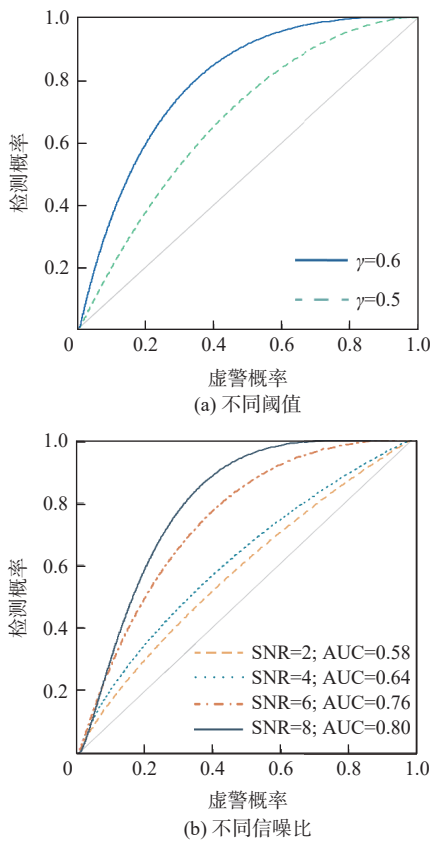


图12 台风检测算法ROC曲线

Fig. 12 ROC curve of typhoon detection algorithm

好。由图可知，2个阈值下的ROC曲线相比，当检测阈值$\gamma=0.6$时，其ROC曲线更靠近(0, 1)点，曲线下的AUC更大，所以检测性能更好。

图12(b)为不同信噪比下台风检测算法检测性能的ROC曲线及相应信噪比下ROC曲线的AUC值。可以看出，本文算法对信噪比的变化比较敏感，信噪比越高检测性能越好；当信噪比较低时，检测性能很差。例如，在SNR=8的条件下，当检测概率为0.33时，虚警概率为0.82，AUC为0.800 8，检测性能较好；随着信噪比的降低，ROC曲线逐渐远离(0, 1)，AUC逐渐减小，说明检测算法的性能逐渐降低。特别地，当SNR=2时，算法的AUC为0.578 8，检测准确性很差，不能用于台风检测。当SNR≥8时，检测性能较好，可用于检测台风事件。

4 台风眼位置估计

4.1 模型匹配联合估计算法

根据2.3节可得到由DDM重构的NBRCS空间域分布，其空间分辨率为1 km×1 km。模型匹配算法流程如图13所示。模型NBRCS数据集由以下方法得到：首先，将台风眼位置以0.1°×0.1°的分辨率划分，根据Willoughby模型可生成不同的风眼位置对应的仿真台风风场，该风场网格的空间分辨率为1 km×1 km。根据双基电磁散射模型可计算不同风眼位置对应的风场中各个网格的NBRCS，由此得到模型NBRCS数据集。将此数据集作为模型匹配的参考数据。

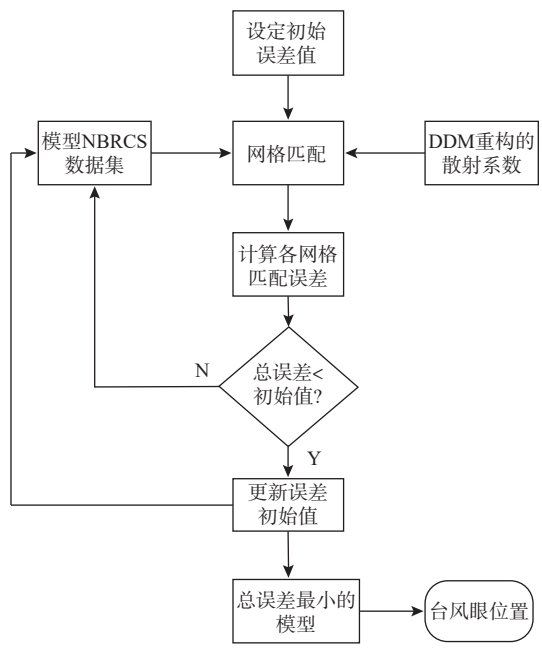


图13 模型匹配算法流程

Fig. 13 Flow chart of model matching algorithm

由单个DDM估计风眼位置的过程如下：以图4(a)中A点对应的台风风场DDM为例，其重构所得空间域散射系数分布如图9所示，将其输入模型匹配算法中，与其匹配的部分模型NBRCS数据集如图14所示，匹配结果如表4所示。

由表4可知，图14(c)匹配误差最小，所以由此DDM估计的风眼位置为(124.253 7°, 17.251 1°)。当穿过台风风场的镜面反射点序列靠近台风眼时，其中某段DDM序列的空间观测区域可能会将台风眼包含在内，将此段DDM序列记为{DDM1, DDM2, …, DDMl}(不同CYGNSS轨道l可能不

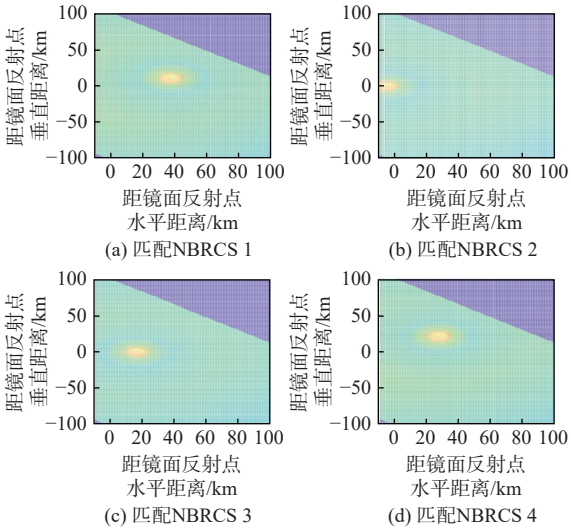


图 14　部分模型 NBRCS 示意图

Fig. 14　Part of the model NBRCS schematic

表 4　匹配结果

Table 4　Matching result

图像	对应风眼经度/(°)	对应风眼纬度/(°)	图像的均方根误差
图14(a)	124.453 7	17.351 1	51.48
图14(b)	124.053 7	17.251 1	49.45
图14(c)	124.253 7	17.251 1	48.73
图14(d)	124.353 7	17.451 1	52.64

同)。按照 3.1 节的步骤把{DDM1, DDM2, …, DDMl}重构得到的空间域散射系数分布序列作为模型匹配算法的输入, 可得到对应的风眼位置估计序列$\{(x_1,y_1),(x_2,y_2),\cdots,(x_l,y_l)\}$。

将这段位置估计序列的结果取平均, 最终得到台风眼位置估计结果:

$$\begin{cases} x=\sum_{i=1}^{l} x_i \\ y=\sum_{i=1}^{l} y_i \end{cases} \tag{38}$$

4.2　位置估计仿真验证

国际气候管理最佳轨迹档案(international best track archive for climate stewardship, IBTrACS)是一个全球热带气旋最佳轨迹追踪数据集, 其数据包含台风的发生频次、发生位置、发生强度、移动路径等参数。IBTrACS 通过一个区域提供多个源的数据, 简化了机构间的数据比较, 是包含最完整的全球历史热带气旋数据集。故本文选择 IBTrACS 数据作为台风数据的参考数据。

以表 3 中的台风样本为例, 筛选包含台风眼的 DDM 时间序列并利用模型匹配联合估计算法来估计各台风的风眼位置。为验证定位结果的准确度, 将其与 IBTrACS 数据集对比并做出误差统计, 结果如表 5 所示。

表 5　时序 DDM 联合估计结果

Table 5　Results of time series DDM joint estimation

台风	平均误差/(°)		均方根误差/(°)	
	经度	纬度	经度	纬度
1	0.269	0.077	0.457	0.410
2	0.322	0.259	0.564	0.419
3	−0.075	0.575	0.141	0.591
4	0.257	0.000	0.470	0.800
5	−0.144	0.606	0.179	0.655
6	−0.275	0.100	0.534	0.660
7	−0.270	−0.530	0.395	0.269
8	−0.275	0.567	0.534	0.568
9	0.295	−0.195	0.542	0.380
10	0.490	0.148	0.715	0.415

以台风“潭美”为例, 不同信噪比下风眼估计结果的误差绝对值与均方根误差随信噪比的变化如表 6 所示。

表 6　风眼估计结果误差绝对值与均方根误差随信噪比的变化

Table 6　Change of root mean square error and deviation absolute value with SNR

信噪比	误差绝对值/(°)		均方根误差/(°)	
	经度	纬度	经度	纬度
SNR=8	0.275	0.100	0.534	0.269
SNR=6	0.125	0.338	0.534	0.411
SNR=4	0.113	0.138	0.627	0.582
SNR=2	0.638	0.663	0.649	0.664

不同信噪比下表 3 中所有台风样本风眼估计的平均偏差与均方根误差的均值如表 7 所示。

表 7　不同信噪比下风眼位置估计误差

Table 7　Estimation deviation of typhoon eye position under different SNR

信噪比	平均误差/(°)		均方根误差/(°)	
	经度	纬度	经度	纬度
SNR=8	0.116	0.160	0.430	0.446
SNR=6	0.087	0.143	0.398	0.500
SNR=4	0.022	0.141	0.399	0.594
SNR=2	0.143	0.278	0.473	0.645

由表 7 知, SNR=8 时, 经度平均误差≈0.116°, 纬度平均误差≈0.160°; 经度均方根误差≈0.430°, 纬度平均均方根误差≈0.446°。在地球赤道上, 1°约为 111 km, 据此可计算出时序 DDM 联合估计的经纬向平均误差大约小于 20 km, 均方根误差小于 49.5 km。随着信噪比的增大, 经纬度的均方根误差减小, 而

经纬度的平均误差在SNR≥4时基本稳定在0.2°以内。

分析误差来源，一是反卷积是个病态问题，受噪声影响严重，即在信噪比较低时反卷积得到的空间域散射系数失真比较严重；二是在时延-多普勒域映射到空间域时，无模糊线很难精准确定，所以空间模糊度问题难以完美解决。

5 结 论

针对星载GNSS-R技术在空间域检测台风事件及风眼位置估计问题，依据台风场景下的镜面反射点序列得到相应的DDM序列，反卷积后得到时延-多普勒域散射系数的分布，进而映射到空间域可得空间域NBRCS分布。基于此定义台风敏感特征观测量来检测台风事件，并利用模型匹配算法来估计台风眼的位置。

1）台风检测：依据空间域NBRCS分布来计算台风敏感特征观测量D_σ时间序列，基于置信区间距离半径的滑动窗口时序异常检测方法来检测台风事件。检测结果的ROC曲线表明，检测性能对信噪比敏感：当SNR=8时，AUC≈0.80；当SNR=2时，AUC≈0.58。信噪比越高，该算法的检测性能越好。

2）风眼位置估计：模型匹配算法的结果表明：4种信噪比下，基于DDM风眼位置估计算法经纬向定位平均误差均值<0.3°，均方根误差<0.6°，可以较好地实现台风眼位置的定位。

参考文献（References）

[1] 国家气候中心. 2023年汛期全国气候趋势滚动预测[EB/OL]. (2023-05-08) [2023-06-18]. http://ncc-cma.net/climate/pred?h_id=100117.
National Climate Centre. Rolling forecast of national climate trend in flood season in 2023[EB/OL]. (2023-05-08) [2023-06-18]. http://ncc-cma.net/climate/pred?h_id=100117(in Chinese).

[2] 白雪梅, 赵琳娜, 王彬雁, 等. 台风灾害影响评估系统业务进展[C]//第32届中国气象学会年会S23第五届研究生年会会议论文集, 2015: 266-267.
BAI X M, ZHAO L N, WANG B Y, et al. Progress of typhoon disaster impact assessment system[C]//Proceedings of the 32nd Annual meeting of the Chinese Meteorological Society S23 The 5th Annual Meeting of Graduate Students, 2015: 266-267(in Chinese).

[3] 梁必骐, 梁经萍, 温之平. 中国台风灾害及其影响的研究[J]. 自然灾害学报, 1995, 4(1): 84-91.
LIANG B Q, LIANG J P, WEN Z P. Study of typhoon disasters and its affects in China[J]. Journal of Natural Disasters, 1995, 4(1): 84-91 (in Chinese).

[4] 李晴. 多参数海洋浮标监测系统研究[D]. 上海: 上海海洋大学, 2017: 12-16.
LI Q. Research on multi-parameter marine buoy monitoring system[D]. Shanghai: Shanghai Ocean University, 2017: 12-16(in Chinese).

[5] 刘宏苏. 星载GNSS-R监测台风变化过程的研究[D]. 南京: 南京信息工程大学, 2021: 9-11.
LIU H S. Study on monitoring typhoon change process by satellite-borne GNSS-R[D]. Nanjing: Nanjing University of Information Science & Technology, 2021: 9-11(in Chinese).

[6] LI X F, ZHANG J A, YANG X F, et al. Tropical cyclone morphology from spaceborne synthetic aperture radar[J]. Bulletin of the American Meteorological Society, 2013, 94(2): 215-230.

[7] MEI W, LIEN C C, LIN I I, et al. Tropical cyclone-induced ocean response: a comparative study of the South China Sea and tropical northwest Pacific*, +[J]. Journal of Climate, 2015, 28(15): 5952-5968.

[8] 陈皓一. 基于深度强化学习算法的多尺度气旋监测方法研究[D]. 天津: 天津大学, 2019: 9-11.
CHEN H Y. Research on multi-scale cyclone monitoring method based on deep reinforcement learning algorithm[D]. Tianjin: Tianjin University, 2019: 9-11(in Chinese).

[9] GLEASON S, GEBRE-EGZIABHER D. GNSS 应用与方法[M]. 杨东凯, 译. 北京: 电子工业出版社, 2011.
GLEASON S, GEBRE-EGZIABHER D. GNSS applications and methods[M]. YANG D K, translated. Beijing: Publishing House of Electronics Industry, 2011(in Chinese).

[10] 杨东凯, 李晓辉, 王峰. GNSS反射信号海洋遥感应用现状分析[J]. 无线电工程, 2019, 49(10): 843-848.
YANG D K, LI X H, WANG F. Analysis of application status of GNSS reflected signal in ocean remote sensing[J]. Radio Engineering, 2019, 49(10): 843-848(in Chinese).

[11] 万玮, 陈秀万, 李国平, 等. GNSS-R遥感国内外研究进展[J]. 遥感信息, 2012, 27(3): 112-119.
WAN W, CHEN X W, LI G P, et al. GNSS reflectometry: a review of theories and empirical applications in ocean and land surfaces[J]. Remote Sensing Information, 2012, 27(3): 112-119(in Chinese).

[12] 万玮, 陈秀万, 彭学峰, 等. GNSS遥感研究与应用进展和展望[J]. 遥感学报, 2016, 20(5): 858-874.
WAN W, CHEN X W, PENG X F, et al. Overview and outlook of GNSS remote sensing technology and applications[J]. Journal of Remote Sensing, 2016, 20(5): 858-874(in Chinese).

[13] 李杰, 杨东凯, 洪学宝, 等. GNSS线极化天线干涉信号反演土壤湿度算法研究与测试[J]. 北京航空航天大学学报, 2024, 50(3): 874-885.
LI J, YANG D K, HONG X B, et al. Research and test of soil moisture retrieval algorithm based on GNSS linearly polarized antenna interference signal[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(3): 874-885(in Chinese).

[14] CAMPS A, PARK H, ALONSO-ARROYO A. Wind speed maping from the ISS using GNSS-R? a simulation study[C]//Proceedings of the IEEE International Geoscience and Remote Sensing Symposium-IGARSS. Piscataway: IEEE Press, 2013: 382-385.

[15] WANG F, ZHANG G D, YANG D K, et al. Single-pass tropical cyclone detector and scene-classified wind speed retrieval model for spaceborne GNSS reflectometry[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 4202716.

[16] PARK H, CAMPS A, PASCUAL D, et al. Simulation study on tropicial cyclone tracking from the ISS using GNSS-R measure-

ments[C]//Proceedings of the IEEE Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2014: 4062-4065.

[17] LI C, HUANG W M, GLEASON S. Dual antenna space-based GNSS-R ocean surface mapping: oil slick and tropical cyclone sensing[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(1): 425-435.

[18] LI C, HUANG W M, WANG W. Simulation based tropical cyclone sensing using space-based GNSS-R[C]//Proceedings of the Oceans-St. John's. Piscataway: IEEE Press, 2014: 1-4.

[19] MORRIS M, RUF C S. Estimating tropical cyclone integrated kinetic energy with the CYGNSS satellite constellation[J]. Journal of Applied Meteorology and Climatology, 2017, 56(1): 235-245.

[20] MORRIS M, RUF C S. Determining tropical cyclone surface wind speed structure and intensity with the CYGNSS satellite constellation[J]. Journal of Applied Meteorology and Climatology, 2017, 56(7): 1847-1865.

[21] MAYERS D, RUF C. Determining tropical cyclone center location with CYGNSS wind speed measurements[C]//Proceedings of the IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2019: 7529-7532.

[22] PARK H, CAMPS A, VALENCIA E, et al. Retracking considerations in spaceborne GNSS-R altimetry[J]. GPS Solutions, 2012, 16(4): 507-518.

[23] 张益强. 基于 GNSS 反射信号的海洋微波遥感技术[D]. 北京: 北京航空航天大学, 2008: 46-52.
ZHANG Y Q. Ocean microwave remote sensing using GNSS reflection signal[D]. Beijing: Beihang University, 2008: 46-52(in Chinese).

[24] WILLOUGHBY H E, DARLING R W R, RAHN M E. Parametric representation of the primary hurricane vortex. part II: a new family of sectionally continuous profiles[J]. Monthly Weather Review, 2006, 134(4): 1102-1120.

[25] VORONOVICH A G, ZAVOROTNY V U. Bistatic radar equation for signals of opportunity revisited[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(4): 1959-1968.

[26] ZAVOROTNY V U, VORONOVICH A G. Scattering of GPS signals from the ocean with wind remote sensing application[J]. IEEE Transactions on Geoscience and Remote Sensing, 2000, 38(2): 951-964.

[27] ULABY F T. Microwave remote sensing active and passive[J]. Theory to Applications, 1986: 2059-2081.

[28] MARCHAN-HERNANDEZ J F, CAMPS A, RODRIGUEZ-ALVAREZ N, et al. An efficient algorithm to the simulation of delay-Doppler maps of reflected global navigation satellite system signals[J]. IEEE Transactions on Geoscience and Remote Sensing, 2009, 47(8): 2733-2740.

[29] SHAH R, GARRISON J L, GRANT M S. Demonstration of bistatic radar for ocean remote sensing using communication satellite signals[J]. IEEE Geoscience and Remote Sensing Letters, 2012, 9(4): 619-623.

[30] 王峰, 杨东凯, 张波, 等. 星载 GNSS 海洋反射信号建模与仿真[J]. 北京航空航天大学学报, 2022, 48(3): 419-429.
WANG F, YANG D K, ZHANG B, et al. Modeling and simulation of spaceborne GNSS ocean-reflectometry[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(3): 419-429 (in Chinese).

[31] VALENCIA E, CAMPS A, MARCHAN-HERNANDEZ J F, et al. Ocean surface's scattering coefficient retrieval by delay-Doppler map inversion[J]. IEEE Geoscience and Remote Sensing Letters, 2011, 8(4): 750-754.

[32] MARTIN-NEIRA M, D'ADDIO S, BUCK C, et al. The PARIS in-orbit demonstrator[C]//Proceedings of the IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2009: II-322-II-325.

[33] 田腾, 石茂林, 宋学官, 等. 基于滑动窗口的时间序列异常检测方法[J]. 仪表技术与传感器, 2021(7): 112-116.
TIAN T, SHI M L, SONG X G, et al. Anomaly detecting method for time series based on sliding windows[J]. Instrument Technique and Sensor, 2021(7): 112-116(in Chinese).

Detection of typhoons and estimation of eye position using satellite-based GNSS-reflectometry

ZHEN Jiahuan[1]，ZHU Yunlong[1]，YANG Dongkai[1]，ZHANG Guodong[2]，WANG Feng[1, 3, *]

(1. School of Electronic and Information Engineering，Beihang University，Beijing 100191，China；
2. Institute of Telecommunication and Navigation Satellites，China Academy of Space Technology，Beijing 100094，China；
3. State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081, China)

Abstract：The singularity of the delay-Doppler map (DDM) is the primary focus of study on typhoons using global navigation satellite system (GNSS) reflectometry. However, due to the presence of ambiguity functions, this singularity is blurred, which affects detection performance. The normalized bistatic radar cross section (NBRCS) is reconstructed from deconvolving DDM and mapped to the spatial domain, based on which the detection of typhoon events and estimation algorithm of eye position is studied. The distribution sequence of NBRCS in the spatial domain can be obtained by reconstructing the DDM time series in the typhoon field. A detector of typhoon features is defined, and a sliding window anomaly detection method based on the confidence interval distance radius is used to detect typhoon events. A typhoon eye position estimation algorithm is proposed, which matches the reconstructed spatial scattering coefficient with the model dataset through a model-matching algorithm to obtain the eye position. The findings demonstrate how the signal-to-noise ratio (SNR) affects the typhoon detection algorithm's detection performance, with higher SNRs indicating greater detection performance. When the SNR is 8, the area under curve (AUC) of the receiver operating characteristic (ROC) curve of the typhoon detection performance is 0.80. The average deviation of the eye position estimation under four SNRs does not exceed 0.3°, and the root mean square deviation does not exceed 0.6°.

Keywords：global navigation satellite system-reflectometry；cyclone detector；cyclone center；CYGNSS；tropical cyclone

Received：2023-06-19；**Accepted**：2023-09-15；**Published Online**：2023-10-16 14：06

URL：link.cnki.net/urlid/11.2625.V.20231012.1316.002

Foundation item：The Open Grants of the State Key Laboratory of Severe Weather (2024LASW-B20); China National Postdoctoral Program for Innovative Talents (BX20200039)

* **Corresponding author**. E-mail：wangf.19@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0406

μHT-1推力器工作参数宽范围调节下束流特性实验研究

黄丹[1, 2, 3, 4, 5]，龙建飞[1, 2, 3, 4, 5, *]，成烨[1, 2, 3, 4, 5]，王嘉彬[1, 2, 3, 4, 5]，徐禄祥[1, 2, 3, 4, 5]，杨威[6]

(1. 国科大杭州高等研究院 基础物理与数学科学学院，杭州 310024；2. 中国科学院国家空间科学中心，北京 100190；
3. 中国科学院大学，北京，100049；4. 国科大杭州高等研究院 引力波太极实验室（杭州），杭州 310024；
5. 国科大杭州高等研究院 浙江省引力波精密测量重点实验室，杭州 310024；6. 重庆科技学院 电气工程学院，重庆 401331)

摘　　要：面向空间引力波探测任务需求，设计了瓦级微功率霍尔推力器μHT-1，并对该推力器的束流特性进行实验研究。采用法拉第探针结合三维移动机构进行诊断，获取阳极电压700～1 200 V、阳极工质流量0.1～0.5 sccm宽范围工况下束流离子电流密度分布，并进一步分析总束流值、阳极电流、电流利用率、发散角等参数变化趋势。测试结果表明：μHT-1推力器可在宽范围工况下稳定工作，束流呈现出较好的轴对称分布特性；离子电流密度沿轴向逐渐减小，沿径向双极扩散；阳极电压和阳极工质流量分别通过影响电子平均温度和通道中性原子密度分布的方式，使得电流利用率与发散角呈现出增长的趋势；推力器总束流与阳极工质流量、阳极电压均呈现线性增长特性。

关 键 词：微功率霍尔推力器；束流特性；离子电流密度；等离子体诊断；实验研究
中图分类号：V439.4
文献标志码：A　　**文章编号**：1001-5965(2025)06-2119-10

2015年美国地面激光干涉引力波天文台aLIGO成功探测到一例由双黑洞合并导致的引力波事件GW150914，证实了引力波的存在，也打开了引力波探测和引力波天文学的新纪元[1]。随之世界各国争相开展空间引力波探测计划，欧空局提出了LISA计划，日本提出了DECIGO计划，中国也正在开展“天琴计划”和“太极计划”[2-3]。空间引力波探测任务中，编队卫星平台需要处于超稳和超静的状态，尽量减少非保守力的干扰，为了使卫星平台克服非保守力的影响，必须采用无拖曳控制技术[4-5]。微推力器作为无拖曳控制技术的执行机构，需具备推力宽范围调节（0.1～250 μN）、高分辨率（<0.1 μN）、低噪声（<0.1 μN/Hz$^{1/2}$）、快响应（<100 ms）的能力[6-7]，同时为了保证任务周期，一般需要推力器实现工作寿命大于10 000 h，这便对微推力器提出了严苛的挑战。

通过分析空间引力波探测任务特点，2017年LISA列选了4种可能的微推力器类型：胶体推力器、会切型霍尔推力器、冷气微推力器和射频离子推力器。这些推力器均具备产生微牛级推力和宽范围调节的能力，然而，由于各自的结构设计和工作原理差异，均面临一些短期内难以克服的技术挑战[8-11]。相比而言，霍尔类推力器在结构和性能兼顾方面具有一定优势。

收稿日期：2023-06-21；**录用日期**：2023-10-13；**网络出版时间**：2024-01-05 15:25
网络出版地址：link.cnki.net/urlid/11.2625.V.20240104.1315.002
基金项目：国家重点研发计划 (2021YFC2202700)；湖南省自然科学基金 (2021JJ30564)；湖南省教育厅重点项目 (19A440)；重庆市教委科学技术研项目 (KJZD-K202101506)；国科大杭州高等研究院自主项目 (2022ZZ01009)
* **通信作者**. E-mail：ljf510@163.com

引用格式：*黄丹，龙建飞，成烨，等. μHT-1推力器工作参数宽范围调节下束流特性实验研究 [J]. 北京航空航天大学学报，2025，51（6）：2119-2128.*
HUANG D，LONG J F，CHENG Y，et al. Experimental study on beam characteristics of μHT-1 thruster under wide range adjustment of operating parameters [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（6）：2119-2128（in Chinese）.

国内外研究人员在微功率霍尔推力器方面进行了大量研究。2018年Hey[12]对会切磁场霍尔推力器开展微型化研究，研制出Full Size NG-μHEMPT和Miniaturised NG-μHEMPT两款微推力器。通过实验测试，分别实现了6～4 800 μN和29～86 μN范围的推力连续调节。于达仁等[6-7]也开展了微牛级会切霍尔电推进技术研究，针对空间引力波探测任务，设计的μCFT推力器实现了1.8～112.8 μN的连续调节输出，推力器闭环控制可有效降低噪声，具备应用于无拖曳控制任务的可能。国科大杭州高等研究院联合新加坡南洋理工大学，开展了微牛级双模霍尔电推进技术研究，并研制了霍尔微推力器HMTs[13-14]。2019年8月，该推力器搭载“太极一号”成功发射，第一阶段在轨测试数据表明，推力范围为5～100 μN，推力分辨率达到0.7 μN，噪声小于0.6 μN/Hz$^{1/2}$(0.01～1.0 Hz)。该推力器在国际上刷新了霍尔推力器在轨应用的功率下限、推力下限和推力调节比[15]。针对空间引力波探测等科学任务应用，现有微功率霍尔推力器仍需要进一步提升性能，即在宽范围电压和流量变化工况下，实现高效电离和加速效率，同时满足微牛级推力宽范围调节、高分辨率等指标要求。微功率霍尔推力器开发与研制仍是该领域研究的热点与难点。

Mikellides等[16]指出，霍尔推进器的性能参数、寿命等与羽流存在密不可分的关系，因此，国内外针对束流特性开展了大量研究。Hofer等对5 kW霍尔推进器羽流进行诊断[17]，得到了该推力器束流分布参数，并进一步分析了真空背景压力对束流特性影响规律；朱悉铭等利用快速测算方法针对HET-P70[18]霍尔推力器的束流诊断，基于束流分布参数，进一步分析了束流聚焦特性影响参数；陈新伟等[19]采用法拉第对LHT60霍尔推力器束流诊断，测试得到了该推力器束流分布特性。不同推力器，其束流分布特性具有一定的差异性。另外显然，当前束流特性研究主要针对结构大、功率高的推力器，而微功率霍尔推力器束流分布研究较少，由于结构更小，空间电流密度更低，束流特性测量存在难度。

国科大杭州高等研究院微推进团队持续开展微功率霍尔电推进技术研究。在前期研究的基础上，探索性设计了新微功率推力器μHT-1，并完成了加工与调试，目前尚未开展微推力测量。现阶段主要针对该推力器参数宽范围调节下的羽流分布特性进行实验研究，一方面，通过羽流参数分析，有利于研究推力器内部离子能量转换、分布、损失情况，结合电流利用率等特性，进一步改进推力器的性能；另一方面，获得该推力器束流发散角，以及总束流随阳极电压、阳极工质流量的变化规律，探索该推力器宽范围性能调节能力。

本文利用自研的等离子体探针对μHT-1推力器羽流特性进行诊断，首次获得该推力器束流离子电流分布，并进一步分析总束流值、阳极电流、电流利用率、发散角等参数变化趋势。研究结果可为后续该推力器优化提供有力的数据支撑。

1 实验系统及装置

1.1 μHT-1推力器

μHT-1推力器是基于传统霍尔推力器上开发而来的微功率推力器，并参考了HMTs推力器结构，设计阳极功率为1～10 W可调。一方面，充分继承了霍尔推力器推功比大、可靠性高等优势；另一方面，在磁路结构、放电通道、阴极结构等方面进行了优化，集成化程度得到明显提升。该推力器放电通道外径约为20 mm，内径直径约为15 mm，采用永磁体结构，通道出口附近最大磁场强度约为200 Gs。研制过程中，采用有限元方法仿真分析了气体分配器挡板导流小孔直径、双腔比等内部关键尺寸，提升了放电通道内气体的均匀程度[20-21]，为微小流量调节奠定了基础；同时采用环式分压气路绝缘结构[22-23]，显著提升了供气管路的耐压性能，解决了推力器小尺寸结构下的高压击穿问题。

μHT-1推力器工作原理及氪气工质点火状态如图1所示。推力器工作时，位于环形通道底部的阳极加载高电压，在通道内产生轴向加速电场，通道

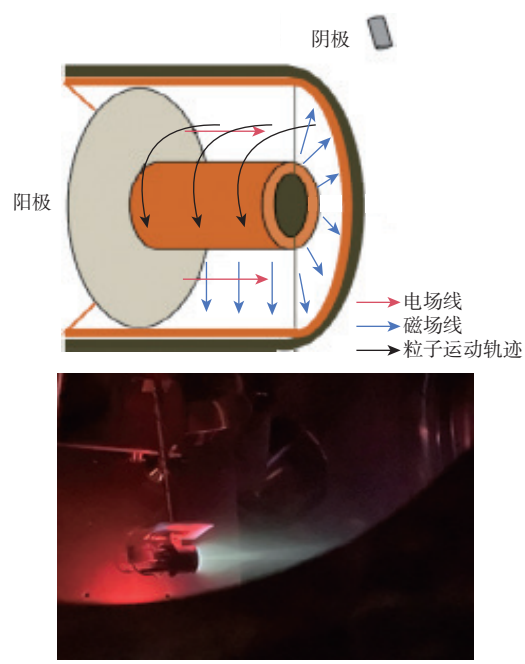


图1 μHT-1推力器工作原理及氪工质点火状态

Fig. 1 The ignition state of krypton gas of uHT-1 thruster

外部阴极产生的部分电子在阳极高压吸引下进入通道, 向阳极移动过程中轰击原子并产生等离子体。推进器径向磁场使电子螺旋旋转形成周向霍尔漂移, 由于离子质量远大于电子质量, 受到径向磁场的影响微弱, 在轴向电场的加速下形成推力束流。阴极不仅是电子发射源, 而且还能中和离子, 保持推进器的电中性。微功率霍尔推力器选用钡钨热阴极方案, 其发射面为平面结构, 发射面直径为 3.5 mm, 1 200 ℃ 条件下可发射电流超过 100 mA, 满足微功率霍尔推力器应用要求。

1.2　实验真空系统

实验研究在国科大杭州高等研究院太极实验室 CS-1 600 微牛级电推进实验平台上进行。该实验平台真空舱直径为 1.8 m, 长度为 5.0 m, 真空舱具有较大空间可以有效减小羽流的相互作用。真空系统配置了 2 台分子泵+多级罗茨泵、1 套大型螺杆+罗茨机械泵组, 最大抽速为 34 000 L/s, 极限真空度可达到 1.0×10^{-6} Pa 量级, 满足本实验条件要求。此外, 实验平台还配置了高精度高压电源、高精度微小流量贮供单元等, 可满足推力器地面实验电气参数调节要求。CS-1600 微牛级电推进实验平台如图 2 所示。



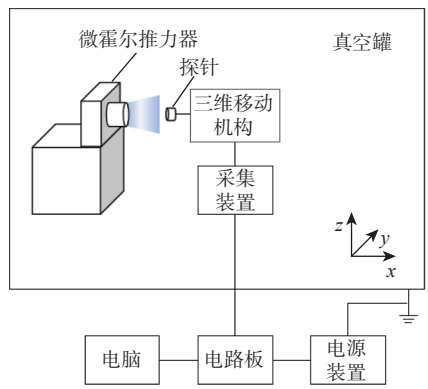
图 2　CS-1 600 微牛级电推进实验平台

Fig. 2　The CS-1 600 micro Hall electric propulsion experimental platform

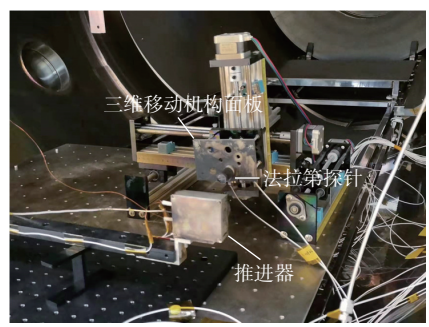
1.3　等离子体诊断系统

为了获得μHT-1 推力器羽流中离子电流密度、发散角和总束流等参数, 采用法拉第探针结合三维移动机构对该推力器羽流进行诊断。

等离子体诊断系统连接及实物如图 3 所示。将推力器固定在安装支架上, 法拉第探针和三维移动机构置于推力器出口正下方, 两者间采用陶瓷进行绝缘, 通过控制三维移动机构带动法拉第探针进行离子密度空间分布测量。采用自编写程序控制三维移动机构可实现 x、y、z 这 3 个方向的直线移动, 每个方向的移动范围为 0～200 mm, 最小移动精度为 0.1 mm。实验过程中, 以推力器出口端面的中心为原点建立空间坐标, 且假定沿着推力器出口轴向为 x 轴, 径向为 y 轴, 垂直方向为 z 轴, 满足右手定则。初始状态下, 推力器出口端面中心点坐标固定为(0, 0, 0)。



(a) 系统硬件示意图



(b) 系统硬件实物图

图 3　等离子体诊断系统连接示意图及实物

Fig. 3　The schematic diagram of the connection of the plasma diagnostic system and the physical object

法拉第探针结构如图 4 所示, 其中收集极为直径 7.0 mm 的圆形平面, 考虑二次电子发射的影响, 收集极采用金属钼, 该材料二次电子发射系数小, 可减少测试误差。保护环为金属圆柱体状, 外径为 9 mm, 内径为 8 mm, 采用成本低、易加工的不透钢材料, 可屏蔽非轴向的低能量离子。收集极和保护环间距为 0.5 mm, 该间距与推力器羽流等离子鞘层厚度在同一数量级, 并且满足小于等于 10 个德拜长度, 能够屏蔽束流中非轴向等离子体[7]。测试过程中, 将收集极和保护环都采用负偏置连接, 以排斥电子并确保均匀的势场。通过高精度电流表可

实时读出离子电流值，再用电流值除以收集极面积便可得到离子电流密度。实验中，电流表数据通过采集卡实时采集与存储，采集频率为2.0 kHz，采集精度优于0.1%。

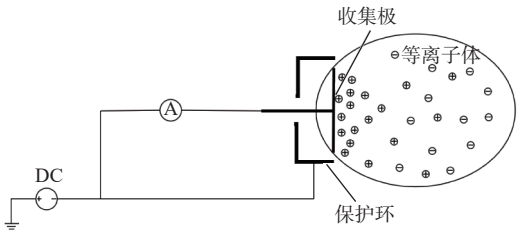


图4 法拉第筒示意图

Fig. 4 Schematic of Faraday cup

根据推力器羽流分布轴对称性，首先确定x轴位置，然后在y方向进行移动扫描，从而测得羽流中离子电流密度的径向分布，测试方法[24]如下。在此基础上，通过积分求和计算出推力器总束流值。束流总值计算[12]如下：

$$I_{\mathrm{b}}=\frac{2\pi\Delta y\sum I(y_{\mathrm{i}})y_{\mathrm{i}}}{\pi R^2} \tag{1}$$

式中：y为距推进器中心线的径向距离；y_i为第i个测量位置处距原点的径向距离；Δy为相邻2个测量位置距离的增量；$I(y_{\mathrm{i}})$为法拉第探针在第i个位置测量的离子电流；R为法拉第筒半径。

电流利用率定义为总束流I_{b}除以阳极电流I_{d}，计算公式如下：

$$\eta_{\mathrm{b}}=\frac{I_{\mathrm{b}}}{I_{\mathrm{d}}} \tag{2}$$

根据离子电流密度分布，进一步可以计算出推力器束流发散角，公式如下：

$$\theta=2\arctan\frac{y_{90\%}-\dfrac{D}{2}}{L} \tag{3}$$

式中：$y_{90\%}$为总离子束电流90%的位置距原点的径向距离；D为推进器喷口直径；L为法拉第探针中心距坐标轴原点的轴向距离。

1.4 实验误差分析

μHT-1推力器束流密度测试实验中，测试误差主要来源于法拉第探针和辅助仪器误差，其中辅助仪器主要包括微安电流表、三维移动机构等。

法拉第探针误差主要包括[25]：收集极二次电子误差、收集极与防护套间接误差，以及收集极与束流球面计算误差。高能离子轰击收集极时会产生二次电子，从而影响收集的离子电流偏差。本文实验中收集极采用金属钼，μHT-1推力器最大阳极电压为1 200 V，考虑二价离子能量约为1.6 keV左右，对应钼的二次电子发射系数约为0.02，即二次电子发射误差约为2.0%；收集极直径为8.0 mm，接收面测试距离为125～175 mm，根据法拉第筒误差公式[26]计算，束流球面误差约为0.43%；为有效避免收集极与保护套间距误差，应该使得该间接小于5个德拜长度λ_{d}。根据估算束流离子中德拜长度约为0.1 mm，本实验中法拉第筒收集极与保护套间接为0.5 mm，满足5个德拜长度，可以较好地消除该项误差。因此，通过以上分析，法拉第探针误差u_{f}主要来自于材料的二次电子发射，约为2.0%。

辅助仪器误差主要来源于三维移动机构和微安电流表。实验中，三维移动机构采用高精度步进电机进行驱动，控制精度为0.1 mm；离子电流采用高精度微安电流表进行采集，控制精度优于0.01 μA。针对仪器不确定误差，按B类不确定度计算，微安电流表不确定度u_{i}和三维移动机构不确定度u_{x}满足：

$$\begin{cases}u_{\mathrm{i}}=\dfrac{\Delta i}{C}\\ u_{\mathrm{x}}=\dfrac{\Delta x}{C}\end{cases} \tag{4}$$

式中：Δi为微安电流表的允许最大误差（误差限）；Δx为三维移动机构的允许最大误差（误差限）；C为包含因子，取决于仪器误差的分布规律，仪器说明书无特殊说明，取矩形分布，$C=\sqrt{3}$。

进一步，由于法拉第探针和辅助仪器误差之间互不相关，结合误差传递函数，系统误差计算公式满足：

$$u_{\mathrm{all}}=\sqrt{u_{\mathrm{f}}^2+u_{\mathrm{i}}^2+u_{\mathrm{x}}^2} \tag{5}$$

式中：u_{all}为系统误差；u_{f}为法拉第探针误差。

实验中，法拉第探针测试误差为2%，微安电流表控制精度为0.01 μA，三维移动机构控制精度为0.1 mm，因此，系统测试误差约为2.43%。

2 测试数据分析

2.1 束流离子电流密度分布诊断

为了全面获得该推力器束流空间分布，法拉第探针在不同轴向位置进行径向扫描。测试过程中，推力器阳极工质流量固定为0.5 sccm，阳极电压1 200 V保持不变，法拉第探针从初始位置（0, 0, 125 mm）首先沿轴向（x轴）后移，每次移动5 mm。然后法拉第探针再沿径向（y轴）从−50 mm扫描至50 mm，每次移动步长为1 mm，停留时间为3 min，采集系统对收集极电流值自动采集，再用该电流值除以接收面积计算得到离子电流密度。为了减少测量误差，每个参数下共测试5次，取平均值为测试结果，并将最大值与最小值的差值作为测试误差棒。测试中沿着轴向共移动11次，即最后移动到

(0, 0, 175 mm)处, 共采集数据 5 500 个。μHT-1 推力器束流离子电流密度分布如图 5 所示。

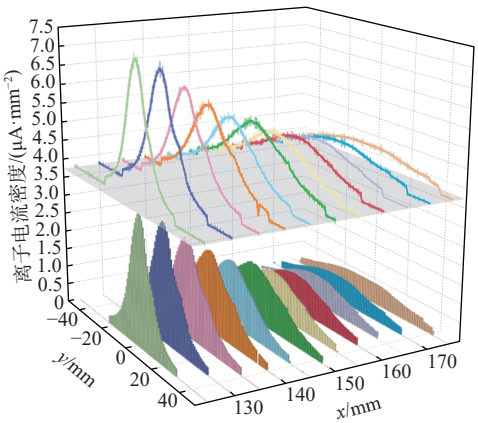


图 5　束流离子电流密度分布随轴向距离变化

Fig. 5　The variation curves of ion current density under different axial distances

三维移动机构从轴向 125 mm 移动到 175 mm 的过程中, 束流离子电流密度分布平直度越来越好。这与束流离子的空间扩散过程相关, 推力器中心离子电流密度最大值从 3.516 μA/mm^2 逐渐减少到 0.586 μA/mm^2, 减小幅度为 83.3%。

根据粒子数守恒, 不同轴向位置的离子通量保持一致, 球面坐标(R,θ)及立体角(Ω)下可表示为

$$R_1^2 j(R_1,\theta)\mathrm{d}\Omega = R_2^2 j(R_2,\theta)\mathrm{d}\Omega \tag{6}$$

根据推力器束流模型[27], 可得:

$$j(R,\theta) = \frac{\Gamma(\theta)}{R^2} \tag{7}$$

式中: $\Gamma(\theta)$表示 $j(R,\theta)$依赖于角度θ的变化关系。

由于束流具有一定的发散角度, 随着三维移动机构远离推力器出口端面(即 z 越大), 对应束流离子空间分布越来越稀疏, 进而使得离子密度减小。同时, 电子与离子在运动过程中相结合, 也使得空间中的离子数量随着轴向距离增加而减少。

2.2　阳极电压参数调节

阳极电压作为 μHT 推力器主要调节参数, 对推力器束流分布具有重要影响。实验过程中, 阳极工质流量设定为 0.5 sccm, 阳极电压调节范围选取 700～1 200 V, 调节间隔为 100 V, 法拉第探针初始位置(0, 0, 150 mm), 即法拉第探针收集极位于推力器出口端下游 150 mm 处, 法拉第探针沿 y 轴从 −50 mm 移至 50 mm, 每次移动步长为 1 mm, 停留时间为 3 min, 实验中每个参数共测试 5 次, 取平均值为测试结果, 并将最大值与最小值的差值作为测试误差棒。通过测试, 获得不同阳极电压下离子电流密度分布。

离子电流密度分布随阳极电压变化如图 6 所示。推力器束流离子电流密度分布总体表现出较好的轴对称性。这种轴向对称分布间接验证了该推力小尺寸结构下供气、磁场分布等轴向均匀性。

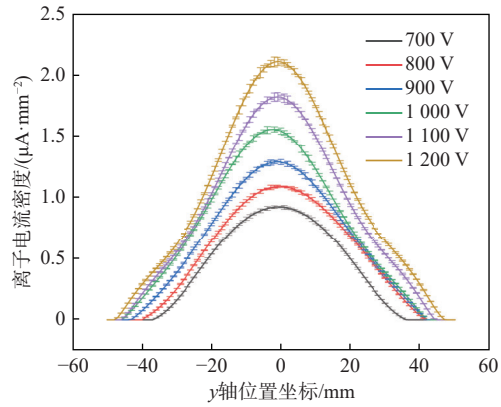


图 6　离子电流密度分布随阳极电压变化

Fig. 6　The variation curves of ion current density under different anode voltage

束流离子在空间中加速:

$$eU = \frac{1}{2}mv^2 \tag{8}$$

式中: e为单位电荷; U 为通道内电压; m 为粒子质量; v 为粒子速度。

由离子加速过程可知, 喷出离子速度v_i、离子电流密度 j、阳极电压V_d满足:

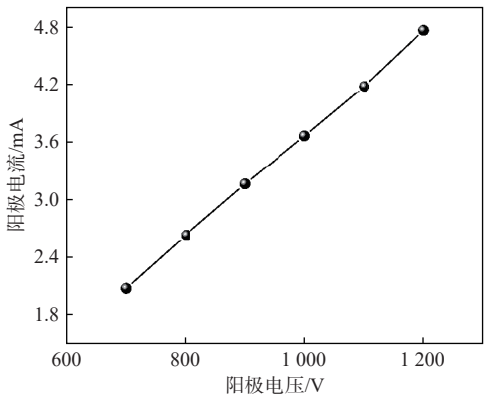
$$\begin{cases} v_\mathrm{i} = \sqrt{\dfrac{2eV_\mathrm{d}\eta_\mathrm{v}}{m_\mathrm{i}}} \\ j = n_\mathrm{i} e v_\mathrm{i} \end{cases} \tag{9}$$

式中: m_i为离子质量; η_v为电压效率; n_i为离子数密度。

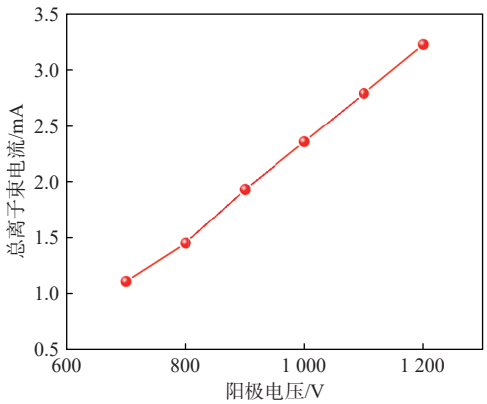
因此, 随着阳极电压从 700 V 增大到 1 200 V, 通道内加速电场强度变大, 离子所获得加速能量更多, 同时通道内电离率和二价离子成分可能增大等(该部分参数未能测试), 多重效果下使得束流离子电流密度整体增大, 不同电压下束流曲线整体趋势如锥形嵌套。

图 7 为不同阳极电压下束流和阳极电流变化关系。当阳极电压从 700 V 逐渐增大到 1 200 V 时, 阳极电流从 2.07 mA 增加到 4.76 mA, 而束流也从 1.10 mA 增加到 3.22 mA, 总束流值和阳极电流均呈现线性增长趋势。由于总束流根据式(1)积分计算得到, 因此, 总束流值也将随着阳极电压增大而增大。测试结果可为推力宽范围调节提供较好的数据支撑。

图 8 为电流利用率和束流发散角随阳极电压变化关系。电流利用率从 53.4% 增长到 67.7%, 发散角全角从 41.3°增大到 56.1°。束流发散角和电流利用率均呈现出先快速增大后趋于缓和的趋势。



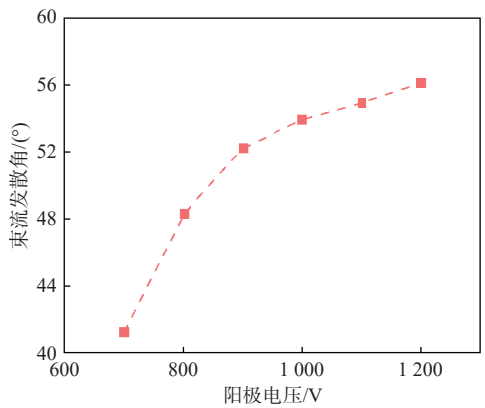
(a) 阳极电流变化



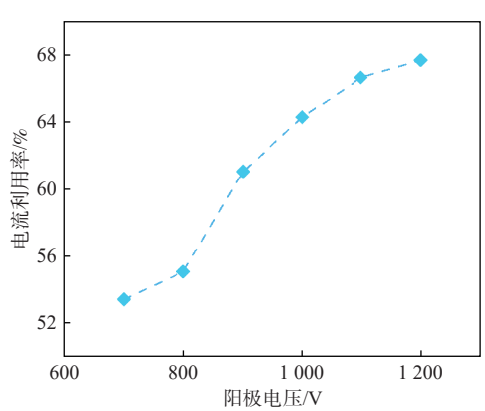
(b) 总束流变化

图 7 不同阳极电压下束流和阳极电流变化关系

Fig. 7 The curves of anode current and total ion beam current under anode voltage



(a) 束流发散角变化



(b) 电流利用率变化

图 8 不同阳极电压下电流利用率和束流发散角变化关系

Fig. 8 The curves of beam divergence angle and current utilization efficiency under anode voltage

分析认为：随着阳极电压的增大，通道内电子平均温度增大使得电离率提高，因此，电流利用率也相应增大；同时，通道壁面能量损耗增大使得电流利用率在获得迅速增大后便逐渐趋于饱和[28]。束流发散角主要与通道内加速电场分布相关，随着阳极电压大范围提高，磁场对高能电子约束能力逐渐减弱，影响了加速电场的聚焦性[29]，使得发散角先增大后逐渐趋于稳定。

2.3 阳极工质流量参数调节

阳极工质流量作为推力器另一个重要调节参数，对束流分布同样具有重要影响。因此，实验过程中，阳极电压设定为 1 200 V，三维移动机构初始位置依然在(0, 0, 150 mm)处。测试时每组测试数据 5 次，对测试结果取平均并进行误差棒处理。阳极工质流量变化共测试 5 种工况(0.1 ～0.5 sccm)，每种工况下采集数据 1 500 个。

离子电流密度分布随阳极工质流量变化如图 9 所示。测试数据显示，与阳极电压变化规律类似，不同流量下的离子电流密度也呈现出中心对称分布的特性。不同阳极工质流量主要影响通道电离过程，其中通道内原子和离子一维稳态连续性如下：

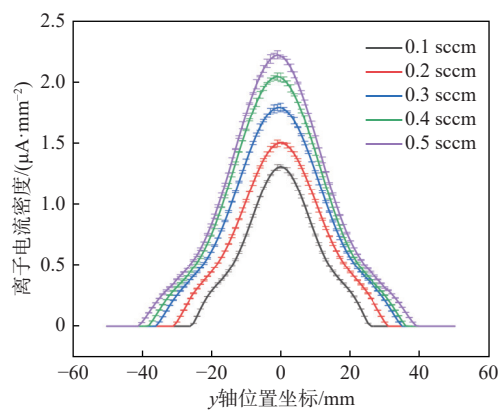


图 9 离子电流密度分布随阳极工质流量变化

Fig. 9 The variation curves of ion current density under different krypton work flow

$$\begin{cases} v_0 \dfrac{\mathrm{d}n_0}{\mathrm{d}z} = -n_e(z)n_0(z)\beta_{iz}(z) \\ v_i \dfrac{\mathrm{d}n_i}{\mathrm{d}z} = n_e(z)n_0(z)\beta_{iz}(z) \end{cases} \tag{10}$$

式中：$n_e(z)$为通道轴向电子数密度；$\beta_{iz}(z)$为电离速率系数；z 为轴向坐标。

当阳极工质流量增大时，放电通道原子数量增多，使得电离碰撞产生的离子数量变多，经加速电场喷出后收集到的离子数量增多，即束流离子电流

密度整体增大。

图 10 为不同阳极工质流量下总束流和阳极电流变化关系。当阳极工质流量从 0.1 sccm 逐渐增大到 0.5 sccm 时, 总束流由最小值 0.83 mA 增长到 3.19 mA, 而阳极电流也从 1.45 mA 增长到 4.80 mA, 均表现出较好的线性增长趋势。

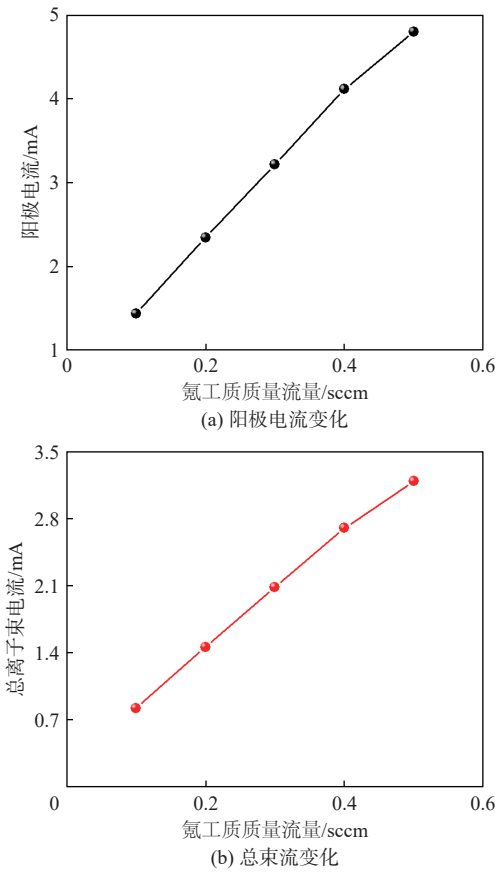


图 10　不同阳极工质流量下总束流和阳极电流变化关系

Fig. 10　The curves of anode current and total ion beam current under krypton work flow

霍尔推进器的阳极效率与离子质量和阳极电流的关系如下:

$$\frac{M}{e}I_{\mathrm{d}}\eta_{\mathrm{b}}=\dot{m}_{\mathrm{i}} \tag{11}$$

式中: e 为单位电荷; M 为原子质量; $\dot{m}_{\mathrm{i}}$ 为离子质量流量。

总质量利用效率可以表示如下:

$$\eta_{\mathrm{m}}=\frac{\dot{m}_{\mathrm{i}}}{\dot{m}_{\mathrm{p}}}=\frac{\dot{m}_{\mathrm{i}}}{\dot{m}_{\mathrm{a}}+\dot{m}_{\mathrm{c}}} \tag{12}$$

式中: η_{m} 为质量利用效率; $\dot{m}_{\mathrm{p}}$ 为推进器总质量流量; $\dot{m}_{\mathrm{a}}$ 为阳极质量流量; $\dot{m}_{\mathrm{c}}$ 为阴极质量流量。

根据上述推力器工作原理, 束流 I_{b} 与阳极工质流量 $\dot{m}_{\mathrm{a}}$ 满足下式:

$$I_{\mathrm{b}}=\frac{e\dot{m}_{\mathrm{a}}\eta_{\mathrm{m}}}{M} \tag{13}$$

由上式可知, 束流 I_{b} 主要与阳极工质流量、工质利用率相关。当阳极电压固定时, 工质利用率较为稳定, 因此, 束流与阳极工质流量具有较好的线性关系。阳极电流的增加不仅与阳极工质流量的增大有关, 还需要考虑壁面二次电子[27] 等因素。束流和阳极工质流量线性增长规律可为推力宽范围调节提供数据支撑。

图 11 为不同阳极工质流量下电流利用率和发散角变化关系。随着阳极工质流量增大, 电流利用率和发散角均呈现出先快速增大后趋于缓和的变化趋势。

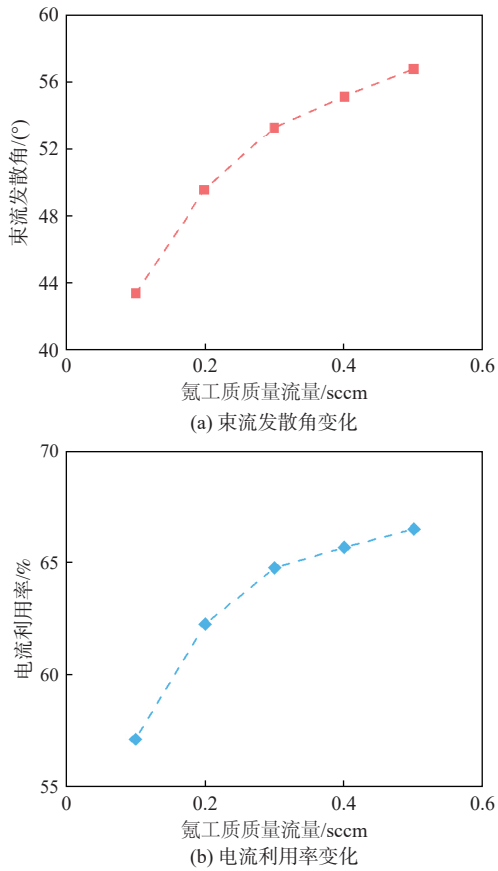


图 11　不同阳极工质流量下电流利用率和发散角变化

Fig. 11　The curves of beam divergence angle and current utilization efficiency under krypton work flow

阳极工质流量主要影响通道内中性原子密度分布, 而当原初电子的数量和电子温度相对稳定时, 电流利用率将随着电离率饱和而趋于稳定。此外, 随着阳极工质流量增加, 通道内电离区逐渐移向出口[28], 电离碰撞产生离子的初始位置更多地聚集在出口附近, 从而影响束流发散角。

3　结　论

采用法拉第探针结合三维移动机构首次对 μHT-1 推力器束流特性进行诊断, 通过实验数据分析, 得到主要结论如下:

1）随着束流离子在空间中的扩散运动以及电子与离子的复合作用，μHT-1推力器束流离子密度沿轴向距离的增加呈现逐渐减小的趋势，且束流平直度随距离增大而显著改善。

2）μHT-1推力器在宽范围工况下（阳极电压为700～1 200 V，阳极工质流量为0.1～0.5 sccm）可实现稳定工作，束流呈现出较好的轴对称分布特性。

3）随着阳极电压增大，离子获得能量增加，束流离子电流密度整体增大，呈现出锥形嵌套形式，且具有较好的轴对称性。同时，电子平均温度升高进一步导致电离率提高，即电流利用率从53.4%增长到67.7%，也使得磁场对高能电子约束能力减弱，影响电场聚焦性，表现为发散角从41.3°增大到56.1°。

4）随着阳极工质流量增大，离子电流密度分布变化规律与阳极电压变化相似。但阳极工质流量变化主要对通道内中性原子密度分布有影响，这种密度分布和电子温度间接使得电流利用率在57.1%～66.8%范围内波动。而影响发散角的原因主要是离子碰撞区向出口转移，从而呈现出增长速度由快及慢增长，从43.4°增加到56.7°。

5）推力器总束流值与阳极工质流量、阳极电压均呈现线性变化，可为后续推力宽范围调节、推力分辨率分析等提供数据基础。后续将开展多工况下的推力测量，进一步对该推力器性能进行综合评估。

参考文献（References）

[1] ABBOTT B P, ABBOTT R, ABBOTT T D, et al. Observation of gravitational waves from a binary black hole merger[J]. Physical Review Letters, 2016, 116(6): 061102.

[2] 罗子人, 张敏, 靳刚, 等. 中国空间引力波探测“太极计划”及“太极1号”在轨测试[J]. 深空探测学报, 2020, 7(1): 3-10.
LUO Z R, ZHANG M, JIN G, et al. Introduction of Chinese space-borne gravitational wave detection program “taiji” and “taiji-1” satellite mission[J]. Journal of Deep Space Exploration, 2020, 7(1): 3-10 (in Chinese).

[3] 涂良成. 天琴计划的回顾与小结[J]. 中山大学学报 (自然科学版), 2021, 60(增刊 1): 253.
TU L C. Review and summary of Tianqin Project[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2021, 60(Sup 1): 253 (in Chinese).

[4] 胡越欣, 张立华, 高永, 等. 空间引力波探测航天器关键技术分析[J]. 航天器工程, 2022, 31(4): 1-7.
HU Y X, ZHANG L H, GAO Y, et al. Analysis of key technologies of spacecraft for gravitational waves detection in space[J]. Spacecraft Engineering, 2022, 31(4): 1-7 (in Chinese).

[5] 祝竺, 赵艳彬, 尤超蓝, 等. 面向空间引力波探测的非接触式卫星平台无拖曳控制技术[J]. 南京航空航天大学学报, 2022, 54(增刊 1): 9-13.
ZHU Z, ZHAO Y B, YOU C L, et al. Drag-free control technology of non-contact satellite platform for space gravitational wave detection[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2022, 54(Sup 1): 9-13 (in Chinese).

[6] 于达仁, 崔凯, 刘辉, 等. 用于引力波探测的微牛级霍尔电推进技术[J]. 哈尔滨工业大学学报, 2020, 52(6): 171-181.
YU D R, CUI K, LIU H, et al. Micro-Newton hall electric propulsion technology for gravitational wave detection[J]. Journal of Harbin Institute of Technology, 2020, 52(6): 171-181 (in Chinese).

[7] 于达仁, 牛翔, 王泰卜, 等. 面向空间引力波探测任务的微推进技术研究进展[J]. 中山大学学报 (自然科学版), 2021, 60(增刊 1): 194-212.
YU D R, NIU X, WANG T B, et al. The developments of micro propulsion technology based on space gravitational wave detection task[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2021, 60(Sup 1): 194-212 (in Chinese).

[8] ZIEMER J, MARRESE-READING C, DUNN C, et al. Colloid microthruster flight performance results from space technology 7 disturbance reduction system[C]//Proceedings of International Electric Propulsion Conference. Reston: AIAA, 2017.

[9] LEITER H, LOTZ B, FEILI D, et al. Design development and test of the RIT-ÁX mini ion engine system[C]//Proceedings of the 31st International Electric Propulsion Conference. Michigan: IEPC, 2009: 179.

[10] 李永, 刘旭辉, 汪旭东, 等. 空间极小推力宽范围可调推进技术研究进展[J]. 空间控制技术与应用, 2019, 45(6): 1-12,19.
LI Y, LIU X H, WANG X D, et al. Review and prospect on the large-range thrust throttling technology with extremely small thrust[J]. Aerospace Control and Application, 2019, 45(6): 1-12,19 (in Chinese).

[11] AMARO-SEOANE P, AUDLEY H, BABAK S, et al. Laser interferometer space antenna[EB/OL]. (2017-02-01) [2013-05-28]. https://arxiv.org/abs/1702.00786, 2017.

[12] HEY F G. Micro Newton thruster development[M]. Wiesbaden: Springer Fachmedien Wiesbaden, 2018.

[13] XU S Y, XU L X, CONG L X, et al. First result of orbit verification of Taiji-1 hall micro thruster[J]. International Journal of Modern Physics A, 2021, 36: 2140013-21400S9.

[14] POTRIVITU G C, SUN Y F, BIN ROHAIZAT M W A, et al. A review of low-power electric propulsion research at the space propulsion centre Singapore[J]. Aerospace, 2020, 7(6): 67.

[15] 杭观荣, 李诗凝, 康小录, 等. 霍尔电推进空间应用现状及未来展望[J]. 推进技术, 2023, 44(6): 38-51.
HANG G R, LI S N, KANG X L, et al. Current space application status and future prospect of Hall electric propulsion[J]. Journal of Propulsion Technology, 2023, 44(6): 38-51 (in Chinese).

[16] MIKELLIDES I G, KATZ I, GOEBEL D M, et al. Hollow cathode theory and experiment. II. A two-dimensional theoretical model of the emitter region[J]. Journal of Applied Physics, 2005, 98(11): 113303.1-113303.14.

[17] HOFER R R, GOEBEL D M, MIKELLIDES I G, et al. Magnetic shielding of a laboratory Hall thruster. II. Experiments[J]. Journal of Applied Physics, 2014, 115(4): 043304.

[18] 朱悉铭, 宁中喜, 于达仁. HEP-70霍尔推力器的发射光谱诊断研究[C]//第十八届全国等离子体科学技术会议摘要集. 合肥: 中国

科学技术大学出版社, 2017: 240.
ZHU X M, NING Z X, YU D R. Emission spectroscopy diagnostics of HEP-70 Hall thruster [C]//Proceedings of the 18th National Conference on Plasma Science and Technology. Hefei: University of Science and Technology of China Press, 2017: 240(in Chinese).

[19] 陈新伟, 高俊, 顾左, 等. 变工况下自励磁模式 LHT-60 霍尔推力器放电特性试验研究[J]. 真空与低温, 2022, 28(1): 106-114(in Chinese).
CHEN X W, GAO J, GU Z, et al. Experimental study on discharge characteristics of self-field mode LHT-60 Hall thruster under variable operating conditions[J]. Vacuum and Cryogenics, 2022, 28(1): 106-114.

[20] LU S X, LUO W, LONG J F, et al. Numerical simulation optimization of neutral flow dynamics in low-power Hall thruster[J]. Results in Physics, 2023, 46: 106268.

[21] 罗威, 龙建飞, 徐禄祥, 等. 霍尔推力器放电通道中性气体分布及检测技术研究进展[J]. 固体火箭技术, 2023, 46(1): 158-166.
LUO W, LONG J F, XU L X, et al. Research progress of distribution and detection technology neutral gas in Hall thruster discharge channel[J]. Journal of Solid Rocket Technology, 2023, 46(1): 158-166 (in Chinese).

[22] 龙建飞, 徐禄祥, 吴铭钐, 等. 一种霍尔推力器供气管路的气路分压绝缘方法及其应用: 中国, CN114458565B[P]. 2022-07-12.
LONG J F, XU L X, WU M S, et al. A method of gas circuit pressure divider insulation for Hall thruster supply line and its application: China, CN114458565B[P]. 2022-07-12(in Chinese).

[23] 龙建飞, 徐禄祥, 吴铭钐, 等. 一种霍尔推力器环式分压气路绝缘结构: 中国, CN114458564B[P]. 2022-07-12.
LONG J F, XU L X, WU M S, et al. An annular voltage-division gas path insulation structure for Hall thruster: China, CN114458564B [P]. 2022-07-12(in Chinese).

[24] 宁中喜. 霍尔推力器羽流发散角的定向探针测量方法[J]. 推进技术, 2011, 32(6): 895-899.
NING Z X. Directional probe measurement of plume divergence angle in Hall thrusters[J]. Journal of Propulsion Technology, 2011, 32(6): 895-899 (in Chinese).

[25] 龙建飞, 张天平, 吴辰宸, 等. LIPS-200 离子推力器放电室出口离子密度分布研究[J]. 推进技术, 2018, 39(5): 1194-1200.
LONG J F, ZHANG T P, WU C C, et al. Study on ion density distribution in discharge chamber exit of LIPS-200 ion thruster[J]. Journal of Propulsion Technology, 2018, 39(5): 1194-1200 (in Chinese).

[26] 刘星宇, 李鸿, 毛威, 等. 霍尔推力器能量损失系统性评价方法[J]. 推进技术, 2022, 43(7): 470-480.
LIU X Y, LI H, MAO W, et al. Systematic evaluation method for power loss of Hall thruster[J]. Journal of Propulsion Technology, 2022, 43(7): 470-480 (in Chinese).

[27] 商圣飞, 顾左, 贺碧蛟, 等. 离子推力器束流密度分布模型[J]. 真空科学与技术学报, 2015, 35(12): 1414-1419.
SHANG S F, GU Z, HE B J, et al. Modelling of ion beam current density of ion thruster[J]. Chinese Journal of Vacuum Science and Technology, 2015, 35(12): 1414-1419 (in Chinese).

[28] 卿绍伟, 鄂鹏, 段萍. 壁面二次电子发射对霍尔推力器放电通道绝缘壁面双鞘特性的影响[J]. 物理学报, 2013, 62(5): 279-286.
QING S W, E P, DUAN P. Effect of wall secondary electron emission on the characteristics of double sheath near the dielectric wall in Hall thruster[J]. Acta Physica Sinica, 2013, 62(5): 279-286 (in Chinese).

[29] 杨三祥, 王倩楠, 高俊, 等. 径向磁场对霍尔推力器性能影响的数值模拟研究[J]. 物理学报, 2022, 71(10): 347-355.
YANG S X, WANG Q N, GAO J, et al. Numerical study of the effect of radial magnetic field on performance of Hall thruster[J]. Acta Physica Sinica, 2022, 71(10): 347-355 (in Chinese).

Experimental study on beam characteristics of μHT-1 thruster under wide range adjustment of operating parameters

HUANG Dan[1, 2, 3, 4, 5], LONG Jianfei[1, 2, 3, 4, 5, *], CHENG Ye[1, 2, 3, 4, 5], WANG Jiabin[1, 2, 3, 4, 5], XU Luxiang[1, 2, 3, 4, 5], YANG Wei[6]

(1. School of Fundamental Physics and Mathematical Sciences, Hangzhou Institute for Advanced Study, UCAS, Hangzhou, 310024, Zhejiang, China;
2. National Space Science Centre, Chinese Academy of Sciences, Beijing 100190, China;
3. University of Chinese Academy of Sciences, Beijing 100049, China;
4. Gravitational Wave Universe Taiji Laboratory, Hangzhou Institute for Advanced Study, UCAS, Hangzhou 310024, China;
5. Key Laboratory of Gravitational Wave Precision Measurement of Zhejiang Province, UCAS, Hangzhou 310024, China;
6. School of Electrical Engineering, Chongqing University of Science and Technology, Chongqing 401331 China)

Abstract: The beam properties of the micro Hall thruster μHT-1, which is intended for space gravitational wave detection, are being experimentally investigated for the first time. A Faraday probe combined with a three-dimensional mobile mechanism was used for diagnosis, and the beam ion current density distribution under a wide range of the anode voltage from 700 to 1 200 V and the anode mass flow from 0.1 to 0.5 sccm was obtained. Moreover, the variation trend of total ion beam current, anode current, current utilization efficiency, beam divergence angle and other parameters were further analyzed. According to test data, the spatial distribution of beam ions changes from dense to sparse as one moves away from the axial direction of the μHT-1 thruster. Additionally, the beam becomes more flat due to the diffusion motion of beam ions in space and the binding motion of electrons and ions. The μHT-1 thruster can work stably under a wide range of conditions (anode voltage 700～1 200 V, anode mass flow 0.1～0.5 sccm), and the beam current presents a good axis-symmetric distribution. With the increase of anode voltage, the average temperature of electrons can be increased, which further leads to the increase of ionization rate, that is, the current utilization efficiency increases from 53.4% to 67.7%; and the magnetic field restraint ability of high-energy electrons is weakened, affecting the electric field focusing, and the beam divergence angle increases from 41.3 ° to 56.1 °. With the increase of anode mass flow, the neutral atom density distribution in the channel is affected, the current utilization efficiency fluctuates in the range of 57.1 % ～ 66.8 %; and the ion collision zone is transferred to the outlet, making the beam divergence angle increase from 43.4 ° to 56.7 °. The total ion beam current of the thruster changes linearly with anode mass flow and anode voltage, which provides the data basis for the subsequent thrust wide range adjustment and thrust resolution analysis.

Keywords: micro Hall thruster; beam characteristics; ion current density; plasma diagnosis; experimental study

Received: 2023-06-21; **Accepted**: 2023-10-13; **Published Online**: 2024-01-05 15:25

URL: link.cnki.net/urlid/11.2625.V.20240104.1315.002

Foundation items: National Key R&D Program of China (2021YFC2202700); Natural Science Foundation of Hunan Province (2021JJ30564); Key Project of Hunan Provincial Department of Education (19A440); Science and Technology Research Project of Chongqing Municipal Education Commission (KJZD-K202101506); Independent Project of Hangzhou Institute of Advanced Studies, National University of Science and Technology (2022ZZ01009)

*** Corresponding author**. E-mail: ljf510@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0312

等离子体激励对火星条件下翼型气动特性影响

杨香港[1]，高永新[1]，汪忠明[1]，李益文[2]，姚程[1,*]

(1. 合肥工业大学 土木与水利工程学院，安徽 合肥 230009；2. 空军工程大学 等离子体动力学重点实验室，陕西 西安 710038)

摘 要：由于火星大气密度低、气压小，火星无人机翼型气动性能亟待进一步提高。采用等离子体激励主动流动控制技术提高火星条件下的翼型升力、降低翼型阻力。在火星低雷诺数条件下研究了等离子体激励的作用位置、激励功率及来流攻角对翼型升力和阻力的影响。结果表明：等离子体激励在下表面尾缘区域增升，最大增升率为37%；在下表面前缘区域减阻，最大减阻率为8%；激励功率越大，来流攻角越小，翼型升阻比提升越明显。等离子体激励诱导压力波，在激励的上、下游分别形成增压区和减压区，导致翼型表面形成增压面和减压面。当激励位置靠近尾缘，增压面扩大，翼型上、下表面压差增大，从而实现增升；当激励位置靠近前缘，减压面扩大，翼型压差阻力降低，从而实现减阻。

关 键 词：火星无人机；低雷诺数；等离子体激励；增升；减阻

中图分类号：V211.4

文献标志码：A **文章编号**：1001-5965(2025)06-2129-08

无人机可以有效提高火星探索效率，弥补火星车的不足[1]。然而，火星大气密度极低，约为地球大气密度的1/70，且火星气压仅为地球标准气压的1%[2]。稀薄的大气环境使得火星无人机翼型在高转速、低雷诺数条件下工作。例如，美国“机智号”火星无人机[3]翼型转速约为2 500 r/min，最大雷诺数约为2.4×10^4。在相同转速条件下，火星无人机翼型产生的升力仅为地球大气环境中的1%左右。火星无人机翼型升力具有很大的提升空间，亟待采用流动控制技术增强火星无人机翼型气动性能。

为提升飞行器的气动特性，国内外学者针对流动控制技术开展了广泛研究。流动控制技术主要分为被动流动控制和主动流动控制[4]。被动流动控制方式包括襟翼系统[5]、鼓包构型[6]和涡流发生器[7]等。这些方法一般是通过改变外表特征实现流动控制，只能针对特定流场环境，灵活性不足。主动流动控制方法包括抽吸[8]、吹除[9]、合成射流[10]和等离子体激励[11]等。其中，发展最迅速的是合成射流和等离子体激励主动流动控制技术。合成射流流动控制环境适应性不足、能量效率不高[12]。等离子体激励主动流动控制技术具有结构简单、响应迅速、无活动部件等优点[13]。

针对火星环境下的流动控制，国内外学者进行了一定研究。陆政旭等[14]采用局部振动的流动控制方法，抑制了火星翼型的层流分离，提高了火星翼型的升阻比。Tabuchi等[15]研究了磁流体流动控制对火星直接入轨和轨道入轨的影响。Takagaki等[16]通过实验评估了火星环境下的等离子体激励流动控制的可行性。

等离子体激励主要通过气体放电产生压力、温度、物性变化，从而对气流施加可控扰动[17]。等离子体激励方式包括介质阻挡放电[18]、火花放电[19]和电弧放电[20]等。其中，电弧放电等离子体通过表面电弧放电产生局部能量沉积，对流场形成温度、

收稿日期：2023-06-05；录用日期：2023-07-21；网络出版时间：2023-08-18 12:27
网络出版地址：link.cnki.net/urlid/11.2625.V.20230817.1606.002
基金项目：国家自然科学基金(51906054,51776222)
*通信作者. E-mail：yaocheng@hfut.edu.cn

引用格式：杨香港，高永新，汪忠明，等. 等离子体激励对火星条件下翼型气动特性影响[J]. 北京航空航天大学学报，2025，51(6)：2129-2136. YANG X G，GAO Y X，WANG Z M，et al. Effect of plasma excitation on aerodynamic characteristics of airfoil in Martian atmosphere[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2129-2136 (in Chinese).

压力扰动，能量转化效率为 5%～10%，具有激励强度大、激励频带宽的显著优势[21]。

在火星环境下进行等离子体放电，与地球有一定差异。地球大气的主要成分为 78% 的 N_2 和 21% 的 O_2。火星大气的主要成分是 CO_2，约占总量的 96%[22]。火星大气中存在许多活跃的带电粒子，大气电导率比地球高 2 个数量级，使得等离子体放电更易于发生[23]。并且，低温、低压的火星环境也有利于等离子体放电[24]，更适合采用等离子体激励开展翼型流动控制。

本文采用计算流体力学方法，研究火星条件下电弧放电等离子体激励对翼型气动特性的影响。分析了激励位置、激励功率及来流攻角对翼型升力、阻力的作用效果。探究了低雷诺数下等离子体激励增升减阻的作用机理。

1 数值方法

为验证连续介质假设对火星稀薄大气的适用性，采用克努森数 Kn 进行分析。根据文献 [25] 给出的克努森数计算方法，可计算出火星大气在 223 K 时，气体流动的克努森数 $Kn = 8.75\times10^{-9} < 0.01$，属于连续介质的范围。

基于唯象学描述方法，不关注放电的具体过程和所涉及的化学反应，将等离子体激励简化为局部区域的能量注入，激励区域简化为矩形[26]。通过在能量方程中添加能量源项的方式模拟等离子体激励对流场的影响。采用可压缩 N-S 方程描述气体流动，如式 (1)～(4) 所示。

$$\frac{\partial\rho}{\partial t}+\nabla(\rho\boldsymbol{u})=0 \tag{1}$$

$$\frac{\partial(\rho u)}{\partial t}+\nabla(\rho u\boldsymbol{u})=\nabla\boldsymbol{\sigma}-\frac{\partial p}{\partial x} \tag{2}$$

$$\frac{\partial(\rho v)}{\partial t}+\nabla(\rho v\boldsymbol{u})=\nabla\boldsymbol{\sigma}-\frac{\partial p}{\partial y} \tag{3}$$

$$\frac{\partial(\rho e)}{\partial t}+\nabla\left[(\rho e+p)\boldsymbol{u}\right]=\nabla(\boldsymbol{u}\boldsymbol{\sigma})+\nabla(\kappa\nabla T)+Q \tag{4}$$

式中：ρ 为气体密度；$\boldsymbol{u}$ 为速度矢量；u 和 v 为速度矢量 $\boldsymbol{u}$ 在 x 和 y 方向的分量；$\boldsymbol{\sigma}$ 为黏性应力张量；p 为气体静压；T 为气体静温；κ 为热传导系数；Q 为电弧放电产生焦耳热的功率密度。

$\boldsymbol{\sigma}$ 表达式为

$$\sigma=\mu\left[\nabla\boldsymbol{u}+(\nabla\boldsymbol{u})^{\mathrm{T}}\right]-\frac{2}{3}(\nabla\boldsymbol{u})\boldsymbol{I} \tag{5}$$

式中：μ 为动力黏度；$\boldsymbol{I}$ 为单位矩阵。

Q 表达式为

$$Q=Q_0G(x/L,y/L)S(t/\tau_e) \tag{6}$$

式中：Q_0 为热能注入的幅值；$G(x/L,y/L)$ 为热能的空间分布函数；L 为特征长度；$S(t/\tau_e)$ 为热能的时间分布函数，电弧放电达到稳定状态，则 $S(t/\tau_e)=1$[27]。

本文所述的雷诺数为基于翼型弦长的参考雷诺数。计算为非定常计算，时间步长为 10^{-6} s，并采用形成稳定流场后再施加激励的计算方法。采用的等离子体激励为持续激励，激励区域尺寸为 4 mm×0.5 mm。在模拟可压缩气体流动时，采用密度基隐式算法。气体密度由理想气体模型给出。动力黏度采用 Sutherland 公式。通量项采用 Roe 格式进行离散。动量、湍能耗散率和湍流动能均采用二阶迎风格式。湍流模型采用 SST k-ω 模型。计算网格采用三角形单元和矩形单元进行划分。在翼型表面附近生成矩形网格，第一层网格高度为 7.8×10^{-2} mm，逐层增长率为 1.2，$y^+\leqslant1$。以 3 种网格划分方案进行网格无关性验证。

2 模型验证

本文选取文献 [28] 中雷诺数 $Re=2\times10^4$ 的实验数据进行模型验证。计算模型如图 1 所示。翼型弦长 c 为 100 mm，弦高比为 0.06。在火星环境下，取来流速度为 132.44 m/s，静压为 727.87 Pa，环境温度为 223 K。经计算雷诺数为 2×10^4。翼型壁面条件采用无滑移边界。自由来流边界为压力远场，与翼型的距离为 $500c$。

在攻角 $\alpha=5°$ 和 $Re=2\times10^4$ 条件下进行网格无关性验证，验证结果见表 1 所列。从表 1 可以看出，网格Ⅱ的升力系数 C_l 和阻力系数 C_d 接近精细

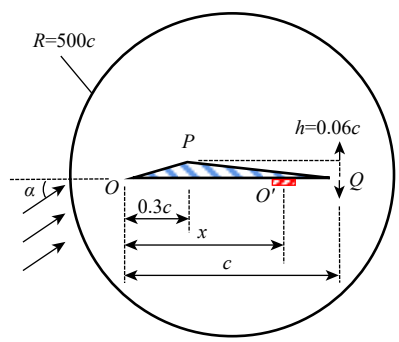


图 1 模型示意图

Fig. 1 Schematic diagram of model

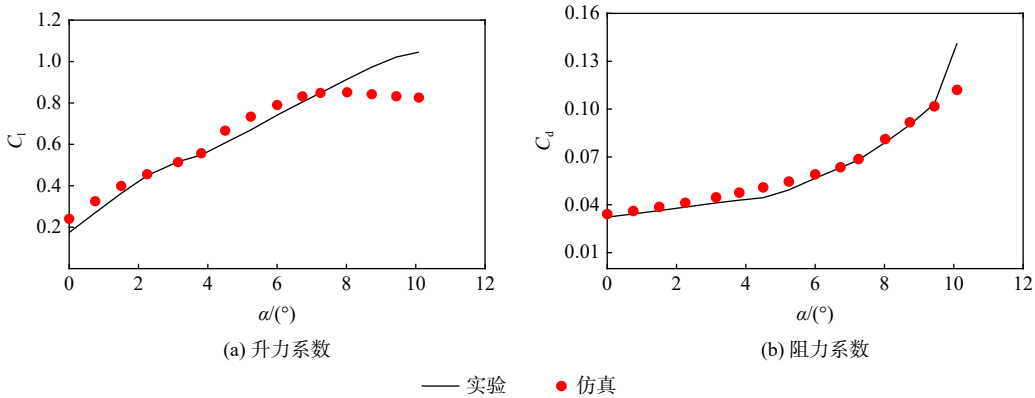
表 1 网格无关性

Table 1 Mesh independence

编号	网格数	C_l	C_d	y^+
Ⅰ	50 107	0.671 9	0.048 7	≤1
Ⅱ	105 669	0.673 0	0.049 5	≤1
Ⅲ	233 701	0.673 1	0.050 1	≤1

化网格Ⅲ。故本文最终采用网格Ⅱ进行计算。

翼型升力系数和阻力系数的实验数据与数值仿真结果如图 2 所示。从图 2 可以看出, 在 0°～8°攻角范围内, 仿真结果与实验数据基本吻合。



(a) 升力系数　　(b) 阻力系数

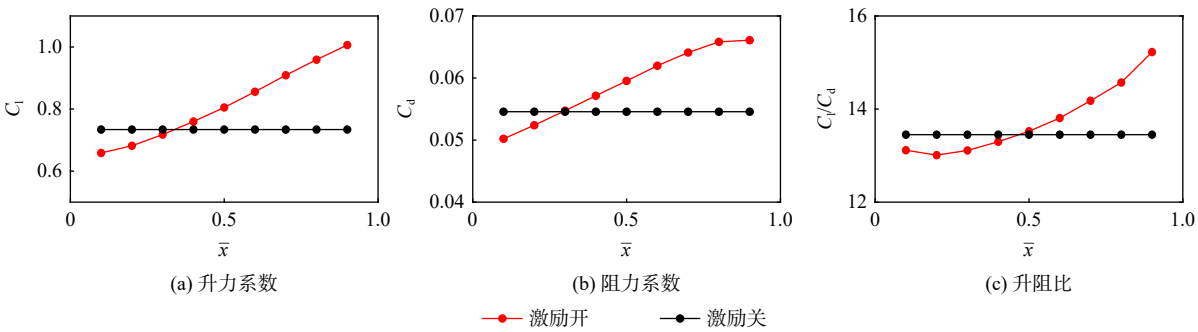
图 2　仿真数据和实验数据的比较

Fig. 2　Comparison between simulated data and experiment data

3　结果与分析

3.1　激励位置的影响

前期研究表明, 在翼型上表面施加等离子体激励无法实现增升减阻[29]。因此, 在翼型下表面 0.1c～0.9c 等间距的 9 个位置施加等离子体激励。记激励位置 $\bar{x}$= x/c。激励功率为 2 000 W。当来流攻角为 5°时, 激励位置对翼型升阻特性的影响如图 3 所示。



(a) 升力系数　　(b) 阻力系数　　(c) 升阻比

图 3　不同激励位置的翼型气动特性

Fig. 3　Aerodynamic characteristics of airfoil with different excitation positions

由图 3(a) 可知, 等离子体激励在翼型前缘使升力减小, 在翼型尾缘使升力增大。最佳增升激励位置位于 0.9c 处, 最大增升率为 37%。因此, 在翼型尾缘附近施加等离子体激励有利于提高翼型升力。

由图 3(b) 可知, 等离子体激励在翼型前缘使阻力减小, 在翼型尾缘使阻力增大。最佳减阻激励位置位于 0.1c 处, 最大减阻率为 8%。因此, 在翼型前缘附近施加激励有利于降低翼型阻力。

由图 3(c) 可知, 等离子体激励在翼型前缘使升阻比降低, 在翼型尾缘使升阻比增大。激励位置位于 0.9c 时, 翼型升阻比提升最大, 升阻比最大提升率为 13%。

通过以上分析可得, 等离子体激励在翼型前缘附近起到减升、减阻效果, 在翼型尾缘附近起到增升、增阻效果。

3.2　激励功率及攻角的影响

取激励位置 $\bar{x}$= 0.9, 改变激励输入功率和来流攻角, 得到翼型升阻特性的变化如图 4 所示。由图 4(a) 可知, 激励功率越大, 对翼型升力的提升越大。随着来流攻角的增大, 等离子体激励对升力的提升效果逐渐降低。由图 4(b) 可知, 激励功率越大, 对翼型阻力的增加越大。来流攻角越大, 等离子体激励对翼型阻力的增加越显著。由图 4(c) 可知, 在 0°～6°攻角范围内, 激励功率越大, 翼型升阻比越高。在 6°～10°攻角范围内, 激励功率越大, 翼型升阻比越低。通过以上分析可得, 激励功率越大, 来流攻角越小, 等离子体的激励效果越好。

3.3　增升减阻机理

选取激励功率为 2 000 W, 来流攻角为 5°的工况进行增升减阻机理分析。无激励的工况记为 P_0。对激励位置 $\bar{x}$= 0.1、$\bar{x}$= 0.5 和 $\bar{x}$= 0.9 分别施加激

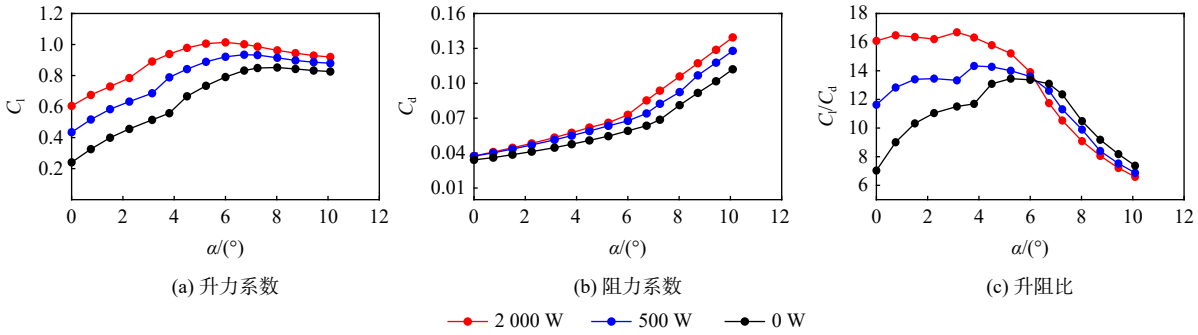


图4 不同激励功率及攻角的翼型气动特性

Fig. 4 Aerodynamic characteristics with different excitation power and angle of attack

励，记为 P_1、P_2 和 P_3。将翼型表面分为3个区域：上表面 OPQ；下表面激励位置上游 OO'；下表面激励位置下游 $O'Q$，如图1所示。

3种工况下翼型的表面压力系数 C_p 如图5所示。在激励作用下，翼型上表面 OPQ 的压力减小，形成第1减压面。下表面激励上游区域 OO' 的压力增大，形成增压面；激励下游区域 $O'Q$ 的压力减小，形成第2减压面。即翼型整个上表面为第1减压面，翼型下表面位于激励位置下游的部分为第2减压面。激励位置越靠近尾缘，增压面越大，第2减压面越小，翼型上、下表面的压差随之增大。因此，等离子体激励在尾缘区域实现增升效果。

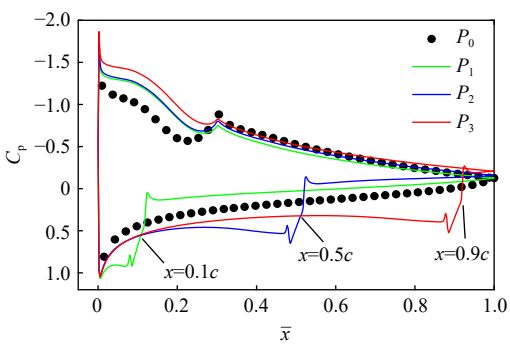


图5 翼型表面压力分布

Fig. 5 Pressure distributions on surface of airfoil

等离子体激励的减阻效果同样受到增压面和减压面的影响。以9组不同激励位置下的翼型阻力数据展开分析，如表2所示。翼型阻力由压差阻力和摩擦阻力构成。由表2可得，激励位置越靠近前缘，翼型压差阻力越小，摩擦阻力越大。

当激励位置靠近前缘时，第二减压面增大，翼型压差阻力因此减小。翼型的摩擦阻力与表面切应力 τ 有关，表面切应力表达式为

$$\tau = \mu \left.\frac{\partial U}{\partial y}\right|_{y=0} \tag{7}$$

式中：$\partial U/\partial y|_{y=0}$ 为近壁流向速度的法向梯度。

表2 不同激励位置下的翼型阻力系数

Table 2 Drag coefficient with excitation positions

激励位置	压差阻力系数	摩擦阻力系数	总阻力系数
无激励	0.037 8	0.016 8	0.054 6
0.1c	0.019 8	0.030 4	0.050 2
0.2c	0.023 5	0.028 9	0.052 4
0.3c	0.027 0	0.027 8	0.054 8
0.4c	0.030 5	0.026 7	0.057 2
0.5c	0.034 1	0.025 4	0.059 5
0.6c	0.037 9	0.024 1	0.062 0
0.7c	0.041 7	0.022 4	0.064 1
0.8c	0.045 4	0.020 4	0.065 8
0.9c	0.048 6	0.017 4	0.066 0

气体黏度系数随温度的升高而增大。P_1、P_2 和 P_3 对翼型流场温度的影响如图6所示。等离子体激励能量的注入，引起激励下游温度的升高，形成高温区。该区域内的气体黏度增大。

P_2 工况下翼型下表面5个位置的法向速度梯度如图7所示。激励区域的法向速度梯度增大，激励上游和激励下游高温区的法向速度梯度降低。

等离子体激励对高温区气体黏度和法向速度梯度的作用是相反的。3种工况下翼型下表面的壁面摩擦系数 C_f 如图8所示。在激励下游的高温区，翼型壁面摩擦系数增大。这表明：气体黏度对摩擦阻力的影响占主导，使得摩擦阻力增大。

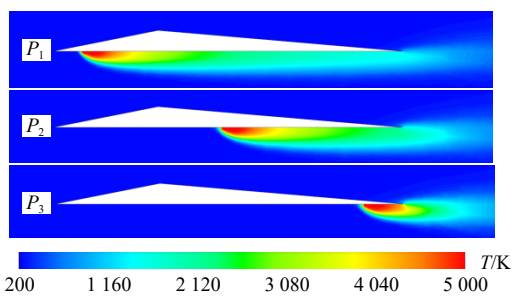


图6 翼型流场温度分布

Fig. 6 Temperature distribution around airfoil flow field

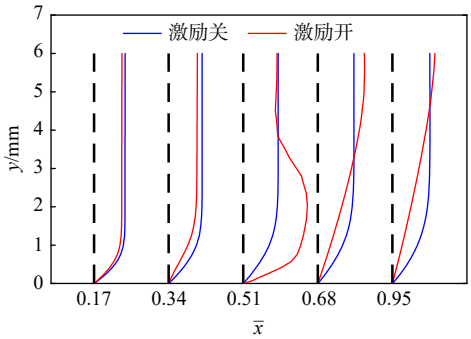


图 7　翼型下表面法线方向的速度分布 (P_2 工况)

Fig. 7　Velocity distribution along normal direction on lower surface of airfoil (P_2 case)

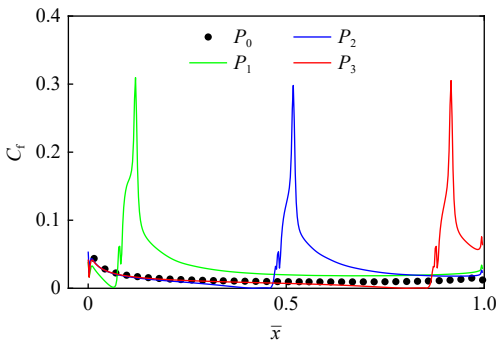


图 8　翼型下表面壁面摩擦系数分布

Fig. 8　Skin friction coefficient distribution on lower surface of airfoil

等离子体激励位置对翼型压差阻力和摩擦阻力的影响是相反的。激励位置越靠近前缘, 翼型阻力越小。这表明: 翼型阻力中的压差阻力占主导地位。在前缘区域, 等离子体激励可以通过降低压差阻力达到减阻效果。

3.4　等离子体全局影响

上述分析表明, 翼型的增升减阻效果主要受到等离子体激励在翼型表面形成的增压面和减压面的影响。翼型表面的压力分布是等离子体激励全局影响的结果。以 P_3 工况为例, 进一步探究等离子体激励对翼型流场的全局影响。

激励启动瞬间, 激励位置的压力迅速升高, 诱导出压力波。在火星条件下, 压力波以 230 m/s 的速度向周围扩散, 随后快速衰减, 如图 9 所示。

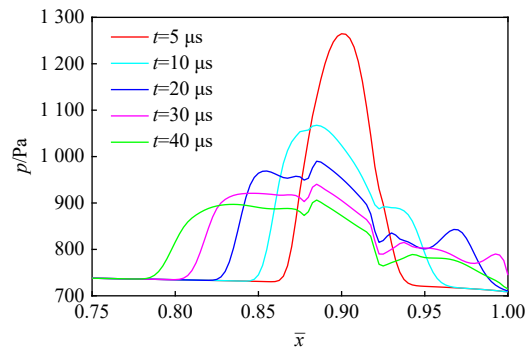


图 9　压力波发展过程 (P_3 工况)

Fig. 9　Pressure wave propagation for P_3 case

等离子体激励诱导的压力波在流场中形成一个启动涡[30], 启动涡发展过程如图 10 所示。当 t = 5 μs 时, 激励位置的压力显著升高, 形成启动涡。当 t = 30 μs 时, 压力波越过翼型尾缘, 沿上表面向上游发展, 此时启动涡半径为 12 cm。随后压力波持续传播, 影响区域逐渐扩大, 最终对翼型流场造成全局扰动。

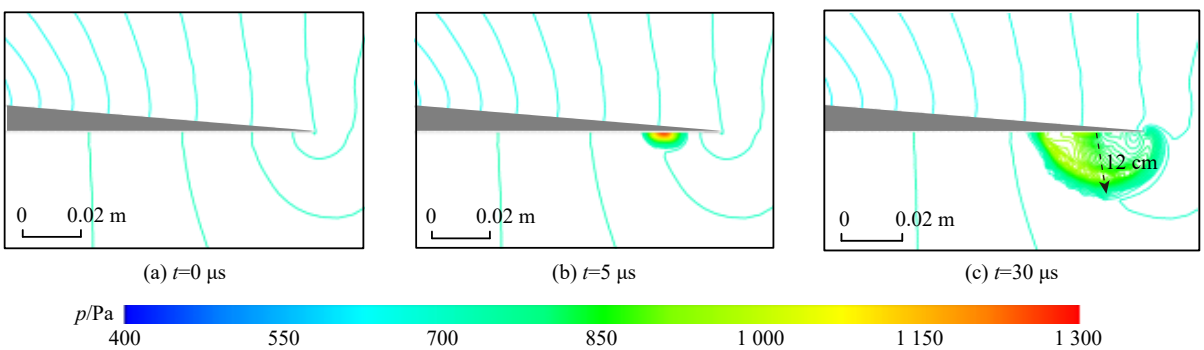
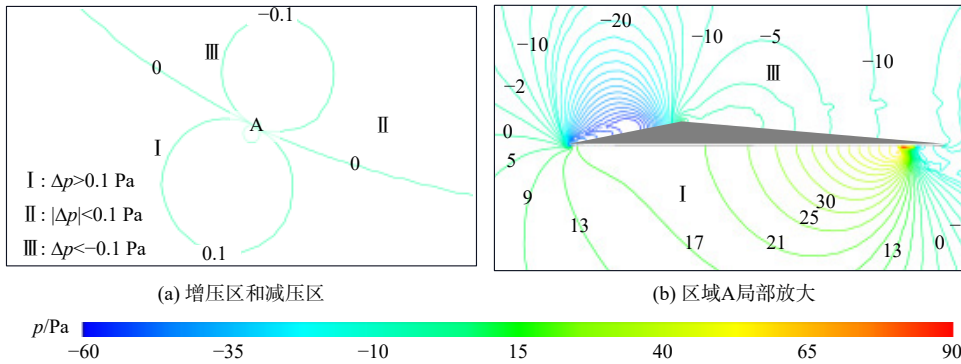


图 10　启动涡发展过程 (P_3 工况)

Fig. 10　Strat-up vortex development process for P_3 case

当 t > 1 s 后, 压力波消散, 全局扰动趋于稳定, 翼型周围流场形成稳定的增压区和减压区。用激励后的压力分布减去激励前的压力分布得到流场压力变化量的分布 $\Delta p(x,y)$, 如图 11 所示。

图 11 中区域 Ⅰ 视为增压区, 直径约为 10 m, 最大压升为 90 Pa。区域 Ⅱ 视为无影响区。区域 Ⅲ 视为减压区, 直径约为 8 m, 最大压降为 60 Pa。在翼型下表面, 激励位置上游 OO' 位于区域 Ⅰ, 压力增大; 上表面 OPQ 和下表面激励位置下游 $O'Q$ 在区域 Ⅲ 中, 压力减小。由此可得, 翼型表面的增压面与减压面, 是等离子体激励对翼型流场全局扰动的结果。改变激励位置, 增压区和减压区随之变化,



(a) 增压区和减压区　　(b) 区域A局部放大

图 11　等离子体激励形成的增压区和减压区 (P_3 工况, t > 1 s)

Fig. 11　Pressurized region and depressurized region induced by plasma excitation (P_3 case, t > 1 s)

从而实现翼型的增升、减阻。

4　结　论

本文研究了等离子体激励在火星低雷诺数条件下对翼型气动特性的影响及其作用机理。得到如下结论:

1) 等离子体激励在下表面尾缘区域增升, 最大增升率为 37%; 在下表面前缘区域减阻, 最大减阻率为 8%。

2) 等离子体激励诱导压力波, 在翼型表面形成 1 个增压面和 2 个减压面。

3) 激励位置靠近下表面尾缘时, 增压面扩大, 翼型上、下表面压差增大, 翼型升力因此提升。

4) 激励位置靠近下表面前缘时, 增压面减小, 第二减压面增大, 导致压差阻力降低, 翼型阻力因此降低。

参考文献 (References)

[1] 赵鹏越, 全齐全, 邓宗全, 等. 旋翼式火星无人机技术发展综述[J]. 宇航学报, 2018, 39(2): 121-130.
ZHAO P Y, QUAN Q Q, DENG Z Q, et al. Overview of research on rotary-wing Mars unmanned aerial vehicles[J]. Journal of Astronautics, 2018, 39(2): 121-130 (in Chinese).

[2] KONING W J F, JOHNSON W, GRIP H F. Improved Mars helicopter aerodynamic rotor model for comprehensive analyses[J]. AIAA Journal, 2019, 57(9): 3969-3979.

[3] CHAHAT N, MILLER J, DECROSSAS E, et al. The Mars helicopter telecommunication link: antennas, propagation, and link analysis[J]. IEEE Antennas and Propagation Magazine, 2020, 62(6): 12-22.

[4] 李益文, 王宇天, 庞垒, 等. 进气道等离子体/磁流体流动控制研究进展[J]. 力学学报, 2019, 51(2): 311-321.
LI Y W, WANG Y T, PANG L, et al. Research progress of plasma/MHD flow control in inlet[J]. Chinese Journal of Theoretical and Applied Mechanics, 2019, 51(2): 311-321 (in Chinese).

[5] 王万波, 姜裕标, 黄勇, 等. 大型飞机襟翼吹气增升风洞试验[J]. 航空学报, 2023, 44(13): 127870.
WANG W B, JIANG Y B, HUANG Y, et al. Lift enhancement wind tunnel test with flap blowing for large aircraft[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(13): 127870 (in Chinese).

[6] 章胜华, 邓枫, 覃宁, 等. 激波控制鼓包对跨声速抖振影响的数值研究[J]. 航空学报, 2022, 43(11): 526806.
ZHANG S H, DENG F, QIN N, et al. Numerical study on impact of shock control bump on transonic buffet[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(11): 526806 (in Chinese).

[7] 李超, 黄河峡, 罗金玲, 等. 基于微型叶片式涡流发生器的前体压缩面低能流掺混机理[J]. 空气动力学学报, 2022, 40(1): 119-128.
LI C, HUANG H X, LUO J L, et al. Mixing of low-momentum fluids over the forebody ramp based on micro-vane vortex generators[J]. Acta Aerodynamica Sinica, 2022, 40(1): 119-128 (in Chinese).

[8] 邹琳, 左红成, 柳迪伟, 等. 基于定常吹吸气的波浪型圆柱主动控制研究[J]. 力学学报, 2022, 54(11): 2970-2983.
ZOU L, ZUO H C, LIU D W, et al. Active flow control of wavy cylinder based on steady blowing and suction[J]. Chinese Journal of Theoretical and Applied Mechanics, 2022, 54(11): 2970-2983 (in Chinese).

[9] SU W Y, CHEN Y, ZHANG F R, et al. Control of pseudo-shock oscillation in scramjet inlet-isolator using periodical excitation[J]. Acta Astronautica, 2018, 143: 147-154.

[10] YU Y, XU J L, GAN N G. Effects of parameters on continuously working plasma synthetic jet[J]. Engineering Applications of Computational Fluid Mechanics, 2014, 8(1): 55-69.

[11] 王宏宇, 闵夫, 解真东, 等. 直流放电控制高速带斜坡锥体气动力的有效性[J/OL]. 空气动力学学报. (2023-1-10)[2023-02-16]. https://kns.cnki.net/kcms/detail/51.1192.TK.20230216.1037.002.html.
WANG H Y, MIN F, XIE Z D, et al. Effectiveness of DC discharge in controlling aerodynamic force of high-speed cone with slope [J/OL]. Acta Aerodynamica Sinica. (2023-1-10)[2023-02-16]. https://kns.cnki.net/kcms/detail/51.1192.TK.20230216.1037.002.html (in Chinese).

[12] 罗振兵, 夏智勋, 邓雄, 等. 合成双射流及其流动控制技术研究进展[J]. 空气动力学学报, 2017, 35(2): 252-264,251.
LUO Z B, XIA Z X, DENG X, et al. Research progress of dual synthetic jets and its flow control technology[J]. Acta Aerodynamica Sinica, 2017, 35(2): 252-264,251 (in Chinese).

[13] 张鑫, 王勋年. 正弦交流介质阻挡放电等离子体激励器诱导流场研究的进展与展望[J]. 力学学报, 2023, 55(2): 285-298.
ZHANG X, WANG X N. Research progress and outlook of flow

field created by dielectric barrier discharge plasma actuators driven by a sinusoidal alternating current high-voltage power[J]. Chinese Journal of Theoretical and Applied Mechanics, 2023, 55(2): 285-298 (in Chinese).

[14] 陆政旭, 陈肇麟, 肖天航, 等. 火星旋翼无人机翼型局部振动的数值模拟[C]//2021 中国力学大会. 南京: 南京航空航天大学, 2022: 480-492.
LU Z X, CHEN Z L, XIAO T H, et al. Numerical simulation of airfoil for Martian rotor with local oscillation[C]//Proceedings of the 2021 China Mechanics Congress. Nanjing: Nanjing University of Aeronautics and Astronauties, 2022: 480-492(in Chinese).

[15] TABUCHI K, SUMITOMO R, TANAKA K, et al. Numerical analysis of magnetohydrodynamic flow control in Mars direct and orbital entries[C]// AIAA SCITECH 2023 Forum. Reston: AIAA, 2023: AIAA2023-2559.

[16] TAKAGAKI M, ISONO S, NAGAI H, et al. Evaluation of plasma actuator performance in Martian atmosphere for applications to Mars airplanes[C]// 4th Flow Control Conference. Reston: AIAA, 2008: 2008-3762.

[17] 李应红, 吴云, 梁华, 等. 等离子体激励气动力学探索与展望[J]. 力学进展, 2022, 52(1): 1-32.
LI Y H, WU Y, LIANG H, et al. Exploration and outlook of plasma-actuated gas dynamics[J]. Advances in Mechanics, 2022, 52(1): 1-32 (in Chinese).

[18] 黄广靖, 戴玉婷, 杨超. 低雷诺数俯仰振荡翼型等离子体流动控制[J]. 力学学报, 2021, 53(1): 136-155.
HUANG G J, DAI Y T, YANG C. Plasma-based flow control on pitching airfoil at low Reynolds number[J]. Chinese Journal of Theoretical and Applied Mechanics, 2021, 53(1): 136-155 (in Chinese).

[19] 刘汝兵, 孙伟, 牛中国, 等. 火花放电等离子体射流实验研究[J]. 推进技术, 2015, 36(3): 372-377.
LIU R B, SUN W, NIU Z G, et al. Experimental investigation on spark discharge plasma jet[J]. Journal of Propulsion Technology, 2015, 36(3): 372-377 (in Chinese).

[20] 王浩, 程邦勤, 纪振伟, 等. 局部电弧丝状放电控制激波/边界层干扰的数值研究[J]. 推进技术, 2017, 38(11): 2431-2438.
WANG H, CHENG B Q, JI Z W, et al. Numerical simulation of localized arc filament plasma actuator for shock wave/boundary layer interaction control[J]. Journal of Propulsion Technology, 2017, 38(11): 2431-2438 (in Chinese).

[21] 吴云, 李应红. 等离子体流动控制研究进展与展望[J]. 航空学报, 2015, 36(2): 381-405.
WU Y, LI Y H. Progress and outlook of plasma flow control[J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(2): 381-405 (in Chinese).

[22] OGLOBLINA P, MORILLO-CANDAS A S, SILVA A F, et al. Mars in situ oxygen and propellant production by non-equilibrium plasmas[J]. Plasma Sources Science Technology, 2021, 30(6): 065005.

[23] 钱茂程. 模拟火星条件下 CO2 辉光放电产物及发射光谱研究[D]. 济南: 山东大学, 2021.
QIAN M C. Study on CO2 glow discharge products and emission spectra under simulated Mars conditions[D]. Jinan: Shandong University, 2021 (in Chinese).

[24] GUERRA V, SILVA T, PINHÃO N, et al. Plasmas for in situ resource utilization on Mars: Fuels, life support, and agriculture[J]. Journal of Applied Physics, 2022, 132(7): 070902.

[25] 杨婷婷. 火星无人机梯形桨叶空气动力学特性分析及实验研究[D]. 哈尔滨: 哈尔滨工业大学, 2018.
YANG T T. Analysis and experimental study on aerodynamic characteristics of trapezoidal blades of Mars UAV[D]. Harbin: Harbin Institute of Technology, 2018 (in Chinese).

[26] 王宇天, 张百灵, 李益文, 等. 等离子体激励控制激波与边界层干扰流动分离数值研究[J]. 航空动力学报, 2018, 33(2): 364-371.
WANG Y T, ZHANG B L, LI Y W, et al. Numerical investigation for control of shock wave and boundary layer interactions flow separation with plasma actuation[J]. Journal of Aerospace Power, 2018, 33(2): 364-371 (in Chinese).

[27] 林敏, 徐浩军, 梁华, 等. 等离子体气动激励减小翼型跨音速阻力的数值模拟[J]. 空军工程大学学报 (自然科学版), 2012, 13(2): 6-10.
LIN M, XU H J, LIANG H, et al. Numerical simulation of plasma aerodynamic actuation for airfoil transonic drag reduction[J]. Journal of Air Force Engineering University (Natural Science Edition), 2012, 13(2): 6-10 (in Chinese).

[28] HIDAKA H, OKAMOTO M. An experimental study of triangular airfoils for Mars airplane[J]. Transactions of the Japan Society of Aeronautical and Space Sciences, Aerospace Technology Japan, 2014, 12(20): 21-27.

[29] 刘加伟, 柳兆涛, 丁仕洪, 等. 等离子体热效应对 NACA0012 翼型增升减阻的研究[J]. 推进技术, 2020, 41(5): 1055-1062.
LIU J W, LIU Z T, DING S H, et al. Study on lift enhancement and drag reduction of NACA0012 airfoil under plasma thermal effect[J]. Journal of Propulsion Technology, 2020, 41(5): 1055-1062 (in Chinese).

[30] LI Y W, LI Y H, ZHOU Z W, et al. Experimental investigation on induced flow velocity of plasma aerodynamic actuation[J]. Transactions of Nanjing University of Aeronautics & Astronautics, 2009, 26(1): 23-28.

Effect of plasma excitation on aerodynamic characteristics of airfoil in Martian atmosphere

YANG Xianggang[1]，GAO Yongxin[1]，WANG Zhongming[1]，LI Yiwen[2]，YAO Cheng[1,*]

(1. College of Civil Engineering，Hefei University of Technology，Hefei 230009，China；
2. Science and Technology on Plasma Dynamics Laboratory，Air Force Engineering University，Xi'an 710038，China)

Abstract：The aerodynamic characteristics of the airfoil for the Martian unmanned aerial vehicles (UAVs) need to be improved because of the low density and the low pressure in the Martian atmosphere. The active flow control technology with plasma excitation was used to enhance the airfoil lift and reduce the airfoil drag in the Martian atmosphere. Effects of plasma excitation positions, excitation power, and angle of attack on the lift and drag of the airfoil were studied at a low Reynolds number on Mars. It is found that the plasma excitation increases the airfoil lift in the region of the trailing edge of the lower surface with a maximum increase of 37% and reduces the airfoil drag in the region of the leading edge of the lower surface with a maximum drag reduction of 8%. The lift-drag ratio of the airfoil significantly raises when increasing the excitation power and decreasing the angle of attack. The pressure wave induced by the plasma excitation generates a pressurized zone and a depressurized zone in the upstream and downstream regions of the excitation, respectively. Therefore, pressurized and depressurized surfaces appear on the airfoil. When the excitation gets close to the trailing edge, the pressurized surface enlarges, which leads to a higher pressure difference across the upper and lower surfaces of the airfoil and increases the airfoil lift. When the excitation is located near the leading edge, the depressurized surface expands, which reduces the pressure difference of the airfoil and decreases the airfoil drag.

Keywords：Martian UAV; low Reynolds number; plasma excitation; lift enhancement; drag reduction

Received：2023-06-05；**Accepted**：2023-07-21；**Published Online**：2023-08-18 12：27
URL：link.cnki.net/urlid/11.2625.V.20230817.1606.002
Foundation items：National Natural Science Foundation of China (51906054,51776222);
* **Corresponding author**. E-mail：yaocheng@hfut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0316

双参数退化型新研产品质保期设计与优化

高帅[1,2]，李艳宏[3]，孙富强[1,2,*]

(1. 北京航空航天大学 可靠性与系统工程学院，北京 100191；2. 可靠性与环境工程技术国防科技重点实验室，北京 100191；
3. 中国航发商用航空发动机有限责任公司，上海 200241)

摘 要：由于新研产品缺乏外场故障数据和历史保修索赔记录，难以开展科学合理的质保成本预测及质保期优化。考虑到产品不同性能参数退化过程之间的相互影响，提出了一种基于Copula理论的双参数退化型新研产品质保期设计与优化方法。根据实验室加速退化试验数据建立产品单参数性能退化模型，采用Copula方法量化退化过程之间的相关性，同时考虑产品在外场的动态运行环境，给出其外场可靠性模型。采用维修改善因子模型量化维修过程中的不完美维修情形，并基于蒙特卡罗仿真计算产品的预计失效数，建立质保成本模型。在此基础上，利用Glickman-Berger模型量化质保期对产品销售量的影响，构建以制造商利润最大化为目标的质保期优化模型。以某型电子组件为例，开展产品质保期设计优化与敏感性分析，验证了模型的有效性和适用性。

关 键 词：新研产品；双参数退化；Copula函数；加速退化试验；质保期；敏感性分析
中图分类号：V57; TB114.3
文献标志码：A **文章编号**：1001-5965(2025)06-2137-11

产品质保是由制造商/销售商为消费者购买的产品提供在其质保期内和正常的使用条件下能够实现预定功能的担保[1]。在制定质保策略时，制造商需要综合考虑产品的质保期与质保成本。质保成本指在给定质保期的情况下，未来会产生的维修费用。通常，质保成本 $A(W)=C_S\times E(M(W))$，其中 C_S 为每次维修或更换的费用，$E(M(W))$ 为在质保期 W 内产品预计失效数的期望。因此，开展质保成本的分析和优化的前提是获取产品在质保期内的预计失效数。新研产品在推出市场前，一般是利用前一代产品的索赔数据，对产品的外场可靠性进行评估，以计算预计失效数[2]。但对于缺乏继承性的新研产品，由于没有现场故障数据和历史保修索赔记录，设计质保策略和预测质保成本变得非常具有挑战性[3]。

对此，有研究人员利用新研产品在研发阶段开展的加速寿命试验 (accelerated life test, ALT) 或加速退化试验 (accelerated degradation test, ADT) 数据，来预测其外场可靠性与预计失效数。Tseng 等[4]采用层次模型对实验室寿命试验数据和现场索赔数据联合建模，预测保修成本。Yang 等[5]讨论了三水平加速寿命试验方案优化问题，目标是质保成本的渐近方差最小。Shang 等[6]基于退化数据拟合了逆高斯过程的退化模型，计算质保成本，并从制造商和消费者的角度确定最佳质保政策。Zhao 和 Xie[7]采用 ALT 实验室数据，考虑外场环境协变量建立外场可靠性模型，计算预计失效数，并在此基础上建立质保成本模型。Li 等[8]基于维纳(Wiener)过程建立退化模型，采用截断分布来模拟随机失效阈值，设计了3种带有随机失效阈值的保修策略。

然而，上述文献中所建立的质保成本模型，只涉及一个退化性能指标。随着新研产品结构愈发复杂与工作条件的多样，产品通常存在多个关键性能参数会随着时间显著退化，同时由于共同的结构

收稿日期：2023-06-05；**录用日期**：2023-10-18；**网络出版时间**：2023-11-01 12:14
网络出版地址：link.cnki.net/urlid/11.2625.V.20231101.1119.002
基金项目：重点实验室稳定支持项目 (WDZC20220101)
*通信作者. E-mail：sunfuqiang@buaa.edu.cn

引用格式：高帅，李艳宏，孙富强. 双参数退化型新研产品质保期设计与优化 [J]. 北京航空航天大学学报，2025，51(6)：2137-2147.
GAO S，LI Y H，SUN F Q. Design and optimization of warranty period of new products with two-parameter degradation [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2137-2147 (in Chinese).

材料属性和使用经历，产品多个性能参数的退化过程往往会相互影响[9-10]，存在一定的相关性。近年来，利用Copula函数进行多变量退化建模的研究得到了迅速发展[11]。Chen等[12]用多种随机过程来描述机械密封性能退化过程，性能指标之间的相关性用不同类型的Copula函数来描述。Fang等[13]提出了一种双变量随机过程的模型框架，利用Copula函数分析加速退化数据。Chi等[14]使用Copula函数拟合惯性测量单元多变量相依退化的加速退化试验数据，评估产品的可靠性和剩余寿命。Jiang等[15]使用了4种Copula函数研究了飞机舱门锁机构两种退化过程间的相关性，提出了产品剩余寿命预测模型。

此外，质保成本与维修方案息息相关。维修方案会影响故障产品维修后的使用可靠性，使得在相同质保期内的产品预计失效数有较大差异，影响产品的质保成本[16]。一般来说，修复方式可分为3种：完全修复、不完全修复和最小修复。目前质保的研究多是完美维修或最小修复。然而，由于维修成本或技术能力的限制，这2个假设表达的维修效果都是极端的情况。实际上，不完善修复的应用更为广泛和贴近实际[17]。Chen等[18]讨论了不完善维护对降解系统的正面和负面影响。Zhao等[19]假设单次维修的保修成本与不完善维修的改进因子有关，通过最小化每个产品的预期保修成本，获得最优不完全维修策略。Wang等[20]基于虚拟年龄模型，假设不完善的修复是为了减少最后一次预防性维修间隔中的虚拟年龄，开展产品二维质保政策优化。

综上所述，在工程实际中，采用实验室加速退化试验数据对产品外场可靠性进行预测还存在以下问题：①由于共同的结构材料属性和使用经历，产品多个性能参数的退化过程往往会相互影响，存在一定的相关性；②用户使用习惯和场景的差异，导致产品的运行环境具有较大的随机性；③质保成本模型中的单次维修成本和预计失效数均会受到维修效果的影响，目前研究大都假设产品的维修效果是完美维修。但在实践中，维修效果一般处于“修复如新”与“修复如旧”之间的不完美维修状态。为此，本文针对双参数退化型新研产品质保期优化问题，提出了一种新的基于实验室加速退化数据的质保期优化模型。该模型采用Copula函数量化性能参数退化的相依关系，考虑产品实际动态运行环境，建立外场可靠性模型，并基于维修改善因子模型考虑产品的不完美维修情景，开发了一种新的预计失效数预测方法，全面系统地建立了新研产品质保期优化模型，解决了双参数退化型新研产品缺少历史数据等带来的质保期设计难题。

本文的技术为：首先，利用新研产品在研发阶段的实验室加速退化试验数据，基于Wiener过程和Copula理论建立二元相关退化模型。然后，考虑外场动态运行环境的不确定性影响，基于环境应力正态分布假设给出产品的外场可靠性模型。进一步，采用维修改善因子表征不完美维修效果[6]，建立维修效果与性能失效阈值的定量关系，采用蒙特卡罗仿真方法计算产品的预计失效数。在此基础上，考虑质保期对产品制造方的影响，采用Glickman-Berger销售量模型表征销售量与质保期的函数关系[21]，建立制造商利润模型。最后，构造以制造商利润最大为目标的质保期优化模型，确定最优质保期，为产品质保政策的制定提供决策依据。

1 动态环境二元相依退化模型构建

对于2个性能参数发生退化的产品，将其退化轨迹记为$X_1(t)$和$X_2(t)$，对应的失效阈值记为H_1和H_2，退化过程首次达到阈值的时间为T_{H1}或T_{H_2}，当产品的任一性能参数超过其退化失效阈值时，产品即发生失效，退化失效过程如图1所示。

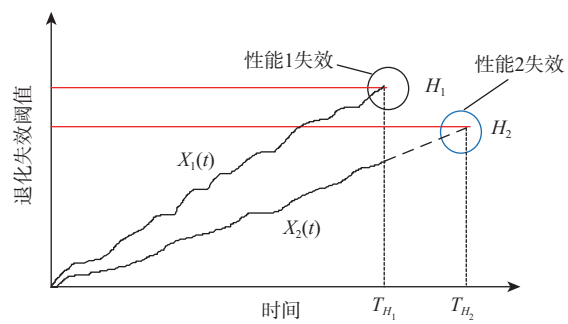


图1 两性能参数退化过程失效机制

Fig. 1 Two performance parameter degradation process failure mechanism

1.1 单性能参数加速退化模型

目前在性能退化分析建模领域，随机过程模型应用最为广泛，其能够灵活地描述退化过程在时间维度的不确定性和产品所处工作环境的影响。本文采用常用的Wiener过程[22]来表征单性能参数的退化轨迹，其表达式为

$$Xk(t)=Xk(0)+\mu_k(z)t+\sigma_k B(t)\ \ (k=1,2) \tag{1}$$

式中，$X_k(t)$表示第k个性能参数在t时刻的退化量；$X_k(0)$为性能退化量的初始值；σ_k为维纳过程的扩散系数；$B(\bullet)$为标准布朗运动；$\mu_k(z)$为维纳过程的漂移参数，表示产品的退化率，是关于应力z的加速模型，加速应力为电应力时一般用逆幂律模型建立

加速方程, 而加速应力为热应力时选用 Arrhenius 模型[23]。

利用产品的实验室加速退化试验数据, 采用极大似然估计方法, 对加速退化模型式 (1) 中的未知参数进行估计, 分别得到产品 2 个性能参数的退化模型。根据布朗运动的独立增量特性, 非重叠区间 Δt上的退化增量ΔX_k呈正态分布, 均值为 $\mu_k\Delta t$, 方差为 $\sigma_k^2\Delta t$。其概率密度函数为

$$f(\Delta Xk)=\frac{1}{\sqrt{2\pi{\sigma_k}^2\Delta t}}\exp\left\{-\frac{[\Delta Xk-\mu_k\Delta t]^2}{2{\sigma_k}^2\Delta t}\right\} \tag{2}$$

假设采用 n 个产品进行 h 个应力水平的加速退化试验, 每个应力水平下有 n_l个试样。试验过程中, 所有试样每隔Δt时间测量一次, 每个应力水平开展 M_l次检测。则第 k 个性能参数在第 l 个应力水平下第 i 个试样的第 j 次测量的退化可以表达为$X_k\left(t_{lij}\right)$。根据式 (2), 第 k 个性能参数的对数似然函数为

$$\begin{aligned}\ln L_k=-\frac{1}{2}\sum_{l=1}^{h}\sum_{i=1}^{n_l}\sum_{j=1}^{M_l-1}\Big\{&\ln(2\pi\sigma_k^2\Delta t)+\\&\frac{[X_k(t_{li(j+1)})-X_k(t_{lij})-\mu_{kl}\Delta t]^2}{\sigma_k^2\Delta t}\Big\}\end{aligned}\qquad k=1,2 \tag{3}$$

随后, 采用遗传算法[24]求解式 (3) 中的待估参数, 开展极大似然估计, 分别得到产品 2 个性能参数的退化模型。

一般来说, 当产品性能参数超过其退化失效阈值后, 即认为产品发生了退化失效, 假设产品的退化失效阈值为 H_k, 则退化过程 $X_k(t)$ 首次穿越 H_k 的时间T_{Hk}对应了产品的寿命, 其失效密度函数和累积分布函数为

$$f_k(t)=\frac{H_k}{\sqrt{2\pi\sigma_k^2t^3}}\exp\left(\frac{-(\mu_k(z)t-H_k)^2}{2\sigma_k^2t}\right)\qquad k=1,2 \tag{4}$$

$$\begin{aligned}F_k(t)=P(T_{Hk}\leqslant t)=&\Phi\left(\frac{\mu_k(z)t-H_k}{\sigma_k\sqrt{t}}\right)+\\&\exp\left(\frac{2\mu_k(z)H_k}{\sigma_k^2}\right)\Phi\left(\frac{-(\mu_k(z)t+H_k)}{\sigma_k\sqrt{t}}\right)\\&k=1,2\end{aligned} \tag{5}$$

式中: $\Phi(\cdot)$代表正态分布的累积失效分布函数。

1.2　相依关系表征

在产品实际使用过程中, 不同性能参数的退化过程之间往往存在相关性。而 Copula 函数作为一种刻画变量间相关性的工具, 应用灵活且被广泛地用于各种复杂产品的寿命评估与预测, 并得到了很好的验证[25], 因此, 本文采用 Copula 函数量化退化过程的相互依赖关系, 其表达式为

$$H(x_1,x_2)=C(F_1(x_1),F_2(x_2))=C(u_1,u_2) \tag{6}$$

式中: $u_k=F_k(x_k)$为边缘分布函数; H 为联合分布函数; C 为 Copula 函数。

常用的二元 Copula 概率密度函数如表 1 所示。其中, θ 为 Copula 函数的参数向量。

表 1　常见的二元 Copula 函数

Table 1　Common binary Copula functions

Copula函数	分布函数$C(u_1,u_2,\theta)$
Clayton	$(u_1^{-\theta}+u_2^{-\theta}-1)^{-\frac{1}{\theta}}$
Frank	$-\frac{1}{\theta}\ln\left(1-\frac{\left(1-\mathrm{e}^{-\theta u_1}\right)\left(1-\mathrm{e}^{-\theta u_2}\right)}{1-\mathrm{e}^{-\theta}}\right)$
Gumbel	$\exp\left\{-\left[(-\ln u_1)^\theta+(-\ln u_2)^\theta\right]^{1/\theta}\right\}$

由于不同的 Copula 函数对退化情况拟合效果不同, 因此采用赤池信息量准则 (Akaike information criterion, AIC) 来评价模型拟合效果的优劣。AIC 信息准则是确定最佳模型的常用准则, AIC 值越小说明所选择的模型越合适[26], 定义为

$$A=2K-2\ln(L) \tag{7}$$

式中: A 赤池信息量准则; K 为参数个数; L 为似然函数。

根据 AIC 信息准则, 选择最佳的 Copula 函数, 根据 Copula 函数的性质, 可以计算得到产品在给定应力条件 z 下的失效概率密度函数为

$$f(t|z)=c(F_1(t),F_2(t))f_1(t)f_2(t) \tag{8}$$

式中, $c(u,v)=\dfrac{\partial C(u,v)}{\partial u\partial v}$。

1.3　外场可靠性模型

产品实际使用过程中, 其所处的工作环境会受到不同任务需求、使用条件等诸多因素的影响, 具有较大的随机性。为此, 本文假设产品外场所受环境应力 z_0 为服从均值为μ_{z_0}, 方差为σ_{z_0}的正态分布随机变量, 其概率密度函数记做$f_Z(z_0)$。基于全概率公式, 可得到动态环境下产品的失效概率密度函数为

$$\begin{aligned}f(t)=&\int_{-\infty}^{+\infty}f\left(t|z_0\right)f_z(z_0)\mathrm{d}z=\\&\int_{-\infty}^{+\infty}f\left(t|z_0\right)\frac{1}{\sqrt{2\pi}\sigma_{z_0}}\exp\left(-\frac{\left(z_0-\mu_{z_0}\right)}{2\sigma_{z_0}^2}\right)\mathrm{d}z\end{aligned} \tag{9}$$

则二元相依退化型产品在动态运行环境下的可靠度为

$$\begin{aligned}R(t)=&\int_0^t f(t)\mathrm{d}t=\\&\int_0^t\int_{-\infty}^{+\infty}f\left(t|z_0\right)\frac{1}{\sqrt{2\pi}\sigma_{Z_0}}\exp\left(-\frac{\left(z_0-\mu_{z_0}\right)}{2\sigma_{z_0}^2}\right)\mathrm{d}z\mathrm{d}t\end{aligned} \tag{10}$$

2　不完美维修条件下的质保期优化

质保期的设计与优化对于增强消费者满意度、

提高制造商利润具有重要意义。一般而言，质保期与产品质保成本成正比，而与产品销量成反比[21]。此外，在工程实践中的大部分维修效果都处于不完美维修状态[17]，即维修虽然可以改善设备的退化状态、延长设备的使用寿命，但难以使设备恢复至全新状态。

因此，本文在不完美维修条件下分别量化质保成本、产品销量与质保期的相关关系，建立制造商利润模型。以制造商利润最大为目标函数，约束制造商利润和相关参数的取值范围，进行质保期优化。

2.1 质保成本模型

质保成本可以用质保期内产品预计失效数的期望和单次维修成本的乘积表示：

$$A(W)=C_{\mathrm{S}}E(M(W)) \tag{11}$$

式中：$A(W)$为单个产品的质保成本；C_{S}为单次维修成本；$E(M(W))$为质保期内预计失效数的期望。

式(11)中单次维修成本和预计失效数均会受到维修效果的影响，因此，本节基于建立的两参数退化型产品可靠性模型，在不完美维修情形下，分别量化产品在给定质保期内的预计失效数和单次维修成本，建立产品质保成本模型。

2.1.1 不完美维修条件下的预计失效数

本文引入维修改善因子来量化产品的不完美维修效果[19]，具体为：考虑维修过程的不确定性，令第m次维修的维修改善因子α_m服从截断正态分布[19]，即$\alpha_m\sim\psi\left(u,v^2,\ 0,1\right)$，其概率密度函数为

$$f_{\alpha_m}(\alpha_m;u,v,0,1)=\frac{\phi\left(\dfrac{\alpha_m-u}{v}\right)}{v\left(\Phi\left(\dfrac{1-u}{v}\right)-\Phi\left(\dfrac{0-u}{v}\right)\right)} \tag{12}$$

式中：u为维修改善因子的均值；v为维修改善因子的标准差；$\phi(\cdot)$为正态分布的概率密度函数；α_m取值范围为0～1。

对于两参数退化型产品而言，其首次失效可能是第1个性能的退化量超过失效阈值造成的退化失效，也可能是第2个性能的退化量超过失效阈值。假设产品首次失效时，2个性能退化量分别为$X_{11}(T_1)$和$X_{21}(T_1)$，T_1为进行第1次不完美维修的时间，退化失效阈值记做H_{11}和H_{21}（$H_{11}=H_1$，$H_{21}=H_2$）。

假设产品第1次故障是由于第1个性能参数超过退化失效阈值造成的，则针对产品故障部件进行第1次不完美维修后，以维修改善因子α_1表征对产品故障部件进行第1次维修后的不完美维修效果。将维修后产品的2个参数性能值记为$X_{12}(T_1)$和$X_{22}(T_1)$，则维修前后的性能退化量有如下关系：

$$\begin{cases}X_{12}(T_1)=(1-\alpha_1)X_{11}(T_1)\\ X_{22}(T_1)=X_{21}(T_1)\end{cases} \tag{13}$$

式中：$0<\alpha_1\leqslant 1$。当$\alpha_1=1$时，维修是完美的，产品退化量降低为0。

对于性能参数服从线性退化轨迹的产品，可以假设维修前后性能退化量的变化为性能阈值的改变。令第1次不完美维修后产品性能初始值为0，退化失效阈值分别为H_{12}和H_{22}的产品，其退化失效阈值满足：

$$\begin{cases}H_{12}=H_1-X_{12}(T_1)=H_1-(1-\alpha_1)X_{11}(T_1)\\ H_{22}=H_2-X_{22}(T_1)=H_2-X_{21}(T_1)\end{cases} \tag{14}$$

同样地，将产品故障部件开展第$m(m\geqslant 1)$次维修的时刻记为T_m，第$m(m\geqslant 1)$次维修的维修改善因子记做α_m，维修后产品的第k个性能参数的退化阈值记做$H_{k,m+1}$，如图2所示，产品在第m次维修后退化失效阈值可表示为

$$\begin{aligned}H_{k,m+1}&=H_k-X_{k,m+1}(T_m)=\\ &H_k-(1-\alpha_m)X_{k,m+1}(T_m)\qquad i=1,2\end{aligned} \tag{15}$$

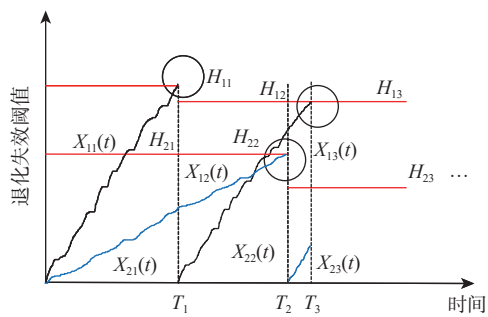


图2 两性能退化型产品不完美维修策略

Fig. 2 Imperfect maintenance strategy for products with two performance parameter degradation

产品进行第1次维修前的可靠度函数可以利用式(10)表示。在此基础上，结合式(13)可以得到产品进行第m次维修，退化阈值变为$H_{i,m+1}$后，在$[T_m,\ T_{m+1})$之间的可靠度函数：

$$R_{m+1}^{*}\left(t^{*}\right)=P\left(X_{1,m+1}\left(t^{*}\right)<H_{1,m+1},X_{2,m+1}\left(t^{*}\right)<H_{2,m+1}\right) \tag{16}$$

式中：$t^{*}=t-T_m$表示第m次维修后产品的工作时间。

产品在规定质保期W内进行m次不完美维修的概率可表示为：在质保期W时，产品共发生了m次失效，并在失效后进行了不完美维修，但并未发生第$(m+1)$次失效的概率：

$$P(M(W)=m)=\int_0^W f_m(t)\cdot R_{m+1}^{*}(W-t)\mathrm{d}t \tag{17}$$

式中：$M(W)$为质保期W内进行不完美维修的次数；

m 为不完美维修次数, m=1,2,…; $f_m(t)$为发生 m 次失效的概率密度函数。

因此, 可给出规定质保期 W 内的预计失效数期望:

$$E(M(W))=\sum_{m=1}^{\infty} m\cdot P(M(W)=m) \tag{18}$$

由于第 m 次维修概率密度函数与第 (m-1) 次不完美维修时间相关, 是一个递推关系, 积分形式为卷积, 导致式 (18) 形式过于复杂, 故采用蒙特卡罗仿真方法计算。仿真步骤流程图如图 3 所示。具体步骤如下:

步骤 1　输入 μ_1、σ_1、μ_2、σ_2、H_1、H_2、μ_{z0}、σ_{z0}等退化相关参数, 以及质保期 W 和仿真次数 N_{mc}, 设置维修次数 m_i、维修时刻 T_m 的初始值为 0, 其中 i=1,2,...,N_{mc} 为蒙特卡罗仿真循环次数, m=0,1,...为第 i 次仿真中的维修次数。

步骤 2　根据式 (13) 采用舍选抽样法[27] 生成 T_m 的数值, 令每次维修间的累积时间 T_i=T_i+T_m。

步骤 3　判断 T_i 是否超过质保期 W: 若否, 继续步骤 4; 若是, 记录维修次数 m_i=m, 转至步骤 5。

步骤 4　根据式 (12) 生成维修改善因子 α_m, 计算在 T_i 时刻性能 1 和性能 2 的产品可靠度 R_1 和 R_2。若 R_1>R_2 则维修产品性能 2: $X_{2,m+1}$=(1-α_m)$X_{2,m}$、$X_{1,m+1}$=$X_{1,m}$; 若 R_1<R_2 则则维修产品性能 1: $X_{1,m+1}$=(1−α_m)$X_{1,m}$、$X_{2,m+1}$= $X_{2,m}$; 若 R_1=R_2 则维修产品性能 1 和性能 2: $X_{1,m+1}$=(1−α_m)$X_{1,m}$、$X_{2,m+1}$=(1−α_m)$X_{2,m}$; 令 m=m+1, 返回步骤 2。

步骤 5　判断循环次数 i 是否等于 N_{mc}: 若是, 根据 $\left(\sum_{i=1}^{N} m_i\right)\Big/N$得到产品在质保期为 W 下的预计失效数算法结束; 若否, i=i+1, 返回步骤 2。

2.1.2　单次维修成本

在得到规定质保期内单个产品的外场预计失效数后, 还需要确定单次维修成本。单次维修成本由固定成本和变动成本构成。其中, 固定成本包括维修过程中检测、人工、备件等每次维修必须支出的费用。变动成本是指维修过程中使用不同的设备、材料造成的波动, 与维修效果相关。维修效果越好, 需要采用的设备就要越精密, 使用的材料就越昂贵, 产生的维修费用就越高, 因此, 可构建变动成本和维修改善因子的函数关系。

根据假设, 维修改善因子$\alpha_m \sim \psi\left(u,v^2,\ 0,1\right)$, 是

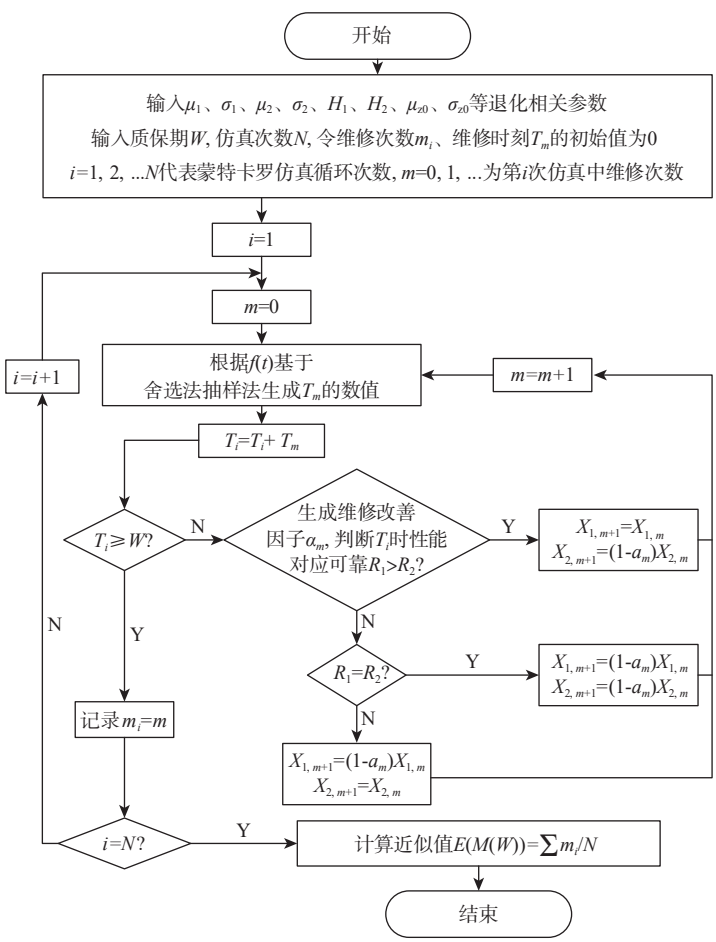


图 3　两性能退化型产品的预计失效数仿真算法流程图

Fig. 3　Flow chart of simulation algorithm for predicted number of failures of products with two performance parameter degradation

一个关于 u 和 v 的随机变量，u 表示维修改善效果均值，u 越大，表示维修效果越好，维修变动成本越高；v 表示维修改善效果标准差，v 越大，表示维修效果波动性越大，会造成维修变动成本变低。综上所述，u 和 v 都会影响变动成本。一般而言，对于较长质保周期内，经历多次维修的产品，其质保期内每次维修对维修成本的累积影响，可以使用维修改善因子的期望来表达。为更好地阐述模型方法，采用维修改善因子的期望来近似计算每次维修的变动成本[19]，即维修改善因子的期望和平均变动成本的乘积。由于维修改善因子服从截断正态分布，其期望为

$$E(\alpha_m;u,v)=u-v\frac{\phi((1-u)/v)-\phi(-u/v)}{\Phi((1-u)/v)-\Phi(-u/v)} \tag{19}$$

因此，可以得到单次维修成本的表达式为

$$C_S=C_F+C_V=C_F+C_AE(\alpha_m;u,v)=C_F+C_A\left[u-v\frac{\phi((1-u)/v)-\phi(-u/v)}{\Phi((1-u)/v)-\Phi(-u/v)}\right] \tag{20}$$

式中：C_S 为单次维修成本；C_F 为固定成本；C_V 为变动成本；C_A 为平均变动成本。

2.2 制造商利润模型

在实践中，当客户无法准确判断产品质量信息时，制造商往往会通过延长质保服务，向消费者传递产品高质量的信号，以吸引更多客户，达到对产品的促销作用[28]。基于此，本文考虑质保期对产品的促销作用，采用在经济管理领域获得广泛认可的 Glickman-Berger 模型[21]，建立销售量和质保期的函数关系：

$$q=k_1p^{-\gamma}(W+k_2)^{\eta} \tag{21}$$

式中：q 为销售量；$k_1>0$ 为幅度因子常数；$k_2\geqslant 0$ 为时间替代常数，即时间为 0 时的需求量；$\gamma>1$ 为价格弹性指数；$0<\eta<1$ 为质保期弹性指数；p 为产品单价；W 为规定质保期。

制造商利润模型可表示为销量和单个产品的利润的乘积，其中单个产品的利润可以用产品单价与制造、销售等成本、质保成本之差表示，具体表述为

$$p_r=p-C_M-A(W) \tag{22}$$

式中：p_r 为单件产品的利润；C_M 为单个产品的制造、销售等其他成本。

则质保期 W 内的制造商利润可以由产品销量和单个产品利润的乘积表示：

$$P_r(W)=qp_r=k_1p^{-\gamma}(W+k_2)^{\eta}\cdot(p-C_M-A(W))=k_1p^{-\gamma}(W+k_2)^{\eta}\cdot(p-C_M-C_SE(M(W))) \tag{23}$$

式中：k_1、k_2、γ、η、p、C_M、C_S 均为固定参数，C_S 为与维修相关的参数，k_1、k_2、γ、η、p、C_M 为与产品相关的参数，预计失效数期望 $E(M(W))$ 和质保期 W 相关。

2.3 质保期优化问题

利润最大化是制造商生产、销售产品的根本目的。因此，本文以制造商利润最大为目标函数，对质保期进行优化。通过式 (23) 可知，制造商利润模型 $P_r(W)$ 可以看做以质保期 W 为自变量的函数，为开口向下的类抛物线模型，存在最优的质保期 W^* 使得制造商利润达到最大值。

优化模型的约束条件从制造商利润约束、质保期约束和维修改善因子约束 3 方面确定。

2.3.1 制造商利润约束

制造商利润是产品销量和单件产品利润的乘积，代表制造商的根本利益，在制定新研产品的质保策略中，制造商一般会根据新研产品的研发等成本确定最低利润，以保证自身权益，因此可以表示为

$$P_r(W)\geqslant P_M \tag{24}$$

式中：$P_r(W)$ 为在规定质保期 W 时的制造商利润；P_M 为制造商制定的最低利润。

2.3.2 质保期约束

质保期是制造商提供给消费者的一种质量保证承诺，不仅受到产品本身可靠性和质量的影响，也受到市场上同类产品质保策略的影响。在市场上竞争中，制造商希望在花费较低成本的基础上为消费者提供质保，与此同时必须保证产品不会带来安全隐患。因此，需要给质保期设定一个合适的取值区间，可以表示为

$$W\in[W_{\min},W_{\max}] \tag{25}$$

式中：$W_{\min}$ 为消费者可接受或者法律规定的最低质保期；$W_{\max}$ 为制造商可接受的最高质保期。

2.3.3 维修改善因子约束

维修改善因子 α 为一个服从 u 为均值，v 为标准差的截断正态的变量，表征维修效果的好坏。在质保期内，产品发生故障后，一般是制造商或委托第三方进行维修，为保证后续产品的使用可靠性，保证维修后的产品下次故障时间控制在消费者可接受的范围内，需要对 u 和 v 这 2 个参数给出一定的约束。维修改善因子是由 u 和 v 共同控制的，基本取值范围为 $[\max(u\text{-}v,0),\min\{u+v,1\}]$[6]，因此，对维修改善因子相关参数的约束可表示为

$$\begin{cases}u\in[0.5,1)\\ v\in[0,0.3]\end{cases} \tag{26}$$

综上所述，得到以“制造商利润最大”为目标的质保期优化设计模型：

$$\begin{cases} \max & P_{\mathrm{r}}(W) \\ \text{s.t.} & \begin{cases} P_{\mathrm{r}}(W) \geqslant P_{\mathrm{M}} \\ W \in [W_{\min}, W_{\max}] \\ u \in [0.5\ ,\ 1) \\ v \in [0\ ,\ 0.3] \end{cases} \end{cases} \tag{27}$$

式中, 需要优化的变量为质保期 W。

3　案例分析

针对本文提出的双参数退化型新研产品质保成本优化方法, 采用文献 [29] 中微波电子组件的开展质保期优化分析。微波组件用于发射/接收信号, 故障机理分析和历史资料表明, 温度是影响其性能退化的主要变量。文献 [29] 对某型微波组件开展了高温恒定应力加速退化试验,并对功率和噪声这 2 个关键性能参数进行了测量。试验中, 正常工作条件为 25 °C, 加速应力水平为 S_1=70 °C、S_2=80 °C 和 S_3=85 °C。每小时测量一次微波组件的功率和噪声, 在加速应力水平 S_1、S_2 和 S_3 下的监测时间分别为 174 h、155 h 和 114 h。功率和噪声的加速退化试验数据如图 4 和图 5 所示。根据文献的分析结果, 2 个性能参数的退化具有明显的相关性, 而本文所提出的 ADT 建模方法可以通过 Copula 函数处理不同的相依情况。此外, 假设该微波组件与其终端产品为单渠道供应链, 即案例对象的订货量与制造商的生产量相同, 且该微波组件的质保期作为终端产品质保期中整体保修的一部分, 可基于 Glickman-Berger 模型量化其对终端产品销量的影响。

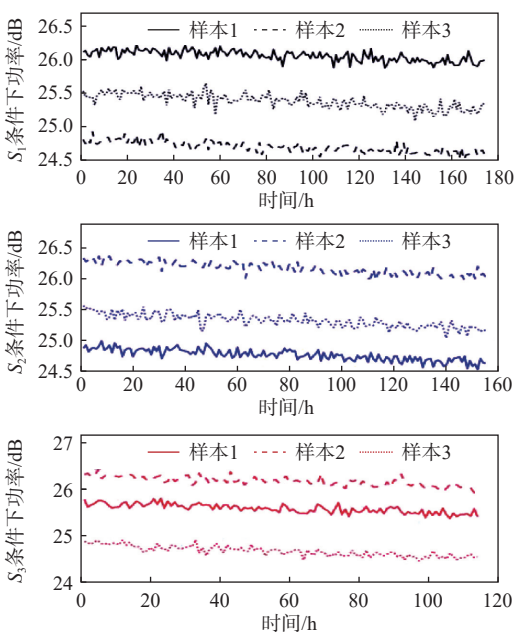


图 4　性能参数功率的 ADT 数据

Fig. 4　ADT data of performance parameter power

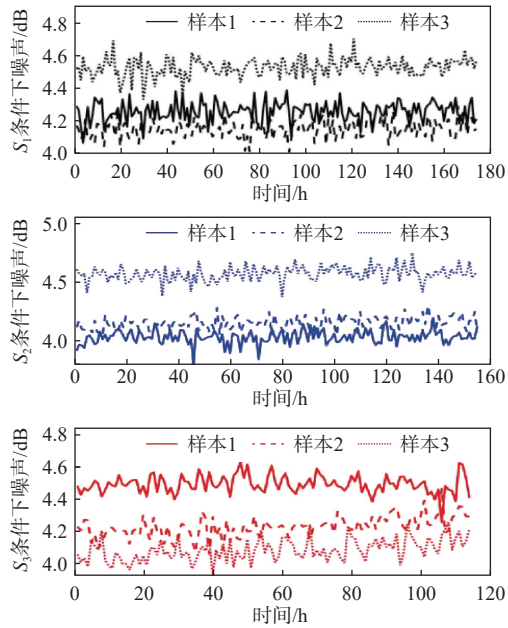


图 5　性能参数噪声的 ADT 数据

Fig. 5　ADT data of performance parameter noise

3.1　质保期优化

根据文献 [29] 中对 ADT 数据特征的分析, 可知该数据具有线性的波动退化特征, 可采用 Wiener 过程对产品单个性能参数的退化进行建模。此外, 由于产品的 ADT 以温度作为敏感应力, 故选择 Arrhenius 模型作为加速模型, 根据式(1)可以得到产品 2 个性能参数的退化轨迹为

$$X_k(t) = X_k(0) + \exp\left(a_k + \frac{b_k}{T_{\mathrm{emp}}}\right) t + \sigma_k B(t) \qquad k = 1,2 \tag{28}$$

式中: a_k 和 b_k 为 Arrhenius 模型中待定参数; T_{emp} 为 ADT 温度应力。

根据图 4 和图 5 所示的退化数据, 采用式 (3) 开展极大似然估计得到退化模型的参数, 如表 2 所示。随后, 采用 Matlab2022 中的 Copula 工具箱来估计 Copula 参数 θ, 并基于式 (7) 开展 AIC 判断, 选用合适的 Copula 函数表示 2 个参数退化过程的相关性, 结果如表 3 所示, 其中 Frank Copula 函数量化性能参数退化相依性效果最好。最后, 考虑产品功率和噪声的相对失效阈值分别为 H_1=12 和 H_2=9, 外场环境应力服从正态分布 $z_0 \sim N(25,2)$, 根据式 (9) 和式 (10) 得到产品失效概率密度函数和可靠度函数, 如图 6 和图 7 所示。

表 2　退化模型参数表

Table 2　Parameters of degradation model

退化参数	a_k	b_k	σ_k
功率	6.873 9	−4 639.611 9	0.087 6
噪声	18.229 8	−9 230.483 6	0.082 4

表3 Copula函数选取

Table 3 Copula function selection

Copula参数	θ	AIC	排序
Gumbel Copula	8.370 1	−36 317	2
Clayton Copula	10.721 2	−23 038	3
Frank Copula	35.701 4	−39 081	1

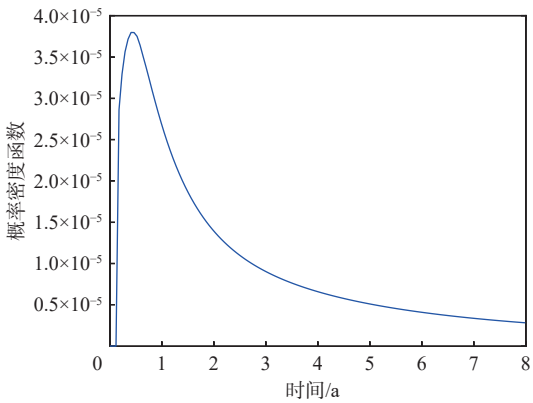


图6 两参数退化型产品失效概率密度函数

Fig. 6 Failure probability density function of products with two-parameter degradation

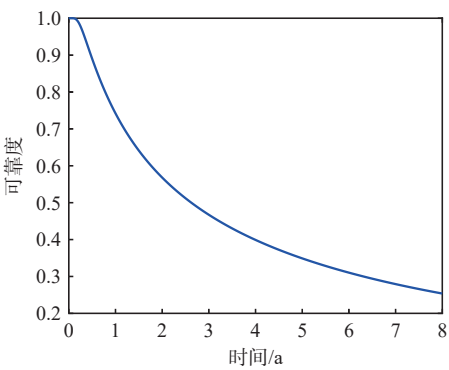


图7 第1次维修前两参数退化型产品可靠度函数

Fig. 7 Reliability function of products with two-parameter degradation before first maintenance

不完美维修条件下，当 u=0.8, v=0.2 时，通过如图4所示的仿真算法得到预计失效数。假设模型各参数值如表4所示，根据式(23)可以得到制造商利润函数，如图8所示。根据制造商利润函数，可以得到产品最优质保期为5.63年，此时制造商利润最大。

此外，为了对比不同情形下产品最优质保期的变化，分别与单独考虑功率退化、单独考虑噪声退化以及两性能独立退化的情况对比，如图8所示。两性能独立退化时产品的可靠度为

$$R_{\text{indep}}(t)=R_1(t)\times R_2(t)=P(X_1(t)<H_1)\times P(X_2(t)<H_2) \tag{29}$$

从图8可以看出，考虑两性能参数相依退化的最优质保期在单独考虑功率退化与单独考虑噪声退化计算出的最优质保期之间。同时，考虑两性能退化相依性的最优质保期明显超出考虑两性能独立退化情形下的最优质保期。由此可见，对于两参数相依退化的产品，量化其性能参数间相关性是十分必要的，这对最优质保期的选择影响十分显著。

表4 制造商利润模型相关参数表

Table 4 Parameters of manufacturer's profit model

参数	数值	来源	参数	值	来源
p	6	假设	k_1	1.2	假设
C_M	2	假设	k_2	2	假设
C_F	0.28	假设	γ	1.2	假设
C_A	0.35	假设	η	0.8	假设

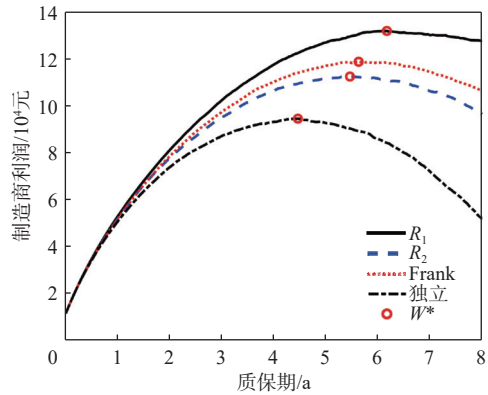


图8 不同情况下的质保期优化

Fig. 8 Warranty period optimization under different conditions

3.2 敏感性分析

为了分析质保期相关参数对制造商利润的影响，本文对退化失效阈值 H_1 和 H_2、维修改善因子均值 u 进行敏感性分析，结果如图9～图12所示。

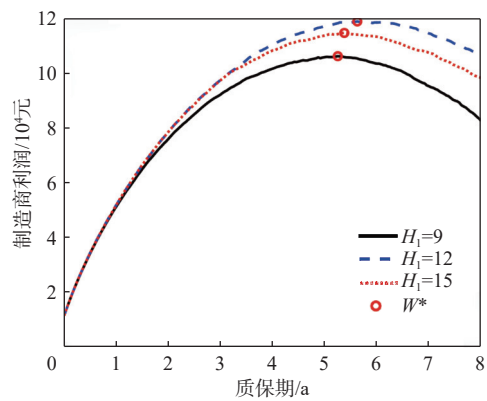


图9 参数 H_1 的敏感性分析

Fig. 9 Sensitivity analysis of parameter H_1

3.2.1 退化失效阈值 H_1 和 H_2 的敏感性分析

由于产品有2个性能参数对应的退化阈值为 H_1 和 H_2，为了分析每个退化失效阈值对产品质保期的影响，分别在 H_1 变化 H_2 固定、H_1 固定 H_2 变化和 H_1、H_2 同时变化3种情况下，分别进行敏感性

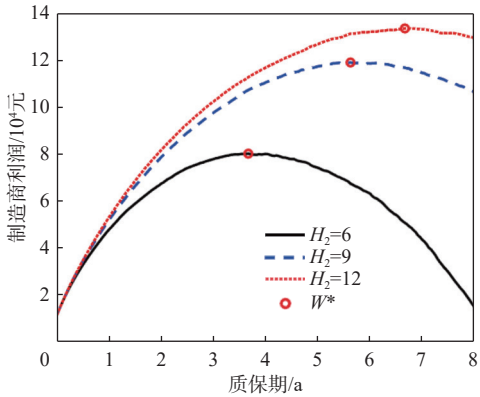


图 10　参数 H_2 的敏感性分析

Fig. 10　Sensitivity analysis of parameter H_2

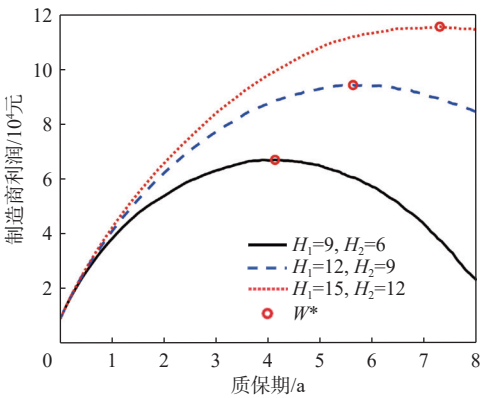


图 11　参数 H_1 与 H_2 的敏感性分析

Fig. 11　Sensitivity analysis of parameters H_1 and H_2

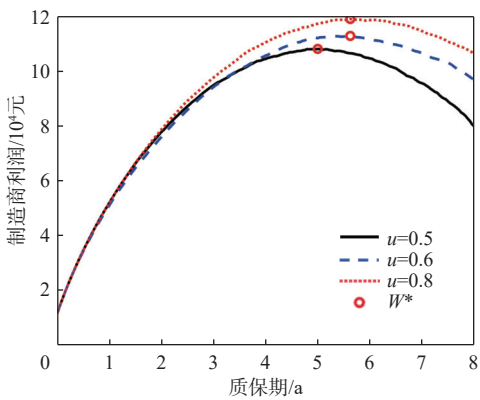


图 12　参数 u 的敏感性分析

Fig. 12　Sensitivity analysis of parameter u

分析。

从图 9～图 11 可以看出，性能退化失效阈值对最优质保期的选择有显著的影响，且退化失效阈值 H_2 对最优质保期的影响显著高于退化失效阈值 H_1 对最优质保期的影响。随着 H_2 的增长（H_2 从 6 增长到 12），最优质保期从 3.67 年增加到 6.67 年。对于该产品而言，应当重点关注第 2 个性能参数 N 的退化过程，提高该性能的失效阈值 H_2 对于提升产品的最优质保期有着显著的作用。

3.2.2　维修改善因子均值 u 的敏感性分析

从图 12 可以看出，维修改善因子均值 u 对最优质保期选择的影响并非十分显著，随着 u 的增长（u 从 0.5 增长到 0.8），最优质保期并不是一直变大，从 4.96 年到 5.63 年后仍旧是 5.63 年。从实际维修的角度，u 越大代表着维修效果越好，降低下次产品故障的概率，但同时单次维修的成本也会有所上升。因此，维修改进因子均值 u 的最优取值要结合产品本身的可靠性与制造商利润模型进行综合量化。

4　结　论

本文针对两参数退化型新研产品，提出了一种基于 Copula 理论与不完美维修因子的产品质保期优化模型，得到了考虑产品相依退化、动态外场环境及不完美维修情景下的产品最优质保期并开展敏感性分析。该方法的提出为缺乏外场故障数据和历史保修索赔记录的两参数退化型新研产品质保期优化提供了科学的分析方法，也可以为其他性能退化过程相互依赖的产品制定质保策略提供依据和参考准则。未来可以从以下 3 个方面进一步研究：

1）考虑外场随机冲击和多退化过程相依性 2 种情形，建立新研产品的质保成本模型，更能贴近产品在制定质保策略时的实际情况。

2）本文仅考虑了维修效果对于维修成本的影响，即为达到制造商预期的维修效果，所需付出的维修成本，以保证产品后续使用的可靠性要求。后续可以从维修成本对维修效果的影响角度进行深入的研究分析。

3）由于本文的案例对象来自于实际的产品，在研究过程中未设计对比试验，后续可以针对该问题继续开展研究。

参考文献（References）

[1] SHANG L J, LIU B L, QIU Q G, et al. Three-dimensional warranty and post-warranty maintenance of products with monitored mission cycles[J]. Reliability Engineering & System Safety, 2023, 239: 109506.

[2] ZHAO X J, GAUDOIN O, DOYEN L, et al. Optimal inspection and replacement policy based on experimental degradation data with covariates[J]. IISE Transactions, 2019, 51(3): 322-336.

[3] KLEYNER A, ELMORE D. Warranty data maturity: effect of observation time on reliability prediction and the warranty management process[J]. Quality and Reliability Engineering International, 2022, 38(5): 2388-2404.

[4] TSENG S T, HSU N J, LIN Y C. Joint modeling of laboratory and field data with application to warranty prediction for highly reliable

products[J]. IIE Transactions, 2016, 48(8): 710-719.

[5] YANG D, HE Z, HE S G. Warranty claims forecasting based on a general imperfect repair model considering usage rate[J]. Reliability Engineering & System Safety, 2016, 145: 147-154.

[6] SHANG L J, SI S B, SUN S D, et al. Optimal warranty design and post-warranty maintenance for products subject to stochastic degradation[J]. IISE Transactions, 2018, 50(10): 913-927.

[7] ZHAO X J, XIE M. Using accelerated life tests data to predict warranty cost under imperfect repair[J]. Computers & Industrial Engineering, 2017, 107: 223-234.

[8] LI T, HE S G, ZHAO X J. Optimal warranty policy design for deteriorating products with random failure threshold[J]. Reliability Engineering & System Safety, 2022, 218: 108142.

[9] SUN F Q, FU F Y, LIAO H T, et al. Analysis of multivariate dependent accelerated degradation data using a random-effect general Wiener process and D-vine Copula[J]. Reliability Engineering & System Safety, 2020, 204: 107168.

[10] SHI Y, XIANG Y S, LIAO Y, et al. Optimal burn-in policies for multiple dependent degradation processes[J]. IISE Transactions, 2021, 53(11): 1281-1293.

[11] WANG F, LI H. On the use of the maximum entropy method for reliability evaluation involving stochastic process modeling[J]. Structural Safety, 2021, 88: 102028.

[12] CHEN R T, ZHANG C, WANG S P, et al. Bivariate-dependent reliability estimation model based on inverse Gaussian processes and copulas fusing multisource information[J]. Aerospace, 2022, 9(7): 392.

[13] FANG G Q, PAN R, HONG Y L. Copula-based reliability analysis of degrading systems with dependent failures[J]. Reliability Engineering & System Safety, 2020, 193: 106618.

[14] CHI B J, WANG Y S, HU J W, et al. Reliability assessment for micro inertial measurement unit based on accelerated degradation data and copula theory[J]. Eksploatacja i Niezawodność – Maintenance and Reliability, 2022, 24(3): 554-563.

[15] JIANG D Y, CHEN T Y, XIE J Z, et al. A mechanical system reliability degradation analysis and remaining life estimation method: with the example of an aircraft hatch lock mechanism[J]. Reliability Engineering & System Safety, 2023, 230: 108922.

[16] FAN P P, YUAN Y P, GAO J X, et al. Reliability modelling and evaluating of wind turbine considering imperfect repair[J]. Scientific Reports, 2023, 13(1): 5323.

[17] ZHAO X J, LIU B, XU J Y, et al. Imperfect maintenance policies for warranted products under stochastic performance degradation[J]. European Journal of Operational Research, 2023, 308(1): 150-165.

[18] CHEN Y X, GONG W J, XU D, et al. Imperfect maintenance policy considering positive and negative effects for deteriorating systems with variation of operating conditions[J]. IEEE Transactions on Automation Science and Engineering, 2018, 15(2): 872-878.

[19] ZHAO X J, HE S G, XIE M. Utilizing experimental degradation data for warranty cost optimization under imperfect repair[J]. Reliability Engineering & System Safety, 2018, 177: 108-119.

[20] WANG R C, CHENG Z H, RONG L Q, et al. Availability optimization of two-dimensional warranty products under imperfect preventive maintenance[J]. IEEE Access, 2021, 9: 8099-8109.

[21] ARORA R, TANDON A, AGGARWAL A G, et al. Genetic algorithm-based price and warranty optimization in software systems[J]. Expert Systems, 2024, 41(7): e13334.

[22] LIU K, ZOU T J, XIN M C, et al. RUL prediction based on two-phase Wiener process[J]. Quality and Reliability Engineering International, 2022, 38(7): 3829-3843.

[23] KUCINSKIS G, BOZORGCHENANI M, FEINAUER M, et al. Arrhenius plots for Li-ion battery ageing as a function of temperature, C-rate, and ageing state–an experimental study[J]. Journal of Power Sources, 2022, 549: 232129.

[24] YI X J, WANG Z Z, LIU S L, et al. An accelerated degradation durability evaluation model for the turbine impeller of a turbine based on a genetic algorithms back-propagation neural network[J]. Applied Sciences, 2022, 12(18): 9302.

[25] ZHENG Y D, ZHANG Y. Reliability analysis for system with dependent components based on survival signature and copula theory[J]. Reliability Engineering & System Safety, 2023, 238: 109402.

[26] KUIPER R M. How to evaluate theory-based hypotheses in meta-analysis using an AIC-type criterion[J]. Entropy, 2022, 24(11): 1525.

[27] 王丙参, 魏艳华, 孙永辉. 利用舍选抽样法生成随机数[J]. 重庆师范大学学报 (自然科学版), 2013, 30(6): 86-91.
WANG B C, WEI Y H, SUN Y H. Generate random number by using acceptance rejection method[J]. Journal of Chongqing Normal University (Natural Science), 2013, 30(6): 86-91 (in Chinese).

[28] ZHANG Z C, XU H Y, ZHAO Y X, et al. Extended warranty service provision: a strategic analysis for the E-commerce platform supply chain[J]. Transportation Research Part E: Logistics and Transportation Review, 2023, 177: 103250.

[29] SUN F Q, WANG N, LI X Y, et al. A time-varying copula-based prognostics method for bivariate accelerated degradation testing[J]. Journal of Intelligent & Fuzzy Systems, 2018, 34(6): 3707-3718.

Design and optimization of warranty period of new products with two-parameter degradation

GAO Shuai[1, 2], LI Yanhong[3], SUN Fuqiang[1, 2, *]

(1. School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China;
2. Science & Technology on Reliability and Environmental Engineering Laboratory, Beijing 100191, China;
3. AECC Commercial Aircraft Engine Co., Ltd, Shanghai 200241, China)

Abstract: Due to the lack of outfield failure data and historical warranty claim records of new products, it is difficult to carry out scientific and reasonable warranty cost prediction and warranty period optimization. Considering the interaction between the degradation processes of different product performance parameters, this paper proposed a method for the design and optimization of the warranty period of new products with two-parameter degradation based on Copula theory. Firstly, a single parameter performance degradation model was established according to the laboratory accelerated degradation test data. Copula theory was used to quantify the correlation between degradation processes. In addition, the outfield reliability model was given by quantifying the dynamic operating environment of the outfield. Secondly, the maintenance improvement factor model was used to quantify the imperfect maintenance situation in the process of maintenance, and the Monte Carlo simulation was employed to calculate the predicted number of product failures. Moreover, the warranty cost model was established. Then, the Glickman-Berger model was used to quantify the impact of the warranty period on product sales, and an optimization model of the warranty period was constructed to maximize the manufacturer's profit. Finally, by taking a certain type of electronic component as an example, the design and optimization of the warranty period of products and sensitivity analysis were carried out to verify the validity and applicability of the model.

Keywords: new product; two-parameter degradation; Copula function; accelerated degradation test; warranty period; sensitivity analysis

Received: 2023-06-05; **Accepted**: 2023-10-18; **Published Online**: 2023-11-01 12:14

URL: link.cnki.net/urlid/11.2625.V.20231101.1119.002

Foundation item: Stable Supporting Fund of Science &Technology on Reliability and Environmental Engineering Laboratory (WDZC20220101)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0343

曲线梁肋机翼结构气动弹性优化设计

周泉知，杨佑绪*，孙录斌，张兴翠，吴逸飞，霍梦文
(南昌航空大学 飞行器工程学院，南昌 330063)

摘 要：传统飞机机翼内部一般采用直线梁肋结构，使用曲线梁肋(SpaRibs)结构能够大大拓展机翼结构设计空间，进一步提升机翼的气动弹性性能。针对链接形状法(LSM)不易自动建模问题，提出使用投影映射方法进行2次空间转换，实现曲线梁肋机翼自动建模。基于遗传算法提出1种曲线梁肋机翼气动弹性综合优化设计方法。使用超声速偶极子格网法计算非定常气动力，采用模态法进行静气动弹性分析，在考虑颤振速度、静气动弹性变形约束的情况下，开展优化设计。某飞翼飞行器综合优化设计算例表明，使用曲线梁肋结构的机翼在重量增幅为1.321%的基础上，颤振速度可提高20.34%；在曲线梁肋构型的基础上进一步进行尺寸参数优化，在满足特定约束条件下，相较于初始构型减重21.76%；综合曲线梁肋构型参数和尺寸参数进行综合(一步)优化，相较于初始构型减重可达26.44%。使用曲线梁肋设计优化和尺寸优化相结合的情况下机翼能够有效地减轻机翼重量，为飞翼式飞行器的结构总体设计提供了一种快速有效的气动弹性综合优化设计方法。

关 键 词：曲线梁肋；气动弹性；飞翼；优化设计；构型设计；初步设计
中图分类号：V221.47
文献标志码：A **文章编号**：1001-5965(2025)06-2148-09

在满足特定约束的前提下，减轻重量始终是飞机结构设计的终极目标。传统飞机机翼一般采为直线梁/肋结构，通过螺栓、铆钉或焊接等方式将梁、肋、蒙皮等典型组件连接起来形成整体，在设计和制造中具有零部件数量较多、制造工艺复杂等局限性。Mulani等[1]于2010年提出曲线梁肋(curvilinear spars and ribs, SpaRibs)新型机翼结构概念，将曲线梁肋机翼当作一个整体部件进行设计，利用摩擦焊、电子束自由成形和增材制造等新的加工方式，在兼顾成本的同时开展曲线梁肋机翼的一体化设计和制造。使用曲线梁肋结构不仅能减少机翼组件数量、节约制造时间和制造成本，而且能增加设计灵活性，拓展机翼结构设计空间。

国外对曲线梁肋机翼相关技术进行了持续研究。Locatelli等[2]提出使用B样条曲线表达曲线梁肋结构的3种参数化方法，分别是边界盒基曲线法(bounding box and base curves method, BBCM)、链接形状法(linked shape method, LSM)和独立形状法(independent shape method, ISM)。使用边界盒基曲线法在考虑屈曲和应力的情况下开发了EBF3SSWingOpt两级优化框架。使用链接形状法对EBF3SSWingOpt进行改进，开发了全局-局部优化框架[3-6]。

Liu等[7-8]进一步探究曲线梁肋这一概念的潜力，考虑静气动弹性、颤振和屈曲等约束，基于Python语言和粒子群优化方法在全局-局部优化框架下对美国航空航天局(National Aeronautics and Space Administration, NASA)的亚音速机翼模型(common research model, CRM)进行优化。结果显示，与基准模型相比减重40%，并且所用计算时间

收稿日期：2023-06-12；**录用日期**：2023-10-27；**网络出版时间**：2023-11-29 10:21
网络出版地址：link.cnki.net/urlid/11.2625.V.20231128.1320.002
基金项目：江西省“双千计划”人才项目(CK202006470)
*通信作者. E-mail：zgdy_1@163.com

*引用格式：周泉知，杨佑绪，孙录斌，等. 曲线梁肋机翼结构气动弹性优化设计[J]. 北京航空航天大学学报，2025，51(6)：2148-2156.
ZHOU Q Z，YANG Y X，SUN L B，et al. Aeroelastic optimization design of SpaRibs wing structure[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2148-2156 (in Chinese).*

减少了 89%。

美国 M4 工程公司的 Robinson 等[9] 提出了集合曲线梁肋结构参数化定义、建模、分析和优化的设计方法, 研究了曲线梁肋机翼所带来的气动弹性收益。通过对 NASA ATw4 模型的直线梁肋和曲线梁肋版本的对比发现, 在重量相当的情况下, 曲线梁肋 ATw4b 机翼的颤振速度高 6%, 同时最大应力降低 4.5%。

Zhao 等[10-12] 提出一种用四节点非均匀有理 B 样条 (NURBS) 曲线参数化曲线梁肋的广义数学模型。将曲线梁肋和复合材料铺层结构综合考虑进行优化设计, 通过优化多控制面偏转角组合实现翼根载荷最小, 与基准模型相比, 重量减小 21.8%, 与直梁/肋机翼相比, 重量下降 7.3%。

De 等[13] 在 EBF3WingOpt 优化框架的基础上提出了一种新的参数化方法, 在创建曲线梁肋时可以从翼盒的任何边缘开始并在任何其他边缘结束, 大大拓展了曲线梁肋的设计空间。

国内对于曲线梁肋相关研究较少, 徐元铭和王东[14]、张坤鹏等[15] 对曲线加筋壁板相关技术开展研究, 尚未有曲线梁肋机翼相关的文献发表。

综上所述, 相较于直线梁肋, 曲线梁肋具有更大的设计空间, 能有效提升机翼气动弹性性能, 在机翼减重方面具有良好的表现, 因而探究曲线梁肋概念的潜力和发展相关技术具有必要性。本文使用曲线梁肋设计飞行器机翼, 研究参数化方法和优化设计方法, 进一步探究曲线梁肋在机翼设计中的潜力, 为飞机新结构形式提供理论支持和预先研究。

1　理论基础

1.1　参数化方法

参数化方法是进行曲线梁肋机翼内部结构设计的基础。基于 LSM 参数化方法[5], 使用三阶 B 样条曲线在单位空间内参数化表达曲线梁肋。三阶 B 样条曲线是二次分段连续多项式函数, 根据文献 [16] 可以用矩阵形式表示:

$$P(t)=\frac{1}{2}\begin{bmatrix} t^2 & t & 1 \end{bmatrix}\boldsymbol{N}\begin{bmatrix} P_0 \\ P_1 \\ P_2 \end{bmatrix} \tag{1}$$

式中: $P(t)$ 为曲线的参数方程; t 为无量纲参数, 取值范围: $t\in(0,1)$; P_0、P_1 和 P_2 为控制 B 样条曲线的 3 个点; 矩阵 $\boldsymbol{N}$ 为混合函数系数矩阵, 其具体形式为

$$\boldsymbol{N}=\begin{bmatrix} 1 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 1 & 0 \end{bmatrix} \tag{2}$$

将矩阵 $\boldsymbol{N}$ 代入式 (1), 将式 (1) 在二维空间内展开并整理后得

$$P(t)=\{x(t),y(t)\}=\begin{cases} \left(t^2-2t+1\right)x_A+\left(-2t^2+2t\right)x_{P_1}+\left(t^2\right)x_B \\ \left(t^2-2t+1\right)y_A+\left(-2t^2+2t\right)y_{P_1}+\left(t^2\right)y_B \end{cases} \tag{3}$$

式 (3) 是定义曲线梁肋形状的 B 样条的数学形式, B 样条曲线如图 1 所示。

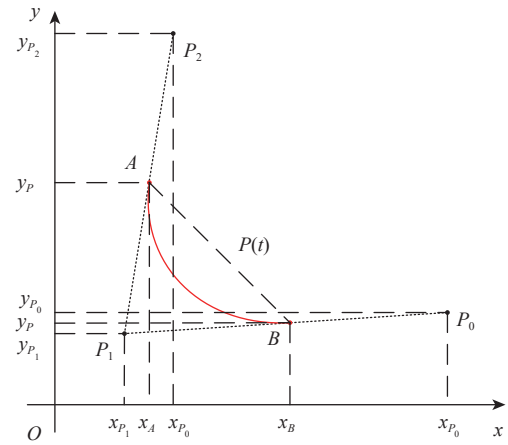


图 1　B 样条曲线

Fig. 1　B-spline curve

图 1 中, 点 A 为线段 P_1P_2 的中点, 点 B 为线段 P_1P_0 的中点, A、B 和 P_1 三点构成一条 B 样条曲线, 点 A、B 为该曲线的起始点和终点, 点 P_1 为该曲线的控制点。

在 LSM 参数化方法中, 单位空间内多条 B 样条曲线组合形成一组曲线梁肋, 单位空间内曲线梁肋布置方式如图 2 所示。

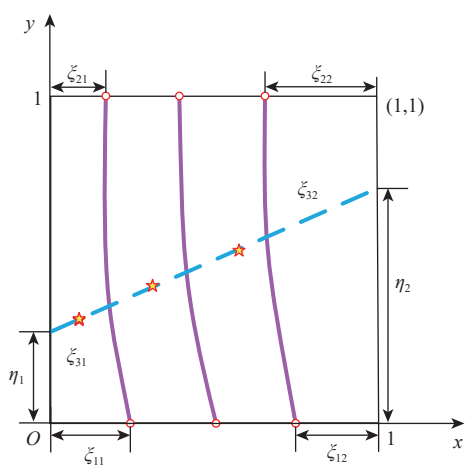


图 2　单位空间内曲线梁肋布置方式

Fig. 2　Layout of SpaRibs in unit space

曲线的起始点和终点分别在不同的边上, 所有控制点位于控制点连线上, 包括端点在内所有相邻点的间距按照等比数列的方式排列。LSM 参数化方法以 6 个参数定义一组曲线梁肋, 表 1 为参数的

表 1 定义一组曲线梁肋与单位空间关系的参数

Table 1 Parameters defining relationship between a set of SpaRibs to unit space

参数	定义
p_1	曲线梁肋的数量
p_2	η_1
p_3	η_2
p_4	ξ_{12}/ξ_{11}
p_5	ξ_{22}/ξ_{21}
p_6	ξ_{32}/ξ_{31}

定义。

p_1 为一组曲线梁肋的数量，p_2 和 p_3 定义了控制点连线的位置，p_4、p_5 和 p_6 定义了起始点、控制点和终点的间距。

1.2 转换矩阵

LSM 参数化方法提出的建模方式是从单位矩形向物理空间直接转换。这种直接从单位空间生成几何模型的方法难度较大，在实际转换过程中复杂程度较高，不易实现。本文提出一种两次空间转换方法，两次转换方法包括单位矩形向机翼平面四边形的转换和机翼平面四边形向三维空间的转换。

单位矩形向机翼平面四边形的转换方法如图 3 所示。

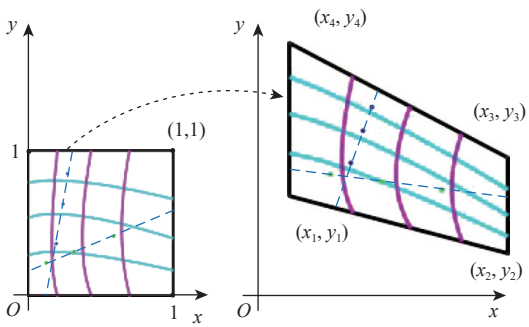


图 3 单位空间向机翼平面空间转换

Fig. 3 Transformation from unit space to wing plane space

将实际机翼平面形状简化成一个四边形，单位空间内的曲线梁肋布置通过透视变换 (perspective transformation) 方法转换到实际机翼四边形内。透视变换[17]方法包含的转换方式如图 4 所示，包括平移、旋转、斜切、缩放和透视。

透视变换公式为

$$[\,x' \quad y' \quad w'\,] = [\,u \quad v \quad w\,] \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{4}$$

式中：(u, v) 为单位矩形内的坐标；(x, y) 为机翼平面四边形内点的坐标，其中：$x = x'/w'$，$y = y'/w'$。

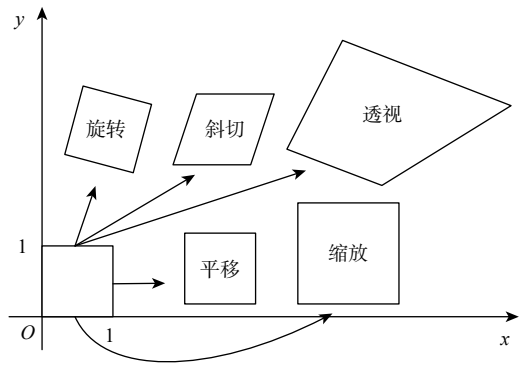


图 4 透视变换

Fig. 4 Perspective transformation

$$\boldsymbol{T} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}_1 & \boldsymbol{T}_2 \\ \boldsymbol{T}_3 & \boldsymbol{T}_4 \end{bmatrix} \tag{5}$$

其中，

$$\boldsymbol{T}_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$\boldsymbol{T}_2 = \begin{bmatrix} a_{13} & a_{23} \end{bmatrix}^{\mathrm{T}}$$

$$\boldsymbol{T}_3 = \begin{bmatrix} a_{31} & a_{31} \end{bmatrix}$$

$$\boldsymbol{T}_4 = [a_{33}]$$

式中，$\boldsymbol{T}$为透视变换矩阵，可拆分为 4 部分。其中，$\boldsymbol{T}_1$表示线性变换，比如缩放、斜切和旋转；$\boldsymbol{T}_3$表示平移；$\boldsymbol{T}_2$表示透视变换；$\boldsymbol{T}_4$=1。将式 (4) 展开后得

$$x = \frac{x'}{w'} = \frac{a_{11}u + a_{21}v + a_{31}}{a_{13}u + a_{23}v + a_{33}} \tag{6}$$

$$y = \frac{y'}{w'} = \frac{a_{12}u + a_{22}v + a_{32}}{a_{13}u + a_{23}v + a_{33}} \tag{7}$$

透视变换矩阵$\boldsymbol{T}$中存在 8 个未知量，在机翼平面四边形内选择 4 个已知点 (一般是角点) 和单位矩形上的 4 个点形成对应关系代入式 (6) 和式 (7)，组合形成方程组，求出方程组的解即得变换矩阵，完成由单位矩形向机翼平面四边形的变换。

机翼平面空间向三维空间的转换如图 5 所示，主要方法是将机翼平面内的点和线投影至机翼上下蒙皮，由上下蒙皮上的相关点线连接生成曲线梁肋面，其中机翼上下蒙皮由机翼的翼型和翼展弦长

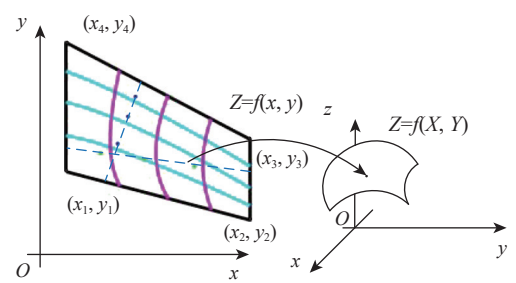


图 5 机翼平面空间向三维空间的转换

Fig. 5 Transformation from wing plane space to three-dimensional space

等参数构成。

1.3　模态分析

频率和振型通过求解特征值问题得到, 模态分析方程一般表示为

$$(\boldsymbol{K}-\lambda\boldsymbol{M})\boldsymbol{x}=\boldsymbol{0} \tag{8}$$

式中: $\boldsymbol{K}$为结构刚度矩阵; $\boldsymbol{M}$为结构质量矩阵; λ为特征值, $\lambda=\omega^2$, ω为圆频率; $\boldsymbol{x}$为特征向量。

1.4　静气动弹性分析

静气动弹性方程一般表示为

$$\left[\boldsymbol{K}-\frac{1}{2}\rho V^2\boldsymbol{Q}_u\right]\bar{u}+\boldsymbol{M}\ddot{\boldsymbol{u}}_R=\boldsymbol{P}_0 \tag{9}$$

式中: $\boldsymbol{K}$为结构刚度矩阵; ρ为飞行高度对应的大气密度; V为来流速度; $\boldsymbol{Q}_u$为结构弹性变形引起的气动力影响系数矩阵; $\boldsymbol{M}$为结构质量矩阵; $\ddot{\boldsymbol{u}}_R$为结构刚体自由度的加速度; $\boldsymbol{P}_0$为外部载荷。

1.5　颤振分析

典型颤振分析方程为

$$\bar{\boldsymbol{M}}\ddot{\boldsymbol{q}}+\bar{\boldsymbol{K}}\boldsymbol{q}=\frac{1}{2}\rho V^2\boldsymbol{A}\boldsymbol{q} \tag{10}$$

式中: $\bar{\boldsymbol{M}}$为模态坐标下的广义质量矩阵; $\bar{\boldsymbol{K}}$为广义刚度矩阵; $\boldsymbol{q}$为广义坐标列阵; $\boldsymbol{A}$为广义气动力影响系数矩阵。广义非定常气动力矩阵$\boldsymbol{A}$为关于马赫数与减缩频率$k=\dfrac{\omega b}{V}$的复函数, b为参考半弦长。

利用 p-k 法求解颤振方程, 令

$$\boldsymbol{q}=\boldsymbol{q}_0\mathrm{e}^{st} \tag{11}$$

式中: $s=(\gamma+i)\omega$; γ为衰减率。将式 (11) 代入式 (10) 中得到如下广义特征值问题:

$$\left[s^2\bar{\boldsymbol{M}}+\bar{\boldsymbol{K}}-\frac{1}{2}\rho V^2\boldsymbol{A}\right]\boldsymbol{q}_0=0 \tag{12}$$

在不同速度下求解该广义特征值问题可得到颤振速度 ($\gamma=0$时对应的速度) 和颤振频率。

2　优化方法

遗传算法是 1962 年美国 Michigan 大学 Holland 教授提出的模拟自然界遗传机制和生物进化论而成的一种并行随机搜索最优化方法[18]。遗传算法具有全局搜索能力强、通用性好的特点。遗传算法具体流程如图 6 所示。

采用遗传算法进行寻优, 以质量最小为目标, 以颤振速度和气动弹性变形为约束, 包括翼尖变形和扭角, 优化曲线梁肋机翼的构型参数和尺寸参数, 包括曲线梁肋位置分布、曲线梁肋腹板和蒙皮的厚度、凸缘的横截面积等。采用罚函数法[19]将优化问题转化为无约束优化问题:

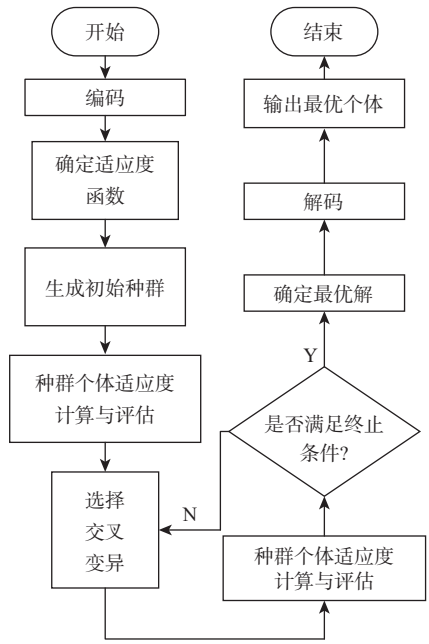


图 6　遗传算法流程

Fig. 6　Process of genetic algorithm

$$\begin{cases}\min\ \phi f(x,\ r)=f(\boldsymbol{x})+\\ \quad\displaystyle\sum_{i=1}^{n_{\mathrm{e}}}r_i^{\mathrm{e}}h_i^2(\boldsymbol{x})+\sum_{j=1}^{n_{\mathrm{g}}}r_i^{\mathrm{g}}[\max(\mathrm{g}_j(\boldsymbol{x})),0]^2\\ \quad\text{s. t. } x_i^{\mathrm{lower}}\leqslant x_i\leqslant x_i^{\mathrm{upper}}\quad i=1,2,\cdots,n_{\mathrm{d}}\end{cases} \tag{13}$$

式中: r_i^{e}和r_i^{g}分别为等式和不等式约束的罚值; x_i^{lower}和x_i^{upper}分别为第 i 个设计变量的上下界; n_{d}为设计变量的数量。

在优化中, 外部罚函数法的惩罚量是和约束的背离程度成正比的, 并且随着优化的进行, 优化变量在不可行区域不断接近最优解。

在遗传算法中, 罚函数法用于在每一代中维持一定数量的不可行解, 从而使遗传搜索从可行区域和不可行区域 2 个方向搜索最优解[20]。

遗传算法采用实数编码, 以式 (13) 的目标函数直接作为适应度函数。在优化过程中, 当曲线梁肋的位置发生变化时, 机翼曲线梁肋结构会发生变化, 必然影响计算结构质量和结构动力学特性的数据输入。这种情况具体体现在优化过程中每生成一次个体, 曲线梁肋结构模型都要更新变化, 需要程序不断修改输入文件。同样的, 在尺寸参数优化过程中需要程序不断修改变化的参数。以上参数修改过程都通过程序自动进行, 使得优化程序能够顺利进行。

3　优化算例

3.1　结构与气动模型

美国当地时间 2022 年 12 月 2 日, 由诺斯罗普·格鲁曼公司研制的美国新一代 B-21 “突袭者” 战

略轰炸在加利福尼亚州亮相，与现役B-2隐形轰炸机相似，采用飞翼布局[21-23]。使用曲线梁肋对飞翼布局飞行器机翼进行设计，曲线梁肋在结构上不同于直线梁肋，其典型特征是不严格区分梁肋，并且梁肋甚至是蒙皮的加筋在翼盒内以曲线形式存在，在提供更有效的承载机制的同时，其梁肋曲率的存在便于耦合机翼的弯曲和扭转变形。

在建立有限元模型时，上下蒙皮、曲线梁肋采用板单元，前后梁缘条、上下桁条采用杆单元，根据文献[24-26]构造基准模型，飞翼初始构型如图7所示，其内外翼有明显的转折，气动网格如图8所示。

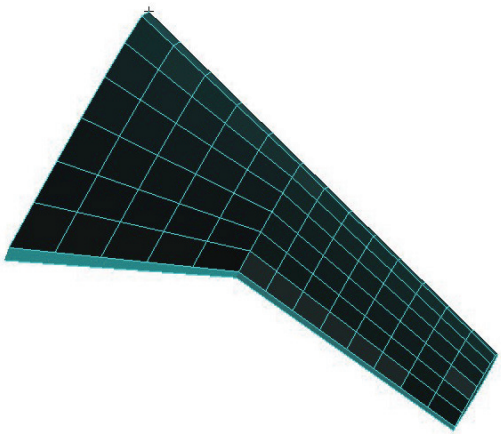


图7 飞翼有限元模型

Fig. 7 Finite element model of flying wing

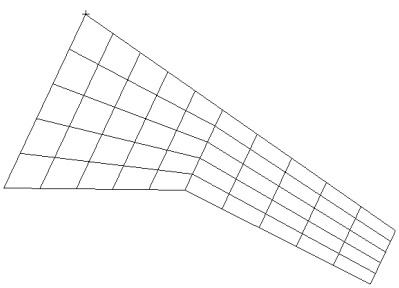


图8 飞翼气动网格

Fig. 8 Aerodynamic mesh of flying wing

按照设计变量组合的复杂性，将飞翼结构优化设计问题分为

1）机翼曲线梁肋形状优化设计。

2）基于最优形状构型的尺寸优化设计。

3）机翼曲线梁肋形状和尺寸综合优化。

为减轻计算复杂程度，本模型为简化模型，仅考虑曲线梁肋、蒙皮和凸缘等基本结构。模型采用NACA0012对称翼形，采用多梁式结构，一共有7个梁，其中内外翼段分别布6个翼肋和9个翼肋。蒙皮、梁肋腹板的厚度均设定为1.5×10^{-3} m，凸缘截面积为7.9×10^{-6} m^2。材料属性：质量密度=2 700 kg/m^3，弹性模量=70 GPa，泊松比=0.3，初始构型的总质量为960 kg。其他数据如表2所示。

表2 飞翼平面形状几何参数

Table 2 Geometric parameters of flying wing plane shape

定义	参数
半翼展/m	4.57
翼根弦长/m	5.53
转折处弦长/m	2.55
翼尖弦长/m	1.7
转折处翼展/m	2.02
前缘后掠角/(°)	35

在优化中主要考虑机翼的颤振和静气动弹性特性，设定工况：动压为2.59×10^5 Pa，迎角为3°。基准模型的颤振速度大约为815.24 m/s，在给定飞行状态下翼尖静变形大约为翼展的3.512%，翼尖扭角大约为−1.273 4°(低头)，单元最大应力约为216.6 MPa。

在以下的优化算例中，遗传参数设置为：交叉概率为0.4，变异概率为0.2，群体大小和进化代数根据不同的优化算例有所不同，终止条件为运行到最大代数。

3.2 算例1 曲线梁肋形状优化

结合表1，针对基准模型内外翼段，分成4组曲线梁肋，机翼的平面形状和曲线梁肋数量不变，仅改变机翼内部曲线梁肋形状，即每一组曲线梁肋中的p_1为常数，参考文献[27−28]，其余每个变量的变化范围如表3所示。

表3 曲线梁肋形状参数变化范围

Table 3 Boundary of shape parameters of SpaRibs

数值类型	p_1	p_2	p_3	p_4	p_5	p_6
最小值	定值	0.05	0.05	0.25	0.25	0.25
最大值	定值	0.95	0.95	4	4	4

注：p_1表示曲线梁肋的数量，是预先设定的定值，没有变化范围，具体值根据翼盒中梁肋的数量来确定。

此次优化中种群大小为20，进化代数为40，计算动压为2.59×10^5 Pa，迎角为3°，翼梢静变形小于翼梢弦长的5%、翼梢扭转角小于1.5°和单元应力小于材料许用应力为约束。满足约束条件的情况下适应度等于颤振速度，在不满足约束的情况下还包括背离约束的加权叠加，在优化迭代结束后，得到最优个体，最优的曲线梁肋构型如图9所示。

优化后重量为972.685 6 kg，颤振速度为981.160 4 m/s，翼梢变形为半展长的3.471%，翼梢扭转角为−1.294 4°，单元最大应力为205.987 MPa。最优构型相较于初始构型重量增加12.685 6 kg，在重量增幅为1.321%的基础上，颤振速度的增幅为20.34%。很明显，仅仅将直线梁肋更换为曲线梁肋，必然导致结构重量的增加。通过此优化，在尽

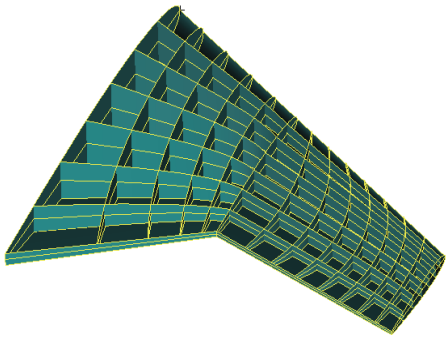


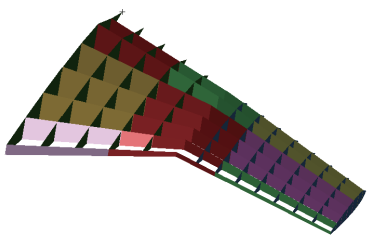
图 9　最优的曲线梁肋构型图

Fig. 9　Optimal SpaRibs configuration

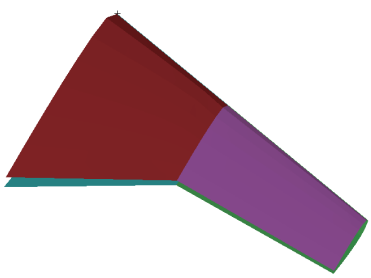
可能增加较少的结构重量或者重量相当的前提下, 曲线梁肋相较于直线梁肋明显提升了飞翼的颤振速度, 与文献 [29] 的结论一致。

3.3　算例 2 曲线梁肋尺寸优化

基于算例 1 最优构型, 进一步优化梁肋腹板厚度和凸缘截面积。将机翼的蒙皮、梁肋和凸缘等分成 28 个区域, 每个区域采用不同的尺寸参数, 具体分块区域如图 10 所示, 将曲线梁肋区域分成 12 块, 蒙皮区域分成 4 块, 凸缘区域分成 12 块, 每个部分参数的取值范围如表 4 所示。



(a) 梁肋区域划分



(b) 蒙皮区域划分

图 10　机翼结构部分划分

Fig. 10　Division of aircraft wing structure

种群大小为 20, 进化代数为 40, 计算动压为 2.59×10^5 Pa, 迎角为 3°, 以飞翼重量最小为目标, 颤振速度大于 800 m/s、翼梢静变形小于翼梢弦长的 5% 、翼梢扭转角小于 1.5°和单元应力小于材料许用应力为约束开展优化设计。

个体适应度进化历程如图 11 所示, 优化后其最小重量为 751.141 3 kg, 颤振速度为 801.266 45 m/s, 翼梢变形为半展长的 4.290%, 翼梢扭转角为−1.446 8°, 单元最大应力为 267.626 MPa。重量相较于初始构型减少 208.858 7 kg, 减重了 21.76%, 减重效果明显。

表 4　尺寸参数参考范围

Table 4　Boundary of size parameters　　mm

类型	最小值	最大值
曲线梁肋	0.5	3
蒙皮	0. 5	2.5
凸缘	0.5	3

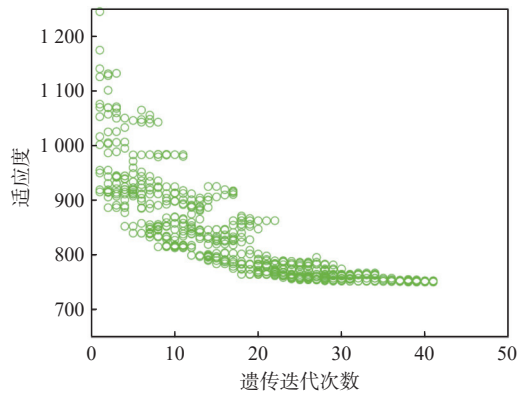


图 11　个体适应度进化历程

Fig. 11　Fitness evaluation history of individuals

3.4　算例 3 曲线梁肋形状尺寸综合优化

算例 1 和算例 2 将飞翼的形状和尺寸参数分成两步独立进行, 在最优形状构型的基础上开展尺寸优化, 相当于两步优化。

开展曲线梁肋形状和尺寸综合(一步)优化。曲线梁肋形状参数数量仍然选取算例 1 的参数; 尺寸参数仍按照算例 2 的分区方式, 同样采用表 4 的尺寸参数变化范围。分别对种群数量和进化代数进行分类组合。挑选出 4 组组合方式, 这 4 种方式的计算成本均比算例 1 和算例 2 要高很多, 并且种群数量和进化代数越多计算成本越高。同时, 进行曲线梁肋构型和尺寸参数优化相对于初始构型能得到良好的曲线梁肋构型和尺寸参数组合。为了避免陷入局部最优解, 在整个优化过程中需要进行多次优化取最优结果。表 5 为 4 次综合优化的最佳结果, 图 12 为优化 D 的最优曲线梁肋构型。

随着进化代数的增加, 减重效果更加明显, 优化 D 显示当遗传代数增加到 70 时, 能够得到比遗传代数少的优化更优的结果, 即能在满足气动弹性要求的前提下, 可以得到更轻的结构重量。但付出的计算成本约是优化 C 的 2 倍, 约是优化 B 的 3 倍, 约是优化 A 的 5 倍。

最优结果的颤振速度为 800.49 m/s, 翼梢变形为半展长的 4.03%, 翼梢扭转角为−1.46°, 最终的结构重量为 706.17 kg, 相对于初始构型减重达 26.44%, 相较于算例 2 减重 5.99%。

表5 形状尺寸综合优化最优个体

Table 5 Optimal individual of comprehensive shape and size optimization

优化结果	最大遗传代数	种群大小	颤振速度/(m·s^{-1})	翼尖变形/%	翼尖扭角/(°)	质量/kg	最大应力/MPa
A	20	20	816.09	4.24	−1.47	741.90	270.60
B	40	20	800.71	4.14	−1.41	780.08	261.38
C	40	30	802.27	4.07	−1.44	715.47	238.10
D	70	30	800.49	4.03	−1.46	706.17	280.63

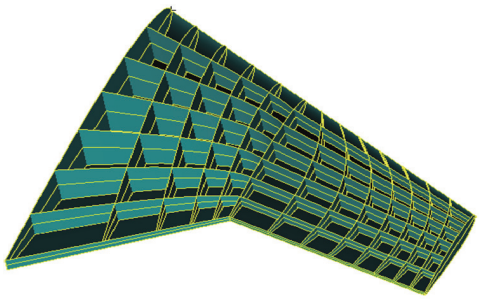


图12 优化D的最优曲线梁肋构型

Fig. 12 Optimal SpaRibs configuration of optimization D

4 结 论

本文发展了一种使用曲线梁肋结构构型综合设计方法，以飞翼飞行器为例，在早期设计阶段考虑气动弹性约束，综合气动结构的耦合影响，减小结构重量。设计空间同时包括曲线梁肋构型参数和尺寸参数，结果显示：

1）在梁肋数量不变的前提下，使用曲线梁肋结构的机翼在重量增幅为1.321%的基础上，颤振速度可提高20.34%；

2）在曲线梁肋构型的基础上进一步进行尺寸参数优化，在满足特定约束条件下，相较于初始构型减重21.76%；

3）进行曲线梁肋构型参数和尺寸参数综合（一步）优化，相较于初始构型减重可达26.44%。

使用曲线梁肋形状优化和尺寸优化相结合的情况下机翼能够有效地减轻机翼重量，为曲线梁肋机翼的结构设计提供了一种快速有效的气动弹性综合优化设计方法。

参考文献（References）

[1] MULANI S, LOCATELLI D, KAPANIA R. Algorithm development for optimization of arbitrary geometry panels using curvilinear stiffeners[C]// Processings of the th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston: AIAA, 2010.

[2] LOCATELLI D, MULANI S B, KAPANIA R K. Wing-Box weight optimization using curvilinear spars and ribs (SpaRibs)[J]. Journal of Aircraft, 2011, 48(5): 1671-1684.

[3] LOCATELLI D, MULANI S, KAPANIA R, et al. A multidisciplinary analysis optimization (MDAO) environment for wings having SpaRibs[C]//Processings of the 53th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference. Reston: AIAA, 2012.

[4] LOCATELLI D. Optimization of supersonic aircraft wing-box using curvilinear SpaRibs[D]. Blacksburg: Virginia Polytechnic Institute and State University, 2012: 22-75.

[5] LOCATELLI D, MULANI S B, KAPANIA R K. Parameterization of curvilinear spars and ribs for optimum wing structural design[J]. Journal of Aircraft, 2014, 51(2): 532-546.

[6] LOCATELLI D, YEILAGHI TAMIJANI A, MULANI S B, et al. Multidisciplinary optimization of supersonic wing structures using curvilinear spars and ribs (SpaRibs)[C]// Processings of the 54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston: AIAA, 2013.

[7] LIU Q, MULANI S B, KAPANIA R K. Global/local multidisciplinary design optimization of subsonic wing[C]// Processings of the 10th AIAA Multidisciplinary Design Optimization Conference. Reston: AIAA, 2014.

[8] LIU Q, JRAD M, MULANI S B, et al. Integrated global wing and local panel optimization of aircraft wing[C]// Processings of the 56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston: AIAA, 2015.

[9] ROBINSON J H, DOYLE S, OGAWA G, et al. Aeroservoelastic optimization of wing structure using curvilinear spars and ribs (SpaRibs) [C]// Processings of the 17th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference. Reston: AIAA, 2016.

[10] ZHAO W, KAPANIA R K. Multiobjective optimization of composite flying-wings with SpaRibs and multiple control surfaces[C]// Processings of the 2018 Multidisciplinary Analysis and Optimization Conference. Reston: AIAA, 2018.

[11] ZHAO W. Optimal design and analysis of Bio-Inspired, curvilinearly stiffened composite flexible wings[D]. Blacksburg: Virginia Polytechnic Institute and State University, 2017.

[12] ZHAO W, KAPANIA R K. Bilevel programming weight minimization of composite flying-wing aircraft with curvilinear spars and ribs[J]. AIAA Journal, 2019, 57(6): 2594-2608.

[13] DE S, JRAD M, KAPANIA R K. Structural optimization of internal structure of aircraft wings with curvilinear spars and ribs[J]. Journal of Aircraft, 2019, 56(2): 707-718.

[14] 徐元铭, 王东. 曲线片型加筋壁板的稳定性优化设计[J]. 北京航空航天大学学报, 2015, 41(4): 567-573.
XU Y M, WANG D. Design and optimization of improving stability of curvilinear blade-stiffened panels[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(4): 567-573 (in Chinese).

[15] 张坤鹏, 郝鹏, 段于辉, 等. 基于深度学习的多级曲线加筋壁板布

局优化设计[J]. 中国舰船研究, 2021, 16(4): 86-95.
ZHANG K P, HAO P, DUAN Y H, et al. Layout optimization design of hierarchical curvilinearly stiffened panels based on deep learning[J]. Chinese Journal of Ship Research, 2021, 16(4): 86-95 (in Chinese).

[16] SALOMON D. Curves and surfaces for computer graphics[M]. New York: Springer, 2006.

[17] 李一航, 周东兴, 韩东升. 一种基于透视变换的远距离双目测距方法[J]. 电子测量技术, 2021, 44(7): 93-99.
LI Y H, ZHOU D X, HAN D S. Long-distance binocular ranging method based on perspective transformation[J]. Electronic Measurement Technology, 2021, 44(7): 93-99 (in Chinese).

[18] HOLLAND J H. Adaptation in natural and artificial systems[M]. Cambridge: The MIT Press, 1992: 1-50.

[19] HAFTKA R T, GURDAL Z. Penalty function method, elements of structural optimization [M]. London: Kluwer Academic Publishers, 1992: 186-198.

[20] GEN M, CHENG R. Genetic algorithm and engineering design[M]. New York: John Wiley&Sons, 1997: 21-30.

[21] 陆依斐. 美新型轰炸机 B-21 亮相[N]. 解放日报, 2022-12-05(008).
LU Y F. US new bomber B-21 appears[N]. Liberation Daily, 2022-12-05(008)(in Chinese).

[22] 马俊, 郭媛丹, 李雪, 等. 美防长冲着中国炫耀军力[N]. 环球时报, 2022-12-05(1).
MA J, GUO Y D, LI X, et al. US defense minister showing off military power to China [N]. Global Times, 2022-12-05(1)(in Chinese).

[23] 魏伟, 肖东风. 大国争相研制新一代隐形轰炸机[N]. 环球时报, 2022-12-03(4).
WEI W, XIAO D F. Great powers vying to develop a new generation of stealth bombers[N]. Global Times, 2022-12-03(4)(in Chinese).

[24] 杨佑绪, 吴志刚, 杨超. 飞翼结构构型气动弹性优化设计方法[J]. 航空学报, 2013, 34(12): 2748-2756.
YANG Y X, WU Z G, YANG C. An aeroelastic optimization design approach for structural configuration of flying wings[J]. Acta Aeronautica et Astronautica Sinica, 2013, 34(12): 2748-2756 (in Chinese).

[25] YANG Y X, WU Z G, YANG C. Equivalent plate modeling for complex wing configurations[J]. Procedia Engineering, 2012, 31: 409-415.

[26] 杨佑绪. 面向机翼构型的气动弹性综合设计方法研究[D]. 北京: 北京航空航天大学, 2013: 57-58.
YANG Y X. Wing Configuration oriented aeroelastic integrated design method study[D]. Beijing: Beihang University, 2013: 57-58(in Chinese).

[27] 万志强, 刘东岳, 唐长红, 等. 大型飞机翼梁位置对气动弹性优化的影响研究[J]. 中国科学: 技术科学, 2012, 42(3): 352-358.
WAN Z Q, LIU D Y, TANG C H, et al. Study on the influence of wing beam position on aeroelastic optimization of large aircraft[J]. Scientia Sinica (Technologica), 2012, 42(3): 352-358 (in Chinese).

[28] JRAD M, DE S, KAPANIA R K. Global-local aeroelastic optimization of internal structure of transport aircraft wing[C]// Processings of the 18th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference. Reston: AIAA, 2017.

[29] LIU Q. EBF3GLWingOpt: a framework for multidisciplinary design optimization of wings using SpaRibs[D]. Blacksburg: Virginia Polytechnic Institute and State University, 2017.

Aeroelastic optimization design of SpaRibs wing structure

ZHOU Quanzhi，YANG Youxu*，SUN Lubin，ZHANG Xingcui，WU Yifei，HUO Mengwen

(School of Aircraft Engineering，Nanchang Hangkong University，Nanchang 330063，China)

Abstract：Traditionally, the internal structure of aircraft wings is generally straight spars and ribs. The use of curvilinear spars and ribs (SpaRibs) can greatly broaden the design space of the aircraft wing structure and further improve the aeroelastic performance of the aircraft wing. Since the linked shape method (LSM) is not easy to automatically model, the projection mapping method was proposed to perform two space transformations to realize the automatic modelling of the SpaRibs. Based on genetic algorithm, an aeroelastic comprehensive optimization design method was proposed for aircraft wings with SpaRibs. The supersonic doublet-lattice method was used to calculate the unsteady aerodynamics, and the modal method was used for the static aeroelastic analysis. The optimal design was carried out by considering the flutter velocity and static aeroelastic deformation constraints. A comprehensive optimization design calculation example of a flying wing aircraft shows that the flutter velocity can be increased by 20.34% on the basis of a weight increase of 1.321% for a wing with a SpaRibs structure. The size parameters are further optimized on the basis of the SpaRibs configuration. Under certain constraints, the weight reduction is 21.76% compared with the initial configuration; the comprehensive (one-step) optimization of the SpaRibs configuration parameters and size parameters is combined, and the weight reduction can reach 26.44% compared with the initial configuration. The weight of the aircraft wing can be effectively reduced by combining SpaRibs design optimization and size optimization, which provides a fast and effective aeroelastic comprehensive optimization design method for the overall design of the flying wing aircraft structure.

Keywords：curvilinear spars and ribs；aeroelasticity；flying wing；optimization design；configuration design；preliminary design

Received：2023-06-12；**Accepted**：2023-10-27；**Published Online**：2023-11-29 10：21

URL：link.cnki.net/urlid/11.2625.V.20231128.1320.002

Foundation item：Jiangxi province "Double Thousand Plan" talent project (CK202006470)

* **Corresponding author**. E-mail：zgdy_1@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0400

基于改进卡尔曼滤波的星内校时守时方法

黄跃[1,2]，王慧泉[1,2,*]，涂实磊[1,2]，金仲和[1,2]

(1. 浙江大学，微小卫星研究中心，浙江 杭州 310027；2. 浙江大学 浙江省微纳卫星重点实验室，浙江 杭州 310027)

摘 要：为解决微纳卫星星上校时守时系统的稳定性差、受环境温度影响大的问题，提出了提高系统性能的新息加权自适应卡尔曼滤波算法。建立温补晶振的频率随温度变化的模型，采用卡尔曼滤波算法滤除输入噪声并实现晶振的校准，采用新息加权技术滤除野值，利用自适应技术减少系统噪声对滤波结果的影响。实验结果表明：所提算法可以在600 s左右实现收敛并且在校时期间可以实时调整，校时过程中可有效地减小输入野值和系统噪声的影响，在环境温度变化时守时精度可达到178 μs/d，有效地提高了系统的稳定性和守时精度。

关 键 词：微纳卫星；卡尔曼滤波；晶振模型；时间同步；新息加权

中图分类号：V11

文献标志码：A **文章编号**：1001-5965(2025)06-2157-08

微纳卫星具有重量轻、研制周期短、机动性强、造价和发射成本低等特点[1]，在空间试验以及商业化应用等方面有着巨大的应用潜力，高精度的星内时间源[2]是实现载荷任务、轨道递推、姿态控制等功能的关键因素。星内普遍采用原子钟作为时间源[3]，考虑到微纳卫星的功耗与成本，温补晶振因其低成本低功耗的特点逐渐受到关注。但是温补晶振的频率受到温度和老化等因素的影响(温补晶振采用温度补偿电路来补偿晶体谐振器的频率漂移，但是补偿后的晶振频率值仍与环境温度相关)，加上自身的固有漂移，亟需一种晶振的驯服以及保持技术[4]的研究。

目前普遍采用的方法是利用全球卫星导航系统 (global navigation satellite system, GNSS) 接收机的秒脉冲 (one pulse per second, 1 PPS) 信号给星上时间系统提供参考[5]，进而实现校时，当GNSS接收机关闭时，采用校时后的时间系统实现守时。但是在工程运用中，由于噪声的干扰，校时用的1 PPS信号输出并不稳定，需要一种滤波方法，滤除校时输入信号的噪声和野值。现有研究中，处理时差测量值的算法有平均滤波器[6]、卡尔曼滤波器[7]、线性最小均方估计滤波器[8]、无迹粒子滤波器[9]等。在守时期间，晶振会因为温度变化和器件老化产生频率漂移，所以需要合适的算法计算补偿量，目前处理的方式有：建立晶振模型[10]，通过最小二乘法拟合模型参数[11]，工作时根据温度值计算出对应的补偿量；通过硬件电路调整实现频率的稳定[12]，直接将补偿值反馈到实际电路中；采用神经网络法预测出补偿值，实现频率的补偿[13]等。目前在基于温补晶振的守时性能研究中，采用分级校正算法[6]可以保证24 h内时间漂移小于1 ms，采用卡尔曼滤波算法可以实现24 h内漂移在700 μs左右。

上述研究算法中，一方面在校时期间，缺乏对输入信号异常值以及系统波动的处理，另一方面缺乏对晶振模型的动态调整。本文尝试采用基于新息加权自适应卡尔曼滤波算法，该算法不仅可以很好地滤除输入信号的噪声和野值，应对系统的不稳定情况，还可以对晶振的模型参数实现实时的校正，进而提高系统的稳定性和守时精度。

本文首先对星内校时守时系统的架构与机制

收稿日期：2023-06-21；录用日期：2023-09-22；网络出版时间：2023-10-26 14:20
网络出版地址：link.cnki.net/urlid/11.2625.V.20231026.1050.002
*通信作者. E-mail：hqwang@zju.edu.cn

引用格式：黄跃，王慧泉，涂实磊，等. 基于改进卡尔曼滤波的星内校时守时方法 [J]. 北京航空航天大学学报，2025，51（6）：2157-2164.
HUANG Y，WANG H Q，TU S L，et al. On-board timekeeping method based on improved Kalman filter [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（6）：2157-2164 （in Chinese）.

进行阐述，再分析晶振的温度与频率关系和老化与频率关系，建立晶振频率模型，基于该模型建立卡尔曼滤波算法模型，进而将新息加权算法和自适应算法融合到卡尔曼滤波算法中，最后通过实验证明所提算法的有效性。

1 星内校时守时系统介绍

星内校时守时系统主要由核心处理器现场可编程逻辑门阵列（field programmable gate array, FPGA）、上位机数字信号处理芯片（digital signal processing, DSP）、主备份GNSS接收机、温度传感器和温补晶振组成，如图1所示。

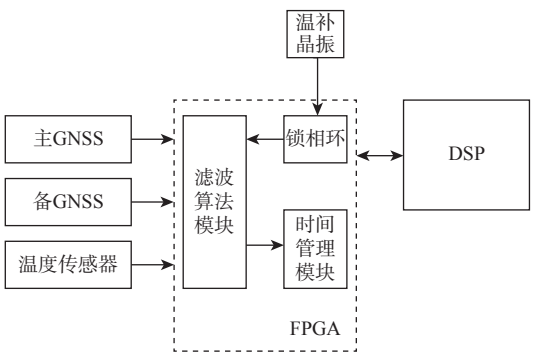


图1 系统架构图

Fig. 1 System architecture diagram

在校时期间，FPGA接收上位机DSP的校时指令，此时GNSS接收机已经开始工作，产生1 PPS信号。1 PPS信号是在每秒的起始发出脉冲信号，FPGA以每周期采样的方式捕获脉冲信号，每当捕获到脉冲信号就开始内部计数，直到捕获到下一个脉冲，记录当前计数周期值，同时计数器清零，开始下一次计数，则每次记录的计数周期值就是这一秒系统时钟的频率值。如此一来，系统每秒都会获得一个计数周期值，该值的稳定性取决于1 PPS秒脉冲和系统时钟的稳定性。因此，需要加入滤波算法，以每秒记录的计数周期值作为输入，进行滤波处理，实时调整晶振模型参数，并得到最优的频率估计值，该值就可作为当前系统的频率值，并将得到的最优估计值传给时间管理模块。时间管理模块以该值作为一秒的计数周期，每当计数达到该周期时，则时间就增加1 s，其中时间的精度就取决于系统的频率，本文的系统时钟采用的是60 MHz，最高精度是16.67 ns左右（实际频率并不是严格的60 MHz）。时间管理模块主要是维护系统时间并将时间信息以总线的方式传递给上位机DSP，工作模式主要分为2种：①校时；②守时。校时分为频率校正和相位校正，频率校正如上文所述，以接收滤波之后的计数周期值作为当前系统的频率，相位校正是以每次捕获的秒脉冲开始作为秒的零时刻；守时主要是以校时期间获取到的计数周期值作为计数周期，并实时监测温度传感器的变化，根据晶振模型调整该计数周期值。外部温补晶振作为系统的时钟源，输入到FPGA引脚，FPGA内部锁相环IP核将时钟倍频到60 MHz，作为系统时钟。

2 卡尔曼数学模型建立

2.1 晶振频率模型

晶振温度频率特性指的是晶振频率随着温度变化的特性，本文采用的是AT切晶体的温补晶振，其频率随着温度的变化呈三次函数[10]分布，其温度频率特性可表示为

$$f_T = a_3T^3 + a_2T^2 + a_1T + a_0 \tag{1}$$

式中：T为当前时刻温度值；a_3为三次项系数；a_2为二次项系数；a_1为一次项系数；a_0为温度为零时频率值。

晶振老化频率特性指的是晶振频率随着生产时间变化的特性，其老化特性[14]随时间变化呈指数函数分布，其频率老化特性可表示为

$$f_t = A\ln(at + c) + B\ln(bt + d) \tag{2}$$

式中：t为时间变量；A、a、c为应力驰豫效应的系数；B、b、d为质量负载效应的系数。2种效应共同作用成晶振老化特性，2种效应都是随着时间推移越来越稳定，考虑到星内晶振是长期上电的，老化效应在短时间内对晶振频率的影响很小，可以通过每次校时后进行估计，结合温度频率特性，将老化特性转成一个可变的参数，形式如下：

$$f_T = a_3T^3 + a_2T^2 + a_1T + e \tag{3}$$

式中：a_3、a_2、a_1与温度频率特性表示相同；e为a_0与f_t的和。

2.2 卡尔曼滤波模型

假设1 PPS信号是标准的1 s时长，FPGA在每一次1 PPS信号的上升沿开始采样并计数，直到下一次1 PPS信号的上升沿到来，计数结束，将此时的计数周期值当作当前系统时钟的频率值。将系统离散化，根据式(3)，将n时刻系统的预测频率表示为

$$N_{\mathrm{f}}[n] = a_3T^3[n] + a_2T^2[n] + a_1T[n] + e \tag{4}$$

$n-1$时刻的预测频率表示为

$$N_{\mathrm{f}}[n-1] = a_3T^3[n-1] + a_2T^2[n-1] + a_1T[n-1] + e \tag{5}$$

联合式(4)和式(5)得

$$N_{\mathrm{f}}[n] = N_{\mathrm{f}}[n-1] + a_3\Delta\tau_3 + a_2\Delta\tau_2 + a_1\Delta\tau_1 \tag{6}$$

式中：$\Delta\tau_k = T^k[n] - T^k[n-1]$。

建立卡尔曼状态模型:

$$\bar{\boldsymbol{X}}[n]=\boldsymbol{A}[n]\hat{\boldsymbol{X}}[n-1]+\boldsymbol{u}[n] \tag{7}$$

式中: 状态矩阵为

$$\boldsymbol{X}[n]=\begin{bmatrix} N_f[n] \\ a_3 \\ a_2 \\ a_1 \end{bmatrix} \tag{8}$$

$\hat{\boldsymbol{X}}[n-1]$为$n-1$时刻的最优估计状态矩阵, $\bar{\boldsymbol{X}}[n]$为n时刻的预测状态矩阵。

状态转移矩阵:

$$\boldsymbol{A}[n]=\begin{bmatrix} 1 & \Delta\tau_3 & \Delta\tau_2 & \Delta\tau_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

状态噪声:

$$\boldsymbol{u}[n]=\begin{bmatrix} N_w[n] \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{10}$$

式中: $N_w[n]\sim N(0,{\delta_w}^2)$, ${\delta_w}^2$为噪声方差。

卡尔曼观测模型为

$$\boldsymbol{Z}[n]=\boldsymbol{H}\boldsymbol{X}[n]+\boldsymbol{V}[n] \tag{11}$$

式中:

$$\boldsymbol{Z}[n]=\begin{bmatrix} Z_m[n] \\ Z_s[n] \end{bmatrix} \tag{12}$$

$Z_{\mathrm{m}}[n]$和$Z_{\mathrm{s}}[n]$分别为主备份 GNSS 接收机n时刻的观测输入。

状态向量到测量向量的转换矩阵:

$$\boldsymbol{H}=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{13}$$

测量噪声矩阵:

$$\boldsymbol{V}[n]=\begin{bmatrix} N_m[n] \\ N_s[n] \end{bmatrix} \tag{14}$$

式中: $N_m[n]\sim(0,\delta_{vm}^2)$, $N_s[n]\sim(0,\delta_{vs}^2)$, δ_{vm}^2和δ_{vs}^2分别为主备份 GNSS 接收机观测的噪声方差。

n时刻的先验估计协方差方程:

$$\bar{\boldsymbol{P}}[n]=\boldsymbol{A}[n]\hat{\boldsymbol{P}}[n-1]\boldsymbol{A}^{\mathrm{T}}[n]+\boldsymbol{Q}[n] \tag{15}$$

式中:

$$\boldsymbol{Q}[n]=\begin{bmatrix} \delta_w^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{16}$$

$\bar{\boldsymbol{P}}[n]$为n时刻的预测协方差, $\hat{\boldsymbol{P}}[n-1]$为$n-1$时刻的最优估计协方差。

卡尔曼增益方程:

$$\boldsymbol{K}[n]=\bar{\boldsymbol{P}}[n]\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\bar{\boldsymbol{P}}[n]\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R}[n])^{-1} \tag{17}$$

式中: $\boldsymbol{K}[n]$为n时刻的卡尔曼增益。

$$\boldsymbol{R}[n]=\begin{bmatrix} \delta_{vm}^2 & 0 \\ 0 & \delta_{vs}^2 \end{bmatrix} \tag{18}$$

修正估计方程:

$$\hat{\boldsymbol{X}}[n]=\bar{\boldsymbol{X}}[n]+\boldsymbol{K}[n](\boldsymbol{Z}[n]-\boldsymbol{H}\bar{\boldsymbol{X}}[n]) \tag{19}$$

n时刻的后验估计协方差方程:

$$\hat{\boldsymbol{P}}[n]=(\boldsymbol{I}-\boldsymbol{K}[n]\boldsymbol{H})\bar{\boldsymbol{P}}[n] \tag{20}$$

式中:

$$\boldsymbol{I}=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{21}$$

2.3　自适应因子

自适应因子是源自于导航定位算法中[15]的概念, 合理的自适应因子能平衡状态模型信息与观测模型信息的权比, 可以控制状态模型误差对系统参数解的影响。在观测值可靠的情况下, 预测残差向量能反映状态模型的误差大小。本文采用预测残差协方差矩阵表示法表示自适应因子。

预测残差向量可以表示为

$$\bar{\boldsymbol{V}}[n]=\boldsymbol{A}[n]\bar{\boldsymbol{X}}[n]-\boldsymbol{Z}[n] \tag{22}$$

预测残差的理论协方差为

$$\bar{\boldsymbol{P}}_{\bar{V}}[n]=\boldsymbol{A}[n]\bar{\boldsymbol{P}}[n]\boldsymbol{A}^{\mathrm{T}}[n]+\boldsymbol{R}[n] \tag{23}$$

采用单历元预测残差估计法, 预测残差的实际协方差估计值为

$$\hat{\boldsymbol{P}}_{\bar{V}}[n]=\bar{\boldsymbol{V}}_n\bar{\boldsymbol{V}}_n^{\mathrm{T}} \tag{24}$$

最后, 将自适应因子表示为

$$\alpha[n]\approx\begin{cases} 1 & \text{其他} \\ \dfrac{\mathrm{tr}(\overline{\boldsymbol{P}}_{\bar{V}}[n])}{\mathrm{tr}(\hat{\boldsymbol{P}}_{\bar{V}}[n])} & \mathrm{tr}(\hat{\boldsymbol{P}}_{\bar{V}}[n])>\mathrm{tr}(\overline{\boldsymbol{P}}_{\bar{V}}[n]) \end{cases} \tag{25}$$

将自适应因子代入卡尔曼增益方程和后验估计协方差方程中:

$$\boldsymbol{K}[n]=\alpha[n]^{-1}\bar{\boldsymbol{P}}[n]\boldsymbol{H}^{\mathrm{T}}(\alpha[n]^{-1}\boldsymbol{H}\bar{\boldsymbol{P}}[n]\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R}[n])^{-1} \tag{26}$$

$$\hat{\boldsymbol{P}}[n]=(\boldsymbol{I}[n]-\boldsymbol{K}[n]\boldsymbol{H})\alpha[n]^{-1}\bar{\boldsymbol{P}}[n] \tag{27}$$

由方程可知, 当系统状态模型发生较大误差时, 通过自适应因子膨胀先验协方差矩阵, 增加卡尔曼增益, 进一步增大观测信息对系统的影响, 降低预测信息的权重, 从而减小系统模型误差造成的影响。

2.4　新息加权因子

在校时期间, 1 PPS 的信号存在钟差不稳导致

跳变野值的情况以及晶振本身存在不确定性波动，都会引起观测输入的较大偏差，这些偏差加入系统中容易导致系统参数偏移甚至发散。本文通过引入新息加权技术调整修正估计方程，当观测值存在野值时，可滤除这些野值，减小对系统带来的影响。

新息向量[16]可表示为

$$\boldsymbol{e}[n]=\boldsymbol{Z}[n]-\bar{\boldsymbol{Z}}[n] \tag{28}$$

式中：

$$\bar{\boldsymbol{Z}}[n]=\boldsymbol{H}\bar{\boldsymbol{X}}[n] \tag{29}$$

新息加权因子表示为

$$\phi[n]\approx\begin{cases}1 & \Psi[n]\leqslant C[n]/\lambda[n]\\ \dfrac{C[n]}{\sqrt{\Psi[n]\lambda[n]}} & \Psi[n]>C[n]/\lambda[n]\end{cases} \tag{30}$$

式中：

$$\Psi[n]=\boldsymbol{e}[n]^{\mathrm{T}}(\bar{\boldsymbol{P}}[n]+\boldsymbol{R}[n])^{-1}\boldsymbol{e}[n] \tag{31}$$

$$\lambda[n]=\max(\boldsymbol{K}[n](\bar{\boldsymbol{P}}[n]+\boldsymbol{R}[n])\boldsymbol{K}^{\mathrm{T}}[n]) \tag{32}$$

式中：max表示求矩正最大特征值；$C[n]$为门限值，可根据工程经验选取。

将新息加权因子加入修正估计方程中：

$$\hat{\boldsymbol{X}}[n]=\bar{\boldsymbol{X}}[n]+\phi[n]\boldsymbol{K}[n](\boldsymbol{Z}[n]-\boldsymbol{H}\bar{\boldsymbol{X}}[n]) \tag{33}$$

由式(33)可知，当观测值出现野值时，$\Psi[n]$增大，$\phi[n]$减小，减小观测值影响的权重，进而减小野值造成的影响。

3 实验结果分析

3.1 晶振模型参数拟合效果分析

在地面测试时，可以通过GNSS模拟器提供高稳定的频率源(默认是标准的秒脉冲)，在温度循环实验中测出晶振模型的温度漂移数据，通过本文算法实现实验数据的参数拟合，拟合结果如图2所示。

图3为时钟频率相对25 ℃频率值的偏移曲线图，横坐标是时间变化，每20 s进行一次温度值的递增，递增精度是1 ℃；纵坐标是FPGA内部在2个脉冲之间的时钟计数周期值，为方便表示，图中画出的是相对25 ℃时的计数值。$Z_{\mathrm{m}}[n]$表示主份1 PPS的输入，$Z_{\mathrm{s}}[n]$表示备份1 PPS的输入，参考值表示理论时钟频率相对偏移值，$X[n]$表示n时刻的算法最优估计值。由图可知，拟合的时钟频率偏移曲线基本上符合三次函数分布，估计值与理论值只有在温度变化时存在误差，当温度稳定后，两者重合。图2为三次函数对应系数值的收敛效果图，3个系数的初值都是1，可见在600 s左右，3个系数逐渐收敛到一个稳定的值。

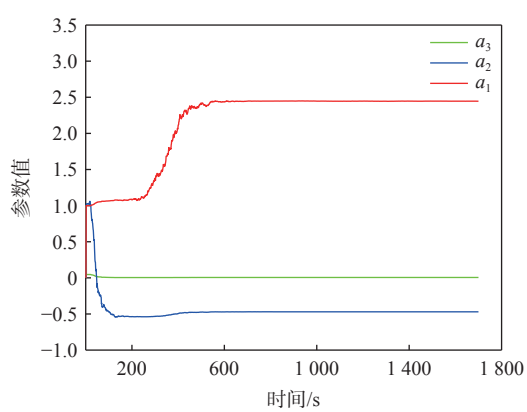


图2 参数收敛图

Fig. 2 Parameter convergence graph

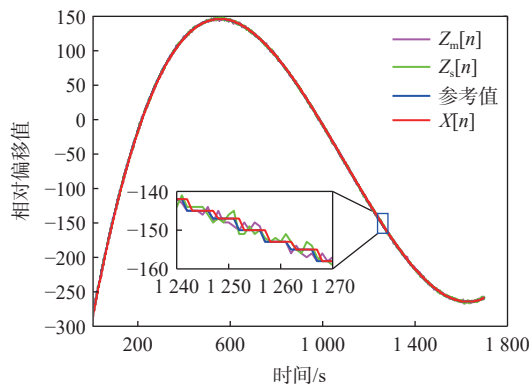


图3 时钟频率变化图

Fig. 3 Clock frequency change graph

通过2个实验效果图可以说明本文算法可以实时校正晶振模型参数，实现晶振模型函数曲线的拟合。

3.2 新息加权自适应因子效果分析

当晶振模型训练好之后，就可以应用到实际的校时系统当中。在实际工作中，会出现GNSS的卫星接收信号差、钟差不稳的情况，比如受到其他信号干扰或者卫星姿态不稳定，都会导致输出的1 PPS信号出现野值或较大误差的值，如果仅依赖上位机监测不稳定状态，会出现滞后性，此时误差已经带入到校时系统中；系统晶振本身也会发生不确定性漂移，与上述1 PPS野值共同导致算法中的观测输入出现较大偏差。另外，温度传感器的采样值也会发生偶然的跳变，导致系统的预测状态出现较大的误差。针对上述两类误差来源，本文采用新息加权自适应算法来增强系统的稳定性。

新息加权因子主要是滤除观测输入的误差，根据上文分析误差来源分别为1 PPS信号的波动和晶振自身频率的漂移，下面通过实验数据发现实际误差的相关特性。

首先是1 PPS信号野值的收集，采用GNSS模拟器设备，将接收卫星信号的通道功率调低，模拟

卫星信号减弱, 进而收集 1 PPS 信号值的变化。如图 4 是在调整 GNSS 模拟器信号强度时收集到的 1 PPS 信号波动数据, 横坐标是时间, 纵坐标是相对初始时刻的计数值。在前 500 s, 信号正常, 系统正常工作, 之后通过降低接收信号强度, 可观察到 1 PPS 信号值发生波动(图中仅展示波动低于 100 个计数周期的信号值大小), 可以发现, 信号波动带来的影响是比较大的, 除去较大的波动值, 波动在 20 个计数周期内。

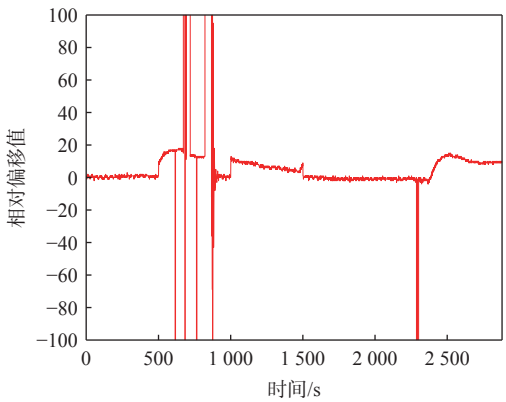


图 4　1 PPS 信号频率变化图

Fig. 4　1 PPS signal frequency change graph

进一步分析晶振自身频率的漂移, 如图 5 是在恒定温度且 1 PPS 信号稳定的情况下测得的晶振频率变化图, 横坐标是时间, 纵坐标是相对初始时刻的计数周期值, 可见晶振频率会存在一些超大的野值和 11 个周期内的波动。

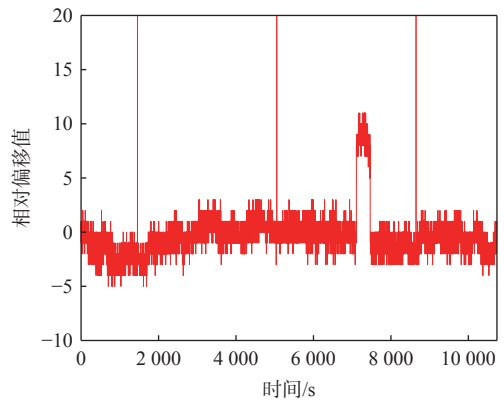


图 5　晶振频率变化图

Fig. 5　Crystal oscillator frequency change graph

下面进行仿真实验, 在校时期间, 加入模拟的信号波动, 持续 15 s, 波动幅度在 20 个周期内, 比较平均滤波、卡尔曼滤波以及本文算法的滤波效果。

由图 6 和图 7 可知, 在出现波动的 15 s 内, 系统的最优估计值均受到较大的影响, 其中平均滤波器的最大误差达到了 10 个计数周期, 恢复到正常值需要在波动结束后 8 s 左右; 普通卡尔曼滤波器最大误差相差 6 个计数周期, 恢复到正常值需要在波动结束后 14 s 左右; 而本文采用的算法仅造成短暂的 2 个周期的偏移, 如图 8 所示, 在波动结束后, 4 s 左右即迅速恢复到正常值, 该性能可以通过调节式 (30) 中的 $C[n]$门限值, 做适应性的变动, 本次的仿真结果采用的是 0.2。显然采用本文算法后, 校时的稳定性以及波动结束后的恢复速度都比传统算法优越, 如果在波动发生时, 上位机停止校时, 此时没有恢复时间, 系统将会使用当前的最优估计

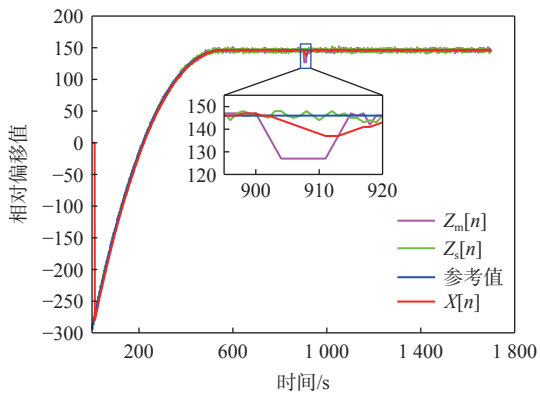


图 6　平均滤波算法滤波效果图

Fig. 6　Average filter algorithm filtering effect diagram

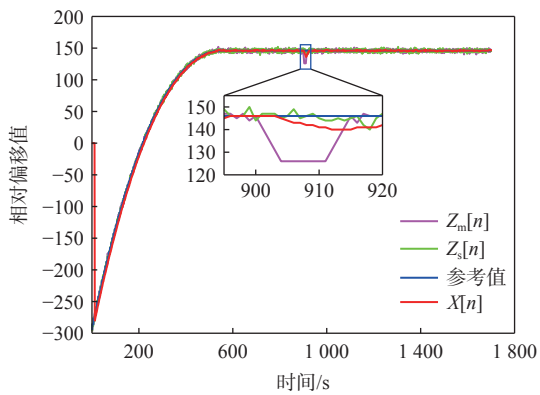


图 7　卡尔曼滤波算法滤波效果图

Fig. 7　Kalman filter algorithm filtering effect diagram

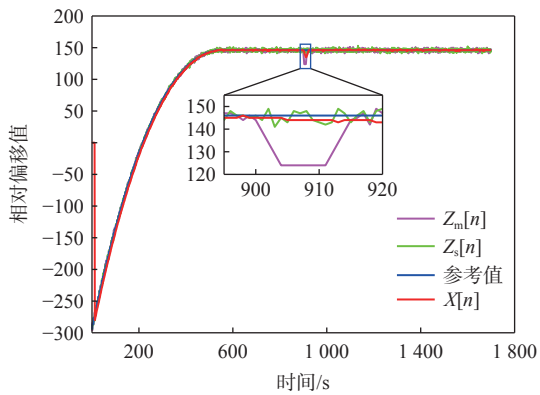


图 8　新息加权自适应卡尔曼滤波算法滤波效果图

Fig. 8　The filtering effect of the Innovation-weighted adaptive Kalman filter algorithm

值作为守时的计数周期值，本文算法因为受波动影响很小，所以停止校时后的守时精度误差也会很小。

自适应因子主要是可以应对温度采样不稳定带来的预测误差。如图 9 为温度循环过程中晶振温度的采样值，由图可知，采样值会概率性出现波动，此时算法中当前时刻的估计值将会出现较大偏差，进而影响最优估计值。

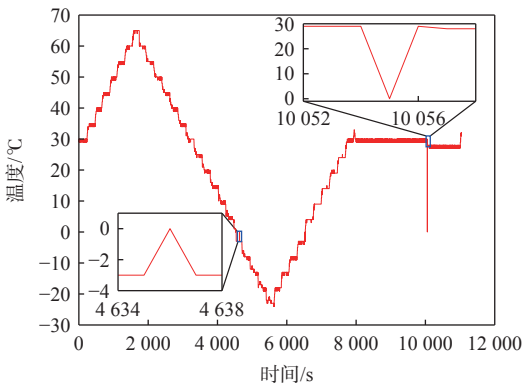


图 9 温循晶振温度采样图

Fig. 9 Crystal oscillator temperature change plot during temperature cycling

下面进行仿真实验，在校时过程中第 900 s 时加入 3 ℃ 温度值误差，观察自适应因子引入前后的系统变化。

如图 10 所示，不含自适应因子的卡尔曼滤算法的最优估计值受到较大的影响，在第 900 s 后误差达到了 8 个计数周期，需要 20 s 左右才能恢复到正常状态；而加入自适应因子后，如图 11 所示，在第 900 s 时产生了轻微波动，误差是 2 个计数周期，随后便恢复正常状态。可见引入自适应因子可以大大减小因温度采样误差带来的影响。

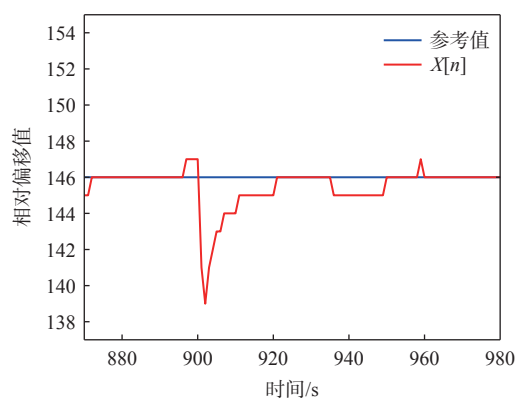


图 10 不含自适应因子频率估计曲线

Fig. 10 Frequency estimation curve without adaptive factor

3.3 晶振守时效果分析

校时完成后，关闭 GNSS 接收机，进入守时阶段，假设在守时期间温度变化趋势是：在 24 h 内，温度值从 10 ℃ 增长到 28 ℃ 再降低到 10 ℃（变化趋势是参考在轨卫星实测的热控分系统遥测数据），以此来验证守时效果，并与没有晶振模型守时做对比。

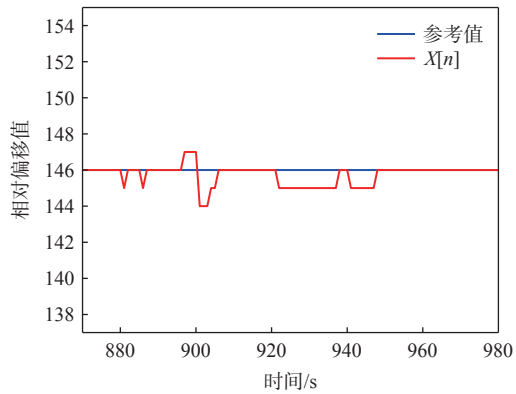


图 11 含自适应因子频率估计曲线

Fig. 11 Frequency estimation curve with adaptive factor

图 12 为 24 h 内，伴随着温度变化的守时精度图，可见没有晶振模型的偏移是毫秒级，而有晶振模型的漂移是微秒级，有晶振模型的守时精度大大增强，具体偏移量见表 1。其中 24 h 内的守时精度，相比较分级校正算法的 1 ms 精度和卡尔曼滤波算法的 737 μs，本文将守时精度提高到了 200 μs 左右。

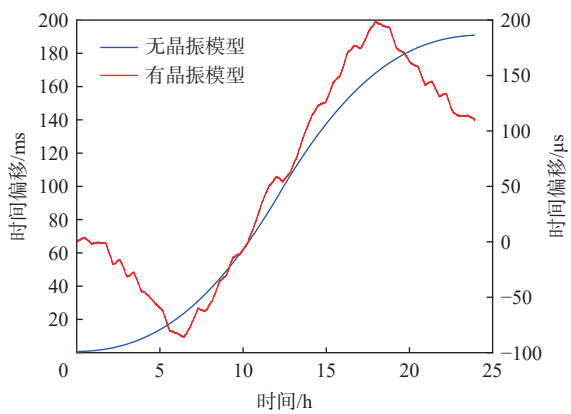


图 12 关闭 GNSS 后时间偏移曲线

Fig. 12 Time offset curve after turning off GNSS

表 1 有无晶振模型时间偏移表

Table 1 Time offset table with or without crystal oscillator model

时长/h	无晶振模型偏移/ms	有晶振模型偏移/μs
0.5	0.174 2	5.616 7
1	0.473 2	6.366 7
3	3.205 4	33.466 7
6	12.835 5	82.133 3
24	107.301 7	178.300 0

使用本文算法建立晶振模型，可大大提高守时的精度，并且可通过每次 GNSS 开机的校时结果，

记录不同温度下的频率偏移值, 在有需求时进行参数校正, 在每次参数校正完时, 计算出式 (3) 中的 e值, 弥补晶振老化带来的偏移影响。

4　结　论

本文为提高微纳卫星校时守时系统的稳定性和守时精度, 提出了一种新息加权自适应卡尔曼滤波算法, 经过实验验证, 得出以下结论:

1) 在地面测试时, 本文算法可以实现晶振模型参数的标校; 在轨运行时, 可以利用历史数据实现晶振模型参数的修正, 补偿因温度变化和老化因素导致的晶振频率漂移;

2) 通过与平均滤波算法、卡尔曼滤波算法比较, 本文算法在校时期间可以很好地滤除输入野值, 且能减少温度采样值波动带来的影响, 提高系统的稳定性;

3) 在温度变化的环境下, 本文算法的守时精度提升到 24 h 内偏移 200 μs 左右, 优于传统算法。

本文算法有望运用到低成本低功耗的微纳卫星时间同步系统中, 为系统提供高可靠的时钟源。

参考文献 (References)

[1] ZHAO X Y, ZHAO C J, LI J L, et al. Research on design, simulation, and experiment of separation mechanism for micro-nano satellites[J]. Applied Sciences, 2022, 12(12): 5997.

[2] XU J L, ZHANG C J, WANG C H, et al. Approach to inter-satellite time synchronization for micro-satellite cluster[J]. Journal of Systems Engineering and Electronics, 2018, 29(4): 805-815.

[3] YAO J, YOON S, STRESSLER B, et al. GPS satellite clock estimation using global atomic clock network[J]. GPS Solutions, 2021, 25(3): 106.

[4] BIANCALANA F, SKRYABIN D V, YULIN A V. Theory of the soliton self-frequency shift compensation by the resonant radiation in photonic crystal fibers[J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2004, 70(1 Pt 2): 016615.

[5] GIRALO V, D'AMICO S. Distributed multi-GNSS timing and localization for nanosatellites[J]. Navigation, 2019, 66(4): 729-746.

[6] 苏星, 王慧泉, 金仲和. 基于 GPS 校准的皮卫星高精度时间系统方案[J]. 传感技术学报, 2016, 29(8): 1200-1204.
SU X, WANG H Q, JIN Z H. The high-precisiontime system design of pico-satellite based on GPS receiver[J]. Chinese Journal of Sensors and Actuators, 2016, 29(8): 1200-1204 (in Chinese).

[7] GIORGI G, NARDUZZI C. Performance analysis of Kalman-filter-based clock synchronization in IEEE 1588 networks[J]. IEEE Transactions on Instrumentation and Measurement, 2011, 60(8): 2902-2909.

[8] PALLIER D, LE CAM V, PILLEMENT S. Energy-efficient GPS synchronization for wireless nodes[J]. IEEE Sensors Journal, 2021, 21(4): 5221-5229.

[9] 李振海, 唐金锐, 熊斌宇, 等. 基于新息加权自适应 UPF 算法的智能变电站 PTP 主时钟源校准新方法[J]. 电网技术, 2022, 46(5): 1662-1671.
LI Z H, TANG J R, XIONG B Y, et al. A novel calibration method for the master PTP clock in smart substations using new information sequence weighting and adaptive unscented Kalman particle filter algorithms[J]. Power System Technology, 2022, 46(5): 1662-1671 (in Chinese).

[10] KOSYKH A V, IONOV B P. Dynamic temperature model and dynamic temperature compensation of crystal oscillators[J]. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2002, 41(3): 370-374.

[11] LIU Z, CHENG Y H, WANG P, et al. A method for remaining useful life prediction of crystal oscillators using the Bayesian approach and extreme learning machine under uncertainty[J]. Neurocomputing, 2018, 305: 27-38.

[12] TAN F, LIAO S, XU L, et al. New method for 100-MHz high-frequency temperature-compensated crystal oscillator[J]. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 2020, 67(12): 2745-2749.

[13] ANDREJEVIC M V, LITOVSKI V. Electronic circuits modeling using artificial neural networks[J]. Journal of Automatic Control, 2003, 13(1): 31-37.

[14] VIG J R. Quartz crystal resonators and oscillators for frequency control and timing applications. A tutorial [J]. Nasa Sti/recon Technical Report N, 1994, 95: 19519.

[15] MOHAMED A H, SCHWARZ K P. Adaptive Kalman filtering for INS/GPS[J]. Journal of Geodesy, 1999, 73(4): 193-203.

[16] BEN SASSI H, FATIMA E, ES-SBAI N. State of charge estimation by multi-innovation unscented Kalman filter for vehicular applications[J]. Journal of Energy Storage, 2020, 32: 101978.

On-board timekeeping method based on improved Kalman filter

HUANG Yue[1, 2], WANG Huiquan[1, 2, *], TU Shilei[1, 2], JIN Zhonghe[1, 2]

(1. Micro-Satellite Research Center, Zhejiang University, Hangzhou 310027, China;
2. Key Laboratory of Micro-Nano Satellite Research, Zhejiang Province, Hangzhou 310027, China)

Abstract: A new innovation-weighted adaptive Kalman filtering algorithm is proposed to improve system performance by addressing the issues of poor stability of the internal calibration punctual system of micro-nano satellites and low punctuality accuracy when large changes occur due to the influence of ambient temperature. A model of frequency change of temperature-compensated crystal oscillator with temperature is established, the Kalman filter algorithm is used to filter out the input noise and realize the calibration of the crystal oscillator model parameters, the input field value is filtered out by new innovation-weighted technology, and the influence of system noise on the filtering results is reduced by adaptive technology. The experimental results show that the algorithm can achieve the convergence of model parameters in about 600s and can be adjusted in real time during the calibration period. The influence of input field value and system noise can be successfully controlled throughout the calibration process; when the ambient temperature changes, the punctual accuracy can reach 178 μs/day. The proposed algorithm effectively improves the stability and punctual accuracy of the system.

Keywords: micro-nano satellite; Kalman filter; crystal oscillator model; time synchronization; innovation-weighted

Received: 2023-06-21; **Accepted**: 2023-09-22; **Published Online**: 2023-10-26 14:20
URL: link.cnki.net/urlid/11.2625.V.20231026.1050.002

* **Corresponding author**. E-mail: hqwang@zju.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0403

基于强化学习的舰载机着舰直接升力控制技术

柳仁地，江驹*，张哲，刘翔
(南京航空航天大学 自动化学院，南京 211106)

摘　　要：针对舰载机自动着舰过程中受甲板运动及舰尾流扰动很容易发生触舰危险的问题，提出了基于近端策略优化(PPO)算法的舰载机自动着舰直接升力控制方法。PPO控制器以俯仰角、高度、航迹倾斜角、俯仰角速率、高度误差和航迹倾斜角速率等6个状态变作为输入，以襟翼的舵偏角增量作为输出，实现舰载机在着舰时航迹倾斜角的快速响应。与传统控制器相比，PPO控制器中的Actor-Critic框架大大提高了控制量的计算效率，降低了参数优化的难度。仿真实验基于MATLAB/Simulink中的F/A-18飞机动力学/运动学模型。利用PyCharm平台上构建的深度强化学习训练环境，通过UDP通信实现2个平台之间的数据交互。仿真结果表明：所提方法具有响应速度快、动态误差小的特点，能够将着舰的高度误差稳定在±0.2 m以内，具有较高的控制精度。

关 键 词：舰载机自动着舰；深度强化学习；近端策略优化算法；直接升力控制；UDP通信
中图分类号：V249.1
文献标志码：A　　**文章编号**：1001-5965(2025)06-2165-11

舰载机着舰是一项非常复杂的任务，由于着舰过程中受大气扰动、舰尾流、甲板运动、波浪和海况等环境因素的影响，舰载机的着舰精度与成功率受到很大影响[1]。

自从直接升力控制技术被应用在着舰过程中[2]，着舰的安全性和稳定性得到了提高。目前直接升力控制主要以PID控制、模糊控制和自抗扰等方法为主[3]。文献[4]通过基于直接升力的非线性综合解耦控制器进行着舰控制。文献[5]提出了一种动态分配控制方法，使襟翼像其他舵面一样能够快速驱动。文献[6]提出了基于移动路径跟踪(moving path following, MPF)法则的着舰引导律和一种新的集成式直接升力控制方法。文献[7]利用动态逆反馈控制方法设计飞行控制器，通过控制分配技术实现特定飞行操纵。虽然上述方法具有良好的控制效果，但对于六自由度飞机模型，都需要进行精确的数学建模和参数设计，特别是当被控对象处于强耦合、强非线性和强时变环境时，参数的调节和优化非常困难，控制效果受到很大影响。

随着AlphaGo的问世，人工智能技术得到了迅速发展，其中以强化学习为代表的算法雨后春笋般涌现[8]。再加上智能计算、并行计算等计算能力的高效提升，基于智能体与环境交互的强化学习算法在任务分配、路径规划、机动决策[9]、协同对抗、飞行控制等领域得到了广泛的应用。文献[10]采用深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法实现了固定翼飞机在平飞状态下的俯仰姿态控制。文献[11]设计了基于DDPG算法的控制器实现航天器的姿态控制。文献[12]在预设性能控制器的基础上引入强化学习算法，用于评估系统性能。文献[13]利用建立好的专家经验数据集，通过模仿强化学习设计了固定翼飞机的姿态控制器。文献[14]提出了一种基于TD3算法的无人机近距空战格斗自主决策模型。文献[15]针对舰载无人机，通过TD3算法设计了全自动着舰控制器。

收稿日期：2023-06-21；**录用日期**：2023-11-10；**网络出版时间**：2023-11-24 17:51
网络出版地址：link.cnki.net/urlid/11.2625.V.20231123.1537.005
*通信作者. E-mail：jiangju@nuaa.edu.cn

引用格式：*柳仁地，江驹，张哲，等. 基于强化学习的舰载机着舰直接升力控制技术[J]. 北京航空航天大学学报，2025，51(6)：2165-2175. LIU R D，JIANG J，ZHANG Z，et al. Direct lift control technology of carrier aircraft landing based on reinforcement learning[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2165-2175 (in Chinese).*

基于上述分析，本文将深度强化学习应用于直接升力控制中，提出一种基于近端策略优化(proximal policy optimization, PPO)算法[16-17]的自动着舰纵向控制器设计方法。在构建完Actor-Critic框架后，设计适当的奖励函数引导舰载机学习襟翼控制策略。仿真结果表明，训练后的纵向自动着舰控制器能够根据不同状态输出相应的襟翼偏转指令，实现飞行轨迹的快速变化，从而使舰载机安全稳定地着舰。

1 舰载机、甲板运动及舰尾流建模

1.1 舰载机模型建立

F/A-18舰载机的部分气动参数以及外形参数如表1所示。

表1 气动参数及外形参数

Table 1 Pneumatic parameters and shape parameters

参数	数值
机翼面积 S_w/m^2	37.16
平均气动弦长 $\bar{c}$/m	3.51
翼展 b/m	11.43
飞机质量 m/kg	16 651
绕O_bx_b轴转动惯量 I_{xx}/(kg·m^2)	31 184
绕O_by_b轴转动惯量 I_{yy}/(kg·m^2)	205 130
绕O_bz_b轴转动惯量 I_{zz}/(kg·m^2)	230 415
绕O_bx_b轴惯性积I_{xz}/(kg·m^2)	−4 028

舰载机在着舰过程中，主要受重力、气动力以及发动机推力的作用。假设发动机的安装角为0°，则推力公式为

$$T=\delta_T T_{\max} \tag{1}$$

式中：δ_T为油门开度；$T_{\max}=f(h,Ma)$为最大推力，与舰载机当前的高度和空速有关。

气动力在气流坐标系中可以分解为阻力D、侧力Y、升力L，气动力矩在机体坐标系中可以分解为滚转力矩$\bar{L}$、俯仰力矩M、偏航力矩N，气动力和气动力矩的表达式如下：

$$\begin{cases} L=\dfrac{1}{2}\rho V^2 S_w C_L \\ D=\dfrac{1}{2}\rho V^2 S_w C_D \\ Y=\dfrac{1}{2}\rho V^2 S_w C_Y \end{cases} \tag{2}$$

$$\begin{cases} \bar{L}=\dfrac{1}{2}\rho V^2 S_w b C_l \\ M=\dfrac{1}{2}\rho V^2 S_w \bar{c} C_m \\ N=\dfrac{1}{2}\rho V^2 S_w b C_n \end{cases} \tag{3}$$

式中：ρ为空气密度；V为舰载机的空速；C_L、C_D、C_Y分别为舰载机的总升力系数、总阻力系数、总侧力系数；C_l、C_m、C_n分别为舰载机的总滚转力矩系数、总俯仰力矩系数、总偏航力矩系数。气动力系数和气动力矩系数如下：

$$\begin{cases} C_L=C_{L_0}+C_{L_\alpha}\alpha+C_{L_{\delta_e}}\delta_e+C_{L_{\delta_f}}\delta_f \\ C_D=C_{D_0}+C_{D_\alpha}\alpha+C_{D_{\delta_e}}\delta_e+C_{D_{\delta_f}}\delta_f \\ C_Y=C_{Y_\beta}\beta+C_{Y_{\delta_r}}\delta_r+C_{Y_{\delta_a}}\delta_a \end{cases} \tag{4}$$

$$\begin{cases} C_l=C_{l_\beta}\beta+C_{l_{\delta_a}}\delta_a+C_{l_{\delta_r}}\delta_r+\dfrac{b}{2V}C_{l_p}p+\dfrac{b}{2V}C_{l_r}r \\ C_m=C_{m_0}+C_{m_\alpha}\alpha+C_{m_{\delta_e}}\delta_e+C_{m_{\delta_f}}\delta_f+\dfrac{\bar{c}}{2V}C_{m_q}q \\ C_n=C_{n_\beta}\beta+C_{n_{\delta_a}}\delta_a+C_{n_{\delta_r}}\delta_r+\dfrac{b}{2V}C_{n_p}p+\dfrac{b}{2V}C_{n_r}r \end{cases} \tag{5}$$

式中：p、q、r分别为滚转角速度、俯仰角速度、偏航角速度；δ_a、δ_e、δ_r、δ_f分别为副翼、升降舵、方向舵、襟翼的舵偏角，气动导数由文献[18]给出，不再赘述。

假设飞机受到的合力转化到机体轴之后，分量为F_x、F_y、F_z，得到力方程组为

$$\begin{cases} \dot{u}=vr-wq-g\sin\theta+\dfrac{F_x}{m} \\ \dot{v}=wq-ru+g\cos\theta\sin\phi+\dfrac{F_y}{m} \\ \dot{w}=uq-pv+g\cos\theta\cos\phi+\dfrac{F_z}{m} \end{cases} \tag{6}$$

式中：u、v、w为3轴机体轴速度分量；m为舰载机的质量。

由于舰载机关于$O_bx_bz_b$平面对称，$I_{xy}=I_{yx}=I_{yz}=I_{zy}=0$，其他转动惯量为固定常量，已在表1中给出，可以得到舰载机的角动力学模型如下：

$$\begin{cases} \dot{p}=(c_1r+c_2p)q+c_3\bar{L}+c_4N \\ \dot{q}=c_5pr-c_6\left(p^2-r^2\right)q+c_7M \\ \dot{r}=(c_8p+c_2r)q+c_4\bar{L}+c_9N \end{cases} \tag{7}$$

式中：$c_1=\left(I_{yy}-I_{zz}\right)I_{zz}-I_{xz}^2\Big/\sum$；$c_2=\left(I_{xx}-I_{yy}+I_{zz}\right)I_{xz}\Big/\sum$；$c_3=I_{zz}\Big/\sum$；$c_4=I_{xz}\Big/\sum$；$c_5=I_{zz}-I_{xx}/I_{yy}$；$c_6=I_{xz}/I_{yy}$；$c_7=1/I_{yy}$；$c_8=I_{xx}\left(I_{xx}-I_{yy}\right)+I_{xz}^2\Big/\sum$；$c_9=I_{xx}\Big/\sum$，$\sum=I_{xx}I_{zz}-I_{xz}^2$。

根据舰载机的3个角速度分量(p,q,r)及其姿态角速度的关系，可以得到其运动方程组：

$$\begin{cases} \dot{\phi}=p+(r\cos\phi+q\sin\phi)\tan\theta \\ \dot{\theta}=q\cos\phi-r\sin\phi \\ \dot{\psi}=\dfrac{1}{\cos\theta}(r\cos\phi+q\sin\phi) \end{cases} \tag{8}$$

式中：$\dot{\psi}$、$\dot{\theta}$、$\dot{\phi}$为舰载机的姿态角速度。

将舰载机的机体坐标系中的状态量转化到气流坐标系中，可以得到舰载机相对气流的速度和角

度状态，如下：

$$\begin{cases} V=\sqrt{u^2+v^2+w^2} \\ \beta=\arcsin(v/V) \\ \alpha=\arctan(w/u) \end{cases} \tag{9}$$

将舰载机在机体坐标系中的线运动转化到地面坐标系中，可以得到其导航方程组，如下：

$$\begin{cases} \dot{x}_g=u\cos\theta\sin\psi+v(\sin\phi\sin\theta\sin\psi+\cos\phi\cos\psi)+ \\ \qquad w(-\sin\phi\cos\psi+\cos\phi\sin\theta\sin\psi) \\ \dot{y}_g=u\cos\theta\sin\psi+v(\sin\phi\sin\theta\sin\psi+\cos\phi\cos\psi)+ \\ \qquad w(-\sin\phi\cos\psi+\cos\phi\sin\theta\sin\psi) \\ \dot{h}=u\sin\theta-v\sin\phi\cos\theta-w\cos\phi\cos\theta \end{cases} \tag{10}$$

式中：$\dot{x}_g$、$\dot{y}_g$、$\dot{h}$为地面坐标系 3 个轴上的线速度。

根据上述分析，可以确定舰载机的纵向控制输入为 $u=[\delta_e,\delta_f,\delta_T]^T$，纵向状态向量为 $x=[V,\alpha,q,\theta,h,\gamma,\dot{\gamma}]^T$。

1.2　甲板运动模型建立

为使后续的着舰仿真更加真实，需要建立合理的甲板运动模型。本文采用文献 [19] 提出的基于 Conolly 线性理论的甲板运动模型。由于航母尺寸和质量很大，且舰载机着舰时间较短，甲板运动可以看作一个平稳的随机过程。考虑俯仰、滚转、沉浮运动对着舰的影响，将白噪声通过成型滤波器生成 3 种运动模型，如下：

$$\begin{cases} \dfrac{h_S}{N}=\dfrac{1.16s(s+0.04)}{(s^2+0.16s+0.16)(s^2+0.22s+0.3025)} \\ \dfrac{\theta_S}{N}=\dfrac{0.334\,059s}{(s^2+0.22s+0.302\,5)(s^2+0.384s+0.409\,6)} \\ \dfrac{\varphi_S}{N}=\dfrac{0.238\,350s}{(s^2+0.088\,8s+0.139\,6)(s^2+0.12s+0.25)} \end{cases} \tag{11}$$

式中：h_S、θ_S、ϕ_S分别为甲板的沉浮、俯仰、滚转运动；N为白噪声。

根据式 (11)，设定船速为 15.43 m/s 可以得到甲板 3 种运动的仿真波形，分别如图 1～图 3 所示。

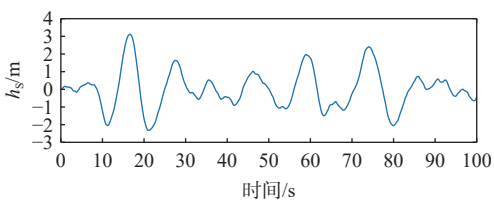


图 1　甲板沉浮运动波形

Fig. 1　Deck ups and downs movement

垂直方向上的高度变化是影响着舰点误差的主要因素，图 4 为理想着舰点垂直高度的变化（将俯仰、滚转运动全部折算成理想着舰点的高度变化）。

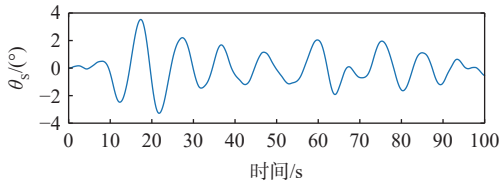


图 2　甲板俯仰运动波形

Fig. 2　Deck pitch movement

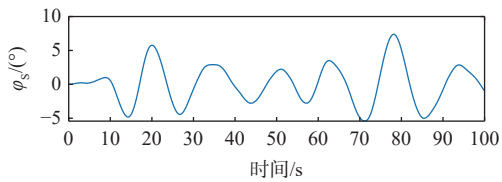


图 3　甲板滚转运动波形

Fig. 3　Deck roll movement

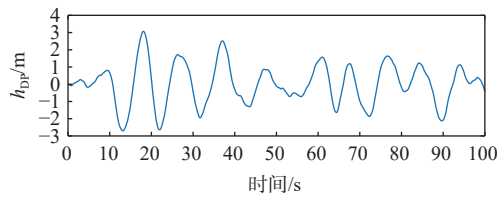


图 4　理想着舰点垂直高度变化波形

Fig. 4　Vertical height variation of ideal landing Point

为减小由甲板运动引起的着舰误差，在距舰 800 m 时需要引入甲板运动补偿器。甲板运动补偿器是一个超前滞后网络，能够补偿系统在 $\omega_s=0.2\sim0.7$ rad/s频段的相位滞后和幅值衰减，使飞机能够实时跟踪理想着舰点的垂直高度变化。甲板运动补偿器为

$$G_{DMC}(s)=66\times\left[\frac{s+3}{s+7}\right]\left[\frac{0.2s^2+0.7s+1}{s^3+9s^2+27s+27}\right] \tag{12}$$

1.3　舰尾流模型建立

本文采用文献 [20] 中的舰尾流模型，根据 MIL-F-8785C 军用规范，舰载机在x方向上距舰大约 800 m 时，开始受到舰尾流干扰，舰尾流主要由 4 个分量组成：①自由大气紊流分量u_1、v_1、w_1；②雄鸡尾流u_2、w_2；③尾流的周期性分量u_3、w_3；④尾流的随机分量u_4、v_4、w_4。

假设u_g、v_g、w_g分别表示 3 轴方向上受到的舰尾气流，表示为

$$\begin{cases} u_g=u_1+u_2+u_3+u_4 \\ v_g=v_1+v_4 \\ w_g=w_1+w_2+w_3+w_4 \end{cases} \tag{13}$$

根据舰尾流模型，设定航母船速 15.43 m/s，得到 3 级海况下，舰尾流的仿真波形，如图 5 所示。

舰尾流对舰载机着舰过程的影响主要体现在空速、迎角以及侧滑角上，从而对舰载机的姿态和

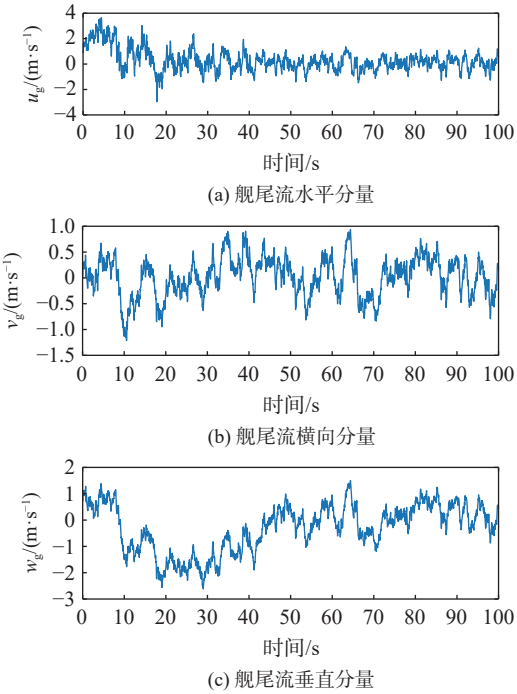


图5 舰尾流波形

Fig. 5 Airwake of the ship

航迹产生影响。将舰尾流3轴分量$[u_g, v_g, w_g]^T$经过多次坐标转换，将其转换到机体坐标系中，得到$[u'_g, v'_g, w'_g]^T$，将其与无扰动情况下的飞机线速度$[u, v, w]^T$相加得到舰尾流扰动下的机体速度$[u_{bg}, v_{bg}, w_{bg}]^T$，那么舰尾流扰动下的空速、迎角以及侧滑角如下：

$$\begin{cases} V' = \sqrt{u_{bg}^2 + v_{bg}^2 + w_{bg}^2} \\ \beta' = \arcsin\left(v_{bg} / V'\right) \\ \alpha' = \arctan\left(w_{bg} / u_{bg}\right) \end{cases} \tag{14}$$

2 PPO算法

PPO算法基于Actor-Critic框架，Actor网络输出动作而Critic网络输出状态价值函数。该算法在一块信任区域内更新策略以保证策略性能的安全性。PPO算法作为一种连续状态、连续动作空间的算法，对于舰载机自动着舰的直接升力控制是很适用的。

假设θ_{rl}表示策略$\pi_{\theta_{rl}}$的参数，PPO算法的优化目标为

$$L_{\theta_{rl}}(\theta'_{rl}) = -E_{\pi_{\theta'_{rl}}}\left[\sum_{t=0}^{\infty}\gamma_{rl}^t(\gamma_{rl}V^{\pi_{\theta_{rl}}}(s_{t+1}) - V^{\pi_{\theta_{rl}}}(s_t))\right] + E_{s_t\sim\nu^{\pi_{\theta'_{rl}}}}E_{a\sim\pi_{\theta'_{rl}}(\cdot|s_t)}\left[\frac{\pi_{\theta'_{rl}}(a|s)}{\pi_{\theta_{rl}}(a|s)}A^{\pi_{\theta_{rl}}}(s,a)\right] \tag{15}$$

式中：θ'_{rl}为相对于θ_{rl}来说更好的参数；s为某一状态；a为某一状态下采取的动作；γ_{rl}为折扣因子；$\nu^{\pi}(s) = (1-\gamma_{rl})\sum_{t=0}^{\infty}\gamma_{rl}^t P_t^{\pi}(s)$为状态访问分布，$1-\gamma_{rl}$为用来使概率加和为1的归一化因子，$P_t^{\pi}(s)$为采取策略$\pi$使得智能体在$t$时刻状态为$s$的概率。

PPO算法对上述优化目标采用PPO-截断的方式，在目标函数中进行限制，以保证新旧参数的差距不大，即

$$\arg\max E_{s_t\sim\nu^{\pi_{\theta_k}}}E_{a\sim\pi_{\theta_k}(\cdot|s_t)}\left[\min\left(\frac{\pi_{\theta'_{rl}}(a|s)}{\pi_{\theta_{rl}}(a|s)}A^{\pi_{\theta_k}}(s,a),\ \mathrm{clip}\left(\frac{\pi_{\theta'_{rl}}(a|s)}{\pi_{\theta_{rl}}(a|s)}, 1-\varepsilon, 1+\varepsilon\right)A^{\pi_{\theta_k}}(s,a)\right)\right] \tag{16}$$

式中：$\mathrm{clip}(x,l,u)$表示把x限制在$[l,u]$区间内。式(16)中的ε为一个超参数，表示PPO-截断的范围。

优势函数A_t通过广义优势估计(generalized advantage estimation, GAE)的方法进行计算，优势函数的表达式为

$$A_t = \delta_t + (\lambda\gamma_{rl})\delta_{t+1} + \cdots + (\lambda\gamma_{rl})^k\delta_{t+k} \tag{17}$$

式中：$\delta_t = r_t + \gamma_{rl}V(s_{t+1}) - V(s_t)$为时序差分误差，$V(s_t)$为Critic网络输出的状态价值函数。

PPO算法的伪代码如下：

for episode = 1 → M do:

for actor = 1 → N do:

使用策略$\pi_{\theta_{rl}}$在环境中执行k个时间步

计算优势函数$A_1, A_2, \cdots, A_k$

end for

计算目标函数$L_{\theta_{rl}}(\theta'_{rl})$

更新网络参数$\theta_{rl} \leftarrow \theta'_{rl}$

end for

3 纵向自动着舰控制器设计

舰载机自动着舰采用飞行轨迹增量控制模式。该模式通过控制飞机的航迹倾斜角来改变飞行轨迹，动态响应更快，有利于航迹的精确控制。

3.1 自动着舰期望高度指令的求解

首先设定飞机初始位置到理想着舰点的距离$X_0 = -4\,578.30$ m，对舰下滑道是一条从理想着舰点发出的射线，其角度规定为$\gamma = -3.5°$，初始高度为

$$H_0 = -X_0\tan(-\gamma/57.3) \tag{18}$$

求得初始高度$H_0 = 280$ m。假设无甲板运动的情况下，飞机到理想着舰点x方向的剩余距离为$R_X(t)$，舰载机已经飞过的横向距离为$X_f(t)$，航母行进的距离为$X_S(t)$，那么$R_X(t)$的表达式为

$$R_X(t) = -X_0 - X_f(t) + X_S(t) \tag{19}$$

根据横向剩余距离可以求解舰载机飞行基准

轨迹，如下：

$$h_{\mathrm{REF}}(t)=R_{\mathrm{X}}(t)\cdot\tan(-\gamma) \tag{20}$$

理想着舰点受甲板运动影响，在垂直方向上运动的波形$h_{\mathrm{DP}}(t)$前文已经给出，假设$h_{\mathrm{DP}}(t)$经过甲板运动补偿器之后为$h_{\mathrm{DMC}}(t)$，那么期望高度指令的表达式为

$$H_{\mathrm{c}}(t)=\begin{cases}h_{\mathrm{REF}}(t) & R_{\mathrm{X}}(t)>800\\ h_{\mathrm{REF}}(t)+h_{\mathrm{DMC}}(t) & R_{\mathrm{X}}(t)\leqslant 800\end{cases} \tag{21}$$

3.2　自动着舰基准航迹倾斜角指令的求解

基准航迹倾斜角指令通过几何计算得到。其不同于对舰下滑道的倾斜角。在航母前进过程中，对舰下滑道会随之移动，导致舰载机的实际航迹倾斜角小于对舰下滑道的夹角，几何关系如图 6 所示。

V_{S}表示航母行驶速度矢量，ψ_{d}表示斜甲板的角度，$V_{\mathrm{S}}\cos\psi_{\mathrm{d}}$表示航母行驶速度在斜甲板上的分量。$V_{\mathrm{p}}$为飞机地速，$\gamma_{\mathrm{ref}}$为飞机的基准航迹倾斜角，$\gamma$为对舰下滑道的夹角。根据图 6 的三角关系，可得

$$\frac{\sin(\gamma+\gamma_{\mathrm{ref}})}{\sin\gamma}=\frac{V_{\mathrm{S}}\cos\psi_{\mathrm{d}}}{V_{\mathrm{p}}} \tag{22}$$

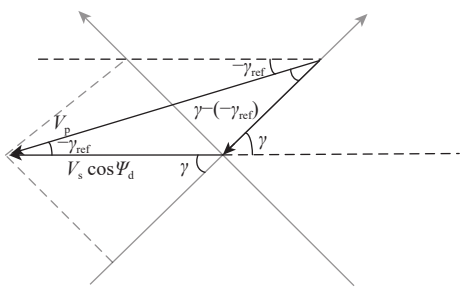


图 6　航迹倾斜角参考指令求解

Fig. 6　The solution graph for γ_{ref}

舰载机在自动着舰状态下，对舰下滑道的夹角γ取为−3.5°，由于γ和γ_{ref}都很小，有

$$\sin(\gamma+\gamma_{\mathrm{ref}})\approx\gamma+\gamma_{\mathrm{ref}}\quad\sin\gamma\approx\gamma \tag{23}$$

代入式 (22)，可得

$$\gamma_{\mathrm{ref}}=\frac{V_{\mathrm{S}}\cos\psi_{\mathrm{d}}-V_{\mathrm{p}}}{V_{\mathrm{p}}}\gamma \tag{24}$$

因此，在飞机地速$V_{\mathrm{p}}=70.016\ 8$ m/s，对舰下滑道的夹角$\gamma=-3.5°$，船速$V_{\mathrm{S}}=15$ m/s，$\psi_{\mathrm{d}}=11°$时，通过式 (24) 可以计算出基准航迹倾斜角指令$\gamma_{\mathrm{ref}}=-2.75°$。

3.3　基于 PID 的直接升力控制器设计

基于 PID 的飞行轨迹增量控制模式的框图如图 7 所示。舰载机通过襟翼的偏转产生直接升力，可以迅速改变航迹倾斜角，实现对高度指令的跟踪。

这里需要注意的是，当襟翼偏转时，会产生俯仰力矩，因此需要升降舵来平衡俯仰力矩，同时保持迎角恒定。

图 7 中，控制系统首先根据高度误差计算航迹倾斜角增量指令，并与航迹倾斜角参考指令进行比较，然后通过控制器得到期望航迹倾斜角速率，再与实际航迹倾斜角速率作差后经过控制器得到襟翼舵偏值。

3.4　基于深度强化学习的直接升力控制器设计

为了降低控制器参数调节与优化的难度以及对数学模型精度的依赖性，本文提出了一种基于 PPO 算法的纵向直接升力控制器，控制框架如图 8 所示。

将舰载机看作待训练的智能体，其交互环境包括甲板运动、舰尾流等，飞机模型以及交互环境均在 MATLAB/Simulink 中搭建。

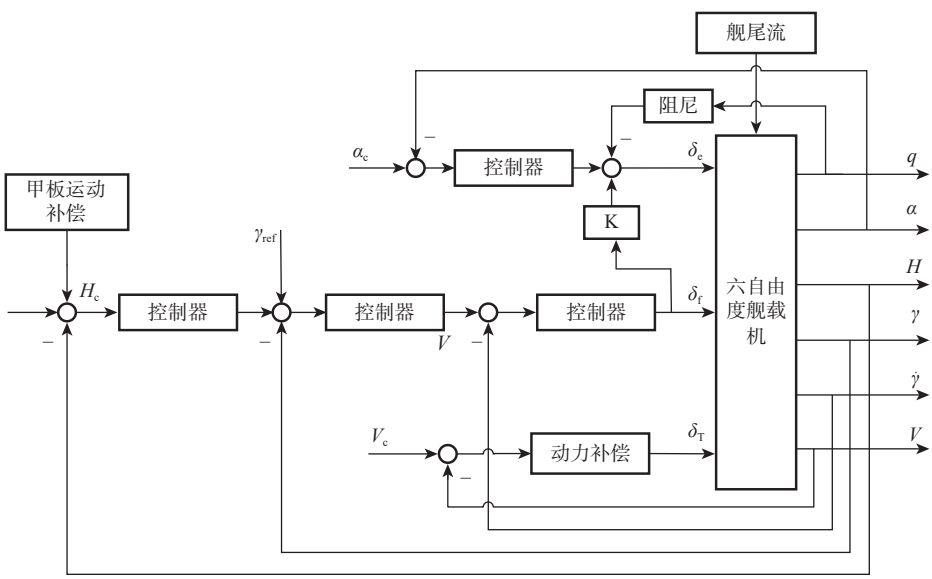


图 7　基于 PID 的直接升力控制框图

Fig. 7　Traditional direct lift control block diagram

在强化学习理论中，舰载机在着舰过程中的状态转移可以看作马尔可夫决策过程(Markov decision process, MDP)，即舰载机从一种状态转移到下一种状态的概率只与当前状态和所采取的动作有关，与之前的历史状态和动作无关，如下：

$$P(S_{t+1}) = E[S_{t+1} = s_{t+1} | S_t = s_t, A_t = a_t] \tag{25}$$

式中：s_t为飞机某一时刻的状态；s_{t+1}为飞机下一时刻的状态；a_t为飞机某一时刻采取的动作。飞机在t时刻采取动作a_t后会获得相应的奖励，如下：

$$r_t = r(s_t, a_t) \tag{26}$$

根据PPO算法的流程，飞机智能体与着舰环境交互多次，根据当前状态s_t采取动作a_t，获得奖励r_t并且状态转移至s_{t+1}，将元组(s_t, a_t, r_t, s_{t+1})存放在Mini-Batch中，当元组个数达到一定数量时，批量计算A_t以及目标函数$L_{\theta_{\mathrm{n}}}(\theta'_{\mathrm{n}})$，并更新网络参数。基于强化学习的直接升力控制器训练框图如图9所示。

3.4.1 状态-动作空间设计

本文选取一个6维的状态空间，如下：

$$\boldsymbol{S} = (H, H_{\mathrm{e}}, \theta, \gamma, \dot{\gamma}, q) \tag{27}$$

式中：H为舰载机的高度；$H_{\mathrm{e}} = H_{\mathrm{c}} - H$为实际高度与期望下滑道的纵向偏差；$\theta$为舰载机的俯仰角；$\gamma$为舰载机的航迹倾斜角；$\dot{\gamma}$为舰载机的航机倾斜角速率；$q$为舰载机的俯仰角速度。

动作空间的维度为1，为飞机襟翼的偏转，如下：

$$\boldsymbol{a} = (\delta_{\mathrm{f}}) \tag{28}$$

3.4.2 奖励函数设计

奖励函数在智能体训练的过程中起着至关重要的作用，合理的奖励函数可以有效地引导智能体进行学习。

奖励函数分为连续型和离散型，连续型奖励函数计算量较大但可以提高训练过程的收敛性，对网

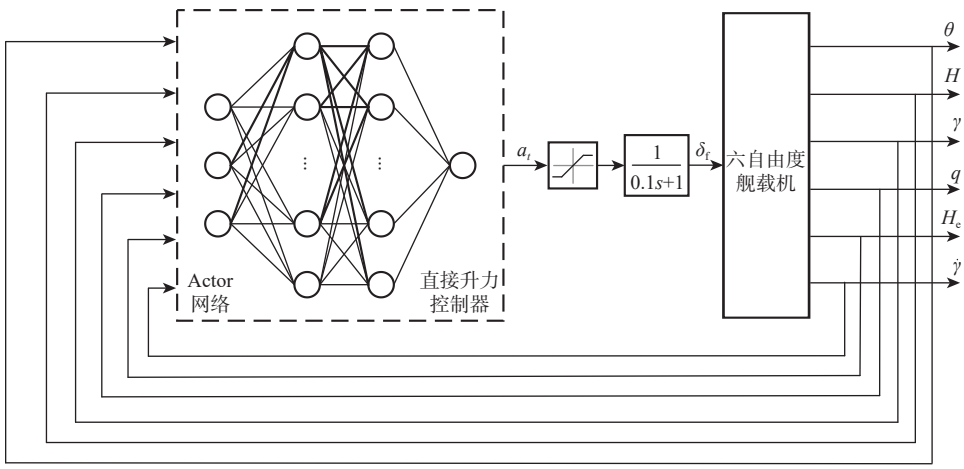


图8 基于深度强化学习的直接升力控制框图

Fig. 8 Deep Reinforcement Learning control block diagram

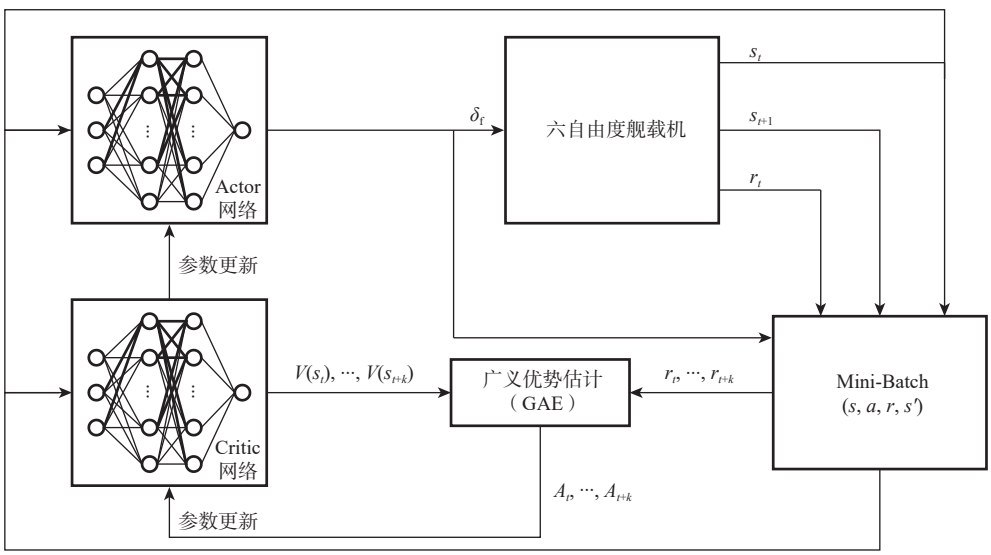


图9 控制器训练框图

Fig. 9 Controller training block diagram

络结构的复杂度较低。而离散型奖励函数由于变化的离散性，使得收敛速度变慢但可以引导智能体在状态空间中找到更好的奖励区域。基于此，本文采用混合型奖励函数，离散函数用于驱动智能体远离不良状态，连续函数通过在目标状态附近提供平滑奖励来提高收敛性。

还有一点值得注意的是，在训练过程中，智能体只会学习如何最大化收益，如果想让其完成某些指定任务，就必须将目标融入到奖励函数中。

高度误差的奖励函数为

$$R_{H_e}=\begin{cases}-0.2H_e^2 & |H_e|>10\\ 63.25\sqrt{\dfrac{1}{|H_e|}} & 0.04<|H_e|\leqslant 10\\ 400 & |H_e|\leqslant 0.04\end{cases} \tag{29}$$

舵面偏转的奖励函数为

$$R_{\delta_f}=-0.5\delta_f^2 \tag{30}$$

航迹倾斜角的奖励函数为

$$R_\gamma=\begin{cases}-(\gamma_{ref}-\gamma)^2 & |\gamma_e|<0.2\\ 15 & |\gamma_e|\geqslant 0.2\end{cases} \tag{31}$$

式中：$\gamma_e=\gamma_{ref}-\gamma$。

俯仰角速度的奖励函数为

$$R_q=-0.2(0-q)^2 \tag{32}$$

训练步进奖励为

$$R_{step}=30 \tag{33}$$

终止训练奖励函数为

$$R_{final}=\begin{cases}10\,000 & H\leqslant 0,0\leqslant H_e<1.8\\ 10\,000 & H_c\leqslant 0,-1\leqslant H_e\leqslant 0\\ -30\,000 & H\leqslant 0,H_e\geqslant 1.8\\ -30\,000 & H_c\leqslant 0,H_e<-1\\ -30\,000 & \theta>20\text{或}\theta<-20\\ -30\,000 & |H_e|\geqslant 25\\ -30\,000 & H>H_0+5\\ -30\,000 & H\leqslant 0,(|H_d|\geqslant 5\text{或}(\theta>10\text{或}\theta\leqslant 0))\end{cases} \tag{34}$$

式中：H_c为期望高度；H_d为舰载机的下沉率；H_0为着舰的初始高度。

总奖励函数的表达式为

$$R_{all}=R_{H_e}+R_{\delta_f}+R_\gamma+R_q+R_{step}+R_{final} \tag{35}$$

3.4.3　超参数设计

PPO 算法中 Actor-Critic 网络的结构以及超参数设计如表 2 所示。

Actor 网络与 Critic 网络均采用全连接结构，隐含层激活用函数为 ReLu 函数。Actor 网络中，动作均值的激活函数为 tanh 函数，动作方差的激活函数为 Softplus 函数，根据概率分布进行高斯采样得到具体的舵面偏转值。Loss function 均采用 Adam 优化器进行梯度更新。

表 2　算法参数设置

Table 2　PPO algorithm parameter setting table

参数	数值
PyCharm平台IP端口	127.0.0.1.500 00(地址)
Simulink平台IP端口	127.0.0.1.500 01(地址)
Actor网络结构	6×128×128×2
Critic网络结构	6×128×128×128×1
Actor网络学习率	1×10^{-3}
Critic网络学习率	5×10^{-3}
回合数	20 000
步数	1 500
γ_{rl}	0.99
λ	0.95
ε	0.2
训练前交互步数	2 048
交叉熵损失函数权重	0.01
小批次数据量	256
平均奖励计算回合数	10
动作界限值	30

3.4.4　控制律设计

假设 Actor 网络输出的动作均值为μ_f，方差为σ_f^2。均值与方差均通过 Actor 网络计算得出。Actor 网络的输入为 6 个状态的列向量$\boldsymbol{x}=\left[H,H_e,\theta,\gamma,\dot{\gamma},q\right]^{\mathrm{T}}$，经过 2 个隐含层之后得到输出$\boldsymbol{y}=[\mu_f,\sigma_f^2]^{\mathrm{T}}$，约定：

1）变量z_i^l为第l层的第i个神经元，变量a_i^l为经过激活函数后的第l层的第i个神经元。

2）$\boldsymbol{W}^l$表示权重矩阵，权重$w_{i,j}^l$表示第$l-1$层的第j个神经元与第l层第i个神经元之间的权重。

Actor 网络第 1 个隐含层的输入$\boldsymbol{z}^1=[z_1^1,z_2^1,\cdots,z_{127}^1,z_{128}^1]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{z}^1=\boldsymbol{W}^1\boldsymbol{x}=\begin{cases}z_1^1=w_{1,1}^1H+w_{1,2}^1H_e+w_{1,3}^1\theta+w_{1,4}^1\gamma+\\ \qquad w_{1,5}^1\dot{\gamma}+w_{1,6}^1q\\ z_2^1=w_{2,1}^1H+w_{2,2}^1H_e+w_{2,3}^1\theta+w_{2,4}^1\gamma+\\ \qquad w_{2,5}^1\dot{\gamma}+w_{2,6}^1q\\ \vdots\\ z_{128}^1=w_{128,1}^1H+w_{128,2}^1H_e+w_{128,3}^1\theta+\\ \qquad w_{128,4}^1\gamma+w_{128,5}^1\dot{\gamma}+w_{128,6}^1q\end{cases} \tag{36}$$

Actor 网络第 1 个隐含层的输出$\boldsymbol{a}^1=[a_1^1,a_2^1,\cdots,a_{127}^1,a_{128}^1]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{a}^1=f_{\mathrm{ReLU}}(\boldsymbol{z}^1)=\begin{cases}a_1^1=f_{\mathrm{ReLU}}(z_1^1)\\ a_2^1=f_{\mathrm{ReLU}}(z_2^1)\\ \vdots\\ a_{128}^1=f_{\mathrm{ReLU}}(z_{128}^1)\end{cases} \tag{37}$$

Actor网络第2个隐含层的输入$\boldsymbol{z}^2=[z_1^2,z_2^2,\cdots,z_{127}^2,z_{128}^2]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{z}^2=\boldsymbol{W}^2\boldsymbol{a}^1=\begin{cases}w_{1,1}^2a_1^1+w_{1,2}^2a_2^1+\cdots+w_{1,127}^2a_{127}^1+\\ \quad w_{1,128}^2a_{128}^1\\ w_{2,1}^2a_1^1+w_{2,2}^2a_2^1+\cdots+w_{2,127}^2a_{127}^1+\\ \quad w_{2,128}^2a_{128}^1\\ \vdots\\ w_{128,1}^2a_1^1+w_{128,2}^2a_2^1+\cdots+w_{128,127}^2a_{127}^1+\\ \quad w_{128,128}^2a_{128}^1\end{cases} \tag{38}$$

Actor网络第2个隐含层的输出$\boldsymbol{a}^2=[a_1^2,a_2^2,\cdots,a_{127}^2,a_{128}^2]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{a}^2=f_{\mathrm{ReLU}}(\boldsymbol{z}^2)=\begin{cases}a_1^2=f_{\mathrm{ReLU}}(z_1^2)\\ a_2^2=f_{\mathrm{ReLU}}(z_2^2)\\ \vdots\\ a_{128}^2=f_{\mathrm{ReLU}}(z_{128}^2)\end{cases} \tag{39}$$

Actor网络输出层的输入$\boldsymbol{z}^3=[z_1^3,z_2^3]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{z}^3=\boldsymbol{W}^3\boldsymbol{a}^2=\begin{cases}w_{1,1}^3a_1^2+w_{1,2}^3a_2^2+\cdots+w_{1,127}^3a_{127}^2+w_{1,128}^3a_{128}^2\\ w_{2,1}^3a_1^2+w_{2,2}^3a_2^2+\cdots+w_{2,127}^3a_{127}^2+w_{2,128}^3a_{128}^2\end{cases} \tag{40}$$

Actor网络输出层的输出$\boldsymbol{y}=[\mu_{\mathrm{f}},\sigma_{\mathrm{f}}^2]^{\mathrm{T}}$，计算表达式为

$$\boldsymbol{y}=\begin{bmatrix}\mu_{\mathrm{f}}\\ \sigma_{\mathrm{f}}^2\end{bmatrix}=\begin{cases}f_{\tanh}(z_1^3)\\ f_{\mathrm{Softplus}}(z_2^3)\end{cases} \tag{41}$$

在某一时刻，智能体的状态到达s，此状态下动作a服从均值为μ_{f}、方差为σ_{f}^2的正态分布：

$$a\sim N(\mu_{\mathrm{f}},\sigma_{\mathrm{f}}^2) \tag{42}$$

则动作a的概率密度为

$$f(a)=\frac{1}{\sqrt{2\pi}\sigma_{\mathrm{f}}}\mathrm{e}^{-\frac{(a-\mu_{\mathrm{f}})^2}{2\sigma_{\mathrm{f}}^2}} \tag{43}$$

襟翼偏转的表达式为

$$\delta_{\mathrm{f}}=\delta_{\mathrm{f}_0}+\mathrm{sample}\left[N(\mu_{\mathrm{f}},\sigma_{\mathrm{f}}^2)\right] \tag{44}$$

式中：δ_{f_0}为襟翼配平值；sample为高斯采样函数，即依照概率进行动作的选取。也就是说，某个具体动作的概率密度越大，被选中的概率就越大。在训练结束后，方差σ_{f}^2很小，那么随机选取的动作将十分逼近均值μ_{f}。

纵向的升降舵及油门的控制仍采用经典的PID控制器，控制律为

$$\delta_{\mathrm{e}}=\delta_{\mathrm{e}_0}+K_qq+\left(K_p+\frac{1}{s}K_i\right)(\alpha_{\mathrm{c}}-\alpha)+K\Delta\delta_{\mathrm{f}} \tag{45}$$

$$\delta_{\mathrm{T}}=\delta_{\mathrm{T}_0}+\left(K_p+\frac{1}{s}K_i\right)(V_{\mathrm{c}}-V) \tag{46}$$

式中：δ_{e_0}为升降舵配平值；$K=-M_{\delta_{\mathrm{f}}}/M_{\delta_{\mathrm{e}}}$为襟翼与升降舵的解耦信号；$M_{\delta_{\mathrm{f}}}$和$M_{\delta_{\mathrm{e}}}$分别为襟翼和升降舵的俯仰力矩系数；$\delta_{\mathrm{T}_0}$为油门开度的配平值。

4 仿真验证

本文的环境在Simulink中搭建，包括舰载机的动力学/运动学模型、着舰引导律、甲板运动模型、舰尾流模型等，飞机参数使用F/A-18的数据，训练框架使用Python3.8.10与PyTorch1.13.0。基于深度强化学习的直接升力控制器的回合奖励曲线如图10所示。

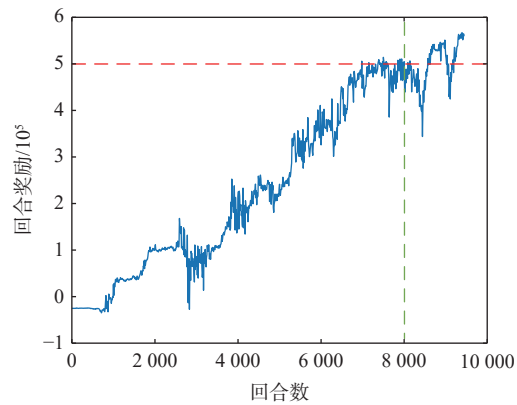


图10 回合奖励收敛曲线

Fig. 10 Episode reward curve

从图10中可以看出，训练8 000回合后，回合奖励基本已经维持在500 000以上并逐渐开始收敛，在9 440回合左右，达到540 000的收敛均值。控制器的平均奖励曲线如图11所示。

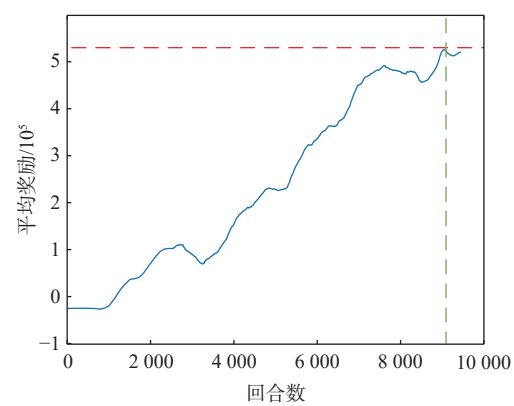


图11 平均奖励收敛曲线

Fig. 11 Average reward curve

从图11中可以看出，训练到9 440回合时，奖励均值(取训练过程中最新50个episodes的奖励求平均)已经达到540 000的收敛值。图12为下滑道跟踪曲线，图13为下滑道跟踪误差曲线。

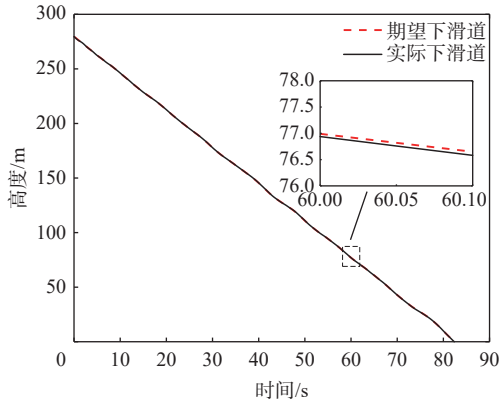


图 12　下滑道跟踪曲线

Fig. 12　Glide track curve

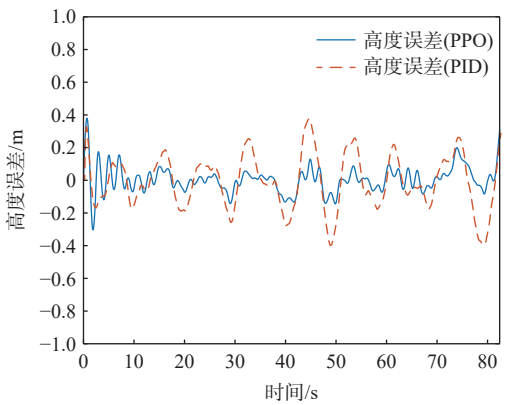


图 13　高度误差曲线

Fig. 13　Height error curve

本文设置的着舰的初始高度 $H_0 = 280$ m，图 12 中红色虚线为期望下滑道，黑色实线为实际下滑道。根据图中的细节可以看出在 PPO 控制器的作用下，下滑道有 0.05 m 左右的误差。通过图 13 的对比可以明显看出，PPO 控制器作用下的高度误差比 PID 控制器更小，对于着舰过程中舰尾流以及甲板运动的影响有更强的抑制作用。图 14 为 PPO 控制器作用下的襟翼偏转曲线图。

本文采用的是增量控制，即舵偏值为减去配平数值后的大小。从图 14 中可以看出，对于 PPO 控制器，开始下滑的前 10 s 为了快速跟踪标准下滑道，襟翼偏转角度较大，达到 20°。10 s 之后趋于稳定，在 0 值附近有一些微调。对于 PID 控制器，稳定后也处于 0 值附近，但在着舰后期，受甲板运动与舰尾流的影响，襟翼舵偏幅度比 PPO 控制器作用下的襟翼舵偏幅度大。图 15 为不同控制器作用下的航迹倾斜角曲线。

在着舰的前 10 s，PID 控制器作用下的航迹倾斜角调节幅度比 PPO 控制器更小，10 s 之后，PID 控制器对−2.75°的角度保持效果更差。在舰载机距舰 800 m 时，受到甲板运动以及舰尾流的影响，

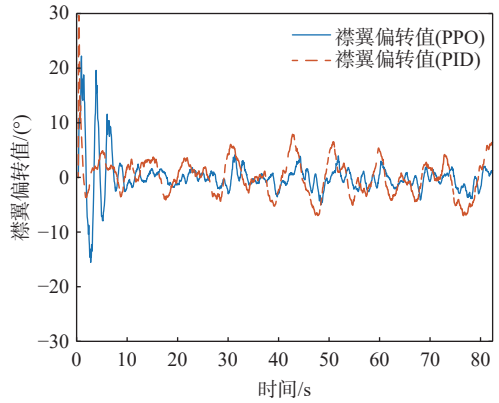


图 14　襟翼偏转曲线

Fig. 14　Flap deflection curve

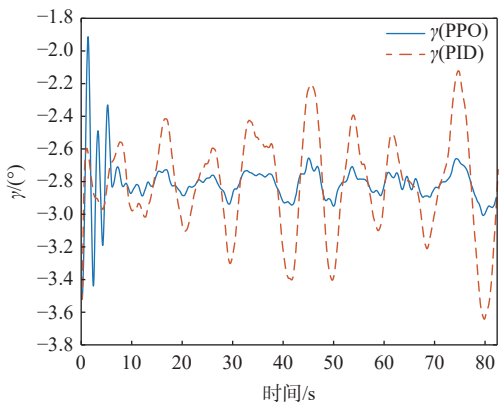


图 15　航迹倾斜角曲线

Fig. 15　Flight path angle

PID 控制器作用下的航迹倾斜角出现了更大幅度的抖动。

舰载机的迎角曲线如图 16 所示。

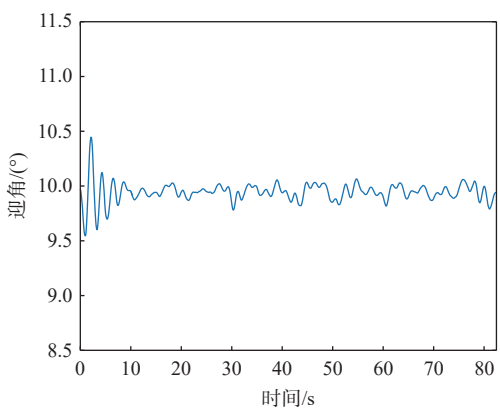


图 16　迎角曲线

Fig. 16　Attack angle curve

从图 16 中可以看出，在升降舵的作用下，迎角基本保持在 9.928 9°的期望值附近。

舰载机的空速响应如图 17 所示。

从图 17 中可以看出，在油门的控制作用下，飞机空速保持在70 m/s的期望值附近。

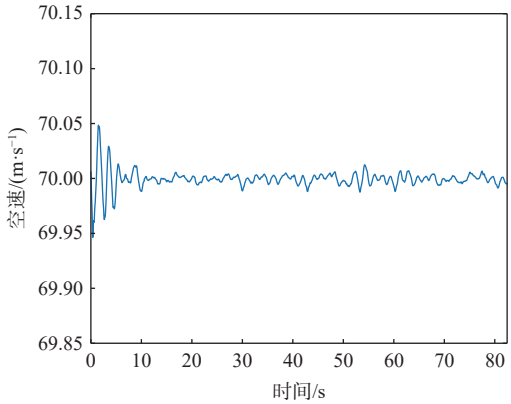


图 17 空速曲线

Fig. 17 Airspeed curve

5 结 论

1）训练收敛后的PPO控制器，实现了自动着舰的精准控制，能够较好地跟踪期望下滑道，最大跟踪误差基本控制在±0.2 m以内。

2）PPO控制器对舰尾流有更强的抑制作用，在距舰800 m，受舰尾流干扰时，航迹倾斜角的抖动幅度更小，控制在(−3,−2.6)的区间内。

3）在PPO算法的作用下，控制器的调节时间更短，从高度误差、航迹倾斜角的结果中均可看出，比起PID控制器调节时间缩短了2 s左右。

4）PPO算法作用下的襟翼舵面偏转幅度更小，稳定后舵偏角度控制在5°以内，这对舵面有更好的保护作用。

为使PPO算法对自动着舰具有更好的控制效果，仍需继续优化奖励函数以及算法的超参数，在提高训练收敛速度的同时提高控制精度。同时，可以考虑将升降舵以及油门开度的控制也替换为强化学习的方法。

参考文献（References）

[1] 张守权, 王华明. 舰载机全自动着舰综述[J]. 飞机设计, 2022, 42(3): 20-24.
ZHANG S Q, WANG H M. Summary report on automatic carrier landing system[J]. Aircraft Design, 2022, 42(3): 20-24 (in Chinese).

[2] 甄子洋, 王新华, 江驹, 等. 舰载机自动着舰引导与控制研究进展[J]. 航空学报, 2017, 38(2): 020435.
ZHEN Z Y, WANG X H, JIANG J, et al. Research progress in guidance and control of automatic carrier landing of carrier-based aircraft[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(2): 020435 (in Chinese).

[3] 张守权. 基于直接力控制的人工着舰技术综述[J]. 飞机设计, 2022, 42(2): 21-25.
ZHANG S Q. A review of manual carrier landing technology based on direct force control[J]. Aircraft Design, 2022, 42(2): 21-25 (in Chinese).

[4] WU W H, SONG L T, ZHANG Y, et al. Nonlinear comprehensive decoupling controller based on direct lift control for carrier landing[J]. IEEE Access, 2022, 10: 113875-113887.

[5] YAN Y D, YANG J, LIU C J, et al. On the actuator dynamics of dynamic control allocation for a small fixed-wing UAV with direct lift control[J]. IEEE Transactions on Control Systems Technology, 2020, 28(3): 984-991.

[6] GUAN Z Y, LIU H, ZHENG Z W, et al. Moving path following with integrated direct lift control for carrier landing[J]. Aerospace Science and Technology, 2022, 120: 107247.

[7] 罗飞, 张军红, 耿延升, 等. 动态逆反馈控制框架下直接升力控制的控制分配研究[J]. 航空科学技术, 2022, 33(8): 51-60.
LUO F, ZHANG J H, GENG Y S, et al. Study on control allocation technology of direct lift control under dynamic inversion feedback control framework[J]. Aeronautical Science & Technology, 2022, 33(8): 51-60 (in Chinese).

[8] 魏毅寅, 郝明瑞, 范宇. 人工智能技术在宽域飞行器控制中的应用[J]. 宇航学报, 2023, 44(4): 530-537.
WEI Y Y, HAO M R, FAN Y. The application of artificial intelligence technology in wide-field vehicle control[J]. Journal of Astronautics, 2023, 44(4): 530-537 (in Chinese).

[9] 孙智孝, 杨晟琦, 朴海音, 等. 未来智能空战发展综述[J]. 航空学报, 2021, 42(8): 525799.
SUN Z X, YANG S Q, PIAO H Y, et al. A survey of air combat artificial intelligence[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(8): 525799 (in Chinese).

[10] 付宇鹏, 邓向阳, 何明, 等. 基于强化学习的固定翼飞机姿态控制方法[J]. 控制与决策, 2023, 38(9): 2505-2510.
FU Y P, DENG X Y, HE M, et al. Reinforcement learning based attitude controller design[J]. Control and Decision, 2023, 38(9): 2505-2510 (in Chinese).

[11] 张瑞卿, 钟睿, 徐毅. 基于强化学习的航天器姿态控制器设计[J]. 上海航天 (中英文), 2023, 40(1): 80-85.
ZHANG R Q, ZHONG R, XU Y. Satellite attitude control based on reinforcement learning method[J]. Aerospace Shanghai (Chinese & English), 2023, 40(1): 80-85 (in Chinese).

[12] 金磊, 杨绍龙. 基于强化学习的航天器姿态预设性能容错控制[J]. 北京航空航天大学学报, 2024, 50(8): 2404-2412.
JIN L, YANG S L. Fault-tolerant control of spacecraft attitude with prescribed performance based on reinforcement learning[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(8): 2404-2412 (in Chinese).

[13] 付宇鹏, 邓向阳, 朱子强, 等. 基于模仿强化学习的固定翼飞机姿态控制器[J]. 海军航空大学学报, 2022, 37(5): 393-399.
FU Y P, DENG X Y, ZHU Z Q, et al. Imitation reinforcement learning based attitude controller for fixed-wing aircraft[J]. Journal of Naval Aviation University, 2022, 37(5): 393-399 (in Chinese).

[14] 周攀, 黄江涛, 章胜, 等. 基于深度强化学习的智能空战决策与仿真[J]. 航空学报, 2023, 44(4): 126731.
ZHOU P, HUANG J T, ZHANG S, et al. Intelligent air combat decision making and simulation based on deep reinforcement learning[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(4): 126731 (in Chinese).

[15] 黄江涛, 刘刚, 周攀, 等. 基于深度强化学习技术的舰载无人机自主着舰控制研究[J]. 南京师范大学学报 (工程技术版), 2022,

22(3): 63-71.
HUANG J T, LIU G, ZHOU P, et al. Research on autonomous landing control of carrier-borne UCAV based on deep reinforcement learning technology[J]. Journal of Nanjing Normal University (Engineering and Technology Edition), 2022, 22(3): 63-71 (in Chinese).

[16] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimizationalgorithms[J/OL].(2017-08-28)[2023-06-14].https://doi.org/10.48550/arXiv.1707.06347.

[17] GU Y, CHENG Y H, YU K, et al. Anti-martingale proximal policy optimization[J]. IEEE Transactions on Cybernetics, 2023, 53(10): 6421-6432.

[18] CHAKRABORTY A, SEILER P, BALAS G J. Susceptibility of F/A-18 flight controllers to the falling-leaf mode: linear analysis[J]. Journal of Guidance, Control, and Dynamics, 2011, 34(1): 57-72.

[19] 张永花. 舰载机着舰过程甲板运动建模及补偿技术研究[D]. 南京: 南京航空航天大学, 2012: 9-10.
ZHANG Y H. Research on deck motion modeling and compensation technology of carrier-based aircraft landing process[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012: 9-10 (in Chinese).

[20] WOODCPCK T J. Background information and user guide for MIL-F-8785C[R]. Washington, D. C. : Air Force Wright Aeronautical, 1982.

Direct lift control technology of carrier aircraft landing based on reinforcement learning

LIU Rendi，JIANG Ju*，ZHANG Zhe，LIU Xiang

(College of Automation Engineering，Nanjing University of Aeronautics and Astronautics，Nanjing 211106，China)

Abstract：The direct lift control method of automatic landing based on Proximal Policy Optimization (PPO) algorithm was proposed to solve the problem that it is easy to touch ship due to disturbance of deck movement and carrier air wake during automatic landing of carrier aircraft. The PPO controller takes six state variables of pitch angle, height, flight path angle, pitch angle rate, height error and flight path angle rate as input and output as flap deflection angle, realizing the rapid response of carrier aircraft in different landing states of flight path angle. Compared with traditional PID controller, the Actor-Critic network in PPO controller greatly improves the calculation efficiency of control quantity, and also reduces the difficulty of parameter optimization. The simulation experiment in this paper is based on the dynamics/kinematics model of F/A-18 aircraft constructed in Matlab/Simulink. The intensive learning and training environment built on PyCharm platform is used to realize the data interaction between the two platforms through user datagram protocol (UDP) communication. The simulation results show that the proposed method has the characteristics of fast response speed and small dynamic error, and can stabilize the landing height error within ±0.2 m, with high control accuracy.

Keywords：carrier aircraft landing; reinforcement learning; proximal policy optimization algorithm; direct lift control; UDP communication

Received：2023-06-21；**Accepted**：2023-11-10；**Published Online**：2023-11-24 17:51
URL：link.cnki.net/urlid/11.2625.V.20231123.1537.005

* **Corresponding author**. E-mail：jiangju@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0409

基于ISCA-DBN的飞机地面空调能耗预测

刘涵，林家泉*

(中国民航大学 电子信息与自动化学院，天津 300300)

摘 要：为提升飞机客舱使用地面空调制冷时地面空调能耗预测精度，提出一种改进正余弦算法(ISCA)优化深度置信网络(DBN)的地面空调能耗预测模型。与标准正余弦优化算法相比，ISCA提出一种改进Logistic混沌映射，提高了种群多样性；引入余弦调节因子，构建了一种新的非线性振荡调整因子，以平衡算法的全局搜索和局部寻优能力；基于变异进化思想提出一种学习策略，避免算法陷入局部最优。将ISCA-DBN模型应用于波音737-800飞机地面空调能耗预测中，与反向传播(BP)、支持向量机(SVM)、DBN等算法进行性能对比，仿真结果表明：基于ISCA-DBN的地面空调能耗预测模型在预测精度和实时性上有一定的提升。

关 键 词：飞机客舱；地面空调；能耗预测；正余弦优化；深度置信网络

中图分类号：V351.3

文献标志码：A **文章编号**：1001-5965(2025)06-2176-09

在航空领域，航空煤油的燃烧是主要的环境污染源之一。为了减少航空煤油燃烧对环境造成的影响，中国民用机场采取了一项重要的措施，即使用地面专用设备代替飞机上的辅助动力装置(auxiliary power unit, APU)。通过这种方式，飞机在停靠期间可以使用地面设备提供的电力，从而减少航空煤油的使用。这一举措不仅可以减少空气中的污染物排放，还可以节约燃料和降低航空公司的运营成本。然而，要实现有效的绿色减排，仅仅依靠地面专用设备是不够的。为了更好地规划和管理机场的能源消耗，设计准确的地面空调能耗预测模型至关重要。

然而，地面空调能耗数据的非线性、非平稳和多季节性[1]，以及其对天气条件、客舱内环境等因素的依赖性，使得准确预测能耗成为一项具有挑战性的任务。王修岩等[2]提出使用广义回归神经网络(generalized regression neural network, GRNN)来构建飞机客舱能耗预测模型，并使用协同粒子群算法对模型参数进行优化，提高了模型的预测精度，但所需的训练时间较长，无法满足实时预测的需求；林家泉等[3]提出使用具有短期记忆功能的Elman神经网络建立飞机地面空调能耗模型，用以准确预测出能耗并缩短预测耗时，但该模型的网络结构单一，在不同任务场景下模型的泛化性能较弱；周璇等[4]提出使用改进长短时记忆(long short-term memory, LSTM)神经网络对飞机地面空调能耗进行预测，使其能够有效利用数据中的时间规律，仿真结果表明该算法有较高的预测精度，但其计算量大，同样难以满足实时性的需求。以上关于飞机地面空调能耗预测的研究具有一定价值，但泛化性和实时性仍有提升空间[5]。

Hinton等[6]提出深度置信网络(deep belief network, DBN)具有较强的数据分析和特征提取的能力。DBN是深度学习理论中的经典算法，因其优异的特征提取和训练方法，成功解决了诸如电力分析和故障诊断[7-8]等问题。DBN的预训练过程采用无监督学习的方法，不依赖于标记样本。这意味着在预训练过程中，DBN可以从未标记数据中自动提取特征，而不需要人工标注样本，从而减少了人为因素带来的不确定性[9]。此外，DBN具有广泛的适

收稿日期：2023-06-23；**录用日期**：2023-08-14；**网络出版时间**：2023-08-25 12:10
网络出版地址：link.cnki.net/urlid/11.2625.V.20230824.1900.001
*通信作者. E-mail：jqlin@cauc.edu.cn

引用格式：*刘涵，林家泉. 基于ISCA-DBN的飞机地面空调能耗预测[J]. 北京航空航天大学学报，2025，51(6)：2176-2184.*
LIU H，LIN J Q. Energy consumption prediction of aircraft ground air conditioning based on ISCA-DBN[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(6)：2176-2184 (in Chinese).

应性和强大的映射能力, 能够近似任意连续函数, 并且常被用于模拟多变量非线性系统。

梁智等[10]提出一种基于变分模态分解技术优化 DBN 网络的短期负荷预测方法, 仿真结果表明该方法在电力系统负荷预测上具有较好的性能; 徐先峰等[11]提出一种融合多层感知机和 DBN 网络的模型, 对光伏发电进行预测, 提高了光伏发电的预测精度。然而 DBN 网络的结构配置对模型的预测精度具有重大的影响[12], 因此选取一种智能算法来提高 DBN 的性能十分重要。Shan 等[13]使用粒子群优化 (particle swarm optimization, PSO) 算法优化 DBN 网络参数, 对配电网可靠性进行分析, 仿真结果的评价指标表明该模型具有更高的准确性; 李益兵等[12]提出一种基于 PSO-DBN 模型的轴承故障诊断模型, 结果表明该模型相较于标准 DBN 模型具有更高的准确性。但粒子群算法需要设置参数较多, 精度和收敛性容易受到内部参数的影响, 存在稳定性不佳的问题。正余弦算法 (sine cosine algorithm, SCA) 是一种新型的随机优化算法, 由 Mirjalili 于 2016 年提出[14]。该方法具有初始参数少、结构简单、易于实现等优点[15], 且已有研究证明[16], 其优化性能均优于遗传算法 (genetic algorithm, GA)、粒子群算法 (particle swarm optimiazation, PSO) 和花授粉算法等, 已被广泛用于数据分类、光谱特征峰定位和电力系统调度等不同领域[17-18]。因此, 将 SCA 应用于飞机地面空调能耗预测问题的研究值得深入探索。

结合上述分析, 本文提出一种基于 ISCA 优化的 DBN 算法, 以及基于该算法的飞机地面空调能耗预测模型。首先, 对 DBN 算法采用自适应时刻估计法 (adaptive moment estimation, Adam) 对层间神经元权重进行微调, 提高 DBN 算法的预测精度; 然后, 采用 ISCA 算法对 DBN 算法的重要参数进行学习训练, 将训练得到的最优参数赋给 DBN 算法, 得到 ISCA-DBN 模型。将 ISCA-DBN 模型用于对 Boeing737-800 飞机地面空调能耗预测, 并与反向传播 (back propagation, BP)、支持向量机 (support vector machine, SVM)、DBN、GA-DBN 和 SCA-DBN 等多种模型进行比较, 结果表明, 该模型在均方误差 (mean-square error, MSE)、平均绝对值误差 (mean absolute percentage error, MAPE) 以及决定系数 R^2 上都具有显著的优势。

1　相关理论

1.1　深度置信网络

DBN 是一个具有多个隐藏层的前馈神经网络, 由多个受限玻尔兹曼机 (restricted Bolzmann machine, RBM) 堆叠而成, 低层表示原始数据信号, 高层表示数据的特征[19]。DBN 的内部结构如图 1所示, 该模型由 3 个 RBM 堆叠而成, 单层 RBM 由 2 层网络组成, 分别是可视层 v 和隐藏层 h, 层间神经元通过权重连接, 层内神经元无连接。其运算过程如式 (1)～式 (7) 所示。

$$\boldsymbol{\theta}=\{\boldsymbol{a},\boldsymbol{b},\boldsymbol{w}\} \tag{1}$$

$$E(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{\theta})=-\sum_{i=1}^{n}a_i v_i-\sum_{j=1}^{m}b_j h_j-\sum_{i=1}^{n}\sum_{j=1}^{m}v_i w_{ij}h_j \tag{2}$$

$$Z(\boldsymbol{\theta})=\sum_{v}\sum_{h}\exp(-E(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{\theta})) \tag{3}$$

$$\theta^*=\arg\max L(\theta)=\arg\max\sum_{n=1}^{N}\ln p(\boldsymbol{v}^n|\theta) \tag{4}$$

$$p(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{\theta})=\frac{1}{Z(\theta)}\exp(-E(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{\theta})) \tag{5}$$

$$p(h_j=1|\boldsymbol{v},\boldsymbol{\theta})=\frac{1}{1+\exp\left(-b_j-\sum_i v_i w_{ij}\right)} \tag{6}$$

$$p(v_i=1|\boldsymbol{h},\boldsymbol{\theta})=\frac{1}{1+\exp\left(-a_i-\sum_j w_{ij}h_j\right)} \tag{7}$$

式中: $\boldsymbol{\theta}$为 RBM 中参数; $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 分别为可视层和隐

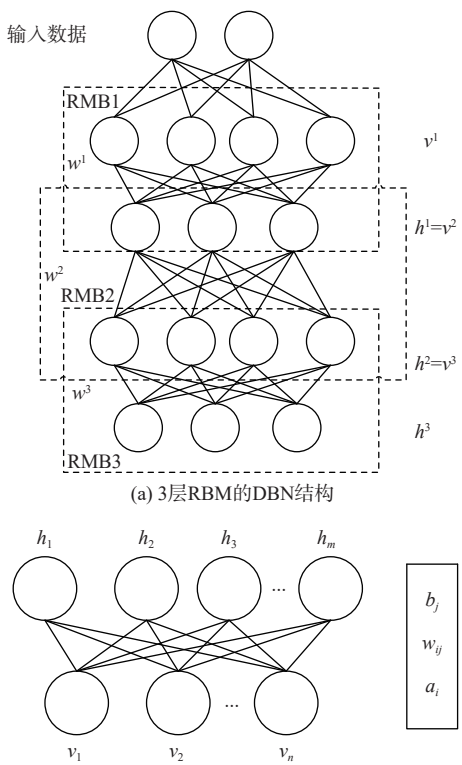


图 1　DBN 模型结构

Fig. 1　Model structure of DBN

含层的偏置向量；$\boldsymbol{w}$为可视层和隐含层间的权重矩阵；$E(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{\theta})$为RBM的能量函数，$(\boldsymbol{v},\boldsymbol{h})$为可视层与隐含层的状态；$Z(\boldsymbol{\theta})$为归一化因子，用于计算能量函数的联合概率分布；$\theta^*$为参数估计结果，用于确定参数集$\boldsymbol{\theta}$；$p(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{\theta})$为状态$(\boldsymbol{v},\boldsymbol{h})$的联合概率分布；$p(h_j=1|\boldsymbol{v},\boldsymbol{\theta})$为可视层状态$\boldsymbol{v}$下隐含层第$j$个神经元被激活概率；$p(v_i=1|\boldsymbol{h},\boldsymbol{\theta})$为隐含层状态$\boldsymbol{h}$下可视层第$i$个神经元被激活概率。

DBN的训练过程由RBM层间训练与反向微调2个阶段组成[20]。RBM层间训练通过第一层RBM的可视层v^1连接输入数据，将数据映射到隐含层h^1。随后隐含层h^1又作为第二层RBM的可视层v^2，数据就像这样逐层传递，从而在高层形成能够表示样本数据的特征；反向微调是通过在DBN的最后一层设置BP神经网络对参数进行微调，使得模型收敛到最优点。

1.2 SCA算法

正余弦算法是一种新型的元启发式算法，仅利用正余弦函数的性质实现对搜索空间的全局搜索和局部寻优。搜索空间中个体的位置对应目标问题的解，对其进行迭代进化可以得到最优解。假设N代表种群规模，t代表迭代次数，SCA算法运算步骤如下：

步骤1 随机初始化N个个体位置。

步骤2 计算个体适应度，并进行比较，选出最优个体，将其位置记为P_i^t。

步骤3 在新一轮迭代中，个体位置按照式(8)进行更新。

步骤4 如果未达到设定迭代数目，将返回步骤2，否则输出P_i^t。

SCA算法中个体更新公式如下：

$$X_i^{t+1}=\begin{cases}X_i^t+r_1\times\sin(r_2)\times|r_3P_i^t-X_i^t| & r_4<0.5\\ X_i^t+r_1\times\cos(r_2)\times|r_3P_i^t-X_i^t| & r_4\geqslant 0.5\end{cases} \tag{8}$$

式中：X_i^t为第t次迭代中解在第i维的值；P_i^t为第t次迭代为止获得的全局最优解的第i维值；r_2、r_3、r_4为随机数，$r_2\in[0,2\pi]$，$r_3\in[0,2]$，$r_4\in[0,1]$。

$$r_1=a-a\frac{t}{T_{\max}} \tag{9}$$

式中：t为当前迭代次数；$T_{\max}$为最大迭代次数；a为常数，通常取2。

2 基于ISCA-DBN优化算法设计

基于ISCA改进DBN网络主要包含3个部分：①DBN神经网络的优化；②改进SCA算法；③ISCA算法优化DBN网络结构。

2.1 DBN网络优化

在微调DBN网络参数时，常采用随机梯度下降(stochastic gradient descent, SGD)算法进行参数更新。SGD算法具有计算复杂度低和训练速度快的优点，一定程度上解决了算法收敛不稳的问题。然而，SGD算法对学习率的取值要求严格，过大的学习率可能导致发散，过小的学习率则会使训练过程缓慢或者陷入局部最小值[12]。

为了解决上述问题，本文引入Adam算法进行参数更新。Adam算法的优点在于可以根据每个参数的历史梯度自适应调整学习率，使得在不同参数上进行有效的学习。Adam更新公式如下：

$$m_t=\beta_1m_{t-1}+(1-\beta_1)g_t \tag{10}$$

$$v_t=\beta_2v_{t-1}+(1-\beta_2)g_t^2 \tag{11}$$

式中：m_t和v_t分别为梯度的一阶矩估计和二阶矩估计；β_1为一阶矩衰减率；β_2为二阶矩衰减率；g_t为当前时间步的梯度。由于m_t和v_t初始值为零向量，所以需要进行校正，网络内部结构如图2所示，校正后的Adam更新公式如式(12)和式(13)所示。

$$\hat{m}_t=\frac{m_t}{1-\beta_1^t} \tag{12}$$

$$\hat{v}_t=\frac{v_t}{1-\beta_2^t} \tag{13}$$

式中：$\hat{m}_t$和$\hat{m}_t$分别为校正后梯度的一阶矩估计和二阶矩估计。

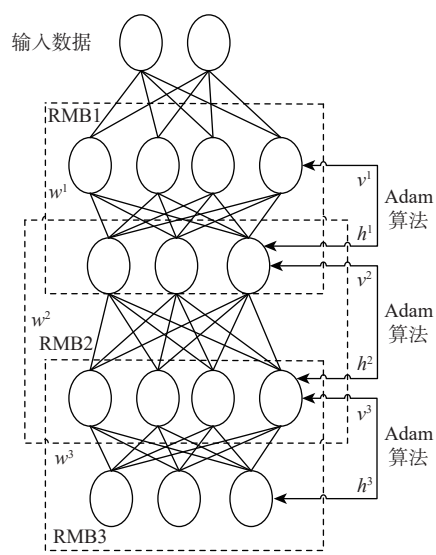


图2 Adam-DBN内部结构

Fig. 2 Internal structure of Adam-DBN

使用相同训练样本，对比SGD算法和Adam算法的网络训练误差变化情况，来选取更合适的优化方法。在不同学习率下，随着迭代次数的增加，模型训练误差变化情况如图3所示。

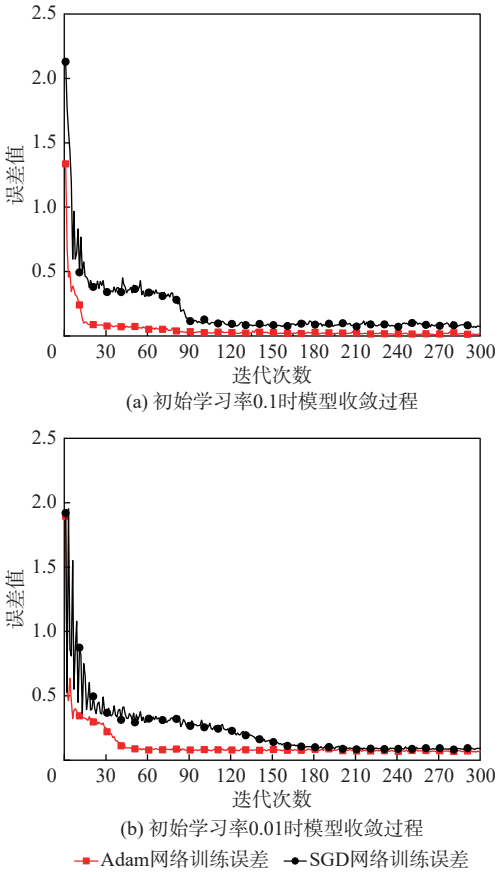


图 3　不同学习率下模型收敛过程

Fig. 3　Comparison of clustering accuracy of algorithms under different model depth conditions

由图 3 可以看到, 当网络设置的初始学习率改变时, Adam 算法不容易受到影响, 而且收敛相对较快。

2.2　ISCA 算法

2.2.1　改进的 Logistic 混沌映射

在正余弦算法中, 种群初始化采用随机分配方式, 导致无法将种群均匀散布在搜索空间中, 降低了多样性, 引起算法过早收敛[21]。Logistic 混沌映射作为经典的混沌映射方式之一, 因其具有随机性和遍历性等特点, 而被广泛用于群智能算法的种群初始化中, 但其仍存在明显缺陷, 如产生的混沌变量分布性不佳等问题。标准 Logistic 混沌映射的表达式为

$$x(n+1)=ux(n)(1-x(n)) \tag{14}$$

式中: $x(n)\in(0,1)$, n=0, 1, 2,…; u 为控制参数, $u\in[0,4]$。

根据 Logistic 混沌映射存在的上述问题, 结合文献 [22] 进行改进, 提出一种混沌映射公式, 如下:

$$\begin{cases} x(n+1)=L(u,x(n))E(k)-\text{floor}(L(u,x(n))E(k)) \\ L(u,x(n))=ux(n)(1-x(n)) \\ E(k)=(e+1)^k \qquad k\in Z^+,k\geqslant 8 \end{cases} \tag{15}$$

式中: $x(n)$ 和 u 的定义同式 (14); floor 为向下取整函数; k 为新增控制参数, 设置的 k 越大, 输出序列的不可预测性越高, 但所需的计算时间也会增加。

假定 $x(0)$=0.5, u=3.9, k=8, n=30 000, 对标准映射和改进映射进行迭代, 然后对映射结果进行统计, 得到如图 4 所示的序列分布。

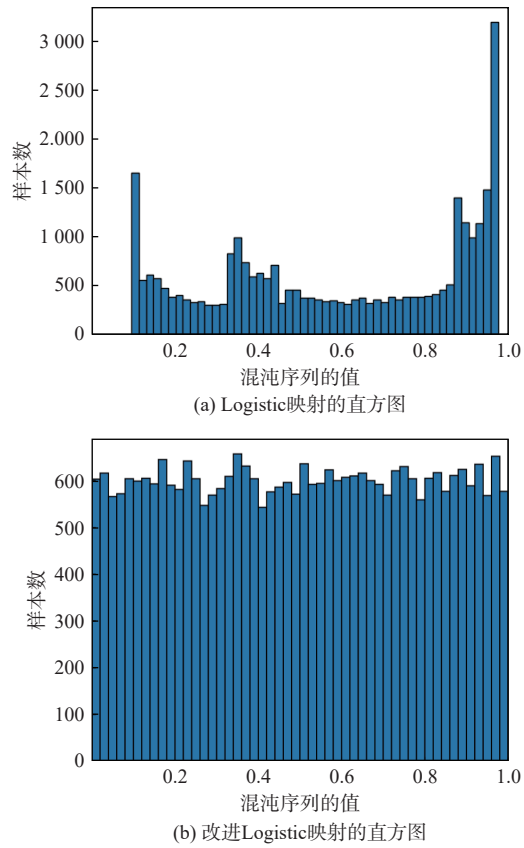


图 4　不同映射下的直方图对比

Fig. 4　Comparison of Histogram comparison on different maps

由图 4 可以看出, 与标准 Logistic 映射相比, 改进 Logistic 映射具有以下优点: ①将分布于两端的混沌变量转移到中间区域, 使其分布较为均衡; ②改进后的映射仍可满足遍历性的需求。

2.2.2　非线性振幅调整因子

分析位置更新式 (8) 可以看出, r_1 主要用于平衡全局探索和局部开发, 当 $r_1>1$ 时, 算法将扩大搜索范围, 体现全局搜索能力, 当 $r_1<1$ 时将缩小搜索范围, 在局部进行精细搜索。经典 SCA 算法中 r_1 的值随着迭代次数从 2 线性递减至 0, 为了弥补线性递减策略在平衡算法全局搜索与局部寻优方面存在的不足, 本文运用余弦函数对公式中的 r_1 进行非线性变化, 其数学表达式为

$$r_1=\cos(\pi\,(t/T_{\max}))+1 \tag{16}$$

改进后振幅调整因子 r_1 的变化曲线如图 5 所示。

从图 5 可知: ①非线性振幅调整因子 r_1 仍符合收缩尺度变化需求; ②改进振幅调整因子 r_1 在迭代

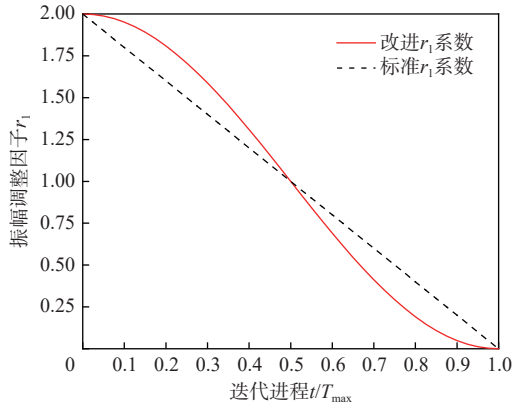


图5 振幅调整因子调节过程

Fig. 5 Adjustment process of amplitude adjustment factor

前期较大且缓慢减小，提高了算法的全局搜索能力；在迭代后期迅速减小，提高了算法的局部寻优性能。

2.2.3 学习策略

在寻优过程中，解空间中各区域个体逐渐向全局最优位置靠近，降低了种群多样性，使标准SCA算法在迭代计算中后期容易陷入局部最优。为了避免SCA算法出现过早收敛的现象，提出一种学习策略。通过分析适应度最优个体是否连续 n 次（本文将 n 设置为5）未发生改变，判断种群可能陷入了停滞，基于变异进化思想，设计了新的个体不仅需要向最优个体进行学习，同时需要向前辈学习（即不同于当前全局最优位置的上一代全局最优个体）的学习策略，帮助算法跳出局部最优，数学表达式为

$$X_{\text{new}}^{t}=X_{\text{old}}^{t}+K(X_{\text{best}}-X_{\text{old}}^{t})+K(X_{\text{pre_best}}-X_{\text{old}}^{t}) \tag{17}$$

式中：X_{best} 为适应度最优个体；$X_{\text{pre_best}}$ 为不同于当前全局最优位置的上一代全局最优个体；K 为随机数，取值范围为[0,1]。

2.3 ISCA算法复杂度

假设种群数目为 N，个体维度为 d，迭代数目为 T。SCA算法的时间复杂度主要为个体位置的更新，其时间复杂度为 $O(N\times d\times T)$。相比标准SCA算法，ISCA算法引入了改进的Logistic混沌映射、非线性振幅调整因子和学习策略。改进的Logistic混沌映射的时间复杂度为 $O(N\times d)$，非线性振幅调整因子的时间复杂度为 $O(T)$，学习策略的时间复杂度为 $O(N\times d\times T)$，与SCA算法的时间复杂度大致相同。

2.4 ISCA-DBN模型的训练流程

ISCA-DBN能耗预测模型的训练流程如图6所示，具体训练流程如下：

1）对地面空调能耗数据进行预处理。为了保证原始数据相对一致性，使用归一化对数据进行预处理，归一化公式如下：

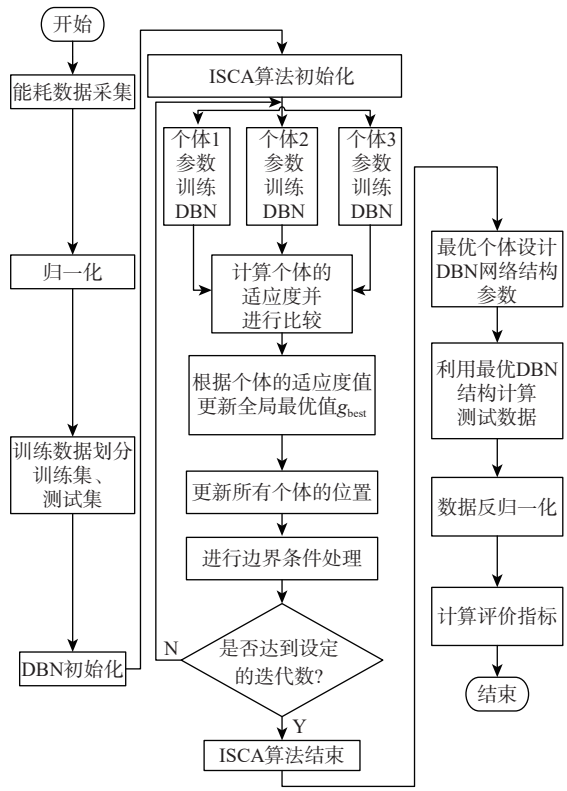


图6 ISCA-DBN训练流程

Fig. 6 The training flow chart of ISCA-DBN

$$X_{\text{e}}=\frac{x-x_{\min}}{x_{\max}-x_{\min}} \tag{18}$$

式中：$x_{\max}$ 和 $x_{\min}$ 为样本数据中各变量的最大值与最小值；X_{e} 为归一化后的数据；x 为样本数据。

2）将样本数据按照7∶3的比例划分，得到训练样本210个，测试样本90个。

3）初始化DBN，并将DBN网络结构参数如预训练学习率、微调学习率、预训练次数、微调次数、批大小、隐含层层数和隐含层节点数作为需要优化的参数，并设置各自的搜索范围。

4）设置ISCA算法的种群大小 N、迭代次数 T 和问题维度 d。

5）初始化个体所在的位置，用训练集对网络结构进行训练，选用预测值与真实值的MSE为适应度函数，并计算所有个体的适应度值。

6）根据个体的适应度求得当前全局最优位置 P_i^t。

7）根据式(8)和式(17)，更新个体的位置。

8）判断迭代次数是否达到设定值。满足条件退出ISCA算法，否则继续迭代。

9）根据ISCA算法得出的全局最优位置对DBN网络结构进行配置。

10）对测试集进行仿真，输出预测结果。

11）与测试集中的能耗标签进行对比，并结合评价指标对模型进行分析。

3　实例分析

3.1　数据来源

将采集夏季停靠在某机场，使用地面空调为飞机客舱制冷的相关数据作为数据集。该飞机型号为 Boeing737-800，使用制冷量为 150 kW 的地面空调对客舱温度和湿度进行调节。该数据共 300 组，每组数据包括乘客下机后的客舱温度和湿度，乘客登机时的客舱温度和湿度以及该过程地面空调所用的耗电量。将下机后客舱温度、下机后客舱湿度、登机时客舱温度、登机时客舱湿度作为模型的输入特征，耗电量作为模型的输出。部分能耗数据数据如表 1 所示。

表 1　部分样本数据

Table 1　Part of sample data

编号	乘客下机的客舱温度/℃	乘客下机的客舱湿度/%	乘客登机的客舱温度/℃	乘客登机的客舱湿度/%	耗电量/kW
54	33.31	73.88	24.65	37.24	170.43
55	30.11	61.45	27.66	36.48	148.00
56	34.61	58.52	25.72	29.82	172.94
57	30.18	63.06	24.56	32.93	168.12
58	34.27	52.49	24.49	36.29	184.36

3.2　评价指标

选择均方误差 S_{ME}、平均绝对百分比误差 A_{MPE} 和 R^2 作为模型预测效果的评价指标。计算过程如下：

$$S_{\mathrm{ME}}=\frac{1}{m}\sum_{i=1}^{m}(y_i-f(x_i))^2 \tag{19}$$

$$A_{\mathrm{MPE}}=\frac{100\%}{m}\sum_{i=1}^{m}\left|\frac{y_i-f(x_i)}{f(x_i)}\right| \tag{20}$$

$$R^2=1-\frac{\sum_{i=1}^{m}(y_i-f(x_i))^2}{\sum_{i=1}^{m}(y_i-\bar{y}_i)^2} \tag{21}$$

式中：$\bar{y}_i$为真实值的平均值；y_i为真实值；$f(x_i)$为预测值；m 为样本数。

3.3　超参数优化过程

本文提出的 ISCA-DBN 模型中的超参数包含预训练学习率、微调学习率、预训练次数、微调次数、批大小、隐含层层数和隐含层节点数，其配置结果决定着模型预测性能的优劣。本文设置的超参数搜索范围如表 2 所示。

选取 SCA-DBN、GA-DBN 和 ISCA-DBN 作为对比模型，在训练集上训练各模型的超参数。将迭代数目设为 100，种群数目设为 8。SCA、GA 和 ISCA 算法迭代过程的对比如图 7 所示。

表 2　超参数搜索范围

Table 2　The search range of hyperparameters

超参数	搜索空间
预训练学习率	0.01～0.1
微调学习率	0.01～0.1
预训练次数	1～20
微调次数	1～200
批大小	1～256
隐含层层数	1～3
隐含层节点数	10～100

从图 7 可以看出：①SCA 算法相比于 GA 算法具有更好的搜索能力；②ISCA 算法的寻优能力要优于 SCA 算法和 GA 算法。3 种算法搜寻结果如表 3 所示。

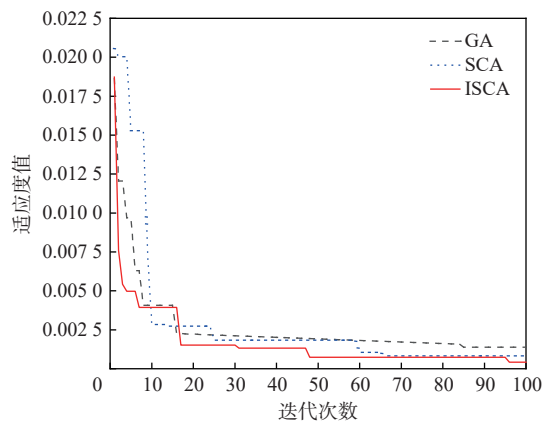


图 7　优化算法迭代过程

Fig. 7　Iterative process of optimization algorithms

表 3　3 种算法搜寻结果

Table 3　Three algorithms search results

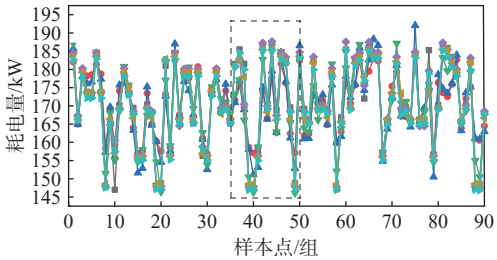
算法	预训练学习率	微调学习率	预训练次数	微调次数	批大小	隐含层层数	隐含层节点数
GA	0.018 5	0.036 2	19	192	12	3	[35,61,31]
SCA	0.051 2	0.055 6	8	100	5	3	[82,44,28]
ISCA	0.042 4	0.082 6	7	147	15	3	[67,46,34]

3.4　仿真结果与分析

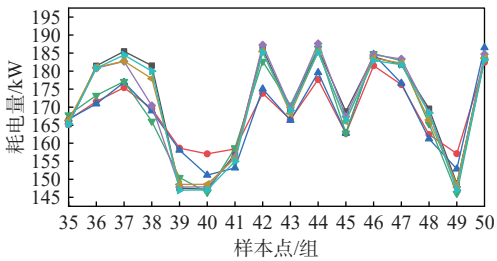
根据算法搜寻结果，对模型网络参数进行赋值。在测试集上仿真，将仿真结果与按照经验调参的 BP 网络、SVM 网络和 DBN 网络的仿真结果进行对比，不同模型预测值和真实值的对比曲线如图 8 所示，图 8(b) 为图 8(a) 的局部放大图。

为进一步评价各模型的性能，在测试集上使用 S_{ME}、A_{MPE}、R^2、泛化误差和时间等评价指标，评价结果如表 4 所示。

由表 4 可以看出：①相较于 BP 网络和 SVM 网



(a) 第1~第90组样本点预测与真实值对比曲线



(b) 第35~第50组样本点预测与真实值对比曲线



图 8 不同模型下预测值与真实值的对比

Fig. 8 Comparison of predicted and real values on different models

表 4 不同模型评价结果

Table 4 Evaluation results for different models

模型	S_{ME}	A_{MPE}/%	R^2	泛化误差	时间/s
BP	36.77	2.732	0.733	0.024	2.01
SVM	35.31	2.831	0.743	0.023	1.33
DBN	13.56	1.540	0.901	0.016	2.13
GA-DBN	4.91	0.731	0.964	0.024	7.55
SCA-DBN	4.29	0.719	0.968	0.050	6.97
ISCA-DBN	3.27	0.706	0.976	0.012	2.78

络，DBN 网络具有更好的预测性能；②利用算法对超参数进行优化相比于经验调参，模型具有更高的预测精度；③ISCA-DBN 模型相比于 GA-DBN 和 SCA-DBN 模型具有更好的预测精度和实时性；④ISCA-DBN 模型相比其他模型具有更强的泛化能力；⑤ISCA 算法相比 SCA 算法具有更强的搜索能力。

4 结 论

设计准确的地面空调能耗预测模型对机场绿色减排的推动起重要作用。基于此，本文提出利用 ISCA 算法优化 DBN 模型参数，建立了 ISCA-DBN 模型，并对仿真结果进行分析，得到以下结论：

1）提出 3 种优化方法改进 SCA 算法，首先提出一种改进 Logistic 混沌映射，对初始种群进行更新；然后引入余弦调节因子对个体位置更新公式进行改进；最后提出一种学习策略，对陷入局部最优的个体进行更新，进而建立了 ISCA 算法。

2）ISCA-DBN 方法能够自适应选取预训练学习率、微调学习率、预训练次数、微调次数、批大小、隐含层层数和隐含层节点数，无需专门花费大量时间反复调整网络层结构，节省了大量的时间和精力。

3）ISCA-DBN 相比于 BP、SVM、DBN、GA-DBN 以及 SCA-DBN 对飞机地面空调的能耗预测，在预测精度和泛化性上均有明显的提高。

参考文献（References）

[1] WANG J Q, DU Y, WANG J. LSTM based long-term energy consumption prediction with periodicity[J]. Energy, 2020, 197: 117197.

[2] 王修岩, 刘艳敏, 张革文, 等. 基于改进协同 PSO 的神经网络飞机客舱能耗预测[J]. 系统仿真学报, 2018, 30(4): 1535-1541.
WANG X Y, LIU Y M, ZHANG G W, et al. Prediction of aircraft cabin energy consumption based on improved cooperative PSO neural network[J]. Journal of System Simulation, 2018, 30(4): 1535-1541 (in Chinese).

[3] 林家泉, 孙凤山, 李亚冲, 等. 基于 IPSO-Elman 神经网络的飞机客舱能耗预测[J]. 航空学报, 2020, 41(7): 323614.
LIN J Q, SUN F S, LI Y C, et al. Prediction of aircraft cabin energy consumption based on IPSO-Elman neural network[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(7): 323614 (in Chinese).

[4] 周璇, 林家泉. 基于改进长短时记忆网络的地面空调能耗预测[J]. 北京航空航天大学学报, 2023, 49(10): 2750-2760.
ZHOU X, LIN J Q. Prediction of ground air conditioner energy consumption based on improved long short-term memory neural network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(10): 2750-2760 (in Chinese).

[5] 刘艳敏. 基于组合预测方法的波音 737 飞机客舱能耗预测研究[D]. 天津: 中国民航大学, 2017: 33-38.
LIU Y M. Research on energy consumption prediction of Boeing 737 cabin based on combined forecasting method[D]. Tianjin: Civil Aviation University of China, 2017: 33-38 (in Chinese).

[6] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.

[7] 张程, 林谷青, 黄靖, 等. 基于 AMBOA-DBN 结合相似日的短期光伏功率预测[J]. 太阳能学报, 2023, 44(6): 290-299.
ZHANG C, LIN G Q, HUANG J, et al. Short-term pv power prediction based on amboa-dbn combined with similar days[J]. Acta Energiae Solaris Sinica, 2023, 44(6): 290-299 (in Chinese).

[8] SHAO H D, JIANG H K, ZHANG H Z, et al. Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network[J]. IEEE Transactions on Industrial Electronics, 2017, 65(3): 2727-2736.

[9] HAN D Y, ZHAO N, SHI P M. A new fault diagnosis method based on deep belief network and support vector machine with Teager–Kaiser energy operator for bearings[J]. Advances in Mechanical Engineering, 2017, 9(12): 168781401774311.

[10] 梁智, 孙国强, 李虎成, 等. 基于 VMD 与 PSO 优化深度信念网络的短期负荷预测[J]. 电网技术, 2018, 42(2): 598-606.
LIANG Z, SUN G Q, LI H C, et al. Short-term load forecasting

based on VMD and PSO optimized deep belief network[J]. Power System Technology, 2018, 42(2): 598-606 (in Chinese).

[11] 徐先峰, 蔡路路, 张丽. 融合 MLP 和 DBN 的光伏发电预测算法[J]. 计算机工程与应用, 2021, 57(3): 266-272.
XU X F, CAI L L, ZHANG L. Photovoltaic power generation prediction algorithm based on MLP and DBN[J]. Computer Engineering and Applications, 2021, 57(3): 266-272 (in Chinese).

[12] 李益兵, 王磊, 江丽. 基于 PSO 改进深度置信网络的滚动轴承故障诊断[J]. 振动与冲击, 2020, 39(5): 89-96.
LI Y B, WANG L, JIANG L. Rolling bearing fault diagnosis based on DBN algorithm improved with PSO[J]. Journal of Vibration and Shock, 2020, 39(5): 89-96 (in Chinese).

[13] SHAN H T, SUN Y Y, ZHANG W J, et al. Reliability analysis of power distribution network based on PSO-DBN[J]. IEEE Access, 2020, 8: 224884-224894.

[14] MIRJALILI S. SCA: a sine cosine algorithm for solving optimization problems[J]. Knowledge-Based Systems, 2016, 96: 120-133.

[15] 徐松金, 龙文. 求解高维优化问题的改进正弦余弦算法[J]. 计算机应用研究, 2018, 35(9): 2574-2577.
XU S J, LONG W. Improved sine cosine algorithm for solving high-dimensional optimization problems[J]. Application Research of Computers, 2018, 35(9): 2574-2577 (in Chinese).

[16] 何庆, 徐钦帅, 魏康园. 基于改进正弦余弦算法的无线传感器节点部署优化[J]. 计算机应用, 2019, 39(7): 2035-2043.
HE Q, XU Q S, WEI K Y. Enhanced sine cosine algorithm based node deployment optimization of wireless sensor network[J]. Journal of Computer Applications, 2019, 39(7): 2035-2043 (in Chinese).

[17] NENAVATH H, JATOTH R K. Hybridizing sine cosine algorithm with differential evolution for global optimization and object tracking[J]. Applied Soft Computing, 2018, 62: 1019-1043.

[18] ATTIA A F, EL SEHIEMY R A, HASANIEN H M. Optimal power flow solution in power systems using a novel sine-cosine algorithm[J]. International Journal of Electrical Power & Energy Systems, 2018, 99: 331-343.

[19] 李正明, 梁彩霞, 王满商. 基于 PSO-DBN 神经网络的光伏短期发电出力预测[J]. 电力系统保护与控制, 2020, 48(8): 149-154.
LI Z M, LIANG C X, WANG M S. Short-term power generation output prediction based on a PSO-DBN neural network[J]. Power System Protection and Control, 2020, 48(8): 149-154 (in Chinese).

[20] 刘帼巾, 刘达明, 缪建华, 等. 基于变分模态分解和改进灰狼算法优化深度置信网络的自动转换开关故障识别[J]. 电工技术学报, 2024, 39(4): 1221-1233.
LIU G J, LIU D M, MIAO J H, et al. Fault identification of automatic transfer switching equipment based on VMD-WPE and IGWO optimized DBN[J]. Transactions of China Electrotechnical Society, 2024, 39(4): 1221-1233 (in Chinese).

[21] 孙凤山, 范孟豹, 曹丙花, 等. 基于混沌映射与差分进化自适应教与学优化算法的太赫兹图像增强模型[J]. 仪器仪表学报, 2021, 42(4): 92-101.
SUN F S, FAN M B, CAO B H, et al. The terahertz image enhancement model based on adaptive teaching-learning based optimization algorithm with chaotic mapping and differential evolution[J]. Chinese Journal of Scientific Instrument, 2021, 42(4): 92-101 (in Chinese).

[22] 刘公致, 吴琼, 王光义, 等. 改进型 Logistic 混沌映射及其在图像加密与隐藏中的应用[J]. 电子与信息学报, 2022, 44(10): 3602-3609.
LIU G Z, WU Q, WANG G Y, et al. A improved logistic chaotic map and its application to image encryption and hiding[J]. Journal of Electronics & Information Technology, 2022, 44(10): 3602-3609 (in Chinese).

Energy consumption prediction of aircraft ground air conditioning based on ISCA-DBN

LIU Han，LIN Jiaquan*

(College of Electronic Information and Automation，Civil Aviation University of China，Tianjin 300300，China)

Abstract：An improved sine-cosine optimization (ISCA) deep belief network (DBN) prediction model for ground air conditioning energy consumption is suggested in order to increase the prediction accuracy of ground air conditioning energy consumption when the aircraft cabin is cooled by ground air conditioning. In contrast to the standard sine-cosine optimization algorithm, the improved sine-cosine algorithm introduces a cosine adjustment factor to create a new non-linear oscillation adjustment factor to balance the algorithm's overall performance. It also suggests an improved logistic chaotic map, which increases population diversity. In order to prevent the algorithm from reaching a local optimum, a learning technique based on the concept of mutation evolution is finally suggested.Search and local optimization capabilities; finally, a learning strategy is proposed based on the idea of mutation evolution to avoid the algorithm from falling into local optimum. The ISCA-DBN model is applied to the prediction of ground air-conditioning energy consumption of Boeing 737-800 aircraft, and the performance is compared with back propagation (BP)、support vector machine (SVM)、DBN algorithms. There is a certain improvement in both prediction accuracy and real-time performance.

Keywords：aircraft cabin；ground air condition；energy demands prediction；sine cosine algorithm；deep belief network

Received：2023-06-23；**Accepted**：2023-08-14；**Published Online**：2023-08-25 12：10

URL：link.cnki.net/urlid/11.2625.V.20230824.1900.001

* **Corresponding author**. E-mail：jqlin@cauc.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51卷　第6期　2025年6月

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)

(Monthly, Started in 1956)

Vol.51 No.6　June 2025

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

9 771001 596250 06>

国内定价：50.00元／期

微信公众号　超星电子书　QQ群