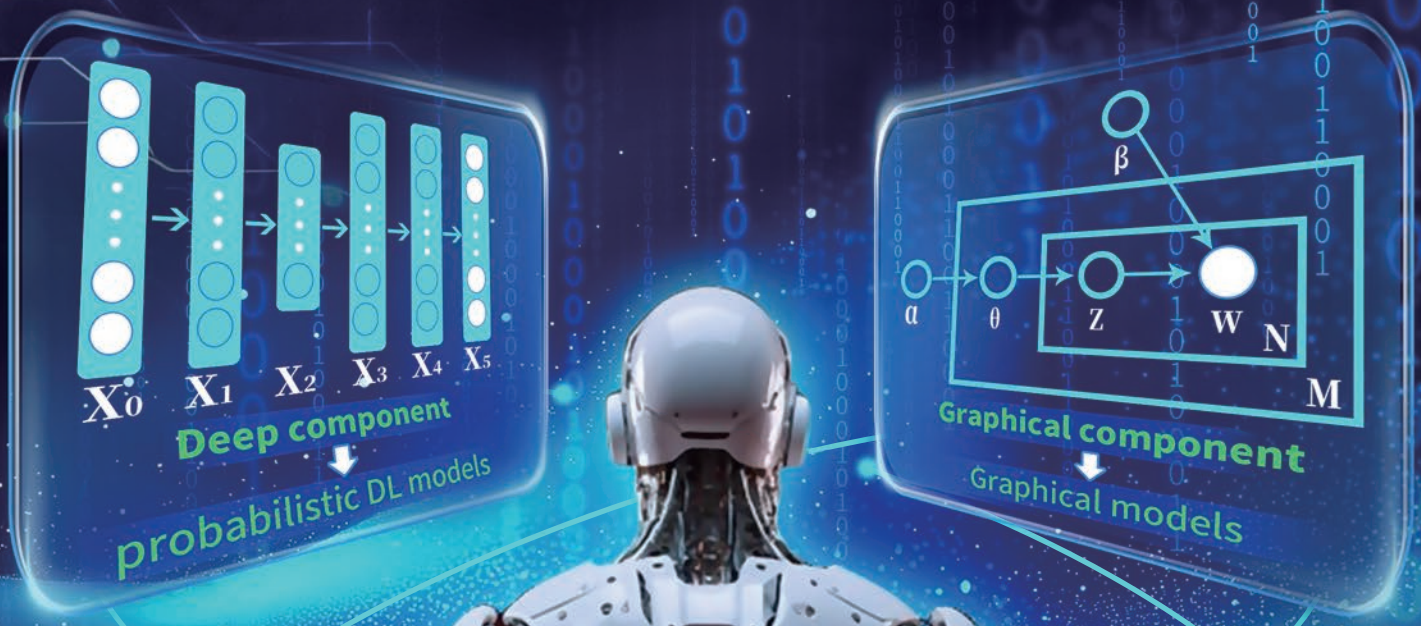
主管单位：中华人民共和国工业和信息化部
主办单位：北京航空航天大学

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



贝叶斯网络结构学习综述

孟光磊　丛泽林　宋　彬　李婷斑　王晨光　周铭哲

9 2025
Vol.51

北 京 航 空 航 天 大 学 学 报

第 51 卷　第 9 期　（总第 391 期）　2025 年 9 月

目　　次

贝叶斯网络结构学习综述 …… 孟光磊，丛泽林，宋彬，李婷琰，王晨光，周铭哲（2829）

柔性支承特性对惯性参考单元性能影响的分析 …… 李醒飞，韩佳欣，拓卫晓，刘雅琪，王天宇（2850）

基于 MPC 的随机风场下航空器纵向自主间隔控制 …… 汤新民，陆晓娜（2860）

采用多级渐进式门控卷积网络的歌声分离方法 …… 罗庆予，张天骐，熊天（2872）

双三相起动发电系统电源变换器功率管开路故障诊断 ……
…… 王化吉，刘禹，孙高阳，李德洪，徐金全（2882）

基于非均匀大气光修正模型的夜间图像去雾 …… 林森，查子月（2894）

强噪声环境下基于 MSDCNN 的滚动轴承故障诊断方法 ……
…… 雷春丽，史佳硕，马淑珍，缪成翔，万会元，李建华（2906）

基于 MPC 的飞机牵引车轨迹跟踪 …… 张军，黄明辉，王玥琳，阳星，叶敏，贾永乐（2916）

航空货运安保供应链弹性测度及优化策略 …… 赵振武，邢肖肖，耿硕（2927）

基于演化博弈的航空器自主冲突解脱效率 …… 王红勇，郭宇鹏（2937）

一种结冰飞机失速迎角实时估计方法 …… 江飞鸿，刘贞报，薛源，孔满昭，赵甜（2947）

临地协同观测中的高空气球组网轨迹优化 …… 曲艺，王生，冯慧，刘强（2955）

基于滑动时间窗的机载传感器多任务调度算法 …… 冉华明（2968）

基于贝叶斯优化 LSTM 神经网络的飞机货舱火源定位 …… 张伟，常本强，杨旭，熊枭（2979）

基于滑模控制和舵轴驱动的多间隙折叠舵颤振抑制方法 ……
…… 任浩源，程涛，张程，蔡毅鹏，刘飞，张炜群（2987）

虑及装配中多源误差的航发转子不平衡量预测 …· 周天一，胡磊，赵所，靳潇潇，王诗雨，高航（3001）

基于 GBDT-GS 方法的民机重着陆风险预测 …… 王向章，牟瑞芳，王赫，许博浩（3011）

基于 U-Net++和特征融合的塑料齿轮复杂黑点检测方法 …… 方一鸣，石照耀，宋辉旭（3020）

氧燃比对自燃推进剂模型火箭发动机燃烧稳定性的影响 ……
…… 任永杰，徐伯起，楚威，郭康康，仝毅恒，聂万胜（3030）

动态区域聚焦的反无人机红外长时跟踪算法 ……
…… 谢学立，席建祥，卢瑞涛，杨小冈，张涛，夏文新（3039）

导星辅助运动暗弱天体检测系统设计与仿真 …… 张威，冯文全，孙国童，王春雷（3052）

面向单子集焦元冲突证据的融合方法 …… 周诚，徐达，曹振地（3062）

尾座式垂直起降无人机过渡轨迹优化方法 …… 邹旭，刘贞报，赵闻，王莉娜（3071）

长时间气动力/热作用下累积热变形演化规律 ……
…… 刘深深，蒋波，韩青华，余婧，杨肖峰，魏东，朱言旦，桂业伟（3086）

线束拓扑图纸快速签收方法 …… 雷璐，吴明宇，郭智博，乔晓利（3098）
基于资源供需差与情境理解度的情境意识模型 ……
…… 王一行，刘双，完颜笑如，冯传宴，周孙夏，钱春颖（3108）
冷云辐射对超压气球热动力学特性影响 …… 贺澜，刘强，杨燕初，祝榕辰，周江华（3117）
轨控发动机含冰晶多相羽流辐射特性 …… 罗吴迪，任军学，李志辉，汤海滨（3130）
基于双天线 TDCP 的 GNSS/IMU 组合导航航向角增强算法 …… 吉莉，孙蕊，王媛媛，戴晔莹（3141）
基于 STF-Net 的信号调制波形识别方法 …… 哈晖，高翔，姚秀娟，付降寅，李伟，张晓燕（3150）
民机货舱氮气-细水雾协同灭火数值仿真研究 …… 黄苏平，朱培，刘全义，邵荃（3161）
格网误差建模辅助的 GNSS/IMU 组合导航综合质量控制算法 ……
…… 戴晔莹，孙蕊，邓思瑜，吉莉，王媛媛，黄薛东（3174）
基于直接概率积分的混合可靠性分析方法 …… 朱启滔，李洪双（3183）
繁忙终端空域连续爬升与下降运行的调度优化 …… 杜卓铭，张军峰，归旭豪（3193）
基于改进 TCN-Elman 神经网络的电离层杂波抑制方法 …… 刘强，尚尚，乔铁柱，祝健，石依山（3203）
基于子弹击穿理论的直升机机身弹伤快速分析方法 ……
…… 邢肖兵，谭剑锋，杨宇霄，雍天，于领军，武瑞（3212）
基于混合注意力的全天候机场跑道异物检测 …… 张敬博，任杰，王美琪（3222）

期刊基本参数：CN 11-2625/V * 1956 * m * A4 * 404 * zh * P * ¥ 50.00 * 400 * 37 * 2025-09

（**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. 9 (Sum 391) September 2025

CONTENTS

A review of Bayesian network structure learning
MENG Guanglei, CONG Zelin, SONG Bin, LI Tingting, WANG Chenguang, ZHOU Mingzhe (2829)

Influence analysis of flexure support characteristics on performance of inertial reference unit
LI Xingfei, HAN Jiaxin, TUO Weixiao, LIU Yaqi, WANG Tianyu (2850)

Longitudinal autonomous separation control of aircraft in random wind fields based on MPC
TANG Xinmin, LU Xiaona (2860)

Singing voice separation method using multi-stage progressive gated convolutional network
LUO Qingyu, ZHANG Tianqi, XIONG Tian (2872)

Open-circuit fault diagnosis of power tubes in power converter of dual three-phase starter-generator system
WANG Huaji, LIU Yu, SUN Gaoyang, LI Dehong, XU Jinquan (2882)

Nighttime image dehazing based on non-uniform atmospheric light correction model
LIN Sen, ZHA Ziyue (2894)

Rolling bearing fault diagnosis method based on MSDCNN in strong noise environment
LEI Chunli, SHI Jiashuo, MA Shuzhen, MIAO Chengxiang, WAN Huiyuan, LI Jianhua (2906)

Trajectory tracking for aircraft tug based on MPC
ZHANG Jun, HUANG Minghui, WANG Yuelin, YANG Xing, YE Min, JIA Yongle (2916)

Resilience measurement and optimization strategy of air cargo secure supply chain
ZHAO Zhenwu, XING Xiaoxiao, GENG Shuo (2927)

Efficiency of aircraft autonomous conflict resolution based on evolutionary games
WANG Hongyong, GUO Yupeng (2937)

A real-time estimation method for stall angle of attack of icing aircraft
JIANG Feihong, LIU Zhenbao, XUE Yuan, KONG Manzhao, ZHAO Tian (2947)

Trajectory optimization of high-altitude balloon in nearspace-ground collaborative observation
QU Yi, WANG Sheng, FENG Hui, LIU Qiang (2955)

Airborne sensor multi-task scheduling algorithm based on slide time window
RAN Huaming (2968)

Fire source localization for long short-term memory networks based on Bayesian optimization
ZHANG Wei, CHANG Benqiang, YANG Xu, XIONG Xiao (2979)

Flutter suppression method for multi-freepalys folding fin based on sliding mode control and fin shaft drive
REN Haoyuan, CHENG Tao, ZHANG Cheng, CAI Yipeng, LIU Fei, ZHANG Weiqun (2987)

Prediction of aeroengine rotor unbalance considering multi-source errors in assembly process
ZHOU Tianyi, HU Lei, ZHAO Suo, JIN Xiaoxiao, WANG Shiyu, GAO Hang (3001)

Hard landing risk prediction of civil aircraft based on GBDT-GS method
WANG Xiangzhang, MOU Ruifang, WANG He, XU Bohao (3011)

Detection method for complex dark spots on plastic gears based on U-Net++ and feature fusion
FANG Yiming, SHI Zhaoyao, SONG Huixu (3020)

Effects of oxygen-fuel ratio on combustion stability of a model rocket engine with hypergolic propellant
REN Yongjie, XU Boqi, CHU Wei, GUO Kangkang, TONG Yiheng, NIE Wansheng (3030)

Long-term infrared object tracking algorithm based on dynamic region focusing for anti-UAV
XIE Xueli, XI Jianxiang, LU Ruitao, YANG Xiaogang, ZHANG Tao, XIA Wenxin (3039)

Design and simulation of guide star assisted moving faint object detection system
ZHANG Wei, FENG Wenquan, SUN Guotong, WANG Chunlei (3052)

Conflict evidence fusion method for single-subset focal elements
…… ZHOU Cheng，XU Da，CAO Zhendi （3062）

Optimization method of transition trajectory for tail-sitter unmanned aerial vehicles
…… ZOU Xu，LIU Zhenbao，ZHAO Wen，WANG Lina （3071）

Cumulative thermal deformation evolution under prelonged aerodynamic and thermal loads
…… LIU Shenshen，JIANG Bo，HAN Qinghua，YU Jing，YANG Xiaofeng，WEI Dong，ZHU Yandan，GUI Yewei （3086）

Fast sign-off method for wiring harness topology drawings
…… LEI Lu，WU Mingyu，GUO Zhibo，QIAO Xiaoli （3098）

Situation awareness model based on resource supply-demand difference and understanding
…… WANG Yihang，LIU Shuang，WANYAN Xiaoru，FENG Chuanyan，ZHOU Sunxia，QIAN Chunying （3108）

Influence of cold cloud radiation on thermal-dynamic characteristics of super-pressure balloon
…… HE Lan，LIU Qiang，YANG Yanchu，ZHU Rongchen，ZHOU Jianghua （3117）

Radiation characteristics of multi-phase plumes containing ice crystals from orbit-control engines
…… LUO Wudi，REN Junxue，LI Zhihui，TANG Haibin （3130）

Heading enhancement algorithm of GNSS/IMU integrated navigation based on dual-antenna TDCP
…… JI Li，SUN Rui，WANG Yuanyuan，DAI Yeying （3141）

Signal modulation waveform recognition method based on STF-Net
…… HA Hui，GAO Xiang，YAO Xiujuan，FU Jiangyin，LI Wei，ZHANG Xiaoyan （3150）

Numerical simulation study on synergetic fire suppression of nitrogen/water-mist in aircraft cargo compartment
…… HUANG Suping，ZHU Pei，LIU Quanyi，SHAO Quan （3161）

Grid error modeling aided GNSS/IMU integrated navigation comprehensive quality control algorithm
…… DAI Yeying，SUN Rui，DENG Siyu，JI Li，WANG Yuanyuan，HUANG Xuedong （3174）

A mixed reliability analysis method based on direct probability integral
…… ZHU Qitao，LI Hongshuang （3183）

Scheduling optimization for continuous climb and descend operations in busy terminal area
…… DU Zhuoming，ZHANG Junfeng，GUI Xuhao （3193）

Ionospheric clutter suppression method based on improved TCN-Elman neural network
…… LIU Qiang，SHANG Shang，QIAO Tiezhu，ZHU Jian，SHI Yishan （3203）

An efficient analysis method of helicopter fuselage bullet damage based on bullet penetration theory
…… XING Xiaobing，TAN Jianfeng，YANG Yuxiao，YONG Tian，YU Lingjun，WU Rui （3212）

All-weather airport runway foreign object debris detection based on mixed attention
…… ZHANG Jingbo，REN Jie，WANG Meiqi （3222）

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0445

贝叶斯网络结构学习综述

孟光磊[1,2,*]，丛泽林[1,2]，宋彬[1,2]，李婷斑[2]，王晨光[2]，周铭哲[1,2]

(1. 沈阳航空航天大学 自动化学院，沈阳 110136；2. 航空科学技术空战系统技术重点实验室，沈阳 110136)

摘　　要：贝叶斯网络作为概率论与图论结合的工具，具备高效处理不确定性推理和数据分析的能力，被广泛应用于各领域解决复杂工程问题。此外，还可以结合先验知识和训练样本学习模型，克服了单纯依靠专家知识建立模型的局限性。基于此，回顾了贝叶斯网络的发展历程，分别从基于约束的方法、基于评分搜索的方法、混合约束和评分搜索的方法3个方面对已提出的贝叶斯网络结构学习方法进行分类归纳，并对各类方法研究的现状进行了总结分析。由于现实应用中的数据往往具有非完备性，从缺失数据处理和隐变量学习2个维度阐释了非完备贝叶斯网络结构学习的研究现状。对贝叶斯网络在不同领域中的应用情况进行阐述，并进行总结，讨论了未来贝叶斯网络结构学习方法研究的发展趋势。

关　键　词：机器学习；人工智能算法；贝叶斯网络；结构学习；隐变量

中图分类号：TP181

文献标志码：A　　**文章编号**：1001-5965(2025)09-2829-21

数据挖掘[1]技术作为一门新兴学科，融合了当今时代先进的计算机技术、人工智能技术和概率统计学技术，其对采集到的数据进行自主分析和归纳推理，发现潜在的数据模型，从大量的数据中提取有效的、具有潜在价值的信息，以达到帮助决策者做出正确决定、降低风险的目的。贝叶斯网络(Bayesian network, BN)算法相比于其他决策模型，将概率推理和图形模型有机结合，对多元不确定性知识具有可视化的模型推理能力，已经被证实为是一种可进行不确定性问题推理分析的有效算法模型。

条件独立的随机变量之间的定性和定量关系是贝叶斯网络学习的基本内容。定性内容指描述变量间因果关系的网络结构；定量内容则主要指描述变量间依赖程度的条件概率分布。在变量结构模型已知的条件下，通过已知变量间的定性关系来分析各变量间的定量关系，即从数据样本中学习得到节点间的条件概率分布的过程被称为贝叶斯网络参数学习；当变量结构模型未知时，贝叶斯网络学习同时揭示各变量间的定性和定量关系，该过程被称为结构学习。军事、医疗、金融、工业生产等很多领域中都积累了海量的数据，贝叶斯网络学习为用户挖掘数据中潜在的规律，找出相对于数据集拟合度最高的网络结构模型辅助相关领域决策，提供了有效的技术途径。

目前，贝叶斯网络结构学习方法大体可以分为3种类型：基于约束的方法、基于评分搜索的方法及混合约束和评分搜索的方法。基于约束的方法是最早提出的贝叶斯网络结构学习方法，利用条件独立性(conditional independence, CI)测试[2]来学习贝叶斯网络结构，尽可能多地去除变量之间多余的依赖关系，如PC算法[3-4]、三阶段依赖分析算法(three-phase dependency analysis algorithm, TPDA)[5]等。由于需要大量样本来测试独立性，基于约束的方法时间代价较大，且有时不可靠的CI测试会使

收稿日期：2023-07-07；**录用日期**：2023-08-04；**网络出版时间**：2023-09-05 10:36
网络出版地址：link.cnki.net/urlid/11.2625.V.20230904.1958.001
基金项目：国家自然科学基金(61973222)；辽宁省“兴辽英才计划”(XLYC2007144)；沈阳市自然科学基金(22-315-6-09)
* **通信作者**. E-mail：mengguanglei@yeah.net

引用格式：孟光磊，丛泽林，宋彬，等. 贝叶斯网络结构学习综述[J]. 北京航空航天大学学报，2025，51(9)：2829-2849.
MENG G L，CONG Z L，SONG B，et al. A review of Bayesian network structure learning[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2829-2849 (in Chinese).

输入级的错误被放大传导到输出级，导致网络推理结果准确性降低。基于评分搜索的方法应用较为广泛，其可以灵活地将专家知识[6]作为结构先验引入搜索学习过程，利用评分函数评估每个可能的网络结构，如K2算法[7]、BIC算法[8]等。随着节点数量增加，基于评分搜索的方法学习网络结构空间数量呈指数级增长，使得最优结构的搜索代价过高。因此，基于评分搜索的方法一般采用启发式策略[9]来减少搜索空间，但这种方式有时会导致网络陷入局部最优，相对而言，更具有自适应性的基于自然选择的群体智能(swarm intelligence, SI)[10]优化算法成为近年来的研究热点。此外，研究学者也提出了一系列混合方法用以解决基于约束的方法和基于评分搜索的方法的缺陷问题。例如，通过低阶CI测试得出节点的排序，减少后续结构搜索代价，再使用K2算法学习网络，搜索所有可能网络中的最优结构。

随着贝叶斯网络算法不断改进，数据的处理能力也随之提高，逐步适用于各种复杂技术领域，并发挥重要作用。本文梳理了贝叶斯网络结构学习方法的发展脉络，对现有成果进行总结，为研究人员进行文献追踪、迅速跟进该领域的研究前沿提供参考。首先，简单介绍了贝叶斯网络的发展背景和原理；其次，对贝叶斯网络结构学习方法进行分类，并进行系统性的总结，针对现实应用中变量(如数据缺失)的情况，对非完备信息下贝叶斯网络结构学习进行详细分类阐述；然后，梳理总结了贝叶斯网络结构学习方法在社会各领域的应用情况；最后，总结了未来贝叶斯网络结构学习理论研究的发展趋势。

1 贝叶斯网络

1.1 背景介绍

贝叶斯网络又称信念网络，由图灵奖获得者Judea Pearl提出[11]，他也因此被誉为“贝叶斯网络之父”。贝叶斯网络是一种用于计算复杂逻辑、推理因果关系中模糊概率的图模型，能够表示随机变量及变量间的依赖关系，其以因子分解的方式定义了联合概率分布的数据结构，并给出了分布中的一系列CI假设。

贝叶斯网络拓扑结构是一个有向无环图(directed acyclic graph, DAG)。其中，节点表示随机变量，弧表示它们之间的条件概率依赖关系。“有向”指的是用来连接不同节点的边的方向是固定的，起点和终点顺序不能调换，说明由因到果的逻辑关系不能颠倒；“无环”指的是从任意节点出发，无法经过若干条边再次回到该点，即在图中不存在任何环路。文献[12]强调应按照因果关系来决定变量顺序。非条件独立的2个随机变量节点之间连接的起始节点为“因”，终点节点为“果”。有向图的结构可以模拟建模域的因果结构，当结构是因果关系时，其为变量之间的相互作用提供了可视化信息，并能预测外部操纵带来的影响。

1.2 数学模型

贝叶斯网络N由结构G和网络参数Θ表示为二元组$N=(G,\Theta)$。在图论中，可以令有向无环结构图$G=(X,R)$表示一组变量之间的独立关系，X为图形中所有节点的非空集合$X=\{X_1,X_2,\cdots,X_i,\cdots,X_n\}$，$X_i$可以为观测变量，也可以为隐变量，$R$为DAG中不同变量之间有向连接线段的集合，$X_j\to X_i$表示节点之间的直接依赖关系。

$$R=\{X_j\to X_i|X_j\in \mathrm{pa}(X_i),i=1,\cdots,n\} \tag{1}$$

式中：$\mathrm{pa}(X_i)$表示节点X_i的“因”，也称为父节点集合。根据马尔可夫独立性，给定父节点集$\mathrm{pa}(X_i)$，X_i独立于其非子代节点集$\mathrm{nd}(X_i)$。因此，若干遵循马尔可夫规则的节点X_i的联合概率分布可以表示为

$$P(X_1,\cdots,X_n)=\prod_{i=1}^{n}P(X_i|\mathrm{pa}(X_i)) \tag{2}$$

网络参数$\Theta=\{\Theta_1,\Theta_2,\cdots,\Theta_n\}$表示每个节点$X_i$在其父节点集$\mathrm{pa}(X_i)$已知时的条件概率表(conditional probability table, CPT)。当一个贝叶斯网络的网络结构G和网络参数Θ已知时，就能求得节点X_i的联合概率分布。贝叶斯网络算法因节点间存在条件独立的性质，所以其计算效率远高于其他计算联合概率的算法。

以糖尿病诊断为例，对贝叶斯网络模型进行直观说明。图1模拟了5点贝叶斯网络模型图，利用“1和0”表示事件发生与否，通过分析节点$A\sim E$之间的独立性和相关性可以得出以下结论。

1）节点的联合概率分布可以表示为

$$P(A,B,C,D,E)=P(E|C)P(D|B,C)P(C|A)P(B|A)P(A) \tag{3}$$

2）当C为已知条件下，A和E相互独立。

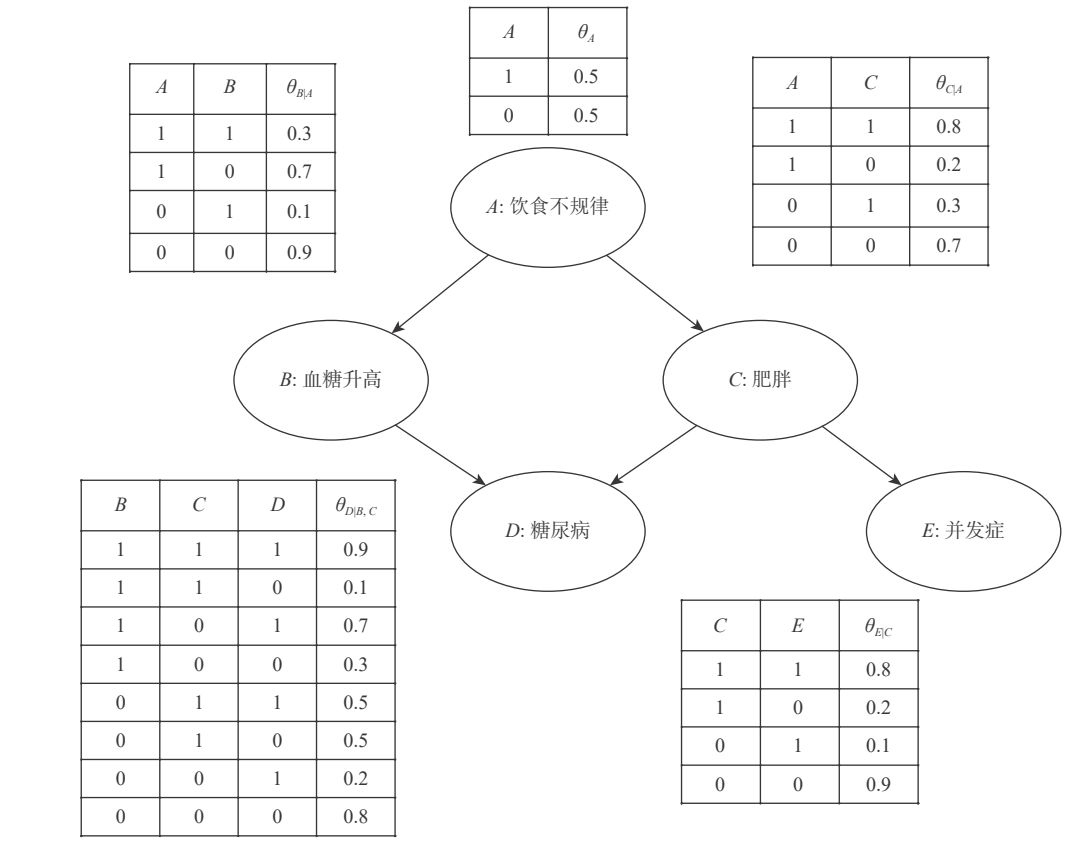
3）当D为未知条件下，B和C相互独立。

结论1由式(2)可得。结论2和结论3则可根据有向分离中3类节点连接类型(见图2)推断得出：

1）顺连。有向连接$a\to c\to b$被称为顺连，其中，节点c被称为头对尾节点。

2）分连。有向连接$a\leftarrow c\to b$被称为分连，其中，节点c被称为尾对尾节点。

3）汇连。有向连接$a\to c\leftarrow b$被称为汇连，其



A	θ_A
1	0.5
0	0.5

A	B	$\theta_{B\|A}$
1	1	0.3
1	0	0.7
0	1	0.1
0	0	0.9

A	C	$\theta_{C\|A}$
1	1	0.8
1	0	0.2
0	1	0.3
0	0	0.7

B	C	D	$\theta_{D\|B,C}$
1	1	1	0.9
1	1	0	0.1
1	0	1	0.7
1	0	0	0.3
0	1	1	0.5
0	1	0	0.5
0	0	1	0.2
0	0	0	0.8

C	E	$\theta_{E\|C}$
1	1	0.8
1	0	0.2
0	1	0.1
0	0	0.9

图 1　糖尿病诊断贝叶斯网络模型

Fig. 1　Bayesian network model for diabetes diagnosis

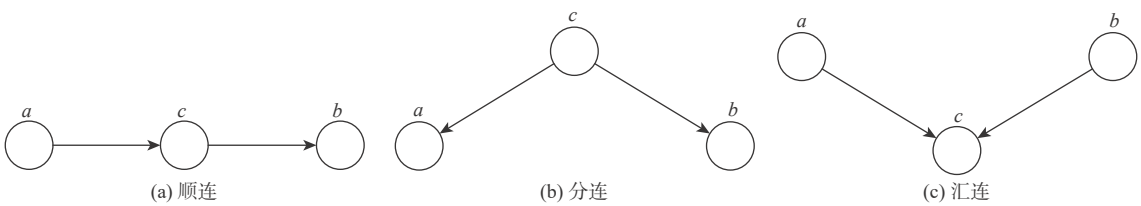


图 2　有向分离中 3 类节点连接类型

Fig. 2　Three types of node connection in d-separation

中, 节点 c 被称为头对头节点。

根据 CI 假设可知, 在顺连中, 当节点 c 未知时, 有: $P(a,b,c)=P(a)P(c|a)P(b|a)$, 但无法推出 $P(a,b)=P(a)P(b)$, 即 a、b 之间存在相关性。当节点 c 已知时, 有

$$P(a,b|c)=P(a,b,c)/P(c)=P(a)P(c|a)P(b|c)/P(c)=P(a,c)P(b|c)/P(c)=P(a|c)P(b|c) \quad (4)$$

即 a、b 之间被 c 阻断, 称为头对尾条件独立。

同理, 在分连中, 当 c 未知时, a、b 节点相关; 当 c 已知时, a、b 节点被 c 阻断, 称为尾对尾条件独立。例如, 图 1 中, 当肥胖未知时, 饮食不规律和并发症之间存在相关性; 而当肥胖已知时, 饮食不规律和并发症关于肥胖条件独立(见结论 2)。

在汇连中, 当节点 c 已知时, 节点 a、b 相关, 可以理解为在结果已知的因果关系中, 各原因之间是存在相关性的; 而当节点 c 未知时, a、b 关于 c 条件独立, 称为头对头条件独立。例如, 图 1 中, 当糖尿病未知时, 血糖升高和肥胖之间条件独立(见结论 3)。

2　贝叶斯网络结构学习方法

贝叶斯网络结构构造有专家咨询和数据学习 2 种方法。在贝叶斯网络发展初期, 通常采用专家咨询的构造方法, 依据专家经验得到网络参数, 再手工建立网络结构模型, 通过减少参数的个数来提高咨询结果的质量, 从而得到一些简单的结构模型。因果独立和环境独立是专家咨询中 2 种常见的局部结构[13]。专家咨询的方式适用于先验知识丰富、模型复杂度低的领域, 多被用于构建没有历史统计数据的网络结构。例如, 文献 [14-15] 通过咨询相应领域的专家经验去判断影响水质各因素之间的联系。对于参数多、复杂度高的大型贝叶斯

网络，仅依靠专家经验是难以建立的。

随着信息技术的进步，很多领域都积累了大量的数据，数据正逐步演变成重要的市场要素，大数据挖掘成为医疗、金融、保险、军事等各社会发展领域前进的驱动力。相比于专家咨询的建模方法，贝叶斯网络结构学习方法可以充分挖掘数据中蕴含的逻辑关联，为建立复杂问题的可解释模型提供了具有潜力的技术途径，受到了学者们的广泛关注。

2.1 基于约束的方法

2.1.1 原理概述

基于约束的方法认为结构学习是约束满足性问题，服从马尔可夫假设和忠诚性假设。基于约束的方法将贝叶斯网络看作是具有独立变量的结构模型，先通过互信息 (mutual information, MI) 或 CI 检验的方式判断变量之间的独立关系，学习变量间的骨架图，再利用 V-结构和定向规则对变量间的无向边进行定向，确定变量间的因果方向，从而学习得到完整的贝叶斯网络结构。

基于约束的方法的精确性通常会随着测试数量和约束集的增多而降低。在构造初始无向图时，通常使用卡方统计量或互信息进行 CI 检验。

1）节点间的卡方统计量公式为

$$\chi^2=\sum_{i=1}^{k}\frac{(W_i-Q_i)^2}{Q_i} \tag{5}$$

式中：W_i为样本观察值；Q_i为理论值。

假设数据集近似服从自由度为$k-1$的卡方分布，根据卡方统计量做 CI 假设检验，确定节点间边的存在与否，最终确定贝叶斯网络的初始框架。

2）节点间的互信息量公式为

$$I(x;y\mid z)=H(x\mid z)-H(x\mid z,y) \tag{6}$$

式中：条件熵$H(x\mid z)$和$H(x\mid z,y)$之差称为给定z时，x和y之间的条件互信息。通过评价节点间的互信息值，可以度量在已知y的观察值时，对x取值的依赖程度。如果互信息量小于给定阈值，反映在真实网络中的 2 个节点极大可能不相邻，在构造初始无向图时，2 个节点间不存在相邻边。

2.1.2 文献分析

1991 年，最早的基于约束的方法，即 SGS 算法[16]被提出，其利用特定的因果模型来判断变量之间的 CI 假设，不需要预先知道节点顺序。然而，SGS 算法具有不稳定性，容易影响后续过程，产生较大误差，且学习贝叶斯网络结构过程中需要进行指数级的 CI 测试，存在工作效率低的问题。Spirtes 和 Glymour 等在 SGS 算法基础上进行改进，提出了 PC 算法[3]，从顶点集上生成的完全无向图开始，利用 d-分离[17]判断进行 CI 检测，递归删除边，直至生成需要的网络结构模型。对于稀疏和离散的网络，PC 算法不需要测试高阶独立性关系，大大减少了测试次数，具有计算量小、复杂程度低的特点，算法在保证独立性判断准确的情况下，结果也具有一定程度的可靠性，但缺点是收敛速度慢。

2002 年，文献 [18] 提出了 TPDA 算法，通过绘图、增厚和削减 3 种操作，计算各节点之间的互信息量，从而确定节点间的条件独立性来构造 DAG 模型。算法将信息论和统计测试结合，以变量间的互信息检测代替了 CI 检验，可以处理已知节点序或未知节点序的模型，生成非常接近底层逻辑的信念网络，在准确度和时间复杂度方面均优于 PC 算法。

2006 年，文献 [19] 提出了贝叶斯网络结构学习的递归自治识别 (recursive autonomy identification, RAI) 算法，从一个完全无向图出发，按照 CI 检验、边的有向性确定、自治子结构识别的顺序来学习贝叶斯网络模型。算法中，CI 测试的顺序与 PC 算法类似，都是在高阶测试之前先执行低阶测试（即使用小条件集进行测试），通过优先执行更可靠的测试，从而降低测试失败的概率。与其他基于约束的方法先对结构进行 d-分离，再引导产生无向图的学习方式不同，RAI 算法从开始就将这 2 个过程结合起来，使得结构学习只需要较少的高阶 CI 测试，大大降低了算法的复杂性和运行时间。

2010 年，文献 [20] 提出了混合父子 (hybrid parents and children, HPC) 算法，用来学习贝叶斯网络中目标变量的父、子节点，寻找父子节点在贝叶斯网络结构学习、降维和特征选取等应用中也是重要的一步。该算法结合了增量法和分治法的思想，能够处理较大的邻域，在小样本内仍保持良好的性能。

2013 年，文献 [21] 提出了轻互最小值 (light mutual min, LMM) 算法，用于解决在小样本数据情况下网络结构学习正确率低的问题，通过考虑节点对的条件独立性来改善对候选边缘节点的评估，并采用基于约束的自适应放宽方式，选择性地删除节点的邻接，从而减少由于多次 CI 测试而导致真实边缘被移除的现象。并行计算技术被引入挖掘基于约束的局部因果关系，从而提高了网络结构学习效率。

2014 年，文献 [22] 提出了一种优化的零一阶超结构 (optimized zero-first-order super-structure, Opt01-SS) 算法。采用零一阶 CI 测试解决了 2 类问题：①小样本条件下近似确定性关系导致的 CI 测试结

果不可靠; ②由于数据稀疏性导致的对于同一组变量测试结果不一致。Opt01SS 算法适用于在有限数据情况下对超级结构进行安全有效的估计, 解释在小样本条件下的近似确定性关系, 解决 CI 测试不一致性的问题。

2016 年, 文献 [23] 提出了一种改进的 PC 算法, 即加速 PC (accclcratcd PC, APC) 算法, 采用“分而治之”的思想, 对一系列局部邻域结构进行推导, 获得贝叶斯网络结构。APC 算法继承了 PC 算法优先执行低阶 CI 测试的优点, 能够从已执行的测试中推导相关拓扑信息, 决定测试顺序, 有效避免无效测试, 减少计算量。实验结论得出, 与 PC 算法相比, APC 算法可以减少高达 50% 的计算量, 同时实现了质量相同的学习效果。

2018 年, 文献 [24] 提出了高斯核概率密度估计贝叶斯网络精确学习方法, 在高斯核概率密度的熵估计方法基础上, 利用互信息计算各变量之间的依赖关系。该算法分别描述了离散型、连续型和混合型数据的具体计算过程。报警网络的实验表明, 该算法在不影响学习效率的前提下, 提高了结构学习的精度, 在处理稀疏的连续数据时, 也有较好的学习效果。

2021 年, 文献 [25] 提出了适用于连续时间 PC (continuous-time PC, CTPC) 贝叶斯网结构学习方法, 包含一组适合测试 CI 的统计量和一个基于 PC 的启发式算法。其优势在于: 与传统贝叶斯网络相比, 可以将模拟的基本随机过程通过已知的图形结构进行表达; 有向图中的因果方向表示可以跟随时间的变化而变化; 专家先验知识可以通过黑名单和白名单的方式整合到结构学习中。贝叶斯网络的无环性对整个网络施加了全局约束, 但 CTPC 算法可以学习独立的父集数据集合, 提高了对局部网络学习的准确性。实验表明, CTPC 算法比连续时间搜索和得分 (continous-time search and score, CTSS) 算法具有更好的结构重构精度, 但随着变量的状态增多, 如变量的状态数大于 20 时, CTPC 算法和 CTSS 算法的性能都会迅速下降。

2022 年, 文献 [26] 提出了一种改进的 PC 算法, 即复杂调查 PC (PC complex surveys, PC-CS) 算法。该算法基于重采样技术与改进的 CI 检验来解释样本选择机制, 解决了复杂调查数据引入的选择偏差。该算法从离散变量数据集进行评估, 并推断网络结构。通过实验比较了 PC 算法和 PC-CS 学习网络的能力, PC-CS 算法的汉明距离仅有 3.4, 表明学习出的网络结构的准确性随着样本数量的增加而提高。

2.1.3　方法总结

在概率分布中, 可能存在 DAG 未反映的条件独立关系, 这种情况称为 DAG 和概率分布彼此不忠诚, 即含有隐藏或潜在变量, 统称为非完备信息。表 1 总结了基于约束的方法的特点及应用场景。

从文献分析可知, 学者们主要从以下角度发展基于约束的贝叶斯网络结构学习方法:

1) 改进互信息和 CI 检验方式, 减少无效测试和高阶测试数量, 提高算法学习效率。

2) 改进 V-结构的定向规则, 提高算法的准确性。

3) 改进算法对于大数据样本或有限小数据样本的灵敏度, 提高模型学习算法的适用性和准确性。

2.2　基于评分搜索的方法

2.2.1　原理概述

基于评分搜索的方法代表了贝叶斯网络结构学习的另一个主要类别, 主要由模型选择和模型优化 2 部分构成。其核心思想是: 将所有可能的结构作为定义域, 将衡量特定结构质量的标准作为评分函数, 将寻找最佳结构的过程看作在定义域范围内寻找评分函数最优值的问题。基于评分搜索的方法由一个初始网络开始, 通过搜索算法对网络结构进行优化操作; 使用评分函数对网络结构打分, 将新网络与旧网络结构的分值进行比较, 若评分高于旧网络, 则对新网络继续进行优化, 直到网络模型不再优化或达到搜索迭代次数限制。

2.2.2　评分函数

基于评分搜索的方法使用评分函数来评估贝叶斯网络结构与给定数据集的拟合程度, 评分函数的适合度直接影响网络模型选择的准确性。评分函数主要包含 2 类: 贝叶斯评分函数和信息论评分函数。

1) 贝叶斯评分函数。对于贝叶斯网络$N=(G,\Theta)$, 关于结构 G 的先验知识概括为概率分布$P(G)$, 称为结构先验分布; 对于给定的结构 G, 关于参数Θ的先验知识被概括为另一个概率分布$P(\Theta|G)$, 称为参数先验分布。N的先验分布可表达为

$$P(G,\Theta)=P(G)P(\Theta|G) \tag{7}$$

当给定观测数据集$K=\{K_1,K_2,\cdots,K_n\}$后, 计算后验概率分布$P(G|K)$。此时, 只考虑最大的后验概率分布对应搜索空间中的结构模型G^*:

$$G^*=\arg\max_G P(G|K) \tag{8}$$

由于

$$P(G|K)=\frac{P(K|G)P(G)}{P(K)} \tag{9}$$

表1 基于约束的方法的特点及应用场景

Table 1 Characteristics and application scenarios of constraint-based algorithm

时间	算法	算法特点	算法应用
1991	SGS	误差较大、效率低，具有一定的稳定性，是后续约束型算法的建立基础	只适用于完备的小型数据集
1997	PC	利用d-分离判断进行CI检测，递归删除边，算法较为准确，效率高，稳定性不如SGS算法，是经典的约束型算法	适用于完备的高阶CI测试集（大型父节点集）
2000	FCI	基于PC算法，通过验证部分祖先图中的2个邻接顶点的CI测试，进而寻找超集D-Sep(A,B)来消除隐变量带来的影响，相比于PC算法更具有普适性，是用于学习含有隐变量数据集的经典约束型算法	适用于非完备的大型数据集
2002	TPDA	TPDA基于SGS算法，使用V-结构和定向相位来定向边缘，减少了所需CI测试数量的限制，相较于传统PC算法具有更高的准确性	适用于完备的大型数据集
2006	RAI	CI测试方式与PC算法类似，在高阶数据集中，相较于PC算法具有更快的运算速度	适用于完备的大型数据集
2010	HPC	区别于大多基于约束的全局搜索算法，采用局部搜索方式，CI测试每次添加新变量时都执行消除步骤，能够更好地处理小样本数据集，但在高阶数据集中学习效率低	只适用于完备的小样本集
2011	RFCI	基于FCI算法，仅利用邻接阶段节点父级的条件集进行学习，并且修改了V-结构的定向规则，从而减少了高阶CI测试量，有效提高了FCI算法的运算速度	适用于非完备的大型数据集
2013	LMM	专用于解决小样本数据情况下网络结构学习正确率低的问题	只适用于完备的数据有限的小样本集
2013	FCI+	基于FCI算法，使用邻接节点父级的条件集的同时能够有效识别祖先集，因此，可以在复杂度更高的数据集中更有效地学习稀疏因果图	适用于非完备的大型数据集
2014	PC-Stable	通过CI测试前重新计算邻接概率，降低了传统PC算法由于样本量有限而导致的错误CI测试决策的影响，在高维图中，PC-Stable算法相较于PC算法具有较高的准确性和稳定性，但运算速度有所降低	适用于完备的数据量有限的小样本集，在大型数据集中准确率和运算速度有所下降
2014	Opt01SS	只执行零阶和一阶CI测试，用于学习数据有限情况下小样本数据的结构	只适用于完备的小样本集
2014	OptHPC	建立在局部搜索的HPC算法的基础上，是一种有效的全局超结构恢复方法，与Opt01-SS算法与HPC算法相比，OptHPC算法在大样本量下具有良好的特异性与准确性	适用于完备的大型数据集
2016	APC	继承了PC算法的高数据效率特性，动态调整CI测试顺序，节省了计算量，因此，APC算法相较于PC算法具有更快的运算速度	适用于完备的大型数据集
2018	MGM-FCI-MAX	基于回归的测试来检测不同变量类型的条件独立性，以及使用最高p值的Sepset来识别V-结构，提高了算法的精确度，另外，算法的并行化缩短了其运行时间	适用于非完备的大型数据集
2021	CTPC	包含一组适合测试CI的统计量和一个基于PC的启发式算法，能够学习连续时间的网络结构，具有较好的精确性和学习效率	适用于完备的连续时间数据集；在大型数据集中精度有所下降
2022	PC-CS	利用重采样技术与改进的CI检验，解决了复杂调查数据引入的选择偏差，PC-CS算法相较于PC-Stable算法具有更高的准确性与运算速度	适用于完备的大型数据集

而$P(K)$不依赖于G，选择后验概率最大的结构就是选择使如下函数达到最大的结构：

$$\log P(G,K)=\log P(K|G)+\log P(G) \tag{10}$$

$\log P(G,K)$称为结构G的贝叶斯评分，先验分布$P(G)$一般被假设为均匀分布，在评分函数中，对数函数log默认以自然对数e为底数，且使用其他底时，通过换底公式转换后，评分函数的相对大小关系不变，因此，不影响网络结构的优化结果。利用贝叶斯评分函数进行结构学习的实质过程是：先找出后验概率最大的结构G^*，再对G^*进行贝叶斯参数估计。

与贝叶斯评分函数选择模型等价的评分函数有CH评分[27]、贝叶斯狄利克雷(Bayesian-Dirichlet, BD)评分[28]、分数等效贝叶斯狄利克雷(Bayesian-Dirichlet equivalent, BDe)评分[29]和贝叶斯狄利克雷等效一致性(Bayesian-Dirichlet equivalent uniform, BDeu)评分等[30]。

2）信息论评分函数。信息论评分函数基于编码长度和信息论的概念将数据集进行压缩，贝叶斯网络的得分与网络对数据集所能实现的压缩程度有关。信息论评分函数包括最小描述长度(minimum description length, MDL)准则、BIC准则、Akaike信息准则(Akaike information criterion, AIC)、归一化极大似然(normalized maximum likelihood, NML)函数和相互信息测试分数(mutual information tests, MIT)[31]等。

经过大量试验分析，贝叶斯评分函数能够应对样本数较大的DAG结构，而信息论评分函数对于小样本数据搜索学习到的网络结构更优[32]。对贝叶斯评分函数和信息论评分函数进行梳理，将其基本表达式和特性进行整理归纳，如表2所示。表中：n为样本变量数目，q_i为X_i父节点个数，r_i为X_i取值数目，m_{ijk}表示在父节点组合j下取值为k的次数，$m_{ij*}=\sum_{k=1}^{r_i}m_{ijk}$为数据中父节点组合$j$出现的次数，$N_{ijk}$为其父节点组合为$j$的样本数，$N_{ij}$为总样本数，$\alpha_{ijk}$为参数先验分布中的超参数，$\alpha_{ij*}=\sum_{k=1}^{r_i}\alpha_{ijk}$，$\mathrm{MIK}\left(X_i,\mathrm{paG}(X_i)\right)$为$X_i$与其父变量间的互信息，$\chi_{\alpha,l_{i\sigma_i(j)}}$表示卡方分布值，$C(G)=0.5\log m\sum_{i=1}^{n}(r_i-1)q_i$为惩罚函数。

表 2　基于评分函数的方法的基本表达式和特性

Table 2　Basic expressions and characteristics of scoring functions based methods

类别	评分函数	评分表达式	特性
贝叶斯评分函数	贝叶斯评分	$\log P(G,K)=\log P(K\|G)+\log P(G)$	最优网络结构满足$P(G\|K)$后验概率分布值最大
	BD	$\mathrm{BD}(G\|K)=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij*})}{\Gamma(\alpha_{ij*}+m_{ij*})}\prod_{k=1}^{r_i}\frac{\Gamma(\alpha_{ijk}+m_{ijk})}{\Gamma(\alpha_{ijk})}$	参数的先验分布符合狄利克雷分布, 但精确度较低
	BDeu	$\mathrm{BDeu}(G\|K)=\log P(G)\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma\left(\frac{\alpha}{q_i}\right)}{\Gamma\left(\frac{\alpha}{q_i}+m_{ij}\right)}\prod_{k=1}^{r_i}\frac{\Gamma\left(\frac{\alpha}{r_iq_i}+m_{ijk}\right)}{\Gamma\left(\frac{\alpha}{r_iq_i}\right)}$	评分算法假设数据D是完整的, 服从均匀分布
	K2	$\mathrm{K2}(G\|K)=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{(r_i-1)!}{(N_{ij}+r_i-1)!}\prod_{k=1}^{r_i}N_{ijk}!$	利用节点序、CH评分来衡量结构的优劣性, 并使用贪婪搜索得到评分最高的网络结构, 是一种经典的评分函数
	CH	$\mathrm{CH}(G\|K)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\left[\log\left(\frac{\Gamma(\alpha_{ij*})}{\Gamma(\alpha_{ij*}+m_{ij*})}\right)+\sum_{k=1}^{r_i}\log\left(\frac{\Gamma(\alpha_{ijk}+m_{ijk})}{\Gamma(\alpha_{ijk})}\right)\right]$	结构先验均匀分布, 需要提前选定超参数a_{ijk}, 当$a_{ijk}=1$时, 则CH评分转变为K2评分
信息论评分函数	BIC	$\mathrm{BIC}(G\|K)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i}m_{ijk}\log\theta_{ijk}-\frac{1}{2}\sum_{i=1}^{n}q_i(r_i-1)\log m$	能够度量结构与数据的拟合度, 具有关于模型复杂度的罚项, 用于大样本数据中, 是一种被广泛使用的评分函数
	MDL	$\mathrm{MDL}(G\|K)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i}m_{ijk}\log\left(\frac{m_{ijk}}{m_{ij}}\right)-\frac{1}{2}C(G)\log m$	对数据进行压缩, 从而降低数据的编码长度
	AIC	$\mathrm{AIC}(G\|K)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i}m_{ijk}\log\left(\frac{m_{ijk}}{m_{ij}}\right)-\sum_{i=1}^{n}(r_i-1)q_i$	当实例数据服从多项分布时, 对MDL进行简化而得, 能够提高模型拟合度
	MIT	$\mathrm{MIT}(G\|K)=\sum_{i=1,\mathrm{paG}(X_i)\neq\varnothing}^{n}\left(2N\cdot\mathrm{MIK}(X_i,\mathrm{paG}(X_i))-\max_{\sigma_{r_i}}\sum_{j=1}^{q_i}\chi_{\alpha,J_{w_{\sigma_i}(j)}}\right)$	基于信息论和CI的评分函数, 通过度量网络结构与观测数据之间的KL距离来评价网络结构的优劣

通过评分函数公式可以发现, 评分函数具有分数可分解性, DAG 的分数由每个变量 X_i 与其父集 $\mathrm{pa}(X_i)$得到的分数的总和构成:

$$\mathrm{score}(G,K)=\sum\mathrm{score}(X_i,\mathrm{pa}(X_i),K)\tag{11}$$

2.2.3　搜索策略及文献分析

模型优化是通过搜索策略在所有可行结构组成的空间中寻找评分最高网络结构的过程。实际上, 这是一个 NP 难题, 即使每个节点的最大父节点数为固定值, 搜索空间也是巨大的。有学者引入专家知识指导搜索过程, 以达到对搜索空间进行限制的目的, 如 K2 算法、K3 算法等; 还有学者采取了启发式搜索策略, 利用启发信息引导搜索过程, 减少搜索范围, 降低问题复杂度, 如贪心算法、模拟退火 (simulated annealing, SA)[33] 算法、粒子群优化 (particle swarm optimization, PSO) 算法等。

1992 年, Cooper 和 Herskovits 提出了著名的 K2 算法[7], 先根据专家经验确定节点顺序, 通过贝叶斯评分准则评估候选结构, 再使用爬山 (hill-climbing, HC) 搜索策略寻找评分最高的模型。K2 算法的精度取决于节点序的优劣, 其使用先验信息优化搜索过程, 该思路被后续贝叶斯网络结构学习方法广为借鉴。1994 年, Bouckaert 提出 K3 算法[34], 采用 MDL 准则, 增加了约束网络复杂程度的罚项函数, 避免学习模型时出现过拟合。2022 年, 为解决 K2 算法学习效果强烈依赖于节点序的问题, 一种新的混合结构学习算法——双重 K2 算法[35] 被提出。首先, 通过 K2 算法的搜索策略得到初始网络结构; 然后, 利用拓扑排序修正节点序; 最后, 使用 K2 算法通过修正后的节点序学习得到最优的网络模型。

在采用启发式搜索策略的算法中, 爬山算法[36] 是一种局部搜索算法, 其核心思想是每次都选择局部最优解, 分数到达稳定状态后结束搜索, 输出最优网络。该算法的缺点是全局开发能力弱, 容易产生局部最优解。为解决局部最优的问题, 学者们提出了基于单个解的优化算法进行改进, 包括禁忌搜索 (tabu search, TS)[37]、模拟退火等。

近年来, 基于自然选择的群体智能优化算法引起了学者的广泛关注。这些优化算法包括遗传算法 (genetic algorithm, GA)[38]、进化规划[39]、蚁群优化 (ant colony optimization, ACO) 算法[40]、粒子群优化算法[41-42]、布谷鸟优化[43]、象群优化[44]、水循环优化[45]、人工蜂群 (artificial bee colony, ABC) 算法[46] 和细菌觅食优化 (bacterial foraging optimization, BFO) 算法[47-48] 等。基于群体智能的优化算法的共同特点是: 使用元启发式搜索机制对贝叶斯网络结构空间进行探索, 通过多次计算对局部模型进行状态更新, 在每次计算时都会创建新的种群来寻找局部最优解。随着循环次数的增加, 解的质量不断提高, 最终可以接近最优模型。

表 3[48] 给出了 5 种算法在 10% 和 20% 噪声数据比例下 Alarm-3000 数据集中学习的详细对比结果。其中, HKS 表示在所有试验中获得的最高

表3 5种算法在不同噪声数据比例下的对比实验结果[48]

Table 3 Comparative experiment results of five algorithms under different noise data proportions[48]

噪声比/%	算法	得分			结构			时间/s		
		HKS	LKS	AKS	SSD	BSD	ASD	SET	LET	AET
10	ACO-B	−14 804.11	−14 810.40	−14 805.79±2.28	2	13	5.55±2.58	358	567	402.5±86.5
	ABC-B	−14 804.11	−14 810.40	−14 804.67±1.46	2	9	4.25±1.26	117	169	141.9±17.0
	BFO-B	−14 804.11	−14 813.26	−14 805.75±2.44	2	6	3.65±1.06	343	477	401.0±36.0
	TS-B	−14 805.81	−14 805.81	−14 805.81±0	2	2	2.00±0	238	245	240.1±2.5
	HC-B	−14 953.06	−14 953.06	−14 953.06±0	11	11	11.00±0	7	8	7.1±0.2
20	ACO-B	−15 247.18	−15 257.07	−15 248.67±2.18	5	14	8.80±2.44	332	495	389.9±47.8
	ABC-B	−15 247.18	−15 256.65	−15 247.70±2.06	8	13	8.50±1.36	114	182	154.9±18.0
	BFO-B	−15 247.18	−15 256.92	−15 252.30±3.50	6	17	10.35±2.74	324	439	376.5±34.9
	TS-B	−15 293.55	−15 293.55	−15 293.55±0	20	20	20.00±0	251	262	254.7±3.1
	HC-B	−15 324.16	−15 324.16	−15 324.16±0	14	14	14.00±0	7	8	7.1±0.3

K2分数；LKS表示在所有试验中获得的最低K2分数；AKS表示在所有试验中获得的平均K2分数；BSD表示在所有试验中获得的最大结构差异；SSD表示在所有试验中结构差异最小；ASD表示在所有试验中收敛的平均结构差异；SET表示在所有试验中获得的最小执行时间；LET表示在所有试验中获得的最长执行时间；AET表示所有试验的平均执行时间。结果以K2分数为准，结构差异最小的网络并不对应得分最高的网络。

与非群体智能优化算法相比，3种群体智能优化算法在每项指标下几乎一致获得最佳结果（10%噪声数据下得分−14 804.11；20%噪声数据下得分−15 247.18）。基于贪婪搜索的算法性能一般，在大多数情况下明显不如3种群体智能优化算法，并且在20%噪声数据下，SSD和BSD从2下降到20，说明在噪声的影响下，结构差异越来越大。基于爬山搜索的算法对噪声的增加不敏感，几乎在所有情况下性能指标都是最差，但从执行时间上是最快的，远远优于其他4种算法，原因是：群体智能优化算法是一种基于群体的优化算法，每个个体在每次迭代中都要独立执行优化任务，并完成个体之间的信息交换，因此，时间成本较高，同时也获得了较好的鲁棒性。

2011年，文献[49]提出了一种基于改进遗传算法的贝叶斯网络结构学习方法，增加了优化变异和修正非法图2个新的算子，代替经典遗传算法中的交叉、选择、变异等操作，新算子不但保持了贝叶斯网络结构学习的多样性和正确性，还能缩小网络结构的搜索空间，提高搜索效率。

2012年，文献[50]提出一种K2与模拟退火相结合的贝叶斯网络结构学习方法——K2SA算法，该算法的主体为模拟退火算法，对贝叶斯网络结构模型的扰动方式进行了修改，通过随机交换当前节点序中2个节点的位置以产生新的节点序，再利用K2算法学习贝叶斯网络结构，作为模拟退火算法中的新状态，以此提高模型全局扰动能力。算法记录了在模拟退火过程中得到的最优网络结构，并在模拟退火过程结束后再利用爬山算法对其作进一步优化。该算法一定程度上改善了模拟退火算法因遍历模型空间能力不足而难以获得全局最优网络结构的问题。实验结果表明，在样本量较为充足的情况下，该算法能够获得近似全局最优的网络结构，具有较好的学习效果。

2021年，文献[51]针对基于粒子群优化算法的贝叶斯网络结构搜索容易陷入局部最优的缺点，提出一种结合灰狼算法的粒子群优化算法。首先，对每次迭代数据中不符合迭代规则的粒子进行优化；其次，由于迭代过程中整个粒子群中最优粒子具有极强的引导搜索能力，对最优粒子进行干扰，增大寻找全局最优的可能性；最后，结合灰狼算法，引导粒子群进行全局包围搜索，寻找全局最优解。

2022年，文献[52]提出了一种基于评分搜索的离散萤火虫优化算法(discrete firefly algorithm, DFA-B)用于学习贝叶斯网络结构。该算法使得每个萤火虫在离散空间中基于重定义的策略进行移动，同时采用变异算子以防算法过早停止而陷入局部最优，利用局部优化器提高萤火虫的使用效率，获得最佳可行解。实验表明，DFA-B算法学习得到的K2分数的标准差为15.2，随着信用网络数据集样本量的增加，算法学习到的结构差异性减小。

2022年，文献[53]提出了一种模型平均爬山(model averaging hill-climbing, MAHC)算法。该算法应用修剪搜索策略对DAG搜索空间进行预处理，在爬山搜索空间上应用模型平均确定最佳网络

结构, 同时采用 BIC 评分函数对网络进行打分。与传统的爬山算法不同, MAHC 算法通过对比所有邻域图, 以此来确定最佳网络结构。实验表明, MAHC 算法在数据噪声存在下的复杂网络中仍具有较好的学习性能。

2.2.4　方法总结

除文献分析外, 表 4[7,47,52-73] 总结了部分基于评分搜索的方法使用的搜索空间、评分函数及搜索策略, 帮助读者在建立贝叶斯网络时更加灵活及有针对性地选择学习方法。

表 4　基于评分搜索的方法的搜索空间、评分函数及搜索策略[7,47,52-73]

Table 4　Search space, scoring function, and search strategy based on method of scoring search[7,47,52-73]

时间	算法	搜索空间	评分函数	搜索策略
1968	Chow-Liu[54]	Tree	Entropy	
1992	K2[7]	DAG	K2	爬山
1995	TS[55]	DAG	K2 & BIC	禁忌搜索
1995	HC[56]	DAG	BDe	爬山
1996	GA[57]	Orderings	K2	遗传算法
2001	K2SN[58]	Orderings	K2	爬山
2002	GES[59]	Equivalence	BDeu	贪婪搜索
2002	ACO-B[60]	DAG	K2	ACO+局部搜索
2005	SGA[61]	Orderings	K2	遗传算法
2025	Hybrid MCMC[62]	DAG	BDeu	MCMC
2008	ACO-K2SN[63]	Orderings	K2	ACO+局部搜索
2009	BN-BQPSO[64]	DAG	BIC	粒子群优化
2009	I-ACO-B[65]	DAG	K2	ACO+局部搜索
2010	IBPSO[66]	DAG	MDL	粒子群优化
2012	GIES[67]	Equivalence	BIC	贪婪搜索
2013	PSO-K2[68]	Orderings	K2	粒子群优化
2017	FGES[69]	Equivalence	BIC	贪婪搜索
2016	BFO-B[47]	DAG	K2	细菌觅食优化
2017	Partition MCMC[70]	DAG	BIC	MCMC
2020	CIGAR[71]	Orderings	BDe	遗传算法
2020	K2-I[72]	DAG	BIC	爬山
2023	FGES-M[73]	Equivalence	BIC	贪婪搜索
2022	MAHC[53]	DAG	BIC	爬山
2022	DFA-B[52]	DAG	K2	萤火虫优化算法

基于评分搜索的方法与基于约束的方法中利用 CI 测试及 V-结构定向的学习策略不同, 需要确定具体的搜索策略与评分函数。目前, 学者们主要从以下方面提升算法性能:

1) 改进经典启发式搜索策略, 如禁忌搜索、爬山算法等, 使其更好地避免陷入局部最优值, 进而学习得到最佳的网络结构。

2) 融合不同类型群体智能优化算法, 如粒子群优化算法、蚁群优化算法等, 达到优势互补的目的, 优化算法性能, 克服收敛速度慢、精确度低等主要问题。

3) 评分函数选择方面, 综合考虑模型对于数据样本的拟合度和模型复杂度, 以提高模型质量和学习效率。

4) 搜索策略设计方面, 综合考虑算法的全局搜索能力和时间复杂度, 以减小在搜索过程中出现局部最优解的几率, 提高算法的搜索效率。

2.3　混合约束和评分搜索的方法

2.3.1　原理概述

事实上, 基于约束的方法实现相对简单, 但性能随着约束集的增多而降低, 对于高阶 CI 检验, 需要足够多的样本数据才能保证学习结果的可靠性。基于评分搜索的方法虽然目前应用较多, 但存在计算复杂、收敛速度较慢等问题。随着各种智能优化算法的发展, 涌现了一批混合约束和评分搜索的贝叶斯网络结构学习方法。

混合方法可以结合基于约束和基于评分搜索方法的优点, 通过 CI 检验缩小、限制搜索空间, 并采用基于评分搜索的方法学习最优网络结构, 对大型网络学习速度有明显提高。

2.3.2 文献分析

早期的混合方法在每次迭代中都有一个约束条件和分数最大化的步骤，Singh等提出了混合方法中的第1个方法——CB算法，利用PC算法确定节点次序及条件设置，再最大化步骤定向无向边，得到邻接节点最高K2评分，重复限制和最大化步骤，增加条件设置大小，直到生成的DAG分数不再提高，完成网络结构学习[74]。随后，基本图搜索(essential graph search, EGS)算法[75]被提出，使用PC算法搜索多个基础图，将每个基础图随机转换为DAG，并使用贝叶斯度量对DAG进行评分，最终使用贝叶斯评分度量搜索评分最大的DAG。类似地，重启I-Map(I-Map restart, IMAPR)算法[76]使用CI测试来寻找一个更优的DAG生成起点，以重新开始局部结构搜索。

最著名的混合方法是最大-最小爬山(max-min hill climbing, MMHC)算法[77]。在约束阶段，MMHC算法采用基于最大-最小父子(max-min parents and children, MMPC)算法使变量进入目标节点的候选父子(candidate parents and children, CPC)节点集之后，再移除多余变量，得到每个节点的父子节点集，从而构建出网络框架，并在最大化阶段融合禁忌搜索和K2评分得到最优DAG结构。

近年来，学者们提出了一些新型混合方法应用于贝叶斯网络结构学习。2017年，文献[78]提出分离与重聚(separation and reunion, SAR)算法，学习大型贝叶斯网络的任务时，先将其分解为相对较小的贝叶斯网络，确切地说，只使用少部分CI检验构造前视部分图[79]，再将其分解成若干大小合适的小子图，使用基于评分搜索的方法从每个子图中学习一个小型贝叶斯网络，通过将这些小型贝叶斯网络统一起来，形成实际的贝叶斯网络整体。

2018年，文献[80]提出了一种连续变量的自适应限制贪婪等价搜索(adaptively restricted GES, ARGES)算法，使用MMPC算法作为限制步骤，并使用改进的贪婪等价搜索(greedy equivalence search, GES)算法作为最大化步骤，构建网络框架。ARGES算法允许在学习完全部分有向无环图(completed partially directed acyclic graph, CPDAG)的同时在V-结构上添加额外边，随着迭代次数的增加，这些额外的屏蔽边缘会自动删除，从而不会出现在最终的CPDAG中，因此，该算法被称为“自适应限制”。实验表明，“自适应限制”意味着ARGES算法具有高维的一致性。

2019年，文献[81]提出了2种数据驱动的对称校正方法，即基于分数的方法symG和基于约束的方法symC，用于确定节点对之间的对称校正问题，并给出了一种结构学习与数据驱动对称修正方法相结合的混合框架，用于提高贝叶斯网络的网络性能，解决非对称的相邻节点对和马尔可夫毯在局部结构之间的冲突问题。

2020年，文献[82]提出一种基于局部拓扑信息分解的贝叶斯网络结构学习方法，主要运用模型分解的思想，且考虑了局部邻域结构的知识。算法主要步骤分为4个阶段：①使用马尔可夫覆盖发现方法绘制无向独立图；②将整个网络拆分为一系列子图；③从观测数据中学习结构规模较小的贝叶斯网络；④通过组合这些小贝叶斯网络来获得最终的贝叶斯网拓扑结构。

2021年，文献[83]将PC算法和粒子群优化算法结合为贝叶斯网络结构学习提供合理的解决方案。首先，使用PC算法生成初始DAG；然后，在搜索过程中结合粒子群优化算法和遗传算法找到最优结构。该算法使用了2种新的变异算子，其中一种被称为添加和删除均匀变异算子，另一种被称为均匀交叉变异算子。该算法将PC算法的输出作为先验结构，改进了贝叶斯网络结构学习的初始解。实验表明，该算法在CANCER网络中的BIC评分为−1 091.88，接近标准BIC评分，能够得到较优的网络结构。

2023年，文献[84]提出了基于PC-SSA算法的贝叶斯网络混合结构学习方法。首先，采用PC算法生成初始网络图作为结构先验，并基于它们生成初始解；然后，采用BIC评分和麻雀搜索算法(sparrow search algorithm, SSA)进行贝叶斯网络结构寻优。与其他启发式搜索算法相比，麻雀搜索算法的时间代价小，设置参数少，在标准网络上的测试分数误差仅为0.2，具有较高的精度。

2.3.3 方法总结

混合方法可以提高模型学习质量和效率，是优化贝叶斯网络结构学习方法的有效手段。通过文献分析，混合方法研究主要集中在以下方面：

1）探索不同人工智能算法的融合能力，发挥智能算法优势，突破传统算法学习短板，提高贝叶斯网络结构学习效率与学习精度。

2）强化对大规模复杂网络的学习能力，利用贝叶斯网络本身特点与数据所提供的信息，缩小搜索空间，用启发式信息指导搜索过程，减少搜索盲目性，提高学习效率。

3 非完备信息下贝叶斯网络结构学习方法

结构学习方法在一般情况下通常假设数据集

是完整的, 但真实的数据集往往存在 2 类情况: ①部分观测变量数据缺失的情况; ②存在无法被观测的变量, 这类变量被称为隐结构变量。针对上述 2 种情况, 学者们研究了非完备信息下的贝叶斯网络结构学习方法。由于现实各领域积累的数据往往具有非完备性, 该方向的研究具有良好的工程应用前景。

3.1　数据缺失的贝叶斯网络结构学习

3.1.1　原理概述

无论是因为偶然性缺失还是系统性缺失, 缺失的数据都可以分为 3 类: 随机缺失 (missing at random, MAR)、完全随机缺失 (missing completely at random, MCAR) 和非随机缺失 (missing not at random, MNAR)[85]。MAR 类型数据的缺失依赖于其他已知变量。MCAR 类型数据的缺失是随机发生的, 且丢失概率与数据变量已知或未知无关。MNAR 类型数据缺失的概率取决于已观测到和未观测到的数据本身。

期望最大化 (expectation maximization, EM) 算法[86]是非完备信息下使用最广泛的网络模型学习算法之一。该算法从网络参数 Θ 的某个初始值 Θ_0 出发开始迭代, 假设进行了 t 次迭代得到估计 Θ_t, 则 $t+1$ 次迭代过程由 2 个步骤组成: 首先, "E-step" 利用估计值 Θ_t 对数据进行补全; 然后, "M-step" 利用补全后的完整数据计算网络参数 Θ 的最大似然估计, 得出 Θ_{t+1}。EM 算法重复以上 2 个步骤, 直至模型不再优化或达到最大迭代次数。

如何将缺失数据合理完备化是非完备信息下贝叶斯网络结构学习的关键, 具体学习流程如图 3 所示。

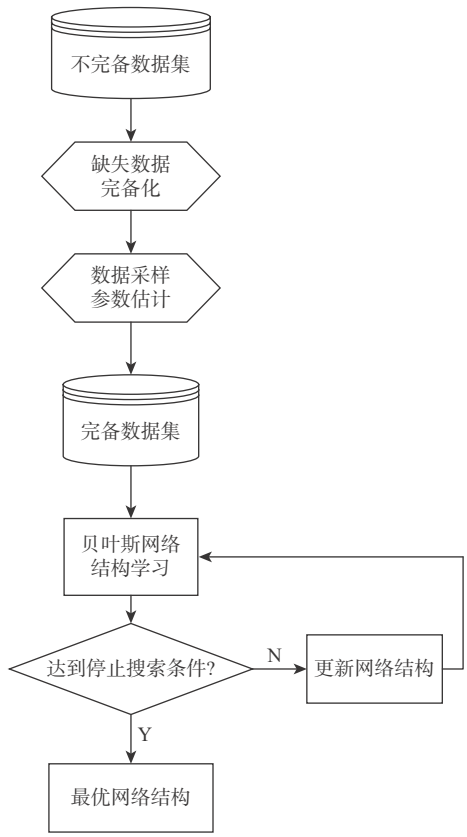


图 3　数据缺失的贝叶斯网络结构学习流程

Fig. 3　Flow of Bayesian network structure learning with incomplete data

3.1.2　文献分析

EM 算法具有对初始化数据敏感、计算复杂、收敛较慢、不适用于大规模数据集和高维数据、容易出现陷入局部最优的问题。为避免以上问题的出现, Friedman 在 EM 算法基础上提出了结构期望最大化 (structural expectation maximization, SEM)[87]算法, 将 EM 算法与结构搜索模型相结合, 学习更优的结构和更加精确的参数。文献 [88] 表明, SEM 算法采用 BIC 和 MDL 评分函数取得了比 EM 算法更好的模型学习结果。

2008 年, 文献 [89] 将进化算法 (evolution algorithm, EA) 与传统 EM 算法相结合, 提出了一种处理具有缺失值的数据库的新方法, 并应用进化算法搜索更优的贝叶斯网络, 实验结果证明了进化贝叶斯网络 (evolutionary Bayesian network, EBN) 结构学习方法可以从非完备的数据集中学习得到更优的网络结构, 同时还探索了 EBN 方法在数据挖掘中应用的可行性。

2018 年, 文献 [90] 提出融合使用 SEM 算法作为贝叶斯网络结构学习的输入, 同时在 EM 算法的最大化步骤中采用 k-MAX 算法, 用于学习有界树宽贝叶斯网络, 实验证明, 采用 k-MAX 算法作为 SEM 算法中的结构学习方法与基于随机森林的方法具有相同估算精度, 但迭代速度更快, 使得 SEM 算法可以应用到含有更多变量的复杂问题中。

除了在 EM 算法框架上进行改进, 学者们近期还提出了其他一些从含有缺失值的数据中进行的结构学习方法。2019 年, 文献 [91] 提出一种基于 Gibbs 采样[92]的马尔可夫链蒙特卡罗 (Markov chain Monte Carlo, MCMC)[93]算法。Gibbs 采样是一种在节点间局部交互的图形结构中执行的采样过程, 用于在难以直接采样时从某一多变量概率分布中近似抽取样本序列, 计算其联合概率分布或积分。该算法用于检测变化点的经验贝叶斯信息准则 (empirical Bayesian information criterion, emBIC), 能同时估计变化点配置和相关参数。首先, 使用经验贝叶斯和最大似然原理推导出模型选择或变化点选择标准; 当观测的后验分布复杂且高维时, EM 算法只能得到后验分布的众数, 此时可以利用 Gibbs

采样对贝叶斯网络的后验分布进行抽样，得到 Gibbs 序列，使其收敛到一个独立于初始值的平稳分布，从而求得如后验均值、方差和分位线等更多的后验量。然后，通过在变化点指示器上诱导多变量离散分布，并使用 Gibbs 采样从诱导分布生成马尔可夫链来估计最佳变化点配置。该算法相比于 EM 算法具有更快的收敛速度，且不易陷入局部最优。

2021 年，文献 [94] 提出了一种基于评分搜索学习下的离散节点平均似然 (node-average likelihood, NAL) 的方法，该方法在结构精度与计算效率上与基于 SEM 的算法相比更有竞争力，对离散贝叶斯网络模型具有可识别性，并证明了 NAL 具有可辨识性和一致性。

2022 年，文献 [95] 提出了基于爬山搜索评分的结构学习与配对删除 (pairwise deletion, PD) 和逆概率加权 (inverse probability weighting, IP-W) 方法之间的融合算法。该算法对缺失数据集加权的情况下执行配对删除，并利用原始配对删除的数据集和标准评分函数计算当前最优 DAG 和相邻 DAG 之间的得分差，能够显著降低由数据缺失引起的偏差。通过实验对比可知，该算法在结构学习效率上优于 SEM 算法，且具有较高的学习准确性。

3.2 含有隐变量的贝叶斯网络结构学习

3.2.1 原理概述

隐变量指在贝叶斯网络结构学习中隐藏或潜在的变量，这些变量可以是实际存在但不能被观测的变量，也可以是为了贝叶斯网络结构学习的合理性而被提出的虚拟变量。例如，在某次医学案例中，训练数据集包含医生观察到的症状（如发烧、头痛、血压等）和处方药物，这些已知变量被称为观测变量；而病人自身不能被观测到的症状就是隐变量。这些隐变量与大多数观测变量是相互独立的，却又能够对患病“原因”和治疗“结果”的解释产生深远的影响。因此，可以通过学习隐藏变量，避免产生对数据的过拟合，提高贝叶斯网络的可解释性，从而更有效地进行推理。

隐变量包括 2 种：①取值个数已知但数值未知的变量，称作值隐变量；②取值个数与数值均未知的变量，称作势隐变量。包含至少一个隐变量的贝叶斯网络模型称为隐变量模型。在实际应用中，往往需要考虑多个隐变量存在的情况，并通过它们与观测变量及自身之间的关系，通过数据分析得到模型中隐变量的个数和势、网络模型结构、模型参数，这一过程称为隐结构模型学习。

含有隐变量的网络结构学习大多采用数据降维的方法，将具有复杂依赖关系的节点变量分解为少数无法直接观测到的公共因子（隐藏变量），以此来确定网络中含有的隐变量和与之对应的观测变量的位置、数量；再通过 EM 算法等补全隐变量的参数。在复杂网络结构中研究隐含变量，能够在简化网络的同时，更深入地了解节点间相互影响的方式。

含有隐变量的贝叶斯网络结构学习方法特性总结如图 4 所示。

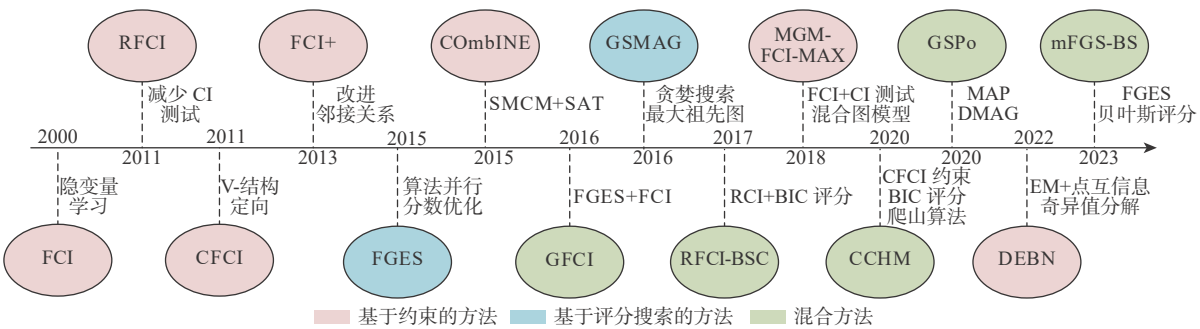


图 4 含有隐变量的贝叶斯网络结构学习方法特性

Fig. 4 Characteristics of Bayesian network structure learning algorithm with hidden variables

3.2.2 基于约束的隐变量学习方法文献分析

最初，在含有隐变量的 DAG 搜索算法中，通常将结构分解算法[96]和信息瓶颈算法[97]与结构搜索、参数估计结合在一起，用于学习含有隐变量的贝叶斯网络。但是，这些算法只是启发式搜索，不能准确确定贝叶斯网络中所含有的隐变量个数及位置，它们的输出取决于输入的初始 DAG。

2000 年，Spirtes 等提出的快速因果推理 (fast causal inference, FCI) 算法[98]是一种基于约束的方法，其以样本数据和可选背景知识作为输入，能够从大样本中学习含有隐变量的等价类因果贝叶斯网络，但该算法难以保证邻接节点识别和有向弧定向的准确性。2011 年，Colombo 等提出的非常真正快速因果推理 (really fast causal inference, RFCI) 和保守快速因果推理 (conservative fast causal inference, CFCI) 算法[99]，以及 2013 年 Claasen 等提出的 FCI+ 算法[100]，都是基于 FCI 的改进算法，通过更精确的初始图来提高 FCI 算法的邻接阶段和定向阶段的

准确性, 并且可以输出一类对观测变量具有马尔可夫等效的图形对象和称为部分祖先图(partial ancestry graph, PAG)的图形模型。PAG 表示等价类中所有 DAG 的共有特性, 在大样本情况下, 给定了无反馈的因果马尔可夫假设和因果忠实假设[101], 可以输出一个含有实际隐变量的 PAG。FCI 早期相关算法的主要限制是在小样本中算法性能差、最终输出网络会出现不同程度的偏差、在大样本中算法运行时间过长等问题。

2015 年, 文献 [102] 提出了一种基于 CFCI 的 COmbINE 算法, 以学习不同数据集的共同特征和 CI 测试结果。COmbINE 算法将数据间的依赖性和独立性转换为半马尔可夫因果模型 (semi Markov causal model, SMCM) 中的路径约束, 并编码为布尔可满足性 (Boolean satisfiability, SAT) 实例。实验结果表明, 在相对稀疏的数据集中, COmbINE 算法具有较好的精度和召回率。

2018 年, 文献 [103] 提出了一种结合混合图形的最大快速因果推理 (mixed graphical models-fast causal inference-max, MGM-FCI-MAX) 算法。算法引入了回归测试, 对邻接节点的所有可能子集执行 CI 检测, 将 CI 检测中 p 值最大的子集作为真正的分离集来识别 V-结构, 以此提高了边的定向精度。该算法还使用了混合图形模型无向图作为邻接阶段的输入, 用于表示混合变量类型之间的条件独立关系。结合以上创新, 通过实验表明, MGM 相比 CFCI 和 FCI, 能够更好地平衡网络学习精度和召回率, 缩短运算时间。

2022 年, 文献 [104] 提出了一种隐变量学习的动态嵌入技术在贝叶斯网络参数学习中的应用。首先, 重建 EM 算法中的 E-step、M-step, 并提出使用动态嵌入来计算分数样本的权重, 以降低网络学习的计算复杂度; 然后, 构建点互信息 (point mutual information, P-MI) 矩阵, 表示从更新的参数转换而来的有向加权图 (directed weighted graph, DWG); 最后, 采用增量奇异值分解 (singular value decomposition, SVD) 生成动态嵌入值, 同时捕获更新的参数, 从而形成贝叶斯网络的图形结构。实验结果表明, 该方法在含有隐变量的贝叶斯网络结构学习上具有较好的性能。

3.2.3　基于评分搜索的隐变量学习方法文献分析

2015 年, Ramsey 提出了一种基于分数的贪婪搜索算法——快速贪婪搜索 (fast greedy search, FGES) 算法[105], 通过算法并行和缓存分数优化 GES 算法的增添、删除边阶段, 提高算法精度。

2016 年, 文献 [106] 提出了一种基于贪婪的最大祖先图搜索 (greedy search for maximal ancestral graph, GSMAG) 算法。首先, 在多元正态性假设下, 采用贪婪搜索策略, 从连续数据中识别出最适合的最大祖先图 (maximal ancestral graph, MAG), 其中, 有向和双向边在邻接之间执行添加、删除、反转或转换操作。然后, 使用 Richardson 等提出的用于 MAG 因子分解的 BIC 评分函数[107], 检查评分最低的网络的局部邻域, 避免重新拟合整个 MAG。在含有隐变量的随机合成高斯网络上评估了 GSMAG 算法与基于 FCI 和 CFCI 约束的算法, 实验表明, GSMAG 算法具有更好的召回率和精度, 但运算速度略有下降。

3.2.4　基于混合的隐变量学习方法文献分析

从观测数据中推断因果 DAG 的马尔可夫等价类问题的主要困难是: 解决在观测变量之间存在的隐变量时, 产生额外的邻接边以及边的定向错误。2016 年, 文献 [108] 提出由 FGES 和 FCI 组成的贪婪快速因果推理 (greedy fast causal inference, GFCI) 算法, 使用 CI 测试删除初始框架中的无关邻接边, 通过修改 FCI 定向规则以产生 PAG 结构图。实验表明, GFCI 算法相较于 FCI 算法提高了学习精度和学习效率, 并在小样本容量下仍具有较好的性能。

2017 年, 文献 [109] 提出了一种基于 RFCI 算法的因果贝叶斯约束评分 (Bayesian scoring of constraints, BSC) 算法的混合方法, 用于学习可能包含隐变量的因果贝叶斯网络(causal Bayesian network, CBN), 该算法调整了 RFCI 算法, 以使用 BSC 评分进行模型评分搜索。当 RFCI 算法请求 CI 测试时, 利用 BSC 来确定独立性成立的概率, 并以对独立性是否成立进行采样, 能同时对多个可能含有隐变量的因果模型进行评分, 并将该结果返回给 RFCI 算法, 最终生成 PAG。将该算法与 RFCI 算法进行对比, 在小样本限制下, 输出模型结构的汉明距离相较于 RFCI 算法有明显改善。

因果不足的结构学习方法倾向于重构贝叶斯网络的原始图, 其中, 有向边表示节点间直接原始关系。2020 年, 文献 [110] 将基于约束的方法与爬山算法相结合提出了保守规则和因果效应爬坡 (conservative rule and causal effect hill-climbing, CCHM) 算法。CCHM 算法假设贝叶斯网络是线性结构方程模型, 其中, 数据遵循多元高斯分布及 CFCI 的约束算法, 最大化 BIC 等效分数。采用 Pearl 的微积分来观测因果效应, 通过爬山搜索结合因果效应定位 MAG 中的无向边。实验表明, CCHM 算法在重建原始图方面具有较好的高效性和准确性。

2020年，文献[111]针对因果关系不足的问题，提出了贪婪稀疏偏序集(greedy sparsest poset, GSPo)算法，偏序集是与有向最大祖先图(directed maximal ancestral graph, DMAG)关联的偏节点排序。GSPo算法提供了从偏序集到最小I-MAP-DMAG的映射，并通过遍历偏序集空间关联至新的DMAG集，每次新的关联都会产生相同或具有更少边的DMAG。实验表明，GSPo算法在高斯网络中的表现比FCI和FCI+算法具有更好的结构精度和更快的运行速度。

2023年，文献[112]提出一种混合多规则及快速贪婪等价搜索与贝叶斯评分(majority rule and fast greedy equivalence search with Bayesian scoring, mFGS-BS)算法，能够从缺乏因果性的离散数据集中进行含有隐变量的结构学习，并生成PAG。利用mFGS-BS算法进行结构学习依赖于混合方法和贝叶斯评分，将该算法与多种算法进行对比，实验结果表明，mFGS-BS算法能够快速学习得到更精确的网络结构。

3.3 非完备信息下贝叶斯网络结构学习方法总结

由于现实应用中使用的数据通常具有非完备性，而非完备信息的处理极大增加了贝叶斯网络结构学习方法设计的难度，该领域的研究引起了学者们的广泛关注。通过文献分析，非完备信息条件下的贝叶斯网络结构学习主要集中于以下方面：

1）关于缺失数据学习，采用对缺失数据进行补全的思路进行网络学习，补全数据主要采用EM算法及其基础框架上衍生出的一系列算法。目前，通过文献可以证实，缺失数据类型下的贝叶斯网络结构学习方法较为成熟，但如何克服已有算法短板，继续改进EM算法框架或提出准确度更高的融合算法，将成为未来该领域的主要研究方向。

2）关于隐变量学习，主要针对在确定隐变量个数及位置信息后，贝叶斯网络节点信息在因果不足下产生错误的邻接关系这一问题进行改进学习。通过改进约束限制、评分条件、定向规则等方式，提高网络结构学习的精确性和完整性。目前，对于隐变量的结构学习，仍存在算法运行效率低、复杂网络学习效果差的问题。混合方法使用可以一定程度上提高算法处理效率及准确度，而融合其他智能算法进行含有隐变量的贝叶斯网络结构学习，为未来学者们进一步的研究提供了新的解决思路。

4 贝叶斯网络结构学习方法应用

贝叶斯网络可以直观地反映出复杂数据模型中各变量之间的因果概率关系，具有良好的可解释性。在工程领域广泛运用贝叶斯网络进行问题域建模，应用一系列推理算法来分析计算模型中各个节点变量之间的关系及状态变化。目前，贝叶斯网络算法的典型应用领域包括军事领域、医学领域、生态及生物信息学应用、工业应用等。

4.1 军事领域

在军事作战领域中，通常因为电子通信干扰、传感器噪声等原因造成部分信息缺失。贝叶斯网络具备在非完备数据、不确定信息情况下进行双向推理的能力，适用于战场作战环境，因此，被广泛应用于战场态势评估、目标毁伤效能评估、智能攻击决策、目标识别等领域。

文献[113]为处理空战环境中的时序模糊信息序列并有效评估态势，构建了一种直觉模糊动态贝叶斯网络(intuitive fuzzy dynamic Bayesian network, IFDBN)。在模型中，将直觉模糊推理机作为虚拟节点嵌入到动态贝叶斯网络中，并提出一种将直觉模糊推理输出转化为概率的新方法，该概率可以作为虚拟信息输入到动态贝叶斯网络中，并给出了详细的推理算法，以空战态势评估为例，验证了模型的可行性。

文献[114]利用贝叶斯网络处理不确定复杂问题的优点，将动态贝叶斯网络与威胁评估模型相结合，建立了用于评估威胁等级的模型。仿真结果表明，动态贝叶斯网络能够较好地对具有实时性、不确定性和不可预知性的战场信息进行威胁等级评估。

文献[115]构建了一种自主机动决策系统，该系统将空战博弈视为马尔可夫过程，利用贝叶斯推理理论对空战态势进行了有效的计算，根据态势评估结果，自适应调整机动决策因子权重，使目标函数更加合理，保证无人战机的优势态势，实验结果证明了所设计机动决策方法的优越性。

4.2 医学领域

近年来，贝叶斯网络已在医学研究中的多个领域得到应用，包括诊断筛检[116]、临床试验[117]、Meta分析[118]等，尤其是临床试验已经贯穿了包括先验信息的优化设计、样本量估算、监测、中期分析和预测等全过程。

2016年，*Machine Learning*杂志刊登了一期关于“医疗保健和医学机器学习”的特刊[119]，讨论了如何开发利用已知的健康监测数据改善对患者的护理模式。机器学习(machine learning, ML)和人工智能(artificial intelligence, AI)方法已被提出用于从遗传学和基因学[120]到治疗方法选择[121]及结果预测[122]的各种医学问题。与来自数据的纯ML方法相比，贝叶斯网络的显著优势是：在数据有限的情况下，可以利用

专家知识对数据进行补充, 从已知的健康监测数据中进行数据分析, 帮助医生实现精准的临床试验决策, 这可能是目前最有效的精确医疗的方法之一。

文献 [123] 提出一个动态的高维贝叶斯推理框架用于学习 COVID-19 模型异质性。使用关于确诊、住院和死亡病例数量的报告数据作为先验概率, 建立不规则的房室模型, 该模型考虑了 COVID-19 在长期护理 (long term care, LTC) 设施内外的时变传播和严重程度的异质性, 提出了一种与维数无关的投影 Stein 变分梯度下降方法用于解决高维推理的维数问题, 专门用于在 COVID-19 传播严重程度不确定条件下进行建模、推理, 得出了基于后验概率的预测和验证结果。该方法用于对 COVID-19 未来传播轨迹进行预测, 以及在风险规避、随机最优控制问题中寻找关于 COVID-19 的缓解策略方面进行了有益的探索。

表 5[124-131] 介绍了贝叶斯网络在各种不同类型疾病中的病情预测及治疗决策等方面的应用。

表 5　贝叶斯网络在各种疾病中的应用[124-131]

Table 5　Application of Bayesian networks in various diseases[124-131]

年份	疾病情况	贝叶斯应用实例	文献引用
2017	中枢神经损伤	使用智能分类器算法预测脊髓损伤患者的疼痛类型	[124]
2018	慢性疾病(糖尿病、关节炎等)	使用DBN的逐步隐变量方法预测慢性疾病并发症	[125]
2015	癌症	深度信念网络用于乳腺癌的预测与诊断	[126]
2017	癌症	使用贝叶斯网络和支持向量机进行乳腺癌手术后生存概率预测	[127]
2016	癌症	使用贝叶斯网络建立乳房病变模型图预测良性肿瘤癌变概率	[128]
2015	心脏疾病	通过电子健康记录数据建立贝叶斯网络模型预测心血管病风险	[129]
2017	心脏疾病	高维数据集中使用贝叶斯网络建立因高血压引起的并发症模型	[130]
2018	传染病	使用贝叶斯网络建立登革热感染诊断模型	[131]

4.3　生态及生物信息学应用

文献 [132] 将水、能源与生物生态关系耦合到贝叶斯网络中, 提供了系统联合概率行为的形式化表示, 分析了过去 50 年锡尔河流域的用水状况, 为水资源管理和决策提供了优化方案, 并证明了该技术能够缓解流域内的生态危机。

文献 [133] 从生态数据中学习贝叶斯网络, 对岩石海岸生态系统的物种数据应用贝叶斯网络推理, 建立具有记录链接的系统, 将贝叶斯网络评估作为一种变量选择方法, 训练网络中每个分组物种的神经网络, 证明了贝叶斯网络在解释生态网络方面具有很大潜力, 以及在处理复杂非线性数据时的强大预测能力。

文献 [134] 提出一种对联结构造贝叶斯网络 (pair-copula constructions Bayesian network, PCCBN) 模型, 对在生物学和生物医学中广泛应用的基因调控网络 (gene regulatory network, GRN) 进行建模。在基因正态性假设下, 通过考虑基因间的条件独立性, 对逆协方差矩阵 (或精度矩阵) 的非零元素进行评估, 构建了关于 GRN 的贝叶斯网络模型, 并将 PCCBN 算法与高斯贝叶斯网络进行了比较。

4.4　工业应用

近年来, 贝叶斯网络还与其他数据分析方法相结合, 广泛应用于预测分析、可靠性分析或故障诊断等领域, 用于保障工业系统正常运行及提升系统的安全性。例如, 利用贝叶斯网络进行故障诊断, 是贝叶斯网络在工业中的典型应用。对每个故障源及系统建立对应的贝叶斯网络故障诊断模型, 利用贝叶斯网络对系统进行反向推理, 可以推算出系统出现故障时, 每个故障源发生故障的概率。贝叶斯网络在故障诊断、风险评估典型的应用案例如表 6[135-139] 所示。

表 6　贝叶斯网络在故障诊断、风险评估等工业应用[135-139]

Table 6　Industrial applications of Bayesian networks in fault diagnosis and risk assessment[135-139]

年份	故障类型	概述	文献引用
2015	智能故障诊断	提出了一种两级流程操作故障诊断技术, 利用贝叶斯网络分析当存在未被监控的流程变量时, 无法准确定位故障原因的限制条件	[135]
2016	智能故障诊断	通过结合历史和在线信息, 开发了一种贝叶斯故障诊断系统, 以解决异步测量问题, 可用于事件树分析	[136]
2018	工业故障诊断	通过故障树和事件序列图(FT-ESD)分析, 建立了系统的紧急断开故障和后果模型图, 将FT-ESD模型映射到贝叶斯网络中, 提出了一种利用贝叶斯网络进行紧急断开操作故障概率分析的方法	[137]
2018	工业故障诊断	提出了一种冷冻机组故障诊断方法DR-BN, 并将该方法用于评价特征选取结果	[138]
2016	动态风险评估	提出了一种基于贝叶斯网络对液化天然气运输船卸载过程动态风险评估的新方法, 能够识别卸载操作中的危险事件, 并对关键事件进行失效分析建模	[139]

4.5　贝叶斯网络结构学习方法应用总结

贝叶斯网络被证实为是一种表达、推理不确定知识的强有力工具, 其应用已经涉及预测、智能推理、诊断、决策、可靠性分析等众多领域。经文献

分析可知，未来贝叶斯网络将广泛应用于人类生产生活等众多方面，这对于贝叶斯网络的深入研究和社会经济发展都具有重要意义。

5 结束语

贝叶斯网络因其“白盒性”、工程易用性引起了学者和工程技术人员的广泛关注。回顾贝叶斯网络的发展历程，针对贝叶斯网络在复杂问题域中存在的模型维数高、学习难度大、易陷入局部最优解等难题，学者们从不同角度提出了多种算法进行改进，并在多个领域的应用中验证了贝叶斯网络的研究和推广价值。

本文回顾了贝叶斯网络的发展历程，介绍了其基本原理。将贝叶斯网络结构学习方法归为3类：基于约束的方法、基于评分搜索的方法及混合方法，并对其各自发展情况进行了梳理和分析。概括来说，基于约束的方法以变量间的CI检验为基础，识因果依赖关系，在数据样本完备的条件下，能够学习获得质量较高的贝叶斯网络模型；基于评分搜索的方法设计评分函数作为贝叶斯网络模型选择指标，多以启发式知识指导模型搜索过程，通过多轮迭代持续优化模型；混合方法可以结合基于约束和基于评分搜索的方法的优点，通过CI检验缩小搜索空间，并采用基于评分搜索的方法进行网络结构学习，对大型网络结构学习速度有明显提高。针对现实应用中的数据往往具有非完备性，从缺失数据处理和隐变量学习2个维度阐释了非完备贝叶斯网络结构学习的研究现状。总结了贝叶斯网络在多个领域中的工程应用情况，揭示了其研究意义和具有的广阔应用前景。

综合分析，未来贝叶斯网络结构学习具有如下发展趋势：

1）非完备信息条件下的贝叶斯网络结构学习。在军事决策、大型设备故障诊断等复杂应用领域，非完备信息的存在极大增加了贝叶斯网络结构学习的难度，如何准确地将缺失数据完备化，以及提高含有隐变量模型的学习精确度是未来贝叶斯网络结构学习研究的热点之一。

2）大型或高维度贝叶斯网络结构学习。目前，贝叶斯网络结构学习涉及范围逐渐广泛，复杂领域的需求使得贝叶斯网络算法逐渐向着善于学习大型网络模型的方向发展。网络中节点的增加会导致贝叶斯网络结构学习准确性降低。因此，在高维度贝叶斯网络算法研究中通常会出现局部最优解的情况，导致最终结果偏差较大。如何解决这些问题，寻求适应大型贝叶斯网络结构的学习方法也是未来的研究重点。

3）融合人工智能算法的贝叶斯网络结构学习。随着人工智能技术的进步，很多学者已经尝试研究融合不同智能算法的贝叶斯网络结构混合学习方法，以发挥不同智能算法的优点，达到学习复杂程度更高的模型、提升算法性能的目的。可以预期，随着智能算法研究的不断深入，通过融合先进算法来不断完善贝叶斯网络结构学习方法将成为未来研究的发展趋势。

参考文献（References）

[1] LEE S J, SIAU K. A review of data mining techniques[J]. Industrial Management & Data Systems, 2001, 101(1): 41-46.

[2] HECKERMAN D, MAMDANI A, WELLMAN M P. Real-world applications of Bayesian networks[J]. Communications of the ACM, 1995, 38(3): 24-26.

[3] SPIRTES P, GLYMOUR C, SCHEINES R. Reply to Humphreys and Freedman’s review of causation, prediction, and search[J]. The British Journal for the Philosophy of Science, 1997, 48(4): 555-568.

[4] COLOMBO D, MAATHUIS M H. Order-independent constraint-based causal structure learning[EB/OL]. (2013-09-27)[2023-07-01]. https://arxiv.org/abs/1211.3295v2.

[5] LIU J X, TIAN Z L. Verification of three-phase dependency analysis Bayesian network learning method for maize carotenoid gene mining[J]. BioMed Research International, 2017, 2017: 1813494.

[6] MASEGOSA A R, MORAL S. An interactive approach for Bayesian network learning using domain/expert knowledge[J]. International Journal of Approximate Reasoning, 2013, 54(8): 1168-1181.

[7] COOPER G F, HERSKOVITS E. A Bayesian method for the induction of probabilistic networks from data[J]. Machine Learning, 1992, 9(4): 309-347.

[8] NEATH A A, CAVANAUGH J E. The Bayesian information criterion: background, derivation, and applications[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2012, 4(2): 199-203.

[9] LIU H, CAO Y H. Study of heuristic search and exhaustive search in search algorithms of the structural learning[C]//Proceedings of the 2nd International Conference on Multimedia and Information Technology. Piscataway: IEEE Press, 2010: 169-171.

[10] CHAKRABORTY A, KAR A K. Swarm intelligence: a review of algorithms[M]//PATNAIK S, YANG X S, NAKAMATSU K. Nature-inspired computing and optimization. Berlin: Springer, 2017: 475-494.

[11] GEFFNER H, DECHTER R, HALPERN J Y. Probabilistic and causal inference: the works of Judea Pearl[M]. New York: ACM, 2022.

[12] PEARL J. Causality: models, reasoning and inference[M]. Cambridge: Cambridge University Press, 2000.

[13] DAGUM P, GALPER A. Additive belief-network models[C]//Proceedings of the Uncertainty in Artificial Intelligence. Amsterdam: Elsevier, 1993: 91-98.

[14] BATCHELOR C, CAIN J. Application of belief networks to water management studies[J]. Agricultural Water Management, 1999,

40(1): 51-57.

[15] TICEHURST J L, NEWHAM L T H, RISSIK D, et al. A Bayesian network approach for assessing the sustainability of coastal lakes in New South Wales, Australia[J]. Environmental Modelling & Software, 2007, 22(8): 1129-1139.

[16] SPIRTES P, GLYMOUR C, SCHEINES R. From probability to causality[J]. Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 1991, 64(1): 1-36.

[17] GEIGER D, VERMA T, PEARL J. D-separation: from theorems to algorithms[M]//HENRION M, SHACHTER R D, KANAL L N, et al. Uncertainty in artificial intelligence. Amsterdam: Elsevier, 1990: 139-148.

[18] CHENG J, GREINER R, KELLY J, et al. Learning Bayesian networks from data: an information-theory based approach[J]. Artificial Intelligence, 2002, 137(1-2): 43-90.

[19] YEHEZKEL R, LERNER B. Bayesian network structure learning by recursive autonomy identification[C]//Proceedings of the Structural, Syntactic, and Statistical Pattern Recognition. Berlin: Springer, 2006: 154-162.

[20] DE MORAIS S R, AUSSEM A. An efficient and scalable algorithm for local Bayesian network structure discovery[C]//Proceedings of the Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2010: 164-179.

[21] MAHDI R, MEZEY J. Sub-local constraint-based learning of Bayesian networks using a joint dependence criterion[J]. Journal of Machine Learning Research, 2013, 14(1): 1563-1603.

[22] VILLANUEVA E, MACIEL C D. Efficient methods for learning Bayesian network super-structures[J]. Neurocomputing, 2014, 123: 3-12.

[23] MINN S, 傅顺开. 贝叶斯网络结构加速学习算法[J]. 计算机科学, 2016, 43(2): 263-268.
MINN S, FU S K. Accelerating structure learning of Bayesian network[J]. Computer Science, 2016, 43(2): 263-268(in Chinese).

[24] JIANG Y L, LIANG Z Z, GAO H, et al. An improved constraint-based Bayesian network learning method using Gaussian kernel probability density estimator[J]. Expert Systems with Applications, 2018, 113: 544-554.

[25] BREGOLI A, SCUTARI M, STELLA F. A constraint-based algorithm for the structural learning of continuous-time Bayesian networks[J]. International Journal of Approximate Reasoning, 2021, 138: 105-122.

[26] MARELLA D, VICARD P. Bayesian network structural learning from complex survey data: a resampling based approach[J]. Statistical Methods & Applications, 2022, 31(4): 981-1013.

[27] ZGUROVSKII M Z, BIDYUK P I, TERENT'EV A N. Methods of constructing Bayesian networks based on scoring functions[J]. Cybernetics and Systems Analysis, 2008, 44(2): 219-224.

[28] SIMONS K T, KOOPERBERG C, HUANG E, et al. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions[J]. Journal of Molecular Biology, 1997, 268(1): 209-225.

[29] YANG S L, CHANG K C. Comparison of score metrics for Bayesian network learning[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2002, 32(3): 419-428.

[30] SCUTARI M. An empirical-Bayes score for discrete Bayesian networks[C]//Proceedings of the Conference on Probabilistic Graphical Models. [S. l.]: PMLR, 2016: 438-448.

[31] DE CAMPOS L M, FRIEDMAN N. A scoring function for learning Bayesian networks based on mutual information and conditional independence tests[J]. Journal of Machine Learning Research, 2006, 7(10): 2149-2187.

[32] DE CAMPOS L M, FERNÁNDEZ-LUNA J M, GÁMEZ J A, et al. Ant colony optimization for learning Bayesian networks[J]. International Journal of Approximate Reasoning, 2002, 31(3): 291-311.

[33] VAN LAARHOVEN P J M, AARTS E H L. Simulated annealing [M]//VAN LAARHOVEN P J M, AARTS E H L. Simulated annealing: theory and applications. Berlin: Springer, 1987: 7-15.

[34] BOUCKAERT R R. Properties of Bayesian belief network learning algorithms[C]//Proceedings of the 10th Annual Conference on Uncertainty in Artificial Intelligence . Amsterdam: Elsevier, 1994: 102-109.

[35] 李晓晴, 于海征. 贝叶斯网络结构学习的双重 K2 算法[J]. 科学技术与工程, 2022, 22(24): 10602-10610.
LI X Q, YU H Z. Double K2 algorithm for Bayesian network structure learning[J]. Science Technology and Engineering, 2022, 22(24): 10602-10610(in Chinese).

[36] SELMAN B, GOMES C P. Hill-climbing search[J]. Encyclopedia of Cognitive Science, 2006, 81: 82.

[37] GLOVER F, LAGUNA M. Tabu search[M]//DU D Z, PARDALOS P M. Handbook of combinatorial optimization. Berlin: Springer, 1998: 2093-229.

[38] 刘大有, 王飞, 卢奕南, 等. 基于遗传算法的 Bayesian 网结构学习研究[J]. 计算机研究与发展, 2001, 38(8): 916-922.
LIU D Y, WANG F, LU Y N, et al. Research on learning Bayesian network structure based on genetic algorithms[J]. Journal of Computer Research and Development, 2001, 38(8): 916-922(in Chinese).

[39] CUI G, WONG M L, LUI H K. Machine learning for direct marketing response models: Bayesian networks with evolutionary programming[J]. Management Science, 2006, 52(4): 597-612.

[40] GÁMEZ J A, PUERTA J M. Searching for the best elimination sequence in Bayesian networks by using ant colony optimization [J]. Pattern Recognition Letters, 2002, 23(1-3): 261-277.

[41] GHEISARI S, MEYBODI M R. BNC-PSO: structure learning of Bayesian networks by particle swarm optimization[J]. Information Sciences, 2016, 348: 272-289.

[42] SAHIN F, DEVASIA A. Distributed particle swarm optimization for structural Bayesian network learning[M]//CHAN F, TIWARI M. Swarm intelligence, focus on ant and particle swarm optimization. London: IntechOpen Limited, 2007: 505-532.

[43] ASKARI M B A, AHSAEE M G. Bayesian network structure learning based on cuckoo search algorithm[C]//Proceedings of the 6th Iranian Joint Congress on Fuzzy and Intelligent Systems. Piscataway: IEEE Press, 2018: 127-130.

[44] KAREEM S W, OKUR M C. Evaluation of Bayesian network structure learning using elephant swarm water search algorithm [M]//SHI C. Handbook of research on advancements of swarm intelligence algorithms for solving real-world problems. Hershey:

Engineering Science Reference, 2020: 139-159.

[45] WANG J Y, LIU S Y. Novel binary encoding water cycle algorithm for solving Bayesian network structures learning problem[J]. Knowledge-Based Systems, 2018, 150: 95-110.

[46] JI J Z, WEI H K, LIU C N. An artificial bee colony algorithm for learning Bayesian networks[J]. Soft Computing, 2013, 17(6): 983-994.

[47] YANG C C, JI J Z, LIU J M, et al. Structural learning of Bayesian networks by bacterial foraging optimization[J]. International Journal of Approximate Reasoning, 2016, 69: 147-167.

[48] JI J Z, YANG C C, LIU J M, et al. A comparative study on swarm intelligence for structure learning of Bayesian networks[J]. Soft Computing, 2017, 21(22): 6713-6738.

[49] 张亮, 章兢. 改进遗传优化的贝叶斯网络结构学习[J]. 计算机系统应用, 2011, 20(9): 68-72.
ZHANG L, ZHANG J. Structure learning of BN based on improved genetic algorithm[J]. Computer Systems & Applications, 2011, 20(9): 68-72(in Chinese).

[50] 金焱, 胡云安, 张瑾, 等. K2与模拟退火相结合的贝叶斯网络结构学习[J]. 东南大学学报(自然科学版), 2012, 42(增刊1): 82-86.
JIN Y, HU Y A, ZHANG J, et al. Bayesian network structure learning based on K2 and simulated annealing[J]. Journal of Southeast University (Natural Science Edition), 2012, 42(Sup 1): 82-86(in Chinese).

[51] 潘成胜, 张斌, 吕亚娜, 等. 改进灰狼优化算法的K-Means文本聚类[J]. 计算机工程与应用, 2021, 57(1): 188-193.
PAN C S, ZHANG B, LYU Y N, et al. K-Means text clustering based on improved gray wolf optimization algorithm[J]. Computer Engineering and Applications, 2021, 57(1): 188-193(in Chinese).

[52] WANG X C, REN H J, GUO X X. A novel discrete firefly algorithm for Bayesian network structure learning[J]. Knowledge-Based Systems, 2022, 242: 108426.

[53] CONSTANTINOU A C, LIU Y, KITSON N K, et al. Effective and efficient structure learning with pruning and model averaging strategies[J]. International Journal of Approximate Reasoning, 2022, 151: 292-321.

[54] CHOW C, LIU C. Approximating discrete probability distributions with dependence trees[J]. IEEE Transactions on Information Theory, 1968, 14(3): 462-467.

[55] BOUCKAERT R R. Bayesian belief networks: from construction to inference[D]. Utrecht: Utrecht University, 1995.

[56] HECKERMAN D, GEIGER D, CHICKERING D M. Learning Bayesian networks: the combination of knowledge and statistical data[C]//Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence. Amsterdam: Elsevier, 1994: 293-301.

[57] LARRANAGA P, KUIJPERS C M H, MURGA R H, et al. Learning Bayesian network structures by searching for the best ordering with genetic algorithms[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 1996, 26(4): 487-493.

[58] DE CAMPOS L M, PUERTA J M. Stochastic local and distributed search algorithms for learning belief networks[C]//Proceedings of the Ⅲ International Symposium on Adaptive Systems: Evolutionary Computation and Probabilistic Graphical Model. Berlin: Springer, 2001: 109-115.

[59] CHICKERING D M. Optimal structure identification with greedy search[J]. Journal of Machine Learning Research, 2002, 3: 507-554.

[60] DE CAMPOS L M, GÁMEZ J A, PUERTA J M. Learning Bayesian networks by ant colony optimisation: searching in two different spaces[J]. Mathware & Soft Computing, 2002, 9: 2-3.

[61] SHETTY S, SONG M. Structure learning of Bayesian networks using a semantic genetic algorithm-based approach[C]//Proceedings of the 3rd International Conference on Information Technology: Research and Education. Piscataway: IEEE Press, 2005: 454-458.

[62] CHAN L S H, CHU A M Y, SO M K P. A hybrid Markov chain Monte Carlo approach for structural learning in Bayesian networks based on variable blocking[EB/OL]. [2023-07-01]. https://doi.org/10.1214/25-BA1521.

[63] SALEHPOUR A A, MIRMOBIN B, AFZALI-KUSHA A, et al. An energy efficient routing protocol for cluster-based wireless sensor networks using ant colony optimization[C]//Proceedings of the International Conference on Innovations in Information Technology. Piscataway: IEEE Press, 2008: 455-459.

[64] ZHAO J, SUN J, XU W B, et al. Structure learning of Bayesian networks based on discrete binary quantum-behaved particle swarm optimization algorithm[C]//Proceedings of the 5th International Conference on Natural Computation. Piscataway: IEEE Press, 2009: 86-90.

[65] JI J Z, ZHANG H X, HU R B, et al. A Bayesian network learning algorithm based on independence test and ant colony optimization [J]. Acta Automatica Sinica, 2009, 35(3): 281-288.

[66] LI X L. A particle swarm optimization and immune theory-based algorithm for structure learning of Bayesian networks[J]. International Journal of Database Theory and Application, 2010, 3(2): 61-69.

[67] HAUSER A, BÜHLMANN P. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs[J]. Journal of Machine Learning Research, 2012, 13(1): 2409-2464.

[68] AOUAY S, JAMOUSSI S, AYED Y B. Particle swarm optimization based method for Bayesian network structure learning[C]//Proceedings of the 5th International Conference on Modeling, Simulation and Applied Optimization. Piscataway: IEEE Press, 2013: 1-6.

[69] RAMSEY J, GLYMOUR M, SANCHEZ-ROMERO R, et al. A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images[J]. International Journal of Data Science and Analytics, 2017, 3(2): 121-129.

[70] SCANAGATTA M, CORANI G, ZAFFALON M. Improved local search in Bayesian networks structure learning[C]//Proceedings of the Advanced Methodologies for Bayesian Networks. [S. l.]: PMLR, 2017: 45-56.

[71] JOSE S, LOUIS S J, DASCALU S M, et al. Bayesian network structure learning using case-injected genetic algorithms[C]//Proceedings of the IEEE 32nd International Conference on Tools with Artificial Intelligence. Piscataway: IEEE Press, 2020: 572-579.

[72] BEHJATI S, BEIGY H. Improved K2 algorithm for Bayesian network structure learning[J]. Engineering Applications of Artificial Intelligence, 2020, 91: 103617.

[73] BERNAOLA N, MICHIELS M, LARRAÑAGA P, et al. Learning massive interpretable gene regulatory networks of the human brain by merging Bayesian networks[J]. PLoS Computational Biology, 2023, 19(12): e1011443.

[74] SINGH M, VALTORTA M. Construction of Bayesian network structures from data: a brief survey and an efficient algorithm[J]. International Journal of Approximate Reasoning, 1995, 12(2): 111-131.

[75] DASH D, DRUZDZEL M J. A hybrid anytime algorithm for the construction of causal models from sparse data[EB/OL]. (2013-01-23)[2023-07-01]. https://arxiv.org/abs/1301.6689.

[76] DE CAMPOS L M, FERNÁNDEZ-LUNA J M, PUERTA J M. An iterated local search algorithm for learning Bayesian networks with restarts based on conditional independence tests[J]. International Journal of Intelligent Systems, 2003, 18(2): 221-235.

[77] TSAMARDINOS I, BROWN L E, ALIFERIS C F. The max-min hill-climbing Bayesian network structure learning algorithm[J]. Machine Learning, 2006, 65(1): 31-78.

[78] LIU H, ZHOU S G, LAM W, et al. A new hybrid method for learning Bayesian networks: separation and reunion[J]. Knowledge-Based Systems, 2017, 121: 185-197.

[79] CASTELO R, ROVERATO A. A robust procedure for Gaussian graphical model search from microarray data with p larger than n [J]. Journal of Machine Learning Research, 2006, 7(12): 2621-2650.

[80] NANDY P, HAUSER A, MAATHUIS M H. High-dimensional consistency in score-based and hybrid structure learning[J]. The Annals of Statistics, 2018, 46(6A): 3151-3183.

[81] ZHAO J J, HO S S. Improving Bayesian network local structure learning via data-driven symmetry correction methods[J]. International Journal of Approximate Reasoning, 2019, 107: 101-121.

[82] DAI J G, REN J, DU W C. Decomposition-based Bayesian network structure learning algorithm using local topology information[J]. Knowledge-Based Systems, 2020, 195: 105602.

[83] SUN B D, ZHOU Y, WANG J J, et al. A new PC-PSO algorithm for Bayesian network structure learning with structure priors[J]. Expert Systems with Applications, 2021, 184: 115237.

[84] 仝兆景, 李金香, 乔征瑞. 一种具有结构先验的贝叶斯网络结构学习算法 [J]. 电子科技, 2023, 36(11): 1-7.
TONG Z J , LI J X, QIAO Z R. A structural learning algorithm for Bayesian networks with structural priori[J]. Electronic Science and Technology, 2023, 36(11): 1-7(in Chinese).

[85] RESSEGUIER N, GIORGI R, PAOLETTI X. Sensitivity analysis when data are missing not-at-random[J]. Epidemiology, 2011, 22(2): 282.

[86] DEMPSTER A P, LAIRD N M, RUBIN D B. Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1977, 39(1): 1-22.

[87] FRIEDMAN N. The Bayesian structural EM algorithm[EB/OL]. (2013-01-30)[2023-07-01]. http://arxiv.org/abs/1301.7373.

[88] FRIEDMAN N. Learning belief networks in the presence of missing values and hidden variables[C]//Proceedings of the Fourteenth International Conference on Machine Learning. New York: ACM, 1997: 125-133.

[89] WONG M L, GUO Y Y. Learning Bayesian networks from incomplete databases using a novel evolutionary algorithm[J]. Decision Support Systems, 2008, 45(2): 368-383.

[90] SCANAGATTA M, CORANI G, ZAFFALON M, et al. Efficient learning of bounded-treewidth Bayesian networks from complete and incomplete data sets[J]. International Journal of Approximate Reasoning, 2018, 95: 152-166.

[91] QIAN G Q, WU Y H, XU M. Multiple change-points detection by empirical Bayesian information criteria and Gibbs sampling induced stochastic search[J]. Applied Mathematical Modelling, 2019, 72: 202-216.

[92] GELFAND A E. Gibbs sampling[J]. Journal of the American Statistical Association, 2000, 95(452): 1300.

[93] BROOKS S. Markov chain Monte Carlo method and its application[J]. Journal of the Royal Statistical Society: Series D (the Statistician), 1998, 47(1): 69-100.

[94] BODEWES T, SCUTARI M. Learning Bayesian networks from incomplete data with the node-average likelihood[J]. International Journal of Approximate Reasoning, 2021, 138: 145-160.

[95] LIU Y, CONSTANTINOU A C. Greedy structure learning from data that contain systematic missing values[J]. Machine Learning, 2022, 111(10): 3867-3896.

[96] ELIDAN G, LOTNER N, FRIEDMAN N, et al. Discovering hidden variables: a structure-based approach[C]//Proceedings of the Advances in Neural Information Processing Systems. [S.l.]: NIPS, 2000.

[97] ELIDAN G, FRIEDMAN N. Learning hidden variable networks: the information bottleneck approach[J]. Journal of Machine Learning Research, 2005, 6: 81-127.

[98] SPIRTES P, GLYMOUR C N, SCHEINES R, et al. Causation, prediction, and search[M]. Cambridge: MIT Press, 2000.

[99] COLOMBO D, MAATHUIS M H, KALISCH M, et al. Learning high-dimensional DAGs with latent and selection variables[C]// Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence. [S. l.]: DBLP, 2011: 850-850.

[100] CLAASSEN T, MOOIJ J, HESKES T. Learning sparse causal models is not NP-hard[EB/OL]. (2013-09-26)[2023-07-01]. http://arxiv.org/abs/1309.6824.

[101] SCHEINES R. An introduction to causal inference[EB/OL]. [2023-07-01]. https://www.academin.edu/3135379/An_introduction_to_causal_inference.

[102] TRIANTAFILLOU S, TSAMARDINOS I. Constraint-based causal discovery from multiple interventions over overlapping variable sets[J]. The Journal of Machine Learning Research, 2015, 16(1): 2147-2205.

[103] RAGHU V K, RAMSEY J D, MORRIS A, et al. Comparison of strategies for scalable causal discovery of latent variable models from mixed data[J]. International Journal of Data Science and Analytics, 2018, 6(1): 33-45.

[104] QI Z W, YUE K, DUAN L, et al. Dynamic embeddings for efficient parameter learning of Bayesian network with multiple latent variables[J]. Information Sciences, 2022, 590: 198-216.

[105] RAMSEY J D. Scaling up greedy causal search for continuous

variables[EB/OL]. (2015-10-11)[2023-07-01]. https://arxiv.org/abs/1507.07749v2.

[106] TRIANTAFILLOU S, TSAMARDINOS I. Score-based vs constraint-based causal learning in the presence of confounders [C]//Proceedings of the Uncertainty in Artificial Intelligence. New York:[s.n.], 2016: 59-67.

[107] DRTON M, EICHLER M, RICHARDSON T S. Computing maximum likelihood estimates in recursive linear models with correlated errors[J]. Journal of Machine Learning Research, 2009, 10: 2329-2348.

[108] OGARRIO J M, SPIRTES P, RAMSEY J. A hybrid causal search algorithm for latent variable models[C]//Proceedings of the Conference on probabilistic graphical models. [S. l.]: PMLR, 2016: 368-379.

[109] JABBARI F, RAMSEY J, SPIRTES P, et al. Discovery of causal models that contain latent variables through Bayesian scoring of independence constraints[C]//Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2017: 142-157.

[110] CHOBTHAM K, CONSTANTINOU A C. Bayesian network structure learning with causal effects in the presence of latent variables[C]//Proceedings of the International Conference on Probabilistic Graphical Models. [S. l.]: PMLR, 2020: 101-112.

[111] BERNSTEIN D, SAEED B, SQUIRES C, et al. Ordering-based causal structure learning in the presence of latent variables[C]// Proceedings of the International Conference on Artificial Intelligence and Statistics. [S. l.]: PMLR, 2020: 4098-4108.

[112] CHOBTHAM K, CONSTANTINOU A C, KITSON N K. Hybrid Bayesian network discovery with latent variables by scoring multiple interventions[J]. Data Mining and Knowledge Discovery, 2023, 37(1): 476-520.

[113] WANG L, YANG H Y, ZHANG S W, et al. Intuitionistic fuzzy dynamic Bayesian network and its application to terminating situation assessment[J]. Procedia Computer Science, 2019, 154: 238-248.

[114] 晏师励, 李德华. 基于动态贝叶斯网络的空战目标威胁等级评估[J]. 计算机与数字工程, 2015, 43(12): 2150-2154.
YAN S L, LI D H. Threat level assessment of the air combat target based on DBN[J]. Computer & Digital Engineering, 2015, 43(12): 2150-2154(in Chinese).

[115] HUANG C Q, DONG K S, HUANG H Q, et al. Autonomous air combat maneuver decision using Bayesian inference and moving horizon optimization[J]. Journal of Systems Engineering and Electronics, 2018, 29(1): 86-97.

[116] BOURS M J. Bayes' rule in diagnosis[J]. Journal of Clinical Epidemiology, 2021, 131: 158-160.

[117] 刘桂芬, 孟海英, 张岩波. Bayes 线性混合效应模型多中心临床试验应用[J]. 中国卫生统计, 2005, 22(4): 200-203.
LIU G F, MENG H Y, ZHANG Y B. Bayes analysis of mixed model applied in amulti-center trials[J]. Chinese Journal of Health Statistics, 2005, 22(4): 200-203(in Chinese).

[118] 周旭毓, 方积乾. Meta 分析中随机效应模型的 Gibbs 抽样及其应用[J]. 中国卫生统计, 2002, 19(4): 204-207.
ZHOU X Y, FANG J Q. Gibbs sampling in random effects model for meta-analysis with application[J]. Chinese Journal of Health Statistics, 2002, 19(4): 204-207(in Chinese).

[119] WIENS J, WALLACE B C. Special issue on machine learning for health and medicine[J]. Machine Learning, 2016, 102(3): 305-307.

[120] LIBBRECHT M W, NOBLE W S. Machine learning applications in genetics and genomics[J]. Nature Reviews Genetics, 2015, 16(6): 321-332.

[121] ALIZADEHSANI R, ABDAR M, JALALI S M J, et al. Comparing the performance of feature selection algorithms for wart treatment selection[C]//Proceedings of the International Workshop on Future Technology. [S.l.: s.n.], 2019: 6-18.

[122] PARMAR C, GROSSMANN P, RIETVELD D, et al. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer[J]. Frontiers in Oncology, 2015, 5: 272.

[123] CHEN P, WU K Y, GHATTAS O. Bayesian inference of heterogeneous epidemic models: application to COVID-19 spread accounting for long-term care facilities[J]. Computer Methods in Applied Mechanics and Engineering, 2021, 385: 114020.

[124] NASRHEIDARABADI N, HAKEMI L, KOLIVAND P, et al. Comparing performances of intelligent classifier algorithms for predicting type of pain in patients with spinal cord injury[J]. Electronic Physician, 2017, 9(7): 4847-4852.

[125] YOUSEFI L, TUCKER A, AL-LUHAYBI M, et al. Predicting disease complications using a stepwise hidden variable approach for learning dynamic Bayesian networks[C]//Proceedings of the IEEE 31st International Symposium on Computer-Based Medical Systems. Piscataway: IEEE Press, 2018: 106-111.

[126] KHADEMI M, NEDIALKOV N S. Probabilistic graphical models and deep belief networks for prognosis of breast cancer[C]//Proceedings of the IEEE 14th International Conference on Machine Learning and Applications. Piscataway: IEEE Press, 2015: 727-732.

[127] ALJAWAD D A, ALQAHTANI E, AL-KUHAILI G, et al. Breast cancer surgery survivability prediction using Bayesian network and support vector machines[C]//Proceedings of the International Conference on Informatics, Health & Technology. Piscataway: IEEE Press, 2017: 1-6.

[128] REFAI A, MEROUANI H F, AOURAS H. Maintenance of a Bayesian network: application using medical diagnosis[J]. Evolving Systems, 2016, 7(3): 187-196.

[129] BANDYOPADHYAY S, WOLFSON J, VOCK D M, et al. Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data[J]. Data Mining and Knowledge Discovery, 2015, 29(4): 1033-1069.

[130] BUKHANOV N, BALAKHONTCEVA M, KOVALCHUK S, et al. Multiscale modeling of comorbidity relations in hypertensive outpatients[J]. Procedia Computer Science, 2017, 121: 446-450.

[131] SA-NGAMUANG C, HADDAWY P, LUVIRA V, et al. Accuracy of dengue clinical diagnosis with and without NS1 antigen rapid test: comparison between human and Bayesian network model decision[J]. PLoS Neglected Tropical Diseases, 2018, 12(6): e0006573.

[132] SHI H Y, LUO G P, ZHENG H W, et al. Coupling the water-energy-food-ecology nexus into a Bayesian network for water resources analysis and management in the Syr Darya River Basin[J]. Journal of Hydrology, 2020, 581: 124387.

[133] HUI E, STAFFORD R, MATTHEWS I M, et al. Bayesian networks as a novel tool to enhance interpretability and predictive power of ecological models[J]. Ecological Informatics, 2022, 68: 101539.

[134] CHATRABGOUN O, HOSSEINIAN-FAR A, DANESHKHAH A. Constructing gene regulatory networks from microarray data using non-Gaussian pair-copula Bayesian networks[J]. Journal of Bioinformatics and Computational Biology, 2020, 18(4): 2050023.

[135] YU H Y, KHAN F, GARANIYA V. Modified independent component analysis and Bayesian network-based two-stage fault diagnosis of process operations[J]. Industrial & Engineering Chemistry Research, 2015, 54(10): 2724-2742.

[136] JIANG Q C, HUANG B, DING S X, et al. Bayesian fault diagnosis with asynchronous measurements and its application in networked distributed monitoring[J]. IEEE Transactions on Industrial Electronics, 2016, 63(10): 6316-6324.

[137] CHANG Y J, CHEN G M, WU X F, et al. Failure probability analysis for emergency disconnect of deepwater drilling riser using Bayesian network[J]. Journal of Loss Prevention in the Process Industries, 2018, 51: 42-53.

[138] WANG Z W, WANG Z W, GU X W, et al. Feature selection based on Bayesian network for chiller fault diagnosis from the perspective of field applications[J]. Applied Thermal Engineering, 2018, 129: 674-683.

[139] YEO C, BHANDARI J, ABBASSI R, et al. Dynamic risk analysis of offloading process in floating liquefied natural gas (FLNG) platform using Bayesian network[J]. Journal of Loss Prevention in the Process Industries, 2016, 41: 259-269.

A review of Bayesian network structure learning

MENG Guanglei[1, 2, *], CONG Zelin[1, 2], SONG Bin[1, 2], LI Tingting[2], WANG Chenguang[2], ZHOU Mingzhe[1, 2]

(1. School of Automation, Shenyang Aerospace University, Shenyang 110136, China;
2. Aviation Science and Technology Key Laboratory of Air Combat System Technology, Shenyang 110136, China)

Abstract: Bayesian networks, as a tool combining probability theory and graph theory, have the ability to efficiently handle uncertain reasoning and data analysis, and are widely used in various fields to solve complex engineering problems. Furthermore, the model can be learned by combining prior knowledge and training samples, overcoming the limitations of establishing the model solely relying on expert knowledge. Based on this, the development history of Bayesian networks was reviewed. The proposed Bayesian network structure learning algorithms were classified and summarized from three aspects: constraint-based methods, rating-based methods, and hybrid search algorithms respectively, and the current research status of various algorithms was summarized and analyzed. Since the data in practical applications often have incompleteness, the research status of incomplete Bayesian network structure learning is explained from two dimensions: missing data processing and latent variable learning. The application of Bayesian networks in different fields is expounded and summarized, and the development trend of future research on Bayesian network structure learning algorithms is discussed.

Keywords: machine learning; artificial intelligence algorithm; Bayesian network; structure learning; hidden variable

Received: 2023-07-07; **Accepted**: 2023-08-04; **Published Online**: 2023-09-05 10:36

URL: link.cnki.net/urlid/11.2625.V.20230904.1958.001

Foundation items: National Natural Science Foundation of China (61973222); Liaoning Talent Program (XLYC2007144); Shenyang Natural Science Foundation (22-315-6-09)

* **Corresponding author**. E-mail: mengguanglei@yeah.net

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0461

柔性支承特性对惯性参考单元性能影响的分析

李醒飞[1,2]，韩佳欣[1]，拓卫晓[1,2,*]，刘雅琪[1]，王天宇[1]

(1. 天津大学 精密测试技术及仪器国家重点实验室，天津 300072；2. 深海技术科学太湖实验室，无锡 214000)

摘 要：复合轴跟瞄与惯性参考单元(IRU)相结合的方式是当前抑制运动平台振动干扰、提升跟瞄精度的主要技术手段。IRU通常采用柔性支承结构，以规避框架式结构转动惯量大的缺陷，柔性支承特性是决定IRU扰动抑制带宽的关键因素。基于此，针对柔性支承特性对IRU性能的影响进行量化分析，基于质量-刚度-阻尼模型建立IRU系统的传递函数模型，并验证其准确性；利用参数分析方法研究柔性支承扰动传递特性和物理参数对IRU系统扰动抑制能力的影响；搭建IRU“速度内环-位置外环”双闭环控制系统的Simulink仿真模型，通过仿真分析给定振动环境下柔性支承转动刚度特性对IRU扰动抑制能力和稳定精度的影响。结果表明：柔性支承的扰动传递特性决定了系统对高频扰动的抑制能力，降低柔性支承的转动刚度能够有效提升IRU系统的扰动抑制能力，但会加剧传感器静态测量噪声、驱动器纹波对IRU系统稳定精度的负面影响。

关 键 词：惯性参考单元；柔性支承；扰动抑制能力；转动刚度；参数化分析
中图分类号：V19
文献标志码：A **文章编号**：1001-5965(2025)09-2850-10

在激光通信、深空探测及精确制导等领域，为实现运动目标的高精度跟瞄，需要将系统指向精度控制在微弧度甚至亚微弧度量级，由此产生了捕获、跟踪、瞄准(acquisition, tracking and pointing, ATP)的概念[1-2]。随着目标的变化和任务拓展，需要将ATP系统安装在机载、车载、星载、船载等运动平台上。然而，运动平台的高机动性会引入极难克服的振动干扰，影响视轴稳定精度[3]。复合轴跟瞄与惯性参考单元(inertial reference unit, IRU)相结合的方式是当前抑制运动平台振动干扰、实现微弧度甚至亚微弧度级跟瞄的主要技术手段[4-5]，并已成功应用于美国的月球激光通信演示验证(lunar laser communication demonstration, LLCD)试验[6]。IRU能够提供一束不受载体振动影响的惯性参考光束，该光束穿过库德光路后注入主望远镜，为ATP系统提供一个虚拟参考星，最终通过辅助快速反射镜(fast steering mirror, FSM)来抑制载体振动对系统视轴的负面影响[7]。

IRU的概念最早由美国麻省理工学院Draper实验室提出[8]。传统的IRU多采用框架式结构实现单轴或多轴的转动，但其转动惯量较大，可实现的扰动抑制带宽非常有限，稳定精度仅能达到几十微弧度[9-10]。采用无摩擦、无空回、线性度好的柔性支承为系统提供转动自由度的IRU，因其兼具宽频带和高稳定精度的特点而被普遍采用。自20世纪90年代开始，美国麻省理工学院[11]、BlueHalo公司[12]和日本三菱电机先进技术中心[13]等研究机构先后开展了基于柔性支承的IRU的研制。截至目前，综合性能最佳的IRU是BlueHalo公司的OIRU-500型，可以实现1～1 000 Hz频带内优于−40 dB的扰动抑制比和500 nrad的稳定精度[14]。国内对IRU的研究开展相对较晚，中国科学院光电技术研究

收稿日期：2023-07-14；**录用日期**：2023-09-28；**网络出版时间**：2023-10-19 18:56
网络出版地址：link.cnki.net/urlid/11.2625.V.20231019.1348.001
基金项目：国家自然科学基金(62203322)；中国博士后科学基金(2022M712372)；深海技术科学太湖实验室“揭榜挂帅”项目(2022JBGS03001)
*通信作者. E-mail：tuoweixiao@tju.edu.cn

引用格式：李醒飞，韩佳欣，拓卫晓，等. 柔性支承特性对惯性参考单元性能影响的分析[J]. 北京航空航天大学学报，2025，51(9)：2850-2859. LI X F，HAN J X，TUO W X，et al. Influence analysis of flexure support characteristics on performance of inertial reference unit[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2850-2859 (in Chinese).

所、电子科技大学、天津大学从事 IRU 的基础理论、关键技术和原理样机的研发。截至目前, 中国科学院光电技术研究所研发的 IRU 原理样机带宽为 1～500 Hz[15]。天津大学研制的 IRU 原理样机实现了±5 mrad 的转角范围和小于 1 μrad 的角分辨率[16], 扰动抑制带宽达到 60 Hz, 但受限于所用惯性传感器的精度, 仅实现了约 20 μrad 的稳定精度[17]。目前, 国内研制的 IRU 带宽与以 BlueHalo 公司为代表的国际先进水平存在较大差距, 难以满足运动平台 ATP 系统的应用需求。

目前, 已有专利和文献[18-19]均明确指出柔性支承是决定 IRU 系统扰动抑制带宽的关键因素, 因而建立包含结构特性的 IRU 控制系统模型, 分析柔性支承结构特性与 IRU 系统扰动抑制能力、稳定精度等特性的关系, 对于系统的设计与优化具有重要意义。

IRU 系统采用的柔性支承结构形式与 FSM 系统类似。设计的一般思路是: 在系统转动惯量/质量固定的条件下, 根据系统高低阶谐振频率与工作带宽的关系, 得到柔性支承各向刚度的要求, 进而结合刚度模型实现柔性支承的参数化设计。然而, FSM 负载仅包括一面反射镜, 相较 IRU 系统轻 1～2 个数量级[20-21], 这对柔性支承的承载能力和各向刚度提出了更高的要求。此外, FSM 系统的柔性支承主要影响其控制带宽[22], 然而, IRU 系统的柔性支承是影响其扰动抑制能力的关键因素。研究柔性支承扰动传递特性和物理参数对 IRU 系统扰动抑制能力及稳定精度的影响, 能够为 IRU 系统的设计和优化提供指导。

本文结合课题组自主研发的 IRU 系统, 在建立系统数学模型的基础上, 分析系统扰动抑制能力的影响因素, 建立基于速度内环-位置外环稳定性解耦控制器的仿真模型, 以此为基础分析柔性支承转动刚度对扰动抑制能力和稳定精度的影响。研究成果能够指导 IRU 系统的结构优化与设计, 进而为中国突破运动平台高精度跟瞄领域的国际封锁提供支撑。

1　IRU 工作原理及理论建模

1.1　系统工作原理

IRU 系统主要由用于敏感平台角扰动的惯性传感器、驱动平台转动的执行器、两自由度柔性支承、测量相对角位置的电涡流线位移传感器及控制器 5 个部分组成。两自由度柔性支承通常采用具有集中柔度式的柔性铰链单元组合而成, 其在 IRU 系统的工作方向(绕 z 轴转动、绕 y 轴转动)上具有较低的刚度, 而在其他非工作方向上具有较高的刚度。当存在基座角扰动时, 该扰动通过柔性支承传至平台, 安装在台面上的 MEMS 陀螺仪与磁流体动力学(magnetohydrodynamic, MHD)角速度传感器作为惯性角速度传感器组, 组合测量基准光源所受的角扰动, 反馈控制音圈电机(voice coil motor, VCM)反方向推动平台, 最终使安装在平台上的激光光源发射的光束被稳定于惯性空间, 进而为 ATP 系统提供一束不受基座角扰动影响的参考光束。针对单速度环控制存在的积分漂移问题, IRU 一般采用如图 1 所示的双闭环控制结构。此外, 平台与基座之间还安装有电涡流线位移传感器, 将测得的相对角位置信息输出给 IRU 载体, 如框架伺服系统或载体姿态控制系统, 防止系统转动角度超出允许的角行程范围。

1.2　数学模型

IRU 通常采用具有宽频带、低纹波的线性驱动器作为音圈电机的功率放大器。由于线性驱动器的带宽(10 kHz)远大于 IRU 系统的控制带宽, 输入电压和输出电流间的关系可以被等效为比例环

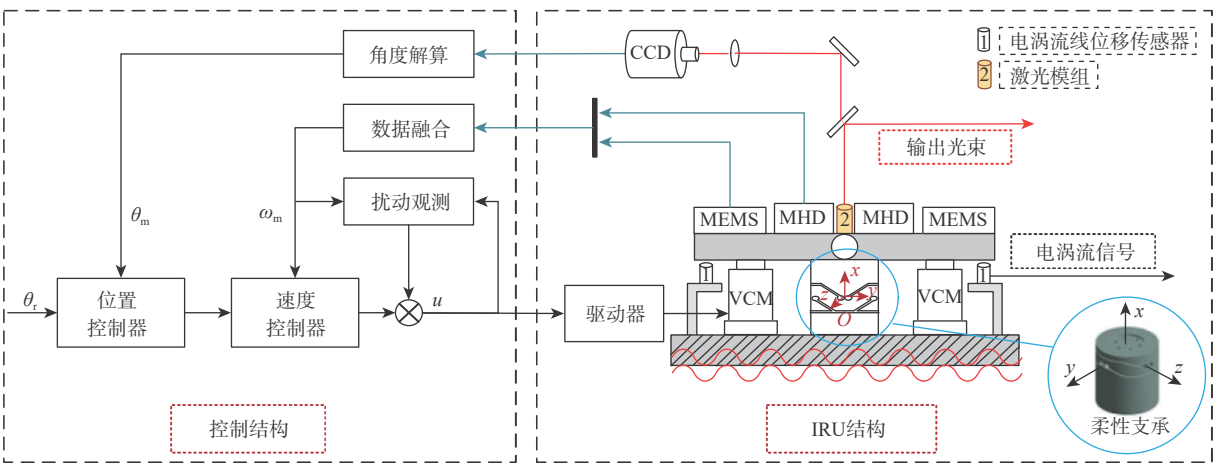


图 1　IRU 工作原理框图

Fig. 1　Block diagram of IRU working principle

节。可以得出，音圈电机输出力 F 与驱动器输入电压 U_d 间的关系为

$$F = k_f I_d = \frac{k_f U_d}{K} \tag{1}$$

式中：k_f 为音圈电机的力常数；I_d 为驱动器的输出电流；K 为驱动器的增益系数。

在加工精度较高的情况下，柔性支承轴间耦合影响较小，可以忽略不计，因而，可以将IRU系统简化为2个完全一致的单自由度质量-刚度-阻尼模型，如图2所示。图中，L 为音圈电机安装半径，m_c 为音圈电机动子的质量，c 为机械阻尼系数，K_θ 为柔性支承转动刚度，θ_m 和 θ_d 分别为平台的偏转角和基座所受的角扰动。

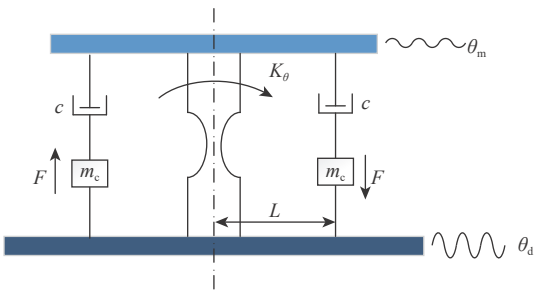


图2 质量-刚度-阻尼模型

Fig. 2 Mass-stiffness-damping model

根据图2所示的模型，可以建立IRU系统的微分方程为

$$2FL = J\frac{d^2\theta_m}{dt^2} + 2cL^2\frac{d(\theta_m-\theta_d)}{dt} + K_\theta(\theta_m-\theta_d) \tag{2}$$

式中：J 为系统的转动惯量。

将式(1)代入式(2)可得

$$2\frac{k_f U_d}{K}L = J\frac{d^2\theta_m}{dt^2} + 2cL^2\frac{d(\theta_m-\theta_d)}{dt} + K_\theta(\theta_m-\theta_d) \tag{3}$$

对式(3)进行拉普拉斯变换并整理，可以建立系统输出转角与输入电压及基座扰动之间的传递函数，如下：

$$\theta_m(s) = \frac{2Lk_f}{K(Js^2+2cL^2s+K_\theta)}U_d(s) + \frac{2cL^2s+K_\theta}{Js^2+2cL^2s+K_\theta}\theta_d(s) \tag{4}$$

系统被控对象特性为

$$G_\theta(s) = \frac{\theta_m(s)}{U_d(s)} = \frac{2Lk_f}{K(Js^2+2cL^2s+K_\theta)} \tag{5}$$

可见，系统被控对象是二阶振荡环节，柔性支承工作方向上的低刚度会引入低频机械谐振。此外，系统组件间的非刚性连接及非工作方向的耦合作用，会在系统开环特性的高频段附加一个或多个高阶谐振。实际的IRU系统的开环幅频响应曲线如图3所示。

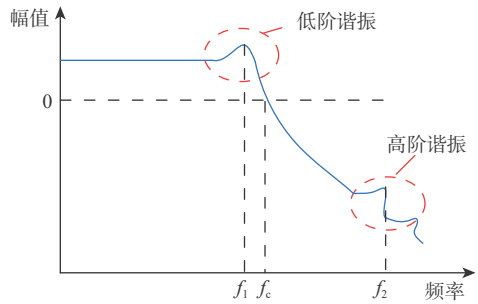


图3 IRU系统开环幅频响应曲线

Fig. 3 Open-loop frequency response curve of IRU system

当控制回路未使能时，系统扰动传递特性为

$$G_d(s) = \frac{\theta_m(s)}{\theta_d(s)} = \frac{2cL^2s+K_\theta}{Js^2+2cL^2s+K_\theta} \tag{6}$$

由式(5)与式(6)可以看出，IRU系统的被控对象特性和扰动传递特性都是二阶振荡环节，且具有相同的极点，参数 K_θ、J、c 及 L 均会对二者产生影响。令式(6)中 s=jω 并取模，可得到IRU系统的扰动传递率为

$$|G_d(\omega)| = \sqrt{\frac{1+4\xi^2\left(\frac{\omega}{\omega_n}\right)^2}{\left[1-\left(\frac{\omega}{\omega_n}\right)^2\right]^2+4\xi^2\left(\frac{\omega}{\omega_n}\right)^2}} \tag{7}$$

式中：ω_n 为无阻尼固有频率；ξ 为阻尼比。

$$\omega_n = \sqrt{\frac{K_\theta}{J}} \tag{8}$$

$$\xi = \frac{cL^2}{\sqrt{JK_\theta}} \tag{9}$$

可以看出，式(7)在低频段增益近似为1，而在频率高于 $\sqrt{2}\omega_n$ 的高频段，扰动传递率逐渐衰减。IRU系统的扰动传递特性对高频扰动有一定的抑制作用，即使控制回路不工作，高频角扰动传递至平台时，幅值也将被衰减。然而，由于闭环带宽的限制，系统控制回路对高频扰动的抑制能力十分有限。IRU系统的扰动抑制能力等于控制回路抑制特性(主动抑制能力)与扰动传递特性(被动抑制能力)之和，如图4所示。图中，a 和 b 分别表示增益

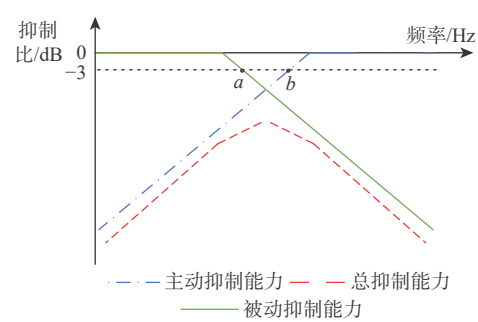


图4 IRU系统扰动抑制特性

Fig. 4 Disturbance rejection characteristics of IRU system

为−3 dB 时被动抑制曲线与主动抑制曲线对应的交点。

2　模型辨识与参数化分析

2.1　模型辨识

为验证理论建模的准确性, 采用 Levy 法对 IRU 系统的被控对象模型进行辨识, 辨识实验系统框图如图 5 所示。利用数据采集卡产生频率范围为 0.1～600 Hz、幅值为 0.5 V、频率间隔为 0.1 Hz、相位随机的多谐差相信号, 输入至线性驱动器, 驱动音圈电机使台面转动, 利用电涡流线位移传感器测量基座与台面间的相对角位置, 并由数据采集卡传回至上位机。

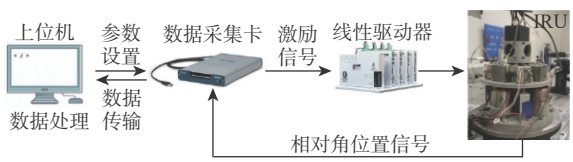


图 5　辨识实验系统框图

Fig. 5　Block diagram of identification experiment system

由于 Levy 法无噪声抑制能力, 需要对采集到的电涡流线位移传感器输出转角信号经去趋势项、初级滤波等数据处理。经实验测试发现, 多次拟合得到的转动角度和激励电压间的传递函数一致性较好(<5%), 且均能保证高于 0.99 的拟合优度。其中, 一组拟合优度为 0.993 的传递函数为

$$G(s)=\frac{1.302\,2\times10^{9}}{(s+4\,756)(s^{2}+6.47s+5.051\times10^{4})} \tag{10}$$

课题组研制的 IRU 系统采用 BEI15-16-024A 型音圈电机和 Glentek 的 SMA5005 型驱动器, 所设计系统的音圈电机分布半径为 0.055 m, 所设计柔性支承转动刚度为 269.03 (N·m)/rad, 利用 Inventor 计算得到的转动惯量为 5.684×10^{-3} kg·m^2, 系统参数化辨识得到的阻尼系数为 5.875 (N·s)/m。依据上述说明, 将系统设计参数(L=0.055 m, K=1 V/A, K_θ=269.03 (N·m)/rad, c=5.875 (N·s)/m, k_f=12.68 N/A, J=5.684×10^{-3} kg·m^2)代入式 (5) 中, 可以得到 IRU 系统被控对象理论传递函数。将理论建模结果和实验辨识结果的频率响应曲线进行对比, 如图 6 所示。可以看出, 理论建模与实验辨识结果趋势基本一致, 理论模型能够较好地反映 IRU 系统的真实被控对象特性。但由于加工和装配误差的存在, IRU 系统的实测谐振频率为 35.8 Hz, 与理论计算值(34.5 Hz)存在 3.6% 的偏离。此外, 在理论建模过程中, 将线性驱动器等效为比例环节, 而真实的线性驱动器应为惯性环节, 理论建模过程中忽略的高频相位滞后是导致实测频率响应曲线与理论建模结果高频存在偏离的主要原因。

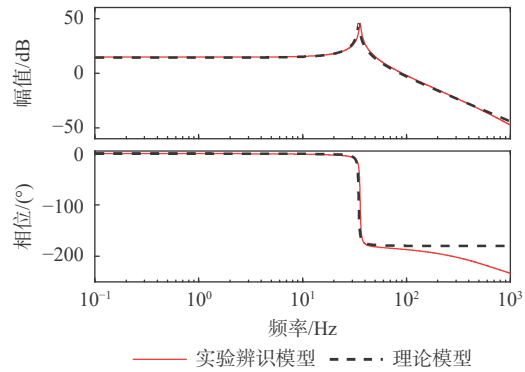


图 6　理论建模与实验辨识结果对比

Fig. 6　Comparison of theoretical modeling and experimental identification results

2.2　参数化分析

根据 1.2 节的分析可知, 柔性支承转动刚度 K_θ、系统转动惯量 J、机械阻尼系数 c 及音圈电机安装半径 L 均是决定 IRU 系统被控对象和扰动传递特性的主要因素。本节将在音圈电机选定且安装位置固定的情况下, 量化柔性支承转动刚度和系统转动惯量对 IRU 系统的影响。

在其他参数保持不变的情况下, 将转动惯量从 3×10^{-3} kg·m^2 分别降低至 1×10^{-3} kg·m^2 和提高至 5×10^{-3} kg·m^2 时, 系统的被控对象特性和扰动传递特性如图 7 所示。

可以看出, 结构的谐振频率与系统转动惯量成

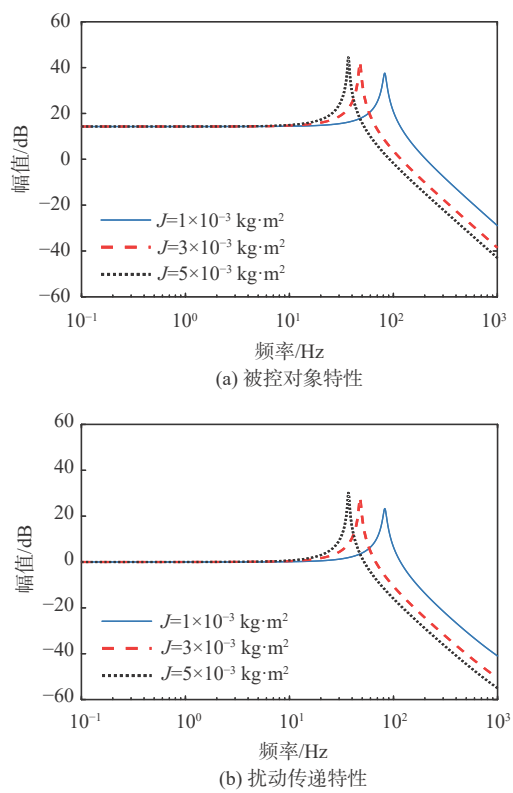


图 7　转动惯量对 IRU 系统的影响

Fig. 7　Influence of moment of inertia on IRU system

负相关，增加转动惯量能够有效提高IRU系统对高频扰动的抑制效果。但转动惯量的增加会放大系统谐振峰附近频率的扰动，增加主动控制实现谐振抑制的负担。此外，在完成IRU系统负载（惯性传感器、激光光源、电机运动部件）的选型和台面的优化布局后，系统的转动惯量由柔性支承转动中心与平台下端面的距离决定。增加此距离将有效提高系统转动惯量，但会加剧系统的轴间耦合，不利于实现系统的高精度稳定控制。综上所述，结构转动惯量对IRU系统的影响较为复杂，为提高IRU系统对扰动的综合抑制能力，设计过程中需在满足稳定精度和电机性能约束的条件下，尽量增加转动惯量。

在其他参数保持不变的情况下，将柔性支承工作方向的转动刚度从269.03 (N·m)/rad分别降低至69.03 (N·m)/rad和提高至469.03 (N·m)/rad时，系统的被控对象特性和扰动传递特性如图8所示。可以看出，柔性支承转动刚度越小，结构谐振频率和谐振峰值越低。降低柔性支承转动刚度能够提高系统中高频扰动抑制能力的同时，降低主动控制的压力，这对于提高系统的综合扰动抑制能力来说是非常有利的。然而，柔性支承工作方向的转动刚度与非工作方向的刚度呈正相关性，降低转动刚度的同时也会导致非工作方向刚度下降。这将造成系统高阶谐振频率下降，不利于主动控制带宽的提升，但可通过扰动传递特性进行弥补。

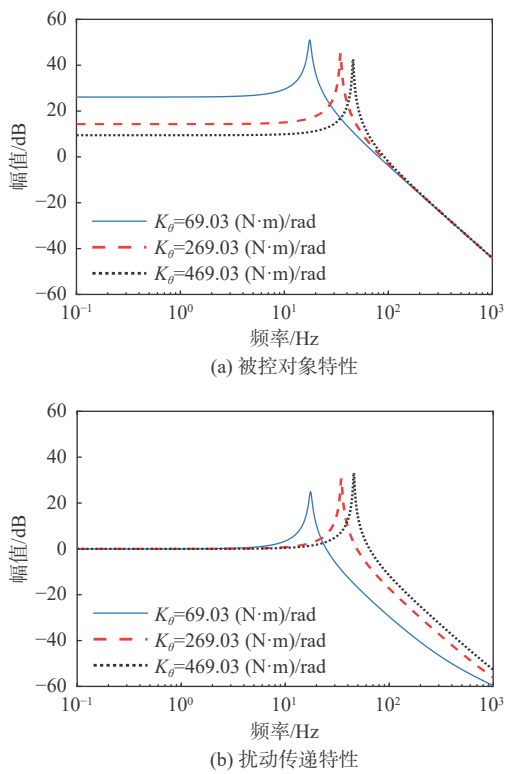


图8 转动刚度对IRU系统的影响

Fig. 8 Influence of rotational stiffness on IRU system

3 仿真分析

3.1 仿真模型

第2节在控制器未使能的条件下，从理论上分析了参数变化对被控对象和扰动传递特性的影响，然而根据1.2节可知，IRU系统的扰动抑制能力由扰动传递特性（被动抑制能力）和控制回路抑制特性（主动抑制能力）共同决定，为进一步探究参数变化对实际IRU系统的影响，需要在控制器使能下，以目前研制的IRU系统为基础进行闭环仿真分析。

IRU系统正常工作时，面临载体振动干扰等外部扰动和负载质量变化、电机力矩波动、温度变化等造成的内部扰动。由于图像位置传感器CCD的时域滞后和惯性传感器的积分漂移，单一的速度或位置闭环回路均无法实现较高的控制带宽。针对上述问题，提出了一种以MHD角速度传感器和MEMS陀螺仪作为速度反馈元件，以CCD作为位置反馈元件的“速度-位置”双闭环控制结构，如图9所示。图中，G_d与G_v分别表示系统的扰动传递特性和速度环被控对象特性，在速度环内利用陷波器G_f实现IRU系统带宽内机械谐振的抑制，利用频段估计器实现外部扰动频段的观测，结合扩张状态观测器（extended state observer, ESO）实现窄带扰动频率附近观测增益的动态调节[23-24]。为解决内外环路间的稳定性耦合及速度内环在频率极低频段内（<1 Hz）扰动难以抑制的问题，内部速度环控制器C_v采用不包含积分结构的Lead-Lag控制器，并辅以高通滤波器Q_H将直流至极低频范围内扰动转移至外部位置环，位置环控制器C_p采用经典的PI控制器。

本节针对图9所示的双闭环控制方案，仿真分析柔性支承转动刚度特性对IRU系统性能的影响。搭建如图10所示的Simulink仿真模型，考虑到实际系统数字传递的特性，依据控制带宽与采样速率的关系设置速度内环采样频率为1 250 Hz，采样间隔为0.8 ms，位置外环采样间隔设置为2.4 ms。此外，由于实际CCD传感器的输出存在延迟，为模拟真实情况，在CCD的输出端加入5 ms的延时。仿真中，被控对象采用式(5)所示的微分，扰动传递特性采用式(6)所示模型，模型各参数见2.1节。

3.2 双闭环回路的扰动抑制能力仿真分析

本节针对课题组目前研制的IRU系统，对图10所示双闭环控制回路在给定振动环境下的扰动抑制能力进行仿真分析。在基座施加幅值为1 mrad、频率范围为0.1～200 Hz的正弦角扰动，采用稳态

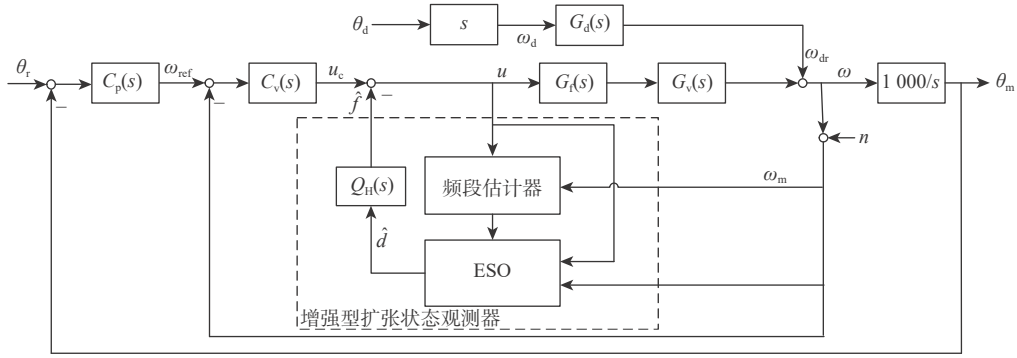


图 9　IRU 系统控制结构示意图

Fig. 9　Control structure diagram of IRU system

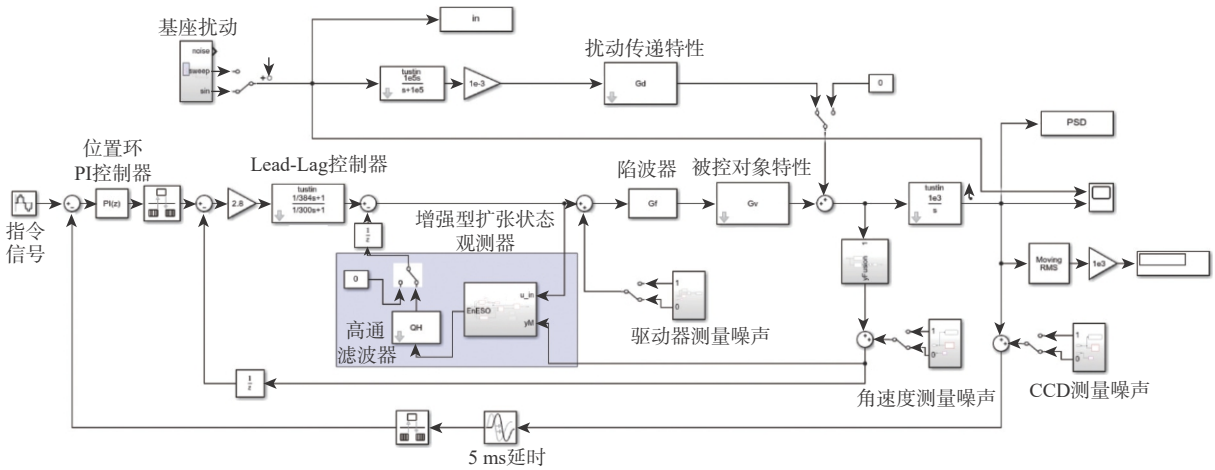


图 10　IRU 系统仿真模型

Fig. 10　Simulation model of IRU system

正弦法测试 IRU 系统的扰动抑制能力。控制回路使能前后, IRU 系统的扰动抑制能力如图 11 所示。

可以看出, 控制回路未使能时, IRU 系统的扰动抑制能力完全由扰动传递特性决定, 频域上表现为低通。控制回路使能后, 有效提升了 IRU 系统对中低频扰动的抑制能力, 其中, 1 Hz 处的扰动抑制比由 0 dB 提升至−34 dB, 谐振频率 34.5 Hz 处的扰动抑制比由 30 dB 提高至−5.4 dB。虽然控制系统的“水床效应”会对 IRU 系统的高频抑制能力造成负面影响, 但系统总的扰动抑制能力得到了明显提升。

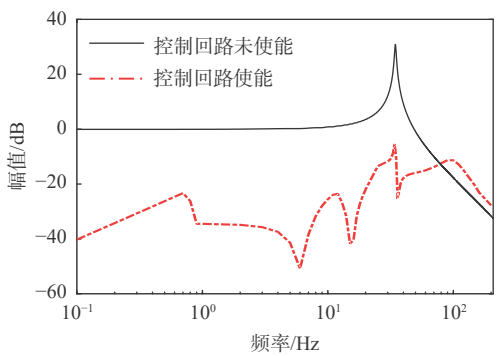


图 11　控制回路对 IRU 系统的影响

Fig. 11　Influence of control loop on IRU system

3.3　转动刚度对系统扰动抑制能力的影响分析

在音圈电机选定及 IRU 系统其他参数不变的情况下, 柔性支承工作方向上的转动刚度从 269.03 (N·m)/rad 分别降低至 69.03 (N·m)/rad 和提高至 469.03 (N·m)/rad 时, IRU 系统总的扰动抑制能力如图 12 所示, 进一步对低频(1 Hz)、中频(谐振频率)、高频(100 Hz)处 IRU 系统的扰动抑制比进行分析, 其结果如表 1 所示。

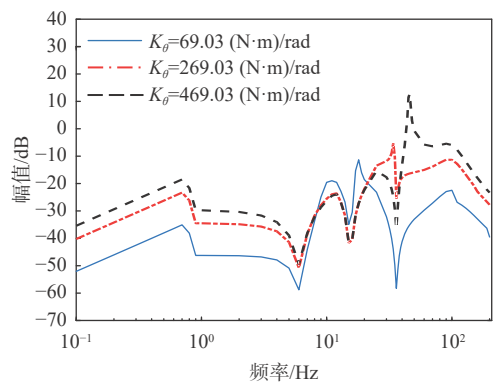


图 12　不同转动刚度下 IRU 系统的扰动抑制能力

Fig. 12　Disturbance rejection ability of IRU system under different rotational stiffness

表 1 不同转动刚度时 IRU 系统扰动抑制能力的影响

Table 1 Influence of different rotational stiffness on disturbance rejection ability of IRU system

转动刚度/(N·m·rad^{-1})	谐振频率/Hz	扰动频率/Hz	扰动抑制比/dB
69.03	17.47	1	−46.25
		17.47	−10.18
		100	−22.45
269.03	34.52	1	−34.5
		34.52	−4.49
		100	−11.34
469.03	45.78	1	−29.76
		45.78	14.314
		100	−5.93

可以看出，随着柔性支承工作方向上转动刚度的减小，IRU 系统的谐振频率逐渐降低，谐振点也逐步左移。由于谐振点会引起附近一定频率范围内曲线的幅值偏大，在转动刚度为 69.03 (N·m)/rad 时，系统中频段的扰动抑制能力略差于另外 2 种情况，但整体而言系统总的扰动抑制能力有所提升。因此，在进行 IRU 系统结构设计时，最小化柔性支承的转动刚度对于提高 IRU 系统的扰动抑制能力是非常有益的。

3.4 转动刚度对系统稳定精度的影响分析

3.3 节的结论表明，柔性支承的转动刚度与 IRU 系统的扰动抑制能力呈负相关。然而，降低柔性支承的转动刚度将导致 IRU 系统对传感器噪声、电机力矩波动等因素更为敏感，影响 IRU 系统的稳定精度。课题组研制的 IRU 采用 CCD 作为位置反馈元件，其静态测量噪声曲线如图 13(a)所示，计算可得其均方根为 3.063 6 μrad。此外，还采用 MHD 角速度传感器和 MEMS 陀螺仪作为速度反馈元件，为实现宽频角扰动的高精度测量，对二者进行了数据融合，融合输出的静态噪声曲线如图 13(b)所示。

对融合输出信号的静态噪声进行功率谱分析，利用 MATLAB 中的 Pwelch 函数计算融合输出信号的噪声功率谱，其结果如图 14 所示。经计算可知，融合输出信号在 0.1～625 Hz 频段范围内的等效噪声角速度为 5.459 8 mrad/s，为实现评价指标的统一，对其进行角度换算，得到的等效噪声角度为 83.004 μrad。在图 10 所示的仿真模型中加入图 13 所示的传感器静态测量噪声，在其他参数不变的情况下，柔性支承 3 组不同的转动刚度(69.03 (N·m)/rad、269.03 (N·m)/rad、469.03 (N·m)/rad)所对应的 IRU 系统平台输出转角如图 15(a) 所示，进一步计算平台输出转角的功率谱密度，其结果如图 15(b) 所示。

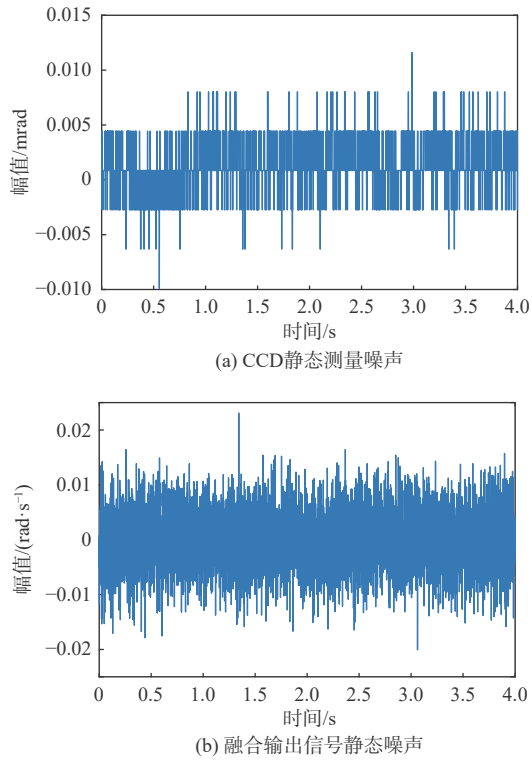


图 13 传感器静态测量噪声

Fig. 13 Static measurement noise of sensor

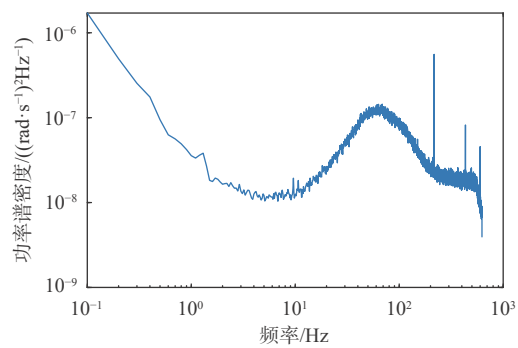
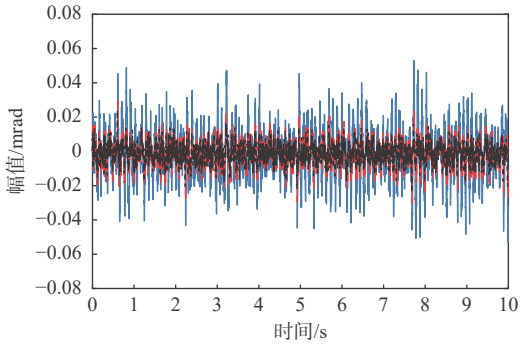


图 14 角速度噪声功率谱密度

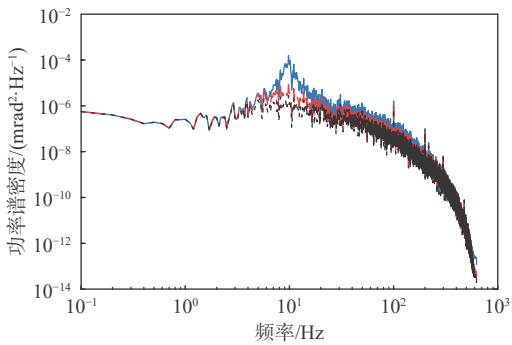
Fig. 14 Noise power spectral density of angular velocity

平台输出转角的均方根和 0.1～625 Hz 频段内等效角位移如表 2 所示。可以看出，随着柔性支承转动刚度的降低，IRU 系统对传感器静态测量噪声更为敏感，转动刚度分别取 69.03 (N·m)/rad 与 469.03 (N·m)/rad 时，系统等效角位移几乎翻倍，影响系统稳定精度。

除受传感器静态测量噪声影响外，目前使用的线性驱动器(SMA5005)存在输出纹波，实测的驱动器纹波如图 16 所示。在图 10 所示的仿真模型中加入图 16 所示的驱动器纹波，在其他参数不变的情况下，柔性支承 3 组不同的转动刚度(69.03 (N·m)/rad、269.03 (N·m)/rad、469.03 (N·m)/rad)所对应的 IRU 系统平台输出转角如图 17(a) 所示，进一步计算输出转角的功率谱密度，如图 17(b) 所示。



(a) 输出转角时域信号



(b) 输出转角的功率谱密度

— K_θ=69.03 (N·m)/rad -- K_θ=269.03 (N·m)/rad - - K_θ=469.03 (N·m)/rad

图 15　不同转动刚度下传感器静态测量噪声对输出转角的影响

Fig. 15　Influence of sensor static measurement noise on output angle under different rotational stiffness

表 2　不同转动刚度下传感器静态测量噪声对稳定精度的影响

Table 2　Influence of sensor static measurement noise on stabilization precision under different rotational stiffness

转动刚度/(N·m·rad^{-1})	输出转角均方根/μrad	等效角位移/μrad
69.03	15.975	15.975
269.03	7.442	7.437 1
469.03	5.536 7	5.530 6

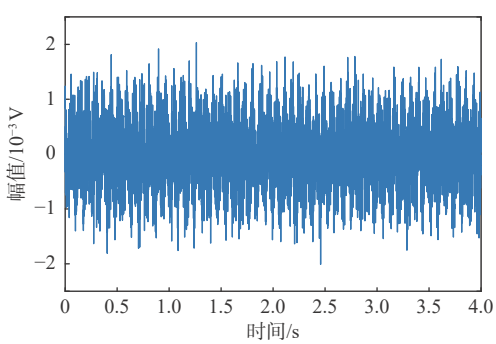
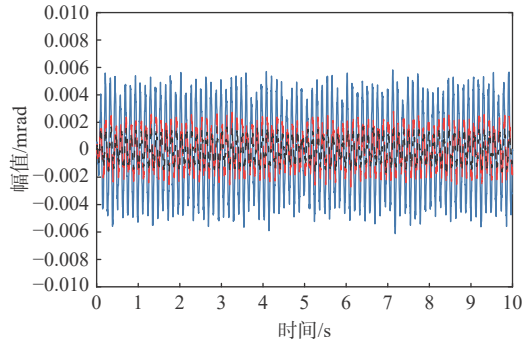


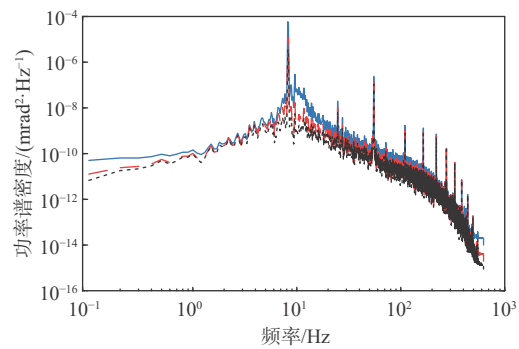
图 16　驱动器纹波

Fig. 16　Driver ripple

平台输出转角的均方根和 0.1～625 Hz 频段内的等效角位移如表 3 所示。经计算, 转动刚度为 69.03 (N·m)/rad 时系统输出转角的等效角位移相比 469.03 (N·m)/rad 时扩大了近 289.7%。可以看出, 柔



(a) 输出转角时域信号



(b) 输出转角的功率谱密度

— K_θ=69.03 (N·m)/rad -- K_θ=269.03 (N·m)/rad - - K_θ=469.03 (N·m)/rad

图 17　不同转动刚度下驱动器纹波对输出转角的影响

Fig. 17　Influence of driver ripple on output angle under different rotational stiffness

表 3　不同转动刚度下驱动器纹波对稳定精度的影响

Table 3　Influence of driver ripple on stabilization precision under different rotational stiffness

转动刚度/(N·m·rad^{-1})	输出转角均方根/μrad	等效角位移/μrad
69.03	3.299 2	3.299 3
269.03	1.449 5	1.449 5
469.03	0.846 6	0.846 7

性支承工作方向转动刚度越低, 系统对驱动器纹波越敏感, 平台稳定精度受驱动器纹波影响越大。

根据上述分析可知, 随着柔性支承转动刚度的降低, IRU 系统的稳定精度对传感器静态测量噪声和驱动器纹波更为敏感, 因而柔性支承转动刚度的设计存在一定的下限。在进行 IRU 结构的优化设计时, 需要将稳定误差对支承刚度的要求作为约束条件, 在保证稳定精度的前提下, 以最小化柔性支承转动刚度为目标, 实现系统扰动抑制能力的提升。

4　结　论

两自由度柔性支承为 IRU 系统提供 2 个工作方向上的转动自由度, 是决定系统扰动抑制带宽、稳定精度等指标的关键因素。本文建立了基于柔性支承的 IRU 系统数学模型, 分析了系统扰动抑制

能力是由控制回路和柔性支承扰动传递特性共同决定的，搭建了系统闭环仿真模型，量化了柔性支承对IRU系统性能的影响。分析结果表明：

1）降低柔性支承转动刚度能够提高IRU系统对高频扰动的抑制能力，且能降低系统对电机输出力矩的要求，助力系统小体积和低功耗设计。

2）柔性支承转动刚度的降低会导致IRU系统对传感器静态测量噪声、驱动器纹波更为敏感，稳定精度随之下降，且应力集中效应愈发明显。

综上所述，在进行IRU系统柔性支承结构设计时，需要将最小化柔性支承转动刚度作为优化目标，将最大应力、传感器静态测量噪声、驱动器纹波等作为约束条件，实现以扰动抑制能力为目标的定量优化设计。

参考文献（References）

[1] 马佳光. 捕获跟踪与瞄准系统的基本技术问题[J]. 光学工程, 1989, 16(3): 1-42.
MA J G. The basic technologies of the acquisition, tracking and pointing systems[J]. Opto-Electronic Engineering, 1989, 16(3): 1-42 (in Chinese).

[2] LI Q, LIU L, YANG H J. High accuracy and multi-target acquisition, pointing and tracking under satellite micro-vibrations[J]. Microgravity Science and Technology, 2020, 32(4): 715-727.

[3] 任彦, 刘正华, 周锐. 滑模干扰观测器在低速光电跟踪系统中的应用[J]. 北京航空航天大学学报, 2013, 39(6): 835-840.
REN Y, LIU Z H, ZHOU R. Application of low speed opto-electronic tracking systems based on sliding mode distutbance observer[J]. [J]. Journal of Beijing University of Aeronautics and Astronautics, 2013, 39(6): 835-840(in Chinese).

[4] 王伟, 刘云清, 董岩, 等. 空间激光通信中复合跟踪技术研究[J]. 激光与红外, 2020, 50(4): 403-406.
WANG W, LIU Y Q, DONG Y, et al. Research on composite tracking technology in space laser communication[J]. Laser & Infrared, 2020, 50(4): 403-406(in Chinese).

[5] 唐涛, 马佳光, 陈洪斌, 等. 光电跟踪系统中精密控制技术研究进展[J]. 光电工程, 2020, 47(10): 200315.
TANG T, MA J G, CHEN H B, et al. A review on precision control methodologies for optical-electric tracking control system[J]. Opto-Electronic Engineering, 2020, 47(10): 200315(in Chinese).

[6] BURNSIDE J W, CONRAD S D, PILLSBURY A D, et al. Design of an inertially stabilized telescope for the LLCD[C]//Proceedings of the Free-Space Laser Communication Technologies XXIII. Bellingham: SPIE, 2011.

[7] BURNSIDE J W, MURPHY D V, KNIGHT F K, et al. A hybrid stabilization approach for deep-space optical communications terminals[J]. Proceedings of the IEEE, 2007, 95(10): 2070-2081.

[8] LUNIEWICZ M F, GILMORE J P, CHIEN T T, et al. Comparison of wideband inertial line-of-sight stabilization reference mechanizations[C]//Proceedings of the Acquisition, Tracking, and Pointing VI. Bellingham: SPIE, 1992: 378-398.

[9] XIA X Q, ZHANG B, LI X T. High precision low-speed control for permanent magnet synchronous motor[J]. Sensors, 2020, 20(5): 1526.

[10] 谢航. 动基座光电稳定平台视轴稳定技术研究[D]. 长春: 中国科学院大学 (中国科学院长春光学精密机械与物理研究所), 2022.
XIE H. Research on line-of-sight stabilization technology of photoelectric stabilization platform with moving base[D]. Changchun: University Chinese Academy of Sciences (Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences), 2022(in Chinese).

[11] LUNIEWICZ M F, MURPHY J, O'NEIL E, et al. Testing the inertial pseudo-star reference unit[C]//Proceedings of the Acquisition, Tracking, and Pointing VIII. Bellingham: SPIE, 1994: 638-649.

[12] ECKELKAMP-BAKER D, SEBESTA H R, BURKHARD K. Magnetohydrodynamic inertial reference system[C]//Proceedings of the Acquisition, Tracking, and Pointing XIV. Bellingham: SPIE, 2000: 99-110.

[13] HARUNA M, KODEKI K, SHIMIZU S, et al. Evaluation of developing inertial stabilization unit[C]//Proceedings of the Free-Space Laser Communication and Atmospheric Propagation XXVII. Bellingham: SPIE, 2015.

[14] BLUEHALO. Optical inertial reference units (OIRU) [EB/OL]. (2022-02-27)[2023-07-01]. https://bluehalo.com/product/optical-inertial-reference-unit-oiru.

[15] DENG J Q, XUE W C, ZHOU X, et al. On dual compensation to disturbances and uncertainties for inertially stabilized platforms[J]. International Journal of Control, Automation and Systems, 2022, 20(5): 1521-1534.

[16] TUO W X, LI X F, JI Y, et al. Mechanical design and determination of bandwidth for a two-axis inertial reference unit[J]. Mechanical Systems and Signal Processing, 2022, 172: 108962.

[17] ZHOU Z, LI X F, TUO W X. Design of disturbance suppression controller for optical inertial reference unit[C]//Proceedings of the International Conference on Optical Instruments and Technology: Optoelectronic Measurement Technology and Systems. Bellingham: SPIE, 2022.

[18] 胡浩军. 运动平台捕获、跟踪与瞄准系统视轴稳定技术研究[D]. 长沙: 国防科学技术大学, 2005.
HU H J. Research on line-of-sight stabilization technology of acquisition, tracking and aiming system of moving platform[D]. Changsha: National University of Defense Technology, 2005(in Chinese).

[19] JACKA N, WALTER R, LAUGHLIN D, et al. Design of stabilized platforms for deep space optical communications (DSOC) [C]//Proceedings of the Free-Space Laser Communication and Atmospheric Propagation XXIX. Bellingham: SPIE, 2017.

[20] XIAO R J, XU M L, SHAO S B, et al. Design and wide-bandwidth control of large aperture fast steering mirror with integrated-sensing unit[J]. Mechanical Systems and Signal Processing, 2019, 126: 211-226.

[21] SUN J J, DING Y L, ZHANG H W, et al. Conceptual design and image motion compensation rate analysis of two-axis fast steering mirror for dynamic scan and stare imaging system[J]. Sensors, 2021, 21(19): 6441.

[22] ZHANG W F, YUAN J, YAN C X, et al. Multi-objective optimiza-

tion design of natural frequency of two-degree-of-freedom fast steering mirror system[J]. IEEE Access, 2021, 9: 33689-33703.

[23] 夏笠城, 王姝旸, 张晶, 等. 基于双频扩张状态观测器的无人机抗扰控制[J]. 北京航空航天大学学报, 2023, 49(5): 1201-1208.
XIA L C, WANG S Y, ZHANG J, et al. Bi-bandwidth extended state observer based disturbance rejection control method and its application on UAV[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(5): 1201-1208(in Chinese).

[24] LI J, ZHANG L Y, LUO L, et al. Extended state observer based current-constrained controller for a PMSM system in presence of disturbances: design, analysis and experiments[J]. Control Engineering Practice, 2023, 132: 105412.

Influence analysis of flexure support characteristics on performance of inertial reference unit

LI Xingfei[1, 2], HAN Jiaxin[1], TUO Weixiao[1, 2, *], LIU Yaqi[1], WANG Tianyu[1]

(1. State Key Laboratory of Precision Measurement Technology and Instruments, Tianjin University, Tianjin 300072, China;
2. Taihu Laboratory of Deepsea Technological Science, Wuxi 214000, China)

Abstract: The combination of compound axis tracking and aiming with an inertial reference unit (IRU) is the main technical means to reject vibration interference of moving platform and improve tracking and aiming accuracy. IRU usually adopts flexure support structure to avoid the defect of large moment of inertia of frame structure. The characteristics of the flexure support structure are the key factors to determine the bandwidth of IRU disturbance rejection. The influence of flexure support characteristics on IRU performance was quantitatively analyzed. Firstly, based on the mass-stiffness-damping model, the transfer function model of IRU system was established, and its accuracy was verified. Then, the parametric analysis method was used to study the influence of disturbance transmission characteristics and physical parameters of flexure support on the disturbance rejection ability of IRU system. Finally, a Simulink simulation model of IRU double-loop (velocity inner loop-position outer loop) control system was built. The influences of flexure support stiffness characteristics on IRU disturbance rejection ability and stabilization precision under a given vibration environment were simulated and analyzed. The results show that the disturbance transmission characteristics of the flexure support determine the ability of the system to reject high-frequency disturbance. Reducing the rotational stiffness of the flexure support can effectively improve the disturbance rejection ability of the IRU system but aggravate the negative impact of sensor noise and driver ripple on the stabilization precision of the IRU system.

Keywords: inertial reference unit; flexure support; disturbance rejection ability; rotational stiffness; parametric analysis

Received: 2023-07-14; **Accepted**: 2023-09-28; **Published Online**: 2023-10-19 18:56

URL: link.cnki.net/urlid/11.2625.V.20231019.1348.001

Foundation items: National Natural Science Foundation of China (62203322); China Postdoctoral Science Foundation (2022M712372); Enlisting and Leading Program of the Taihu Laboratory of Deepsea Technological Science (2022JBGS03001)

* **Corresponding author**. E-mail: tuoweixiao@tju.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0414

基于MPC的随机风场下航空器纵向自主间隔控制

汤新民[1,2,*]，陆晓娜[1]

(1. 南京航空航天大学 民航学院，南京 211106；2. 中国民航大学 交通科学与工程学院，天津 300300)

摘　　要：针对航迹随机扰动下的纵向自主间隔保持问题，高空风这一随机因素往往导致两机纵向间隔鲁棒性较差的情况，基于此，提出基于模型预测控制(MPC)的航空器纵向自主间隔控制方法。建立两机所受风场差值与纵向间隔的非线性运动学微分方程，推导线性时变预测模型，选取两机的纵向间隔和航路偏差距离为优化目标，高空风为随机扰动量，前机的真空速和偏航角作为观测量，并在空中安全和航空器性能约束中加入终端等式约束保持系统的稳定性。为验证所提方法的有效性，在规定的120 s仿真时间内，设置3组不同期望纵向间隔分别为12,13,14 km，通过设计的MPC控制器，在滚动时域周期内通过控制后机真空速和偏航角，两机纵向间隔曲线较为平滑且始终不低于最小安全间隔10 km，分别在第74,90,118 s稳定在期望的目标纵向间隔，在第58,74,95 s开始回归航路，最终回到航路中心线；设置了2组风场对照组，一组为预报风强度增大2倍，另一组为紊流风的扰动强度增大8倍，均能分别在第61,72 s平滑且稳定地建立期望的纵向间隔12 km。

关　键　词：空中交通管理；飞行跟驰模型；自主间隔控制；模型预测控制；随机扰动
中图分类号：V249.12; TB553
文献标志码：A　　**文章编号**：1001-5965(2025)09-2860-12

20世纪以来，随着航路上航空器数量的增多，空域容量接近极限水平，新航行系统带来的优势越来越不明显。整个航空器间隔调整的过程采用传统的集中式管理模式，即管制员通过自动化系统采用无线通话的方式向飞行机组传达间隔指令，机组根据指令要求操纵飞机完成间隔调配。这种模式不仅增加了管制员的工作负荷，也增加了机组操纵失误的可能性，导致航空器间的间隔不能精准实施。

为满足未来国家空域系统对灵活性和容量的需求，需改变飞行机组与地面管制之间传统的任务分配方式。早在1999年，美国国家航空航天局(National Aeronautics and Space Administration, NASA)提出了一种称为分布式空中/地面交通管理(distributed air/ground traffic management, DAG-TM)的方法[1]，目前主要由地面对飞机进行管制指挥的管理方式，转变为由飞机之间自主维持间隔的交通管理方式，解放服务效率对流量的限制，航空公司和飞行员可以提高机上主动性和灵活性[2]。中国民用航空局在2016年下发的信息通告《中国民航现代化空中交通管理系统体系架构》中也提出了分布式管理的概念：多模式间隔管理是在集中式管理的基础上，在各方条件都允许的情况下，把部分的间隔管理任务授权给机组来完成[3]。

基于航空器与航空器相对间隔的分布式空中交通管理意味着，当地面管制员为在同一固定航路上一前一后飞行的航空器配备了规定的管制间隔，在前机受到扰动导致运动状态发生改变时，两机的纵向间隔随之改变，不满足规定安全间隔时，后机通过由自主保持间隔系统(autonomous separation

收稿日期：2023-06-28；录用日期：2023-10-08；网络出版时间：2023-10-24 10:23
网络出版地址：link.cnki.net/urlid/11.2625.V.20231023.1041.003
基金项目：国家自然科学基金(61773202,52072174)
*通信作者. E-mail：tangxinmin@nuaa.edu.cn

引用格式：汤新民，陆晓娜. 基于MPC的随机风场下航空器纵向自主间隔控制[J]. 北京航空航天大学学报，2025，51(9)：2860-2871. TANG X M，LU X N. Longitudinal autonomous separation control of aircraft in random wind fields based on MPC[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2860-2871 (in Chinese).

assurance system, ASAS)实时计算提供间隔保障。目前, 关于航空器纵向自主间隔控制的研究主要集中在航空器纵向间隔建模、间隔保持算法 2 个方面。

在航空器纵向间隔建模方面, 中国民航大学王莉莉等针对近距平行跑道配对进近方式下的跟驰现象, 通过量化人为因素及尾流传播特性对配对进近两机纵向距离的影响建立配对进近微观跟驰模型[4], 并基于航路实际运行情况, 根据飞机流的跟驰效应, 分别从宏观和微观角度建立了纵向间隔的计算模型, 得出航路运行间隔的动态变化范围[5]。中国民航大学王超和朱明构建微观交通流模型, 分析了航路单向交通流和终端区进场交通流航空器时距分布特性[6]。南京航空航天大学华明壮等针对高速基于性能的导航(performance based navigation, PBN)航路的运行特点, 根据飞机的安全间隔和速度约束, 利用车辆跟驰理论, 建立了飞行管理系统的减速约束模型[7]。

在航空器间隔保持算法方面, 名古屋大学 Nakamura 等研究了 2 种类型在高密度空中走廊中运行的飞机的自主间隔控制算法: 基于相对速度矢量和基于横向位置[8-9], 并在此基础上进一步拓展到了考虑飞行员飞行意图[10], 以及飞行员手动控制[11]的场景。南京航空航天大学汤新民和郑鹏程针对航路序贯飞行中存在风场随机扰动因素下的 2 架航空器的间隔保持问题, 提出了基于滚动时域控制的航向角调整和真空速调整的混合间隔保持策略[12]。南京航空航天大学田勇等基于航空器动力学模型和比例积分控制理论, 提出了一种包含自主间隔、“车道通过”和“车道切换行为”的并行流道的自主间隔过程和算法[13]。

针对航路高空风导致沿航路飞行的前机与后机的航迹均会发生随机扰动, 间隔控制的鲁棒性较差的问题, 本文提出了基于模型预测控制(model predictive control, MPC) 的随机风场下航空器纵向自主间隔控制方法。首先, 将两机纵向间隔运动学微分方程看做一个状态变量为两机纵向间隔、控制变量为后机的真空速和偏航角、观测量为前机真空速和偏航角、随机扰动量为高空风的非线性控制系统。然后, 使用针对状态轨迹的线性化方法和前项欧拉法, 得到离散化线性化的预测模型。目标函数建立为在滚动时域周期内, 预测输出纵向间隔与期望输出纵向间隔之间的累计差值的加权平方和最小, 并加入终端等式约束保持系统的稳定性。最后, 建立仿真场景, 验证了 MPC 控制器的有效性。

1　航空器纵向间隔模型

1.1　航空器纵向间隔跟随规则

对于航空器纵向间隔的建模, 最初来源于研究单一车道车辆排队行为的地面车辆跟驰理论, 但不同于地面跟驰理论, 也不同于传统的集中式管理下的飞行跟驰模型, 同条航路航空器纵向自主间隔跟随模型具有以下特性:

1) 控制对象。当跟随航空器接收到前机扰动的消息, 立即进行自主计算跟随行为的纵向间隔保持控制律并执行机动飞行, 不再有管制员进行指挥和协调, 因此, 纵向间隔模型不需要考虑管制员和飞行员的反应时间。

2) 控制速度。在高空跟随过程中, 航空器的速度始终不为 0, 航空器性能差异较小, 在进行纵向间隔保持的前后, 前后机都以相同的巡航速度继续飞行。考虑高空风的存在及其随机性, 在纵向间隔模型中采用 2 架航空器的地速差 $\boldsymbol{V}_{\mathrm{G,f}}-\boldsymbol{V}_{\mathrm{G}}$ 而非传统的真空速差 $\boldsymbol{V}_{\mathrm{TAS,f}}-\boldsymbol{V}_{\mathrm{TAS}}$ 作为控制的速度量。

3) 控制纵向间隔。正常情况下, 两机纵向间隔保持不变。当前机受扰动减速时, 后机的跟随过程是一个动态响应的过程, 在此过程中, 不改变高度, 只进行水平方向间隔的改变, 两机纵向间隔会在一定范围内浮动, 最终建立合理的安全间隔。

1.2　随机风场下的航空器纵向间隔模型

根据同条航路航空器纵向间隔跟随规则, 建立航空器纵向间隔场景, 假设:

1) 平坦、不旋转的地面。

2) 标准大气。

3) 完全协调的飞行, 没有侧向力, 侧滑角始终为 0。

4) 在短时间机动飞行中, 飞机被视为一个质量恒定的点, 围绕其重心的转动运动动态可以忽略不计。

如图 1 所示, 跟随航空器在与前机进行间隔保持的过程中, 两者均受到高空风的影响, 由于所处位置的不同, 两者同一时刻受到的高空风风速和风向也不同。图中: $|\boldsymbol{V}_{\mathrm{TAS}}|$、$|\boldsymbol{V}_{\mathrm{G}}|$、$|\boldsymbol{V}_{\mathrm{W}}|$ 和 H_{W} 分别为跟随航空器的真空速、地速、所受到的高空风的风速和风向; $|\boldsymbol{V}_{\mathrm{TAS,f}}|$、$|\boldsymbol{V}_{\mathrm{G,f}}|$、$|\boldsymbol{V}_{\mathrm{W,f}}|$ 和 $H_{\mathrm{W,f}}$ 分别为前机的真空速、地速、所受到的高空风的风速和风向; φ 为跟随航空器的偏航角; φ_{f} 为前机的偏航角; N_H 为磁北方向, 简化计算航线角设置为 90°。

以沿前机的飞行方向(即前机地速 $|\boldsymbol{V}_{\mathrm{G,f}}|$ 方向)为横轴 x 正方向, 以初始时刻 t_0 跟随航空器所在位置与 x 轴的垂直交点为坐标轴原点 O, 机翼左侧并

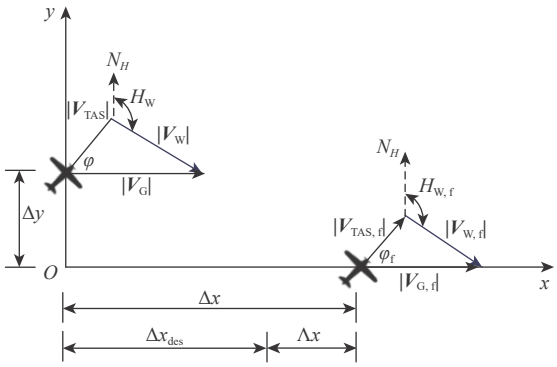


图1 随机风场下的航空器纵向间隔模型

Fig. 1 Aircraft longitudinal separation model in random wind fields

垂直于横轴的方向为纵轴y正方向。

初始时刻t_0，两机以相同的巡航速度飞行，并保持一定的安全间隔$d_0 > d_{\min}$，$d_{\min}$为管制规定的最小安全间隔。前机受到扰动运动状态改变，后机则需要做出机动行为，避免发生小于最小安全间隔的情况。

任意时刻，前机的真空速为$|\boldsymbol{V}_{\mathrm{TAS,f}}(t)|$，地速为$|\boldsymbol{V}_{\mathrm{G,f}}(t)|$，横向位移为$x_{\mathrm{f}}(t)$，纵向位移为$y_{\mathrm{f}}(t)$；后机的真空速为$|\boldsymbol{V}_{\mathrm{TAS}}(t)|$，地速为$|\boldsymbol{V}_{\mathrm{G}}(t)|$，横向位移为$x(t)$，纵向位移为$y(t)$。

根据2架航空器的相对位置关系，对如下概念进行定义。

两机的实时纵向间隔Δx为

$$\Delta x(t) = x_{\mathrm{f}}(t) - x(t) \tag{1}$$

两机的实时纵向间隔与期望纵向间隔Δx_{des}的差值$\Lambda x(t)$为

$$\Lambda x(t) = \Delta x(t) - \Delta x_{\mathrm{des}}(t) \tag{2}$$

两机的实时地速差$\Delta \boldsymbol{V}_{\mathrm{G}}(t)$为

$$\Delta \boldsymbol{V}_{\mathrm{G}}(t) = \boldsymbol{V}_{\mathrm{G,f}}(t) - \boldsymbol{V}_{\mathrm{G}}(t) \tag{3}$$

由航行速度三角形可知：

$$\boldsymbol{V}_{\mathrm{G}} = \boldsymbol{V}_{\mathrm{TAS}} + \boldsymbol{V}_{\mathrm{W}} \tag{4}$$

$$\boldsymbol{V}_{\mathrm{G,f}} = \boldsymbol{V}_{\mathrm{TAS,f}} + \boldsymbol{V}_{\mathrm{W,f}} \tag{5}$$

两机的纵向间隔为两机地速差的积分，则有风情况下的两机实时水平间隔运动学方程为

$$\begin{aligned}\int_{t_0}^{t}(\boldsymbol{V}_{\mathrm{G,f}}(t)-\boldsymbol{V}_{\mathrm{G}}(t))\mathrm{d}t &= \int_{t_0}^{t}[(\boldsymbol{V}_{\mathrm{TAS,f}}(t)+\boldsymbol{V}_{\mathrm{W,f}}(t))-\\ &(\boldsymbol{V}_{\mathrm{TAS}}(t)+\boldsymbol{V}_{\mathrm{W}}(t))]\mathrm{d}t = \int_{t_0}^{t}[(\boldsymbol{V}_{\mathrm{TAS,f}}(t)-\boldsymbol{V}_{\mathrm{TAS}}(t))+\\ &(\boldsymbol{V}_{\mathrm{W,f}}(t)-\boldsymbol{V}_{\mathrm{W}}(t))]\mathrm{d}t\end{aligned} \tag{6}$$

将两机的真空速和所受风速沿x轴和y轴分解，得到两机的纵向间隔和航路偏差距离运动学方程为

$$\begin{aligned}\Delta x(t) = \int_{t_0}^{t}&[(|\boldsymbol{V}_{\mathrm{TAS,f}}(t)|\cos\varphi_{\mathrm{f}}(t)-|\boldsymbol{V}_{\mathrm{TAS}}(t)|\cos\varphi(t))+\\ &(|\boldsymbol{V}_{\mathrm{W,f}}(t)|\cos H_{\mathrm{W,f}}(t)-|\boldsymbol{V}_{\mathrm{W}}(t)|\cos H_{\mathrm{W}}(t))]\mathrm{d}t\end{aligned} \tag{7}$$

$$\begin{aligned}\Delta y(t) = \int_{t_0}^{t}&[(|\boldsymbol{V}_{\mathrm{TAS,f}}(t)|\sin\varphi_{\mathrm{f}}(t)-|\boldsymbol{V}_{\mathrm{TAS}}(t)|\sin\varphi(t))+\\ &(|\boldsymbol{V}_{\mathrm{W,f}}(t)|\sin H_{\mathrm{W,f}}(t)-|\boldsymbol{V}_{\mathrm{W}}(t)|\sin H_{\mathrm{W}}(t))]\mathrm{d}t\end{aligned} \tag{8}$$

1.3 随机风场模型

在实际飞行中，航空器会遇到大气的扰动。大气扰动一般可以分解为平均风和紊流风，在大气的运动过程中，风速一般围绕平均风值在相对较小的范围内进行波动[14]。为研究两机在大气扰动下的间隔保持，有必要建立合理的时变风场模型，模拟大气运动过程。

本文将两机之间的风速差设置为平均风和紊流风2个部分的和。平均风为两机预报平均风的差值，紊流风为在平均风值上下较小范围内服从一定概率分布的波动。

在某一时刻，两机的高空风的风速$|\boldsymbol{V}_{\mathrm{W}}|$、$|\boldsymbol{V}_{\mathrm{W,f}}|$和风向$H_{\mathrm{W}}$、$H_{\mathrm{W,f}}$信息可以通过航空天气预报获得。设沿$x$轴和$y$轴的随机风场可以建模为一个预报风矢量（风速$|\hat{\boldsymbol{V}}_{\mathrm{W}}|$和风向$\hat{H}_{\mathrm{W}}$已知）和一个在预报风上下随机扰动的高空风误差的和，随机扰动服从某一相对误差水平$\delta_{\mathrm{w},i}$，均值为0、方差为1的高斯分布随机数。式(9)为加入随机扰动后的风场表达式：

$$V^{\delta}_{\mathrm{W},i}(t) = \hat{V}_{\mathrm{W},i}(t) + \delta_{\mathrm{w},i}R^{n}_{i}(t)\hat{V}_{\mathrm{W},i}(t) \tag{9}$$

式中：$V^{\delta}_{\mathrm{W},i}(t)$为添加了高斯分布噪声的风场数据；$\hat{V}_{\mathrm{W},i}(t)$为某一方向的预报风，$i = x, y$；$\delta_{\mathrm{w},i}$为相对误差水平，通过改变$\delta_{\mathrm{w},i}$的大小，可以增大或减小风场的扰动；$R^{n}_{i} \sim N(\mu,\sigma^2)$为$\mu = 0$、$\sigma = 1$的标准高斯分布随机数。

根据1.2节得到的沿x轴和y轴风向的两机所受预报风速差值，2个方向加入随机扰动后的风场表达式为

$$V^{\delta}_{\mathrm{W},x}(t) = \hat{V}_{\mathrm{W},x}(t) + \delta_{\mathrm{w},x}R^{n}_{x}(t)\hat{V}_{\mathrm{W},x}(t) \tag{10}$$

$$V^{\delta}_{\mathrm{W},y}(t) = \hat{V}_{\mathrm{W},y}(t) + \delta_{\mathrm{w},y}R^{n}_{y}(t)\hat{V}_{\mathrm{W},y}(t) \tag{11}$$

式中：

$$\hat{V}_{\mathrm{W},x}(t) = |\hat{\boldsymbol{V}}_{\mathrm{W,f}}(t)|\cos\hat{H}_{\mathrm{W,f}}(t) - |\hat{\boldsymbol{V}}_{\mathrm{W}}(t)|\cos\hat{H}_{\mathrm{W}}(t) \tag{12}$$

$$\hat{V}_{\mathrm{W},y}(t) = |\hat{\boldsymbol{V}}_{\mathrm{W,f}}(t)|\sin\hat{H}_{\mathrm{W,f}}(t) - |\hat{\boldsymbol{V}}_{\mathrm{W}}(t)|\sin\hat{H}_{\mathrm{W}}(t) \tag{13}$$

在进行间隔保持的过程中，风场持续一段时间以指数衰减函数$y = ae^{-bx}$的形式周期性减弱，其中，a为峰值强度，b为衰减系数。

1.4 纵向间隔非线性运动学模型的线性化离散化

由1.2节和1.3节可得到随机风场下两机纵向间隔和航路偏差距离运动学方程为

$$\begin{cases}\Delta x=\int_{t_0}^{t}(|\boldsymbol{V}_{\mathrm{TAS,f}}(t)|\cos\varphi_{\mathrm{f}}(t)-|\boldsymbol{V}_{\mathrm{TAS}}(t)|\cos\varphi(t)+V_{\mathrm{W},x}^{\delta}(t))\mathrm{d}t\\ \Delta y=\int_{t_0}^{t}(|\boldsymbol{V}_{\mathrm{TAS,f}}(t)|\sin\varphi_{\mathrm{f}}(t)-|\boldsymbol{V}_{\mathrm{TAS}}(t)|\sin\varphi(t)+V_{\mathrm{W},y}^{\delta}(t))\mathrm{d}t\end{cases} \tag{14}$$

根据两机的纵向间隔和航路偏差距离的运动学方程可知, 通过控制跟随航空器的真空速$|\boldsymbol{V}_{\mathrm{TAS}}|$、偏航角$\varphi$与前机保持间隔, 可将纵向间隔控制看作如下控制系统: 控制输入为$\boldsymbol{u}=[|\boldsymbol{V}_{\mathrm{TAS}}|,\varphi]^{\mathrm{T}}$, 状态变量为$\boldsymbol{x}=\left[\Delta x,\Delta y\right]^{\mathrm{T}}$, 前机的真空速$|\boldsymbol{V}_{\mathrm{TAS,f}}|$和偏航角$\varphi_{\mathrm{f}}$作为观测量, 在每一个滚动时域周期内为常数项$\boldsymbol{d}=\left[|\boldsymbol{V}_{\mathrm{TAS,f}}|\cos\varphi_{\mathrm{f}},|\boldsymbol{V}_{\mathrm{TAS,f}}|\sin\varphi_{\mathrm{f}}\right]^{\mathrm{T}}$; 沿 x 轴和 y 轴的高空风为随机扰动量$\boldsymbol{\omega}=[\hat{V}_{\mathrm{W},x},\hat{V}_{\mathrm{W},y}]^{\mathrm{T}}$的控制系统。

可以看到, 纵向间隔模型的是一个非线性系统, 不能直接用于线性时变预测模型, MPC 中一般采用近似的线性化方法。 本文使用针对状态轨迹的线性化方法, 相比于存在参考系统的线性化方法, 优势在于不需要预先得到期望跟踪路径的状态量和控制量[15]。

考虑系统的某个工作点为$[\boldsymbol{\xi}_0,\boldsymbol{u}_0]$, $\dot{\boldsymbol{\xi}}_0(k)$为始终施加控制量为$\boldsymbol{u}_0$后得到的系统状态量, 则存在以下关系:

$$\begin{cases}\dot{\boldsymbol{\xi}}_0=f(\boldsymbol{\xi}_0(k),\boldsymbol{u}_0)\\ \boldsymbol{\xi}_0(0)=\boldsymbol{\xi}_0\end{cases} \tag{15}$$

对于航空器的参考轨迹, 其上每一轨迹点也必定满足式 (15) 的运动学方程, 因此, 任意时刻的参考状态方程可表示为

$$\dot{\boldsymbol{\xi}}_0=f(\boldsymbol{\xi}_0,\boldsymbol{u}_0) \tag{16}$$

式中: $\boldsymbol{\xi}_0$和$\boldsymbol{u}_0$分别表示参考轨迹任意时刻的参考状态量$\boldsymbol{\xi}=\left[\Delta x,\Delta y\right]^{\mathrm{T}}$和参考控制量$\boldsymbol{u}=[|\boldsymbol{V}_{\mathrm{TAS}}|,\varphi]^{\mathrm{T}}$。

利用泰勒公式将式 (16) 在工作点$[\boldsymbol{\xi}_0,\boldsymbol{u}_0]$处进行展开, 并且仅保留常数项和一阶项, 得到如下结果:

$$\dot{\boldsymbol{\xi}}=f(\boldsymbol{\xi}_0,\boldsymbol{u}_0)+\frac{\partial f(\boldsymbol{\xi},\boldsymbol{u})}{\partial\boldsymbol{\xi}}(\boldsymbol{\xi}-\boldsymbol{\xi}_0)+\frac{\partial f(\boldsymbol{\xi},\boldsymbol{u})}{\partial\boldsymbol{u}}(\boldsymbol{u}-\boldsymbol{u}_0) \tag{17}$$

将式 (16) 与式 (17) 相减, 可得

$$\dot{\tilde{\boldsymbol{\xi}}}=\begin{bmatrix}\dot{\boldsymbol{x}}-\dot{\boldsymbol{x}}_0\\ \dot{\boldsymbol{y}}-\dot{\boldsymbol{y}}_0\end{bmatrix}=\frac{\partial f(\boldsymbol{\xi},\boldsymbol{u})}{\partial\boldsymbol{\xi}}\begin{bmatrix}\Delta x-\Delta x_0\\ \Delta y-\Delta y_0\end{bmatrix}+\frac{\partial f(\boldsymbol{\xi},\boldsymbol{u})}{\partial\boldsymbol{u}}\begin{bmatrix}|\boldsymbol{V}_{\mathrm{TAS}}|-|\boldsymbol{V}_{\mathrm{TAS},0}|\\ \varphi-\varphi_0\end{bmatrix} \tag{18}$$

令$\dot{\tilde{\boldsymbol{\xi}}}=\dot{\boldsymbol{\xi}}-\dot{\boldsymbol{\xi}}_0$、$\tilde{\boldsymbol{\xi}}=\boldsymbol{\xi}-\boldsymbol{\xi}_0$、$\tilde{\boldsymbol{u}}=\boldsymbol{u}-\boldsymbol{u}_0$分别表示模型各变量的误差。$\frac{\partial f(\boldsymbol{\xi},\boldsymbol{u})}{\partial\boldsymbol{\xi}}$和$\frac{\partial f(\boldsymbol{\xi},\boldsymbol{u})}{\partial\boldsymbol{u}}$为二元函数的雅可比矩阵, 将式 (16) 代入式 (18), 可得纵向间隔的线性化系统状态空间方程为

$$\dot{\tilde{\boldsymbol{\xi}}}(t+1)=\boldsymbol{A}(t)\tilde{\boldsymbol{\xi}}(t)+\boldsymbol{B}(t)\tilde{\boldsymbol{u}}(t)+\boldsymbol{G}(t)\boldsymbol{\omega}(t)+\boldsymbol{F}(t)\boldsymbol{d}(t) \tag{19}$$

式中: $\boldsymbol{A}(t)=\begin{bmatrix}0&0\\0&0\end{bmatrix}$; $\boldsymbol{B}(t)=\begin{bmatrix}-\cos\varphi(t)&|\boldsymbol{V}_{\mathrm{TAS}}(t)|\sin\varphi(t)\\-\sin\varphi(t)&-|\boldsymbol{V}_{\mathrm{TAS}}(t)|\cos\varphi(t)\end{bmatrix}$; $\boldsymbol{G}(t)=\begin{bmatrix}1+\delta_{\mathrm{w},x}R_x^n(t)&0\\0&1+\delta_{\mathrm{w},y}R_y^n(t)\end{bmatrix}$; $\boldsymbol{F}(t)=\begin{bmatrix}-1&0\\0&-1\end{bmatrix}$。

由此可得到线性化的状态方程。对于线性时变系统, 其系数矩阵不再是常数系统, 矩阵元素随着时间变化而变化。

由于该状态方程是连续的, 不能直接用于 MPC 控制器的设计, 有必要对其进行离散化处理。采用前项欧拉法对该模型进行离散化处理, 如下:

$$\dot{\tilde{\boldsymbol{\xi}}}(k+1)=\frac{\tilde{\boldsymbol{\xi}}(k+1)-\tilde{\boldsymbol{\xi}}(k)}{\Delta T}=\boldsymbol{A}(k)\tilde{\boldsymbol{\xi}}(k)+\boldsymbol{B}(k)\tilde{\boldsymbol{u}}(k)+\boldsymbol{G}(k)\boldsymbol{\omega}(k)+\boldsymbol{F}(k)\boldsymbol{d}(k) \tag{20}$$

式中: ΔT为采样周期。

可得

$$\boldsymbol{A}_{k,x}=\boldsymbol{I}+\Delta T\cdot\boldsymbol{A}(k) \tag{21}$$

$$\boldsymbol{B}_{k,u}=\Delta T\cdot\boldsymbol{B}(k) \tag{22}$$

$$\boldsymbol{G}_{k,\omega}=\Delta T\cdot\boldsymbol{G}(k) \tag{23}$$

$$\boldsymbol{F}_{k,d}=\Delta T\cdot\boldsymbol{F}(k) \tag{24}$$

离散化后的系统状态空间方程表达式如下:

$$\tilde{\boldsymbol{\xi}}(k+1)=\boldsymbol{A}_{k,x}(k)\tilde{\boldsymbol{\xi}}(k)+\boldsymbol{B}_{k,u}(k)\tilde{\boldsymbol{u}}(k)+\boldsymbol{G}_{k,\omega}(k)\boldsymbol{\omega}(k)+\boldsymbol{F}_{k,d}(k)\boldsymbol{d}(k) \tag{25}$$

式中: $\boldsymbol{A}_{k,x}$、$\boldsymbol{B}_{k,u}$、$\boldsymbol{G}_{k,\omega}$和$\boldsymbol{F}_{k,d}$分别为状态矩阵、输入矩阵、随机扰动系数矩阵和可测系数矩阵, 各矩阵的表达式为: $\boldsymbol{A}_{k,x}(k)=\begin{bmatrix}1&0\\0&1\end{bmatrix}$, $\boldsymbol{B}_{k,u}(k)=\begin{bmatrix}-\Delta T\cos\varphi(k)&|\boldsymbol{V}_{\mathrm{TAS}}(k)|\Delta T\sin\varphi(k)\\-\Delta T\sin\varphi(k)&-|\boldsymbol{V}_{\mathrm{TAS}}(k)|\Delta T\cos\varphi(k)\end{bmatrix}$, $\boldsymbol{F}_{k,d}(k)=\begin{bmatrix}-\Delta T&0\\0&-\Delta T\end{bmatrix}$, $\boldsymbol{G}_{k,\omega}(k)=\begin{bmatrix}(1+\delta_{\mathrm{w},x}R_x^n(k))\Delta T&0\\0&(1+\delta_{\mathrm{w},y}R_y^n(k))\Delta T\end{bmatrix}$。

2　模型预测控制纵向间隔控制设计

2.1　模型预测控制问题描述

MPC 是近年来被广泛讨论的一种反馈控制策略, 遵循预测模型、滚动优化和反馈校正的基本原理[16]。

通过第 1 节的分析, 航空器纵向间隔保持问题可描述为: $k=0,1,2,\cdots$为仿真的每一个采样时刻, 系统通过广播式自动相关监视 (automatic depentent surveillance broadcast, ADS-B) 实时采集前机的状态信息, 并通过航空天气预报获得当前时刻的预报风信息; 在每一采样时刻k, 在系统规定的滚动时域周期N内通过纵向间隔模型$\tilde{\boldsymbol{\xi}}(k+1)=F(\tilde{\boldsymbol{\xi}}(k),\tilde{\boldsymbol{u}}(k))$作为

预测模型，求解出滚动时域周期内的一系列最优的控制变量，使得目标函数$J(\tilde{\boldsymbol{\xi}}(k),\tilde{\boldsymbol{u}}(k))$最小，并将$\boldsymbol{U}_k^*$的第一个分量$\boldsymbol{u}_k^*$作用于控制系统，并且在$k+1$时刻以重新得到的前机的状态信息作为初始条件，重新预测系统未来输出，刷新优化问题，如此循环往复，滚动优化[17]。

2.2 系统控制目标与约束

本文考虑后机间隔保持的性能指标及约束条件，在第1节建立的随机风场下的纵向间隔模型的基础上，将优化目标转化为一个带约束的二次规划问题，实现纵向间隔的多目标控制。

图2为基于MPC的纵向间隔控制设计示意图。

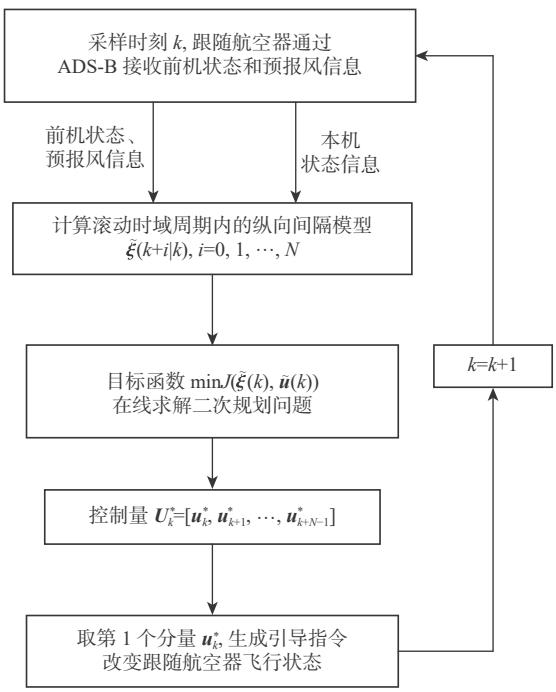


图2 基于模型预测控制的纵向间隔控制设计

Fig. 2 Design of longitudinal separation control based on MPC

1）控制目标。为使跟随航空器在检测到前机受扰动时能够及时进行机动飞行与前机保持间隔，使两机纵向间隔逐渐趋向于期望的安全间隔Δx_{des}，并且航路偏差距离Δy均收敛于0：

$$\begin{cases}\Delta x-\Delta x_{\text{des}}=0\\ \Delta y=0\end{cases} \tag{26}$$

2）系统约束。状态量约束：①考虑空中交通运行的安全性，两机的纵向间隔Δx既不小于管制规定的最小安全间隔，也不能过大造成空域的浪费，②考虑航路的宽度的限制，两机的航路偏差距离Δy不能超出航路的宽度：

$$\begin{cases}\Delta x_{\min}\leqslant\Delta x\leqslant\Delta x_{\max}\\ -\Delta y_{\max}\leqslant\Delta y\leqslant\Delta y_{\max}\end{cases} \tag{27}$$

控制量约束：考虑航空器自身的性能，跟随航空器的真空速$|\boldsymbol{V}_{\text{TAS}}|$和偏航角$\varphi$有如下约束：

$$\begin{cases}V_{\text{TAS,min}}\leqslant|\boldsymbol{V}_{\text{TAS}}|\leqslant V_{\text{TAS,max}}\\ \varphi_{\min}\leqslant\varphi\leqslant\varphi_{\max}\end{cases} \tag{28}$$

2.3 目标函数的建立

根据本文的控制目标，选择两机实时纵向间隔与期望纵向间隔的差值$\Lambda x(k)$,和航路偏差距离$\Delta y(k)$作为系统的输出量$\boldsymbol{\eta}(k)=\left[\Lambda x(k),\Delta y(k)\right]^{\mathrm{T}}$，输出方程表达式为

$$\boldsymbol{\eta}(k)=\boldsymbol{C}_{k,y}\tilde{\boldsymbol{\xi}}(k)-z \tag{29}$$

$$\begin{cases}\boldsymbol{C}_{k,y}=\begin{bmatrix}1&0\\0&1\end{bmatrix}\\ z=\begin{bmatrix}\Delta x_{\text{des}}(k)\\0\end{bmatrix}\end{cases} \tag{30}$$

对于不稳定的约束线性系统，在有限时域滚动优化问题中加入终端等式约束$\boldsymbol{\eta}(k+N)=\boldsymbol{0}$，迫使滚动时域终端状态回到平衡点，即为达到系统的最终目标，需满足：①两机达到期望纵向间隔$\Delta x=\Delta x_{\text{des}}$；②后机回到航路中心线，$\Delta y=0$。而对于非线性系统或不稳定的有约束线性系统，加入终端约束也是必须的，以保证系统的Lyapunov稳定性[16]。

由此建立如下目标函数：

$$\begin{cases}J(\tilde{\boldsymbol{\xi}}(t),\tilde{\boldsymbol{u}}(t))=\int_{t_0}^{t_{\text{f}}}F(\tilde{\boldsymbol{\xi}}(t),\tilde{\boldsymbol{u}}(t),t)\mathrm{d}t+\phi(\tilde{\boldsymbol{\xi}}(t_{\text{f}}),t_{\text{f}})\\ \dot{\tilde{\boldsymbol{\xi}}}(t)=F(\tilde{\boldsymbol{\xi}}(t)\tilde{\boldsymbol{u}}(t),t)\\ \tilde{\boldsymbol{\xi}}(t_0)=\tilde{\boldsymbol{\xi}}_0\end{cases} \tag{31}$$

式中：$F(\tilde{\boldsymbol{\xi}}(t),\tilde{\boldsymbol{u}}(t),t)$为积分指标；$\phi(\tilde{\boldsymbol{\xi}}(t_{\text{f}}),t_{\text{f}})$为终端指标；$\dot{\tilde{\boldsymbol{\xi}}}(t)=F(\tilde{\boldsymbol{\xi}}(t),\tilde{\boldsymbol{u}}(t),t)$为系统线性化后的连续状态空间方程；$t_0$和$\tilde{\boldsymbol{\xi}}_0$分别为初始时刻和初始状态；$t_{\text{f}}$和$\tilde{\boldsymbol{\xi}}_{t_{\text{f}}}$分别为终端时刻和终端状态。使用设计的MPC控制器对纵向间隔模型分为多个时域进行滚动优化，采样周期为ΔT。目标函数建立在滚动时域周期内，使两机实时纵向间隔与期望纵向间隔的差值$\Lambda x(k)$和航路偏差距离$\Delta y(k)$的加权平方和最小，最终目标为：$\Lambda x=0$，$\Delta y=0$。本文采用离散装置控制连续时间受控系统，连续化的目标函数也要化为针对离散系统的形式[18]：

$$\begin{aligned}&\min J(\tilde{\boldsymbol{\xi}}(k),\tilde{\boldsymbol{u}}(k),N,\boldsymbol{Q},\boldsymbol{P})=\\&\sum_{i=0}^{N-1}F(\tilde{\boldsymbol{\xi}}(k+i|k),\tilde{\boldsymbol{u}}(k),k)+\phi\left(\tilde{\boldsymbol{\xi}}(k+N|k),k+N\right)=\\&\sum_{i=0}^{N-1}\left[q_1\Lambda x(k+i|k)^2+q_2\Delta y(k+i|k)^2\right]+\\&\left[p_1\Lambda x(k+N|k)^2+p_2\Delta y(k+N|k)^2\right]\end{aligned} \tag{32}$$

式中：$\boldsymbol{Q}=\mathrm{diag}(q_1,q_2)$为两机纵向间隔误差$\Lambda x(k)$

和航路偏差距离$\Delta y(k)$的权重系数矩阵; $\boldsymbol{P}=\mathrm{diag}(p_1,p_2)$为终端约束项的权重矩阵, 表示 2 个性能指标所占比例; $q_1+q_2=1$, $p_1+p_2=1$。通过设置适当的权重系数, 可以同时满足系统对期望的纵向间隔的跟随能力和回归航路中心线飞行的能力。

将目标函数中的加权和写成矩阵形式为

$$\sum_{i=0}^{N-1}\boldsymbol{\eta}(k+i|k)^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\eta}(k+i|k)+\boldsymbol{\eta}(k+N|k)^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\eta}(k+N|k)=$$

$$\sum_{i=0}^{N-1}\|\boldsymbol{\eta}(k+i|k)\|_{\boldsymbol{Q}}^2+\|\boldsymbol{\eta}(k+N|k)\|_{\boldsymbol{P}}^2=$$

$$\begin{bmatrix}\boldsymbol{\eta}(k\mid k)\\ \boldsymbol{\eta}(k+1\mid k)\\ \vdots\\ \boldsymbol{\eta}(k+N\mid k)\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\boldsymbol{Q}&&&\\ &\boldsymbol{Q}&&\\ &&\ddots&\\ &&&\boldsymbol{P}\end{bmatrix}\begin{bmatrix}\boldsymbol{\eta}(k\mid k)\\ \boldsymbol{\eta}(k+1\mid k)\\ \vdots\\ \boldsymbol{\eta}(k+N\mid k)\end{bmatrix}=$$

$$\sum_{i=0}^{N}\|\boldsymbol{\eta}(k+i|k)\|_{\tilde{\boldsymbol{Q}}}^2 \tag{33}$$

$$\tilde{\boldsymbol{Q}}=\mathrm{diag}(\boldsymbol{Q},\boldsymbol{Q},\cdots,\boldsymbol{Q},\boldsymbol{P})_{N\times N} \tag{34}$$

该优化问题满足离散化系统运动学方程:

$$\begin{cases}\tilde{\boldsymbol{\xi}}(k+1)=F(\tilde{\boldsymbol{\xi}}(k),\tilde{\boldsymbol{u}}(k),k)\\ \boldsymbol{\eta}(k)=h(\tilde{\boldsymbol{\xi}}(k),\tilde{\boldsymbol{u}}(k),k)\\ \tilde{\boldsymbol{\xi}}(k_0)=\tilde{\boldsymbol{\xi}}_0\end{cases} \tag{35}$$

满足终端状态约束:

$$\boldsymbol{\eta}(k+N|k)=\boldsymbol{0} \tag{36}$$

满足状态变量和控制量约束:

$$\begin{cases}\Delta x_{\min}\leqslant\Delta x(k+i|k)\leqslant\Delta x_{\max}\\ -\Delta y_{\max}\leqslant\Delta y(k+i|k)\leqslant\Delta y_{\max}\end{cases}\quad i=0,1,2,\cdots,N \tag{37}$$

$$\begin{cases}V_{\mathrm{TAS},\min}\leqslant|\boldsymbol{V}_{\mathrm{TAS}}(k+i|k)|\leqslant V_{\mathrm{TAS},\max}\\ \varphi_{\min}\leqslant\varphi(k+i|k)\leqslant\varphi_{\max}\end{cases}\quad i=0,1,2,\cdots,N \tag{38}$$

式(33)～式(38)的功能是使跟随航空器能够尽快且平稳地与前机保持纵向间隔。

最终可以将目标函数化成二次规划标准形式, 对于此类含有等式和不等式的二次规划问题, 可以调用 MATLAB 中的 quadprog 求解器。每个滚动时域周期对式(33)～式(38)进行求解, 得到一系列最优的控制变量:

$$\boldsymbol{U}_k^*=[\boldsymbol{u}_k^*,\boldsymbol{u}_{k+1}^*,\cdots,\boldsymbol{u}_{k+N-1}^*] \tag{39}$$

将式(39)最优控制变量第 1 个元素作用系统:

$$\boldsymbol{u}(k)=\boldsymbol{u}_k^* \tag{40}$$

每个滚动时域周期中, 通过计算式(39)和式(40)控制跟随航空器的真空速和偏航角, 实现纵向间隔控制。

3　仿真验证及结果分析

3.1　仿真场景建立和参数设置

为验证跟随航空器的纵向间隔控制系统的鲁棒性, 设置如下仿真场景。

本文假设在同一航路飞行的航空器性能相近, 均以 B737-800 机型为例, 机型性能数据来源于飞行计划和性能手册[19]。

在间隔保持的过程中, 航空器视为质量恒定的质点, 航空器巡航高度为 9.144 km, 此高度的高空风速一般在几十米每秒。

初始时刻t_0, 2 架航空器在同一固定航路上以相同的巡航速度$|\boldsymbol{V}_{\mathrm{TAS,f}}(t_0)|=|\boldsymbol{V}_{\mathrm{TAS}}(t_0)|=250$ m/s飞行, 并保持一定的安全间隔, 且两机均未发生偏航, 前机与后机的初始纵向间隔为$d_0=10$ km, 管制规定的最小安全间隔$d_{\min}$设置为 10 km, 航路宽度为航路中心线两侧各 10 km。当后机接收到前机的扰动信息后, 立即实时滚动计算纵向间隔控制律, 通过控制真空速和偏航角执行间隔保持。

1) 前机运动状态参数设置。为真实模拟前机运动状态的扰动及验证本文方法的有效性, 在规定的仿真时间内, 对前机的真空速在(250±25) m/s的范围内进行加速、减速、匀速运动的变化。图 3 为模拟的前机的真空速与时间图。

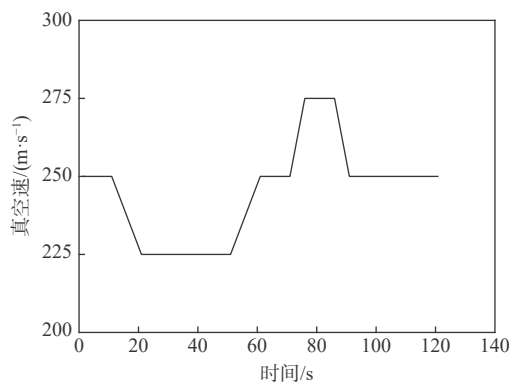


图 3　前机真空速与时间的关系

Fig. 3　Relationship between vacuum velocity and time of front aircraft

2) 随机风场参数设置。本文设置通过航空天气预报获得的前机和后机在初始时刻的预报风的风速和风向分别为: $|\hat{\boldsymbol{V}}_{\mathrm{W,f}}|=40$ m/s, $\hat{H}_{\mathrm{W,f}}=40°$, $|\hat{\boldsymbol{V}}_{\mathrm{W}}|=30$ m/s, $\hat{H}_{\mathrm{W}}=45°$。

在初始时刻, 沿x轴和y轴风向的两机所受预报风速的差值计算为: $\hat{V}_{\mathrm{W},x}=|\hat{\boldsymbol{V}}_{\mathrm{W,f}}|\cos\hat{H}_{\mathrm{W,f}}-|\hat{\boldsymbol{V}}_{\mathrm{W}}|\cos\hat{H}_{\mathrm{W}}=9.428\,6$ m/s, $\hat{V}_{\mathrm{W},y}=|\hat{\boldsymbol{V}}_{\mathrm{W,f}}|\sin\hat{H}_{\mathrm{W,f}}-|\hat{\boldsymbol{V}}_{\mathrm{W}}|\sin\hat{H}_{\mathrm{W}}=4.498\,3$ m/s。

随机风场模型参数如表 1 所示。

表 1 随机风场模型参数

Table 1 Parameters of random wind model

参数	x轴	y轴
平均风预报风差值/(m·s^{-1})	9.428 6	4.498 3
紊流风的概率分布	$R_x^n \sim N(0,1)$	$R_y^n \sim N(0,1)$
紊流风的相对误差水平	0.1	0.08
起始峰值强度	9.428 6	4.498 3
衰减系数	0.1	0.15
衰减时间/s	40～50,80～120	40～50,80～120

设置前机的运动状态和随机风场模型的各项参数后，基于建立的纵向间隔模型，在 MATLAB 中调用 quadprog 求解器进行求解，使用计算后机跟随策略的 MPC 控制器的参数设置如图 4 和表 2 所示。

控制系统控制过程中的控制量约束和输出量约束范围如表 3 所示。

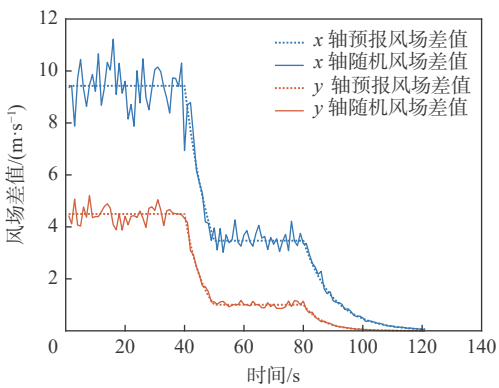


图 4 x、y 方向风场差值与时间的关系

Fig. 4 Relationship between wind field difference in x and y directions and time

表 2 MPC 控制器参数

Table 2 Model predictive controller parameters

参数	设定值
状态量个数N_x	2
控制量个数N_u	2
输出量个数N_y	2
随机扰动量个数N_w	2
可测扰动量个数N_f	1
仿真步长	120
采样周期ΔT/s	1
滚动时域周期N/s	10
跟踪项加权系数矩阵$\boldsymbol{Q}$	[0.8,0.2]
终端项加权系数矩阵$\boldsymbol{P}$	[0.8,0.2]

表 3 控制量约束和输出量约束范围

Table 3 Control quantity and output quantity constraint range

系统量	参数	约束范围
控制量	真空速/(m·s^{-1})	[220,280]
	偏航角/(°)	[−10,10]
输出量	纵向间隔/km	[10,15]
	航路偏差距离/km	[−10,10]

3.2 基于 MPC 的纵向间隔仿真结果验证与分析

3.2.1 不同期望纵向间隔下的方法验证

如图 5 所示，初始时刻，在后机接收到前机速度扰动的信息时，如果后机不执行间隔保持，第 26 s 开始，两机的纵向间隔为 9.997 km<10 km，开始突破最小安全间隔；从初始时刻受到持续了一定时间的高空风的作用，并开始逐渐偏离航路中心线，偏离的最大距离为 237.091 6 m。

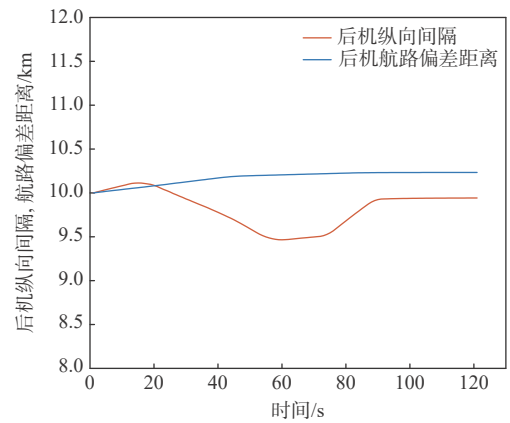


图 5 不执行间隔保持情况下的后机纵向间隔和航路偏差距离曲线

Fig. 5 Longitudinal separation and route deviation curves of rear aircraft without separation maintenance

后机必须立即进行自主计算跟随控制律，通过控制真空速和偏航角与前机保持间隔。最终目标为后机达到与前机相同的巡航速度，并以建立的新的安全间隔沿着航路继续飞行。

为验证本文方法的有效性与鲁棒性，本文设置 3 组不同期望纵向间隔：Δx_{des} = 12 km、Δx_{des} = 13 km、Δx_{des} = 14 km，分别进行仿真实验。图 6～图 10 分别为不同目标期望纵向间隔下纵向间隔-时间图、两机实时纵向间隔与期望纵向间隔的差值-时间图、航路偏差距离-时间图、后机真空速-时间图和后机偏航角-时间图。

从图 6 和图 7 可知，后机通过设计的 MPC 控制器，不断调整真空速与偏航角的 120 s 的机动时间内，两机纵向间隔始终不低于最小安全间隔 10 km，

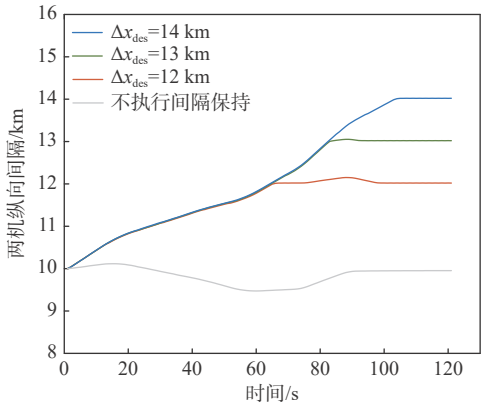


图 6　不同期望纵向间隔下的两机纵向间隔-时间曲线

Fig. 6　Longitudinal separation-time curves under different expectation longitudinal separation between two aircrafts

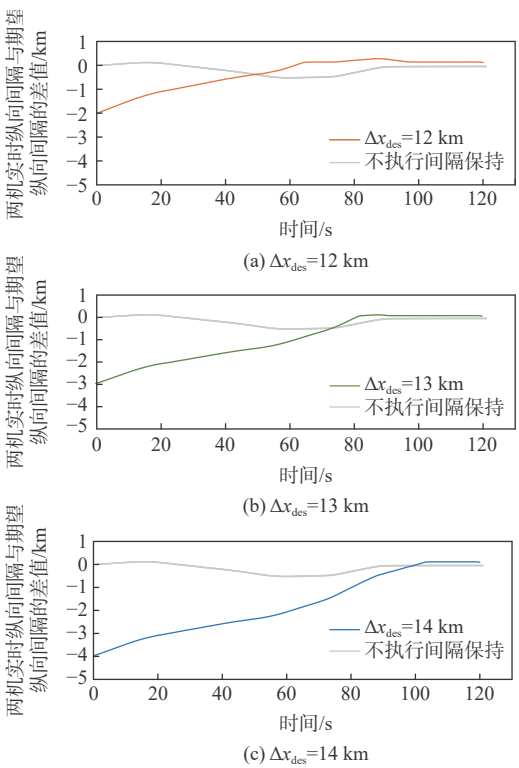


图 7　不同期望纵向间隔下的两机实时纵向间隔与期望纵向间隔的差值-时间曲线

Fig. 7　Difference between real-time longitudinal separation and expectation longitudinal separation-time curves under different expectation longitudinal separations between two aircrafts

两机的纵向间隔曲线较为平滑，且最终稳定在期望的纵向间隔。从表 4 结果也可以看出，Δx_{des} = 12,13, 14 km 这 3 种设置的期望纵向间隔对应的纵向间隔建立时间分别为 74 s、90 s、118 s。期望纵向间隔越大，纵向间隔稳定所需的时间也越长。

从图 8 可知，后机通过设计的 MPC 控制器，不断调整真空速与偏航角的 120 s 的仿真时间内，Δx_{des} = 12,13,14 km这 3 种设置的期望纵向间隔对应的最大航路偏差距离分别为 880.46 m(第 14 s)，

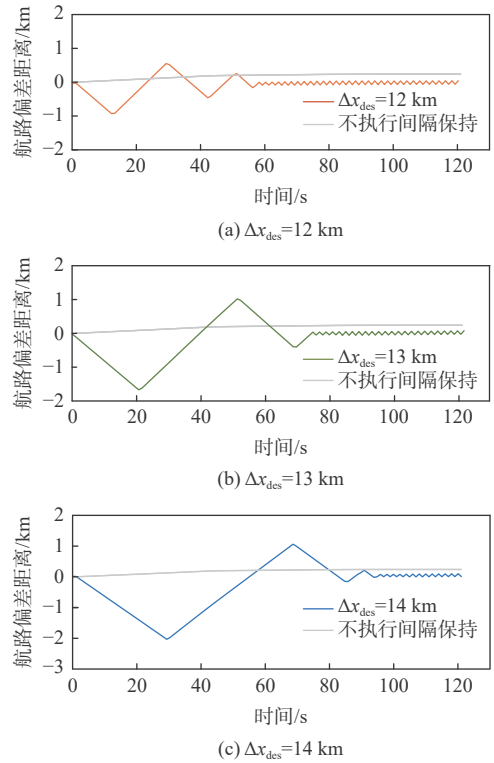


图 8　不同期望纵向间隔下的航路偏差距离-时间曲线

Fig. 8　Route deviation distance-time curves under different expectation longitudinal separations

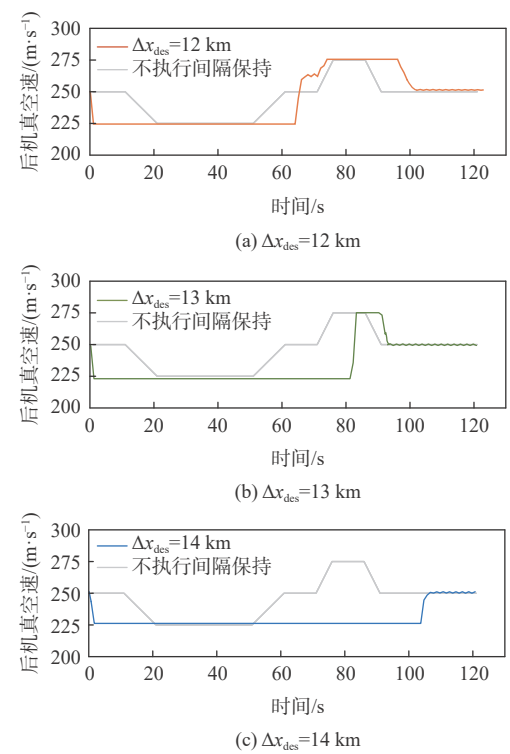
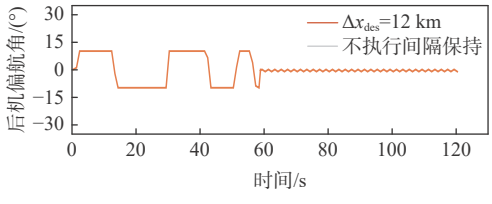


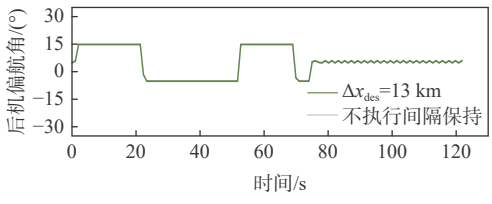
图 9　不同期望纵向间隔下的后机真空速-时间曲线

Fig. 9　Vacuum speed-time curves of the following aircraft under different expectation longitudinal separations

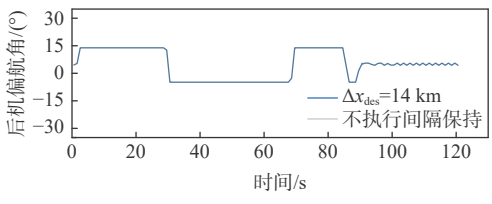
1 538.11m(第 23 s)，2 127.46 m(第 31 s)。后机偏离航路的距离不会超过规定航路左右两侧 10 km，且



(a) Δx_{des}=12 km



(b) Δx_{des}=13 km



(c) Δx_{des}=14 km

图 10 不同期望纵向间隔下的后机偏航角-时间曲线

Fig. 10 Yaw angle-time curves of the following aircraft under different expectation longitudinal separations

表 4 纵向间隔保持仿真结果

Table 4 Simulation results of longitudinal separation maintenance

期望纵向间隔/km	建立纵向间隔时间/s	开始回归航路时间/s	最大航路偏差距离/km	达到最大航路偏差距离时间/s
12	74	58	0.880 46	14
13	90	74	1.538 11	23
14	118	95	2.127 46	31

最终能回到航路中心线。期望纵向间隔越大，最大航路偏差距离也会越大。

3.2.2 不同随机风场条件下的方法验证

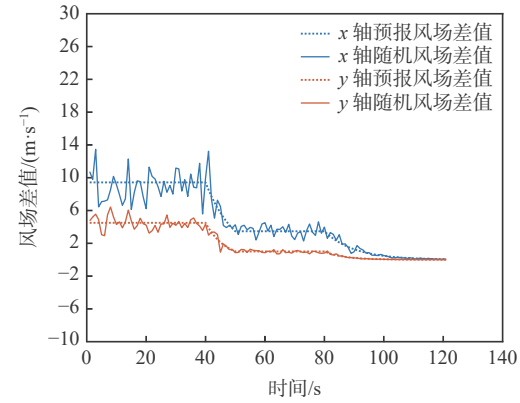
为验证在不同的随机风场条件下执行 MPC 控制器依然能够使两机迅速且平稳地保持纵向间隔，设置了 3 组不同的随机风场模型对照组，参数如表 5 所示，第 1 组和第 2 组为风场强度对照组，第 1 组和第 3 组为扰动强度对照组。

图 11～图 13 分别为对照组的两方向风场差值-时间图、不执行间隔保持情况下的纵向间隔与最小间隔差值/航路偏差距离-时间图和两机纵向间隔-时间图。

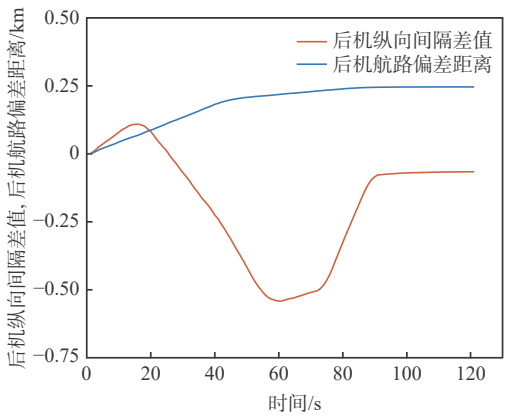
表 5 随机风场对照组参数

Table 5 Parameters of random wind field control groups

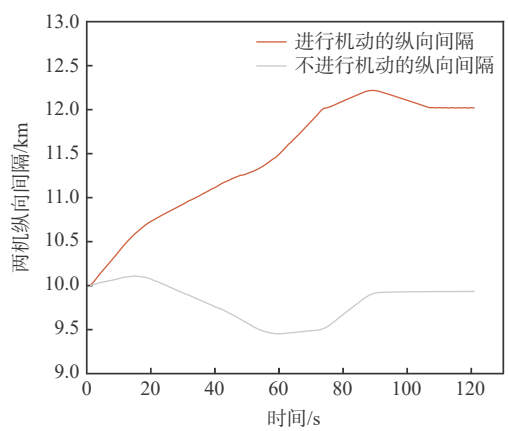
对照组序号	平均风预报风数值/(m·s^{-1})		紊流风相对误差水平	
	前机	后机	x轴	y轴
1	40	30	0.1	0.08
2	80	60	0.1	0.08
3	40	30	0.8	0.64



(a) 两方向的风场差值-时间曲线



(b) 不执行间隔保持情况下的后机纵向间隔差值和航路偏差距离-时间曲线

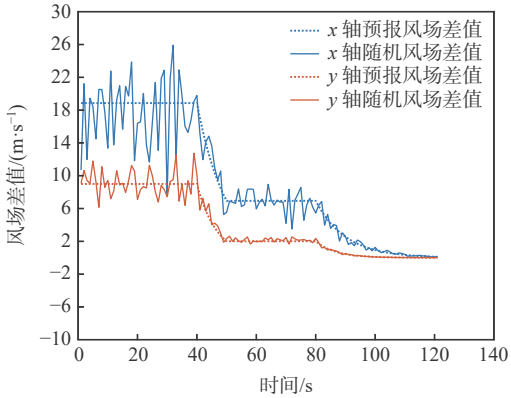


(c) 两机纵向间隔-时间曲线

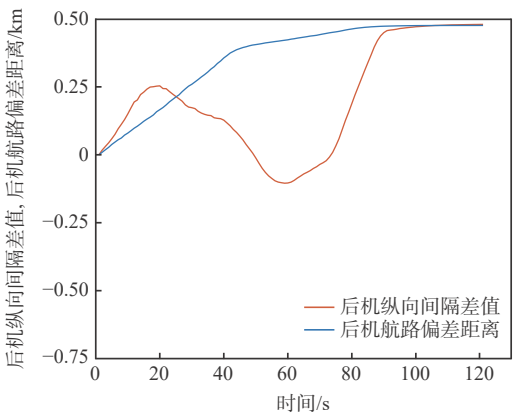
图 11 随机风场模型对照组 1

Fig. 11 Random wind field model control group 1

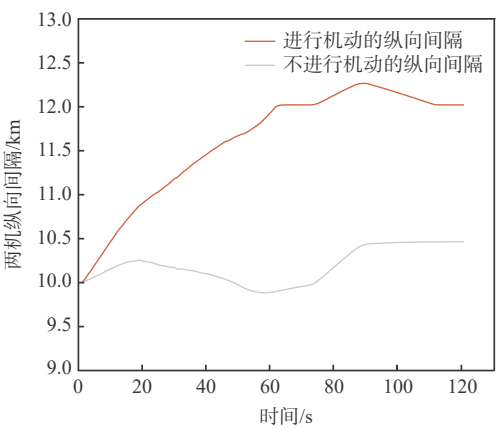
当不执行间隔保持策略时，从图 11(b)和图 12(b)可以看出，如果两机受到的预报风强度加倍，紊流风的扰动强度不变时，在 0～30 s，50～80 s 之间，纵向间隔差值曲线的弯曲幅度增大，说明两机的纵向间隔与最小安全间隔的差值变化更加迅速；但从图 11(c)和图 12(c)可以看出，当执行 MPC 控制器，在预报风强度加倍的情况下，依然



(a) 两方向的风场差值-时间曲线



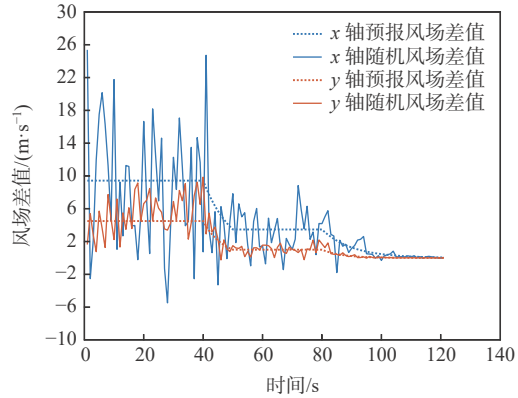
(b) 不执行间隔保持情况下的后机纵向间隔差值和航路偏差距离-时间曲线



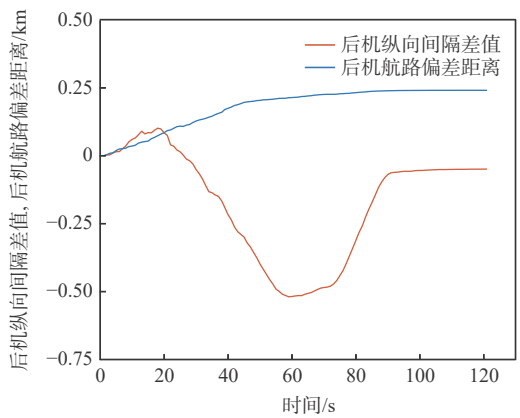
(c) 两机纵向间隔-时间曲线

图 12　随机风场模型对照组 2

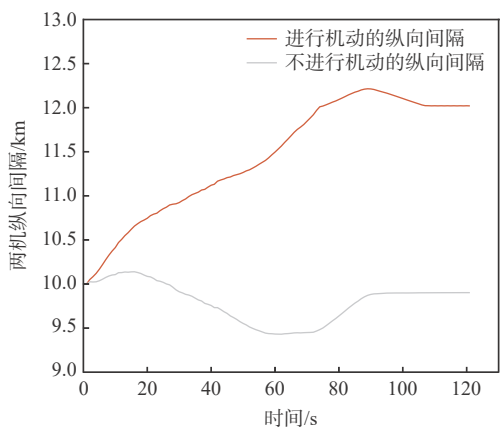
Fig. 12　Random wind field model control group 2



(a) 方向风场差值-时间曲线



(b) 不执行间隔保持情况下的后机纵向间隔差值和航路偏差距离-时间曲线



(c) 两机纵向间隔-时间曲线

图 13　随机风场模型对照组 3

Fig. 13　Random wind field model control group 3

能够在规定的仿真时间内建立期望的间隔(第 61 s), 且两机纵向间隔曲线持续平滑, 始终不低于最小安全间隔 10 km。

当不执行间隔保持策略时, 从图 11(b)和图 13(b)可以看出, 虽然两机纵向间隔差值曲线变化趋势相同, 但在紊流风的扰动强度增大 8 倍的情况下, 0～45 s 曲线抖动非常严重, 但从图 11(c) 和图 13(c) 可以看出, 当执行 MPC 控制器, 即使在紊流风的扰动强度增大 8 倍的情况下, 相比对照组 1, 略有抖动, 依然较为平滑稳定地在第 72 s 建立期望纵向间隔 12 km, 始终不低于最小安全间隔 10 km。

4　结　论

针对高空预报风这一随机因素, 导致两机纵向间隔保持的鲁棒性较差, 本文提出基于 MPC 的航空器纵向自主间隔控制方法。

1) 设置了 3 组不同的期望纵向间隔 12,13, 14 km 下, 在规定的 120 s 仿真时间内, 两机的纵向间隔曲线较为平滑, 且始终不低于规定的最小安全间隔 10 km, 分别在第 74, 90, 118 s 开始建立并最终

稳定在期望纵向间隔；后机偏离航路的距离始终不超过航路宽度 10 km，分别在第 58，74，95 s 开始回归航路且最终均能回到航路中心线。

2）设置了 3 组随机风场模型对照组，在规定的 120 s 仿真时间内，在预报风强度增大 2 倍的情况下，未执行机动的纵向间隔曲线的弯曲幅度增大，当执行 MPC 控制器，依然能够在第 61 s 平滑迅速地建立期望的纵向间隔 12 km；在紊流风的扰动强度增大 8 倍的情况下，未执行机动的纵向间隔曲线抖动非常严重，当执行 MPC 控制器，依然第 72 s 平滑稳定地建立期望的纵向间隔 12 km。

本文研究还存在以下不足：从航空器的飞行阶段来说，只着眼于航路飞行阶段，未来可将航空器纵向间隔建模场景拓展到终端区进近着陆阶段；从航空器的驾驶模式来说，只适用于具有先进机载设备 ASAS 的、能够自主进行间隔保持的航空器，未来可以进一步讨论驾驶员介入操纵航空器时的间隔控制模型。

参考文献（References）

[1] BALLIN M, WING D, HUGHES M, et al. Airborne separation assurance and traffic management-research of concepts and technology[C]//Proceedings of the Guidance, Navigation, and Control Conference and Exhibit. Reston: AIAA, 1999.

[2] 章学锋, 李洪伟, 冯涛. 机载间隔保持系统建模及仿真[J]. 中国民航飞行学院学报, 2020, 31(1): 51-55.
ZHANG X F, LI H W, FENG T. Model and simulation of airborne separation assurance system[J]. Journal of Civil Aviation Flight University of China, 2020, 31(1): 51-55(in Chinese).

[3] 中国民用航空局空中交通管理局. 中国民航现代化空中交通管理系统体系架构: IB-TM-2016-003[S]. 北京: 中国民用航空局空中交通管理局, 2016.
Civil Aviation Administration of China Air Traffic Management Bureau. Architecture of CAAC modern air traffic management system: IB-TM-2016-003[S]. Beijing: Civil Aviation Administration of China Air Traffic Management Bureau, 2016(in Chinese).

[4] 王莉莉, 朱博, 位放. 近距平行跑道配对进近微观跟驰模型研究[J]. 安全与环境学报, 2017, 17(3): 985-988.
WANG L L, ZHU B, WEI F. Microscopic tracing model for the paired approach to the narrow-spaced parallel runways[J]. Journal of Safety and Environment, 2017, 17(3): 985-988(in Chinese).

[5] 王莉莉, 王坤. 飞机流宏观与微观同高度纵向间隔研究[J]. 安全与环境学报, 2016, 16(5): 78-82.
WANG L L, WANG K. Study of the longitudinal interval for the airplanes to keep away at the same height both from the macro- and micro point of view[J]. Journal of Safety and Environment, 2016, 16(5): 78-82(in Chinese).

[6] 王超, 朱明. 空中交通流微观尾随时距分布模型[J]. 计算机仿真, 2018, 35(5): 55-59.
WANG C, ZHU M. Microscopic aircraft-following headway distribution model of air traffic flow[J]. Computer Simulation, 2018, 35(5): 55-59(in Chinese).

[7] HUA M Z, ZHANG M, TANG X M, et al. Structural modelling and deceleration algorithm for a follow aircraft on performance-based navigation airway based on multi-agent technique[J]. Cybernetics and Information Technologies, 2015, 15(6): 46-56.

[8] TAKEICHI N, NAKAMURA Y, FUKUOKA K. Fundamental characteristics of decentralized air traffic flow control in high density corridor[C]//Proceedings of the 28th Congress of the International Council of the Aeronautical Sciences. [S. l.]: ICAS, 2012.

[9] NAKAMURA Y, TAKEICHI N, KAGEYAMA K. A self-separation algorithm using relative speed for a high-density air corridor[J]. Transactions of the Japan Society for Aeronautical and Space Sciences, 2014, 57(6): 336-342.

[10] TAKEICHI N, NAKAMURA Y, KAGEYAMA K. Aircraft self-separation algorithm for high density air corridor operation based on flight intent[J]. Transactions of the Japan Society for Aeronautical and Space Sciences, 2014, 57(3): 179-185.

[11] FUKUOKA K, TAKEICHI N, NAKAMURA Y. A self-separation algorithm in a high density air corridor feasible for a human pilot control[J]. Journal of the Japan Society for Aeronautical and Space Sciences, 2014, 62(3): 107-115.

[12] 汤新民, 郑鹏程. 航路序贯飞行条件下的航空器自主间隔控制[J]. 南京航空航天大学学报, 2019, 51(6): 742-748.
TANG X M, ZHENG P C. Aircraft autonomous separation control under sequential flying conditions[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(6): 742-748(in Chinese).

[13] TIAN Y, DONG Y L, YE B J, et al. A framework for the assessment of distributed self-separation procedures for air traffic in flow corridors[J]. IEEE Access, 2019, 7: 123544-123557.

[14] 王瑜嘉, 王永国, 鲁鹏. 翼身融合飞机在风场中的建模与 Simulink 仿真[J]. 科技风, 2021(29): 4-6.
WANG Y J, WANG Y G, LU P. Modeling and Simulink simulation of blended wing body aircraft in wind field[J]. Science and Technology Wind, 2021(29): 4-6(in Chinese).

[15] 龚建伟, 姜岩, 徐威. 无人驾驶车辆模型预测控制[M]. 北京: 北京理工大学出版社, 2014: 49-50.
GONG J W, JIANG Y, XU W. Model predictive control for unmanned vehicles[M]. Beijing: Beijing Institute of Technology Press, 2014: 49-50(in Chinese).

[16] 陈虹. 模型预测控制[M]. 北京: 科学出版社, 2013: 1-3.
CHEN H. Model predictive control[M]. Beijing: Science Press, 2013: 1-3(in Chinese).

[17] 万兵, 苏析超, 汪节, 等. 基于模型预测控制算法的精确着舰控制方法[J]. 北京航空航天大学学报, 2024, 50(4): 1197-1207.
WAN B, SU X C, WANG J, et al. Research on accurate landing control based on model predictive control algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(4): 1197-1207(in Chinese).

[18] 徐哲, 胡趁义, 龙永文, 等. 车辆纵向跟车舒适性模型预测控制算法研究[J]. 重庆理工大学学报 (自然科学), 2022, 36(12): 9-17.
XU Z, HU C Y, LONG Y W, et al. Research on model predictive control algorithm for longitudinal vehicle following comfort[J]. Journal of Chongqing University of Technology (Natural Science),

2022, 36(12): 9-17(in Chinese).

[19] 中国民用航空局. 运输类飞机适航标准: CCAR-25-R4[S]. 北京: 中国民用航空局, 2016.
Civil Aviation Administration of China. Airworthiness standards for transport aircraft: CCAR-25-R4[S]. Beijing: Civil Aviation Administration of China, 2016(in Chinese).

Longitudinal autonomous separation control of aircraft in random wind fields based on MPC

TANG Xinmin[1, 2, *], LU Xiaona[1]

(1. College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
2. College of Transportation Science and Engineering, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In view of the fact that the high altitude wind, a random factor, often leads to poor robustness of the longitudinal separation between two aircraft in the longitudinal autonomous separation maintenance problem under the random disturbance of track, this paper proposed a longitudinal autonomous separation control method based on model predictive control (MPC). Firstly, the linear time-varying prediction model was developed by establishing the nonlinear kinematics differential equation of the longitudinal separation and the wind field difference between the two aircraft. The longitudinal separation and route deviation distance of the two aircraft was selected as the optimization objectives, the vacuum speed and yaw angle of the front aircraft were taken as the measurable disturbances, and the high-altitude wind was the random disturbance. Terminal equality constraints were added to the air safety and aircraft performance constraints to maintain the stability of the system. To verify the effectiveness of the proposed method, within the specified 120-second simulation time, this article set three sets of different expected separations of 12 km, 13 km, and 14 km. Through the design of an MPC controller, the vacuum speed and yaw angle of the following aircraft were controlled during the rolling time domain cycle. The separation curve between the two aircraft is relatively smooth and always not less than the minimum safety separation of 10 km. It stabilized at the expected target separation in the 74th second, 90th second, and 118th second, and returned to the route starting from the 58th second, 74th second, and 95th second. Two sets of wind field control groups were set up. Two times as much wind was forecast in one group, while eight times as much turbulent wind was disturbed in the other. Both groups were able to establish the expected interval of 12 km smoothly and stably in the 61th second and 72th second, respectively.

Keywords: air traffic management; flight following model; autonomous separation control; model predictive control; random disturbance

Received: 2023-06-28; **Accepted**: 2023-10-08; **Published Online**: 2023-10-24 10:23
URL: link.cnki.net/urlid/11.2625.V.20231023.1041.003
Foundation items: National Natural Science Foundation of China (61773202,52072174)
* **Corresponding author**. E-mail: tangxinmin@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0419

采用多级渐进式门控卷积网络的歌声分离方法

罗庆予，张天骐*，熊天
(重庆邮电大学 通信与信息工程学院，重庆 400065)

摘　　要：针对目前基于卷积神经网络 (CNN) 的歌声分离方法对高低层特征融合时存在语义差异，以及忽视语音特征在通道维度上潜在价值的问题，提出了一种堆叠式的多级渐进式门控卷积网络来实现歌声分离。在每级子网络中设计一种门控自适应卷积 (GAC) 单元来充分学习并提取歌曲的时频特征，增强特征通道间的竞争合作关系；为减少浅层与深层网络信息融合时的语义误差，在子网络的编解码层间引入门控注意力机制；在各级子网络间提出一种监督注意力 (SA) 模块来选择性地传递有效信息流，并实现多级网络的渐进式学习。在公开的2个数据集上进行综合对比实验，结果表明：所提方法相比于近年来的代表性模型，在分离歌声与伴奏时均具有一定的优越性。

关 键 词：歌声分离；神经网络；多级渐进式；自适应；监督注意力
中图分类号：TN912.3
文献标志码：A　　**文章编号**：1001-5965(2025)09-2872-10

歌声分离 (singing voice separation, SVS) 是音乐信息检索的基础研究之一，也是音乐源分离 (music source separation, MSS) 的一种特殊形式，其旨在将音乐中的歌声与伴奏进行有效分离，分离出的信号源可进一步用于其他任务，如歌曲检索、歌词识别、歌手识别和乐器识别等。

早期的歌声分离方法主要是无监督方法，如重复模式提取技术 (repeating pattern extraction technique, REPET)[1]、鲁棒主成分分析 (robust principal component analysis, RPCA)[2] 和非负矩阵分解 (non-negative matrix factorization, NMF)[3] 等。这些单一的传统方法往往存在不能有效将歌声和伴奏彻底分离开、分离的歌声或伴奏中含有较多残留的伴奏或歌声的问题。

随着近年来深度学习的再次兴起，其在语音应用领域得到了广泛应用。Uhlich 等[4] 通过使用深度神经网络 (deep neural network, DNN)，成功从多种乐器混合演奏的声源中分离出目标源，但 DNN 对于音频任务存在一种缺陷，即不能充分利用音频的时序信息。因此，Sprechmann 等[5] 提出将循环神经网络 (recurrent neural network, RNN) 应用于音频任务，在单通道语音分离任务中加入区分性训练来考虑各个声源间的差异，从而进一步提高模型的分离性能。随后，文献 [6] 将区分性训练的 RNN 成功应用到歌声分离任务中，考虑了分离前后歌声与伴奏之间的差异及相关性，取得了不错的分离效果。随着 RNN 出现无法处理序列长期依赖性及训练时易发生梯度爆炸和消失的现象，文献 [7] 提出将长短时记忆 (long short-term memory, LSTM) 网络模型应用到单通道语音分离任务中，取得了比 RNN 更优的效果。除此之外，文献 [8] 还将 LSTM 的变体，即双向长短时记忆 (bidirectional long short-term memory, BiLSTM) 网络应用于歌声分离任务，进一步取得了不错的分离效果。

随着卷积神经网络 (convolutional neural network, CNN) 在图像领域中盛行，学者们也开始将其应用

收稿日期：2023-06-28；**录用日期**：2023-11-03；**网络出版时间**：2023-11-24 11:02
网络出版地址：link.cnki.net/urlid/11.2625.V.20231123.1413.004
基金项目：国家自然科学基金 (61771085)；重庆市自然科学基金 (cstc2021jcyj-msxmX0836)；重庆市教育委员会科研项目 (KJ1600427, KJ1600429)
* **通信作者**. E-mail：zhangtq@cqupt.edu.cn

引用格式：*罗庆予，张天骐，熊天. 采用多级渐进式门控卷积网络的歌声分离方法 [J]. 北京航空航天大学学报，2025，51（9）：2872-2881. LUO Q Y，ZHANG T Q，XIONG T. Singing voice separation method using multi-stage progressive gated convolutional network [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（9）：2872-2881（in Chinese）.*

到音频处理领域。由于音频信号的幅度谱图可以看作二维的单通道图像, 可将音频信号转换成图像输入到 CNN 中处理, 并且 CNN 在参数量与计算量方面相比于 DNN、RNN 及 LSTM 都占优势, 目前绝大多数歌声分离任务都是基于 CNN 设计开发的。Jansson 等[9] 将医学图像分割任务中的编解码式网络结构 U-Net 引入歌声分离任务, 基于信号谱图进行音乐分离; 而文献 [10-12] 在 U-Net 的基础上提出了一系列端到端的音乐源分离方法, 解决了未充分利用音乐信号相位的问题, 得到了更优的歌声分离性能。近年, 文献 [13] 将自然图像中的人体姿态进行评估, 设计了堆叠沙漏网络 (stacked hourglass network, SHN) 用于音乐源分离任务, 其以堆叠式的编解码结构逐步实现目标源分离, 取得了与当时最先进方法相当的分离效果。此后, 文献 [14-16] 在基线 SHN 基础上进行研究并衍生, 对模型性能实现了一定的改进。但也存在一些问题: ①此类卷积模型都是直接将编码层底层低级特征与解码层高级特征进行融合, 并未考虑语义差距; ②随着网络层数的增加, 特征通道数也随之增加, 浅层网络或深层网络中歌声或伴奏特征通道之间的关系被逐渐忽略。这都会导致训练阶段误差变大, 影响歌声分离方法的最终性能。

针对上述问题, 本文为获取更全面的歌声或伴奏特征来提升网络分离性能, 以 SHN 为基线模型, 设计了一种多级渐进式门控卷积网络用于歌声分离。首先, 提出一种门控自适应卷积 (gated adaptive convolution, GAC) 模块作为基础单元来传输特征信息, 并同时强调深层网络中通道间的竞争合作关系, 以获取更全面的语音上下文信息。然后, 将 GAC 作为基础单元, 搭建一种编码器-解码器式结构的门控卷积子网络, 并且在编解码层间引入门控注意力 (gating attention, GA) 机制来实现浅层特征与深层特征的有效融合。最后, 提出一种监督注意力 (supervised attention, SA) 模块来连接每一级子网络, 以实现网络渐进学习的目的。在权威 MIR-1K 数据集上进行多种对比与消融实验, 证明本文方法在各项分离指标上取得了较为显著的提升, 并具有良好的歌声分离性能。

1　方法描述

本文提出的歌声分离方法利用神经网络模型, 同时对混合音乐信号中的歌声与伴奏进行建模, 具体流程如图 1 所示。通过对时域混合音乐信号经过短时傅里叶变换 (short-time Fourier transformer, STFT) 后得到时频谱图, 即幅度谱, 由于人耳对相位的感知并不敏感, 取幅度谱作为该网络模型的输入。先利用一个初始卷积层提取初始特征并增加通道数, 再输入到后续子网络中估计掩蔽值。

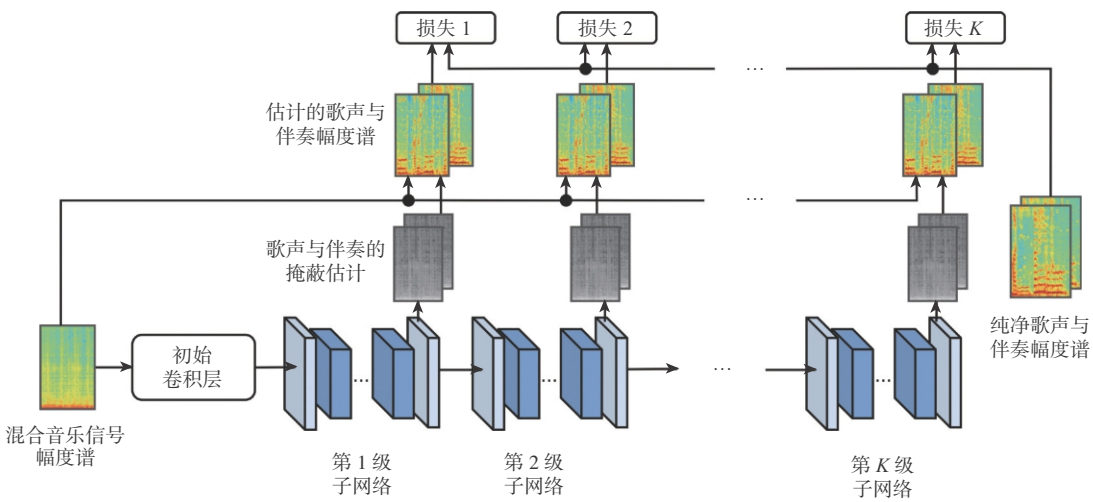


图 1　多级渐进式门控卷积网络

Fig. 1　Multi-stage progressive gated convolutional network

在网络训练阶段, 模型中的每一级子网络均会分别输出估计歌声与估计伴奏的掩蔽, 将掩蔽分别与混合音乐信号幅度谱作点乘运算得到估计的歌声与伴奏的幅度谱, 以每级子网络的估计值和对应的真实值之间的总误差最小化作为目标来训练该模型; 在网络测试阶段, 采用最后一级子网络预测出歌声与伴奏的幅度谱, 分别与输入音乐的相位谱相结合, 再经过逆短时傅里叶变换 (inverse short-time Fourier transformer, ISTFT), 最终得到网络分离出预测的歌声与伴奏的时域波形。

2　模型结构

2.1　门控注意力机制

由于传统 U-Net 网络中实现编码层与解码层的

融合直接使用卷积映射的残差连接的方式，并没有考虑编码层中的底层特征与解码层中的高层特征间的语义差距，因此，本文在残差卷积映射之后将注意力机制[17]与门控机制相结合，提出一种 GA 机制来缓解该问题。

语音的频谱中含有丰富的频率成分，其中，共振峰在低频区域占主导，在高频处分布稀疏，因此，引入 GA 机制来对不同的区域分配不同的权重。该模块能够自动学习聚焦不同形状和大小的目标结构，并且抑制输入特征图中不相关的背景区域，同时突出对特定任务有用的显著特征。如图 2 所示，GA 由 2 部分组成：①求门控注意力权重；②每层编码层的输出特征与门控注意力权重逐点相乘，求得最终的门控注意力输出。公式如下：

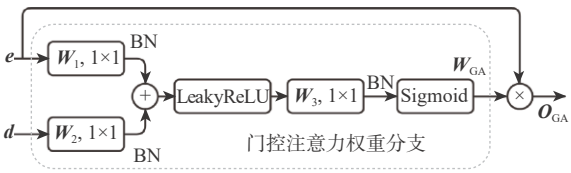


图 2 门控注意力机制

Fig. 2 Gated attention mechanism

$$\boldsymbol{W}_{\mathrm{GA}}=\sigma_1[\boldsymbol{W}_3*\sigma_2(\boldsymbol{W}_1*\boldsymbol{e}\oplus\boldsymbol{W}_2*\boldsymbol{d})] \tag{1}$$

$$\boldsymbol{O}_{\mathrm{GA}}=\boldsymbol{e}\otimes\boldsymbol{W}_{\mathrm{GA}} \tag{2}$$

式中：$\boldsymbol{e}$为每层编码层输出特征；$\boldsymbol{d}$为每层解码层的输出特征；$\boldsymbol{W}_1$、$\boldsymbol{W}_2$和$\boldsymbol{W}_3$为卷积核大小为 1×1 的卷积；$\boldsymbol{W}_{\mathrm{GA}}$为门控注意力权重；“$*$”表示卷操作，“$\oplus$”和“$\otimes$”分别表示逐点相加和相乘；$\sigma_1$和$\sigma_2$分别为 leakyReLU 和 Sigmoid 激活函数表示；$\boldsymbol{O}_{\mathrm{GA}}$为 GA 模块的输出。

GA 模块能有效解决编码器-解码器架构引起的信息丢失问题，并稳定网络优化过程。

2.2 门控自适应卷积单元

对于 CNN 来说，绝大部分特征信息都包含在信道中。随着网络层数的加深，网络会倾向于收集空间视图信息而忽视语音在通道维度上的潜在价值，故本文利用视觉任务中提出的门控通道转换(gated channel transformation, GCT)[18]的优势，提出 GAC 单元来代替普通卷积层，并作为编解码部分的基础模块，以增强对语音多维信息的提取与传递，具体结构如图 3 所示。

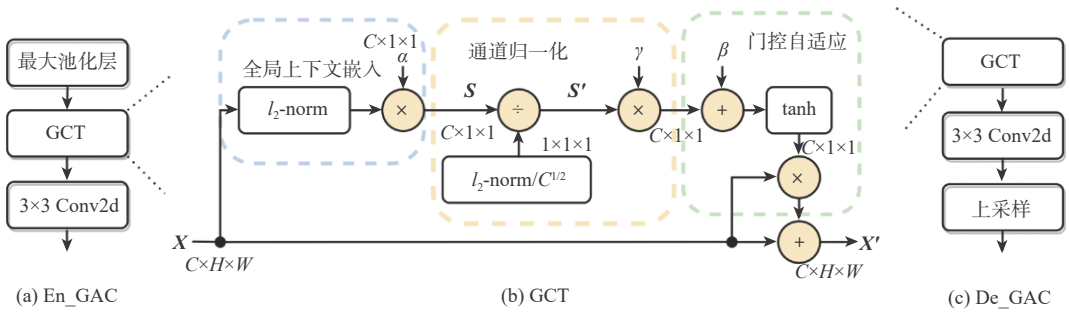


图 3 门控自适应卷积单元

Fig. 3 Gated adaptive convolution unit

将编码器中的 GAC 单元表示为 En_GAC，其输入特征先通过最大池化层，将特征维度减半以控制网络参数量，再通过 GCT 模块以动态地为输入通道分配不同权重，加强特征通道注意力，最终将层次信息丰富的语音特征输入卷积核大小为 3×3 的 2 维卷积 (Conv2d)，以进一步提取上下文信息。解码器中的 GAC 单元表示为 De_GAC，其为 En_GAC 的逆过程，但不同的是，输入特征在解码部分先通过 GCT 模块来突出重要特征信息，再依次输入卷积核大小为 3×3 的 Conv2d 与上采样，对语音特征信息补偿恢复。另外，卷积层后均有 BN 与 Leaky-ReLU 激活函数操作。

本文利用 GCT 模块来模拟不同通道间信息关系，如图 3(b) 所示，其由 3 个部分组成：全局上下文嵌入、通道归一化和门控自适应。设输入特征为 $\boldsymbol{X}\in\mathbf{R}^{C\times H\times W}$，其中，$C$、$H$和$W$分别为输入特征的通道数、高度和宽度。对$\boldsymbol{X}$中的每个通道进行$l_2$范数操作以代替全局平均池化，实现具有大感受野的上下文嵌入。以第c通道为例，利用嵌入权重因子α_c来调节其权重，嵌入后输出$\boldsymbol{s}_c$为

$$[\boldsymbol{s}_c]_{ij}=\alpha_c\|\boldsymbol{x}_c\|_2=\alpha_c\left\{\left[\sum_{i=1}^{H}\sum_{j=1}^{W}\left(x_c^{i,j}\right)^2+\varepsilon\right]\right\}^{\frac{1}{2}} \tag{3}$$

式中：$\boldsymbol{x}_c$和$x_c^{i,j}$分别为第c通道特征向量和每个时频点上的值；$\|\cdot\|_2$表示l_2范数；ε为避免零点求导的一个极小常数。

对所有通道特征向量作l_2范数归一化及跨通道操作后输出$\hat{\boldsymbol{s}}_c$，具体操作可以表示为

$$\hat{\boldsymbol{s}}_c=\frac{\sqrt{C}\boldsymbol{s}_c}{\|\boldsymbol{S}\|_2}=\frac{\sqrt{C}\boldsymbol{s}_c}{\left[\left(\sum_{c=1}^{C}\boldsymbol{s}_c^2\right)+\varepsilon\right]^{\frac{1}{2}}} \tag{4}$$

式中: $\boldsymbol{S}=[\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_C]$, $\boldsymbol{s}_c$ 为第 c 个通道的向量。

$$\hat{\boldsymbol{x}}_c = \boldsymbol{x}_c(1+\tanh(\gamma_c\hat{\boldsymbol{s}}_c+\boldsymbol{\beta}_c)) \quad (5)$$

式中: $\sqrt{C}$用于对$\hat{\boldsymbol{s}}_c$的尺度进行归一化, 避免当C很大时$\hat{\boldsymbol{s}}_c$的尺度过小。

为了网络在训练过程中加强特征通道间的竞争和合作, 结合门控机制, 使用激活函数 tanh, 并引入门控权重因子γ_c与偏执权重因子$\boldsymbol{\beta}_c$调整最终输出$\hat{\boldsymbol{x}}_c$, 如式 (5) 所示。GCT 中, 所有权重因子均可训练, 共同决定了输入特征中每个通道的行为。

GAC 单元代替普通卷积层, 不仅能有效缓解由普通卷积卷积核小而固定引起的局部模糊问题, 在加深网络层数而增加感受野的同时, 还不会带来大量计算资源的浪费。

2.3　监督注意力模块

在堆叠式的多级网络中, 前级子网络很大可能会出现语音信息丢失而迭代给下一级子网络, 从而造成整体网络性能下降, 故本文设计了一种 SA 模块来对子网络输出重新注入原始语音信息, 并对其细化后输入下一级子网络, 以缓解该问题。如图 4 所示, 在每级子网络之间设置 SA 模块, 其提供原始输入, 并生成注意系数矩阵以抑制当前子网络的信息特征, 并只允许有用的特征流传输到下一级子网络进行处理, 以实现模型渐进式学习的目的。

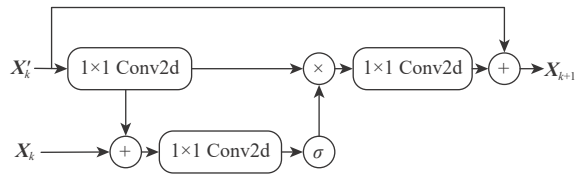


图 4　监督注意力模块

Fig. 4　Supervised attention module

设第k个子网络的输入特征为$\boldsymbol{X}_k$, 输出特征为$\boldsymbol{X}'_k$, 输入特征先被送到一个 1×1 卷积层以生成残差特征图, 再将残差特征图添加到原始输入特征, 由 1×1 卷积层融合信息后用σ表示的 Sigmoid 激活函数生成注意力掩蔽矩阵, 并与残差特征图相乘, 最终用 1×1 卷积层重新校准处理后, 与输出特征相加传递到下一级子网络, 作为第$k+1$个子网络的输入特征$\boldsymbol{X}_{k+1}$。

2.4　门控卷积子网络

由 GA、SA、En_GAC、De_GAC 等模块搭建出的门控卷积子网络如图 5 所示, 其主要由 4 部分组成: 编码层、中间层、解码层和输出层。

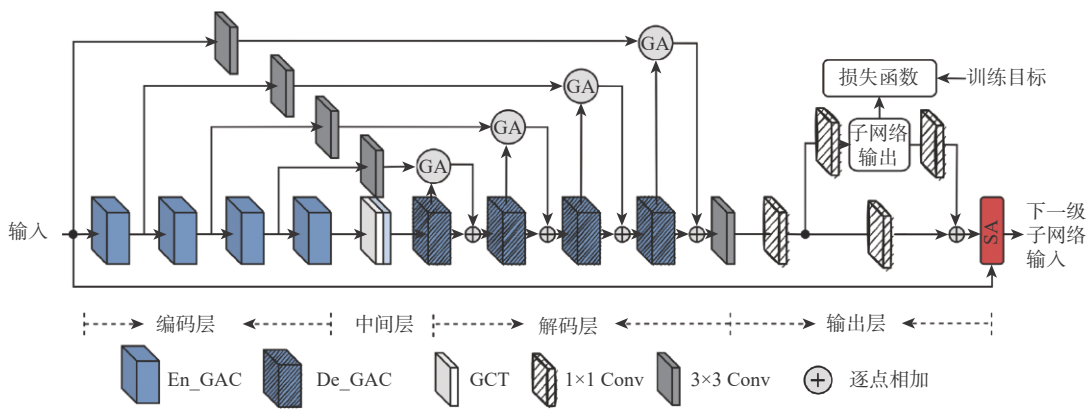


图 5　门控卷积子网络

Fig. 5　Gated convolutional subnetwork

编码层主要由 4 个 GAC 单元 (En_GAC) 组成, 每单元捕获语音特征进行压缩, 并补偿感受野损失, 为传输层提供含有丰富全局上下文信息的低维特征; 中间层则由一个 GCT 与卷积核大小为 3×3 的二维卷积 (Conv2d) 组成, 其作为过渡层来传递特征并保持特征图尺寸大小, 通道数增加; 解码层有 4 个 GAC 单元 (De_GAC) 组成, 将特征图维度逐层恢复, 并且在与编码部分的对应层级间采用跳跃连接, 每个跳跃连接上引入了卷积与 GA 模块, 每个 GA 模块的输出与上一解码单元输出进行逐点相加后输入给下一解码单元, 通过减小低分辨率特征与高分辨率特征间融合时的语义差距, 为解码层提供了更具鲁棒性的特征, 以增强网络恢复效果。

在子网络末端的输出层中, 先输出特征, 通过由卷积核为 3×3 与 1×1 的 2 个卷积层作为网络的输出卷积层, 再将其输出传入上下两分支: 上分支由单个 1×1 卷积来将通道数降低为 2, 并生成网络估计的歌声幅度谱与伴奏幅度谱的掩蔽, 同时网络计算纯净目标值与子网络估计值之间的误差损失; 下分支则通过 1×1 卷积, 再同样通过 1×1 卷积并通道数还原为 64 的上分支输出相加融合, 最终与该子网络输入一起传入 SA 模块, 将重新调整并细化的语音信息输入下一级门控卷积子网络。

在第1级子网络前，需要将幅值特征输入到初始化卷积层进行语音特征的初步提取，其由4层的2维卷积组成，并在保持输入特征维度下将通道维数增加到64。另外，最后一级子网络的输出只有掩蔽输出支路，而没有下一级子网络的输出支路。本文网络中的子网络结构参数如表1所示，其中，只给出了初始卷积层和第1个改进沙漏子网络的参数，堆叠的子网络结构参数设置均相同。表中的输入输出维度按照$H\times W\times C$格式，d_c为卷积核大小，s_c为卷积步长，p为最大池化层的窗大小，ε为GCT中的极小常数值，表中所有二维卷积(Conv2d)后均有BN与LeakyReLU激活函数操作。

表1 子网络结构参数

Table 1 Subnetwork structure parameters

层结构	层参数	层输入维度	层输出维度
初始卷积层	$d_c=7,1,1,\ s_c=1$	512×64×1	512×64×64
En_GAC_1	$\varepsilon=10^{-5},d_c=3,\ s_c=1,\ p=2$	512×64×64	512×32×64
En_GAC_2	$\varepsilon=10^{-5},d_c=3,\ s_c=1,\ p=2$	256×32×64	128×16×128
En_GAC_3	$\varepsilon=10^{-5},d_c=3,\ s_c=1,\ p=2$	128×16×128	64×8×192
En_GAC_4	$\varepsilon=10^{-5},d_c=3,\ s_c=1,\ p=2$	64×8×192	32×4×256
GCT+Conv2d	$d_c=3,\ s_c=1,\ \varepsilon=10^{-5}$	32×4×256	32×4×320
De_GAC_4	$d_c=3,\ s_c=1,\ \varepsilon=10^{-5}$	32×4×320	64×8×256
De_GAC_3	$d_c=3,\ s_c=1,\ \varepsilon=10^{-5}$	64×8×256	128×16×192
De_GAC_2	$d_c=3,\ s_c=1,\ \varepsilon=10^{-5}$	128×16×192	256×32×128
De_GAC_1	$d_c=3,\ s_c=1,\ \varepsilon=10^{-5}$	256×32×128	512×64×64
输出卷积层	$d_c=3,1,\ s_c=1$	512×64×64	512×64×64
Conv2d_1	$d_c=1,\ s_c=1$	512×64×64	512×64×2
Conv2d_2	$d_c=1,\ s_c=1$	512×64×2	512×64×64
Conv2d_3	$d_c=1,\ s_c=1$	512×64×64	512×64×64
SA	$d_c=1,\ s_c=1$	512×64×64	512×64×64

2.5 损失函数

本文网络的损失函数为纯净歌声与伴奏和预测歌声和伴奏之间的最小绝对值误差。假设给定输入特征谱图为$\boldsymbol{Y}$、第i个纯净目标源$\boldsymbol{T}_i$和第j级门控卷积子网络中的第i个源生成的掩蔽$\boldsymbol{M}_{i,j}$，则第j级门控卷积子网络中的第i个源的损失定义为

$$L(i,j)=\left\|\boldsymbol{T}_i-\boldsymbol{Y}\odot\boldsymbol{M}_{i,j}\right\|_1 \tag{6}$$

式中："⊙"表示矩阵元素乘法；$\|\cdot\|_1$表示计算矩阵中元素绝对值的和。

整个网络的损失函数可以表示为

$$L=\sum_{i=1}^{M}\sum_{j=1}^{K}L(i,j) \tag{7}$$

式中：M为要分离源的数量；K为网络中堆叠的子网络个数。

3 实验结果与分析

3.1 数据集及参数设置

为充分验证本文模型对歌声分离的效果，在小、大2个规模的公开权威数据集上分别对本文方法进行训练与测试，数据集具体构成如下：

1）MIR-1K数据集。该数据集在110首中文歌曲中提取1 000首歌曲片段，其中，包含混合音轨和音乐伴奏音轨，且伴奏与歌声分别录制为左右声道，每个片段时长为4～13 s，且所有音频采样率均为16 kHz。本文采用男(Abjones)和女(Amy)所唱歌曲片段作为训练集，共包含175个音乐片段，剩余825个音乐片段用于网络模型评估。

2）MUSDB18数据集。该数据集旨在比较源分离系统在相同数据和指标上的性能，其中，共包含150首不同类型的立体声格式的歌曲，总时长约为11 h，并为每首歌曲提供了人声、贝斯、鼓及其他乐器(伴奏)的独立声道。实验中，将人声以外的其他声源视为伴奏，所有音频采样率为16 kHz。该数据集中的100首歌用于训练，50首歌用于测试。

本文实验均是在RTX3090的GPU上基于CUDA11.2与CUDNN的开发环境下搭建的Pytorch模型。为加快网络处理速度，所有音频信号均下采样到8 kHz，采用窗长约为64 ms、窗移为16 ms的汉明窗进行短时傅里叶变换。由于频谱具有共轭对称性，采取一半频率维度计算，网络输入的特征时间维度为64，频率维度为512。实验中，批次大小为4，训练迭代次数为20 000，并采用Adam来优化网络模型参数，初始学习率为0.001。另外，将GCT中的嵌入权重因子初始化为1，门控权重因子与偏执权重因子初始化为0。所有对比模型都保持其文献中的最佳配置，同时采用本文数据集进行训练与测试。

3.2 实验性能评价指标

结果评估阶段中使用Mir_eval工具箱来评估网络分离歌声与伴奏性能，主要指标包括全局归一化信号偏差比(global normalized signal-to-distortion ratio, GNSDR)、全局信号干扰比(global signal-to-interference ratio, GSIR)和全局源到伪影比(global source-to-artifacts ratio, GSAR)，分别为归一化的信号偏差比(normalized signal-to-distortion ratio, NSDR)、信号干扰比(signal-to-interference ratio, SIR)和源到伪影比(source-to-artifacts ratio, SAR)的加权平均值，且值越高，表示网络歌声与伴奏分离性能越好。其中，归一化信号偏差比O_{nSDR}、信号干扰比O_{SIR}、系统误差比O_{SAR}分别定义为

$$O_{\mathrm{nSDR}}=10\lg\left(\frac{\left\|\boldsymbol{s}_{\mathrm{target}}(t)\right\|^2}{\left\|\boldsymbol{e}_{\mathrm{interf}}(t)+\boldsymbol{e}_{\mathrm{artif}}(t)\right\|^2}\right) \tag{8}$$

$$O_{\mathrm{SIR}}=10\lg\left(\frac{\left\|\boldsymbol{s}_{\mathrm{target}}(t)\right\|^2}{\left\|\boldsymbol{e}_{\mathrm{interf}}(t)\right\|^2}\right) \tag{9}$$

$$O_{\mathrm{SAR}}=10\lg\left(\frac{\left\|\boldsymbol{s}_{\mathrm{target}}(t)+\boldsymbol{e}_{\mathrm{interf}}(t)\right\|^2}{\left\|\boldsymbol{e}_{\mathrm{artif}}(t)\right\|^2}\right)\tag{10}$$

式中: $\boldsymbol{s}_{\mathrm{target}}(t)$为预测信号中属于源信号的部分; $\boldsymbol{e}_{\mathrm{interf}}(t)$为预测信号中不属于源信号的部分; $\boldsymbol{e}_{\mathrm{artif}}(t)$为网络本身产生的系统噪声误差。

主观评价指标的平均意见得分 (mean opinion score, MOS) 是一种使用主观听感打分的形式来评价语音质量的标准。MOS 采取满分 5 分制的形式, 其中, 5 分表示听感优质, 4 分表示听感良好, 3 分表示听感尚可, 2 分表示听感较差, 1 分表示听感劣质。MOS 为被选听众对每一段分离出的歌声与伴奏所给出分数的算术平均值, 其表达式如下:

$$O_{\mathrm{MOS}}=\frac{\sum_{n=1}^{N}R_n}{N}\tag{11}$$

式中: N为选取听众的人数; R_n为每段歌声和伴奏的 MOS 得分。

3.3　实验结果及对比分析

3.3.1　消融实验

本文在 MIR-1K 数据集上分别探讨网络在不同子网络个数、不同模块下的消融实验, 以优化网络并证明所提模块的有效性。表 2 比较了在不同子网络个数k下总模型对歌声分离性能的影响。随着k的依次增加, 网络的歌声分离性能也随之提升。当$k=4$时, 网络总模型的歌声分离性能各指标几乎均有明显提高, 而当$k=5,6$时, 评价指标稍有下降。出于对网络参数量的考虑, 本文将子网络个数$k=4$设置为模型固定参数。

表 2　子网络个数对网络分离性能的影响

Table 2　Effect of number of subnetworks on network separation performance　　dB

子网络数量	GNSDR		GSIR		GSAR	
	歌声	伴奏	歌声	伴奏	歌声	伴奏
1	12.02	10.63	18.41	14.76	13.55	12.83
2	12.14	10.67	18.62	14.78	13.61	12.96
3	12.21	10.68	18.76	14.81	13.68	13.17
4	12.24	10.73	18.85	14.88	13.68	13.31
5	12.22	10.70	18.86	14.86	13.60	13.30
6	12.13	10.65	18.71	14.72	13.50	13.11

表 3 探究了不同模块对网络分离性能的影响, 其中, SHN 为本文的基线模型, GAC-SHN、GA-SHN、GAC-GA-SHN、GAC-GA-SA-SHN 分别表示在 SHN 的基础上将编解码层中基础单元换为 GAC、在跳跃连接上嵌入 GA 机制、依次加入 GAC 与 GA、在此基础上每级子网络间加入 SA 模块(即本文方法)。可以看出, 本文方法与基线相比, 整体提升了 1.35 dB, 伴奏分离性能有微小提升, 整体提升了 0.81 dB。这说明相比于固定尺寸大小的卷积滤波器, 将 GAC 替换普通卷积层更能捕获到歌声的全局信息特征, 以提高网络歌声与伴奏的分离性能。通过比较 GA-SHN 与基线可以看出, 歌声分离性能平均提升了 0.75 dB, 伴奏分离性能平均提升了 0.29 dB, 这说明将 GA 机制的引入能够一定程度上缩小底层特征与深层特征之间的语义差距, 进一步提升分离性能。可以看出, 在 GAC-GA-SHN 基础上, 在每级间引入 SA 模块后的本文方法在分离效果上达到了最佳。

表 3　不同模块对网络分离性能的影响

Table 3　Effects of different modules on network separation performance　　dB

网络模型	GNSDR		GSIR		GSAR	
	歌声	伴奏	歌声	伴奏	歌声	伴奏
SHN[13]	10.51	9.88	16.01	14.24	12.53	12.36
GAC-SHN	11.98	10.65	18.35	14.73	13.45	12.78
GA-SHN	10.89	10.32	17.64	14.36	12.76	12.66
GAC-GA-SHN	12.01	10.69	18.25	14.78	13.66	12.91
GAC-GA-SA-SHN(本文)	12.24	10.73	18.85	14.88	13.64	13.31

3.3.2　对比实验

表 4 横向对比了在基线 SHN 模型上衍生的一系列网络模型与多种经典网络模型的歌声分离性能[9,14-16,19-21]。可以看出, 除了伴奏的 GSIR 指标略低, 总体上来看, 本文方法的整体分离性能要优于所有对比方法, 尤其是歌声分离指标, 超过了之前表现最好的 SA-SHN, 平均增幅有 0.69 dB, 伴奏分离性能指标中的 GNSDR 值和 GSAR 值也高于 SA-SHN, 平均增幅有 0.14 dB。与此同时, 本文方法相比将 SHN 作为基线的几种方法的歌声分离性能指标都要高, 其中, 包括文献 [15-18] 中的较先进方法, 这也说明本文方法在 MIR-1K 数据集上达到了较为先进的歌声分离性能。通过与表 4 中其他非基于 SHN 的方法做比较, 本文方法的歌声分离性能指标得到了全面且较大幅度的提升。

除了上述客观评价指标外, 在 MIR-1K 数据集中随机选取 5 首音乐片段, 本文方法与基线网络分离得到歌声与伴奏 MOS 得分, 如表 5 所示。选取了 50 个听众, 将其中 25 个男生与 25 个女生分为 2 组来分别对其打分。可以看出, 本文方法分离出的歌声得分均优于基线网络, 更能有效提升分离歌声的听感, 而伴奏上的得分两者几乎相近, 这是因为伴奏中含有各种乐器音源, 微小的变化可能导致人耳难以察觉。

表 4 MIR-1K 数据集上不同方法的分离性能比较[9,14-16,19-21]

Table 4 Comparison of separation performance of different methods on MIR-1K dataset[9,14-16,19-21] dB

模型	GNSDR		GSIR		GSAR	
	歌声	伴奏	歌声	伴奏	歌声	伴奏
MLRR[19]	3.85	4.19	5.63	7.80	10.70	8.22
DRNN[20]	7.45		13.08		9.68	
ModGD[21]	7.50		13.73		9.45	
U-Net[9]	7.43	7.45	11.79	11.43	10.42	10.38
U-Net-SE[16]	7.49	7.46	11.78	11.42	10.38	10.41
FC-Net[16]	10.85	9.91	16.95	14.09	12.54	12.64
TSMS-G-CS[16]	10.93	9.96	17.02	14.13	12.59	12.68
SHN[15]	10.51	9.88	16.01	14.24	12.53	12.36
SE-SHN[16]	10.48	9.86	15.99	14.21	12.52	12.35
SA-SHN[14]	11.66	10.60	17.65	14.92	13.38	13.17
P-SHN[15]	10.83	9.89	16.54	14.01	12.67	12.65
本文	12.24	10.73	18.85	14.88	13.68	13.31

表 6 对比了近年多种先进网络模型的歌声分离性能[22-26]，其中，文献 [22-24] 的模型是将 CNN 分别与 LSTM、不同大小的池化操作、密集连接相结合。可以看出，本文方法在大部分分离指标上优于或接近对比方法，其中，歌声分离性能指标 SDR 值平均增幅 1.36 dB。除了伴奏分离性能指标中的 SIR，本文方法在各项指标上值也高于将 RPCA 与递归 CNN 组合的 RPCA-DRNN 模型。另外，从表 6 中可以看出，本文方法与采用 2 层嵌套的 U 型结构的 FC-U^2-Net 模型相比，各指标总体上取得了相近的数值，其中，歌声分离指标 NSDR、SAR 值更是分别提升了 0.15 dB、0.12 dB，伴奏分离性能指标 SAR 提高 1.24 dB。总体上来看，本文所采用的复杂堆叠式门控卷积网络结构所带来的整体分离性能优于所有对比方法。

图 6 用语谱图的形式直观比较了本文方法与基线 SHN 网络分离出歌声与伴奏之间的细微差别。随机选取了 MIR-1K 数据集中一首音乐片段 (annar_4_09) 为例，可以看出，本文方法分离出歌声语谱图在 0～1.5 s 的静音段中如图中黑框所示，将伴奏部分误分离到歌声中的部分明显减少，在 2～5 s 中的 1～2.5 kHz 处能够恢复基线网络难以恢复的歌声细节特征，并且在有声段 2.5～3 kHz 有更加清晰的语音共振峰。对于伴奏而言，本文方法与基线网络相差不大。但从整体上看，本文方法分离出的伴奏在 0～2 kHz 频段的能量与纯净伴奏更接近，比基线模型分离出的伴奏有更高的能量。

图 7 分析了不同子网络个数 k 下本文与基线网络 SHN 间的参数量对比。可以看出，虽然本文方法参数量相比于基线网络在不同子网络个数下增加了 1×10^6～2×10^6，但相对于自身参数量平均仅增加了约 4%。但歌声与伴奏分离性能指标却比基线网络分别平均提升了 1.91 dB、0.81 dB。故综上所述，本文在相近的网络参数下，能对分离效果进行显著的提升，证明了该方法的有效性。

表 5 MIR-1K 数据集中 5 首音乐片段下的歌声与伴奏的 MOS 得分

Table 5 MOS scores of singing voice and accompaniment for 5 music segments in MIR-1K dataset

歌曲名称	SHN						本文					
	歌声			伴奏			歌声			伴奏		
	男	女	Avg	男	女	Avg	男	女	Avg	男	女	Avg
annar_5_03	3.32	3.31	3.31	3.43	3.42	3.42	3.51	3.53	3.52	3.45	3.42	3.43
Bobon_2_06	3.25	3.28	3.26	3.39	3.4	3.39	3.42	3.43	3.42	3.4	3.41	3.40
Kenshin_3_08	3.26	3.23	3.24	3.46	3.42	3.44	3.36	3.34	3.35	3.43	3.47	3.45
Khair_6_06	3.36	3.34	3.35	3.32	3.36	3.33	3.48	3.5	3.49	3.32	3.33	3.32
Yifen_1_07	3.38	3.40	3.39	3.43	3.45	3.44	3.56	3.58	3.57	3.42	3.45	3.43

表 6 MUSDB18 数据集上不同先进方法的分离性能比较[22-26]

Table 6 Comparison of separation performance of different advanced methods on MUSDB18 dataset[22-26] dB

模型	NSDR		SIR		SAR	
	歌声	伴奏	歌声	伴奏	歌声	伴奏
Conv-Tasnet[22]	6.81	12.69	14.30	17.61	6.87	13.77
E-MRP-CNN[23]	6.36	12.99	13.68	16.18	6.60	14.41
D3Net[24]	7.24	13.52				
RPCA-DRNN[25]	6.41	8.70	19.53	24.77	6.87	15.78
FC-U^2-Net[26]	8.22	13.39	20.70	21.26	8.33	14.57
本文	8.37	12.81	19.83	21.13	8.45	15.81

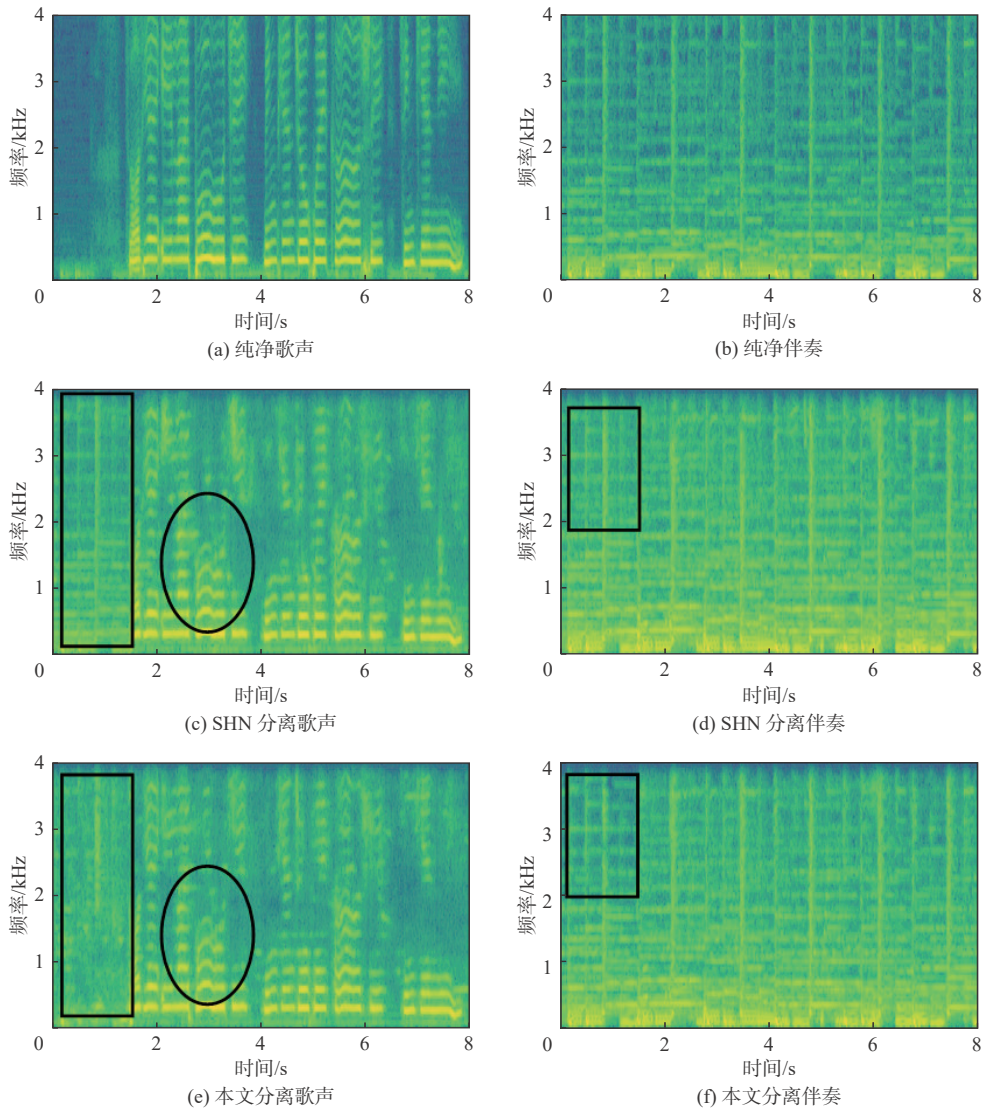


图 6　语谱图

Fig. 6　Spectrograms

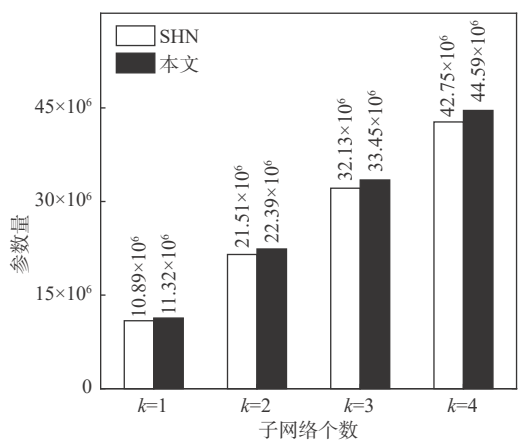


图 7　参数量对比

Fig. 7　Comparison of parameter quantities

4　结　论

1）本文构造出一个适用于歌声分离的门控自适应卷积单元来代替普通卷积层，以更高效地学习与提取声音的时频信息，并强调了神经网络在训练过程中易被忽视的特征通道关系。在 MIR-1K 数据集上，证实该单元模块能为网络有效预测出高精度的歌声与伴奏幅度谱。

2）本文方法在各级子网络间设置的监督注意力模块有效防止了迭代过程中重要语音信息流的丢失，实验表明，监督注意力提高了模型的分离精度，使得分离的歌声与伴奏在 GNSDR、GSIR 指标上显著提升。

3）本文提出一种基于多级渐进式门控卷积网络的歌声分离方法，在综合对比实验中表明，几乎所有分离评价指标均优于其他相关对比方法，说明该方法的语音失真相对较小，且具有较好的分离能力。

目前的分离模型仍难以对语音信号的相位精准建模，尽管人耳对语音的相位并不敏感，但分离出的歌声与伴奏质量可能仍会受到影响。后续研

究考虑将进一步优化网络结构,并考虑对频域模型中相位进性直接或间接估计,以提高分离伴奏与歌声的纯净度。

参考文献(References)

[1] RAFII Z, PARDO B. Repeating pattern extraction technique (REPET): a simple method for music/voice separation[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2013, 21(1): 73-84.

[2] HUANG P S, CHEN S D, SMARAGDIS P, et al. Singing-voice separation from monaural recordings using robust principal component analysis[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2012: 57-60.

[3] GRAIS E M, ERDOGAN H. Single channel speech music separation using nonnegative matrix factorization with sliding windows and spectral masks[C]//Proceedings of the Interspeech 2011. Copenhagen: ISCA, 2011: 1773-1776.

[4] UHLICH S, GIRON F, MITSUFUJI Y. Deep neural network based instrument extraction from music[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2015: 2135-2139.

[5] SPRECHMANN P, BRUNA J, LECUN Y. Audio source separation with discriminative scattering networks[C]//Proceedings of the Latent Variable Analysis and Signal Separation. Berlin: Springer, 2015: 259-267.

[6] 张天骐, 熊梅, 张婷, 等. 结合区分性训练深度神经网络的歌声与伴奏分离方法[J]. 声学学报, 2019, 44(3): 393-400.
ZHANG T Q, XIONG M, ZHANG T, et al. A separation method of singing and accompaniment combining discriminative training deep neural network[J]. Acta Acustica, 2019, 44(3): 393-400(in Chinese).

[7] CHEN J T, WANG D L. Long short-term memory for speaker generalization in supervised speech separation[J]. The Journal of the Acoustical Society of America, 2017, 141(6): 4705.

[8] 张天. 单通道音乐信号中的人声伴奏分离方法研究[D]. 重庆: 重庆邮电大学, 2020: 43-57.
ZHANG T. Research on separation method of vocal accompaniment in single channel music signal[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2020: 43-57(in Chinese).

[9] JANSSON A, HUMPHREY E, MONTECCHIO N, et al. Singing voice separation with deep U-Net convolutional networks[C]//Proceedings of the 18th International Society for Music Information Retrieval Conference. [S. l.]: DBLP, 2017: 745-751.

[10] STOLLER D, EWERT S, DIXON S. Wave-U-Net: a multi-scale neural network for end-to-end audio source separation[EB/OL]. (2018-06-08)[2023-06-01]. https://arxiv.org/abs/1806.03185v1.

[11] DÉFOSSEZ A, USUNIER N, BOTTOU L, et al. Demucs: deep extractor for music sources with extra unlabeled data remixed[EB/OL]. (2019-09-03)[2023-06-01]. https://arxiv.org/abs/1909.01174v1.

[12] IBTEHAZ N, RAHMAN M S. MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation[J]. Neural Networks, 2020, 121: 74-87.

[13] PARK S, KIM T, LEE K, et al. Music source separation using stacked hourglass networks[EB/OL]. (2018-06-22)[2023-06-01]. https://arxiv.org/abs/1805.08559v2.

[14] YUAN W T, WANG S B, LI X R, et al. A skip attention mechanism for monaural singing voice separation[J]. IEEE Signal Processing Letters, 2019, 26(10): 1481-1485.

[15] BHATTARAI B, PANDEYA Y R, LEE J. Parallel stacked hourglass network for music source separation[J]. IEEE Access, 2020, 8: 206016-206027.

[16] 买峰. 基于深度卷积神经网络的音乐源分离算法及其应用研究[D]. 成都: 电子科技大学, 2021: 20-70.
MAI F. Research on music source separation algorithm based on deep convolution neural network and its application[D]. Chengdu: University of Electronic Science and Technology of China, 2021: 20-70(in Chinese).

[17] SUBAKAN C, RAVANELLI M, CORNELL S, et al. Attention is all you need in speech separation[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2021: 21-25.

[18] 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: 11791-11800.

[19] YANG Y H. Low-rank representation of both singing voice and music accompaniment via learned dictionaries[C]//Proceedings of the 14th International Society for Music Information Retrieval Conference. Curitiba: [s. n.], 2013: 427-432.

[20] HUANG P S, KIM M, HASEGAWA-JOHNSON M, et al. Joint optimization of masks and deep recurrent neural networks for monaural source separation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(12): 2136-2147.

[21] SEBASTIAN J, MURTHY H A. Group delay based music source separation using deep recurrent neural networks[C]//Proceedings of the International Conference on Signal Processing and Communications. Piscataway: IEEE Press, 2016: 1-5.

[22] DÉFOSSEZ A, USUNIER N, BOTTOU L, et al. Music source separation in the waveform domain[EB/OL]. (2021-04-28)[2023-06-01]. https://arxiv.org/abs/1911.13254v2.

[23] YUAN W T, DONG B F, WANG S B, et al. Evolving multi-resolution pooling CNN for monaural singing voice separation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 807-822.

[24] TAKAHASHI N, MITSUFUJI Y, TAKAHASHI N, et al. D3Net: densely connected multidilated DenseNet for music source separation[EB/OL]. (2021-05-27)[2023-06-01]. https://arxiv.org/abs/2010.01733v4.

[25] LAI W H, WANG S L. RPCA-DRNN technique for monaural singing voice separation[J]. EURASIP Journal on Audio, Speech, and Music Processing, 2022, 2022: 4.

[26] NI X, REN J. FC-U^2-Net: a novel deep neural network for singing voice separation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 30: 489-494.

Singing voice separation method using multi-stage progressive gated convolutional network

LUO Qingyu，ZHANG Tianqi*，XIONG Tian

(School of Communications and Information Engineering，Chongqing University of Posts and Telecommunications，Chongqing 400065，China)

Abstract：To solve the problems that current singing voice separation algorithms based on convolutional neural network (CNN) have semantic differences in the fusion of high- and low-layer features and ignore the potential value of speech features in channel dimension, this paper proposed a stacked multi-stage progressive gated convolutional network to achieve singing voice separation. Firstly, a gated adaptive convolution (GAC) unit was designed in each level of subnetwork to fully learn and extract the time-frequency features of songs and enhance competition and cooperation between the feature channels. Then, to reduce the semantic errors in the fusion of shallow and deep network information, a gated attention mechanism was introduced between the codec layers of the subnetwork. Finally, supervised attention (SA) was proposed for different levels of subnetwork to selectively deliver effective information flow and realize progressive learning of multi-stage networks. Comprehensive comparative experiments were carried out on a large dataset and a small dataset publicly available. The results show that compared with the representative models in recent years, the algorithm has certain advantages in separating singing voice and background accompaniment.

Keywords：singing voice separation; neural network; multi-stage progressive; adaptive; supervised attention

Received：2023-06-28；**Accepted**：2023-11-03；**Published Online**：2023-11-24 11：02

URL：link.cnki.net/urlid/11.2625.V.20231123.1413.004

Foundation items：National Natural Science Foundation of China (61771085); Natural Science Foundation of Chongqing, China (cstc2021jcyj-msxmX0836); Research Project of Chongqing Educational Commission (KJ1600427,KJ1600429)

* **Corresponding author**. E-mail：zhangtq@cqupt.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2025.0425

双三相起动发电系统电源变换器功率管开路故障诊断

王化吉[1]，刘禹[2]，孙高阳[3]，李德洪[1]，徐金全[2,*]

(1. 中国直升机设计研究所，景德镇 333001；2. 北京航空航天大学 自动化科学与电气工程学院，北京 100191；
3. 苏州汇川联合动力系统股份有限公司，苏州 215000)

摘 要：为解决航空双三相永磁辅助式同步磁阻起动发电系统电源变换器的功率管开路故障诊断问题，提出了基于z_1-z_2坐标系电流和直流母线电压的功率管单相开路故障诊断与定位方法。采用空间矢量解耦方法建立电机本体和电源变换器模型；基于该模型分析电源变换器单相单管和单相双管开路故障特性及其在z_1-z_2坐标系电流轨迹和直流母线电压上反映出的故障特征；提出基于z_1-z_2坐标系电流和直流母线电压的单相单管及单相双管开路故障诊断与定位方法，检测故障发生，确定故障位置。通过实验平台验证所提方法的有效性，结果表明：所提方法可对电源变换器单相开路故障进行诊断与定位，进而为故障发生后的容错运行提供指导。

关 键 词：故障诊断；功率管开路；起动发电机；电源变换器；谐波电流
中图分类号：V242.2
文献标志码：A **文章编号**：1001-5965(2025)09-2882-12

常规直升机由于构型影响，限制了其飞行速度等主要性能指标，高速化已成为未来直升机发展的重要趋势之一，而大功率电驱动技术的出现则为直升机高速化设计带来了新的变革[1]。然而，随着电驱动技术的广泛应用，飞机用电量快速增长，进而对电源系统的可靠性提出了更严格的要求。发电系统作为飞机电源系统的主要组成部分，负责将发动机的机械能转换为电能，是飞机用电设备在正常工作情况下全部电能的来源[2]。因此，高可靠的发电系统是先进直升机电源系统的基础。随着航空发电系统功率等级的提高与现代电力电子技术的发展，将起动功能集成于发电系统的起动发电系统应运而生[3]。永磁辅助式同步磁阻电机(permanent magnet assisted synchronous reluctance motor, PMA-SynRM)凭借其高功率密度、高可靠性等优势及较小的短路电流成为了航空起动发电机的重要发展方向。而与PMA-SynRM相配套的电源变换器则是保证起动发电系统正常运行的关键设备，如何在电源变换器发生故障时及时确定故障类型与故障位置，对系统出现故障后的容错处置具有重要意义。

目前，针对电源变换器的故障诊断方法可以分为基于模型的方法、基于特征的方法和基于数据驱动的方法。基于模型的方法[4-5]以模型估计值和实测值之差为故障判断依据。基于特征的方法[6-7]以故障状态下发电系统电压、电流等信号的特征变化为依据进行故障诊断。基于数据驱动的方法[8-9]着眼于对系统运行过程数据进行分析处理，从而在无需得到精确的发电系统模型的前提下，利用机器学习、统计分析等数据处理技术，挖掘数据分布规律与故障模式间的隐含映射机制，从而精准识别故障。

Pang等[10]采用基于模型的方法重构了三级式同步起动发电机旋转整流器的电流，并使用重构电流二次谐波与基波幅值之比来实现二极管开路与短路故障的诊断与定位。Jiao等[11]在分析了正常

收稿日期：2025-07-01；录用日期：2025-07-11；网络出版时间：2025-07-16 15:58
网络出版地址：link.cnki.net/urlid/11.2625.V.20250716.1454.001
*通信作者. E-mail：xujinquan@buaa.edu.cn

引用格式：王化吉，刘禹，孙高阳，等. 双三相起动发电系统电源变换器功率管开路故障诊断[J]. 北京航空航天大学学报，2025，51(9)：2882-2893. WANG H J，LIU Y，SUN G Y，et al. Open-circuit fault diagnosis of power tubes in power converter of dual three-phase starter-generator system[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2882-2893 (in Chinese).

情况与故障情况下主励磁机转子电流轨迹特征后，提出了一种基于特征的三级式同步起动发电机旋转整流器二极管故障诊断与定位方法。为诊断开关磁阻电机功率变换器故障，张竞文等[12]通过引入多尺度排列熵对直流母线电流进行变分模态分解，实现了开关磁阻电机不对称功率变换器开路短路的故障诊断。Najafabadi 等[13]使用基于电机模型的自适应观测器对电机电阻、电流和磁链进行估算，分别用于相电流传感器、速度传感器和直流母线电压传感器的故障诊断。Salehifar 等[14]利用五相永磁同步电机模型对电机各相电流进行估算，并计算出相电流估测值与实际值的差值来指示缺相故障和功率管开路故障。Duran 等[15]通过分析双三相电机 z_1-z_2 坐标系电流特征，实现电机缺相故障的诊断。Arafat 等[16]提取出五相同步磁阻电机相电流的对称分量，根据对称分量中基波电流的幅值和相角变化量实现电机缺相故障的诊断。Cai 等[17]为解决由传感器偏差、噪声等多种原因引起的故障诊断不确定性问题，提出了一种基于贝叶斯网络的三相逆变器故障诊断方法。Xia 等[18]基于极限学习机和随机向量函数链接网络 2 种机器学习技术，提出了一种混合集成学习方案，用于提取故障模式与所选特征之间的映射关系，实现了感应电机驱动器开路故障诊断。

为提高起动发电系统的容错能力，起动发电机往往采用多相结构。然而，目前对多相发电系统电源变换器的故障诊断研究相对较少，且关于电源变换器故障诊断的研究大多针对起动发电机的电动运行阶段，而未考虑发电运行阶段。起动发电机工作在发电状态时，具有与电动状态相反的功率流向，并且由于功率器件自身存在反并联二极管，功率管开路后，电流可能仍可通过二极管续流，导致部分故障特征被掩盖，进而为故障诊断带来了困难。对此，本文以双三相永磁辅助式同步磁阻起动发电（dual three-phase permanent magnet assisted synchronous reluctance starter-generator, DTPPMA-SynRSG）系统中的电源变换器为对象，基于空间矢量解耦（vector space decomposition, VSD）的方法建立电机本体模型和电源变换器模型，分析了电源变换器分别在单相单管开路故障、单相双管开路故障情况下的 z_1-z_2 坐标系相电流谐波轨迹特征和直流母线电压波动特征；根据分析得到的故障特征，提出了基于 z_1-z_2 坐标系电流和直流母线电压的电源变换器功率管开路故障诊断方法，通过搭建实验平台验证了所提方法的有效性。

1　DTPPMA-SynRSG 系统模型结构分析

1.1　DTPPMA-SynRSG 本体模型

DTPPMA-SynRSG 绕组结构如图 1 所示。转子分为 4 极，每极采用 3 层 U 形磁障结构，每层磁障由水平磁障和倾斜磁障组成，磁障边缘设置内磁桥与外磁桥，磁障与磁障之间的铁心部分称为磁肋，利用转子的凸极性可以得到磁阻转矩。为提高功率因数、功率密度和效率，在水平磁障中嵌入永磁体。为降低每相电流并增强起动发电系统的容错能力，电机定子采用 30°相移双三相 Y 形绕组。

在静止坐标系下，电机的电压方程、功率方程和转矩方程别为[19]

$$p_{\mathrm{M}} = \boldsymbol{u}_{\mathrm{s}}^{\mathrm{T}} \boldsymbol{i}_{\mathrm{s}} \tag{1}$$

$$\boldsymbol{u}_{\mathrm{s}} = R_{\mathrm{s}} \boldsymbol{i}_{\mathrm{s}} + \dot{\boldsymbol{\psi}}_{\mathrm{s}} \tag{2}$$

$$T_{\mathrm{c}} = \frac{\partial w'_{\mathrm{M}}(\boldsymbol{i}_{\mathrm{s}}, \theta_{\mathrm{m}})}{\partial \theta_{\mathrm{m}}} \tag{3a}$$

或

$$T_{\mathrm{c}} = \frac{\partial w'_{\mathrm{M}}(\boldsymbol{\psi}_{\mathrm{s}}, \theta_{\mathrm{m}})}{\partial \theta_{\mathrm{m}}} \tag{3b}$$

式中：$\boldsymbol{u}_{\mathrm{s}} = [u_A, u_B, u_C, u_D, u_E, u_F]^{\mathrm{T}}$ 为定子电压矢量；$\boldsymbol{i}_{\mathrm{s}} = [i_A, i_B, i_C, i_D, i_E, i_F]^{\mathrm{T}}$ 为定子电流矢量（以流入电机为正方向）；p_{M} 为电机输出功率；$\boldsymbol{\psi}_{\mathrm{s}} = [\psi_A, \psi_B, \psi_C, \psi_D, \psi_E, \psi_F]^{\mathrm{T}}$ 为定子磁链矢量；R_{s} 为定子相电阻；T_{e} 为电机电磁转矩；w'_{M} 为电机磁共能；θ_{m} 为转子机械角度（等于电角度 θ_{re} 除以极对数）。

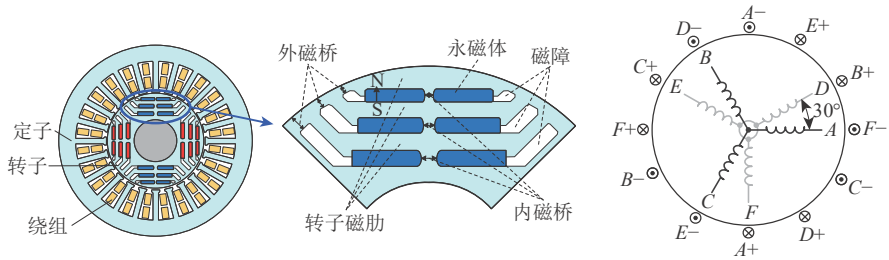


图 1　DTPPMA-SynRSG 转子结构与定子绕组

Fig. 1　DTPPMA-SynRSG rotor structure and stator winding

为将电机模型变换到旋转坐标系中以便于后续分析电压、电流特性，进一步对上述模型进行VSD变换。VSD变换矩阵为

$$\boldsymbol{T}_{\mathrm{VSD}}=\begin{bmatrix}\boldsymbol{P}(\theta_{\mathrm{re}}) & \boldsymbol{0}_{2\times2}\\ \boldsymbol{0}_{2\times2} & \boldsymbol{I}_{2\times2}\end{bmatrix}\cdot \frac{1}{3}\begin{bmatrix}1 & -\frac{1}{2} & -\frac{1}{2} & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & 0\\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & \frac{1}{2} & \frac{1}{2} & -1\\ 1 & -\frac{1}{2} & -\frac{1}{2} & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & 0\\ 0 & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & \frac{1}{2} & \frac{1}{2} & -1\end{bmatrix} \tag{4}$$

式中：$\boldsymbol{P}(\theta_{\mathrm{re}})=\begin{bmatrix}\cos\theta_{\mathrm{re}} & \sin\theta_{\mathrm{re}}\\ -\sin\theta_{\mathrm{re}} & \cos\theta_{\mathrm{re}}\end{bmatrix}$为Park变换矩阵；$\boldsymbol{I}_{2\times2}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$。

采用VSD变换后，电机方程变为

$$\boldsymbol{u}_{\mathrm{VSD}}=R_{\mathrm{s}}\boldsymbol{i}_{\mathrm{VSD}}+\dot{\boldsymbol{\psi}}_{\mathrm{VSD}}+\omega_{\mathrm{re}}\boldsymbol{J}_{\mathrm{VSD}}\boldsymbol{\psi}_{\mathrm{VSD}} \tag{5}$$

$$p_{\mathrm{M}}=3\boldsymbol{u}_{\mathrm{VSD}}^{\mathrm{T}}\boldsymbol{i}_{\mathrm{VSD}} \tag{6}$$

式中：$\boldsymbol{u}_{\mathrm{VSD}}=[u_d,u_q,u_{z_1},u_{z_2}]^{\mathrm{T}}$；$\boldsymbol{i}_{\mathrm{VSD}}=[i_d,i_q,i_{z_1},i_{z_2}]^{\mathrm{T}}$；$\boldsymbol{\psi}_{\mathrm{VSD}}=[\psi_d,\psi_q,\psi_{z_1},\psi_{z_2}]^{\mathrm{T}}$；$\omega_{\mathrm{re}}$为电角速度；$\boldsymbol{J}_{\mathrm{VSD}}=\begin{bmatrix}\boldsymbol{J} & \\ & \boldsymbol{0}_{2\times2}\end{bmatrix}$，$\boldsymbol{J}=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$为正交旋转矩阵。

通过VSD变换，DTPPMA-SynRSG电气量中的基波和$12k\pm1(k=1,2,3)$次谐波被映射到d-q坐标系，$6k\pm1(k=1,3,5)$次谐波被映射到z_1-z_2坐标系。

1.2 电源变换器模型

DTPPMA-SynRSG系统电源变换器所用全控式全桥电路具有功率双向流通的能力，因此，可凭借一套主功率电路实现起动和发电双重功能，从而提高电源变换器的功率密度。此外，全控式全桥电路控制灵活，可使DTPPMA-SynRSG的磁阻转矩和永磁转矩均得到有效控制，进而使DTPPMA-SynRSG的性能得到充分发挥。基于全控式全桥电路拓扑的电源变换器及其在直升机电气系统中的连接关系如图2所示。电源变换器双三相全桥电路由6个半桥电路并联组成，其交流侧与电机的双三相绕组相连接，直流侧与直流母线电容相连接。直流母线电容用于稳定直流母线电压，向外与配电网络连接。

定义电机定子电流矢量为$\boldsymbol{i}_{\mathrm{s}}=[i_A,i_B,i_C,i_D,i_E,i_F]^{\mathrm{T}}$，参考方向以流出电机为正向。忽略死区时间，同一相桥臂上下功率管的驱动信号是互补的，定义k相桥臂的开关状态函数如下：

$$s_k=\begin{cases}1 & \text{上桥臂功率管导通下桥臂功率管关断}\\ 0 & \text{下桥臂功率管导通下桥臂功率管关断}\end{cases} \tag{7}$$

式中：$k=A,B,C,D,E,F$。则六相开关矢量表达式为$\boldsymbol{s}=[s_A,s_B,s_C,s_D,s_E,s_F]^{\mathrm{T}}$。

进而由图2可以计算得到全控式全桥电路输出电流i_{out}为

$$i_{\mathrm{out}}=\boldsymbol{s}^{\mathrm{T}}\boldsymbol{i}_{\mathrm{s}} \tag{8}$$

由于脉冲宽度调制(pulse width modulation, PWM)作用，功率管高频开关会导致i_{out}中包含大量调制频率相关的谐波分量。记i_{out}中的直流分量与谐波分量分别为i_{outDC}和i_{outH}，则i_{out}为

$$i_{\mathrm{out}}=i_{\mathrm{outDC}}+i_{\mathrm{outH}} \tag{9}$$

由于电源变换器通常距离负载和起动电源较远，线路寄生电感远大于直流母线电容的内部等效串联电感，因此，i_{out}中的谐波分量i_{outH}大部分流经直流母线电容C_{DC}，而电源变换器输出电流i_{DC}中谐波含量较少。根据图2和式(10)，直流母线电容电压u_{DC}表示为

$$u_{\mathrm{DC}}=\frac{1}{C_{\mathrm{DC}}}\int(i_{\mathrm{out}}-i_{\mathrm{DC}})\mathrm{d}t=\frac{1}{C_{\mathrm{DC}}}\int(i_{\mathrm{outDC}}-i_{\mathrm{DC}})\mathrm{d}t+\frac{1}{C_{\mathrm{DC}}}\int i_{\mathrm{outH}}\mathrm{d}t \tag{10}$$

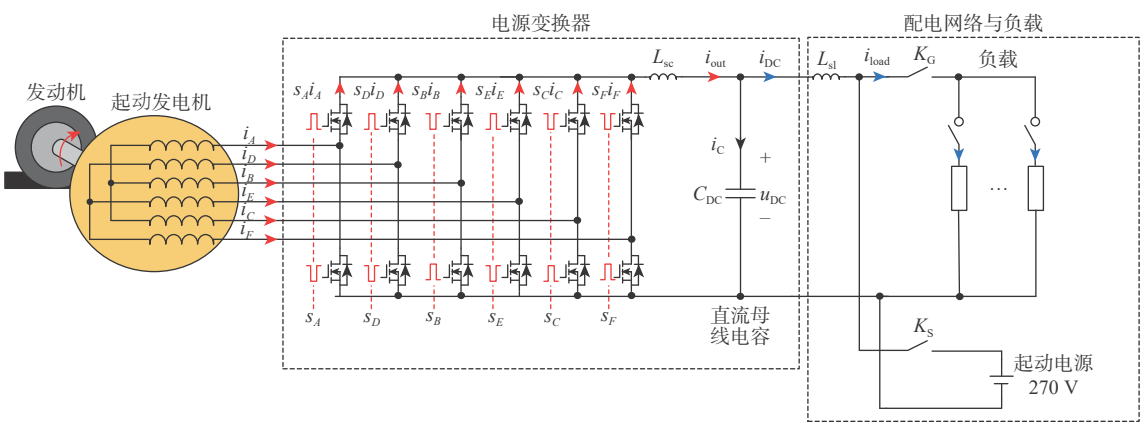


图2 电源变换器连接关系

Fig. 2 Power converter connection relationship

1.3　整体控制结构

起动发电系统控制架构如图 3 所示，其分为内外两环，通过位置传感器实时监测 DTPPMA-SynRSG 的转子位置θ_{re}、转子速度ω_{re}和直流母线电压u_{DC}，并将其反馈给外环。由外环根据起动、发电运行阶段执行相应的外环控制算法。内环为采用 VSD 变换控制的电流环，通过采集电机 2 套绕组的相电流$\boldsymbol{i}_{ABC}$和$\boldsymbol{i}_{DEF}$对i_d和i_q给出控制指令u_d^*和u_q^*，进行闭环控制，而对i_{z_1}和i_{z_2}进行控制指令为$u_{z_1}^* = u_{z_2}^* = 0$ V的开环控制。通过引入最大转矩电流比(maximum torque per ampere, MTPA)电流分配策略，将外环的起动控制器和发电控制器输出的电流指令分配至内环 d、q 轴电流指令i_d^*、i_q^*。

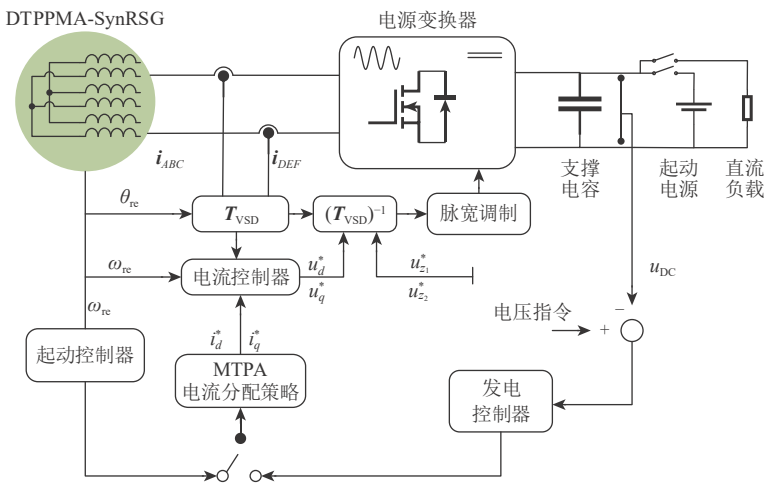


图 3　起动发电系统控制结构

Fig. 3　Control structure of starter-generator system

2　功率管开路故障特征分析

2.1　单相单管开路故障

电源变换器单相单管开路故障对起动发电系统的起动运行和发电运行都会产生影响。单相单管开路故障发生后，受闭环调节作用，双三相电源变换器剩余正常功率管虽能维持起动发电系统继续运行，但电机绕组和电源变换器功率管可能因相电流不对称而出现局部过热问题。此外，电机的起动转矩或发电功率也可能出现严重脉动。因此，需要及时完成对单相单管开路故障的诊断与定位才能执行后续相应的容错控制算法。当单相单管开路故障发生后，起动发电系统 z_1-z_2 坐标系电流和直流母线电压将会出现相应的故障特征，基于这些特征即可实现单相单管开路故障的诊断和定位。首先，分析 z_1-z_2 坐标系电流特征，图 4 为桥臂分别在正常和上、下功率管开路故障时的电流路径。

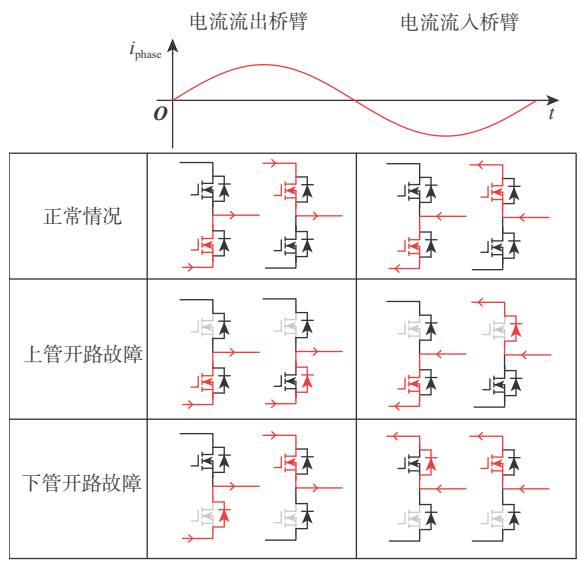


图 4　正常和上、下功率管开路故障情况下的桥臂电流路径

Fig. 4　Bridge arm current path under normal conditions and open-circuit fault of upper and lower power transistors

由图 4 可见，在正常情况下，无论相电流方向，半桥电路均能输出高、低 2 种电平，即$0.5u_{DC}$和$-0.5u_{DC}$(以直流母线电容虚拟中点为参考点)。然而，对于上管开路故障，当电流流出桥臂时，半桥电路只能输出$-0.5u_{DC}$；对于下管开路故障，当电流流入桥臂时，半桥电路只能输出$0.5u_{DC}$。因此，上管开路故障会导致在一半时间内半桥电路输出电压值偏低，下管开路故障则会导致在一半时间内半桥电路输出电压值偏高，即上管和下管的开路故障均会导致特定极性的偏差电压。忽略三次相电压谐波后，DTPPMA-SynRSG 相、线电压满足如下关系：

$$\begin{bmatrix} u_{AB} \\ u_{BC} \\ 0 \\ u_{DE} \\ u_{EF} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} u_A \\ u_B \\ u_C \\ u_D \\ u_E \\ u_F \end{bmatrix} \tag{11}$$

将式(11)求逆后,得

$$\begin{bmatrix} u_A \\ u_B \\ u_C \\ u_D \\ u_E \\ u_F \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & -2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -1 & -2 \end{bmatrix}\begin{bmatrix} u_{AB} \\ u_{BC} \\ u_{DE} \\ u_{EF} \end{bmatrix} \tag{12}$$

将式(12)代入VSD变换矩阵中,可以得到z_1-z_2坐标系电压与线电压之间关系为

$$\begin{bmatrix} u_{z_1} \\ u_{z_2} \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 2 & 1 & -\sqrt{3} & 0 \\ 0 & -\sqrt{3} & 1 & 2 \end{bmatrix}\begin{bmatrix} u_{AB} \\ u_{BC} \\ u_{DE} \\ u_{EF} \end{bmatrix} \tag{13}$$

因此,单相单管开路故障等效于电源变换器供给的线电压u_{AB}、u_{BC}、u_{DE}、u_{EF}中的1个或2个出现错误。例如,当发生A相单管开路故障时,u_{AB}出现错误,而当发生B相单管开路故障时,u_{AB}、u_{BC}都会出现错误。对于引起2个线电压错误的单相单管开路故障情况,可以把式(14)代入(13)中,进而使之转换为单个线电压错误的情况。

$$\begin{cases} u_{AB} + u_{BC} + u_{CA} = 0 \\ u_{DE} + u_{EF} + u_{FD} = 0 \end{cases} \tag{14}$$

以A相单管开路故障为例,故障后u_{AB}可以视为其正确值u_{AB}^{C}与误差值Δu_{AB}之和,即

$$u_{AB} = u_{AB}^{\mathrm{C}} + \Delta u_{AB} \tag{15}$$

将式(15)代入式(13),且考虑到DTPPMA-SynRSG控制系统对于u_{z_1}和u_{z_2}采用的是控制指令为$u_{z_1}^* = u_{z_2}^* = 0$ V的开环控制,则可以计算出A相单管开路故障情况下的z_1-z_2坐标系电压$\boldsymbol{u}_{z_1z_2}$为

$$\boldsymbol{u}_{z_1z_2} = \begin{bmatrix} u_{z_1} \\ u_{z_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{3}\Delta u_{AB} \\ 0 \end{bmatrix} \tag{16}$$

由于单相单管开路故障会导致特定极性的偏差电压,使得故障下$\boldsymbol{u}_{z_1z_2}$电压矢量的轨迹为起始于原点的射线,射线的斜率与故障位置有关。由式(16)可知,A相上管开路故障和A相下管开路故障引起的$\boldsymbol{u}_{z_1z_2}$电压矢量轨迹A_{U}和A_{L}的斜率为0,方向分别为z_1轴负向和正向。按照同样的方法,可以计算出其余各位置单管开路故障的z_1-z_2坐标系直流母线电压轨迹,如图5所示。

根据式(4)所示VSD变换矩阵,可得到经VSD变换后的DTPPMA-SynRSG电压方程,如下:

$$\begin{bmatrix} u_d \\ u_q \\ u_{z_1} \\ u_{z_2} \end{bmatrix} = \begin{bmatrix} R_{\mathrm{s}} & & & \\ & R_{\mathrm{s}} & & \\ & & R_{\mathrm{s}} & \\ & & & R_{\mathrm{s}} \end{bmatrix}\begin{bmatrix} i_d \\ i_q \\ i_{z_1} \\ i_{z_2} \end{bmatrix} + \begin{bmatrix} \dot{\psi}_d \\ \dot{\psi}_q \\ \dot{\psi}_{z_1} \\ \dot{\psi}_{z_2} \end{bmatrix} + \omega_{\mathrm{re}}\begin{bmatrix} -\psi_q \\ -\psi_d \\ 0 \\ 0 \end{bmatrix} \tag{17}$$

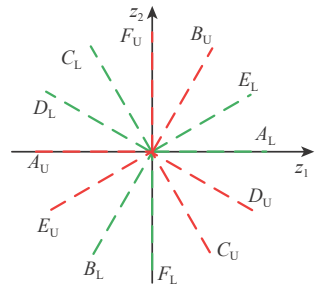


图5 不同单相单管开路故障下的z_1-z_2坐标系直流母线电压轨迹

Fig. 5 DC bus voltage trajectory of z_1-z_2 coordinate system under different single-phase single-tube open-circuit faults

由式(17)可知,z_1-z_2坐标系电流与d-q坐标系电流解耦,假设z_1-z_2坐标系电流与z_1-z_2坐标系直流母线电压呈线性关系,即

$$\boldsymbol{u}_{z_1z_2} = R_{\mathrm{s}}\boldsymbol{i}_{z_1z_2} + L_1\frac{\mathrm{d}\boldsymbol{i}_{z_1z_2}}{\mathrm{d}t} \tag{18}$$

式中:L_1为z_1-z_2坐标系下的电感。则在正常情况下,由于所采用的控制方式保证了$\boldsymbol{u}_{z_1z_2} = \boldsymbol{0}$,稳态时,$\boldsymbol{i}_{z_1z_2} \approx \boldsymbol{0}$。而在单相单管开路故障下,$\boldsymbol{u}_{z_1z_2}$会产生相应斜率的误差电压矢量,并导致$\boldsymbol{i}_{z_1z_2}$也产生同样斜率的电流矢量。因此,可以选取$\boldsymbol{i}_{z_1z_2}$电流作为单相单管开路的故障特征,并根据$\boldsymbol{i}_{z_1z_2}$的幅值和相位实现单相单管故障诊断与定位。

分析单相单管开路故障情况下的直流母线电压特征。假设A相上桥臂功率管出现开路故障,由前文分析可知,当电流流出桥臂时,桥臂只能输出低电平。而维持电流流出桥臂,通常需要桥臂输出高电平,这将导致相电流在原本应该流出桥臂时被钳位为0 A。假设A相上桥臂功率管出现开路故障后,A相电流变为(电流参考正方向为流出桥臂)

$$i_{A_{\mathrm{OCU}}} = \begin{cases} 0 & 0 < \omega_{\mathrm{re}}t < \pi \\ I\sin(\omega_{\mathrm{re}}t) & -\pi < \omega_{\mathrm{re}}t < 0 \end{cases} \tag{19}$$

式中:I为电流幅值。

根据前文分析,当A相上桥臂功率管出现开路故障后,若电流流入桥臂,则桥臂仍然能够提供正常的电压。因此,当$\pi/2 < \omega_{\mathrm{re}}t < 3/2\pi$时,$u_A$仍然为$U\cos(\omega_{\mathrm{re}}t + \varphi)$,$U$为电压幅值,$\varphi$为功率因数角。则$A$相上桥臂功率管出现开路故障后,$A$相绕组的输出功率为

$$P_{A_\mathrm{OCU}} = \begin{cases} 0 & 0 < \omega_{\mathrm{re}}t < \pi \\ UI\sin(\omega_{\mathrm{re}}t)\sin(\omega_{\mathrm{re}}t + \varphi) & -\pi < \omega_{\mathrm{re}}t < 0 \end{cases} \tag{20}$$

因此,A相上桥臂功率管开路故障会导致A相绕组输出功率产生如下变化:

$$\delta P_A = P_{A_OCU} - P_{Anorm} = \begin{cases} -\dfrac{1}{2}UI\left(\cos\varphi - \cos(2\omega_{re}t + \varphi)\right) & 0 < \omega_{re}t < \pi \\ 0 & -\pi < \omega_{re}t < 0 \end{cases} \tag{21}$$

式中：P_{Anorm}为正常情况下 A 相绕组输出功率；δP_A为以2π为周期的函数，利用傅里叶变换可将其展开为

$$\delta P_A = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n\cos(n\omega_{re}t) + b_n\sin(n\omega_{re}t)\right) \tag{22}$$

$$a_n = \frac{1}{\pi}\int_0^{\pi}\delta P_A\cos(n\omega_{re}t)\mathrm{d}(\omega_{re}t) \tag{23}$$

$$b_n = \frac{1}{\pi}\int_0^{\pi}\delta P_A\sin(n\omega_{re}t)\mathrm{d}(\omega_{re}t) \tag{24}$$

n 次谐波的幅值为 $\sqrt{a_n^2+b_n^2}$，据此，以一次谐波（$n=1$）幅值为基准，分别计算出功率因数$\cos\varphi$等于 1, 0.9, 0.8, 0.7, …, 0 时，δP_A中直流分量幅值比 H_{DC}、一次谐波分量幅值比H_1及其他各次谐波幅值比结果如图 6 所示。可见，当出现单相单管开路故障时，故障相绕组功率变化量的谐波含量较为丰富，其中，最显著的谐波分量为基频分量与二次谐波分量。考虑到通常电机运行功率因数不低于 0.7，因此，基频分量幅值大于二次谐波分量。即使在功率因数低于 0.7 的情况下，二次谐波分量幅值最高也不超过基频分量幅值的 6/5 倍。

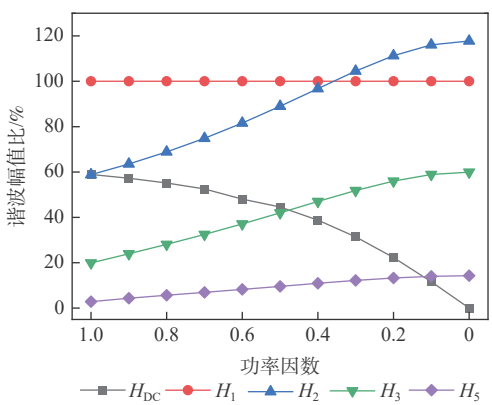


图 6　单相单管开路故障下绕组输出功率变化量各次谐波幅值比

Fig. 6　Amplitude ratio of each harmonic component of change in winding output power under single-phase single-tube open circuit fault

单相单管开路故障产生的基频和二倍频功率脉动会导致直流母线电压分别产生基频和二倍频的电压脉动。因此，直流母线电压也可作为单相单管开路故障特征。上述分析基于 A 相上桥臂功率管开路故障的假设，对于其他位置的功率管开路故障情况，上述分析方法与结论同样适用。

2.2　单相双管开路故障

与单相单管开路故障类似，电源变换器单相双管开路故障对起动发电系统的起动和发电运行都将产生影响。单相双管开路故障可被视为电源变换器给正常的电机提供了错误的电压，并使电机某相电流为 0 A。单相双管开路故障发生后，受闭环调节作用，起动发电系统的剩余五相虽能继续维持系统运行，但电机绕组和电源变换器功率管同样可能因相电流不均而出现局部过热的问题。此外，电机的起动转矩或发电功率也可能出现严重脉动。因此，需要对单相双管开路故障进行诊断并定位，从而执行相对应的容错控制算法。

与单相单管开路故障相同，单相双管开路故障也会导致错误的$\boldsymbol{u}_{z_1z_2}$电压，并最终使$\boldsymbol{i}_{z_1z_2}$出现异常，因此，单相双管开路故障也可将 z_1-z_2 坐标系电流作为故障特征。不同于单相单管开路故障的是，单相双管开路故障不会导致特定极性的偏差电压，因此，故障下$\boldsymbol{u}_{z_1z_2}$电压矢量的轨迹关于原点对称。与单相单管开路故障情况相同，轨迹的斜率由故障位置决定。以 A 相双管开路故障为例，$\boldsymbol{u}_{z_1z_2}$的轨迹为 z_1-z_2 坐标系下一条斜率为 0 的直线。按照单相单管开路故障分析方法，可以计算出 B、C、D、E、F 各相故障情况下的$\boldsymbol{u}_{z_1z_2}$电压，其轨迹分别为 z_1-z_2 坐标系上斜率为 $\sqrt{3}$、$-\sqrt{3}$、$-\sqrt{3}/3$、$\sqrt{3}/3$、∞的直线，如图 7 所示。

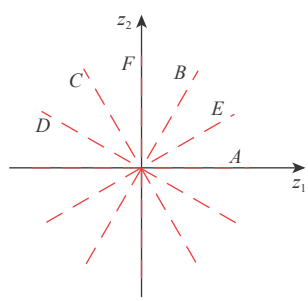


图 7　各相双管开路故障下的 z_1-z_2 坐标系电流轨迹

Fig. 7　Current trajectory in z_1-z_2 coordinate system under open-circuit faults in each phase double-tube

单相双管开路故障同样会引起电机输出功率脉动，并导致直流母线电压脉动。以 A 相为例，正常情况下，A 相绕组输出功率为P_{Anorm}：

$$P_{Anorm} = u_A i_A \tag{25}$$

忽略电压与电流中的谐波，$i_A = I\sin(\omega_{re}t)$，$u_A = U\sin(\omega_{re}t+\varphi)$，可以计算出正常情况下 A 相绕组的输出功率为

$$P_{Anorm} = \frac{1}{2}UI\left(\cos\varphi - \cos(2\omega_{re}t+\varphi)\right) \tag{26}$$

当发生 A 相双管开路故障时，A 相电流恒等于 0 A，即$i_A \equiv 0$，因此，单相双管开路故障下，A 相绕组

的输出功率为

$$P_{A\text{lost}}=0$$

因此，A相双管开路故障会导致A相绕组输出功率产生如下变化：

$$\delta P_A=P_{A\text{lost}}-P_{A\text{norm}}=-\frac{1}{2}UI\cos\varphi+\frac{1}{2}UI\cos(2\omega_{\text{re}}t+\varphi) \tag{27}$$

由此可见，A相双管开路故障会使A相输出功率产生直流分量的降幅及二倍基频的脉动。由于电压闭环调节作用，直流分量的功率降低会由其余健全相绕组提高输出功率补偿，并不会造成直流母线电压稳态值的变化。而二倍基频功率脉动将会导致直流母线电压产生同频率的电压脉动。因此，直流母线电压中的二倍基频脉动幅值可作为单相双管开路故障特征。

3 功率管开路故障诊断与定位方法

单相双管开路故障与单相单管开路故障都会使$\boldsymbol{i}_{z_1z_2}$的幅值增大，且使其轨迹处于特定斜率的直线上，这为2类故障的同步诊断提供了便利。然而，在单相双管开路故障和单相单管开路故障下，$\boldsymbol{i}_{z_1z_2}$的区别并不明显，仅仅依靠$\boldsymbol{i}_{z_1z_2}$信息来区分2类故障并定位故障位置容易导致误诊断。为提高故障诊断与定位的准确性，本文提出一种基于直流母线电压和z_1-z_2坐标系电流的单相单管开路和单相双管开路故障诊断与定位方法。

单相双管开路故障下，绕组功率变化量的基频分量为0，主要谐波分量为二次谐波分量，而单相单管开路故障情况下，主要谐波分量为基波分量，且其幅值不低于二次谐波分量幅值的5/6倍。因此，当检测到单相双管或单相单管开路故障后，可以利用直流母线电压中基波和二次谐波幅值关系来区分2类故障。本文方法分为3步：第1步检测单相单管或单相双管开路故障的发生，第2步判断故障类型为单相双管开路故障还是单相单管开路故障，第3步定位故障功率管或故障相位置。

由于发电系统中各种非线性因素，如电机磁饱和及PWM死区等，DTPPMA-SynRSG正常运行电流中包含了显著的5、7次谐波电流分量[20]。5、7次谐波电流会映射到z_1-z_2坐标系，导致$\boldsymbol{i}_{z_1z_2}\neq\boldsymbol{0}$。5、7次谐波电流对应的$\boldsymbol{i}_{z_1z_2}$为交变的电流矢量，将会影响对$\boldsymbol{i}_{z_1z_2}$矢量幅值与斜率的判断。图8中绿色轨迹为正常情况下，某幅值和角度下的5、7次谐波电流对应的$\boldsymbol{i}_{z_1z_2}$轨迹。

因此，在利用$\boldsymbol{i}_{z_1z_2}$进行单相单管开路和单相双管开路故障的诊断和定位时，应先抑制$\boldsymbol{i}_{z_1z_2}$中的5、7次谐波分量。考虑到5次和7次谐波分别形成正向5次和反向7次的旋转矢量，可先对$\boldsymbol{i}_{z_1z_2}$进行Park变换，将电流中的5、7次谐波分量变换为正向6次和反向6次的谐波，再用带阻滤波器（band stop filter, BSF）对6次谐波进行滤波，最终将滤波结果进行反Park变换，从而得到抑制了5次和7次谐波电流的z_1-z_2坐标系电流，所提5、7次谐波电流滤波方法如图9所示。

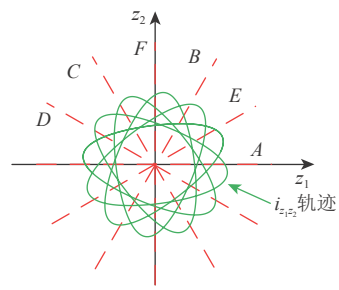


图8 5、7次谐波电流对应的z_1-z_2坐标系电流轨迹

Fig. 8 Current trajectories in the z_1-z_2 coordinate system corresponding to the 5th and 7th harmonic currents

图9 5、7次谐波电流滤波方法

Fig. 9 The 5th and 7th harmonic current filtering method

根据前文分析，故障发生后，$\boldsymbol{i}_{z_1z_2}$的幅值将增大，因此，故障发生的判断条件设置为滤波后的z_1-z_2坐标系电流幅值大于某一阈值：

$$\left|\boldsymbol{i}_{z_1z_2_\text{F}}\right|>I_{z_1z_2_\text{TH}}$$

确定故障发生后，计算直流母线电压中的基频分量幅值U_{DCH1}和二倍频率分量幅值U_{DCH2}，若$U_{\text{DCH1}}<5U_{\text{DCH2}}/3$，则认为故障类型为单相双管开路故障，若$U_{\text{DCH1}}>5U_{\text{DCH2}}/3$，则认为故障类型为单相单管开路故障。考虑到单相单管开路故障的功率脉动基波幅值，即使在最极端的情况下，即功率因数为0时，仍然不低于二次谐波幅值的5/6倍。本文所研究的DTPPMA-SynRSG额定工作阶段的功率因数为0.8，且功率脉动引起的电压脉动幅值反比于功率脉动的频率，因此，在计算U_{DCH2}时乘以系数5/3。

定位具体单相单管或单相双管开路故障位置时，依靠$\boldsymbol{i}_{z_1z_2_\text{F}}$轨迹取向算法，在每个采样周期内分别对$\boldsymbol{i}_{z_1z_2_\text{F}}$进行角度为30°、60°、90°、120°和150°的Park变换，即

$$\begin{cases}\boldsymbol{i}_{z_1z_2_\text{F}_30^\circ}=\boldsymbol{P}(30^\circ)\boldsymbol{i}_{z_1z_2_\text{F}}\\ \boldsymbol{i}_{z_1z_2_\text{F}_60^\circ}=\boldsymbol{P}(60^\circ)\boldsymbol{i}_{z_1z_2_\text{F}}\\ \boldsymbol{i}_{z_1z_2_\text{F}_90^\circ}=\boldsymbol{P}(90^\circ)\cdot\boldsymbol{i}_{z_1z_2_\text{F}}\\ \boldsymbol{i}_{z_1z_2_\text{F}_120^\circ}=\boldsymbol{P}(120^\circ)\boldsymbol{i}_{z_1z_2_\text{F}}\\ \boldsymbol{i}_{z_1z_2_\text{F}_150^\circ}=\boldsymbol{P}(150^\circ)\boldsymbol{i}_{z_1z_2_\text{F}}\end{cases} \tag{28}$$

分别对每个采样周期内的$i_{z_1_F}$、$i_{z_1_F_30°}$、$i_{z_1_F_60°}$、$i_{z_1_F_90°}$、$i_{z_1_F_120°}$、$i_{z_1_F_150°}$及$|i_{z_1_F}|$、$|i_{z_1_F_30°}|$、$|i_{z_1_F_60°}|$、$|i_{z_1_F_90°}|$、$|i_{z_1_F_120°}|$、$|i_{z_1_F_150°}|$进行累加求和计算。根据故障类型判断结果, 对上述累加求和结果进行大小比较, 进而确定出单相单管或单相双管开路故障位置。单相双管开路故障和单相单管开路故障的定位方法分别如表 1 和表 2 所示。

表 1　单相双管开路故障定位

Table 1　Single-phase double-tube open circuit fault location

最大值	单相双管开路故障位置
$\sum\|i_{z_1_F}\|$	A
$\sum\|i_{z_1_F_30°}\|$	E
$\sum\|i_{z_1_F_60°}\|$	B
$\sum\|i_{z_1_F_90°}\|$	F
$\sum\|i_{z_1_F_120°}\|$	C
$\sum\|i_{z_1_F_150°}\|$	D

表 2　单相单管开路故障定位

Table 2　Single-phase single-tube open circuit fault location

最大值	单管开路故障位置
$\sum i_{z_1_F}$	A_{L}
$\sum i_{z_1_F_30°}$	E_{L}
$\sum i_{z_1_F_60°}$	B_{U}
$\sum i_{z_1_F_90°}$	F_{U}
$\sum i_{z_1_F_120°}$	C_{L}
$\sum i_{z_1_F_150°}$	D_{L}
$-\sum i_{z_1_F}$	A_{U}
$-\sum i_{z_1_F_30°}$	E_{U}
$-\sum i_{z_1_F_60°}$	B_{L}
$-\sum i_{z_1_F_90°}$	F_{L}
$-\sum i_{z_1_F_120°}$	C_{U}
$-\sum i_{z_1_F_150°}$	D_{U}

4　实验结果及分析

为验证本文所提电源变换器故障诊断方法, 搭建了如图 10 所示实验平台进行实验验证。实验平台包括原动机、DTPPMA-SynRSG、电源变换器、负载箱、上位机等。

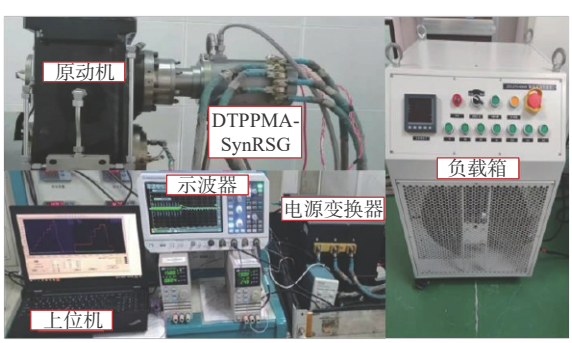


图 10　实验平台

Fig. 10　Experimental platform

为验证使用滤波后的z_1-z_2坐标系电流$\boldsymbol{i}_{z_1z_2_F}$来定位单相双管开路故障与单相单管开路故障具有更优的性能, 进行如下实验。起动发电系统工作在发电状态下, 电机转速为 12 000 r/min, 负载为 20 kW。为模拟单相双管开路故障, 运行时人为地将某相桥臂的上、下管驱动信号置零; 为模拟单相单管开路故障, 运行时人为地将某相桥臂的上管或下管驱动信号置零。本文所提 5、7 次谐波滤波方法中的带阻滤波器基于式(29)进行设计:

$$y(n)=\sum_{i=0}^{N}\lambda_i x(n-i)-\sum_{i=1}^{M}\eta_i y(n-i)\qquad n\geqslant 0 \qquad (29)$$

式中: $y(n)$为输出序列; $x(n)$为输入序列; η_i和λ_i分别为反馈滤波系数和前向滤波系数; M和N分别为反馈滤波阶数和前向滤波阶数。该型滤波器采用 4 阶 IIR 结构, 下限频率为 2 200 Hz, 上限频率为 2 600 Hz, 滤波器幅频特性曲线如图 11 所示。

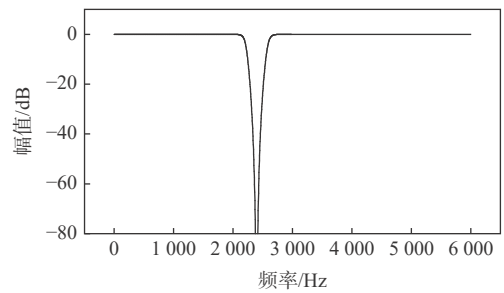


图 11　滤波器幅频特性曲线

Fig. 11　Amplitude-frequency characteristic curve of filter

各单相双管开路故障情况下滤波前z_1-z_2坐标系电流$\boldsymbol{i}_{z_1z_2}$轨迹示于图 12。可见, 由于$\boldsymbol{i}_{z_1z_2}$中包含幅值较高的 5、7 次谐波电流, $\boldsymbol{i}_{z_1z_2}$轨迹较为弯曲(扭曲程度较高)。由于轨迹弯曲, $\boldsymbol{i}_{z_1z_2}$在一些时刻偏离理论故障轨迹甚至与其他故障轨迹相交, 这可能导致单相双管开路故障定位错误。

图 13 为各相双管开路故障情况下采用 5、7 次谐波滤波后的$\boldsymbol{i}_{z_1z_2_F}$电流轨迹。可见, 在各相双管开路故障情况下, $\boldsymbol{i}_{z_1z_2_F}$的轨迹线性度提高, 说明$\boldsymbol{i}_{z_1z_2_F}$中的 5、7 次谐波分量得到有效抑制。此外, $\boldsymbol{i}_{z_1z_2_F}$的轨迹与理论故障轨迹一致度更好, $\boldsymbol{i}_{z_1z_2_F}$始终在理论故障轨迹附近, 说明 5、7 次谐波滤波器并没有使$\boldsymbol{i}_{z_1z_2}$中的有效故障信号产生明显的畸变或相位滞后。因此, 采用本文方法滤波后的z_1-z_2坐标系电流可用于单相双管开路故障的定位。

单相单管开路故障情况下, z_1-z_2坐标系电流轨迹示于图 14。考虑到三相系统的对称性, 图 14 中只给出了 A 相与 D 相桥臂上、下功率管开路故障情况下的电流轨迹。由图 14 可见, 经 5、7 次谐波电流滤波后, z_1-z_2坐标系电流轨迹扭曲程度降低,

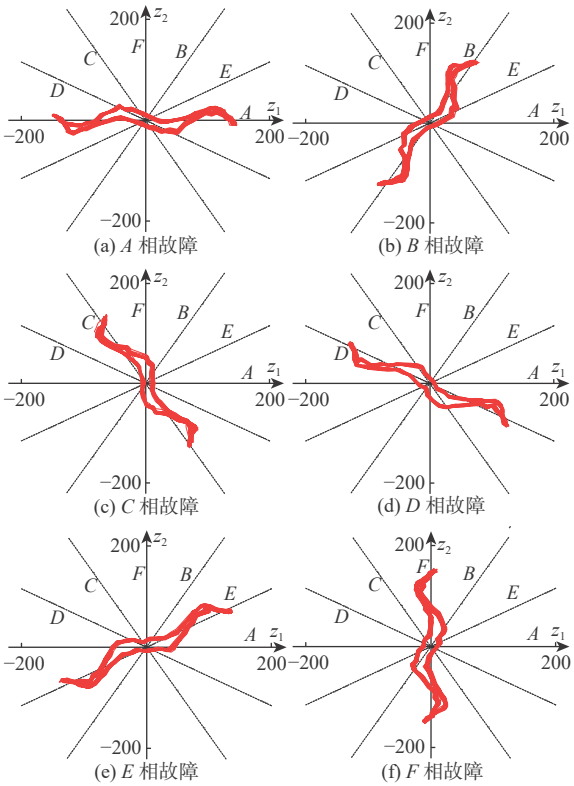


图 12 滤波前的单相双管开路故障下$i_{z_1z_2}$电流轨迹

Fig. 12 Current trajectories of $i_{z_1z_2}$ under single-phase double-tube open circuit fault before filtering

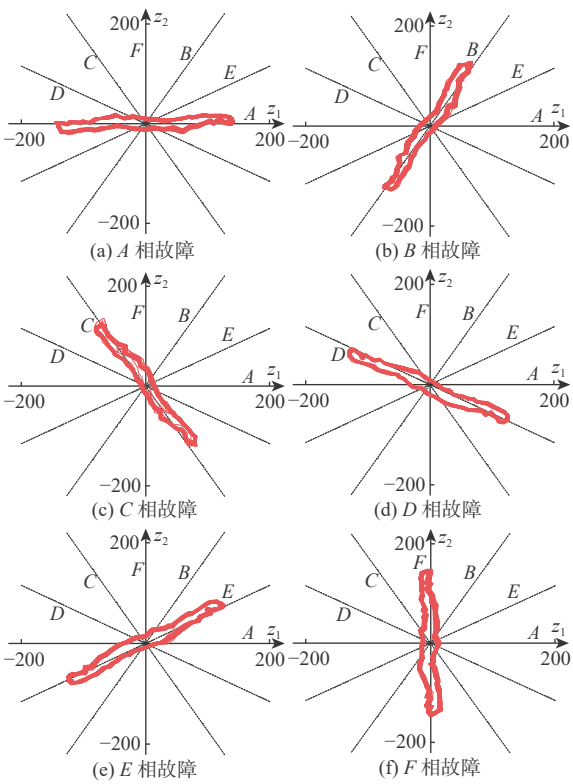


图 13 滤波后的单相双管开路故障下$i_{z_1z_2}$电流轨迹

Fig. 13 Current trace of $i_{z_1z_2}$ under single-phase double-tube open circuit fault after filtering

与理论故障轨迹的一致度较高，说明所提滤波方法有效降低了 z_1-z_2 坐标系电流中的 5、7 次谐波分量，

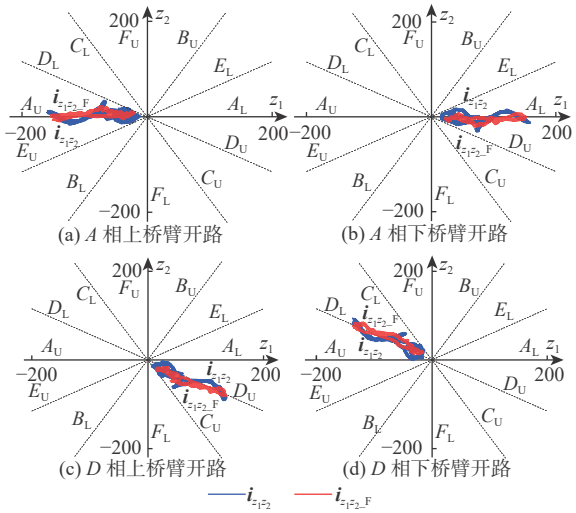


图 14 单相单管开路故障情况下$i_{z_1z_2}$与$i_{z_1z_2_F}$电流轨迹

Fig. 14 Current trajectory of $i_{z_1z_2}$ and $i_{z_1z_2_F}$ under single-phase single-tube open circuit fault

且并未使有效故障信号发生明显的畸变或相位滞后。因此，采用本文方法滤波后的 z_1-z_2 坐标系电流可用于单相单管开路故障的定位。

为验证所提基于直流母线电压和 z_1-z_2 坐标系电流的单相单管和单相双管开路故障诊断与定位方法的有效性，在上文所示实验条件下分别进行 A 相双管开路故障和 A 相上桥臂功率管开路故障实验，采集实验数据并使用前文所述故障诊断算法进行数据处理。图 15 为单相双管和单相单管开路故障下的直流母线电压波形。在单相单管开路故障和单相双管开路故障下，直流母线电压都产生了显著的脉动，而单相双管开路故障情况下电压脉动频率比开路故障频率更高。通过快速傅里叶变换分析可知，单相双管开路故障下直流母线电压基波频率分量幅值 U_{DCH1} 为 0.063 V，二倍基频分量幅值 U_{DCH2} 为 1.308 V；而单相单管开路故障下直流母线电压基波频率分量幅值 U_{DCH1} 为 1.295 V，二倍基频分量幅值 U_{DCH2} 为 0.595 V。因此，可以用 U_{DCH1} 与 $5U_{\mathrm{DCH2}}/3$ 的大小关系作为单相双管开路故障与单相

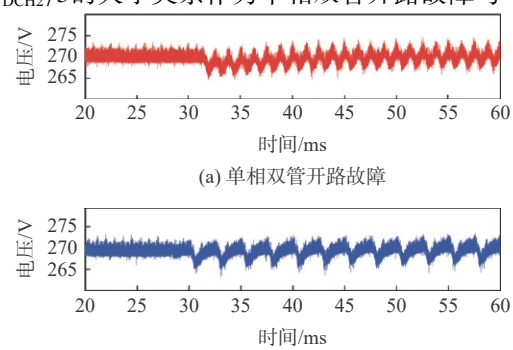


图 15 单相双管与单相单管开路故障下的直流母线电压

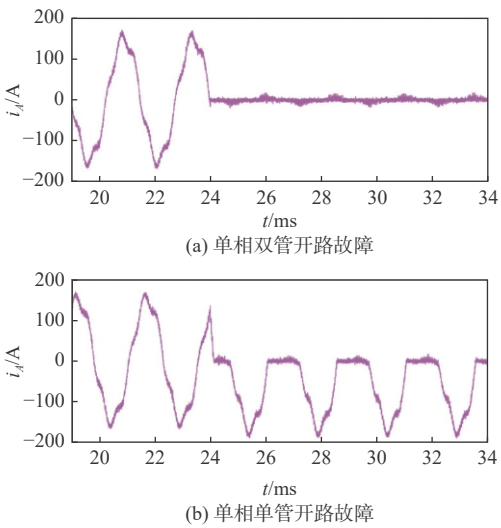
Fig. 15 DC bus voltage under single-phase double-tube and single-phase single-tube open circuit fault

单管开路故障的判断依据。

图 16 为单相双管开路故障与单相单管开路故障下的故障相电流波形。由图 16(a)可见，单相双管开路故障下相电流几乎保持为 0 A，仅在少许时刻存在较小的非 0 A 电流。相电流无法始终保持为 0 A 的原因为：由于桥臂功率管反并联二极管仍具有导通能力，当相电势低于直流负母线电势或高于直流正母线电势时，仍会产生相电流。由图 16(b)可见，当 A 相上桥臂功率管开路故障后，A 相电流波形出现上半部削波现象，而 A 相电流波形下半部仍相对完整。综上，本文对单相单管开路故障下的相电流波形函数所作的假设是合理的。

图 17 和图 18 为在诊断故障类型后，利用 z_1-z_2 坐标系电流累加值作为指示值来定位单相双管开路故障位置与单相单管开路故障位置的波形图。图 17 和图 18 中，各故障位置与 z_1-z_2 坐标系电流累加值之间的对应关系示于表 1 和表 2 中。

由图 17 可见，当发生 A 相双管开路故障后，各相故障指示值都开始增加，而 A 相故障指示值增长



(a) 单相双管开路故障

(b) 单相单管开路故障

图 16　单相双管与单相单管开路故障下的故障相电流波形

Fig. 16　Fault phase current waveforms under single-phase double-tube and single-phase single-tube open-circuit faults

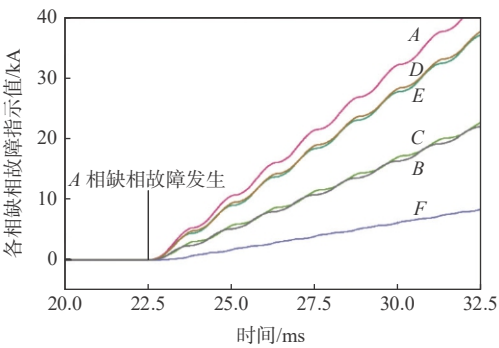


图 17　A 相双管开路故障情况下各相缺相故障指示值

Fig. 17　Phase loss fault indication values of each phase in case of double-tube open circuit fault in phase A

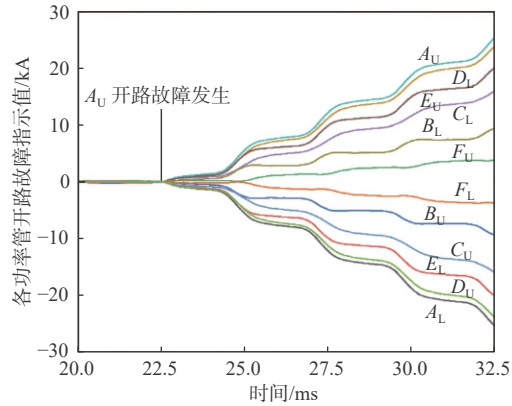


图 18　A 相上桥臂功率管开路故障情况下各功率管开路故障指示值

Fig. 18　Open circuit fault indication values of each power tube when upper bridge arm power tube of phase A has an open circuit fault

速度最快，因此，可以通过各相故障指示值的最大值来定位单相双管开路故障位置。由图 18 可见，当发生 A 相上桥臂功率管开路故障后，各相故障指示值都开始增加或减少，而 A_U 相故障指示值增长速度最快，因此，可以通过各相故障指示值的最大值来定位功率管开路故障位置。

5　结　论

本文提出了一种基于 z_1-z_2 坐标系电流和直流母线电压的双三相永磁辅助式同步磁阻起动发电系统电源变换器的故障诊断方法，通过系统模型、故障特性、诊断与定位方法方面的研究，解决了起动发电系统在起动和发电阶段的电源变换器功率管开路故障诊断问题。

1）建立了双三相永磁辅助式同步磁阻电机与全控式全桥电源变换器数学模型，分别分析了电源变换器单相双管开路和单相单管开路故障下对直流母线电压和 z_1-z_2 坐标系电流的影响，并提出以直流母线电压和 z_1-z_2 坐标系电流作为故障诊断特征。

2）提出了基于直流母线电压和 z_1-z_2 坐标系电流特征的单相开路故障诊断与定位方法，并搭建了实验平台，验证了所提方法能够有效解决电源变换器单相单管与单相双管开路故障的诊断与定位问题，为双三相永磁辅助式同步磁阻起动发电系统电源变换器功率管开路故障诊断问题提供了解决途径。

参考文献（References）

[1] 邓景辉. 高速直升机关键技术与发展[J]. 航空学报, 2024, 45(9): 529085.
DENG J H. Key technologies and development for high-speed helicopters[J]. Acta Aeronautica et Astronautica Sinica, 2024, 45(9):

529085(in Chinese).

[2] 戴卫力, 王慧贞, 严仰光, 等. 航空起动/发电系统的发展趋势与研究现状[J]. 航空科学技术, 2010, 21(5): 28-32.
DAI W L, WANG H Z, YAN Y G, et al. Development trend and current situation of starter-generator for aircraft engine[J]. Aeronautical Science and Technology, 2010, 21(5): 28-32(in Chinese).

[3] BHANGU B S, RAJASHEKARA K. Electric starter generators: their integration into gas turbine engines[J]. IEEE Industry Applications Magazine, 2013, 20(2): 14-22.

[4] CAMPOS-DELGADO D U, ESPINOZA-TREJO D R. An observer-based diagnosis scheme for single and simultaneous open-switch faults in induction motor drives[J]. IEEE Transactions on Industrial Electronics, 2011, 58(2): 671-679.

[5] AN Q T, SUN L, SUN L Z. Current residual vector-based open-switch fault diagnosis of inverters in PMSM drive systems[J]. IEEE Transactions on Power Electronics, 2015, 30(5): 2814-2827.

[6] CAMPOS-DELGADO D U, PECINA-SÁNCHEZ J A, ESPINOZA-TREJO D R, et al. Diagnosis of open-switch faults in variable speed drives by stator current analysis and pattern recognition[J]. IET Electric Power Applications, 2013, 7(6): 509-522.

[7] SHI T C, HE Y G, WANG T, et al. An improved open-switch fault diagnosis technique of a PWM voltage source rectifier based on current distortion[J]. IEEE Transactions on Power Electronics, 2019, 34(12): 12212-12225.

[8] JI Z D, LIU W. Open-circuit fault detection for three-phase inverter based on backpropagation neural network[J]. Neural Computing and Applications, 2019, 31(9): 4665-4674.

[9] XU L, CAO M Y, SONG B Y, et al. Open-circuit fault diagnosis of power rectifier using sparse autoencoder based deep neural network[J]. Neurocomputing, 2018, 311: 1-10.

[10] PANG J, LIU W G, WEI Z H, et al. Online diode fault detection in rotating rectifier of the brushless synchronous starter generator [J]. IEEE Transactions on Industrial Informatics, 2020, 16(11): 6943-6951.

[11] JIAO N F, HAN X, WEI Z H, et al. Online fault diagnosis for rotating rectifier in wound-rotor synchronous starter-generator based on geometric features of current trajectory[J]. IEEE Transactions on Industrial Electronics, 2021, 68(4): 2952-2963.

[12] 张竞文, 熊立新, 马宏昌, 等. 基于 VMD-MPE 的开关磁阻电机功率变换器故障诊断[J]. 北京航空航天大学学报, 2022, 48(6): 1022-1029.
ZHANG J W, XIONG L X, MA H C, et al. Fault diagnosis of switched reluctance motor power converter based on VMD-MPE[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(6): 1022-1029(in Chinese).

[13] NAJAFABADI T A, SALMASI F R, JABEHDAR-MARALANI P. Detection and isolation of speed-, DC-link voltage-, and current-sensor faults based on an adaptive observer in induction-motor drives[J]. IEEE Transactions on Industrial Electronics, 2011, 58(5): 1662-1672.

[14] SALEHIFAR M, ARASHLOO R S, MORENO-EGUILAZ M, et al. Observer-based open transistor fault diagnosis and fault-tolerant control of five-phase permanent magnet motor drive for application in electric vehicles[J]. IET Power Electronics, 2015, 8(1): 76-87.

[15] DURAN M J, GONZALEZ-PRIETO I, RIOS-GARCIA N, et al. A simple, fast, and robust open-phase fault detection technique for six-phase induction motor drives[J]. IEEE Transactions on Power Electronics, 2017, 33(1): 547-557.

[16] ARAFAT A, CHOI S, BAEK J. Open-phase fault detection of a five-phase permanent magnet assisted synchronous reluctance motor based on symmetrical components theory[J]. IEEE Transactions on Industrial Electronics, 2017, 64(8): 6465-6474.

[17] CAI B P, ZHAO Y B, LIU H L, et al. A data-driven fault diagnosis methodology in three-phase inverters for PMSM drive systems[J]. IEEE Transactions on Power Electronics, 2017, 32(7): 5590-5600.

[18] XIA Y, XU Y, GOU B. A data-driven method for IGBT open-circuit fault diagnosis based on hybrid ensemble learning and sliding-window classification[J]. IEEE Transactions on Industrial Informatics, 2019, 16(8): 5223-5233.

[19] BIANCHI N, ALBERTI L. MMF harmonics effect on the embedded FE analytical computation of PM motors[J]. IEEE Transactions on Industry Applications, 2010, 46(2): 812-820.

[20] HU M J, HUA W, MA G T, et al. Improved current dynamics of proportional-integral-resonant controller for a dual three-phase FSPM machine[J]. IEEE Transactions on Industrial Electronics, 2021, 68(12): 11719-11730.

Open-circuit fault diagnosis of power tubes in power converter of dual three-phase starter-generator system

WANG Huaji[1]，LIU Yu[2]，SUN Gaoyang[3]，LI Dehong[1]，XU Jinquan[2,*]

(1. China Helicopter Research and Development Institute，Jingdezhen 333001，China；
2. School of Automation Science and Electrical Engineering，Beihang University，Beijing 100191，China；
3. Suzhou Inovance United Power System Co.，Ltd.，Suzhou 215000，China)

Abstract：The aerospace dual three-phase permanent magnet assisted synchronous reluctance starter-generator system's power converter had an open-circuit fault diagnosis issue. To address this, a single-phase open-circuit fault diagnosis and localization method for power tubes based on the bus voltage and z_1-z_2 coordinate system current was proposed. Firstly, the starter-generator and power converter model were established by using the vector space decomposition method. Next, using this model, the single-phase double-tube open-circuit and single-phase single-tube open-circuit fault characteristics represented in the power converter's DC bus voltage and z_1-z_2 coordinate system current trajectory were examined. A single-phase single-tube open circuit and single-phase double-tube open circuit fault diagnosis and location method based on z_1-z_2 coordinate system current and DC bus voltage was proposed to detect the occurrence of faults and determine the fault location. The experimental platform ultimately confirmed the method's efficacy. The experimental findings show that the technique is capable of identifying and diagnosing power converter single-phase power tube open-circuit failures. This in turn provides important guidance for fault-tolerant operation after the fault occurs.

Keywords：fault diagnosis；power tube open circuit；starter-generator；power converter；harmonic current

Received：2025-07-01；**Accepted**：2025-07-11；**Published Online**：2025-07-16 15：58
URL：link.cnki.net/urlid/11.2625.V.20250716.1454.001

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0437

基于非均匀大气光修正模型的夜间图像去雾

林森*，查子月
(沈阳理工大学 自动化与电气工程学院，沈阳 110159)

摘　　要：针对夜间有雾图像受人工光源影响，存在亮度低、辉光及色偏问题，提出一种基于非均匀大气光修正模型的夜间图像去雾算法。结合非均匀大气光成像特性，通过加入表面照射光和辉光项修正传统大气散射模型；根据图像亮度和饱和度分量引入近光源系数，准确估计辉光项以校正非均匀大气光；基于暗通道先验(DCP)及带色彩恢复的多尺度Retinex增强算法求解模型，使用引导滤波解决透射率块状效应，采用线性融合方法解决过度增强问题；利用高斯高通滤波增强图像细节，得到最终的清晰图像。实验结果表明：所提算法恢复的夜间图像颜色自然、细节丰富且可见度高；相比暗通道先验等经典算法，以及联合对比度增强和曝光融合(CEEF)等新颖算法，色差降低约15%，峰值信噪比提升约15%，平均梯度提升约8%，在定性和定量实验中更具优势。

关　键　词：夜间图像去雾；大气散射模型；大气光估计；辉光；线性融合
中图分类号：TP391.41
文献标志码：A　　**文章编号**：1001-5965(2025)09-2894-12

雾天环境下，空气中大量悬浮颗粒使光线在传播过程中被吸收或散射，导致成像设备采集的图像质量发生退化，从而对依赖清晰图像的计算机视觉任务，如航空导航、户外监控及目标识别等产生严重影响[1-2]。此外，夜间自然光线弱及人工光源的影响使图像存在细节不清晰、色偏严重、光照不均等现象，相较于白天图像，夜间图像去雾更具挑战性。

目前，去雾算法主要应用于白天场景，如基于直方图均衡化[3]、Retinex[4]和小波变换[5]等图像增强算法，通过加强图像特征，实现视觉效果提升，但并未解决雾化问题。Liu等[6]提出联合对比度增强和曝光融合(contrast enhancement and exposure fusion, CEEF)框架，通过伽马校正和保色对比度自适应直方图均衡化(color preservation-contrast adaptive histogram equalization, CP-AHE)提高图像对比度，自适应融合保留更多细节信息；He等[7]提出暗通道先验(dark channel prior, DCP)算法，估计模型参数还原退化图像，已成为目前公认的有效模型参数估计方法，众多学者就此开展延拓性的研究[8-9]。然而，夜间复杂的照明环境使大气光不再均匀分布，图像存在对比度、亮度低及色偏严重等问题，导致上述算法难以直接应用于夜间图像去雾。

现有夜间图像去雾算法主要分为2类：基于光学模型和基于深度学习的算法。基于光学模型的算法以大气散射模型为基础，通过分析夜间光学成像特点求解相关参数。Pei等[10]利用颜色转移方法，将夜间有雾图像转为白天有雾图像，使用DCP算法去雾，提高图像可见性，但未考虑人工光源的光照特性，导致整体颜色不真实；Zhang等[11]在大气散射模型中引入光照强度，通过Retinex算法估计入射光，使用伽马校正平衡图像亮度，颜色校正后使用DCP算法去雾，有效增强对比度，但未考虑夜间大气光分布不均匀的特性，导致容易出现失真和噪声放大问题；Li等[12]建立含有光晕项的夜间雾

收稿日期：2023-07-04；录用日期：2023-12-01；网络出版时间：2024-01-15 18:35
网络出版地址：link.cnki.net/urlid/11.2625.V.20240115.1203.003
基金项目：国家重点研发计划(2018YFB1403303)；辽宁省教育厅高等学校基本科研项目(LJKMZ20220615)
*通信作者. E-mail：lin_sen6@126.com

引用格式：林森，查子月. 基于非均匀大气光修正模型的夜间图像去雾[J]. 北京航空航天大学学报，2025，51(9)：2894-2905.
LIN S，ZHA Z Y. Nighttime image dehazing based on non-uniform atmospheric light correction model[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2894-2905 (in Chinese).

天成像模型, 利用相对平滑约束分层算法去除光晕, 颜色恢复相对自然, 但未考虑人工光源低照度特性, 导致图像仍存在局部过暗现象。上述算法虽根据夜间有雾环境建立成像模型, 但未深入分析大气光分布情况, 且未考虑人工光源对大气光的影响, 导致去雾不彻底。

近年来, 深度学习技术逐渐被应用于夜间图像去雾算法中。Liao 等[13] 基于卷积神经网络提出雾密度预测网络(haze density prediction-network, HDP-Net), 估计并去除雾密度图像获得清晰结果, 在一定程度上去雾, 但未解决图像亮度低和色偏问题; Kuanar 等[14] 引入去除光晕-去雾迭代架构, 通过递归网络去除光晕, 使用 DehazeNet 模型有效去雾, 但结果图像依然存在亮度低的问题; Zhang 等[15] 提出基于最佳尺度融合去雾(optimal-scale fusion dehazing, OSFD)的夜间去雾算法, 估计变化的大气光照和透射率, 实现快速去雾, 但处理颜色单一的图像会引入伪影。虽然基于深度学习的算法具有较好的去雾效果, 但目前能达到训练标准的夜间数据集较少且质量一般, 难以满足模型训练的要求。

此外, 针对不均匀光照环境, 鲁丹等[16] 根据光源的发光特性对光照进行补偿, 但未考虑辉光和大气光分布不均匀的影响, 导致图像出现局部过亮和细节模糊等现象。学者们对非均匀大气光也进行了相关研究[17-19], 解决了图像光晕现象, 但这些研究仅分析了大气光传播过程中被吸收和散射造成的非均匀分布, 未对夜间人造光源的光照特性进行分析, 从而导致结果图像仍存在色偏和亮度低的问题。针对以上问题, 本文提出了基于非均匀大气光修正模型的夜间图像去雾算法。在大气光项中引入随场景变化的辉光项, 建立非均匀大气光修正模型; 使用低通滤波获取大气光, 并利用近光源系数获得准确的辉光图像, 实现非均匀大气光校正; 利用 DCP 和带色彩恢复的多尺度 Retinex(multi-scale Retinex with color restoration, MSRCR)[20] 增强算法实现图像去雾和颜色恢复; 使用高斯高通滤波增强图像细节。本文算法有效解决了夜间有雾图像中出现的辉光、颜色失真和细节信息缺失等问题, 增强亮度和对比度, 在视觉效果改善和质量提高方面具有显著优势。

1　非均匀大气光修正模型

目前, 简化的大气散射模型[21] 被广泛应用于图像去雾领域, 由直接衰减项和大气光项构成:

$$\boldsymbol{I}^c(x)=\boldsymbol{J}^c(x)t(x)+\boldsymbol{A}^c(1-t(x)) \tag{1}$$

式中: $c\in\{\mathrm{R,G,B}\}$为任一颜色通道; x为像素点位置; $\boldsymbol{I}^c(x)$为有雾图像; $\boldsymbol{J}^c(x)$为无雾图像; $\boldsymbol{A}^c$为全局大气光; $t(x)$为透射率; $\boldsymbol{J}^c(x)t(x)$为直接衰减项; $\boldsymbol{A}^c(1-t(x))$为大气光项。

由于夜间图像受人工光源影响, 存在大气光分布不均匀、亮度低和色偏严重问题, 大气散射模型不再适用。因此, 根据夜间雾天光学成像特性, 分析大气光分布规律, 非均匀大气光修正模型为

$$\boldsymbol{I}^c(x)=\boldsymbol{J}^c(x)\boldsymbol{L}^c(x)t(x)+\boldsymbol{A}^c(1-t(x))+\boldsymbol{G}^c(x) \tag{2}$$

式中: $\boldsymbol{J}^c(x)\boldsymbol{L}^c(x)t(x)$为修正的直接衰减项; $\boldsymbol{A}^c(1-t(x))+\boldsymbol{G}^c(x)$为非均匀大气光项; $\boldsymbol{L}^c(x)$为表面照射光; $\boldsymbol{G}^c(x)$为辉光项。

根据 Retinex 理论[4,20], 相同场景下物体所反射的光线亮度及颜色分布等均保持不变。然而, 人工光源的光照强度较小且往往只发出单一颜色的光线, 夜间成像受表面照射光的影响, 图像呈现亮度低及色偏现象, 难以获取真实的反射属性。结合 Retinex 模型, 修正的直接衰减项定义为

$$\boldsymbol{E}_0^c(x)=\boldsymbol{J}^c(x)\boldsymbol{L}^c(x)t(x) \tag{3}$$

式中: $\boldsymbol{E}_0^c(x)$为场景高频分量。

夜间大气光主要来源于人工光源, 在有雾环境下, 经过多重散射成像为一圈辉光[12]。光的衰减会导致辉光强度下降, 且辉光会和物体反射光叠加成像, 随场景和光源位置而不同(见图 1), 因此, 大气光呈现非均匀性。

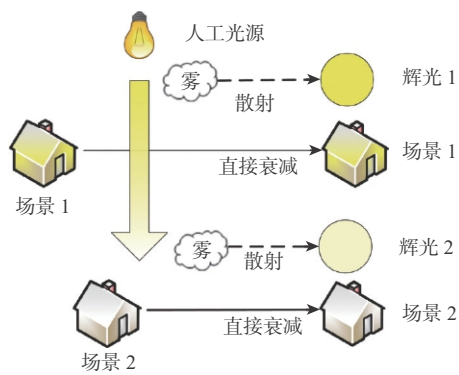


图 1　辉光成像模型

Fig. 1　Glow imaging model

加入辉光项$\boldsymbol{G}^c(x)$和近光源系数矩阵$\boldsymbol{K}$, 非均匀大气光项定义为

$$\boldsymbol{A}_{\mathrm{night}}^c(x)=\boldsymbol{A}^c(1-t(x))+\boldsymbol{G}^c(x)=\boldsymbol{A}^c(1-t(x))+\min(\boldsymbol{K}(x),\boldsymbol{O}-\boldsymbol{K}(x))\boldsymbol{A}_\phi^c(x) \tag{4}$$

式中: $\boldsymbol{A}_{\mathrm{night}}^c(x)$为大气光低频分量; $\boldsymbol{A}_\phi^c(x)$为经过高斯滤波后的粗略大气光; $\boldsymbol{O}$ 为全 1 矩阵; $\boldsymbol{K}$中元素取值范围为[0,1], 表示每个像素点辉光在大气光中的占比。距离光源越近, 受辉光影响越大, $\boldsymbol{K}(x)$值越大; 距离越远, $\boldsymbol{K}(x)$值越小。光源区域的像素点受辉光影响最大, $\boldsymbol{K}(x)$接近于 1, 为避免光源被误去除及有效

估计其他区域的辉光，进行$\min(\boldsymbol{K}(x),\boldsymbol{O}-\boldsymbol{K}(x))$运算。

如图2所示，基于简化大气散射模型的结果图像有一定的去雾效果，但未改善原始夜间有雾图像的色偏、亮度低及辉光现象，不再适用于夜间图像去雾；而本文提出的非均匀大气光修正模型充分考虑了人工光源光照特性、大气光分布规律及夜间雾天光学成像特性，弥补了简化大气散射模型应用于夜间有雾图像的不足，有效去雾的同时，能够解决亮度低、色偏及辉光问题。

(a) 原始夜间有雾图像　(b) 基于简化大气散射模型的结果图像　(c) 基于非均匀大气光修正模型的结果图像

图2　不同模型的结果对比

Fig. 2　Comparison of results from different models

2　本文算法

基于第1节的修正模型，本文算法整体流程如图3所示。首先，利用低通滤波获取粗略大气光，并根据图像亮度和饱和度分量确定近光源系数求取辉光项，校正非均匀大气光；其次，基于DCP算法进行去雾操作，并使用引导滤波解决透射率的块状效应；然后，使用MSRCR增强算法进行增强处理，为避免过度增强，与增强前的图像线性融合得到清晰图像；最后，使用图像锐化技术增强细节得到最终结果。

2.1　辉光去除

为准确获取辉光项$\boldsymbol{G}(x)$，利用高斯滤波的平滑能力处理夜间有雾图像，去除场景高频分量，保留大气光低频分量，得到粗略大气光为

$$\boldsymbol{A}_{\phi}^{c}(x)=\phi(\boldsymbol{I}^{c}(x))=\phi(\boldsymbol{E}_{0}^{c}(x)+\boldsymbol{A}_{\text{night}}^{c}(x))=\phi(\boldsymbol{A}_{\text{night}}^{c}(x)) \tag{5}$$

式中：$\phi(\cdot)$为高斯滤波器；$\phi(\boldsymbol{E}_{0}^{c}(x))\approx\boldsymbol{0}$。

根据图4展示的完整有雾、近光源区域和光源区域3类图像亮度和饱和度的概率分布直方图，确定图像亮度、饱和度和近光源系数矩阵$\boldsymbol{K}$的关系。从图4可知，像素点越靠近光源区域，其亮度越大，饱和度越小，若用亮度减去饱和度，则差值逐渐增大，与$\boldsymbol{K}$的变化规律相同。

将有雾图像转换到HSI色彩空间，分别获取亮度分量$\boldsymbol{I}_{\text{image}}$和饱和度分量$\boldsymbol{S}_{\text{image}}$，则$\boldsymbol{K}$及其归一化过程分别表示为

$$\boldsymbol{K}(x)=\boldsymbol{I}_{\text{image}}(x)-\boldsymbol{S}_{\text{image}}(x) \tag{6}$$

$$\boldsymbol{K}(x)=\frac{\boldsymbol{K}(x)-\boldsymbol{K}_{\min}}{\boldsymbol{K}_{\max}-\boldsymbol{K}_{\min}} \tag{7}$$

式中：$\boldsymbol{K}_{\max}$为最大近光源系数；$\boldsymbol{K}_{\min}$为最小近光源系数。

辉光项$\boldsymbol{G}(x)$的获取过程如图5所示，图中，⊗表示2项相乘，方框标出了光源及近光源区域的亮度、饱和度及辉光情况。观察可知，辉光从有雾图像中被分离出来，通过$\min(\boldsymbol{K},\boldsymbol{O}-\boldsymbol{K})$计算，避免了光

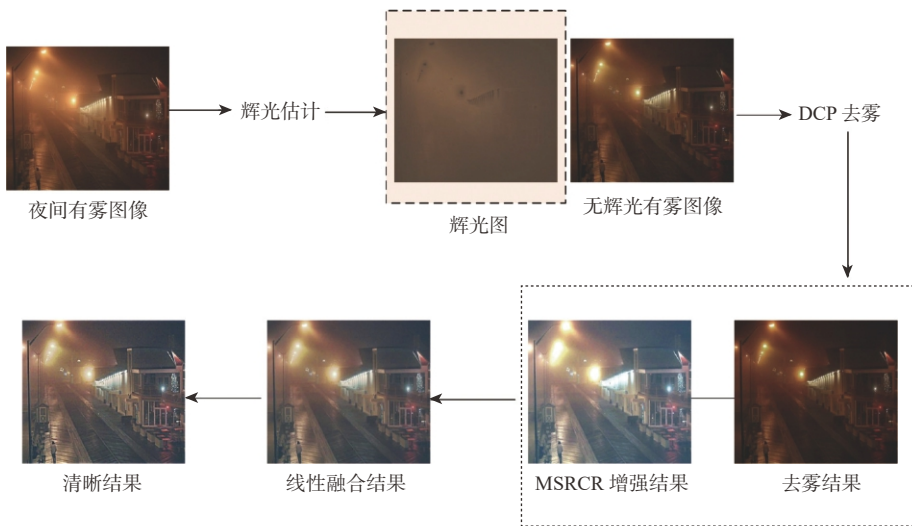


图3　本文算法流程

Fig. 3　Flow chart of the proposed algorithm

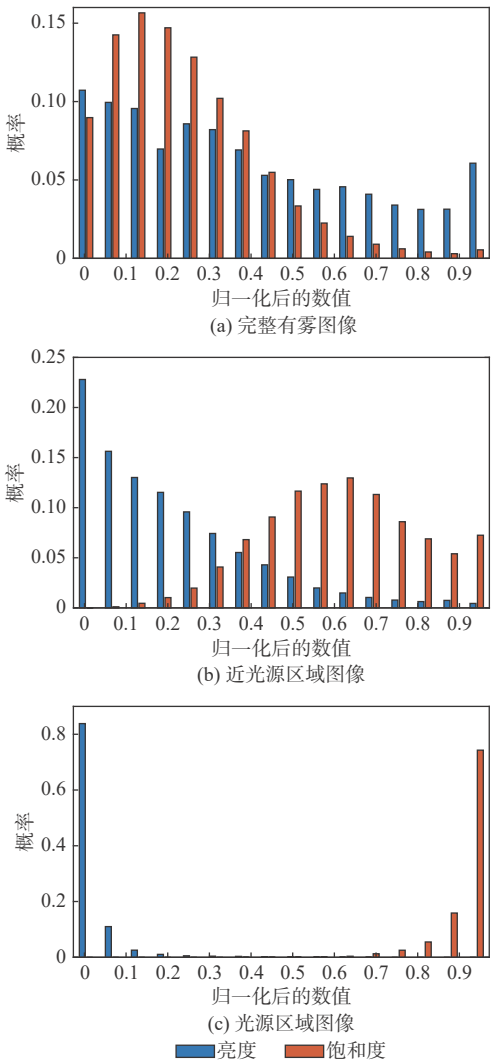


图 4　亮度和饱和度的统计数据

Fig. 4　Statistics of intensity and saturation

源被误判为辉光。通过辉光的去除来校正非均匀大气光, 具体定义为

$$\tilde{\boldsymbol{I}}^c(x)=\boldsymbol{J}^c(x)\boldsymbol{L}^c(x)t(x)+\boldsymbol{A}^c(1-t(x)) \tag{8}$$

式中: $\tilde{\boldsymbol{I}}^c(x)$为非均匀大气光校正后的夜间有雾图像。

2.2　夜间图像去雾和增强

受 DCP 算法启发, 统计发现夜间雾天环境下, 光源区域的大气光最强, 即全局大气光所在区域。选取夜间有雾图像中前 0.1% 个最亮像素点的平均值作为全局大气光值, 图 6 为大气光的估计结果。

根据 DCP 算法, 透射率为

$$t(x)=\max\left(1-\min_{z\in\Omega(x)}\left(\min_{c\in\{\mathrm{R,G,B}\}}\frac{\tilde{\boldsymbol{I}}^c(z)}{\boldsymbol{A}^c}\right),0.1\right) \tag{9}$$

式中: $\Omega(x)$表示以 x 为中心的方形窗口; $\min\limits_{c\in\{\mathrm{R,G,B}\}}$ 表示图像三通道中的最小值; $\min\limits_{z\in\Omega(x)}$表示以 x 为中心的窗口进行最小值滤波, 为避免透射率过小而设置最小值 0.1[7], 若最小值设置过大, 透射率估计值会偏大, 导致结果图像失真, 若最小值设置过小, 会导致算法去雾效果不佳。使用引导滤波解决透射率的块状效应, 根据文献 [22] 将参数设置为: 滤波窗口 $r=15$, 正则化参数 $e=0.001$。若 r 过大, 去雾效果过于强烈, 使得图像细节和纹理被过度平滑, 导致图像不自然; 若 r 过小, 会导致去雾效果不明显。若 e 设置过小, 算法更注重细节保留, 导致去雾效果不明显; 若 e 设置过大, 算法更注重平滑图像, 导致图像细节损失。

将估计出的透射率 t 和全局大气光 $\boldsymbol{A}$ 代入式 (8) 中, 无雾图像为

$$\boldsymbol{J}^c(x)\boldsymbol{L}^c(x)=\frac{\tilde{\boldsymbol{I}}^c(x)-\boldsymbol{A}^c(1-t(x))}{t(x)} \tag{10}$$

为解决图像亮度低和色偏问题, 去除照射分量 $\boldsymbol{L}$ 的影响, MSRCR 增强算法[20] 通过加入色彩恢复因子 $\boldsymbol{C}$, 调节原始图像中 3 个颜色通道的比例关系(见式 (11)), 凸显较暗区域的信息。如图 7(c) 所示, 图像对比度和亮度提高, 颜色更加自然真实。

$$\boldsymbol{J}^c(x)=Q\left\{\boldsymbol{C}^c(x)\left[\frac{1}{n}\sum_n\left(\ln\boldsymbol{J}^c_{\mathrm{night}}(x)\right)-\ln\left(\boldsymbol{J}^c_{\mathrm{night}}(x)*F_n(x)\right)\right]+b\right\} \tag{11}$$

式中: Q 为增益常数; $\boldsymbol{C}$ 为色彩恢复因子; n 为尺度个数; $\boldsymbol{J}_{\mathrm{night}}$ 为原始夜间图像; $F(x)$为中心环绕函数; 偏

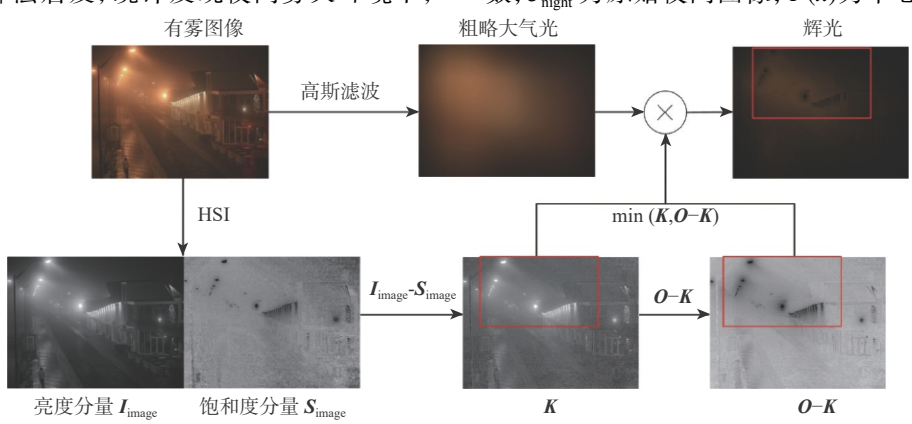


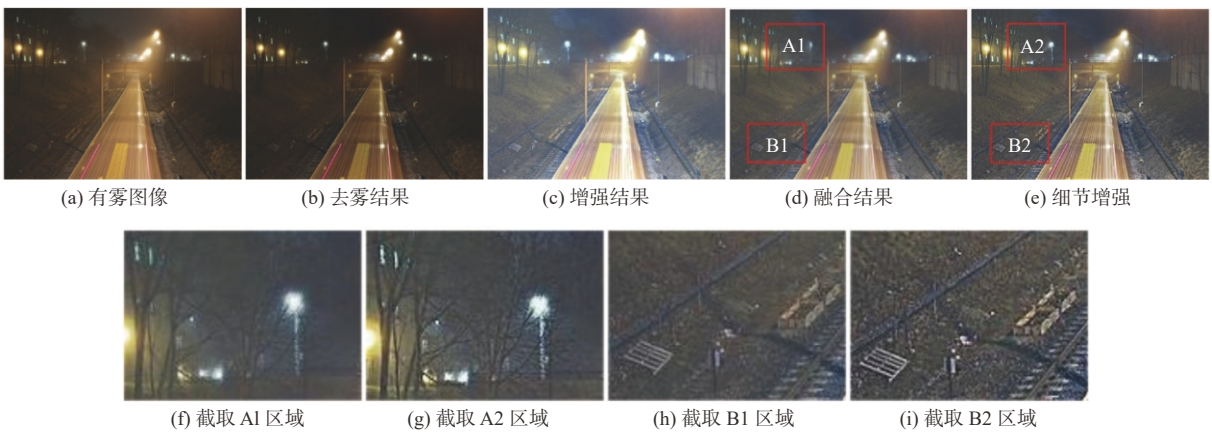
图 5　获取辉光流程

Fig. 5　Process of obtaining glow

图6 全局大气光所在区域

Fig. 6 Area of global atmospheric light

差b用来修正图像，所有参数值均参考文献[20]。Q控制算法的整体增益，影响图像的亮度，若Q设置较小，图像偏暗，反之图像过亮。$\boldsymbol{C}$解决色偏问题，若$\boldsymbol{C}$设置较小，无法有效纠正色偏，反之，会导致颜色失真。n影响图像的细节保留和模糊程度，若n较小，图像较为模糊，若n较大，图像存在噪声。b影响图像的亮度和对比度，若b较小，无法有效修正图像，若b较大，会导致图像过亮或过暗。



(a) 有雾图像 (b) 去雾结果 (c) 增强结果 (d) 融合结果 (e) 细节增强

(f) 截取A1区域 (g) 截取A2区域 (h) 截取B1区域 (i) 截取B2区域

图7 各阶段图像结果

Fig. 7 Image results of each stage

MSRCR增强算法改善了图像色偏问题，但亮度过度增强，整体泛白，视觉观感较差。将增强前后的图像进行线性融合：

$$\tilde{\boldsymbol{J}}^c(x)=\alpha \boldsymbol{J}^c(x)\boldsymbol{L}^c(x)+\beta \boldsymbol{J}^c(x) \tag{12}$$

式中：$\tilde{\boldsymbol{J}}$为融合后的结果；α和β为权重系数。权重系数用来平衡算法的色偏改善和亮度增强效果，通过实验对比，当参数α和β分别设置为0.4和0.6时，效果最佳。结果如图7(d)所示，图像亮度自然，整体视觉观感更好，但细节信息有待进一步增强。

本文采用高斯高通滤波方法（见式(13)）处理融合后的结果，如图7(f)～(i)对比可知，图像细节信息更加丰富：

$$\boldsymbol{J}_{\text{last}}^c(x)=\tilde{\boldsymbol{J}}^c(x)+\gamma\left(\tilde{\boldsymbol{J}}^c(x)-\phi\left(\tilde{\boldsymbol{J}}^c(x)\right)\right) \tag{13}$$

式中：$\boldsymbol{J}_{\text{last}}^c(x)$为经高斯高通滤波处理后的融合结果；$\gamma$为锐化系数，用来调整图像锐化程度，根据实验结果，将其设置为1。

3 实验结果与分析

为验证本文模型和对应算法的有效性，挑选3个数据集中所有夜间有雾图像进行实验，分别为：包含真实和人工合成有雾图像的数据集3R[15]、人工合成数据集Nighttime with Haze 4K[23]和真实有雾图像数据集RESIDE[24]。与5种去雾算法进行对比，分别为He2011[7]、CEEF2022[6]、Zhang2014[11]、Li2015[12]、OSFD2020[15]，通过主观分析和客观指标评价算法效果。

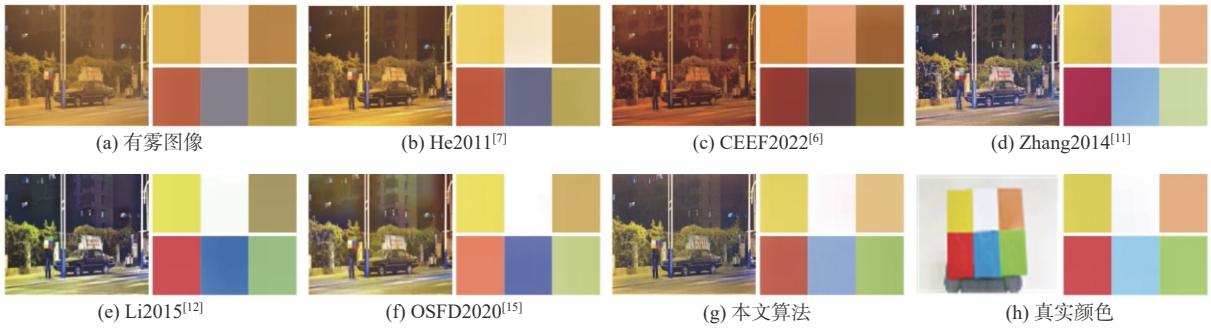
3.1 主观评价

为验证算法颜色校正能力，基于一张真实颜色图像[11]及标准色卡进行颜色恢复实验研究。为避免噪声影响，截取图像中色卡各颜色平滑部分进行对比，结果如图8所示。

观察可知，CEEF2022算法处理后颜色失真更严重，He2011算法未做颜色校正处理，恢复效果较差；Zhang2014算法能较好地恢复褐色，但图像整体偏亮偏红，其他颜色都有不同程度的失真；Li2015算法结果图中白、红和绿色恢复较好，但颜色过饱和导致其他颜色偏差较大；OSFD2020算法能较好地恢复黄色和白色，但绿色、红色和褐色偏暗；本文算法仅有蓝色偏差稍大，其他颜色都接近真实颜色，有着优秀的视觉效果。

为验证算法的去雾效果，部分人工合成和真实夜间有雾图像去雾结果分别如图9和图10所示。

观察图9可知，He2011算法去雾效果较好，但图像暗区域处理后变得更暗，可见度不高；CEEF-2022算法和OSFD2020算法去雾不彻底，部分图像仍有残雾，如图9中第2行和第3行图像所示；Zhang-2014算法去雾后出现失真和伪影问题，如图9中第1行图像颜色失真；Li2015算法去雾效果差且图像亮度过大；本文算法在有效去雾的同时增强亮

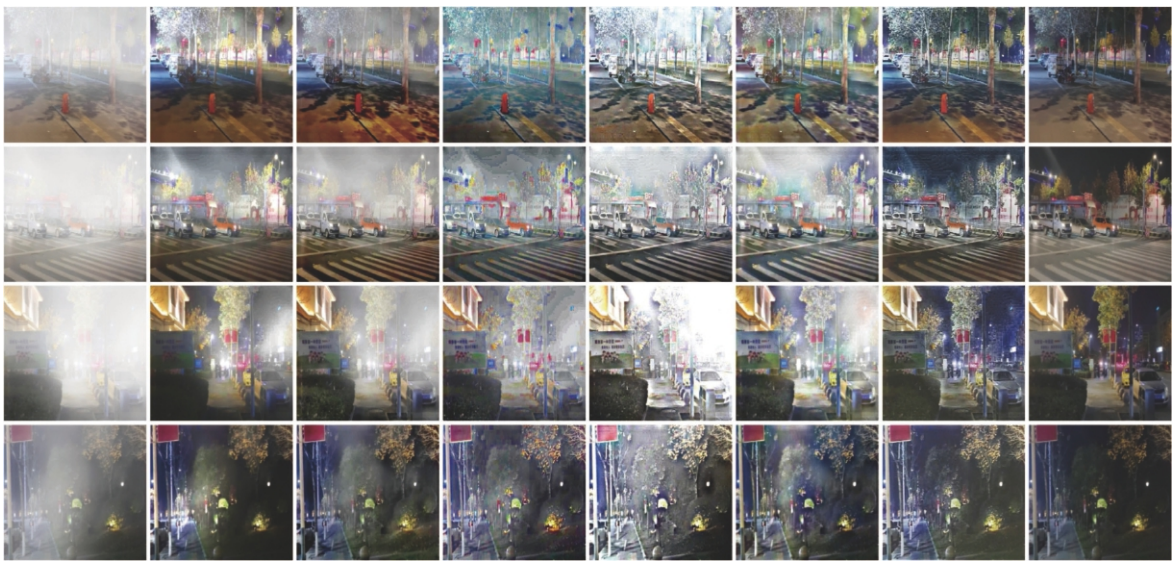


(a) 有雾图像　(b) He2011[7]　(c) CEEF2022[6]　(d) Zhang2014[11]

(e) Li2015[12]　(f) OSFD2020[15]　(g) 本文算法　(h) 真实颜色

图 8　颜色恢复实验结果

Fig. 8　Results of color restoration experiment



(a) 人工合成有雾图像　(b) He2011[7]　(c) CEEF2022[6]　(d) Zhang2014[11]　(e) Li2015[12]　(f) OSFD2020[15]　(g) 本文算法　(h) 去雾图像

图 9　人工合成夜间有雾图像去雾结果对比

Fig. 9　Comparison of dehazing results of synthetic nighttime hazy images

度, 图像具备更好的视觉感受。

由图 10 可知, He2011 算法去雾效果一般, 未能有效解决图像亮度低和色偏问题; CEEF2022 算法结果图像虽细节明显, 但整体色偏严重; Zhang-2014 算法结果图整体偏亮, 视觉效果较好, 但图 10 中第 1 行和第 2 行图像出现过饱和现象, 第 4 行图像存在颜色失真问题; Li2015 算法结果图像对比度较高, 但存在明亮区域更亮及暗区域更暗的问题, 导致细节丢失, 特别处理颜色较单一的图像时, 会出现严重色偏现象, 如图 10 中第 5 行图像所示; OSFD2020 算法去雾效果较好, 一定程度上也改善了色偏, 但结果图像仍存在亮度低、细节不明显等问题, 如图 10 中第 1 行和第 3 行图像所示; 本文算法很好地平衡了图像的亮度和对比度, 有效去除图 10 中第 3 行、第 5 行、第 6 行和第 8 行图像的色偏, 结果图像颜色自然且细节丰富, 暗区域信息清晰可见。综合对比可知, 本文算法处理夜间有雾图像效果最好, 给人更好的视觉观感。

3.2　客观评价

为了更加客观地评价算法颜色恢复效果, 使用 CIEDE2000 标准[25] 衡量待测颜色与标准颜色之间的差距, 取值范围为 [0,100], 色差值越小, 图像颜色恢复效果越好, 公式为

$$V_{\mathrm{CIEDE}}=\left[\left(\frac{\Delta L'}{k_L S_L}\right)^2+\left(\frac{\Delta C'}{k_C S_C}\right)^2+\left(\frac{\Delta H'}{k_H S_H}\right)^2+R_T\left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)\right]^{\frac{1}{2}} \tag{14}$$

式中: $\Delta L'$、$\Delta C'$和$\Delta H'$分别为明度差、彩度差和色相差; R_T为旋转函数; k_L、k_C和k_H分别为明度权重因子、彩度权重因子和色相权重因子, 根据实验要求定义数值; S_L、S_C和S_H分别为校正明度、校正彩度和校正色相。

表 1 展示了图 8 中各算法 6 种颜色恢复结果的色差值及色差平均值(最优值用粗体标出)。对于不同颜色, 本文算法色差值都较低, 仅蓝色色差值较高, 而其他算法也无法较好地恢复蓝色信息, 充分说明此算法在夜间去雾中具有更好的颜色恢复

(a) 真实有雾图像 (b) He2011[7] (c) CEEF2022[6] (d) Zhang2014[11] (e) Li2015[12] (f) OSFD2020[15] (g) 本文算法

图 10 真实夜间有雾图像去雾结果对比

Fig. 10 Comparison of dehazing results of real nighttime hazy images

表 1 CIEDE2000 评测值对比

Table 1 Comparison of CIEDE2000 metrics

算法	色差值						色差值均值
	黄色	白色	褐色	红色	蓝色	绿色	
He2011[7]	10.384	14.362	12.356	13.031	33.204	19.878	17.203
Zhang2014[11]	8.543	11.497	10.409	24.261	**30.797**	12.618	16.354
Li2015[12]	6.604	4.915	16.213	**9.109**	32.103	**9.080**	13.004
OSFD2020[15]	3.800	3.637	12.518	13.427	40.042	11.388	14.135
CEEF2022[6]	26.813	26.068	15.886	9.713	40.018	31.475	24.996
本文算法	**1.955**	**2.321**	**9.936**	10.780	31.684	9.297	**10.996**

能力，且本文算法平均色差较次优算法 Li2015 降低约 15%。

为进一步验证算法有效性，选用结构相似性（structural similarity index measurement, SSIM）[26] 和峰值信噪比（peak signal-to-noise ratio, PSNR）[27] 2 个全参考指标进行客观评价。SSIM 和 PSNR 指标以

真实清晰图像为基准, 分析评价结果图的恢复效果, 值越大, 代表图像结构信息恢复越完整, 失真程度越小, 越接近真实清晰图像。真实夜间有雾图像没有对应的无雾图像, 而人工合成夜间有雾图像有其对应的无雾图像, 因此, 对于所有人工合成夜间有雾图像去雾结果的 SSIM 和 PSNR 指标均值如表 2 所示, 图 9 中 4 幅图像的指标对比如图 11 所示。

表 2　人工合成夜间有雾图像数据集 SSIM 和 PSNR 均值

Table 2　Average values of SSIM and PSNR on artificial synthetic nighttime haze image dataset

算法	SSIM	PSNR/dB
He2011[7]	0.636	12.766
Zhang2014[11]	0.433	10.933
Li2015[12]	0.326	6.867
OSFD2020[15]	0.399	8.389
CEEF2022[6]	0.543	10.104
本文算法	0.487	14.702

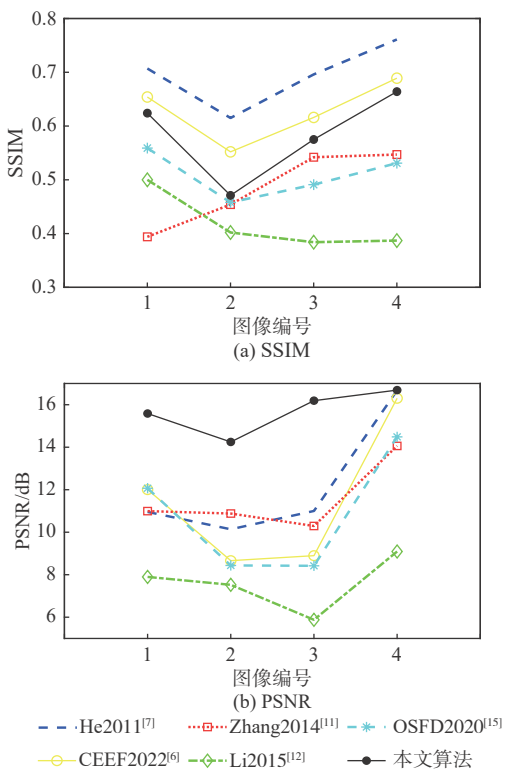


图 11　图 9 中 4 幅图像的有参考指标曲线

Fig. 11　Curves of two reference indexes of four images in Fig. 9

由于人工合成的夜间有雾图像和真实夜间有雾图像差别较大, 前者更接近白天环境下的雾, 本文算法的 SSIM 指标不如 He2011 算法和 CEEF-2022 算法, 但仍大于其他对比算法, 且 PSNR 指标最优, 本文算法 PSNR 较次优算法 He2011 提升约 15%, 证明了其优越的去雾能力。

选用平均梯度(average gradient, AG)、信息熵(information entropy, IE)[28] 和自然图像质量评估(natural image quality evaluator, NIQE)[29] 3 个无参考指标对算法进行定量评估。AG 指标从图像边缘纹理及微小细节的变化方面反映图像清晰程度, 值越大, 表示图像清晰度越高。IE 指标衡量图像所包含的边缘、纹理等细节信息的大小, 值越大, 代表图像细节信息越多, 去雾效果越好。NIQE 指标通过计算待测图像与自然图像多元高斯模型参数的距离来评价图像质量, 值越小, 表示图像视觉效果越好。所有数据集中的夜间有雾图像去雾结果的 AG、IE 和 NIQE 均值如表 3 所示, 图 9 和图 10 中 12 幅图像的指标对比如图 12 所示。

表 3　AG、IE、NIQE 均值

Table 3　Average values of AG, IE, and NIQE

算法	AG	IE	NIQE
He2011[7]	2.716	6.744	3.848
Zhang2014[11]	5.688	7.312	3.042
Li2015[12]	6.355	7.010	3.232
OSFD2020[15]	4.656	7.210	3.270
CEEF2022[6]	3.798	7.121	4.681
本文算法	6.864	6.845	3.032

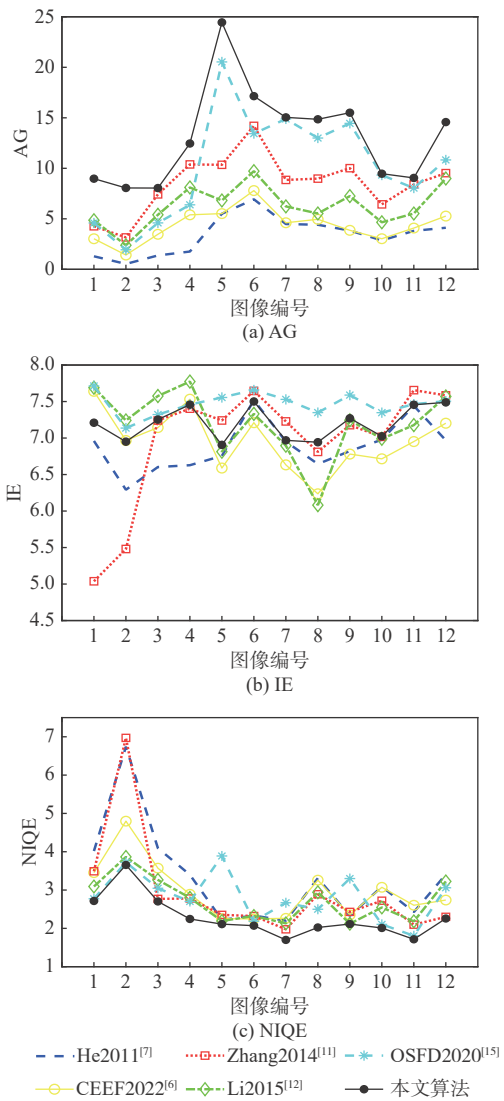


图 12　图 9 和图 10 中 12 幅图像的无参考指标曲线

Fig. 12　Curves of three non-reference indexes of twelve images in Fig. 9 and Fig. 10

本文算法的AG和NIQE指标最优，且本文算法AG较次优算法Li2015提升约8%，进一步客观说明了结果图像更清晰。但IE指标并不理想，与该算法展示更多细节信息的主观评价不一致，考虑是结果图像产生的失真和伪影等被误为细节信息，而导致评测结果偏差。

通过实验对比（见图13），Zhang2014算法的颜色失真问题较为严重，B4中呈现了不同颜色且较为混乱；Li2015算法的伪影问题较为严重，A5和A6中天空和楼房呈现出不规则的块状伪影。这些颜色失真和伪影导致图像出现更多梯度、边缘轮廓等无意义的细节信息，影响了IE指标的评判。

(a) 有雾图像 (b) He2011[7] (c) CEEF2022[6] (d) Zhang2014[11] (e) Li2015[12] (f) OSFD2020[15] (g) 本文算法

图13 失真和伪影情况对比

Fig. 13 Comparison of distortion and artifact

因此，为排除该因素的影响，从结果中挑选出所有失真和伪影相对较少的图像重新进行IE指标的评测实验，再结合表3中IE指标进行对比，结果如图14所示。

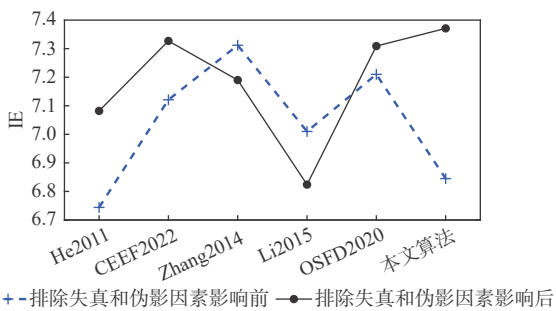


图14 排除失真和伪影因素影响前后图像IE值对比

Fig. 14 Comparison of image IE before and after eliminating distortion and artifact

在排除了失真和伪影影响后，Zhang2014算法和Li2015算法的指标值均有不同程度的下降，这验证了上述论断。排除干扰因素后，本文算法的IE值最大，表明其能够恢复更多的细节信息。

3.3 算法时间对比

为验证算法的效率，在Win10系统下对3种不同尺寸的退化图像进行测试，硬件环境为：Intel i5-12400CPU、16 GB内存和GTX3060显卡，软件平台为：MATLAB R2018b和PyCharm。实验结果如表4所示，本文算法运行效率达先进水平，其中，处理300像素×300像素的图像效率最高，比He2011算法快2.028 s，综合考虑各算法的主观结果和其他客观指标，本文算法更具优势。

3.4 消融实验

为验证算法各组成部分的有效性，在数据集上

表4 不同算法的运行时间

Table 4 Running time of different algorithms

图像尺寸/像素	运行时间/s					
	He2011[7]	Zhang2014[11]	Li2015[12]	OSFD2020[15]	CEEF2022[6]	本文算法
300×300	2.104	0.845	2.073	0.101	0.117	0.076
600×600	9.217	7.012	9.002	0.243	0.624	0.259
1 200×1 200	43.030	26.877	38.878	1.042	2.723	1.113

随机选取图像进行消融研究, 实验包括以下 4 个部分: 未去除表面照射光、未进行融合和增强、未去除辉光及包含所有部分的本文算法, 实验结果如图 15 和表 5 所示。

(a) 原始有雾图像　(b) 未去除表面照射光　(c) 未进行融合和增强　(d) 未去除辉光　(e) 本文算法

图 15　消融实验结果

Fig. 15　Results of ablation experiment

表 5　消融实验的客观评价指标

Table 5　Objective evaluation indexes of ablation experiment

方法	AG	IE	NIQE
原始有雾图像	4.426	7.069	2.786
未去除表面照射光	5.199	6.513	2.791
未进行融合和增强	7.100	7.316	2.266
未去除辉光	10.692	7.373	2.109
本文算法	11.034	7.381	2.073

由图 15 的视觉比较可知, 与原始有雾图像相比, 未去除表面照射光的结果图有一定的去雾效果, 但存在亮度低和色偏问题, 不能去除人造光源低照度和单色光的影响; 未进行融合和增强的结果图亮度和色偏得到改善, 但存在过度增强和局部细节不清晰现象; 未去除辉光的结果图对比度得到有效增强, 但仍存在局部区域色偏及过亮现象, 不能修正大气光分布不均匀特性; 包含所有模块的本文算法的结果图颜色恢复自然, 同时亮度和对比度得到显著提高, 细节清晰, 具有最佳视觉效果。

表 5 展示了消融实验的客观评价指标, 实验结果更加客观地表明了各组成部分对算法性能的影响, 各部分都有助于提高算法性能, 且包含所有部分的本文算法实现了最佳性能。

4　结　论

1) 本文针对夜间雾天环境下人工光源干扰及大气散射模型存在的问题, 提出一种基于非均匀大气光修正模型的夜间图像去雾算法。

2) 加入表面照射光项修正直接衰减项, 避免了人工光源低照度和单色光的影响; 分析夜间成像中大气光分布不均匀特性, 加入辉光项修正大气光项。

3) 根据图像亮度和饱和度分量引入近光源系数, 准确去除辉光; 基于 DCP 和 MSRCR 算法, 采用引导滤波、线性融合及高斯高通滤波方法实现图像去雾、颜色恢复、亮度及细节增强。

4) 实验结果表明, 本文算法不仅在主观视觉上具有更好的效果, 且 PSNR、AG、NIQE 客观评价指标最优, 算法效率与 OSFD2020 算法相当, 高于其他 4 种对比算法。综上所述, 本文算法高效去雾的同时, 图像质量恢复更优。

未来研究中, 将进一步提高算法处理大尺寸图像的效率, 此外, 还将实现算法参数的自适应调整。

参考文献 (References)

[1] 杨勇, 邱根莹, 黄淑英, 等. 基于改进大气散射模型的单幅图像去雾方法[J]. 北京航空航天大学学报, 2022, 48(8): 1364-1375.

YANG Y, QIU G Y, HUANG S Y, et al. Single image dehazing method based on improved atmospheric scattering model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(8): 1364-1375(in Chinese).

[2] 杨燕, 张得欣, 岳辉. 基于最小值通道与对数衰减的图像融合去雾算法[J]. 北京航空航天大学学报, 2020, 46(10): 1844-1852.
YANG Y, ZHANG D X, YUE H. Image fusion dehazing algorithm based on minimum channel and logarithmic attenuation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(10): 1844-1852(in Chinese).

[3] PAUL A. Adaptive tri-plateau limit tri-histogram equalization algorithm for digital image enhancement[J]. The Visual Computer, 2023, 39(1): 297-318.

[4] FU H Y, ZHENG W K, MENG X Y, et al. You do not need additional priors or regularizers in Retinex-based low-light image enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 18125-18134.

[5] 刘备, 边少锋, 纪兵, 等. 小波变换在船载海洋重力测量数据处理与分析中的应用[J]. 系统工程与电子技术, 2023, 45(3): 654-659.
LIU B, BIAN S F, JI B, et al. Application of wavelet transform in shipborne ocean gravimetry data processing and analysis[J]. Systems Engineering and Electronics, 2023, 45(3): 654-659(in Chinese).

[6] LIU X N, LI H, ZHU C. Joint contrast enhancement and exposure fusion for real-world image dehazing[J]. IEEE Transactions on Multimedia, 2022, 24: 3934-3946.

[7] HE K M, SUN J, TANG X O. Single image haze removal using dark channel prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(12): 2341-2353.

[8] LI C, YUAN C J, PAN H B, et al. Single-image dehazing based on improved bright channel prior and dark channel prior[J]. Electronics, 2023, 12(2): 299.

[9] CUI G M, MA Q, ZHAO J F, et al. Image dehazing algorithm based on optimized dark channel and haze-line priors of adaptive sky segmentation[J]. Journal of the Optical Society of America A, 2023, 40(6): 1165-1182.

[10] PEI S C, LEE T Y. Nighttime haze removal using color transfer preprocessing and dark channel prior[C]//Proceedings of the 19th IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2012: 957-960.

[11] ZHANG J, CAO Y, WANG Z F. Nighttime haze removal based on a new imaging model[C]//Proceedings of the IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2014: 4557-4561.

[12] LI Y, TAN R T, BROWN M S. Nighttime haze removal with glow and multiple light colors[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 226-234.

[13] LIAO Y H, SU Z, LIANG X G, et al. HDP-Net: haze density prediction network for nighttime dehazing[C]//Proceedings of the Pacific Rim Conference on Multimedia. Berlin: Springer, 2018: 469-480.

[14] KUANAR S, RAO K R, MAHAPATRA D, et al. Night time haze and glow removal using deep dilated convolutional network[EB/OL]. (2019-02-03)[2023-07-01]. https://arxiv.org/abs/1902.00855v1.

[15] ZHANG J, CAO Y, ZHA Z J, et al. Nighttime dehazing with a synthetic benchmark[C]//Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM, 2020: 2355-2363.

[16] 鲁丹. 不均匀光照下单幅雾天图像复原算法研究[D]. 合肥: 合肥工业大学, 2015: 17-44.
LU D. Research on restoration algorithm of single foggy image under uneven illumination[D]. Hefei: Hefei University of Technology, 2015: 17-44(in Chinese).

[17] 何弘. 非均匀大气光视频图像去雾算法及其FPGA实现[D]. 长沙: 湖南大学, 2021: 14-16.
HE H. Defogging algorithm for non-uniform atmospheric light video image and its FPGA implementation[D]. Changsha: Hunan University, 2021: 14-16(in Chinese).

[18] HE Y F, LI C L, LI X. Remote sensing image dehazing using heterogeneous atmospheric light prior[J]. IEEE Access, 2023, 11: 18805-18820.

[19] LAI Y, LIU Y. Improved single image dehazing with heterogeneous atmospheric light estimation[C]//Proceedings of the 2nd CCF Chinese Conference. Berlin: Springer, 2017: 102-112.

[20] 赖欣, 王储, 陈航. 低照度下人脸检测MSRCR光频分段滤波增强算法[J]. 电子测量与仪器学报, 2022, 36(2): 96-106.
LAI X, WANG C, CHEN H. MSRCR optical frequency segmented filter enhancement algorithm in low-light face detection[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(2): 96-106 (in Chinese).

[21] 杨燕, 张金龙, 张浩文. 基于区间估计与透射率自适应约束的去雾算法[J]. 北京航空航天大学学报, 2022, 48(1): 15-26.
YANG Y, ZHANG J L, ZHANG H W. Dehazing algorithm based on interval estimation and adaptive constraints of transmittance[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(1): 15-26(in Chinese).

[22] HE K M, SUN J, TANG X O. Guided image filtering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(6): 1397-1409.

[23] ZHENG Z R, REN W Q, CAO X C, et al. Ultra-high-definition image dehazing via multi-guided bilateral learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 16180-16189.

[24] LI B Y, REN W Q, FU D P, et al. Benchmarking single-image dehazing and beyond[J]. IEEE Transactions on Image Processing, 2019, 28(1): 492-505.

[25] SHARMA G, WU W, DALAL E N. The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations[J]. Color Research and Application, 2005, 30(1): 21-30.

[26] 杨紫骞, 王艳秋, 郑福, 等. 多维点云结构相似性定量化评价[J]. 光学精密工程, 2023, 31(4): 533-542.
YANG Z Q, WANG Y Q, ZHENG F, et al. Quantitative evaluation method for structural similarity of multidimensional point cloud[J]. Optics and Precision Engineering, 2023, 31(4): 533-542(in Chinese).

[27] ZHAO B, LIU Z Y, DING S X, et al. Motion artifact correction for MR images based on convolutional neural network[J]. Optoelectronics Letters, 2022, 18(1): 54-58.

[28] 王殿伟, 韩鹏飞, 范九伦, 等. 基于光照-反射成像模型和形态学操

作的多谱段图像增强算法[J]. 物理学报, 2018, 67(21): 104-114.
WANG D W, HAN P F, FAN J L, et al. Multispectral image enhancement based on illuminance-reflection imaging model and morphology operation[J]. Acta Physica Sinica, 2018, 67(21): 104-114 (in Chinese).

[29] 罗国强. Gamma 变换与多尺度细节增强的红外图像增强[J]. 激光与红外, 2023, 53(2): 253-260.
LUO G Q. Infrared image enhancement based on Gamma transformation and multi-scale detail enhancement[J]. Laser & Infrared, 2023, 53(2): 253-260(in Chinese).

Nighttime image dehazing based on non-uniform atmospheric light correction model

LIN Sen*，ZHA Ziyue

(School of Automation and Electrical Engineering，Shenyang Ligong University，Shenyang 110159，China)

Abstract：To address the problems of low brightness, glow, and color distortion caused by artificial light sources in nighttime hazy images, this paper proposed a nighttime image dehazing algorithm based on a non-uniform atmospheric light correction model. Firstly, according to the imaging characteristics of non-uniform atmospheric light, the traditional atmospheric scattering model was modified by adding surface illumination light and glow terms. Secondly, the near-light source coefficient was determined based on image intensity and saturation components, and the glow term was accurately estimated to correct the non-uniform atmospheric light. Then, based on the dark channel prior (DCP) and multi-scale Retinex enhancement algorithm with color restoration ability, the study adopted guided filtering to solve the block effect of transmittance and the linear fusion to solve the problem of excessive enhancement. Finally, Gaussian high-pass filter was used for enhancing the details of the image to obtain the final clear image. The experimental results show that nighttime hazy images restored by the proposed algorithm have natural color, rich details, and high visibility. Compared to classical algorithms such as dark channel prior, as well as novel algorithms like contrast enhancement and exposure fusion (CEEF), the proposed algorithm reduces the color difference by about 15%, improves the peak signal-to-noise ratio by about 15%, and enhances the average gradient by about 8%, having more advantages in qualitative and quantitative experiments.

Keywords：nighttime image dehazing；atmospheric scattering model；estimation of atmospheric light；glow；linear fusion

Received：2023-07-04；**Accepted**：2023-12-01；**Published Online**：2024-01-15 18：35

URL：link.cnki.net/urlid/11.2625.V.20240115.1203.003

Foundation items：National Key Research and Development Program of China (2018YFB1403303); Basic Scientific Research Project of Higher Education Institutions of the Educational Department of Liaoning Province (LJKMZ20220615)

* **Corresponding author**. E-mail：lin_sen6@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0456

强噪声环境下基于MSDCNN的滚动轴承故障诊断方法

雷春丽[1,*]，史佳硕[1]，马淑珍[2]，缪成翔[1]，万会元[1]，李建华[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：针对传统基于深度学习的轴承故障诊断方法存在抗噪性能差、计算复杂度高和泛化性能不足的问题，提出了一种基于多尺度动态卷积神经网络(MSDCNN)的滚动轴承故障诊断方法。采用傅里叶变换将滚动轴承一维振动信号转换到频域进行表示，并通过宽卷积核进一步提取特征；提出一种多尺度动态卷积结构，利用改进的通道注意力机制，对不同大小的卷积核提取的特征信息赋予不同的权重；设计一种自校准空间注意力机制(SCSAM)，将提取的特征信息输入到空间注意力机制中，捕获不同区域的重要程度；通过小卷积核进一步提取特征，利用Softmax分类器进行故障类别分类。使用2种不同数据集验证所提模型的故障诊断性能，实验结果表明：与多尺度深度卷积神经网络(MSD-CNN)、宽卷积核卷积神经网络(WKCNN)等智能模型相比，所提模型在强噪声背景下具有更高的分类精度、更好的泛化能力和更强的鲁棒性。

关 键 词：故障诊断；傅里叶变换；多尺度动态卷积；注意力机制；滚动轴承

中图分类号：TH133.33

文献标志码：A **文章编号**：1001-5965(2025)09-2906-10

在现代工业的各个领域中，旋转机械起着极其重要的作用。旋转机械一旦发生故障，将导致工业生产率下降，如果不及时解决故障，甚至会使工人处于危险之中[1-2]。因此，对旋转机械的及时准确监测和故障诊断是非常重要的[3]。在旋转机械中，滚动轴承是支撑旋转机械运转的核心部件之一，通常工作在变速、变载和高温等极端工况下。相关资料显示，45%～55%的旋转机械失效案例是由滚动轴承失效导致的[4]。因此，滚动轴承的故障诊断对保证旋转机械的正常运行和现代工业的安全生产具有重要意义[5]。工业现场测量的轴承振动信号不可避免地受到嘈杂环境的影响，在强噪声干扰下，振动信号中的轴承故障特征往往不明显，提取难度较大，在噪声环境下实现准确和稳定的故障诊断是一个非常艰巨的任务。

近年来，通过振动数据识别故障类型已成为故障诊断领域的主流，其中，基于深度学习的智能故障诊断方法受到了大量学者的青睐[6]。卷积神经网络(convolutional neural network, CNN)作为深度学习代表算法之一，因其独特的卷积结构和强大的特征提取能力，被广泛应用于故障诊断领域，并取得了重大进展[7-8]，目前已衍生出大量优秀的网络模型，如一维CNN[9]、残差网络[10]和Inception网络[11]等。周奇才等[12]提出了一种一维深度CNN的故障诊断模型，采用改进的一维卷积核和池化层以适应一维时域信号，并通过轴承和齿轮箱数据集进行验证，取得了良好的故障诊断结果。Liu等[13]提出了一种基于改进递归图和CNN的滚动轴承故障诊断

收稿日期：2023-07-12；**录用日期**：2023-07-28；**网络出版时间**：2023-08-23 16:48

网络出版地址：link.cnki.net/urlid/11.2625.V.20230823.1327.002

基金项目：国家自然科学基金(51465035)；甘肃省自然科学基金(20JR5RA466)；甘肃省教育厅研究生“创新之星”项目(2023CXZX-411)；兰州理工大学红柳一流学科建设项目

* **通信作者**. E-mail：lclyq2004@163.com

引用格式：雷春丽，史佳硕，马淑珍，等. 强噪声环境下基于MSDCNN的滚动轴承故障诊断方法[J]. 北京航空航天大学学报，2025，51(9)：2906-2915. LEI C L，SHI J S，MA S Z，et al. Rolling bearing fault diagnosis method based on MSDCNN in strong noise environment[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2906-2915 (in Chinese).

方法, 在 Paderborn 轴承数据集上验证了方法的有效性。但上述网络都使用单尺度特征学习方案, 因此, 难以提取振动信号的多尺度特征, 在面对复杂工况和噪声干扰时, 识别准确率偏低。

由于轴承通常受到载荷变化、环境噪声等因素的影响, 振动信号的特征频率变化很大, 导致振动信号在多个时间尺度上具有复杂的特征, 即振动信号的多尺度特征。一些学者已经使用多尺度学习方案来获得不同尺度的特征信息, 这有助于提高神经网络的特征学习能力[14]。张玺君等[15]提出了一种加入通道注意力的多尺度 CNN 抗噪模型, 在不同信噪比的干扰下能够有效提取轴承故障特征。卞景艺等[16]通过改进一维深度 CNN, 构建出一种多尺度深度 CNN, 能够准确识别轴承故障类型。Peng 等[17]将传统的滤波方法与 CNN 相结合, 提出了一种多分支多尺度 CNN, 可以更准确地诊断高速列车轴承故障。Wang 等[18]提出了一种基于宽卷积和多尺度卷积的新型胶囊网络故障诊断方法, 直接将一维振动信号作为输入, 实现端对端的故障诊断, 实验有效验证了模型的有效性和推广性, 表明所提方法在低噪声水平下具有非常好的准确性。虽然以上研究人员将 CNN 和多尺度特征相结合, 利用多尺度 CNN 来学习振动信号的多尺度特征, 取得了许多令人鼓舞的成果, 然而, 大多数多尺度 CNN 仍然存在如下问题需要解决: ①多尺度 CNN 的卷积核分布相对单一, 不能根据不同尺度特征自适应地调整权值。②多尺度 CNN 使用多个并行网络结构, 导致网络结构复杂, 网络训练困难。

针对以上问题, 本文提出了一种基于多尺度动态卷积神经网络 (multi-scale dynamic convolutional neural network, MSDCNN) 的滚动轴承故障诊断方法, 该方法构造了多尺度动态卷积层, 能够更好地适应不同尺度的特征, 增强模型的特征提取能力。本文的主要贡献如下:

1) 通过结合改进的通道注意力 (improved channel attention, ICA) 机制和多尺度 CNN 构建 MSDCNN。具体而言, 该网络能够根据输入特征的不同, 使用 ICA 机制动态调整不同尺度下卷积层的权重, 选择性地聚焦更重要的特征信息, 从而提高本文方法的故障识别能力。

2) 利用深度可分离卷积计算复杂度低和空洞卷积在参数量不变的条件下能够扩大感受野的优点, 提出了一种深度可分离空洞卷积神经网络 (deep separable dilated convolution neural network, DSDCNN) 算法, 并将其应用到多尺度动态卷积中, 能够大量减少网络的参数量。

3) 设计了一种自校准空间注意力机制 (self-calibrating spatial attention mechanism, SCSAM), 能够使模型自动捕获重要区域的特征, 实现对重要信息的捕捉, 更好地实现特征提取。

1　原　理

1.1　改进的通道注意力机制

SE 注意力机制[19]作为经典的注意力机制, 通过显式建模通道之间的相互依赖性来自适应地重新校准通道特性响应。虽然 SE 注意力机制以略微的额外计算成本为模型带来了性能上的显著改善, 但其仍存在一些不足, 本文提出了一种 ICA 机制, 具体结构如图 1 所示。图中: C、H 和 W 分别为特征图的长度、高度和宽度, r 为比例因子, 本文取值为 8。针对 SE 注意力机制用全连接层来实现通道信息交互会造成参数量过大的问题, ICA 机制用逐点卷积代替全连接层实现跨通道信息的交互。此外, 针对 SE 注意力机制用全局平均池化压缩特征时会导致特征丢失的问题, ICA 机制从全局特征和局部特征 2 个方面来提取, 使获取到的特征信息更充分。图 1 左侧分支提取的是全局特征信息, 右侧分支提取的是局部特征信息。全局特征的计算公式为

$$L_1(x)=\mathrm{BN}(\mathrm{PWConv2}(\delta(\mathrm{BN}(\mathrm{PWConv1}(\mathrm{GAP}(x))))))\tag{1}$$

式中: x为输入特征; GAP(·)表示全局平均池化操作; PWConv(·)表示逐点卷积运算; BN(·)表示批量归一化; δ(·)表示 Sigmoid 激活函数; $L_1(x)$为全局输出特征。

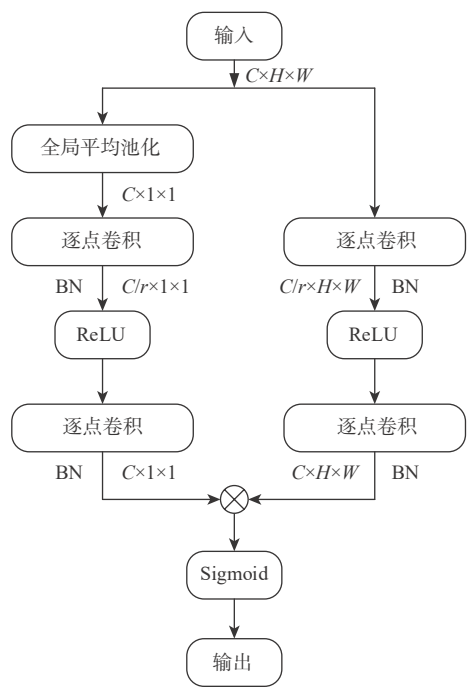


图 1　改进的通道注意力机制

Fig. 1　Improved channel attention mechanism

局部特征的计算公式为

$$L_2(x)=\mathrm{BN}(\mathrm{PWConv2}(\delta(\mathrm{BN}(\mathrm{PWConv1}(x)))))\tag{2}$$

式中：$L_2(x)$为局部输出特征。

1.2 自校准空间注意力机制的设计

本文设计的SCSAM结构如图2所示。SCSAM能够自动捕获重要区域特征，对重要的区域赋予更大的权重。前2个分支对输入特征图通过逐点卷积进一步提取特征，对第1个分支提取的特征进行重塑形状和转置，对第2个分支提取的特征进行重塑形状，以便于2个矩阵相乘，相乘后的结果通过Softmax函数可获得位置注意力图，对输入特征图不同区域赋予不同的权重。矩阵的乘法相当于使每个像素点之间都产生了联系，计算2个位置之间的相似度。其中，2个位置相似度越高，赋予的权重就越大，具体计算公式为

$$\partial_{ji}=\frac{\exp(H_i\times C_j)}{\sum_{i=1}^{N}\exp(H_i\times C_j)}\tag{3}$$

式中：∂_{ji}为生成特征图中第i行第j列所赋予的权重；H_i为第1个分支生成特征图的第i行；C_j为第2个分支生成特征图的第j列；N为特征图的列数。

第3个分支先经过全局平均池化将输入转换为低维嵌入，下采样倍数为4，特征尺寸变为$C\times H/4\times W/4$，再经过逐点卷积进行特征提取，将经过卷积后的特征图进行上采样操作，使其尺寸大小恢复成$C\times H\times W$。这种自校准操作能够实现对重要信息的捕捉，更好地实现特征提取。具体运算过程如下：

$$x_3=\mathrm{UP}(\mathrm{PWConv}(\mathrm{GAP}(x)))\tag{4}$$

式中：GAP(·)表示全局平均池化操作；UP(·)为上采样函数；x_3为第3个分支的输出特征。

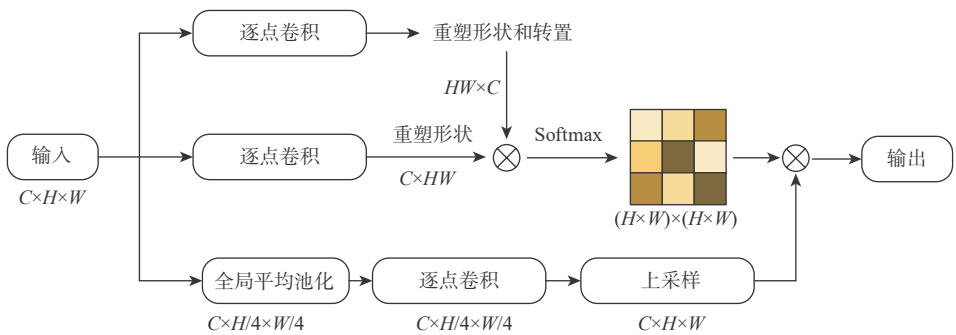


图2 自校准空间注意力机制

Fig. 2 Self-calibrating spatial attention mechanism

1.3 深度可分离空洞卷积神经网络

本文利用深度可分离卷积参数量少和空洞卷积在卷积核不变的情况下能够扩大感受野的优点，提出了一种DSDCNN算法，结构如图3所示。

1）深度空洞卷积是向深度卷积的卷积层中引入一个新的参数扩张率(dilated rate, DR)，利用添加空洞扩大感受野，使每个卷积输出都包含较大范围的信息。采用不同的扩张率能够挖掘出更深层、更丰富的多尺度特征信息，提高网络的特征学习能力。

2）在深度空洞卷积后添加逐点卷积，能够使不同通道之间的特征信息进行交互，提高模型的泛化能力。

3）在深度空洞卷积和逐点卷积后添加BN层，能够加快模型的训练和收敛速度，控制模型梯度爆炸并防止梯度消失，在一定程度上避免了过拟合。添加ReLU激活函数能够增强模型的非线性表达能力，更好地解决复杂的问题。

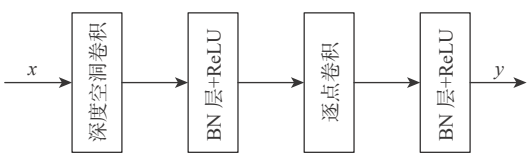


图3 深度可分离空洞卷积神经网络

Fig. 3 Deep separable dilated convolution neural network

假设输入特征图的尺寸为$h\times w$，通道数为c，卷积核大小为$k\times k$，卷积核的数量为n。若常规卷积获得的感受野和DSDCNN一样大，则DSDCNN和常规卷积的计算量比值为

$$e=\frac{h\times w\times c\times k\times k+c\times n\times h\times w}{h\times w\times c\times[k+(k-1)(E_{\mathrm{DR}}-1)]^2\times n}=\frac{k\times k+n}{[k+(k-1)(E_{\mathrm{DR}}-1)]^2\times n}\tag{5}$$

式中：k一般取3、5、7；E_{DR}为扩张率，$E_{\mathrm{DR}}\geqslant 1$，则$k+(k-1)\cdot(E_{\mathrm{DR}}-1)$的值大于$k$。$n$的值往往很大，因此，$e$的值远远小于1，即DSDCNN算法的计算量小于常规卷积。

2 基于MSDCNN的滚动轴承故障诊断模型

2.1 MSDCNN故障诊断模型的构建

MSDCNN模型结构如图4所示。首先，将滚动轴承振动信号通过快速傅里叶变换(fast Fourier

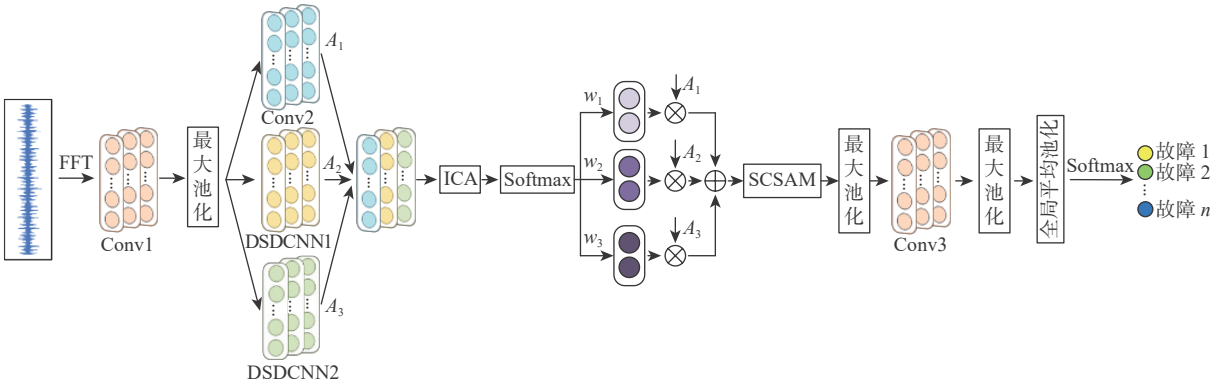


图 4　MSDCNN 模型结构

Fig. 4　MSDCNN model structure

transform, FFT) 转换到频域增强特征表示，并通过宽卷积核进一步提取特征，避免无关信号的干扰。其次，通过 3 种不同大小的卷积核进一步提取特征，得到特征图 A_1、A_2 和 A_3。特征图 A_2、A_3 是通过 DSDCNN 得到的，为避免 DSDCNN 造成特征丢失，特征图 A_1 仍采用常规卷积。由于不同尺度的卷积核提取的特征信息不同，重要程度也不同，设计了一种 ICA 机制，能够根据不同的输入动态地调整多尺度卷积层的权重，提高多尺度卷积层的学习效率，将得到的权重和多尺度特征图 A_1、A_2、A_3 相乘后进行特征融合。然后，将融合后的特征图输入 SCSAM 获取特征图不同区域的重要程度。最后，通过小卷积核进一步提取特征，采用全局平均池化代替全连接层，进一步减少参数并防止过拟合，通过 Softmax 分类器实现故障诊断和分类。

2.2　MSDCNN 模型的结构参数

MSDCNN 模型的结构参数如表 1 所示。每个卷积层都由卷积、BN 层和 ReLU 激活函数组成，使用 BN 层加快网络的训练和收敛速度，使用 ReLU 激活函数增强模型的非线性表达能力。

表 1　MSDCNN 模型的结构参数

Table 1　Structure parameters of MSDCNN model

特征层	卷积核数量	卷积核大小	输出尺寸/像素
输入			1 024×1
Conv1	8	32	1 024×8
最大池化			512×8
Conv2	32	5	512×32
DSDCNN1	32	5(E_{DR}=2)	512×32
DSDCNN2	32	5(E_{DR}=3)	512×32
特征融合			512×96
ICA			512×96
特征融合			512×32
最大池化			256×32
SCSAM			256×32
Conv3	32	5	256×32
最大池化			128×32
全局平均池化			1×32
Softmax分类器			7

3　实验验证与分析

3.1　实验数据集

本文所用数据集为美国凯斯西储大学 (Case Western Reserve University, CWRU) 的滚动轴承数据集和团队所在实验室的 MFS 深沟球轴承数据集。CWRU 采集试验台[20] 如图 5 所示，轴承的故障形式为点蚀，滚动体、内圈、外圈的故障均是由电火花加工形成的单点故障。本文使用的轴承型号为 SKF6205，故障直径为 0.18 mm 和 0.36 mm，试验台在采样频率 12 kHz、负载 0.746～2.238 kW(1～3 hp) 下采集振动信号。每种尺寸都由内圈、外圈、滚动体 3 种故障组成，并加 1 种正常状态，共计 7 种故障类型。

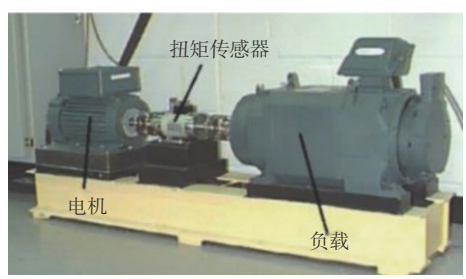


图 5　美国凯斯西储大学数据采集试验台

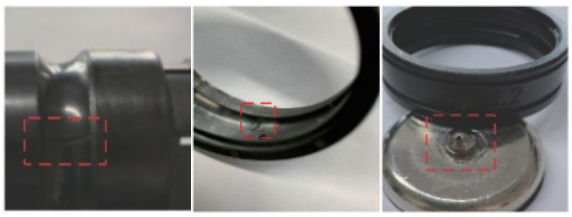
Fig. 5　Data acquisition test bench of CWRU

笔者实验室 MFS 试验台如图 6 所示。采用深沟球轴承的型号为 ER-16K，故障由激光刻蚀技术加工而成。轴承数据在采样频率为 15.36 kHz、转速为 1 200, 1 300, 1 400 r/min 条件下，通过信号采集器采集而得，共有故障直径分别为 0.6 mm 和 1.2 mm 的内圈、外圈、滚动体的点蚀故障和正常状态等 7 种类型。故障部位如图 7 所示。

为方便表示，将 CWRU 数据集负载为 0.746, 1.492, 2.238 kW(1, 2, 3 hp) 下的数据集分别记作数据集 A、

图6 MFS机械故障仿真试验台

Fig. 6 MFS mechanical fault simulation test bench



(a) 内圈缺陷 (b) 外圈缺陷 (c) 滚动体缺陷

图7 ER-16K滚动轴承故障缺陷

Fig. 7 ER-16K rolling bearing fault defects

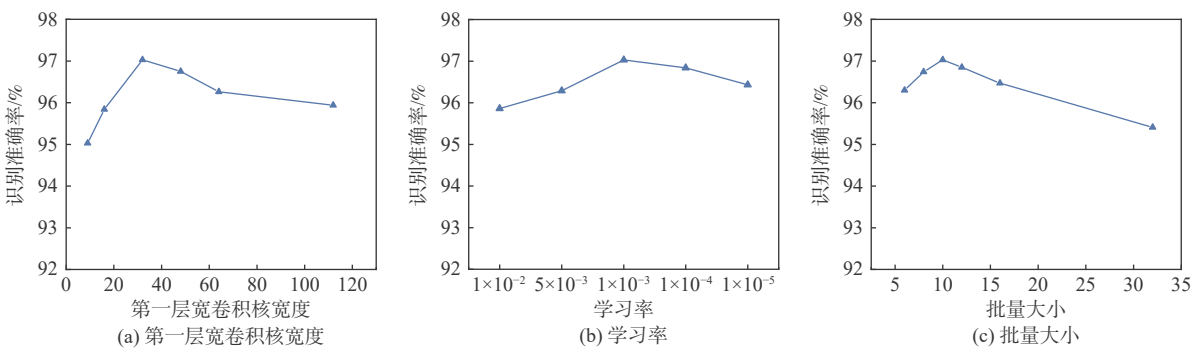
数据集B和数据集C，将MFS数据集转速为1 200，1 300，1 400 r/min下的数据集分别记作数据集D、数据集E和数据集F。2种数据集实验中，训练集每类故障状态选取60个样本，共420个样本，样本长度为1 024，训练集比例为80%。测试集选取100个样本数量。

3.2 模型最优参数的选择

为优化MSDCNN模型的结构参数并获得更好的分类结果，本文在信噪比为−4 dB下C-B工况时进行了多次实验，分析了第1层宽卷积核大小、不同学习率和批量大小对模型故障诊断性能的影响，测试结果如图8所示。测试结果均取10次实验的平均值，每次测试时，除测试参数外，其他参数均保持不变。

第1层宽卷积核的宽度对于模型的性能有着重要影响。当第1层宽卷积核的宽度太小时，无法提取到足够的特征，而且易受到无关信号的干扰。当第1层宽卷积核的宽度太大时，不利于提取信号的局部信息，且增加了模型的参数量，导致训练速度变慢。从图8(a)可以看出，当第1层宽卷积核宽度为32时，模型的识别效果最好，因此，本文第1层宽卷积核的宽度选择为32。

学习率作为模型中重要的超参数，其决定着模型目标函数能否收敛到最小值及收敛到最小值的速度。当学习率太小时，模型的收敛过程将变得十分缓慢。当学习率太大时，模型可能会在最小值附近来回振荡，甚至可能无法收敛。从图8(b)可知，当学习率选择1×10^{-3}时，模型的故障诊断性能最好。



(a) 第一层宽卷积核宽度 (b) 学习率 (c) 批量大小

图8 不同参数对模型性能的影响

Fig. 8 Influence of different parameters on model performance

批量大小直接决定模型的训练速度，批量尺寸太大会造成模型的性能降低，批量尺寸太小会造成模型的收敛速度太慢。如图8(c)所示，当批量大小选择为10时，模型的故障诊断精度达到了97.03%，随着批量大小的增大，模型的识别准确率开始降低，因此，本文选择批量尺寸的大小为10。

3.3 强噪声环境下跨工况故障诊断

本文实验采用Pycharm2020.1.2软件，软件环境为Tensorflow，硬件平台为Intel(R) Xeon-(R) Silver 4110CPU @2.10 GHz 2.10 GHz双处理器和NVID-IA Quadro P4000显卡。设置批量大小batch size=10，迭代次数epoch=35，学习率为1×10^{-3}。使用Adam优化算法优化模型的结构参数，并在模型全局平均池化层后设置dropout值为0.5，在模型训练时，使某个神经元的激活值以一定的概率停止工作，这样可以使模型泛化能力更强，防止模型过拟合。3种不同的实验验证均是在信噪比SNR=−4 dB的强噪声环境下进行的。

3.3.1 CWRU数据集故障诊断

在轴承实际运转过程中，轴承的载荷经常发生变化，因此，在不同负载轴承数据集下验证模型的诊断性能和泛化能力是非常有必要的。为验证本文模型的有效性，与近年来的深度学习模型进行了对

比分析，包括多尺度深度卷积神经网络（multi-scale deep convolutional neural network, MSD-CNN）[21]、宽卷积核卷积神经网络（wide convolutional kernel convolutional neural network, WKCNN）[22]、多尺度注意力机制卷积神经网络（multi-scale attention mechanism convolutional neural network, MSACNN）[23]、多尺度级联中点残差卷积神经网络（multi-scale cascade midpoint residual convolutional neural network, MSC-MpResCNN）[24]。其中，MSD-CNN 是一种多尺度深度 CNN；WKCNN 是一种宽卷积核故障诊断模型，前 2 个卷积层采用宽卷积核来快速提取特征；MSACNN 为一种注意力 CNN 模型；MSC-MpResCNN 为一种多尺度级联中点残差 CNN，引入一种新的多尺度级联结构来提取原始数据中包含的多分辨率特征。实验结果如表 2 所示，其中，A-B 表示数据集 A 作为训练集，数据集 B 作为测试集。从表 2 中可以看出，本文模型在信噪比为−4 dB 时，6 种不同工况下平均识别准确率达到 97.23%，相较于其他 4 种方法，故障识别准确率最高。在 6 种不同的工况中，B-A 的故障识别准确率最高，为 97.51%，而效果表现最差的 MSACNN 模型，B-A 的识别准确率为 92.29%，高出 5.22%。综上所述，本文模型在噪声环境下有着更高的故障诊断性能、更强的泛化能力。

表 2　强噪声环境下变负载识别准确率

Table 2　Variable load recognition accuracy under strong noise environment　　%

模型	识别准确率						均值
	A-B	A-C	B-A	B-C	C-A	C-B	
本文模型	97.43	96.86	97.51	97.26	97.29	97.03	97.23
MSD-CNN	94.57	93.68	95.05	93.63	93.77	92.69	93.90
WKCNN	93.37	92.06	95.17	92.29	90.11	91.20	92.37
MSACNN	92.69	91.29	92.29	93.79	94.00	93.14	92.87
MSC-MpResCNN	96.23	95.48	96.85	96.54	96.31	95.83	96.21

为更清楚地展示故障数据的分类情况，引入 t-SNE 非线性降维算法[25]，故障数据分类情况如图 9 所示。可以看出，大部分的故障数据都得到了聚类，只有少部分的故障出现了错分的情况。以 A-B 为例，少部分的 0.36 mm 滚动体故障被模型错分为 0.36 mm 外圈故障，说明此时这 2 种故障类型容易混淆，而其他故障类型的数据都得到了准确聚类，说明本文模型有着较高的故障识别准确率和抗噪性能。

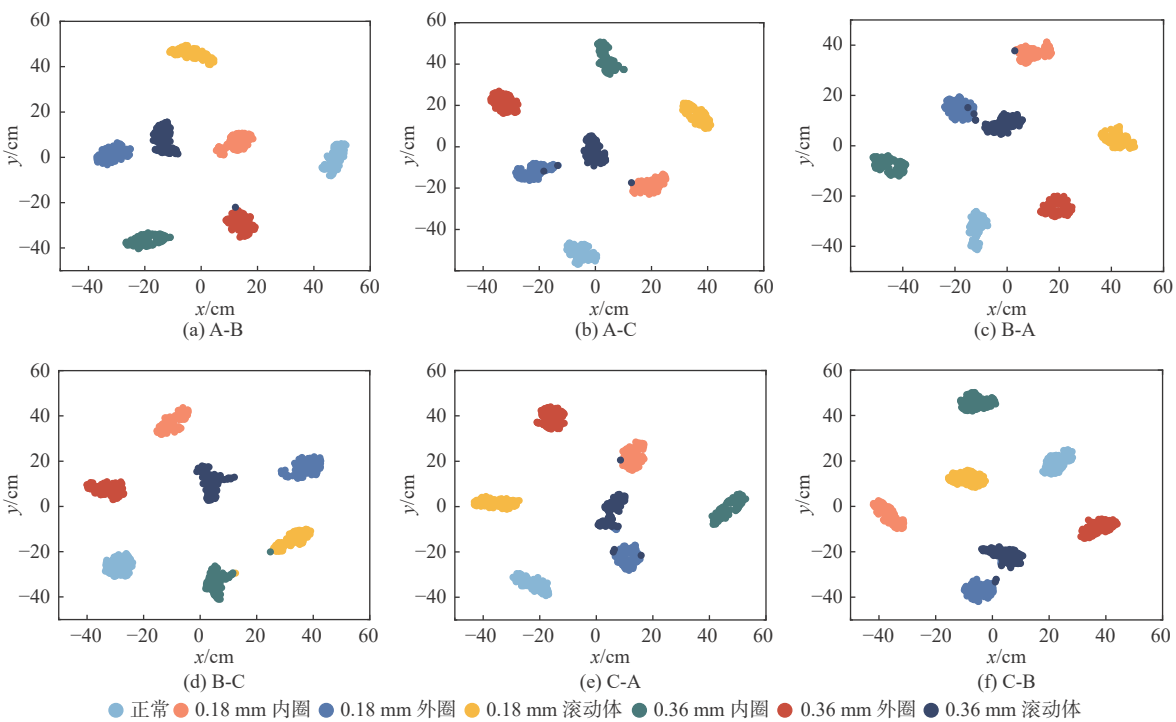


图 9　t-SNE 可视化

Fig. 9　t-SNE visualization

3.3.2 MFS 数据集故障诊断

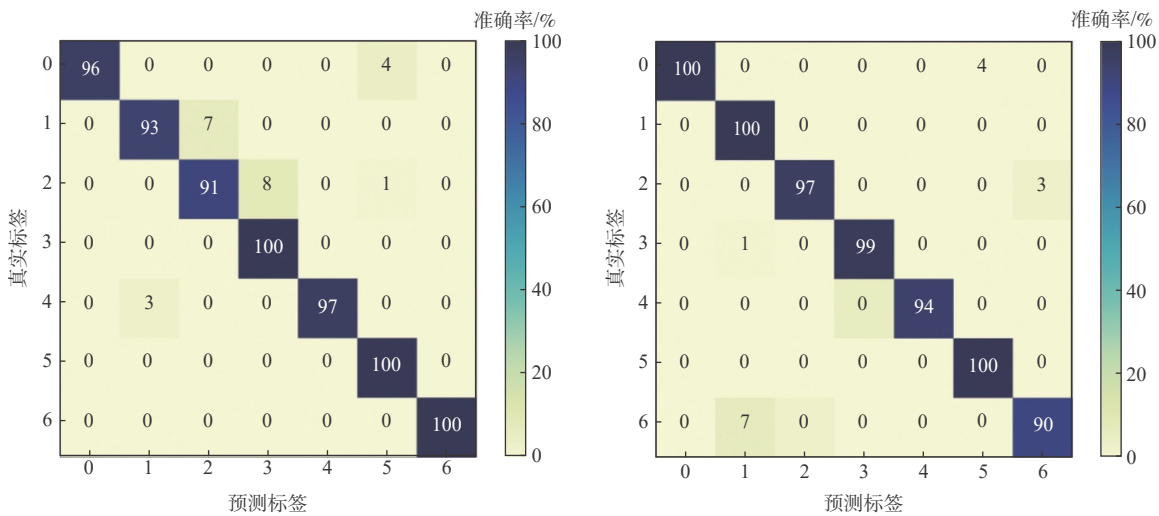
为更进一步验证本文模型的故障诊断性能和泛化能力，对 MFS 轴承数据集进行诊断和测试，结果如表 3 所示。其中，D-E 表示将数据集 D 作为训练集，数据集 E 作为测试集。可以看出，本文模型在 6 种不同转速下的识别准确率均高于其他 4 种方法，平均识别准确率为 97.06%，比效果最好的 MSC-MpResCNN 高出了 1.83%，比效果最差的 MSACNN 高出了 4.96%，相较于其他 4 种方法效果提升明显，说明本文方法在噪声环境下有更强的抗干扰能力。在 D-F 工况下，本文模型的识别准确率为 96.69%，而效果最差的 MSACNN 识别准确率为 90.74%，高出 5.95%。综上所述，本文模型在 MFS 变转速轴承数据集下也有着较好的故障分类识别准确率和泛化能力。

表 3 强噪声环境下变转速识别准确率

Table 3 Variable speed recognition accuracy under strong noise environment %

模型	识别准确率						均值
	D-E	D-F	E-D	E-F	F-D	F-E	
本文模型	97.20	96.69	97.34	97.03	96.89	97.22	97.06
MSD-CNN	93.63	91.77	94.09	93.29	92.89	93.20	93.15
WKCNN	92.71	91.37	93.54	92.69	91.94	93.26	92.59
MSACNN	92.23	90.74	93.03	91.94	92.40	92.28	92.10
MSC-MpResCNN	95.25	94.54	95.74	96.20	94.46	95.17	95.23

为更直观地反映故障数据的聚类和错分情况，引入混淆矩阵进行分析。图 10 展示了 D-F 工况和 F-E 工况下的故障分类结果的混淆矩阵。从图 10(a) 可以看出，在强噪声环境下，700 个样本只有 23 个被错分，其中，0.18 mm 内圈故障和 0.18 mm 外圈故障被错分的数量最多。从图 10(b) 可以看出，700 个样本只有 20 个被错分，进一步验证了本文方法的故障诊断性能和泛化能力。图中，坐标数值 0 表示正常状态，1、2、3 分别表示 0.18 mm 内圈故障、0.18 mm 外圈故障、0.18 mm 滚动体故障，4、5、6 分别表示 0.36 mm 内圈故障、0.36 mm 外圈故障、0.36 mm 滚动体故障。



(a) 变转速 D-F 工况下混淆矩阵 (b) 变转速 F-E 工况下混淆矩阵

图 10 故障分类混淆矩阵

Fig. 10 Fault classification confusion matrix

箱线图可以反映数据的分布情况，故可用于说明不同模型的诊断性能和稳定性，本文将不同方法变转速工况下的故障识别准确率绘制成箱线图，结果如图 11 所示。可以看出，本文模型的识别准确率的均值和中位线最高，均高于其他 4 种模型，且上四分位数与下四分位数的差最小，即方差最小，而 MSD-CNN 和 MSACNN 还伴有异常值的出现，因此，本文模型故障识别准确率最好，有更强的鲁棒性。

3.3.3 早期故障识别性能分析

由于机械设备具有多样性，导致滚动轴承故障发生的时间往往难以确定，若能精确识别轴承早期故障，对设备故障预警和提高轴承甚至整个机械设备的使用寿命具有重要意义。为此，采用西安交通

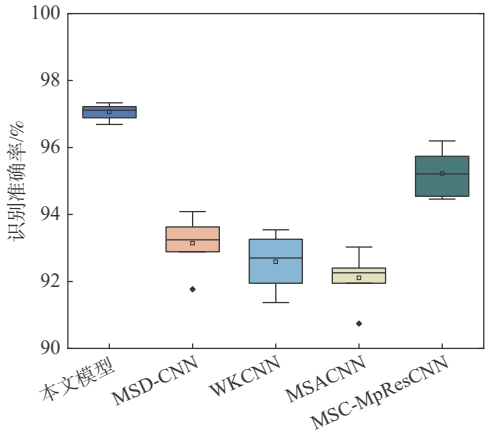


图 11　不同模型变转速工况箱线图

Fig. 11　Box plot of different models under variable speed conditions

大学和浙江长兴昇阳科技有限公司轴承数据集(XJTU-SY)早期的初始故障信号进行了实验验证。文献 [26] 通过计算 JS 散度相关系数矩阵清晰地看出不同状态的跃迁，并结合相邻时刻的平均相似性，得到了早期故障发生时间点，采用该文献计算所得的初始故障时间对应的数据作为实验的故障数据集，具体故障样本分布如表 4 所示，该样本既包含了缓慢失效故障信号，又包含了突发失效故障信号。实验设置训练集样本量为 48，测试集样本量为 100。

表 4　XJTU-SY 数据集早期故障样本分布

Table 4　Early fault sample distribution of XJTU-SY data set

故障类型	轴承编号	总实验时长/min	初始故障发生时间点/min	转速/(r·min^{-1})
外圈	Bearing1_1	123	77	2 100
混合故障	Bearing1_5	52	33	2 100
内圈	Bearing2_1	491	454	2 250
保持架	Bearing2_3	533	325	2 250

为更好地观察不同模型早期故障诊断效果形成差异的原因，通过训练好的模型对测试样本进行故障分类，与实际标签对比，计算出各类健康状态识别准确率，实验中信噪比设置为−4 dB，实验结果如表 5 所示。可以看出，本文模型的识别准确率均值为 97.9%，比其他 4 种对比模型分别高 18.7%、36.75%、14.25%、6.3%。表 5 中，各方法均能对内圈故障实现准确的识别，形成模型诊断性能不同的原因主要是对外圈、混合故障和保持架的分类效果上存在较大差异，如本文模型外圈故障的识别准确率为 94.8%，而 WKCNN 模型外圈的识别准确率只有 41.4%。综上，验证了本文模型能够实现滚动轴承早期故障的精准识别，且对各类健康状态的分类识别具有更高的准确率，表明本文模型具有更强的泛化能力。

表 5　不同模型对早期故障的识别准确率

Table 5　Recognition accuracy of different models forearly faults　%

模型	识别准确率				识别准确率均值
	外圈	混合故障	内圈	保持架	
本文模型	94.8	98.8	100	98	97.9
MSD-CNN	62	75.2	100	79.6	79.2
WKCNN	41.4	41.6	99.6	62	61.15
MSACNN	47	96.6	100	91	83.65
MSC-MpResCNN	87.2	93	100	86.2	91.6

4　结　论

本文针对传统深度学习方法参数量多、在受到噪声环境干扰时识别准确率低和泛化性能不足的问题，提出了一种基于多尺度动态卷积神经网络的滚动轴承故障诊断方法。具体结论如下：

1）第 1 层采取宽卷积核进行特征提取，能够快速提取特征并抑制无关信号的干扰。

2）设计了一种深度可分离空洞卷积神经网络算法，在此基础上提出了一种多尺度动态卷积神经网络，能够根据不同的输入动态调整多尺度卷积层的权重，提高模型特征学习能力的同时，大大减少模型的参数量。

3）提出了一种自校准空间注意力机制，使模型能够自动捕获重要区域的特征。

4）使用 CWRU 轴承数据集、MFS 数据集和 XJTU-SY 数据集对本文模型进行了验证分析，实验结果表明，本文模型相较于其他方法在噪声环境下有着更好的分类识别准确率和更强的鲁棒性。

参考文献（References）

[1] LI W, ZHONG X, SHAO H D, et al. Multi-mode data augmentation and fault diagnosis of rotating machinery using modified ACGAN designed with new framework[J]. Advanced Engineering Informatics, 2022, 52: 101552.

[2] CHEN Z Y, GRYLLIAS K, LI W H. Mechanical fault diagnosis using convolutional neural networks and extreme learning machine[J]. Mechanical Systems and Signal Processing, 2019, 133: 106272.

[3] SONG L Y, WANG H Q, CHEN P. Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery[J]. IEEE Transactions on Instrumentation and Measurement, 2018, 67(8): 1887-1899.

[4] WANG B, LEI Y G, LI N P, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings[J]. IEEE Transactions on Reliability, 2020, 69(1): 401-412.

[5] XIAO Y M, SHAO H D, HAN S Y, et al. Novel joint transfer network for unsupervised bearing fault diagnosis from simulation domain to experimental domain[J]. IEEE/ASME Transactions on Mechatronics, 2022, 27(6): 5254-5263.

[6] CHEN Z Y, LI W H. Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network[J]. IEEE Transactions on Instrumentation and Measurement, 2017, 66(7): 1693-1702.

[7] ZHAO J, YANG S P, LI Q, et al. A new bearing fault diagnosis method based on signal-to-image mapping and convolutional neural network[J]. Measurement, 2021, 176: 109088.

[8] CHOUDHARY A, MIAN T, FATIMA S. Convolutional neural network based bearing fault diagnosis of rotating machine using thermal images[J]. Measurement, 2021, 176: 109196.

[9] 曲建岭, 余路, 袁涛, 等. 基于一维卷积神经网络的滚动轴承自适应故障诊断算法[J]. 仪器仪表学报, 2018, 39(7): 134-143.
QU J L, YU L, YUAN T, et al. Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network[J]. Chinese Journal of Scientific Instrument, 2018, 39(7): 134-143(in Chinese).

[10] 袁彩艳, 孙洁娣, 温江涛, 等. 多域信息融合结合改进残差密集网络的轴承故障诊断[J]. 振动与冲击, 2022, 41(4): 200-208.
YUAN C Y, SUN J D, WEN J T, et al. Bearing fault diagnosis based on information fusion and improved residual dense networks[J]. Journal of Vibration and Shock, 2022, 41(4): 200-208 (in Chinese).

[11] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2017: 4278-4284.

[12] 周奇才, 刘星辰, 赵炯, 等. 旋转机械一维深度卷积神经网络故障诊断研究[J]. 振动与冲击, 2018, 37(23): 31-37.
ZHOU Q C, LIU X C, ZHAO J, et al. Fault diagnosis for rotating machinery based on 1D depth convolutional neural network[J]. Journal of Vibration and Shock, 2018, 37(23): 31-37(in Chinese).

[13] LIU X P, XIA L J, SHI J, et al. A fault diagnosis method of rolling bearing based on improved recurrence plot and convolutional neural network[J]. IEEE Sensors Journal, 2023, 23(10): 10767-10775.

[14] AN Z H, LI S M, WANG J R, et al. Generalization of deep neural network for bearing fault diagnosis under different working conditions using multiple kernel method[J]. Neurocomputing, 2019, 352: 42-53.

[15] 张玺君, 尚继洋, 余光杰, 等. 基于注意力的多尺度卷积神经网络轴承故障诊断[J]. 吉林大学学报 (工学版), 2024, 54(10): 3009-3017.
ZHANG X J, SHANG J Y, YU G J, et al. Bearing fault diagnosis based on attention for multi-scale convolutional neural network[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(10): 3009-3017(in Chinese).

[16] 卞景艺, 刘秀丽, 徐小力, 等. 基于多尺度深度卷积神经网络的故障诊断方法[J]. 振动与冲击, 2021, 40(18): 204-211.
BIAN J Y, LIU X L, XU X L, et al. Fault diagnosis method based on a multi-scale deep convolutional neural network[J]. Journal of Vibration and Shock, 2021, 40(18): 204-211(in Chinese).

[17] PENG D D, WANG H, LIU Z L, et al. Multibranch and multiscale CNN for fault diagnosis of wheelset bearings under strong noise and variable load condition[J]. IEEE Transactions on Industrial Informatics, 2020, 16(7): 4949-4960.

[18] WANG Y, NING D J, FENG S L. A novel capsule network based on wide convolution and multi-scale convolution for fault diagnosis[J]. Applied Sciences, 2020, 10(10): 3659.

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

[20] SMITH W A, RANDALL R B. Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study[J]. Mechanical Systems and Signal Processing, 2015, 64: 100-131.

[21] SHAO Z H, LI W Q, XIANG H, et al. Fault diagnosis method and application based on multi-scale neural network and data enhancement for strong noise[J]. Journal of Vibration Engineering & Technologies, 2024, 12(1): 295-308.

[22] SONG X D, CONG Y Y, SONG Y F, et al. A bearing fault diagnosis model based on CNN with wide convolution kernels[J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(8): 4041-4056.

[23] 丁雪, 邓艾东, 李晶, 等. 基于多尺度和注意力机制的滚动轴承故障诊断[J]. 东南大学学报 (自然科学版), 2022, 52(1): 172-178.
DING X, DENG A D, LI J, et al. Fault diagnosis of rolling bearing based on multi-scale and attention mechanism[J]. Journal of Southeast University (Natural Science Edition), 2022, 52(1): 172-178(in Chinese).

[24] CHAO Z Q, HAN T. A novel convolutional neural network with multiscale cascade midpoint residual for fault diagnosis of rolling bearings[J]. Neurocomputing, 2022, 506: 213-227.

[25] BELKINA A C, CICCOLELLA C O, ANNO R, et al. Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets[J]. Nature Communications, 2019, 10(1): 5415.

[26] 黄如意, 李霁蒲, 王震, 等. 基于多任务学习的装备智能诊断与寿命预测方法[J]. 中国科学: 技术科学, 2022, 52(1): 123-137.
HUANG R Y, LI J P, WANG Z, et al. Intelligent diagnostic and prognostic method based on multitask learning for industrial equipment[J]. Scientia Sinica (Technologica), 2022, 52(1): 123-137(in Chinese).

Rolling bearing fault diagnosis method based on MSDCNN in strong noise environment

LEI Chunli[1,*]，SHI Jiashuo[1]，MA Shuzhen[2]，MIAO Chengxiang[1]，WAN Huiyuan[1]，LI Jianhua[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：To address the poor anti-noise performance, high computational complexity, and insufficient generalization performance of traditional bearing fault diagnosis methods based on deep learning, this paper proposed a rolling bearing fault diagnosis method based on multi-scale dynamic convolutional neural network (MSDCNN). Firstly, the one-dimensional vibration signal of the rolling bearing was transformed into frequency domain by Fourier transform, and the features werefurther extracted by wide convolution kernel. Secondly, a multi-scale dynamic convolution structure was presented, and an improved channel attention mechanism wasutilized to assign different weights to the feature information extracted by convolution kernels of different sizes. Then, a self-calibrating spatial attention mechanism (SCSAM) was designed to capture the importance of different regions by inputting the extracted feature information into the spatial attention mechanism. Finally, the features were further extracted by the small convolution kernel, and the Softmax classifier was used to classify faults. Two different data sets were used to verify the fault diagnosis performance of the proposed model. The experimental results show that the proposed model has higher classification accuracy, better generalization ability, and stronger robustness under strong noise background than other intelligent modelssuch as multi-scale deep convolutional neural network (MSD-CNN) and wide convolutional kernel convolutional neural network (WKCNN).

Keywords：fault diagnosis；Fourier transform；multi-scale dynamic convolution；attention mechanism；rolling bearings

Received：2023-07-12；**Accepted**：2023-07-28；**Published Online**：2023-08-23 16：48

URL：link.cnki.net/urlid/11.2625.V.20230823.1327.002

Foundation items：National Natural Science Foundation of China (51465035); Natural Science Foundation of Gansu Province (20JR5RA466); Graduate Innovation Star Project of the Education Department of Gansu Province (2023CXZX-411); Hongliu First-class Disciplines Development Program of Lanzhou University of Technology

*** Corresponding author**. E-mail：lclyq2004@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0431

基于MPC的飞机牵引车轨迹跟踪

张军[1,*]，黄明辉[2]，王玥琳[3]，阳星[4]，叶敏[1]，贾永乐[5]

(1. 长安大学 道路施工技术与装备教育部重点实验室，西安 710064；2. 中国路桥工程有限责任公司，北京 100010；
3. 长安大学 道路基础设施数字化教育部工程研究中心，西安 710064；4. 西安爱生技术集团有限公司，西安 710065；
5. 西安睿诺航空装备有限责任公司，西安 710060)

摘 要：为能满足物流机场短时间、高频次的快捷飞机牵引需求，提出了基于无人驾驶技术的快速牵引方法。采用“理论建模-算法设计-算例测试和仿真优化-样机实验”的技术路线和方法，以10 t飞机牵引车为对象，构建牵引车的运动学模型，确定牵引车的约束条件和控制量，通过增加防碰撞处理、最小转弯半径和路径平滑的方式改进A*算法，生成牵引车运动轨迹；设计模型预测控制(MPC)的轨迹跟踪控制器，构建MATLAB/Simulink和ADAMS联合仿真模型，通过轨迹跟踪仿真实验优化MPC的控制参数，并在改造的电传动飞机牵引车样机上开展轨迹跟踪实验。结果表明：改进的A*算法满足飞机牵引车工作路径规划和最小转弯半径要求，联合仿真方法优化了MPC控制器，在样机上实现了较好的跟踪精度，弯道和直线跟踪误差的标准差分别为0.362 m和0.128 m，实现了飞机牵引车的无人驾驶功能，为智慧物流机场的无人牵引飞机奠定技术基础。

关 键 词：电传动飞机牵引车；无人驾驶技术；路径规划；改进A*算法；轨迹跟踪；模型预测控制

中图分类号：V351.34; TP23

文献标志码：A **文章编号**：1001-5965(2025)09-2916-11

飞机在机库、廊桥、停机坪和维修车间移动均由牵引车牵引实现移动，牵引路线和效率多依赖于操作人员的经验和熟练度。现有的人工操控飞机牵引方式无法满足短时间内高频次的要求，如物流机场，短时间内有大量飞机起落，需要快速高效的牵引配合。因此，迫切需要针对飞机牵引车的特点，开展牵引车的避障和轨迹跟踪控制研究，提高牵引车移动精确性及工作效率，为智慧机场助力。

路径规划是通过路径搜索过程来确定最优的路径，可分为基于图搜索和基于随机采样的2类路径规划方法[1]。前者包括Dijkstra算法、A*算法和D*算法[2]，这类算法环境建模复杂，运行在栅格地图上，适合做全局规划；后者包括快速探索随机树(rapidly-exploring random tree, RRT)算法[3]、改进的RRT*算法[4]和Anytime RRT*算法[5]，这类算法计算量适中，存在路径不是最优、路径平滑性差等问题。路径规划研究集中在无人驾驶汽车、自动引导运输车(automated guided vehicle, AGV)小车上，并被引入到飞机牵引车上。孙家玮等[6]构建了舰载机牵引系统模型，结合闭环快速搜索随机树(close loop rapidly-exploring random tree, CL-RRT)算法和模型预测控制(model predictive control, MPC)算法，提出了一种舰载机牵引系统的路径规划算法，解决了随机采样算法带来的路线冗余问题。张智等[7]针对舰载机牵引系统的甲板调运过程，引入行为动力学方法，实现牵引系统在复杂障碍物环境中的避

收稿日期：2023-07-03；录用日期：2024-02-23；网络出版时间：2024-04-09 19:23
网络出版地址：link.cnki.net/urlid/11.2625.V.20240408.1252.001
基金项目：陕西省2022年自然科学基础研究计划(2022JM-249)；陕西省重点研发计划(2023-YBSF-104)；长安大学本科实践类课程建设项目(2025-05)
*通信作者. E-mail：zhangjun@chd.edu.cn

*引用格式：张军，黄明辉，王玥琳，等. 基于MPC的飞机牵引车轨迹跟踪[J]. 北京航空航天大学学报，2025，51(9)：2916-2926.
ZHANG J，HUANG M H，WANG Y L，et al. Trajectory tracking for aircraft tug based on MPC[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2916-2926 (in Chinese).*

碰路径规划。Meng 等[8] 采用目标偏置方案和快速碰撞检测技术, 利用 RRT 算法规划出无碰撞风险下的最快工作路径。Liu 等[9] 考虑牵引过程中姿态变化, 将牵引车-飞机系统转化为拖车系统, 基于滚动时域控制理论和离线最优控制算法提出了一种在线跟踪方法, 但未考虑机场环境特点及整体碰撞风险。Liu 等[10] 采用离线路径规划法在甲板场景下开展了仿真实验, 但避障所使用的范围过大。张震男[11] 提出一种基于几何理论扩展的 A*路径规划算法, 用 Reeds-Shepp 曲线优化了规划结果, 解决了曲率突变问题。目前, 牵引车的路径规划仍更多地被视为平面移动机器人的避碰问题, 由于在机场环境中, 场地及障碍物相对固定, 具有计算量小、路径平滑等优点的 A*算法可作为飞机牵引车全局路径优化方法。

轨迹跟踪控制的主要任务是根据决策规划层输出的参考路径, 控制车辆的驱动、制动和转向等动作, 从而准确跟踪参考路径。现有研究多依据预瞄理论和模型预测理论进行控制算法设计, 如纯追踪算法[12]、线性二次型调节器 (linear quadratic regulator, LQR) 算法[13] 及 MPC 算法[14]。张攀等[15] 基于纯追踪算法对飞机牵引车系统的路径进行跟踪控制, 最终规划出偏差小、控制精度更高的轨迹。Yu 等[16] 设计了一种双层闭环控制器, 用于在不完全约束和各种物理条件的约束下, 实现甲板上飞机牵引车的轨迹跟踪。Werner 等[17] 提出了一种针对牵引车-拖车控制的 LQR 调节器, 能够对拖车位置进行控制, 但当跟踪曲率较大的曲线时, 需要引入前馈控制进行补偿。Bin 等[18] 基于混合逻辑动力学对牵引车-拖车系统进行建模, 提出了一种基于 MPC 算法的控制框架, 设计了一种倒车最优控制器。Backman 等[19] 将非线性模型预测控制(nonlinear model predictive control, NMPC)应用到牵引车-拖车系统的轨迹跟踪控制中, 实现了牵引车和拖车对指定直线和曲线的跟踪控制。与纯追踪算法和 LQR 算法相比, MPC 算法拥有较强的约束处理能力、优良的滚动优化控制机制、综合考虑系统当前状态和输出信息的校正思想, 广泛应用于自动驾驶车辆轨迹跟踪控制。

为提高飞机牵引车的智能化水平, 本文建立了飞机牵引车的运动学模型, 结合飞机防碰撞处理及最小转弯半径的安全性限制改进了 A*算法, 获得牵引车的全局路径。采用 MPC 算法进行在线轨迹跟踪控制, 在 MATLAB/Simulink 与 ADAMS 机电仿真模型中进行算例验证和控制器参数优化, 采用北斗 RTK GPS 获取样机的定位、姿态及运动信息, 使用 CAN 总线通信技术对样机进行改造, 使其具备无人驾驶功能。开展轨迹跟踪实验, 实现了飞机牵引车的无人驾驶功能。本文方法可为智慧物流机场的牵引车无人驾驶提供参考。

1　牵引车路径规划

1.1　牵引车运动学建模

本文采用的 10 t 电传动飞机牵引车与文献 [20] 一致, 2 个前轮为万向惰轮, 2 个后轮为驱动轮。根据牵引车的运动特征, 建立笛卡儿直角坐标系 xOy下的运动学模型, 如图 1 所示, 其中, 坐标原点为分析牵引车运动学的坐标系参考点。图中: M_1为牵引车后轴中心, M_2为飞机后轴中心, β为牵引车转向下自身的航向角, θ为飞机与牵引车之间形成的夹角, λ为飞机转向时自身的转动角, α为牵引车的万向前轮偏角, L为牵引车驱动轮轮距, d_1为飞机轴距, d_2为牵引车轴距, O_1为飞机的瞬时转向中心, O_2为牵引车的瞬时转向中心, O_1M_2为飞机转弯半径, 记为R_1, O_2M_1为牵引车转弯半径, 记为R。

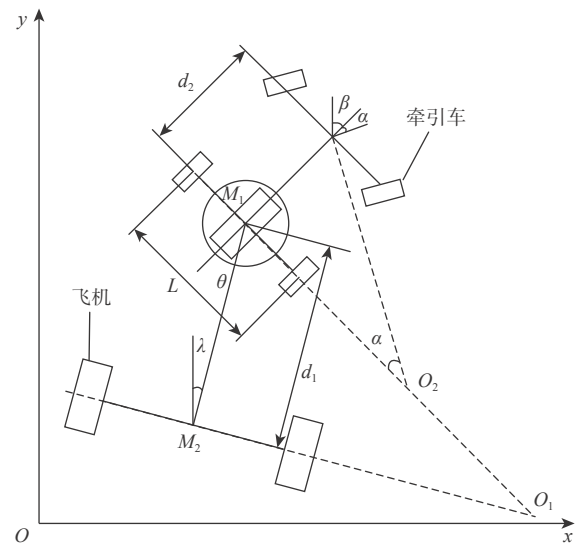


图 1　牵引车运动学模型

Fig. 1　Motion model of tug

由图 1 可知, 电传动飞机牵引系统运动学方程可表示为

$$\begin{cases} \dot{x} = v\sin\beta \\ \dot{y} = v\cos\beta \\ \dot{\beta} = \dfrac{v\tan\alpha}{d_1} \\ \dot{\theta} = \dfrac{v\sin\beta}{d_2} + \dfrac{v\tan\alpha}{d_1} \\ \beta = \theta + \lambda \end{cases} \tag{1}$$

式中: x、y分别为牵引车后轴中心点M_1的横、纵坐标; v为牵引车后轴中心点M_1的速度。

进一步, 将运动学模型表示为一般形式:

$$\dot{\boldsymbol{\chi}} = f(\boldsymbol{\chi}, \boldsymbol{u}) \tag{2}$$

式中：状态量$\boldsymbol{\chi}=[x,y,\beta]^{\mathrm{T}}$；控制量$\boldsymbol{u}=[v,\omega]^{\mathrm{T}}$，$\omega$为牵引车航向角速度。

为实现牵引车控制，将轨迹跟踪系统输出的控制量$[v,\omega]$进一步转换至车辆左右两侧电机车速V_{L}、V_{R}，再转化为对应的电机转速n_{L}、n_{R}。

由图1可得牵引车的相对牵引角速度ω为

$$\omega=v\left(\frac{\sin\theta}{d_1}-\frac{1}{R}\right) \tag{3}$$

式(3)中，被牵引飞机前轮的左右转向角度$\theta\leqslant 40°$，以$\omega\geqslant 0$和飞机轴距$d_1=4\,754$ mm代入式(3)，可获得牵引状态下最小转弯半径$R_{\min}\geqslant 7\,395.9$ mm，取$R_{\min}=8.0$ m；而在非牵引状态下$R_{\min}=0$，可原地转向。

1.2 A*算法

A*算法是采用启发式方法的图搜索路径规划算法，实现流程步骤如下：

步骤1 设置创建2个表Open List与Close List，其中，Open List用于存放可扩展节点，可扩展节点指当前节点周围除障碍物、边界及已被扩展节点外的节点，Close List用于存放已扩展过的节点，并将起点S放入Open List中。

步骤2 寻找起点S附近可扩展节点并放入Open List中，当存在可扩展节点时，将点S弹出至Close List中，并利用代价函数$f(n)$计算节点附近所有可扩展节点代价值，选取估计代价最低节点作为下一移动节点；若起点S附近无可扩展节点，则无法规划可行路径。

步骤3 判断当前Close List是否已包含终点，包含终点则路径规划成功，否则进行步骤4。

步骤4 寻找当前节点n附近所有可扩展节点，并利用代价函数$f(n)$计算节点n附近所有可扩展节点代价值，选择代价值最小节点m作为下一移动节点，并将当前节点n移出Open List，存入Close List。

步骤5 搜寻Open List中是否存在节点m，若存在，比较前后2次节点m的实际代价$s'(m)$与$s(m)$，取二者中小值替换$s(m)$；若不存在，则将节点m存入Open List。循环步骤3～步骤5，直至搜索到终点或Open List为空。

1.3 改进A*算法

为确保转向过程中飞机的安全性，对传统A*算法进行防碰撞处理及最小转弯半径限制处理，最小转弯半径在后续的路径平滑算法中也进行约束，使得所规划路径满足飞机牵引车防碰撞安全要求及运动约束条件。

1.3.1 防碰撞处理

防碰撞处理是保证所规划路径安全性至关重要的部分，先对飞机牵引车工作场地及障碍物进行栅格化处理，再将连续的空间离散化为离散状态的栅格，在栅格地图中体现飞机牵引车工作场地(如机场跑道、廊桥等)道路特征。道路及障碍物所占栅格数量由其实际占地面积决定。

基于上述内容及飞机牵引车-飞机系统实际尺寸12 000 mm×9 480 mm×4 800 mm，本文围绕飞机牵引车及机体构建半径为6 m的防碰撞安全范围，如图2所示。通过在栅格地图中障碍物原本位置及范围上，加上防碰撞安全区域，由此可在栅格地图中将飞机牵引车-飞机系统视为其质心位置的质点以进行路径规划。

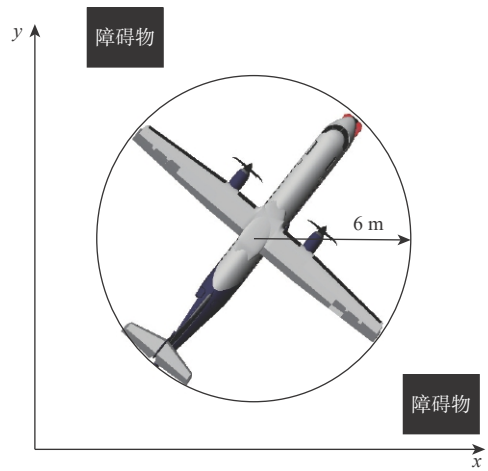


图2 防碰撞安全范围

Fig. 2 Anti-collision safety range

1.3.2 最小转弯半径限制

1.1节计算了飞机牵引车在牵引状态下的最小转弯半径，因此，在进行路径规划时，将$R\geqslant 8$ m设为路径规划的硬约束，确保路径各处最小转弯半径均满足要求。

1.3.3 路径平滑算法

为提升路径的平顺性，对得到的离散路径点进行曲线拟合，以获得使飞机牵引车平稳前进的路径。

设曲线平滑前的点$\boldsymbol{x}(i)$，$\boldsymbol{x}(i)\in\boldsymbol{X}$，平滑后的点$\boldsymbol{y}(i)$，$\boldsymbol{y}(i)\in\boldsymbol{Y}$，定义路径平滑的代价函数$F$为

$$F=a_1\|\boldsymbol{x}(i)-\boldsymbol{y}(i)\|+a_2\|\boldsymbol{y}(i)-\boldsymbol{y}(i+1)\| \tag{4}$$

式中：a_1和a_2为平滑路径权重系数。式(4)第1项为衡量平滑后的点偏离原始点的程度，第2项为衡量平滑点之间的距离。这2项相互制衡，平滑的过程就是最小化代价函数的过程。

采用共轭梯度下降法求解代价函数的最小值，平滑前设定初值$\boldsymbol{y}(i)=\boldsymbol{x}(i)$。遍历所有点，对$\boldsymbol{Y}$进行更新，更新方法如下：

$$\boldsymbol{y}(i)=\boldsymbol{x}(i)+a_1(\boldsymbol{x}(i)-\boldsymbol{y}(i))+a_2(\boldsymbol{y}(i-1)-2\boldsymbol{y}(i)+\boldsymbol{y}(i+1)) \tag{5}$$

将式(5)中a_1和a_2均设置为 0.5, 右侧$\boldsymbol{y}(i)$为更新前对应的$\boldsymbol{x}(i)$值, 通过不断迭代求解目标函数最优解, 实现路径平滑, 最终获得平滑曲线$\boldsymbol{Y}$。

2　轨迹跟踪控制器设计

2.1　MPC 数学模型建立

MPC 算法中, 普遍采用线性时变模型来进行预测控制以减小算法运算量。根据牵引车的运动学模型, 当前牵引车的连续系统模型$\dot{\boldsymbol{\chi}}_i$可表示为

$$\dot{\boldsymbol{\chi}}_i=f(\boldsymbol{\chi}_i,\boldsymbol{u}_i) \tag{6}$$

式中: i为当前连续时间轴时刻; $\boldsymbol{\chi}_i=[x_i,y_i,\beta_i]^{\mathrm{T}}$为当前状态量; $\boldsymbol{u}_i=[v_i,\omega_i]^{\mathrm{T}}$为当前控制量。

采用泰勒展开的方法对式(6)进行处理, 系统的误差模型$\dot{\tilde{\boldsymbol{\chi}}}$可表示为

$$\dot{\tilde{\boldsymbol{\chi}}}=\begin{bmatrix}\dot{x}-\dot{x}_i\\ \dot{y}-\dot{y}_i\\ \dot{\beta}-\dot{\beta}_i\end{bmatrix}=\begin{bmatrix}0&0&-v_i\sin\beta_i\\0&0&v_i\cos\beta_i\\0&0&0\end{bmatrix}\begin{bmatrix}x-x_i\\y-y_i\\\beta-\beta_i\end{bmatrix}+\begin{bmatrix}\cos\beta_i&0\\\sin\beta_i&0\\0&1\end{bmatrix}\begin{bmatrix}v-v_i\\\omega-\omega_i\end{bmatrix} \tag{7}$$

为实现牵引车的控制, 对式 (7) 按采样时间T进行离散化处理, 可得到 MPC 控制器的离散误差模型为

$$\tilde{\boldsymbol{\chi}}(k+1)=\boldsymbol{A}_{k,t}\tilde{\boldsymbol{\chi}}(k)+\boldsymbol{B}_{k,t}\tilde{\boldsymbol{u}}(k) \tag{8}$$

式中: $\boldsymbol{A}_{k,t}=\begin{bmatrix}1&0&-v_iT\sin\beta_i\\0&1&v_iT\cos\beta_i\\0&0&1\end{bmatrix}$; $\boldsymbol{B}_{k,t}=\begin{bmatrix}T\cos\beta_i&0\\T\sin\beta_i&0\\0&T\end{bmatrix}$; k为当前离散时间轴时刻, t为离散时间轴的某一时刻。

2.2　状态预测模型建立

为使轨迹跟踪过程中牵引车轨迹更加贴合规划路径, 构建状态预测模型对牵引车未来路径误差进行预测, 取$\boldsymbol{\xi}(k|t)=\begin{bmatrix}\tilde{\boldsymbol{\chi}}(k|t)\\ \tilde{\boldsymbol{u}}(k-1|t)\end{bmatrix}$可得

$$\begin{cases}\boldsymbol{\xi}(k+1|t)=\tilde{\boldsymbol{A}}_{k,t}\boldsymbol{\xi}(k|t)+\tilde{\boldsymbol{B}}_{k,t}\Delta\boldsymbol{U}(k|t)\\ \boldsymbol{\eta}(k|t)=\tilde{\boldsymbol{C}}_{k,t}\boldsymbol{\xi}(k|t)\end{cases} \tag{9}$$

式中: $\tilde{\boldsymbol{A}}_{k,t}=\begin{bmatrix}\boldsymbol{A}_{k,t}&\boldsymbol{B}_{k,t}\\ \boldsymbol{0}_{m\times n}&\boldsymbol{I}_m\end{bmatrix}$; $\tilde{\boldsymbol{B}}_{k,t}=\begin{bmatrix}\boldsymbol{B}_{k,t}&\boldsymbol{I}_m\end{bmatrix}^{\mathrm{T}}$; $\tilde{\boldsymbol{C}}_{k,t}=\begin{bmatrix}\boldsymbol{C}&\boldsymbol{0}_{n\times m}\\ \boldsymbol{0}_{m\times n}&\boldsymbol{0}_{m\times m}\end{bmatrix}$, $\boldsymbol{C}=\mathrm{diag}(1,1,1)$; n为状态量$\boldsymbol{\chi}=[x,y,\beta]^{\mathrm{T}}$的维度, 即$n=3$; m为控制量$\boldsymbol{u}=[v,\omega]^{\mathrm{T}}$的维度, 即$m=2$; $\Delta\boldsymbol{U}(k|t)$为预测时刻 k 相对于前一时刻 k–1 的控制输入增量; $\boldsymbol{\eta}$ 为系统的输出向量, $\boldsymbol{\eta}(k|t)$ 为基于 t 时刻的信息对 k 时刻的预测; $k+1|t$表示基于t时刻参数对$k+1$时刻系统变量进行预测。

取N为预测时长, 令

$$\begin{cases}\boldsymbol{A}_{k,t}=\boldsymbol{A}_{t,t}\\ \boldsymbol{B}_{k,t}=\boldsymbol{B}_{t,t}\end{cases}\quad k=1,2,\cdots,t+N-1 \tag{10}$$

推导可得状态预测模型为

$$\boldsymbol{G}(t)=\boldsymbol{\psi}_t\boldsymbol{\xi}(t|t)+\boldsymbol{\Theta}_t\Delta\boldsymbol{U}(t) \tag{11}$$

式中: $\boldsymbol{G}(t)=\begin{bmatrix}\boldsymbol{\eta}(t+1|t)\\ \boldsymbol{\eta}(t+2|t)\\ \vdots\\ \boldsymbol{\eta}(t+N_{\mathrm{c}}|t)\\ \vdots\\ \boldsymbol{\eta}\left(t+N_{\mathrm{p}}|t\right)\end{bmatrix}$; $\boldsymbol{\psi}_t=\begin{bmatrix}\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}\\ \tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^2\\ \vdots\\ \tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{c}}}\\ \vdots\\ \tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{p}}}\end{bmatrix}$;

$\Delta\boldsymbol{U}(t)=\begin{bmatrix}\Delta u(t|t)\\ \Delta u(t+1|t)\\ \vdots\\ \Delta u(t+N_{\mathrm{c}}|t)\end{bmatrix}$; N_{c}为输入预测模型的时域数量; N_{p}为模型预测未来输出的时域数量。

$$\boldsymbol{\Theta}_t=\begin{bmatrix}\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{B}}_{t,t}&0&0&\cdots&0&0\\ \tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{B}}_{t,t}&0&\cdots&0&0\\ \vdots&\vdots&\vdots&&\vdots&\vdots\\ \tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{c}}-2}\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{c}}-3}\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{c}}-4}\tilde{\boldsymbol{B}}_{t,t}&\cdots&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{B}}_{t,t}&0\\ \tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{c}}-1}\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{c}}-2}\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{c}}-3}\tilde{\boldsymbol{B}}_{t,t}&\cdots&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{B}}_{t,t}\\ \tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{c}}}\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{c}}-1}\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{c}}-2}\tilde{\boldsymbol{B}}_{t,t}&\cdots&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^2\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}\tilde{\boldsymbol{B}}_{t,t}\\ \vdots&\vdots&\vdots&&\vdots&\vdots\\ \tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{p}}-1}\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{p}}-2}\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{p}}-3}\tilde{\boldsymbol{B}}_{t,t}&\cdots&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{p}}-N_{\mathrm{c}}+1}\tilde{\boldsymbol{B}}_{t,t}&\tilde{\boldsymbol{C}}_{t,t}\tilde{\boldsymbol{A}}_{t,t}^{N_{\mathrm{p}}-N_{\mathrm{c}}}\tilde{\boldsymbol{B}}_{t,t}\end{bmatrix}$$

根据系统在t时刻的运动状态量及路径误差, 预测未来N_{p}个时域内系统的状态量, 可实现飞机牵引车轨迹跟踪过程的状态预测。

2.3　目标函数构建

为保证控制过程的精确性和稳定性, 控制目标为牵引车在轨迹跟踪过程中保持尽可能小的轨迹误差及控制量波动, 目标函数如下:

$$J(k)=\sum_{i=1}^{N_{\mathrm{p}}}\|\boldsymbol{\eta}(k+i|t)-\boldsymbol{\eta}_{\mathrm{ref}}(k+i|t)\|_{\boldsymbol{Q}}^2+\sum_{i=1}^{N_{\mathrm{c}}-1}\|\Delta\boldsymbol{U}(k+i|t)\|_{\boldsymbol{P}}^2+\rho\varepsilon^2 \tag{12}$$

式中：ρ为权重系数；ε为松弛因子；$\boldsymbol{Q}$和$\boldsymbol{P}$分别为状态权重矩阵和控制权重矩阵。

目标函数$J(k)$由2项组成，前1项反映了状态量误差及对轨迹点的跟踪效果，后1项反映了控制量的变化情况。为避免硬约束造成优化问题无法求解的情况出现，目标函数添加软约束。根据牵引车技术要求和运动限制，对轨迹跟踪系统输出的速度、角速度及其增量进行约束，牵引过程车速限制为$-1\ \text{m/s} \leqslant v \leqslant 1\ \text{m/s}$，车速变化限制为$-0.2\ \text{m/s} \leqslant \Delta v \leqslant 0.2\ \text{m/s}$；牵引过程最小转弯半径$R_{\min} \geqslant 8\ \text{m}$，车辆角速度输出限制为：$-0.125\ \text{rad/s} \leqslant \omega \leqslant 0.125\ \text{rad/s}$，$-0.025\ \text{rad/s} \leqslant \Delta\omega \leqslant 0.025\ \text{rad/s}$。

2.4 轨迹跟踪控制算法

牵引车轨迹跟踪MPC控制器组成如图3所示，系统由数学模型、约束条件和目标函数构成。约束条件包括牵引车速度、加速度及与飞机夹角角度等；数学模型是用式(11)描述被控制的牵引车系统，目标函数按照式(12)设计，并在MATLAB/Simulink中构建MPC控制器代码。

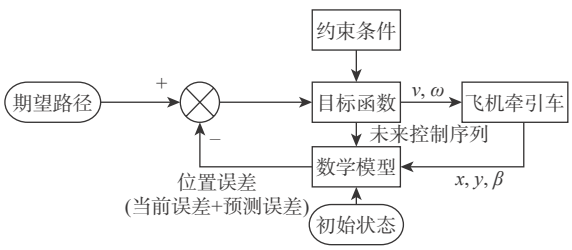


图3 MPC控制器组成

Fig. 3 MPC controller composition

基于MPC的轨迹跟踪具体算法步骤如下：①由牵引车路径规划系统生成满足飞机牵引车防碰撞及最小转弯半径要求的路径；②将规划路径离散化成一系列的离散路径点；③从第一个离散路径点开始，MPC控制器通过计算当前位置与期望位置偏差，跟踪选取的路径点，同时判断当前位置是否为路径终点，到达终点即轨迹跟踪过程结束。

3 牵引车联合仿真实验

3.1 联合仿真模型构建

以.stp格式将飞机牵引车和飞机的三维模型导入到ADAMS软件中，添加轮胎、地面等元素并与原三维模型部分相重合，获得图4所示的牵引车动力学模型。其中，轮胎采用ADAMS中的PAC89轮胎模型，路面模型采用2d_flat.rdf模型。对模型中非必要组成成分进行删除简化，定义轮胎与车体之间为转动、飞机起落架前轮与抱轮装置之间为固定、飞机机体与起落架之间为转动等约束。定义重力方向，对牵引车2个后驱动轮添加Motion模块，

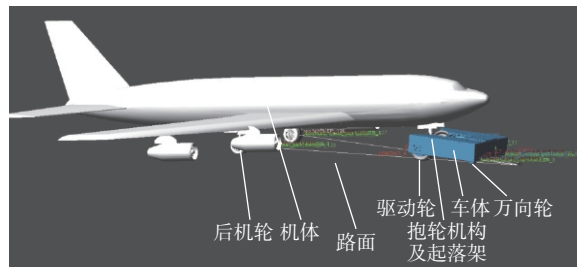


图4 电传动无杆飞机牵引车动力学模型

Fig. 4 Dynamics model of electric drive aircraft tug

定义转速及转动方向等属性。设置Control模块，定义ADAMS与MATLAB/Simulink联合仿真的输入输出变量，其中，输入为左右驱动轮转速n_{L}和n_{R}，输出为车体坐标x和y、线速度v、角速度ω等属性，变量名与前文运动学分析的变量一致。

将ADAMS与MATLAB/Simulink进行联合建模，建立如图5所示的牵引车-飞机系统的机电联合仿真模型。运动控制模块的输入为MPC轨迹跟踪控制系统输出的车体角速度ω及线速度v；其输出是根据运动学模型将这2个变量转化为飞机牵引车两驱动轮转速，再输入至ADAMS模块输入口。对角速度等进行积分处理以获得飞机牵引车及飞机的航向角变化，以二者之差来观测飞机牵引车-飞机系统在牵引过程中连接角度变化情况。

3.2 路径规划实验

设置2个典型场景：场景1为机场弯道场景(如飞机在航站楼登机后到滑行道的转弯)，场景2为直线滑行过程存在障碍的道路场景。仅考虑全局路径规划，采用改进A*算法规划所得离散路径点，分别利用最小二乘法与梯度下降法进行曲线拟合，并进行实验对比。

2种场景的轨迹规划结果如图6所示。从图中对比可知，经过梯度下降法平滑的路径与原路径偏差小，更为平滑，贴合初始曲线，符合飞机牵引车驾驶特点。

将平滑后路径与初始路径的数据进行统计分析，通过方差、标准差及路径本身平滑性来衡量所平滑路径的优劣，结果如表1所示。梯度下降法在2个场景下的性能最优，因此，将梯度下降法用于改进A*算法的路径平滑。上述结果也表明改进A*算法所规划路径在保持了与路沿、障碍物间充分的安全距离及满足最小转弯半径要求的前提下，能实现最优路线规划。

3.3 轨迹跟踪实验

在MPC控制方法中，预测步长N_{p}及控制步长N_{c}决定了目标轨迹跟踪的精度及行驶平滑性，这

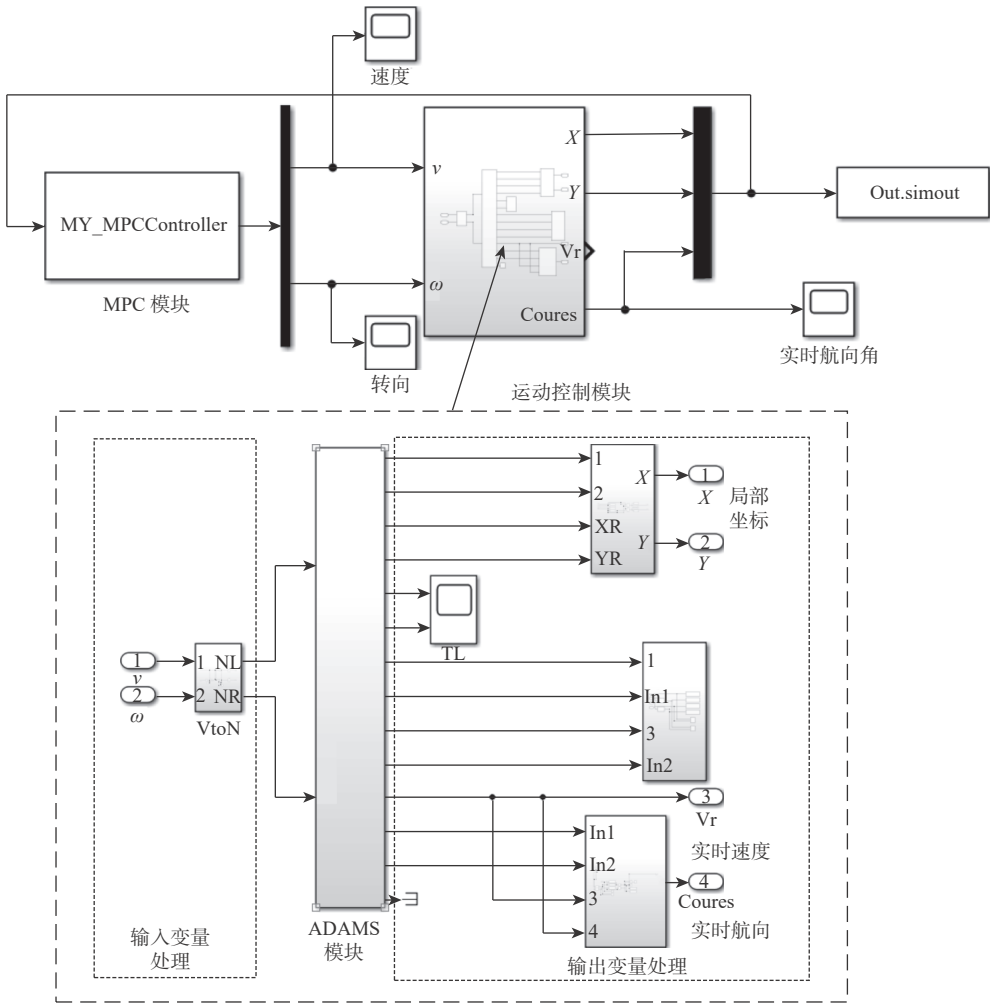


图 5　牵引车-飞机系统的机电联合仿真模型

Fig. 5　Co-simulation model of tug-aircraft system

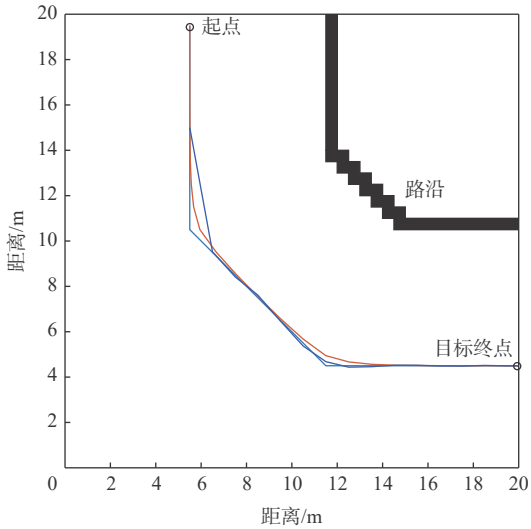
2 个参数对飞机牵引车的无人驾驶控制至关重要, 过小的预测步长会使得车辆在运行至轨迹变化较大处出现较大误差的情况, 而过大的预测步长会导致系统计算量过大; 同理, 合适的控制步长可以更好地贴合规划路径, 过大的控制步长也会导致过大的计算量, 因此, 需要通过实验确定合适的控制参数。

在 MATLAB 2021 与 ADAMS 2019 中进行联合仿真, 硬件环境为 Intel Core i5-9300HF @ 2.40 GHz CPU, 16 GB RAM, NVDIA GTX 1660 TI GPU。以仿真场景 1 的弯道转向为例, 取多个不同的预测步长及控制步长进行仿真实验, 通过对比控制结果的标准差和耗时来选取合适的控制参数范围。MPC 控制中, $N_p \geqslant N_c$, 通过单因素实验, 固定 N_p=60, 将 N_c 从 5 逐步增大到 60 进行实验, 发现控制系统误差随 N_c 增大而减小, 仿真时间随 N_c 增大而增大; 当 $N_c \geqslant$ 30 时, 仿真时间大于 7 s, 运算时间过长。为获得合适的控制器参数组合, 设定 $N_c \leqslant$30, 在此基础上进行 N_p 参数对比实验, 考虑了 8 种实验参数组合, 结果如表 2 所示。结果表明, 最适合飞机牵引车-飞机系统的控制参数组合为组合 8, 即 N_p=5, N_c=5, 因此, 采取该组参数作为 MPC 控制器的参数。

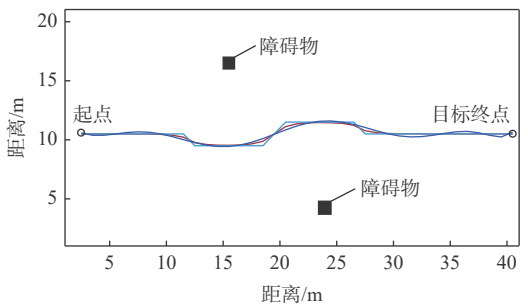
将优化的全局轨迹和 MPC 控制器用于联合仿真控制实验, 在 2 个场景下进行实验, 轨迹跟踪结果如图 7 所示。可知, 设计的 MPC 控制器可有效实现轨迹跟踪, 验证了本文算法的有效性。

图 7 所示 2 个场景仿真结果都很好地避开了路沿及障碍物, 并保持了一定的安全距离。对 2 个场景的仿真跟踪结果与规划路径进行误差分析, 结果如表 3 所示, 2 种场景在仿真情况下皆展示出较小的轨迹误差。

为进一步控制算法的有效性, 避免飞机存在转向角超过前起落架极限的情况, 在 ADAMS 模型输出端监测牵引车与飞机间连接处夹角。以角度变化最大的弯道场景为例, 结果如图 8 所示, 弯道转向过程中有一次角度波动, 连接夹角 θ 最大为 21.19°, 满足工作过程中 $\theta \leqslant$40°的技术要求, 符合牵引飞机的安全性要求。



(a) 仿真场景1



(b) 仿真场景2

改进 A* 算法 基于最小二乘法的曲线拟合
基于梯度下降的曲线拟合

图 6 仿真场景轨迹规划结果

Fig. 6 Compared simulated path planning results

表 1 路径平滑仿真结果

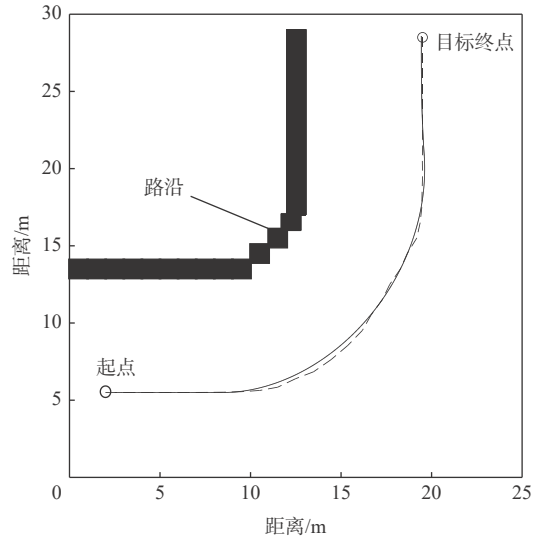
Table 1 Path smoothing simulation results

场景	方法	方差/m^2	标准差/m	平滑性
仿真场景1	最小二乘法	1.851	1.360	一般
	梯度下降法	0.015	0.126	好
仿真场景2	最小二乘法	0.026	0.161	一般
	梯度下降法	0.012	0.111	好

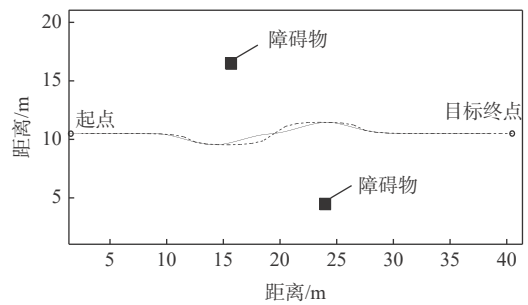
表 2 不同控制参数下的 MPC 控制效果对比

Table 2 Compared results at different MPC control parameters

参数组合	N_c	N_p	标准差/m	仿真时间/s
1	30	60	0.267	6.95
2	30	50	0.247	6.39
3	30	40	0.231	6.23
4	30	30	0.218	5.92
5	20	20	0.204	5.07
6	10	10	0.182	3.73
7	5	10	0.183	2.83
8	5	5	0.157	2.78



(a) 仿真场景1



(b) 仿真场景2

规划轨迹 跟踪轨迹

图 7 联合仿真实验结果

Fig. 7 Results of joint simulation experiments

表 3 联合仿真轨迹结果

Table 3 Joint simulation trajectory results

场景	方差/m^2	标准差/m
仿真场景1	0.021	0.146
仿真场景2	0.016	0.128

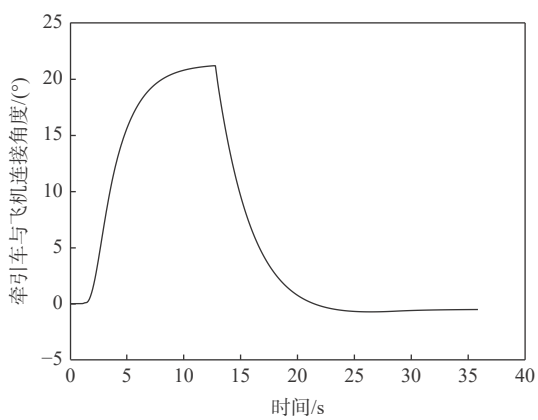


图 8 弯道下牵引车与飞机的连接角度变化

Fig. 8 The connection angle change

4 工程实验

4.1 无人驾驶牵引车平台搭建

采用 CAN 总线通信和上下位机主从控制方法

对 10 t 飞机牵引车进行无人驾驶功能改造。上位机为 MATLAB/Simulink 平台, 负责轨迹规划、MPC 控制、GPS 定位信息和采集下位机转速参数, 实现轨迹跟踪中的速度反馈调节, 提高轨迹跟踪控制精度。下位机为牵引车的 PLC 控制系统, 在保留原有的遥控功能情况下增加了通信指令控制方式。当 PLC 接收上位机的控制指令后, 按照上位机的目标来控制牵引车动作。上位机通过 PCAN-USB 设备与下位机进行 CAN 总线通信, 通信波特率为 250 Kbit/s。

无人驾驶飞机牵引车软件包含 GPS 数据解析及坐标转换、期望速度设置、MPC 轨迹跟踪控制、左右驱动轮的转速解算、电机转速 PID 控制、CAN 通信模块和驱动电机执行部分。驱动电机采用速度模式, 通过 CAN 通信实现电机动作。由于车辆位姿和路径依赖于 GPS 信息, 系统需先判断 GPS 信号源是否稳定, 信号稳定才进行局部坐标系转换, 再按照规划的轨迹运行 MPC 轨迹跟踪, 输出牵引车左右轮的目标转速, 实现车辆控制。

建立样机的上位机控制系统如图 9 所示, 包括 MPC 模块、运动指令转换模块、车辆通信模块及定位通信模块组成。相比图 5 中的仿真模型, 样机控制系统上增加了 CAN 通信、RTK-GPS 数据解析功能和真实采样时间模块, 使得上位机系统时间与真实时间保持一致, 采样时间为 0.1 s。实际车辆的通信模块替换 ADAMS 的动力学模型, 实现车辆车速和转向控制。

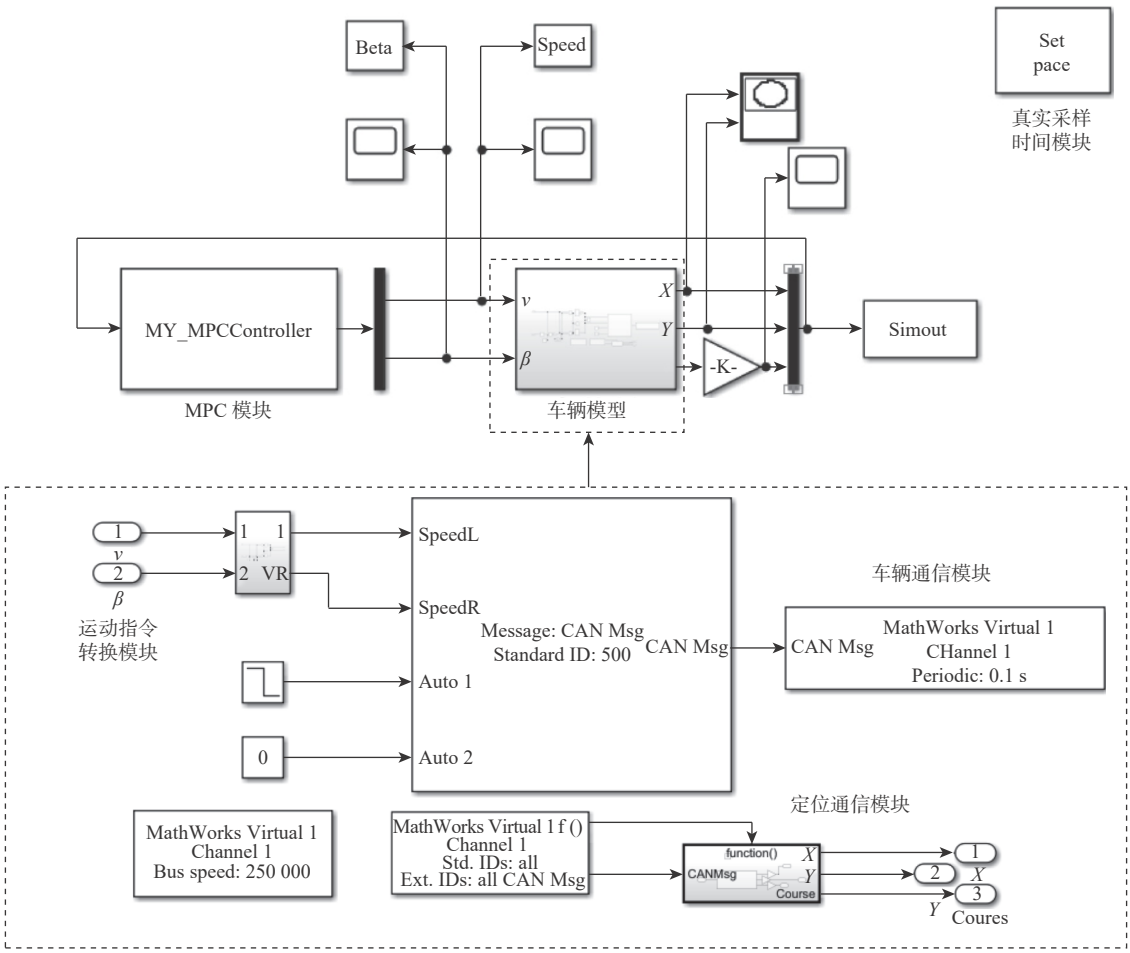


图 9　样机的上位机程序

Fig. 9　Overall structure of upper computer software

RTK-GPS 接收到的数据为 WGS84 坐标系下的 KSXT 数据结构定位信息, 包含了时间、经度、纬度、海拔、俯仰角、航向、速度、差分状态及接收卫星信号数量。为实现轨迹规划, 将 GPS 坐标转化为二维的 UTM 投影坐标系, 再转换为局部坐标系的定位和航向信息, 输入给 MPC 控制器。

按照上述控制策略, 对 10 t 飞机牵引车进行改装, 如图 10 所示。通过沿车体中轴线布置前后 2 个高精度 GPS-RTK 流动站设备的方式, 可以精确

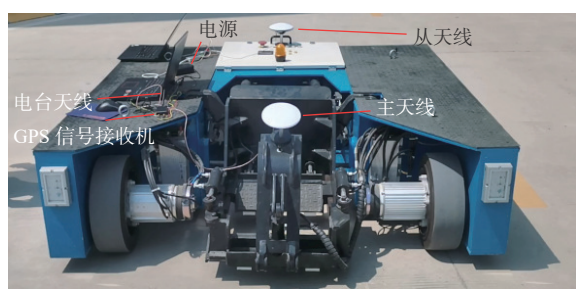


图 10　无人驾驶飞机牵引车

Fig. 10　Autonomous aircraft tug

获取牵引车车体的定位、航向、运动速度及方向。构建局域网 RTK-GPS 系统，2 个 BT300 GPS 天线分别置于车辆中心线上，通过千寻 D300 电台接收机将 GPS 信号传给 PC 机。为减轻 MPC 控制主机的负载，增加一台 PC 主机用于 GPS 信息解析，通过 CAN 发到总线后再被主机接收，获得车辆当前位置信息。

4.2 室内轨迹跟踪实验

实验地点为位于西安市高新区的军民融合产业园，弯道和直线避障场景如图 11 所示。其中，直线避障场景所选取路段的道路宽度为 10 m，长度为 40 m，使用 2 把 0.5 m×0.5 m 的椅子模拟道路中的固定障碍物。

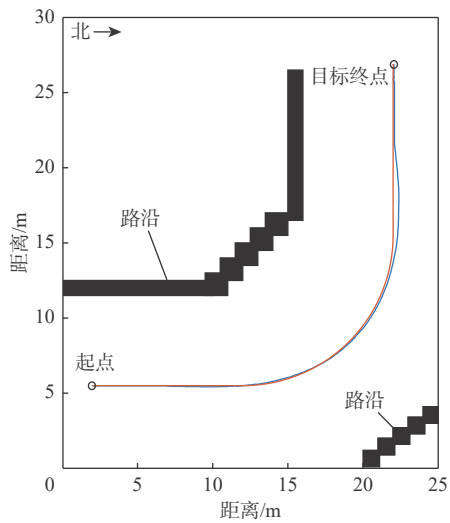
(a) 实验场景1 (b) 实验场景2

图 11 实验场地

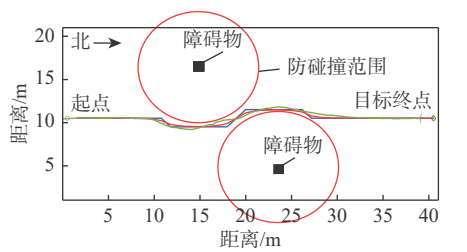
Fig. 11 Experimental site

基于图 11 所示实验场地构建数字地图，采用 UTM 坐标系，y 轴正方向为正东方向，x 轴方向为正北方向，并进行路径规划，MPC 控制器参数取 N_p=5，N_c=5。考虑飞机牵引车牵引工作运动限制，取 v=0.8 m/s 作为轨迹跟踪过程期望速度。

实验结果如图 12 所示。实验跟踪轨迹与规划轨迹整体误差较小，且整体轨迹及移动过程较为平滑稳定，满足飞机牵引车牵引工作的要求。在直线避障中与障碍物保持 6 m(牵引车和飞机总体半径< 6 m)以上的安全距离，达到了预期目标。轨迹跟踪的统计结果如表 4 所示。发现实际实验时弯道转向场景实验轨迹标准差为 0.362 m，大于仿真结果；直线避障场景实验轨迹标准差为 0.128 m，与仿真误差接近。主要原因是：实验场地有一定的局限性，弯道场地较短，直线场景路径较长。由于牵引车转速控制误差会导致控制误差增大，相比于直线场景，在较短的弯道上会出现较大的误差，若增加测试距离，标准差会减小，但园区内无法开展大范围转弯场景实验。现场实验验证了本文所提轨迹跟踪方法的有效性。



(a) 实验场景1 (弯道转向)



(b) 实验场景2 (直线避障)

图 12 轨迹跟踪实验结果

Fig. 12 Field result of path tracking

表 4 轨迹跟踪误差分析

Table 4 Track tracking error analysis

场景	方差/m^2	标准差/m
实验场景1	0.131	0.362
实验场景2	0.016	0.128

5 结 论

1) 建立了 10 t 飞机牵引车的运动学模型，并基于飞机前起落架的转弯角度约束，获得了牵引车在牵引状态下最小转弯半径为 8 m，基于防碰撞需求、最小转弯半径限制和路径平滑对 A*算法进行改进，实现了牵引车牵引过程的轨迹规划功能。

2) 设计了飞机牵引车的 MPC 控制器，并采用 ADAMS 与 MATLAB/Simulink 联合仿真的方法优化了 MPC 控制器参数，通过仿真实验验证了算法的有效性，实现了较好的控制精度。

3) 搭建了无人驾驶牵引车样机平台，并在样机上开展了轨迹跟踪实验，实现了飞机牵引车的无人驾驶控制功能。

后期将开展环境感知实验，加入局部避障策略，以实现牵引车的无人驾驶功能。

参考文献（References）

[1] 李永丹, 马天力, 陈超波, 等. 无人驾驶车辆路径规划算法综述[J]. 国外电子测量技术, 2019, 38(6): 72-79.
LI Y D, MA T L, CHEN C B, et al. Review of path planning algorithm for unmanned vehicles[J]. Foreign Electronic Measurement Technology, 2019, 38(6): 72-79(in Chinese).

[2] 刘军, 冯硕, 任建华. 移动机器人路径动态规划有向 D*算法[J]. 浙江大学学报 (工学版), 2020, 54(2): 291-300.
LIU J, FENG S, REN J H. Directed D* algorithm for dynamic path planning of mobile robots[J]. Journal of Zhejiang University (Engineering Science), 2020, 54(2): 291-300(in Chinese).

[3] 陈秋莲, 蒋环宇, 郑以君. 机器人路径规划的快速扩展随机树算法综述[J]. 计算机工程与应用, 2019, 55(16): 10-17.
CHEN Q L, JIANG H Y, ZHENG Y J. Summary of rapidly-exploring random tree algorithm in robot path planning[J]. Computer Engineering and Applications, 2019, 55(16): 10-17(in Chinese).

[4] NOREEN I, KHAN A, HABIB Z. Optimal path planning using RRT* based approaches: a survey and future directions[J]. International Journal of Advanced Computer Science and Applications, 2016, 7(11): 97-107.

[5] KARAMAN S, WALTER M R, PEREZ A, et al. Anytime motion planning using the RRT[C]//Proceedings of the IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2011: 1478-1483.

[6] 孙家玮, 余明晖, 杨大鹏, 等. 基于 CL-RRT 与 MPC 的舰载机牵引系统路径规划[J]. 系统工程与电子技术, 2024, 46(5): 1745-1755.
SUN J W, YU M H, YANG D P, et al. Path planning of carrier based aircraft traction system based on CL-RRT and MPC[J]. Systems Engineering and Electronic Technology, 2024, 46(5): 1745-1755(in Chinese).

[7] 张智, 林圣琳, 邱兵, 等. 舰载机牵引系统甲板调运避碰路径规划[J]. 系统工程与电子技术, 2014, 36(8): 1551-1557.
ZHANG Z, LIN S L, QIU B, et al. Collision avoidance path planning of carrier aircraft traction system indispatching on deck[J]. Systems Engineering and Electronics, 2014, 36(8): 1551-1557(in Chinese).

[8] MENG X L, WANG N J, LIU Q H. Aircraft parking trajectory planning in semistructured environment based on kinodynamic safety RRT[J]. Mathematical Problems in Engineering, 2021, 2021: 3872248.

[9] LIU J, DONG X Z, WANG J Y, et al. A novel EPT autonomous motion control framework for an off-axle hitching tractor-trailer system with drawbar[J]. IEEE Transactions on Intelligent Vehicles, 2021, 6(2): 376-385.

[10] LIU J, HAN W, PENG H J, et al. Trajectory planning and tracking control for towed carrier aircraft system[J]. Aerospace Science and Technology, 2019, 84: 830-838.

[11] 张震男. 牵引系统路径规划与路径跟踪控制策略研究[D]. 哈尔滨: 哈尔滨工程大学, 2021.
ZHANG Z N. Research on path planning and path tracking control strategy of traction system[D]. Harbin: Harbin Engineering University, 2021(in Chinese).

[12] WALLACE R, STENTZ A, THORPE C, et al. First results in robot road-following[C]//Proceedings of the 9th International Joint Conference on Artificial Intelligence. New York: ACM, 1985: 1089-1095.

[13] 黄海洋, 张建, 王宇, 等. 基于多点预瞄最优控制的智能车辆路径跟踪[J]. 汽车技术, 2018(10): 6-9.
HUANG H Y, ZHANG J, WANG Y, et al. Path traking for intelligent vehicle based on the optimal multipoint preview control[J]. Automobile Technology, 2018(10): 6-9(in Chinese).

[14] CHENG S, LI L, CHEN X, et al. Model-predictive-control-based path tracking controller of autonomous vehicle considering parametric uncertainties and velocity-varying[J]. IEEE Transactions on Industrial Electronics, 2021, 68(9): 8698-8707.

[15] 张攀, 柳阳, 刘新杰, 等. 改进的智能飞机牵引车路径导航纯追踪算法[J]. 计算机工程, 2019, 45(5): 267-271.
ZHANG P, LIU Y, LIU X J, et al. Improved pure pursuit algorithm for intelligent aircraft tractor path navigation[J]. Computer Engineering, 2019, 45(5): 267-271(in Chinese).

[16] YU M H, GONG X, FAN G W, et al. Trajectory planning and tracking for carrier aircraft-tractor system based on autonomous and cooperative movement[J]. Mathematical Problems in Engineering, 2020, 2020: 6531984.

[17] WERNER R, KORMANN G, MUELLER S. Systematic model based path tracking control of actively steered implements in simulation and experiment[J]. IFAC Proceedings Volumes, 2013, 46(18): 85-90.

[18] BIN Y, SHIM T, FENG N L, et al. Path tracking control for backing-up tractor-trailer system via model predictive control[C]//Proceedings of the 24th Chinese Control and Decision Conference. Piscataway: IEEE Press, 2012: 198-203.

[19] BACKMAN J, OKSANEN T, VISALA A. Navigation system for agricultural machines: nonlinear model predictive path tracking[J]. Computers and Electronics in Agriculture, 2012, 82: 32-43.

[20] 张军, 文川, 阳星, 等. 基于 ADRC 的电传动飞机牵引车控制系统设计[J]. 北京航空航天大学学报, 2023, 49(5): 1017-1026.
ZHANG J, WEN C, YANG X, et al. Design of an electric drive-aircraft tug control system based on ADRC[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(5): 1017-1026(in Chinese).

Trajectory tracking for aircraft tug based on MPC

ZHANG Jun[1,*], HUANG Minghui[2], WANG Yuelin[3], YANG Xing[4], YE Min[1], JIA Yongle[5]

(1. Key Laboratory of Road Construction Technology and Equipment of MOE, Chang'an University, Xi'an 710064, China;
2. China Road & Bridge Corporation, Beijing 100010, China;
3. Engineering Research Center of Ministry of Education for Digital Road Infrastructure, Chang'an University, Xi'an 710064, China;
4. Xi'an ASN Technology Group Co., Ltd., Xi'an 710065, China;
5. Xi'an Ruinuo Aviation Equipment Co., Ltd., Xi'an 710060, China)

Abstract: To meet the short and high frequency demands of quick aircraft towing at logistics airports, a rapid towing method based on unmanned driving technology is proposed. Using the technical route and method of "theoretical modeling-algorithm design-case test and simulation optimization-prototype experiment", a kinematic model of the towing vehicle is constructed for a 10-ton aircraft towing vehicle, the vehicle's constraint conditions and control quantities are determined, and the A* algorithm is improved by adding collision avoidance processing, minimum turning radius, and path smoothing to generate the motion trajectory of the towing vehicle. A trajectory tracking controller based on model predictive control (MPC) is designed, and a joint simulation model of MATLAB/Simulink and ADAMS is constructed. The control parameters of MPC are optimized through trajectory tracking simulation experiments, and trajectory tracking experiments are carried out on the modified electric aircraft towing vehicle prototype. The results show that the improved A* algorithm meets the requirements of aircraft towing vehicle work path planning and minimum turning radius, the joint simulation method optimizes the MPC controller, achieves good tracking accuracy on the prototype, with standard deviation of tracking errors for curve paths and straight paths being 0.362 m and 0.128 m respectively, realizes the unmanned driving function of the aircraft towing vehicle, and lays the technical foundation for unmanned aircraft towing at smart logistics airports.

Keywords: electric aircraft towing vehicle; unmanned driving technology; path planning; improved A* algorithm; trajectory tracking; model predictive control

Received: 2023-07-03; **Accepted**: 2024-02-23; **Published Online**: 2024-04-09 19:23

URL: link.cnki.net/urlid/11.2625.V.20240408.1252.001

Foundation items: Natural Science Basic Research Plan in Shaanxi Province of 2022 (2022JM-249); The Key R & D Project in Shaanxi Province (2023-YBSF-104); Undergraduate Practical Course Development Project of Chang'an University (2025-05)

* **Corresponding author**. E-mail: zhangjun@chd.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0499

航空货运安保供应链弹性测度及优化策略

赵振武，邢肖肖，耿硕*
(中国民航大学 安全科学与工程学院，天津 300300)

摘 要：为提升航空货运安保供应链的弹性水平，针对陆路运输环节偏移指定路线的扰动情形，根据弹性三角形理论进行弹性测度。在恢复阶段，为寻求最优弹性，以弹性最大化为目标，求解恢复水平、设备数量、恢复时间约束下的最优恢复策略，即确定在空侧货站需采取的安检方式、各安检方式的货物检查量和使用的安检设备台数。以深圳宝安国际机场为例，选取路线偏移作为扰动事件进行分析。结果表明：随着货物总量或扰动时长增大，航空货运安保供应链弹性降低；安检速率、安检设备数量、违禁品检出率决定分配货物检查量时各安检方式的优先性；降低扰动市场、增大安全速率、提高违禁品检出率可提高航空货运安保供应链弹性。

关 键 词：航空货运；安保供应链；弹性；恢复策略；短途驳运
中图分类号：X951
文献标志码：A **文章编号**：1001-5965(2025)09-2927-10

中国常见的货运安保模式为位于机场的空侧货站完成货物收运、安检、组板、装机等操作。近年来，为提升货运效率，深圳宝安国际机场等多家单位遵循国际民航组织(International Civil Aviation Organization, ICAO)及中国航空货运安保的相关要求，并基于中国“放管服”的国情，在逐步探索短途驳运安保模式；首先，在陆侧货站完成收运、安检、组板等操作；然后，将安全的货物通过道路运输至空侧货站；最终，在空侧货站完成后续检查及装机操作。国际民航组织2011年提出航空货运安保供应链的概念[1]，即可将其视为航空货运安保模式下各环节主体对货物采取的一系列相互关联的安保控制措施[1]。航空货运安保供应链的性能(安保水平)决定了货物处于安全状态，即未遭受非法干扰的概率。因此，航空货物安保供应链包括货物到达陆侧货站、陆侧货站收运、陆侧货站安检、陆侧货站货物存放、陆侧货站货物装车、陆路运输、货物到达空侧货站、空侧货站收运、空侧货站安检、空侧货站货物存放、货物装机、货物到达目的地机场等环节，然而，短途驳运模式安保环节的扩展增加了航空货运安保供应链面临的扰动因素。扰动因素主要来自于人为失误(各类安保人员工作时粗心大意)、设施设备失效(视频监控、安检设备等发生故障)、环境干扰(暴雨等极端天气干扰驾驶员、押运员视线；道路施工路段封闭导致货车偏移至非法干扰风险更高的路线)和管理缺陷(安保管理程序存在漏洞等)。以上扰动事件将降低安保控制措施的持续性和有效性，导致航空货运安保供应链性能下降，无法有效识别、防止非法干扰行为，进而威胁航空货运安全。因此，在以上背景下，对中国航空货运安保供应链的弹性，即扰动事件发生后，抵抗、适应并将其性能快速恢复到稳定状态的能力进行研究具有必要性。

本文的工作主要与2类文献有关。第1类是航空货运安保相关研究。关于航空货运安保供应链的概念，国际民航组织[1]提出航空货运安保供应链

收稿日期：2023-07-29；录用日期：2023-09-25；网络出版时间：2023-10-07 15:35
网络出版地址：link.cnki.net/urlid/11.2625.V.20230928.1903.002
基金项目：天津市教委科研计划项目(2021SK025)；民航安全能力建设项目(14002500000019J013)
*通信作者. E-mail：1248076228@qq.com

*引用格式：赵振武，邢肖肖，耿硕. 航空货运安保供应链弹性测度及优化策略[J]. 北京航空航天大学学报，2025，51(9)：2927-2936.
ZHAO Z W，XING X X，GENG S. Resilience measurement and optimization strategy of air cargo secure supply chain[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2927-2936 (in Chinese).*

的概念，对航空货运安保供应链提出了设施安保、人员安保、培训、检查、对安全货物的保护、保管链、监督与合规7个方面的安保程序。da Cunha等[2]认为航空货运安保供应链的本质为保护航空货物，确保从货物“封装”之时起，不被引入禁运物品，并随后保护货物，确保货物从装载到航空器之前不遭受未经授权的干扰。王赢[3]认为航空货运安保供应链概念为将货物经由已知托运人交给管制代理人或直接交付至航空器运营人的过程中，各环节主体需承担相应的货物安保控制责任，即从运输到装机，货物始终受到监管和保护。关于航空货运安保供应链的安保模式，Domingues等[4]提出欧盟需加强航空货运安保的立法工作，政策需向基于风险评估的安保模式转移，重点为协同产业驱动的供应链安保模式。赖国基等[5]对比分析中国与ICAO、美国、欧盟关于航空货运安保政策法规、安全策略的差异，提出中国应修订货物安保法律法规且对航空货运安保供应链中参与主体落实安保职责建立政府监管框架，以完善中国航空货运安保供应链。目前，专家学者尚未从弹性视角对航空货运安保供应链实际运行开展研究。

第2类文献与弹性研究相关。弹性研究涉及电力、交通等多个领域。虽各系统运行机制有其自身特性，但对系统弹性进行定量研究的方法较集中：首先，描述系统的性能演化过程，横轴表示时间，纵轴表示系统性能，趋势通常是扰动阶段下降，恢复阶段上升；然后，基于弹性三角形理论进行弹性测度[6-11]，系统性能随时间演化所损失的三角形面积即弹性损失，扰动下的性能曲线与假设未发生扰动时的性能曲线，在扰动期间所覆盖图形面积的比值表征系统弹性；最终，通过研究相关参数变化对弹性的影响提出弹性优化策略[6,11-12]。此外，由于不可能通过消除灾害的发生来保护系统，因此，部分学者从恢复的角度，通过建立恢复策略模型以达到系统弹性最优。陈锦渠等[8]考虑修复资源总量、修复队伍数及站点失效程度约束，以最小化弹性损失为目标，构建了城市轨道交通网络修复策略模型。Dui等[13]提出剩余弹性的概念，并通过最小化剩余弹性，求解了自然灾害事件下海运系统港口和线路的最优恢复顺序。

综上所述，以往对航空货运安保供应链的研究侧重于政策层面，尚未从弹性视角研究安保供应链弹性水平。鉴于此，本文识别需重点关注的扰动事件，对航空货运安保供应链从扰动到恢复阶段的弹性进行测量，针对恢复阶段建立基于弹性的最优恢复策略模型并提出弹性优化策略。与已有研究相比，本文的主要创新之处有以下2点：①本文从弹性角度开展研究航空货运安保供应链的安全性，扩展了航空货运安保供应链的研究内容。②本文将弹性理论研究与航空货运安保供应链自身特性相结合，对航空货运安保供应链从弹性、恢复策略、弹性优化等方面进行定量化的数据分析，更准确地为管理者提供决策支持。

1 航空货运安保供应链扰动事件

短途驳运安保模式下，航空货运安保供应链流程如图1所示。各环节均面临人为失误、设施设备失效、环境干扰、管理缺陷方面的扰动事件。从各单位试运行经验来看，扰动事件多集中于陆路运输环节，诸如道路施工路段封闭导致货车无法按既定路线行驶，或者车载监控系统因移动流量卡流量不足而导致后段路程无监控视频图像等。这主要是因为，陆侧货站及空侧货站的安保环节处于人员充足、设施设备冗余度较高的固定封闭场所内，安保程序、设施建设等遵循《民用航空安全检查规则》[14]及《民用运输机场安全保卫设施》[15]等文件的要求。若发生扰动事件，应急响应较迅速。相比之下，陆路运输时货车一直暴露在社会环境中，可控性较弱，一旦发生扰动事件，现场应急响应速度受到制约，其导致的航空货运安保供应链性能下降程度更高，且不能立即恢复。因此，需重点关注陆路运输环节的扰动事件。其中，人为失误、设施设备失效、管理缺陷导致的扰动事件相比环境干扰方面更易于控制。因此，需重点关注陆路运输环节因环境干扰导致的扰动事件。结合深圳宝安国际机场等单位的试运行经验，最终以陆路运输环节偏移指定行驶路线为扰动事件，开展航空货运安保供应链弹性研究。

为确定指定行驶路线，先需评估各路线的风险性。陆路运输过程面临的非法干扰风险主要体现为：非法分子可能靠近货车，然后接连破坏车辆监控、车厢门锁、货物防篡改包装等设施设备，最终在安全货物中混入危险品、违禁品。基于深圳宝安国际机场等单位的运行方案，综合5名航空货运安保专家的意见，从停靠点数量、事故易发路段数量、路线长度、拥堵程度4个方面评估各路线的风险性。以上四方面数值越大，则路线风险性越大。因为货车在停靠点（红绿灯路口、高速收费站）或遭遇交通事故时，将在原地停留一段时间，如此为非法分子实施干扰行为提供了便利。路线越长，则货车暴露在社会环境中的时间越长。道路拥堵时，车辆行驶缓慢，车间距较小。以上各方面均增大了非

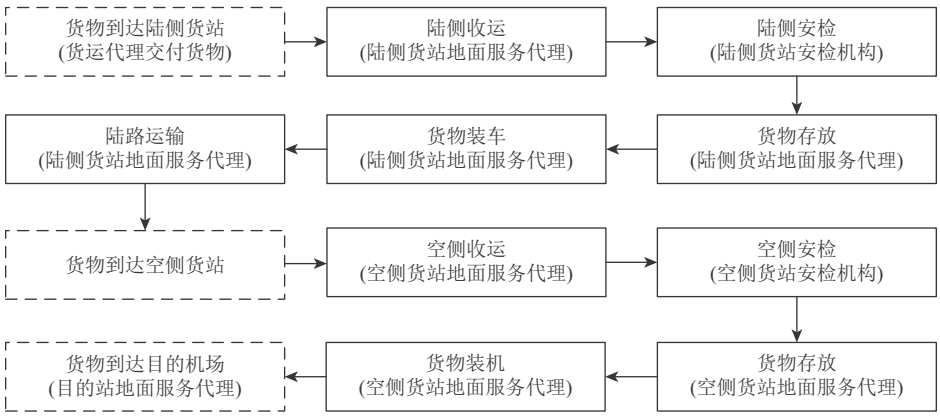


图 1　货运安保供应链流程

Fig. 1　Operation process of air cargo secure supply chain

法干扰风险。

正常情况下, 货车需在风险性最低的指定路线上行驶。若该路线因施工而封闭或突发重大交通事故, 则驾驶员将被迫选择其他风险性更高的路线行驶, 进而导致货物安全性下降。此类事件中, 航空货运安保供应链安保性能变化过程可分为 4 个阶段, 如图 2 所示。为简化情形, 仅描述陆路运输及空侧货站开展恢复工作的环节, 其余环节未在图 2 中体现。图中, $S(t)$ 为航空货运安保供应链性能(安保水平)函数, t_p为货车在偏离路线行驶中安保措施等级与新路线风险性相匹配的时间点, t_f为货车驶离陆侧货站的时间点, t_s为货车行驶路线发生偏离的时间点, t_r为货车到达空侧货站的时间点, t_e为重新安检完全部货物的时间点, t_d为使该批货物能全部装载至预订航班, 应在该时刻前重新安检完全部货物。

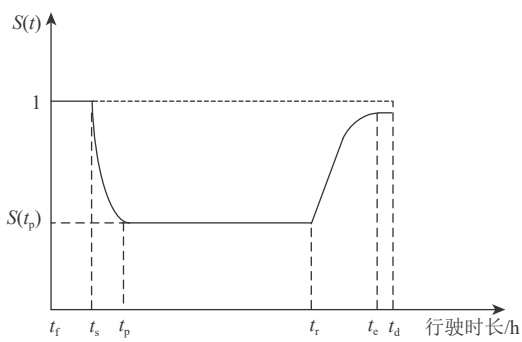


图 2　航空货运安保供应链安保性能变化过程

Fig. 2　Security performance change process of air cargo secure supply chain

1) 原始稳定阶段($t_f \leqslant t < t_s$)

在t_f时刻, 货车驶离陆侧货站, 在指定路线行驶, 并于t_s时刻发生路线偏移。假设该阶段均符合安保管理要求, $S(t) = 1$。

2) 扰动阶段($t_s \leqslant t < t_p$)

在t_s时刻, 因指定道路在进行施工或发生交通事故, 驾驶员选择其他路线行驶, 安保水平开始下降。随后押运员上报事件信息, 运输环节控制部门通过密切关注视频监控和 GPS 信息等措施提高安保等级。在t_p时刻, 安保措施等级与新路线的风险性相匹配, $S(t)$停止下降。

3) 适应阶段($t_p \leqslant t < t_r$)

在t_p时刻, 开始进入适应阶段。为简化情形, 假设社会道路网络结构简单, 交叉口较少, 货车继续在偏离路线上行驶至空侧货站, $S(t) = S(t_p) = S(t_r)$。

4) 恢复阶段($t_r \leqslant t \leqslant t_d$)

在t_r时刻, 空侧货站利用安检设备对货物重新进行安全检查, 以识别因路线偏移而可能混入的危险品、违禁品等, 提升货物安全性。t_e时刻检查完全部货物后, 安保水平恢复至稳定状态$S(t_e)$, $S(t_e)$由空侧货站采取的安检策略决定, $S(t_e) \leqslant 1$。考虑到后续组板和装机所花费的时间, 为使该批货物能全部装载至预订航班, 应在t_d时刻前重新安检完全部货物, 即$t_e \leqslant t_d$, 如此, $S(t_d) = S(t_e) \leqslant 1$。

2　航空货运安保供应链弹性模型

为提升航空货运安保供应链的弹性水平, 需对弹性进行测量。本节先确定扰动阶段、适应阶段、恢复阶段航空货运安保供应链的安保性能函数, 后根据弹性三角形理论确定弹性测度表达式。扰动阶段抵抗能力已知时, 航空货运安保供应链的弹性水平取决于恢复阶段采取的恢复策略, 为寻求最佳弹性值, 建立恢复时间、设备数量等资源约束下以弹性最大化为目标的最优恢复策略模型。

2.1　扰动阶段安保性能函数

为测量弹性水平, 需确定扰动阶段的安保性能函数。该阶段性能下降程度反映了航空货运安保供应链对扰动事件的抵御能力[10]。作为旨在保护货物免遭非法干扰的系统, 航空货运安保供应链的可靠性决定了货物安全性。可靠性$R(t)$指一个系统

在规定条件下和规定时间内完成规定功能的能力[16]，表征系统从初始时刻0至某时刻t仍正常运行的概率。一个系统若不能完成规定的功能，则称其为“失效”。失效率λ指系统工作到某时刻之后，在单位时间内发生失效的概率。当系统处于环境等偶然因素导致的失效阶段时，λ为常数，可靠性与失效率的关系为$R(t)=\mathrm{e}^{-\lambda t}$[17]。在路线偏移情景下，航空货运安保供应链的失效率，表现为下一单位时间内安保措施不再能抵抗非法干扰的概率。

当货车在风险性最低的路线，即指定路线行驶时，航空货运安保供应链的安保措施等级α与指定路线的风险性相匹配，则安保措施有效抵抗非法干扰的概率为100%，失效率为0，代入表达式$R(t)=\mathrm{e}^{-\lambda t}$得可靠性为1，与航空货运安保供应链的初始安保水平1相对应。当行驶路线偏移至风险性更高的路线时，假定道路风险性是原路线的k倍，安保措施等级应为αk。但因安保措施等级未及时加强，仍为α，则在下一单位时间内不再能抵抗非法干扰的概率为$\left(1-\dfrac{\alpha}{\alpha k}\right)$，即认为失效率为$\left(1-\dfrac{1}{k}\right)$。扰动阶段安保性能函数为

$$S(t)=S(t_\mathrm{s})\mathrm{e}^{-(1-1/k)t-t_\mathrm{s}} \qquad t_\mathrm{s}\leqslant t<t_\mathrm{p} \tag{1}$$

2.2 恢复阶段安保性能函数

为测量弹性水平，需明确恢复阶段的安保性能函数。该阶段采取的安检方式、各安检方式下的货物检查量、应用的安检设备台数，决定了航空货运安保供应链的恢复能力。

短途驳运安保模式下所交运的航空货物初期多为性质较单一的普通货物。为便于装卸运输，陆侧货站实施了组板操作，即将多个单元包装件码放捆扎到集装板上，集装板货物可直接装载至航空器。若空侧货站需对货物进行重新安检，则需拆解集装板货物。拆板后，为快速过检，可将多个单元包装件码放到托盘上，以托盘为单位对货物实施检查。目前，常用的货物安检技术包括爆炸物痕量检测技术、爆炸物探测技术。2种技术的原理不同，且前者侧重于检查货物外包装，后者侧重于检查货物内容物，后者对爆炸物等违禁品的检出率更高。鉴于货量较大，受设备数量和恢复时间限制，可采取多种安检技术相结合的恢复策略。综合考虑各技术特性，提出如式(2)所示的3种安检方式。各方式同步运行，且对每件货物仅应用一种方式。

$$l_i=\begin{cases}0\\1\end{cases} \tag{2}$$

式中：l_i表示空测货站是否应用安检方式i，l_i为0时表示未应用安检设备，l_i为1时表示应用安检设备。i表示空测货站可选用的安检方式，$i=1$表示使用爆炸物痕量检测技术检查整个托盘货物的外包装，$i=2$表示使用爆炸物痕量检测技术检查托盘中每个单元包装件的外包装，$i=3$表示使用爆炸物痕量检测技术检查托盘中每个包装件的外包装。

P_i与应用该方式后单件货物j的S_j^i间的关系如下：

$$S_j^i=S(t_\mathrm{r})+(1-S(t_\mathrm{r}))P_i \tag{3}$$

式中：P_i为安检方式i正确探测出爆炸物等违禁品的概率，称为检出率。货物件数以托盘为单位计算，j表示托盘货物编号，$j=1,2,\cdots,Q$（Q表示以托盘为单位计算的货物总量）。S_j^i表示应用安检方式i后货物j的安全性。

对安检方式i，其所应用的d_i台设备同时启动，安检进度一致。当分配至安检方式i的货量Q_i与设备台数d_i间不能完全整除时，安检方式i在时刻t下已安检的货物总量$f_i(t)$为分段函数，如式(4)～式(7)所示。当能整除时，$f_i(t)=v_id_i(t-t_\mathrm{r})$，安检完货量$Q_i$的时间结点为$\hat{T}_i$。

$$f_i(t)=\begin{cases}v_id_i(t-t_\mathrm{r}) & t_\mathrm{r}\leqslant t\leqslant\hat{T}_i\\ f_i(\hat{T}_i)+v_i[Q_i-(Q_i/d_i)d_i](t-\hat{T}_i) & \hat{T}_i<t\leqslant T_i\\ Q_i & T_i<t\leqslant t_\mathrm{e}\end{cases} \tag{4}$$

$$\hat{T}_i=(Q_i/d_i)/v_i+t_\mathrm{r} \tag{5}$$

$$T_i=[(Q_i/d_i)+1]/v_i+t_\mathrm{r} \tag{6}$$

$$t_\mathrm{e}=\max\{T_i\} \tag{7}$$

式中：$f_i(t)$为安检方式i已检查的货物量函数；v_i为安检方式i的检查速率（件/h）；d_i为安检方式i下实际应用的设备台数；$\hat{T}_i$为d_i台安检设备均处于安检状态的时间结点；Q_i为使用安检方式i检查的货物量，以托盘为单位；T_i为安检方式i检查完货量Q_i的时间点。

由Q件托盘货物安全性的算数平均值表征恢复阶段航空货运安保供应链的安保水平，其表达式为

$$S(t)=\sum_{i=1}^{3}l_i\left[(Q_i-f_i(t))S(t_\mathrm{r})+\sum_{j=1}^{f_i(t)}S_j^i\right]\Big/Q \qquad t_\mathrm{r}\leqslant t<t_\mathrm{e} \tag{8}$$

2.3 弹性测度

设置时间间隔T以衡量航空货运安保供应链的弹性水平[11]。重点关注偏离路线后恢复工作的及时性，将该时间间隔设为$T=t_\mathrm{d}-t_\mathrm{s}$。

弹性包含脆弱性[18]（$S(t_\mathrm{s})-S(t_\mathrm{p})$）、破坏速度[19]、

恢复速度[20]和可恢复性[21]（$S(t_d)-S(t_r)$）。脆弱性和破坏速度可以反映航空货运安保供应链的抵抗能力。恢复速度和可恢复性可以反映航空货运安保供应链的恢复能力。基于弹性三角形理论，陆路运输环节路线偏离扰动情形下，航空货运安保供应链的弹性实际上是扰动下的累积性能与原有理想状态下累积性能间的比率[11]，如式（9）所示。该式进行弹性测度时综合考虑了性能变化和恢复时间，增加了弹性测量的准确性。为便于计算，也可先求解弹性损失，如式（10）所示。式（10）中等号右侧第 2 项的分母是未发生扰动事件的累积性能，而分子是因发生扰动事件而导致的性能损失，可用图 2 中的阴影区域表示，主要体现为因货物安全性降低而导致的人财物潜在的累积损失。

$$R=\int_{t_s}^{t_d} S(t)\mathrm{d}t/S(t_s)(t_d-t_s) \qquad R\in[0,1] \tag{9}$$

$$R=1-\int_{t_s}^{t_d}(S(t_s)-S(t))\mathrm{d}t/S(t_s)(t_d-t_s) \tag{10}$$

2.4　弹性最优模型

当扰动阶段抵抗能力已知时，恢复阶段采取何种恢复策略决定了航空货运安保供应链的弹性水平，最优恢复策略如式（11）所示。因此，在恢复水平（见式（12）和式（13））、设备数量（见式（14）和式（15））及恢复时间（见式（16））的有限资源约束下，确定基于弹性的最优恢复策略，即 $\max R$。

$$\max R=1-\int_{t_s}^{t_p}(S(t_s)-S(t))\mathrm{d}t+\int_{t_p}^{t_r}(S(t_s)-S(t))\mathrm{d}t+\int_{t_r}^{t_d}(S(t_s)-S(t))\mathrm{d}t/(t_d-t_s) \tag{11}$$

$$\begin{cases} Q_i=0 & S_j^i<S_0 \\ Q_i\geqslant 0 & S_j^i\geqslant S_0 \end{cases} \qquad i\in\{1,2,3\},Q_i\in\mathbf{N} \tag{12}$$

$$Q=\sum_{i=1}^{3}Q_i \tag{13}$$

$$l_i=\begin{cases} 0 & Q_i=0 \\ 1 & Q_i>0 \end{cases} \tag{14}$$

$$d_i=\begin{cases} Q_i & Q_i<D_i \\ D_i & Q_i\geqslant D_i \end{cases} \tag{15}$$

$$t_e=\max\{[(Q_i/d_i)+1]/v_i+t_r\}\leqslant t_d \tag{16}$$

式（11）表示以时间间隔 $T=t_d-t_s$ 内弹性最大化为目标函数，该函数中包含变量 Q_i、l_i、d_i，R 表示航空货运安保供应链的弹性值。式（12）表示根据应用安检方式 i 后的货物安全性 S_j^i 是否达到最低要求 S_0，确定可选取的安检方式范围，主要通过 Q_i 与 0 的关系来体现。式（13）表示各安检方式被分配的货物检查量 Q_i 之和等于货物总量 Q。式（14）表示根据 Q_i 是否为 0 确定 l_i 的取值，l_i 表示空侧货站是否应用安检方式 i。式（15）表示根据安检方式 i 被分配的货物检查量 Q_i 与设备总数 D_i（D_i 表示安检方式 i 下可应用的设备总台数）之间的关系，确定实际应用的安检设备数量 d_i。当 Q_i 大于 D_i 时，为到达快速恢复的目的，默认为方式 i 配置的设备被全部应用。式（16）表示应在规定的时间点 t_d 前重新安检完全部货物。总体来看，模型中需首要确定的决策变量为各安检方式对应的货物检查量 Q_i，根据式（14）和式（15）可进一步确定 l_i 和 d_i，由此寻求满足式（16）约束下的弹性最大化。

3　实例分析

3.1　扰动情形设置

民航机场生产统计公报显示，近年来深圳宝安国际机场的货邮吞吐量一直居于全国前五。空侧货站场地资源及通道能力不足，促使深圳宝安国际机场探索航空货物短途驳运安保模式，该模式下航空货运安保供应链流程如图 1 所示。由百度地图知，从陆侧货站至空侧货站共有 3 条社会路线，如图 3 所示，从左至右依次为路线 1、路线 3、路线 2。模拟陆路运输环节偏移指定路线的扰动情形时，选取业务较为繁忙的时间段（周一 18:00 从陆侧货站 A 出发）。

为确定指定行驶路线，需计算各路线的风险性。由于停靠点数量、事故易发路段数量、路线长度、拥堵程度 4 个方面的数据均可获取，因此，利用逼近理想值排序法（technique for order preference by similarity to ideal solution, TOPSIS）综合评估各路线的风险性。其中，停靠点数量为路线上红绿灯、高速收费站的总数量，事故易发路段数量由百度地图模拟导航功能获知，路线长度单位为 km。百度地图智慧交通平台采用拥堵指数表征道路交通拥堵程度，拥堵指数为车辆的实际行程时间与道路畅通情形下行程时间的比值。借助百度地图未来出发预测功能确定各路线畅通（路线全程显示为绿色）行程时间。然后，计算周一 18:00 从 A 点出发至终点所需的时间与畅通行程时间的比值，用该比值表征路线的拥堵程度。整理各路线各项指标下的原始数据，3 条路线的停靠点数量、事故易发路段数量、路线长度、拥堵情况数据如表 1 所示。其中，基于专家调查法将各指标的权重设置为 0.25。

用 TOPSIS 计算各路线风险性步骤如下：

步骤 1　构造指标数据矩阵。

设共有 m 条待评价路线，q 个评价指标，将指标数据矩阵构造为

图3 行驶路线

Fig. 3 Driving route

表1 路线数据

Table 1 Route data

路线	停靠点数量	事故易发路段数量	路线长度/km	拥堵长度/km
路线1	14	0	22	1.21
路线2	16	1	22	1.36
路线3	9	2	26	1.68

$$\boldsymbol{U}_0=\begin{bmatrix}u_{11} & u_{12} & \cdots & u_{1q}\\ u_{21} & u_{22} & \cdots & u_{2q}\\ \vdots & \vdots & & \vdots\\ u_{m1} & u_{m2} & \cdots & u_{mq}\end{bmatrix} \tag{17}$$

本文选取3条社会路线，4个评价指标（停靠点数量、事故易发路段数量、路线长度、拥堵情况），即$m=3$，$q=4$。因此，指标数据矩阵为

$$\boldsymbol{U}_0=\begin{bmatrix}14 & 0 & 22 & 1.21\\ 16 & 1 & 22 & 0.36\\ 9 & 2 & 26 & 0.51\end{bmatrix} \tag{18}$$

步骤2 构建标准化矩阵。

利用表达式$u_{mq}^{*}=u_{mq}\Big/\sqrt{\sum_{k=1}^{3}(u_{mq})^2}$，对各指标进行标准化处理，使各指标在同一量纲，计算可知标准化矩阵为

$$\boldsymbol{U}=\begin{bmatrix}0.61 & 0 & 0.54 & 0.49\\ 0.69 & 0.45 & 0.64 & 0.55\\ 0.39 & 0.9 & 0.64 & 0.68\end{bmatrix} \tag{19}$$

步骤3 确定最优方案和最劣方案。

最优方案$\boldsymbol{U}^{+}$由$\boldsymbol{U}$中每列元素的最大值构成：

$$\begin{aligned}\boldsymbol{U}^{+}=&[\max(u_{11},u_{21},u_{31}),\max(u_{12},u_{22},u_{32}),\\ &\max(u_{13},u_{23},u_{33}),\max(u_{14},u_{24},u_{34})]=\\ &[0.69,0.9,0.64,0.68]\end{aligned} \tag{20}$$

最劣方案$\boldsymbol{U}_0^{-}$由$\boldsymbol{U}$中每列元素的最小值构成：

$$\begin{aligned}\boldsymbol{U}^{-}=&[\min(u_{11},u_{21},u_{31}),\min(u_{12},u_{22},u_{32}),\\ &\min(u_{13},u_{23},u_{33}),\min(u_{14},u_{24},u_{34})]=\\ &[0.39,0,0.54,0.49]\end{aligned} \tag{21}$$

标准化数据及最优、最劣方案如表2所示。

表2 路线风险性评估指标标准化数据

Table 2 Standardized data of route risk assessment indicators

路线	停靠点数量	事故易发路段数量	路线长度/km	拥堵长度/km
路线1	0.61	0	0.54	0.49
路线2	0.69	0.45	0.64	0.55
路线3	0.39	0.90	0.64	0.68
最优	0.69	0.90	0.64	0.68
最劣	0.39	0	0.54	0.49

步骤4 计算各路线与最优方案、最劣方案间的风险性差距。

因各指标权重均为0.25，则各路线风险性与最优距离D_m^{+}、最劣距离D_m^{-}分别为

$$\begin{cases}D_m^{+}=\sqrt{\sum_{l=1}^{4}(U_q^{+}-u_{mq})^2/4}\\ D_m^{-}=\sqrt{\sum_{l=1}^{4}(U_q^{-}-u_{mq})^2/4}\end{cases} \tag{22}$$

步骤5 计算各条路线与最优方案间的风险性。贴近度，用C_m^{*}表示。并进行归一化，用C_m表示。$0\leqslant C_m\leqslant 1$，$C_m\to 1$表明该条路线风险性越高。风险性贴近度表达式为

$$\begin{cases}C_m^{*}=D_m^{-}/(D_m^{+}+D_m^{-})\\ C_m=C_m^{*}\Big/\sum_{m=1}^{3}C_m^{*}\end{cases} \tag{23}$$

步骤6 根据贴近度大小对各路线风险性排序，结果如表3所示。

因路线1风险性最低，指定路线1为默认行驶路线。

模拟扰动情形。周一18:00货车从陆侧货站A出发，拟按照路线1行驶，但由于C处临近路段施工，路线1和路线3的后续路段禁止通行，因此，在货车行驶至交叉口B时，驾驶员决定偏移至路线2。由表3可知，路线3的风险性是路线1的3.92倍，取整后为4倍，即令$k=4$。

表 3　路线风险性

Table 3　Routes risk

路线	D_m^+	D_m^-	C_m	风险性次序
路线1	0.46	0.11	0.13	3
路线2	0.23	0.28	0.36	2
路线3	0.15	0.46	0.51	1

3.2　参数设置

弹性最优化模型中时间和资源参数的设置如表 4 所示。基于路线 2 时间路段行驶情况对时间参数t_{f}、t_{s}、t_{r}、t_{d}设置数值，单位为 h。t_{p}值取决于深圳宝安国际机场航空货运安保供应链对扰动事件的抵御能力，即押运员上报事件信息，并升级安保措施等级所需时长，为探究不同扰动时长情形下所采取的恢复策略，令t_{p}取值范围为{0.4 h,0.6 h,0.8 h,1.0 h}。根据深圳宝安国际机场实际运行数据对资源参数D_i、v_i、P_i、S_0、Q设置数值，各安检方式的检查速率v_i的单位为件/h，检出率P_i取值基于中国民航科学技术研究院航空安保研究所[22]对痕量爆炸物设备检出率的研究数据和 Mclay 等[23]对痕量爆炸物设备检出率的设定值，货物总量Q（以托盘为单位计算）取值范围为{20,24,28,32,36}。

表 4　参数设置

Table 4　Parameter settings

类型	参数	取值
时间	t_{f}/h	0
	t_{s}/h	0.2
	t_{r}/h	1.0
	t_{d}/h	1.3
	t_{p}/h	{0.4,0.6,0.8,1.0}
资源	D_1、D_2、D_3	2、2、2
	v_1、v_2、v_3	80、65、70
	P_1、P_2、P_3	0.5、0.7、0.85
	S_0	0.9
	Q	{20,24,28,32,36}

3.3　弹性分析

代入相关参数值，应用 MATLAB R2018a 求解不同货量及不同扰动时长情景下的最优弹性值和需采取的最优恢复策略。按照定积分的定义求解定积分时，积分变量t的取值间隔设为 10^{-6}。得最优结果平均用时 32 s，最优结果如表 5 所示。可知，当$t_{\mathrm{p}}=0.4$时，应用安检方式 1、安检方式 2、安检方式 3 均能满足S_0要求，当货量为 20 时，不选择安检方式 1，而由安检方式 2 检查 6 件托盘货物，安检方式 3 检查 14 件托盘货物，可得最优弹性值 0.897 8。由检查量与D_i间的大小关系，可知各安检方式实际应用的设备台数分别为 0、2、2。总体来看，各情形下的最优弹性值处于 0.75 以上。当货量Q固定时，随着t_{p}增大，可选的安检方式变少，最优弹性值逐渐减小。当扰动时长t_{p}固定时，随着货量Q增大，最优弹性值亦逐渐减小。

针对货物检查量分配策略，将根据可用安检方式的情况进行划分。偏移至路线 2 后，当安检方式 1、安检方式 2、安检方式 3 均可用时，优先考虑安检方式 3 和安检方式 2，且总货量较少时，可选择不应用安检方式 1。当仅安检方式 2、安检方式 3 可用，货物检查量优先分配至安检方式 3。根据式(8)的分子和式(4)中$t_{\mathrm{r}}\leqslant t\leqslant\hat{T}_i$时的函数来解释货量分配策略。式(8)的分子表示各安检方式下的货物安全性随时间变化的函数，其导数$S_i^{'}$推导过程如式(24)所示。当$S(t_{\mathrm{r}})$已知时，$S_i^{'}$由安检速率、应用的设备数量、违禁品检出率决定。代入相关参数值可知，$S_3^{'}>S_2^{'}>S_1^{'}$，即单位时间内应用安检方式 3 后货物安全性增大情况最明显，其次是安检方式 2、安检方式 1。因此，可按照以上顺序分配货物检查量。

$$\begin{aligned}S_i^{'}=&[(Q_i-f_i(t+\Delta t))S(t_{\mathrm{r}})+f_i(t+\Delta t)S_j^i]-\\&[(Q_i-f_i(t))S(t_{\mathrm{r}})+f_i(t)S_j^i]/\Delta t=\\&v_i d_i P_i(1-S(t_{\mathrm{r}}))\qquad t_{\mathrm{r}}\leqslant t<\hat{T}_i\end{aligned}\tag{24}$$

3.4　弹性优化策略

由式(24)可知，v_i、d_i、P_i、$S(t_{\mathrm{r}})$影响恢复阶段货物安全性变化情况。其中，$S(t_{\mathrm{r}})$由t_{p}直接决定。此外，受空侧货站场地面积、设备、人员数量限制及爆炸物探测设备间距影响，不宜通过增大d_i来提升弹性。基于此，运用控制变量法，通过分析t_{p}、v_i、P_i变化对最优弹性值的影响，提出弹性优化策略。

1）提升抵御能力，降低t_{p}

t_{p}直接影响扰动阶段和适应阶段的弹性损失值，并通过影响恢复阶段可选用的安检方式，进而影响恢复阶段的弹性损失值。由表 5 可知，降低t_{p}可增大航空货运安保供应链弹性。提升抵御能力主要是通过制定路线偏移情形下的应急预案，合理配置后台安保人员数量，并通过培训及应急演练提升相关人员资质，保证在事件发生后能快速响应。

2）提升业务能力，增大v_i

v_i决定了空侧货站重新安检货物的速率。以$Q=36$为例，将 3 种安检方式的v_i分别提升 5，以及均提升 5，最优弹性值变化如表 6 所示。由表 6 可见，提升v_i能增大弹性。仅提升一种v_i时，增大v_3带来的效果最明显，而仅提升v_1并不能增大弹性值。增大v_i的途径主要是通过培训增强空侧货站相关工作人员装卸、检查货物外包装、判别图像等业务的熟练度。

表5 最优弹性值和恢复策略

Table 5 Optimal resilience and recovery strategies

条件	最优弹性值					恢复策略				
	Q=20	Q=24	Q=28	Q=32	Q=36	Q=20	Q=24	Q=28	Q=32	Q=36
t_p = 0.4，3种安检方式均可用	0.897 8	0.897 0	0.896 2	0.895 5	0.894 8	[0,6,14]	[0,8,16]	[0,10,18]	[2,12,18]	[4,14,18]
t_p = 0.6，安检方式2、安检方式3可用	0.832 2	0.830 6	0.829 1	0.827 7	0.826 2	[0,6,14]	[0,8,16]	[0,10,18]	[0,14,18]	[0,16,20]
t_p = 0.8，安检方式3可用	0.790 5	0.786 4	0.782 4	0.778 3	0.774 3	[0,0,20]	[0,0,24]	[0,0,28]	[0,0,32]	[0,0,36]
t_p = 1.0，安检方式3可用	0.773 6	0.768 5	0.763 5	0.758 4	0.753 4	[0,0,20]	[0,0,24]	[0,0,28]	[0,0,32]	[0,0,36]

表6 增大v_i后的最优弹性值和恢复策略

Table 6 Optimal resilience and recovery strategy after increasing v_i

条件	最优弹性值					恢复策略				
	原始值	仅v_1提升5	仅v_2提升5	仅v_3提升5	v_i均提升5	原始值	仅v_1提升5	仅v_2提升5	仅v_3提升5	v_i均提升5
t_p=0.4，3种安检方式均可用	0.894 8	0.894 8	0.894 9	0.895 1	0.859 2	[4,14,18]	[4,14,18]	[4,14,18]	[4,12,20]	
t_p=0.6，安检方式1不可用	0.826 2	0.826 2	0.826 6	0.826 9	0.827 1	[0,16,20]	[0,16,20]	[0,16,20]	[0,14,22]	
t_p=0.8，安检方式1、安检方式2不可用	0.774 3	0.774 3	0.774 3	0.776 7	0.776 7	[0,0,36]	[0,0,36]	[0,0,36]	[0,0,36]	[0,0,36]

3）提升设备维护能力，增大P_i

P_i决定了空侧货站可选用的安检方式，并通过影响货物安全性来影响恢复阶段弹性损失值。以Q = 36为例，将3种安检方式的P_i分别提升0.1，以及均提升0.1，弹性值变化如表7所示。由表可见，提升P_i能增大弹性。其中，仅提升一种P_i时，增大P_3带来的效果最明显。此外，将P_2提升0.1能改变原来t_p = 0.8时安检方式2不可用的状况。相关研究表明，随着使用频次的增加，爆炸物痕量检测设备的违禁品检出率将进行严重衰减，而经过维护后，设备的检出率会有较大提升。依此，空侧货站需定期对相关安检设备进行维护测试。

表7 增大P_i后的最优弹性值和恢复策略

Table 7 Optimal resilience and recovery strategy after increasing P_i

条件	最优弹性值					恢复策略				
	原始值	仅P_1提升0.1	仅P_2提升0.1	仅P_3提升0.1	仅P_4提升0.1	原始值	仅P_1提升0.1	仅P_2提升0.1	仅P_3提升0.1	仅P_4提升0.1
t_p=0.4，3种安检方式均可用	0.894 8	0.895 4	0.896 1	0.896 4	0.897 9	[4,14,18]	[8,10,18]	[2,16,18]	[2,12,22]	[6,12,18]
t_p=0.6，安检方式2、安检方式3可用	0.826 2	0.826 2	0.828 8	0.829 5	0.831 8	[0,16,20]	[0,16,20]	[0,16,20]	[0,14,22]	[0,16,20]
t_p=0.8，安检方式3可用	0.774 3	0.774 3		0.780 0		[0,0,36]	[0,0,36]		[0,0,36]	
t_p=0.8，安检方式2、安检方式3可用			0.790 0		0.794 2			[0,16,20]		[0,16,20]

4 结 论

针对航空货运安保供应链陆路运输环节偏移指定行驶路线的扰动情形，本文进行了航空货运安保供应链弹性测度，并求解了弹性最优目标下的恢复策略，以确定空侧货站可应用的安检方式、各方式的货物检查量和使用的安检设备台数。以深圳宝安国际机场为例进行分析得出以下结论。

1）安检速率、安检设备数量、违禁品检出率决定分配货物检查量时各安检方式的优先性。顺序如下：安检方式3优于安检方式2，再优于安检方式1。

2）货物总量或扰动时长增大，航空货运安保供应链弹性降低。当扰动时长较大时仅安检方式3可用，且随着总货量的增加，货物检查量分配策略基本保持一致，即总货量每增加4件，安检方式3被分配的货物检查量将在原来基础上增加4件。

3）提升3种安检方式的检查速率，航空货运安保供应链弹性增加。增大安检方式3的检查速率，航空货运安保供应链弹性增加的最快，仅提升安检

方式 1 的检查速率并不能增大弹性值。

4) 提高 3 种安检方式的检出率, 航空货运安保供应链弹性增加。增大安检方式 3 的检出率, 航空货运安保供应链弹性增加最明显。安检方式 2 的检出率提升 0.1, 即可改变扰动时长较长时不可使用安检方式 2 的情况。

本文选取路线偏移作为扰动事件进行研究, 后续可从设备设施失效、管理缺陷、恶劣天气等方面进行航空货运安保供应链弹性研究。本文提出 3 种优化策略, 即降低扰动时长、增大安检速率、提高违禁品检出率均可提高航空货运安保供应链弹性, 实际应用中机场管理者会在提高弹性水平和投入成本之间权衡, 后续研究可考虑成本限制下弹性水平优化策略。

参考文献（References）

[1] International Civil Aviation Organization. Moving air cargo globally: air cargo and mail secure supply chain and facilitation guidelines[EB/OL].(2011-06-24)[2023-07-26].https://www.icao.int/Security/aircargo/Documents/Moving%20Air%20Cargo%20Globally%20-%20Third%20edition.EN.pdf.

[2] DA CUNHA D A, MACÁRIO R, REIS V. Keeping cargo security costs down: a risk-based approach to air cargo airport security in small and medium airports[J]. Journal of Air Transport Management, 2017, 61: 115-122.

[3] 王赢. 国际航空货运安保模式对中国民航的启示[J]. 民航管理, 2020(8): 88-92.
WANG Y. Enlightenment of international air cargo's security check to Chinese civil aviation[J]. Civil Aviation Management, 2020(8): 88-92(in Chinese).

[4] DOMINGUES S, MACÁRIO R, PAUWELS T, et al. An assessment of the regulation of air cargo security in Europe: a Belgian case study[J]. Journal of Air Transport Management, 2014, 34: 131-139.

[5] 赖国基, 罗嘉欣, 张怡, 等. 国际航空货运安保政策研究[J]. 综合运输, 2020, 42(9): 79-84.
LAI G J, LUO J X, ZHANG Y, et al. On the security policy of international air cargo transportation[J]. China Transportation Review, 2020, 42(9): 79-84(in Chinese).

[6] 陈锦渠, 张帆, 彭其渊, 等. 大客流下城市轨道交通站点韧性评估及划分[J]. 安全与环境学报, 2022, 22(6): 2994-3002.
CHEN J Q, ZHANG F, PENG Q Y, et al. Resilience assessment and partition of an urban rail transit station under large passenger flow[J]. Journal of Safety and Environment, 2022, 22(6): 2994-3002 (in Chinese).

[7] 陈磊, 邓欣怡, 陈红坤, 等. 电力系统韧性评估与提升研究综述[J]. 电力系统保护与控制, 2022, 50(13): 11-22.
CHEN L, DENG X Y, CHEN H K, et al. Review of the assessment and improvement of power system resilience[J]. Power System Protection and Control, 2022, 50(13): 11-22(in Chinese).

[8] 陈锦渠, 左彤, 朱蔓, 等. 城市轨道交通网络失效修复策略[J]. 安全与环境学报, 2022, 22(1): 316-323.
CHEN J Q, ZUO T, ZHU M, et al. Repair strategies for a damaged urban rail transit network[J]. Journal of Safety and Environment, 2022, 22(1): 316-323(in Chinese).

[9] 孔建国, 卢靖宇, 梁海军. 基于级联失效的管制扇区网络韧性评估[J]. 安全与环境学报, 2023, 23(11): 3978-3984.
KONG J G, LU J Y, LIANG H J. Air traffic control sector network resilience evaluation based on cascading failures model[J]. Journal of Safety and Environment, 2023, 23(11): 3978-3984(in Chinese).

[10] 颜克胜, 荣莉莉. 面向韧性提升的相互依赖关键基础设施网络灾后修复模型研究[J]. 运筹与管理, 2021, 30(5): 21-30.
YAN K S, RONG L L. Post-disruption restoration model for enhancing resilience of interdependent critical infrastructure networks[J]. Operations Research and Management Science, 2021, 30(5): 21-30 (in Chinese).

[11] FANG C, CHU Y Z, FU H R, et al. On the resilience assessment of complementary transportation networks under natural hazards[J]. Transportation Research Part D: Transport and Environment, 2022, 109: 103331.

[12] WANG N X, YUEN K F. Resilience assessment of waterway transportation systems: combining system performance and recovery cost[J]. Reliability Engineering & System Safety, 2022, 226: 108673.

[13] DUI H Y, ZHENG X Q, WU S M. Resilience analysis of maritime transportation systems based on importance measures[J]. Reliability Engineering & System Safety, 2021, 209: 107461.

[14] 中国人民共和国交通运输部. 民用航空安全检查规则[J]. 中华人民共和国国务院公报, 2017(12): 77-87.
Ministry of Transport of the People's Republic of China. Civil aviation safety inspection rules[J]. The Bulletin of the State Council of the People's Republic of China, 2017(12): 77-87(in Chinese).

[15] 中国民用航空局. 民用运输机场安全保卫设施[EB/OL]. (2017-07-06)[2022-07-20].http://www.caac.gov.cn/big5/www.caac.gov.cn/XXGK/XXGK/BZGF/HYBZ/202008/P020240712354418443635.pdf.
Civil Aviation Administration of China. Civil transport airport securityfacilities[EB/OL].(2017-07-06)[2022-07-20].http://www.caac.gov.cn/big5/www.caac.gov.cn/XXGK/XXGK/BZGF/HYBZ/202008/P020240712354418443635.pdf(in Chinese).

[16] EBELING C E. An introduction to reliability and maintainability engineering[M]. Long Grove: Waveland Press, 2019.

[17] 杨纪珂. 寿命分布、失效率和可靠性[J]. 微电子学与计算机, 1974(3): 23-37.
YANG J K. Life distribution, failure rate and reliability[J]. Microelectronics & Computer, 1974(3): 23-37(in Chinese).

[18] BERDICA K. An introduction to road vulnerability: what has been done, is done and should be done[J]. Transport Policy, 2002, 9(2): 117-127.

[19] NAN C, SANSAVINI G. A quantitative method for assessing resilience of interdependent infrastructures[J]. Reliability Engineering & System Safety, 2017, 157: 35-53.

[20] BAROUD H, BARKER K, RAMIREZ-MARQUEZ J E, et al. Importance measures for inland waterway network resilience[J]. Transportation Research Part E: Logistics and Transportation Review, 2014, 62: 55-67.

[21] ADAMS T M, BEKKEM K R, TOLEDO-DURÁN E J. Freight resilience measures[J]. Journal of Transportation Engineering, 2012, 138(11): 1403-1409.

[22] 中国民航科学技术研究院航空安保研究所. 2020 年痕量爆炸物设备差异化研究[R]. 北京: 航科院安保所, 2020: 4-7.
China Academy of Civil Aviation Science and Technology Institute of Aviation Security. 2020 trace explosives device differentiation study[R]. Beijing: Institute of Aviation Security, 2020: 4-7(in Chinese).

[23] MCLAY L A, JACOBSON S H, KOBZA J E. A multilevel passenger screening problem for aviation security[J]. Naval Research Logistics, 2006, 53(3): 183-197.

Resilience measurement and optimization strategy of air cargo secure supply chain

ZHAO Zhenwu，XING Xiaoxiao，GENG Shuo*

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

Abstract：To enhance the resilience level of the air cargo security supply chain, elastic measurement is conducted based on the theory of elastic triangles to address the disturbance caused by the deviation of designated routes in the land transportation process. In the recovery phase, in order to seek the optimal elasticity, with the goal of maximizing elasticity, the optimal recovery strategy is solved under the constraints of recovery level, equipment quantity, and recovery time, that is, to determine the security inspection methods to be adopted at the airside cargo terminal, the cargo inspection volume of each security inspection method, and the number of security inspection equipment used. Taking Shenzhen Bao'an International Airport as an example, route deviation is selected as the disturbance event for analysis. The results indicate that as the total amount of goods or disturbance duration increases, the elasticity of the air cargo security supply chain decreases; The priority of each security check method is determined by the speed of security checks, the number of security check equipment, and the detection rate of prohibited items when allocating the amount of goods for inspection; Reducing market disruptions, increasing safety rates, and improving prohibited item detection rates can enhance the resilience of the air cargo security supply chain.

Keywords：air cargo; secure supply chain; resilience; recovery strategy; short-distance lightering

Received：2023-07-29；**Accepted**：2023-09-25；**Published Online**：2023-10-07 15：35

URL：link.cnki.net/urlid/11.2625.V.20230928.1903.002

Foundation items：Scientific Research Plan Project of Education Commission of Tianjin (2021SK025)；Civil Aviation Safety Capacity Building Project (14002500000019J013)

* **Corresponding author**. E-mail：1248076228@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0478

基于演化博弈的航空器自主冲突解脱效率

王红勇[1,*]，郭宇鹏[2]

(1. 中国民航大学 空中交通管理学院，天津 300300；2. 北京市交通委员会，北京 101160)

摘　　要：针对航空器自主路径规划问题，提出一种面向航空器自主冲突解脱的博弈协调方法，并研究了该方法在航空器各类博弈策略下达成均衡状态所需的计算周期。基于演化博弈论构建航空器冲突解脱博弈模型，通过计算博弈轮次间的复制动态方程，促进航空器迭代进化其博弈选择偏好，加速博弈系统达到局部均衡状态；构建雅可比矩阵分析博弈中各局部均衡解的稳定性，并证明博弈系统中有且仅有一个局部均衡解稳定，参与博弈的各航空器均会趋向于该均衡解；使用ZSSSAR01扇区空域数据进行仿真实验，设定多种航空器博弈策略并分析各策略达到均衡状态所需的博弈时间。仿真结果表明：理性博弈策略具有较高的博弈效率，在非极端情况下平均仅需5.31轮次博弈即达到均衡；激进和保守博弈策略分别会加快和减慢博弈均衡过程；不合作博弈策略会显著减慢博弈均衡过程，平均需要110.53轮次博弈，不合作博弈者进行的运行成本补偿策略则会缓解该趋势(平均需要86.87轮次)。

关　键　词：航空运输；基于航迹运行；冲突解脱；演化博弈；空中交通管理

中图分类号：V355; U8

文献标志码：A　　**文章编号**：1001-5965(2025)09-2937-10

在目前的空中交通管理(air traffic management, ATM)系统中，随着基于航迹运行(trajectory based operation, TBO)[1]和自由航路空域(free route airspace, FRA)[2]等理念和技术的快速发展，空中交通管制模式也将发生重大变化。具备TBO能力的航空器在FRA环境下的路径规划和冲突解脱责任将由管制员转移至飞行员；即航空器自主规划路径，并主动保持与其他航空器的安全间隔[3-4]。

ATM系统的工作重点也由集中式控制(地面管制中心)向分布式控制(具备自主运行能力的航空器)转变。通过为航空器规划无冲突的安全路径，并对于航空器间存在的潜在冲突进行协调解脱[5-6]，从而促进空域资源在各航空器之间公平分配。在分布式控制问题的求解过程中，需要在各参与者的个体最优之间寻求平衡，以此实现全局最优[7]。而博弈论是解决此类问题的有效方法。

在交通运输领域中，博弈论常用于智能道路交通和多无人机系统等。诸多学者将博弈论理念引入互联自动驾驶车辆的协同运行场景，以此研究其在多车道环境下的跟随与超越[8]，以及无信号交叉口的通行顺序[9-11]等问题；亦有学者基于博弈论研究了无人机集群的运行组织[12-13]，无人机/通用航空器与运输航空器在低空空域的资源分配共享[14-15]等。

对于航空运输领域的空中交通管理问题，受当前集中式决策模式和航空器间信息交互能力限制，博弈论的应用相对较少。现有应用中大多以避免航空器碰撞作为约束条件，2架(或多架)航空器的运行成本(燃油和时间等)作为效用函数[16]，通过博弈和妥协过程，优化控制各航空器通过某航路点的

收稿日期：2023-07-21；**录用日期**：2023-10-10；**网络出版时间**：2023-10-20 10:42
网络出版地址：link.cnki.net/urlid/11.2625.V.20231020.0917.002
基金项目：天津市自然科学基金(21JCZDJC00840)
*通信作者. E-mail：hy_wang@cauc.edu.cn

引用格式：王红勇，郭宇鹏. 基于演化博弈的航空器自主冲突解脱效率[J]. 北京航空航天大学学报，2025，51(9)：2937-2946.
WANG H Y，GUO Y P. Efficiency of aircraft autonomous conflict resolution based on evolutionary games[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2937-2946 (in Chinese).

时刻或占用某航段的时间[17-18]，以此完成航空器潜在冲突解脱和空域资源均衡分配[19-20]。并最终实现各架航空器间无冲突运行。

但是该应用场景在航空器自主运行模式下会发生较大变化。现有研究普遍将航空器视为“理性程度”固定的有限理性博弈者，鲜有考虑各航空器在相同运行环境，但是不同理性程度或不同策略偏好下，因理性程度-博弈偏好差异而对博弈过程产生的影响；现有研究中多存在地面管制中心作为博弈结果的“仲裁者”，但是自主运行中各航空器需自主判定各博弈结果存在的碰撞风险，以此进行决策并尝试达到平衡；而且潜在冲突解脱过程有极高的时效性要求，现有研究对于博弈达成稳定状态所需时间或博弈轮次的关注相对较少。

由此，针对自主运行环境下航空器冲突解脱过程中的博弈协调问题，本文引入演化博弈论思想，构建不同博弈轮次间航空器决策偏好演化的复制动态方程；由此生成雅可比矩阵，判定并论证了有且仅有唯一的局部均衡状态为进化稳定状态；仿真实验中分别模拟航空器各类博弈策略，计算不同策略不同初始偏好设定下达到稳定平衡状态所需的博弈轮次数，以此论证该博弈策略的计算速度是否满足航空器自主冲突解脱需要。

1 航空器自主冲突解脱博弈模型

1.1 博弈行为分析

在航空器自主路径规划过程中，若2架航空器规划所得的路径存在潜在冲突，则需进行自主协调以此实现冲突解脱。航空器的行为策略可分为2类，即保持原始路径迫使对方规避，或本方主动规避。前者使其实现了更低的运行成本（避免额外时间和燃油消耗），且相较博弈对方，本方获得了更优的“心理成本”；而后者则降低了发生不安全事件（航空器碰撞）的可能性，降低风险成本（即增大安全收益）。

执行自主运行模式的航空器，其机载设备可自主规划多条备选路径，并选择最优[21]。为简化博弈问题的计算复杂性，本文假设选择规避策略的航空器均执行其最优规避路径。且由于FRA环境下航空器路径横向分散，不再集中于现有航线或航路点附近，因此，涉及多架航空器同时同地存在潜在冲突的可能性较低。若出现该问题，则将相关航空器按单循环模式分组，根据航空器优先级排列顺序，逐组进行博弈。基于此，本文设定该博弈问题为2×2的策略博弈，如表1所示。

假设某次航空器潜在冲突事件中涉及的航空器分别为i和j。C_i和C_j表示航空器i和j执行原始路径的成本；ΔC_i和ΔC_j表示航空器i和j执行最优改航路径时产生的额外成本；φ_i和φ_j表示航空器i和j的劣势心理成本；D_i和D_j表示航空器冲突而造成的损失成本。

表1 双航空器路径选择决策-成本

Table 1 Dual aircrafts routing decision - cost

路径规划	航空器i成本		航空器j成本	
	航空器i规避	航空器i不规避	航空器j规避	航空器j不规避
航空器j规避	$C_i+\Delta C_i$	C_i	$C_j+\Delta C_j$	$C_j+\Delta C_j+\varphi_j$
航空器j不规避	$C_i+\Delta C_i+\varphi_i$	D_i	C_j	D_j

由于航空器i和j的型号、质量、速度等均不相同，因此，航空器在该路径片段的运行成本C_i和C_j不同，因改航而产生的额外成本ΔC_i和ΔC_j（燃油消耗和时间延误）亦不同。该成本的计算方法参照文献[16]。航空器碰撞损失成本D_i和D_j远大于原始运行成本、额外成本和劣势心理成本。

1.2 博弈参数设置

假设某次博弈过程涉及航空器i和j，其中，i为中型机，j为重型机。

1）航空器的运行成本C_i与C_j。参考文献[16]的相关内容，对于航路飞行阶段的波音公司某典型型号的中型机和重型机，相同环境下，重型机的单位距离运行成本是中型机的1.32～1.51倍。本文取中值1.42。在相同环境下，同型号航空器运行的时间成本和燃油成本均与其飞行距离直接相关。若航空器i飞行距离为L_i，设定其运行成本$C_i=L_i$；航空器j飞行距离为L_j，则$C_j=1.42C_iL_j/L_i=1.42L_j$。

2）航空器的额外运行成本ΔC_i与ΔC_j。因改航而产生的额外运行成本与其改航路径较原始路径的偏移程度相关。使用文献[21]所述的人工势场-粒子群联合路径规划算法为扇区内各航空器规划最优路径，并为存在潜在冲突的片段规划改航规避路径。假设重型航空器j的最优路径长度为L_j，改航规避路径长度为L_j'，则其额外运行成本$\Delta C_j=1.42(L_j'-L_j)$。航空器$i$的额外运行成本同理。

3）航空器因博弈“失利”而产生的劣势心理成本φ_i与φ_j。在自主路径规划运行模式下，各航空器均会参照本方和对方在此前的历次博弈结果而预估对方的博弈偏好，进而确定本方在本次博弈中的决策选择倾向。因此，若航空器i在本次博弈中选择规避，此决策会对航空器i在后续时间的博弈结果产生影响。假设该影响范围为后续n次博弈，并假设其中m次的博弈中航空器均选择规避而产生额外成本，该额外成本的产生需由本次博弈“失利”

而“承担部分责任”。假设航空器i因后续某次博弈失利而产生的额外成本为ΔC_i^k, 则$\varphi_i=\frac{1}{n}\left(\sum_{k=1}^{m}\Delta C_i^k\right)$。航空器$j$的劣势心理成本同理。

4) 航空器碰撞损失成本D_i与D_j。由于该值难以确定, 本文将根据演化博弈过程中稳定策略的数学特征和航空器安全事故发生概率的历史数据确定该值, 详见第 2.2 节。

由表 1 中数据可知, 此情况不存在纳什均衡解, 2 架航空器之一在任意决策下都可通过改变策略而降低自身成本。由此, 引入演化博弈的思想。

2　基于演化博弈的航空器自主冲突解脱

2.1　演化博弈过程

受航空器间通信能力和运行态势感知能力差异限制, 各航空器在博弈-决策时无法获得完全信息, 也难以实现完全理性。因此, 本文假设各航空器具有不完全理性特征, 在博弈开始时不能准确判断双方的博弈成本, 而是在各轮次博弈之间通过信息交互、经验积累而演化调整其策略, 最终达到群体性平衡。在某一轮次博弈中, 假设航空器i选择规避策略的偏好为p, 选择不规避策略的偏好为$1-p$; 同理航空器j分别为q和$1-q$。则航空器i和j选择规避策略的成本分别为

$$\begin{cases}E_{ai}=q(C_i+\Delta C_i)+(1-q)(C_i+\Delta C_i+\varphi_i)\\ E_{aj}=p(C_j+\Delta C_j)+(1-p)(C_j+\Delta C_j+\varphi_j)\end{cases}\tag{1}$$

航空器i和j选择不规避策略的成本分别为

$$\begin{cases}E_{bi}=qC_i+(1-q)D_i\\ E_{bj}=pC_j+(1-p)D_j\end{cases}\tag{2}$$

航空器i和j的收益可用不同策略对应成本的数学期望表示为

$$\begin{cases}E_i=pE_{ai}+(1-p)E_{bi}=p(C_i+\Delta C_i)+\\ \qquad(1-p)(D_i-qD_i+qC_i)+p(1-q)\varphi_i\\ E_j=qE_{aj}+(1-q)E_{bj}=q(C_j+\Delta C_j)+\\ \qquad(1-q)(D_j-pD_j+pC_j)+q(1-p)\varphi_j\end{cases}\tag{3}$$

根据复制动态的思想[10, 22], 航空器若采用了成本较高的策略, 其会通过学习进化过程调整策略偏好, 从而在下一轮次博弈中更倾向于选择成本更低的策略。因此, 航空器选择规避/不规避策略的偏好在各轮次间会发生变化。当前轮次结束后, 航空器i和j博弈偏好的变化速度分别为

$$\begin{aligned}F_p=\frac{\mathrm{d}p}{\mathrm{d}t}&=p(E_{bi}-E_i)=p[(1-q)D_i+qC_i-p(1-q)\varphi_i-\\ &\quad p(C_i+\Delta C_i)-(1-p)(D_i-qD_i+qC_i)]=\\ &\quad p^2(D_i-qD_i-C_i+qC_i-\Delta C_i-\varphi_i+q\varphi_i)\end{aligned}\tag{4}$$

$$\begin{aligned}F_q=\frac{\mathrm{d}q}{\mathrm{d}t}&=q(E_{bj}-E_j)=q[(1-p)D_j+pC_j-q(1-p)\varphi_j-\\ &\quad q(C_j+\Delta C_j)-(1-q)(D_j-pD_j+pC_j)]=\\ &\quad q^2(D_j-pD_j-C_j+pC_j-\Delta C_j-\varphi_j+p\varphi_j)\end{aligned}\tag{5}$$

该微分方程即为演化博弈过程的复制动态方程, 策略占比 (博弈偏好) 的变化速度与策略间的收益差异有关。收益差异越大, 选择较差策略的航空器越倾向于调整其策略, 也即其博弈偏好的演化速度越快。

若航空器选择其中某种策略的可能性不再发生变化, 即$\frac{\mathrm{d}p}{\mathrm{d}t}=0$且$\frac{\mathrm{d}q}{\mathrm{d}t}=0$, 则该博弈问题达到稳定状态, 此时求得

$$\begin{cases}p^*=\dfrac{D_j-C_j-\Delta C_j-\varphi_j}{D_j-C_j-\varphi_j}\\ q^*=\dfrac{D_i-C_i-\Delta C_i-\varphi_i}{D_i-C_i-\varphi_i}\end{cases}\tag{6}$$

因此, 该演化博弈共存在 5 个局部均衡状态$(p,q)=[(0,0)、(0,1)、(1,0)、(1,1)、(p^*,q^*)]$。但是并非所有的局部均衡状态都是稳定的, 因此, 需讨论各状态的稳定性。

2.2　局部稳定性分析

对于式 (4) 和式 (5) 求解关于p和q的一阶偏导数, 并由此生成雅可比矩阵, 如式 (7) 所示; 计算雅可比矩阵形成的行列式和迹, 如式 (8) 和式 (9) 所示。其中, $\mathrm{tr}(\boldsymbol{J})$表示矩阵$\boldsymbol{J}$的迹。以此判定各局部均衡状态的稳定性情况, 如表 2 所示。

$$\boldsymbol{J}=\begin{bmatrix}\dfrac{\mathrm{d}F_p}{\mathrm{d}p} & \dfrac{\mathrm{d}F_p}{\mathrm{d}q}\\ \dfrac{\mathrm{d}F_q}{\mathrm{d}p} & \dfrac{\mathrm{d}F_q}{\mathrm{d}q}\end{bmatrix}=\begin{bmatrix}2p(C_i+\Delta C_i-D_i+qD_i-qC_i+\varphi_i-q\varphi_i) & p^2(D_i-C_i-\varphi_i)\\ q^2(D_j-C_j-\varphi_j) & 2q(C_j+\Delta C_j-D_j+pD_j-pC_j+\varphi_j-p\varphi_j)\end{bmatrix}\tag{7}$$

$$\begin{aligned}|\boldsymbol{J}|&=4pq(C_i+\Delta C_i-D_i+qD_i-qC_i+\varphi_i-q\varphi_i)\cdot\\ &\quad(C_j+\Delta C_j-D_j+pD_j-pC_j+\varphi_j-p\varphi_j)-\\ &\quad(pq)^2(D_i-C_i-\varphi_i)(D_j-C_j-\varphi_j)\end{aligned}\tag{8}$$

$$\begin{aligned}\mathrm{tr}(\boldsymbol{J})&=2p(C_i+\Delta C_i-D_i+qD_i-qC_i+\varphi_i-q\varphi_i)+\\ &\quad 2q(C_j+\Delta C_j-D_j+pD_j-pC_j+\varphi_j-p\varphi_j)\end{aligned}\tag{9}$$

由表 2 可知, 策略(0,0)、(1,0)、(0,1)、(1,1)均为不稳定状态。本节引入演化博弈论中的进化稳定[10, 22]概念, 并证明鞍点(p^*,q^*)是进化稳定状态: 假设在双方博弈中, S和T分别为博弈者i和j的备选路径策略集合, $E_i(s,t)$表示博弈者i执行策略(s,t)时所需

表2 局部均衡状态稳定性判别

Table 2 Stability discrimination of local equilibrium state

(p,q)	$\lvert\boldsymbol{J}\rvert$	$\lvert\boldsymbol{J}\rvert$的正负性	$\mathrm{tr}(\boldsymbol{J})$	$\mathrm{tr}(\boldsymbol{J})$的正负性	稳定状态
(0, 0)	0		0		不稳定
(0, 1)	0		$2(C_j+\Delta C_j-D_j+\varphi_j)$	负	不稳定
(1, 0)	0		$2(C_i+\Delta C_i-D_i+\varphi_i)$	负	不稳定
(1, 1)	$4\Delta C_i\Delta C_j-D_iD_j-C_iC_j-\varphi_i\varphi_j$	负	$2\Delta C_i+2\Delta C_j$	正	不稳定
$(p*,q*)$	$-\frac{[(D_i-C_i-\varphi_i-\Delta C_i)(D_j-C_j-\varphi_j-\Delta C_j)]^2}{(D_i-C_i-\varphi_i)(D_j-C_j-\varphi_j)}$	负	0		鞍点

的成本。若式(10)和式(11)成立，则(s,t)为进化稳定策略，s'表示集合S中与s不同的策略，t'同理。

$$\begin{cases} E_j(s,t)<E_j(s,t') & \forall t'\in T,t'\neq t \\ E_j(s,t)=E_j(s,t'),E_i(s,t')<E_i(s',t') & \\ \quad \forall t'\in T,t'\neq t,\forall s'\in S,s'\neq s & \end{cases} \tag{10}$$

$$\begin{cases} E_i(s,t)<E_i(s',t) & \forall s'\in S,s'\neq s \\ E_i(s,t)=E_i(s',t),E_j(s',t)<E_j(s',t') & \\ \quad \forall t'\in T,t'\neq t,\forall s'\in S,s'\neq s & \end{cases} \tag{11}$$

式(10)和式(11)表明，进化稳定策略对于博弈系统内的任意策略突变均具有稳定性，即若某轮次博弈中航空器未遵照稳定策略进行选择，后续轮次的复制动态过程也会指引该系统重新回归至进化稳定策略中。

命题 在本文所述演化博弈问题中，策略$(p*,q*)$是进化稳定策略。

证明 假设当前策略为(p',q')，其中，$p'\neq p*$且$q'\neq q*$，则p'与$p*$、q'与$q*$的大小关系如图1所示，根据式(4)和式(5)可求得航空器策略选择偏好在此状态的变化情况。

如图1所示，对于区域Ⅰ，$p'<p*$且$q'>q*$，根据式(4)和式(5)计算可得$\frac{\mathrm{d}p}{\mathrm{d}t}>0$且$\frac{\mathrm{d}q}{\mathrm{d}t}<0$。也即随着演化过程进行，$p'$数值增大，$q'$数值减小，直至$p'=p*$且$q'=q*$时，$\frac{\mathrm{d}p}{\mathrm{d}t}=0$且$\frac{\mathrm{d}q}{\mathrm{d}t}=0$。此时航空器的博弈策略将不再发生演化，系统达到稳定状态。区域Ⅱ、Ⅲ、Ⅳ同理。

证毕

由此可知，$(p*,q*)$策略是本文博弈问题的进化稳定策略。在某一组博弈中，无论2架航空器初始策略偏好(p_0,q_0)取值如何，经过充足时间(充足轮次)博弈进化后，策略组合都会趋向于$(p*,q*)$状态。

据此讨论航空器碰撞损失成本D_i(或D_j)与进化稳定策略$(p*,q*)$的关系，如式(12)所示。式(11)中的航空器运行成本C_i(或C_j)、额外运行成本ΔC_i(或ΔC_j)、劣势心理成本φ_i(或φ_j)均取实验时段内同级别航空器对应数据的平均值，即$\overline{C}_i(\overline{C}_j)$、$\Delta\overline{C}_i(\Delta\overline{C}_j)$、$\overline{\varphi}_i(\overline{\varphi}_j)$。

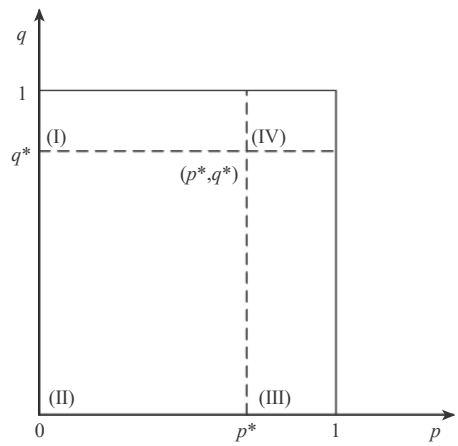


图1 系统进化状态示意图

Fig. 1 System evolution state

$$\begin{cases} D_i=\dfrac{(1-q*)(\overline{C}_i+\overline{\varphi}_i)+\Delta\overline{C}_i}{1-q*}= \\ \qquad \dfrac{(1-q*)\left(\overline{C}_i+\dfrac{m}{n}\Delta\overline{C}_i\right)+\Delta\overline{C}_i}{1-q*} \\ D_j=\dfrac{(1-p*)\left(\overline{C}_j+\overline{\varphi}_j\right)+\Delta\overline{C}_j}{1-p*}= \\ \qquad \dfrac{(1-p*)\left(\overline{C}_j+\dfrac{m}{n}\Delta\overline{C}_j\right)+\Delta\overline{C}_j}{1-p*} \end{cases} \tag{12}$$

本文设定若2架航空器的碰撞风险概率不大于基于历史数据的航空器安全事故发生概率，则判定2架航空器的博弈过程已达到稳定平衡状态。根据国际航空运输协会(International Air Transport Association, IATA)相关数据[23]，航空器发生空中事故的概率约为1/1 010 000，则博弈过程的稳定平衡状态判定标准如式(13)所示，以此计算$D_i(D_j)$。

$$(1-p*)(1-q*)\leqslant 1/1\,010\,000 \tag{13}$$

此处假设2架航空器均为理性博弈者，会通过进化学习过程调整博弈策略，向着风险更低(成本更低)的方向进化。若某架航空器为非理性博弈者，即其博弈策略不执行进化过程，则会显著改变该博弈求解过程和所需时间，第3节将具体讨论。

而且平衡状态$(p*,q*)$并非是一次博弈的结果，需要较多轮次的博弈协调。但是航空器在路径博弈协调过程中仍处于相互接近状态，因此，对于博

弈运算所需的轮次 (博弈时间) 有较为严格的限制要求。第 3 节将通过仿真实验分别讨论航空器执行不同博弈策略时的博弈效率, 也即其达到平衡状态所需的博弈轮次。以此判定相关策略是否可以在自主运行环境下得到执行。

3　仿真实验

本文选取上海区域管制中心 1 号管制扇区 (ZSSSAR01) 作为仿真空域。使用文献 [21] 所述方法, 在 1 h 时间段内随机生成 150 架航空器的运行信息, 包括机型、进出该扇区的坐标位置和时刻, 使用人工势场-粒子群联合算法为各架次航空器规划初始路径; 判断该路径是否与其他航空器存在潜在冲突, 若存在, 则为 2 架航空器的潜在冲突路径片段分别规划规避路径; 统计各航空器初始路径和规避路径长度。为保证该数据具有可靠性, 仿照该过程并重复实验 20 次。共生成 3 000 架次航空器的运行数据, 其中, 存在 309 对 (双) 航空器潜在冲突, 不存在多航空器同时同地发生冲突的情况。

分别设定不同的博弈策略, 对每组存在潜在冲突的航空器进行博弈, 统计其所需的博弈轮次和计算时间。在本节仿真实验中, 单轮次博弈所需计算时间的平均值为 0.039 s, 远小于自动相关监视广播系统 (automatic dependent surveillance broadcast system, ADS-B) 的信息交互周期 (1 s), 因此, 本文以达到稳定平衡状态所需的博弈轮次 (信息交互轮次) 代指博弈所需时间。

博弈过程中航空器初始偏好值(p_0,q_0)的设定会严重影响其达到稳定平衡状态所需的博弈轮次。本文设定航空器初始偏好值的取值区间为 [0.01,0.99], 以 0.01 为单位进行离散, 共有 99 个初始偏好值备选, 即$R_0=\{0.01,0.02,\cdots,0.98,0.99\}$。

3.1　博弈策略—正常博弈者

假设各组博弈中, 涉及的 2 架航空器i和j均为正常博弈者, 即其在博弈过程中严格遵循复制动态方程, 迭代当前轮次的偏好值, 最终达到式 (6) 所述的稳定平衡状态。分别为其设定初始偏好值$\{p_0\in R_0,q_0\in R_0\}$, 有$99\times99=9\,801$组初始值设定, 统计各组初始值设定下达到稳定平衡状态所需的博弈轮次数; 分别为本节中 309 组存在潜在冲突的航空器重复上述过程, 并计算不同航空器组在各初始值设定下所需博弈轮次的平均值, 如图 2 所示。

假设式 (4) 所述的复制动态方程中, $B=D_i-qD_i-C_i+qC_i-\Delta C_i-(1-q)\varphi_i$, $A=p^2$, 则复制动态方程与式 (14) 等价。航空器j及其复制动态方程F_q同理。

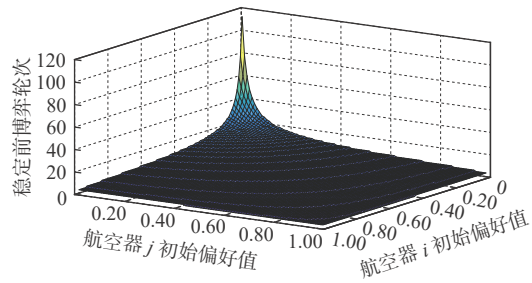


图 2　理性博弈者达到稳定平衡状态所需博弈轮次

Fig. 2　Number of game rounds required for rational players to reach a stable equilibrium state

$$F_p=p^2[D_i-qD_i-C_i+qC_i-\Delta C_i-(1-q)\varphi_i]=AB \quad (14)$$

在非极端状况$R_0=\{0.11,0.12,\cdots,0.89\}$下, 航空器平均仅需 5.31 轮次博弈即可达到均衡。如图 2 所示, 当航空器的博弈偏好初始值较小时, 随初始值逐渐增加, 达到稳定平衡状态所需的博弈轮次骤降, $(p_0,q_0)=(0.20,0.20)$仅为 (0.01,0.01)状态所需博弈轮次的 10.74%; 随数值继续增加, 所需博弈轮次仍呈下降趋势, 但是下降速度逐渐减缓至趋于稳定。当航空器的偏好值较小时, 式 (14) 所述的复制动态方程中的A项较小, 进化速度较慢, 所需博弈轮次较多; 随着航空器偏好值增大, 进化速度相对较快; 但是随偏好值持续增大, 式 (14) 的B项大幅减小, 进化速度下降, 使得所需博弈轮次在相邻初始偏好值(p_0,q_0)之间趋于稳定, 不会无限制降低至 0 或 1。

本节仅讨论了一般情况下博弈所需轮次与初始偏好值之间的关系, 3.2～3.5 节针对不同博弈策略进行仿真实验, 并分析各博弈策略较一般策略的变化趋势和相关原因。

3.2　博弈策略—激进和保守博弈者

在博弈过程中, 不同博弈者对于偏好值的进化过程存在不同的反应, 3.1 节内容的对象为严格遵照复制动态方程执行进化过程的正常博弈者; 本节额外定义“激进博弈者”为对于进化过程反应较为激烈的航空器, 其进化速度加快; “保守博弈者”反之。假设激进博弈者的复制动态方程满足式 (15); 保守博弈者的动态复制方程满足式 (16)。

$$F_{ap}=2p^2[D_i-qD_i-C_i+qC_i-\Delta C_i-(1-q)\varphi_i] \quad (15)$$

$$F_{bp}=0.5p^2[D_i-qD_i-C_i+qC_i-\Delta C_i-(1-q)\varphi_i] \quad (16)$$

为简化计算, 本节假设航空器i和j的偏好初始值相等, 即$p_0=q_0$。分别设定航空器i为正常、激进和保守策略博弈者, 设定航空器j为正常博弈者。设定不同偏好初始值$\{p_0\in R_0,q_0\in R_0,p_0=q_0\}$并逐一进行实验。对于 309 组存在潜在冲突的航空器分

别重复上述过程，计算不同航空器组在相同初始值设定下所需博弈轮次的平均值，如表3和图3所示，其中，正常、激进和保守博弈者所需轮次平均值分别为11.18、6.40和18.30。

表3 激进和保守博弈者在不同初始状态所需博弈轮次

Table 3 Number of game rounds required by aggressive and conservative players in different initial states

(p_0,q_0)	正常博弈者所需轮次	激进博弈者所需轮次	激进较正常博弈者的变化比例/%	保守博弈者所需轮次	保守较正常博弈者的变化比例/%
(0.01,0.01)	113.95	81.59	−28.39	160.04	40.44
(0.10,0.10)	18.85	13.48	−28.48	28.01	48.59
(0.20,0.20)	13.25	10.29	−22.34	18.91	42.71
(0.30,0.30)	10.34	7.64	−26.11	16.09	55.60
(0.40,0.40)	8.50	7.41	−12.82	11.72	37.88
(0.50,0.50)	7.74	5.79	−25.19	9.64	24.54
(0.60,0.60)	7.46	5.28	−29.22	8.98	20.37
(0.70,0.70)	5.98	5.01	−16.38	7.88	31.77
(0.80,0.80)	5.23	4.04	−22.75	6.83	30.59
(0.90,0.90)	5.08	3.59	−29.33	5.96	17.32
(0.99,0.99)	3.62	3.29	−9.12	4.21	16.29

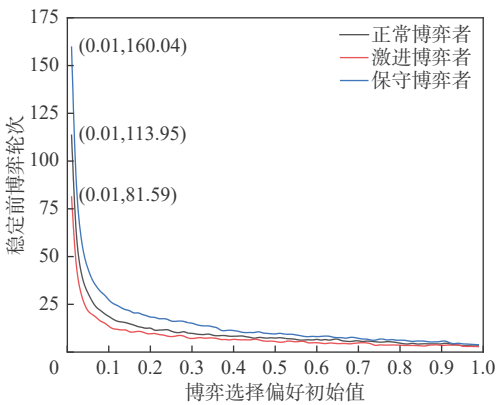


图3 激进和保守博弈者所需博弈轮次

Fig. 3 Number of rounds required to reach a stable equilibrium state of aggressive and conservative players

如表3和图3所示，相较正常博弈者，激进博弈者所需博弈次数平均减少22.74%，保守博弈者所需博弈次数平均增加33.28%。该变化量在不同初始值区间的分布不均匀，随偏好初始值增加，不同策略间所需博弈次数的差异逐渐减小。其原因在于随偏好初始值增加，所需博弈次数快速下降并趋向于稳定；且3.1节已论述了博弈轮次不会无限制下降至0或1，因此，3种策略对应的博弈轮次之间的差异随初始偏好值增加而逐渐缩小。

3.3 博弈策略—不合作博弈者

在自主运行环境下，不同航空器的博弈偏好和策略受多种因素影响而存在很大差异，例如航空器自主运行能力，航空器运行优先程度等。对于某些高优先级的航空器(如重要旅客包机等)，因其更追求时效性而倾向于选择原始路径运行，且妥协意愿小，也即其偏好值不随博弈过程进化学习。本文设定此类博弈者为“不合作博弈者”。对于存在潜在冲突的某航空器i和j，假设航空器i为不合作者，航空器j为正常博弈者，为其设定不同的偏好初始值$\{p_0=\{0.3,0.5,0.7,0.9\},q_0\in R_0\}$并逐一进行实验；计算各航空器在相同初始值下所需博弈轮次，如表4和图4所示，其中，表4中第2列～第6列博弈轮次均值分别为11.18、16.99、40.81、95.26、288.58(平均110.41)，第3列～第6列博弈轮次平均变化比例分别为51.98%、264.96%、751.88%、2 480.85%。

由表4和图4可知，在博弈过程中，若航空器i执行不合作策略，则航空器组i和j达到稳定平衡状态所需的博弈轮次远高于正常博弈策略。其原因在于2架航空器在正常博弈策略下均会根据每轮次计算所得的复制动态方程而调整其博弈偏好值；但是若航空器i执行不合作策略，则只有航空器j的选择偏好随博弈过程进行而不断调整，进化速度相对较慢；另一方面若不合作博弈者的偏好值恒定较小(如$p=0.3$)，则q相同下的碰撞风险值$(1-p)\cdot(1-q)$较高，因此，为达到稳定平衡状态，航空器j也需更多轮次的博弈学习过程。

3.4 博弈策略—成本补偿

由3.3节可知，若航空器组内某架航空器执行不合作策略，将使得达到稳定平衡状态所需的博弈轮次大幅度增加，安全风险较高。而另一方面，执行不合作策略的航空器，其时间成本亦相对较高，多为时间敏感型而非价格敏感型博弈者，因此，本文假设此类航空器有意愿通过对博弈对方提供一定程度的成本补偿而换取对方更大程度的妥协，以此减少达到平衡所需的博弈轮次并加快博弈进程。该补偿过程可由航空器所在航司之间相互结算，本文不做具体设定。

表 4　不合作博弈者在不同初始状态所需博弈轮次

Table 4　Number of rounds required by non-cooperative players in different initial states

q_0	正常博弈者 $p_0=q_0$	不合作博弈者 $p_0=p=0.9$	不合作博弈者 $p_0=p=0.7$	不合作博弈者 $p_0=p=0.5$	不合作博弈者 $p_0=p=0.3$
0.01	113.95	18.43	42.50	99.60	301.03
0.10	18.85	16.42	41.15	99.48	299.33
0.20	13.25	16.93	41.47	98.61	298.44
0.30	10.34	17.29	42.00	97.98	297.18
0.40	8.50	16.99	41.77	97.89	294.66
0.50	7.74	16.78	41.67	96.83	293.12
0.60	7.46	17.78	41.12	96.05	290.49
0.70	5.98	17.58	41.31	94.74	288.53
0.80	5.23	16.43	40.27	93.13	283.68
0.90	5.08	16.76	39.70	90.46	275.83
0.99	3.62	15.56	35.95	83.17	252.16

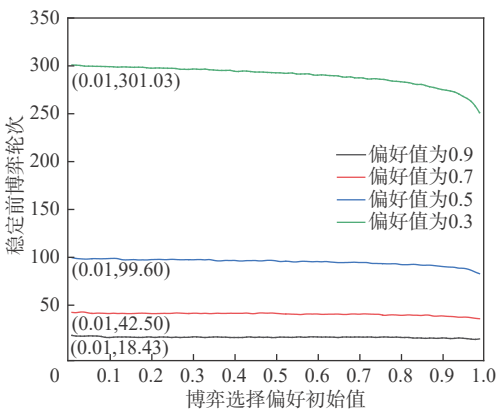


图 4　不合作博弈者所需博弈轮次

Fig. 4　Number of rounds required to reach a stable equilibrium state of non-cooperative players

本节分别假设成本补偿的数值为航空器运行成本的不同倍数, 假设航空器i为不合作博弈者, 其偏好值恒定 $p=p_0=0.5$, 航空器 j为正常博弈者。计算各补偿策略设定下航空器组所需博弈轮次的平均值, 如表 5 和图 5 所示, 由表 5 可知, 第 2 列～第 6 列博弈轮次的均值分别为 95.26、94.02、85.80、83.89、81.47(平均 86.87), 第 3 列～第 6 列博弈轮次的平均变化比例分别为−1.31%、−9.93%、−11.94%、−14.49%。

如表 5 和图 5 所示, 成本补偿策略可缩减达成稳定平衡状态所需的博弈轮次, 且博弈轮次的缩减程度随着补偿数额增大而增大。但是另一方面, 成本补偿后所需的博弈轮次仍远多于正常博弈策略。其原因在于$D_j >> C_j$, 补偿的成本数值对于式 (14) 中B部分整体数值的影响较小, 也即对于进化速度的影响较小, 因此, 不会显著缩减达成稳定状态所需的博弈轮次。

3.5　博弈策略—碰撞损失成本变化

在未来环境中, 随着无人机技术的快速发展, 大型 (货运) 无人机具备与有人驾驶航空器相似的飞行性能, 并实现混合运行。但是此类无人机的碰撞损失成本远小于有人驾驶航空器。本节模拟此类运行场景, 假设航空器i为无人机, 其碰撞损失成本D_i较传统航空器得到一定缩减; 航空器 j为传统

表 5　成本补偿策略在不同初始状态所需博弈轮次

Table 5　Number of game rounds required by cost compensation strategy in different initial states

q_0	不合作博弈者, 不补偿	不合作博弈者, 补偿ΔC_j	不合作博弈者, 补偿C_j	不合作博弈者, 补偿$2C_j$	不合作博弈者, 补偿$4C_j$
0.01	99.60	98.17	90.19	87.68	86.22
0.10	99.48	98.47	89.59	86.99	85.17
0.20	98.61	96.17	89.12	87.99	84.29
0.30	97.98	97.25	88.24	86.41	84.30
0.40	97.89	96.16	87.52	86.84	83.80
0.50	96.83	95.11	87.29	85.29	83.23
0.60	96.05	94.94	86.16	84.30	82.14
0.70	94.74	93.51	84.89	83.43	81.29
0.80	93.13	92.21	84.17	82.36	79.01
0.90	90.46	89.28	81.99	79.34	77.16
0.99	83.17	82.95	74.69	72.15	69.54

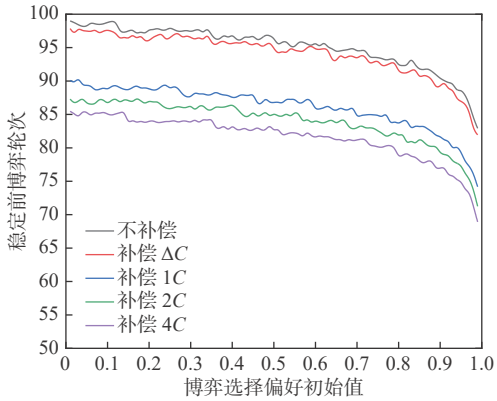


图5 成本补偿策略所需博弈轮次

Fig. 5 Number of game rounds required to reach a stable equilibrium state of cost compensation strategy

航空器,航空器i和j均执行正常博弈策略。为简化计算,本节假设航空器i和j的偏好初始值相等,即$p_0=q_0$。设定不同偏好初始值$\{p_0\in R_0, q_0\in R_0, p_0=q_0\}$并逐一进行实验,结果如表6和图6所示。由表6可知,第2列~第7列博弈轮次的均值分别为11.18、11.90、14.59、18.87、25.83、45.91,第3列~第7列博弈轮次的平均变化比例分别为6.41%、30.49%、68.79%、131.01%、310.59%。

由表6和图6可知,随航空器碰撞损失成本下降,达成稳定平衡状态所需的博弈次数大幅度上升。由式(14)可知,碰撞损失成本下降导致B部分数值下降,航空器偏好值的进化速度减慢,需要更多次博弈才可逐渐趋于稳定平衡状态。以心理学的角度考虑,碰撞损失成本下降使得博弈者在最劣情况(航空器碰撞)面临的成本下降,因此,航空器较正常博弈情况更倾向于选择激进(高风险)策略,需要更多轮次的博弈才能达到稳定平衡状态。

表6 碰撞损失成本调整在不同初始状态所需博弈轮次

Table 6 Number of rounds required by collision loss cost adjustment in different initial states

q_0	碰撞损失成本为D	碰撞损失成本为0.9D	碰撞损失成本为0.7D	碰撞损失成本为0.5D	碰撞损失成本为0.3D	碰撞损失成本为0.1D
0.01	113.95	118.82	134.53	159.93	201.07	328.30
0.10	18.85	20.34	23.03	27.31	34.71	57.47
0.20	13.25	13.08	15.58	18.23	23.28	39.39
0.30	10.34	10.23	12.63	14.79	19.15	31.97
0.40	8.50	9.52	10.41	14.10	16.40	27.87
0.50	7.74	8.92	8.83	11.23	16.12	24.47
0.60	7.46	7.36	7.83	10.69	13.97	20.91
0.70	5.98	6.34	7.32	8.79	12.49	18.95
0.80	5.23	5.82	7.26	7.64	10.57	15.12
0.90	5.08	4.20	5.86	7.93	8.05	12.07
0.99	3.62	3.25	4.22	3.98	5.33	5.51

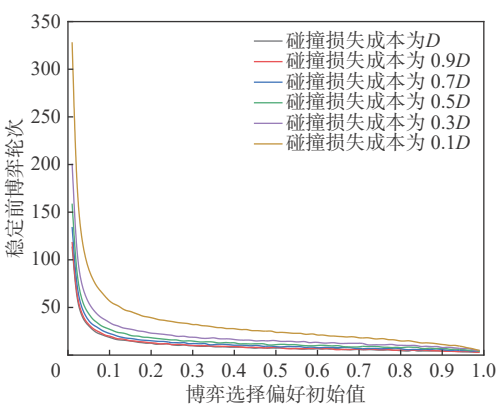


图6 碰撞损失成本调整所需博弈轮次

Fig. 6 Number of rounds required to reach a stable equilibrium state of collision loss cost adjustment

4 结 论

1)基于进化博弈理论,航空器自主路径规划-冲突解脱问题中仅有$(p*,q*)$为进化稳定状态,无论航空器路径偏好初始值如何,在博弈中均会通过进化学习过程,使其偏好值向稳定状态趋近。

2)在正常博弈策略下航空器组内达到稳定状态所需博弈轮次和博弈时间较短,具有较强的时效性。

3)航空器博弈策略严重影响达到稳定状态所需的博弈轮次,持不合作策略的航空器所需博弈轮次远多于正常策略,成本补偿策略可缓解该现象;碰撞损失成本较低的航空器所需博弈轮次也远多于正常策略。

后续研究中可拓展航空器的备选路径集合,将本问题丰富为双航空器多改航路径的博弈问题,且可模拟航空器在博弈中相向运动,导致备选路径及其成本时变的场景,使得本文方法具有更强的实用性。

参考文献(References)

[1] International Civil Aviation Organization. Global air traffic management operational concept[Z]. Montreal: ICAO, 2005.

[2] Federal Aviation Administration. NextGen implementation plan[Z].

Washington, D. C. : FAA, 2017.

[3] 蔡开泉, 赵鹏. 空中交通自主间隔管控技术研究[J]. 南京航空航天大学学报, 2022, 54(4): 688-699.
CAI K Q, ZHAO P. Key technologies on air traffic self-separation management[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2022, 54(4): 688-699(in Chinese).

[4] HAO S Q, CHENG S W, ZHANG Y P. A multi-aircraft conflict detection and resolution method for 4-dimensional trajectory-based operation[J]. Chinese Journal of Aeronautics, 2018, 31(7): 1579-1593.

[5] 陈雨童, 胡明华, 杨磊, 等. 受限航路空域自主航迹规划与冲突管理技术[J]. 航空学报, 2020, 41(9): 324045.
CHEN Y T, HU M H, YANG L, et al. Autonomous trajectory planning and conflict management technology in restricted airspace[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(9): 324045(in Chinese).

[6] CHEN Y T, HU M H, YANG L. Autonomous planning of optimal four-dimensional trajectory for real-time en-route airspace operation with solution space visualisation[J]. Transportation Research Part C: Emerging Technologies, 2022, 140: 103701.

[7] DEGAS A, KADDOUM E, GLEIZES M P, et al. Cooperative multi-agent model for collision avoidance applied to air traffic management[J]. Engineering Applications of Artificial Intelligence, 2021, 102: 104286.

[8] STRYSZOWSKI M, LONGO S, D'ALESSANDRO D, et al. A framework for self-enforced optimal interaction between connected vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(10): 6152-6161.

[9] WANG H, MENG Q, CHEN S K, et al. Competitive and cooperative behaviour analysis of connected and autonomous vehicles across unsignalised intersections: a game-theoretic approach[J]. Transportation Research Part B: Methodological, 2021, 149: 322-346.

[10] 马庆禄, 聂振宇. 基于博弈论的无信号交叉口冲突消解方法[J]. 重庆理工大学学报 (自然科学), 2021, 35(10): 144-151.
MA Q L, NIE Z Y. Conflict resolution method based on game theory at unsignalized intersection[J]. Journal of Chongqing University of Technology (Natural Science), 2021, 35(10): 144-151(in Chinese).

[11] 雷爱国, 胡启洲, 李慧慧, 等. 无信号交叉口过街行人与司机演化博弈行为研究[J]. 南京理工大学学报, 2020, 44(6): 705-714.
LEI A G, HU Q Z, LI H H, et al. Evolutionary game behavior between crossing pedestrians and drivers at unsignalized intersections[J]. Journal of Nanjing University of Science and Technology, 2020, 44(6): 705-714(in Chinese).

[12] AGNEL TONY L, GHOSE D, CHAKRAVARTHY A. Correlated-equilibrium-based unmanned aerial vehicle conflict resolution[J]. Journal of Aerospace Information Systems, 2022, 19(4): 283-304.

[13] 张宏宏, 甘旭升, 孙静娟, 等. 针对合作型无人机的最优防相撞策略[J]. 计算机工程与应用, 2022, 58(4): 290-297.
ZHANG H H, GAN X S, SUN J J, et al. Optimal anti-collision strategy for cooperative UAV[J]. Computer Engineering and Applications, 2022, 58(4): 290-297(in Chinese).

[14] ALBABA B M, MUSAVI N, YILDIZ Y. A 3D game theoretical framework for the evaluation of unmanned aircraft systems airspace integration concepts[J]. Transportation Research Part C: Emerging Technologies, 2021, 133: 103417.

[15] PANG Y T, ZHAO X Y, HU J M, et al. Bayesian spatio-temporal grAph tRansformer network (B-STAR) for multi-aircraft trajectory prediction[J]. Knowledge-Based Systems, 2022, 249(5): 108998.

[16] PARK S G, MENON P K. Game-theoretic trajectory-negotiation mechanism for merging air traffic management[J]. Journal of Guidance, Control, and Dynamics, 2017, 40(12): 3061-3074.

[17] TANG X M, LU X N, ZHENG P C. Aircraft autonomous separation assurance based on cooperative game theory[J]. Aerospace, 2022, 9(8): 421.

[18] 韩丹, 蒋豪, 陈亚青. 微分对策理论在配对进近防撞安全间隔研究中的应用[J]. 安全与环境学报, 2022, 22(5): 2617-2623.
HAN D, JIANG H, CHEN Y Q. Application of differential game in collision safety limit for paired approach[J]. Journal of Safety and Environment, 2022, 22(5): 2617-2623(in Chinese).

[19] QIAN X W, MAO J F, CHEN C H, et al. Coordinated multi-aircraft 4D trajectories planning considering buffer safety distance and fuel consumption optimization via pure-strategy game[J]. Transportation Research Part C: Emerging Technologies, 2017, 81: 18-35.

[20] 蒋旭瑞, 吴明功, 温祥西, 等. 基于合作博弈的多机飞行冲突解脱策略[J]. 系统工程与电子技术, 2018, 40(11): 2482-2489.
JIANG X R, WU M G, WEN X X, et al. Conflict resolution of multi-aircraft based on the cooperative game[J]. Systems Engineering and Electronics, 2018, 40(11): 2482-2489(in Chinese).

[21] 王红勇, 郭宇鹏. 终端区离场航空器自主路径规划研究[J]. 北京航空航天大学学报, 2025, 51(2): 446-456.
WANG H Y, GUO Y P. Research on autonomous path planning of departing aircraft in terminal area[J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(2): 446-456(in Chinese).

[22] FREY E. Evolutionary game theory: theoretical concepts and applications to microbial communities[J]. Physica A: Statistical Mechanics and Its Applications, 2010, 389(20): 4265-4298.

[23] International Air Transport Association. 2022 airline safety performance[EB/OL]. (2023-03-07)[2023-07-15]. https://www.iata.org/en/pressroom/2023-releases/2023-03-07-01/.

Efficiency of aircraft autonomous conflict resolution based on evolutionary games

WANG Hongyong[1,*], GUO Yupeng[2]

(1. College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China;
2. Beijing Municipal Commission Of Transport, Beijing 101160, China)

Abstract: To solve the problem of autonomous path planning for aircraft, this article proposes a game coordination method for aircraft autonomous conflict resolution, and discusses the game efficiency of this method under various game strategy settings for aircraft. First, based on evolutionary Game theory, a game model of aircraft conflict resolution is constructed. By calculating the replication dynamic equation between adjacent game rounds, the aircraft iterative evolution of its game selection preference is promoted, and the game system is accelerated to reach Partial equilibrium; Construct a Jacobian matrix and determinant to analyze the stability of each Partial equilibrium solution in the game, and prove that there is one and only one Partial equilibrium solution in the game system is stable, and all aircraft participating in the game will tend to this equilibrium solution; Conduct simulation experiments using ZSSSAR01 sector airspace data, set multiple aircraft game strategies, and analyze the game time required for each strategy to reach an equilibrium state. The simulation results show that rational gamers have high game efficiency and can reach equilibrium on average within 5.31 rounds of the game; Radical and conservative gamers will accelerate and slow down the equilibrium process of the game, respectively; Non cooperative gamers will significantly slow down the game equilibrium process, requiring an average of 110.53 rounds of gaming. The operating cost compensation strategy based on non-cooperative gamers will accelerate this process (with an average of 86.87 rounds).

Keywords: air transportation; trajectory based operation; conflict resolution; evolutionary game; air traffic management

Received: 2023-07-21; **Accepted**: 2023-10-10; **Published Online**: 2023-10-20 10:42

URL: link.cnki.net/urlid/11.2625.V.20231020.0917.002

Foundation item: Tianjin Natural Science Foundation (21JCZDJC00840)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0420

一种结冰飞机失速迎角实时估计方法

江飞鸿[1,2,*]，刘贞报[1]，薛源[2]，孔满昭[2]，赵甜[2]

(1. 西北工业大学 民航学院，西安 710072；2. 航空工业第一飞机设计研究院，西安 710089)

摘 要：基于某运输机带冰风洞试验数据和结冰升力特性分析，得出结冰对小迎角(AOA)范围内升力系数影响较小的结论。设计一种基于该先验知识增强的升力系数多项式在线主成分估计方法，克服由于激励不足和回归量之间共线性导致的数据矩阵病态问题。利用结冰后升力曲线斜率变化量与最大升力系数呈现的良好线性关系估计了最大升力系数和失速迎角。与现有方法相比，所设计方法具有成熟度高、实时性好、不依赖主动激励的显著优点。仿真结果表明：所设计方法估计精度可以满足迎角保护要求，具有良好的工程应用前景。

关 键 词：结冰飞机；失速迎角；主成分估计；实时估计；迎角保护

中图分类号：V212.11

文献标志码：A **文章编号**：1001-5965(2025)09-2947-08

结冰是造成空难的主要隐患之一，小到中等程度结冰导致上升率、航程、久航时间、最大速度和加速度等性能降低；严重结冰将导致尾翼失速、机翼失速等严重危害[1]。

1998年，美国国家航空航天局(National Aeronautics and Space Administration, NASA)、伊利诺伊大学、俄亥俄州立大学联合发起了智能结冰保护系统(smart icing protection systems, SIS)研究项目[1-2]。通过传感器检测结冰的状况及防冰系统的工作状况，将数据反馈到结冰管理系统，修改飞行包线和控制律，使得飞机可以按新的控制规律继续安全飞行。SIS概念的提出开创了容冰飞行控制新领域。Hossain等[3]在假定结冰飞机模型已知条件下通过在线数值积分的方法快速预测当前输入条件下系统响应，若迎角超限则调整输入幅值，该方法可同时适用于人工操纵和自动驾驶模态。美国田纳西大学与Bihrle研究公司合作开发了结冰边界保护系统[4]，实时估计飞机结冰严重程度参数，以此估计飞行包线，并向驾驶员提供安全飞行包线提示。实时参数估计中需要多轴正交输入激励。

国内相对起步较晚，对结冰后的飞行控制律重构、飞行包线估计和保护等研究还比较薄弱，总体处于跟踪研究状态。南京航空航天大学团队在国内外飞机结冰研究基础上建立了简化的飞机结冰气动模型，对于人工驾驶、自动驾驶飞行2种情况，分别设计了开环、闭环飞行包线保护控制律[5]；复旦大学团队研究了基于自组织特征映射网络的结冰定性检测、基于H_∞理论的结冰定量检测方法并设计了边界保护控制律[6-7]；空军工程大学团队提出了一种通过减小关键参数变化速率的最优边界保护算法[8]，基于相平面法[9]和正规形法[10]构建了结冰飞机的非线性稳定域，以此确定结冰后飞机安全边界，研究中假设结冰飞机气动数据已知。

总体而言，容冰飞行控制尚属新兴研究领域，目前对结冰后的失速迎角估计方法研究较少。主要通过H_∞参数辨识方法对结冰前后俯仰力矩导数进行估计，将估计结果与输入信号一起输入分类神经网络，通过神经网络检测分类后得出结冰严重程度系数。然后利用气动试验数据确定结冰严重程度系数与结冰升力导数的关系，根据结冰前后升力

收稿日期：2023-06-28；录用日期：2023-09-14；网络出版时间：2023-10-25 17:18
网络出版地址：link.cnki.net/urlid/11.2625.V.20231025.1107.002
*通信作者. E-mail：370812450@qq.com

引用格式：江飞鸿，刘贞报，薛源，等. 一种结冰飞机失速迎角实时估计方法[J]. 北京航空航天大学学报，2025，51(9)：2947-2954.
JIANG F H，LIU Z B，XUE Y，et al. A real-time estimation method for stall angle of attack of icing aircraft[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2947-2954 (in Chinese).

导数变化量预测最大升力系数和失速迎角。结冰严重程度系数定义为飞机结冰后实际阻力系数增量与最大连续结冰条件下的阻力系数增量之比，相关性研究工作还远没有完成，需要整合更多参数提高成熟度和估计精度[2-3,5-7]。此外，由于辨识基于飞机动力学响应过程，需要特定输入激励和递归计算，给实际应用带来困难。

本文在分析某运输机带模拟冰型风洞试验数据基础上，结合结冰对飞机升力特性影响机理，得出结冰对小迎角升力系数影响较小的结论。利用这一先验知识增强通过飞机法向过载间接实测的升力系数，并运用主成分估计(principal components estimation, PCE)方法克服三次多项式回归中存在的共线性问题，实时在线估算结冰后升力曲线斜率变化量。再利用结冰前后升力曲线斜率变化量与结冰后最大升力系数具有较好线性关系这一特征，估计最大升力系数和失速迎角。本文方法具有以下显著优点：

1）通过测量法向过载得到升力系数测量方程，直接估计结冰前后升力曲线斜率变化，不依赖于构造复杂且成熟度不高的结冰严重程度系数；

2）将最小化特定输入条件下结冰动力学模型和飞机实际动态响应误差的结冰严重程度系数辨识问题简化为升力系数多项式回归问题，参数估计过程中无需迭代计算，实时性好；

3）经先验知识增强实测数据后对实时测试数据需求小，辨识过程中无需专门的输入激励，具有工程实用性。

1 结冰升力特性

干净飞机小迎角范围内，升力系数与迎角成较好线性关系。随着迎角增加，驻点后移，气流沿翼型上表面运动的距离更长，流速更快，但能量不足，产生局部分离。超过失速迎角后，机翼表面大面积气动分离，升力大幅下降。机翼前缘冰聚集将引起当地的局部气流分离，导致飞机流动特性恶化，引起翼型压强再分布，升力系数明显减小。同时，由于气动外形遭到破坏，促使气流提早分离，因而使得失速迎角下降，冰型不同对失速迎角的影响也不一样[11-12]。

为保证飞机能够在恶劣结冰气象条件下安全飞行，并确定飞机结冰后的性能和品质，某运输机按照《运输类飞机适航标准》(CCAR-25-R4)[13]开展了适航符合性验证工作。主要包括临界冰型确定、带冰风洞试验、模拟冰型试飞和自然结冰试飞等内容。其中，临界冰型是整个结冰验证工作的基础，为后续工作提供基本输入，综合考虑气象条件、飞行阶段进行计算并结合冰风洞试验确定[14]。经过综合性能、操稳和动力专业对于结冰临界性的不同要求，对比评估各种冰型对飞机影响的严重程度，最终确定待机冰型、失效冰型和延迟冰型为临界冰型。

图1为飞机低速巡航构型带冰风洞试验得到的升力曲线，模型比例为1∶10，其中，α为迎角，C_y为升力系数。为验证结冰防护区域设置合理性，试验中除研究临界冰型外，还考虑了机翼展向不同部位加装冰型后的气动力特性影响。

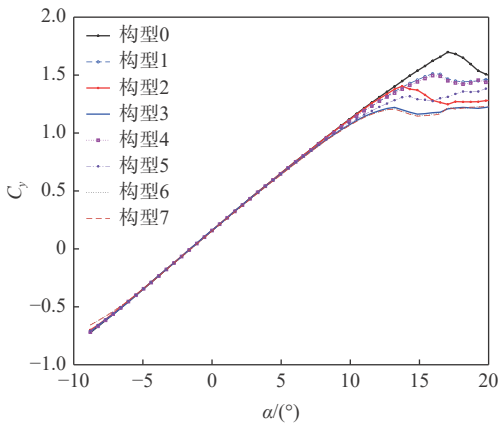


图1 不同结冰构型升力系数

Fig. 1 Lift coefficients of different icing configurations

其中：构型0为无冰干净构型。构型1为待机冰型，根据适航相关咨询通告建议，需要考虑45 min的待机飞行结冰。在待机冰型确定时，认为飞机的结冰防护系统正常工作，待机冰型主要考虑飞机非防护表面的结冰情况。构型2为延迟冰型，指在结冰气象条件下，飞行员接到结冰告警后启动防除冰系统，防除冰系统启动前的结冰情况，包括了告警延迟时间、系统响应时间的冰积聚。构型3为失效冰型，指飞机结冰防护系统失效后，防护表面的结冰冰型。根据适航咨询通报建议，失效冰型的结冰时间取最大待机结冰的1/2，即22.5 min。构型4在待机冰型基础上减去机翼翼尖冰型。构型5在构型4基础上增加机翼中段冰型。构型6在构型5基础上增加外翼冰型，不含翼尖。构型7在构型6基础上增加机翼翼尖冰型。

由于构型2即延迟冰型结冰时间短，冰型较薄，模拟冰型难于与机体等比例加工，实际模拟冰型相对较厚，因此，试验数据可信度较弱，在后续分析中不予使用。构型1和构型4、构型6和构型7升力曲线差异极小，说明机翼翼尖区域结冰对飞机升力特性影响可忽略不计。构型3和构型7在机翼结冰构型相同，但构型3水平安定面带冰，而构

型 7 水平安定面无冰, 二者升力曲线差别细微, 最大升力系数相差 1% 左右, 说明水平安定面结冰对全机升力系数影响小。

对图 1 进一步分析发现, 小迎角时冰型对升力系数影响较小。小迎角范围内, 虽然气动力部件前缘出现冰积聚, 但结冰引起的局部分离流动很小, 影响范围也很小, 仅限于前缘区域, 因此, 小迎角升力特性和干净构型升力特性一致性好。随着迎角增加, 分离区逐渐向后扩散, 引起整个翼面的分离, 造成大的升力损失。对于本节研究对象, 迎角大于 7°后, 各构型升力曲线开始出现较明显分离。

2　结冰失速迎角实时估计

2.1　升力曲线模型结构

本节先建立用于结冰失速迎角实时估计的升力曲线模型。根据失速迎角实时估计需要, 升力曲线拟合范围选为失速迎角前的某段迎角, 此处失速迎角主要考虑升力失速。这一范围内, 升力曲线一般可采用两次或三次多项式逼近[15-16]。针对构型 0、构型 3 和构型 5 的升力曲线, 用二次和三次多项式进行拟合, 如图 2～图 4 所示。

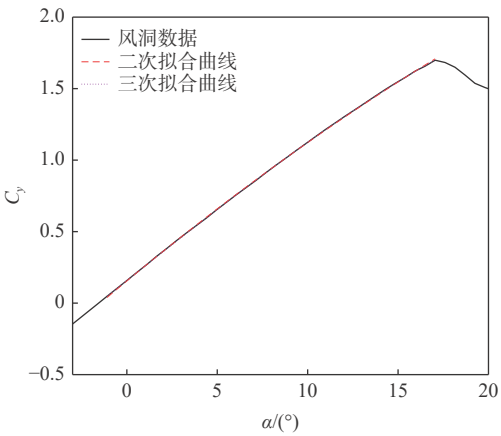


图 2　构型 0 升力系数多项式拟合曲线

Fig. 2　Polynominal fitting curves of lift coefficient of configuration 0

不难看出, 对构型 0 即干净构型二次、三次多项式均能较好拟合吹风升力曲线, 随着结冰强度的增加, 大迎角范围内升力曲线非线性特征表现更为明显, 此时三次多项式拟合精度上明显优于二次多项式, 与风洞试验数据几乎重合, 因此, 本文选用三次多项式对升力曲线进行拟合:

$$C_y = \theta_0 + \theta_1\alpha + \theta_2\alpha^2 + \theta_3\alpha^3 \tag{1}$$

式中: θ_0、θ_1、θ_2和θ_3为相关多项式系数。

特定迎角α处斜率C_y^α通过对多项式求导得到

$$C_y^\alpha = \theta_1 + 2\theta_2\alpha + 3\theta_3\alpha^2 \tag{2}$$

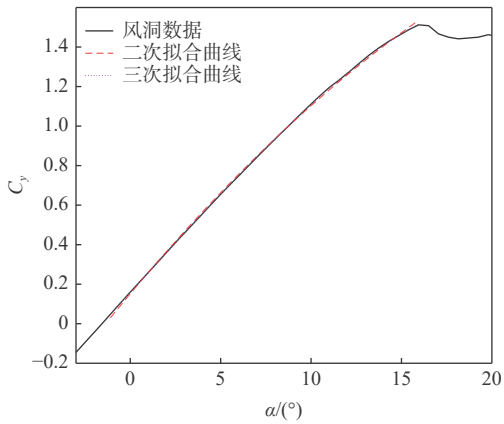


图 3　构型 3 升力系数多项式拟合曲线

Fig. 3　Polynominal fitting curves of lift coefficient of configuration 3

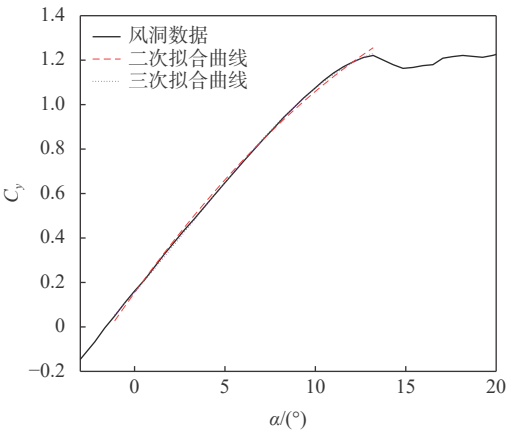


图 4　构型 4 升力系数多项式拟合曲线

Fig. 4　Polynominal fitting curves of lift coefficient of configuration 4

2.2　升力曲线实时估计

在 2.1 节建立的升力曲线模型基础上, 本节介绍模型参数向量$\hat{\boldsymbol{\theta}}$的估计方法。考虑到测量噪声影响, 升力系数测量方程可表达为

$$z(i) = \theta_0 + \theta_1\alpha + \theta_2\alpha^2 + \theta_3\alpha^3 + v(i) \tag{3}$$

式中: $z(i)$为 C_y的第 i次间接测量值, $z(i) = C_y(i) = mgn_y(i)/qS$, m为飞机质量, $n_y(i)$为法向过载第i次测量值, $i = 1,2,\cdots,N$; q为飞机动压; S为飞机机翼参考面积; $v(i)$为测量误差; N为数据点的数量。

称式 (3) 为回归方程, α、α^2、α^3称为回归量。对于线性回归问题, 通常基于最小二乘估计(least-squares estimation, LSE)求出参数$\boldsymbol{\theta} = [\theta_0,\theta_1,\theta_2,\theta_3]^{\mathrm{T}}$的估计值$\hat{\boldsymbol{\theta}}$。

LSE 下的$\boldsymbol{\theta}$最佳估计量可通过使测量值和模型值之间误差的平方和最小而获得:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{z} \tag{4}$$

式中: $\boldsymbol{X} = [\boldsymbol{I},\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3]$为$N\times4$阶矩阵, $\boldsymbol{I}$为全 1 列向量, $\boldsymbol{x}_1$、$\boldsymbol{x}_2$和$\boldsymbol{x}_3$为N次回归量α、α^2和α^3测量值构成

的列向量；$\boldsymbol{z}=[z(1), z(2), \cdots, z(N)]^{\mathrm{T}}$。

升力系数一般为飞行速度和迎角的函数，在实际飞行过程中高度实时变化，在选定辨识数据时长时应满足考虑以下3个方面：

1）在一个辨识周期内飞行速度相对固定，辨识时长不应超过飞机短周期响应时间。短周期内飞行速度变化不大，可认为该时间段内升力系数固定；

2）所选时间尺度应远小于结冰时间尺度，在该时间内结冰状态可认为是稳定的，无变化的。单位时间内机体表面所形成的冰层厚度用结冰强度表示（见表1）。

表1 飞机结冰强度等级划分

Table 1 Classification of aircraft icing intensity

结冰等级	单位时间结冰厚度/(mm·min^{-1})
微量结冰	<0.6
轻度结冰	0.6~1.0
中等结冰	1.1~2.0
严重结冰	>2.0

3）辨识时间不能太短，否则可能由于回归数据矩阵信息量不足而导致拟合精度欠佳，尤其在建立大迎角的快速机动过程中，为实现精确的升力系数测量这一点显得更为重要。

综上所述，辨识时长以飞机短周期运动时长为宜，针对研究对象飞机取采样周期为200 ms，采样时长为4 s，采样数据总数为20。

飞机在空中飞行时，存在3种典型状态：机动、扰动和定常飞行。实时在线辨识存在2方面的挑战：

1）机动和大气扰动对辨识是有利的，飞机处于受激状态，但对于定常飞行，由于缺乏激励，可能导致较长时间内飞机气动参数小范围变化，从而引起数据矩阵病态问题；

2）各回归量之间近线性相关，导致多项式系数LSE方差变大，随着多项式阶次的增加该问题将变得更加严重。

对于由于激励不足导致数据矩阵$\boldsymbol{X}$变态问题，采用先验知识进行增强，避免矩阵$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$奇异不可逆。利用小迎角范围内升力系数作为先验知识增强测量数据，理论上只要在该范围内选取不少于3个采样点，加上稳定状态，便可辨识出升力曲线。

对于回归量共线性问题，拟采用在LSE基础上改进的PCE[17-18]对升力系数多项式进行回归。

2.3 主成分估计基本原理

对$\boldsymbol{X}$中非全1向量去中心化处理，将式(3)改写为

$$\boldsymbol{z}=\theta_{0\mathrm{c}}\boldsymbol{I}+\boldsymbol{X}_{\mathrm{c}}\boldsymbol{\theta}_{\mathrm{c}}+\boldsymbol{v} \tag{5}$$

式中：$\boldsymbol{X}_{\mathrm{c}}$为去中心化后的回归量状态矩阵；$\theta_{0\mathrm{c}}$和$\boldsymbol{\theta}_{\mathrm{c}}$为中心化处理后对应多项式系数；$\boldsymbol{v}=[v(1), v(2), \cdots, v(N)]^{\mathrm{T}}$。记$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$为$\boldsymbol{X}_{\mathrm{c}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{c}}$的特征根，$\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$为对应的标准正交化特征向量，$\boldsymbol{U}=[\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n]$为$n\times n$正交矩阵。$\boldsymbol{U}$可通过对$\boldsymbol{X}_{\mathrm{c}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{c}}$进行奇异值分解得到

$$\boldsymbol{X}_{\mathrm{c}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{c}}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}} \tag{6}$$

式中：$\boldsymbol{\Lambda}$为特征值对角阵。再记$\tilde{\boldsymbol{X}}_{\mathrm{c}}=\boldsymbol{X}_{\mathrm{c}}\boldsymbol{U}$，$\tilde{\boldsymbol{\theta}}_{\mathrm{c}}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\theta}_{\mathrm{c}}$，则式(5)变为

$$\boldsymbol{z}=\theta_{0\mathrm{c}}\boldsymbol{I}+\tilde{\boldsymbol{X}}_{\mathrm{c}}\tilde{\boldsymbol{\theta}}_{\mathrm{c}}+\boldsymbol{v} \tag{7}$$

令

$$\tilde{\boldsymbol{X}}_{\mathrm{c}}=(\tilde{\boldsymbol{x}}_{\mathrm{c}1}, \tilde{\boldsymbol{x}}_{\mathrm{c}2}, \cdots, \tilde{\boldsymbol{x}}_{\mathrm{c}n}) \tag{8}$$

则有

$$\tilde{\boldsymbol{x}}_{\mathrm{c}i}=\boldsymbol{X}_{\mathrm{c}}\boldsymbol{u}_i \qquad i=1,2,\cdots,n \tag{9}$$

于是$\tilde{\boldsymbol{X}}_{\mathrm{c}}$的第$i$列$\tilde{\boldsymbol{x}}_{\mathrm{c}i}$是原来$n$个自变量的线性组合，其组合系数为$\boldsymbol{X}_{\mathrm{c}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{c}}$的第$i$个特征值对应的特征向量$\boldsymbol{u}_i$。在统计学上，称这些新的自变量为主成分。由式(8)和式(9)可得

$$\tilde{\boldsymbol{X}}_{\mathrm{c}}^{\mathrm{T}}\tilde{\boldsymbol{X}}_{\mathrm{c}}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{c}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{c}}\boldsymbol{U}=\boldsymbol{\Lambda} \tag{10}$$

当设计矩阵$\boldsymbol{X}_{\mathrm{c}}$呈病态时，$\boldsymbol{X}_{\mathrm{c}}^{\mathrm{T}}\boldsymbol{X}_{\mathrm{c}}$的特征根有一部分很小，设$n-r$个特征根很小，即$\lambda_r, \lambda_{r+1}, \cdots, \lambda_n \approx 0$，这意味着新的自变量$\tilde{\boldsymbol{x}}_{(cr+1)}, \tilde{\boldsymbol{x}}_{(cr+2)}, \cdots, \tilde{\boldsymbol{x}}_{cn}$在$N$次试验中波动很小，其作用可以并入常数项即这些新变量可以从模型中删除。

对$\boldsymbol{\Lambda}$、$\tilde{\boldsymbol{\theta}}_{\mathrm{c}}$、$\tilde{\boldsymbol{X}}_{\mathrm{c}}$和$\boldsymbol{U}$做分块，得

$$\boldsymbol{\Lambda}=\begin{bmatrix}\boldsymbol{\Lambda}_1 & \\ & \boldsymbol{\Lambda}_2\end{bmatrix} \tag{11}$$

$$\tilde{\boldsymbol{\theta}}_{\mathrm{c}}=\begin{bmatrix}\tilde{\boldsymbol{\theta}}_{\mathrm{c}1}\\ \tilde{\boldsymbol{\theta}}_{\mathrm{c}2}\end{bmatrix} \tag{12}$$

$$\tilde{\boldsymbol{X}}_{\mathrm{c}}=\begin{bmatrix}\tilde{\boldsymbol{X}}_{\mathrm{c}1} & \tilde{\boldsymbol{X}}_{\mathrm{c}2}\end{bmatrix} \tag{13}$$

$$\boldsymbol{U}=\begin{bmatrix}\boldsymbol{U}_1 & \boldsymbol{U}_2\end{bmatrix} \tag{14}$$

删除$n-r$个对应变量影响较小的主成分后，可得

$$\boldsymbol{z}\approx\theta_{0\mathrm{c}}\boldsymbol{I}+\tilde{\boldsymbol{X}}_{\mathrm{c}1}\tilde{\boldsymbol{\theta}}_{\mathrm{c}1}+\boldsymbol{v} \tag{15}$$

对式(15)应用LSE，得到

$$\hat{\theta}_{0\mathrm{c}}=\frac{1}{N}\sum_{i=1}^{N}z_i \tag{16}$$

$$\hat{\tilde{\boldsymbol{\theta}}}_{\mathrm{c}1}=\left(\tilde{\boldsymbol{X}}_{\mathrm{c}1}^{\mathrm{T}}\tilde{\boldsymbol{X}}_{\mathrm{c}1}\right)^{-1}\tilde{\boldsymbol{X}}_{\mathrm{c}1}^{\mathrm{T}}\boldsymbol{z}=\boldsymbol{\Lambda}_1^{-1}\tilde{\boldsymbol{X}}_{\mathrm{c}1}^{\mathrm{T}}\boldsymbol{z} \tag{17}$$

从而可以得到原$\boldsymbol{\theta}_{\mathrm{c}}$的PCE值为

$$\hat{\hat{\boldsymbol{\theta}}} \approx \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \hat{\hat{\boldsymbol{\theta}}}_{\mathrm{c1}} \\ \boldsymbol{0} \end{bmatrix} = \boldsymbol{U}_1 \boldsymbol{\Lambda}_1^{-1} \boldsymbol{U}_1^{\mathrm{T}} \boldsymbol{X}_{\mathrm{c}}^{\mathrm{T}} \boldsymbol{z} \tag{18}$$

通常通过定义主成分累计贡献率t, 比如$t \geqslant 95\%$来进行主成分筛选:

$$t = \sum_{i=1}^{r} \lambda_i \Big/ \sum_{i=1}^{n} \lambda_i \tag{19}$$

PCE 是一种有偏估计, 适当保留主成分个数可使 PCE 比 LSE 有较小的均方误差。

2.4　失速迎角估计

本节基于 2.1 节、2.2 节和 2.3 节建立的升力曲线模型和参数估计方法, 进行结冰失速迎角估计。大量的风洞试验结果表明最大升力系数$C_{y\max}$可以表示为在某一较小迎角(为便于表述, 将该迎角定义为结冰特征迎角, 记为α_{if})下升力系数增量或阻力系数增量的线性函数[3,5-7], 根据这一试验结果可以建立结冰后最大升力系数计算的近似数学模型。

根据图 1, 计算 7°、8°、9°、10°和 11°这 5 个迎角处的干净构型和 6 种结冰构型(不含延迟冰型)下的升力曲线斜率增量, 得到升力系数斜率变化量ΔC_y^{α}与最大升力系数的关系, 如图 5 所示。

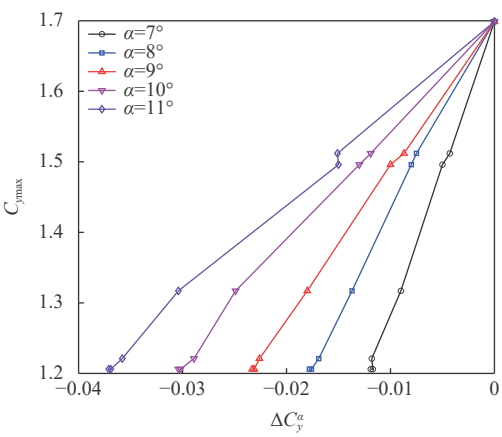


图 5　ΔC_y^{α}与$C_{y\max}$的线性关系

Fig. 5　Linear relationship between ΔC_y^{α} and $C_{y\max}$

可以发现, ΔC_y^{α}与$C_{y\max}$具有良好的线性关系, 根据式 (20) 可得结冰飞机最大升力系数$C_{y\max}$:

$$C_{y\max} = C_{y\max}\big|_{\mathrm{clean}} - k\Delta C_y^{\alpha} \tag{20}$$

$$\Delta C_y^{\alpha} = C_y\big|_{\mathrm{clean},\alpha=\alpha_{\mathrm{if}}} - C_y^{\alpha}\big|_{\alpha=\alpha_{\mathrm{if}}} \tag{21}$$

式中: $C_{y\max}\big|_{\mathrm{clean}}$为干净构型下的最大升力系数; $C_y^{\alpha}\big|_{\mathrm{clean},\alpha=\alpha_{\mathrm{if}}}$和$C_y^{\alpha}\big|_{\alpha=\alpha_{\mathrm{if}}}$为干净构型和结冰构型下结冰特征迎角处的升力曲线斜率; k为$C_{y\max}$随ΔC_y^{α}变化的变化率。

通过在线辨识方法, 可以得到当前升力曲线多项式, 通过α_{if}处结冰前后的升力系数增量ΔC_y^{α}估算$C_{y\max}$, 进而由式 (1) 可得到结冰后失速迎角。

结冰特征迎角的选择应遵循以下原则:

1) α_{if}越小越好, 这样可以较早识别出结冰现象, 并在小迎角处开始采取措施进行迎角限制;

2) α_{if}不能太小, 小α_{if}导致ΔC_y^{α}小, 可能导致结冰虚警, 且k值过大, 造成$\alpha_{\max}$对辨识误差敏感。

从图 1 不难发现, 迎角为 7°时各结冰构型升力系数差异不够明显, 特征迎角选为 11°又太晚, 造成失效冰型下需要在 2°迎角增量范围内限制迎角不超出 13.2°, 给迎角保护控制律设计带来困难。综合考虑, 对象飞机结冰特征迎角选为 9°, 对应k值为 21.3。

3　仿真分析

针对典型的失效冰型, 首先, 在不同迎角定常飞行中考核基于 PCE 的升力曲线多项式各参数估计精度及失速迎角预测精度, 再结合机动过程考核失速迎角预测精度和迎角限制效果。

3.1　定常飞行

取稳定迎角为 7°、9°、11°和 13.2°这 4 种工况, 先验升力系数在迎角为−4°、−1°、+3°处提取, 应用 PCE 得到三次多项式及其他估计参数如表 2 所示。表中, $\bar{C}_y^{\alpha}\big|_{\alpha=9^\circ}$为升力系数斜率相对真值时的比值, $\Delta\bar{C}_y^{\alpha}\big|_{\alpha=9^\circ}$为$\Delta C_y^{\alpha}\big|_{\alpha=9^\circ}$相对真值的比值。不难发现, 小迎角范围内预测精度不高, 受直线段先验值影响$C_y^{\alpha}\big|_{\alpha=9^\circ}$偏大但仍小于无冰时的升力曲线斜率, 因此, 可作为结冰定性判据, 此外, 小迎角时对迎角限制无需求, 因此, 精度可接受。随着大迎角范围内信息量的增加, 估计精度逐步逼近真值, 失速迎角为 13.2°时升力曲线斜率估计误差小于 2%, 升力曲线斜率变化量估计误差约为 5%, 对应的失速迎角预测值$\hat{\alpha}_{\mathrm{m}}$为 13.4°, 可以满足结冰定量检测及预测失速迎角的要求。

表 2　不同稳态迎角下的参数估计值

Table 2　Parameter estimates at different steady-state angles of attack

稳态迎角/(°)	$C_y^{\alpha}\big\|_{\alpha=9^\circ}$	$\bar{C}_y^{\alpha}\big\|_{\alpha=9^\circ}$	$\Delta C_y^{\alpha}\big\|_{\alpha=9^\circ}$	$\Delta\bar{C}_y^{\alpha}\big\|_{\alpha=9^\circ}$	$\hat{\alpha}_{\mathrm{m}}$/(°)
7	0.066 5	0.921 0	0.033 6	1.204 3	12.414
9	0.078 5	1.087 2	0.021 6	0.774 1	14.069
11	0.079 3	1.098 3	0.020 8	0.745 5	14.180
13.2	0.073 6	1.019 3	0.026 5	0.949 8	13.394

在对升力曲线在线辨识过程中, 矩阵$\boldsymbol{X}$的条件数决定了估计值对输入扰动的灵敏性[19], 相同扰动水平下大的条件数将导致大的估计偏差。图 6 为考虑 2% 随机噪声水平下, 基于普通 LSE 和 PCE 得到的多项式各参数相对变化率(相对无噪声时基准值的比值)在 50 次试验中的变化曲线。

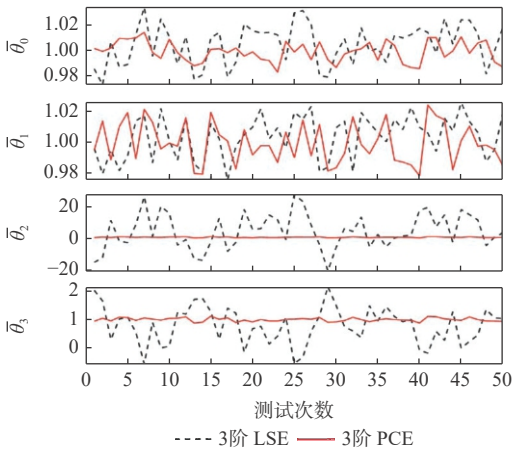


图6 多项式各参数相对变化曲线

Fig. 6 Relative change curves of polynominal parameters

可以看出，在反映升力曲线主要信息的θ_0和θ_1这2项参数上，在测量噪声下LSE和PCE精度相当，但在非主要参数θ_2和θ_3估计上PCE明显优于LSE。这是PCE对矩阵$\boldsymbol{X}$进行了降维，降低了数据矩阵条件数从而降低了估计参数对测量噪声的敏感性。

对应的失速迎角估计曲线如图7所示，PCE明显好于LSE。

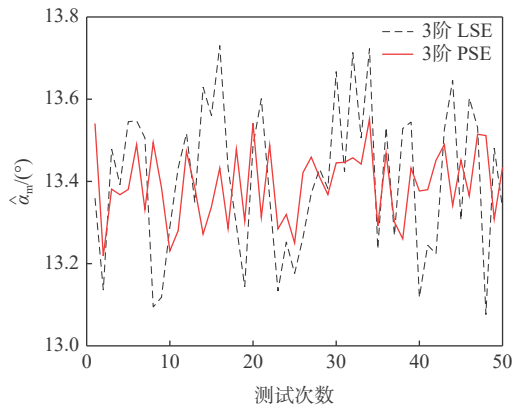


图7 随机试验中失速迎角估计值

Fig. 7 Estimated stall angle of attack in a randomized trial

3.2 机动飞行

《中国民用航空规章》第25部《运输类飞机适航标准》(CCAR-25)[13]中的第25.201条失速演示中明确要求采用直线飞行和30°坡度转弯演示失速，其中，直线飞行失速以每秒1节减速率进行失速演示，转弯飞行失速以每秒3节减速率进行失速演示。

某飞机巡航构型，气压高度为3 000 m，指示空速为360 km/h，驾驶杆位移范围为[−150,100] mm，控制律采用两级限制[20]，在油门慢车的情况下分别进行水平飞行和30°坡度转弯失速迎角保护仿真。图8为水平飞行失速迎角保护仿真曲线，图9为30°坡度转弯失速迎角保护仿真曲线。由图可知，在2种仿真工况下，成功的实现了对失速迎角的实时在线估计，随着迎角的逐步增加，失速迎角估计值分别逐步收敛到13.05°和13.09°，当拉杆到底时，飞机迎角达到且不超过失速迎角，失速迎角估计精度能够满足迎角限制要求。

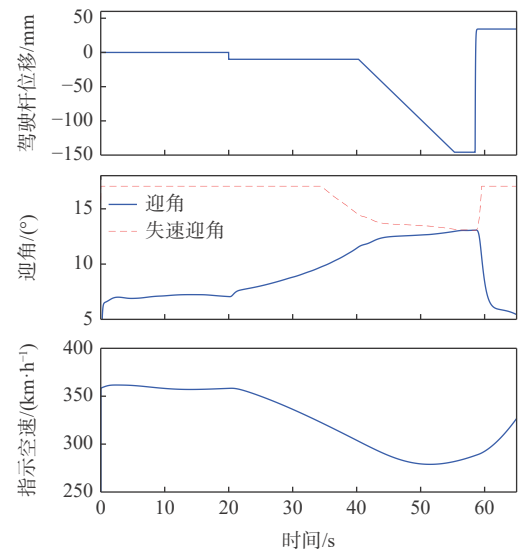


图8 水平飞行迎角保护仿真曲线

Fig. 8 Simulation curves of angle of attack protection for horizontal flight

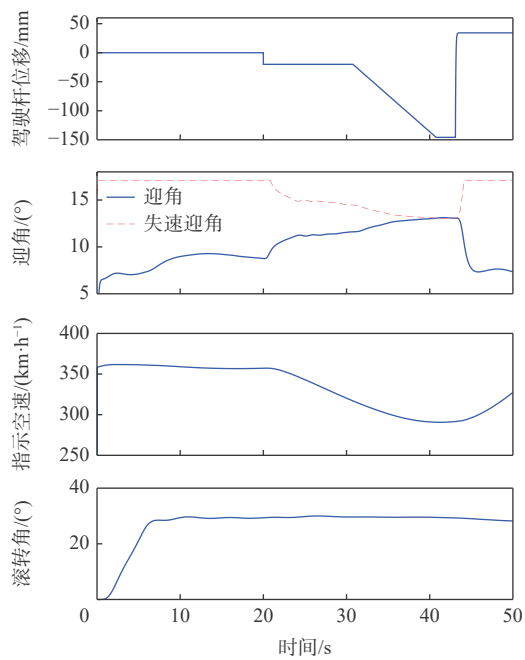


图9 30°坡度转弯迎角保护仿真曲线

Fig. 9 Simulation curves of angle of attack protection for 30° bank turn

4 结 论

1）本文提出一种基于先验知识增强的升力曲线在线PCE方法，利用结冰后升力曲线斜率变化量与最大升力系数的良好线性关系预测了最大升力

系数和失速迎角。

2）与国内外现有结冰检测方法研究相比，本文方法实时性好，不依赖于在线估计精度，可以满足迎角限制要求，具有良好的工程应用前景。

通过法向过载间接测量升力系数时，隐含了飞机质量及其他升力面升力系数已知的假设，这些假设条件与预测精度的关系需要进一步研究。另外，结冰对飞机性能影响十分广泛，不同构型下构建的先验知识并不相同，需要深入的理论分析和模拟冰型风洞试验数据支持。

参考文献（References）

[1] MELODY J W, HILLBRAND T, BAŞAR T, et al. H∞ parameter identification for inflight detection of aircraft icing: the time-varying case[J]. Control Engineering Practice, 2001, 9(12): 1327-1335.

[2] BRAGG M, PERKINS W, SARTER N, et al. An interdisciplinary approach to inflight aircraft icing safety[C]//Proceedings of the 36th AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 1998.

[3] HOSSAIN K, SHARMA V, BRAGG M, et al. Envelope protection and control adaptation in icing encounters[C]//Proceedings of the 41st Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2003.

[4] GINGRAS D R, BARNHART B, RANAUDO R, et al. Envelope protection for in-flight ice contamination[C]//Proceedings of the 47th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition. Reston: AIAA, 2009.

[5] 张智勇. 结冰飞行动力学特性与包线保护控制律研究[D]. 南京: 南京航空航天大学, 2006.
ZHANG Z Y. Study on icing flight dynamics and envelope protection control law[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2006(in Chinese).

[6] 应思斌. 飞机容冰飞行控制系统设计的理论与方法研究[D]. 上海: 复旦大学, 2010.
YING S B. Research on theory and method of aircraft ice-holding flight control system design[D]. Shanghai: Fudan University, 2010 (in Chinese).

[7] 应思斌, 艾剑良. 飞机结冰包线保护对开环飞行性能影响与仿真[J]. 系统仿真学报, 2010, 22(10): 2273-2275.
YING S B, AI J L. Simulation of aircraft flight envelope protect in icing encounters effects on open loop dynamic[J]. Journal of System Simulation, 2010, 22(10): 2273-2275(in Chinese).

[8] 周莉, 徐浩军, 杨哲, 等. 飞机在结冰条件下的最优边界保护方法[J]. 上海交通大学学报, 2013, 47(8): 1217-1221.
ZHOU L, XU H J, YANG Z, et al. Optimal boundary protection method for aircraft under icing conditions[J]. Journal of Shanghai Jiao Tong University, 2013, 47(8): 1217-1221(in Chinese).

[9] 屈亮, 李颖晖, 袁国强, 等. 基于相平面法的结冰飞机纵向非线性稳定域分析[J]. 航空学报, 2016, 37(3): 865-872.
QU L, LI Y H, YUAN G Q, et al. Longitudinal nonlinear stabilizing region for icing aircraft based on phase-plane method[J]. Acta Aeronautica et Astronautica Sinica, 2016, 37(3): 865-872(in Chinese).

[10] 郑无计, 李颖晖, 屈亮, 等. 基于正规形法的结冰飞机着陆阶段非线性稳定域[J]. 航空学报, 2017, 38(2): 520724.
ZHENG W J, LI Y H, QU L, et al. Nonlinear stability region of icing aircraft during landing phase based on normal form method[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(2): 520724(in Chinese).

[11] 张锡金, 宋文滨, 张森. 型号空气动力学设计[M]. 上海: 上海交通大学出版社, 2020: 23-25.
ZHANG X J, SONG W B, ZHANG M. Aircraft aerodynamic design[M]. Shanghai: Shanghai Jiao Tong University Press, 2020: 23-25(in Chinese).

[12] 林贵平, 卜雪琴, 申晓斌, 等. 飞机结冰与防冰技术[M]. 北京: 北京航空航天大学出版社, 2016: 11-12.
LIN G P, BU X Q, SHEN X B, et al. Aircraft icing and anti-icing technology[M]. Beijing: Beijing University of Aeronautics & Astronautics Press, 2016: 11-12(in Chinese).

[13] 中国民用航空局. 运输类飞机适航标准: CCAR-25-R4[M]. 北京: 中国民用航空局, 2011.
Civil Aviation Administration of China. Airworthiness standards for transport category airplanes: CCAR-25-R4[M]. Beijing: Civil Aviation Administration of China, 2011.

[14] 张强. 民用飞机临界冰型确定策略浅析[J]. 民用飞机设计与研究, 2019(1): 53-58.
ZHANG Q. Determination method of critical ice shapes for large civil aircraft[J]. Civil Aircraft Design & Research, 2019(1): 53-58 (in Chinese).

[15] 张培田, 韩意新, 张喆. 飞机系统辨识理论与实践[M]. 北京: 航空工业出版社, 2019: 97-98.
ZHANG P T, HAN Y X, ZHANG Z. Theory and practice of aircraft system identification[M]. Beijing: Aviation Industry Press, 2019: 97-98(in Chinese).

[16] VLADISLAV K, EUGENE A M. Aircraft system identification: theory and practice[M]. Reston: American Institute of Aeronautics and Astronautics, 2006: 138-139.

[17] 林建忠. 回归分析与线性统计模型[M]. 2 版. 上海: 上海交通大学出版社, 2022: 134-137.
LIN J Z. Regression analysis and linear statistical model[M]. 2nd ed. Shanghai: Shanghai Jiao Tong University Press, 2022: 134-137 (in Chinese).

[18] MASSY W F. Principle components regression in exploratory statistical research[J]. Journal of the American Statistical Association, 1965, 60(309): 234-256.

[19] CALAFIORE G C, GHAOUI L E. 最优化模型[M]. 北京: 机械工业出版社, 2022: 162-165.
CALAFIORE G C, GHAOUI L E. Optimization models[M]. Beijing: China Machine Press, 2022: 162-165(in Chinese).

[20] 袁志鹏, 薛源, 巩磊, 等. 大型飞机迎角保护控制律设计及试飞技术研究[J]. 飞行力学, 2020, 38(1): 90-94.
YUAN Z P, XUE Y, GONG L, et al. Research on design and flight test technologies of large aircraft angle of attack protection control law[J]. Flight Dynamics, 2020, 38(1): 90-94(in Chinese).

A real-time estimation method for stall angle of attack of icing aircraft

JIANG Feihong[1, 2, *], LIU Zhenbao[1], XUE Yuan[2], KONG Manzhao[2], ZHAO Tian[2]

(1. School of Civil Aviation, Northwestern Polytechnical University, Xi'an 710072, China;
2. The First Aircraft Institute of AVIC, Xi'an 710089, China)

Abstract: Based on the wind tunnel test data of a transport aircraft with ice and the analysis of icing lift characteristics, it is concluded that icing has little effect on the lift coefficient in the small angle of attack (AOA) range. An online principal component estimation algorithm of lift coefficient polynomial based on prior knowledge enhancement was designed to overcome the ill-conditioned problem of data matrix due to insufficient excitation and collinearity between regression variables. The maximum lift coefficient and stall AOA were estimated by using the good linear relationship between the slope variation of the lift curve and the maximum lift coefficient after icing. Compared with traditional methods, the designed method has significant advantages such as high maturity, good real-time performance, and no need for active excitation. Simulation results show that the estimation accuracy can meet the requirements of AOA protection, and the designed method has good engineering application prospects.

Keywords: icing aircraft; stall angle of attack; principal component estimation; real-time estimation; angle of attack protection

Received: 2023-06-28; **Accepted**: 2023-09-14; **Published Online**: 2023-10-25 17:18
URL: link.cnki.net/urlid/11.2625.V.20231025.1107.002
* **Corresponding author**. E-mail: 370812450@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0471

临地协同观测中的高空气球组网轨迹优化

曲艺[1,2]，王生[1,2,*]，冯慧[1]，刘强[1]

(1. 中国科学院空天信息创新研究院，北京 100094；2. 中国科学院大学，北京 100049)

摘 要：高空气球与地基平台组网开展协同观测的需求日益迫切，覆盖一致性是其关键问题之一，面临网络拓扑时变、气球调控能力受限、临近空间风场复杂等多重挑战。为改善高空气球与地基站网协同观测的覆盖一致性，以带有副气囊且不配备动力装置的高空气球为研究对象，从垂直方向和水平方向两方面分析高空气球轨迹调控方法，研究高空气球对地观测覆盖区域，针对高空气球位置时变的特征，设计了临地协同观测覆盖一致性评价指标，提出基于鲸鱼算法的高空气球组网轨迹优化方法，并针对多种输入条件对该方法进行了仿真验证。仿真结果表明：所提方法能够大幅改善临地协同观测的覆盖一致性，尤以夏秋季准零风层存在的情况下改善效果更为明显。

关 键 词：高空气球；轨迹优化；协同观测；覆盖一致性；准零风层

中图分类号：V273

文献标志码：A **文章编号**：1001-5965(2025)09-2955-13

高空气球是利用浮力原理升空的飞行器，具有发放灵活、成本低廉、载重量大、可开展原位探测等优势，很早就被作为临近空间科学观测平台(以下简称“临基平台”)，并在大气环境、地球科学、天体物理、宇宙科学等诸多领域取得了丰硕成果[1-5]。随着技术的发展，高空气球的性能逐步提升，高空气球与地基站网构建临地协同观测网络、实现优势互补的需求日益迫切[6-9]。

临地协同观测的关键问题之一是临基平台与地基站网的覆盖一致性问题，该问题与高空气球的区域驻留能力和轨迹控制能力密切相关。早期的高空气球只能被动地随风飞行，区域驻留能力较弱，难以与地基站网开展协同规划。为改善高空气球的区域驻留能力和轨迹控制能力，近年来学者们开展了多项研究。

Bellemare 等以带有副气囊的高空气球为研究对象，利用强化学习和自我校正等手段改进了气球的自主导航算法，提高了气球的轨迹控制能力，实现了气球在 50 km 半径区域内的长时间驻留[10]。Ramesh 等对带有副气球的双球系统和带有气动帆的高空气球系统分别进行了分析，表明上述 2 类气球系统均可调整自身的运行轨迹，实现在数百千米范围内的驻留[11]。Jiang 等提出一种带有副气球的双球系统，设计了通过调整双球之间系绳长度进而改变气球运行轨迹的方法，并讨论了此类气球系统在特定区域内驻留的可行性[12]。王益平等为带有副气囊的高空气球建立了动力学模型和热力学模型，阐述了利用准零风层上下纬向风方向相反这一特点改变气球运行轨迹的方法，模拟了气球在预定区域内驻留的场景[13]。邓小龙等设计了带有副气囊和螺旋桨推进系统的临近空间快速响应浮空器，并为此类浮空器设计了东西方向利用纬向风场、南北方向利用推进系统动力的区域驻留策略[14]。柏方超等针对无动力装置和有动力装置的高空气球，分别设计了区域驻留控制器，并以平均区域驻留半径、区域驻留有效时间等指标评价了 2 类气球的区域驻留性能[15]。李魁等采用 Fourier 级数和反向传

收稿日期：2023-07-19；**录用日期**：2023-11-27；**网络出版时间**：2023-12-20 10:18
网络出版地址：link.cnki.net/urlid/11.2625.V.20231219.1651.002
基金项目：国家重点研发计划 (2022YFB3901800,2022YFB3901805)
* **通信作者**. E-mail：shengwang@aoe.ac.cn

引用格式：曲艺，王生，冯慧，等. 临地协同观测中的高空气球组网轨迹优化 [J]. 北京航空航天大学学报，2025，51（9）：2955-2967.
QU Y，WANG S，FENG H，et al. Trajectory optimization of high-altitude balloon in nearspace-ground collaborative observation [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（9）：2955-2967（in Chinese）.

播(back propagation, BP)神经网络2种方法建立了风场预报模型,比较了不同风场预测模型对浮空器轨迹控制的影响[16]。林康等采用反步法设计了浮空器高度控制器,利用状态观测器对浮空器的各种误差进行估计,实现了浮空器的高度控制[17]。

上述研究提出了多种调控高空气球轨迹的技术手段,提高了高空气球区域驻留的性能,但是没有兼顾高空气球与地基平台的协同观测需求,没有量化分析临地协同观测的覆盖一致性,也没有考虑多个高空气球组网观测的协同调控需求。

而在实际应用中,高空气球与地基平台的覆盖一致性对于科学观测质量有至关重要的影响,且面临观测网络拓扑时变、气球调控能力有限、临近空间风场环境复杂等多重挑战,亟待解决。为此,本文以带有副气囊且不配备动力装置的高空气球为研究对象,分析了此类高空气球的轨迹调控方法,设计了临地协同观测的覆盖一致性评价指标,提出临地协同观测中高空气球轨迹优化方法,并在多种输入条件下对该方法进行了仿真验证。结果表明,本文方法可以大幅改善临地协同观测的覆盖性能。

1 高空气球轨迹调控方法

本文研究的高空气球主要由球体和吊舱组成,如图1所示。球体包括主气囊和副气囊,其中,主气囊填充氦气为高空气球提供浮力,副气囊填充空气调节气球质量。球体底部安装有排气阀和风机,可以向副气囊中充入空气或排出副气囊中的空气。科学观测载荷安装在吊舱当中。

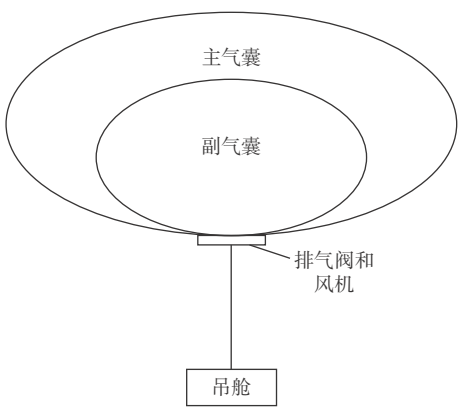


图1 高空气球结构示意图

Fig. 1 Structure illustration of high-altitude balloon

由于本文研究的高空气球没有配备动力装置,故而其轨迹调控主要依靠风机和排气阀调整高空气球的质量,从而改变高空气球的飞行高度,再利用风场在不同高度上的差异性,间接实现高空气球的轨迹控制。因此,本节的高空气球轨迹调控分析主要从垂直方向和水平方向2个维度展开。

1.1 垂直方向调控

高空气球在飞行过程中,垂直方向主要受浮力、重力、惯性附加力和气动力的影响,如下:

$$\boldsymbol{F}_{\mathrm{v}}=\boldsymbol{B}+\boldsymbol{G}+\boldsymbol{A}+\boldsymbol{D} \tag{1}$$

式中:$\boldsymbol{F}_{\mathrm{v}}$为气球在垂直方向受力之和;$\boldsymbol{B}$为浮力;$\boldsymbol{G}$为重力;$\boldsymbol{A}$为气球惯性附加力;$\boldsymbol{D}$为气动力。

气球所受浮力$\boldsymbol{B}$与所在高度的大气密度、排开空气的体积、重力加速度等因素相关,如下:

$$\boldsymbol{B}=\rho_{\mathrm{air}}(h)V\boldsymbol{g} \tag{2}$$

式中:h为高空气球所在高度;$\rho_{\mathrm{air}}(h)$为高度h处的空气密度;V为高空气球排开的空气的体积;$\boldsymbol{g}$为重力加速度。气球所受重力$\boldsymbol{G}$与主气囊内填充的氦气质量、副气囊内填充的空气质量、气球及载荷质量等因素相关,如下:

$$\boldsymbol{G}=(m_{\mathrm{He}}+m_{\mathrm{air}}(h)+m_{\mathrm{balloon}})\boldsymbol{g} \tag{3}$$

式中:m_{He}为气球主气囊内填充的氦气质量;$m_{\mathrm{air}}(h)$为高度h处气球副气囊内填充的空气质量;m_{balloon}为除填充气体外的高空气球及其携带载荷的总质量。气球所受气动力$\boldsymbol{D}$与其阻力系数、运动速度、参考面积等因素相关,如下:

$$\boldsymbol{D}=\frac{1}{2}C_D\rho_{\mathrm{air}}(h)\boldsymbol{v}_{\mathrm{v}}^2S_{\mathrm{ref}} \tag{4}$$

式中:C_D为气球阻力系数;$\boldsymbol{v}_{\mathrm{v}}$为气球在垂直方向的运动速度;$S_{\mathrm{ref}}$为气球参考面积。气球所受惯性附加力$\boldsymbol{A}$与所在高度的空气密度、气球体积、气球运动加速度等因素相关,如下[18]:

$$\boldsymbol{A}=\frac{1}{2}\rho_{\mathrm{air}}(h)V\dot{\boldsymbol{v}}_{\mathrm{v}} \tag{5}$$

若要降低气球飞行高度,可利用风机向副气囊内充入一定质量的空气,随着副气囊内空气质量的增加,气球所受重力将逐渐加大,气球向下运动,而随着气球高度的下降,空气密度增加,导致气球所受浮力增大,直至浮力与重力在新的高度建立平衡关系。反之,若排出副气囊内的部分空气,则气球所受重力减小,气球的飞行高度将会上升。

气球排气过程中副气囊内的空气质量变化可表述为[19]

$$\frac{\mathrm{d}m_{\mathrm{air}}(h)}{\mathrm{d}t}=-\sqrt{\frac{2\Delta P\rho_{\mathrm{air}}(h)(\pi R_{\mathrm{air}}^2)^2}{K}} \tag{6}$$

式中:ΔP为气球与大气环境的压差;R_{air}为阀门开口半径;K为阀系数。

气球充气过程中副气囊内的空气质量变化可

表述为[19]

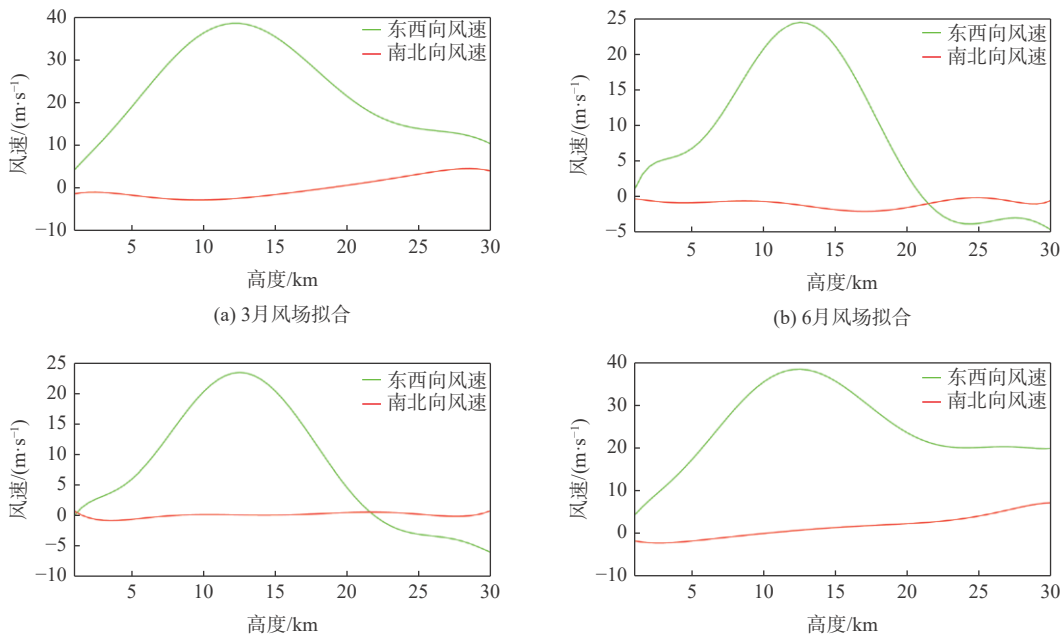
$$\frac{\mathrm{d}m_{\mathrm{air}}(h)}{\mathrm{d}t}=\rho_{\mathrm{air}}(h)V_{\mathrm{in}}(h) \tag{7}$$

式中: $V_{\mathrm{in}}(h)$ 为气球在高度 h 处的单位时间进气量。

1.2　水平方向调控

由于本文研究的高空气球没有配备动力装置, 因此, 其在水平方向主要遵循风场分布进行运动。

风场随经纬度、高度、季节、时间等发生显著变化。大体而言, 风速在对流层随高度的增加而增加, 到高度约为 10～15 km 达到极值, 而后又逐渐降低, 到高度约为 18～25 km 降至低点[20]。冬夏转换期间, 平流层底部的纬向风发生方向翻转, 在东风区和西风区之间往往存在空间范围较大、时间较为稳定的准零风层。准零风层上下的纬向风方向相反, 南北风分量也很小, 是目前高空气球驻空的主要高度[14,21-23]。图 2 为某地 3 月份、6 月份、9 月份、12 月份在 1～30 km 高度的风场拟合, 可见风场随时空变化明显, 不同高度风场差异较大, 6 月份、9 月份的纬向风场在 21 km 高度附近发生了方向翻转。



(a) 3月风场拟合　(b) 6月风场拟合

(c) 9月风场拟合　(d) 12月风场拟合

图 2　某地 1～30 km 高度的拟合风场

Fig. 2　Fitting of wind field from 1 km to 30 km for a region

因此, 若能够在垂直方向上合理控制高空气球的飞行高度, 则可以借助不同高度的风场调节气球在水平方向的轨迹, 进而改善临地协同观测的性能。

2　临地协同观测性能指标

假设在临地协同观测中, 利用 n_{G} 个地面观测站构建地基观测网络, 每个地面观测站的观测覆盖区域不随时间发生变化; 利用 n_{B} 个高空气球构建临基观测网络, 每个高空气球的观测覆盖区域随高空气球的位置发生改变。

为简化分析, 忽略高空气球在飞行过程中的姿态变化, 将其视为质点; 忽略高空气球吊舱与球体之间的高度差异, 将高空气球飞行高度视为科学观测载荷所在高度。

2.1　高空气球对地观测覆盖区域

记 t 时刻第 j 个高空气球的位置为 $\boldsymbol{P}_j(t)=(\chi_j(t),\lambda_j(t),h_j(t))$, 其中, $\chi_j(t)$、$\lambda_j(t)$、$h_j(t)$ 分别表示高空气球 j 在 t 时刻的经度、纬度、高度, 若高空气球所携带的科学观测载荷的观测角度为 2θ, 则 t 时刻第 j 个高空气球的观测覆盖半径 $r_j(t)$ 可按照式 (8) 和式 (9) 计算, 高空气球对地覆盖区域示意图如图 3 所示。

$$\frac{h_j(t)+R_{\mathrm{e}}}{\sin(\pi-\theta-\beta_j(t))}=\frac{R_{\mathrm{e}}}{\sin\theta} \tag{8}$$

$$r_j(t)=R_{\mathrm{e}}\beta_j(t) \tag{9}$$

式中: R_{e} 为地球半径; $\beta_j(t)$ 为 t 时刻第 j 个高空气球覆盖半径对应的地心角。

2.2　临地协同观测覆盖一致性

由于高空气球位置处于持续变化当中, 导致临基观测网络拓扑时变, 临基观测网络覆盖区域不定, 因此, 本文从空间和时间 2 个维度评价临地协同观测的覆盖一致性。

在空间维度, 将某时刻临地协同观测的覆盖一

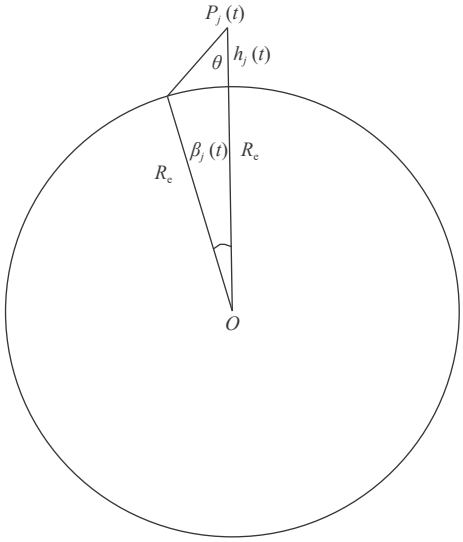


图3 高空气球对地覆盖区域示意图

Fig. 3 Schematic diagram of high-altitude balloon coverage area on ground

致性定义为该时刻临基网络观测区域与地基网络观测区域的交集在地基网络观测区域中所占的比例，如式(10)～式(12)所示；在时间维度，将整个临地协同观测过程的覆盖一致性定义为各个时刻协同覆盖一致性的平均值，如式(13)所示：

$$M = M_1 \cup M_2 \cup \cdots \cup M_{n_G} \tag{10}$$

$$S(t) = S_1(t) \cup S_2(t) \cup \cdots \cup S_{n_B}(t) \tag{11}$$

$$c(t) = (M \cap S(t))/M \tag{12}$$

$$c = \sum_{t=t_0}^{t_e} c(t)/t_g \tag{13}$$

式中：M为地基网络覆盖区域，M_i为第i个地面观测站的覆盖区域；$S(t)$为t时刻临基网络覆盖区域，$S_j(t)$为第j个高空气球在t时刻的覆盖区域；$c(t)$为t时刻临地协同观测覆盖一致性(以下简称协同覆盖一致性)；‘∩’表示覆盖区域的交集，‘∪’表示覆盖区域的并集，二者可按照计算机图形学规则进行求解；c为整个临地协同观测过程的覆盖一致性；t_0为观测起始时刻；t_e为观测结束时刻；t_g为预期观测时长。

按照上述定义，在空间维度扩大临基网络覆盖范围，可以增大协同覆盖一致性$c(t)$，改善观测过程的覆盖一致性c；在时间维度延长气球在预定空域内的飞行时间，可以增大观测结束时刻t_e，也可改善观测过程的覆盖一致性c。本文将观测过程覆盖一致性c作为后文高空气球轨迹优化的目标。

图4～图6分别给出了地基网络覆盖区域M、t时刻临基网络覆盖区域$S(t)$和临基网络观测区域与地基网络观测区域交集$M \cap S(t)$的示例。

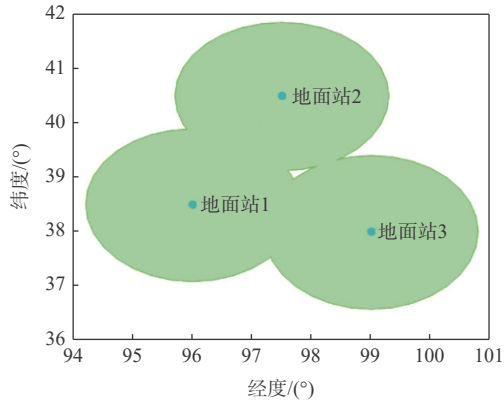


图4 地基网络覆盖区域M示例

Fig. 4 Example of ground based network coverage area M

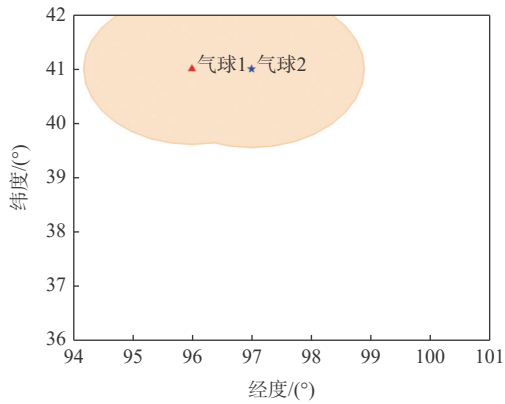


图5 t时刻临基网络覆盖区域$S(t)$示例

Fig. 5 Example of coverage area $S(t)$ of nearspace base network at time t

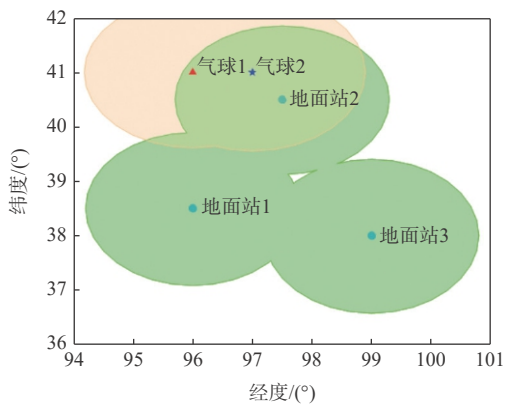


图6 地基网络覆盖区域与t时刻临基网络覆盖区域的交集$M \cap S(t)$示例

Fig. 6 Example of intersection $M \cap S(t)$ between coverage area of ground based network and coverage area of nearspace based network at time t

3 临地协同观测中的高空气球组网轨迹优化

临地协同观测中的高空气球组网轨迹优化指

在预定的空域范围内, 为参与协同观测的每个高空气球寻求从观测起始时刻到观测结束时刻的轨迹调控策略, 使之既满足各项约束条件, 又能够彼此兼顾, 达到整体最优的临地协同观测性能指标。

3.1　高空气球组网轨迹优化算法

从第 1 节的分析可知, 对于高空气球飞行轨迹的调控, 主要是调控其副气囊内的空气质量, 进而改变高空气球的飞行高度, 再利用不同高度的风场差异, 实现高空气球在水平方向上的轨迹转移, 最终改变高空气球的观测覆盖区域。为此, 本文将高空气球组网轨迹优化问题转化为高空气球副气囊内空气质量调整量的组合优化问题, 并以临空协同观测过程的覆盖一致性 c 为优化目标, 寻求每个高空气球副气囊内空气质量调整量的最优组合。

启发式算法是常用的优化问题求解方法, 其中鲸鱼算法更是由于其鲁棒性好、搜索能力强、使用方便而得到了广泛应用[24-28]。鲸鱼算法的优化迭代策略主要包括收缩包围、螺旋更新和随机搜索 3 种, 3 种策略的选择由随机概率因子 p 和系数 A_{W}、C_{W} 共同决定, 随机概率因子 p 是 [0,1] 之间的随机数, 而系数 A_{W} 和 C_{W} 的定义分别为[24]

$$a = 2 - (2n_i / N_{\max}) \tag{14}$$

$$A_{\mathrm{W}} = 2ar - a \tag{15}$$

$$C_{\mathrm{W}} = 2r \tag{16}$$

式中: n_i 为当前优化迭代次数; $N_{\max}$ 为算法预期优化迭代次数; 变量 a 的取值随着迭代次数的增加从 2 减少到 0; r 为 [0,1] 之间的随机数。

鲸鱼算法的 3 种优化迭代策略中, 收缩包围策略模拟鲸鱼围捕猎物的过程, 具体可描述为种群中的所有个体都以当前种群的最优个体位置为目标更新自身位置, 对应的数学模型为[24]

$$D_{\mathrm{W}} = |C_{\mathrm{W}} X_{\mathrm{best}}(n_i) - X(n_i)| \tag{17}$$

$$X(n_i + 1) = X_{\mathrm{best}}(n_i) - A_{\mathrm{W}} D_{\mathrm{W}} \tag{18}$$

式中: $X_{\mathrm{best}}(n_i)$ 为第 n_i 次优化迭代中的种群最优个体位置; $X(n_i)$ 为第 n_i 次优化迭代中的当前个体位置; $A_{\mathrm{W}} D_{\mathrm{W}}$ 为当前个体在本次优化迭代中的位置改变量。

螺旋更新策略模拟鲸鱼捕食的过程, 具体可描述为, 首先, 计算种群中所有个体与当前种群最优个体之间的距离, 然后, 种群中的所有个体都以螺旋线方式更新自身位置, 对应的数学模型为[24]

$$D'_{\mathrm{W}} = |X_{\mathrm{best}}(n_i) - X(n_i)| \tag{19}$$

$$X(n_i + 1) = D'_{\mathrm{W}} \mathrm{e}^{bl} \cos(2\pi l) + X_{\mathrm{best}}(n_i) \tag{20}$$

式中: b 为常数, 用于确定螺旋线形状; l 为 [−1,1] 之间的随机数。

随机搜索策略模拟鲸鱼随机寻找食物的过程, 具体可描述为当前个体在种群内随机选择一个个体作为目标并向其靠拢, 对应的数学模型为[24]

$$D_{\mathrm{W}} = |C_{\mathrm{W}} X_{\mathrm{rand}}(n_i) - X(n_i)| \tag{21}$$

$$X(n_i + 1) = X_{\mathrm{rand}}(n_i) - A_{\mathrm{W}} D_{\mathrm{W}} \tag{22}$$

式中: $X_{\mathrm{rand}}(n_i)$ 为当前个体从当前种群中随机选择的靠拢目标位置。

鲸鱼算法的基本流程如图 7 所示。

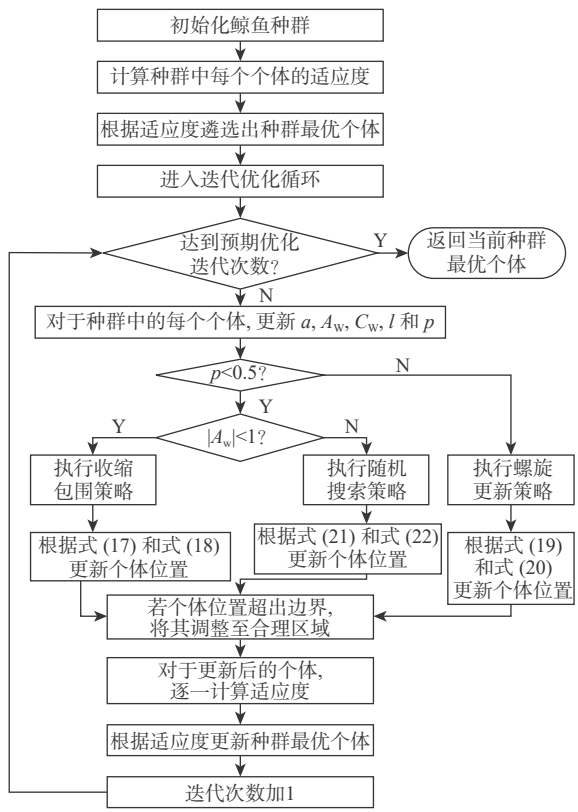


图 7　鲸鱼算法流程

Fig. 7　Flow of whale algorithm

3.2　优化约束

高空气球轨迹优化过程中, 除需要将气球副气囊内的空气质量控制在合理范围之外, 还需要考虑诸多空间约束与时间约束。

飞行高度约束。考虑到高空气球内外压差、填充氦气量等因素, 每个高空气球都有各自可行的飞行高度范围, 轨迹优化必须在此范围内开展。即若第 j 个气球的飞行高度上限为 $h_{j\max}$, 飞行高度下限为 $h_{j\min}$, t_0 时刻飞行高度为 $h_j(t_0)$, t 时刻该气球高度调整量为 $\Delta h_j(t)$, 则对于任意 $j \in [1, n_{\mathrm{B}}]$、$t \in [t_0, t_{\mathrm{e}}]$, 都应满足:

$$h_j(t_0)+\sum_{k=t_0}^{t}\Delta h_j(k)\in[h_{j\min},h_{j\max}] \tag{23}$$

飞行空域约束。高空气球一旦飞出空域就将实施吊舱切割，无法继续开展观测，此后该气球的覆盖区域可视为0，进而影响临地协同观测覆盖一致性。即若第j个高空气球的切割时间为t_{jc}，t_0时刻该高空气球所在经纬度分别为$\chi_j(t_0)$、$\lambda_j(t_0)$，则对于任意$t>t_{jc}$，$S_j(t)=0$，而对于任意$t\leqslant t_{jc}$，都应满足

$$\chi_j(t_0)+\sum_{k=t_0}^{t}v_{je}(k)\in[\chi_{\min},\chi_{\max}] \tag{24}$$

$$\lambda_j(t_0)+\sum_{k=t_0}^{t}v_{jn}(k)\in[\lambda_{\min},\lambda_{\max}] \tag{25}$$

式中：$\chi_{\min}$、$\chi_{\max}$、$\lambda_{\min}$、$\lambda_{\max}$分别为高空气球飞行空域的最小经度、最大经度、最低纬度、最高纬度；$v_{je}(k)$为k时刻高空气球j所在位置的纬向风速；$v_{jn}(k)$为k时刻高空气球j所在位置的经向风速。

时间间隔约束。由于临近空间大气稀薄，且排气阀与风机能力有限，因此，高空气球质量调整所需时间较长，尤其是充气过程，往往需要持续数十分钟乃至更长时间，具体用时可按式(6)和式(7)计算。气球轨迹优化调整的时间间隔应不少于气球排气充气所需时间。

4 仿真分析

本文以3个地基测站和2个高空气球开展协同观测为背景进行仿真，仿真工具采用MATLAB 2018b。

4.1 仿真参数设置

1）大气模型。采用美国标准大气模型[29]。

2）风场模型。为降低计算量，本文采用统一的风场进行仿真，除第4.2.3节和第4.2.4节外，均采用图2中的6月风场进行仿真。

3）参与观测的地基测站参数。测站参数如表1所示。

表1 地基站网仿真参数

Table 1 Simulation parameters of ground-based platforms

地基测站	经度/(°)	纬度/(°)	观测覆盖半径/km
地基测站1	96.0	38.5	200
地基测站2	97.5	40.0	200
地基测站3	99.0	38.0	200

4）高空气球飞行空域范围。飞行空域范围为94°E～101°E，36°N～42°N。

5）预计观测时间。预计观测时间为96 h，采样时间间隔为2 h。

6）高空气球仿真参数。仿真参数如表2所示。

表2 高空气球仿真参数

Table 2 Simulation parameters of high-altitude balloons

气球	初始经度/(°)	初始纬度/(°)	初始飞行高度/km	最低飞行高度/km	最高飞行高度/km	副气囊内初始空气质量/kg	阀门开口半径/m	单位时间进气量/($m^3\cdot s^{-1}$)
高空气球1	96	41	21	20	24	144.7	0.15	0.3
高空气球2	97	41	22	20	24	111.1	0.15	0.3

4.2 仿真结果

4.2.1 轨迹优化结果与无轨迹优化结果对比

图8为从气球副气囊内空气质量、飞行高度、飞行轨迹、协同覆盖一致性等方面对轨迹优化仿真结果与无轨迹优化仿真结果进行了对比。

由图8可知，若不采用轨迹优化算法，则气球副气囊内的空气质量不发生改变，气球始终在初始高度飞行，飞行轨迹近似直线，一旦飞出预定空域，将减少临基网络覆盖区域，降低协同观测覆盖一致性。以气球2为例，若不采用轨迹优化算法，则气球始终在22 km高度飞行，约47 h后飞出空域，无法继续开展观测，不能与气球1的覆盖区域形成配合，导致临基网络与地基网络的协同覆盖一致性较差。

若采用本文的轨迹优化方法，则通过控制每个气球副气囊内的空气质量，调整气球的飞行高度，再利用不同高度的风场差异，改变气球的飞行轨迹，延长气球在预定空域内的飞行时间；此外，由于本文方法对全部气球的空气质量进行统筹规划，使得气球之间相互配合，形成良好的网络整体覆盖，提高了协同覆盖一致性。

仿真以2 h为采样间隔，分别计算每个采样时刻的协同覆盖一致性，如图8(d)所示，再求得整个观测过程的临地协同观测覆盖一致性，结果如表3所示。从数据可知，轨迹优化结果的性能比未经优化结果的性能高了67.35%，改善效果显著。

4.2.2 高空气球网络初始布局对于临地协同观测性能的影响

为考察高空气球网络初始布局对临地协同观测性能的影响，本节对4种气球网络初始布局进行分析。

布局1。2个高空气球处于相同的初始水平位置和高度(仅限于仿真分析)；

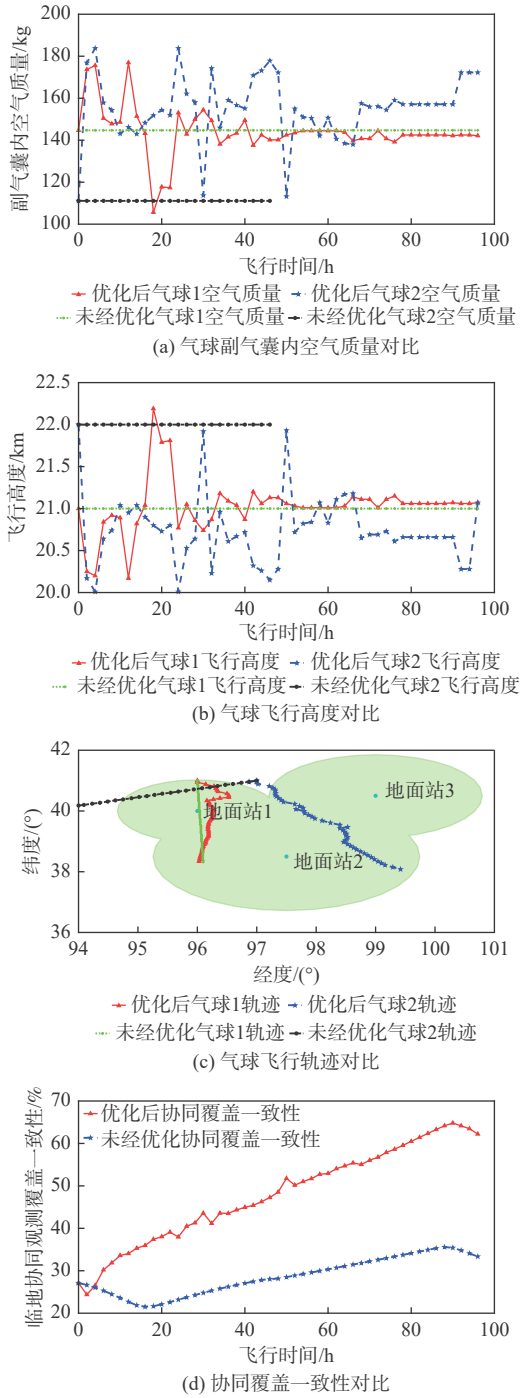


图 8　轨迹优化与无轨迹优化的仿真结果对比

Fig. 8　Comparison of optimized results and unoptimized results

表 3　轨迹优化结果与无轨迹优化结果对比

Table 3　Performance comparison between optimized results and unoptimized results

结果类型	临地协同观测覆盖一致性/%
轨迹优化结果	24.45
无轨迹优化结果	14.61

布局 2。2 个高空气球处于相同的初始水平位置, 但是初始高度不同;

布局 3。2 个高空气球处于相同的初始高度, 但是初始水平位置不同;

布局 4。2 个高空气球处于不同的初始水平位置与高度。

高空气球初始网络布局参数如表 4 所示。

表 4　高空气球初始网络布局

Table 4　Initial network layout of high-altitude balloons

布局	气球1初始经度/(°)	气球1初始纬度/(°)	气球1初始高度/km	气球2初始经度/(°)	气球2初始纬度/(°)	气球2初始高度/km
布局1	97	41	22	97	41	22
布局2	97	41	20	97	41	24
布局3	97	41	22	100	38	22
布局4	97	41	20	100	38	24

4 种初始网络布局下的轨迹优化结果如图 9 所示。

从图 9 和表 5 中可以看到, 在本文的仿真条件与优化方法下, 若不采用轨迹优化方法, 则水平位置不同的初始网络布局性能远优于水平位置相同的初始网络布局; 若采用优化方法, 则各种初始布局之间的性能差异不大, 水平位置不同的初始网络布局性能略优于水平位置相同的初始网络布局。

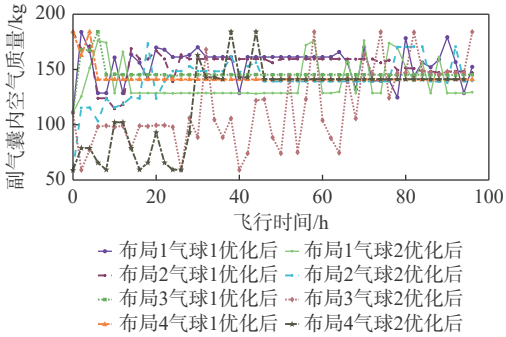
4.2.3　风场对于临地协同观测的影响

为考查风场对临地协同观测性能的影响, 本节分别利用图 2 所示的 3 月、6 月、9 月、12 月风场进行高空气球轨迹优化仿真, 结果如图 10 所示。

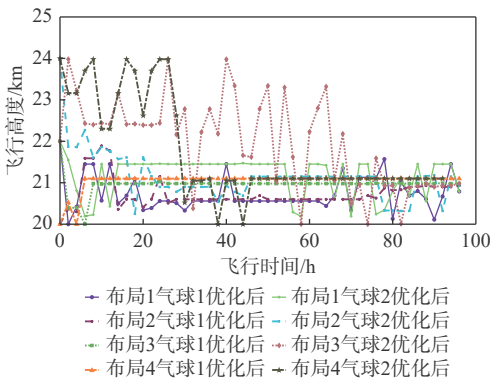
从表 6 和图 10 可以看到, 风场对于临地协同观测效果的影响极大, 最优性能与最劣性能相差可达数十倍。3 月和 12 月, 20～24 km 高度处的纬向风速可达 14～24 m/s, 且不存在风向翻转, 因此, 气球在目标区域内驻留时间较短, 即使经过轨迹优化, 临地协同观测一致性依然较差。而在 6 月和 9 月, 20～24 km 高度的纬向风速绝对值小于 5 m/s, 经向风速的绝对值也较小, 且纬向风场存在风向翻转, 气球轨迹优化的效果更好。

从表 6 和图 10 还可以看到, 虽然轨迹优化方法在上述 4 个风场中都能发挥作用, 但是其优化机理不同。在没有准零风层的 3 月和 12 月, 优化程序倾向于将气球调整到更高的飞行高度以便使每个气球覆盖更大的区域, 进而从空间维度提高协同观测覆盖一致性。而在存在准零风层的 6 月和 9 月, 优化程序的主要策略是持续调整副气囊内的空气质量, 借助不同高度的水平风场, 调整气球的纬向飞行轨迹, 既延长了气球的观测时间, 又使临基网络覆盖区域与地基网络覆盖区域更好地吻合, 从空间和时间 2 个维度提高协同观测覆盖一致性。

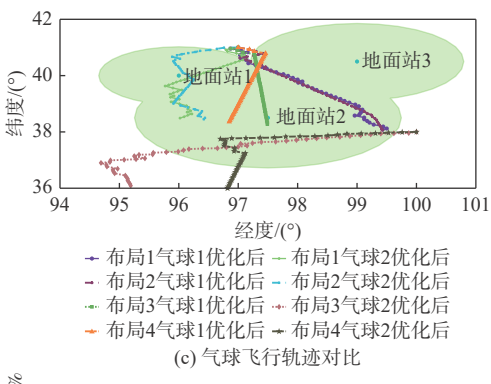
从图 10 还可以看到, 存在准零风层的 6 月和 9 月, 气球的纬向轨迹可以通过飞行高度进行调整, 但是由于气球没有配备动力装置, 导致气球的经向



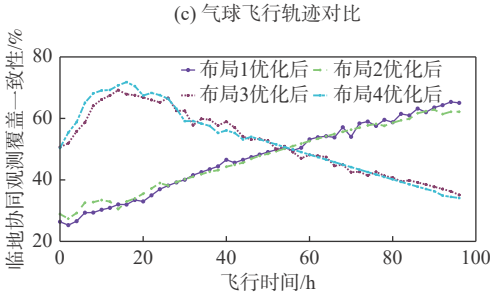
(a) 气球副气囊内空气质量对比



(b) 气球飞行高度对比



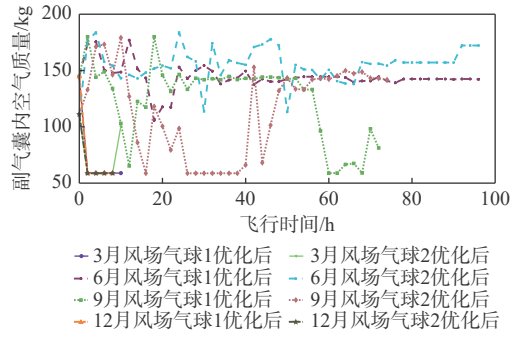
(c) 气球飞行轨迹对比



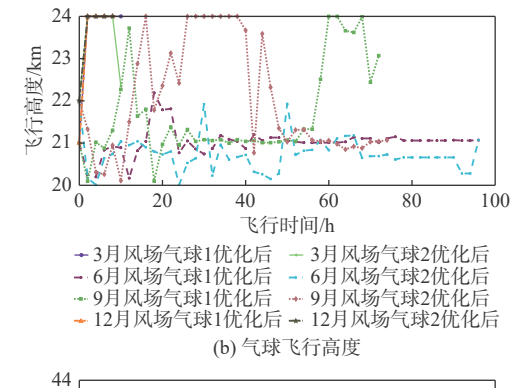
(d) 协同覆盖一致性对比

图 9 不同初始布局的优化结果对比

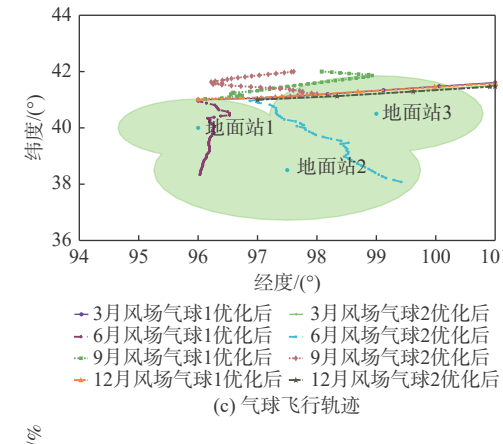
Fig. 9 Comparison of optimization results for different initial layouts



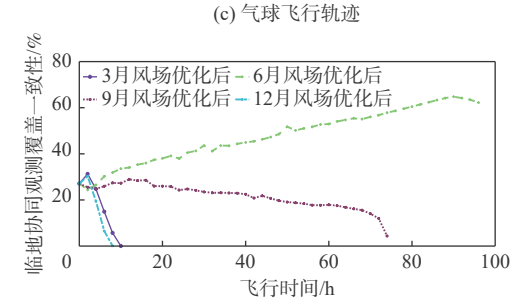
(a) 气球副气囊内空气质量



(b) 气球飞行高度



(c) 气球飞行轨迹



(d) 协同覆盖一致性

图 10 不同风场的仿真结果对比

Fig. 10 Comparison of simulation results in different wind fields

表 5 高空气球初始网络布局对协同观测性能的影响对比

Table 5 Comparison of influence of initial network layout of high-altitude balloons on collaborative observation performance

布局	轨迹优化后的临地协同观测覆盖一致性/%	无轨迹优化的临地协同观测覆盖一致性/%
布局1	23.95	4.00
布局2	23.97	4.97
布局3	26.64	13.09
布局4	26.83	12.26

表 6 不同风场的轨迹优化结果对比

Table 6 Comparison of trajectory optimization results for different wind fields

风场	轨迹优化后的临地协同观测覆盖一致性/%	无轨迹优化的临地协同观测覆盖一致性/%
3月	1.08	0.88
6月	24.45	14.61
9月	8.56	7.48
12月	0.87	0.77

轨迹暂时无法更好地调控, 因而该场景下可根据经向风场设计气球网络的初始水平布局以改善协同观测一致性。以 9 月为例, 若将 2 个气球的初始纬度从 41°N 分别改成 37°N 和 38°N, 其他条件不变, 则临地协同观测一致性可从表 6 中的 8.56% 提升至 28.38%, 仿真对比如图 11 所示。

4.2.4　高空气球高度调节能力对于临地协同观测的影响

为验证气球高度调节能力对于临地协同观测性能的影响, 本节假设高空气球的高度调节能力增强, 气球飞行高度可由表 2 的 20～24 km 扩大到 18～26 km, 依然利用图 2 所示的 3 月、6 月、9 月、12 月的风场开展仿真优化, 结果如图 12～图 15 和表 7 所示。从图 12~图 15 和表 7 可以看到, 提升高空气球的高度调节能力有助于改善临地协同观测一致

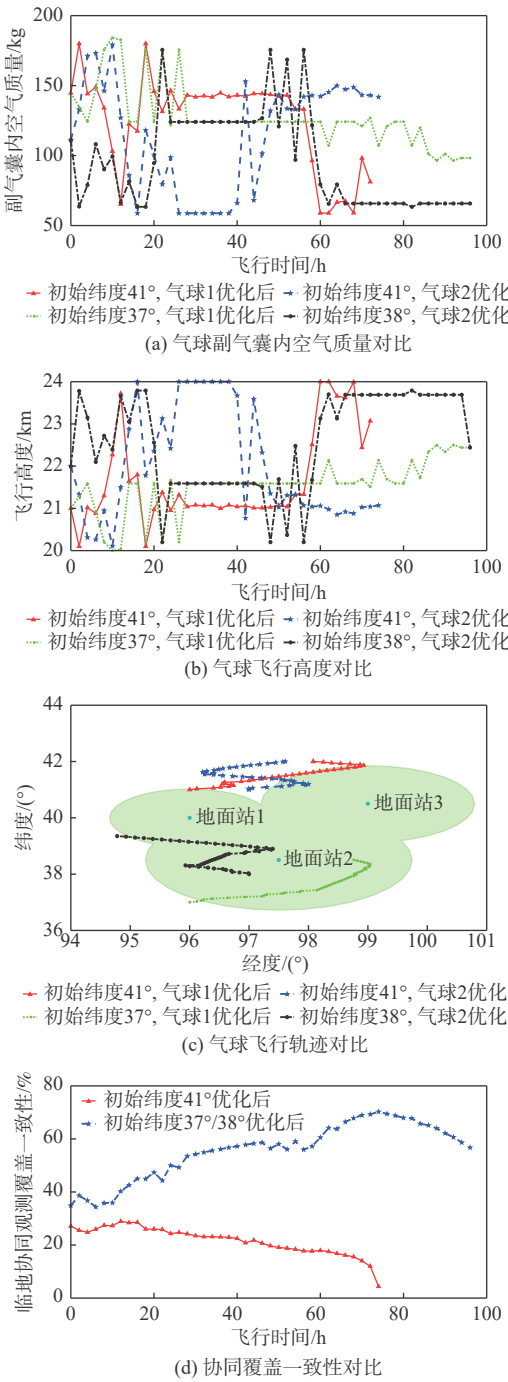


图 11　9 月风场下不同初始水平布局的轨迹优化对比

Fig. 11　Comparison of trajectory optimization with different initial horizontal layouts in wind fields of September

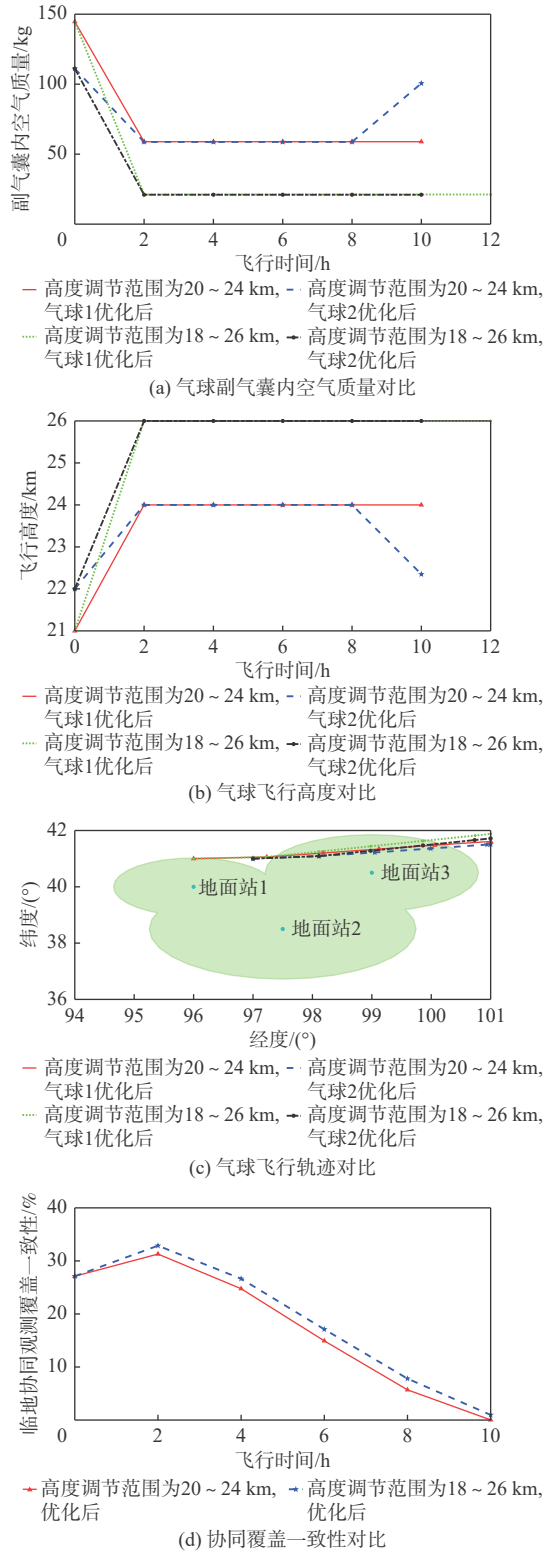
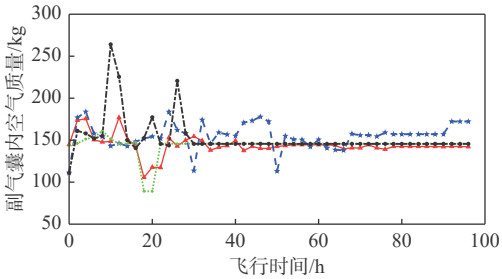


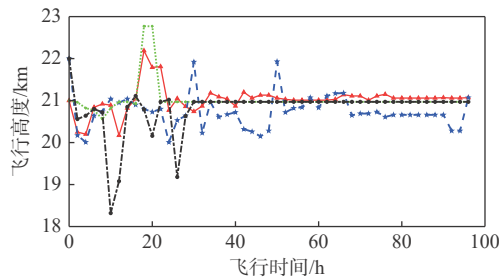
图 12　不同高度调节能力的仿真结果对比(3 月风场)

Fig. 12　Comparison of simulation results for different height adjustment capabilities (March wind field)



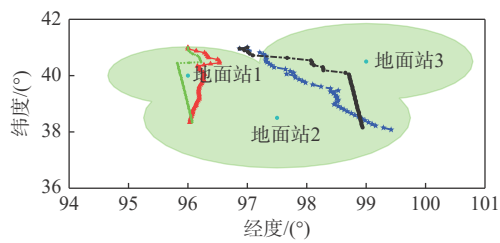
高度调节范围为20～24 km，气球1优化后　高度调节范围为20～24 km，气球2优化后
高度调节范围为18～26 km，气球1优化后　高度调节范围为18～26 km，气球2优化后

(a) 气球副气囊内空气质量对比



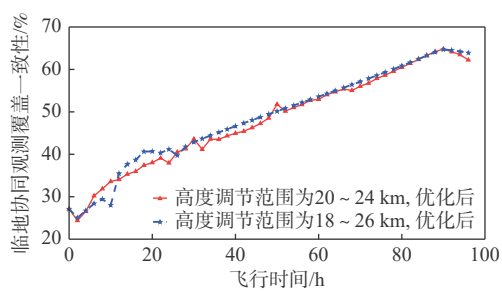
高度调节范围为20～24 km，气球1优化后　高度调节范围为20～24 km，气球2优化后
高度调节范围为18～26 km，气球1优化后　高度调节范围为18～26 km，气球2优化后

(b) 气球飞行高度对比



高度调节范围为20～24 km，气球1优化后　高度调节范围为20～24 km，气球2优化后
高度调节范围为18～26 km，气球1优化后　高度调节范围为18～26 km，气球2优化后

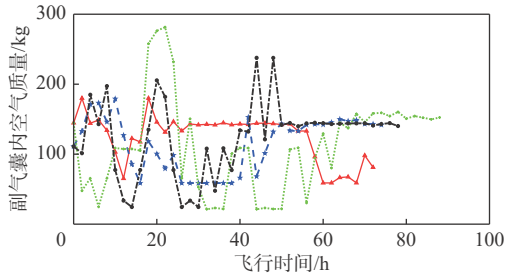
(c) 气球飞行轨迹对比



(d) 协同覆盖一致性对比

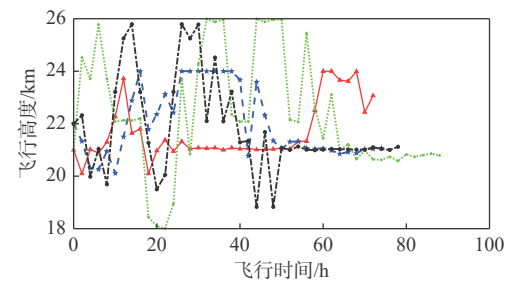
图13　不同高度调节能力的仿真结果对比(6月风场)

Fig. 13　Comparison of simulation results for different height adjustment capabilities (June wind field)



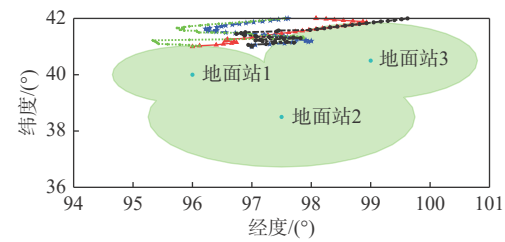
高度调节范围为20～24 km，气球1优化后　高度调节范围为20～24 km，气球2优化后
高度调节范围为18～26 km，气球1优化后　高度调节范围为18～26 km，气球2优化后

(a) 气球副气囊内空气质量对比



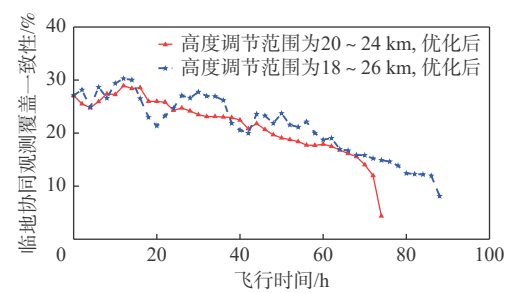
高度调节范围为20～24 km，气球1优化后　高度调节范围为20～24 km，气球2优化后
高度调节范围为18～26 km，气球1优化后　高度调节范围为18～26 km，气球2优化后

(b) 气球飞行高度对比



高度调节范围为20～24 km，气球1优化后　高度调节范围为20～24 km，气球2优化后
高度调节范围为18～26 km，气球1优化后　高度调节范围为18～26 km，气球2优化后

(c) 气球飞行轨迹对比



(d) 协同覆盖一致性对比

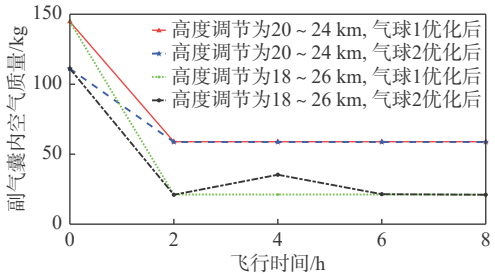
图14　不同高度调节能力的仿真结果对比(9月风场)

Fig. 14　Comparison of simulation results for different height adjustment capabilities (September wind field)

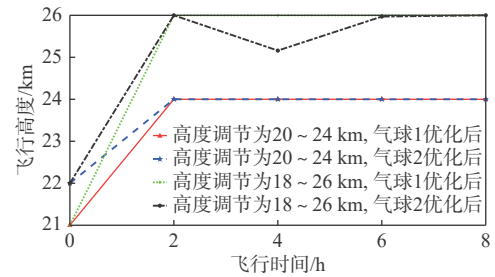
性，但是改善程度有限。

对于不存在准零风层的3月和12月(见图12、图15和表7)，即使扩大高空气球的飞行高度范围，也无法改变气球可用的纬向风场方向，也就难以大幅度调整气球的纬向飞行轨迹，因此，临地协同覆盖一致性的改善主要来源于提升高空气球的飞行高度，扩大每个高空气球的观测覆盖区域半径。在本文仿真条件下，气球飞行高度从24 km提升至26 km，覆盖区域半径可扩大数千米，临地协同观测一致性可得到小幅提升。

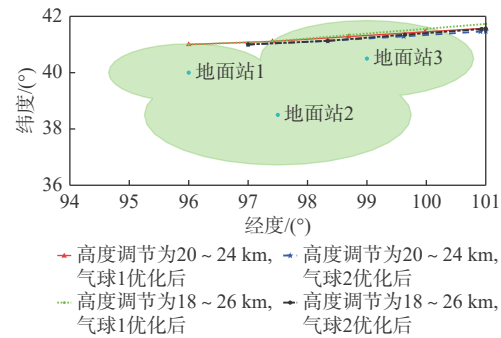
而对于存在准零风层的6月和9月(见图13、图14和表7)，临地协同观测覆盖一致性的改善主



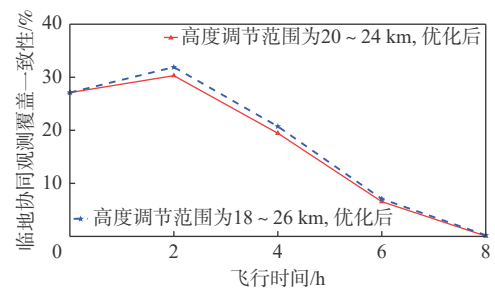
(a) 气球副气囊内空气质量对比



(b) 气球飞行高度对比



(c) 气球飞行轨迹对比



(d) 协同覆盖一致性对比

图 15　不同高度调节能力的仿真结果对比(12 月风场)

Fig. 15　Comparison of simulation results for different height adjustment capabilities (December wind field)

表 7　不同高度调控能力的轨迹优化结果对比

Table 7　Comparison of trajectory optimization results with different height control capabilities

风场	气球高度调控范围为20～24 km的临地协同观测覆盖一致性/%	气球高度调控范围为18～26 km的临地协同观测覆盖一致性/%
3月	1.08	1.17
6月	24.45	24.77
9月	8.56	10.04
12月	0.87	0.91

要是借助准零风层风速较小的特性和准零风层上下纬向风场反向的特性, 围绕准零风层调整气球的飞行高度与飞行轨迹, 从时间和空间 2 个维度达到改善目的。对于本节列举的 2 种气球飞行高度范围, 都包含了准零风层高度, 高度调节能力强的情况可以得到更好的临地协同观测覆盖一致性, 但是二者相比没有发生显著差别。

5　结　论

1) 本文提出的高空气球轨迹优化方法可以大幅改善临地协同观测覆盖一致性, 在本文的仿真条件下, 可使性能提升 67.35%。

2) 高空气球的初始网络布局对临地协同观测覆盖一致性有所影响。根据本文仿真结果, 水平位置不同的布局性能优于水平位置相同的布局。

3) 风场对于临地协同观测覆盖一致性有巨大影响。仿真结果显示, 存在准零风层的风场比无准零风层的风场更有利于提高临地协同观测性能, 最优性能与最劣性能相差可达数十倍。

4) 提高气球的高度调节能力有助于改善临地协同观测覆盖一致性, 但改善程度有限。

根据已有研究, 部分地区在某些年份存在双准零风层, 对于高空气球的轨迹优化可能存在影响; 此外, 实际应用中, 高空气球风机的充气排气效率受球内外压强差、海拔高度等多种因素影响, 而非本文仿真过程中假设的常数, 对于高空气球的轨迹调控也可能存在影响, 后续可开展进一步的研究。

参考文献 (References)

[1] 顾逸东. 气球科学观测 100 年[J]. 现代物理知识, 2020, 32(2): 3-12.
GU Y D. 100 years of scientific balloon observations[J]. Modern Physics, 2020, 32(2): 3-12(in Chinese).

[2] FESEN R, BROWN Y. A method for establishing a long duration, stratospheric platform for astronomical research[J]. Experimental Astronomy, 2015, 39(3): 475-493.

[3] VOSS P B, HOLE L R, HELBLING E F, et al. Continuous in situ soundings in the Arctic boundary layer: a new atmospheric measurement technique using controlled meteorological balloons[J]. Journal of Intelligent & Robotic Systems, 2013, 70(1): 609-617.

[4] KREMIC T, CHENG A F, HIBBITTS K, et al. Stratospheric balloons for planetary science and the balloon observation platform for planetary science (BOPPS) mission summary[C]//Proceedings of the IEEE Aerospace Conference. Piscataway: IEEE Press, 2015: 1-10.

[5] CARRENO-LUENGO H, CAMPS A, QUEROL J, et al. First results of a GNSS-R experiment from a stratospheric balloon over boreal forests[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(5): 2652-2663.

[6] 宋平. 基于卫星、火箭和气球探测资料的我国临近空间大气环境特征分析[D]. 长沙: 国防科技大学, 2020: 4-7.
SONG P. Analysis of atmospheric environment characteristics in China's near space based on satellite, rocket and balloon detection data[D]. Changsha: National University of Defense Technology,

2020: 4-7(in Chinese).

[7] 荆楠, 李创, 潘越, 等. 临近空间高空气球地基探测能力分析[J]. 光子学报, 2016, 45(1): 104002.
JING N, LI C, PAN Y, et al. Analysis of ground-based detection ability of high altitude balloon in near space[J]. Acta Photonica Sinica, 2016, 45(1): 104002(in Chinese).

[8] 王喆, 王振会, 曹晓钟. 毫米波雷达与无线电探空对云垂直结构探测的一致性分析[J]. 气象学报, 2016, 74(5): 815-826.
WANG Z, WANG Z H, CAO X Z. Consistency analysis for cloud vertical structure derived from millimeter cloud radar and radiosonde profiles[J]. Acta Meteorologica Sinica, 2016, 74(5): 815-826 (in Chinese).

[9] 韩彦霞, 王成刚, 严家德, 等. 新型边界层气象探空系统的开发与应用[J]. 气象科技, 2017, 45(5): 804-810.
HAN Y X, WANG C G, YAN J D, et al. Development and application of new boundary layer meteorological sounding system[J]. Meteorological Science and Technology, 2017, 45(5): 804-810(in Chinese).

[10] BELLEMARE M G, CANDIDO S, CASTRO P S, et al. Autonomous navigation of stratospheric balloons using reinforcement learning[J]. Nature, 2020, 588(7836): 77-82.

[11] RAMESH S S, MA J L, LIM K M, et al. Numerical evaluation of station-keeping strategies for stratospheric balloons[J]. Aerospace Science and Technology, 2018, 80: 288-300.

[12] JIANG Y, LV M Y, LI J. Station-keeping control design of double balloon system based on horizontal region constraints[J]. Aerospace Science and Technology, 2020, 100: 105792.

[13] 王益平, 周飞, 徐明. 临近空间浮空器区域驻留控制策略研究[J]. 中国空间科学技术, 2018, 38(1): 63-69.
WANG Y P, ZHOU F, XU M. Research on control strategy of territory-hovering aerostat in near space[J]. Chinese Space Science and Technology, 2018, 38(1): 63-69(in Chinese).

[14] 邓小龙, 李魁, 于春锐, 等. 准零风层新型临近空间浮空器区域驻留性能[J]. 国防科技大学学报, 2019, 41(1): 5-12.
DENG X L, LI K, YU C R, et al. Station-keeping performance of novel near-space aerostat in quasi-zero wind layer[J]. Journal of National University of Defense Technology, 2019, 41(1): 5-12(in Chinese).

[15] 柏方超, 杨希祥, 邓小龙, 等. 基于深度强化学习的风场中浮空器驻留控制[J]. 北京航空航天大学学报, 2024, 50(7): 2354-2366.
BAI F C, YANG X X, DENG X L, et al. Station keeping control for aerostat in wind fields based on deep reinforcement learning[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(7): 2354-2366(in Chinese).

[16] 李魁, 邓小龙, 杨希祥, 等. 基于平流层风场预测的浮空器轨迹控制[J]. 北京航空航天大学学报, 2019, 45(5): 1008-1018.
LI K, DENG X L, YANG X X, et al. Trajectory control of aerostat based on prediction of stratospheric wind field[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(5): 1008-1018 (in Chinese).

[17] 林康, 马云鹏, 郑泽伟, 等. 基于副气囊的平流层浮空器高度控制[J]. 北京航空航天大学学报, 2022, 48(5): 762-770.
LIN K, MA Y P, ZHENG Z W, et al. Height control of stratospheric aerostat based on secondary airbag[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(5): 762-770(in Chinese).

[18] 龙远, 邓小龙, 杨希祥, 等. 极涡风场中平流层浮空器轨迹仿真研究[J]. 计算机仿真, 2021, 38(8): 37-42.
LONG Y, DENG X L, YANG X X, et al. Trajectory simulation of stratosphere aerostats in polar vortex wind field[J]. Computer Simulation, 2021, 38(8): 37-42(in Chinese).

[19] 邓小龙, 丛伟轩, 李魁, 等. 风场综合利用的新型平流层浮空器轨迹设计[J]. 宇航学报, 2019, 40(7): 748-757.
DENG X L, CONG W X, LI K, et al. Trajectory design of a novel stratospheric aerostat based on comprehensive utilization of wind fields[J]. Journal of Astronautics, 2019, 40(7): 748-757(in Chinese).

[20] MUELLER J B, ZHAO Y J, GARRARD W L. Optimal ascent trajectories for stratospheric airships using wind energy[J]. Journal of Guidance, Control, and Dynamics, 2009, 32(4): 1232-1245.

[21] 常晓飞, 白云飞, 符文星, 等. 基于平流层特殊风场的浮空器定点方案研究[J]. 西北工业大学学报, 2014, 32(1): 12-17.
CHANG X F, BAI Y F, FU W X, et al. Research on fixed-point aerostat based on its special stratosphere wind field[J]. Journal of Northwestern Polytechnical University, 2014, 32(1): 12-17(in Chinese).

[22] 陶梦初, 何金海, 刘毅. 平流层准零风层统计特征及准两年周期振荡对其影响分析[J]. 气候与环境研究, 2012, 17(1): 92-102.
TAO M C, HE J H, LIU Y. Analysis of the characteristics of the stratospheric quasi-zero wind layer and the effects of the quasi-biennial oscillation on it[J]. Climatic and Environmental Research, 2012, 17(1): 92-102(in Chinese).

[23] 胡耀月, 王东海, 吴珍珍, 等. 中国区域临近空间准零风层的观测特征[J]. 空间科学学报, 2022, 42(3): 383-395.
HU Y Y, WANG D H, WU Z Z, et al. Analysis of the observational characteristics of the quasi-zero wind layer in the near space over China[J]. Chinese Journal of Space Science, 2022, 42(3): 383-395 (in Chinese).

[24] MIRJALILI S, LEWIS A. The whale optimization algorithm[J]. Advances in Engineering Software, 2016, 95: 51-67.

[25] MAFARJA M M, MIRJALILI S. Hybrid whale optimization algorithm with simulated annealing for feature selection[J]. Neurocomputing, 2017, 260: 302-312.

[26] XIONG G J, ZHANG J, SHI D Y, et al. Parameter extraction of solar photovoltaic models using an improved whale optimization algorithm[J]. Energy Conversion and Management, 2018, 174: 388-405.

[27] 孙琪, 于永进, 王玉彬, 等. 采用改进鲸鱼算法的配电网综合优化[J]. 电力系统及其自动化学报, 2021, 30(5): 22-29.
SUN Q, YU Y J, WANG Y B, et al. Comprehensive optimization of distribution network with improved whale algorithm[J]. Proceedings of the Chinese Society of Universities for Electric Power System and its Automation, 2021, 30(5): 22-29(in Chinese).

[28] 吴坤, 谭劭昌. 基于改进鲸鱼优化算法的无人机航路规划[J]. 航空学报, 2020, 41(S2): 724286.
WU K, TAN S C. Route planning of UAV based on improved whale optimization algorithm[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(S2): 724286(in Chinese).

[29] National Oceanic and Atmospheric Administration. U. S. standard atmosphere[S]. Washington D. C. : NOAA, NASA, USAF, 1976.

Trajectory optimization of high-altitude balloon in nearspace-ground collaborative observation

QU Yi[1, 2], WANG Sheng[1, 2, *], FENG Hui[1], LIU Qiang[1]

(1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The requirement of collaborative observation by high-altitude balloon network and ground-based network is increasingly urgent. And coverage consistency is one of its critical problems affecting observation performance, which faces many challenges such as time-varying network topology, limited balloon control ability, and complicated wind field in near space. To improve the coverage consistency of high-altitude balloon network and ground-based network in cooperative observation, high-altitude balloon with secondary airbag but without propulsion device is studied, its trajectory control approach is analyzed from both the vertical direction and the horizontal direction, and its coverage in earth observation is discussed. Considering the time-variant position of high-altitude balloons, a performance index is designed to evaluate the coverage consistency, and a trajectory optimization algorithm for high-altitude balloon is proposed based on whale optimization algorithm. Simulations are carried out under many inputs. And simulation results show that the proposed algorithm can improve the coverage consistency significantly, especially in summer and autumn wind fields with quasi-zero wind layer.

Keywords: high-altitude balloon; trajectory optimization; collaborative observation; coverage consistency; quasi-zero wind layer

Received: 2023-07-19; **Accepted**: 2023-11-27; **Published Online**: 2023-12-20 10:18

URL: link.cnki.net/urlid/11.2625.V.20231219.1651.002

Foundation items: National Key Research and Development Program of China (2022YFB3901800,2022YFB3901805)

* **Corresponding author**. E-mail: shengwang@aoe.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0488

基于滑动时间窗的机载传感器多任务调度算法

冉华明[1, 2, *]

(1. 复杂航空系统仿真全国重点实验室，成都 610036； 2. 中国电子科技集团公司第十研究所，成都 610036)

摘 要：针对机载传感器任务管理系统的多任务调度效能受多任务请求之间存在的执行时间窗口冲突所影响的问题，根据各任务请求的可执行时间窗等任务请求信息，描述各任务请求的前滑时间窗、后滑时间窗等任务调度约束信息，并在此基础上设计基于滑动时间窗的机载传感器多任务调度算法。筛选出可在调度周期内执行任务请求，并按优先级排序生成待调度任务列表；计算各待调度任务与各已有调度方案的时间交叠关系，形成多个时间片，通过判断待调度任务能否插入扩展后的时间片，不断更新调度方案集合；从调度方案集合中优选出最佳的机载传感器任务调度方案。仿真结果表明：所设计算法调度效能的中位数可达到最优算法调度效能的96.52%以上，且调度效能和计算耗时受调度规模和任务时间精细度的影响较少，具有较强的适应性。

关 键 词：机载传感器；多任务调度；前后滑动时间窗；多任务请求；执行时间窗口冲突；任务调度约束信息

中图分类号：V243.2; TN959.73; TP272

文献标志码：A **文章编号**：1001-5965(2025)09-2968-11

随着先进机载传感器技术(如多功能孔径(multi-function apertures, MFA))的不断发展，使得机载传感器具备多任务能力，同时随着空基平台使用环境的复杂度不断提升，导致机载传感器需响应的任务数量及复杂度也大幅增加，而大多任务存在时间包含、时间交叉等冲突情况，如何设计合理的机载传感器任务调度算法以使传感器自动高效地完成多个任务一直是专家学者们研究的热点[1-2]。传感器任务调度就是研究利用有限的传感器资源，通过合理编排执行各任务的时间窗口来提高传感器的资源利用率[3-4]。在高动态、强对抗的复杂环境中，传感器任务调度算法的优劣决定了机载传感器任务管理系统性能的高下，直接影响任务目标的达成。

传感器任务调度策略总体上可分为静态的模板类策略和动态的自适应任务调度策略两大类[4]。模板类策略[5-6]主要包括固定模板法、多模板法和部分模板法。其基本思想是将需要调度的任务按照一定的顺序和组合方式设计成固定的模板输入给传感器以完成任务的调度，该类策略虽然计算方便、易于实施，但动态适应性差、效率低下。Shih等提出最小空闲模版优先[7]和基于模版的综合周期[8]的离线模板调度算法，可提升传感器的调度效率，但存在动态适应性不足的问题。Gopalakrishnan等提出在线设计模版算法[9]，可提升传感器调度的动态适应性，但存在计算复杂、实现困难的问题。

随着现代机载传感器使用环境的复杂性、对抗性及动态性越来越强烈，使得自适应任务调度策略受到越来越多的关注。自适应调度策略指在满足不同任务优先级的条件下，在传感器能力范围内，实时地平衡各类任务请求所需的时间和资源，为一个调度周期选择一个最佳任务调度序列的调度策略[4]。文献[10-13]基于启发式算法实现对常见任务调度问题的求解。Izquierdo-Fuente等提出一种

收稿日期：2023-07-27；录用日期：2023-09-22；网络出版时间：2023-10-13 16:04
网络出版地址：link.cnki.net/urlid/11.2625.V.20231012.1107.001
*通信作者. E-mail：ranhuaming7245@163.com

引用格式：冉华明. 基于滑动时间窗的机载传感器多任务调度算法[J]. 北京航空航天大学学报，2025，51(9)：2968-2978.
RAN H M. Airborne sensor multi-task scheduling algorithm based on slide time window[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2968-2978 (in Chinese).

基于脉冲交错的雷达传感器动态调度算法[14]。Mir 等提出一种扩展任务交错调度算法, 并分析了多种任务交错形式下的调度问题, 可提高任务调度的时间利用率, 但对传感器的实时处理能力提出了更高的要求[15]。Huizing 等提出任务调度时间窗的概念, 基于固定规则使任务实际执行时间可以在一定时间范围内前后移动, 提升了任务调度的成功率和时间利用率, 但其未考虑更新已调度任务的时间窗, 灵活性较低[16]。文献 [17-18] 提出一种动态时间窗算法, 提升了各任务时间窗的灵活性, 但其计算复杂度较高。文献 [19-20] 针对多雷达调度请求, 提出基于整数规划技术的最优调度方法, 但存在耗时较长的问题。Mcguffin 针对 MFA 调度问题, 设计了一种可为软时间约束或硬时间/软时间混合约束任务请求序列提供近乎最优调度的算法, 但该算法的实现速度太慢, 无法实时运行[21]。相较于模板类策略, 自适应任务调度策略可增加任务调度的灵活性, 但普遍存在计算复杂、实时性低的问题。

本文基于任务调度时间窗、任务持续时间等任务请求信息, 提出了任务前后滑动时间窗调度概念, 设计了一种多任务请求下的高效任务调度算法, 通过在调度过程中不断调整已调度任务的前后滑动时间窗, 生成最佳的任务调度方案, 确定要响应调度的任务请求及各任务的执行次序和执行时间窗。通过多次仿真实验对本文算法进行了验证分析。

1　多传感器任务请求描述

1.1　传感器任务请求定义

在一个任务调度周期内, 机载传感器任务管理系统会收到多个用户发送任务请求, 任务请求主要包含任务编号、最早开始时间、最晚结束时间、任务持续时间、任务可行时间窗、任务优先级 6 个维度的信息, 如图 1 所示。

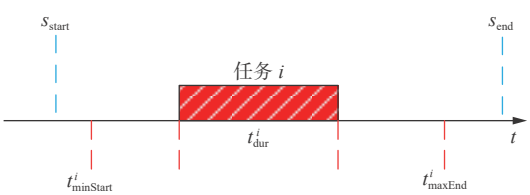


图 1　传感器任务请求信息

Fig. 1　Sensor task request informations

图中, s_{start} 和 s_{end} 分别为任务调度周期的开始时间和结束时间, 表示机载传感器只响应处于任务调度周期内的任务请求; 各任务请求的信息如下:

1) 任务编号 i。进行多任务调度时, 该任务的唯一编号。

2) 最早开始时间 $t^i_{minStart}$。其表示该任务不能早于该时间执行。

3) 最晚结束时间 t^i_{maxEnd}。其表示该任务不能晚于该时间结束。

4) 任务持续时间 t^i_{dur}。其表示完成该任务需要持续的时间。

5) 任务可行时间窗 $[t^i_{minStart}, t^i_{maxEnd}]$。其表示可以执行该任务的时间范围。

6) 任务优先级 t^i_{pri}。其表示该任务的优先级。

1.2　传感器任务调度约束信息

机载传感器任务管理系统收到的处于同一任务调度周期内的各个任务请求的可行时间窗存在相交、覆盖、分离等关系, 根据各任务请求与其前后任务请求在可行时间窗上的关系, 生成各任务请求的调度约束信息, 如图 2 所示。

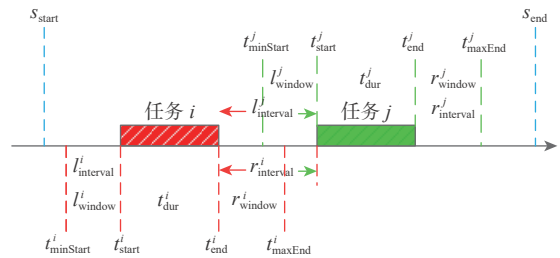


图 2　任务调度约束信息

Fig. 2　Task scheduling constraint informations

各任务请求的调度约束信息主要包含实际开始时间、实际结束时间、实际执行时间窗、前一任务的编号、后一任务的编号、与前一任务的间隔、与后一任务的间隔、前滑时间窗、后滑时间窗 9 个维度的信息:

1) 实际开始时间。执行任务的实际开始时间, 如 t^i_{start}、t^j_{start} 分别表示任务 i 和 j 的实际开始时间。

2) 实际结束时间。执行任务的实际结束时间, 如 t^i_{end}、t^j_{end} 分别表示任务 i 和 j 的实际结束时间。

3) 实际执行时间窗。其表示任务的实际执行时间范围, 如 $[t^i_{start}, t^i_{end}]$、$[t^j_{start}, t^j_{end}]$ 分别表示任务 i 和 j 的实际执行时间窗。

4) 前一任务的编号 l^i_{ID}。其表示执行该任务之前需执行的任务编号, 若该任务为第一个执行的任务, 则该值为 0。

5) 后一任务的编号 r^i_{ID}。其表示执行完该任务之后需执行的任务编号, 若该任务为最后一个执行的任务, 则该值为 0。

6) 与前一任务的间隔 $l^i_{interval}$。其表示该任务实际开始时间与前一任务实际结束时间的差值, 若该任务为第 1 个任务, 则该值为 0:

$$l^i_{\text{interval}}=\begin{cases}t^i_{\text{start}}-t^{l^i_{\text{ID}}}_{\text{end}} & l^i_{\text{ID}}\neq 0\\ 0 & l^i_{\text{ID}}=0\end{cases} \tag{1}$$

7）与后一任务的间隔r^i_{interval}。其表示后一任务实际开始时间与该任务实际结束时间的差值，若该任务为最后一个任务，则该值为0：

$$r^i_{\text{interval}}=\begin{cases}t^{r^i_{\text{ID}}}_{\text{start}}-t^i_{\text{end}} & r^i_{\text{ID}}\neq 0\\ 0 & r^i_{\text{ID}}=0\end{cases} \tag{2}$$

8）前滑时间窗l^i_{window}。其表示在实际开始时间基础上该任务可提前开始的最大时间：

$$l^i_{\text{window}}=\begin{cases}\min\left(t^i_{\text{start}}-t^i_{\min\text{Start}},l^i_{\text{interval}}+l^{l^i_{\text{ID}}}_{\text{window}},t^i_{\text{start}}-s_{\text{start}}\right) & l^i_{\text{ID}}\neq 0\\ \min\left(t^i_{\text{start}}-t^i_{\min\text{Start}},t^i_{\text{start}}-s_{\text{start}}\right) & l^i_{\text{ID}}=0\end{cases} \tag{3}$$

9）后滑时间窗r^i_{window}。其表示在实际结束时间基础上该任务可延迟完成的最大时间：

$$r^i_{\text{window}}=\begin{cases}\min\left(t^i_{\max\text{End}}-t^i_{\text{end}},r^{r^i_{\text{ID}}}_{\text{window}}+r^i_{\text{interval}},s_{\text{end}}-t^i_{\text{end}}\right) & r^i_{\text{ID}}\neq 0\\ \min\left(t^i_{\max\text{End}}-t^i_{\text{end}},s_{\text{end}}-t^i_{\text{end}}\right) & r^i_{\text{ID}}=0\end{cases} \tag{4}$$

2 传感器多任务最佳调度方案生成

机载传感器任务管理系统进行机载传感器多任务调度的逻辑框如图3所示，其主要调度逻辑如下。

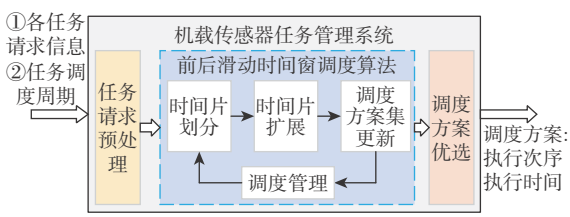


图3 机载传感器多任务调度逻辑框

Fig. 3 Logic block of airborne sensor multi-task scheduling

首先，任务请求预处理模块根据接收到的各任务请求的可行时间窗和任务调度周期，剔除不能在该调度周期内执行的任务，并将可执行任务按任务优先级进行排序，生成待调度任务列表。

然后，多次迭代调用本文的前后滑动时间窗调度算法，依次判断各待调度任务能否插入到已有调度方案集合中的各个方案中，不断调整各方案中各已确定要调度任务的调度约束信息，以最大限度地为待调度任务留下足够的执行时间窗，更新已有调度方案集合。

最后，综合可完成任务数量、任务优先级等因素优选最佳的任务调度方案，确定可调度执行的任务请求及各任务的执行次序和执行时间。

2.1 传感器多任务最佳调度方案生成流程

根据机载传感器多任务调度逻辑框，设计传感器多任务最佳调度方案生成流程如图4所示，具体步骤如下：

1）选择可在任务调度周期内执行的任务请求，并将其加入待调度任务列表，选择条件如下：

$$\begin{cases}t^i_{\max\text{End}}-t^i_{\text{dur}}\geqslant s_{\text{start}}\\ t^i_{\min\text{Start}}+t^i_{\text{dur}}\leqslant s_{\text{end}}\end{cases} \tag{5}$$

2）对待调度任务列表中的任务按照其优先级从高到低进行排序，初始化当前待调度任务的编号$i=1$，初始化调度方案集合为空；

3）调用前后滑动时间窗调度算法，根据任务方案集合各方案信息和当前待调度任务的任务请求信息，计算该任务的调度约束信息，更新方案集合中插入该任务的方案信息；

4）若所有待调度任务都已经计算完成，从方案集合中选择最佳的任务调度方案，结束，否则，待调度任务的编号加1，转到第3步。

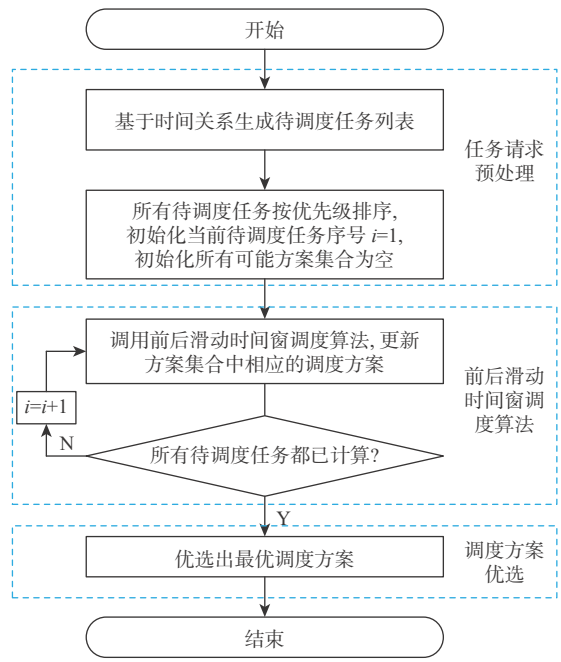


图4 传感器多任务最佳调度方案生成流程

Fig. 4 Flow-process for sensor generating the optimal scheduling scheme for multi-tasks

2.2 前后滑动时间窗调度算法

在确定待调度任务列表后，依次针对每个待调度任务，调用前后滑动时间窗调度算法，最终生成一个包含多个可行调度方案的集合，参考图3，本文算法的主要逻辑如下：

首先，通过调度管理模块依次选择已有方案集合中的各个调度方案；

然后，通过时间片划分模块，根据当前调度方

案中各任务的调度约束信息, 将待调度任务的可行时间窗划分为多个时间片;

接着, 通过时间片扩展模块, 基于各时间片前后已调度任务的约束信息, 将该时间片的时间范围扩展到最大;

最后, 将待调度任务插入时间范围大于待调度任务的持续时间的时间片中, 生成新的调度方案, 并更新调度方案集。

2.2.1　时间片划分

根据待调度任务请求的可行时间窗、当前方案集合中各方案已编排调度任务的任务调度约束信息, 将待调度任务请求的可执行时间窗划分为多个时间片, 如图 5 所示。

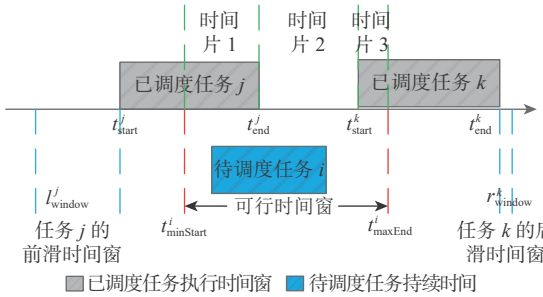


图 5　时间片划分示意图

Fig. 5　Schematic diagram of time slice division

图 5 中, 基于待调度任务 i、已调度任务 j和 k的信息, 将待调度任务 i的可行时间窗分为 3 个时间片, 各时间片的起止时间分别为 $[t^i_{\text{min Start}}, t^j_{\text{end}}]$, $[t^j_{\text{end}}, t^k_{\text{start}}]$, $[t^k_{\text{start}}, t^i_{\text{max End}}]$, 其中, t^j_{end}表示前一个已调度任务 j的实际结束时间, t^k_{start}表示后一个已经调度任务 k的实际开始时间, l^j_{window}表示前一个已调度任务 j的实际执行时间可向前滑动的时间窗, r^k_{window}表示后一个已调度任务 k的实际执行时间可向后滑动的时间窗。

2.2.2　时间片扩展

针对每个时间片, 基于其前一任务的前滑时间窗、后一任务的后滑时间窗等信息, 尽可能地扩大该时间片的时间范围, 如图 6 所示, 时间片 2 扩大后的最大可用时间范围为

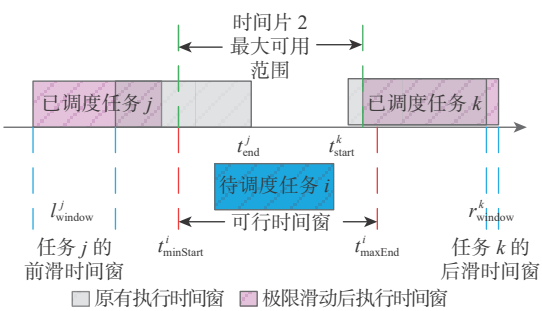


图 6　时间片滑动扩展示意图

Fig. 6　Schematic diagram of time slice sliding extension

$$\left[\max\left(t^i_{\text{min Start}}, t^j_{\text{end}} - l^j_{\text{window}}\right), \min(t^i_{\text{max End}}, t^k_{\text{start}} + r^k_{\text{window}})\right] = \left[t^i_{\text{min Start}}, t^k_{\text{start}} + r^k_{\text{window}}\right] \tag{6}$$

2.2.3　调度方案集更新

当确定各时间片可扩展后的最大可用时间范围后, 根据待调度任务的任务持续时间, 判断待调度任务能否在该时间片中调度, 若能调度, 则确定待调度任务的实际执行时间窗、同时更新其他已调度任务的调度约束信息, 更新任务调度方案, 具体更新方法如下。

1) 待调度任务的实际执行时间窗

以时间片的最大可用范围的起点作为待调度任务的实际开始时间, 以实际开始时间加上待调度任务的任务持续时间作为待调度任务的实际结束时间, 如图 7 所示, 针对时间片 2, 待调度任务 i的实际执行时间窗为: $[t^i_{\text{min Start}}, t^i_{\text{min Start}} + t^i_{\text{dur}}]$。

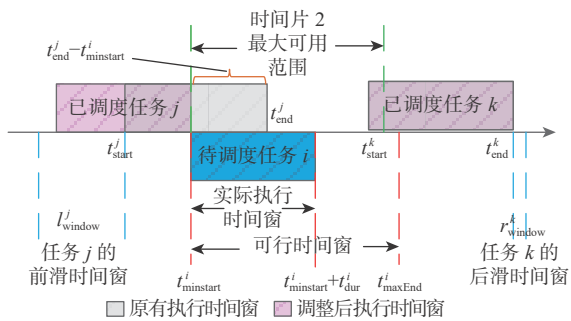


图 7　任务实际执行时间窗更新示意图

Fig. 7　Schematic diagram of updating of actual execution time window for task

2) 方案中已调度任务的实际执行时间窗

针对方案中在时间片之前调度的任务, 按照时间轴从后向前的顺序, 计算各任务的实际结束时间与其后一任务的实际开始时间的差值, 若该差值大于 0, 则将该任务的实际开始时间和实际结束时间都减去该差值, 否则, 该任务的实际执行时间窗不变, 如图 7 所示, 在时间片 2 之前已调度任务 j的实际执行时间窗调整为

$$\left[t^j_{\text{start}} - (t^j_{\text{end}} - t^i_{\text{min Start}}), t^j_{\text{end}} - (t^j_{\text{end}} - t^i_{\text{min Start}})\right] = \left[t^j_{\text{start}} - (t^j_{\text{end}} - t^i_{\text{min Start}}), t^i_{\text{min Start}}\right] \tag{7}$$

针对方案中在时间片之后调度的任务, 按照时间轴从前向后的顺序, 计算各任务的实际开始时间与其前一任务的实际结束时间的差值, 若该差值小于 0, 则将该任务的实际开始时间和实际结束时间都减去该差值, 否则, 该任务的实际执行时间窗不变; 如图 7 所示, 在时间片 2 之后已调度任务 k的实际执行时间窗不变。

2.2.4　前后滑动时间窗调度算法流程

通过 2.2.1 节～2.2.3 节的方法, 可根据当前待

调度任务信息对已有调度方案集合中所有调度方案进行更新，如图8所示，具体流程如下：

1）计算已有调度方案集合中方案的个数M，若$M=0$，则根据待调度任务的任务请求信息，生成第1个调度方案，并将其加入已有调度方案集合中，流程结束；若$M>0$，将所有方案的状态置为未更新，初始化待更新的方案编号$m=1$，转到第2步；

2）若$m\leqslant M$，根据第m个方案中各已调度任务的任务约束信息，将待调度任务的可执行时间窗划分为N个时间片，计算每个时间片的起止时间等信息，初始化待判断的时间片编号$n=1$，转到第3步；若$m>M$，流程结束；

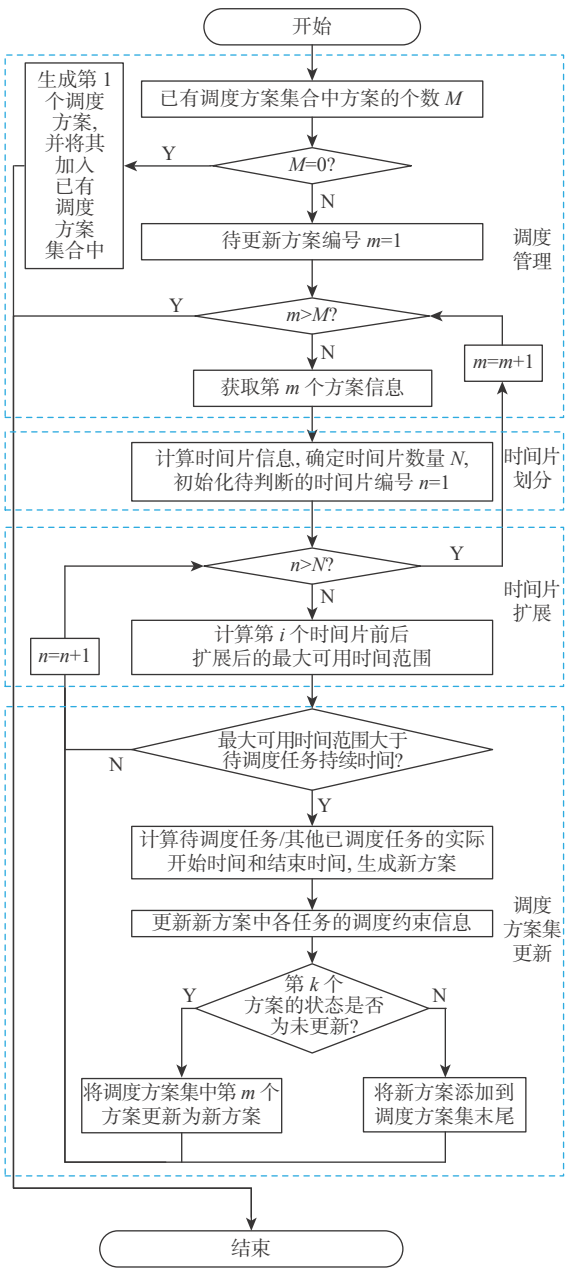


图8 前后滑动时间窗调度算法流程

Fig. 8 Flow-process for forward-back slide time window scheduling algorithm

3）若$n>N$，则增大待更新的方案编号，即$m=m+1$，转到第2步；若$n\leqslant N$，计算第n个时间片前后扩展后的最大可用时间范围，转到第4步；

4）若第n个时间片的最大可用时间范围大于待调度任务的任务持续时间，转到第5步；否则，增加时间片编号，即$n=n+1$，转到第3步；

5）计算待调度任务及方案中已调度任务的实际开始时间和结束时间，生成新的调度方案；

6）根据新方案中各任务的请求信息、实际执行时间窗，更新各任务的调度约束信息；

7）若方案集合中第m个方案的状态为未更新，将第m个方案更新为新生成的方案，并将其状态更新为已更新，增加时间片编号，即$n=n+1$，转到第3步；否则，将新生成的方案添加到方案集合的最后，增加时间片编号，即$n=n+1$，转到第3步。

3 仿真结果与分析

影响机载传感器多任务调度效能的因素主要包括：

1）任务调度周期和单个任务请求信息。详见1.1节中相关定义。

2）需要调度的任务数量。

3）任务时间精细度。即所有时间相关变量的最小颗粒度，如时间精细度为0.5 s时，各任务请求中所有与时间相关的请求信息应能被0.5整除。

针对上述影响要素，设计了3组不同的仿真场景对本文算法进行验证评价，每组仿真场景设置的任务时间精细度不同，如表1所示。

表1 各仿真场景时间精细度参数

Table 1 Time granularity parameters for each simulation scene

场景	任务时间精细度/s
场景1	1.0
场景2	0.5
场景3	0.1

针对每个场景又细分为5个测试集，每个测试集设置不同的任务调度周期和需要调度的任务数量，具体仿真参数如表2所示。

表2 各测试集仿真参数

Table 2 Simulation parameters for each test set

测试集编号	调度开始时间/s	调度结束时间/s	任务数
1	0	20	20
2	0	40	40
3	0	60	60
4	0	80	80
5	0	100	100

针对每个测试集, 分别进行 100 次测试, 每次测试按照表 3 设置的参数范围, 随机生成满足相应仿真场景及测试集仿真参数约束的调度任务请求序列。

表 3　每次测试时随机生成任务请求序列的参数范围

Table 3　Parameters range for randomly generating scheduling task request sequence during each test

持续时间范围/s	滑动时间窗范围/s	任务优先级范围
1～8	0～10	1～10

同时, 在 3 组仿真场景中, 对本文算法和文献 [21] 提出的软时间调度算法进行对比分析。该算法通过将一个软时间约束调度任务(存在滑动时间窗)按一定时间颗粒度离散为多个任务可行时间窗存在差异的硬时间约束调度任务(不存在滑动时间窗), 然后, 通过其提出的最优硬时间调度算法, 可实现对软时间约束任务请求序列的近乎最优调度。若其离散的颗粒度不大于任务时间精细度, 该算法可以得到最优的调度方案。

通过统计本文算法和对比算法的调度方案生成时间, 本文算法调度效能和对比算法调度效能比值等数据, 对本文算法进行分析, 其中, 调度效能定义为调度方案中所有任务的优先级之和。

1) 场景 1

该场景下所有待调度任务的时间精细度为 1 s, 对比算法离散时的时间颗粒度设为 1 s。

在场景 1 的 5 个测试集中, 统计本文算法调度效能与对比算法调度效能比值的分布情况, 如表 4 和图 9 所示, 可以看出, 在不同调度规模下, 本文算法生成的调度方案能达到一个较优的调度效能(中位数在最优调度方案调度效能的 96.52% 以上)。

表 4　本文算法和对比算法在场景 1 中调度效能比值

Table 4　Scheduling efficiency ratio distribution of the proposed algorithm and comparison algorithm in scene 1

测试集编号	调度效能比值/%						离群值个数
	最大值	上四分位数	中位数	下四分位数	最小值	最小离群值	
1	100.00	100.00	99.35	95.39	88.82	86.99	2
2	100.00	98.50	97.02	95.36	91.24		0
3	100.00	98.95	97.73	96.00	91.72	87.26	2
4	100.00	98.19	97.40	95.95	92.65		0
5	99.44	97.68	96.52	95.58	93.56	91.69	2

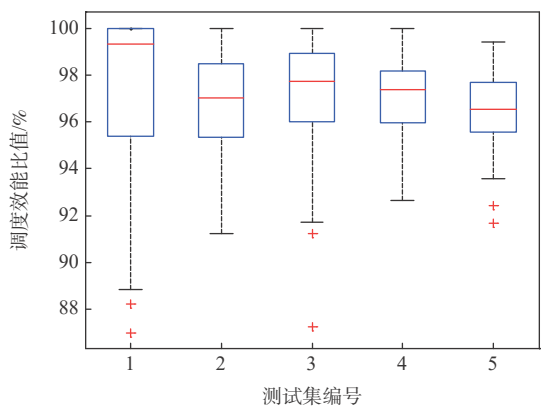


图 9　本文算法在场景 1 中调度效能比值分布

Fig. 9　Scheduling efficiency ratio distribution of the proposed algorithm in scene 1

在场景 1 的 5 个测试集中, 本文算法和对比算法的计算耗时如表 5 和图 10 所示, 可以看出, 随着调度规模的不断增加, 本文算法的计算耗时增加量远小于对比算法的计算耗时增加量, 统计 1 500 次仿真实验中本文算法最大计算耗时为 1.58 s。

在场景 1 中的某次测试实验中, 其调度周期为 0～20 s, 共有 20 个传感器任务请求, 各请求的具体信息如表 6 和图 11 所示。

基于本文算法和对比算法对所有任务请求进行调度, 具体的调度方案如图 12 所示, 图上方数字为调度的任务编号, 调度方案中各任务的调度顺序、实际开始时间、实际结束时间等调度信息如表 7 所示。

从任务调度仿真结果可以看出, 本文算法和对比算法所调度的任务完全相同, 得到的调度效能均为 62.89, 只是部分任务的调度顺序存在差异, 如 2 种算法对任务 11 和任务 15 的调度顺序刚好相反, 但这种顺序差异并未影响调度效能, 本文算法在该次仿真测试中得到了最优调度方案。

表 5　本文算法和对比算法在场景 1 中计算耗时

Table 5　Calculation time for the proposed algorithm and comparison algorithm in scene 1

测试集编号	本文算法计算耗时/s			对比算法[21]计算耗时/s		
	最大值	平均值	最小值	最大值	平均值	最小值
1	0.22	0.05	0.03	0.75	0.16	0.03
2	0.39	0.23	0.12	6.97	1.44	0.34
3	1.14	0.63	0.43	18.05	5.49	2.49
4	1.40	0.90	0.67	33.05	11.99	5.87
5	1.58	1.13	0.90	42.46	20.96	11.12

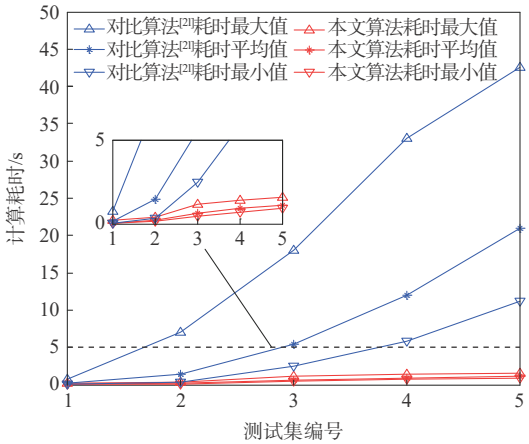


图 10 本文算法和对比算法在场景 1 中计算耗时

Fig. 10 Calculation time for the proposed algorithm and comparison algorithm in scene 1

表 6 场景 1 某次实验的任务请求信息表

Table 6 Task request information tables for one test in scene 1

任务编号	优先级	最早开始时间/s	最晚结束时间/s	持续时间/s
1	2.995	10	20	8
2	9.394	1	10	1
3	8.186	0	5	2
4	6.517	5	18	7
5	6.238	2	10	6
6	5.267	8	13	2
7	9.659	0	7	3
8	7.284	2	10	8
9	4.519	15	20	1
10	4.443	13	18	5
11	5.201	13	19	2
12	1.123	13	17	3
13	5.678	5	8	2
14	3.023	7	12	1
15	9.685	9	19	2
16	2.278	5	15	4
17	5.556	1	4	3
18	6.614	0	10	8
19	9.832	7	19	7
20	5.626	5	13	5

2) 场景 2

该场景下所有待调度任务的时间精细度为 0.5 s, 对比算法离散时的时间颗粒度设为 0.5 s。

在场景 2 的 5 个测试集中, 统计本文算法调度效能与对比算法调度效能比值的分布情况, 如图 13 和表 8 所示, 由图和表可知, 在不同调度规模下, 本文算法生成的调度方案能达到较优的调度效能(中位数在最优调度方案调度效能的 96.67% 以上)。

在场景 2 的 5 个测试集中, 本文算法和对比算

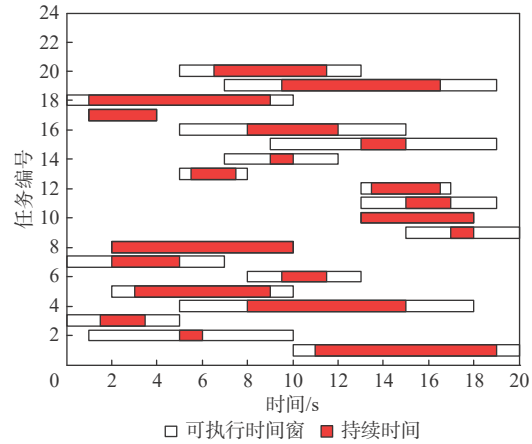


图 11 场景 1 某次实验中各任务的请求信息

Fig. 11 Request informations for each task for one test in scene 1

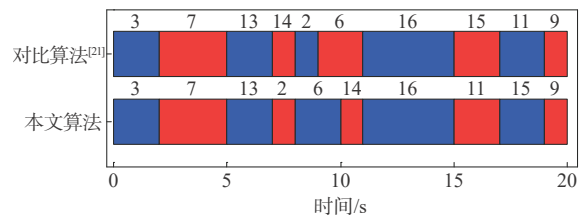


图 12 场景 1 某次实验的任务调度方案

Fig. 12 Task scheduling scheme for one test in scene 1

表 7 场景 1 某次实验的任务调度方案信息

Table 7 Task scheduling scheme information for one test in scene 1

执行顺序	任务编号		实际开始时间/s		实际结束时间/s	
	本文算法	对比算法[21]	本文算法	对比算法[21]	本文算法	对比算法[21]
1	3	3	0	0	2	2
2	7	7	2	2	5	5
3	13	13	5	5	7	7
4	2	14	7	7	8	8
5	6	2	8	8	10	9
6	14	6	10	9	11	11
7	16	16	11	11	15	15
8	11	15	15	15	17	17
9	15	11	17	17	19	19
10	9	9	19	19	20	20

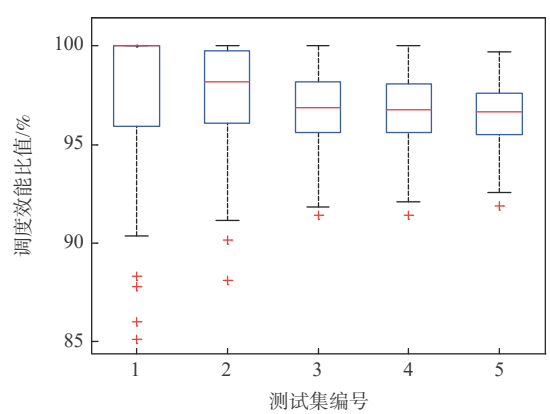


图 13 本文算法在场景 2 中调度效能比值分布

Fig. 13 Scheduling efficiency ratio distribution of the proposed algorithm and comparison algorithm in scene 2

表 8　本文算法与对比算法在场景 2 中调度效能比值分布

Table 8　Scheduling efficiency ratio distribution of the proposed algorithm and comparison algorithm in scene 2

测试集合编号	调度效能比值/%						离群值个数
	最大值	上四分位数	中位数	下四分位数	最小值	最小离群值	
1	100.00	100.00	100.00	95.93	90.37	85.14	4
2	100.00	99.75	98.18	96.06	91.17	88.10	2
3	100.00	98.21	96.87	95.63	91.81	91.40	2
4	100.00	98.06	96.79	95.59	92.11	91.41	2
5	99.69	97.61	96.67	95.53	92.58	91.90	2

法的计算耗时如图 14 和表 9 所示, 可以看出, 随着调度规模的不断增加, 本文算法的计算耗时增加量远小于对比算法的计算耗时增加量。

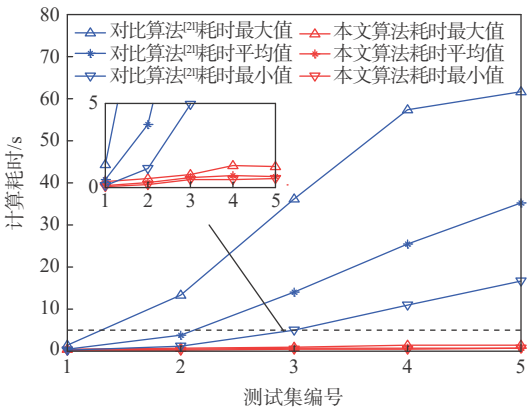


图 14　本文算法和对比算法在场景 2 中计算耗时

Fig. 14　Calculation time for the proposed algorithm and comparison algorithm in scene 2

表 9　本文算法和对比算法在场景 2 中计算耗时

Table 9　Calculation time for the algorithm in this paper and the comparison algorithm in scene 2

测试集编号	本文算法计算耗时/s			对比算法[21]计算耗时/s		
	最大值	平均值	最小值	最大值	平均值	最小值
1	0.33	0.08	0.03	1.33	0.46	0.07
2	0.51	0.27	0.16	13.22	3.70	1.11
3	0.76	0.55	0.42	35.98	13.88	4.94
4	1.28	0.71	0.47	99.10	25.83	10.81
5	1.24	0.63	0.50	61.44	35.02	16.52

3) 场景 3

该场景下所有待调度任务的时间精细度为 0.1 s, 对比算法离散时的时间颗粒度设为 0.1 s。

在场景 3 的 5 个测试集中, 统计本文算法调度效能与对比算法调度效能比值的分布情况, 如图 15 和表 10 所示, 可以看出, 在不同调度规模下, 本文算法生成的调度方案能达到一个较优的调度效能(中位数在最优调度方案调度效能的 96.61% 以上)。

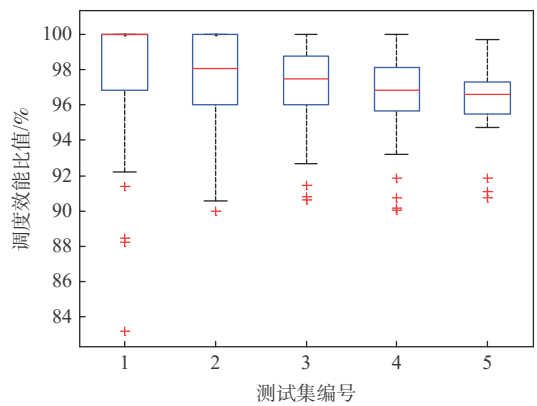


图 15　本文算法与对比算法在场景 3 中调度效能比值分布

Fig. 15　Statistical graph of scheduling efficiency ratio distribution of the proposed algorithm and comparison algorithm in scene 3

在场景 3 的 5 个测试集中, 本文算法和对比算法的计算耗时如图 16 和表 11 所示, 可以看出, 随着调度规模的不断增加, 本文算法的计算耗时增加量远小于对比算法的计算耗时增加量。

在场景 3 中的某次测试实验中, 其调度周期为 0～20 s, 共有 20 个传感器任务请求, 各请求的具体信息如表 12 和图 17 所示。

表 10　本文算法与对比算法在场景 3 中调度效能比值分布

Table 10　Scheduling efficiency ratio distribution of the proposed algorithm and comparison algorithm and comparison algorithm in scene 3

测试集编号	调度效能比值/%						离群值个数
	最大值	上四分位数	中位数	下四分位数	最小值	最小离群值	
1	100.00	100.00	100.00	96.80	92.20	83.22	4
2	100.00	99.98	98.07	95.99	90.58	90.00	1
3	100.00	98.78	97.45	96.03	92.65	90.63	3
4	100.00	98.11	96.81	95.65	93.19	90.04	4
5	99.69	97.33	96.61	95.49	94.71	90.77	9

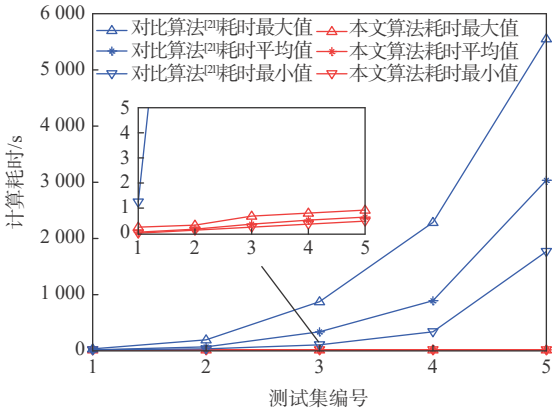


图 16 本文算法和对比算法在场景 3 中计算耗时

Fig. 16 Calculation time for the proposed algorithm and comparison algorithm in scene 3

表 11 本文算法和对比算法在场景 3 中计算耗时

Table 11 Calculation time for the proposed algorithm and comparison algorithm in scene 3

测试集编号	本文算法计算耗时/s			对比算法[21]计算耗时/s		
	最大值	平均值	最小值	最大值	平均值	最小值
1	0.26	0.06	0.02	32.76	6.89	1.24
2	0.35	0.18	0.12	194.34	67.16	21.13
3	0.69	0.37	0.25	869.38	330.02	105.92
4	0.81	0.53	0.37	2 285.13	880.06	333.94
5	0.94	0.65	0.52	5 550.48	3 027.08	1 764.79

表 12 场景 3 某次实验的任务请求信息

Table 12 Task request information table for one test in scene 3

任务编号	优先级	最早开始时间/s	最晚结束时间/s	持续时间/s
1	1.86	12.20	17.20	4.10
2	7.37	3.80	5.20	1.00
3	6.90	9.60	17.40	3.70
4	5.78	5.30	11.10	5.70
5	8.37	8.90	18.50	4.30
6	1.03	5.70	9.00	2.00
7	4.10	8.70	18.70	2.50
8	8.03	1.90	14.70	5.90
9	7.58	6.70	17.90	7.40
10	1.23	0.20	12.20	6.80
11	7.06	13.00	16.40	1.40
12	3.12	1.80	13.90	4.70
13	5.02	4.60	15.50	1.20
14	8.58	7.20	11.50	4.20
15	2.99	11.50	19.00	1.60
16	4.61	3.00	12.10	2.80
17	8.45	11.80	19.70	3.70
18	7.42	2.90	17.40	6.00
19	8.30	0.70	10.20	1.10
20	9.56	5.70	18.00	6.40

基于本文算法和对比算法对所有任务请求进行调度，具体的调度方案如图 18 所示，调度方案中各任务的调度顺序、实际开始时间、实际结束时间等调度信息如表 13 所示。

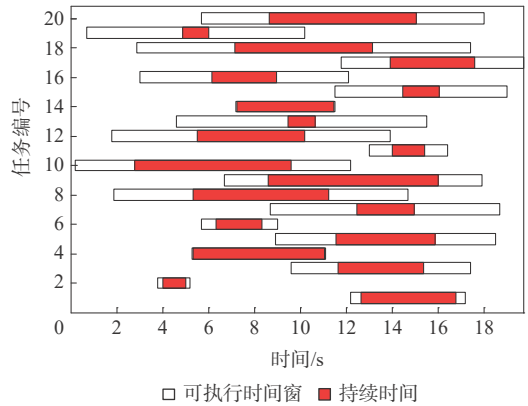


图 17 场景 3 某次实验中各任务的请求信息

Fig. 17 Request information for each task for one test in scene 3

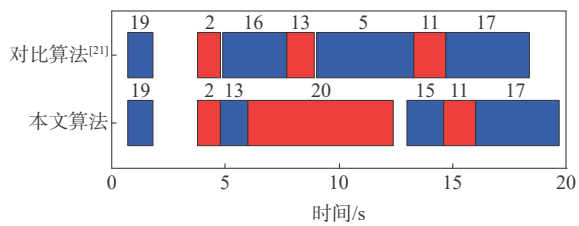


图 18 场景 3 某次实验的任务调度方案

Fig. 18 Task scheduling scheme for one test in scene 3

表 13 场景 3 某次实验的任务调度方案信息

Table 13 Task scheduling scheme information for one test in scene 3

执行顺序	任务编号		实际开始时间/s		实际结束时间/s	
	本文算法	对比算法[21]	本文算法	对比算法[21]	本文算法	对比算法[21]
1	19	19	0.70	0.70	1.80	1.80
2	2	2	3.80	3.80	4.80	4.80
3	16	13	4.90	4.80	7.70	6.00
4	13	20	7.70	6.00	8.90	12.40
5	5	15	9.00	13.00	13.30	14.60
6	11	11	13.30	14.60	14.70	16.00
7	17	17	14.70	16.00	18.40	19.70

从任务调度仿真结果可以看出，本文算法和对比算法均调度了 7 个任务请求，但本文算法得到的调度效能为 48.76，为最优方案调度效能的 99.13%(对比算法得到的调度效能为 49.19)。2 种算法调度结果的主要差距在于本文算法调度了任务 20 和任务 15，而对比算法调度了任务 16 和任务 5，导致本文算法取得的调度效能略逊于对比算法。

图 19 为本文算法和对比算法的平均计算耗时随任务时间精细度变化的情况，图 20 为本文算法

调度效能与对比算法调度效能比值的中位数随任务时间精细度变化的情况, 从中可以看出, 本文算法的计算耗时和调度效能不受任务时间精细度的影响, 而对比算法的计算耗时随着任务时间精细度的减少而急剧增加, 以此可以说明, 本文算法具有较强的适应性。

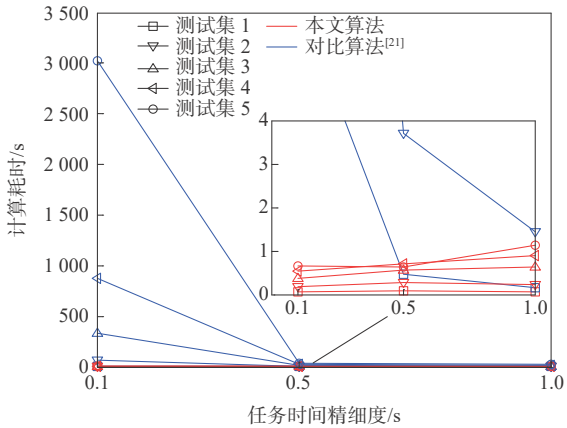


图 19　2 种算法计算耗时随任务时间精细度变化

Fig. 19　Computational time of two algorithms versus granularity of task time

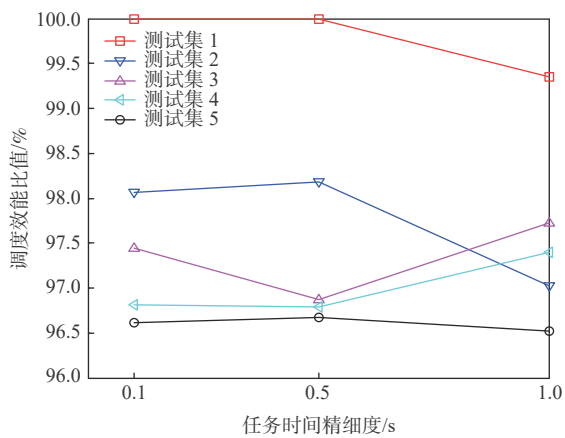


图 20　2 种算法调度效能比值随任务时间精细度变化

Fig. 20　Scheduling efficiency ratio of two algorithms versus granularity of the task time

4　结　论

1) 本文提出的多传感器任务请求下的前后滑动时间窗调度算法, 可生成调度效能较优的调度方案, 通过 1 500 次仿真实验, 统计本文算法调度效能与最优算法调度效能的比值, 其中位数在 96.52% 以上。

2) 本文算法的计算耗时较短, 且不受任务时间精细度的影响, 受任务调度周期长短、调度任务数量的影响也较小, 统计 1 500 次仿真实验中本文算法最大计算耗时为 1.58 s, 远小于对比的最优算法。

3) 在 1 500 次仿真实验中, 也偶尔出现调度效能不佳的方案, 如与最优方案调度效能比值为 83.22%, 说明本文算法还有一定的提升空间, 这也是作者后续需要分析研究的工作。

随着机载传感器技术的不断发展, 单一传感器同时调度多个不同类型的任务将成为可能, 如何实现这种更为复杂条件下的机载传感器多任务高效调度, 是作者后续研究的重要方向。

参考文献（References）

[1] HUIZING A G, BLOEMEN A A F. An efficient scheduling algorithm for a multifunction radar[C]//Proceedings of the International Symposium on Phased Array Systems and Technology. Piscataway: IEEE Press, 1996: 359-364.

[2] 黄佳沁, 钮伟, 郑世友. 基于孔径动态重构的多功能雷达任务调度方法[J]. 指挥控制与仿真, 2021, 43(6): 128-134.
HUANG J Q, NIU W, ZHENG S Y. Multifunction radar task scheduling method based on dynamic aperture reconstruction[J]. Command Control & Simulation, 2021, 43(6): 128-134(in Chinese).

[3] 胡子军, 翟海涛. 基于任务驱动的机载相控阵雷达 TAS 调度算法[J]. 系统工程与电子技术, 2017, 39(3): 536-541.
HU Z J, ZHAI H T. Task-driven TAS scheduling algorithm for airborne phased array radar[J]. Systems Engineering and Electronics, 2017, 39(3): 536-541(in Chinese).

[4] 杨善超, 田康生, 刘仁争, 等. 基于价值优化的相控阵雷达任务调度算法[J]. 电子与信息学报, 2020, 42(2): 465-471.
YANG S C, TIAN K S, LIU R Z, et al. Scheduling algorithm based on value optimization for phased array radar[J]. Journal of Electronics & Information Technology, 2020, 42(2): 465-471(in Chinese).

[5] ORMAN A J, POTTS C N, SHAHANI A K, et al. Scheduling for a multifunction phased array radar system[J]. European Journal of Operational Research, 1996, 90(1): 13-25.

[6] MIRANDA S L C, BAKER C J, WOODBRIDGE K, et al. Phased array radar resource management: a comparison of scheduling algorithms[C]//Proceedings of the IEEE Radar Conference. Piscataway: IEEE Press, 2004: 79-84.

[7] SHIH C S, GANTI P, GOPALAKRISHNAN S, et al. Synthesizing task periods for dwells in multi-function phased array radars[C]//Proceedings of the IEEE Radar Conference. Piscataway: IEEE Press, 2004: 145-150.

[8] SHIH C S, GOPALAKRISHNAN S, GANTI P, et al. Scheduling real-time dwells using tasks with synthetic periods[C]//Proceedings of the 24th IEEE Real-Time Systems Symposium. Piscataway: IEEE Press, 2003: 210-219.

[9] GOPALAKRISHNAN S, CACCAMO M, SHIH C S, et al. Finite-horizon scheduling of radar dwells with online template construction[J]. Real-time Systems, 2006, 33(1): 47-75.

[10] MOSLEHI G, MAHNAM M. A Pareto approach to multi-objective flexible job-shop scheduling problem using particle swarm optimization and local search[J]. International Journal of Production Economics, 2011, 129(1): 14-22.

[11] 张浩为, 谢军伟, 师俊朋, 等. 饱和时序下防空相控阵雷达动态优先级调度算法[J]. 北京航空航天大学学报, 2016, 42(12): 2722-2729.

ZHANG H W, XIE J W, SHI J P, et al. Dynamic priority scheduling algorithm for air defense phased array radar in overload situations[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(12): 2722-2729(in Chinese).

[12] RAHMATI S H A, ZANDIEH M. A new biogeography-based optimization (BBO) algorithm for the flexible job shop scheduling problem[J]. The International Journal of Advanced Manufacturing Technology, 2012, 58(9): 1115-1129.

[13] 陶孙杰, 宋竹. 一种测控数传一体化站网资源调度算法[J]. 电讯技术, 2018, 58(7): 760-767.
TAO S J, SONG Z. A ground-station resource scheduling algorithm based on integration of TT&C and data transmission[J]. Telecommunication Engineering, 2018, 58(7): 760-767(in Chinese).

[14] IZQUIERDO-FUENTE A, CASAR-CORREDERA J R. Approach to multifunction radar scheduling simulation[C]//Proceedings of the IEEE National Telesystems Conference. Piscataway: IEEE Press, 1994: 67-70.

[15] MIR H S, WILKINSON J D. Task scheduling algorithm for an air and missile defense radar[C]//Proceedings of the IEEE Radar Conference. Piscataway: IEEE Press, 2008: 1629-1634.

[16] HUIZING A G, BLOEMEN A A F. A high-level multifunction radar simulation for studying the performance of multisensor data fusion systems[C]//Proceedings of the SPIE Conference on Signal Processing, Sensor Fusion, and Target Recognition VII. Bellingham: SPIE, 1998: 129-138.

[17] YING C L, WANG Y, HE J X. Study on time window of track tasks in multifunction phased array radar tasks scheduling[C]//Proceedings of IET International Radar Conference. London: IET, 2009: 1-4.

[18] JANG D S, CHOI H L, ROH J E. A time-window-based task scheduling approach for multi-function phased array radars[C]//Proceedings of the International Conference on Control, Automation and Systems. Piscataway: IEEE Press, 2011: 1250-1255.

[19] SPASOJEVIC Z, DEDEO S, JENSEN R. Dwell scheduling algorithms for phased array antenna[J]. IEEE Transactions on Aerospace and Electronic Systems, 2013, 49(1): 42-54.

[20] MIR H S, GUITOUNI A. Variable dwell time task scheduling for multifunction radar[J]. IEEE Transactions on Automation Science and Engineering, 2014, 11(2): 463-472.

[21] MCGUFFIN B. Priority scheduling for multi-function apertures with hard- and soft-time constraints[C]//Proceedings of the IEEE Aerospace Conference. Piscataway: IEEE Press, 2021: 1-11.

Airborne sensor multi-task scheduling algorithm based on slide time window

RAN Huaming[1, 2, *]

(1. National Key Laboratory of Complex Aviation System Simulation，Chengdu 610036，China；
2. The 10th Research Institute of China Electronics Technology Group Corporation，Chengdu 610036，China)

Abstract: Aiming at the problem that the multi-task scheduling efficiency of the airborne sensor task management system was reduced due to the execution time window conflicts among multiple task requests, the task scheduling constraint information, include the forward slide time window and the back slide time window of each task, is depicted according to the task request information, such as the executable time window of each task request, and a multi-task scheduling algorithm for the airborne sensor is designed based on the slide time window. Firstly, task requests that can be executed within the scheduling period are selected, and a list of tasks to be scheduled is generated based on priority sorting. Then, multiple time slices are formed by calculating the time overlap relationship between each task to be scheduled and each existing scheduling scheme, and the scheduling scheme set is continuously updated by determining whether the task to be scheduled can be inserted into the extended time slice. Finally, the optimal airborne sensor task scheduling scheme is selected from the scheduling scheme set. The simulation results show that the median scheduling efficiency of the designed algorithm can reach over 96.52% of the optimal algorithm's scheduling efficiency, and the scheduling efficiency and computational time of the designed algorithm are rarely affected by the scheduling scale and granularity of the task time, the designed algorithm has strong adaptability.

Keywords: airborne sensor; multi-task scheduling; forward-back slide time window; multiple task requests; execution time window conflicts; task scheduling constraint information

Received：2023-07-27；**Accepted**：2023-09-22；**Published Online**：2023-10-13 16：04
URL：link.cnki.net/urlid/11.2625.V.20231012.1107.001

* **Corresponding author**. E-mail：ranhuaming7245@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0482

基于贝叶斯优化LSTM神经网络的飞机货舱火源定位

张伟[1]，常本强[2]，杨旭[2]，熊枭[3,*]

(1. 中国民航大学 交通科学与工程学院，天津 300300；2. 中国民航大学 安全科学与工程学院，天津 300300；
3. 深圳市城市公共安全技术研究院有限公司，深圳 518024)

摘 要：民航飞机货舱火灾多发于高空低温低压的环境，对飞机安全飞行造成了巨大的威胁。为快速定位货舱火灾源点和采取针对性区域灭火措施，提出一种基于贝叶斯优化（BO）的长短期记忆（LSTM）神经网络火源定位模型（BO-LSTM）。该模型使用LSTM神经网络充分挖掘多种火灾特征时序数据（烟雾、温度、CO浓度）与火灾源点的时空关联特性，同时采用贝叶斯算法搜寻LSTM神经网络的最优超参数组合以提高模型的鲁棒性和准确性。通过仿真研究验证BO-LSTM模型，使用Pyrosim火灾模拟软件以1∶1比例建立了8个常用民航飞机货舱模型，并在每个模型中随机选取10个火源点进行低温低压环境的火灾仿真。实验结果表明：所建模型预测火源中心点距离实际火源中心点的直线距离误差皆小于0.1 m，预测火源二维坐标皆处于真实火源的范围内。贝叶斯优化过的LSTM神经网络极大提高了传统LSTM神经网络的性能，适用于低温低压状态下的飞机货舱火源定位。

关 键 词：飞机货舱；低温低压；火源定位；贝叶斯优化；LSTM神经网络；Pyrosim软件
中图分类号：V244.1+2
文献标志码：A **文章编号**：1001-5965(2025)09-2979-08

随着科学技术和经济水平的迅速提高，中国民航业得到了持续稳定的发展，越来越多的人们将飞机作为出行交通工具的第一选择。飞机货舱火灾具有易燃物多、火势蔓延迅速、救援困难等特点，对机组和乘客的人身安全及财产构成巨大威胁[1]。而以哈龙为代表的飞机常用灭火剂残留物的污染问题也一直是航空业待攻克的难题。如果能在飞机货舱火灾早期阶段快速准确定位到火源的位置并定点投放灭火剂，就可以在使用较少灭火剂的情况下及时灭火并减少环境污染。所以研究火灾早期火源定位技术不仅可以保障民航客机的安全运行，而且对建设绿色民航也具有现实意义。

为解决火灾早期的火源定位问题，达到在火灾初期及时灭火的目的，诸多学者研究了火灾发生与传播的规律。在火灾发生的初始阶段，火源进行不完全燃烧反应，生成较多的CO和稀少的CO_2，同时生成极少含量的烟雾，周围温度变化较小。随着热解作用的持续进行，释放出大量的CO和CO_2气体，可见烟雾开始出现，温度持续上升并产生明火。当火势迅速发展时，产生大量的烟、热和辐射，随后形成烟气羽流。

根据火灾发生规律，Andrew等使用烟气传感器采集室内建筑早期火灾气体数据，提取数据特征建立分类模型用于早期火灾探测和预测[2]，但其监测方式过于单一，火灾漏报的概率较大。王洋等提出一种基于红外阵列传感器的火源快速定位方法[3]，

收稿日期：2023-07-21；录用日期：2023-10-13；网络出版时间：2023-11-08 12:41
网络出版地址：link.cnki.net/urlid/11.2625.V.20231107.1342.006
基金项目：中央高校基本科研业务费专项资金 (3122020048)
*通信作者. E-mail：xiongxiao@szsti.org

引用格式：张伟，常本强，杨旭，等. 基于贝叶斯优化LSTM神经网络的飞机货舱火源定位[J]. 北京航空航天大学学报，2025，51(9)：2979-2986. ZHANG W，CHANG B Q，YANG X，et al. Fire source localization for long short-term memory networks based on Bayesian optimization[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2979-2986 (in Chinese).

利用火源的红外图像，计算火源重心坐标实现精准快速定位，但红外阵列传感器对环境条件比较敏感，温度变化和周围热源等因素影响传感器的性能。刘帅等提出基于单摄像头视频图像处理的火灾探测及定位技术，并在全尺寸试验隧道内开展了验证[4]，但视频图像处理需要较长的时间来分析和处理图像，这可能会延迟火灾的检测和报警。胡淋翔等提出一种基于反向传播神经网络的单室火源位置识别模型，识别模型对火源位置的识别精度满足要求，并对新场景具有适应性[5]。

国内外火源定位研究虽已取得了很大进展，但大多都是针对公路隧道和楼宇建筑等场所，定位方法存在误报率高、对所处环境条件要求高、检测方式单一及定位精度低等缺点，不适用于复杂特殊的民机货舱火灾探测。本文提出基于贝叶斯优化（Bayesian optimization, BO）的长短期记忆（long short-term memory, LSTM）神经网络火源定位预测模型（简称BO-LSTM模型），选取火灾早期变化较大的CO浓度、温度、烟雾浓度作为火灾特征参数，建立了随时间变化的火灾特征参数与源点的时空关系，并应用BO解决LSTM神经网络模型中超参数优化问题，以提高火灾模型的整体性能。限于实验条件，本文采用Pyrosim软件进行火灾仿真建模对BO-LSTM模型进行评估验证。Pyrosim软件是火灾学研究的专用工具，相关研究表明，Pyrosim软件能够有效地模拟火灾发生、发展过程和各项特征参数的空间时间变化。其模拟仿真的输出数据可用于诸如灾害评估、火源定位、人员疏散等相关研究[6-8]。

研究结果表明，本文模型相较于传统的LSTM模型提高了对火源坐标的识别精度，火源位置的直线距离误差小于0.1 m，实现了高精度预测，更加适用于民航客机货舱的早期火灾溯源。

1 火源定位模型

1.1 LSTM神经网络理论基础

LSTM神经网络是一种特殊的循环神经网络（recurrent neural networks, RNN），不易受梯度消失和梯度爆炸的影响。相较于普通RNN，LSTM神经网络可以有效筛选关键特征，能够通过“门”的控制保留早期信息，在更长的序列中具有更好的表现，具备更好的特征存储能力和更准确的历史信息查询优势[9]。

LSTM神经网络单个细胞层结构如图1所示，主要包含遗忘门、输入门和输出门。其中，$\boldsymbol{X}_t$为当前状态的数据输入，$\boldsymbol{h}_{t-1}$和$\boldsymbol{h}_t$分别为上一个时间点和本时刻的隐藏层输出值，$\boldsymbol{C}_{t-1}$和$\boldsymbol{C}_t$分别为上一时刻和本时刻的细胞状态，$\boldsymbol{f}_t$为遗忘门输出值，$\boldsymbol{i}_t$和$\tilde{\boldsymbol{C}}_t$分别为激活函数σ与激活函数tanh输出值，$\boldsymbol{o}_t$为当前时刻输出门的输出结果。

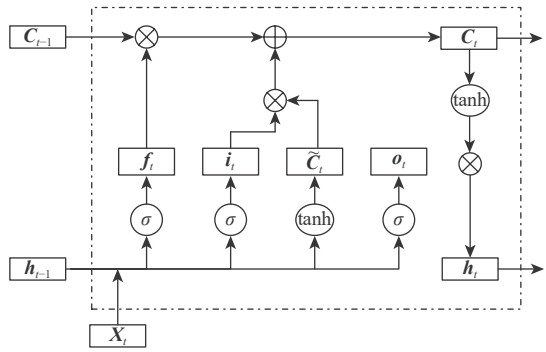


图1 LSTM神经网络层结构

Fig. 1 LSTM neural networks layer structure

1.2 贝叶斯优化算法理论基础

在LSTM神经网络中，超参数的取值会直接影响到模型的性能与预测效果。而贝叶斯算法作为一种全局优化算法，可以找到当前模型的最优超参数组合。本文以寻找最大值为例，可表示为

$$\boldsymbol{x}_{\text{opt}} = \arg\max_{\boldsymbol{x}\in X} f(\boldsymbol{x}) \tag{1}$$

式中：$\boldsymbol{x}$为LSTM神经网络的超参数组合；X为超参数集合；$f(\boldsymbol{x})$为超参数组合$\boldsymbol{x}$到模型泛化性能的映射；$\boldsymbol{x}_{\text{opt}}$为最优超参数组合。

BO算法主要由2个核心部件构成：①概率代理模型，用来拟合黑箱目标函数；②采集函数，根据已知数据递推估算最优值。该算法通过不断更新基于高斯过程的概率代理模型，快速评估目标函数以更新优化函数的后验概率，再经采集函数评估得到最优的超参数组合[10-11]。

1.2.1 概率代理模型

概率代理模型是基于高斯过程（Gaussian processes, GP），而高斯过程中任意随机变量的线性组合都会服从多维高斯分布[12]，而对变量$\boldsymbol{x}=[x_1,x_2,\cdots,x_t]$，有

$$f(\boldsymbol{x}) \sim \text{GP}(\mu(\boldsymbol{x}), k(\boldsymbol{x},\boldsymbol{x})) \tag{2}$$

式中：$\mu(\boldsymbol{x})$为均值函数；$k(\boldsymbol{x},\boldsymbol{x})$为协方差函数。

假设模型中已知的数据$\boldsymbol{D}_{1:t}=\left[x_i,f_i\right]$，$i=1,2,\cdots,t$，简记为$\boldsymbol{D}$。其中，$f_i$为在参数$x_i$训练下得到的测试集精度，记$f_i=f(x_i)$，下一个搜寻值为$x_{t+1}$，协方差矩阵为

$$\boldsymbol{K}=\begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_t) \\ \vdots & & \vdots \\ k(x_t,x_1) & \cdots & k(x_t,x_t) \end{bmatrix} \tag{3}$$

且当均值为0时，已知的数据$\boldsymbol{D}$和新观测点(x_{t+1},f_{t+1})的联合高斯分布满足：

$$\begin{bmatrix} f_{1:t} \\ f_{t+1} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \boldsymbol{K} & \boldsymbol{k}_{\text{o}} \\ \boldsymbol{k}_{\text{o}}^{\text{T}} & k(x_{t+1},x_{t+1}) \end{bmatrix}\right) \tag{4}$$

其中：$\boldsymbol{k}_{\text{o}}$表示为

$$\boldsymbol{k}_{\mathrm{o}}=\begin{bmatrix}k(x_{t+1},x_1)\\k(x_{t+1},x_2)\\\vdots\\k(x_{t+1},x_t)\end{bmatrix} \tag{5}$$

而f_{t+1}的后验概率为

$$P\left(f_{t+1}|\boldsymbol{D},x_{t+1}\right)=N\left(\mu_t(x_{t+1}),\sigma_t^2(x_{t+1})\right) \tag{6}$$

式中: $\mu_t(x_{t+1})$为f_{t+1}的期望; $\sigma_t^2(x_{t+1})$为f_{t+1}的方差, 计算式为

$$\mu_t(x_{t+1})=\boldsymbol{k}_{\mathrm{o}}^{\mathrm{T}}\boldsymbol{K}^{-1}f_{1:t} \tag{7}$$

$$\sigma_t^2(x_{t+1})=k(x_{t+1},x_{t+1})-\boldsymbol{k}_{\mathrm{o}}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{k}_{\mathrm{o}} \tag{8}$$

由式 (7) 和式 (8) 可得出x_{t+1}在任意处的正态分布, 然后, 使用采集函数去搜寻下一个最优的采样点。

1.2.2　采集函数

在 BO 过程中, 为解决局部最优解的问题, 通过采集函数指示目标函数值大于当前最优值的概率, 作为下一个采样点取值的参考依据。本文采用蔡文杰[13]的高斯过程上置信界(Gaussian process-upper confidence bound, GP-UCB)策略作为采集函数, 其计算式为

$$\mathrm{UCB}(\boldsymbol{x})=\mu(\boldsymbol{x})+\lambda\sigma(\boldsymbol{x}) \tag{9}$$

式中: λ为均值与方差之间的调节参数, 用以控制开发与探索之间的平衡。

1.3　火源定位模型流程建立

针对飞机货舱火灾溯源预测, 本文设定 LSTM 网络层数为 2 层。为避免在训练中出现过拟合现象, 在 2 层网络中添加 dropout 层。应用烟雾浓度、温度、CO 浓度 3 种火灾特征参数预测火源二维坐标值, 故设定输入维度为 3, 输出维度为 2。将第 2 个 Dropout 层的最后一个隐藏状态作为整个 LSTM 网络结构的输出单元, 输出单元接全连接层, 得到最终的定位坐标预测值, 网络结构如图 2 所示。

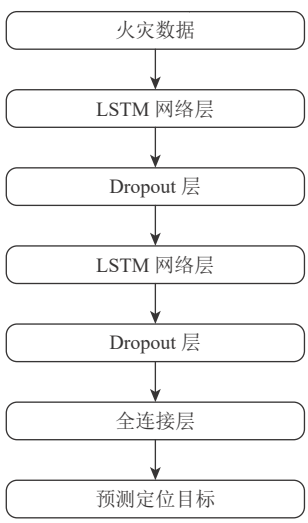


图 2　LSTM 网络结构

Fig. 2　LSTM networks structure

对于本文模型而言, 其神经元数量(units)、Dropout 层的弃权系数(dropout)、批处理大小(batch_size)、激活函数(activation)、优化器(optimizer)、训练步数(Epochs)均为需要优化的超参数, 这些参数作为 LSTM 网络模型的域空间, 其取值范围如表 1 所示。采用一种基于树状结构帕森估计器(tree of Parzen estimators, TPE)的优化算法作为搜索算法, 并将 LSTM 网络模型训练结果中的 R^2 值作为目标函数。

BO-LSTM 模型的建立流程如图 3 所示。

表 1　LSTM 网络域空间的超参数

Table 1　Hyperparametric domain space of LSTM networks

超参数	域空间
神经元数量1(units1)	(50,100)
神经元数量2(units2)	(1,50)
弃权系数1(dropout1)	(0,0.5)
弃权系数2(dropout2)	(0,0.5)
批处理大小(batch_size)	(0,100)
训练步数(Epochs)	(300,600)
优化器(optimizer)	['Adam','RMSProp','AdaGrad']
激活函数(activation)	['sigmoid','relu']

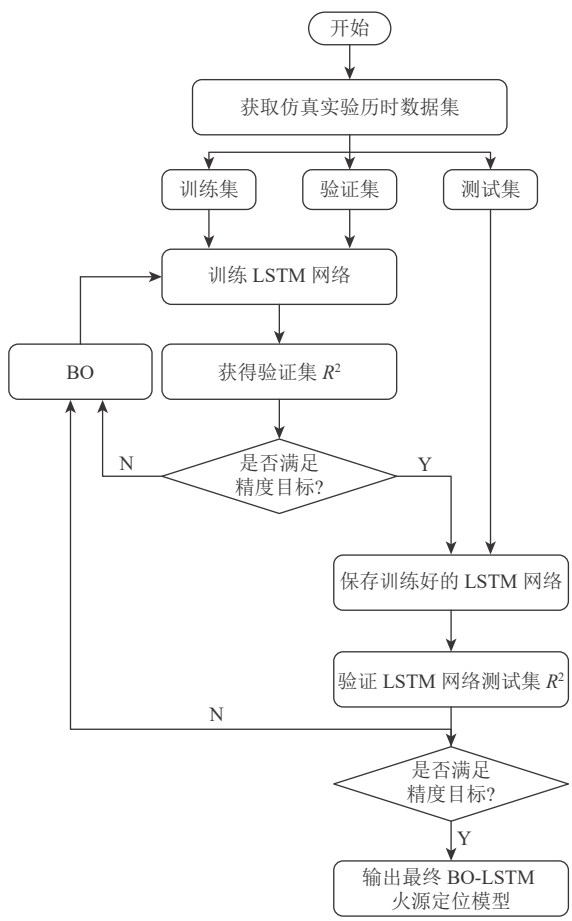


图 3　BO-LSTM 模型建立流程

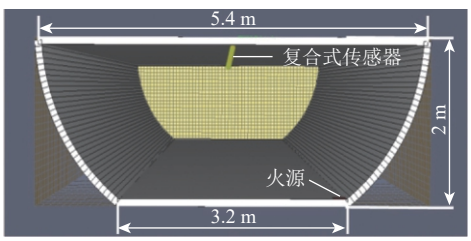
Fig. 3　BO-LSTM model building process

2 实验方法

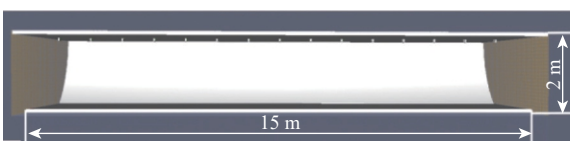
本文使用 PyroSim 火灾模拟软件对常用机型货舱进行建模和火灾仿真实验。为使所得数据具有代表性，选取波音与空客 2 个公司旗下的 4 架主流飞机（A330-300、A330-200F、B747-8I、B777-F）作为仿真对象，其中，A330-300 与 B747-8I 为客机，A330-200F 与 B777-F 为民用货机。4 个机型的前后下货舱共进行 8 组实验，每组实验在货舱底板随机选取 10 个火源点，共获得 80 组实验数据，用以火源定位预测模型的建立。使用均方误差（mean square error, MSE）和相关指数 R^2 作为定位预测模型的评价指标。

2.1 货舱模型数据

本文建立货舱模型的数据皆来源于中国国际货运航空有限公司于 2015 年发布的《中国国际货运航空有限公司机型手册》[14]。以 A330-300 机型前下货舱为例，该机型货舱天花板净宽度为 5.4 m、底板为 3.2 m、高为 2 m、货舱总长为 15 m，建立的 1∶1 仿真模型如图 4 所示。根据飞机货舱的实际尺寸设定网格，其单元尺寸设为 0.1 m×0.1 m×0.1 m。



(a) 正视图



(b) 侧视图

图 4 A330-300 前下货舱模型

Fig. 4 A330-300 front-lower cargo compartment model

2.2 模型参数设置

2.2.1 运行环境设置

飞机在高空巡航阶段时处于低温低压的环境状况，为还原真实场景，结合《中国国际货运航空有限公司机型手册》中的货舱信息[14]，把运行环境温度设为 5 ℃。而飞机从海平面到巡航高度（约为 10 000 m）的货舱压力变化范围在 101 kPa 到 70 kPa 之间[15]，故将运行环境气压设为 80 kPa。考虑到低温低压对火焰燃烧特性的影响，做了如下设计：①由于低温低压对燃烧反应的发生有一定的抑制作用，在模拟中将热解时间做了延迟处理。②低温会增大点火半径，低压会使火焰点火极限上升，造成点火性能变差。这是因为气动力不足、液体黏度力增大及表面张力增大、雾化质量差等因素不利于火焰的传播，从而导致点火性能比常温情况下差。综合考虑将火源大小设定为 0.2 m×0.2 m，燃烧方式为聚氨酯燃烧反应。

2.2.2 探测器与火源的设置

由于目前飞机火灾探测存在误报问题，本文选用感烟-感温-感 CO 三合一复合式传感器作为飞机货舱火灾探测设备，并设置当烟雾、温度和 CO 传感器同时接收到火灾信息后才进行火源定位，避免单一传感器因故障原因而发生火灾误报警的情况。根据文献 [15] 中的研究结论，复合式传感器应置于天花板下方 0.12 m 处，并沿着货舱天花板中线建立横坐标 x 轴，垂直中线的方向建立纵坐标 y 轴，传感器以 1 m 为间隔依次排布，具体探测点位数量视货舱长度而定。火源位置随机设置，每组实验设置 10 个火源点。

2.3 实验效果展示

在设置好所有实验参数之后，运行火灾动力学模拟（fire dynamics simulator, FDS）的计算，并调用 Smokeview 来查看运行过程，图 5 为以烟雾场形式展现的烟雾运动轨迹图。由图 5 可知，当货舱内发生火情时，烟雾扩散到飞机货舱天花板的用时为 4 s，之后烟雾在火羽流的驱动下持续撞击天花板并向其四周扩散，排布好的复合式传感器则相继接收具有时间序列的数据。

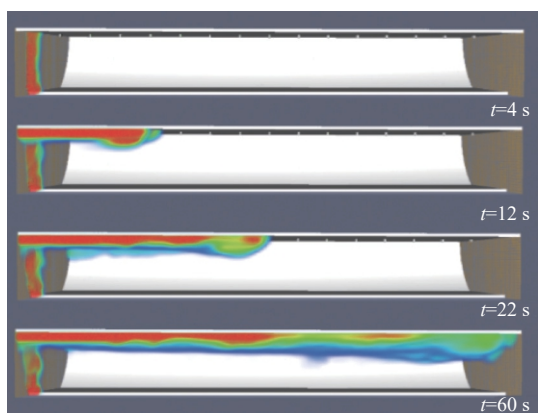


图 5 A330-300 前下货舱烟雾场

Fig. 5 A330-300 front-lower cargo compartment smoke field

2.4 数据集生成与划分

本文建立了 4 种机型的前和后下货舱模型，针对 8 个货舱进行 8 组实验，每组实验进行 10 次，所获数据用以定位预测模型数据集的建立。设定运行时间为 60 s[16]，每次实验 3 s 可以输出 50 组数据（温度、烟雾及 CO 浓度值），共 1 000 组。由于复合传感器距离货舱底板有一定距离，复合式传感器大约在火灾发生 4 s 后接收到烟雾信号。考虑到复合

传感器接收到火灾信息的延迟及反应时间, 除去前 4 s 的数据, 数据样本大小由 1 000 组变为 934 组。为建立高效实用的 LSTM 网络定位预测模型, 再将每个样本数据切割成 884 个数据样本, 每个数据样本大小为 50 组(温度、烟雾及 CO 浓度值), 在火灾发生的过程中收集 3 s 内的数据就可以准确定位火源位置。

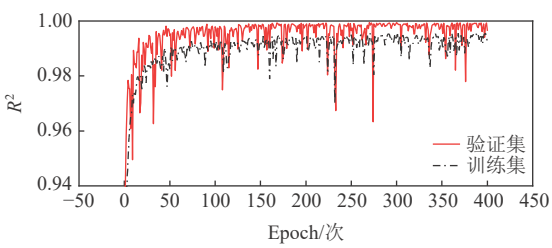
将这 884 个数据样本所在组的火源坐标值赋予标签, 10 次实验共生成 8 840 个数据样本作为模型的输入部分, 10 次实验的火源坐标构成标签值作为模型的输出部分。然后, 将所有标注的样本随机分为训练数据集、验证数据集和测试数据集, 其比例分别为 60%、25% 和 15%。训练时使用训练数据集对模型进行训练, 使用验证数据集对拟合模型进行评估, 训练后使用测试数据集对拟合模型的质量进行评估。

3　网络训练与参数分析

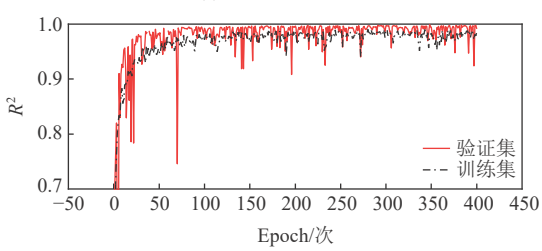
3.1　单组实验预测结果分析

以 A330-300 的前下货舱为例进行分析, 仿真模型中随机设置 10 个火源点, 将仿真获得的实验数据导入到 BO-LSTM 模型中进行训练和验证, 并选用 R^2 和 MSE 作为评价指标来评价该模型, 评价结果如图 6～图 8 所示。图 6 为 10 个火源点的横坐标值与纵坐标值的 R^2 历时演进分析, 图 7 为 BO-LSTM 模型横纵坐标的整体 R^2 历时演进分析, 图 8 为 BO-LSTM 模型横纵坐标的整体损失历时演进分析。

由图 6 可知, 训练初始时火源横坐标的 R^2 值高于纵坐标的 R^2 值, 且收敛速度快于纵坐标。这是



(a) 横坐标 R^2 历时分析



(b) 纵坐标 R^2 历时分析

图 6　横纵坐标 R^2 历时分析

Fig. 6　R^2 duration analysis of horizontal and vertical coordinate

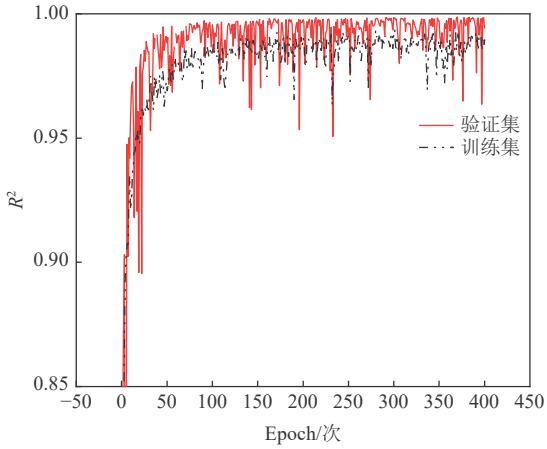


图 7　本文模型总体 R^2 历时分析

Fig. 7　Overall R^2 duration analysis of the proposed model

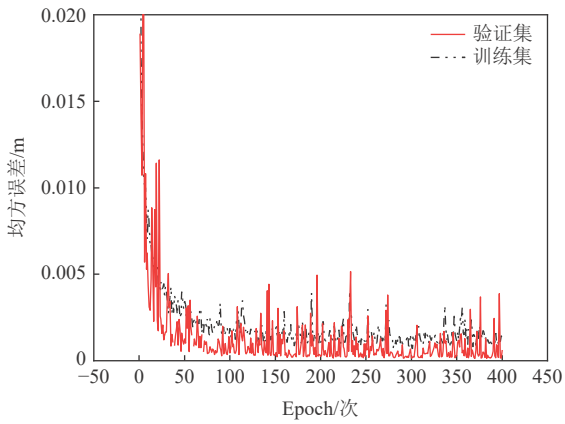


图 8　本文模型总体损失历时分析

Fig. 8　Overall loss duration analysis of the proposed model

由实验中传感器的排布方向导致的, 因传感器是沿 x 轴方向进行排布的, 所以沿 x 轴方向的识别速度要快于沿 y 轴方向的识别速度。总体来看, R^2 值最终都收敛接近于 100%, 可见模型对于横坐标和纵坐标的识别都得到一个有效的训练。

图 7 和图 8 为从 BO-LSTM 模型总体 R^2 值与 MSE 的角度进行模型性能分析, 由图 7 和图 8 可知, 验证集的性能在任何阶段都高于训练集的性能, 这表明本文模型训练是成功有效的。由图 7 可知, 验证集的损失值和训练集的损值相差较小, 这意味着本文模型在训练效果较好的情况下没有出现过拟合的问题。图中, 验证集和训练集的相关曲线均出现了部分波动, 但这是机器学习的一个常见后果[17]。综合来看, BO-LSTM 模型能够在较少的训练步数内趋于收敛, 且 R^2 值都收敛于 100%, MSE 值都收敛于 0, 这说明此次模型训练是成功的, 本文模型可以应用测试集进行后续的火源点预测。

本文选用总体数据集的 15% 作为测试集, 得到的模型预测效果如图 9 所示。可以看到, 除少数点没有落在标签处, 绝大部分点完美预测了真实坐标

值。经BO-LSTM模型预测，测试集中x标签的数据R^2为0.996 2，y标签的R^2为0.976 5，模型总体的R^2为0.987 8，模型总体的损失为0.001 3。计算1 326组预测值与真实值两点间的直线距离，总体平均绝对误差为0.069 6 m。

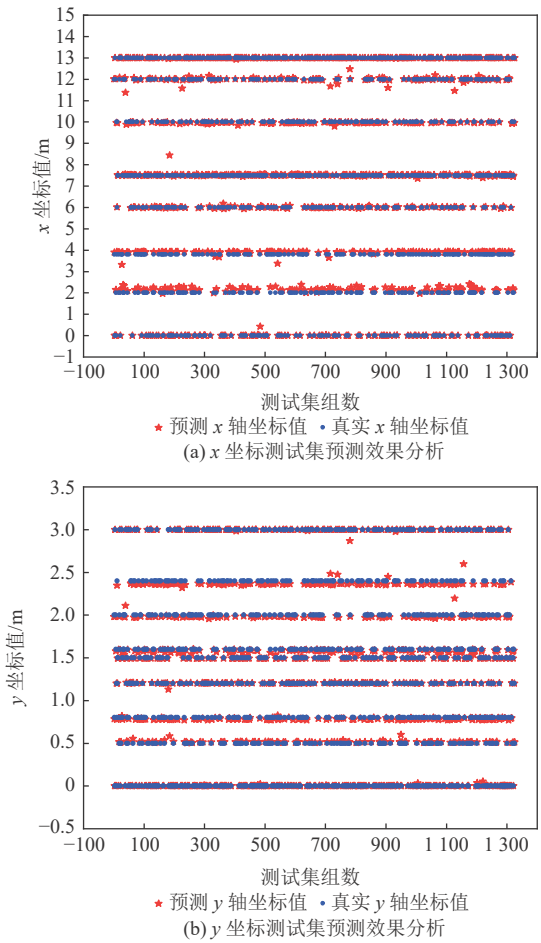


图9 测试集预测效果分析

Fig. 9 Prediction effect analysis of the test set

为进一步证实BO-LSTM模型的准确性与优越性，使用同样的数据将未经贝叶斯算法优化的LSTM模型与BO-LSTM模型对比，结果如图10所示。以评价指标R^2为例进行结果分析，传统LSTM模型火源横坐标x的测试集R^2为0.941 7，纵坐标y的R^2为0.866 1，模型总体测试集R^2为0.911 1；BO-LSTM模型火源横坐标x的测试集R^2为0.996 2，纵坐标y的R^2为0.976 5，模型总体测试集R^2为0.987 8。由图可知，BO-LSTM模型的横纵坐标的R^2值均高于传统LSTM模型的R^2值，尤其明显提高了纵坐标y的R^2值，克服了由于传感器排布方式带来的对火源纵坐标识别精度不够的问题，使模型更加适用于飞机货舱的火灾溯源。

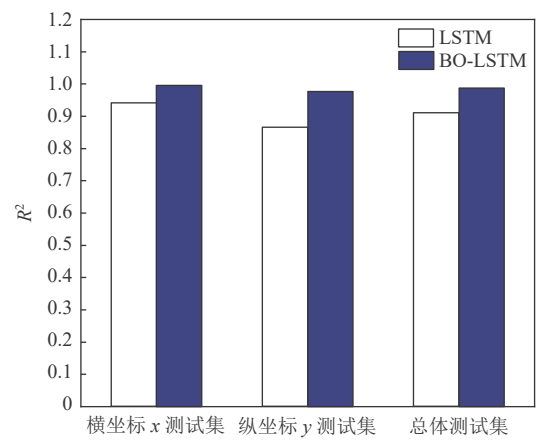


图10 LSTM与BO-LSTM对比分析

Fig. 10 Comparative analysis of LSTM and BO-LSTM

3.2 整体实验预测结果分析

将8组实验数据全部导入到BO-LSTM模型中，得到的模型效果如图11所示。计算并统计不同机型每个舱室的平均绝对误差（预测点与实际火源中心点的直线距离），最小定位误差为0.045 m，最大绝对误差为0.099 4 m，所有舱室的火源预测点距离实际火源中心点的距离皆小于0.1 m。而实验火源大小设置为0.2 m×0.2 m，实际火源中心点到火源边线的距离为0.1 m，说明本文BO-LSTM模型能够满足精准定位飞机货舱火源的需求。

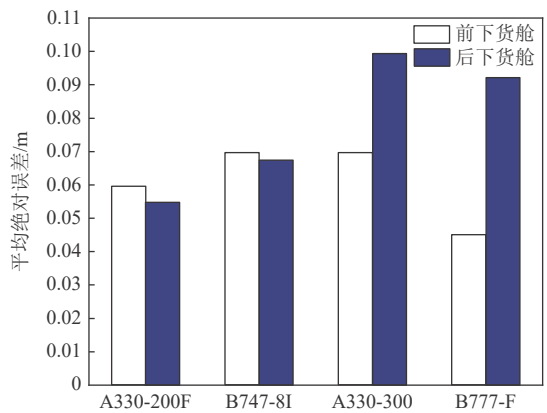


图11 不同机型前后舱的平均绝对误差分析

Fig. 11 Mean absolute error analysis of front and lower cabins of different aircraft models

4 结 论

本文BO-LSTM模型用以快速定位飞机货舱火灾源点，经实验验证表明：

1）与传统LSTM神经网络预测模型相比，本文BO-LSTM火源定位模型的拟合与预测性能更优。

2）本文BO-LSTM火源定位模型性能稳定，预测定位精度高，满足精准定位飞机货舱火源的需求。

3）本文BO-LSTM火源定位模型适用于现有市面常用机型的不同舱室，可应用性强。

参考文献（References）

[1] 刘鹏. 新时期民航飞机火灾的特点和扑救战术原则探讨[J]. 湖南

安全与防灾, 2019(1): 48-51.
LIU P. Discussion on the characteristics of civil aviation aircraft fire and the principle of fighting tactics in the new period[J]. Hunan Safety and Disaster Prevention, 2019(1): 48-51(in Chinese).

[2] ANDREW A M, ZAKARIA A, SAAD S M, et al. Multi-stage feature selection based intelligent classifier for classification of incipient stage fire in building[J]. Sensors, 2016, 16(1): 31.

[3] 王洋, 李向国, 梅志千, 等. 基于红外阵列传感器的火源定位方法[J]. 系统仿真技术, 2022, 18(2): 73-76.
WANG Y, LI X G, MEI Z Q, et al. Fire source positioning method based on infrared array sensor[J]. System Simulation Technology, 2022, 18(2): 73-76(in Chinese).

[4] 刘帅, 吴梦军, 左远正. 基于图像处理的大跨隧道火灾定位技术试验研究[J]. 公路交通技术, 2018, 34(S1): 41-44.
LIU S, WU M J, ZUO Y Z. Experimental study on fire locating technology based on image processing in large-span tunnel[J]. Technology of Highway and Transport, 2018, 34(S1): 41-44(in Chinese).

[5] 胡淋翔, 安子樱, 李伟, 等. 基于神经网络的单室火源位置识别模型[J]. 建筑热能通风空调, 2022, 41(8): 53-57.
HU L X, AN Z Y, LI W, et al. Compartment fire location identification model based on neural network[J]. Building Energy & Environment, 2022, 41(8): 53-57(in Chinese).

[6] WAHLQVIST J, VAN HEES P. Implementation and validation of an environmental feedback pool fire model based on oxygen depletion and radiative feedback in FDS[J]. Fire Safety Journal, 2016, 85: 35-49.

[7] 杨建忠, 邵资焱, 陈希远. 通风对飞机货舱烟雾探测影响研究[J]. 中国安全科学学报, 2019, 29(2): 69-75.
YANG J Z, SHAO Z Y, CHEN X Y. Research on influence of ventilation on detection of smoke in an aircraft cargo compartment[J]. China Safety Science Journal, 2019, 29(2): 69-75(in Chinese).

[8] BROHEZ S, CARAVITA I. Fire induced pressure in airthigh houses: experiments and FDS validation[J]. Fire Safety Journal, 2020, 114: 103008.

[9] 涂芝润. 基于 LSTM 的 VOD 精炼炉终点预报模型研究[D]. 西安: 西安理工大学, 2021.
TU Z R. Research on endpoint prediction model of VOD refining furnace based on LSTM[D]. Xi'an: Xi'an University of Technology, 2021(in Chinese).

[10] FRAZIER P I. A tutorial on Bayesian optimization[EB/OL]. (2018-07-08)[2023-07-21]. https://arxiv.org/abs/1807.02811.

[11] SHIELDS B J, STEVENS J, LI J, et al. Bayesian reaction optimization as a tool for chemical synthesis[J]. Nature, 2021, 590(7844): 89-96.

[12] VARGAS HERNÁNDEZ R A. Bayesian optimization for calibrating and selecting hybrid-density functional models[J]. The Journal of Physical Chemistry A, 2020, 124(20): 4053-4061.

[13] 蔡文杰. 基于机器学习的广告推荐技术研究[D]. 南京: 南京邮电大学, 2021.
CAI W J. Research on advertising recommendation technology based on machine learning[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2021(in Chinese).

[14] 中国国际货运航空有限公司. 中国国际货运航空有限公司机型手册[EB/OL]. (2015-04-13)[2023-07-21]. https://max.book118.com/html/2017/0701/119584966.shtm.
Air China Cargo Co. , Ltd. Aircraft type manual of air China cargo Co., Ltd[EB/OL]. (2015-04-13)[2023-07-21]. https://max.book118.com/html/2017/0701/119584966.shtm(in Chinese).

[15] 熊枭, 陈达, 韩宙, 等. 基于传感器十字交叉式排布的飞机货舱早期火源定位研究[J]. 消防科学与技术, 2022, 41(12): 1632-1636.
XIONG X, CHEN D, HAN Z, et al. Early fire source location in aircraft cargo hold based on crossover arrangement of sensors[J]. Fire Science and Technology, 2022, 41(12): 1632-1636(in Chinese).

[16] Federal Government. Code of federal regulations: CFR Part 25(Title 14)-2022[S]. Washington, D. C, : Government Printing Office, 2022: 858.

[17] QU C, HOUSTON P L, YU Q, et al. Machine learning classification can significantly reduce the cost of calculating the Hamiltonian matrix in CI calculations[J]. Journal of Chemical Physics, 2023, 159(7): 071101.

Fire source localization for long short-term memory networks based on Bayesian optimization

ZHANG Wei[1], CHANG Benqiang[2], YANG Xu[2], XIONG Xiao[3,*]

(1. School of Transportation Science and Engineering, Civil Aviation University of China, Tianjin 300300, China;
2. School of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China;
3. Shenzhen Urban Public Safety Technology Research Institute Co., Ltd., Shenzhen 518024, China)

Abstract: Civil aviation aircraft cargo compartment fires often occur in high-altitude, low-temperature and low-pressure environments, posing a significant threat to aircraft safety. A Bayesian optimization (BO) based long short-term memory (LSTM) neural network model (BO-LSTM) is proposed to quickly locate fire source and take specific means for fire suppression. In BO-LSTM model, time series data are fully analyzed to investigate the spatiotemporal correlation between fire characteristics (smoke, temperature, CO concentration) and fire source by using LSTM network. Meanwhile, the Bayesian algorithm is employed to search for the optimal LSTM network hyperparameter combination to improve model robustness and accuracy. In this paper, simulation study was used to validate the proposed BO-LSTM method: 8 widely used aircraft cargo compartment models were built by Pyrosim® fire simulation software at a scale of 1∶1, and 10 fire source positions were randomly selected in each cargo compartment model to simulate fire data at low-temperature and low-pressure conditions. The experimental results demonstrate that the distance error between predicted fire location and true fire source was less than 0.1 meters, and the predicted two-dimensional fire source localized within the true fire source range. The performance of LSTM neural network has been improved dramatically by the Bayesian optimization method and make it a suitable tool for aircraft cargo compartment fire source localization at the environment of low-temperature and low-pressure.

Keywords: aircraft cargo hold; low temperature and low pressure; fire source localization; Bayesian optimization; LSTM neural network; Pyrosim

Received: 2023-07-21; **Accepted**: 2023-10-13; **Published Online**: 2023-11-08 12:41
URL: link.cnki.net/urlid/11.2625.V.20231107.1342.006
Foundation item: The Fundamental Research Funds for the Central Universities (3122020048)
* **Corresponding author**. E-mail: xiongxiao@szsti.org

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0466

基于滑模控制和舵轴驱动的多间隙折叠舵颤振抑制方法

任浩源*，程涛，张程，蔡毅鹏，刘飞，张炜群
(中国运载火箭技术研究院，北京 100076)

摘　　要：高速飞行器折叠舵因连接结构存在间隙，在飞行过程中易出现非线性动力学行为，气动弹性主动抑制技术是重要的解决方案。由于全动折叠舵只有内舵的舵轴转动位移是输入参数，基于滑模控制网络，提出通过控制舵轴转动来实现多间隙折叠舵颤振主动抑制的新方法。基于三维折叠舵模型和活塞理论非定常气动力计算方法，采用线性弹簧模型描述内/外舵折叠机构连接刚度，推导了四自由度折叠舵动力学模型，阐述了舵轴驱动对线性折叠舵系统颤振抑制的可行性和控制机理；进一步将折叠机构连接刚度模型改为间隙非线性弹簧，分析了多间隙非线性折叠舵气动弹性响应行为和滑模控制的颤振抑制效果。研究结果表明：对于所分析的折叠舵系统，沉浮间隙和扑动间隙有利于提高颤振速度，而俯仰间隙将降低颤振速度；采用滑模控制方法能够有效提高线性和间隙非线性折叠舵面的颤振速度。所提方法和结果为新型折叠舵系统的设计和气动弹性控制提供了一种工程化解决思路。

关　键　词：折叠舵；间隙；气动弹性；滑模控制；颤振抑制；非线性振动
中图分类号：V214; V414.1
文献标志码：A　　**文章编号**：1001-5965(2025)09-2987-14

高速飞行器的外形尺寸包络应满足不同类型储运/发射平台的舱内空间约束[1]，这在一定程度上制约了飞行器翼/舵的展长和面积，进一步引发飞行器升阻比下降、空气舵控制能力不足等设计难题，而翼/舵折叠方案成为解决该问题的关键技术。可依据设计形式将折叠机构划分为梳式折叠机构[2-3]和铰接/锁销折叠机构[4-5]：梳式折叠机构多用于舰载机折叠翼，设计特点是通过多组连杆机构连接内翼和外翼，由液压支杆或电机齿轮进行折展驱动，采用耳片和锁销配合完成锁紧；铰接/锁销折叠机构多用于小型飞行器的空气舵面，通过转轴连接内舵和外舵，利用扭簧或火工品驱动外舵面展开，再通过锁紧销进行展开状态的锁定。

在各类折叠机构设计时，为保证机构正常作动，都需要在部件之间预置间隙，但是空气舵折叠机构与舰载机翼折叠机构的设计思路仍然存在明显差异：舰载机折叠机构包括复杂的液压或电动设备进行折展驱动和多点锁紧，整个展开和锁定过程可以在起飞前完成，可控性更高；而空气舵折叠机构的设计则更加追求方案便捷、结构紧凑和轻质小巧，尤其新型飞行器在高速飞行历程中展开翼舵，为在毫秒量级的时间内完成高速动态展开，并在到位瞬间实现可靠锁紧[6]，就要求折叠机构部件的间隙值相对更大，多间隙效应也更加明显。折叠机构部件之间的预置间隙表现为外舵面绕折叠轴沉浮、俯仰和扑动方向的非线性连接刚度特性，含间隙非线性连接刚度的力学系统在气流激励下易表现出准周期、极限环和混沌等复杂非线性颤振特性，严重影响飞行稳定性和飞行品质，有必要深入研究多间隙影响下的折叠舵非线性气动弹性行为，提出有

收稿日期：2023-07-14；录用日期：2023-11-20；网络出版时间：2023-12-15 12:54
网络出版地址：link.cnki.net/urlid/11.2625.V.20231214.1659.001
*通信作者. E-mail：renhaoyuan@buaa.edu.cn

引用格式：任浩源，程涛，张程，等. 基于滑模控制和舵轴驱动的多间隙折叠舵颤振抑制方法[J]. 北京航空航天大学学报，2025，51(9)：2987-3000. REN H Y，CHENG T，ZHANG C，et al. Flutter suppression method for multi-freepalys folding fin based on sliding mode control and fin shaft drive[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：2987-3000 (in Chinese).

效的颤振抑制方法。

折叠翼/舵非线性气动弹性问题是国内外研究热点[7]。He等[8-9]利用子结构模态综合法和非线性扭转弹簧对折叠舵进行建模，基于活塞流理论建立了非定常气动力模型，讨论了间隙刚度特性对气动弹性稳定性的影响。Yang等[10]开展了含间隙折叠舵的风洞试验研究，比较了含间隙系统与无间隙系统之间颤振速度、颤振动压的差别，并基于子结构综合法和Zona6气动模型进行颤振数值模拟。Zhou等[11]针对双线性刚度折叠翼模型的气动弹性问题进行研究，通过建立参数化模型分析不同折叠角度时的颤振边界，并指出折叠机构中的非线性刚度能够抑制颤振。Castrichini等[12]分析了民航客机的翼尖折叠机构的非线性连接特性对阵风载荷减缓的影响，探讨了非线性刚度与弯矩阈值、折叠机构阻尼、翼尖振动响应之间的关系。Firouz-Abadi等[13-14]构建了考虑沉浮、俯仰和扑动间隙的三维全动舵气动弹性模型，研究了不同飞行马赫数、舵面形状和舵轴位置对极限环振动和混沌振动的影响。

颤振抑制是飞行器设计中的一项关键技术，得到了学者和工程设计师的广泛关注[15-17]。根据颤振抑制的作用机理可将其划分为被动抑制和主动抑制[18]：被动抑制的思想是通过调整结构刚度、阻尼或改变翼面配重特性来提高颤振速度；主动抑制的思想是利用系统的状态反馈，通过闭环控制方法驱动机翼后缘舵面偏转来产生控制力，进而改变机翼的颤振特性。黄锐等[19]综述了飞行器气动伺服弹性研究中的非定常气动力建模、非线性结构动力学建模和控制率设计方法等研究进展和挑战。白刘月等[20]分析了含扭转间隙全动舵面的典型非线性颤振模式及影响因素，提出了调整舵面根部的弯扭刚度比和改变舵面质量特性分布等颤振被动抑制方法，对正向设计舵面颤振特性很有实用意义。Eugeni等[21]提出在飞行器结构部件中嵌入阻尼耗散材料来提高颤振速度，通过敏感性分析来研究不同颤振模式的最优阻尼贴片布置方案。宋晨等[22]针对含双后缘控制面的三维机翼模型，建立了基于低阶滑模观测器的颤振主动抑制方法，模型中未考虑翼面连接结构的间隙特性。Ouyang等[23]提出采用可变形襟翼进行气动弹性主动抑制的方法，分析了弧形襟翼相比传统平面襟翼的优势。

在以上文献综述的颤振主动抑制方法中，主要将机翼-副翼组合结构的副翼偏转角作为控制系统输入量。但是对于全动折叠舵，不存在副翼结构，更难以直接对外舵施加控制力或控制位移，内舵的舵轴处转动位移是折叠空气舵系统的唯一输入量。因此，本文针对这一现实条件发展了通过驱动内舵的舵轴处转动自由度来抑制可折叠外舵颤振的新思路。首先，对折叠舵的内/外舵面引入刚化原理，将折叠舵系统缩减为4自由度非线性动力学模型，基于活塞理论推导了非定常气动力模型，通过与Nastran商业气动弹性分析软件的计算结果进行对比，校核了模型的正确性和精度；其次，分析了不同间隙条件下的非线性动力学响应行为，讨论了间隙对颤振速度的影响；最后，设计了滑模控制器来抑制外舵俯仰自由度的发散，破坏了颤振发生的必要耦合机制，进而实现了对线性折叠舵和间隙折叠舵的颤振主动抑制，该方法具有一定的实际意义和潜在的工程应用价值。

1 折叠舵气动弹性模型描述

1.1 折叠舵与简化模型

本文研究对象为根部折叠舵面，由外舵、内舵和折叠机构组成，如图1所示。外舵和内舵为高温合金钢结构，折叠机构的物理模型和力学特性可参考文献[4-5]。由于折叠机构属于弱刚度环节，连接刚度远小于舵面本体刚度，气动弹性分析时，相互耦合的振型近似为舵面的刚体振型。通过有限元方法进行试算，分别建立考虑舵面刚度的柔性舵面模型和不考虑舵面刚度的刚性舵面模型，计算得到的前两阶固有频率相对误差较小，说明了采用舵面刚化原理建模的有效性，在建模时考虑舵面本身的质量效应(外舵的质量矩阵为 $\boldsymbol{M}_{\mathrm{o}}$，内舵的质量矩阵为 $\boldsymbol{M}_{\mathrm{i}}$)，而不考虑舵面的刚度效应。外舵面保留绕折叠轴转动方向、俯仰方向和沉浮方向位移，分别记作 β、α 和 h，即外舵面具有3自由度；内舵的舵轴与机体伺服作动机构连接，保留内舵的舵轴的转动位移 δ，即内舵面具有1自由度；采用非线性弹簧 (K_β,K_α,K_h) 模拟内舵和外舵之间折叠机构的连接特性。简化后的折叠舵动力学模型可被视为舵轴转动基础激励 δ 和气流激励 F_{air} 共同作用下的4自由度“质量-弹簧”系统。

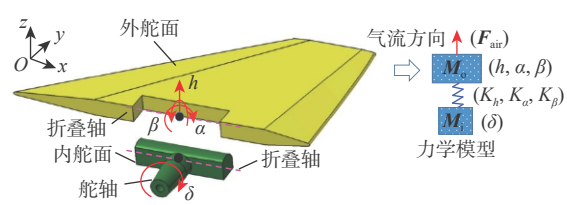


图1 折叠舵结构与等效力学模型

Fig. 1 Folding fin structure and equivalent mechanical model

1.2 气动弹性动力学方程

设舵轴中心轴线通过内舵的质心，定义内舵的质心位置为坐标原点 O，以舵面弦向为 Ox 轴，舵面

展向为 Oy 轴, 建立坐标系。将非线性弹簧力和气动力处理为外力, 采用拉格朗日方程推导折叠舵气动弹性系统的动力学方程[14]。

定义系统的广义坐标: 外舵沿 Oz 方向沉浮位移 h、绕 Oy 轴转动位移 α、绕 Ox 轴转动位移 β、内舵绕 Oy 轴转动位移 δ。需要指出的是, 外舵面广义坐标 β 被定义为与坐标原点的绝对位移, 而不定义为与内舵面 δ 的相对位移, 这样处理便于质量矩阵和气动矩阵解耦。内舵面某点 (x,y,z) 速度矢量 $\boldsymbol{v}_{\text{in}}$ 的表达式为

$$\boldsymbol{v}_{\text{in}}=\dot{\boldsymbol{\theta}}_{\text{in}}\times\boldsymbol{r}_{\text{in}}=(\dot{\delta}\boldsymbol{j})\times(x\boldsymbol{i}+y\boldsymbol{j}+z\boldsymbol{k})=\dot{\delta}z\boldsymbol{i}-\dot{\delta}x\boldsymbol{k} \tag{1}$$

式中: $\dot{\boldsymbol{\theta}}_{\text{in}}$为内舵的角速度矢量; $\boldsymbol{r}_{\text{in}}$ 为坐标原点指向内舵某点的矢径; $\boldsymbol{i}$、$\boldsymbol{j}$ 和 $\boldsymbol{k}$ 分别为沿坐标轴 Ox、Oy 和 Oz 正方向的单位矢量。

在小角度条件下, 外舵某点 (x,y,z) 速度矢量 $\boldsymbol{v}_{\text{out}}$ 的表达式为

$$\begin{aligned}\boldsymbol{v}_{\text{out}}&=\dot{h}\boldsymbol{k}+\dot{\boldsymbol{\theta}}_{\text{out}}\times\boldsymbol{r}_{\text{out}}=\dot{h}\boldsymbol{k}+(\dot{\beta}\boldsymbol{i}+\dot{\alpha}\boldsymbol{j})\times(x\boldsymbol{i}+y\boldsymbol{j}+z\boldsymbol{k})=\\&\dot{h}\boldsymbol{k}+\dot{\beta}y\boldsymbol{k}-\dot{\beta}z\boldsymbol{j}-\dot{\alpha}x\boldsymbol{k}+\dot{\alpha}z\boldsymbol{i}=\dot{\alpha}z\boldsymbol{i}-\dot{\beta}z\boldsymbol{j}+(\dot{h}+\dot{\beta}y-\dot{\alpha}x)\boldsymbol{k}\end{aligned} \tag{2}$$

式中: $\dot{\boldsymbol{\theta}}_{\text{out}}$为外舵的角速度矢量; $\boldsymbol{r}_{\text{out}}$ 为坐标原点指向外舵某点的矢径。

系统的动能表达式为

$$\begin{aligned}T=T_{\text{in}}+T_{\text{out}}=&\frac{1}{2}\iiint_{\Omega}(\rho_{\text{in}}\boldsymbol{v}_{\text{in}}\cdot\boldsymbol{v}_{\text{in}})\,\mathrm{d}x\mathrm{d}y\mathrm{d}z+\\&\frac{1}{2}\iiint_{\Omega}(\rho_{\text{out}}\boldsymbol{v}_{\text{out}}\cdot\boldsymbol{v}_{\text{out}})\,\mathrm{d}x\mathrm{d}y\mathrm{d}z=\\&\frac{1}{2}\iiint_{\Omega}\left[\rho_{\text{in}}\dot{\delta}^2\left(x^2+z^2\right)\right]\mathrm{d}x\mathrm{d}y\mathrm{d}z+\\&\frac{1}{2}\iiint_{\Omega}\rho_{\text{out}}\left[\dot{\alpha}^2z^2+\dot{\beta}^2z^2+\dot{h}^2+\dot{\beta}^2y^2+\dot{\alpha}^2x^2+\right.\\&\left.2\dot{h}\dot{\beta}y-2\dot{h}\dot{\alpha}x-2\dot{\alpha}\dot{\beta}xy\right]\mathrm{d}x\mathrm{d}y\mathrm{d}z\end{aligned} \tag{3}$$

式中: T 为系统的总动能; T_{in} 为内舵的动能; T_{out} 为外舵动能; ρ_{in} 和 ρ_{out} 分别为内舵和外舵的材料密度; Ω 为体积分域。

根据拉格朗日方程列写系统力学平衡方程:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial\dot{q}_j}\right)=Q_j\qquad j=1,2,3,4 \tag{4}$$

式中: q_j 为第 j 个自由度对应的广义坐标; Q_j 为第 j 个自由度对应的广义力。

需要说明的是, 由于舵面本体刚度在文中忽略, 舵面弹性势能为 0; 含间隙条件下, 弹簧边界约束刚度存在分段线性特性, 将弹簧力处理为广义力, 基本方程(4)中不包含弹性势能项。

定义 $\boldsymbol{q}=[\delta,h,\alpha,\beta]^{\mathrm{T}}$, 式(4)左侧表达式展开如下:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial\dot{\delta}}\right)=\iiint_{\Omega}\left[\rho_{\text{in}}\ddot{\delta}\left(x^2+z^2\right)\right]\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{5}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial\dot{h}}\right)=\iiint_{\Omega}\rho_{\text{out}}\left(\ddot{h}+\ddot{\beta}y-\ddot{\alpha}x\right)\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{6}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial\dot{\alpha}}\right)=\iiint_{\Omega}\rho_{\text{out}}\left[\ddot{\alpha}\left(z^2+x^2\right)-\ddot{h}x-\ddot{\beta}xy\right]\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{7}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial\dot{\beta}}\right)=\iiint_{\Omega}\rho_{\text{out}}\left[\ddot{\beta}\left(z^2+y^2\right)+\ddot{h}y-\ddot{\alpha}xy\right]\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{8}$$

式(4)右侧的广义力 Q_j 包括非线性弹簧力和非定常气动力两部分贡献量, 当采用对称间隙非线性刚度模型时, 非线性弹簧力(F_h、F_α 和 F_β)与广义位移的关系式如下:

$$F_h=\begin{cases}K_h(h-h_{\text{s}}) & h\geqslant h_{\text{s}}\\0 & -h_{\text{s}}\leqslant h<h_{\text{s}}\\K_h(h+h_{\text{s}}) & h<-h_{\text{s}}\end{cases} \tag{9}$$

$$F_\alpha=\begin{cases}K_\alpha(\alpha-\delta-\alpha_{\text{s}}) & \alpha-\delta\geqslant\alpha_{\text{s}}\\0 & -\alpha_{\text{s}}\leqslant\alpha-\delta<\alpha_{\text{s}}\\K_\alpha(\alpha-\delta+\alpha_{\text{s}}) & \alpha-\delta<-\alpha_{\text{s}}\end{cases} \tag{10}$$

$$F_\beta=\begin{cases}K_\beta(\beta-\beta_{\text{s}}) & \beta\geqslant\beta_{\text{s}}\\0 & -\beta_{\text{s}}\leqslant\beta<\beta_{\text{s}}\\K_\beta(\beta+\beta_{\text{s}}) & \beta<-\beta_{\text{s}}\end{cases} \tag{11}$$

式中: h_{s}、α_{s} 和 β_{s} 分别为外舵沉浮、俯仰和扑动方向的单边间隙量。

采用活塞理论进行非定常气动力计算, 舵面某点的压强 P 与法向挠度 w 之间的关系如下:

$$P=\frac{\rho|\boldsymbol{U}|^2}{\sqrt{Ma^2-1}}\left(w'+\frac{1}{|\boldsymbol{U}|}\cdot\frac{Ma^2-2}{Ma^2-1}\dot{w}\right) \tag{12}$$

式中: Ma 为飞行马赫数; ρ 为空气密度; $\boldsymbol{U}$ 为气流速度矢量; w'和$\dot{w}$分别为舵面压强计算点的斜率和速度, 可以由舵面外法向矢量 $\boldsymbol{n}$、转动位移矢量 $\boldsymbol{\theta}$、速度矢量 $\boldsymbol{v}$ 和气流方向单位矢量 $\boldsymbol{U}/|\boldsymbol{U}|$表示如下:

$$w'=(\boldsymbol{n}\times\boldsymbol{\theta})\cdot\frac{\boldsymbol{U}}{|\boldsymbol{U}|} \tag{13}$$

$$\dot{w}=\boldsymbol{n}\cdot\boldsymbol{v} \tag{14}$$

将折叠舵表面进行分区, 可依据各分区的表面压强与位置坐标积分运算得到气动合力和力矩。为便于气动压强的计算, 内舵圆弧表面被假设为平面, 不考虑舵轴的气动力。提取舵面气动力和气动力矩在各广义自由度的分量, 得到非定常气动力的广义力表达式如下:

$$Q_\delta=\iint_S-P(zn_1-xn_3)\mathrm{d}S \tag{15}$$

$$Q_h=\iint_S-Pn_3\mathrm{d}S \tag{16}$$

$$Q_\alpha=\iint_S-P(zn_1-xn_3)\mathrm{d}S \tag{17}$$

$$Q_\beta=\iint_S-P(yn_3-zn_2)\mathrm{d}S \tag{18}$$

式中: n_1、n_2、n_3 为舵面外法向矢量 $\boldsymbol{n}$ 的分量; Q_δ 的积分面积为内舵表面; Q_h、Q_α、Q_β 的积分面积为外

舵表面，在求解时可采用第一型曲面积分公式将 $\mathrm{d}S$ 转化为 $\mathrm{d}x\mathrm{d}y$。

综上，将折叠舵系统气弹动力学方程式(4)整理为

$$\boldsymbol{M}\ddot{\boldsymbol{q}}=\boldsymbol{Q}_{\mathrm{k}}\boldsymbol{q}+\boldsymbol{Q}_{\mathrm{c}}\dot{\boldsymbol{q}}+\boldsymbol{F}_{\mathrm{spr}} \tag{19}$$

将式(19)转化为状态空间形式如下：

$$\dot{\boldsymbol{x}}=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{I}\\ \boldsymbol{M}^{-1}\boldsymbol{Q}_{\mathrm{k}} & \boldsymbol{M}^{-1}\boldsymbol{Q}_{\mathrm{c}}\end{bmatrix}\boldsymbol{x}+\begin{bmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{M}^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{F}_{\mathrm{spr}}\end{bmatrix} \tag{20}$$

$$\boldsymbol{x}=[\boldsymbol{q}\quad\dot{\boldsymbol{q}}]^{\mathrm{T}} \tag{21}$$

式中：$\boldsymbol{M}$ 为质量矩阵；$\boldsymbol{Q}_{\mathrm{k}}$ 为气动刚度矩阵；$\boldsymbol{Q}_{\mathrm{c}}$ 为气动阻尼矩阵；$\boldsymbol{F}_{\mathrm{spr}}$ 为非线性弹簧力向量；各矩阵和向量的表达式见附录A。

2 基于滑模控制器的颤振抑制理论

2.1 基准线性折叠舵系统的控制

定义无间隙连接刚度的舵系统为基准线性系统，将外舵位移 (h,α,β) 定义为系统的状态量，将内舵的转动自由度 δ 定义为输入量(或控制量)u，将式(20)整理如下：

$$\begin{bmatrix}\ddot{h}\\ \ddot{\alpha}\\ \ddot{\beta}\end{bmatrix}=\boldsymbol{M}'^{-1}\left(\boldsymbol{Q}'_{\mathrm{k}}+\boldsymbol{K}'_{\mathrm{k}}\right)\begin{bmatrix}h\\ \alpha\\ \beta\end{bmatrix}+\boldsymbol{M}'^{-1}\boldsymbol{Q}'_{\mathrm{c}}\begin{bmatrix}\dot{h}\\ \dot{\alpha}\\ \dot{\beta}\end{bmatrix}+\boldsymbol{M}'^{-1}\boldsymbol{K}_{u}\begin{bmatrix}0\\ 1\\ 0\end{bmatrix}u \tag{22}$$

式中，各矩阵表达式如下：

$$\boldsymbol{M}'=\begin{bmatrix}M_{22} & M_{23} & M_{24}\\ M_{32} & M_{33} & M_{34}\\ M_{42} & M_{43} & M_{44}\end{bmatrix} \tag{23}$$

$$\boldsymbol{Q}'_{\mathrm{k}}=\begin{bmatrix}Q_{\mathrm{k}22} & Q_{\mathrm{k}23} & Q_{\mathrm{k}24}\\ Q_{\mathrm{k}32} & Q_{\mathrm{k}33} & Q_{\mathrm{k}34}\\ Q_{\mathrm{k}42} & Q_{\mathrm{k}43} & Q_{\mathrm{k}44}\end{bmatrix} \tag{24}$$

$$\boldsymbol{Q}'_{\mathrm{c}}=\begin{bmatrix}Q_{\mathrm{c}22} & Q_{\mathrm{c}23} & Q_{\mathrm{c}24}\\ Q_{\mathrm{c}32} & Q_{\mathrm{c}33} & Q_{\mathrm{c}34}\\ Q_{\mathrm{c}42} & Q_{\mathrm{c}43} & Q_{\mathrm{c}44}\end{bmatrix} \tag{25}$$

$$\boldsymbol{K}'_{\mathrm{k}}=\begin{bmatrix}-K_{h} & 0 & 0\\ 0 & -K_{\alpha} & 0\\ 0 & 0 & -K_{\beta}\end{bmatrix} \tag{26}$$

$$\boldsymbol{K}_{u}=\begin{bmatrix}0 & 0 & 0\\ 0 & K_{\alpha} & 0\\ 0 & 0 & 0\end{bmatrix} \tag{27}$$

对方程(22)分析可知：系统的输入量仅有1个，而系统状态量有3个，通过1个舵轴转动输入来同时实现沉浮、俯仰和扑动3个状态的收敛是主要难点。根据气动弹性颤振机理可知：若实现气动能量向结构能量的转换，需要俯仰产生的气动力在沉浮位移或扑动位移作正功，因此，只要抑制俯仰自由度收敛，则沉浮和扑动自由度无法实现能量交换，并在气动阻尼的作用下逐渐收敛，进而实现颤振抑制的目标。基于该思想提出的滑模控制方法相比文献[24]可省去一个输入量，且无需采用文献[22]的状态观测器对模态坐标进行观测。

定义俯仰角度误差 $e=\alpha$，滑模函数 s 表达式为

$$s=\dot{e}+e \tag{28}$$

分析可知，当 $s=0$ 时，误差函数 e 以指数方式收敛，因此，在滑模面系统稳定。为证明在控制器作用下滑模函数 s 能够渐近稳定，定义李雅普诺夫函数为

$$L=(\dot{e}+e)^2/2=s^2/2 \tag{29}$$

对式(29)求导可得

$$\dot{L}=s\dot{s} \tag{30}$$

令 $\dot{s}$ 满足如下关系：

$$\dot{s}=-ks-\varepsilon\mathrm{sign}(s)\qquad k>0,\varepsilon\geqslant 0 \tag{31}$$

将式(22)和式(28)代入式(31)，则：

$$\ddot{\alpha}+\dot{\alpha}=-k(\dot{\alpha}+\alpha)-\varepsilon\mathrm{sign}(\dot{\alpha}+\alpha) \tag{32}$$

$$\ddot{\alpha}=\boldsymbol{A}\begin{bmatrix}h\\ \alpha\\ \beta\end{bmatrix}+\boldsymbol{B}\begin{bmatrix}\dot{h}\\ \dot{\alpha}\\ \dot{\beta}\end{bmatrix}+Cu \tag{33}$$

式中：行向量 $\boldsymbol{A}$ 为矩阵 $\boldsymbol{M}'^{-1}\left(\boldsymbol{Q}'_{k}+\boldsymbol{K}'_{k}\right)$ 的第2行；行向量 $\boldsymbol{B}$ 为矩阵 $\boldsymbol{M}'^{-1}\boldsymbol{Q}'_{c}$ 的第2行；常数 C 为 $\boldsymbol{M}'^{-1}\boldsymbol{K}_{u}[0\quad 1\quad 0]^{\mathrm{T}}$ 的第2个元素，将式(33)代入式(32)可得控制量 u 的表达式为

$$u=\frac{1}{C}\left\{-k(\dot{\alpha}+\alpha)-\varepsilon\mathrm{sign}(\dot{\alpha}+\alpha)-\dot{\alpha}-\boldsymbol{A}\begin{bmatrix}h\\ \alpha\\ \beta\end{bmatrix}-\boldsymbol{B}\begin{bmatrix}\dot{h}\\ \dot{\alpha}\\ \dot{\beta}\end{bmatrix}\right\} \tag{34}$$

通过上述讨论，控制量 u 可使式(31)成立，$\dot{L}$ 显然小于0，被控量 α 的渐进稳定性得证。

2.2 间隙非线性折叠舵系统的控制

对于含间隙非线性系统，将式(20)整理如下：

$$\begin{bmatrix}\ddot{h}\\ \ddot{\alpha}\\ \ddot{\beta}\end{bmatrix}=\boldsymbol{M}'^{-1}\boldsymbol{Q}'_{\mathrm{k}}\begin{bmatrix}h\\ \alpha\\ \beta\end{bmatrix}+\boldsymbol{M}'^{-1}\boldsymbol{Q}'_{\mathrm{c}}\begin{bmatrix}\dot{h}\\ \dot{\alpha}\\ \dot{\beta}\end{bmatrix}+\boldsymbol{M}'^{-1}\boldsymbol{K}_{\mathrm{knl}}\begin{bmatrix}h-h_{\mathrm{s}}\\ \alpha-\alpha_{\mathrm{s}}\\ \beta-\beta_{\mathrm{s}}\end{bmatrix}+\boldsymbol{M}'^{-1}\boldsymbol{K}_{u\mathrm{nl}}\begin{bmatrix}0\\ 1\\ 0\end{bmatrix}u \tag{35}$$

式中：矩阵 $\boldsymbol{K}_{\mathrm{knl}}$ 和矩阵 $\boldsymbol{K}_{u\mathrm{nl}}$ 随系统处于间隙区或线性区的状态而改变，表达式如下：

$$\boldsymbol{K}_{\mathrm{knl}}=\begin{bmatrix}-k_{11} & 0 & 0\\ 0 & -k_{22} & 0\\ 0 & 0 & -k_{33}\end{bmatrix} \tag{36}$$

$$\boldsymbol{K}_{unl}=\begin{bmatrix}0 & 0 & 0\\ 0 & k_{22} & 0\\ 0 & 0 & 0\end{bmatrix} \tag{37}$$

$$\begin{cases} k_{11}=\begin{cases}0 & -h_s<h<h_s\\ K_h & \text{其他}\end{cases}\\ k_{22}=\begin{cases}0 & -\alpha_s<\alpha-\delta<\alpha_s\\ K_\alpha & \text{其他}\end{cases}\\ k_{33}=\begin{cases}0 & -\beta_s<\beta<\beta_s\\ K_\beta & \text{其他}\end{cases}\end{cases} \tag{38}$$

将外舵的俯仰角度 α 作为被控量。需要指出的是, 对于含俯仰方向间隙的折叠舵, 若将俯仰角度 α 的控制目标定为 0, 则在俯仰位移进入间隙区后, 输入信号出现大幅频繁抖动, 因此, 将俯仰角度控制在间隙边界处, 误差函数取为 $e=\alpha-\alpha_s$。采用与线性系统相同的滑模函数和李雅普诺夫函数, 将式 (35) 代入式 (32) 可得控制量 u 的表达式如下:

$$u=\frac{1}{C}\left\{-k(\dot{\alpha}+\alpha-\alpha_s)-\varepsilon\,\mathrm{sign}(\dot{\alpha}+\alpha-\alpha_s)-\dot{\alpha}-\boldsymbol{A}_{nl}\begin{bmatrix}h\\ \alpha\\ \beta\end{bmatrix}-\boldsymbol{B}_{nl}\begin{bmatrix}\dot{h}\\ \dot{\alpha}\\ \dot{\beta}\end{bmatrix}-\boldsymbol{D}_{nl}\begin{bmatrix}h-h_s\\ \alpha-\alpha_s\\ \beta-\beta_s\end{bmatrix}\right\} \tag{39}$$

式中: 行向量 $\boldsymbol{A}_{nl}$ 为矩阵 $\boldsymbol{M}'^{-1}\boldsymbol{Q}'_k$ 的第 2 行; 行向量 $\boldsymbol{B}_{nl}$ 为矩阵 $\boldsymbol{M}'^{-1}\boldsymbol{Q}'_c$ 的第 2 行; 行向量 $\boldsymbol{D}_{nl}$ 为矩阵 $\boldsymbol{M}'^{-1}\boldsymbol{K}_{knl}$ 的第 2 行; 常数 C 为 $\boldsymbol{M}'^{-1}\boldsymbol{K}_{unl}\begin{bmatrix}0 & 1 & 0\end{bmatrix}^{\mathrm{T}}$ 的第 2 个元素。需要指出的是, 式(39)中系数矩阵随 u 改变, 故在实际编程计算时, 控制量 u 需要先自定义初值, 再通过迭代计算得到。为防止间隙状态矩阵奇异, 可将间隙状态时矩阵 $\boldsymbol{K}_{knl}$ 和矩阵 $\boldsymbol{K}_{unl}$ 中的元素 k_{ii}(i=1, 2, 3)取为 1。

3　数值仿真

3.1　气动弹性模型验证

为验证结构动力学方程和非定常气动力模型的正确性, 考虑到当前尚缺乏针对三维非线性折叠舵气动弹性模型的标准计算结果或通用分析软件, 先采用退化模型对本文推导的结构模态特性和颤振特性计算方法进行验证, 即: 固定内舵转轴, 且不考虑系统间隙时, 折叠舵模型退化为经典三自由度线性气动弹性系统。将表 1 所示的数据代入方程, 求解系统的前三阶模态特性。表 2 为模态频率的本文方法和 Nastran 有限元软件结果, 以及舵面各阶振型。由表 2 可知: 模态频率的本文方法结果与 Nastran 结果的相对误差在 0.2% 以内, 证明了结构模型的正确性。

采用高度为 0 km 的气流条件, 分别计算 Ma=3 和 Ma=5 时折叠舵的颤振曲线, 并与 Nastran 基于平板气动模型的颤振计算结果进行对比, V-f 曲线和 V-g 曲线如图 2 和图 3 所示, 表 3 为本文方法与 Nastran 软件计算的颤振速度和颤振频率的对比。由图和表可知, 在所分析的工况条件下采用本文方法与

表 1　折叠舵模型的主要参数

Table 1　Main parameters of folding fin model

参数	数值
内舵相对坐标原点绕y轴转动惯量I_{yyin}/(kg·m^2)	0.003 9
外舵的质量m_{out}/kg	13.60
外舵质心位置的x坐标$\bar{x}_{out}$/m	0.047 1
外舵质心位置的y坐标$\bar{y}_{out}$/m	0.112
外舵相对坐标原点绕x轴转动惯量I_{xxout}/(kg·m^2)	0.246 7
外舵相对坐标原点绕y轴转动惯量I_{yyout}/(kg·m^2)	0.220 5
外舵相对坐标原点xy惯量积I_{xyout}/(kg·m^2)	0.115 4
折叠机构沉浮线弹簧连接刚度K_h/(N·m^{-1})	1×10^8
折叠机构俯仰扭簧连接刚度K_α/((N·m)·rad^{-1})	5×10^4
折叠机构扑动扭簧连接刚度K_β/((N·m)·rad^{-1})	4×10^4

表 2　折叠舵模态频率的本文方法与 Nastran 结果

Table 2　The proposed method and Nastran results of modal frequency of folding fin

模态阶数	模态频率/Hz		误差/%	振型
	本文方法	Nastran结果		
一阶	56.028	55.978	0.089 3	舵面弯曲
二阶	99.162	99.223	−0.061 5	舵面扭转
三阶	787.853	786.24	0.205 2	舵面扑动

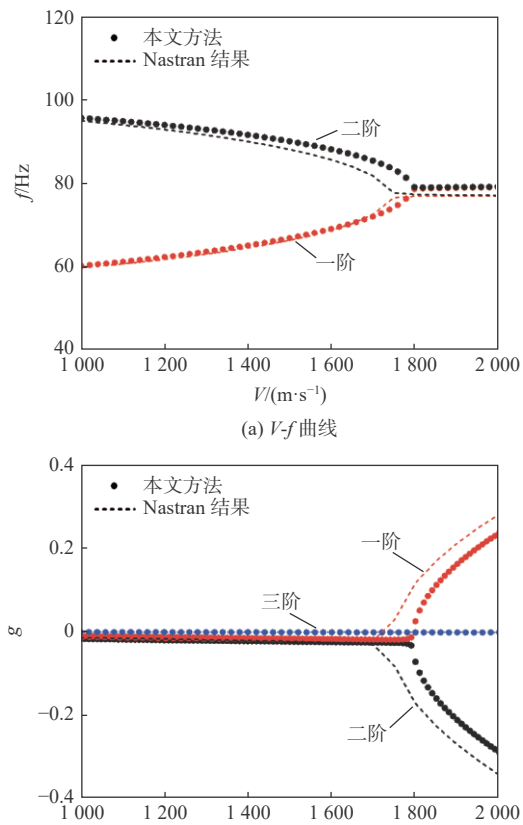
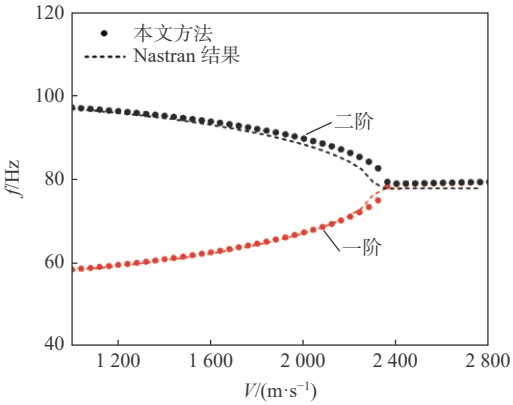
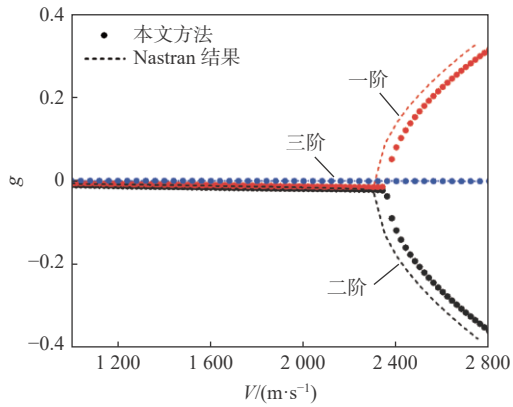


图 2　Ma=3 条件下折叠舵 V-f 与 V-g 曲线

Fig. 2　V-f and V-g curves of folding fin at Ma=3



(a) *V-f* 曲线



(b) *V-g* 曲线

图 3 *Ma*=5 条件下折叠舵 *V-f* 与 *V-g* 曲线

Fig. 3 *V-f* and *V-g* curves of folding fin at *Ma*=5

表 3 折叠舵颤振速度和颤振频率的本文方法与 Nastran 结果

Table 3 The proposed method and Nastran results of flutter speed and frequency for folding fin

Ma	颤振速度/(m·s⁻¹)		频率误差/%	颤振频率/Hz		频率误差/%
	本文方法	Nastran		本文方法	Nastran	
3	1 795	1 720	−4.178	79.34	76.63	−3.416
5	2 370	2 308	−2.616	78.75	77.84	−1.156

Nastran 软件计算的颤振速度相对误差小于 5%，颤振频率相对误差为−3.416%，且随马赫数增大相对误差减小。

图 2 和图 3 给出的阻尼和频率在较大风速时存在差异，误差产生的原因为平板气动模型和活塞理论气动模型的基本假设的差异。Nastran 软件的平板气动模型主要采用超音速偶极子网格法，是基于线性化速度势方程的超音速非定常空气动力学计算方法，速度势方程是在 Euler 方程的基础上引入了正压气体、无旋运动和等熵流动假设形成。超音速条件下，对于超音速偶极子网格法，针对翼面上一点 *H*，在以 *H* 为顶点的倒置马赫锥内的点所发出的扰动能够影响到 *H* 点；而活塞理论非定常气动力模型认为翼型上某一点 *H* 的压力只与该点的下洗速度有关，没有考虑翼面其他点扰动所产生的影响，即忽略了翼面气动的三维效应。因此，在相同马赫数下，2 种气动模型方法得到的颤振曲线存在误差。但由于气动三维效应随来流马赫数的提高而减小，对比 *Ma*=3 和 *Ma*=5 这 2 种来流工况，可以发现 2 种气动模型方法的计算误差随马赫数增大而减小。

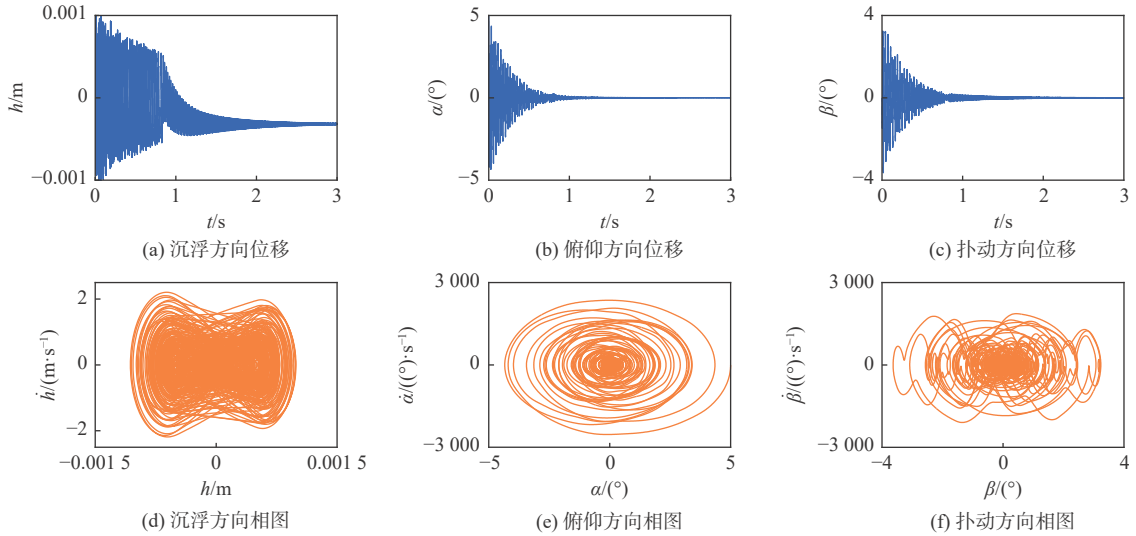
3.2 无控制条件下的非线性气动弹性响应

3.2.1 考虑单间隙时外舵面气动弹性响应

当舵机系统不进行驱动时，可认为舵轴转动自由度被固定，系统模型退化为经典非线性 3 自由度舵面模型，设模型的单边间隙值为 h_s=0.5 mm，α_s=1°，β_s=3°。采用高度为 0 km，*Ma*=3 气流条件，由 3.1 节计算结果可知线性折叠舵系统的颤振速度 V_f=1 795 m/s。假定外舵的初始迎角扰动 α_0=5°，采用自适应步长的 Runge-Kutta 算法对动力学方程进行求解，分别计算 0.6V_f、0.8V_f 和 V_f 共 3 种来流速度条件下的舵面位移时域响应。

图 4～图 12 分别为不同来流速度条件下，单独考虑沉浮 h_s、俯仰 α_s 和扑动 β_s 间隙时，系统的位移响应和相图。由图 4～图 12 可知：①对于所分析的 3 种气流条件，单独存在沉浮间隙或扑动间隙并不会导致系统发散，即间隙的存在提高了系统的颤振速度；②单独存在俯仰间隙时，在线性颤振速度以下系统就呈现出极限环振动现象，如图 8 所示，当速度尚未达到线性颤振速度时系统就存在振动幅值超出设计值的风险，即俯仰间隙降低了系统的颤振速度；③间隙对颤振速度影响的机理为折叠舵的一阶频率为扑动振型，二阶频率为俯仰振型，且表现为扑动和俯仰相耦合的振动模式，在该前提下，扑动方向的间隙减小了扑动方向的等效连接刚度，一阶频率降低，一阶和二阶频率差增大，系统的颤振速度提升；俯仰方向的间隙减小了俯仰方向的等效连接刚度，二阶频率降低，一阶和二阶的频率差减小，系统的颤振速度降低。

为进一步分析含俯仰间隙折叠舵的位移响应，图 13 为不同速度条件下舵面稳态振动的幅值。根据幅值-速度关系曲线可知：①当速度小于 0.73V_f 时，系统响应呈现收敛趋势，稳态振幅为 0；速度在 0.73V_f～V_f 时，系统呈现极限环振动，且极限环幅值随速度增大而增大；当速度大于 V_f 时，舵面位移发散。②在速度达到 0.84V_f 时，舵面极限环振动幅值为 5.33°，超出初始扰动角度，且随着速度增大极限环幅值将进一步增大，尽管在线性颤振速度 V_f 以内舵面未发生气弹发散，但实际上可能已经无法满足舵面使用要求，因此，含俯仰间隙折叠舵的许用飞行速度低于线性颤振边界，在设计时其颤振速度不应该由线性系统确定。

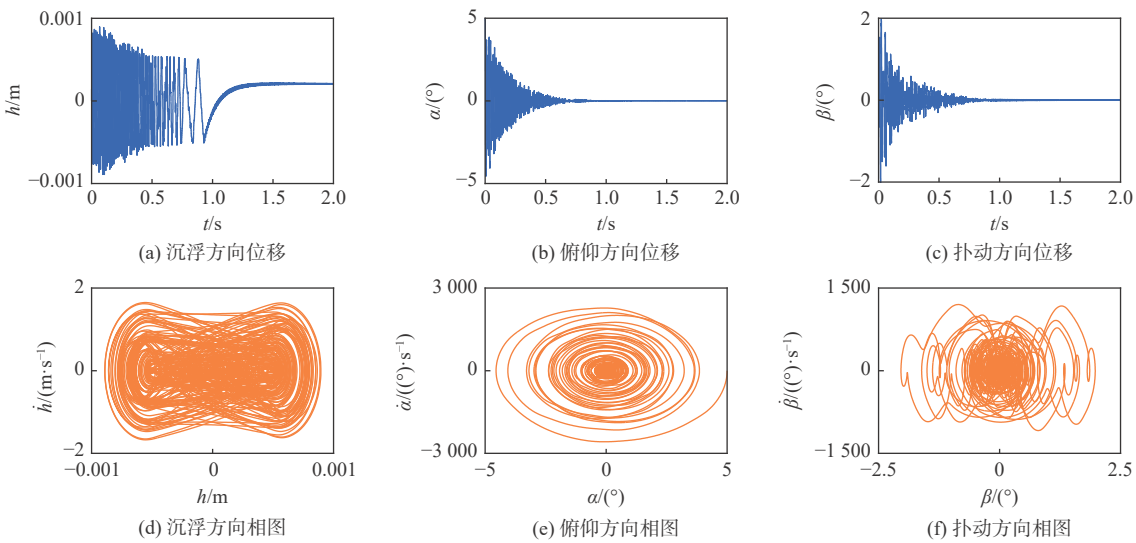


(a) 沉浮方向位移　(b) 俯仰方向位移　(c) 扑动方向位移

(d) 沉浮方向相图　(e) 俯仰方向相图　(f) 扑动方向相图

图 4　$0.6V_f$时含沉浮间隙折叠舵的时域响应和相图

Fig. 4　Time histories and phase plane diagram and of folding fin with plunging free-play at $0.6V_f$

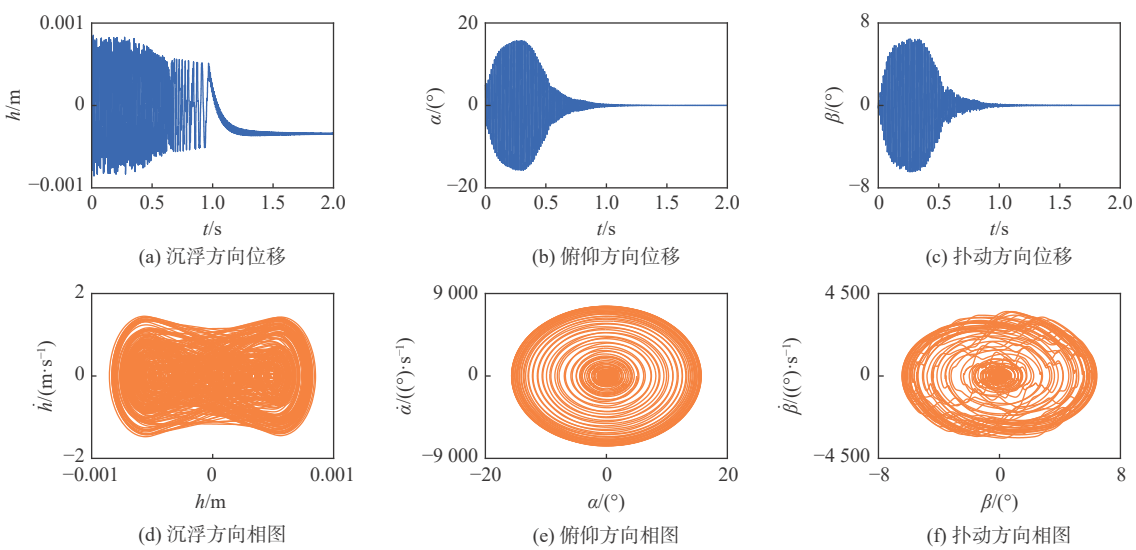


(a) 沉浮方向位移　(b) 俯仰方向位移　(c) 扑动方向位移

(d) 沉浮方向相图　(e) 俯仰方向相图　(f) 扑动方向相图

图 5　$0.8V_f$时含沉浮间隙折叠舵的时域响应和相图

Fig. 5　Time histories and phase plane diagram and of folding fin with plunging free-play at $0.8V_f$

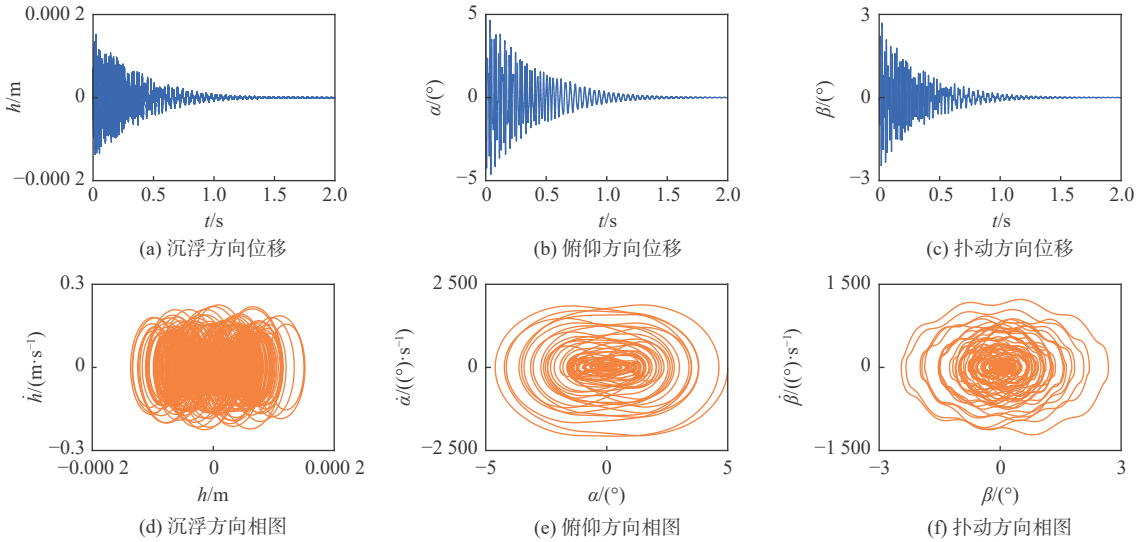


(a) 沉浮方向位移　(b) 俯仰方向位移　(c) 扑动方向位移

(d) 沉浮方向相图　(e) 俯仰方向相图　(f) 扑动方向相图

图 6　V_f时含沉浮间隙折叠舵的时域响应和相图

Fig. 6　Time histories and phase plane diagram and of folding fin with plunging free-play at V_f

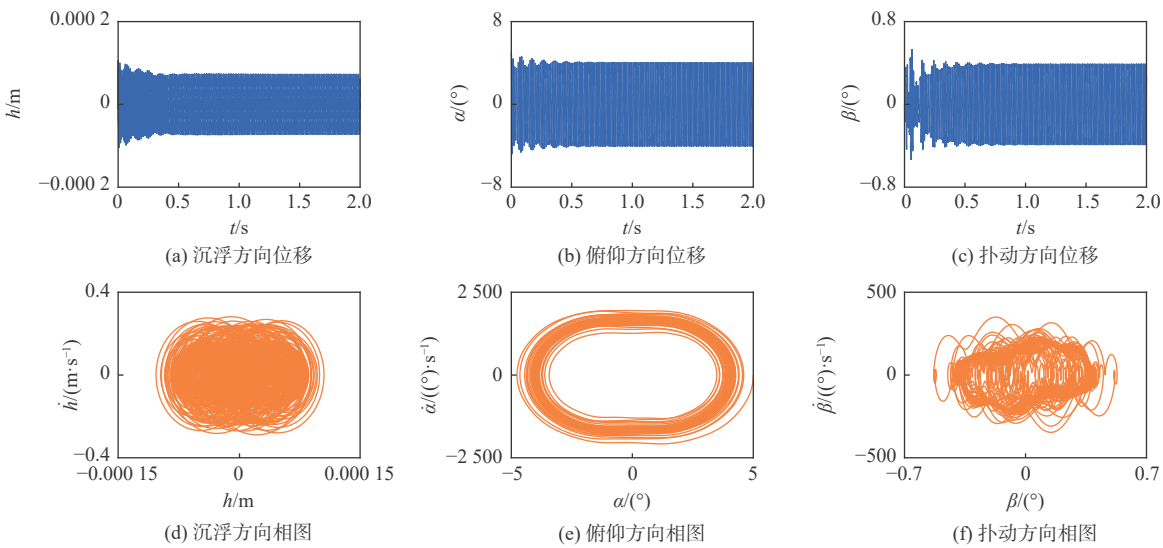


(a) 沉浮方向位移 (b) 俯仰方向位移 (c) 扑动方向位移

(d) 沉浮方向相图 (e) 俯仰方向相图 (f) 扑动方向相图

图 7 $0.6V_f$时含俯仰间隙折叠舵的时域响应和相图

Fig. 7 Time histories and phase plane diagram and of folding fin with pitching free-play at $0.6V_f$

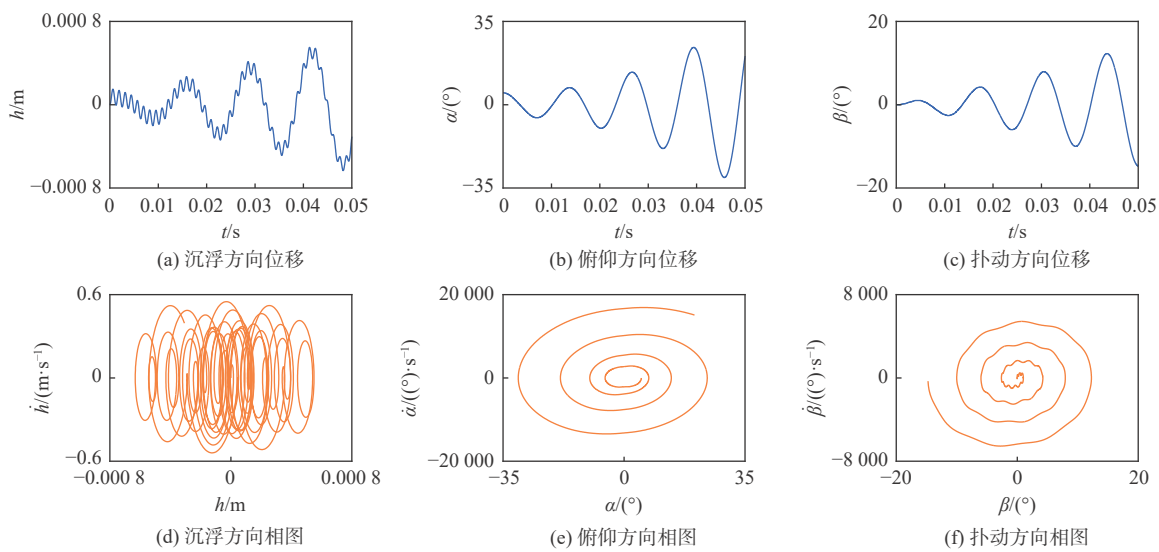


(a) 沉浮方向位移 (b) 俯仰方向位移 (c) 扑动方向位移

(d) 沉浮方向相图 (e) 俯仰方向相图 (f) 扑动方向相图

图 8 $0.8V_f$时含俯仰间隙折叠舵的时域响应和相图

Fig. 8 Time histories and phase plane diagram and of folding fin with pitching free-play at $0.8V_f$

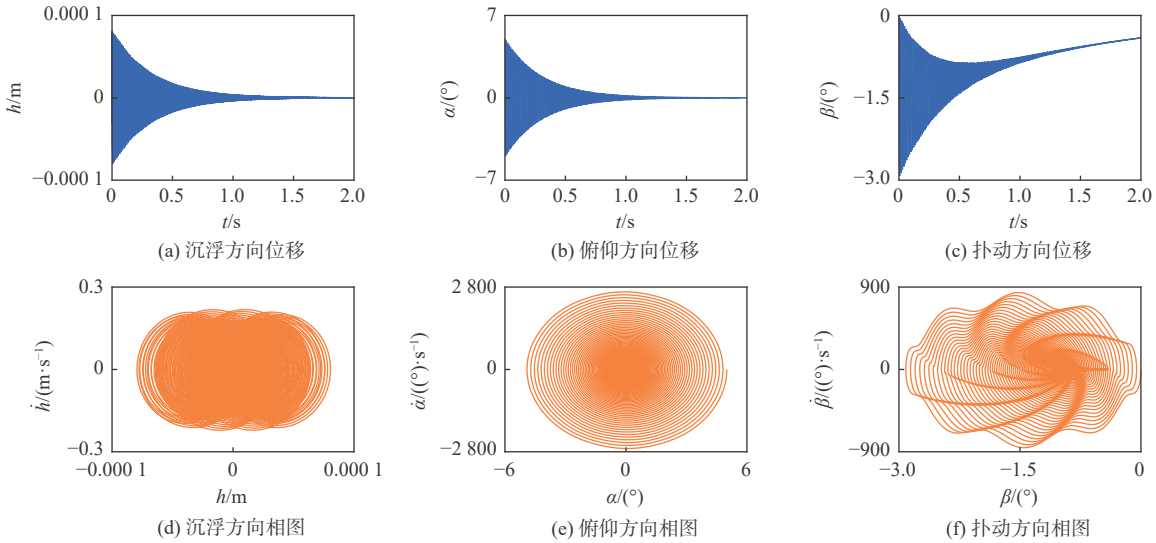


(a) 沉浮方向位移 (b) 俯仰方向位移 (c) 扑动方向位移

(d) 沉浮方向相图 (e) 俯仰方向相图 (f) 扑动方向相图

图 9 V_f时含俯仰间隙折叠舵的时域响应和相图

Fig. 9 Time histories and phase plane diagram and of folding fin with pitching free-play at V_f

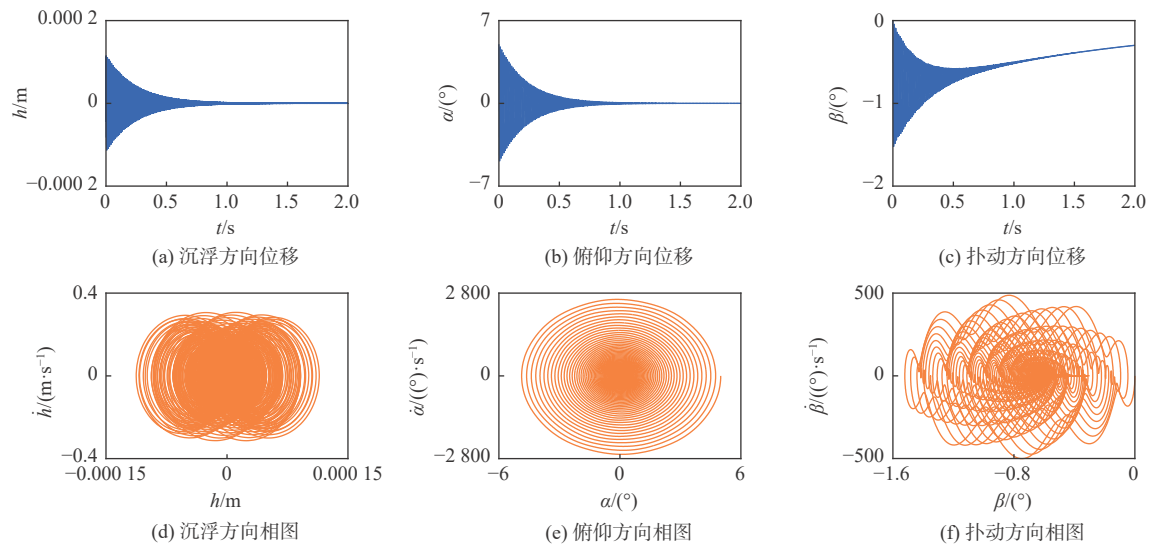


(a) 沉浮方向位移　(b) 俯仰方向位移　(c) 扑动方向位移

(d) 沉浮方向相图　(e) 俯仰方向相图　(f) 扑动方向相图

图 10　$0.6V_f$时含扑动间隙折叠舵的时域响应和相图

Fig. 10　Time histories and phase plane diagram and of folding fin with flapping free-play at $0.6V_f$

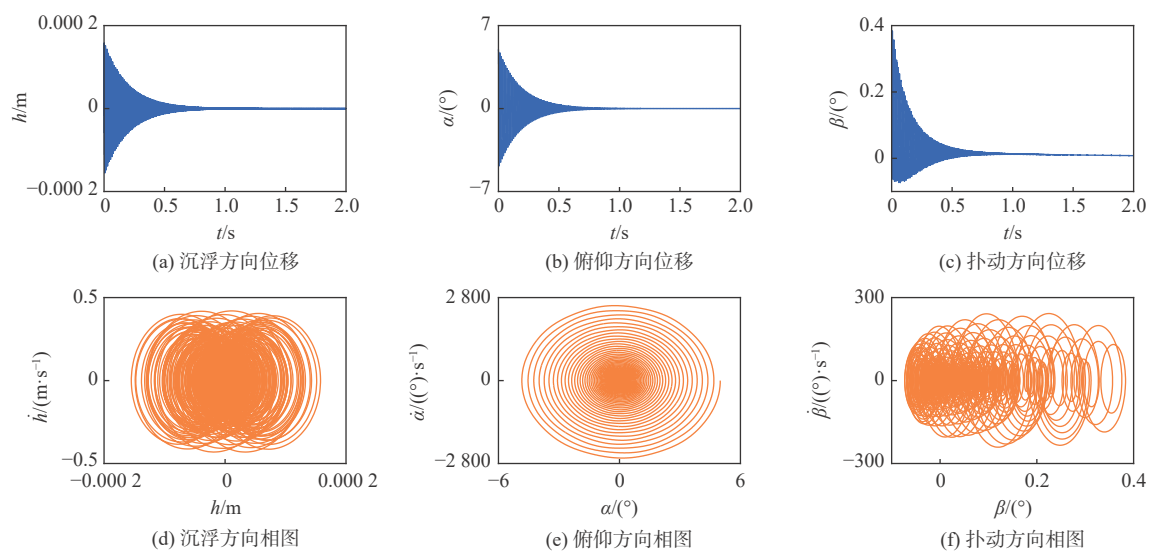


(a) 沉浮方向位移　(b) 俯仰方向位移　(c) 扑动方向位移

(d) 沉浮方向相图　(e) 俯仰方向相图　(f) 扑动方向相图

图 11　$0.8V_f$时含扑动间隙折叠舵的时域响应和相图

Fig. 11　Time histories and phase plane diagram and of folding fin with flapping free-play at $0.8V_f$

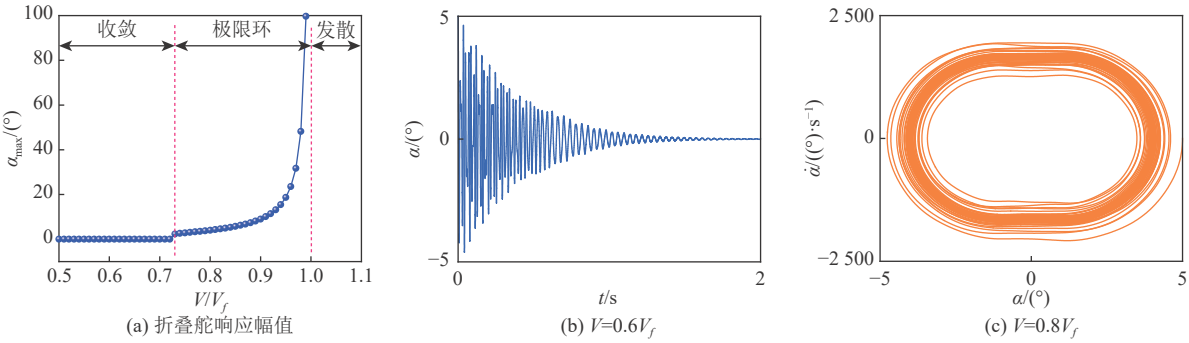


(a) 沉浮方向位移　(b) 俯仰方向位移　(c) 扑动方向位移

(d) 沉浮方向相图　(e) 俯仰方向相图　(f) 扑动方向相图

图 12　V_f时含扑动间隙折叠舵的时域响应和相图

Fig. 12　Time histories and phase plane diagram and of folding fin with flapping free-play at V_f



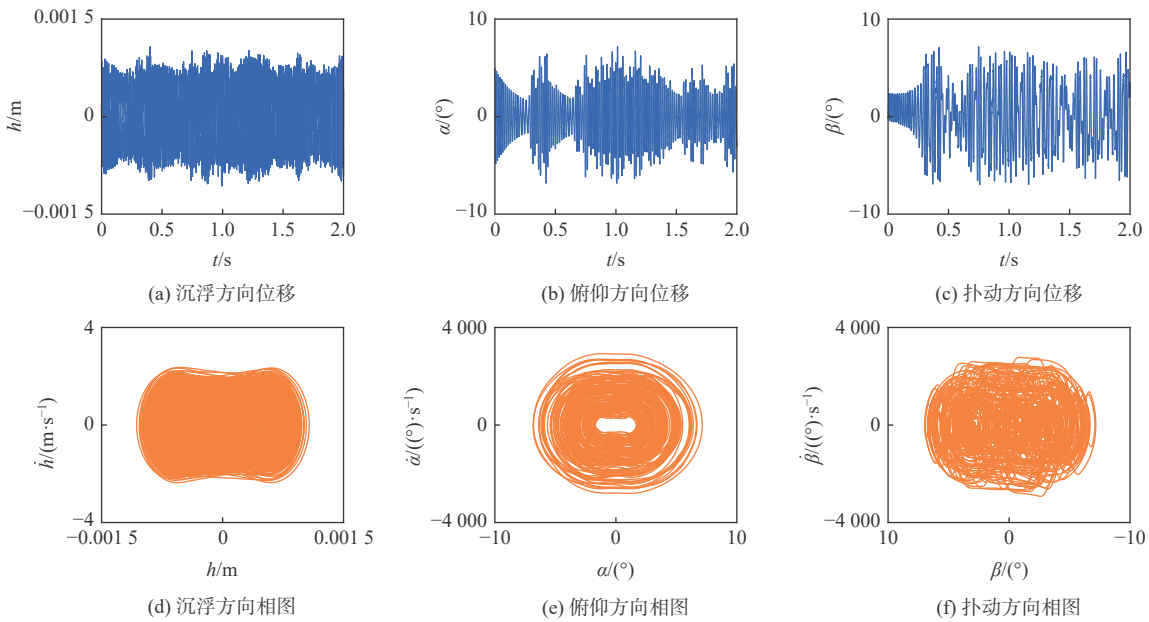
(a) 折叠舵响应幅值 (b) V=0.6V_f (c) V=0.8V_f

图 13 含俯仰间隙折叠舵响应幅值随气流速度变化曲线

Fig. 13 Amplitude response versus airflow velocity of folding fin with pitch free-play

3.2.2 考虑多间隙时外舵面气动弹性响应

采用高度为 0 km, Ma=3 气流条件，图 14 为 V_f 来流速度条件下，同时考虑沉浮 h_s、俯仰 α_s 和扑动 β_s 间隙时各自由度的响应历程和相图。分析可知，含三间隙折叠舵在临近颤振速度条件下未发散也未收敛，呈现较为混乱的振动模式。



(a) 沉浮方向位移 (b) 俯仰方向位移 (c) 扑动方向位移

(d) 沉浮方向相图 (e) 俯仰方向相图 (f) 扑动方向相图

图 14 V_f 时含三间隙折叠舵的时域响应和相图

Fig. 14 Time histories and phase plane diagram and of folding fin with three free-plays at V_f

3.3 有控制条件下的气动弹性响应

3.3.1 线性折叠舵的控制响应

对于基准线性折叠舵系统，将控制量 u 表达式 (34) 中参数设置为 k=5, ε=0.01，采用高度为 0 km, Ma=3 气流条件，设置初始扰动俯仰角度为 1°，计算来流速度为 V_f 时被控系统的动力学响应。图 15 为控制系统作用下外舵面各自由度的时域响应，图 16 为舵轴转动角时域输入参数，图 17 为滑模函数 s 随时间的收敛历程。由图 15～图 17 可知：①滑模控制函数 s 在 0.45 s 时收敛到 0，且未发生振荡，此时系统状态量中的俯仰角 α 达到了滑模面，随后俯仰角 α 以指数方式平滑收敛，达到了控制系统的预期；②控制量 u 在 0～0.6 s 区间明显振荡，其原因为：根据气动弹性动力学方程可知，系统的俯仰和

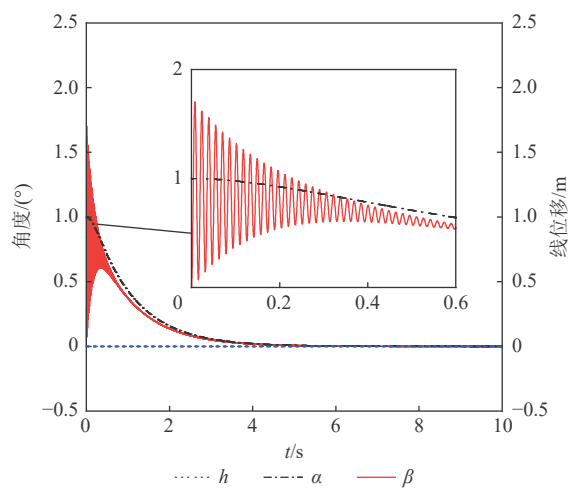


图 15 V_f 条件下被控线性折叠舵系统的位移响应

Fig. 15 Displacement response of controlled linear folding fin system at V_f

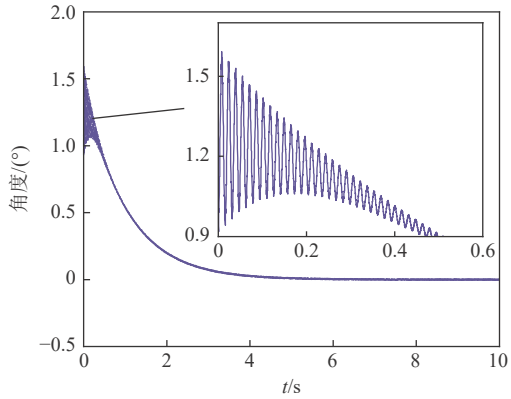


图 16　V_f 条件下线性折叠舵系统控制量 u 随时间变化

Fig. 16　Control parameter u of linear folding fin system versus time at V_f

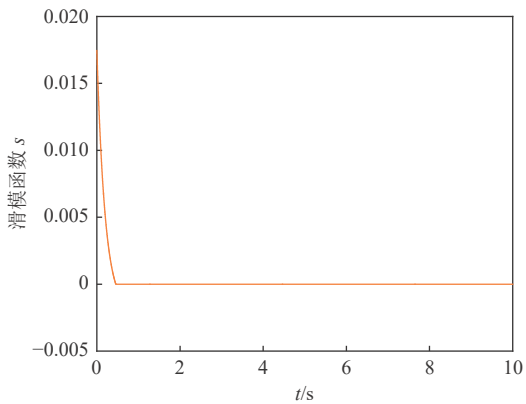


图 17　V_f 条件下线性折叠舵系统滑模函数 s 随时间变化

Fig. 17　Sliding mode function s of linear folding fin system versus time at V_f

扑动自由度存在耦合, 舵轴转动量 u 在初始时刻对俯仰角进行控制的同时, 不可避免地激发了扑动角的振荡, 再基于式(34)可推知扑动进一步导致了控制量 u 产生同步振荡。但随着气动阻尼作用, 扑动自由度仍能够在 5.2 s 完成收敛, 说明了气动弹性抑制机理的正确性和可实现性, 即采用滑模控制方法能够满足线性折叠舵系统的颤振抑制需求。

3.3.2　多间隙折叠舵的控制响应

引入 3 个方向的间隙弹簧刚度, 则被控对象为非线性折叠舵系统。飞行气流条件与线性折叠舵系统相同, 取俯仰角度的初值为 1.5°, 将非线性系统控制量 u 表达式 (39) 中参数设置为 k=5, ε=0.01, 计算被控系统的动力学响应。图 18 为舵轴控制作用下外舵面各自由度的时域响应, 图 19 为舵轴转动角时域输入参数, 图 20 为滑模函数 s 随时间的收敛历程。由图 18～图 20 可知, ①通过滑模控制算法和舵轴驱动, 在临界颤振速度条件下, 多间隙折叠舵系统仍能实现收敛, 符合控制的预期, 验证了本文方法对间隙折叠舵系统的颤振抑制效果。②间隙折叠舵俯仰角和扑动角的收敛值与线性折叠舵

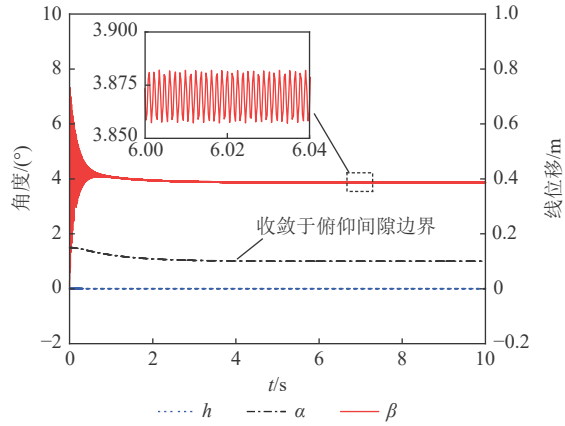


图 18　V_f 条件下被控间隙折叠舵系统的位移响应

Fig. 18　Displacement response of controlled free-paly folding fin system at V_f

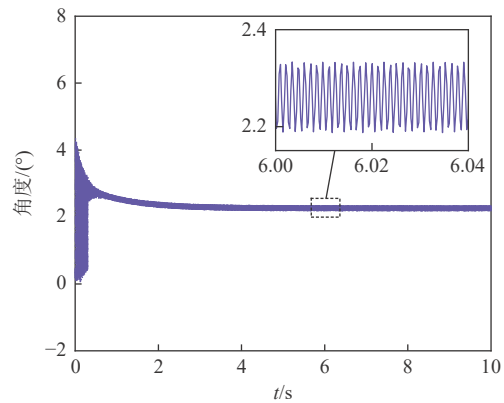


图 19　V_f 条件下间隙折叠舵系统控制量 u 随时间变化

Fig. 19　Control parameter u of free-paly folding fin system versus time at V_f

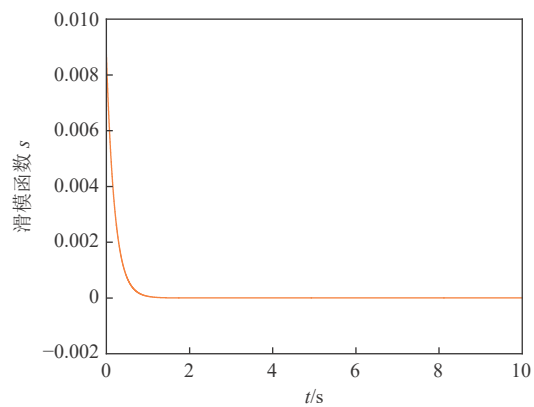


图 20　V_f 条件下间隙折叠舵系统滑模函数 s 随时间变化

Fig. 20　Sliding mode function s of free-paly folding fin system versus time at V_f

存在差异。线性折叠舵的角位移能够收敛到 0°, 而非线性系统的俯仰和扑动角位移分别收敛至 1°和 3.88°。其原因为: 为避免俯仰角进入间隙区后控制输入频繁大幅抖动, 俯仰角的控制量被设计为单边间隙值 1°, 因而俯仰角度最终被稳定控制在设定值; 根据舵面的力学平衡方程, 在 1°俯仰角时翼面

的升力、俯仰力矩和扑动力矩均为有限值，扑动弯矩导致了扑动方向的附加角位移，扑动角的最终收敛值包括间隙值和附加角位移2部分。③进入稳定阶段后，间隙折叠舵系统的控制量 u 仍存在0.2°小幅振荡。其原因为：扑动方向存在间隙导致抖动，且在稳定段也未收敛至0，根据式(39)可知，扑动方向的抖动位移和速度将导致控制量同步振荡，但 u 的振荡幅值在可接受范围内。

4 结 论

1）本文基于舵面刚性假设和活塞理论非定常气动力模型，构建了四自由度多间隙折叠舵气动弹性模型，并通过商业有限元软件进行模型校验，采用本文方法和软件仿真得到的结构模态频率相对误差在0.2%以内，颤振速度相对误差小于5%，表明了模型推导的正确性和可用性。

2）考察了单间隙和多间隙折叠舵在无控条件下的气动弹性响应，仿真结果表明：单独存在沉浮间隙或扑动间隙提高了系统的颤振速度，而单独存在俯仰间隙的折叠舵在线性颤振速度以下就出现极限环振动，存在振动幅值超出许用值的风险，俯仰间隙降低了系统的颤振速度。

3）本文提出了采用舵轴转动实现多间隙折叠舵颤振抑制的思路，通过设计滑模控制器控制外舵的俯仰自由度收敛，破坏颤振发生的必要耦合机制，实现了对线性折叠舵和多间隙折叠舵的颤振主动抑制，仿真结果表明了该方法的可行性。

4）针对多间隙折叠舵的误差函数确定，指出俯仰角位移的控制目标应定为俯仰间隙的边界处，而非俯仰角为0处，避免了外舵面俯仰角进入间隙区导致控制输入的高频大幅振荡。

本文对基于滑模控制的折叠舵颤振抑制方法进行了理论研究和机理阐述，提供了一种工程化解决思路，后续将考虑舵机和伺服传动系统的控制能力，开展飞行稳定控制和舵面颤振抑制的一体化设计分析方法研究，进一步向高速飞行器实际应用靠近。

参考文献（References）

[1] SAYLER K M. Emergency military technologies: background and issues for congress: R46458[R]. Washington, D. C. : Congressional Research Service, 2022: 1-37.

[2] DIAMANTE R G. Wing fold mechanism: US10272988[P]. 2019-04-30.

[3] FOX S J, SAKURAI S, MUNSEN V A. Wing folding system: EP2730499A2[P]. 2014-05-14.

[4] 任浩源, 王毅, 王亮, 等. 航天飞行器折叠翼锁紧机构力学模型[J]. 航空动力学报, 2023, 38(12): 3009-3019.
REN H Y, WANG Y, WANG L, et al. Mechanical model of locking mechanisms of folding wing for spacecraft[J]. Journal of Aerospace Power, 2023, 38(12): 3009-3019(in Chinese).

[5] 陈克, 金玲, 徐倩, 等. 基于高温合金的高速飞行器折叠舵结构设计与研究[J]. 航空兵器, 2022, 29(6): 71-77.
CHEN K, JIN L, XU Q, et al. Design and research of high-speed aircraft folding rudder structure based on high-temperature alloy[J]. Aero Weaponry, 2022, 29(6): 71-77(in Chinese).

[6] 陈克, 金玲, 雷豹, 等. 高速飞行器折叠翼/舵设计技术与进展研究[J]. 强度与环境, 2022, 49(1): 53-59.
CHEN K, JIN L, LEI B, et al. Research on design technology and progress of high-speed aircraft folding wing/rudder[J]. Structure & Environment Engineering, 2022, 49(1): 53-59(in Chinese).

[7] LIU S S, ZHAO R, YU K P, et al. Nonlinear system identification framework of folding fins with freeplay using backbone curves[J]. Chinese Journal of Aeronautics, 2022, 35(10): 183-194.

[8] HE H N, TANG H, YU K P, et al. Nonlinear aeroelastic analysis of the folding fin with freeplay under thermal environment[J]. Chinese Journal of Aeronautics, 2020, 33(9): 2357-2371.

[9] 何昊南, 于开平, 唐宏, 等. 有间隙折叠舵面的振动实验与非线性建模研究[J]. 力学学报, 2019, 51(5): 1476-1488.
HE H N, YU K P, TANG H, et al. Vibration experiment and nonlinear modelling research on the folding fin with freeplay[J]. Chinese Journal of Theoretical and Applied Mechanics, 2019, 51(5): 1476-1488(in Chinese).

[10] YANG N, WANG N, ZHANG X, et al. Nonlinear flutter wind tunnel test and numerical analysis of folding fins with freeplay nonlinearities[J]. Chinese Journal of Aeronautics, 2016, 29(1): 144-159.

[11] ZHOU X H, HUANG R. Subcritical and supercritical nonlinear aeroelastic behavior of a morphing wing with bilinear hinge stiffness[J]. Communications in Nonlinear Science and Numerical Simulation, 2021, 103: 105946.

[12] CASTRICHINI A, HODIGERE SIDDARAMAIAH V, CALDERON D E, et al. Nonlinear folding wing tips for gust loads alleviation[J]. Journal of Aircraft, 2016, 53(5): 1391-1399.

[13] FIROUZ-ABADI R D, ALAVI S M. Effect of thickness and angle-of-attack on the aeroelastic stability of supersonic fins[J]. The Aeronautical Journal, 2012, 116(1182): 777-792.

[14] FIROUZ-ABADI R D, ALAVI S M, SALARIEH H. Analysis of non-linear aeroelastic response of a supersonic thick fin with plunging, pinching and flapping free-plays[J]. Journal of Fluids and Structures, 2013, 40: 163-184.

[15] RICCI S, TOFFOL F, DE GASPARI A, et al. Wind tunnel system for active flutter suppression research: overview and insights[J]. AIAA Journal, 2022, 60(12): 6692-6714.

[16] THEIS J, PFIFER H, SEILER P. Robust modal damping control for active flutter suppression[J]. Journal of Guidance, Control, and Dynamics, 2020, 43(6): 1056-1068.

[17] TANG W, WANG Y, GU J W, et al. LPV modeling and controller design for body freedom flutter suppression subject to actuator saturation[J]. Chinese Journal of Aeronautics, 2020, 33(10): 2679-2693.

[18] 杨智春, 田玮, 谷迎松, 等. 带集中非线性的机翼气动弹性问题研究进展[J]. 航空学报, 2016, 37(7): 2013-2044.

YANG Z C, TIAN W, GU Y S, et al. Advance in the study on wing aeroelasticity with concentrated nonlinearity[J]. Acta Aeronautica et Astronautica Sinica, 2016, 37(7): 2013-2044(in Chinese).

[19] 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).

[20] 白刘月, 吴志刚, 杨超. 含间隙全动舵面的非线性颤振模式及被动抑制方法[J]. 北京航空航天大学学报, 2023, 49(9): 2361-2373.
BAI L Y, WU Z G, YANG C. Nonlinear flutter modes and flutter suppression of an all-movable fin with freeplay[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(9): 2361-2373(in Chinese).

[21] EUGENI M, SALTARI F, MASTRODDI F, et al. Structural damping models for passive aeroelastic control[J]. Aerospace Science and Technology, 2021, 118: 107011.

[22] 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).

[23] OUYANG Y, GU Y S, KOU X P, et al. Active flutter suppression of wing with morphing flap[J]. Aerospace Science and Technology, 2021, 110: 106457.

[24] 贾尚帅, 丁千. 含间隙超音速二元弹翼非线性颤振与主动控制[J]. 中国科学: 物理学 力学 天文学, 2013, 43(4): 390-400.
JIA S S, DING Q. Nonlinear flutter and active control of the supersonic two-dimensional missile wings with a freeplay[J]. Scientia Sinica (Physica, Mechanica & Astronomica), 2013, 43(4): 390-400 (in Chinese).

附录 A:

质量矩阵 $\boldsymbol{M}$ 表达式为

$$\boldsymbol{M}=\begin{bmatrix}\iiint_{\Omega}\rho\left(x^2+z^2\right)\mathrm{d}\Omega & 0 & 0 & 0\\ 0 & \iiint_{\Omega}\rho\mathrm{d}\Omega & \iiint_{\Omega}-\rho x\mathrm{d}\Omega & \iiint_{\Omega}\rho y\mathrm{d}\Omega\\ 0 & \iiint_{\Omega}-\rho x\mathrm{d}\Omega & \iiint_{\Omega}\rho\left(x^2+z^2\right)\mathrm{d}\Omega & \iiint_{\Omega}-\rho xy\mathrm{d}\Omega\\ 0 & \iiint_{\Omega}\rho y\mathrm{d}\Omega & \iiint_{\Omega}-\rho xy\mathrm{d}\Omega & \iiint_{\Omega}\rho\left(y^2+z^2\right)\mathrm{d}\Omega\end{bmatrix} \tag{A1}$$

式中: 矩阵 $\boldsymbol{M}$ 中第 1 个元素的体积分域为内舵, 其他项的体积分域为外舵, 对于复杂工程结构对象可以通过有限元软件计算。

气动刚度矩阵 $\boldsymbol{Q}_{\mathrm{k}}$ 表达式为

$$\boldsymbol{Q}_{\mathrm{k}}=-\frac{\rho u}{\sqrt{Ma^2-1}}\cdot\begin{bmatrix}\iint_S(-n_3u_1+n_1u_3)(zn_1-xn_3)\mathrm{d}S & 0 & 0 & 0\\ 0 & 0 & \iint_S(-n_3u_1+n_1u_3)n_3\mathrm{d}S & \iint_S(n_3u_2-n_2u_3)n_3\mathrm{d}S\\ 0 & 0 & \iint_S(-n_3u_1+n_1u_3)(zn_1-xn_3)\mathrm{d}S & \iint_S(n_3u_2-n_2u_3)(zn_1-xn_3)\mathrm{d}S\\ 0 & 0 & \iint_S(-n_3u_1+n_1u_3)(yn_3-zn_2)\mathrm{d}S & \iint_S(n_3u_2-n_2u_3)(yn_3-zn_2)\mathrm{d}S\end{bmatrix} \tag{A2}$$

气动刚度矩阵 $\boldsymbol{Q}_{\mathrm{c}}$ 表达式为

$$\boldsymbol{Q}_{\mathrm{c}}=-\frac{\rho u\left(Ma^2-2\right)}{\left(Ma^2-1\right)^{3/2}}\cdot\begin{bmatrix}\iint_S(n_1z-n_3x)(zn_1-xn_3)\mathrm{d}S & 0 & 0 & 0\\ 0 & \iint_S(n_3)n_3\mathrm{d}S & \iint_S(zn_1-xn_3)n_3\mathrm{d}S & \iint_S(yn_3-zn_2)n_3\mathrm{d}S\\ 0 & \iint_S(n_3)(zn_1-xn_3)\mathrm{d}S & \iint_S(zn_1-xn_3)(zn_1-xn_3)\mathrm{d}S & \iint_S(yn_3-zn_2)(zn_1-xn_3)\mathrm{d}S\\ 0 & \iint_S(n_3)(yn_3-zn_2)\mathrm{d}S & \iint_S(zn_1-xn_3)(yn_3-zn_2)\mathrm{d}S & \iint_S(yn_3-zn_2)(yn_3-zn_2)\mathrm{d}S\end{bmatrix} \tag{A3}$$

非线性弹簧力向量 $\boldsymbol{F}_{\mathrm{spr}}$ 表达式为

$$
\begin{cases}
\boldsymbol{F}_{\mathrm{spr}}=\begin{bmatrix}F_{\mathrm{spr1}}\\F_{\mathrm{spr2}}\\F_{\mathrm{spr3}}\\F_{\mathrm{spr4}}\end{bmatrix}\\
F_{\mathrm{spr1}}=\begin{cases}K_{\alpha}(\alpha-\delta-\alpha_{\mathrm{s}}) & \alpha-\delta\geqslant\alpha_{\mathrm{s}}\\0 & -\alpha_{\mathrm{s}}\leqslant\alpha-\delta<\alpha_{\mathrm{s}}\\K_{\alpha}(\alpha-\delta+\alpha_{\mathrm{s}}) & \alpha-\delta<-\alpha_{\mathrm{s}}\end{cases}\\
F_{\mathrm{spr2}}=\begin{cases}-K_{h}(h-h_{\mathrm{s}}) & h\geqslant h_{\mathrm{s}}\\0 & -h_{\mathrm{s}}\leqslant h<h_{\mathrm{s}}\\-K_{h}(h+h_{\mathrm{s}}) & h<-h_{\mathrm{s}}\end{cases}\\
F_{\mathrm{spr3}}=\begin{cases}-K_{\alpha}(\alpha-\delta-\alpha_{\mathrm{s}}) & \alpha-\delta\geqslant\alpha_{\mathrm{s}}\\0 & -\alpha_{\mathrm{s}}\leqslant\alpha-\delta<\alpha_{\mathrm{s}}\\-K_{\alpha}(\alpha-\delta+\alpha_{\mathrm{s}}) & \alpha-\delta<-\alpha_{\mathrm{s}}\end{cases}\\
F_{\mathrm{spr4}}=\begin{cases}-K_{\beta}(\beta-\beta_{\mathrm{s}}) & \beta\geqslant\beta_{\mathrm{s}}\\0 & -\beta_{\mathrm{s}}\leqslant\beta<\beta_{\mathrm{s}}\\-K_{\beta}(\beta+\beta_{\mathrm{s}}) & \beta<-\beta_{\mathrm{s}}\end{cases}
\end{cases}
\tag{A4}
$$

Flutter suppression method for multi-freepalys folding fin based on sliding mode control and fin shaft drive

REN Haoyuan*, CHENG Tao, ZHANG Cheng, CAI Yipeng, LIU Fei, ZHANG Weiqun

(China Academy of Launch Vehicle Technology, Beijing 100076, China)

Abstract: The folding fin of a high-speed vehicle demonstrates nonlinear dynamic behavior during flight due to freeplays in the connection structure, and the aeroelastic active suppression technology is an important solution. Because the axis rotation displacement of the inner fin is the only input parameter for the full-motion folding fin, a new method is proposed to realize the active flutter suppression of multi-gap folding fin, by controlling the rotation of the rudder axis based on the sliding mode control. Firstly, based on the three-dimensional folding fin model and the piston theory aerodynamic method, the linear spring model is used to describe the connection stiffness of the folding mechanism, and a four-degree-of-freedom folding fin dynamic model is derived. The feasibility and control mechanism of the fin shaft drive for the flutter suppression of linear folding fin system are illustrated. Further, the connection stiffness model of the folding mechanism is changed to the clearance nonlinear spring, and the behavior of the multi-clearance nonlinear folding fin aeroelasticity response is analyzed. The flutter suppression effect of multi-clearance nonlinear folding rudder is analyzed. The results show that: for the folding fin in this paper, the plunging and flapping clearances increase the flutter speed, while the pitching clearances reduce the flutter speed; the sliding mode control can effectively improve the flutter speed of linear and clearance nonlinear folding fin. The methods and results of this paper provide an engineering solution for the design and aeroelastic control of the folding fin system.

Keywords: folding fin; free-play; aeroelasticity; sliding mode control; flutter suppression; nonlinear vibration

Received: 2023-07-14; **Accepted**: 2023-11-20; **Published Online**: 2023-12-15 12:54

URL: link.cnki.net/urlid/11.2625.V.20231214.1659.001

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0438

虑及装配中多源误差的航发转子不平衡量预测

周天一[1,*]，胡磊[2]，赵所[2]，靳潇潇[2]，王诗雨[2]，高航[3]

(1. 大连交通大学 机械工程学院，大连 116028；2. 中国航空工业集团公司 沈阳飞机设计研究所，沈阳 110035；
3. 大连理工大学 机械工程学院，大连 116024)

摘 要：针对航发转子在复杂装配系统下产品不平衡量难预测问题，提出虑及多源误差的不平衡量数值模拟预测方法，旨在实现工件组精度及其装配系统精度的可靠溯源设计。阐明了转子装配过程中工艺参数、多源误差、相对位姿及不平衡量之间的传递原理，分别建立了典型转子零件制造误差、法兰止口位姿测量误差和多自由度安装机构运动误差的量化表征模型。基于所建模型设计了转子不平衡量蒙特卡罗模拟算法，实现不同装配条件下的不平衡量预测，以低压涡轮转子模拟件为实例，通过敏感度综合分析实现了其装配系统的精度优化反求，模拟获得不平衡量合格率达99.8%，提供了可靠的精度设计依据。进行装配验证实验，结果表明：不平衡量均值相对预测结果的偏离率均<10%，所提方法具有较高准确度。

关 键 词：航空发动机；转子；装配；不平衡量；蒙特卡罗；精度设计
中图分类号：V232.2
文献标志码：A **文章编号**：1001-5965(2025)09-3001-10

航空发动机转子的不平衡量对整机的气动特性、磨损状况和振动水平等性能造成重要影响[1-4]，为制造行业内的严控指标。装配作为转子制造过程中的关键环节，涉及组件众多，且工艺复杂[5-6]，多种高随机性的误差源均对不平衡量造成影响，如零件自身的制造误差[7-8]、组件初始位姿测量误差[9-10]及装配运动误差[11-12]等，导致装配体的不平衡量难控。尤其在现装配行业内，逐渐采用自动化机械系统代替人工进行组件的测量、调姿和安装[13]，导致新设计的专用系统对于不同精度等级零件的精度适用性难以预知。因此，有必要对影响不平衡量的多源误差建模方法进行研究，并形成一套拟合实际工况的装配模拟算法，实现不同输入条件下的不平衡量预测，为装配系统的精度设计提供反求依据。

近年来，国内外学者针对装配过程中主要误差源的建模方法作出了相关研究：①对于零件自身制造误差建模，Whitney 等采用齐次变换矩阵实现零件形位误差的表征[14]，Ding 等针对转子制造误差建立了数值模型[15]。然而，装配前转子零件的配套涉及齐套性，从合格组中随机选件，现仍缺少一个通用的统计学模型，满足对装配后与不平衡量相关联的各制造误差项的统一表征。②对于装配前的测量误差建模，齐次转换矩阵同样具有适用性，Zhou 等通过齐次变换矩阵对回转体的特征测点进行表征[16]，孙汕民等建立了融合测点偏移、测件偏心量及偏角的转子圆轮廓测量误差模型[17]。然而，现仍缺少专门针对转子法兰止口等特征结构的测量位姿误差模型。③对于装配运动误差建模，目前多采用矩阵法实现高自由度机构的复杂误差建模，Yang 等用微分矩阵对一种五自由度机构进行了误差建模[18]，Zhou 等基于多参数 D-H 方法对一种六自由度机构进行误差分析[19]。而对于用作转子零件调姿安装的多自由度运动机构，当装夹零件的装配轨迹给定时，模型需根据新产生的微分矩阵对运

收稿日期：2023-07-04；**录用日期**：2023-09-03；**网络出版时间**：2024-01-05 21:01
网络出版地址：link.cnki.net/urlid/11.2625.V.20240105.1321.003
基金项目：国家自然科学基金 (51621064)
* **通信作者**. E-mail：ztydlut@foxmail.com

引用格式：周天一，胡磊，赵所，等. 虑及装配中多源误差的航发转子不平衡量预测 [J]. 北京航空航天大学学报，2025，51（9）：3001-3010.
ZHOU T Y，HU L，ZHAO S，et al. Prediction of aeroengine rotor unbalance considering multi-source errors in assembly process [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（9）：3001-3010（in Chinese）.

动误差进行实时预估，同时，模型需便于在复杂数值模拟算法中进行集成和解算。

基于上述误差建模框架，国内外学者开展了多级零件装配过程中的不平衡量预测方法研究，Hussain 等开展了多种装配工艺方法下航发转子装配误差的数值分析[20]，在此基础上，Memon 等建立了转子装配误差与装配序列之间的关联模型[21]，Bolotov 等建立了低压涡轮转子不平衡量与转子跳动误差、对接周向角和装配几何误差之间的数值关联模型[22]，Qureshi 和 Lee 等分别基于不同统计学模型开发了虑及多零件尺寸误差的装配精度模拟程序[23-24]，胡明模拟了大规模装配生产过程中的转子不平衡量[25]。上述方法在以转子零件自身制造误差为单一误差源的条件下，取得了较好的预测效果。然而，现代航发转子不平衡量要求更为严格，而新型自动化测控系统的应用导致不平衡量的影响因素多样且复杂。因此，有必要将多主因误差模型集成进同一套模拟算法中，使不平衡量预测可以虑及多因素的同步影响。同时，新型装配系统下的装配模拟算法需与实际高随机性工况和复杂工艺流程相匹配，确保预测精度。

为此，本文开展虑及装配中多源误差的航发转子不平衡量预测研究。首先，分析影响不平衡量的主因误差，阐明误差传递规律，提出各误差源的普适性建模方法。然后，基于误差模型，设计不平衡量的数值模拟算法。最后，以低压涡轮转子模拟件为实例，分析不同装配条件下的误差敏感度，提出装配系统中各关联精度项的优化设计方案，通过装配实验对本文方法的预测精度进行验证。

1 转子装配过程中的误差传递原理

在航发转子装配过程中，由于零件批产精度随机性影响，以及复杂自动化测量与安装等运动执行系统的应用，引入了对装配体不平衡量产生影响的不同类型误差源，阐明其传递原理，并对主因误差源进行量化表征，是误差溯源分析和不平衡量预测的必要前提。所述的主要误差源包括：①各转子零件连接与定位等结构处的形位误差，其对相邻两零件间的相对位姿偏差产生影响；②各转子零件自身的不平衡量，由于材料质量分布不均等影响，导致单个零件自身的重心相对名义转轴存在偏离；③测量零件位姿等信息时的测量误差，由于测量点云坐标的示值误差，造成拟合的几何特征相对实际位姿产生偏移和偏转，通过含有测量位姿误差信息而计算出来的装配轨迹偏离于正确轨迹；④调姿安装机构的运动误差，用于调整零件位姿和进行对接运动的多自由度机构含有机械误差，造成零件执行装配轨迹运动后的终位姿不准确。误差分类及传递示意图如图 1 所示。

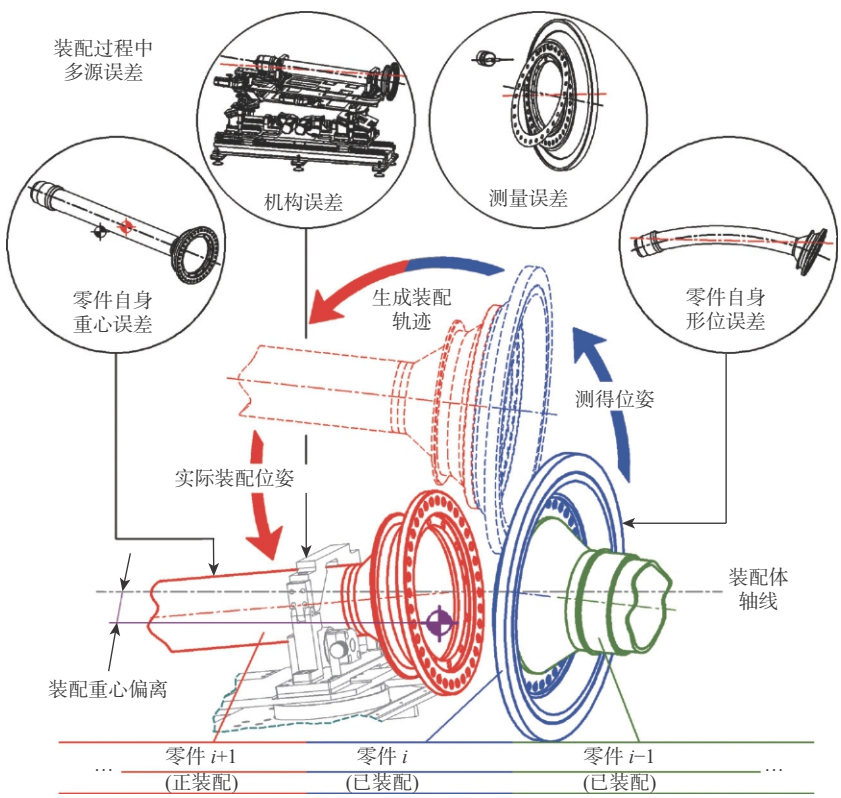


图 1 航发转子装配过程中误差分类和传递

Fig. 1 Error classification and transmission during aeroengine rotor assembly process

通过齐次坐标变换位姿表征矩阵和误差转换矩阵, 对装配过程中的误差传递过程进行描述。考虑第 i 个转子零件自身不平衡量的影响, 其重心相对于全局基准坐标系 O_0xyz 的位置向量可由 ${}^{0}_{Gi}\boldsymbol{P}$ 表述:

$$ {}^{0}_{Gi}\boldsymbol{P} = {}_{RS1}{}^{0}\boldsymbol{T}\left(\prod_{i=1}^{i-1} {}^{RSi}_{FSi}\boldsymbol{T}\cdot {}^{FSi}_{RSi+1}\boldsymbol{T}\right) {}^{RSi}_{GNi}\boldsymbol{T}\cdot {}^{GNi}_{Gi}\boldsymbol{D}_{(gra)}\cdot[0\ \ 0\ \ 0\ \ 1]^{T} \tag{1} $$

式中: ${}^{A}_{B}\boldsymbol{T}$ 和 ${}^{A}_{B}\boldsymbol{D}$ 分别为坐标系 O_Bxyz 相对 O_Axyz 位姿的齐次位姿变换矩阵和误差转换矩阵; i 为转子零件序数; 左角标 R、F、S、G 和 N 分别为零件后端、零件前端、连接止口、重心和名义位置的符号。

考虑形位误差的影响, 在单个零件上的误差传递为

$$ {}^{RSi}_{FSi}\boldsymbol{T} = {}^{RSNi}_{FSNi}\boldsymbol{T}\cdot {}^{RSi}_{FSi}\boldsymbol{D}_{(sha)} \tag{2} $$

式 (1) 中的 ${}_{RSi+1}{}^{FSi}\boldsymbol{T}$ 表示装配过程中相邻转子零件前后止口间的位姿变换, 分析其受测量误差和运动机构误差的影响。经过测量和拟合后的后端止口坐标系 $O_{FSi}xyz$ 相对测量坐标系 O_Mxyz 的位姿变换矩阵为

$$ {}^{M}_{FSi}\boldsymbol{T}_{(fit)} = {}_{RS1}{}^{0}\boldsymbol{T}\cdot {}^{RS1}_{FS1}\boldsymbol{T}\left(\prod_{i=1}^{i-1} {}^{FSi}_{RSi+1}\boldsymbol{T}\cdot {}^{RSi+1}_{FSi+1}\boldsymbol{T}\right)\cdot {}^{0}_{M}\boldsymbol{T}^{-1}\cdot {}^{M}_{FSi}\boldsymbol{D}_{(mea)} \tag{3} $$

引入周向对接相位角 ϕ_i 和防撞间隙 Δ_A, 得到第 i+1 个零件止口结构的目标矩阵 ${}_{RSi+1}{}^{0}\boldsymbol{T}_{(tar)}$ 用以控制装配轨迹的运动指令生成:

$$ {}_{RSi+1}{}^{0}\boldsymbol{T}_{(tar)} = {}^{M}_{FSi}\boldsymbol{T}_{(fit)}\cdot {}^{0}_{M}\boldsymbol{T}\cdot \mathbf{Rot}(O_{FSi\text{-}z},\phi_i)\cdot \mathbf{Trans}(0,0,-\Delta_A) \tag{4} $$

式中: **Rot** 和 **Trans** 分别为旋转和平移的齐次变换矩阵。

同时, 因调姿安装机构的运动误差影响, 机构装夹的第 i+1 个转子零件相对其目标位姿产生了偏离, 即

$$ {}_{RSi+1}{}^{0}\boldsymbol{T} = {}_{RSi+1}{}^{0}\boldsymbol{T}_{(tar)}\cdot {}_{RSi+1}{}^{0}\boldsymbol{D}_{(mec)} \tag{5} $$

设转子装配体的几何转轴由第 i 个和第 j 个转子零件的轴承坐标系 $O_{Bi/j}xyz$ 决定, 则第 k 个转子零件的重心相对几何转轴的偏离向量 ${}^{BiBj}_{Gk}\boldsymbol{P}$ 表达式为

$$ {}^{BiBj}_{Gk}\boldsymbol{P} = \begin{bmatrix} {}^{0}_{Gk}p_x - \left({}^{0}_{Bj}p_x - {}^{0}_{Bi}p_x\right)\lambda - {}^{0}_{Bi}p_x \\ {}^{0}_{Gk}p_y - \left({}^{0}_{Bj}p_y - {}^{0}_{Bi}p_y\right)\lambda - {}^{0}_{Bi}p_y \\ {}^{0}_{Gk}p_z - \left({}^{0}_{Bj}p_z - {}^{0}_{Bi}p_z\right)\lambda - {}^{0}_{Bi}p_z \end{bmatrix} \tag{6} $$

$$ \begin{aligned} \lambda = &\left[\left({}^{0}_{Bj}p_x - {}^{0}_{Bi}p_x\right)\left({}^{0}_{Gk}p_x - {}^{0}_{Bi}p_x\right) + \left({}^{0}_{Bj}p_y - {}^{0}_{Bi}p_y\right)\cdot\right. \\ &\left.\left({}^{0}_{Gk}p_y - {}^{0}_{Bi}p_y\right) + \left({}^{0}_{Bj}p_z - {}^{0}_{Bi}p_z\right)\left({}^{0}_{Gk}p_z - {}^{0}_{Bi}p_z\right)\right] / \\ &\left[\left({}^{0}_{Bj}p_x - {}^{0}_{Bi}p_x\right)^2 + \left({}^{0}_{Bj}p_y - {}^{0}_{Bi}p_y\right)^2 + \left({}^{0}_{Bj}p_z - {}^{0}_{Bi}p_z\right)^2\right] \end{aligned} \tag{7} $$

多级零件装配后的重心相对几何轴的偏离, 如图 2 所示。

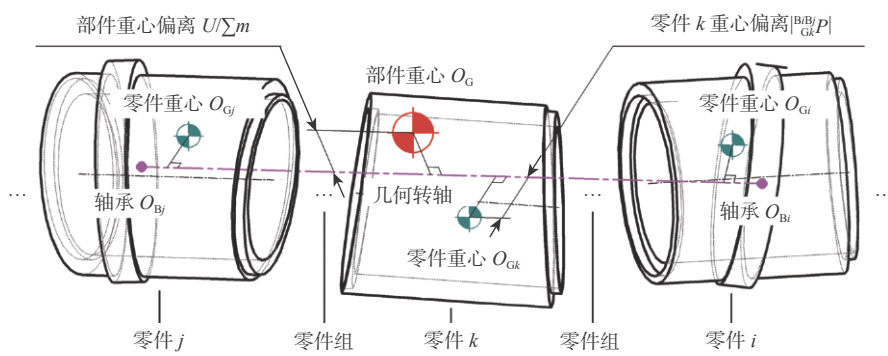


图 2　多级零件装配后的重心相对几何轴的偏离

Fig. 2　Deviation between centers of gravity of assembled parts and geometric axis

则转子装配体的不平衡量 U 可表示为

$$ U = \left|\sum_{k=1}^{N} m_k\cdot {}^{BiBj}_{Gk}\boldsymbol{p}\right| \tag{8} $$

式中: m_k 为第 k 个零件的质量。

2　不平衡量的多源误差统计学建模

分别建立表征零件制造误差、测量误差和装配误差的通用统计学数值模型, 具有适用广、量化表征清晰和便于在数值算法中运算的特点, 适应于不同的装配工件、工艺和工装。

2.1　零件制造误差建模

首先, 考虑转子零件形位误差的影响。前后端面跳动、前后连接止口同心度及轴承安装面同轴度分别影响了相邻零件间的轴向定位与偏转定向、径向定位及装配体几何转轴的位置, 在各个几何特征的名义位置和含误差的实际位置处分别建立空间坐标系, 如图 3 所示。

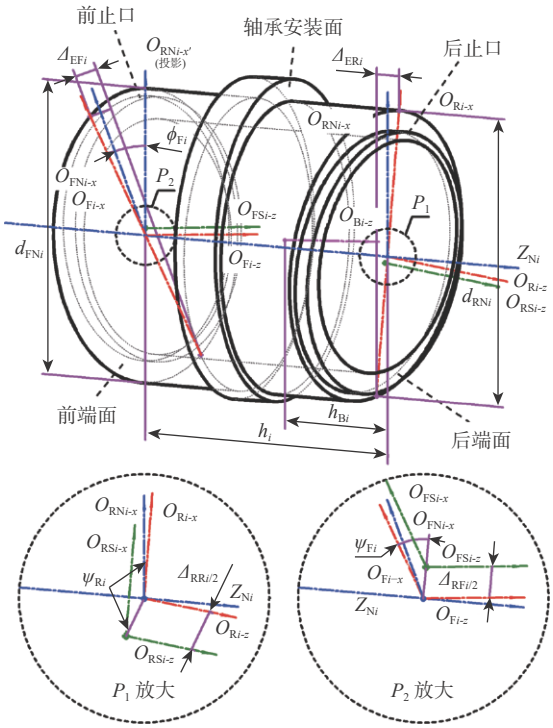


图3 转子零件形位误差示意图

Fig. 3 Schematic of rotor part shape and position errors

通过齐次位姿变换矩阵对零件的各个几何特征的形位误差进行表征：

$$ {}^{0}_{\mathrm{R1}}\boldsymbol{T} = {}^{0}_{\mathrm{RN1}}\boldsymbol{T}\cdot\mathbf{Trans}\left({}^{\mathrm{RN1}}_{\mathrm{R1}}p_x,{}^{\mathrm{RN1}}_{\mathrm{R1}}p_y,{}^{\mathrm{RN1}}_{\mathrm{R1}}p_z\right)\cdot \mathbf{RPY}\left(\phi_{\mathrm{R1}},\theta_{\mathrm{R1}},\varphi_{\mathrm{R1}}\right) \tag{9} $$

$$ {}^{\mathrm{RN}i}_{\mathrm{R}i}\boldsymbol{T} = \mathbf{Rot}\left[O_{\mathrm{RN}i\text{-}y},-\arctan\left(\frac{\Delta_{\mathrm{ER}i}}{d_{\mathrm{RN}i}}\right)\right] \tag{10} $$

$$ {}^{\mathrm{F/R}i}_{\mathrm{F/RS}i}\boldsymbol{T} = \mathbf{Trans}\left(\frac{\Delta_{\mathrm{RF/R}i}\cos\psi_{\mathrm{F/R}i}}{2},\frac{\Delta_{\mathrm{RF/R}i}\sin\psi_{\mathrm{F/R}i}}{2},0\right) \tag{11} $$

$$ {}^{\mathrm{RN}i}_{\mathrm{FN}i}\boldsymbol{T} = \mathbf{Trans}(0,0,-h_{\mathrm{N}i})\mathbf{Rot}(O_{\mathrm{FN}i\text{-}z},\phi_{\mathrm{F}i}) \tag{12} $$

$$ {}^{\mathrm{FN}i}_{\mathrm{F}i}\boldsymbol{T} = \mathbf{Trans}(0,0,h_{\mathrm{N}i}-h_i)\mathbf{Rot}\left(O_{\mathrm{F}i\text{-}y},\arctan\frac{\Delta_{\mathrm{EF}i}}{d_{\mathrm{FN}i}}\right) \tag{13} $$

$$ {}^{\mathrm{RN}i}_{\mathrm{B}i}\boldsymbol{T} = \mathbf{Trans}\left(\frac{\Delta_{\mathrm{B}i}\cos\psi_{\mathrm{B}i}}{2},\frac{\Delta_{\mathrm{B}i}\sin\psi_{\mathrm{B}i}}{2},-h_{\mathrm{B}i}\right) \tag{14} $$

式中：**RPY** 为空间旋转齐次变换矩阵；$\Delta_{\mathrm{EF}i}$和$\Delta_{\mathrm{ER}i}$分别为第 i 个转子零件前端面和后端面的全跳动误差，$\phi_{\mathrm{F}i}$为前、后端面的全跳动高点周向相位角；$\Delta_{\mathrm{RF}i}$、$\Delta_{\mathrm{RR}i}$和$\Delta_{\mathrm{B}i}$分别为第 i 个转子零件前、后止口和轴承安装面的同心度，$\psi_{\mathrm{F}i}$、$\psi_{\mathrm{R}i}$和$\psi_{\mathrm{B}i}$分别为对应同心度误差方向角。

同时，考虑制造误差导致零件质量分布不均的影响，重心与形心间存在偏离：

$$ {}^{\mathrm{GN}i}_{\mathrm{G}i}\boldsymbol{T} = \mathbf{Trans}\left(\frac{U_i\cos\psi_{\mathrm{G}i}}{m_i},\frac{U_i\sin\psi_{\mathrm{G}i}}{m_i},\frac{U_i\tan\theta_{\mathrm{G}i}}{m_i}\right) \tag{15} $$

式中：U_i为第 i 个转子零件的不平衡量，$\psi_{\mathrm{G}i}$、$\theta_{\mathrm{G}i}$为重心方位角。

将误差矩阵中元素与对应设计公差进行关联。其中的系统误差可在制造阶段就进行预补偿，而随机误差多数条件下是由很多相互独立的随机因素所引起，偏差数值可近似为服从正态分布。为更好地表示设计公差与实际误差的关系，分别引入影响形位误差和不平衡量的工艺能力系数 C_{PS} 和 C_{PU}，满足式 (16) 和式 (17)，当工艺能力系数 C_{P}=1.00、1.33 和 1.67 时，分别对应工艺不合格率 0.27%、0.003 2%、0.000 029%。

$$ C_{\mathrm{PS}} = \frac{T_{\mathrm{S}}}{6\sigma_{\mathrm{S}}} \tag{16} $$

$$ C_{\mathrm{PU}} = \frac{U_{\max}}{3\sigma_{\mathrm{U}}} \tag{17} $$

式中：T_{S}为形位误差的设计公差带宽度；$U_{\max}$为不平衡量的设计上限值；σ_{S}和σ_{U}分别为形位误差和不平衡量的标准差。

2.2 测量误差建模

装配过程中的测量对象为相邻两零件的初始位姿${}^{\mathrm{M}}_{\mathrm{FS}i}\boldsymbol{T}$和${}^{\mathrm{M}}_{\mathrm{RS}i+1}\boldsymbol{T}$，包括 2 类待测几何特征：①零件连接法兰的环形端面，其法向量用于计算位姿变换矩阵中的第 2 列姿态列向量${}^{\mathrm{M}}\boldsymbol{a}$；②止口和基准连接孔的圆柱面，相对端面映射变换得到的中心坐标用于计算第 1 列和第 3 列姿态列向量${}^{\mathrm{M}}\boldsymbol{n}$和${}^{\mathrm{M}}\boldsymbol{o}$、第 4 列位置列向量${}^{\mathrm{M}}\boldsymbol{p}$。受多种相互独立的误差因素影响，则各测量坐标轴 O_{M}-x/y/z 的测量示值误差服从以${}^{\mathrm{M}}p_{x/y/z}+\Delta_{\mathrm{MS}x/y/z}$为数学期望 μ，以$\Delta_{\mathrm{MR}x/y/z}/6$为标准差 σ 的正态分布，其中，$\Delta_{\mathrm{MS}x/y/z}$和$\Delta_{\mathrm{MR}x/y/z}$分别为各轴方向上的测量系统误差和测量随机误差。

当 N 个测点均布于 J 个同心圆测量路径时，如图 4(a) 所示，则第 i 个零件的前端法兰端面上第 j 个同心环的第 k 个测点 M_{jk} 的名义和示值坐标分别为

$$ \left[{}^{\mathrm{M}}\boldsymbol{p}^{\mathrm{T}}_{M_{jk}}\quad 1\right]^{\mathrm{T}} = {}^{\mathrm{M}}_{\mathrm{F}i}\boldsymbol{T}\cdot\left\{\mathbf{Rot}(O_{\mathrm{F}i\text{-}z},\theta_k)\cdot\left[{}^{\mathrm{F}i}\boldsymbol{p}^{\mathrm{T}}_{M_{j1}}\quad 1\right]^{\mathrm{T}}\right\} \tag{18} $$

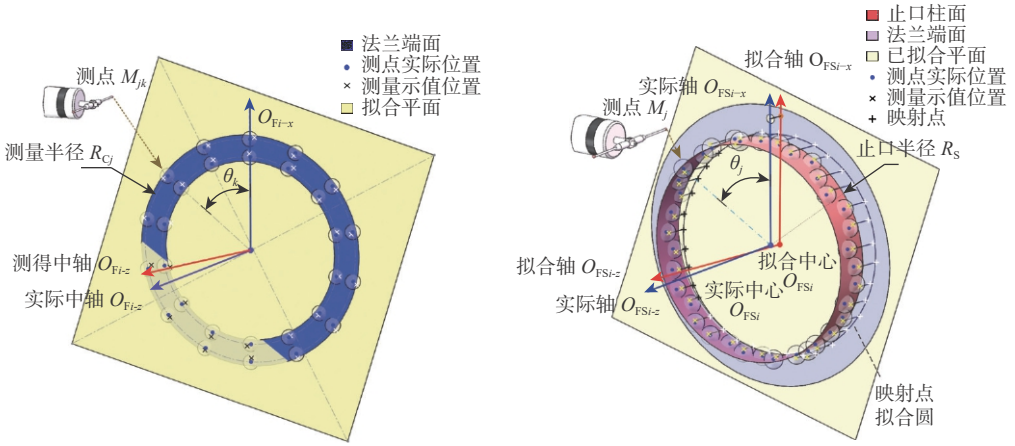
$$ {}^{\mathrm{M}}p_{x/y/zjk(\mathrm{mea})} \sim N\left[{}^{\mathrm{M}}p_{x/y/zjk}+\Delta_{\mathrm{MS}x/y/z},\left(\Delta_{\mathrm{MR}x/y/z}/6\right)^2\right] \tag{19} $$

拟合后，得到端面方程${}^{\mathrm{M}}p_z = a_1 + a_2{}^{\mathrm{M}}p_x + a_3{}^{\mathrm{M}}p_y$，则包含测量误差的拟合位姿矩阵${}^{\mathrm{M}}_{\mathrm{F}i}\boldsymbol{T}$中元素为

$$ {}^{\mathrm{M}}_{\mathrm{F}i}\boldsymbol{a}^{\mathrm{T}}_{(\mathrm{fit})} = \begin{bmatrix}-k_a a_2 & -k_a a_3 & k_a\end{bmatrix} \tag{20} $$

$$ k_a = 1/\sqrt{a_2^2+a_3^2+1} \tag{21} $$

同样地，如图 4(b) 所示，第 i 个零件的前端止口的第 j 个测点 M_j 的名义和示值坐标分别为



(a) 法兰环形端面的测量误差　　(b) 连接止口的测量误差

图 4　转子零件的位姿测量误差示意图

Fig. 4　Schematic of pose measurement errors of rotor parts

$$\left[{}^{\mathrm{M}}\boldsymbol{p}_{M_j}^{\mathrm{T}} \quad 1\right]^{\mathrm{T}} = {}^{\mathrm{M}}_{\mathrm{FS}i}\boldsymbol{T}\cdot\left\{\mathbf{Trans}\left(0,0,h_{\mathrm{M}_j}\right)\cdot \mathbf{Rot}\left(O_{\mathrm{FS}i\text{-}z},\theta_j\right)\cdot\left[{}^{\mathrm{FS}i}\boldsymbol{p}_{M_1}^{\mathrm{T}} \quad 1\right]\right\} \tag{22}$$

$${}^{\mathrm{M}}p_{x/y/z\,j(\mathrm{mea})} \sim N\left[{}^{\mathrm{M}}p_{x/y/z\,j}+\Delta_{\mathrm{MS}x/y/z},\left(\Delta_{\mathrm{MR}x/y/z}/6\right)^2\right] \tag{23}$$

映射与拟合后, 得到投影在平面 $O_{\mathrm{F}i}$-$xy_{(\mathrm{fit})}$ 上的空间圆的另一个联立方程为 ${}^{\mathrm{M}}p_z = \left[b_4^2-\left({}^{\mathrm{M}}p_x-b_1\right)^2-\left({}^{\mathrm{M}}p_y-b_2\right)^2\right]^{1/2}+b_3$, 则对应位姿描述矩阵 ${}^{\mathrm{M}}\boldsymbol{T}_{\mathrm{FS}i}$ 中位置向量元素为

$${}^{\mathrm{M}}_{\mathrm{FS}i}\boldsymbol{p}_{(\mathrm{fit})} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}^{\mathrm{T}} \tag{24}$$

对端跳高点处对应的连接孔作类似运算, 得到连接孔圆心坐标$(b_{\mathrm{H1}},b_{\mathrm{H2}},b_{\mathrm{H3}})$, 最终得到拟合的包含测量误差的连接止口坐标系的位姿变换矩阵为

$${}^{\mathrm{M}}_{\mathrm{FS}i}\boldsymbol{T}_{(\mathrm{fit})} = \begin{bmatrix} k_n(b_{\mathrm{H1}}-b_1) & k_nk_a(-b_{\mathrm{H2}}+b_2-a_3b_{\mathrm{H3}}+a_3b_3) & -k_aa_2 & b_1 \\ k_n(b_{\mathrm{H2}}-b_2) & k_nk_a(-b_1+b_{\mathrm{H1}}+a_2b_{\mathrm{H3}}-a_2b_3) & -k_aa_3 & b_2 \\ k_n(b_{\mathrm{H3}}-b_3) & k_nk_a(-a_3b_1+a_3b_{\mathrm{H1}}-a_2b_{\mathrm{H2}}+a_2b_2) & k_a & b_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{25}$$

$$k_n = 1/\sqrt{(b_{\mathrm{H1}}-b_1)^2+(b_{\mathrm{H2}}-b_2)^2+(b_{\mathrm{H3}}-b_3)^2} \tag{26}$$

2.3　装配运动误差建模

2.3.1　调姿安装机构的运动误差

将调姿安装机构的末端误差看作由各相邻运动部件间的误差传递与合成。首先, 通过多参数 D-H 方法对多自由度机构的运动学进行解析, 则运动学正解如式 (27) 和式 (28) 所示, 并可通过代数封闭法计算运动学逆解。然后, 通过微分变换进行任意位姿条件下的综合误差计算, 对于具有 m 个自由度的调姿对接机构, 其末端坐标系 O_mxyz 的位姿误差矩阵如式 (29) 所示。在此基础上, 转子零件相对于末端夹具含安装误差的相对位姿为 ${}_{\mathrm{RS}i+1}{}^{m}_{}\boldsymbol{T}$, 最终转子零件运动后的后端连接止口坐标系 $O_{\mathrm{RS}i+1}xyz$ 相对基准坐标系 O_0xyz 的实际位姿如式 (30) 所示。

$${}^{i-1}_{i}\boldsymbol{T} = \mathbf{Rot}\left(O_{i\text{-}1\text{-}z},\theta_i\right)\mathbf{Trans}\left(0,0,d_i\right)\mathbf{Trans}\left(a_i,0,0\right)\cdot \mathbf{Rot}\left(O_{i\text{-}x},\alpha_i\right)\mathbf{Rot}\left(O_{i\text{-}y},\beta_i\right) \tag{27}$$

$${}^{j}_{k}\boldsymbol{T} = \prod_{i=j+1}^{k}{}^{i-1}_{i}\boldsymbol{T} \tag{28}$$

$$\Delta^0_m\boldsymbol{T} = \sum_{e=a,\alpha,\theta,d,\beta}\left\{\sum_{i=1}^{m}\left[\left(\prod_{k=1}^{i-1}{}^{k-1}_{k}\boldsymbol{T}\right)\frac{\partial\,{}^{i-1}_{i}\boldsymbol{T}}{\partial s_i}\left(\prod_{k=i+1}^{m}{}^{k-1}_{k}\boldsymbol{T}\right)\Delta e_i\right]\right\} \tag{29}$$

$${}_{\mathrm{RS}i+1}{}^{0}\boldsymbol{T} = \left({}^{0}_{m}\boldsymbol{T}+\Delta^0_m\boldsymbol{T}\right)\cdot{}_{\mathrm{RS}i+1}{}^{m}\boldsymbol{T} \tag{30}$$

式中: Δe_i 为调姿安装机构中 D-H 参数的误差。

多自由度调姿安装机构的末端综合位姿误差受项数众多的部件间运动误差和静态误差影响, 其中, 运动误差为主要影响因素, 因此, 可将各运动部件间误差表示为项数更少的 D-H 运动变量的误差列阵 $\mathrm{d}\boldsymbol{q}$, 通过表征 $\mathrm{d}\boldsymbol{q}$ 相对末端误差之间转换的绝对 Jacobian 矩阵 $\boldsymbol{J}$ 的计算, 实现运动误差的简化求解:

$$\mathrm{d}^0_m\boldsymbol{p}_{\mathrm{p}} = \begin{bmatrix} {}^{0}_{m}\boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{0} & {}^{0}_{m}\boldsymbol{R} \end{bmatrix}\boldsymbol{J}\,\mathrm{d}\boldsymbol{q} \tag{31}$$

$$\boldsymbol{J} = \partial\sum_{i=1}^{m}\left[\begin{pmatrix} {}^{i-1}_{m}\boldsymbol{R}^{\mathrm{T}} & -{}^{i-1}_{m}\boldsymbol{R}^{\mathrm{T}}\cdot\boldsymbol{S}\left({}^{i-1}_{m}\boldsymbol{p}\right) \\ \boldsymbol{0} & {}^{i-1}_{m}\boldsymbol{R}^{\mathrm{T}} \end{pmatrix}\mathrm{d}^{i-1}\boldsymbol{p}_{\mathrm{p}}\right]/\partial\boldsymbol{q} \tag{32}$$

式中: $\boldsymbol{R}$ 和 $\boldsymbol{P}$ 分别为对应位姿变换矩阵 $\boldsymbol{T}$ 中的姿态和位置子矩阵; $\boldsymbol{S}$ 为反对称阵; $\boldsymbol{p}_{\mathrm{p}}$ 为六自由度位姿

列阵。

在运动误差中，系统误差项可以通过机械精度调整或数控系统内进行直接补偿，补偿后的误差通常只受相互独立的随机误差因素影响，因此，D-H 运动变量误差项 $\mathrm{d}q_i$ 服从正态分布，经采样后的 $\mathrm{d}q_i$ 可作为机构运动误差模型的输入项。

2.3.2 对接相位角

在相邻 2 个转子零件的对接工艺中，有 3 种常用的周向对接方式：①随机对接，在未标定对接端面的端跳高低点条件下，同时为提高对接效率，式 (4) 中的对接相位角 ϕ_i 服从均匀分布 (0～2π)；②高低点配位对接，为更高的装配几何精度，零件的端跳高点朝向另一零件的端跳低点对接，即 ϕ_i 被固定为 π；③全局优化对接，为尽可能地减少各零件的形心与重心相对基准轴的综合偏差，通过特定优化算法获取各 ϕ_i 的最优解。

3 不平衡量蒙特卡罗模拟算法设计

3.1 模拟算法流程

基于蒙特卡罗框架对装配过程中的多源误差模型进行集成与运算，在算法中，初始装配条件参数、系统误差和根据精度等级采样的随机误差作为输入，随后大量的转子仿照实际工艺流程进行虚拟装配，最终对转子装配体不平衡量的统计学分布结果进行输出，用以对不同输入条件进行分析和反求精度，设计的算法流程如图 5 所示。

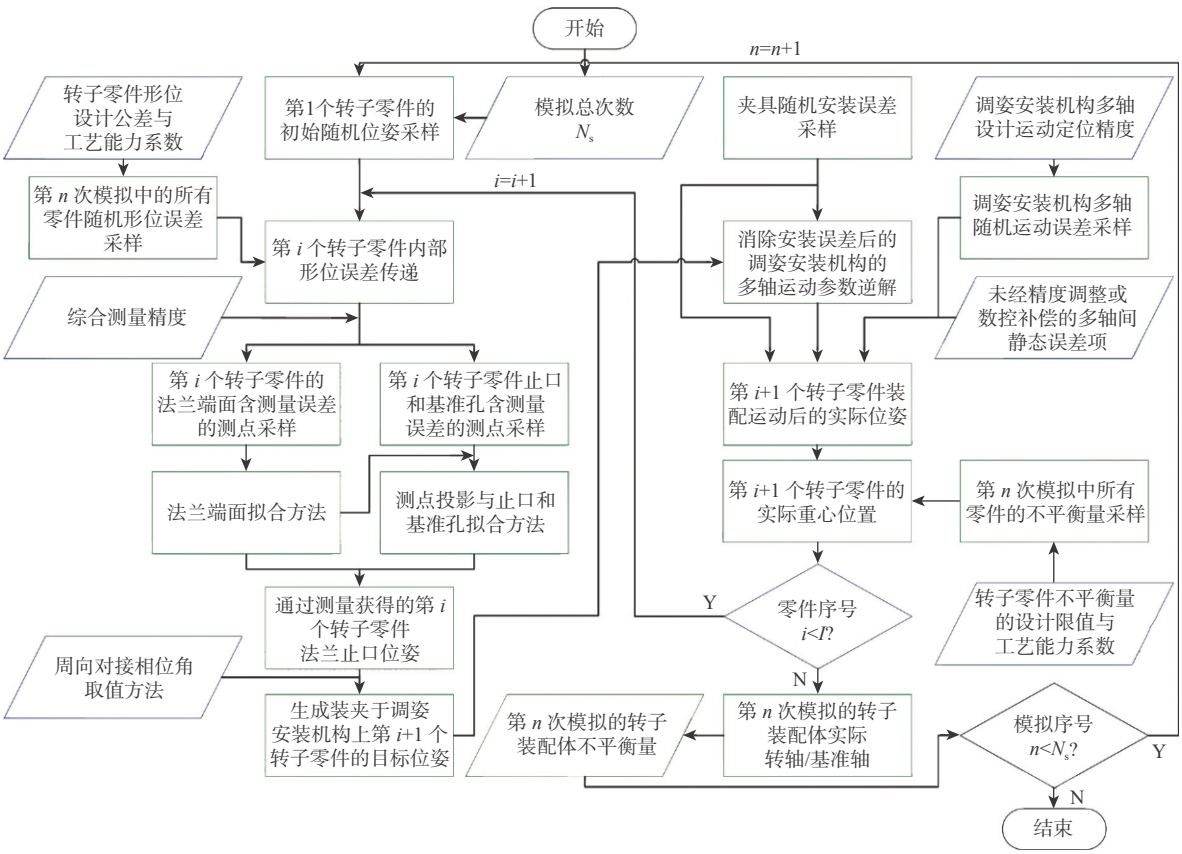


图 5 转子不平衡量蒙特卡罗模拟算法流程

Fig. 5 Monte Carlo simulation algorithm flow of rotor unbalance

3.2 主要输入条件

以某型商用飞机低压涡轮转子模拟件作为模拟对象，其由后轴 (i=1)、锥形盘 (i=2) 和低压涡轮轴 (i=3) 组成。模拟算法的主要输入项包括：

1) 不同精度等级下的误差采样：①零件形位误差：如表 1 所示，工艺能力系数 C_{PS} 由低到高精度分别设为 1.00、1.33 和 1.67；②零件不平衡量：如表 2 所示，工艺能力系数 C_{PU} 由低到高精度分别设为 1.00、1.33 和 1.67；③测量误差：三轴综合测量精度 Δ_{MR} 由低到高精度分别设为 0.005, 0.002, 0.001 mm；④调姿安装机构的各轴运动误差：对应于低、中和高精度等级，分别设定 6 个位置和姿态自由度的 D-H 运动变量误差的上限值 $\max(\Delta q_{x/y/z})$ 为 0.05, 0.02, 0.01 mm, $\max(\Delta q_{\alpha/\beta/\gamma})$ 为 0.3′、0.2′和 0.1′；⑤对接相位角：采用“随机对接”、“高低点配位对接”和“全局优化对接”方法对对接相位角 ϕ_i 进行取值，分别对应于低、中和高的对接工艺精度。

2) 其他随机采样项：零件初始位姿、夹具安装误差等。

表 1　转子零件形位误差输入

Table 1　Input of rotor part shape and position errors

形位公差类别	公差值/mm	误差值的分布	误差方向角的分布
后轴后端面全跳动	0.03	Δ_{ER1}服从 $N_+[0,(0.01/C_{PS})^2]$	
后轴前端面全跳动	0.03	Δ_{EF1}服从 $N_+[0,(0.01/C_{PS})^2]$	ϕ_{F1}服从 $R(0,2\pi)$
后轴前端止口同心度	ϕ0.03	Δ_{RF1}服从 $N[0,(0.005/C_{PS})^2]$	ψ_{F1}服从 $R(0,\pi)$
5支点轴承安装面同轴度	ϕ0.02	Δ_{B1}服从 $N[0,(0.003\ 3/C_{PS})^2]$	ψ_{B1}服从 $R(0,2\pi)$
锥形盘后端面全跳动	0.04	Δ_{ER2}服从 $N_+[0,(0.013\ 3/C_{PS})^2]$	
锥形盘后端止口同心度	ϕ0.03	Δ_{RR2}服从 $N[0,(0.005/C_{PS})^2]$	ψ_{R2}服从 $R(0,\pi)$
锥形盘前端面全跳动	0.04	Δ_{EF2}服从 $N_+[0,(0.013\ 3/C_{PS})^2]$	ϕ_{F2}服从 $R(0,2\pi)$
锥形盘前端止口同心度	ϕ0.03	Δ_{RF2}服从 $N[0,(0.005/C_{PS})^2]$	ψ_{F2}服从 $R(0,\pi)$
低压涡轮轴后端面全跳动	0.04	Δ_{ER3}服从 $N_+[0,(0.013\ 3/C_{PS})^2]$	
低压涡轮轴后端止口同心度	ϕ0.03	Δ_{RR3}服从 $N[0,(0.005/C_{PS})^2]$	ψ_{R3}服从 $R(0,\pi)$
套齿联轴器安装面同心度	ϕ0.02	Δ_{B3}服从 $N[0,(0.003\ 3/C_{PS})^2]$	ψ_{B3}服从 $R(0,2\pi)$

表 2　转子零件不平衡量输入

Table 2　Input of rotor part unbalance

转子零件	不平衡量设计上限/ (g·mm)	对应矩阵元素及其分布
后轴	100	U_1服从 $N_+[0,(33.3/C_{PU})^2]$
锥形盘	250	U_2服从 $N_+[0,(83.3/C_{PU})^2]$
低压涡轮轴	500	U_3服从 $N_+[0,(166.7/C_{PU})^2]$

4　结果与分析

4.1　不平衡量预测与精度反求

本文算例共 243 种输入条件组合, 每种条件均经过 N_S=1×10^5 次模拟, 并对装配体不平衡量 U 进行输出。其中, 当零件制造工艺能力系数 C_{PS} 与 C_{PU}、综合测量精度Δ_{MR}、调姿安装机构各自由度 D-H 运动变量偏差的上限值max (Δq_i)和对接相位角ϕ_i分别以单一变量形式逐级变化时, 其他输入条件保持中精度等级作为基准, 模拟得到 U 的频数分布与累积分布曲线如图 6 所示。

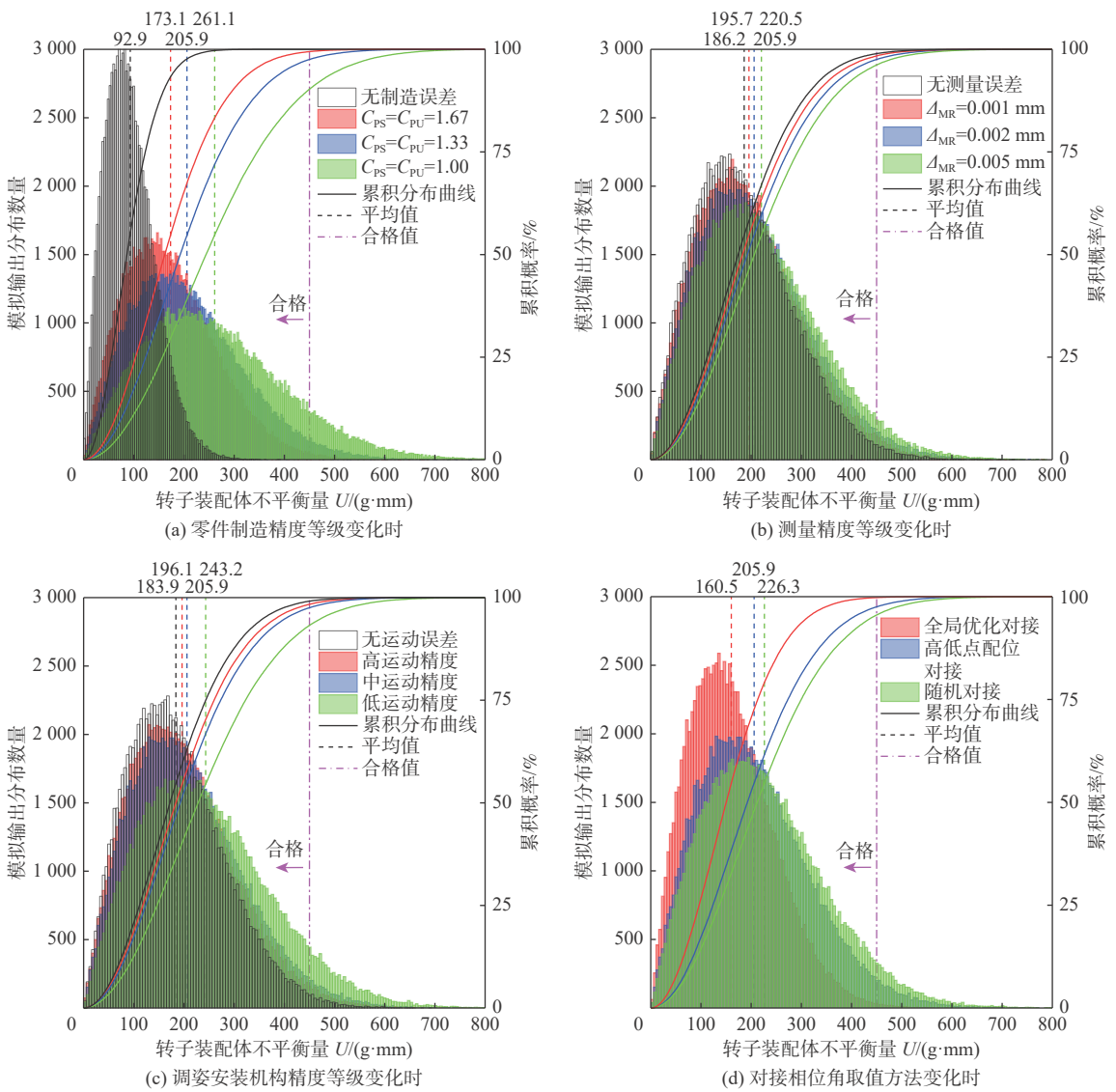


图 6　不同精度等级输入下的转子装配体不平衡量分布

Fig. 6　Distributions of unbalance of rotor assemblies under inputs of different accuracy grades

所有输出的 U 均值右侧的数据具有更高的离散性，偏度>0，各分布是左偏的，所有输出分布形态相比于标准正态分布更加陡峭，峰度>0。与无误差假设相比，各误差源均对 U 造成了负面影响。对于 U 的均值、标准差、95% CI(置信区间)宽度和上限值这 4 项统计量，其均随着零件制造精度等级的提升而下降，随着综合测量精度的提升而下降，随着调姿安装机构的运动定位精度提升而下降，并随着对接方式优化而下降，所述统计量的下降程度各不相同。根据所有条件组合下的各误差源的关联精度等级变化时造成 U 的均值降低量的显著程度，计算敏感度系数如表 3 所示。

表 3　装配系统中各关联精度的敏感度系数

Table 3　Sensitivity coefficient of each related accuracy in assembly system

输入的精度等级	敏感度系数				
	零件形位精度	零件不平衡量精度	测量精度	调姿安装机构精度	对接方式
低→中	0.06	0.19	0.17	0.06	0.09
中→高	0.05	0.09	0.04	0.05	0.20

根据各关联精度项的敏感度系数及不同精度等级的实施难度，对精度方案进行优选，例如，当对接方式由“高低点配位对接”变化为“全局优化对接”，同时其他条件保持基准值时，U 的均值降低率达 22.0%，此时具有最大敏感度系数 S=0.20，而“全局优化对接”可通过智能算法便于实现，因此为优选方案；当综合测量精度 Δ_{MR} 由 0.002 mm 变化到 0.001 mm，U 的均值降低率仅为 5.0%，具有最小敏感度系数 S=0.04，而此时硬件成本和测量耗时明显增加，因此，选取中测量精度等级为最优方案。同理，分析其他条件影响，得到各误差源的优化关联精度方案为：C_{PS}=1.33，C_{PU}=1.67，Δ_{MR}=0.002 mm，$\max\left(\Delta q_{x/y/z}\right)$=0.02 mm，$\max\left(\Delta q_{\alpha/\beta/\gamma}\right)$=0.2′，对接相位角 ϕ_i 取值采用“全局优化对接”方法，在此精度组合条件下，预测得到的转子装配体不平衡量合格($U\leqslant$ 450 g·mm)率达到 99.8%，较未进行精度优化的全局平均合格率增加了约 5%，较低精度组合条件的合格率增加了约 18%，优化方案为实际装配硬件系统的精度设计提供了可靠依据。

4.2　实验验证

图 7 为航发转子装配验证实验。根据本文提出的精度优化设计方案搭建航发转子装配实验系统，如图 7(a) 所示。对 3 组转子零件、测量系统和调姿安装机构的精度进行检验和记录，确保所有精度项达到对应的优化指标，并将对应的测定数值输入至模拟算法中。

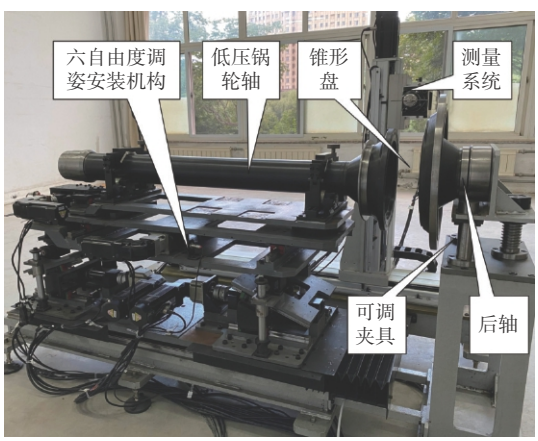
每组转子零件的装配实验次数为 30，每次实验与模拟算法的流程相一致，主要流程如下。

步骤 1　后轴安装与位姿测量。实验中，通过可调夹具对后轴的初始位姿进行随机化设定，并经测量系统测量；模拟中，对 ${}^{0}_{R1}\boldsymbol{T}$ 中元素随机采样，按照输入的后轴形位误差计算 ${}^{0}_{FS1}\boldsymbol{T}$，对测量误差采样并生成虚拟测点，算得 ${}^{M}_{FS1}\boldsymbol{T}_{(fit)}$。

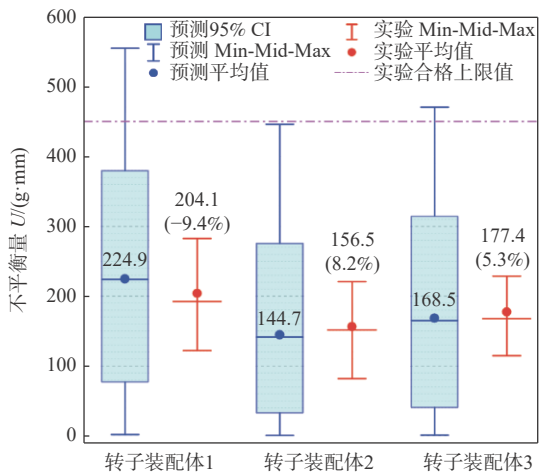
步骤 2　锥形盘的装配。实验中，将锥形盘安装于调姿安装机构上的夹具中，测量初始位姿，根据全局优化对接方法计算其目标位姿 ${}^{0}_{RS2}\boldsymbol{T}_{(tar)}$，进行调姿安装运动、紧固和位姿复测；模拟中，各运动变量误差 Δq_i 根据输入分布进行采样，末端误差通过 Jacobian 矩阵解出，按照输入的锥形盘形位误差计算 ${}^{0}_{FS2}\boldsymbol{T}$，并算得含测量误差的位姿 ${}^{M}_{FS2}\boldsymbol{T}_{(fit)}$。

步骤 3　低压涡轮轴的装配。类比步骤 2。

步骤 4　不平衡量测量。实验中，测量装配体的不平衡量；模拟中，根据输入的各零件不平衡量



(a) 实验条件



(b) 实验与模拟结果对比

图 7　航发转子装配验证实验

Fig. 7　Aeroengine rotor assembly verification experiments

参数, 计算各转子零件重心相对装配体几何转轴的偏离向量 ${}^{Bi-Bj}_{Gk}\boldsymbol{P}$, 算得装配体不平衡量 U。

实验与模拟获得的不平衡量结果如图 7(b) 所示, 共计 90 次实验的不平衡量合格率达到 100%, 每个转子装配体的 30 次实验所得不平衡量极值区间全部位于对应 1×10^5 次算法预测结果的 95% CI 区间中。相对于算法预测的不平衡量均值, 对应实验结果的偏离率均<10%。综上, 本文方法具有较高的预测精度。

5　结　论

本文开展了航发转子不平衡量预测方法研究, 取得如下结论:

1) 阐明了航发转子装配过程中的工艺参数、多源误差、相对位姿及不平衡量之间的传递原理。分别建立了典型转子零件空间形位与重心误差表征模型、基于虚拟测点生成的特征结构测量误差模型、多自由度调姿安装机构通用误差模型和运动误差简化解法。

2) 基于多源误差模型建立了航发转子不平衡量蒙特卡罗模拟算法, 实现了不同装配条件下的不平衡量预测。对于某型低压涡轮转子模拟件实例, 通过敏感度分析实现了零件制造精度、测量精度和装配运动精度的反求设计, 优化方案下的不平衡量合格率预测值达 99.8%。

3) 转子装配验证实验结果表明: 每个转子装配体 30 次实验的不平衡量极值区间全部位于 1×10^5 次算法预测结果 95% CI 区间中, 实验获得的不平衡量均值相对预测结果的偏离率均<10%, 验证了本文方法的预测准确度。

对于后续研究展望, 拟对装配系统中的工件精度、测量精度和运动精度等精度组成模块之间的相对敏感度作深入研究, 细化高敏感性误差模型, 简化低敏感性误差建模过程, 进一步提高不平衡量预测精度。

参考文献 (References)

[1] 陈予恕, 张华彪. 航空发动机整机动力学研究进展与展望[J]. 航空学报, 2011, 32(8): 1371-1391.
CHEN Y S, ZHANG H B. Review and prospect on the research of dynamics of complete aero-engine systems[J]. Acta Aeronautica et Astronautica Sinica, 2011, 32(8): 1371-1391(in Chinese).

[2] 陈果, 李兴阳. 航空发动机整机振动中的不平衡-不对中-碰摩耦合故障研究[J]. 航空动力学报, 2009, 24(10): 2277-2284.
CHEN G, LI X Y. Study on imbalance-misalignment-rubbing coupling faults in aero-engine vibration[J]. Journal of Aerospace Power, 2009, 24(10): 2277-2284(in Chinese).

[3] EWINS D J. Control of vibration and resonance in aero engines and rotating machinery-an overview[J]. International Journal of Pressure Vessels and Piping, 2010, 87(9): 504-510.

[4] 陈立芳, 王维民, 高金吉. 航空发动机自动平衡技术发展综述[J]. 航空动力学报, 2019, 34(7): 1530-1541.
CHEN L F, WANG W M, GAO J J. Summary on the development of auto-balancing technology of aero-engine[J]. Journal of Aerospace Power, 2019, 34(7): 1530-1541(in Chinese).

[5] 张渝, 李琳, 陈津, 等. 航空发动机重要装配工艺分析及研发展望[J]. 航空制造技术, 2019, 62(15): 14-21.
ZHANG Y, LI L, CHEN J, et al. Research current status and prospect on aero-engine assembly process technology[J]. Aeronautical Manufacturing Technology, 2019, 62(15): 14-21(in Chinese).

[6] 石宏. 航空发动机装配工艺技术[M]. 北京: 北京航空航天大学出版社, 2015.
SHI H. Aeroengine assembly technology[M]. Beijing: Beijing University of Aeronautics & Astronautics Press, 2015(in Chinese).

[7] SUN C Z, LIU Z W, LIU Y M, et al. An adjustment method of geometry and mass centers for precision rotors assembly[J]. IEEE Access, 2019, 7: 169992-170002.

[8] SUN Q C, ZHAO B B, LIU X, et al. Assembling deviation estimation based on the real mating status of assembly[J]. Computer-Aided Design, 2019, 115: 244-255.

[9] YU L, ZHANG Y L, BI Q Z, et al. Research on surface normal measurement and adjustment in aircraft assembly[J]. Precision Engineering, 2017, 50: 482-493.

[10] ZHANG M W, LIU Y M, SUN C Z, et al. Precision measurement and evaluation of flatness error for the aero-engine rotor connection surface based on convex hull theory and an improved PSO algorithm[J]. Measurement Science and Technology, 2020, 31(8): 085006.

[11] 何晓煦, 田威, 曾远帆, 等. 面向飞机装配的机器人定位误差和残差补偿[J]. 航空学报, 2017, 38(4): 420538.
HE X X, TIAN W, ZENG Y F, et al. Robot positioning error and residual error compensation for aircraft assembly[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(4): 420538(in Chinese).

[12] LI Z H, TIAN W, WANG M, et al. Positioning error compensation of a flexible track hybrid robot for aircraft assembly based on response surface methodology and experimental study[J]. The International Journal of Advanced Manufacturing Technology, 2022, 119(1): 1313-1330.

[13] ZHU Y G, HUANG X, FANG W, et al. Trajectory planning algorithm based on quaternion for 6-DOF aircraft wing automatic position and pose adjustment method[J]. Chinese Journal of Aeronautics, 2010, 23(6): 707-714.

[14] WHITNEY D E, GILBERT O L, JASTRZEBSKI M. Representation of geometric variations using matrix transforms for statistical tolerance analysis in assemblies[J]. Research in Engineering Design, 1994, 6(4): 191-210.

[15] DING S Y, HE Y H, ZHENG X H. A probabilistic approach for three-dimensional variation analysis in aero-engine rotors assembly[J]. International Journal of Aeronautical and Space Sciences, 2021, 22(5): 1092-1105.

[16] ZHOU Z L, LIU W, WU Q, et al. A combined measurement method for large-size aerospace components[J]. Sensors, 2020, 20(17): 4843.

[17] 孙汕民, 孙传智, 冯硕, 等. 1 种高精度大型阶梯轴圆柱轮廓测量模型[J]. 航空发动机, 2020, 46(2): 61-65.
SUN S M, SUN C Z, FENG S, et al. One kind of high accuracy measurement model of large stepped shaft cylinder profile[J]. Aeroengine, 2020, 46(2): 61-65(in Chinese).

[18] YANG J X, DING H. A new position independent geometric errors identification model of five-axis serial machine tools based on differential motion matrices[J]. International Journal of Machine Tools and Manufacture, 2016, 104: 68-77.

[19] ZHOU W Y, CHEN W Y, LIU H D, et al. A new forward kinematic algorithm for a general Stewart platform[J]. Mechanism and Machine Theory, 2015, 87: 177-190.

[20] HUSSAIN T, MCWILLIAM S, POPOV A A, et al. Geometric error reduction in the assembly of axi-symmetric rigid components: a two-dimensional case study[J]. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 2012, 226(7): 1259-1274.

[21] MEMON M, HUSSAIN T, MEMON Z A. Minimizing assembly errors by selecting optimum assembly sequence in the assembly of a rigid circular structure[J]. Mehran University Research Journal of Engineering and Technology. 2012, 31(4): 743-754.

[22] BOLOTOV M A, PECHENIN V A, YU PECHENINA E, et al. Digital model to optimize the rotor assembly for aircraft engine compressor models[J]. Journal of Physics: Conference Series, 2021, 1745(1): 012033.

[23] QURESHI A J, DANTAN J Y, SABRI V, et al. A statistical tolerance analysis approach for over-constrained mechanism based on optimization and Monte Carlo simulation[J]. Computer-Aided Design, 2012, 44(2): 132-142.

[24] LEE N K S, YU G, JONEJA A, et al. The modeling and analysis of a butting assembly in the presence of workpiece surface roughness and part dimensional error[J]. The International Journal of Advanced Manufacturing Technology, 2006, 31(5): 528-538.

[25] 胡明. 面向质量特性优化的航空发动机转子堆叠装配方法[D]. 哈尔滨: 哈尔滨工业大学, 2019.
HU M. Stacked assembly method of aero-engine rotor oriented to optimization of quality characteristics[D]. Harbin: Harbin Institute of Technology, 2019(in Chinese).

Prediction of aeroengine rotor unbalance considering multi-source errors in assembly process

ZHOU Tianyi[1,*], HU Lei[2], ZHAO Suo[2], JIN Xiaoxiao[2], WANG Shiyu[2], GAO Hang[3]

(1. School of Mechanical Engineering, Dalian Jiaotong University, Dalian 116028, China;
2. Shenyang Aircraft Design & Research Institute, Aviation Industry Corporation of China, Ltd., Shenyang 110035, China;
3. School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: To address the difficulty in predicting the unbalance of aeroengine rotors assembled by complex assembly system, this paper proposed a numerical simulation method for unbalance considering multi-source errors, aiming to achieve the reliable traceability design of the accuracy of workpiece group and its assembly system. Firstly, the transmission principle among process parameters, multi-source errors, relative pose, and unbalance during the rotor assembly process was clarified. Quantitative characterization models were established for typical rotor part manufacturing errors, flange and spigot pose measurement errors, and multi-degree-of-freedom installation mechanism motion errors. Then, a Monte Carlo simulation algorithm for rotor unbalance was designed based on the above models. It realized unbalance prediction under different assembly conditions. For the analysis case of a low-pressure turbine rotor simulator, the accuracy of its assembly system was optimized inversely through comprehensive sensitivity analysis. The qualified rate of corresponding simulated unbalance reaches 99.8%. The proposed simulation method provides reliable accuracy design reference. Finally, assembly verification experiments were carried out. The results show that the deviation rates between the averages of the experimental unbalance and the predicted unbalance are all less than 10%. Thus, the proposed prediction method has high accuracy.

Keywords: aeroengine; rotor; assembly; unbalance; Monte Carlo; accuracy design

Received: 2023-07-04; **Accepted**: 2023-09-03; **Published Online**: 2024-01-05 21:01
URL: link.cnki.net/urlid/11.2625.V.20240105.1321.003
Foundation item: National Natural Science Foundation of China (51621064)
* **Corresponding author**. E-mail: ztydlut@foxmail.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0443

基于 GBDT-GS 方法的民机重着陆风险预测

王向章[1,*]，牟瑞芳[1]，王赫[2]，许博浩[2]

(1. 西南交通大学 交通运输与物流学院，成都 610031；2. 四川航空股份有限公司 飞行部，成都 610200)

摘 要：重着陆可能造成机体结构损伤等征候事件，甚至机毁人亡飞行事故。针对当前重着陆风险评估缺乏物理本质剖析，为有效实施重着陆风险识别和等级判据，以便提高飞行员着陆操作品质，结合飞行状态分析，基于梯度提升决策树（GBDT）算法和网格搜索（GS）法构建重着陆风险预测模型。通过飞机受力分析并构建着陆飞行运动学方程，确定与重着陆关系密切的5项飞行状态参数；从机载快速存取记录器（QAR）记录的数据中提取飞行状态数据构建数据集，并根据QAR参数特征，通过GBDT算法构建重着陆风险预测模型，并利用GS优化模型参数；以某航空公司“成都-沈阳”航线运行为例，选取530个QAR数据对该模型进行训练和测试，并与随机森林、Logistic多元回归、循环神经网络（RNN）等算法结果比较。结果表明：GBDT-GS方法在预测重着陆风险方面的性能较其他算法优异，预测准确率达到92%，验证了所建模型的客观有效性。

关 键 词：飞行状态；重着陆；风险预测；QAR数据；梯度提升决策树

中图分类号：V328.5

文献标志码：A **文章编号**：1001-5965(2025)09-3011-09

进近着陆阶段是飞行运行关键阶段。为保证飞机安全平稳着陆，在进近着陆过程中，飞行员需保持飞机航向、速度、姿态及发动机推力始终处于稳定状态[1]。飞行状态的不稳定意味着飞机能量存在异常，若不能及时控制则可能造成能量意外释放，进而导致飞行事故发生。根据波音公司2007—2016年安全报告，进近着陆飞行事故死亡人数占全球商用喷气式飞机事故总死亡人数的48%，其中，飞行状态不稳定是导致飞行事故的关键因素[2]。2019年国际航空运输协会(International Air Transport Association, IATA)发布的安全报告则进一步指出，重着陆、冲出跑道和飞行失控是进近着陆阶段发生频率较高的飞行事故，这些事故几乎都与飞机飞行状态异常直接相关，其中重着陆占到所有飞行事故的14%[3]。众所周知，重着陆可能造成机体结构损伤等事故征候，甚至机毁人亡飞行事故。重着陆致因较为复杂，包括环境因素、航空器性能、飞行员操作等，但这些因素的影响结果最终都反应到飞机飞行状态变化上[4]。

重着陆是指飞机接地时垂直加速度过大，接地载荷超过了该机型给定的限制值[5]。在实际运行过程中，航空公司主要通过监控快速存取记录器(quick access recorder, QAR)数据中的接地时垂直加速度(vertical acceleration during grounding, VRTG)参数判断飞机是否发生接地载荷大事件。这种监控方式较为单一，属于事后型分析，缺乏及时性，且常常存在误判漏判情况。对此，围绕重着陆诊断和探测，部分学者结合仿真技术或利用QAR数据，采用贝叶斯灵敏性分析方法[6]、支持向量机[7-8]、人工神经网络[9-10]等数据挖掘技术提高重着陆判断的精确度，以便及时发现危害飞机结构的安全事件；针对重着陆风险分析和预测，有的学者主要通过软件、硬件、环境、人(software, hardware, environment,

收稿日期：2023-07-05；**录用日期**：2023-09-10；**网络出版时间**：2023-09-13 16:46
网络出版地址：link.cnki.net/urlid/11.2625.V.20230912.1820.001
基金项目：民航安全能力建设基金项目 (ASSA202019)
* **通信作者**. E-mail：513432965@qq.com

引用格式：王向章，牟瑞芳，王赫，等. 基于GBDT-GS方法的民机重着陆风险预测[J]. 北京航空航天大学学报，2025，51(9)：3011-3019. WANG X Z，MOU R F，WANG H，et al. Hard landing risk prediction of civil aircraft based on GBDT-GS method[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3011-3019 (in Chinese).

liveware, SHELL)模型来识别影响重着陆的风险因素[11-12],有的则基于统计分析方法[5]、正态云概念[13]等构建了重着陆风险分析与评价模型,以评价飞行员着陆操作风险,或者结合实际飞行数据(主要为QAR数据),利用过程控制方法[14]、动态事件规整方法[15]、自适应变异粒子群法[16]等实现了重着陆风险预测。但是,上述研究更多偏重于数学模型和算法的应用,并未从物体运动的角度深入分析重着陆发生的本质原因。

从物理本质上讲,重着陆是飞机受力改变引起飞行状态变化导致的结果,而飞行员操作、大气气流等都可能引起飞机受力发生变化,进而改变飞行状态。飞行状态是飞机在任一瞬时所处的空间位置、飞行速度及飞行姿态。因此,要认识重着陆发生的本质,需要理清飞行状态变化与重着陆之间的关系。目前,国内外专家学者在飞行状态研究方面已开展大量研究。这些研究的基本思路是通过传统的飞行动力学建模,运用卡尔曼滤波[17-18]、约束粒子滤波模型[19]及隐马尔可夫模型[20]等方法识别或预测飞机飞行状态及轨迹变化,以帮助管制员指挥飞机及时调整飞行速度、高度和航向,进而避免发生空中冲突。而随着飞行数据质量和可用性的不断提高,越来越多学者开始利用递归神经网络[21]、聚类分析技术[22]、长短期记忆网络模型[23]等大数据挖掘方法来识别飞行状态,分析飞机运动轨迹。然而,不管是传统机理建模分析,还是通过大数据技术预测,都只是从方法上提高飞行状态估计的准确程度,而对飞行状态变化与飞行事故之间的关联性并没有作进一步剖析。

尽管国内外学者已较为系统地分析了重着陆成因并开展了风险评价研究,但鲜有人剖析飞行状态变化对重着陆的影响。鉴于此,本文在已有研究的基础上,从飞机着陆运动过程入手,通过运动受力分析并构建着陆飞行运动学方程,确定与重着陆关系密切的飞行状态参数;然后,提出利用网格搜索(grid search, GS)法优化梯度提升决策树(gradient boosting decision tree, GBDT)来构建重着陆风险预测模型;最后,结合航班真实QAR数据,采用构建的GBDT-GS方法对重着陆风险进行预测,并验证本文方法的有效性。

1 飞行状态与重着陆风险

1.1 着陆飞行运动学方程

飞机从离跑道端口高15 m左右开始到飞机完全停止运动为着陆阶段。飞行员通常在决断高度前脱开自驾,改为人工操纵飞机进近着陆,并通过飞行显示器监控飞机速度、高度、姿态等飞行状态信息,以便在状态发生偏离时及时恢复飞机稳定状态。根据标称着陆过程,可将着陆划分为3个阶段:下滑阶段、拉平飘落阶段、减速滑跑阶段,如图1所示。其中,γ为下滑角,R为拉平圆周运动半径,H_A为下滑飞行高度,H_B为飞机过跑道端口点高度,H_C为飞行员拉平开始高度,X_1为着陆下滑至拉平点的水平飞行距离,X_2为拉平过程飞行距离,X_3为着陆滑跑距离。

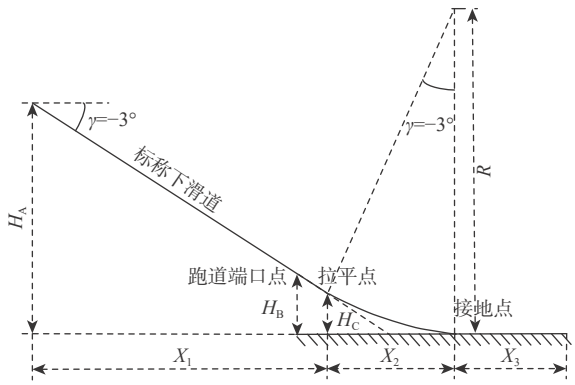


图1 标称进近着陆任务示意图

Fig. 1 Diagram of nominal approach and landing mission

飞机在着陆过程中主要受气动力(气动升力、气动阻力)、发动机推力和重力共同作用,并由气动力产生滚转力矩、俯仰力矩和偏航力矩。气动升力垂直于飞行速度,方向向上;气动阻力平行于飞行速度,方向向后;忽略发动机安装角,则推力平行于飞机纵轴,方向与速度方向一致,如图2所示,其中,L为气动升力,D为气动阻力,θ为俯仰角,α为飞机迎角,V为飞行速度,T为发动机推力,W为重力。着陆过程中,飞机飞行状态变化的物理本质是受力的作用,且主要沿纵向做线性运动及绕机体纵轴和横轴的转动,而横轴的位置变化较小[24]。因此,参考文献[24]建立各个阶段纵向运动学方程。

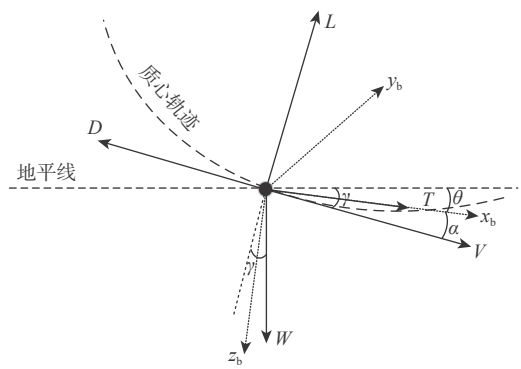


图2 飞机纵向受力示意图

Fig. 2 Diagram of longitudinal force on aircraft

1.1.1 下滑阶段

正常下降时,发动机处于慢车状态,飞机机头略微向下,并沿正常下滑角(3°)减速下滑。当飞

机在离跑道端口 15 m 左右高度时, 飞机速度达到进场速度 V_{app}。通常情况下, 进场速度 V_{app} 与进场基准速度 V_{REF} 相同。下滑阶段飞机的飞行状态方程为

$$\begin{cases} \dfrac{W}{g}\cdot\dfrac{\mathrm{d}V}{\mathrm{d}t}=T\cos\alpha-D+W\sin\gamma \\ T\sin\alpha+L-W\cos\gamma=0 \\ \dfrac{\mathrm{d}H}{\mathrm{d}t}=-V\sin\gamma \\ \dfrac{\mathrm{d}\gamma}{\mathrm{d}t}=0 \\ \dfrac{\mathrm{d}X_1}{\mathrm{d}t}=V\cos\gamma-V_{\mathrm{w}} \end{cases} \tag{1}$$

式中: W 为飞机重力; g 为重力加速度; H 为无线电高度; V_{w} 为风速。

1.1.2　拉平飘落阶段

由于前起落架能承受的撞击力远小于主起落架, 若以正常下滑角着陆, 飞机将直接俯冲向跑道, 造成严重结构损伤。因此, 在下滑到离地约 9 m 高度时, 飞行员通过拉杆和收油门配合操纵, 使飞机逐渐退出下滑状态转为平飘接地状态。在拉平过程中, 下滑角将由 3°逐渐减小为 0°, 飞机俯仰姿态和迎角逐渐增大, 机头由低头状态逐渐转为抬头状态, 迎角保持一定, 形成两点姿势, 然后, 匀速下沉并在接地区域达到接地速度。拉平飘落阶段飞机的飞行状态方程为

$$\begin{cases} \dfrac{W}{g}\cdot\dfrac{\mathrm{d}V}{\mathrm{d}t}=T\cos\alpha-D+W\sin\gamma \\ \dfrac{W}{g}V\dfrac{\mathrm{d}\gamma}{\mathrm{d}t}=T\sin\alpha+L-W\cos\gamma \\ \dfrac{\mathrm{d}H}{\mathrm{d}t}=-V\sin\gamma \\ \dfrac{\mathrm{d}X_2}{\mathrm{d}t}=V\cos\gamma-V_{\mathrm{w}} \end{cases} \tag{2}$$

1.1.3　减速滑跑阶段

飞机主轮接地后, 反推和地面扰流板打开, 飞机阻力增大, 滑跑速度减小。随着速度的减小, 升力也随之减小, 机头自然下俯, 直至前轮接地。待前轮接地后, 飞行人员使用脚踏板进行刹车减速, 并随滑跑速度减小, 再逐渐增大刹车量。减速滑跑阶段飞机的飞行状态方程为

$$\begin{cases} V_t=V-at \\ X_3=Vt-\dfrac{1}{2}at^2 \\ a=\dfrac{g}{W}\left[T\cos\alpha-\mu W-(C_D-\mu C_L)\rho V^2S/2-W\varphi\right] \end{cases} \tag{3}$$

式中: μ 为摩擦系数; ρ 为空气密度; S 为机翼参考面积; C_D 为阻力系数; C_L 为升力系数; V_t 为某时刻飞机的滑跑速度; a 为飞机滑跑加速度; φ 为滚转角。

1.2　重着陆风险因子

飞行状态一般采用飞行速度、飞行高度、俯仰角、滚转角、迎角及侧滑角等来表征。由重着陆定义可知, 重着陆的发生必然伴随接地垂直加速大现象, 而垂直加速度与飞行高度、接地前垂直方向的飞行速度等飞行状态参数直接相关。结合式(1)和式(2)的飞行运动学方程, 从下滑、拉平、平飘、接地等过程, 对飞机着陆运动过程进行分析。

在下滑过程中, 若进跑道端口高度高, 为保证飞机在接地区域接地, 需增大下滑角快速下降, 从而导致下降率增加, 垂直加速度增大; 若拉平前飞行速度过大, 通过提前收油门减速, 可能使飞机升力快速减小, 垂直加速度增大。在拉平过程中, 若拉开始操作的高度较低, 飞机在接地前具有一定下滑角, 在重力 W 作用下将以较大的下降率快速下降, 飞行员带杆修正量不足, 则势必导致垂直加速度过大。在平飘至接地过程中, 接地飞行俯仰角小, 飞机迎角也将变小, 可能导致飞机升力骤减, 进而因 $W>L$ 造成飞机垂直加速度过大。在接地过程中, 受天气或飞行操作影响, 飞机可能存在一定滚转角, 使得接地瞬间存在主轮单点接地情况, 则单侧垂直方向上受力较大, 接地载荷可能瞬间超过机型限制值。

根据上述分析, 飞机在着陆过程中垂直加速度过大与过跑道端口高度、飞行速度、下降率、飞行俯仰角等密切相关。而飞行员对油门、操作杆的操作又直接影响飞行速度、俯仰姿态角, 同时接地前遇到风切变、顺风、乱流等恶劣天气, 也会影响飞行速度和姿态角。当这些因子出现异常时, 都可能直接或间接造成重着陆。但无论是飞行操作原因, 还是天气变化带来的影响, 最终都反应在飞行状态的变化上。因此, 本文以无线电高度 H、飞行速度 V、下降率 I_{VV}、俯仰角 θ、滚转角 φ 等飞行状态参数作为变量来分析状态变化导致的重着陆风险。

2　飞机重着陆风险预测模型

QAR 作为记录飞机飞行参数信息的重要设备, 储存了飞行员全部操作动作及飞机内外部状态参数数据, 常被用来监控飞行操作品质, 分析不安全事件原因。利用 QAR 数据分析飞机飞行状态和预测重着陆风险, 需要从原始 QAR 数据中提取有效信息。实际飞行运行中, 绝大部分 QAR 数据均为正常数据, 而涉及重着陆的飞行状态异常数据较少, 因此, 需要从大量已知数据中识别出异常数据, 即对异常数据进行分类处理以实现风险预测。

2.1 GBDT算法

GBDT是一种集成学习算法，其通过迭代训练多个决策树来提高模型性能。由于GBDT算法在回归预测和分类问题上具有良好的性能，因而受到各行业广泛关注。该算法由多棵决策树组成，主要采用前向分布式迭代方式，通过每次迭代学习基函数并更新权重系数来优化损失函数，从而使算法能够有效处理非线性关系和高维特征。假设t–1次迭代得到的学习器为$f_{t-1}(x)$，损失函数为$L(y,f_{t-1}(x))$，则在第t次迭代后可找到一个能够使损失函数$L(y,f_t(x))$最小的基函数$g(x,\theta_t)$[25]：

$$L(y,f_t(x))=L(y,f_{t-1}(x))+g(x,\theta_t)=$$
$$(y-f_{t-1}(x)-g(x,\theta_t))^2=r-g(x,\theta_t) \tag{4}$$

式中：$r=y-f_{t-1}(x)$为残差。在每次迭代过程中，沿负梯度方向，通过与每棵回归树拟合以减小残差。Friedman提出利用梯度提升法减小残差，其主要通过泰勒公式一阶展开来近似处理损失函数[26]，如式(5)所示。该方法得到的损失函数值与实际更接近，且具有较高的预测准确率。图3为GBDT算法原理图。

$$r\approx-\left(\frac{\partial L(y_i,f(x_i))}{\partial f(x_i)}\right)_{f(x_i)=f_{t-1}(x)} \tag{5}$$

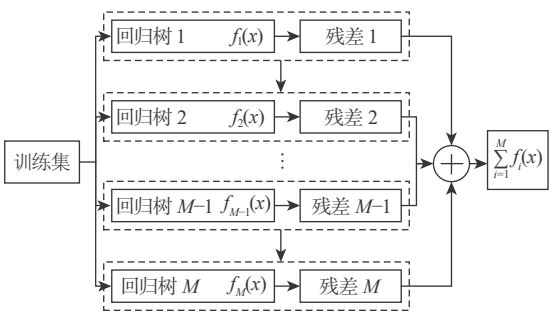


图3 GBDT算法原理图

Fig. 3 Schematic diagram of GBDT algorithm

2.2 基于GBDT-GS方法的重着陆风险预测模型

根据GBDT算法原理，以回归树为基学习器，平方误差为损失函数构建基于GBDT算法的重着陆风险预测模型。该模型通过累加所有基学习器的计算结果得到预测值，再利用下一个基学习器拟合损失函数对预测值的残差得到预测精度。为验证GBDT模型预测结果是否符合精度要求，再采用GS进一步优化模型参数。因此，重着陆风险预测模型包括基于GBDT算法的预测值计算及基于GS的模型参数优化。

2.2.1 基于GBDT算法的预测值计算

GBDT算法的训练过程包括特征选择、分裂点选择和模型更新3个阶段。特征选择阶段，通过计算各飞行状态参数特征的平方差损失，并选择具有最小损失的特征作为分裂特征；分裂点选择阶段，根据选定的分裂特征，计算不同分裂点的分裂增益(Gain)，并选择具有最大增益的分裂点作为最佳分裂点；模型更新阶段，利用贪心算法遍历所有分裂特征的所有分裂点，并通过循环迭代分裂过程，生成多个基学习器。在每次迭代中，GBDT根据学习器的误差率来更新模型的权值分布。Gain计算式为

$$G_{\text{ain}}=\frac{1}{2}\left[\frac{G_{\text{L}}^2}{H_{\text{L}}+\lambda}+\frac{G_{\text{R}}^2}{H_{\text{R}}+\lambda}-\frac{(G_{\text{L}}+G_{\text{R}})^2}{H_{\text{L}}+H_{\text{R}}+\lambda}\right]-\hat{\gamma} \tag{6}$$

式中：G_{L}为左子树梯度值；G_{R}为右子树梯度值；H_{L}为左子树的设置分数；H_{R}为右子树的设置分数；λ表示正则项系数；$\hat{\gamma}$为叶子个数。

无线电高度H、飞行速度V、下降率I_{VV}、俯仰角θ、滚转角φ等飞行状态参数的平方误差损失分别为：h_{loss}、V_{loss}、I_{VVloss}、θ_{loss}、φ_{loss}，其权值更新过程如图4所示。

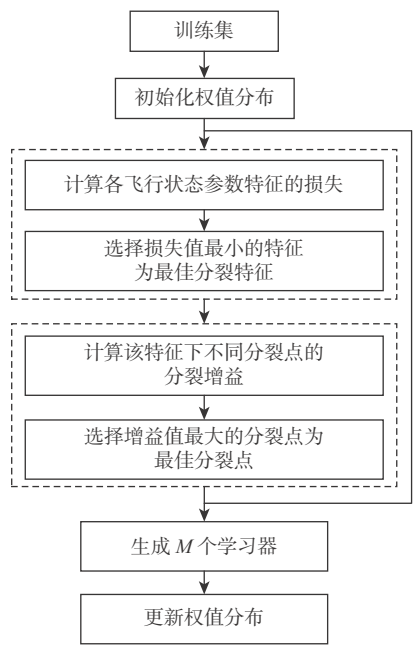


图4 参数特征损失的权值更新过程

Fig. 4 Parameter feature loss weight updating process

2.2.2 基于GS的模型参数优化

模型训练时，模型的学习率l和决策树最大深度d对重着陆风险预测结果影响较大。为找到最佳的参数组合，本文采用GS法对模型参数进行调优[27]。通过将待搜索参数的取值空间划分成一个网格，再对每个参数组合进行模型训练，选择出最佳预测结果的参数组合。根据经验，将l、d的取值空间分别设置为[0.05, 0.15]、[1, 5]，空间步长设置为0.01、1。根据取值范围和空间步长，通过遍历每个待搜索参数组合对模型进行迭代训练和评估。经过多轮优化后，可得到具有通用性的超参数组

合, 即将均方根误差(root mean square error, RMSE)最小的参数组合作为最终模型参数。利用 GS 法优化模型参数的过程如图 5 所示。

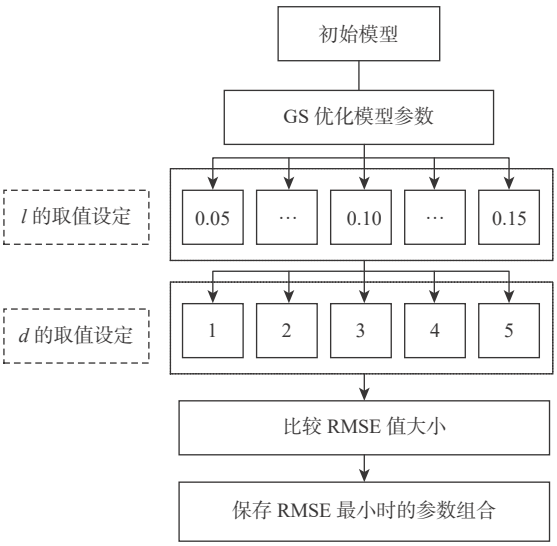


图 5　GS 优化模型参数

Fig. 5　Optimization of model parameters by GS

2.3　评估指标

本文实验主要采用准确率(accuracy, ACC)、RMSE、平均绝对误差(mean absolute error, MAE)和均方误差(mean square error, MSE)对 GBTD-GS 方法的预测性能进行评估。4 个评估指标的计算式分别为

$$A_{CC}=\frac{T_P+T_N}{T_P+T_N+F_P+F_N} \tag{7}$$

$$M_{AE}=\frac{1}{m}\sum_{i=1}^{m}|y_i-\hat{y}_i| \tag{8}$$

$$M_{SE}=\frac{1}{m}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2 \tag{9}$$

$$R_{MSE}=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2} \tag{10}$$

式中: y_i为真实值; $\hat{y}_i$为模型预测值; m 为实验样本数。若 MAE、MSE、RMSE 的值越小, 则说明本文模型的预测性能越好。

3　实例分析与验证

3.1　实例分析流程

首先, 本文根据着陆飞行运动方程分析, 选取与重着陆相关的风险因子作为监测指标, 并从 QAR 数据中提取重着陆相关飞行参数进行预处理; 然后, 建立基于 GBDT 的预测模型进行训练与测试, 得到重着陆风险预测结果; 最后, 将预测结果与循环神经网络(recurrent neural networks, RNN)、Logistic 多元回归、随机森林等算法进行比较, 验证本文方法的有效性。本文方法流程如图 6 所示。

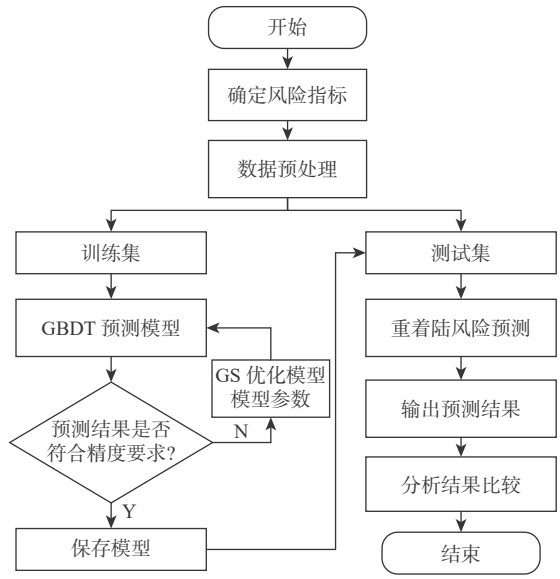


图 6　本文方法流程

Fig. 6　Flow of the proposed method

3.2　指标确定及数据预处理

根据飞行运动学分析, 无线电高度 H、飞行速度 V、下降率 I_{VV}、俯仰角 θ、滚转角 φ 等飞行状态参数出现异常, 将可能导致重着陆发生。飞行员在人工操作飞机着陆过程中, 极易因操作技术、天气原因等导致飞行状态异常。因此, 参考民航咨询通告《飞行品质监控(FOQA)实施与管理》(AC-121/135-FS-2012-45R1)[28] 及航空公司飞行品质监控中的监测标准, 采用 Delphi 法, 在广泛征集一线飞行人员意见后, 以 A320 机型着陆过程中进跑道端口高度 x_1、拉平时(9 m 左右)飞行速度 x_2 及接地前 1 s 的下降率 x_3、俯仰角 x_4、滚转角 x_5 共 5 项飞行状态参数作为重着陆风险监测指标, 并根据 A320 机型着陆载荷大监控标准, 以接地期间的 VRTG 值作为样本等级判据, 并设置标签。飞行数据样本的数据化处理方式如表 1 所示。

从某航空公司 AirFASE 译码系统抽取 2022 年上半年“成都-沈阳”航段 QAR 数据, 共计 580 份。经分析数据结构, 所有 QAR 数据参数的采集频率均为 1 Hz(每秒采样 1 次)。由于飞机传感器敏感

表 1　飞行数据样本的数据化处理

Table 1　Data processing of flight data samples

接地载荷	样本状态	标签
VRTG > 1.75G	高风险	3
1.6G < VRTG < 1.75G	中风险	2
1.5G < VRTG < 1.6G	低风险	1
VRTG < 1.5G	无风险	0

注: G表示飞机“过载”。

度原因，导致部分QAR数据存在缺失。为使样本数据的质量满足预测分析要求，通过剔除缺失严重的数据，最终得到530份QAR数据，然后，从数据样本中选取前文确定的与重着陆相关的飞行状态参数建立数据集，数据集样本如表2所示。由于篇幅原因，本文仅列出其中部分原始数据，并按照7:3的比例建立训练集与测试集，即371个样本作为训练集，159个样本作为测试集。训练集和测试集的数据样本均按照航班运行时间排列，使样本保持时间序列特征。

表2 数据集样本（部分）

Table 2 Data set sample (partial)

样本	x_1/m	x_2/(m·s^{-1})	x_3/(m·s^{-1})	x_4/(°)	x_5/(°)	标签
1	14.32	67.39	−1.204	2.81	0	0
2	14.63	69.45	−1.219	3.52	−1.32	0
3	15.24	69.45	−1.300	6.33	−2.46	0
4	13.72	72.02	−1.382	2.98	−2.46	1
5	19.51	70.48	−2.753	1.41	0	3
6	14.93	70.48	−1.529	4.20	0.40	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
529	18.29	68.42	−1.544	2.30	0.40	2
530	15.85	69.45	−1.199	3.46	1.05	0

由于飞行参数数据单位不统一，数值范围差异较大，为加强预测模型的泛化能力，加块收敛速度，保证模型的训练和预测效果，采用式(11)对训练集和测试集原始数据进行归一化处理，实现一致性分析：

$$x' = \frac{x-(x_{\max}-x_{\min})}{x_{\max}-x_{\min}} \tag{11}$$

式中：x为样本当前值；$x_{\max}$和$x_{\min}$分别为样本数据的最大值和最小值。

3.3 实验结果讨论与分析

3.3.1 实验结果

本次实验运行环境及参数配置：电脑CPU为2.70 GHz，内存32.0 GB，操作系统为Windows11，编程语言为Python3.11。在实验过程中，设置梯度提升决策树模型训练最大迭代次数为200次，利用GS法对GBDT模型进行优化，通过调整学习率和树的深度，得到不同参数条件下模型的预测性能，如表3所示。使用ACC、MSE、MAE和RMSE作为评估指标，经过反复迭代实验，当最大迭代次数为200次，学习率为0.10，最大深度为4时，RMSE的值最小。

利用测试集数据对模型进行验证，由GBDT-GS模型预测测试集数据，预测准确度为92%，结果如图7所示。图中，0级风险、1级风险、2级风险、3级风险分别为无风险、低风险、中风险、高风险。

表3 不同参数下的预测性能（部分）

Table 3 Prediction performance under different parameters (partial)

参数	ACC	RMSE	MAE	MSE
[200, 0.05, 1]	0.868 1	0.125 8	0.419 6	0.176 1
⋮	⋮	⋮	⋮	⋮
[200, 0.05, 5]	0.869 8	0.159 2	0.469 2	0.221 3
⋮	⋮	⋮	⋮	⋮
[200, 0.10, 3]	0.886 4	0.169 8	0.520 0	0.270 4
[200, 0.10, 4]	0.918 5	0.119 5	0.412 1	0.163 5
[200, 0.10, 5]	0.875 4	0.207 6	0.555 1	0.308 2
⋮	⋮	⋮	⋮	⋮
[200, 0.15, 1]	0.865 8	0.251 6	0.537 9	0.289 3
⋮	⋮	⋮	⋮	⋮
[200, 0.15, 5]	0.875 0	0.213 8	0.560 78	0.314 5

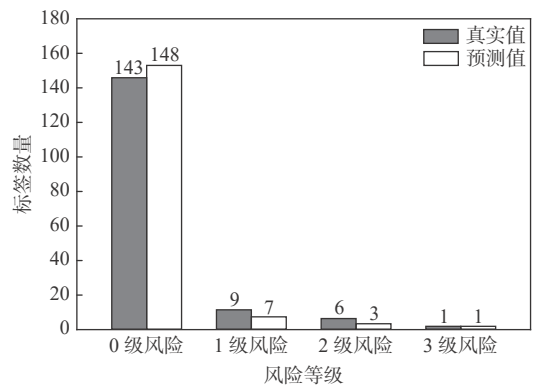


图7 重着陆风险识别结果

Fig. 7 Identification results of hard landing risk

3.3.2 模型对比分析

为验证本文模型的优越性，将GBDT-GS模型与随机森林、Logistic多元回归和RNN构建的模型进行比较。将GBDT-GS的模型参数设定为：最大迭代次数为200，学习率为0.1，最大深度为4；将随机森林的模型参数设定为：树的个数为100，分裂准则为“gini”；将Logistic多元回归的模型参数设定为：L2正则化；将RNN的模型参数设定为：隐含层神经元数目为10，迭代训练次数为10次，核函数采用Adam函数。经过反复实验，以上4种算法训练得到的模型准确率变化对比如图8所示。

由图8可知，GBDT算法结合GS建立的预测模型对重着陆风险预测的准确率比随机森林、Logistic多元回归和RNN模型的准确率高。使用最优参数构建的模型和其进行预测的性能如表4所示。

由表4可知，GBDT-GS模型的RMSE、MAE及MSE值均小于其他3种算法，说明使用GBDT-GS模型预测重着陆风险时，实际值与预测值之间的误差较小，模型的精度和预测效果优于其他算法。从而证

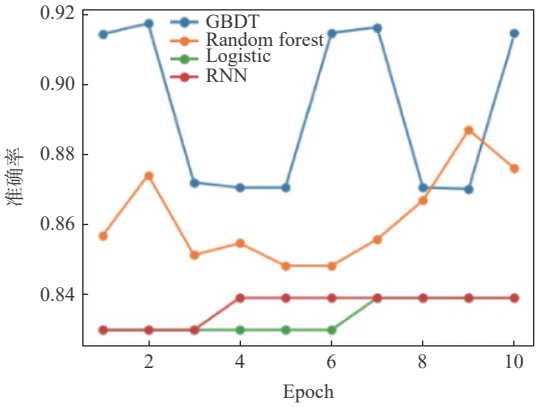


图 8　准确率变化对比

Fig. 8　Accuracy change comparison

表 4　4 种算法预测性能对比

Table 4　Comparison of prediction performance among four algorithms

算法	ACC	RMSE	MAE	MSE
GBDT-GS	0.918 5	0.119 5	0.412 1	0.163 5
随机森林	0.886 8	0.123 2	0.440 4	0.182 4
Logistic多元回归	0.838 9	0.138 4	0.488 9	0.239 0
RNN	0.838 9	0.132 1	0.469 2	0.220 3

明了本文方法的优越性，可以为重着陆风险预测提供有力的方法支持。

提取“成都-沈阳”航段 A320 机型的 QAR 数据，运用 GBTD-GS 模型对最近 2 周 50 个航班重着陆风险进行评估预测，结果如表 5 所示。由表 5 可知，近 2 周 A320 机型运行沈阳机场，有 2 个航班存在重着陆风险，其中，第 25 个航班的重着陆风险等级为“3 级，高风险”，第 49 个样本的风险等级为“1 级，低风险”。针对第 25 个航班，结合具体飞行数据分析，飞机进跑道端口高度偏高（H>18 m），拉平时飞行速度大，将导致飞机能量过大，为在接地区域接地，避免冲出跑道事故，飞行员通过推杆使飞机快速下沉，在后续修正过程中带杆不足，进一步造成飞机下降率大及接地俯仰角小，进而导致飞机垂直加速度大，造成重着陆风险事件。后调取航班航后报告，着陆载荷达到 2.24G，评估结果与实际相符，因此，验证了 GBDT-GS 方法的客观有效性。

表 5　“成都-沈阳”航段近期 QAR 数据

Table 5　Recent QAR data of Chengdu-Shenyang segment

样本	x_1/m	x_2/(m·s^{-1})	x_3/(m·s^{-1})	x_4	x_5	风险等级
1	15.85	70.28	−0.518	4.60	−0.35	0
2	17.68	70.93	−1.463	4.22	−1.05	0
3	16.76	69.19	−0.833	4.92	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
25	23.16	73.89	−2.946	−1.10	−0.35	3
26	14.93	69.79	−1.488	4.04	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
49	16.76	70.74	−1.951	3.03	4.22	1
50	16.46	69.98	−0.523	4.13	0	0

根据调研，绝大多数重着陆事件的发生都与进近着陆飞行操作密切相关，而飞行操作直接影响飞机飞行状态的变化，造成飞机能量异常。因此，在进近着陆过程中，对飞机能量进行控制是降低重着陆风险的有效手段。飞机能量管理主要包括动能管理和势能管理，其中，动能管理主要是对着陆飞行速度及飞行姿态的控制；势能管理主要是对飞机高度及下降率的控制，需要飞行员通过控制操作面（侧杆、方向舵等）和油门杆协调配合完成。在进近阶段，飞行员应在 300 m 左右使飞机处于稳定状态，并在飞行状态参数出现偏差后及时进行修正；下滑过程中，宜以固定的下滑点、合适的下滑曲线、稳定的下滑速度和均匀的下降率下滑，避免过跑道端口时高度过高、下降率过大；在拉平飘落过程中，应综合判断飞机下沉趋势，逐步带杆，柔和增加俯仰；若飞机过早拉平或拉飘后，不应为追求接地点而前推驾驶杆。

4　结　论

1）本文从重着陆事件发生的物理本质出发，针对飞机进近着陆过程，通过飞机受力分析并构建着陆飞行运动学方程，确定过跑道端口高度、飞行速度、下降率、俯仰角、滚转角等 5 项飞行状态参数与重着陆事件密切相关。

2）通过 GBDT 算法构建了重着陆风险预测模型，并利用 GS 优化了模型参数；利用大量真实数据对模型进行训练和测试，并与随机森林、Logistic 多元回归和 RNN 模型对比。结果表明，GBDT-GS 方法在重着陆风险预测中的准确率和预测性能都高于其他模型，验证了本文模型的可用性。

3）通过 QAR 数据，利用 GBDT-GS 模型评估 A320 机型运行沈阳机场的重着陆风险，得出准确率为 92%，进一步验证了本文模型识别和预测重着陆风险的客观有效性。

本文分析了各飞行状态参数对重着陆风险的影响关系，而飞行操作输入、气象环境等数据对重着陆的影响均反映在飞行状态变化上，因此，下一步需要分别建模分析飞行操作风险、气象环境风险对飞机着陆飞行状态的影响，以便深入识别重着陆间接风险因素，为重着陆风险控制提供指导。

参考文献（References）

[1] International Civil Aviation Organization. Aircraft operations[M].

5th ed. Montreal: International Civil Aviation Organization, 2006.

[2] The Boeing Company. Statistical summary of commercial jet airplane accidents: worldwide operations | 1959-2024[R]. Chicago: The Boeing Company, 2025.

[3] International Air Transport Association. IATA safety report 2018[R]. Montreal: International Air Transport Association, 2019.

[4] 王冉. 基于飞行数据的颠簸飞行状态估计初步研究[D]. 南京: 南京航空航天大学, 2020.
WANG R. Preliminary study on bumpy flight state estimation based on flight data[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2020(in Chinese).

[5] 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).

[6] SARTOR P, BECKER W, WORDEN K, et al. Bayesian sensitivity analysis of flight parameters in a hard-landing analysis process[J]. Journal of Aircraft, 2016, 53(5): 1317-1331.

[7] 聂磊, 黄圣国, 舒平, 等. 基于支持向量机 (SVM) 的民用飞机重着陆智能诊断研究[J]. 中国安全科学学报, 2009, 19(7): 149-153.
NIE L, HUANG S G, SHU P, et al. Intelligent diagnosis for hard landing of aircraft based on SVM[J]. China Safety Science Journal, 2009, 19(7): 149-153(in Chinese).

[8] 许桂梅, 黄圣国. 应用 LS-SVM 的飞机重着陆诊断[J]. 系统工程理论与实践, 2010, 30(4): 763-768.
XU G M, HUANG S G. Airplane's hard landing diagnosis using LS-SVM[J]. Systems Engineering-Theory & Practice, 2010, 30(4): 763-768(in Chinese).

[9] 曹海鹏, 舒平, 黄圣国. 基于神经网络的民用飞机重着陆诊断技术研究[J]. 计算机测量与控制, 2008, 16(7): 906-908.
CAO H P, SHU P, HUANG S G. Study of aircraft hard landing diagnosis based on nerual network[J]. Computer Measurement & Control, 2008, 16(7): 906-908(in Chinese).

[10] 于健, 董健康, 李佳. 基于神经网络的飞机着陆垂直载荷大诊断技术研究[J]. 中国民航飞行学院学报, 2010, 21(4): 16-18.
YU J, DONG J K, LI J. Research on large diagnosis technology of aircraft landing vertical load based on neural network[J]. Journal of Civil Aviation Flight University of China, 2010, 21(4): 16-18(in Chinese).

[11] 郑薇. 基于 QAR 数据的重着陆风险评估及预测研究[D]. 天津: 中国民航大学, 2014.
ZHENG W. Research on risk assessment and prediction of heavy landing based on QAR data[D]. Tianjin: Civil Aviation University of China, 2014(in Chinese).

[12] 周进. 基于熵权 TOPSIS 的飞机重着陆风险因素排序[J]. 河南科技, 2019, 38(19): 106-110.
ZHOU J. Risk factors sequencing of aircraft hard landing based on entropy weight and topsis method[J]. Henan Science and Technology, 2019, 38(19): 106-110(in Chinese).

[13] 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).

[14] 孙瑞山, 张思远, 汪磊. 基于统计过程控制的 QAR 数据风险预测方法[J]. 统计与决策, 2015, 31(21): 33-36.
SUN R S, ZHANG S Y, WANG L. QAR data risk prediction method based on statistical process control[J]. Statistics & Decision, 2015, 31(21): 33-36(in Chinese).

[15] 郑磊, 池宏, 许保光, 等. 飞机重着陆预警分析方法[J]. 数学的实践与认识, 2019, 49(3): 56-72.
ZHENG L, CHI H, XU B G, et al. Method of early warning analysis for aircraft hard landing[J]. Mathematics in Practice and Theory, 2019, 49(3): 56-72(in Chinese).

[16] 陈思, 孙有朝, 郑敏. 基于支持向量机的飞机重着陆风险预警模型[J]. 兵器装备工程学报, 2019, 40(9): 154-158.
CHEN S, SUN Y C, ZHENG M. Heavy landing risk pre-warning model based on support vector machine[J]. Journal of Ordnance Equipment Engineering, 2019, 40(9): 154-158(in Chinese).

[17] 阙向东. 飞机飞行状态估计与飞行品质评估研究[D]. 西安: 西北工业大学, 2004.
QUE X D. Research on aircraft flight state estimation and flight quality evaluation[D]. Xi'an: Northwestern Polytechnical University, 2004(in Chinese).

[18] 耿建中, 姚海林. 基于 UKF 的飞机飞行状态估计[C]//2008 系统仿真技术及其应用学术会议论文集. 合肥: 中国科学技术大学出版社, 2008: 68-71.
GENG J Z, YAO H L. Aircraft flight state estimation based on unscented Kalman filter[C]//Proceedings of the 2008 Academic Conference on System Simulation Technology and Applications. Hefei: Press of University of Science and Technology of China, 2008: 68-71(in Chinese).

[19] COOMBES M, LIU C J, CHEN W H. Situation awareness for UAV operating in terminal areas using bearing-only observations and circuit flight rules[C]//Proceedings of the American Control Conference. Piscataway: IEEE Press, 2016: 479-485.

[20] 孙禾. 航空器飞行状态预测方法研究[D]. 天津: 中国民航大学, 2014.
SUN H. Research on prediction method of aircraft flight state[D]. Tianjin: Civil Aviation University of China, 2014(in Chinese).

[21] RAOL J, JATEGAONKAR R. Aircraft parameter estimation using recurrent neural networks——A critical appraisal[C]//Proceedings of the 20th Atmospheric Flight Mechanics Conference. Reston: AIAA, 1995.

[22] LI L S, GARIEL M, HANSMAN R J, et al. Anomaly detection in onboard-recorded flight data using cluster analysis[C]//Proceedings of the IEEE/AIAA 30th Digital Avionics Systems Conference. Piscataway: IEEE Press, 2011: 1-31.

[23] LEE H, PURANIK T G, FISCHER O P, et al. Flight data driven system identification using neural networks for landing safety assessment[C]//Proceedings of the IEEE/AIAA 40th Digital Avionics Systems Conference. Piscataway: IEEE Press, 2021: 1-9.

[24] 王晴昊, 胡剑波, 姚登凯. STPA 在进近着陆飞行安全分析中的研究及应用[J]. 系统工程理论与实践, 2018, 38(10): 2703-2712.
WANG Q H, HU J B, YAO D K. Research and application of STPA to flight safety analysis in approach and landing[J]. Systems Engineering-Theory & Practice, 2018, 38(10): 2703-2712(in Chinese).

[25] 刘琼, 张豹. 基于 GBDT 算法的锂电池剩余使用寿命预测[J]. 电

子测量与仪器学报, 2022, 36(10): 166-172.
LIU Q, ZHANG B. Remaining useful lifetime prediction for lithiumbattery based on GBDT algorithm[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(10): 166-172(in Chinese).
[26] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. The Annals of Statistics, 2001, 29(5): 1189-1232.
[27] 奉国和. SVM 分类核函数及参数选择比较[J]. 计算机工程与应用, 2011, 47(3): 123-124.
FENG G H. Parameter optimizing for support vector machines classification[J]. Computer Engineering and Applications, 2011, 47(3): 123-124(in Chinese).
[28] 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[S]. 北京: 中国民用航空局飞行标准司, 2015-06-30.
Civil Aviation Administration of China. Implementation and management of flight quality control (FOQA): AC-121/135-FS-2012-45R1[S]. Beijing: Flight Standards Department of Civil Aviation Administration of China, 2015-06-30.

Hard landing risk prediction of civil aircraft based on GBDT-GS method

WANG Xiangzhang[1,*], MOU Ruifang[1], WANG He[2], XU Bohao[2]

(1. School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China;
2. Flight Department, Sichuan Airlines Co., Ltd., Chengdu 610200, China)

Abstract: Hard landing may cause structural damage of aircraft or other potential accident causes and even crash and fatal flight accidents. In view of the lack of physical nature analysis in current hard landing risk assessment, combined with flight status analysis, a hard landing risk prediction model based on gradient boosting decision tree (GBDT) and grid search (GS) was proposed to effectively implement hard landing risk identification and grade criteria and improve pilots' landing operation quality. Firstly, the flight kinematics equation of landing was established through the force analysis of aircraft, and five flight status parameters closely related to hard landing were determined. Then, flight status data was extracted from onboard quick access recorder (QAR) data to construct a data set. According to QAR parameter characteristics, the hard landing risk prediction model was constructed by the GBDT algorithm, and model parameters were optimized by GS. Finally, taking the Chengdu-Shenyang route operation of an airline as an example, the study selected 530 pieces of QAR data to train and test the model and compared the result of the model with those of random forest, recurrent neural networks (RNN), and Logistic multiple regression. The results show that the GBDT-GS method is better than other algorithms in predicting hard landing risk, and its prediction accuracy reaches 92%, which verifies the objective validity of the constructed model.

Keywords: flight status; hard landing; risk prediction; QAR data; gradient boosting decision tree

Received: 2023-07-05; **Accepted**: 2023-09-10; **Published Online**: 2023-09-13 16:46
URL: link.cnki.net/urlid/11.2625.V.20230912.1820.001
Foundation item: Civil Aviation Safety Capacity Building Fund project (ASSA202019)
* **Corresponding author**. E-mail: 513432965@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0418

基于U-Net++和特征融合的塑料齿轮复杂黑点检测方法

方一鸣，石照耀*，宋辉旭
(北京工业大学 北京市精密测控技术与仪器工程技术研究中心，北京 100124)

摘 要：传统的缺陷检测算法在检测塑料齿轮表面复杂黑点时效果不佳，主要问题是对齿轮边缘上黑点的大小和位置判别不准确、对浅色黑点的漏检率高、易将点浇口误判为黑点。为此，提出一种基于U-Net++和特征融合的塑料齿轮复杂黑点检测方法。所提方法通过U-Net++预测黑点区域；根据梯度特征对黑点区域进行修正；结合多特征融合分析给出最终判定结果，提高了对复杂黑点检测的准确性和稳定性。测试结果表明：所提方法表征检测结果准确性的P_c值达到了98.93%，表征分割结果准确性的交并比平均值达到了0.864，相比传统的缺陷检测算法和未做修正的深度学习算法，交并比的平均值分别提高了0.478和0.309。

关 键 词：塑料齿轮缺陷；深度学习；U-Net++；多特征融合；局部区域修正
中图分类号：TP249
文献标志码：A **文章编号**：1001-5965(2025)09-3020-10

塑料齿轮在注塑成型和上下料过程中由于原材料混入杂质、局部碳化、模具内部含有油污、传送带/料框不洁净等问题，易在齿轮表面形成与周围区域相比在亮度、颜色等属性上有差异的区域，呈现出黑色点状杂质，这类缺陷称为塑料齿轮的黑点缺陷[1]。此类缺陷即使在注塑工艺稳定后依然频繁出现，是塑料齿轮最常见的缺陷。

随着塑料齿轮的应用愈加广泛，下游产业对塑料齿轮的外观要求也越来越高，塑料齿轮生产企业都有严格的外观检验规范，然而企业目前仍普遍采用人工目视检测黑点，行业缺乏有效的塑料齿轮黑点缺陷在线检测技术。随着机器视觉技术的不断发展，基于机器视觉技术的塑料齿轮黑点缺陷在线检测成为一种可行的方案[2]。

塑料齿轮黑点缺陷现阶段主要有3种检测方法。第1种是基于微分算子，这种方法通过检测图像中的极值点确定缺陷的位置，常见的微分算子有高斯拉普拉斯算子(Laplacian of Gaussian, LoG)、高斯差分算子(difference of Gaussian, DoG)和海森矩阵行列式算子(determinant of Hessian, DoH)[3]。文献[4]通过Hessian矩阵分析获取极值点(特征点)，再通过灰度差分算法对经过矩阵变换配准后的图像进行缺陷区域提取，并对缺陷区域进行Blob分析。然而，基于图像对位置梯度的微分方法，在单一特征尺度下只能检测对应半径的斑点，要检测不同尺寸的斑点，需要进行多尺度分析，但这种方法运算复杂，易受噪声影响。第2种是基于分水岭算法，文献[5]提出采用基于形态学梯度重构和内外标记的分水岭算法对表面缺陷进行分割，对水果表面的黑斑、疤伤等缺陷的识别率能达到96.8%。文献[6]采用分水岭算法准确分割出焊点，并按面积等特征对焊点缺陷进行分类。分水岭算法通过阈

收稿日期：2023-06-28；**录用日期**：2023-10-25；**网络出版时间**：2023-11-01 12:52
网络出版地址：link.cnki.net/urlid/11.2625.V.20231101.0843.001
基金项目：国家自然科学基金重大科研仪器研制项目(52227809)；精密测试技术及仪器国家重点实验室(天津大学)(PILAB2105)；中国科协青年人才托举工程(2021QNRC001)
* **通信作者**. E-mail：shizhaoyao@bjut.edu.cn

引用格式：方一鸣，石照耀，宋辉旭. 基于U-Net++和特征融合的塑料齿轮复杂黑点检测方法[J]. 北京航空航天大学学报，2025，51(9)：3020-3029. FANG Y M，SHI Z Y，SONG H X. Detection method for complex dark spots on plastic gears based on U-Net++ and feature fusion[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3020-3029 (in Chinese).

值来分割斑点, 但是阈值需要人为设定, 通用性不佳; 当黑点位于齿轮边缘轮廓上时, 分割的偏差较大。第 3 种是基于深度学习, 前 2 种方法均属于传统的图像处理方法, 普遍对于特征不明显、形状不规则和干扰较多的缺陷会存在较高的误检率。近年来, 不少基于深度学习的缺陷检测方法被广泛应用在各种工业场景中, 其拥有对颜色、亮度、对比度等更稳定的检测结果[7]。文献 [8] 基于 YOLO-v3 网络实现了对金属齿轮端面凸起、凹陷和划痕 3 种缺陷的快速检测, 平均精度均值和平均召回率分别达到了 93% 和 91%。文献 [9] 提出一种基于改进的 YOLO-v3 网络的齿轮缺陷检测方法, 平均精度均值比 YOLO-v3 网络提高了 3.87%, 对齿轮缺失部分检测的平均精度均值提高了 5.7%。文献 [10] 将 YOLO-v4 网络和改进分水岭算法结合对绝缘子进行精准的识别与定位, 有效弥补了传统方法在复杂背景下识别能力低下的不足。

可以看出, 当下应用于齿轮缺陷检测的深度学习算法主要为目标检测算法(YOLO 算法), 而鲜有采用语义分割算法对齿轮缺陷区域进行精确分割的研究。目标检测算法只能框出缺陷但无法获取缺陷形状, 而 U-Net[11] 作为一种先进的语义分割网络, 不仅能精确定位缺陷位置, 还能分割出缺陷形状, 进而根据缺陷形状分析追溯缺陷来源。Zhou 等[12] 对 U-Net 进行拓展形成了 U-Net++, 通过在编码器和解码器之间加入 Dense block 和卷积层来提高分割精度, 可在数据样本较少的情况下实现高性能和准确性。然而, 由于卷积网络的上采样结果是模糊平滑的, 对图像中的细节区域不敏感, 因此, 对于物体边缘的分割效果不理想, 对于面积较小的缺陷分割效果差。

本文提出基于 U-Net++和特征融合的塑料齿轮复杂黑点检测方法, 首先, 利用 U-Net++预测黑点区域位置。然后, 在该区域基于图像梯度特征进行修正。最后, 对修正后的缺陷区域进行多特征置信度加权融合分析, 根据分析结果完成最终判定。该方法能够有效改善塑料齿轮复杂黑点检测的准确性, 对塑料齿轮工艺溯源工作具有重要意义。

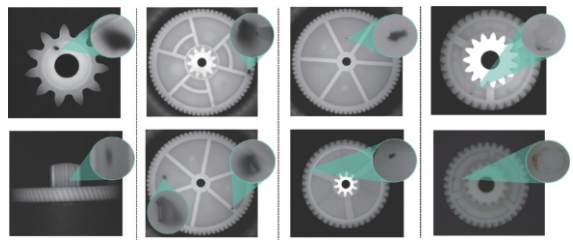
1　基于 U-Net++的黑点检测方法

1.1　黑点检测的复杂性

1.1.1　黑点分类

塑料齿轮黑点缺陷如图 1 所示, 可分为以下 4 类:

1) 成型前混入到原料的内部杂质, 在塑料齿轮表面呈现出黑色斑点, 大小、形状各异, 颜色深浅不一;



(a) 材料含杂　(b) 局部碳化　(c) 金属粉末、料屑　(d) 油污

图 1　塑料齿轮黑点缺陷的 4 种类型(按缺陷成因分类)

Fig. 1　Four types of dark spot defects of plastic gears (classified by the cause of defect)

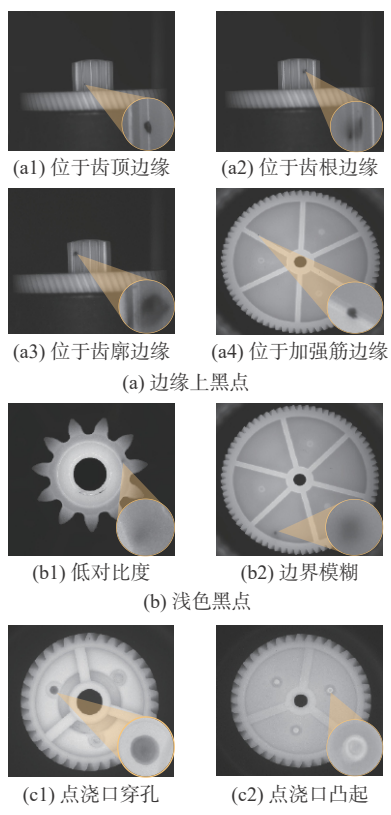
2) 原料长期积料或局部受热焦化/碳化形成的黑点, 这种黑点一般面积较大;

3) 成型后因外界原因产生的异色杂质, 如经过传送带后黏附在塑料齿轮表面的铁粉、铁锈和料屑;

4) 从模具滑块缝隙或顶针缝隙黏附的油污。

1.1.2　黑点检测的难题

塑料齿轮的黑点缺陷尺度变化大、形状各异, 有的与背景差异小、对比度低、边界模糊, 有的周围区域存在大量噪声, 还有一些塑料齿轮独有的特性形成的干扰(如点浇口), 这时传统的缺陷检测方法就会出现大量漏检、误检的情况。图 2 为传统缺陷检测算法面临的 3 个难题。



(a1) 位于齿顶边缘　(a2) 位于齿根边缘

(a3) 位于齿廓边缘　(a4) 位于加强筋边缘

(a) 边缘上黑点

(b1) 低对比度　(b2) 边界模糊

(b) 浅色黑点

(c1) 点浇口穿孔　(c2) 点浇口凸起

(c) 点浇口的干扰

图 2　传统缺陷检测算法面临的 3 个难题

Fig. 2　Three problems faced by traditional defect detection algorithms

1）塑料齿轮黑点的位置具有随机性和不确定性，当其位于塑料齿轮的齿顶、齿根、齿廓和加强筋等边缘轮廓上时，受边界线和背景的干扰，易造成黑点个数、面积和形状的误判；

2）浅色黑点边缘模糊，与背景的对比度低，相比于周围灰度值没有明显的突变，传统的缺陷检测算法很难准确识别边界，也很难通过阈值进行分割；

3）相比金属齿轮，塑料齿轮端面的浇口是其特有的性质，浇口是塑料齿轮浇道的最始端，特别是点浇口，与黑点的大小和形状极为相似，很难通过灰度和形状等特征过滤掉。此外，点浇口还会出现凸起和穿孔等复杂情况，使得筛除的难度更大。

为解决上述3个难题，本文构建了基于深度学习语义分割网络的U-Net++对塑料齿轮黑点缺陷进行检测和分割，并对模型进行优化设计。

1.2 基于U-Net++的塑齿黑点检测模型

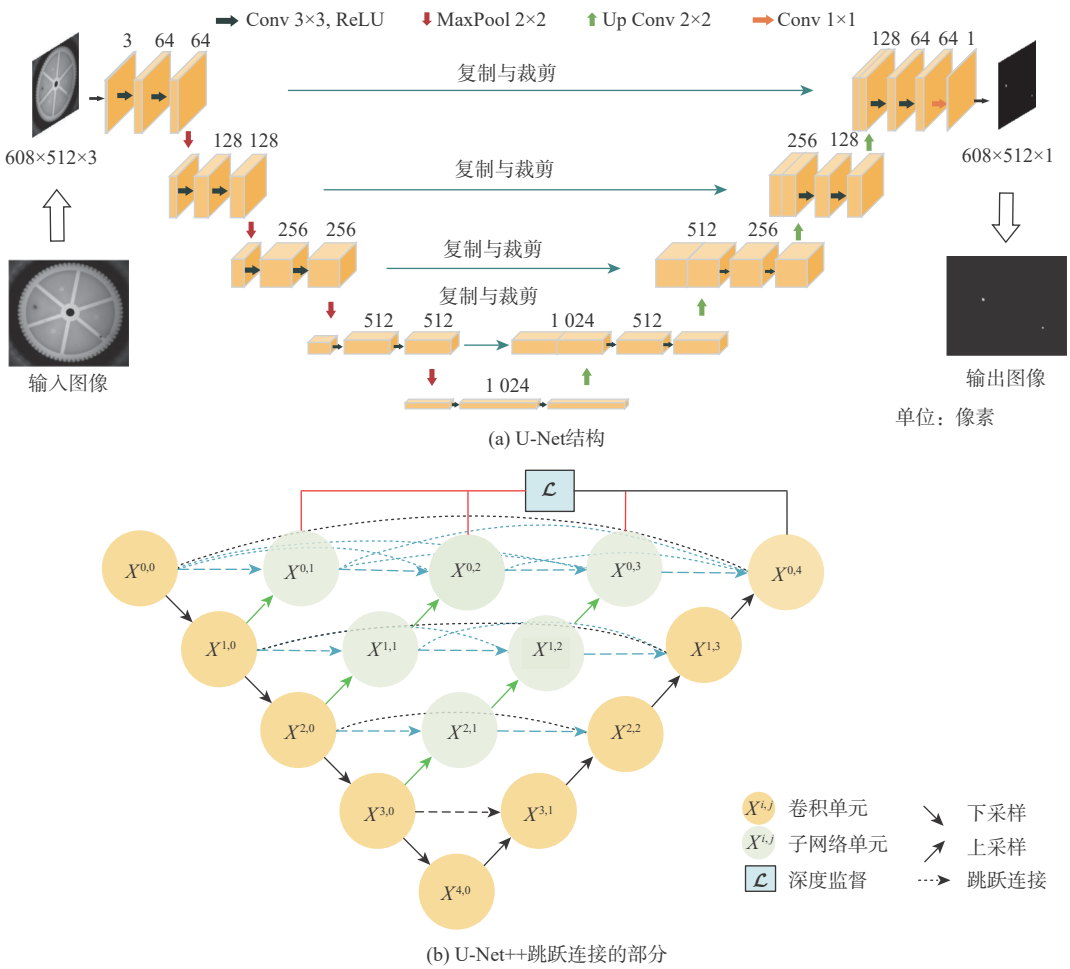
1.2.1 U-Net++模型结构

U-Net++的前身是U-Net，U-Net借鉴了全卷积网络（fully convolutional networks，FCN）[13]，如图3(a)所示，该网络结构包括收缩路径和扩展路径2个对称部分。以本文基于U-Net++的塑料齿轮黑点缺陷检测模型为例，将初始608像素×512像素×3像素的彩色塑料齿轮图像输入U-Net++模型中，网络的左侧部分（压缩路径）由卷积和最大池化构成一系列降采样操作，由4个块（block）组成，每个块使用了3个有效卷积和1个最大池化降采样，每次降采样之后特征图（feature map）的个数乘2；右侧部分（扩展路径）同样由4个块组成，每个块开始之前通过反卷积将特征图的尺寸乘2，同时将其个数减半，然后和左侧对称的压缩路径的特征图合并，最终输出含有缺陷特征的608像素×512像素×1像素的灰度图像。

U-Net++相比U-Net，增加了多个跳跃路径，如图3(b)所示，跳跃路径上的所有卷积层使用大小为3×3的核，图中黑色部分为原始的U-Net结构，绿色部分为嵌套的U-Net子网络，蓝色虚线部分是跳跃路径，红黑连线则是U-Net++补充的深监督机制，使网络能够顺利得到训练[12]。

1.2.2 Z-Score标准化

塑料齿轮黑点缺陷图片训练集一些数据的计算量大、复杂度高、不容易收敛，难以进行统计分析。为对数据做一些符合正态分布的数学分析，有必要对数据进行标准化，把数据映射到0～1范围内。



(a) U-Net结构

(b) U-Net++跳跃连接的部分

图3 用于塑料齿轮黑点缺陷检测的U-Net++结构

Fig. 3 U-Net++ structure for dark spot defect detection in plastic gears

采用 Z-Score 标准化方法对预处理数据进行标准化, 可以将特征数据的分布调整成标准正态分布, 使得数据的均值为 0, 方差为 1。其过程分为 2 步: ①去均值的中心化(均值变为 0); ②方差的规模化(方差变为 1)。

Z-Score 标准化转换表达式为

$$x_* = (x-\mu)/\sigma \tag{1}$$

式中: x 为原始数据; μ 为总体数据的均值; σ 为总体数据的标准差。

Z-Score 标准化的原理如图 4 所示, 图 4 (a) 中蓝色曲线为 Sigmoid 函数, 输入数据用紫色小圆点表示, 当没有经过任何处理的输入数据在梯度很小的区域, 学习率就会很慢甚至陷入长时间的停滞。经过减均值再除以方差后, 数据才被移到中心区域, 这可以看做是一种对抗梯度消失的有效手段。图 4 (b) 体现了 Z-Score 标准化, 在输入层之前给每个神经元做一次标准化, 这个方法应用于数据预处理, 加快了梯度下降求最优解的速度, 并且把有量纲表达式变为无量纲表达式, 去除了数据的单位限制, 便于不同单位或量级的指标能够进行比较和加权, 使数据具有可比性。

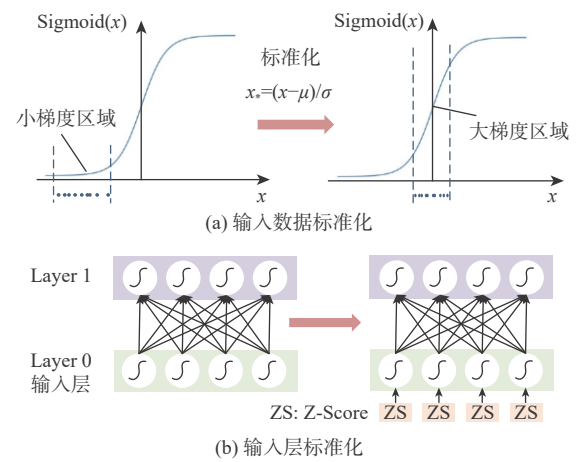


图 4　Z-Score 标准化

Fig. 4　Z-Score standardization

1.2.3　评价方法

利用 U-Net++进行训练时需要一些指标来评价训练结果的好坏。本文采用以下几个指标作为评价指标。

1) 损失函数

为指导训练始终向着正确的方向进行, 需要计算神经网络每次迭代的前向计算结果与真实值的差距, 也就是损失函数。DiceLoss 函数可作为损失函数, 定义为

$$D_{\mathrm{L}} = 1 - \frac{2T_{\mathrm{P}}}{2T_{\mathrm{P}} + F_{\mathrm{N}} + F_{\mathrm{P}}} \tag{2}$$

式中: T_{P}、T_{N}、F_{N} 和 F_{P} 分别为真阳性、真阴性、假阴性和假阳性区域的像素个数。

2) 交并比

交并比(intersection over union, IoU)是指该类的真实标签和预测值的交集和并集的比值, 单类的交并比的计算式为

$$I_{\mathrm{oU}} = \frac{T_{\mathrm{P}}}{T_{\mathrm{P}} + F_{\mathrm{N}} + F_{\mathrm{P}}} \tag{3}$$

交并比反映 U-Net++分割的准确性, 数值越接近于 1, 说明网络预测的区域与实际黑点区域重合度越高, 分割得越准确。

3) 召回率

召回率为样本中所有正例中被正确预测出来的比率, 用于评估检测器对所有待检测目标的检测覆盖率, 表达式为

$$R = \frac{T_{\mathrm{P}}}{T_{\mathrm{P}} + F_{\mathrm{N}}} \tag{4}$$

4) 自定义的 P_{c} 指标

P_{c} 为自定义的一个评价指标, 其用来表示经过 U-Net++分割出来的区域质心落在正确黑点区域内的概率, 表达式为

$$P_{\mathrm{c}} = \frac{C_{\mathrm{T}}}{C_{\mathrm{T}} + C_{\mathrm{F}} + C_{\mathrm{M}}} \tag{5}$$

式中: C_{T} 为质心落在黑点区域内的个数; C_{F} 为质心不在黑点区域的个数; C_{M} 为漏检的个数, 也即实际的黑点区域内没有被检测到的个数。P_{c} 指标是考察将全部的黑点都正确检测出的能力, 这个指标的好坏是后续进行局部区域修正的关键。

2　融合多特征的局部区域修正方法

基于 U-Net++的塑料齿轮复杂黑点检测方法相比传统的缺陷检测方法在检测准确率上得到极大提升, 但是分割结果的准确性一般。由于塑料齿轮的黑点缺陷面积普遍较小, 在图中所占的像素比例过低, 在利用反向传播算法更新网络参数时, 其对参数更新量的影响容易被面积较大缺陷区域的影响覆盖, 小区域缺陷的特征难以被很好的学习。与此同时, 被错误标记的区域由于像素往往较少, 所以在训练中也难以被正确区分。此外, 样本不均衡也会造成分割结果不准。为让最终的分割结果尽可能地接近人工标记的结果, 建立基于多特征融合分析的局部区域修正方法, 来改善黑点的分割效果。

2.1　基于局部图像梯度特征的区域生长法

图像梯度特征是分析缺陷区域的一个重要特征, 但传统的缺陷检测方法很少考虑局部区域的梯度特征, 本文基于局部图像梯度特征的区域生长法对 U-Net++预测的黑点缺陷区域进行修正, 该方法

的思想是将U-Net++预测出的区域外包矩形作为局部图像，再根据梯度特征作二次分割。

首先，将U-Net++预测出的区域作为候选黑点区域，根据候选黑点区域所包含的像素点集合和图像矩的概念计算出候选黑点区域的质心。

图像矩是图像的某些特定像素灰度的加权平均值，假设一幅$M\times N$大小的图像$f(x,y)$，其$p+q$阶几何矩m_{pq}为

$$m_{pq}=\sum_{x=1}^{M}\sum_{y=1}^{N}x^p y^q f(x,y) \tag{6}$$

1阶矩表示横坐标x和对应像素值$f(x,y)$的乘积，以及纵坐标y和对应像素值$f(x,y)$的乘积，表达式为

$$m_{10}=\sum_{x=1}^{M}\sum_{y=1}^{N}x\cdot f(x,y) \tag{7}$$

$$m_{01}=\sum_{x=1}^{M}\sum_{y=1}^{N}y\cdot f(x,y) \tag{8}$$

对于二值图像来说，0阶矩m_{00}等于其面积。假设第i个候选黑点A_i的质心坐标为$(\bar{x}_i,\bar{y}_i)$，则其计算式为

$$(\bar{x}_i,\bar{y}_i)=\left(\frac{m_{10}}{m_{00}},\frac{m_{01}}{m_{00}}\right) \tag{9}$$

训练好U-Net++塑料齿轮黑点检测模型后，将待检的塑料齿轮图片喂进模型里，如图5(a)所示，输出的预测结果将是概率图，本文将输出的概率图转换为热力图，如图5(b)所示，这样有助于发现分割结果与网络结构的关系、找出极值，也能够刻画出数据的整体样貌，方便在数据集之间进行比较。

将每个候选黑点区域A_i(见图5(c))的最小外接矩形作为区域生长法的局部图像$P_i(x,y)$，种子点在$P_i(x,y)$局部图像中选取，可选择候选黑点区域A_i的质心$(\bar{x}_i,\bar{y}_i)$作为种子点，或选择邻域内灰度值在一定范围内的多个像素点作为种子点。本文将A_i的质心$(\bar{x}_i,\bar{y}_i)$作为种子点。

接着，给出相似性准则，即生长准则，如果采用灰度差的评判标准鲁棒性会很差，本文采用的评判标准是梯度特征，包括幅值和方向[14]。将种子点与其八邻域像素点进行梯度特征的相似性比较，确定是否将邻域内的像素点加入种子点集合并形成新的种子点，如图5(d)和图5(e)所示，相似点则作为下次生长的新的种子点，八邻域内没有合并的像素点时，即没有新的种子点时，停止生长。

依次迭代最终将所有符合条件的像素点生长成为一个连通区域，这个连通区域即为最终分割出来的黑点区域B_i，最终得到的分割结果如图5(f)所示，图中用绿色标注的轮廓即为最终的黑点区域B_i的轮廓。

2.2 多特征置信度融合分析

除2.1节所述的图像梯度特征，塑料齿轮黑点区域还包含大量的几何特征、位置特征和形状特征，这些特征既相互独立，又相互关联。在2.1节方法中仅仅考虑了颜色(灰度)特征，而忽视了这些同样重要的特征。在实际使用中，不同的产品齿轮、不同的端面的黑点都会有不同的判定标准，这就要求根据不同的特征进行判定。例如，塑料齿轮通常看不到的部位(底面)一些面积很小的黑点就不需要被判定为缺陷，但在肉眼能看到的、影响成品外观形象的部位(正面、上面)则需要判定为缺陷；还有一些与黑点具有相似灰度、纹理的干扰区域也会被当做是黑点区域被提取出来(如点浇口)，但该区域的形状和位置特征具有一定规律，可以通过多种特征综合判断进行筛除。本文提出一种多特征置信度融合分析方法，根据二次分割出的黑点区域B_i的多特征数值及其置信度加权融合计算的结果完成对黑点缺陷的最终筛选。

所涉及的特征值涵盖面积S、质心$(\bar{x},\bar{y})$、半径r、周长L、质心处灰度值$f(x,y)$、圆度C、凸度V、惯性

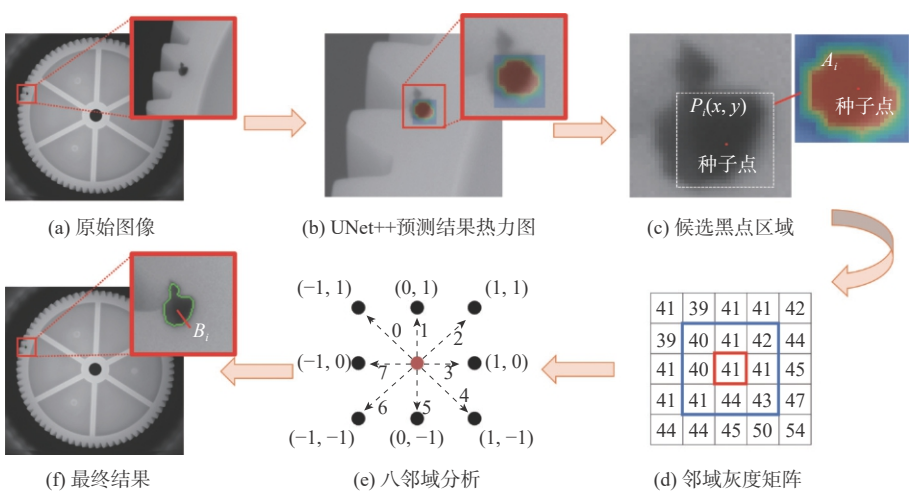


图5 基于局部图像梯度特征的区域生长法

Fig. 5 Region growth method based on local image gradient features

率 I 和偏心率 E 等。任意单一特征为依据均无法很好地筛选塑料齿轮黑点缺陷, 且不同特征对判定黑点的贡献度也有所不同。为弥补通过单一特征检测黑点的局限性, 对黑点缺陷图片数据集的检测结果进行特征分析, 建立特征数据库, 根据特征数据库确定不同特征参数的不同数值范围的概率, 也即区间的置信度, 按置信度高低对各个特征参数分配权重, 将各特征加权融合后的结果作为黑点判定的依据。

假设待判别的黑点区域 B_i 根据显著性特征序列提取的特征集合为 $\{t_1,t_2,\cdots,t_n\}$, 多特征融合的置信度 $c(t_1,t_2,\cdots,t_n)$ 与显著性水平 α_n 成反比关系。当 $c_n=1-\alpha_n>0.5$ 时, 置信度应该得到增强[15], 按照这一理念, 则多特征融合置信度的计算式为

$$c(t_1,t_2,\cdots,t_n)=1-(1-c_1)(1-c_2)\cdots(1-c_n)=1-\alpha_1\alpha_2\cdots\alpha_n \tag{10}$$

所有特征在进行加权融合时有先后顺序, 要先计算最显著的特征的置信度, 若置信度不低于门限, 则提取第 2 个显著特征, 根据式(10)计算融合的置信度; 当置信度低于某一门限时, 置信度更新过程可以提前停止。具体算法流程如图 6 所示, 图中, $c_{\min}$ 为提前结束阈值, c_t 为最终决策阈值。

图 7 提供了 2 个案例分析, 图中, 右下角为 $t_1\sim t_9$

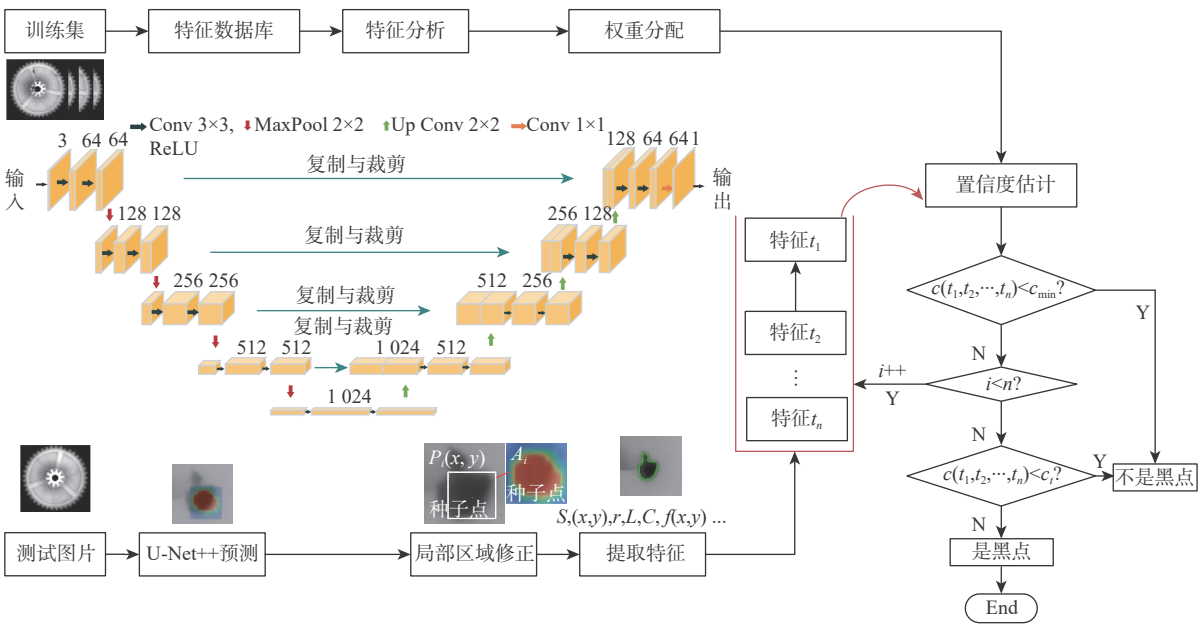


图 6　多特征融合置信度分析

Fig. 6　Confidence analysis of multi-feature fusion

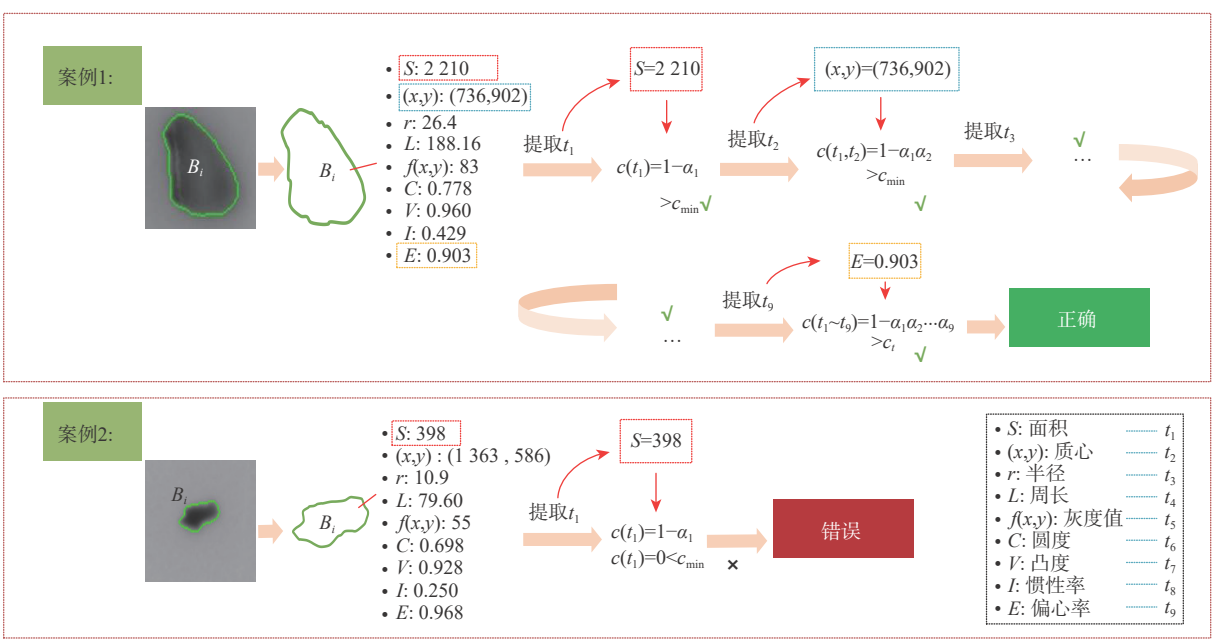


图 7　多特征置信度融合分析案例

Fig. 7　Analysis case of multi-feature confidence fusion

共 9 个特征值的解释，并按照显著性顺序排序。案例 1 在按顺序计算 9 个特征值融合置信度的过程中始终大于用于提前结束计算的阈值$c_{\min}$，最终的融合置信度$c(t_1,t_2,\cdots,t_9)$大于决策判定的阈值c_t，所以该区域最终被判定为是一个黑点缺陷；案例 2 在计算第一个显著性特征“面积”时，就已经小于$c_{\min}$，计算提前结束，并且判定该区域不是黑点缺陷。

3 测试结果

实验测试设备如图 8 所示，正面、底面和侧面的图像采集机构均由工业千兆以太网相机（分辨率 2 448 像素×2 048 像素、23.7 帧/s、工业远心镜头和光源）组成，塑料齿轮通过玻璃转盘依次经过各个图像采集机构，图像采集机构将采集到的塑料齿轮正面、底面和侧面的图像传输给上位机进行存储和处理。

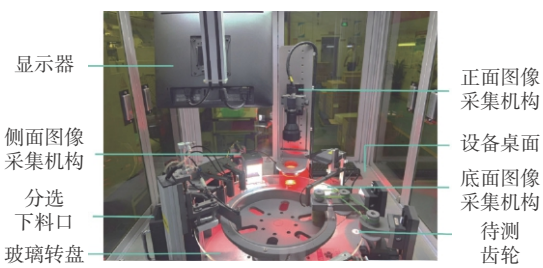


图 8 塑料齿轮黑点检测实验设备

Fig. 8 Experimental equipment for dark spot detection in plastic gears

相比传统的缺陷检测算法，本文方法具有较明显的优势，主要体现在对于 3 个难题的检测准确率有极大提升，图 9 将本文方法与传统的缺陷检测算法（基于局部极值的分水岭算法）作了对比，证实了这一点。

采集了 201 张正面、底面和侧面的塑料齿轮图像组成数据集，数据集涵盖图 1 所示的黑点缺陷 4 种类型，以及图 2 给出的 3 种传统算法难检测的情况。

选择 3 种常用的语义分割网络特征金字塔网络（feature pyramid networks, FPN）[16]、U-Net 和 U-Net++进行比较，结果如表 1 所示，其中，最好的结果用粗体标出。该实验在 Intel Core 2.30 GHz CPU、GPU 显存 8 GB 的 Windows 电脑上运行。数据集由 Labelme 进行标注，分为背景和黑点 2 类，按照训练集 60%、验证集 20% 和测试集 20% 的策略进行分配，训练集 121 张图片，验证集 40 张图片，测试集 40 张图片。使用 Adam 优化器[17-18]，学习率初始值设为 10^{-3}，学习率通过余弦退火法衰减，batch size 设为 1，网络的输入图片大小为 608 像素×512 像素。

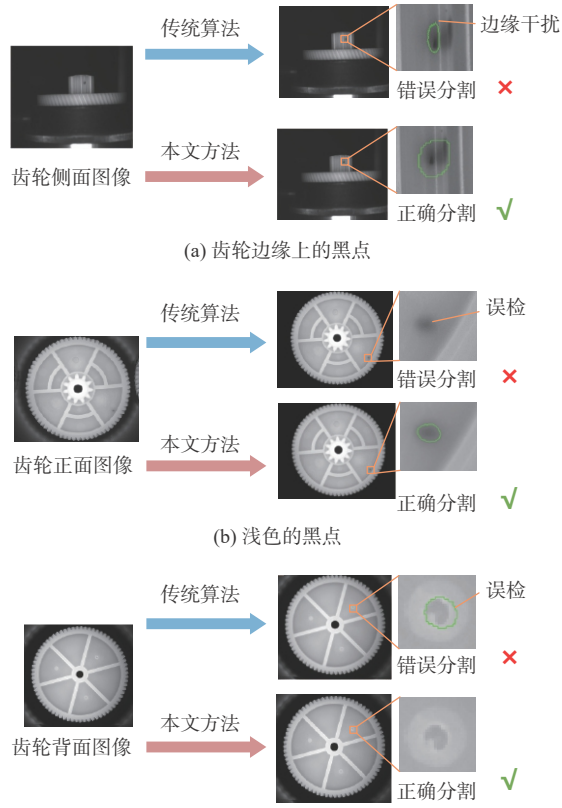


图 9 本文方法与传统算法在 3 个问题上检测效果对比

Fig. 9 Comparison of detection effects between the proposed method and traditional algorithms on three problems

表 1 3 种模型比较

Table 1 Comparison of three models

模型	编码器	图像数量	平均处理时间/s	IoU平均值	P_c/%
FPN	resnet18	201	**0.192**	0.45	97.87
	resnet34		0.200	**0.48**	**97.99**
	efficientnet-b1		0.204	0.44	97.53
U-Net	resnet18	201	**0.196**	0.48	98.02
	resnet34		0.202	**0.51**	**98.52**
	efficientnet-b1		0.227	0.42	97.81
U-Net++	resnet18	201	0.263	0.52	98.32
	resnet34		0.298	**0.60**	**98.93**
	efficientnet-b1		**0.246**	0.54	98.58

由于实际生产中塑料齿轮的缺陷样本较难收集，样本数量较少，在选用 backbone 时应该避免参数量太大的模型，所以尽量选择轻量化的 encoder 调试网络，本文选择了 3 种比较轻量化的 encoder：resnet18、resnet34 和 efficientnet-b1。

本文需要重点关注与检测和分割结果好坏密切相关的 P_c 和 IoU 平均值 2 个指标。由表 1 中 IoU 平均值可知，3 种语义分割网络 FPN、U-Net 和 U-Net++ 的 IoU 平均值都较低，这说明语义分割网络对于塑料齿轮黑点这种小缺陷的分割效果有限，最高的是

U-Net++, 201 张图像的 IoU 平均值达到了 0.60, 此值依然不够理想。

与 IoU 指标类似, 语义分割网络的召回率同样也是通过像素个数来计算的, 因为黑点的特征像素点个数太少, 因此, 结果依然不理想。以上分析说明了为何本文在语义分割预测的结果之后还要通过多特征融合分析进行局部区域修正的原因。

如图 10 所示, 3 种模型均可以很好地收敛, 损失函数值均在 0.05 以下。相比而言, 所有网络的 P_c 指标的结果都非常好, 3 种网络得到的 P_c 值均高于 97.5%, 所以基于深度学习方法检测的结果准确率要优于传统的缺陷检测方法。表 1 中 U-Net++的 IoU 平均值和 P_c 都略优于 FPN 和 U-Net, 并且 U-Net++不会受到不必要的限制性跳接的限制, 故本文选用 U-Net++。

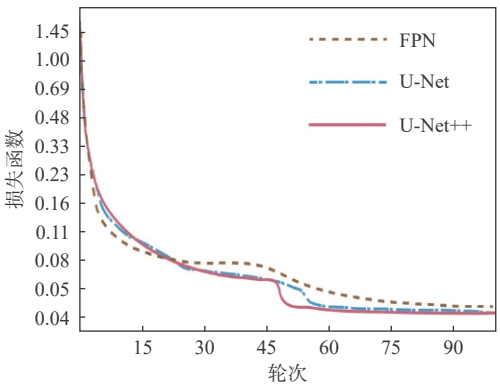


图 10　训练过程中的 Diceloss 衰减曲线

Fig. 10　Decay curves of Diceloss during training

采用基于局部图像梯度特征的区域生长法和多特征置信度分析法对候选黑点区域作修正, 测试效果如图 11 所示。测试选择了包含黑点全部情形的数据进行比较, 包括侧面齿顶处黑点、侧面齿根处黑点、端面加强筋处黑点和端面轮齿上黑点, 可以看出, 无论黑点位于哪种位置, 经过多特征融合分析结果修正以后的 IoU 值都明显提高。

对修正前后的 IoU 值进行比较, 实验选用分好的 40 张测试集图片进行, 典型结果如图 11 所示。图 11 (a), 修正前: I_{oU1}=0.234, I_{oU2}=0.107; 修正后: I_{oU1}=0.882, I_{oU2}=0.887(经过特征融合分析步骤后被筛除, 即融合分析后判断不是黑点缺陷); 图 11 (b), 修正前: I_{oU}=0.469; 修正后: I_{oU}=0.934; 图 11 (c), 修正前: I_{oU}=0.576; 修正后: I_{oU}=0.838(多个种子点); 图 11 (d), 修正前: I_{oU}=0.711; 修正后: I_{oU}=0.852(此处采用多个种子点)。

选用先前分好的 40 张测试集图片进行修正前后的 IoU 的平均值对比实验, 实验结果如图 12 所示。40 张图片修正前的 IoU 平均值为 0.555, 修正

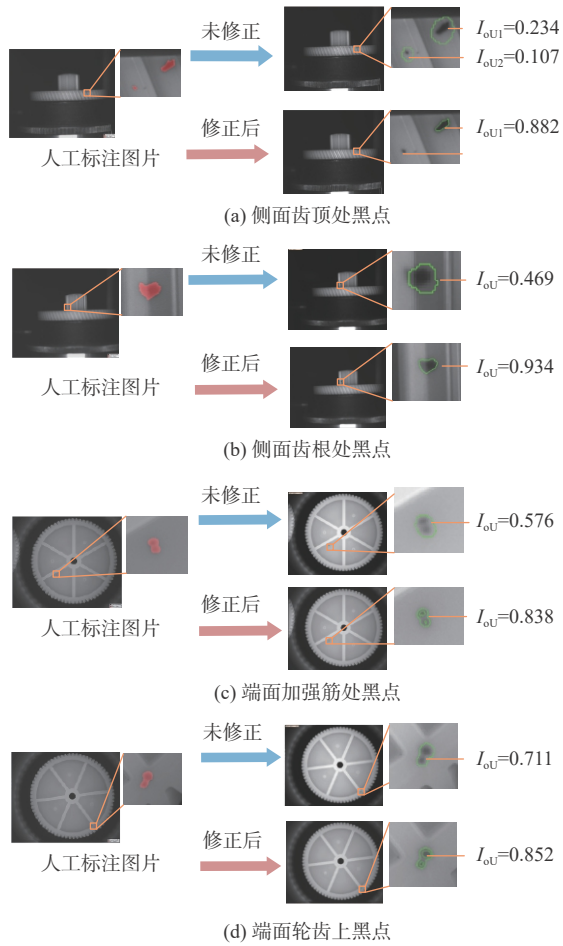


图 11　修正前后黑点分割效果的对比

Fig. 11　Comparison of dark spot segmentation effects before and after correction

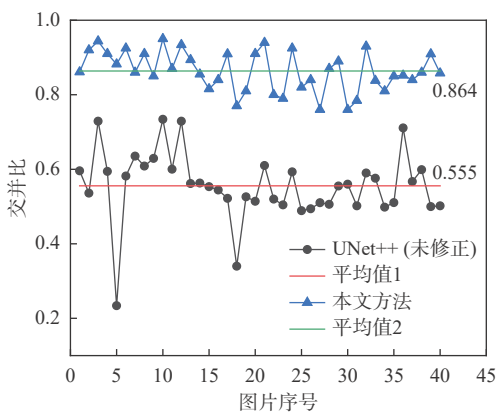


图 12　修正前后的 IoU 的平均值对比

Fig. 12　Comparison of the average values of IoU before and after correction

后的 IoU 平均值为 0.864, 而且修正后的 IoU 值的波动幅度比修正前更小, 提升显著。表 2 为不同检测方法对塑料齿轮黑点缺陷数据集的检测效果对比, 其中, 最好的结果用粗体标出。相比于传统缺陷检测算法和未做修正的 U-Net++检测算法, 本文方法得到的 IoU 平均值分别高出 0.478 和 0.309, P_c 值可达 98.93%, 相比于传统缺陷检测算法要高出 43.36%,

表2 不同检测方法对塑料齿轮黑点缺陷数据集的检测效果对比

Table 2 Comparison of detection effects among different methods on dataset of plastic gears with dark spot defects

方法	IoU平均值	P_c值/%
传统缺陷检测算法（基于局部极值的分水岭算法）	0.386	55.60
U-Net++(未做修正)	0.555	98.93
本文方法	**0.864**	**98.93**

但相比于未做修正的U-Net++检测算法提升不明显。

U-Net++的检测时间已在表1中给出，实际完整的检测时间还要加上多特征融合分析修正的时间，图13为测试集40张塑料齿轮黑点缺陷图片每张的检测时间，平均处理速度低于0.35 s/个。

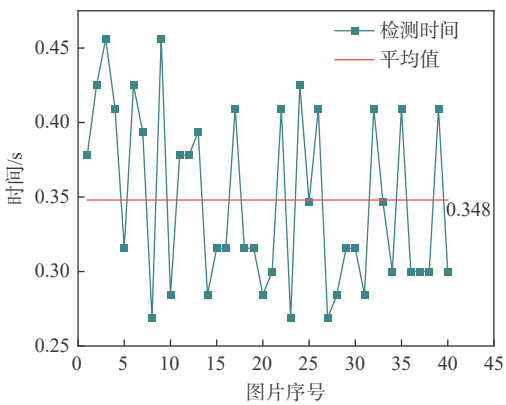


图13 完整算法的检测效率

Fig. 13 Detection efficiency of complete algorithm

4 结 论

本文提出基于U-Net++和特征融合的塑料齿轮复杂黑点检测方法，搭建并优化了U-Net++用于黑点候选区域的预测，再基于局部图像梯度特征的区域生长法修正区域，最后，根据多特征融合的置信度分析结果完成判定，得到准确的黑点缺陷区域。根据测试结果，可以得出以下结论：

1）利用本文方法检测得到的黑点缺陷结果中的P_c值可达98.93%，远超传统的缺陷检测算法，解决了塑料齿轮复杂黑点检测的3个难题，能很好应对塑料齿轮黑点缺陷位置的随机性、形状的不规则性和背景的复杂性。

2）提出基于局部图像梯度特征的区域生长法和多特征置信度融合分析方法，对U-Net++分割后的候选黑点区域进行局部区域的修正，弥补了深度学习方法对小缺陷分割效果欠佳的不足，最终测试图像集的IoU平均值从0.555提高到了0.864，相比传统的缺陷检测算法和未做修正的深度学习算法，IoU分别高出0.478和0.309，增强了检测结果的准确性和稳定性。

3）本文方法检测效率高、准确率高、稳定性好，可应用于塑料齿轮在线检测设备上代替人工检测，并可为塑料齿轮工艺溯源提供大数据。

参考文献（References）

[1] 石照耀, 方一鸣, 王笑一. 齿轮视觉检测仪器与技术研究进展[J]. 激光与光电子学进展, 2022, 59(14): 1415006.
SHI Z Y, FANG Y M, WANG X Y. Research progress in gear machine vision inspection instrument and technology[J]. Laser & Optoelectronics Progress, 2022, 59(14): 1415006(in Chinese).

[2] 欧阳志喜, 石照耀. 塑料齿轮设计与制造[M]. 北京: 化学工业出版社, 2011: 271-276.
OUYANG Z X, SHI Z Y. Design and manufacturing of plastic gears[M]. Beijing: Chemical Industry Press, 2011: 271-276(in Chinese) .

[3] HAN K T M, UYYANONVARA B. A survey of blob detection algorithms for biomedical images[C]//Proceedings of the 7th International Conference of Information and Communication Technology for Embedded Systems. Piscataway: IEEE Press, 2016: 57-60.

[4] 杨亚, 陶红艳, 余成波. SURF与灰度差分在小模数塑料齿轮缺陷检测中的研究与应用[J]. 机械传动, 2018, 42(5): 156-160.
YANG Y, TAO H Y, YU C B. Research and application of SURF and gray difference in detection of small modulus plastic gear defect[J]. Journal of Mechanical Transmission, 2018, 42(5): 156-160(in Chinese).

[5] 李江波, 彭彦昆, 黄文倩, 等. 桃子表面缺陷分水岭分割方法研究[J]. 农业机械学报, 2014, 45(8): 288-293.
LI J B, PENG Y K, HUANG W Q, et al. Watershed segmentation method for segmenting defects on peach fruit surface[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(8): 288-293(in Chinese).

[6] 韦玉科, 陈玉, 田洪金. 基于计算机视觉的焊点缺陷检测系统的设计[J]. 测控技术, 2015, 34(1): 138-141.
WEI Y K, CHEN Y, TIAN H J. Design of solder joint defect detection based on machine vision[J]. Measurement & Control Technology, 2015, 34(1): 138-141(in Chinese).

[7] 陶显, 侯伟, 徐德. 基于深度学习的表面缺陷检测方法综述[J]. 自动化学报, 2021, 47(5): 1017-1034.
TAO X, HOU W, XU D. A survey of surface defect detection methods based on deep learning[J]. Acta Automatica Sinica, 2021, 47(5): 1017-1034(in Chinese).

[8] SU Y T, YAN P. A defect detection method of gear end-face based on modified YOLO-V3[C]//Proceedings of the 10th Institute of Electrical and Electronics Engineers International Conference on Cyber Technology in Automation, Control, and Intelligent Systems. Piscataway: IEEE Press, 2020: 283-288.

[9] 张广世, 葛广英, 朱荣华, 等. 基于改进YOLOv3网络的齿轮缺陷检测[J]. 激光与光电子学进展, 2020, 57(12): 121009.
ZHANG G S, GE G Y, ZHU R H, et al. Gear defect detection based on the improved YOLOv3 network[J]. Laser & Optoelectronics Progress, 2020, 57(12): 121009(in Chinese).

[10] 刘悦, 黄新波. 基于 YOLOv4 和改进分水岭算法的绝缘子爆裂检测定位研究[J]. 电网与清洁能源, 2021, 37(7): 51-57.
LIU Y, HUANG X B. Research on insulator burst fault identification based on YOLOv4 and improved watershed algorithm[J]. Power System and Clean Energy, 2021, 37(7): 51-57(in Chinese).

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

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

[13] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE Press, 2017: 640-651.

[14] VON GIOI R G, JAKUBOWICZ J, MOREL J M, et al. LSD: a line segment detector[J]. Image Processing on Line, 2012, 2: 35-55.

[15] ZHOU J, HUANG X H, PENG G. Recognition of airplanes using multi-feature fusion[J]. Journal of Huazhong University of Science and Technology, 2009, 37(1): 38-41.

[16] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 936-944.

[17] DUCHI J, HAZAN E, SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. The Journal of Machine Learning Research, 2011, 12: 2121-2159.

[18] RUDER S. An overview of gradient descent optimization algorithms [EB/OL]. (2017-06-15)[2023-02-01]. https://arxiv.org/abs/1609.04747v2.

Detection method for complex dark spots on plastic gears based on U-Net++ and feature fusion

FANG Yiming，SHI Zhaoyao*，SONG Huixu

(Beijing Engineering Research Center of Precision Measurement Technology and Instruments，Beijing University of Technology，Beijing 100124，China)

Abstract: Traditional defect detection algorithms exhibit poor performance in accurately detecting complex dark spots on the surface of plastic gears. There are three primary issues: firstly, inaccurately distinguishing the size and position of dark spots on the gear edge; secondly, a high rate of missed detection for light dark spots; thirdly, a tendency to misjudge the point gate as dark spots. This paper proposed an improved detection method for complex dark spots on plastic gears based on U-Net++ and feature fusion. The dark spot area was predicted through U-Net++ and corrected depending on gradient features. Multi-feature fusion analysis was utilized to provide the final judgment result, thus improving the accuracy and stability of complex dark spot detection. The test results demonstrate that the P_c value, which represents the accuracy of the detection results, is as high as 98.93%, and the average value of IoU, representing the accuracy of the segmentation results, reaches 0.864. In comparison to traditional defect detection algorithms and uncorrected deep learning algorithms, the proposed method increases the average value of IoU by 0.478 and 0.309, respectively.

Keywords: plastic gear defects; deep learning; U-Net++; multi-feature fusion; local area correction

Received：2023-06-28；**Accepted**：2023-10-25；**Published Online**：2023-11-01 12：52

URL：link.cnki.net/urlid/11.2625.V.20231101.0843.001

Foundation items：National Natural Science Foundation Major Scientific Research Instrument Development Project (52227809); State Key Laboratory of Precision Measuring Technology and Instruments (Tianjin University) (PILAB2105); Young Elite Scientists Sponsorship Program by CAST (2021QNRC001)

* **Corresponding author**. E-mail：shizhaoyao@bjut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0444

氧燃比对自燃推进剂模型火箭发动机燃烧稳定性的影响

任永杰，徐伯起，楚威，郭康康*，仝毅恒，聂万胜

(航天工程大学 宇航科学与技术系，北京 101416)

摘　　要：为研究氧燃比（O/F）对自燃推进剂模型火箭发动机燃烧稳定性的影响，在液/液双旋流矩形模型发动机中开展不同氧燃比条件下的试验，采用高频压力传感器和光电倍增管（PMT）同步捕捉了燃烧室压力振荡和CH*表征的释热脉动，获得了氧燃比对燃烧稳定性的影响规律。结果表明：点火后燃料集液腔出现了41 Hz的低频脉动，并在关机过程中诱发了燃烧室同步低频振荡。氧燃比从0.933～1.789增长的过程中，燃烧室经历了稳定、轻微不稳定、一阶横向不稳定和二阶横向不稳定的转换过程，在二阶模态下压力振荡幅值仅为平均室压的4.69%。结合光电倍增管信号，发现燃烧不稳定越剧烈压力和释热信号的耦合就越明显。基于试验的瑞利指数分析表明：一阶横向模态下燃烧不稳定的驱动源主要位于燃烧室两侧，燃烧室中间则呈抑制特性；燃烧不稳定的产生可能和推进剂与燃烧室壁面的相互作用有关。

关　键　词：自燃推进剂；氧燃比；高频燃烧不稳定；液/液双旋流；瑞利准则

中图分类号：V434+.3

文献标志码：A　　**文章编号**：1001-5965(2025)09-3030-09

高频燃烧不稳定的出现严重威胁了火箭发动机的可靠工作，是制约大推力火箭发动机发展的关键因素之一[1]。高频燃烧不稳定又称声学不稳定，其是压力振荡和燃烧室本征声学模态耦合的过程，并伴随着剧烈的周期性压力和释热脉动[2]。这种脉动对发动机冷却和机械结构提出了巨大挑战，甚至引起灾难性后果。自20世纪40年代起，对燃烧不稳定问题的研究一直持续至今，但对其产生机理和抑制方法仍未认识清楚[3]。

普渡大学对单喷嘴纵向燃烧不稳定进行了系统的实验[4]、数值计算[5]和理论分析[6]。深入研究了模型燃烧室的纵向燃烧不稳定机理，认为氧管长度、燃烧室突扩面周期性涡脱落及其与壁面的相互作用是影响燃烧稳定性的关键因素。在后续开发的多喷嘴二维矩形燃烧室中，研究了横向燃烧不稳定[7]，并对其进行了丰富的光学诊断测量[8]。采用高频压力传感器和光电倍增管(photomultiplier tubes, PMT)相结合的方法实现了基于实验的瑞利准则定量表征[9]。根据瑞利准则，压力脉动和释热同相时燃烧不稳定就会被驱动[10]，瑞利指数被广泛用于燃烧不稳定的驱动或抑制分析。瑞利指数的计算需要同时获得相同位置的压力和释热信号的脉动值，这种同步测量在低压燃气轮机背景条件下较为常见[11-12]，而火箭发动机燃烧室高温高压，给同步测量带来了很大的挑战。德国宇航中心在进行BKD氢氧模型发动机的热试车中采用光学探针，实现了同步测量[13]，而国内对火箭发动机背景下PMT诊断方法的使用却鲜有报道。

氧化剂和燃料的质量流量比(氧燃比，O/F)是影响燃烧稳定的重要因素之一。在煤油富燃燃气

收稿日期：2023-07-05；录用日期：2023-10-27；网络出版时间：2023-11-23 16:34
网络出版地址：link.cnki.net/urlid/11.2625.V.20231122.1604.002
基金项目：国家科学自然基金 (12002386)
*通信作者. E-mail：guokangkang@sjtu.edu.cn

引用格式：任永杰，徐伯起，楚威，等. 氧燃比对自燃推进剂模型火箭发动机燃烧稳定性的影响 [J]. 北京航空航天大学学报，2025，51（9）：3030-3038. REN Y J，XU B Q，CHU W，et al. Effects of oxygen-fuel ratio on combustion stability of a model rocket engine with hypergolic propellant [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（9）：3030-3038（in Chinese）.

发生器的试验研究中, 发现随着 O/F 的增加燃烧室特征速度几乎与 O/F 成正比, 而与燃烧室的室压和长度关系不大[14]。德国宇航中心在进行 BKD 热试车时发现, O/F 能够改变燃烧室本征声学模态的频率, 从而很大程度上影响燃烧室压力振荡频率[15-16]。但在二维矩形模型发动机中, 采用煤油和催化分解的过氧化氢作为推进剂, 在改变 O/F 热试车时, 发现模型发动机的燃烧动力学特性没有明显改变[7,17], 这可能与二维燃烧室的低稳定性裕度有关。在单喷嘴预混燃烧室中, Buschhagen 等[18] 发现当量比对燃烧不稳定有很大的影响。在当量比较高时(0.72)燃烧室仅出现了一阶纵向燃烧不稳定, 振幅仅为 3%, 在突扩面处呈现轴对称的涡脱落; 在当量比为 0.69 时, 燃烧室振幅增加至 15%, 且同时出现了一阶纵向和一阶切向燃烧不稳定, 突扩面出现非对称涡脱落。

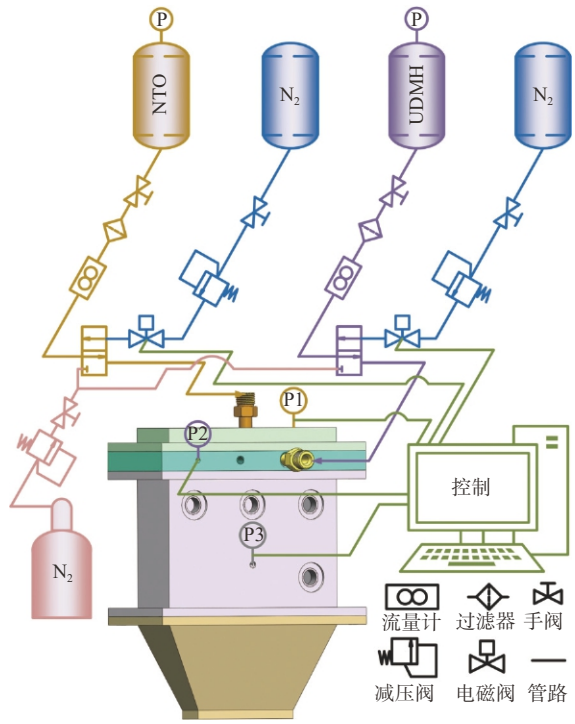
常规自燃推进剂(肼类和四氧化二氮)因其具有剧毒, 试验门槛较高, 针对其燃烧稳定性的试验研究较少。严宇等[19] 研究了混合比对旋转爆震的影响, 发现在总流量较大且偏富燃(O/F<1.7)条件下更易形成旋转爆震。Xue 等[20] 在进行撞击式自燃推进剂模型火箭发动机试验时发现, 燃烧室中间 O/F 能够很大程度上影响燃烧稳定性。发现其燃烧不稳定边界符合休伊特准则, 中心 O/F 在 1.5～2.5 之间更易诱发燃烧不稳定, 并认为这和撞击式喷嘴的混合效率有关。Chu 等[21] 通过数值仿真, 再现了模型发动机中的旋转模态和驻波模态的切向燃烧不稳定, 认为燃烧不稳定的激发与局部的突然非均匀释热有关。Qin 等[22] 通过数值计算发现喷雾粒径是影响燃烧不稳定的重要因素, 在一定范围内燃烧不稳定出现了一阶纵向和一阶切向的模态转变。采用气动声学方法结合压力时滞模型, 初敏等[23] 对某型姿轨控自燃推进剂发动机进行了稳定性分析。发现该方法能够计算发动机固有声学频率, 并能预测高频燃烧稳定性。但这种方法忽略了真实化学反应过程, 具有很大的局限性。由于自燃推进剂双液相特性, 给数值仿真带来了巨大的挑战, 对其喷注过程进行了大量的简化, 其准确性与合理性必须经过大量试验验证。

目前关于偏二甲肼(UDMH)/四氧化二氮(NTO)的燃烧不稳定研究较少, 基于液/液双旋流喷注器的燃烧不稳定研究则更为匮乏, 而撞击式喷注器则有著名的休伊特准则。本文针对 UDMH/NTO 液/液双旋流模型发动机开展了高频燃烧不稳定性研究, 分析了 O/F 对燃烧稳定性的影响, 初步得到了燃烧室动态燃烧特性。采用高频压力传感器和 PMT 同步诊断, 实现了基于试验的瑞利准则定量表征。

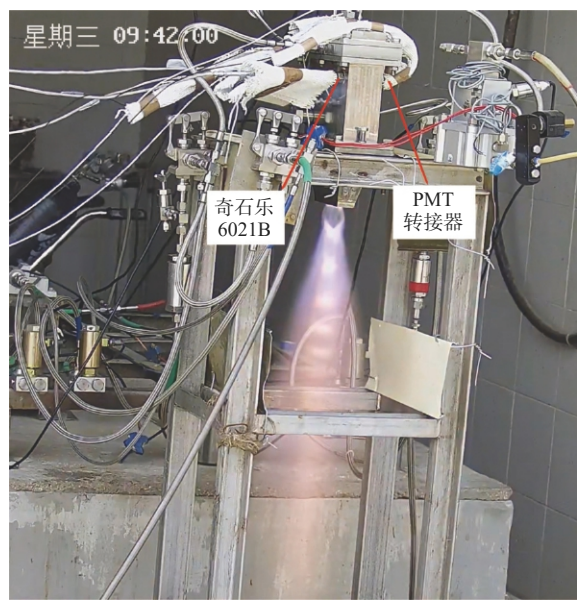
1　试验系统与测量装置

1.1　试验系统

图 1 为模型发动机试验系统及热试车。如图 1(a)所示, 试验装置包含推进剂供应系统、氮气(N_2)吹除系统、测控系统和矩形模型发动机。推进剂供应系统通过氮气挤压的方式为模型发动机提供 UDMH 和 NTO; 氮气吹除系统用于每次试验结束后吹除集液腔中的残留推进剂; 测控系统则控制



(a) 模型发动机试验系统



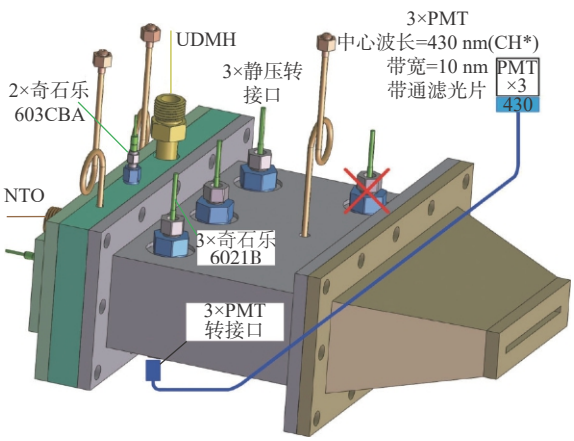
(b) 热试车

图 1　模型发动机试验系统及热试车

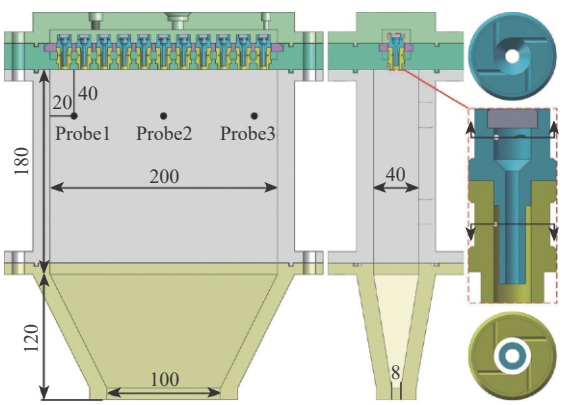
Fig. 1　Schematic of the model engine experiment system and diagram of hot-fire test

着开机和关机，同时测量模型发动机的热试车数据。图1(b)为模型发动机热试车图，在热试车中采用高频压力传感器和PMT等采集了热试车动态数据。

模型发动机及其诊断装置如图2(a)所示。模型发动机包含11个液/液双旋流喷嘴、矩形燃烧室及喷管。为产生自激的高频燃烧不稳定，燃烧室具有较大长宽比，其长度为200 mm，宽度为40 mm，可以看作二维模型燃烧室。模型燃烧室由2部分组成，包含180 mm的等直段和120 mm的收缩段，如图2(b)所示。喷管的长度为100 mm，宽度为8 mm。液/液双旋流喷嘴细节如图2(b)右侧所示，其结构为内收口外敞口式喷嘴组合，内外喷嘴均包含4个切向孔。推进剂通过挤压，在切向孔的导引下推进剂进入喷嘴，为燃烧室提供推进剂。整个模型发动机采用热沉冷却，为保证发动机和传感器的安全，单次热试车时间约为1.7 s。



(a) 模型发动机诊断装置



(b) 模型发动机关键尺寸

图2 模型发动机诊断装置和关键尺寸

Fig. 2 Configuration of model engine diagnostic sensors and key geometric parameters

1.2 测量方法

燃烧室、UDMH和NTO集液腔的静态压力采用稳态压力传感器测量，其采集频率为2 kHz，量程为10 MPa，满量程输出精度为±0.25%。由于燃烧室内部环境恶劣，燃烧不稳定频率较高，对模型发动机的诊断方式主要集中在高频数据测量。燃烧室包含3个耐高温高频压力传感器(长时间700 ℃)，型号Kistler 6021B，命名为Probe 1～Probe 3，其位置如图2(b)所示。Kistler 6021B是差分型，具有加速度补偿的压电式高频压力传感器，满量程测量误差不超过1%。由于氧化剂和燃料集液腔的温度较低(几乎为常温)，则采用了成本较低的Kistler 603CBA压电式高频压力传感器，量程为7 MPa。其输出信号为电压信号，灵敏度为710 mV/MPa，满量程输出精度为±1%。如图2(a)所示，在燃烧室高频压力传感器的对称位置安装了3个对应的光纤PMT探头，型号为Hamamatsu H10723-210。其是一款集成了光电倍增管、高压电源和低噪声放大器的PMT模块，在典型条件下，其峰值噪声仅为输出电压的0.012 5%。在PMT前端安装了带通滤光片(中心波长为430 nm，带宽为±5 nm)，用于探测燃烧室CH*基团的发光强度，根据前人研究CH*基团的发光强度可以表征燃烧室的释热特性[9]。由于燃烧室具有较大长宽比，PMT和高频压力传感器几乎可以认为测量的相同位置，采用PMT表征的释热特性和高频压力传感器的压力脉动特性可以实现基于试验的瑞利准则定量表征。高频数据的采集则采用Kistler Kidaq便携式采集系统。其高频采集板卡为Kistler 5512A，采集频率为100 kHz，采集精度为0.025%。推进剂质量流量的测量采用Rosemount科氏流量计，其质量流量测量精度为±0.2%。

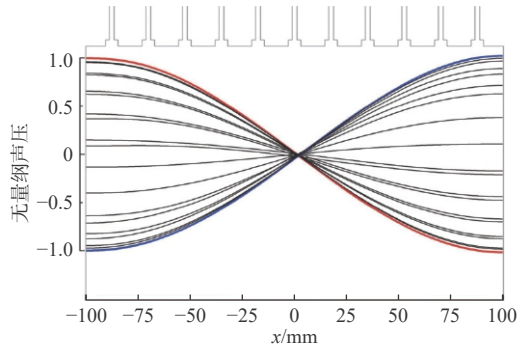
2 结果与讨论

2.1 横向振荡模态

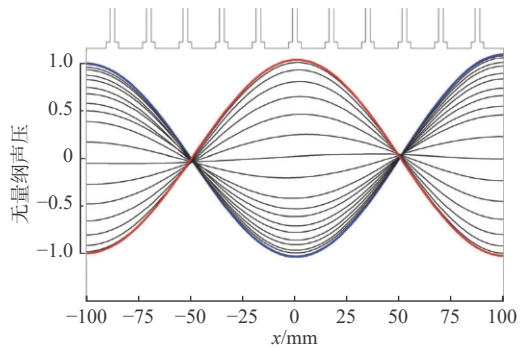
图3为当前矩形模型发动机燃烧室内理论一阶横向压力振荡模态(1W)和二阶横向压力振荡模态(2W)。图中，上框背景为模型发动机燃烧室头部和喷嘴的简要构型，这样结合压力振荡模态对比分析，更加易于理解。通过分析试验中高频压力信号(Probe 1～Probe 3)的相位和幅值关系可以判断燃烧室中出现的横向压力振型。如图3(a)所示，Probe 1和Probe 3监测点在1W模态下压力振荡呈等幅值反相位状态，压力振荡波节在燃烧室中心(Probe 2)。在2W模态下，Probe 1和Probe 3具有相同的振幅和相位，同时Probe 2位置处与前两者压力振荡幅值相同，相位相反。通过对比不同位置高频传感器压力振荡幅值和相位特征，可以判断燃烧室中的模态振型。

2.2 试验概况

为研究O/F对燃烧稳定性的影响，开展了5种O/F条件下的试验，试验编号对应为Test 1～Test 5。每种O/F条件下试验重复3次，共开展15次热试车



(a) 1W 模态



(b) 2W 模态

图 3　燃烧室横向空间模态振型

Fig. 3　Transverse spatial mode shapes of combustion chamber

试验, 结果表明, 试验具有良好的可重复性。为保证试验稳定安全运行, 燃烧室室压不超过 1 MPa。试验简要结果如表 1 所示, 每种 O/F 条件下仅展示 1 次试验结果。其中, P'/P_c 为压力振荡峰峰值 P' 与燃烧室平均压力 P_c 的比值, $\dot{m}_F$ 和 $\dot{m}_O$ 分别为燃料和氧化剂的质量流量。Test 1 虽然压力振荡峰峰值为 5.62%, 但其多数时间处于粗暴燃烧状态, 其振荡主频 1 981 Hz 并不突出, 因此, 认为 Test 1 工况下燃烧室是稳定的。随着 O/F 的增加, 燃烧室压略有增加, 燃烧室经历了稳定到一阶横向(1W)燃烧不稳定再到二阶横向(2W)燃烧不稳定的转变。不稳定发生时伴随着明显的啸叫声, 但高频压力动态数据显示压力振荡幅值较低, 最高仅有 12.42%。当燃烧不稳定从一阶横向(1W)转换到二阶横向(2W)时压力振荡幅值进一步降低, 在纯二阶模态时仅有平均室压的 4.69%。

表 1　试验结果总结

Table 1　Summary of experimental results

试验编号	$\dot{m}_F$/(g·s^{-1})	$\dot{m}_O$/(g·s^{-1})	P_c/MPa	O/F	$\left(\frac{P'}{P_c}\right)$/%	f/Hz	结果
1	231.5	216.1	0.73	0.933	5.62	1 981	稳定
2	188.7	248.8	0.75	1.318	7.85	2 363	轻微
3	183.8	270.0	0.80	1.469	12.42	2 381	1W模态
4	181.3	291.5	0.82	1.608	4.88	2 409	1W-2W模态
5	173.6	310.5	0.86	1.789	4.69	4 628	2W模态

试验中典型的压力信号序列如图 4 所示, 对应的试验工况条件为 Test 3, 其氧燃比为 1.469。推进剂进入燃烧室后迅速建立燃烧室压力, 并未出现明显的点火压力峰。在热试车过程中, 燃料集液腔出现了明显的低频振荡, 振荡峰/峰值约为 0.25 MPa。发动机关机指令发出后, 由于推进剂集液腔残留的推进剂, 燃烧室并未立刻熄火, 而是出现了明显的低频振荡(见图 5), 燃料和氧化剂集液腔也出现了同步振荡。进一步分析表明, 燃烧室低频振荡与燃料集液腔的低频振荡频率一致(41 Hz), 也诱发了氧化剂集液腔的同步脉动。燃料集液腔压力振荡可能是导致关机后燃烧室同步低频振荡的关键因素。高频压力传感器显示, 燃烧室迅速形成了高频振荡, 试车时伴随着明显的啸叫声。

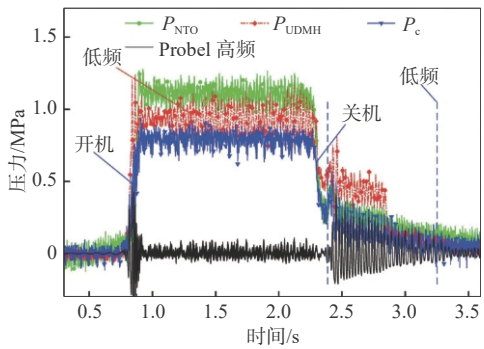


图 4　Test 4 热试车过程典型信号

Fig. 4　Typical signals during hot-fire Test 4

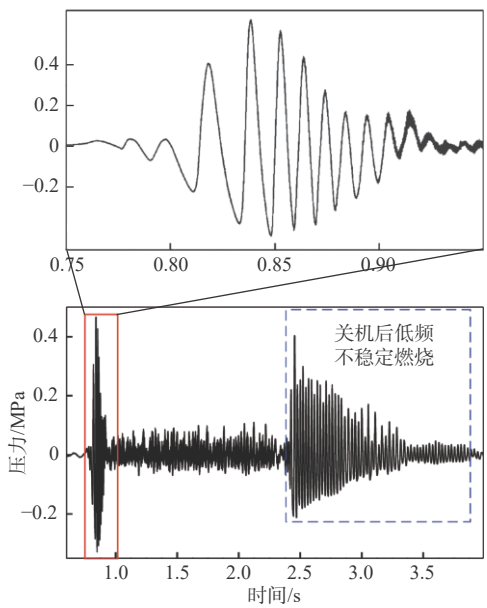


图 5　Probe 1 监测点热试车压力振荡细节(Test 3)

Fig. 5　Detailed pressure oscillations of Probe 1 during hot-fire Test 3

2.3　压力振荡特性

图 5 为高频压力监测点 Probe 1 压力信号的局部放大图。图中实线框部分展示了高频燃烧不稳定的增长过程, 虚线框部分为关机过程中引起的低频燃烧不稳定。如图 5 上方局部放大显示, 本试验

中高频燃烧不稳定的增长过程与前人不同[20]，并不是由低振幅逐渐增长到高振幅而是由前期燃烧室点火引起的压力扰动发展而来。这可能与燃烧不稳定的程度有关，本试验中压力振荡峰峰值较小。燃烧室点火扰动产生低频高振幅的压力扰动，这种低频扰动逐渐衰减。而其中的高频成分可能与燃烧室声学特性耦合，逐渐增长放大，最终形成高频燃烧不稳定。

图 6 为典型压力信号及其 PSD 分析。图 6(a)为 5 组热试车试验的高频压力信号。其中，Test 1 并未出现燃烧不稳定，监测点压力信号以噪声为主，仅有短时间很微弱的周期性信号。如图 6(b)所示，Test 1 的功率谱密度(power spectral density, PSD)图显示仅在 1 981 Hz 处有峰值，其主峰并不突出。随着 O/F 的增加，在 Test 2 试验中出现明显的周期性压力振荡，其中，Probe 1 和 Probe 3 的相位几乎相反，而 Probe 2 的频率则是另外 2 个监测点的 2 倍(在 Probe 1 和 Probe 3 的一个振荡周期内 Probe 2 出现了 2 次峰值)，这是明显的 1W 模态特征。Test 2 试验的压力振荡幅值较低，为平均室压的 7.85%，但频谱特性分析出现了明显的峰值，试车中也出现了明显的啸叫，因此称为轻微燃烧不稳定。继续增加 O/F 至 1.469 时(Test 3)，如图 6(a)所示，模型发动机出现明显的 1W 模态燃烧不稳定，燃烧室振荡幅值达到极限环状态，峰峰值超过了平均室压的 10%。结合图 6(b)，发现 Test 3 PSD 图出现了明显的主峰并伴随着明显的倍频。继续增加 O/F(Test 4)燃烧室出现了 1W-2W 燃烧不稳定模态的转变，振荡峰峰值下降，但燃烧不稳定模态仍以 1W 模态为主，试验中仍然伴随着尖锐的啸叫声。当 O/F 增加至 1.789(Test 5)时，燃烧不稳定变为 2W 模态主导，此时 Probe 1 和 Probe 3 同相，频率也相同，而 Probe 2 与另外两者同频反相。结合图 6(b)，Test 5 的振荡主频为 4 628 Hz，几乎为 Test 3 的 2 倍，是典型的 2W 模态的燃烧不稳定。由于振荡频率高，需要更多的能量来维持高频振荡。因此，Test 5 的振荡峰峰值仅为平均室压的 4.69 %，为低振幅燃烧不稳定。

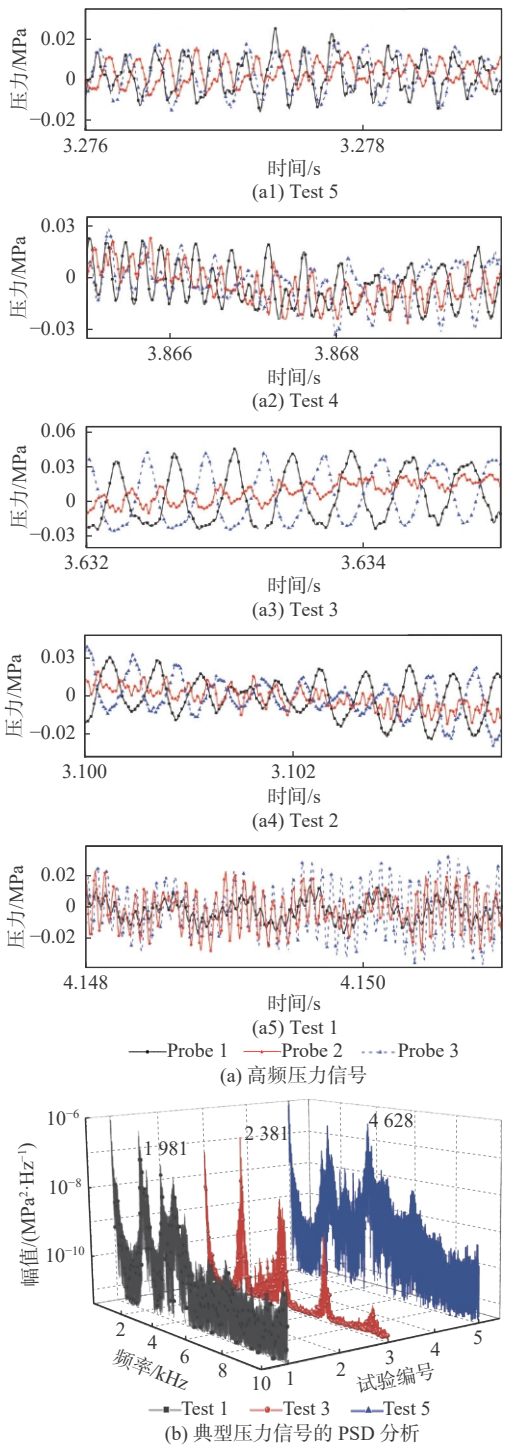


图 6　典型压力信号及其 PSD 分析

Fig. 6　Typical pressure signals and PSD analysis

图 7 为氧燃比对燃烧不稳定压力振荡特性(P'/P_c)和振荡模态的影响。特别的，在氧燃比为 1.608 时压力振荡模态为 1W-2W 的随机切换，在图 7 中位于 1W 和 2W 之间。由图可知，随着氧燃比的增加，压力振荡幅度呈先增加后减小的趋势，横向压力振荡模态由 1W 模态逐渐转变为 2W 模态。压力振荡幅度的降低可能源于燃烧不稳定模态的转变，

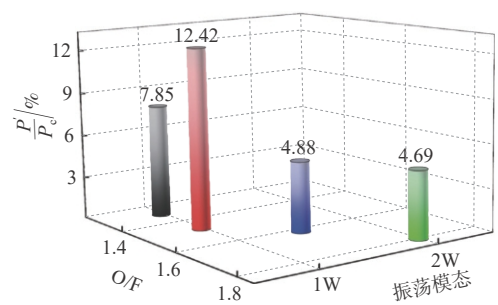


图 7　氧燃比对压力振荡模态和幅值的影响(Test 2～Test 5, Probe 1)

Fig. 7　Effects of O/F ratio on mode and amplitude of pressure oscillation at Probe 1 in Test 2～Test 5

从 1W 模态的振荡主频 2 381 Hz 转变为 2W 模态的 4 628 Hz 时, 燃烧室振荡频率几乎翻倍, 燃烧室声场需要更大的能量维持高频振荡, 因此, 导致振荡幅度降低。

为进一步分析试验中出现的燃烧不稳定模态转换行为, 图 8 为 Test 4 压力振荡模态转换特性。图 8(a)为高频压力监测点 Probe 1 和 Probe 3 的压力振荡特性, 其中, 图 8(a1)和图 8(a2)分别展示了 1W-2W 和 2W-1W 的模态转换特性。在 1W-2W 转换过程中, 压力波动会在 2 个主峰之间产生副峰, 如图 8(a1)红圈所示。在压力振荡过程中, 副峰不断发展其振荡幅值逐渐与主峰相当, Probe 1 和 Probe 3 逐渐形成同相位压力振荡, 模态由 1W 变成 2W, 振荡幅值也有所降低。在 2W 向 1W 模态转换的过程中, 其中一个振荡峰值逐渐减弱, 如图 8(a2)红圈所示。然后, 经过调制形成 1W 模态的压力振荡, 振荡幅值也有所增加。图 8(b)为 Test 4 整个热试过程 Probe 1 高频压力信号的短时傅里叶变换(short-time Fourier transform, STFT)。可以看出, 在整个热试车过程中 1W 模态占主导地位, 1W 和 2W 之间的模态转换贯穿了整个热试车过程。

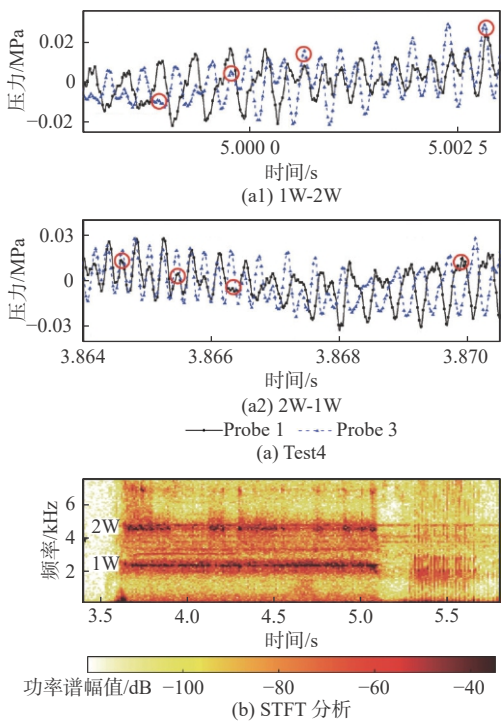


图 8　Test 4 压力模态切换及 STFT 分析

Fig. 8　Shift of pressure mode and STFT analysis of Test 4

2.4　基于试验的瑞利指数

火箭发动机由于燃烧室高温高压, 很难对其进行大面积开窗光学诊断。高频燃烧不稳定发生时, 燃烧室存在剧烈压力和释热脉动, 进一步增加了光学诊断的难度。文献 [24] 中对瑞利准则的分析主要集于数值仿真方面, 本文采用 PMT 和高频压力传感器同步测量燃烧室脉动释热和压力振荡, 能够实现基于试验的瑞利准则的表征。

图 9(a)为试验过程中压力和 CH*发光强度的振荡特性。在 Test 1 条件下, 压力振荡峰峰值仅有燃烧室的 5.62%, 即使在如此微弱的条件下 PMT 也能观测到释热的脉动, CH*脉动信号与压力信号几乎同相。结合图 9(b), Test 2 中 PMT 振荡也得到了

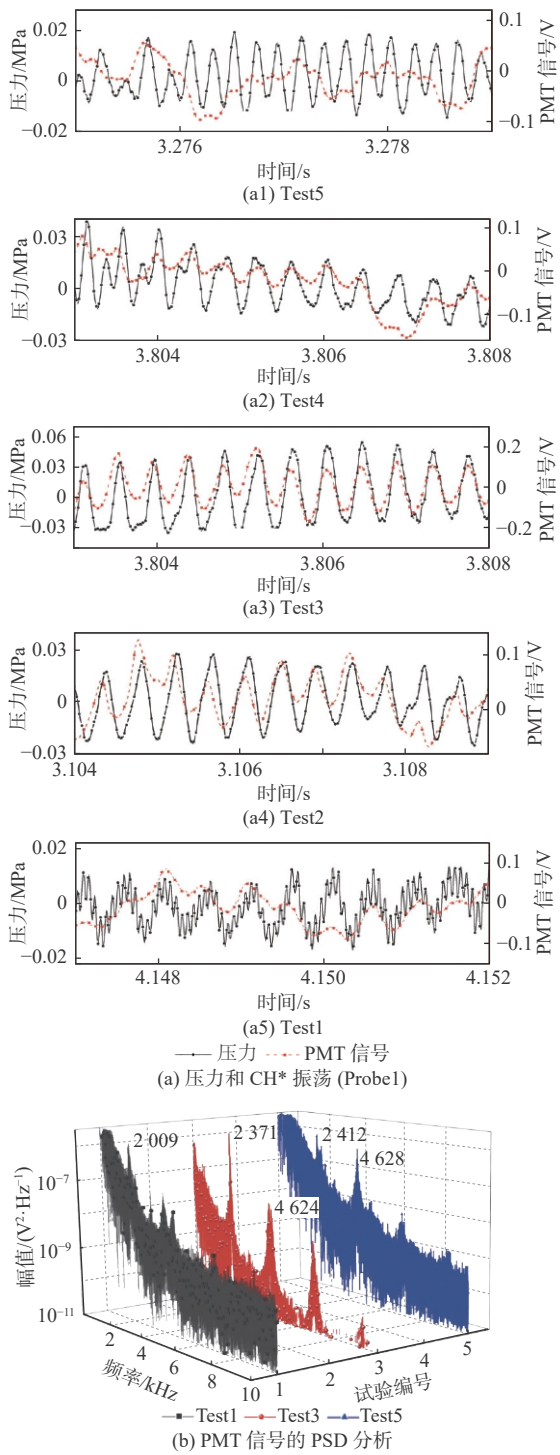
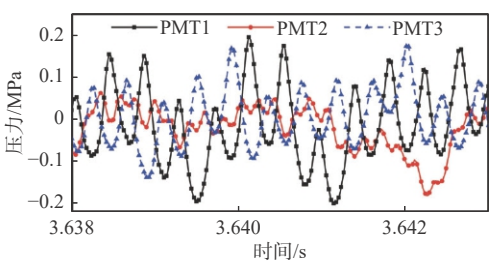


图 9　压力和 CH*发光强度振荡特性和 PMT 信号 PSD 分析

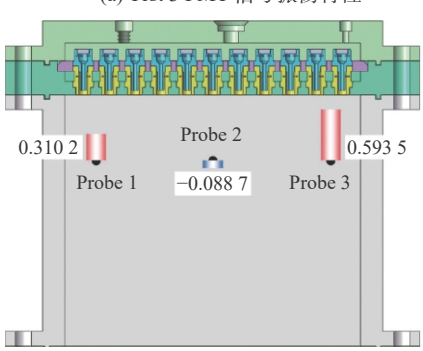
Fig. 9　Pressure and CH* luminescence intensity oscillations and PSD analysis of PMT signals

较小的峰值，其频率为2 009 Hz与压力振荡1 981 Hz几乎相同。PMT对发光强度的变化非常敏感，其分辨率甚至能够达到单个光子。随着O/F的增加CH*的振荡幅值明显增加(Test 2和Test 3)，与压力振荡曲线的耦合也更加明显，说明此时燃烧不稳定得到增强。结合图9(b)，Test 3中CH*脉动主频为2 371 Hz，与压力脉动2 381 Hz几乎完全相同，说明两者达到了高度的耦合。进一步增加O/F，燃烧室压力振荡出现了模态转换，压力振荡幅值明显降低，但在Test 4中仍然能够看到明显的压力与释热脉动的耦合。当O/F增加到1.789时，燃烧不稳定变为2W模态主导，此时压力振荡幅值仅为平均室压的4.69%，压力与PMT信号已经没有明显的耦合特征。但从图9(b)，Test 5中仍然可以看出，CH*脉动具有主峰，其频率为4 628 Hz与压力振荡频率相同。综上可以看出，即使在压力振荡很弱的条件下，PMT仍然能够测得CH*的发光强度脉动，其振荡主频几乎与压力振荡频率完全相同。这说明PMT可以结合高频压力传感器，在模型火箭发动机中实现脉动释热和压力振荡的同步测量。进一步在燃烧室释热区域，压力振荡较为明显时(本文中平均室压的7.85%)，PMT甚至可以作为部分高频压力传感器的替代，采用PMT和高频压力传感器相结合的方式，大幅降低试验成本。

为进一步分析PMT的性能，图10(a)为Test 3中3个PMT信号的时间相位关系。PMT1和PMT3的信号呈现了相位相反的关系，其相位特征与高频压力传感器极为相似。而燃烧室中心处PMT2的信号脉动幅值则较低，说明燃烧室中心释热脉动较弱。仔细观察也可以看出，PMT2的脉动频率几乎是其余两者的2倍。说明在燃烧不稳定发生时PMT能够一定程度上替代高频压力传感器。



(a) Test 3 PMT 信号振荡特性



(b) Test 3 瑞利指数分布

图10 Test 3 PMT信号振荡特性及瑞利指数分布

Fig. 10 Pulsation of PMT signals and distribution of Rayleigh index of test 3

热声耦合燃烧不稳定的驱动或抑制结合瑞利准则[10]可以更好地理解。采用瑞利指数，能够量化空间域局部热声不稳定的抑制或驱动[21]。定义瑞利指数为[9]

$$R_{\mathrm{I}}=\frac{1}{t_2-t_1}\int_{t_1}^{t_2}\frac{\boldsymbol{p}'(x,t)\cdot\boldsymbol{I}'(x,t)}{\overline{p}(x,t)\cdot\overline{I}(x,t)}\mathrm{d}t \tag{1}$$

式中：R_{I}为瑞利指数；$\boldsymbol{p}'$为高频压力传感器信号；$\boldsymbol{I}'$为PMT释热脉动信号；$\overline{p}(x,t)$和$\overline{I}(x,t)$分别为t_1～t_2压力和PMT释热脉动信号的时均值；t为时间；x为信号测量位置。上划线表示变量的平均值，右上一点则表示脉动值(测量值减去平均值)。图10(b)为高频压力脉动信号和CH*发光信号得到的瑞利指数分布，瑞利指数为正说明燃烧不稳定被驱动，为负则相反。由图可见瑞利指数在燃烧室两侧为正值，在燃烧室中心为负值(−0.088 7)。说明在Test 3的1W模态下，燃烧不稳定的驱动区域主要在燃烧室侧面，而燃烧室中心则抑制燃烧不稳定，这与文献[25]在相似模型发动机中得到的结果是吻合的。在1W模态下，压力振荡的波腹位于燃烧室两侧，而燃烧室中心则为波节。在波腹位置出现巨大的释热脉动，进一步当释热脉动和压力振荡同相时也就驱动了燃烧不稳定。结合高频压力振荡和CH*发光强度脉动实现了基于试验的瑞利指数定量化分析。

最边缘喷嘴距离燃烧室壁面仅有9.5 mm，推进剂从喷嘴喷出后，在未完全反应之前有可能已经撞击到侧壁面。推进剂与壁面的相互作用会增强掺混，从而增强燃烧室侧壁附近的释热。热试车中PMT数据也证明了这一点。在发动机点火启动阶段，对燃烧室压力产生巨大扰动，而燃烧室又具有较大长宽比，横向声学频率较低，燃烧室的扰动在横向声学频率上被不断放大，结合燃烧室侧壁面的非均匀释热，最终导致不稳定燃烧。然而，其形成的具体原因仍有待进一步分析。

3 结 论

本文针对自燃推进剂液/液双旋流矩形模型发动机在不同氧燃比条件下进行了燃烧稳定性研究。通过高频压力和PMT释热表征的同步测量，实现了基于试验的瑞利准则定量表征，研究结果表明：

1) 热试车过程中燃料集液腔出现了 41 Hz 低频压力振荡, 这种低频振荡在热试车过程中并未引起燃烧室同步的压力振荡。但在关机后, 燃烧室出现了同步的低频振荡, 也诱发了氧化剂集液腔的同步脉动, 认为燃料腔的低频压力振荡是引起关机后燃烧室同步压力脉动的主要原因。

2) 氧燃比对模型发动机燃烧特性有重大影响。随着氧燃比的增加, 模型发动机经历了稳定、轻微不稳定、一阶横向不稳定和二阶横向不稳定的转换过程。在氧燃比为 1.608 时, 燃烧室出现了一阶横向模态和二阶横向模态的随机转换过程。转化为纯二阶横向模态后, 压力振荡幅值大幅降低, 仅为平均室压的 4.69%。

3) 实现了基于试验的瑞利准则定量表征。同步测量了燃烧室高频压力脉动和 PMT 表征(CH*)的化学释热特性, 初步验证了诊断方法的有效性, 为火箭发动机高频热(释热)声(压力)同步测量开拓了新思路。结合瑞利准则, 发现在一阶横向不稳定模态下, 燃烧不稳定的驱动源主要位于燃烧室两侧, 燃烧室中间则呈抑制特性, 燃烧不稳定可能与推进剂和燃烧室壁面的相互作用有关。

致谢　全体作者由衷感谢航天科技集团北京 101 所: 程永喜、公绪滨、姚亚东、莫昊扬等提供的无私帮助。

参考文献（References）

[1] YANG V, ANDERSON W E. Liquid rocket engine combustion instability[M]. Reston: AIAA, 1995.

[2] CAO P B. Experimental study on the unsteady spray combustion process of a liquid oxygen/methane swirl coaxial injector[J]. 2021, 6(40): 26191-26200.

[3] 冯伟, 聂万胜, 李斌, 等. 模型燃烧室内不稳定燃烧发展过程的数值分析[J]. 北京航空航天大学学报, 2016, 42(6): 1195-1202.
FENG W, NIE W S, LI B, et al. Numerical analysis of unstable combustion developing process in model combustor[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(6): 1195-1202(in Chinese).

[4] YU Y C, SISCO J C, ROSEN S, et al. Spontaneous longitudinal combustion instability in a continuously-variable resonance combustor[J]. Journal of Propulsion and Power, 2012, 28(5): 876-887.

[5] SMITH R, XIA G P, ANDERSON W, et al. Extraction of combustion instability mechanisms from detailed computational simulations[C]//Proceedings of the 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition. Reston : AIAA, 2010.

[6] FREZZOTTI M L, NASUTI F, HUANG C, et al. Extraction of response function from numerical simulations and their use for longitudinal combustion intsbility modeling[C]//Proceedings of the 55th AIAA Aerospace Sciences Meeting. Reston: AIAA, 2017.

[7] GEJJI R, ANDERSON W E, AUSTIN B L. Transverse combustion instabilities in a high pressure multi-element combustor[C]//Proceedings of the AIAA Propulsion and Energy. Reston: AIAA, 2019.

[8] GEJJI R, LEMCHERFI A I, STRELAU R, et al. Combustion response of shear coaxial injectors to transverse combustion instabilities[C]//Proceedings of the AIAA Scitech 2020 Forum. Reston: AIAA, 2020.

[9] KASTHURI P, PAWAR S A, GEJJI R, et al. Coupled interaction between acoustics and unsteady flame dynamics during the transition to thermoacoustic instability in a multi-element rocket combustor[J]. Combustion and Flame, 2022, 240: 112047.

[10] Rayleigh. The explanation of certain acoustical phenomena1[J]. Nature, 1878, 18(455): 319-321.

[11] LUBARSKY E, HADJIPANAYIS M, SHCHERBIK D, et al. Control of tangential combustion instability by asymmetric baffle [C]//Proceedings of the 46th AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2008.

[12] KOBAYASHI T, MURAYAMA S, HACHIJO T, et al. Early detection of thermoacoustic combustion instability using a methodology combining complex networks and machine learning[J]. Physical Review Applied, 2019, 11(6): 064034.

[13] ARMBRUSTER W, HARDI J, OSCHWALD M. Impact of shear-coaxial injector hydrodynamics on high-frequency combustion instabilities in a representative cryogenic rocket engine[J]. International Journal of Spray and Combustion Dynamics, 2022, 14(1-2): 118-130.

[14] AHN K, CHOI H S. Combustion dynamics of swirl coaxial injectors in fuel-rich combustion[J]. Journal of Propulsion and Power, 2012, 28(6): 1359-1367.

[15] GRÖNING S, HARDI J S, SUSLOV D, et al. Injector-driven combustion instabilities in a hydrogen/oxygen rocket combustor[J]. Journal of Propulsion and Power, 2016, 32(3): 560-573.

[16] GRÖNING S, HARDI J, SUSLOV D, et al. Influence of hydrogen temperature on the stability of a rocket engine combustor operated with hydrogen and oxygen[J]. CEAS Space Journal, 2017, 9(1): 59-76.

[17] WIERMAN M, NUGENT N, ANDERSON W. Combustion response of a LOX/LCH$_4$ element to transverse instabilities[C]//Proceedings of the 47th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit. Reston: AIAA, 2011.

[18] BUSCHHAGEN T, GEJJI R, PHILO J, et al. Self-excited transverse combustion instabilities in a high pressure lean premixed jet flame[J]. Proceedings of the Combustion Institute, 2019, 37(4): 5181-5188.

[19] 严宇, 胡洪波, 洪流, 等. 自燃推进剂旋转爆震燃烧实验研究[J]. 推进技术, 2018, 39(9): 1986-1993.
YAN Y, HU H B, HONG L, et al. Experimental investigation on rotated detonation combustion of hypergolic propellants[J]. Journal of Propulsion Technology, 2018, 39(9): 1986-1993(in Chinese).

[20] XUE S J, YANG W D, ZHOU L X, et al. Experimental investigation of self-excited combustion instabilities in a small Earth storable bipropellant rocket combustor[J]. Aerospace Science and Technology, 2020, 105: 106008.

[21] CHU W, GUO K K, TONG Y H, et al. Numerical analysis of self-excited tangential combustion instability for an MMH/NTO rocket combustor[J]. Proceedings of the Combustion Institute, 2023, 39(4):

5053-5061.

[22] QIN J X, ZHANG H Q. Numerical analysis of self-excited combustion instabilities in a small MMH/NTO liquid rocket engine[J]. International Journal of Aerospace Engineering, 2020, 2020(1): 3493214.

[23] 初敏, 徐旭. 高频不稳定燃烧的声学数值仿真[J]. 北京航空航天大学学报, 2015, 41(7): 1215-1222.
CHU M, XU X. Acoustic numerical simulation of high frequency combustion instability[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(7): 1215-1222(in Chinese).

[24] GUO K K, REN Y J, TONG Y H, et al. Analysis of self-excited transverse combustion instability in a rectangular model rocket combustor[J]. Physics of Fluids, 2022, 34(4): 047104.

[25] PHILO J J, GEJJI R M, SLABAUGH C D. Injector-coupled transverse instabilities in a multi-element premixed combustor[J]. International Journal of Spray and Combustion Dynamics, 2020, 12: 1756827720932832.

Effects of oxygen-fuel ratio on combustion stability of a model rocket engine with hypergolic propellant

REN Yongjie, XU Boqi, CHU Wei, GUO Kangkang*, TONG Yiheng, NIE Wansheng

(Department of Aerospace Science and Technology, Space Engineering University, Beijing 101416, China)

Abstract: To estimate the effects of oxygen-fuel (O/F) ratio on the combustion stability of a model rocket engine with hypergolic propellant, experiments were carried out at different O/F ratios in a rectangular model engine with dual-liquid swirl coaxial injectors. The pressure oscillations and CH*-characterized heat release pulsations in the combustion chamber were simultaneously recorded by high-frequency pressure sensors and photomultiplier tubes (PMT). The effects of O/F ratio on combustion stability were obtained. The results show that the low-frequency oscillation of 41 Hz occurs in fuel mainfold after combustion initiation, which induces synchronous low-frequency oscillations in the combustion chamber during shutdown. In the process of increasing O/F ratio from 0.933 to 1.789, the combustion chamber undergoes a combustion stability transition process of stability, mild instability, first-order transverse instability, and second-order transverse instability. The amplitude of pressure oscillations in the 2W mode is only 4.69% of the mean combustion chamber pressure. By incorporating the PMT signal, it is found that the coupling of pressure and heat release signals is more obvious when combustion instability is more intense. Rayleigh index analysis based on the experimental data shows that the driving source of combustion instability in the 1W mode is mainly located at both sides of the combustion chamber, while suppression is found in the middle of the combustion chamber. The analysis suggests that the generation of combustion instability may be related to the interaction between the propellant and the combustion chamber walls.

Keywords: hypergolic propellant; oxygen-fuel ratio; high-frequency combustion instability; dual-liquid swirl coaxial injector; Rayleigh criterion

Received: 2023-07-05; **Accepted**: 2023-10-27; **Published Online**: 2023-11-23 16:34

URL: link.cnki.net/urlid/11.2625.V.20231122.1604.002

Foundation item: National Natural Science Foundation of China (12002386)

* **Corresponding author**. E-mail: guokangkang@sjtu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0446

动态区域聚焦的反无人机红外长时跟踪算法

谢学立，席建祥*，卢瑞涛，杨小冈，张涛，夏文新
(火箭军工程大学 导弹工程学院，西安 710025)

摘 要：无人机（UAV）的滥用催促着反无人机（anti-UAV）技术发展。基于红外探测器的反无人机目标跟踪技术成为反无人机领域的研究热点，但仍面临着因背景干扰所导致的跟踪失败问题。为提高复杂环境下红外反无人机跟踪的准确性和稳定性，提出一种动态区域聚焦的反无人机红外长时跟踪算法。构建基于特征金字塔的孪生主干网络，通过跨尺度特征融合，增强网络对红外无人机的特征提取能力。提出基于时空联合约束的动态区域建议网络，通过联合目标模板表观特征和目标运动约束，在全图范围内预测目标的定位概率分布，并引导先验锚框聚焦于候选区域，实现一种动态搜索区域选取。通过聚焦搜索区域，所提算法融合了局部搜索的抗背景干扰能力与全局搜索的重捕获能力，有效缓解全局搜索带来的负样本干扰，进一步增强目标特征的可辨别性。在Anti-UAV数据集上进行评测，与其他先进跟踪算法相比，所提算法具有更高的性能指标，跟踪精确率、成功率和平均准确度分别达到0.895、0.649和0.656，运行速度达到18.5帧/s，在快速运动、热交叉干扰和相似物干扰等复杂场景下，也具有优越的跟踪效果。

关 键 词：反无人机；红外目标跟踪；孪生神经网络；动态区域聚焦；时空联合约束
中图分类号：TN911.73; TP391
文献标志码：A **文章编号**：1001-5965(2025)09-3039-13

随着无人机应用迅速普及，无人机被滥用的危险性显著增加，给公共安全带来巨大威胁[1]。如何有效地反制非合作无人机入侵，成为亟待解决的社会安全问题。

作为无人机反制体系的重要环节，准确地跟踪非合作无人机是进一步实施反制的感知基础。相比于其他反无人机探测手段，红外成像探测器具有设备小型化、隐蔽效果好和工作全天时等优点，在反无人机系统中广泛应用。基于红外视频的无人机目标跟踪技术已成为反无人机技术领域的研究热点[2]。

红外无人机目标跟踪任务旨在给定红外无人机初始状态条件下(即给定初始红外帧和初始目标框)，自动识别后续红外帧中待跟踪无人机的位置与形状。相比于自然目标跟踪，红外无人机目标跟踪主要具有以下3个难点：①红外无人机目标尺度小，颜色信息缺失，纹理特征稀少，容易受到背景热杂波干扰；②无人机运动快速和消失-重现频繁，容易导致跟踪漂移；③环境热交叉干扰和相似物干扰使得难以区分红外无人机与其他背景目标。基于上述困难，如何设计鲁棒高效的红外反无人机跟踪算法仍面临巨大挑战。

近年来，基于孪生网络的目标跟踪算法因兼具良好的精度和速度而备受关注。文献[3]提出全卷积孪生网络(fully-convolu-tional siamese networks, SiamFC)，将目标跟踪问题转化为目标相似度匹配问题，成为后续孪生目标跟踪算法的设计范式。文献[4]提出孪生区域候选网络(siamese region proposal networks, SiamRPN)，在目标跟踪中引入了区域候选网络(region proposal networks, RPN)模块，实现了目标准确定位与边框精确回归。文献[5]提出的

收稿日期：2023-07-07；**录用日期**：2023-08-12；**网络出版时间**：2023-09-28 11:55
网络出版地址：link.cnki.net/urlid/11.2625.V.20230927.1550.003
基金项目：国家自然科学基金(62176263,62276274)；陕西省杰出青年科学基金(2021JC-35)
*通信作者. E-mail：xijx07@mails.tsinghua.edu.cn

引用格式：*谢学立，席建祥，卢瑞涛，等. 动态区域聚焦的反无人机红外长时跟踪算法[J]. 北京航空航天大学学报，2025，51(9)：3039-3051. XIE X L，XI J X，LU R T，et al. Long-term infrared object tracking algorithm based on dynamic region focusing for anti-UAV[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3039-3051 (in Chinese).*

SiamRPN++,解决了孪生目标跟踪算法存在的骨干网络深度限制问题,进一步提高了目标特征的表达能力。许多学者从目标模型在线更新[6-7]、目标鲁棒特征提取[8]和相似干扰物抑制[9-10]等角度对孪生目标跟踪算法进行改进,不断刷新目标跟踪算法的性能上限。然而,上述算法都建立在一种局部平滑假设上,强制将目标搜索区域限制在一个小领域内,属于短时跟踪算法。当无人机快速运动或消失-重现时,受搜索范围限制,此类算法容易发生跟踪错误和目标丢失问题。

长时跟踪算法具备全局搜索能力,能够在目标丢失后重捕获跟踪目标,更加适合红外无人机目标跟踪任务。当前的红外反无人机长时跟踪算法主要可以分为2类:①以跟踪-检测联合学习(tracking-learning-detection, TLD)算法[11]、粗-精结合目标跟踪(skimming perusal tracking, SPLT)算法[12]和长时跟踪元更新(long-term tracking with meta-updater, LTMU)算法[13]为代表的组合跟踪算法。此类算法以短时跟踪模型为基础,采用从局部到全局的搜索策略。当局部跟踪失效后,算法将开启全局检测进行目标重捕获。但在复杂红外背景下,红外无人机目标特征区分不明显,且环境还存在相似物干扰,此时难以准确判断目标是否丢失,从而导致全局重检测切换不稳定。②以GlobalTrack[14]和孪生区域卷积神经网络(siamese region-based convolutional neural networks, Siam RCNN)[15]为代表的全局跟踪算法。此类算法摒弃了局部搜索机制,直接从整幅图像中搜索目标,是一种全局搜索的孪生目标跟踪算法,具有结构简洁、端对端优化和无切换机制等优点,受到研究者广泛关注。文献[16]采用全局孪生目标跟踪范式,设计了SiamYOLO跟踪算法,利用运动信息抑制相似物干扰,实现了对红外无人机目标的高效跟踪。文献[17]在Siam RCNN中引入一种基于时空注意力的局部跟踪策略,通过自适应切换全局跟踪与局部跟踪,有效融合了局部跟踪与全局检测的优势,实现对红外无人机目标的鲁棒跟踪。针对红外反无人机目标跟踪任务难点,上述算法以全局跟踪算法为基础,引入不同的局部约束策略,提高了红外无人机目标跟踪性能。但这些策略主要是面向跟踪决策结果的后处理方法,没有从特征层面上改善红外无人机的特征表达能力,因此,在复杂红外环境下仍存在跟踪不稳定问题。

主要有2个原因导致上述问题:①固定的搜索区域。局部搜索和全局搜索都不适合红外反无人机跟踪。局部搜索通过聚焦于一个小邻域来提高搜索精度,但代价是搜索范围受限,导致搜索召回率降低。全局搜索通过重捕获丢失的目标来提高搜索召回率,但同时也引入了更多的背景干扰,使红外无人机目标陷入混乱的背景杂波中,导致搜索准确性下降。因此,将局部约束与全球搜索框架相结合,有利于构建更强大的红外反无人机跟踪算法。②目标特征的区分性不足。红外无人机缺乏颜色信息和纹理特征,使得难以区分与环境相似干扰物。此外,搜索区域的扩大会增加大量背景负样本。过多的简单负样本会干扰特征的区分性学习,从而导致目标和相似干扰物(即困难负样本)的辨别变得更加困难。因此,迫切需要加强目标特征的区分性,以便准确区分红外无人机目标。

基于上述分析,本文提出一种动态区域聚焦的反无人机红外长时跟踪算法。该算法采用从全局到局部的搜索策略,将时空约束引入到全局孪生目标跟踪框架中,引导网络聚焦红外无人机目标。具体地,针对红外无人机目标尺度小、表观特征稀疏的问题,构建了基于特征金字塔的孪生卷积骨干网络,通过跨层特征融合,将深层的语义特征与浅层的定位特征相融合,增强网络对红外无人机目标的特征提取能力。针对当前全局跟踪算法对红外无人机辨别力不足的问题,提出时空联合约束的动态区域建议网络(spatio-temporal joint constraints based dynamic region proposal networks, STC-DRPN),通过联合红外无人机的表观特征和运动信息,在全图范围内预测无人机的定位概率分布,并从锚点层面上,以一种软选取形式实现搜索区域的动态聚焦,进一步增强红外无人机特征的可辨别力。在Anti-UAV数据集上的结果表明,本文算法可以有效应对环境干扰,准确地跟踪红外无人机目标,为后续无人机反制奠定数据基础。

1 本文算法

本文算法的框架如图1所示,主要包括孪生特征提取网络和STC-DRPN子网络。其中,孪生特征提取网络用于提取模板图像与搜索图像的深度特征;STC-DRPN子网络用于生成与模板目标时空相似的候选目标,并判别出跟踪结果。

1.1 孪生特征提取网络

由于红外无人机目标尺度小,颜色信息缺失且纹理特征不明显,单尺度特征难以兼具语义信息和定位信息。因此,本文构建了具有特征金字塔结构的孪生骨干网络,旨在通过多尺度特征融合,获得更鲁棒的红外无人机深度描述特征。该骨干网络结构主要包含模板处理分支和搜索处理分支,分别输入模板图像$\boldsymbol{I}_z$和搜索图像$\boldsymbol{I}_s$,其中,模板图像通常为图像序列的第1帧,搜索图像为待处理的后续帧。首先,采用参数共享的骨干网络$\varphi(\cdot)$,提取模板

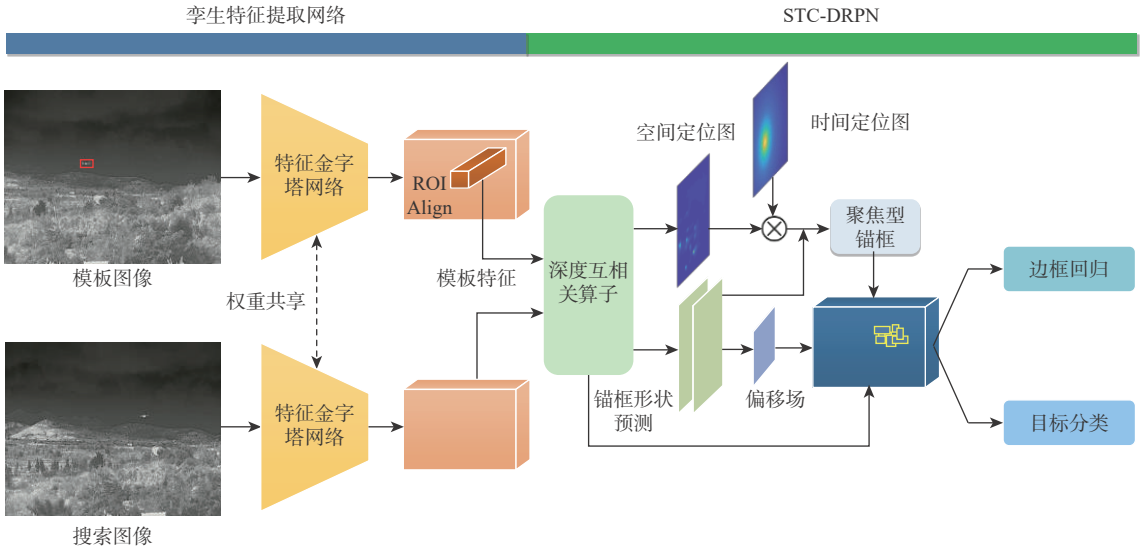


图 1　本文算法框架

Fig. 1　Framework of the proposed algorithm

图像特征 $\boldsymbol{z}=\varphi(\boldsymbol{I}_z)$和搜索图像特征 $\boldsymbol{s}=\varphi(\boldsymbol{I}_s)$; 然后, 构建参数共享的特征金字塔网络$\phi(\cdot)$, 通过融合骨干网络顶层特征与底层特征, 获得多尺度模板图像特征 $\boldsymbol{Z}=[\boldsymbol{Z}_1,\boldsymbol{Z}_2,\cdots,\boldsymbol{Z}_n]=\phi(z)$与多尺度搜索图像特征 $\boldsymbol{S}=[\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_n]=\phi(\boldsymbol{s})$; 最后, 采用 ROIAlign 算子, 根据初始目标框 $\boldsymbol{B}_1$, 从模板图像特征中提取多尺度模板目标特征 $\boldsymbol{F}=[\boldsymbol{F}_1,\boldsymbol{F}_2,\cdots,\boldsymbol{F}_n]=\mathrm{ROI}(\boldsymbol{Z},\boldsymbol{B}_1)$。本文采用 ResNet-50 作为骨干网络, 采用 FPN[18] 构建多尺度特征 P2～P6, 并通过 1×1 卷积将各层特征的通道数统一成 256。考虑到红外无人机目标尺度较小, 选择分辨率较高的 P2、P3、P4 作为输出。

1.2　STC-DRPN 子网络

在红外无人机跟踪过程中, 目标搜索范围的选取对跟踪性能具有重要影响[19]。传统的局部搜索只在前一帧目标位置的小领域内搜索, 是一种基于运动约束的硬选取策略, 如图 2(a) 所示, 虽然有利于减少背景干扰, 但难以应对无人机的快速运动和消失-重现。全局搜索则在全图范围内搜索, 不考虑运动先验, 是一种均匀选取策略, 如图 2(b) 所示。这种方式能够重捕获丢失的无人机目标, 但同时也引入大量的背景负样本, 干扰红外无人机的辨别, 容易引起跟踪跳变问题。考虑到局部搜索与全局搜索之间的互补性, 本文将局部约束引入到全局搜索框架中, 提出 STC-DRPN, 通过融合时空信息, 以一种软选取方式实现目标搜索区域的动态聚焦, 如图 2(c) 所示, 进而缓解背景负样本干扰问题, 提高目标特征的可辨别性。该子网络包括目标位置预测、锚框形状预测和特征适应 3 部分。

1.2.1　目标位置预测

目标位置预测旨在预测待跟踪目标在图像中可能出现的位置。由于待跟踪目标通常与模板目标表观匹配且与其先前运动模式相似, 因此, 本文利用红外无人机的时空线索, 即红外无人机的表观特征和运动信息, 对目标进行了 3 个层次的定位预测, 即空间定位预测、时间定位预测和时空融合预测。

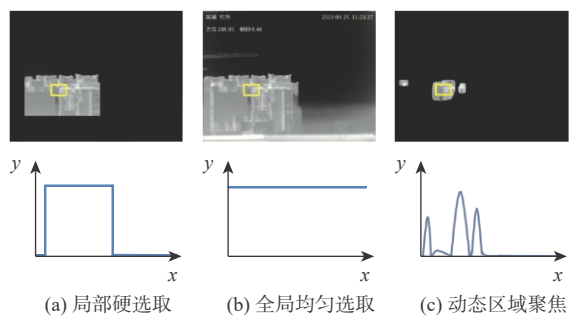


图 2　几类目标搜索区域选取策略

Fig. 2　Several kinds of target search region selection strategies

1)空间定位预测。从空间角度来看, 待跟踪的红外无人机与模板无人机是表观相似的, 因此, 可以利用特征匹配策略, 在搜索图像中寻找相似目标的位置。首先, 采用深度互相关算子[5], 将模板目标特征 $\boldsymbol{F}_\mathrm{T}\in\mathbf{R}^{k\times k\times c}$信息扩散到搜索图像特征 $\boldsymbol{F}_\mathrm{S}\in\mathbf{R}^{w\times h\times c}$中, 得到融合特征 $\boldsymbol{F}_\mathrm{m}\in\mathbf{R}^{w\times h\times c}$。然后, 采用 1×1 卷积整合 $\boldsymbol{F}_\mathrm{m}$的各通道特征, 输出单通道特征图 $\boldsymbol{F}'_\mathrm{m}\in\mathbf{R}^{w\times h\times 1}$; 最后, 采用 Sigmoid 函数将 $\boldsymbol{F}'_\mathrm{m}$压缩到[0,1], 得到空间定位预测图 $\boldsymbol{F}_\mathrm{SL}\in\mathbf{R}^{w\times h\times 1}$。空间定位预测的计算过程为

$$\boldsymbol{F}_\mathrm{SL}=\sigma(\varphi_\mathrm{c}(\boldsymbol{F}_\mathrm{T}\otimes\boldsymbol{F}_\mathrm{S})) \tag{1}$$

式中: ⊗为深度互相关算子; $\varphi_\mathrm{c}(\cdot)$为 1×1 卷积; σ为 Sigmoid 函数。

2)时间定位预测。从时间角度来看, 无人机的运动状态在前后时刻是连续的, 但因为相机抖动与拍摄间隔, 图像中的无人机运动可能呈现非平稳状

态。当无人机平稳运动时，其前后时刻的运动状态相关度高，此时可以通过运动模型预测下一帧目标位置。本文采用卡尔曼滤波算法估计目标运动状态。设目标运动状态$\boldsymbol{X}=[x_c,y_c,\dot{x}_c,\dot{y}_c]$，其中，$(x_c,y_c)$表示目标中心点的坐标，$(\dot{x}_c,\dot{y}_c)$表示中心点的运动速度。目标的状态模型为

$$\boldsymbol{X}_t=\boldsymbol{A}\boldsymbol{X}_{t-1}+\boldsymbol{w}_{t-1} \tag{2}$$

式中：$\boldsymbol{A}$为状态转移矩阵，对应着目标运动模型，本文采用匀速运动模型描述目标实际运动；$\boldsymbol{w}_{t-1}$为过程噪声，其协方差为$\boldsymbol{Q}$；观测模型为$\boldsymbol{O}_t=\boldsymbol{H}\boldsymbol{X}_t+\boldsymbol{v}_t$，$\boldsymbol{H}$为观测矩阵；$\boldsymbol{v}_t$为观测噪声，其协方差为$\boldsymbol{R}$。

预测误差的协方差为

$$\tilde{\boldsymbol{P}}_t=\boldsymbol{A}\boldsymbol{P}_{t-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{Q} \tag{3}$$

式中：$\boldsymbol{P}_{t-1}$为卡尔曼估计误差协方差。根据预测误差协方差$\tilde{\boldsymbol{P}}_t$，将运动预测的位置$\tilde{\boldsymbol{O}}_t=\boldsymbol{H}\boldsymbol{A}\boldsymbol{X}_{t-1}$高斯化，生成时间定位预测图$\boldsymbol{F}_{\mathrm{TL}}\in\mathbf{R}^{w\times h\times 1}$。

当运动突变时，预测结果将变得不可靠，此时取$\boldsymbol{F}_{\mathrm{TL}}=\boldsymbol{I}$，表示目标可能出现在全图任意位置。因此，时间定位预测图$\boldsymbol{F}_{\mathrm{TL}}=[\boldsymbol{f}_{\mathrm{TL}}^i]$为

$$\boldsymbol{f}_{\mathrm{TL}}^i=\begin{cases}\exp\left(-\left\|\boldsymbol{m}_i-\tilde{\boldsymbol{Z}}_t\right\|^2\Big/\alpha p_t\right) & \\ 1 & \text{其他}\end{cases} \tag{4}$$

式中：$\boldsymbol{m}_i$为$\boldsymbol{F}_{\mathrm{TL}}$中第$i$点的坐标；$\|\cdot\|^2$表示两点的欧氏距离；$\alpha$为平衡系数，本文取$\alpha=5$；$p_t$为位置预测误差的方差。

为判断目标是否突然运动，本文提出一种基于伪新息的机动检测方法。在卡尔曼滤波中，新息是指预测观测值$\tilde{\boldsymbol{O}}_t=\boldsymbol{H}\tilde{\boldsymbol{X}}_t$与实际观测值$\boldsymbol{O}_t$的差值。由于在时间定位预测过程中，无法准确获得实际观测值，因此，本文采用空间定位图的最高峰位置为跟踪观测估计值，即

$$\boldsymbol{O}_t'=\mathrm{argmax}(\boldsymbol{F}_{\mathrm{SL}}) \tag{5}$$

定义归一化新息平方ε_t为

$$\varepsilon_t=\boldsymbol{n}_t^{\mathrm{T}}\boldsymbol{L}_t^{-1}\boldsymbol{n}_t \tag{6}$$

式中：$\boldsymbol{n}_t=\boldsymbol{O}_t'-\tilde{\boldsymbol{O}}_t$为伪新息；$\boldsymbol{L}_t=\boldsymbol{H}\tilde{\boldsymbol{P}}_t\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R}$为伪新息协方差。由于$\varepsilon_t$服从卡方分布[20]，可设定机动检测门限为

$$P\{\varepsilon_t\leqslant\varepsilon_{\max}\}=1-\eta \tag{7}$$

其中：η为置信度分数，本文取$\eta=0.05$；$\varepsilon_{\max}$为对应的卡方临界值。若$\varepsilon_t>\varepsilon_{\max}$，则认为目标发生了机动。

3）时空间融合预测。本文采用“乘”的方式融合时/空定位预测图，得到时空目标定位预测图$\boldsymbol{F}_{\mathrm{L}}\in\mathbf{R}^{w\times h\times 1}$为

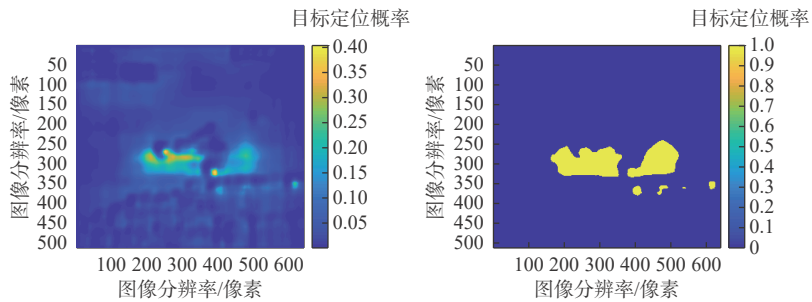
$$\boldsymbol{F}_{\mathrm{L}}=\boldsymbol{F}_{\mathrm{SL}}\odot\boldsymbol{F}_{\mathrm{TL}} \tag{8}$$

式中：$\odot$表示哈达玛积。

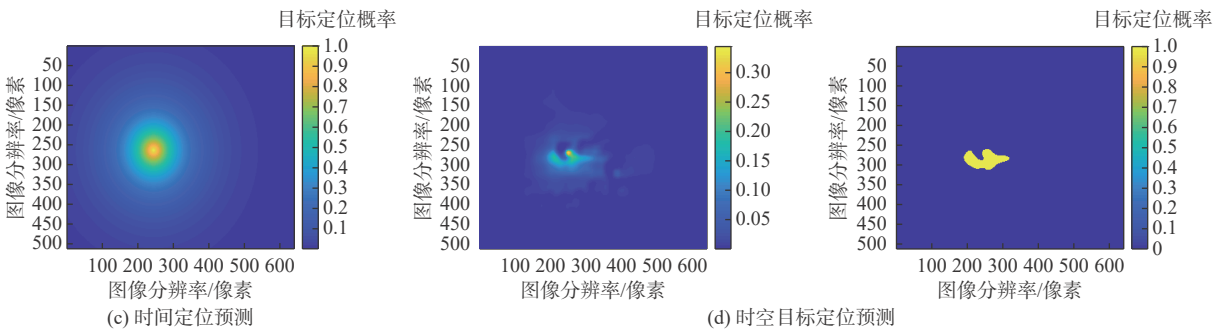
在目标位置预测过程中，无人机存在平稳运动和非平稳运动2种状态。当无人机运动平稳时，在时间定位图的作用下，空间定位图中远离预测点的区域会被逐渐抑制，有助于抑制相似物干扰。当运动突变时，时间定位图变为单位矩阵，此时目标定位概率图退化为空间定位图，防止错误的预测带来反作用。

根据预设的概率阈值，将目标定位预测图二值化，选定高置信区域为目标搜索区域，并逐点设置先验锚点。目标位置预测的示例如图3所示，在时

(a) 红外图像



(b) 空间定位预测



(c) 时间定位预测　(d) 时空目标定位预测

图3　目标位置预测流程示例

Fig. 3　Example of target location prediction process

空信息的双重聚焦下, 算法可以提取一个动态的目标搜索区域。通过比较空间定位预测图和目标定位预测图的二值图, 可以发现在时间定位预测图的作用下, 选定的目标搜索区域更聚焦于真实目标区域。

1.2.2　锚框形状预测

在确定锚点位置后, 需要预测最优的锚框形状, 以获得与真实目标最相似的先验锚框。首先, 采用输出通道不变的 1×1 卷积和 Leak ReLU 激活函数, 对融合特征 $\boldsymbol{F}_{\mathrm{m}}$ 进行非线性变换, 得到形状预测特征 $\boldsymbol{F}_{\mathrm{sh}}$; 然后, 采用 1×1 卷积将 $\boldsymbol{F}_{\mathrm{sh}}$ 映射为形状预测图 $\boldsymbol{F}'_{\mathrm{sh}} \in \mathbf{R}^{w\times h\times c}$。$\boldsymbol{F}'_{\mathrm{sh}}$ 的通道分别用于预测锚框的长和宽。最后, 将预测的锚框形状与先验锚点相结合, 得到区域聚焦的先验锚框。

为提高锚框形状预测的稳定性, 本文采用了一种非线性放缩变换[21], 将所要预测的锚框长 h 与宽 w 放缩为 $\hat{h}$ 和 $\hat{w}$, 变换关系为

$$\begin{cases} w = \tau v \mathrm{e}^{\hat{w}} \\ h = \tau v \mathrm{e}^{\hat{h}} \end{cases} \tag{9}$$

式中: v 为特征图的步长; τ 为尺度因子, 本文取 $\tau = 8$。

1.2.3　特征适应性变换

红外无人机尺度、形状多变, 使得特征图上不同像素所预测的锚框大小不同, 从而存在特征不匹配问题。为提高跟踪的准确性, 本文采用文献 [21] 提出的特征适应模块 $\mathcal{N}$, 增强特征对不同大小锚框的适应性。首先, 采用 1×1 卷积将 i 点的锚框形状信息 (w_i, h_i) 转化为可变形卷积的偏移量, 得到该点的偏移场 $\boldsymbol{f}_{\mathrm{o}}^{i}$。然后, 采用 $\boldsymbol{f}_{\mathrm{o}}^{i}$ 调整可变形卷积的采样位置, 并用来修正特征 $\boldsymbol{f}_{\mathrm{m}}^{i}$, 得到对锚框形状具有适应性的特征 $\boldsymbol{f}_{\mathrm{a}}^{i}$。适应性特征 $\boldsymbol{f}_{\mathrm{a}}^{i}$ 的增强过程为

$$\boldsymbol{f}_{\mathrm{a}}^{i} = \mathcal{N}\left(\boldsymbol{f}_{\mathrm{m}}^{i}, w_i, h_i\right) \tag{10}$$

式中: $\boldsymbol{f}_{\mathrm{m}}^{i}$ 为融合特征 $\boldsymbol{F}_{\mathrm{m}}$ 中第 i 点的特征; (w_i, h_i) 为相应的锚框形状。$\mathcal{N}$ 由 3×3 可变形卷积建模实现。

基于适应性特征 $\boldsymbol{F}_{\mathrm{a}} = \left[\boldsymbol{f}_{\mathrm{a}}^{i}\right] \in \mathbf{R}^{w\times h\times c}$, 对预设锚框进行分类与回归, 生成与模板目标时空相似的高质量候选目标, 并选择分类分数最高的候选目标为最终跟踪结果。分类分支与回归分支的结构与传统 RPN[22] 一致。不同的是, 本文采用掩膜卷积[23] 替换传统卷积。在跟踪时, 掩膜卷积以锚点分布区域为掩膜, 过滤其他无关区域特征, 减少了背景负样本干扰, 提高了运算效率。

1.3　损失函数

本文采用多任务损失函数 $\mathcal{L}_{\mathrm{total}}$ 对本文算法进行端对端训练。$\mathcal{L}_{\mathrm{total}}$ 由空间定位损失 $\mathcal{L}_{\mathrm{loc}}$、形状预测损失 $\mathcal{L}_{\mathrm{shape}}$、分类损失 $\mathcal{L}_{\mathrm{cls}}$ 和回归损失 $\mathcal{L}_{\mathrm{reg}}$ 组成:

$$\mathcal{L}_{\mathrm{total}} = \lambda_1 \mathcal{L}_{\mathrm{loc}} + \lambda_2 \mathcal{L}_{\mathrm{shape}} + \mathcal{L}_{\mathrm{cls}} + \mathcal{L}_{\mathrm{reg}} \tag{11}$$

式中: λ_1 和 λ_2 为损失函数平衡因子。

考虑到无人机目标在图像中占据的像素较少, 训练过程中存在严重样本不平衡问题, 本文采用 Focal loss[24] 作为空间定位损失用于目标位置预测:

$$\begin{aligned} \mathcal{L}_{\mathrm{loc}}\left(p, p^*\right) = \frac{1}{N_p}\sum_{i=1}^{N_p} -\big[&\beta p_i^* (1-p_i)^{\gamma} \ln(p_i) + \\ &(1-\beta)(1-p_i^*) p_i^{\gamma} \ln(1-p_i)\big] \end{aligned} \tag{12}$$

式中: N_p 为像素数量; β 和 γ 为调节系数, 本文取 $\beta = 0.25, \gamma = 2$; p_i 为算法预测的空间定位概率值; p_i^* 为对应的标签值, 其生成方式参考文献 [21]。

针对锚框形状预测, 本文采用文献 [21] 提出的匹配策略, 将锚点与真实目标框相匹配, 并选用 Bounded IOU 损失[25] 作为形状预测损失:

$$\begin{aligned} \mathcal{L}_{\mathrm{shape}}(w, h, w^*, h^*) = &\mathcal{L}_1(1 - \min(w/w^* - w^*/w)) + \\ &\mathcal{L}_1(1 - \min(h/h^* - h^*/h)) \end{aligned} \tag{13}$$

式中: (w, h) 为预测的锚框宽高; (w^*, h^*) 为匹配的真实目标宽高; $\mathcal{L}_1$ 为 Smooth L1 损失函数, 其计算式为

$$\mathcal{L}_1(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & \text{其他} \end{cases} \tag{14}$$

本文采用二值交叉熵损失用于锚框分类, 采用 Smooth L1 损失用于锚框回归。分类损失为

$$\mathcal{L}_{\mathrm{cls}}\left(p, p^*\right) = \frac{1}{N_{\mathrm{cls}}}\sum_{i=1}^{N_{\mathrm{cls}}} \left[p_i^* \ln(p_i) + (1-p_i^*)\ln(1-p_i)\right] \tag{15}$$

式中: N_{cls} 为锚框总数。

回归损失为

$$\mathcal{L}_{\mathrm{reg}}\left(t, t^*\right) = \frac{1}{N_{\mathrm{reg}}}\sum_{i=1}^{N_{\mathrm{reg}}} \mathcal{L}_1\left(t_i - t_i^*\right) \tag{16}$$

式中: N_{reg} 为正样本锚框数量; t_i 为预测的边框参数; t_i^* 为对应的真实边框参数。

1.4　跟踪流程

综合上述对算法关键模块的描述, 本文算法的主要流程如下。

输入: 图像序列 $\{\boldsymbol{I}^1, \boldsymbol{I}^2, \cdots, \boldsymbol{I}^L\}$;

初始帧目标边框 $\boldsymbol{B}^1 = \left[x_{\mathrm{c}}^1, y_{\mathrm{c}}^1, w^1, h^1\right]$。

输出: 每帧的目标边框 $\{\boldsymbol{B}^i | i = 1, 2, \cdots, L\}$。

1. for f=1 to L do

2. if f== 1 then

3. 通过骨干网络, 提取模板图像特征 $\boldsymbol{z}^1 = \varphi\left(\boldsymbol{I}^1\right)$;

4. 通过特征金字塔, 构建多尺度模板图像特征 $\boldsymbol{Z}^1 = \left[\boldsymbol{Z}_i^1 | i = 1, 2, 3\right] = \phi\left(\boldsymbol{z}^1\right)$;

5. 根据 $\boldsymbol{B}_1$，采用 ROIAlign 提取多尺度模板目标特征 $\boldsymbol{F}^1=\left[\boldsymbol{F}_i^1\right]=\mathrm{ROI}\left(\boldsymbol{Z}^1,\boldsymbol{B}^1\right)$；

6. If $1<f<L+1$ then

7. 通过骨干网络，计算多尺度搜索图像特征 $\boldsymbol{Z}^f=\phi\left(\varphi\left(\boldsymbol{I}^f\right)\right)$；

8. 根据式 (1)，计算目标空间定位预测图 $\left\{\boldsymbol{F}_{\mathrm{SL}}^{i,f}\right\}$；

9. 根据式 (4)，计算目标时间定位预测图 $\left\{\boldsymbol{F}_{\mathrm{TL}}^{i,f}\right\}$；

10. 计算目标定位概率图 $\boldsymbol{F}_{\mathrm{L}}^{i,f}=\boldsymbol{F}_{\mathrm{SL}}^{i,f}\odot\boldsymbol{F}_{\mathrm{TL}}^{i,f}$，并在高于阈值的区域设置锚点；

11. 预测各锚点的锚框形状，获得聚焦锚框 $\left\{\boldsymbol{B}_{\mathrm{anchor}}^{i,f}\right\}$；

12. 根据式 (10)，进行特征适应变换，获得适应性特征 $\left\{\boldsymbol{F}_{\mathrm{a}}^{i,f}\right\}$；

13. 基于 $\left\{\boldsymbol{F}_{\mathrm{a}}^{i,f}\right\}$，采用 RPN 对各锚框进行分类与回归，经过非极大值抑制（non-maximum suppression, NMS）后生成候选目标；

14. 选择目标判别分数最高的候选目标，并结合回归参数，输出目标边框 $\boldsymbol{B}^i$。

15. End for

2 实验与分析

2.1 数据集与评价指标

本文采用 Anti-UAV[26] 数据集对本文算法进行综合评测。Anti-UAV 是一个面向真实动态环境的红外反无人机目标跟踪数据集，包含 318 个高质量红外无人机飞行视频，每个视频的长度约为 1 000 帧，具有热交叉干扰、目标消失-重现、尺度变化、目标快速移动和相似物干扰等多种典型跟踪场景，十分具有挑战性。

本文采用一次运行评估（one-pass-evaluation, OPE）测试模式，选取精确率、成功率、运行帧率（frames per second, FPS）和平均准确度 4 个指标对算法进行全面评价。精确率是指在预测结果中，预测目标与真实目标之间的中心定位误差（center location error, CLE）小于给定阈值的帧数百分比。精确率的计算式为

$$p_{\mathrm{recision}}=\frac{1}{M}\sum_{t=1}^{M}\delta(C_t) \tag{17}$$

式中：M 为视频总帧数；T_{c} 为给定的精度阈值；δ 为指示函数，当 $C_t>T_{\mathrm{c}}$ 时，$\delta(C_t)=1$，否则，$\delta(C_t)=0$；C_t 为预测目标与真实目标之间的中心定位误差，其计算式为

$$C_t=\left\|\boldsymbol{L}_{\mathrm{p}}-\boldsymbol{L}_{\mathrm{g}}\right\|_2 \tag{18}$$

其中：$\|\cdot\|_2$ 为向量的二范数；$\boldsymbol{L}_{\mathrm{p}}$ 和 $\boldsymbol{L}_{\mathrm{g}}$ 分别为预测目标和真实目标的中心位置。

成功率是指在预测结果中，预测目标框与真实目标框之间的面积交并比（intersection over union, IOU）大于给定阈值的帧数百分比。成功率的计算式为

$$s_{\mathrm{uccess}}=\frac{1}{M}\sum_{t=1}^{M}\delta\qquad I_{\mathrm{OU}t}>T_0 \tag{19}$$

式中：$I_{\mathrm{OU}t}$ 为预测目标框与真实目标框间的面积交并比；T_0 为给定的阈值。通过计算不同阈值的精确率和成功率，可以绘制出算法的成功率曲线图和精确率曲线图。为进一步定量评价算法性能，本文选取 20 像素误差对应的精确率和成功率曲线的曲线面积（area under curve, AUC）值作为代表性指标。

此外，本文还采用文献 [26] 定义的平均准确度来评价算法的长时跟踪性能。平均准确度的计算式为

$$a_{\mathrm{cc}}=\frac{1}{T}\sum_{t=1}^{T}I_{\mathrm{OU}t}\delta(\kappa_t)+p_t(1-\delta(\kappa_t)) \tag{20}$$

式中：κ_t 为目标可见性标签；δ 为指示函数；当目标真实可见时，$\kappa_t=1$，$\delta(v_t)=1$，当目标不可见时，$\kappa_t=0$，$\delta(v_t)=0$；p_t 为预测指示函数，当预测为空时 $p_t=1$，否则 $p_t=0$。

2.2 实验实施细节

实验平台。操作系统为 Windows 10，编程语言为 python 3.7，深度学习框架为 pytorch 1.1。硬件配置为，CPU：Intel Core i9-9940X，3.30 GHz；GPU：4× NVIDIA RTX 2080Ti，12 GB。

算法参数。输入图像大小设为 640 像素×512 像素。ROIAlign 的输出特征尺寸设为 7 像素×7 像素。训练时，总损失函数中各项损失的权重设为 $\lambda_1=1,\lambda_2=0.1$。测试时，输出置信分数最高的目标。目标消失判别的分数阈值设为 0.1，目标定位概率阈值设为 0.1。其他参数设置参考文献 [14] 与文献 [21]。

训练细节。本文使用在 COCO 数据集上训练好的 Faster-RCNN 权重参数初始化本文的 ResNet-50 和 FPN。针对模型训练，本文在 4 个 NVIDIA 2080Ti GPU 上，采用随机梯度下降（stochastic gradient descent, SGD）算法对模型进行优化，设置动量为 0.95，权重衰减为 10^{-5}，每个 GPU 上的 batch size 为 2。图像对的采样方式与增强方式同 GlobalTrack 一致。为增强对红外图像的适应性，首先，采用 LSOTB-TIR[27] 数据集对模型进行预训练，训练轮次为 32，学习率为 10^{-3}，在第 16 轮和第 28 轮时学习率下降为上一轮的 0.1 倍。然后，采用 Anti-UAV 训练集对模型进行训练，训练轮次为 32，学习率为 10^{-3}，在

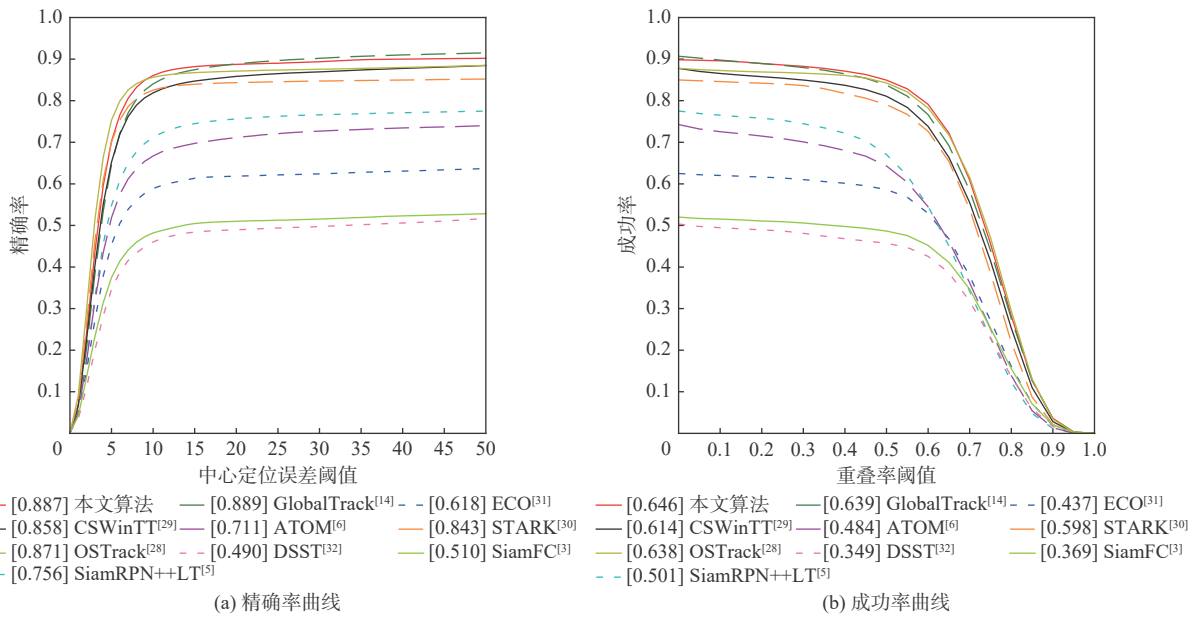
第 16 轮和第 28 轮时学习率下降为上一轮的 0.1 倍。

2.3　实验结果及分析

2.3.1　定量对比与分析

为定量评价算法跟踪红外无人机的性能, 本文在 Anti-UAV 训练集上将本文算法与 GlobalTrack[14]、OSTrack[28]、CSWinTT[29]、STARK[30]、SiamRPN++LT[5]、ATOM[6]、SiamFC[3]、ECO[31] 和 DSST[32] 这 9 种目标跟踪算法进行对比。各对比算法均使用默认的参数设置和公开的训练权重, 并在 Anti-UAV 训练集上训练至收敛。定量对比指标包括精确率、成功率、平均准确度 FPS。

图 4 为 10 种算法在 Anti-UAV 测试集上的精确率曲线和成功率曲线, 展示了不同阈值下各算法的精确率和成功率。可以直观的发现, 本文算法的精确率和成功率均优于其他对比算法。表 1 为 10 种算法的各项指标对比结果。本文算法的精确率、成功率和平均准确度分别达到 0.887、0.646 和 0.656, 运行帧率达到 18.5 帧/s, 总体性能超过 GlobalTrack 算法, 显著超越其他对比算法, 充分说明了本文算法的有效性和先进性。



(a) 精确率曲线　　(b) 成功率曲线

图 4　各跟踪算法在 Anti-UAV 测试集上的精确率曲线和成功率曲线

Fig. 4　Precision plots and success rate plots of different tracking algorithms on Anti-UAV test set

表 1　各算法在 Anti-UAV 上的定量对比结果

Table 1　Quantitative comparison results of tracking algorithms on Anti-UAV

算法	精确率	成功率	平均准确度	运行速率/(帧·s^{-1})
DSST[32]	0.490	0.349	0.354	31.2
SiamFC[3]	0.510	0.369	0.375	60.2
ECO[31]	0.618	0.437	0.444	7.5
ATOM[6]	0.711	0.484	0.490	28.7
SiamRPN++LT[5]	0.756	0.501	0.507	26.3
STARK[30]	0.843	0.588	0.607	33.5
CSWinTT[29]	0.858	0.614	0.623	8.5
OSTrack[28]	0.871	0.638	0.647	22.6
GlobalTrack[14]	0.889	0.639	0.648	10.2
本文算法	0.887	0.646	0.656	18.5

为全面评估算法性能, 本文针对目标消失(out-of-view, OV)、目标遮挡(occlusion, OC)、快速运动(fast motion, FM)、尺度变化(scale variation, SV)、热交叉(thermal crossover, TC)和低目标分辨率(low resolution, LR)6 种挑战属性, 开展算法对比实验。图 5 和图 6 分别为各算法在不同属性上的精确率曲线和成功率曲线。可以发现, 本文算法的成功率和精确率在各个属性上均领先其他算法, 充分体现了本文算法的性能优势。说明在时空聚焦机制的帮助下, 本文算法具有较强的尺度变化适应能力和目标判别能力。

图 7 为各算法在不同属性上的平均准确度对比结果, 本文算法的平均准确度在各属性上均排名第 1, 全面领先其他算法, 进一步展现了本文算法的优越性能。综合各算法在不同挑战属性上的精确率、成功率和平均准确度的排名情况, 可以说明本文算法能够较好地应对红外无人机目标跟踪过程中的各项挑战, 尤其在处理目标尺度变化挑战和低分辨率干扰挑战时, 具有显著的性能优势。

2.3.2　定性对比与分析

为更直观地分析算法跟踪红外无人机目标的性能, 本文列举了 GlobalTrack、ATOM、STARK、Siam-

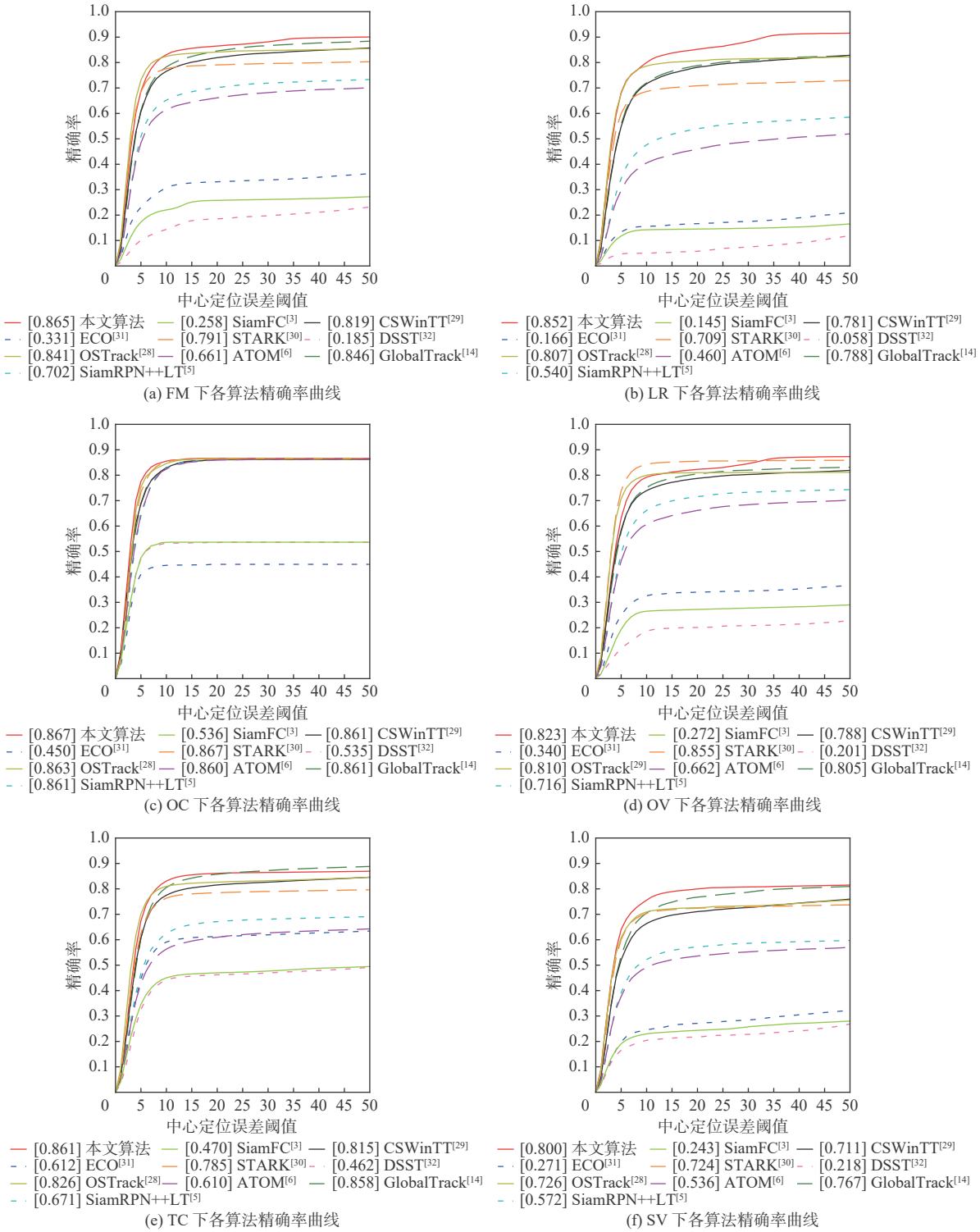


图 5 各跟踪算法在不同跟踪属性上的精确率曲线

Fig. 5 Precision plots of different tracking algorithms on different tracking attributes

RPN++LT、OSTrack、CSWinTT 和本文算法在 4 个典型场景下对红外无人机目标的跟踪结果，如图 8 所示。

图 8(a) 为各跟踪算法对快速运动目标的跟踪效果。在序列的第 53 帧、第 54 帧和第 822 帧、823 帧中，无人机在相邻帧之间发生大尺度位移，且产生运动模糊，给目标搜索带来困难。在该场景下，ATOM、OSTrack 等局部跟踪算法受限于固定的搜索领域导致容易丢失目标，而 GlobalTrack、SiamRPN++LT 和本文算法因具有全局搜索能力能够及时跟踪无人机目标，且其中本文算法的边框回归更精确，这是因为本文算法采用了自适应锚框策略，相比于 GlobalTrack，本文算法对目标尺度变化的适应能力更强。

图 8(b) 为各跟踪算法应对目标遮挡和相似物干扰的能力。在该序列中，无人机由于遮挡在第

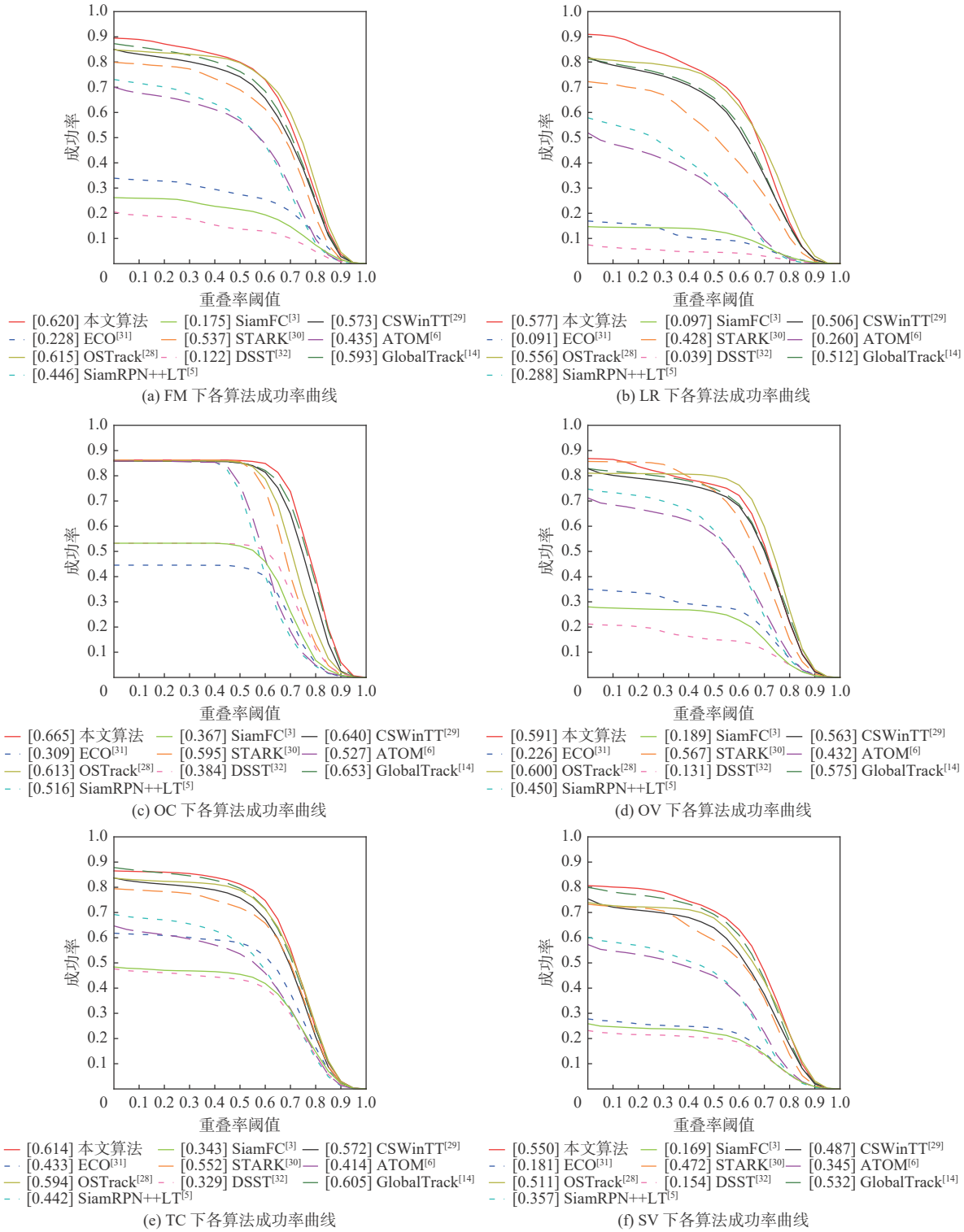


图 6　各跟踪算法在不同跟踪属性上的成功率曲线

Fig. 6　Success rate plots of different tracking algorithms on different tracking attributes

36 帧消失, 在第 102 帧重新出现。在第 435 帧和第 447 帧, 场景出现了其他空中干扰物。跟踪结果表明, 当无人机在第 102 帧重现时, 本文算法能够快速准确地重捕获目标, 而其他算法均表现相对较差。当背景中出现其他空中飞行物时, 如第 435 帧和第 447 帧, SiamRPN++LT、CSWinTT 和 GlobalTrack 因辨别能力不足, 会错误地跟踪干扰目标, 而本文算法具有动态聚焦机制, 能够利用时空线索来抵抗相似物干扰, 避免了跟踪错误问题。

图 8(c) 和图 8(d) 为复杂热交叉和热背景干扰下各算法的跟踪效果。在此场景下, 背景干扰严重, 无人机与背景环境的对比度非常小, 给跟踪算法的辨别能力带来很大挑战。通过观察 2 个视频序列的跟踪效果, 可以发现, 当无人机与背景建筑

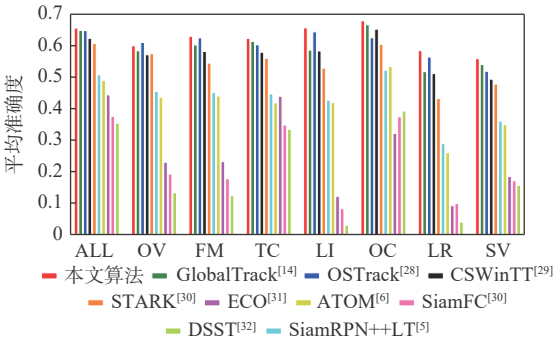


图7 各跟踪算法在不同属性上的平均准确度

Fig. 7 Average accuracy of different tracking algorithms on different attributes

相重叠时，其他算法均发生跟踪跳变和跟踪丢失现象，而本文算法在这种具有挑战性的情况下仍表现出出色的准确性和稳定性，这些结果有效地验证了本文算法比其他跟踪算法具有更强的目标辨别力。

综上，本文算法可以较好地处理目标快速运动、相似物干扰和背景热交叉干扰等多种困难跟踪场景。相比其他几种算法，本文算法能够更好地适应红外无人机目标跟踪任务。

2.3.3 消融实验

为探索算法各部分对算法性能的贡献，本文针对关键子模块和目标位置预测策略开展消融实验。以在Anti-UAV测试集上的精确率、成功率和平均准确度为评价指标。

关键子模块的消融实验的结果如表2所示，主要对FPN模块与STC-DRPN模块进行消融分析，表中每行表示一组模型变体。通过比较变体1和变体2、变体3和变体4的各项指标，可以发现基于STC-DRPN的变体比基于查询引导的区域建议网络(query-guided region proposal network, QG-RPN)[14]的变体有较大的性能提升。平均提升幅度为0.023 5的精确率、0.021的成功率和0.022 5的平均准确率。这归功于本文设计的时空联合约束策略能够有效减少相似物干扰，提高算法的跟踪性能。通过比较变体1和变体3、变体2和变体4的各项指标，

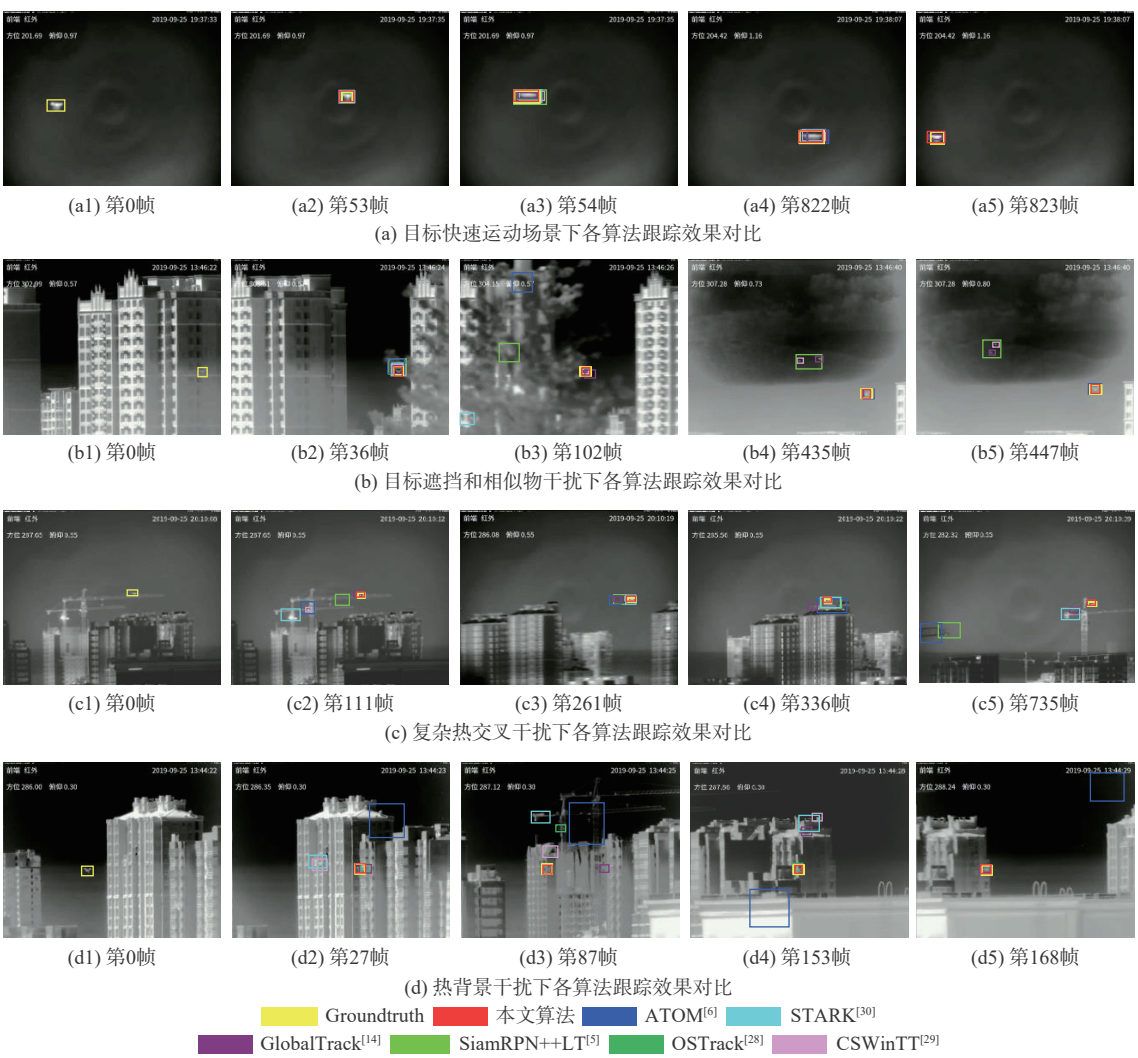


图8 各跟踪算法在不同场景下的定性跟踪结果

Fig. 8 Qualitative results of tracking algorithms in different scenarios

表 2　模型组件性能对比

Table 2　Performance comparison of model components

变体	FPN	STC-DRPN	QG-RPN	精确率	成功率	平均准确度
1			√	0.854	0.612	0.617
2		√		0.877	0.632	0.638
3	√		√	0.871	0.627	0.632
4	√	√		0.895	0.649	0.656

可以发现采用 FPN 后, 各项性能指标也均有一定提升, 说明通过跨层特征融合后, 生成的描述特征质量更高, 更适应红外无人机目标跟踪任务。

目标位置预测策略的消融结果如表 3 所示, 分别对比仅空间定位预测、仅时间定位预测和时空融合定位预测 3 种策略对算法性能的影响。可以看出, 时空融合定位预测的性能最高, 3 项指标比仅空间定位预测分别提高了 0.020、0.017 和 0.018, 而仅时间定位预测的性能最弱。这是因为时间定位预测本质上是一种局部策略, 难以应对无人机的突然运动和消失-重现; 空间定位预测是一种全局策略, 对抗背景干扰和相似物干扰的能力较弱。而时空融合定位预测巧妙结合了局部搜索的抗背景干扰能力与全局搜索的重捕获能力, 更适应红外无人机目标跟踪。

表 3　不同目标位置预测策略的性能对比

Table 3　Performance comparison of different target location prediction strategies

空间定位预测	时间定位预测	精确率	成功率	平均准确度
√		0.875	0.632	0.638
	√	0.713	0.474	0.485
√	√	0.895	0.649	0.656

3　结　论

本文提出一种动态区域聚焦的反无人机红外长时跟踪算法, 实现复杂背景下红外无人机目标的稳定跟踪, 为无人机反制奠定基础。

1) 本文 STC-DRPN 通过联合目标模板特征和运动特征, 引导先验锚点聚焦目标候选区域, 实现了目标搜索区域的动态聚焦, 有效缓解了全局跟踪带来的样本不平衡问题, 进一步增强了目标特征的可辨别性。

2) 本文算法采用从全局到局部的搜索策略, 融合了局部搜索的抗背景干扰能力与全局搜索的重捕获能力, 有效提高了红外无人机跟踪性能。

3) 本文算法实现了较为优异的红外无人机跟踪性能, 在 Anti-UAV 数据集上, 算法的精确率、成功率和平均准确度分别达到 0.895、0.649 和 0.656, 跟踪速度达到 18.5 帧/s。

4) 由于本文算法所采用的骨干特征来自于 FPN 的 P2、P3、P4 层, 其对应的分辨率跨步(Stride)分别为 1/8、1/16 和 1/32, 因此, 对于小于 7×7 的微小目标, 本文算法的跟踪性能受限。

下一步计划设计分辨率保持的特征提取网络和小目标增强模块, 以避免微小目标信号被背景噪声湮没, 进一步提高算法对微小目标的跟踪准确性。

参考文献 (References)

[1] 张瑞鑫, 黎宁, 张夏夏, 等. 基于优化 CenterNet 的低空无人机检测方法[J]. 北京航空航天大学学报, 2022, 48(11): 2335-2344.
ZHANG R X, LI N, ZHANG X X, et al. Low-altitude UAV detection method based on optimized CenterNet[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2335-2344 (in Chinese).

[2] 罗俊海, 王芝燕. 无人机探测与对抗技术发展及应用综述[J]. 控制与决策, 2022, 37(3): 530-544.
LUO J H, WANG Z Y. A review of development and application of UAV detection and counter technology[J]. Control and Decision, 2022, 37(3): 530-544(in Chinese).

[3] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking[M]//Computer Vision-ECCV 2016 Workshops. Berlin: Springer, 2016: 850-865.

[4] 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.

[5] LI B, WU W, WANG Q, et al. SiamRPN: evolution of siamese visual tracking with very deep networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4277-4286.

[6] 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.

[7] 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, 2019: 6181-6190.

[8] FAN H, LING H B. SANet: structure-aware network for visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2017: 2217-2224.

[9] BHAT G, DANELLJAN M, VAN GOOL L, et al. Know your surroundings: exploiting scene information for object tracking[M]//Computer Vision-ECCV 2020. Berlin: Springer , 2020: 205-221.

[10] 柏罗, 张宏立, 王聪. 基于高效注意力和上下文感知的目标跟踪算法[J]. 北京航空航天大学学报, 2022, 48(7): 1222-1232.
BAI L, ZHANG H L, WANG C. Target tracking algorithm based on efficient attention and context awareness[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(7): 1222-1232 (in Chinese).

[11] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.

[12] YAN B, ZHAO H J, WANG D, et al. 'Skimming-perusal' tracking: a framework for real-time and robust long-term tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 2385-2393.

[13] DAI K N, ZHANG Y H, WANG D, et al. High-performance long-term tracking with meta-updater[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 6297-6306.

[14] HUANG L H, ZHAO X, HUANG K Q. GlobalTrack: a simple and strong baseline for long-term tracking[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11037-11044.

[15] VOIGTLAENDER P, LUITEN J, TORR P H S, et al. Siam R-CNN: visual tracking by re-detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 6577-6587.

[16] FANG H Z, WANG X L, LIAO Z K, et al. A real-time anti-distractor infrared UAV tracker with channel feature refinement module [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Piscataway: IEEE Press, 2021: 1240.

[17] CHEN J J, HUANG B, LI J N, et al. Learning spatio-temporal attention based siamese network for tracking UAVs in the wild[J]. Remote Sensing, 2022, 14(8): 1797.

[18] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 936-944.

[19] SHI X R, ZHANG Y, SHI Z G, et al. GASiam: graph attention based Siamese tracker for infrared anti-UAV[C]//Proceedings of the 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications. Piscataway: IEEE Press, 2022: 986-993.

[20] 肖佳平, 蒋建春, 佘春东. 新息序列驱动的无人机控制系统数据攻击检测[J]. 控制理论与应用, 2017, 34(12): 1575-1582.
XIAO J P, JIANG J C, SHE C D. Data attack detection for an unmanned aerial vehicle control system using innovation sequences[J]. Control Theory & Applications, 2017, 34(12): 1575-1582(in Chinese).

[21] WANG J Q, CHEN K, YANG S, et al. Region proposal by guided anchoring[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 2960-2969.

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

[23] SONG G L, LIU Y, JIANG M, et al. Beyond trade-off: accelerate FCN-based face detector with higher accuracy[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7756-7764.

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

[25] TYCHSEN-SMITH L, PETERSSON L. Improving object localization with fitness NMS and bounded IoU loss[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6877-6885.

[26] JIANG N, WANG K R, PENG X K, et al. Anti-UAV: a large-scale benchmark for vision-based UAV tracking[J]. IEEE Transactions on Multimedia, 2021, 25: 486-500.

[27] LIU Q, LI X, YUAN D, et al. LSOTB-TIR: a large-scale high-diversity thermal infrared single object tracking benchmark[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(7): 9844-9857.

[28] YE B T, CHANG H, MA B P, et al. Joint feature learning and relation modeling for tracking: a one-stream framework[M]//Computer Vision-ECCV 2022. Berlin: Springer, 2022: 341-357.

[29] SONG Z K, YU J Q, CHEN Y P, et al. Transformer tracking with cyclic shifting window attention[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 8781-8790.

[30] YAN B, PENG H W, FU J L, et al. Learning spatio-temporal transformer for visual tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 10428-10437.

[31] ZOLFAGHARI M, SINGH K, BROX T. ECO: efficient convolutional network for online video understanding[M]//Computer Vision-ECCV 2018. Berlin: Springer, 2018: 713-730.

[32] DANELLJAN M, HÄGER G, SHAHBAZ KHAN F, et al. Accurate scale estimation for robust visual tracking[C]//Proceedings of the British Machine Vision Conference. Brighton: British Machine Vision Association, 2014: 1-11.

Long-term infrared object tracking algorithm based on dynamic region focusing for anti-UAV

XIE Xueli，XI Jianxiang*，LU Ruitao，YANG Xiaogang，ZHANG Tao，XIA Wenxin

(College of Missile Engineering，Rocket Force University of Engineering，Xi'an 710025，China)

Abstract: The misuse of unmanned aerial vehicles (UAV) is accelerating the development of anti-UAV technologies. Infrared detector-based tracking methods have gained special attention in the anti-UAV field, which, however, still face the problem of tracking failures caused by background interference. To enhance the precision and stability of infrared anti-UAV tracking in complex environments, this paper proposed a long-term infrared object tracking algorithm based on dynamic region focusing. Firstly, the Siamese backbone network based on feature pyramid was constructed to improve the feature extraction capability of the model for infrared UAV by the fusion of cross-scale features. Secondly, a dynamic region proposal network based on spatio-temporal joint constraints was proposed. Under the constraints of template appearance features and target motion information, the location probability distribution of the object was predicted over the entire image, and then the prior anchor box was guided to focus on the candidate regions, realizing a dynamic search region selection mechanism. The anti-background interference capability of local search and the recapture ability of global search were subtly integrated by focusing on the search area, which effectively mitigated the negative sample interference caused by global search and further enhanced the discriminability of target features. Experiments on the Anti-UAV dataset show that the proposed algorithm achieves precision of 0.895, a success rate of 0.649, and average accuracy of 0.656 with a tracking speed of 18.5 FPS. Compared with other advanced tracking algorithms, the proposed algorithm exhibits superior performance and demonstrates its effectiveness in handling complex tracking scenarios such as fast motion, thermal crossover, and similar distractors.

Keywords: anti-UAV; infrared object tracking; siamese neural networks; dynamic region focusing; spatio-temporal joint constraints

Received：2023-07-07；**Accepted**：2023-08-12；**Published Online**：2023-09-28 11：55

URL：link.cnki.net/urlid/11.2625.V.20230927.1550.003

Foundation items：National Natural Science Foundation of China (62176263,62276274); Shaanxi Province Science Funds for Distinguished Young Scholar (2021JC-35)

* **Corresponding author**. E-mail：xijx07@mails.tsinghua.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0503

导星辅助运动暗弱天体检测系统设计与仿真

张威[1,2]，冯文全[1,*]，孙国童[2]，王春雷[1]
(1. 北京航空航天大学 电子信息工程学院，北京 100191；2. 北京空间飞行器总体设计部，北京 100094)

摘 要：暗弱空间运动小目标检测在航天及军事领域有着广泛的应用，为提高空间目标监视系统的探测能力，提出并设计了导星辅助运动暗弱天体检测系统，该系统由星图模拟、图像去噪、质心解算、运动先验信息反向传播和主焦面成像组成。对星表及星空环境进行研究，确保星点坐标和噪声模拟的准确性；通过引入抖动误差、噪声误差和运动先验误差，更真实地模拟望远镜所处星空环境造成的误差影响，同时设计导星辅助的方法消除上述误差，并通过运动先验信息反向传播模块调整系统。实验证明：所设计的系统不仅可以模拟真实星图，而且可以减弱星空环境对望远镜的误差影响。

关 键 词：暗弱目标；空间目标检测；美国海军天文台导航星表；误差分析；系统仿真
中图分类号：TP317.4; P171
文献标志码：A **文章编号**：1001-5965(2025)09-3052-10

随着空间光学探测技术的发展，空间目标识别已成为当今外太空探索的主流[1]。星载大口径、高精度空间天文望远镜具有大气干扰小、观测波长宽和分辨率高的特点，具有较好的应用前景。然而，由于卫星的位姿难以估计，实时数据难以获取，现有研究大多通过天文星表给出的星等与空间位置信息构建数据库[2-3]，同时对真实星空场景进行分析，加入杂散光等噪声模拟真实星图场景。太空环境复杂，望远镜的光轴系统由于轨道运动和热变形等因素，在成像期间极易发生漂移或随机振动等不稳定的情况，导致成像质量低。

本文通过第4代美国海军天文台导航星表(USNO CCD astrograph catalog 4, UCAC4)[4]模拟生成空间星图，同时引入多种噪声，模拟真实场景下的星图，并进行去噪处理。为了满足天文望远镜的观测精度，本文设计一种基于导星辅助的运动暗弱天体检测系统，通过引入抖动误差、噪声误差和运动先验误差，模拟真实场景下望远镜的运行。

1 国内外研究现状

实现精细导星构建的星表是研究空间目标识别工作的数据基础，UCAC2[5]提出以来，众多学者致力于对星表的研究，NASA/IPAC河外星系数据库(NASA/IPAC extragalactic database, NED)包含了银河系内的恒星，NASA/IPAC红外科学档案库(NASA/IPAC infrared science archive, IRSA)包含了众多巡天星表[6]，同时支持时域测光数据，直到UCAC4的出现，几乎包含了可探测的所有恒星的详细数据。针对精细导星构建，朱长征等[7]提出了阈值剔除法，得到次优的最小导航星库。为了实现对精细导星星库的有效搜索，赵玥皎[2]提出混合分区星等过滤算法，但存在小视场的空洞问题。因此，崔祥祥等[3]通过增选部分候选导星和改进球面基准点，提出了新的星表构建方案。

空间图像的噪声模拟和去噪技术直接影响运动暗弱小目标的识别准确率。因此，需要较强的星图预处理技术抑制或消除星图背景中的噪声干扰。王松林等[8]改进了自适应加权中值滤波算法，避免了邻域噪声干扰，邹永宁等[9]提出了自适应中值滤波算法，在不损失边缘信息的情况下提升了去噪性能。

收稿日期：2023-08-02；录用日期：2024-01-26；网络出版时间：2024-03-08 10:13
网络出版地址：link.cnki.net/urlid/11.2625.V.20240306.1005.002
*通信作者. E-mail：buaafwq@buaa.edu.cn

引用格式：张威，冯文全，孙国童，等. 导星辅助运动暗弱天体检测系统设计与仿真[J]. 北京航空航天大学学报，2025，51(9)：3052-3061. ZHANG W，FENG W Q，SUN G T，et al. Design and simulation of guide star assisted moving faint object detection system[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3052-3061 (in Chinese).

由于在单帧灰度图像中, 目标成像信噪比低, 极易淹没在噪声中, 因此, 目前天基空间目标检测常采用基于参考星[10]、多帧时序投影[11]及轨迹检测等方法[12]。基于参考星的方法关键是利用星图中的指定背景恒星作为参考星, 根据 2 帧间的目标运动推断目标位置; 多帧时序投影是利用背景与目标的独立运动关系, 通过消除背景检测目标; 轨迹检测法通过目标逐帧运动轨迹实现连续性检测目标。

2　算法流程

2.1　整体流程

本文设计一种导星辅助的运动暗弱天体检测系统, 如图 1 所示。首先, 由 UCAC4 星表提取规定视场内所有星点信息, 根据星图中的恒星信息在有噪环境下模拟星图, 从模拟得到的星图序列图中选择所需的导星, 进行亚像素质心解算, 利用质心运动情况反算至望远镜运动信息。同时, 将提取得到的恒星信息用于运动拖尾图像模拟, 结合控制模块修正后的望远镜运动信息, 得到主焦面成像信息。最后, 根据主焦面运动拖尾图像计算运动先验误差。

图 1 中, “暗弱天体运动先验信息”是指由于导星和主焦面处于同一望远镜基座下, 因此, 通过导星的运动可以得到望远镜的先验信息, 即输出主焦面的先验信息。“暗弱天体检测模块”通过图像积分加强暗弱天体的检测, 并根据运动与先验误差信息进行调整, 辅助最终的暗弱天体检测。

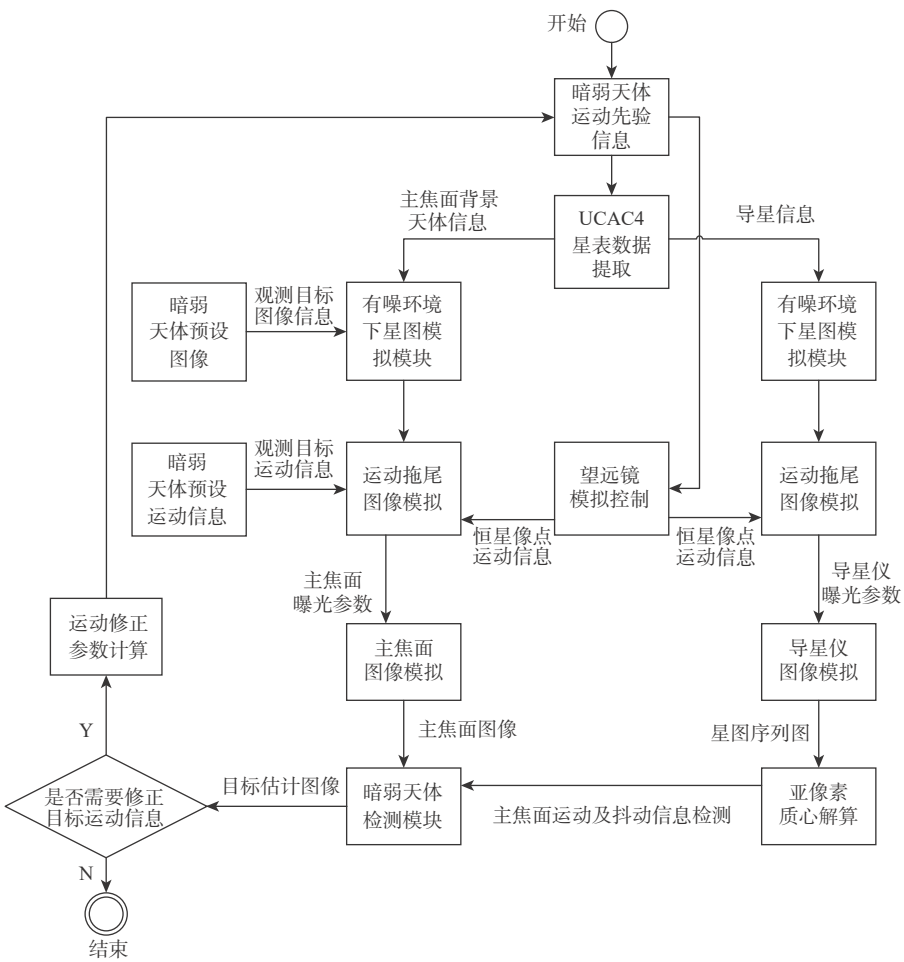


图 1　导星辅助的暗弱天体检测系统整体流程

Fig. 1　Overall flow of guide star-assisted faint object detection system

2.2　星表提取

UCAC4 是目前较为完备的天体测量星表, 共收录了约 1.14 亿颗恒星的赤经、赤纬、星等、自行等信息。该星表采用 J2000.0 历元, 坐标系为国际天球坐标系 (international celestial reference system, ICRS), 星等范围为 8～16 等。经过精度较高的 FK6, Hippocampus、Tycho 等星表的校正, 该星表在 10～14 等星等范围内定位精度可精确到 15～20 mas (1 mas=4.848 136 8×10^{-9} rad)。UCAC4 星表中星密度如图 2 所示。

提取出所需的星表信息后, 需要对星点位置进行修正[13], 利用星表中给出的天体自行数据, 修正

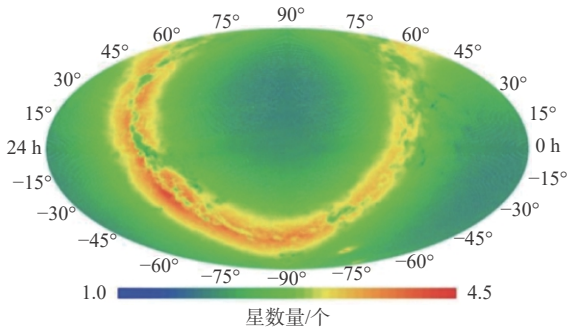


图2 UCAC4星密度分布[2]

Fig. 2 UCAC4 star density distribution[2]

为观测时刻的恒星位置，将ICRS坐标系下的数据通过坐标变换转换到望远镜成像坐标系，即恒星在主焦面上的成像位置，坐标变化如图3所示。

图3说明了将UCAC4星表中提取的导星赤经赤纬坐标转化为主焦面成像坐标的过程。其中，(α,δ)为天球坐标系下导星的赤经赤纬坐标，(x,y,z)为精细导星坐标系下导星的方向矢量坐标，(X,Y)表示经过透视投射后导星成像点坐标。

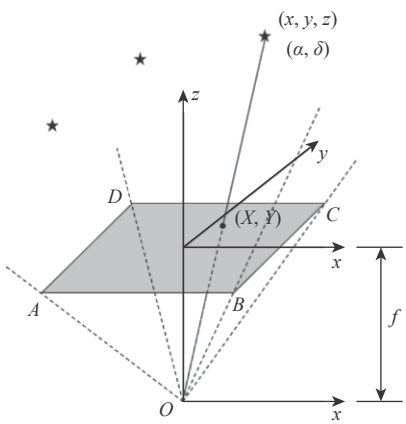


图3 UCAC4坐标变换示意图

Fig. 3 UCAC4 coordinate transformation diagram

2.3 星图模拟与预处理

2.3.1 恒星光源模拟

在实际应用中，恒星光源相对于天文望远镜可以看作无穷远处的点光源，光学焦面上的辐照度分布可以使用点扩散函数来描述[14]，点扩散函数成像为一个圆形的光斑，常在实际应用中用于模拟天体和太空目标。在实际应用中，可以使用零均值高斯分布函数来近似仿真点扩散函数，定义如下：

$$I(\Phi)=\begin{cases} I_0 & \Phi=0 \\ I_0\left[\dfrac{2J_1(\Phi)}{\Phi}\right]^{-2} & \Phi\neq 0 \end{cases} \tag{1}$$

式中：I_0为恒星的中心强度；J_1为一阶贝塞尔函数；$\Phi=\sqrt{(x'-x_0)^2+(y'-y_0)^2}$，恒星中心坐标为$(x_0,y_0)$，$x'$和$y'$分别为恒星周围像素点与中心的距离。

2.3.2 噪声模拟

由于天基空间系统所处工作环境及系统特性的影响，其存在的噪声大致可以分为2部分：来自光学望远镜系统的杂散光噪声和来自图像传感器与电子学系统的噪声[15]。

由于任何系统结构都难以完全消除杂散光的影响，因此，采用非均匀噪声表示系统外部的杂散光噪声。电子学系统噪声通常包含热噪声、光子散粒噪声、读出噪声、暗电流噪声和固定模式噪声(fixed pattern noise, FPN)。由于光电元器件中的热噪声是电子因随机热振动引起的，根据其产生原理通常可以通过降低元器件的工作温度来抑制热噪声的影响。暗电流噪声同样与温度变化相关，通过工艺的提高和制冷机制可以大幅降低该噪声的影响。FPN噪声是材料掺杂浓度或工艺缺陷等原因造成的，在探测器的位置相对固定，且容易受到暗电流噪声和图像放大器的影响。光子散粒噪声与随机起伏光子数和探测时间相关，其生成过程服从泊松分布。虽然通过现有的制冷机制可以降低部分噪声，但仍会干扰拍摄的图像。

综上所述，本文采用泊松噪声模型模拟光子散粒噪声和暗电流噪声，采用高斯噪声模型模拟热噪声和读出噪声，采用椒盐噪声模型模拟FPN噪声，最终，综合非均匀噪声、高斯噪声、泊松噪声和椒盐噪声模拟天基空间系统对拍摄图像的噪声干扰情况。由于星图图像背景占据主导且灰度值较低，同时存在分散但灰度值较高的目标与恒星像元，采用一维自适应中值滤波去除噪声干扰。

2.3.3 星图滤波预处理

导星仪的成像系统在输出图像时不仅包含导星目标，同时往往受到各种噪声源的干扰，根据噪声源的不同来源，可以分别采用不同的算法消除背景噪声的干扰。针对光电传感器产生的高频噪声，通常采用低通滤波的方法抑制[16]。针对空间杂光导致的低频干扰，常使用图像阈值分割法[17]。参考主流噪声去除方案，提出一种新的星图预处理方法。

2.4 亚像素质心解算

由于探测器和光学成像系统的工艺限制，恒星的亚像元尺寸与星图分辨率往往不能很好地适配，无法通过提高星图分辨率达到提高目标恒星质心准确率的目的，因此，现有方案大多通过离焦的方法扩大目标恒星成像范围，从而通过亚像素质心细分方法获取目标天体的质心位置信息。

目前，主流的质心定位方法主要分为灰度检测法[18]和边缘检测法[19]。灰度检测法主要利用天体的灰度分布情况，迭代获取质心坐标。边缘检测法则根据天体的轮廓形状信息，通过计算回归目标质心。

本文采用迭代开窗质心法结合双线性插值法

提取恒星质心, 具体而言, 迭代开窗质心法首先使用较大窗口利用质心法计算质心位置, 记录该质心像素, 再将窗口缩小至该像素附近, 再次计算质心位置, 不断重复该过程, 直到窗口大小等于艾里斑直径。通过迭代开窗法可以获得更精确的质心定位窗口, 使天体质心位置尽可能位于窗口中心, 以便获得最精密的解算结果。

2.5　主焦面成像

理想情况下, 暗弱天体运动与先验信息一致, 主焦面目标天体成像应为点状且边缘锐度较高, 当长时间曝光后, 背景中其他非运动天体的成像为时域拖尾成像, 出现沿先验运动信息方向的条状像斑。而曝光周期内, 相机因空间环境不确定性造成的抖动影响, 会导致主焦面所有天体像点像斑变大、对比度下降, 由于缺乏抖动或振动时间与振幅的先验信息, 导致单帧上无法有效区分目标和恒星。

通过引入导星辅助信息, 可以有效缓解甚至解决上述问题。通过导星解算出相机平台的运动和振动先验信息, 将该先验信息反传给主焦面平台, 同时根据导星图像中条状目标的信息, 并通过目标天体像斑的形状反算天体的真实运动信息, 以修正天体的运动先验。

2.6　运动参数反传

为了修正运动天体所带来的运动先验误差对目标成像的影响, 在导星辅助的运动参数校正主焦面图像后, 设计反传模块进一步提取暗弱目标图像中的残余信息, 精化先验信息, 指导下一次观测的望远镜指向控制, 称为运动先验信息反向传播。该模块主要包含 2 个功能: ①通过多帧导星的运动图像, 计算望远镜运动速度和导星的抖动误差, 消除抖动误差对整个焦面成像的影响; ②根据主焦面降噪后的图像和曝光时间信息, 以及望远镜运动先验信息, 结合主焦面上存在的条状目标角点, 计算曝光时间内该星点与望远镜之间相对运动的速度, 与望远镜的实际运动信息相结合, 即可计算天体的真实运动信息。

3　先验误差与处理

3.1　抖动误差

复杂星空环境下的望远镜光学元件会受到各种空间环境的干扰, 导致光线发生随机抖动, 影响主焦面成像质量。针对该问题, 通过多帧导星辅助减弱甚至消除抖动误差。

由于星图中的导星在成像的极短时间内可以认为是静止的, 且望远镜的运动先验也可以认为是一致的, 同时, 通过 2 帧间导星图像的跟踪残差可以确定运动先验, 因此, 在误差允许范围之外的导星运动可以认为是该时刻望远镜的抖动误差。

3.2　噪声误差

针对暗弱天体检测场景, 模拟星空中真实存在的噪声, 然而由于去噪算法的性能差异, 导致星图的去噪往往不尽如人意, 影响星图的观测, 若需要检测的暗弱天体被当做噪声削弱甚至消除, 会大大影响系统的性能。为了更有效去除星图的噪声, 采用一维自适应中值滤波去噪算法。选择较大的一维中值窗去除非均匀的背景噪声, 设滤波窗口为 W_0, 则背景估计可以表示为

$$\hat{\boldsymbol{B}}_1(i,j)=\underset{(x_b,y_b)\in W_0(x_b,y_b)}{\text{median}}\boldsymbol{F}_0(x_b,y_b) \tag{2}$$

式中: (x_b,y_b)为像素坐标; median(·) 表示中值滤波。减弱非均匀噪声背景的图像为

$$\boldsymbol{F}_1(i,j)=\boldsymbol{F}_0(i,j)-\hat{\boldsymbol{B}}_1(i,j) \tag{3}$$

此时可以去除整体非均匀噪声, 但仍然残留部分局部非均匀噪声, 因此, 需要继续迭代, 计算此时图像$\boldsymbol{F}_1(i,j)$的均值$\boldsymbol{F}_{\text{avg}}$为

$$\boldsymbol{F}_{\text{avg}}=\frac{1}{M\times N}\sum\boldsymbol{F}_1(i,j) \tag{4}$$

式中: M和N分别为行和列的像元个数。若以(i,j)为中心的窗口内灰度中值f_{med}小于灰度平均值f_{avg}, 则输出f_{med}, 否则增大窗口, 重新计算新窗口灰度中值, 直到$f_{\text{med}}\geqslant f_{\text{avg}}$。

第 2 次自适应中值滤波对非均匀噪声背景的精确估计为

$$\hat{\boldsymbol{B}}_2(i,j)=\underset{(x_b,y_b)\in W_{ij}(i,j)}{\text{median}}\boldsymbol{F}_1(x_b,y_b) \tag{5}$$

式中: $W_{ij}(i,j)$为第 2 次中值滤波的滤波窗口。

最终得到去噪后的图像$\boldsymbol{F}(i,j)$为

$$\boldsymbol{F}(i,j)=\boldsymbol{F}_1(i,j)-\hat{\boldsymbol{B}}_2(i,j) \tag{6}$$

实验证明, 该方法不仅能有效去除高斯噪声与光子散粒噪声, 而且能较好地去除非均匀背景噪声, 同时对星图中的目标暗弱天体影响较小。

3.3　运动先验误差

实际观测中暗弱天体的真实运动为未知参数, 望远镜只能根据运动先验信息控制主焦面运动, 运动先验误差主要指在计算望远镜运动速度时, 由于导星的平移、旋转、抖动等不确定因素导致的恒星在导星仪、主焦面图像的不规则运动情况。

对目标天体的图像, 本文仿真中首先根据目标天体运动真实信息反传给主焦面, 结合望远镜运动得到计算后的目标暗弱天体理论运动图像, 针对该图像采用角点检测的方法, 获得高精度亚像素级的

角点定位结果，反算目标天体真实运动的修正量。

对条状目标的信息提取，主要是对其端点进行定位。根据端点位置信息，即可精确解算其相对于望远镜的运动角速度。根据该运动角速度即可计算天体的真实运动情况。

在实际观测中，部分目标在主焦面视场下会成像出具有一定弧度的曲线段，甚至出现像旋。因此，采用角点检测的方法解决存在弧度的目标进行端点定位，首先，利用2.3.3节的方法对主焦面图像进行预处理，对其进行去噪。然后，使用高斯函数的一阶微分作为梯度算子卷积核，对原图像作卷积操作，并计算图像在水平、垂直方向的导数：

$$\begin{cases} P_x = \dfrac{\partial \boldsymbol{P}(x,y)}{\partial x} \approx \boldsymbol{K}_x * \boldsymbol{P}(x,y) \\ P_y = \dfrac{\partial \boldsymbol{P}(x,y)}{\partial y} \approx \boldsymbol{K}_y * \boldsymbol{P}(x,y) \end{cases} \tag{7}$$

式中：$\boldsymbol{P}(x,y)$为主焦面图像；$\boldsymbol{K}_x$和$\boldsymbol{K}_y$分别为x和y方向上的梯度算子。

通过计算图像水平和垂直两维度梯度的积，并对图像进行卷积操作，得到相对稳定的结果$\boldsymbol{D}$。最后计算自相关矩阵与每个像素的Harris响应R：

$$\begin{cases} \boldsymbol{D} = \begin{bmatrix} w*P_x^2 & w*P_xP_y \\ w*P_xP_y & w*P_y^2 \end{bmatrix} \\ R = \det(\boldsymbol{D}) - k\cdot \mathrm{trace}^2(\boldsymbol{D}) \end{cases} \tag{8}$$

式中：w为高斯核（窗口大小为7×7，标准差$\sigma=2$）；$\det(\boldsymbol{D})$为自相关矩阵$\boldsymbol{D}$的行列式；k为常数系数；$\mathrm{trace}(\boldsymbol{D})$为自相关矩阵$\boldsymbol{D}$的迹。

通过阈值判断每个像素的响应是否存在疑似角点。后续采用一定的窗口，对所有像素点进行非极大值抑制操作，即可保留图像中真实的角点位置。获得像素级的角点位置后，为提高定位精度，需要对该位置坐标进行亚像素级的定位修正。可以以角点为中心，在其附近取一个特定大小的窗口，并在此范围内使用灰度质心法，即可获得高精度的亚像素级端点定位结果。

4 仿真模型与验证

4.1 仿真平台搭建

仿真平台主要使用Python进行开发，由5个主要模块构成，分别为UCAC4星表数据提取模块、有噪环境下星图模拟模块、亚像素质心解算模块、控制系统仿真模块和运动拖尾图像模拟模块。

4.2 星表数据提取

将二进制格式的UCAC4星表进行一定规则的划分，提高星图生成的搜索效率，通过对UCAC4星表进行分析，本文系统可根据输入视场中心点的赤经赤纬坐标，自动提取相应视场恒星的赤经、赤纬、星等及自行信息，同时将赤经赤纬转化至望远镜成像坐标系，将星等转化为恒星的亮度信息。

根据所输入视场中心点的赤经赤纬，该仿真系统可以输出所需不同视场下的星图：$[x_{\mathrm{RA}}-\beta_{\mathrm{RA}}, x_{\mathrm{RA}}+\beta_{\mathrm{RA}}]$，$[x_{\mathrm{Dec}}-\beta_{\mathrm{Dec}}, x_{\mathrm{Dec}}+\beta_{\mathrm{Dec}}]$。其中，$x_{\mathrm{RA}}$为输入视场中心赤经坐标，$x_{\mathrm{Dec}}$为输入视场中心赤纬坐标，$\beta_{\mathrm{RA}}$和$\beta_{\mathrm{Dec}}$分别为视场赤经和赤纬范围，默认$\beta_{\mathrm{RA}}=\beta_{\mathrm{Dec}}=0.925$。

由于星表中的星等信息与天体亮度存在特定的转换关系，同时星等的不同级别间亮度相差固定的2.512倍[20]。为了方便，将星等信息转化为等效视星等，即假设星等为a_1的星体照度为$E(a_1)$，另一星等为a_2的星体照度为$E(a_2)$，两者的关系为

$$E(a_2) = E(a_1)\,10^{0.4(a_1-a_2)} \tag{9}$$

最后，通过点扩散函数对每颗恒星进行模拟。

4.3 有噪环境下星图模拟

4.3.1 星图噪声模拟

通过星表提取可以得到输入视场内恒星的亮度和位置信息，由于星图中噪声影响因素繁多，难以获得一个准确且完备的光电系统噪声模型，为了更真实地反映复杂空间真实情况，采用高斯噪声、光子散粒噪声和非均匀噪声模拟真实星图场景。高斯分布的非均匀噪声用来模拟杂散光：

$$I = \frac{A}{2\pi\sigma^2}\exp\left[-\frac{(x-x_{\mathrm{dif}})^2+(y-y_{\mathrm{dif}})^2}{2\sigma^2}\right] \tag{10}$$

式中：A为幅值，控制最大灰度值的大小，高斯函数扩散中心坐标为$(x_{\mathrm{dif}}, y_{\mathrm{dif}})$。

4.3.2 星图去噪统计

通过引入星图背景图像的非均匀、强弱分布不均及高斯噪声和光子散粒噪声模拟真实场景下的星图，为了实现质心解算与主焦面运动拖尾模拟，需要对星图进行去噪，由于杂散光影响，星图中左下角灰度值低，右上角灰度值高，为了保证后续质心解算算法的准确率，同时对噪声进行较高水平的滤波去噪，将两者相折中，本节通过一维自适应中值滤波去噪算法去除由于噪声导致的背景灰度变化误差，抑制背景非均匀性所带来的干扰信息，得到均匀背景的星图，如图4所示，左半部分为背景灰度信息，右半部分为天体星等分布信息。

4.3.3 噪声误差统计

目标天体由于其成像区域像素高亮，背景估计几乎不会受到影响，因此，该方法在去除背景噪声的同时，几乎不会降低弱小目标的信噪比，同时由于背景噪声得到了抑制，相当于提高了弱小目标的信噪比，增强了弱小目标检测性能。去噪星图如图5所示。

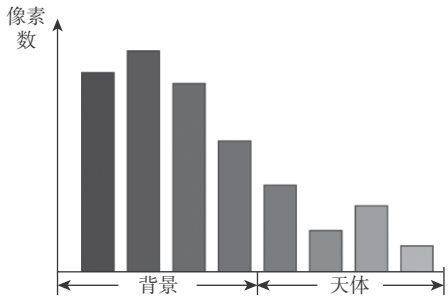


图 4　星图背景和目标灰度分布直方图

Fig. 4　Star image background and target grayscale distribution histogram

(a) 视场内星图位置信息　(b) 加入噪声后的星图　(c) 去噪星图

图 5　星图仿真去噪结果

Fig. 5　Star map simulation denoising results

采用残差分析的方法验证去噪算法的有效性, 则残差图像$\boldsymbol{E}(i,j)$可以表示为

$$\boldsymbol{E}(i,j)=\boldsymbol{B}_{\mathrm{p}}(i,j)-\boldsymbol{B}_{\mathrm{q}}(i,j) \tag{11}$$

式中: $\boldsymbol{B}_{\mathrm{p}}(i,j)$为原始非均匀背景噪声; $\boldsymbol{B}_{\mathrm{q}}(i,j)$为算法估计的背景噪声。残差背景噪声统计如表 1 所示。

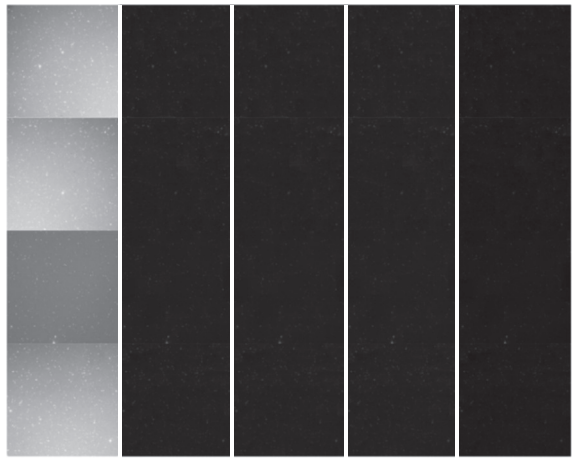
表 1　残差背景噪声统计

Table 1　Residual background noise statistics

图像	均值/像素	标准差/像素
原始图像	63.25	18.73
背景消除图像	0.56	0.77

由表 1 可知, 经过一维自适应中值滤波算法处理后, 可以有效抑制仿真星图中的非线性背景噪声和空间中的各种高斯噪声与光子散粒噪声, 提高弱小目标的检测性能。

为验证本文方法不仅对仿真星图具有一定效果, 同时也适用于实拍星图上的去噪, 在现有实拍星图上进行去噪处理, 如图 6 所示。可知, 原始星图中除了非均匀背景噪声, 还存在其他类型的噪声。对原始星图(见图 6(a))进行非均匀背景矫正, 得到去除非均匀噪声后的矫正星图(见图 6(b)), 采用均值滤波(见图 6(c))、高斯滤波(见图 6(d))等



(a) 实拍星图 (b) 背景矫正 (c) 均值滤波 (d) 高斯滤波 (e) 本文方法

图 6　实拍星图去噪结果

Fig. 6　Real star map denoising results

传统方式对所采用的星图进行去噪处理, 但所获得的星图仍存在一定程度的噪声, 此图像会极大影响后续目标识别及质心定位精度。通过矫正星图的分析, 发现残留的噪声多为局部灰度值较高的噪声, 类似椒盐噪声, 通过本文方法进行矫正后, 能在去噪的同时保留目标信息, 以提高后续的目标识别精度。

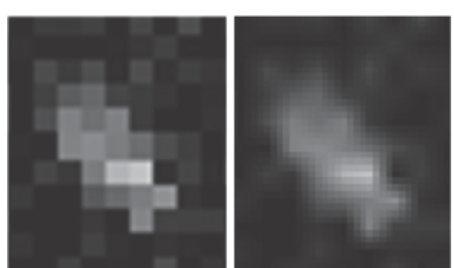
4.4　亚像素质心解算

4.4.1　质心解算

亚像素质心解算模块输入为去噪后的星图及星图内的多颗导星坐标, 此处默认为 2 颗。采用迭代开窗质心法结合双线性插值法提取恒星质心并计算质心位移, 根据多颗导星质心在多张连续时间图像上的位移信息, 最终得到目标天体在主焦面上的移动速度。具体实现流程如下:

1) 质心窗口定位。将去噪后的星图利用阈值分割法提取最小外接矩形的连通域, 对该矩形进行预设像素个数的扩张, 该窗口即星点质心定位的初始窗口。

2) 双线性插值。将窗口内图像上采样所需精度的倍数, 此处默认为 4 倍, 获得插值后的图像如图 7 所示。



(a) 质心定位窗口　(b) 双线性插值

图 7　亚像素质心解算

Fig. 7　Sub-pixel quality centroid solution

3)目标质心解算。利用灰度质心法计算上采样后的质心坐标:

$$x_{\mathrm{i}}=\frac{\sum_{x=1}^{m}\sum_{y=1}^{n}\left[F_{\mathrm{c}}(x,y)-T\right]^{2}x}{\sum_{x=1}^{m}\sum_{y=1}^{n}\left[F_{\mathrm{c}}(x,y)-T\right]} \tag{12}$$

$$y_{\mathrm{i}}=\frac{\sum_{x=1}^{m}\sum_{y=1}^{n}\left[F_{\mathrm{c}}(x,y)-T\right]^{2}y}{\sum_{x=1}^{m}\sum_{y=1}^{n}\left[F_{\mathrm{c}}(x,y)-T\right]} \tag{13}$$

式中:T为预设的阈值。

4)质心定位校准。计算窗口内的坐标中心位置,根据坐标偏差校准窗口位置,在新的中心位置窗口中不断迭代质心位置,最终得到高精度的质心坐标。

不同的初始窗口会导致计算得到的质心位置与真实质心位置误差较大,本文通过迭代开窗法不断缩小质心位置区域,降低质心定位的误差,提高质心定位精度。

根据同一拍摄位置的多张连续星图,按上述流程分别计算每帧视场多导星质心的位置,并与前一时刻视场位置做差,根据每帧采样时刻导星仪解算瞬时速度,反传望远镜的实际运动信息。

4.4.2 抖动误差计算

设前一时刻导星的位置为(x'_0,y'_0),后一时刻导星的位置为(x'_1,y'_1),如图8所示,导星的运动先验为v,先验误差为δ_1,则抖动误差范围可以表示为

$$\delta_2=\frac{\sqrt{(x'_1-x'_0)^2+(y'_1-y'_0)^2}}{t}-v\pm\delta_1 \tag{14}$$

式中:t为时间。

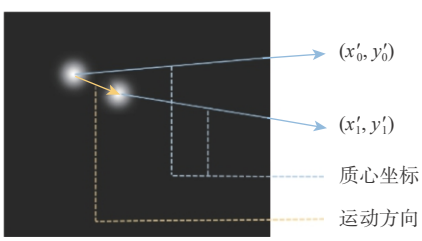


图8 2帧间导星运动示意图

Fig. 8 Schematic diagram of guide star movement between two frames

针对单导星多帧与多导星2帧间设计消融实验,验证导星辅助天体检测对抖动误差的消除影响,实验结果如表2与表3所示。

如表2所示,当选定导星后,由于导星的运动先验信息已知,因此,最终目标定位点的误差主要来源于平台的抖动误差,随着导星曝光帧数的增加,误差逐渐减小。由表3可知,当每颗导星的运动先验已知时,随着导星个数的增加,抖动误差不断减小。实验表明,多颗导星多帧可有效减少抖动误差,但由于模拟的星图较大,帧率会严重影响算法效率,因此,采用多导星2帧的方式减少抖动误差。

表2 单导星多帧消融实验结果

Table 2 Results of ablation experiment for single guide star with multiple frames

导星帧数/帧	运动先验误差/(像素·帧$^{-1}$)	抖动误差/(像素·帧$^{-1}$)
2	1.37	0.631
3	1.37	0.562
4	1.37	0.560

表3 多导星2帧消融实验结果

Table 3 Results of ablation experiment for multiple guide stars with two frames

导星数/个	运动先验误差/(像素·帧$^{-1}$)	抖动误差/(像素·帧$^{-1}$)
1	1.37	0.631
2	1.37/0.751	0.491
3	1.37/0.751/−0.107	0.412

4.5 主焦面仿真

4.5.1 运动拖尾图像模拟

主焦面图像指相机拍摄的包含目标暗弱天体在内的天空视场成像。主焦面仿真模块由于其长时间的曝光积分,导致其图像必然存在一定程度的运动拖尾现象,因此,通过导星计算出的运动先验误差修正信息,进一步降低运动拖尾图像中的天体运动先验信息误差。

运动拖尾图像模拟的输入为UCAC4星表信息、曝光时间及望远镜的实际运动信息。根据UCAC4星表的数据,计算目标天体在主焦面上的成像信息。根据望远镜每个时刻的运动信息,对曝光时间内的每个时刻进行亮度积分,得到积分完成的主焦面星图,如图9所示。

随后按照长曝光条件下不同噪声的影响,对生

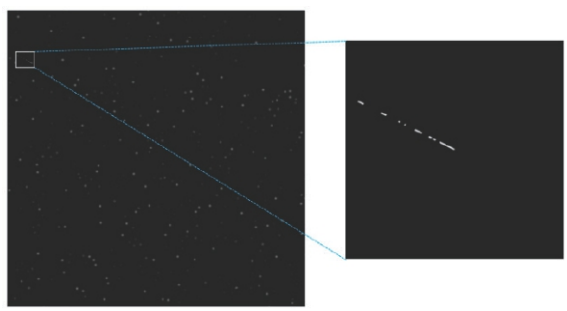


图9 目标天体主焦面成像区域积分

Fig. 9 Integral of the imaging area of the main focal plane of the target celestial body

成完毕的星图进行加噪。对生成的多张主焦面单帧图像进行叠加处理, 进一步提高成像的信噪比。通过检测主焦面上存在的条状目标角点, 计算曝光时间内该星点与望远镜之间相对运动的速度。

同时, 由于天体的运动先验存在一定误差, 因此, 根据主焦面条状目标的信息, 可以反算天体的真实运动信息, 以修正天体的运动先验。对条状目标的信息提取, 主要是对其端点进行定位。根据端点位置信息, 即可精确解算其相对于望远镜的运动角速度。根据该运动角速度即可计算天体的真实运动情况。利用角点检测的方法可以对存在弧度的目标进行端点定位。Harris 角点检测方法采用大小固定的滑动窗口, 检测图像内每个像素邻域的灰度变化情况, 若灰度在某处剧烈波动, 则认为该区域存在角点。

4.5.2　运动先验修正

主焦面角点检测对运动先验修正的影响结果如图 10 所示, 其中, 虚线框表示目标天体运动范围检测框, 实线框为所需目标天体。图 10(a) 为理想情况下主焦面运动拖尾图像, 图 10(b) 为加入噪声与运动先验误差的主焦面真实图像, 图 10(c) 为角点检测结果, 其中, 圆形表示检测到的角点像素位置, 十字表示修正后的目标端点位置, 虚线框内的角点检测结果为检测出的抖动误差, 图 10(d) 为范围检测框内目标天体修正运动先验误差与运动先验后的结果。实验结果表明, 所采用的角点检测不仅可以修正运动后的先验误差, 而且可以有效检测出主焦面抖动误差所导致的角点。

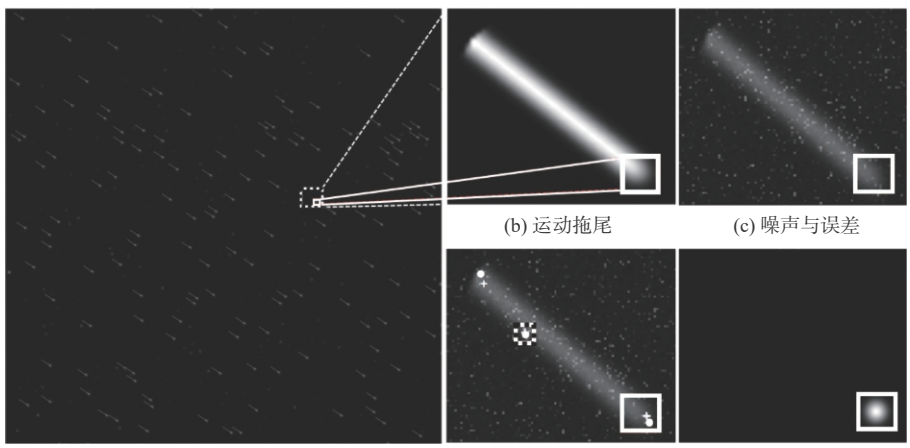


图 10　目标天体主焦面运动先验修正

Fig. 10　A priori correction for the movement of the main focal plane of the target celestial body

表 4 为条状目标端点定位结果。其中, 定位误差是指检测出的端点坐标与仿真图像端点坐标之差。原始定位误差是直接采用 Harris 方法获取的端点坐标的误差; 修正定位误差是通过导星先验信息修正并采用 Harris 方法获取的端点坐标的误差; 定位误差降低量指通过导星先验信息修正后, 端点检测误差的降低量。由此可得, 通过导星的先验信息修正, 可以降低定位误差, 提高对运动暗弱天体的检测。

表 4　条状目标端点定位结果

Table 4　Positioning results of strip target endpoints

目标	坐标	精确坐标	检测坐标	修正坐标	原始定位误差/像素	修正定位误差/像素	定位误差降低量/像素
目标1	*x*坐标	116.03	110.21	119.25	5.82	3.22	2.60
	*y*坐标	89.29	83.51	90.13	5.78	0.84	4.94
目标2	*x*坐标	330.57	340.68	338.51	10.11	7.94	2.17
	*y*坐标	230.41	240.72	237.16	10.31	6.75	3.56
目标3	*x*坐标	793.15	810.35	802.57	17.20	9.42	7.78
	*y*坐标	701.36	715.69	703.86	14.33	2.50	11.83

5　结　论

本文通过导星辅助的运动暗弱天体检测仿真系统, 展示了星空环境下运动暗弱天体检测的设计思路与实现过程, 根据仿真结果可以得到以下结论:

1) 本文设计的多导星辅助系统可以降低抖动误差对主焦面成像质量的影响, 提高空间望远镜的鲁棒性。

2）通过一维自适应中值滤波算法可以很好地消除太空环境所带来的噪声误差。

3）导星辅助方法可以很好地为主焦面暗弱目标检测提供先验信息，结合主焦面角点检测，可降低望远镜控制误差和抖动的影响，提高暗弱天体图像检测的有效性。

4）通过设计星图模拟、图像去噪、质心解算、运动先验信息反向传播和主焦面成像等模块，本文设计的系统可以解耦地精化暗弱目标先验运动信息，提高后续观测的目标图像质量。

本文设计的系统可以系统化处理真实运行场景下可能出现的抖动误差、噪声误差和运动先验误差，仿真算法及系统功能为后续望远镜观测方法的工程实现及在轨数据处理奠定了基础。

参考文献（References）

[1] 王传军. 丽江2.4米望远镜观测控制系统关键技术研究[D]. 北京: 中国科学院大学 (中国科学院云南天文台), 2019.
WANG C J. Research on key technologies of observation and control system of Lijiang 2.4 m telescope[D]. Beijing: University of Chinese Academy of Sciences (Yunnan Observatories, Chinese Academy of Sciences), 2019(in Chinese).

[2] 赵玥皎. 12 mag 精细导星星库构建及解算方法研究[D]. 上海: 中国科学院大学 (中国科学院上海技术物理研究所), 2017: 10-15.
ZHAO Y J. Study on the construction and calculation method of 12 mag fine guide star library[D]. Shanghai: University of Chinese Academy of Sciences (Shanghai Institute of Technical Physics, Chinese Academy of Sciences), 2017: 10-15(in Chinese).

[3] 崔祥祥, 王宏力, 陆敬辉, 等. 适用于小视场星敏感器的导航星表构建方法[J]. 红外与激光工程, 2015, 44(4): 1249-1253.
CUI X X, WANG H L, LU J H, et al. Guide star selection method for star tracker with thin field of view[J]. Infrared and Laser Engineering, 2015, 44(4): 1249-1253(in Chinese).

[4] ZACHARIAS N, FINCH C T, GIRARD T M, et al. The fourth USNO CCD astrograph catalog (UCAC4)[J]. The Astronomical Journal, 2013, 145(2): 44.

[5] ZACHARIAS N, URBAN S E, ZACHARIAS M I, et al. The second US Naval Observatory CCD astrograph catalog (UCAC2)[J]. The Astronomical Journal, 2004, 127(5): 3043-3059.

[6] BERRIMAN G B. The NASA/IPAC infrared science archive (IRSA) as a resource in supporting observatory operations[C]//Observatory Operations: Strategies, Processes, and Systems II. Bellingham: SPIE, 2008, 7016: 379-387.

[7] 朱长征, 居永忠, 杜晓辉. 导航星库制定方法研究[J]. 宇航学报, 2010, 31(5): 1327-1330.
ZHU C Z, JU Y Z, DU X H. Approach to founding a guide star catalogue[J]. Journal of Astronautics, 2010, 31(5): 1327-1330(in Chinese).

[8] 王松林, 蒋峥. 改进的自适应加权中值滤波算法[J]. 传感器与微系统, 2016, 35(11): 128-131.
WANG S L, JIANG Z. Improved adaptive weighted median filtering algorithm[J]. Transducer and Microsystem Technologies, 2016, 35(11): 128-131(in Chinese).

[9] 邹永宁, 姚功杰. 自适应窗口形状的中值滤波[J]. 光学 精密工程, 2018, 26(12): 3028-3039.
ZOU Y N, YAO G J. Median filtering algorithm for adaptive window shape[J]. Optics and Precision Engineering, 2018, 26(12): 3028-3039(in Chinese).

[10] 宿德志. 一种快速稳定的星图识别方法及单星模拟器的调校[D]. 长沙: 国防科学技术大学, 2010: 15-20.
SU D Z. A fast and stable star map recognition method and the adjustment of single star simulator[D]. Changsha: National University of Defense Technology, 2010: 15-20(in Chinese).

[11] 雍杨, 王敬儒, 张启衡. 基于多特征融合的弱小运动目标识别[J]. 量子电子学报, 2006, 23(5): 594-598.
YONG Y, WANG J R, ZHANG Q H. Small moving target recognition based on multi-feature fusion[J]. Chinese Journal of Quantum Electronics, 2006, 23(5): 594-598(in Chinese).

[12] 张兵. 光学图像末制导中的点目标检测与识别算法研究[D]. 长沙: 国防科学技术大学, 2005: 20-25.
ZHANG B. Research on point target detection and recognition algorithm in optical image terminal guidance[D]. Changsha: National University of Defense Technology, 2005: 20-25(in Chinese).

[13] 赵玥皎, 尹达一, 许春, 等. 16 mv 精细导星星库构建与评价[J]. 应用光学, 2017, 38(3): 385-391.
ZHAO Y J, YIN D Y, XU C, et al. Construction and evaluation of 16 mv fine guide star catalog[J]. Journal of Applied Optics, 2017, 38(3): 385-391(in Chinese).

[14] 刘南南, 徐抒岩, 曹小涛, 等. 自适应加权质心算法在高精度星点定位中的应用[J]. 光学学报, 2016, 36(8): 121-126.
LIU N N, XU S Y, CAO X T, et al. Application of adaptive weighted centroid algorithm in high accuracy star localization[J]. Acta Optica Sinica, 2016, 36(8): 121-126(in Chinese).

[15] 庞文. 天基可见光空间目标检测识别研究[D]. 长沙: 国防科学技术大学, 2014: 35-45.
PANG W. Research on detection and recognition of space-based visible light space targets[D]. Changsha: National University of Defense Technology, 2014: 35-45(in Chinese).

[16] 韩思奇, 王蕾. 图像分割的阈值法综述[J]. 系统工程与电子技术, 2002, 24(6): 91-94.
HAN S Q, WANG L. A survey of thresholding methods for image segmentation[J]. Systems Engineering and Electronics, 2002, 24(6): 91-94(in Chinese).

[17] 任剑峰, 饶长辉, 李明全. 一种 Hartmann-Shack 波前传感器图像的自适应阈值选取方法[J]. 光电工程, 2002, 29(1): 1-5.
REN J F, RAO C H, LI M Q. An adaptive threshold selection method for Hartmann-Shack wavefront sensor[J]. Opto-electronic Engineering, 2002, 29(1): 1-5(in Chinese).

[18] YANG J, LIANG B, ZHANG T, et al. A novel systematic error compensation algorithm based on least squares support vector regression for star sensor image centroid estimation[J]. Sensors, 2011, 11(8): 7341-7363.

[19] 贾辉. 高精度星敏感器星点提取与星图识别研究[D]. 长沙: 国防科技大学, 2011: 25-35.
JIA H. Research on star point extraction and star map identification of high-precision star sensors[D]. Changsha: National University of

Defense Technology, 2011: 25-35(in Chinese).

[20] 张己化, 范如玉, 赵宁, 等. 强背景下光电系统空间目标探测能力[J]. 红外与激光工程, 2014, 43(1): 212-216.
ZHANG J H, FAN R Y, ZHAO N, et al. Electro-optic system detection ability to space-object in strong background[J]. Infrared and Laser Engineering, 2014, 43(1): 212-216(in Chinese).

Design and simulation of guide star assisted moving faint object detection system

ZHANG Wei[1, 2], FENG Wenquan[1, *], SUN Guotong[2], WANG Chunlei[1]

(1. Electronic Information Engineering, Beihang University, Beijing 100191, China;
2. Beijing Institute of Spacecraft System Engineering, Beijing 100094, China)

Abstract: The detection of faint moving objects in space is widely used in aerospace and military fields. In order to improve the detection capability of the space object monitoring system, a guide star assisted detection system for moving faint objects is proposed and designed. The system consists of star map simulation, image denoising, centroid calculation, motion backpropagation and main focal plane imaging. Firstly, the star catalog and the starry sky environment are studied to ensure the accuracy of the star point position and noise simulation. Secondly, to realistically simulate the starry sky environment where the telescope is located, jitter errors, noise errors, and motion prior errors are introduced. At the same time, a guide star assisted method is designed to eliminate the above errors, and the system is adjusted through the motion back-transmission module. Experiments prove that the designed system can not only simulate the real star map, but also slow down the error influence of the star environment on the telescope.

Keywords: faint object; space object detection; USNO CCD astrograph catalog; error analysis; system simulation

Received: 2023-08-02; **Accepted**: 2024-01-26; **Published Online**: 2024-03-08 10:13
URL: link.cnki.net/urlid/11.2625.V.20240306.1005.002

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0449

面向单子集焦元冲突证据的融合方法

周诚[1,2]，徐达[1,*]，曹振地[3]

(1. 陆军装甲兵学院 兵器与控制系，北京 100072；2. 陆军炮兵防空兵学院 南京校区，南京 211132；
3. 解放军 75560 部队，屯昌 571600)

摘　　要：为有效融合多源数据以充分利用多源数据信息，针对只包含单子集焦元证据的多源数据融合时存在证据冲突的问题，提出了面向单子集焦元冲突证据的融合方法。综合考虑证据间的关联性和证据本身的不确定性2个方面，引入夹角余弦度量证据间的冲突程度来修正冲突系数和Jousselme距离的不足，利用冲突系数和信息熵的组合构造判断规则对证据进行分组，结合证据的支持度和可信度共同修正证据，进而实现单子集焦元冲突证据的有效融合。以4种典型冲突悖论为数值案例，验证了所提方法有效可行，且融合性能优于传统证据融合方法。

关 键 词：多源数据融合；单子集焦元；冲突证据；夹角余弦；信息熵
中图分类号：TP301.6
文献标志码：A　　**文章编号**：1001-5965(2025)09-3062-09

多源数据融合技术作为一种高新的数据处理技术，近年来发展快速且应用广泛[1]。考虑到多源数据具有不确定性，且不同源数据之间可能相互冲突，引入Dempster-Shafer (DS)证据理论[2]对多源数据进行融合是一种能充分利用多源数据的有效方法。但当多源数据的证据高度冲突时，DS证据理论的Dempster组合规则不能有效处理，易出现与事实相悖的现象[3]。

对于Dempster组合规则不能有效处理高冲突证据的原因，研究人员主要持2种观点：①认为组合规则中冲突的分配不合理；②认为组合规则没问题，缺陷由冲突证据本身造成。相应的，对DS证据理论的改进主要有2个思路：①修改组合规则[4-12]，其将冲突悖论的产生归因于组合规则不合理，通过修改组合规则，虽然在一定程度上解决了高冲突证据的融合问题，但破坏了组合规则的交换律、结合律等良好的数学性质，导致融合大量证据时的效果较差，具有一定的局限性；②修正证据模型[13-26]，其将冲突悖论的产生归因于冲突证据本身，通过确定证据的权重系数并利用权重系数对证据进行修正，减弱了证据间的冲突程度，可以灵活高效地处理高冲突证据。例如，Murphy[14]提出了证据平均加权法，简单赋予每个证据相同的权重后，使用Dempster组合规则对加权平均证据融合$n-1$次，融合结果表明，该方法具有较强的收敛性，但由于权重的确定没有考虑证据间的关联性，使融合结果的准确度较差。Deng等[15]在借鉴Murphy[14]思想的基础上，引入了Jousselme距离度量证据间的冲突程度，进而确定证据的加权系数，一定程度上弥补了平均加权法的不足。Chen等[17]提出了一种基于加权DS证据理论的显著性图融合方法，借助夹角余弦度量证据间的冲突程度，也能很好地解决高冲突证据的融合问题。李捷等[18]也借助夹角余弦表示证据间的相似性，调整置信度在目标属性指向的分配，提出了基于改进DS证据理论多周期数据融合的目标识别方法，在提高目标识别准确性的同时，能够有效解决目标密集和应答欺骗造成的目标误识别问题。Ye等[19]利用Lance距离修正证据的性能优于

收稿日期：2023-07-10；**录用日期**：2023-09-28；**网络出版时间**：2023-10-11 12:45
网络出版地址：link.cnki.net/urlid/11.2625.V.20231010.1005.002
*通信作者. E-mail：ljzjbxy2022@163.com

引用格式：*周诚，徐达，曹振地. 面向单子集焦元冲突证据的融合方法[J]. 北京航空航天大学学报，2025，51(9)：3062-3070.*
ZHOU C，XU D，CAO Z D. Conflict evidence fusion method for single-subset focal elements[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3062-3070 (in Chinese).

Minkowski 距离(包含 Manhattan 距离[20]、Euclidean 距离等)和 Mahalanobis 距离的特点[25], 改进了 DS 证据融合方法, 实现了多传感器冲突信息的可靠融合。雷震烁[21]和张欢[22]等针对传统 DS 证据理论无法有效解决高冲突证据融合的问题, 分别利用皮尔逊相关系数度量证据体的相关性, 仿真结果表明, 所提方法在处理高冲突证据时准确度更高、收敛速度更快。Yuan 等[23]提出了一种结合 Jousselme 距离和信息熵[24]确定证据加权系数的方法, 既考虑了证据间的关联性, 又考虑了证据本身的不确定性, 与只考虑单一方面的证据修正方法相比, 组合方法融合性能更好, 但 Jousselme 距离在度量证据冲突时也会出现结果与事实相悖的情况, 不适合度量单子集焦元证据间的冲突。应文健等[25]为解决测试性评估数据融合中存在的冲突问题, 引入了 Lance 距离、信息熵和专家打分法确定冲突数据的融合权重, 并通过实例表明所提方法能有效减少数据的冲突问题。陶洋等[26]为缓解高冲突数据对融合的影响, 提出了基于皮尔逊系数和信息熵的多传感器数据融合方法, 引入皮尔逊系数度量证据间的冲突程度, 采用信息熵度量证据的不确定性, 并结合数值算例证明了该方法的有效性和可行性。

综上所述, 通过修正证据模型的方式解决证据冲突的问题, 同时, 考虑证据间的关联性和证据本身的不确定性 2 个方面确定证据的权重系数修正证据的性能, 优于只考虑单一方面的证据修正方法。其中, 描述证据本身的不确定性常使用信息熵, 描述证据间的关联性常使用冲突系数k、夹角余弦、Lance 距离、Minkowski 距离和皮尔逊相关系数等, 不同的冲突证据度量方法面对识别框架中子集类型的不同, 其度量性能也不同。选取某一冲突证据度量方法和信息熵加权组合的方式是当前解决证据冲突的有效方法, 但现有文献更多的是一种普遍适用性方法, 虽能解决多源数据的冲突证据融合问题, 但其针对性解决某一类数据融合的性能仍有待进一步提高。本文基于多源维修性先验数据融合这一研究和应用背景, 针对这一类只包含单子集焦元证据的多源数据融合时存在证据冲突的问题, 在借鉴证据加权组合思想的基础上, 提出面向单子集焦元冲突证据的融合方法。利用冲突系数k和信息熵的组合构造了证据分组的判断规则, 依据单个证据的平均冲突系数和冲突系数矩阵的平均冲突系数及单个证据的信息熵和信息熵矩阵的平均信息熵的大小关系, 将证据划分为高可信度证据和一般可信度证据 2 类, 以突出高可信度证据的重要性, 分别确定不同组证据的可信度加权因子; 引入夹角余弦度量证据间的冲突程度, 较其他冲突证据度量方法能更合理地描述单子集焦元证据间的冲突, 进而借助余弦相似度确定证据的支持度加权因子; 结合可信度加权因子和支持度加权因子共同修正证据, 并使用 Dempster 组合规则组合证据; 最后, 结合案例验证所提方法的可行性和有效性。所提方法可用于解决复杂设备多源维修性数据的冲突证据融合的问题, 从而实现小子样条件下复杂设备维修性的有效验证, 提高试验效率, 节约试验经费。同时, 所提方法还适用于如复杂设备测试性验证等领域中多源数据的冲突证据是单子集焦元证据的融合过程。

1　理论基础

1.1　DS 证据理论的基本概念

定义 1　识别框架[2]。设$\Theta=\{A_1,A_2,\cdots,A_n\}$为关于命题$A$的集合, 若集合$\Theta$有限、完备且所包含的$n$个子集互不相容, 则将$\Theta$定义为识别框架。幂集是由$\Theta$中命题的所有可能组合的集合, 记为$2^\Theta$, $2^\Theta=\{\varnothing,\{A_1\},\cdots,\{A_n\},\{A_1,A_2\},\cdots,\{A_1,A_n\},\cdots,\Theta\}$。

定义 2　基本概率赋值(basic probability assignment, BPA)函数[2]。设A为幂集2^Θ中的任一命题, $A\in 2^\Theta$, 函数m为集合$2^\Theta\to[0,1]$的映射, 且满足:

$$\begin{cases} m(\varnothing)=0 \\ \sum\limits_{A\in 2^\Theta} m(A)=1 \\ m(A)\geqslant 0 \qquad \forall A\in 2^\Theta 且 A\neq\varnothing \end{cases} \tag{1}$$

则称m为Θ上的 BPA 函数或 mass 函数, $m(A)$为命题A的 BPA, 表示证据对命题A的支持度。如果$m(A)>0$, 则称命题A为焦元。使用向量$\boldsymbol{m}_i$表示第i个证据的 BPA, $\boldsymbol{m}_i=[\boldsymbol{m}_i(A_1),\boldsymbol{m}_i(A_2),\cdots,\boldsymbol{m}_i(A_q)](i=1,2,\cdots,n)$, $\alpha=\{A_1,A_2,\cdots,A_q\}$为证据$\boldsymbol{m}_i$中所有焦元的集合, $A_1,A_2,\cdots,A_q\subseteq 2^\Theta$, 每个命题$A_u(u=1,2,\cdots,q)$都有确定的$m(A_u)$。当焦元为单子集时, 称为单子集焦元, 当 mass 函数m中的所有焦元都是单子集时, 对应的证据称为单子集焦元证据。

定义 3　组合规则[2]。设m_1、m_2分别为识别框架Θ上 2 个证据的 mass 函数, 则 Dempster 组合规则定义为

$$m(A)=\begin{cases} 0 & A=\varnothing \\ \dfrac{1}{1-k}\sum\limits_{A_i\cap A_j=A} m_1(A_i)m_2(A_j) & A\neq\varnothing \end{cases} \tag{2}$$

式中: k为冲突系数, $k\in[0,1)$, 表示 2 个证据间的冲突程度。

1.2　证据冲突的度量方法

在度量证据间的冲突程度时, 典型证据冲突的

度量方法有时会出现与事实相悖的现象，结合例1对几种典型证据冲突的度量方法进行分析。

例1 设识别框架$\Theta=\{A_1,A_2,A_3,A_4,A_5,A_6\}$，有3组证据，每组证据的BPA如表1所示。

表1 不同组证据的BPA

Table 1 BPA of different sets of evidence

组号	BPA
1	$m_1(A_1)=m_1(A_2)=0.5$ $m_2(A_1)=m_2(A_2)=0.5$
2	$m_1(A_1)=m_1(A_2)=0.5$ $m_2(A_3)=m_2(A_4)=0.5$
3	$m_1(A_1)=m_1(A_2)=m_1(A_3)=1/3$ $m_2(A_4)=m_2(A_5)=m_2(A_6)=1/3$

1）冲突系数k。

$$k=\sum_{A_i\cap A_j=\varnothing}m_1(A_i)m_2(A_j) \tag{3}$$

对于第1组中2个没有冲突、完全一致的证据，使用冲突系数度量冲突程度的计算结果为$k=0.5$，表示2个证据之间存在冲突，结果与事实相悖，表明冲突系数k在度量证据冲突时存在不足。

2）Jousselme距离[27]。

$$d(\boldsymbol{m}_i,\boldsymbol{m}_j)=\sqrt{\frac{1}{2}(\boldsymbol{m}_i-\boldsymbol{m}_j)^{\mathrm{T}}\boldsymbol{D}(\boldsymbol{m}_i-\boldsymbol{m}_j)} \tag{4}$$

式中：$\boldsymbol{m}_i$和$\boldsymbol{m}_j$分别为Θ上2个证据的矢量形式；$\boldsymbol{D}$为$2^n\times2^n$矩阵，其元素$d_{ij}=|A_s\cap A_t|/|A_s\cup A_t|$，$\forall A_s,A_t\subseteq 2^{\Theta}$，$s,t=1,2,\cdots,2^n$。

式(4)可简化为

$$d(\boldsymbol{m}_i,\boldsymbol{m}_j)=\sqrt{\frac{1}{2}\left(\|\boldsymbol{m}_i\|^2+\left\|\boldsymbol{m}_j\right\|^2-2\left\langle\boldsymbol{m}_i,\boldsymbol{m}_j\right\rangle\right)} \tag{5}$$

式中：$\|\boldsymbol{m}_i\|^2=\boldsymbol{m}_i\boldsymbol{m}_i^{\mathrm{T}}$；$\left\|\boldsymbol{m}_j\right\|^2=\boldsymbol{m}_j\boldsymbol{m}_j^{\mathrm{T}}$；$\left\langle\boldsymbol{m}_i,\boldsymbol{m}_j\right\rangle=\sum_{s=1}^{2^n}\sum_{t=1}^{2^n}\boldsymbol{m}_i(A_s)\boldsymbol{m}_j(A_t)\frac{|A_s\cap A_t|}{|A_s\cup A_t|}$。

对于第2组和第3组中2对完全冲突的证据，其冲突程度理论上应均为1，但使用Jousselme距离的计算结果分别为0.707、0.577，结果与事实相悖，表明Jousselme距离在度量证据冲突时也存在不足。此外，由式(4)可知，矩阵$\boldsymbol{D}$的大小与Θ的基数呈指数增长关系，使得证据间Jousselme距离的求解面临计算量较大的问题。

3）余弦相似度。借助夹角余弦表示证据间的相似度：

$$\cos(\boldsymbol{m}_i,\boldsymbol{m}_j)=\frac{\boldsymbol{m}_i\boldsymbol{m}_j^{\mathrm{T}}}{\|\boldsymbol{m}_i\|\cdot\|\boldsymbol{m}_j\|} \tag{6}$$

则证据间的冲突程度可用$1-\cos(\boldsymbol{m}_i,\boldsymbol{m}_j)$度量，对于第1～3组证据，使用余弦定理的计算结果分别为1、0、0，对应的冲突程度度量结果分别为0、1、1，与事实一致。由此可知，对于单子集焦元证据冲突的度量，夹角余弦优于冲突系数k和Jousselme距离。

1.3 信息熵

信息熵[28]是描述一个具体事件所有可能发生的结果所包含信息量的数学期望，Shannon[29]将其引入到证据不确定程度的度量中，后经Deng[30-31]进一步推广。

已知$\boldsymbol{m}_i=[\boldsymbol{m}_i(A_1),\boldsymbol{m}_i(A_2),\cdots,\boldsymbol{m}_i(A_q)]$为第$i$个证据$E_i$的mass函数值，$i=1,2,\cdots,n$，$A_u(u=1,2,\cdots,q)$为识别框架$\Theta$下的第$u$个命题，则证据$E_i$的Deng熵为

$$H_{\mathrm{d}}(\boldsymbol{m}_i)=-\sum_{u=1}^{q}\boldsymbol{m}_i(A_u)\log_2\frac{\boldsymbol{m}_i(A_u)}{2^{|A_u|}-1} \tag{7}$$

式中：$|A_u|$为命题A_u的基数，即包含的元素个数。$|A_u|$的值越大，$H_{\mathrm{d}}(\boldsymbol{m}_i)$的值越大，表示证据的不确定度越大，包含的信息量越多，可信度越低，在Dempster组合中的权重就越小。

当A_u只包含单元素时，即$|A_u|=1$，Deng熵便退化为Shanoon熵，即

$$H_{\mathrm{d}}(\boldsymbol{m}_i)=-\sum_{u=1}^{q}\boldsymbol{m}_i(A_u)\log_2\boldsymbol{m}_i(A_u) \tag{8}$$

为避免由于$H_{\mathrm{d}}(\boldsymbol{m}_i)=0$的出现而使Dempster组合中证据$E_i$的权重为0，对$H_{\mathrm{d}}(\boldsymbol{m}_i)$取指数：

$$\mathrm{e}^{H_{\mathrm{d}}(\boldsymbol{m}_i)}=\mathrm{e}^{-\sum_{u=1}^{q}\boldsymbol{m}_i(A_u)\log_2\frac{\boldsymbol{m}_i(A_u)}{2^{|A_u|}-1}} \tag{9}$$

2 融合方法

为有效融合多源数据的单子集焦元冲突证据，综合考虑证据间的关联性和证据本身的不确定性2个方面因素修正冲突证据，提出一种面向单子集焦元冲突证据的融合方法。结合冲突系数k和信息熵将证据分为高可信度证据和一般可信度证据，分别确定不同组证据的可信度加权因子。引入夹角余弦度量证据间的冲突程度，借助余弦相似度确定证据的支持度加权因子。结合可信度加权因子和支持度加权因子共同修正证据得到加权平均证据，并利用Dempster组合规则融合$n-1$次，得到多源数据的加权系数，进而实现多源数据的有效融合。单子集焦元冲突证据融合方法流程如图1所示。

2.1 可信度加权因子

步骤1 设有n条证据，构成证据矩阵$\boldsymbol{M}$，$\boldsymbol{M}=[\boldsymbol{m}_1,\boldsymbol{m}_2,\cdots,\boldsymbol{m}_n]$。利用式(3)计算证据$\boldsymbol{m}_i$和$\boldsymbol{m}_j$之间的冲突系数$k_{ij}$，$i,j=1,2,\cdots,n$，建立冲突系数矩阵$\boldsymbol{K}$：

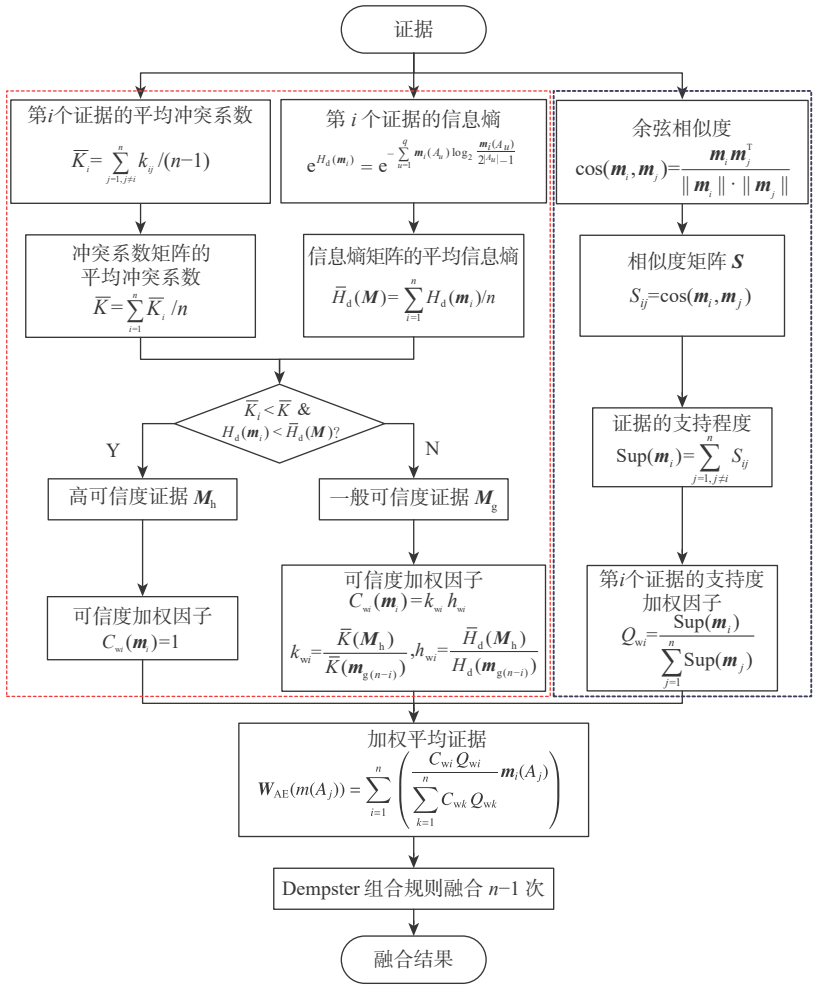


图 1　单子集焦元冲突证据融合方法流程

Fig. 1　Flow chart of conflict evidence fusion method for single-subset focal elements

$$\boldsymbol{K}=\begin{bmatrix}k_{11}&k_{12}&\cdots&k_{1n}\\k_{21}&k_{22}&\cdots&k_{2n}\\\vdots&\vdots&&\vdots\\k_{n1}&k_{n2}&\cdots&k_{nn}\end{bmatrix}\tag{10}$$

步骤 2　计算证据$\boldsymbol{m}_i$的平均冲突系数$\overline{K}_i$:

$$\overline{K}_i=\sum_{j=1,j\neq i}^{n}k_{ij}/(n-1)\tag{11}$$

则冲突系数矩阵$\boldsymbol{K}$的平均冲突系数$\overline{K}$为

$$\overline{K}=\sum_{i=1}^{n}\overline{K}_i/n\tag{12}$$

步骤 3　根据式(8)计算证据$\boldsymbol{m}_i$的信息熵$H_d(\boldsymbol{m}_i)$, 并建立信息熵矩阵$\boldsymbol{H}_d(\boldsymbol{M})$:

$$\boldsymbol{H}_d(\boldsymbol{M})=[H_d(\boldsymbol{m}_1),H_d(\boldsymbol{m}_2),\cdots,H_d(\boldsymbol{m}_n)]\tag{13}$$

则信息熵矩阵$\boldsymbol{H}_d(\boldsymbol{M})$的平均信息熵$\overline{H}_d(\boldsymbol{M})$为

$$\overline{H}_d(\boldsymbol{M})=\sum_{i=1}^{n}H_d(\boldsymbol{m}_i)/n\tag{14}$$

步骤 4　根据判断规则, 将证据分为 2 组: ①高可信度证据矩阵, 记为$\boldsymbol{M}_h$, $\boldsymbol{M}_h=[\boldsymbol{m}_{h1},\boldsymbol{m}_{h2},\cdots,\boldsymbol{m}_{hi}]$ $(i\leqslant n)$, 满足条件$\overline{K}_i<\overline{K}$, 且$H_d(\boldsymbol{m}_i)<\overline{H}_d(\boldsymbol{M})$; ②一般可信度证据矩阵, 记为$\boldsymbol{M}_g$, $\boldsymbol{M}_g=[\boldsymbol{m}_{g1},\boldsymbol{m}_{g2},\cdots,\boldsymbol{m}_{g(n-i)}]$, 不需要满足上述条件。

步骤 5　确定证据的可信度加权因子。高可信度矩阵$\boldsymbol{M}_h$中的证据不需要修正, 其可信度加权因子默认为 1, 即$C_{wi}(\boldsymbol{m}_i)=1$。对于一般可信度矩阵$\boldsymbol{M}_g$中的证据, 采用冲突转换系数$k_{wi}$和信息熵转换系数$h_{wi}$相结合的方式确定其可信度加权因子:

$$C_{wi}(\boldsymbol{m}_i)=k_{wi}h_{wi}\tag{15}$$

式中: $k_{wi}=\dfrac{\overline{K}(\boldsymbol{M}_h)}{\overline{K}(\boldsymbol{m}_{g(n-i)})}$; $h_{wi}=\dfrac{\overline{H}_d(\boldsymbol{M}_h)}{H_d(\boldsymbol{m}_{g(n-i)})}$; $\overline{K}(\boldsymbol{M}_h)$为高可信度证据矩阵$\boldsymbol{M}_h$的平均冲突系数; $\overline{K}(\boldsymbol{m}_{g(n-i)})$为一般可信度证据$\boldsymbol{m}_{g(n-i)}$的平均冲突系数; $\overline{H}_d(\boldsymbol{M}_h)$为高可信度证据矩阵$\boldsymbol{M}_h$的平均信息熵; $H_d(\boldsymbol{m}_{g(n-i)})$为一般可信度证据$\boldsymbol{m}_{g(n-i)}$的信息熵。

为准确描述证据的综合可信度, 综合考虑证据间的冲突属性和证据本身的信息熵属性, 将证据分为高可信度证据和一般可信度证据 2 类。可信度加权因子即为综合上述 2 个属性的证据综合可信

度的比例系数。对于高可信度证据，用常数1表示其冲突属性和信息熵属性的可信度，$C_{wi}(\boldsymbol{m}_i)=1\times1=1$。对于一般可信度证据，以高可信度证据的属性值为基准，将一般可信度证据转化为对应高可信度证据下的比例值，记为转移系数，则$C_{wi}(\boldsymbol{m}_i)$等于冲突转换系数k_{wi}和信息熵转换系数h_{wi}的乘积。

2.2 支持度加权因子

步骤1 利用式(6)计算证据$\boldsymbol{m}_i(i=1,2,\cdots,n)$和$\boldsymbol{m}_j(j=1,2,\cdots,n)$之间的余弦相似度$S_{ij}$，并建立相似度矩阵$\boldsymbol{S}$：

$$\boldsymbol{S}=\begin{bmatrix}1 & S_{12} & \cdots & S_{1n}\\ S_{21} & 1 & \cdots & S_{2n}\\ \vdots & \vdots & & \vdots\\ S_{n1} & S_{n2} & \cdots & 1\end{bmatrix} \tag{16}$$

式中：$\boldsymbol{S}$为对称矩阵，$S_{ij}=S_{ji}$，$0\leqslant S_{ij}\leqslant1$，$S_{ij}$越大，表示证据$\boldsymbol{m}_i$和$\boldsymbol{m}_j$相互支持的程度越高。

步骤2 计算证据$\boldsymbol{m}_i$的支持度。

$$\mathrm{Sup}(\boldsymbol{m}_i)=\sum_{j=1,j\neq i}^{n}S_{ij}\qquad i,j=1,2,\cdots,n \tag{17}$$

则证据$\boldsymbol{m}_i$的标准化支持度加权因子为

$$Q_{wi}=\frac{\mathrm{Sup}(\boldsymbol{m}_i)}{\sum_{j=1}^{n}\mathrm{Sup}(\boldsymbol{m}_j)}\qquad i,j=1,2,\cdots,n \tag{18}$$

2.3 证据的加权融合

步骤1 利用可信度加权因子C_{wi}和支持度加权因子Q_{wi}加权融合证据得到加权平均证据$\boldsymbol{W}_{\mathrm{AE}}(m(A_j))$：

$$\boldsymbol{W}_{\mathrm{AE}}(m(A_j))=\sum_{i=1}^{n}\left(\frac{C_{wi}Q_{wi}}{\sum_{k=1}^{n}C_{wk}Q_{wk}}\boldsymbol{m}_i(A_j)\right) \tag{19}$$

步骤2 通过Dempster组合规则对$\boldsymbol{W}_{\mathrm{AE}}(m(A_j))$进行$n-1$次融合，得到多源数据的综合融合权重，进而实现多源数据的有效融合。

3 典型数值案例分析

为验证所提改进DS（improved Dempster-Shafer, IDS）证据融合方法的可行性，以4种典型高冲突证据悖论为数值案例说明DS证据理论在融合高冲突证据时的不足，并结合多种改进方法进行证据融合结果的对比分析。由于本文引入了信息熵，当某一命题的BPA为0时，取对数计算的结果无意义，为避免这种情况，使用1×10^{-12}这种很小的数替代0，证据融合的结果不变[32]。

3.1 完全冲突悖论

设在识别框架$\Theta=\{A,B,C\}$上有3个证据，每个证据的BPA分别为

$\boldsymbol{m}_1$：$m_1(A)=1,m_1(B)=0,m_1(C)=0$
$\boldsymbol{m}_2$：$m_2(A)=0,m_2(B)=1,m_2(C)=0$
$\boldsymbol{m}_3$：$m_3(A)=0.9,m_3(B)=0.1,m_3(C)=0$

可知，证据$\boldsymbol{m}_1$和$\boldsymbol{m}_3$支持命题A，$\boldsymbol{m}_2$完全支持命题B，直观上3个证据的融合结果应为支持命题A，然而冲突系数$k=1$，此时，Dempster组合规则公式中分母为0，方法失效，不能使用DS证据理论，结果与事实完全相反。表2和图2(a)为不同方法在

表2 4种典型高冲突证据融合结果

Table 2 Fusion results of four kinds of typical high conflict evidence

悖论类型	融合方法	基本概率赋值				目标命题
		命题A	命题B	命题C	识别框架Θ	
完全冲突悖论	Yager[4]	0	0	0	1	Θ
	Murphy[14]	0.837 5	0.162 5	0	0	A
	Deng[15]	0.998 6	0.001 4	0	0	A
	Chen[17]	0.998 6	0.001 4	0	0	A
	Ye[19]	0.925 0	0.075 0	0	0	A
	Yuan[23]	0.998 6	0.001 4	0	0	A
	IDS	0.999 9	0.000 1	0	0	A
0信任悖论	DS[2]	0	0.631 6	0.368 4	0	B
	Yager[4]	0	0.018 0	0.010 5	0.989 5	Θ
	Murphy[14]	0.350 0	0.522 4	0.127 6	0	B
	Deng[15]	0.581 6	0.243 9	0.174 5	0	A
	Chen[17]	0.619 5	0.201 0	0.179 5	0	A
	Ye[19]	0.622 0	0.197 4	0.180 5	0	A
	Yuan[23]	0.731 9	0.079 0	0.189 1	0	A
	IDS	0.731 9	0.079 0	0.189 1	0	A
1信任悖论	DS[2]	0	1	0	0	B
	Yager[4]	0	0.01	0	1	Θ
	Murphy[14]	0.488 0	0.024 1	0.488 0	0	A/C
	Deng[15]	0.488 0	0.024 1	0.488 0	0	A/C
	Chen[17]	0.488 0	0.024 1	0.488 0	0	A/C
	Ye[19]	0.488 0	0.024 1	0.488 0	0	A/C
	Yuan[23]	0.488 0	0.024 1	0.488 0	0	A/C
	IDS	0.488 0	0.024 1	0.488 0	0	A/C
高冲突悖论	DS[2]	0	0.915 3	0.084 7	0	B
	Yager[4]	0	0.000 2	0	1	Θ
	Murphy[14]	0.838 9	0.150 2	0.009 9	0.001 0	A
	Deng[15]	0.952 4	0.038 9	0.007 8	0.000 9	A
	Chen[17]	0.962 3	0.028 9	0.007 8	0.001 0	A
	Ye[19]	0.950 9	0.040 5	0.007 7	0.000 9	A
	Yuan[23]	0.961 0	0.026 8	0.010 5	0.001 7	A
	IDS	0.995 9	0.002 3	0.001 7	0.000 1	A

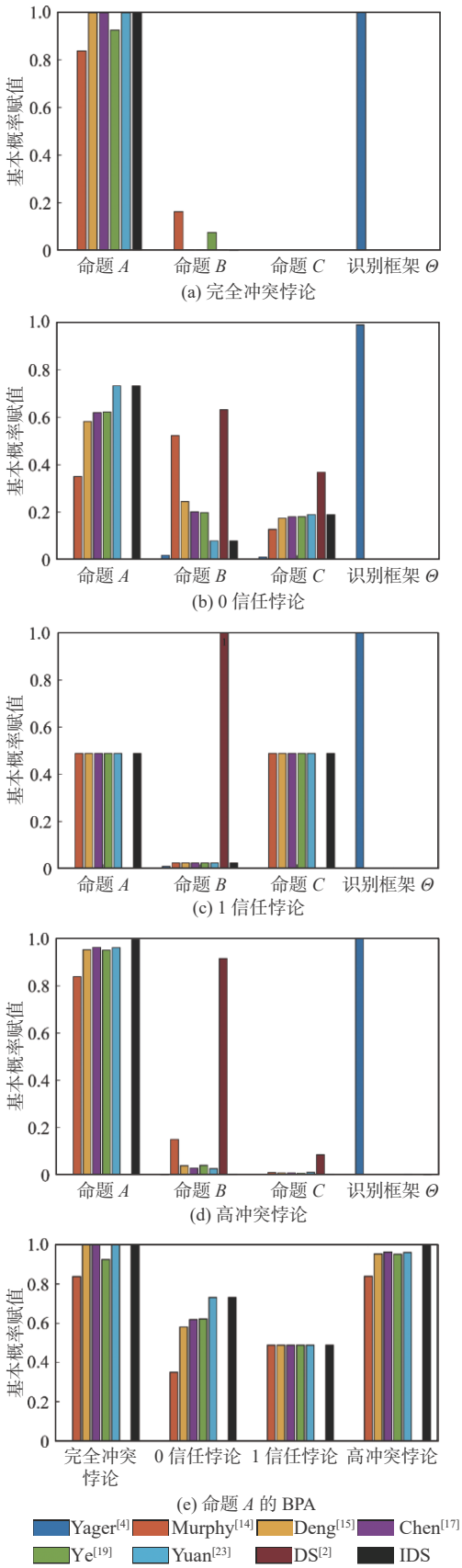


图 2　证据融合结果

Fig. 2　Evidence fusion results

完全冲突悖论下的融合结果。

3.2　0 信任悖论

设在识别框架 $\Theta=\{A,B,C\}$上有 3 个证据, 每个证据的 BPA 分别为

$$\boldsymbol{m}_1:\ m_1(A)=0.5, m_1(B)=0.2, m_1(C)=0.3$$
$$\boldsymbol{m}_2:\ m_2(A)=0, m_2(B)=0.9, m_2(C)=0.1$$
$$\boldsymbol{m}_3:\ m_3(A)=0.55, m_3(B)=0.1, m_3(C)=0.35$$

利用 DS 证据理论对该组证据进行融合, 结果为 $m(A)=0$, $m(B)=0.631\,6$, $m(C)=0.368\,4$, $k=0.973\,1$。其中, 证据 $\boldsymbol{m}_2$ 中命题 A 的 BPA 为 0, 此时, 不管证据 $\boldsymbol{m}_1$ 和 $\boldsymbol{m}_3$ 中命题的 BPA 如何分配、或后续又增加了多少个支持命题 A 的证据, 最终的融合结果 $m(A)$ 均为 0, 即 m_2 对该组证据具有一票否决权, 融合结果与事实相悖, 体现了 DS 证据理论融合此类证据的局限性。表 2 和图 2(b) 为 0 信任悖论下不同方法的融合结果。

3.3　1 信任悖论

设在识别框架 $\Theta=\{A,B,C\}$上有 2 个证据, 每个证据的 BPA 分别为

$$\boldsymbol{m}_1:\ m_1(A)=0.9, m_1(B)=0.1, m_1(C)=0$$
$$\boldsymbol{m}_2:\ m_2(A)=0, m_2(B)=0.1, m_2(C)=0.9$$

可知, 证据 $\boldsymbol{m}_1$ 支持命题 A, $\boldsymbol{m}_2$ 支持命题 C, 2 个证据均以小概率支持命题 B, 但利用 DS 证据理论融合的结果为 $m(A)=0$, $m(B)=1$, $m(C)=0$, $k=0.99$, 完全支持命题 B, 即小概率命题 B 通过 DS 证据理论融合后成为了必然命题, 融合结果与事实相悖, 类似于此类冲突证据产生的悖论均属于 1 信任悖论。表 2 和图 2(c) 为不同方法在 1 信任悖论下的融合结果。

3.4　高冲突悖论

设在识别框架 $\Theta=\{A,B,C\}$上有 5 个证据, 每个证据的 BPA 分别为

$$\boldsymbol{m}_1:\ m_1(A)=0.3, m_1(B)=0.2, m_1(C)=0.1, m_1(\Theta)=0.4$$
$$\boldsymbol{m}_2:\ m_2(A)=0, m_2(B)=0.9, m_2(C)=0.1, m_2(\Theta)=0$$
$$\boldsymbol{m}_3:\ m_3(A)=0.6, m_3(B)=0.1, m_3(C)=0.1, m_3(\Theta)=0.2$$
$$\boldsymbol{m}_4:\ m_4(A)=0.7, m_4(B)=0.1, m_4(C)=0.1, m_4(\Theta)=0.1$$
$$\boldsymbol{m}_5:\ m_5(A)=0.7, m_5(B)=0.1, m_5(C)=0.1, m_5(\Theta)=0.1$$

利用 DS 证据理论对该组证据进行融合, 结果为 $m(A)=0$, $m(B)=0.915\,3$, $m(C)=0.084\,7$, $m(\Theta)=0$, $k=0.8$, 结果以较大概率支持命题 B, 由于证据 $\boldsymbol{m}_2$ 中 $m_2(A)=0$, 使融合结果完全不支持命题 A, 与事实相悖。表 2 和图 2(d) 为不同方法在高冲突悖论下的融合结果。

3.5　讨论分析

在完全冲突悖论中, 已知冲突系数 $k=1$, Dempster 组合规则失效, DS 证据理论不可用。由表 2 和图 2(a) 可知, Yager[4] 赋予识别框架 Θ 的 BPA 为 $m(\Theta)=1$, 反而增加了命题的不确定性, 无法正确识别目标命题 A。同时, 其余方法均能正确识别目标命题 A, 其中, Murphy[14] 和 Ye[19] 的识别精度略低,

命题A的BPA分别为$m(A)=0.837\,5$和$m(A)=0.925\,0$，IDS的识别精度最高，$m(A)=0.999\,9$。

在0信任悖论中，已知证据$\boldsymbol{m}_1$和$\boldsymbol{m}_3$为相对一致的证据，均支持命题A，证据$\boldsymbol{m}_2$支持命题B，而准确的融合结果应与这些相对一致的证据保持一致，远离冲突的证据。由表2和图2(b)可知，DS证据理论[2]和Murphy[14]分别以$m(B)=0.631\,6$和$m(B)=0.522\,4$支持命题B，但此时并不能因此确定命题B就是正确的目标。Yager[4]赋予了识别框架Θ的BPA为$m(\Theta)=0.989\,5$，表示命题是高度不确定的，无法正确识别目标命题。其余方法的融合结果均与相对一致的证据保持一致，认为命题A为目标命题，且Yuan[23]和IDS的识别精度最高。

在1信任悖论中，由表2和图2(c)可知，DS证据理论[2]完全支持命题B，$m(B)=1$，融合结果显然与事实相悖，错误选择了目标命题B。Yager[4]赋予识别框架Θ的BPA为$m(\Theta)=1$，也无法正确识别目标命题。其余方法均以相同概率$m(A)=0.488\,0$和$m(C)=0.488\,0$支持命题A和C，此结果是由本组证据的BPA设置的特殊性造成的，若后续继续增加证据，这些方法就可以准确识别目标命题。

在高冲突悖论中，已知证据$\boldsymbol{m}_1$、$\boldsymbol{m}_3$、$\boldsymbol{m}_4$和$\boldsymbol{m}_5$为相对一致的证据，均支持命题A，证据$\boldsymbol{m}_2$支持命题B，由表2和图2(d)可知，DS证据理论[2]以较大概率支持命题B，$m(B)=0.915\,3$，不能正确识别目标命题A。Yager[4]赋予识别框架Θ的BPA为$m(\Theta)=1$，也无法正确识别目标命题。其余方法均能正确识别目标命题A，其中，Murphy[14]的识别精度略低，$m(A)=0.838\,9$，IDS的识别精度最高，$m(A)=0.995\,9$。

综合上述分析可知，DS证据理论[2]和Yager[4]均处理不了4种典型冲突悖论，其中，DS证据理论[2]在完全冲突悖论中失效，在其余3种冲突悖论中均错误选择了命题B，没有正确识别目标命题A。Yager[4]在4种冲突悖论中，基本均将证据冲突的BPA赋予了识别框架Θ，增加了命题的不确定性，也无法正确识别目标命题。Murphy[14]除了在0信任悖论中错误选择了命题B外，在其余3种冲突悖论中均正确识别了目标命题A。Deng[15]、Chen[17]、Ye[19]、Yuan[23]和IDS在4种冲突悖论中均能正确识别目标命题A。进一步结合图2(e)可知，由于本文在1冲突悖论案例中证据的BPA设置的特殊性，使Murphy[14]、Deng[15]、Chen[17]、Ye[19]、Yuan[23]和IDS这6种修正证据方法的融合结果相同，在其余3种冲突悖论中，Murphy[14]由于只采用相同权重加权的方式修正证据，既没有考虑证据间的相关性，也没有考虑证据本身的不确定性，其融合精度最低，IDS方法的融合精度最高，验证了所提方法的可行性和有效性。

4 结 论

1）引入夹角余弦度量单子集焦元证据间的冲突程度，与冲突系数k和Jousselme距离这2种典型证据冲突度量方法相比，度量结果更加准确，可以有效避免度量结果与事实相悖现象的出现。

2）引入信息熵很好地从考虑证据本身的方面度量了证据的不确定程度，且综合冲突系数k和信息熵构造的证据可信度分组模型进一步实现了高可信度证据和一般可信度证据的有效分组，由此确定的证据可信度加权因子有助于修正证据，减弱证据间的冲突程度。

3）综合考虑证据间的关联性和证据本身的不确定性2个方面，提出了面向单子集焦元冲突证据的融合方法。以4种典型高冲突证据悖论为数值案例，与DS等7种证据修正方法的融合结果进行对比，发现IDS方法的融合精度最高，验证了所提方法的可行性和有效性。

参考文献（References）

[1] PARAI M, SRIMANI S, GHOSH K, et al. Multi-source data fusion technique for parametric fault diagnosis in analog circuits[J]. Integration, 2022, 84: 92-101.

[2] DEMPSTER A P. Upper and lower probabilities induced by a multivalued mapping[J]. The Annals of Mathematical Statistics, 1967, 38(2): 325-339.

[3] ZHAO Y X, JIA R F, SHI P. A novel combination method for conflicting evidence based on inconsistent measurements[J]. Information Sciences, 2016, 367: 125-142.

[4] YAGER R R. On the Dempster-Shafer framework and new combination rules[J]. Information Sciences, 1987, 41(2): 93-137.

[5] SMETS P. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem[J]. International Journal of Approximate Reasoning, 1993, 9(1): 1-35.

[6] 孙全, 叶秀清, 顾伟康. 一种新的基于证据理论的合成公式[J]. 电子学报, 2000, 28(8): 117-119.
SUN Q, YE X Q, GU W K. A new combination rules of evidence theory[J]. Acta Electronica Sinica, 2000, 28(8): 117-119(in Chinese).

[7] SMETS P. Belief functions on real numbers[J]. International Journal of Approximate Reasoning, 2005, 40(3): 181-223.

[8] 邓勇, 施文康, 朱振福. 一种有效处理冲突证据的组合方法[J]. 红外与毫米波学报, 2004, 23(1): 27-32.
DENG Y, SHI W K, ZHU Z F. Efficient combination approach of conflict evidence[J]. Journal Infrared Millimeter and Waves, 2004, 23(1): 27-32(in Chinese).

[9] LIN Z, XIE J Y. Research on improved evidence theory based on multi-sensor information fusion[J]. Scientific Reports, 2021, 11(1): 9267.

[10] 李弼程, 王波, 魏俊, 等. 一种有效的证据理论合成公式[J]. 数据

采集与处理, 2002, 17(1): 33-36.
LI B C, WANG B, WEI J, et al. An effective formula for synthesizing evidence theory[J]. Journal of Data Acquisition and Processing, 2002, 17(1): 33-36(in Chinese).

[11] 韩德强, 韩崇昭, 邓勇, 等. 基于证据方差的加权证据组合[J]. 电子学报, 2011, 39(S1): 153-157.
HAN D Q, HAN C Z, DENG Y, et al. Weighted evidence combination based on evidence variance[J]. Acta Electronica Sinica, 2011, 39(S1): 153-157(in Chinese).

[12] LI Y B, CHEN J, YE F, et al. The improvement of DS evidence theory and its application in IR/MMW target recognition[J]. Journal of Sensors, 2016, 2016: 1903792.

[13] HAEIVIVI R. Are alternatives to Dempster's rule of combination real alternative?: comments on "about the belief function combination and the conflict management problem" [J]. Information Fusion, 2002, 3(3): 237-239.

[14] MURPHY C K. Combining belief functions when evidence conflicts[J]. Decision Support Systems, 2000, 29(1): 1-9.

[15] DENG Y, SHI W K, ZHU Z F, et al. Combining belief functions based on distance of evidence[J]. Decision Support Systems, 2004, 38(3): 489-493.

[16] YANG Y, HAN D Q, HAN C Z. Discounted combination of unreliable evidence using degree of disagreement[J]. International Journal of Approximate Reasoning, 2013, 54(8): 1197-1216.

[17] CHEN B C, TAO X, YANG M R, et al. A saliency map fusion method based on weighted DS evidence theory[J]. IEEE Access, 2018, 6: 27346-27355.

[18] 李捷, 杨雪洲, 周亮. 基于改进 DS 理论多周期数据融合的目标识别方法[J]. 火力与指挥控制, 2019, 44(7): 43-48.
LI J, YANG X Z, ZHOU L. Research on target identification based on improved DS evidence of multi-period fusion method[J]. Fire Control & Command Control, 2019, 44(7): 43-48(in Chinese).

[19] YE F, CHEN J, LI Y B. Improvement of DS evidence theory for multi-sensor conflicting information[J]. Symmetry, 2017, 9(5): 69.

[20] LI S B, LIU G K, TANG X H, et al. An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis[J]. Sensors, 2017, 17(8): 1729.

[21] 雷震烁, 刘松涛, 葛杨, 等. 基于平均证据和焦元距离的高冲突证据融合方法[J]. 电光与控制, 2021, 28(4): 6-10.
LEI Z S, LIU S T, GE Y, et al. High conflict evidence fusion method based on average evidence and focal distance[J]. Electronics Optics & Control, 2021, 28(4): 6-10(in Chinese).

[22] 张欢, 陆见光, 唐向红. 面向冲突证据的改进 DS 证据理论算法[J]. 北京航空航天大学学报, 2020, 46(3): 616-623.
ZHANG H, LU J G, TANG X H. Improved DS evidence theory algorithm for conflict evidence[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(3): 616-623(in Chinese).

[23] YUAN K J, XIAO F Y, FEI L G, et al. Conflict management based on belief function entropy in sensor fusion[J]. SpringerPlus, 2016, 5(1): 638.

[24] XIAO F Y. Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy[J]. Information Fusion, 2019, 46: 23-32.

[25] 应文健, 程雨森, 王旋, 等. 基于研制阶段数据融合的舰炮制导弹药测试性评估方法[J]. 系统工程与电子技术, 2024, 46(8): 2730-2737.
YING W J, CHENG Y S, WANG X, et al. Testability evaluation method of naval gun guided ammunition based on data fusion in development stage[J]. Systems Engineering and Electronics, 2024, 46(8): 2730-2737(in Chinese).

[26] 陶洋, 祝小钧, 杨柳. 基于皮尔逊相关系数和信息熵的多传感器数据融合[J]. 小型微型计算机系统, 2023, 44(5): 1075-1080.
TAO Y, ZHU X J, YANG L. Multi-sensor data fusion based on Pearson coefficient and information entropy[J]. Journal of Chinese Computer Systems, 2023, 44(5): 1075-1080(in Chinese).

[27] JOUSSELME A L, GRENIER D, BOSSÉ É. A new distance between two bodies of evidence[J]. Information Fusion, 2001, 2(2): 91-101.

[28] KHAN M N, ANWAR S. Paradox elimination in Dempster-Shafer combination rule with novel entropy function: application in decision-level multi-sensor fusion[J]. Sensors, 2019, 19(21): 4810.

[29] SHANNON C E. A mathematical theory of communication[J]. ACM SIGMOBILE Mobile Computing and Communications Review, 2001, 5(1): 3-55.

[30] DENG Y. Deng entropy[J]. Chaos, Solitons & Fractals, 2016, 91(3): 549-553.

[31] DENG Y. Deng entropy: a generalized Shannon entropy to measure uncertainty[EB/OL]. (2015-02-24)[2023-07-07]. http://vixra.org/abs/1502.0222.

[32] GUO X Z, XIN X L. Partial entropy and relative entropy of fuzzy sets[J]. Fuzzy System & Mathematics, 2005, 19(2): 97-102.

Conflict evidence fusion method for single-subset focal elements

ZHOU Cheng[1,2], XU Da[1,*], CAO Zhendi[3]

(1. Weapons and Control Department, Army Academy of Armored Forces, Beijing 100072, China;
2. Nanjing Campus, Army Academy of Artillery and Air Defence, Nanjing 211132, China;
3. Unit 75560 of PLA, Tunchang 571600, China)

Abstract: To effectively fuse multi-source data and fully utilize multi-source data information, a conflict evidence fusion method for single-subset focal elements was proposed in response to evidence conflicts in fusion of multi-source data that only contains single-subset focal element evidence. Comprehensively considering the correlation between evidence and the uncertainty of the evidence itself, this study introduced the cosine of the included angle to measure the degree of conflict between evidence for correcting the deficiency of conflict coefficient and Jousselme distance. The combination of conflict coefficient and information entropy was used to construct judgment rules for evidence grouping, and then the evidence was modified with its support and credibility, so as to achieve effective fusion of single-subset focal element conflict evidence. With four typical conflict paradoxes as numerical cases, the proposed method is verified to be effective and feasible, and its fusion performance is better than that of the traditional evidence fusion method.

Keywords: multi-source data fusion; single-subset focal element; conflict evidence; cosine of included angle; information entropy

Received: 2023-07-10; **Accepted**: 2023-09-28; **Published Online**: 2023-10-11 12:45
URL: link.cnki.net/urlid/11.2625.V.20231010.1005.002
* **Corresponding author**. E-mail: ljzjbxy2022@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0458

尾座式垂直起降无人机过渡轨迹优化方法

邹旭[1]，刘贞报[1,2]，赵闻[1,*]，王莉娜[1]

(1. 西北工业大学 民航学院，西安 710072；2. 西北工业大学 深圳研究院，深圳 518057)

摘 要：针对目前基于传统最优化方法得到的过渡轨迹在尾座式垂直起降无人机实际飞行过程中可行性低和鲁棒性差的问题，提出一种基于过渡走廊的过渡轨迹优化方法。以一种双发尾座式垂直起降无人机为研究对象，通过分析机翼不同区域之间的迎角差异，构建非线性动力学模型。基于倾转旋翼飞行器过渡走廊研究思路，设计一种针对尾座式垂直起降无人机的过渡走廊，并通过限制爬升速率和俯仰角速率来提高过渡走廊的可行性。通过分析模型误差对过渡走廊的影响，得到一条具有最大安全裕度的目标过渡轨迹。将过渡过程视为轨迹优化问题，求解得到最接近目标过渡轨迹且保留足够作动器裕度的最优过渡轨迹。仿真和实际飞行结果表明，所提方法能够引导飞机快速安全地完成过渡，避免出现高度增加过大、过渡时间过长及作动器饱和等不利情况。

关 键 词：尾座式垂直起降无人机；动力学模型；最优化方法；过渡走廊；过渡轨迹
中图分类号：V212.5
文献标志码：A **文章编号**：1001-5965(2025)09-3071-15

近年来，垂直起降无人机已成为无人机家族中崭新的关键力量，其不仅能实现垂直起降机动，还能保持长时间高速飞行[1-3]。根据结构差异，垂直起降无人机主要分为3类：复合式[4]、倾转式[5]和尾座式[6]。相比其他2类，尾座式不需要复杂的机械结构，仅通过控制舵面的偏转即可实现模式的转换。这使尾座式垂直起降无人机在相同任务条件下具有更轻的质量和更小的能量消耗，因此，具有巨大的应用前景。然而，过渡阶段存在的大迎角使尾座式垂直起降无人机的气动力和控制效果表现出高度非线性特性，该特性给尾座式垂直起降无人机的过渡飞行控制带来极大的挑战。

目前，对于尾座式垂直起降无人机过渡阶段的研究主要聚焦于过渡控制器设计[7-10]和过渡轨迹优化[11-16]2个方面。其中，过渡轨迹优化的主要研究内容是寻找可行的最优控制输入，使无人机从初始状态转移到目标状态的过程中，性能指标(如过渡持续时间、过渡高度变化及过渡能量消耗)最小化。对过渡轨迹进行优化不仅能够降低尾座式垂直起降无人机对过渡控制器性能的要求，还能提高过渡效率，稳定机体姿态并保证过渡飞行安全。因此，进行尾座式垂直起降无人机过渡轨迹优化方法研究非常必要。

在尾座式垂直起降无人机过渡轨迹优化问题上，国内外已取得了出色的成果。开源PX4原生固件中的过渡方法是目前使用最为广泛的方法，该方法通过设置一个固定的推力输入和一个由经验确定的线性变化的俯仰角设定值来引导飞机完成过渡，因此，最为简单并易于实现。然而，PX4原生固件中的过渡方法在实际过渡时会导致较大的飞行高度变化和作动器输出，不利于飞行安全。文献[12]将飞机的平衡条件转化为约束优化问题，在平衡状态下进行线性化分析以确保稳定性和可控性，并通过构建新的代价函数实现高度变化最小的过渡机

收稿日期：2023-07-12；**录用日期**：2023-10-13；**网络出版时间**：2023-11-30 17:04
网络出版地址：link.cnki.net/urlid/11.2625.V.20231130.1113.001
基金项目：国家自然科学基金(52072309)；陕西省重点研发计划(2019ZDLGY14-02-01)；陕西省自然科学基础研究计划(2021JQ-086)；深圳基础研究项目(JCYJ20190806152203506)；航空科学基金(ASFC-2018ZC53026)
*通信作者. E-mail：zhaowen@nwpu.edu.cn

引用格式：邹旭，刘贞报，赵闻，等. 尾座式垂直起降无人机过渡轨迹优化方法[J]. 北京航空航天大学学报，2025，51(9)：3071-3085.
ZOU X，LIU Z B，ZHAO W，et al. Optimization method of transition trajectory for tail-sitter unmanned aerial vehicles[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3071-3085 (in Chinese).

动。文献[14]设计了针对尾座式飞机的动态平衡条件，并限制过渡机动在动态平衡条件下，以保证过渡轨迹的鲁棒性，该方法将飞机高度变化作为过渡轨迹优化问题的唯一指标，并在最大高度变化值的基础上，引入了飞机的高度变化均值，因此，具有出色的定高效果。文献[15]提出了能量最优过渡方法，该方法同时考虑能量消耗指标和俯仰角跟踪指标，因此，兼顾了无人机高度变化要求和能量消耗要求，是一种实用的过渡方法，但在实际飞行中，出现了飞机无法快速跟踪过渡轨迹的情况。

最优化理论是解决尾座式垂直起降无人机过渡轨迹规划问题的常用方法。然而，该方法是基于已知飞机模型推导最优过渡轨迹的方法，很难得到准确的无人机数学模型。现有大多数过渡方法很难同时兼顾飞行安全、能耗及失速风险，因此，可行性较低。此外，这些方法在追求高度变化最小或过渡时间最少时，没有考虑到作动器的饱和会降低飞机的抗干扰能力。因此，在研究基于最优化理论的过渡方法时，如何提高可行性并保留足够的作动器裕度以提高鲁棒性仍然具有很大的研究价值。

近年来，过渡走廊被广泛应用在倾转旋翼飞行器的设计中。倾转旋翼飞行器的过渡走廊是指其有效飞行区间，由包围该区间的包络线与包络面共同组成[17]。

有关倾转旋翼飞行器过渡走廊的研究[17-20]为本文解决尾座式垂直起降无人机过渡轨迹优化问题提供了重要的思路。将尾座式垂直起降无人机过渡轨迹限制在过渡走廊内部，不仅可以保证过渡轨迹的可行性和安全性，还可以根据过渡轨迹在过渡走廊中的位置评估其鲁棒性。目前，尾座式垂直起降无人机过渡走廊的研究还很有限，通过分析倾转旋翼飞行器倾转过渡走廊的构建方法[17-20]，可以得知构建过渡走廊的关键点在于利用其变速和变体等特点约束过渡走廊的范围。然而，这种构建方法不能直接用于尾座式垂直起降无人机，因为两者的过渡原理不同，导致针对倾转旋翼飞行器过渡走廊的构建方法不适用于尾座式垂直起降无人机的过渡模型。此外，在过渡时，尾座式垂直起降无人机相比倾转旋翼飞行器有着更大的姿态变化，需要更多的约束条件。

因此，本文基于倾转旋翼飞行器过渡走廊的设计思路，研究了针对尾座式垂直起降无人机过渡走廊的构建方法，并将其应用在尾座式垂直起降无人机的过渡轨迹优化问题中。以一种双发尾座式垂直起降无人机为研究对象，通过分析机翼不同区域之间的迎角差异，构建出飞机的非线性动力学模型，并设计一种尾座式垂直起降无人机过渡走廊。通过分析模型误差对过渡走廊的影响，使用最优化方法得到一条具有最大安全裕度的目标过渡轨迹。同时，将尾座式垂直起降无人机的过渡过程视为轨迹优化问题，求解该问题得到最接近目标过渡轨迹且能够保留足够作动器裕度的最优过渡轨迹。仿真和实际飞行结果表明，所设计的过渡方法能够在保留一定作动器裕度的前提下，引导飞机快速安全地完成过渡，避免飞机出现高度增加过大、过渡时间过长及作动器饱和等对过渡飞行不利的情况。

本文的创新点总结如下：

1）目前，尾座式垂直起降无人机过渡走廊的研究很少，受传统倾转旋翼飞行器过渡走廊设计思路启发，设计了一种特别针对尾座式垂直起降无人机的过渡走廊。

2）将尾座式垂直起降无人机过渡走廊引入尾座式垂直起降无人机过渡轨迹优化问题。与传统方法相比，本文方法具有更高的可行性和鲁棒性。

3）为保证建模的准确性，不同于传统的整体建模方法，采用“分区域”建模思路构建尾座式垂直起降无人机的非线性动力学模型，并进行了小型风洞试验。

1 动力学分析及建模

1.1 坐标系定义

如图1所示，双发尾座式垂直起降无人机的坐标系包括地面坐标系、机体坐标系及螺旋桨坐标系。其中，地面坐标系采用北-东-地的常规定义形式。机体坐标系以飞机重心位置为原点，x^{B}轴指向机头方向，y^{B}轴指向翼尖，z^{B}轴指向机腹。螺旋桨坐标系的原点位于2个螺旋桨的中心处，且与机体坐标系平行。

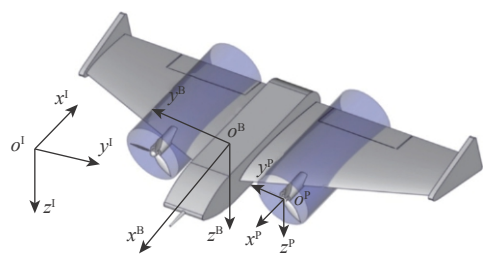


图1 尾座式垂直起降无人机坐标系定义示意图

Fig. 1 Coordinate frames of the tail-sitter UAV

1.2 动力学模型

尾座式垂直起降无人机的过渡阶段以纵向运动为主，因此，仅考虑飞机的纵向动力学模型。完整6自由度运动方程可以表示为

$$\begin{cases} m\dot{\boldsymbol{V}}_{\mathrm{g}}^{\mathrm{I}}=\boldsymbol{R}_{\mathrm{B}}^{\mathrm{I}}(\boldsymbol{F}_{\mathrm{e}}^{\mathrm{B}}+\boldsymbol{F}_{\mathrm{d}}^{\mathrm{B}}) \\ \boldsymbol{J}\dot{\boldsymbol{\omega}}^{\mathrm{B}}=-[\boldsymbol{\omega}^{\mathrm{B}}]\times\boldsymbol{J}\boldsymbol{\omega}^{\mathrm{B}}+\boldsymbol{M}_{\mathrm{e}}^{\mathrm{B}}+\boldsymbol{M}_{\mathrm{d}}^{\mathrm{B}} \\ \dot{\boldsymbol{P}}=\boldsymbol{V}_{\mathrm{g}}^{\mathrm{I}} \end{cases} \tag{1}$$

式中: m 为无人机质量; $\boldsymbol{P}$为位置; $\boldsymbol{V}_{\mathrm{g}}^{\mathrm{I}}$为地速; $\boldsymbol{R}_{\mathrm{B}}^{\mathrm{I}}$为旋转矩阵; $\boldsymbol{\omega}^{\mathrm{B}}$为旋转角速度向量; $\boldsymbol{J}$为惯性矩阵; $\boldsymbol{F}_{\mathrm{d}}^{\mathrm{B}}$和$\boldsymbol{M}_{\mathrm{d}}^{\mathrm{B}}$分别为未知的力和力矩; $\boldsymbol{F}_{\mathrm{e}}^{\mathrm{B}}$和$\boldsymbol{M}_{\mathrm{e}}^{\mathrm{B}}$分别为已知合力和合力矩, 其表达形式为

$$\begin{cases} \boldsymbol{F}_{\mathrm{e}}^{\mathrm{B}}=\boldsymbol{F}_{\mathrm{aero}}+\boldsymbol{F}_{\mathrm{T}}+\boldsymbol{G}^{\mathrm{B}} \\ \boldsymbol{M}_{\mathrm{e}}^{\mathrm{B}}=\boldsymbol{M}_{\mathrm{aero}}+\boldsymbol{M}_{\mathrm{T}} \end{cases} \tag{2}$$

其中: $\boldsymbol{G}^{\mathrm{B}}$为重力向量; $\boldsymbol{F}_{\mathrm{aero}}$和$\boldsymbol{M}_{\mathrm{aero}}$分别为气动力向量和气动力矩向量; $\boldsymbol{F}_{\mathrm{T}}$和$\boldsymbol{M}_{\mathrm{T}}$分别为推力向量和推力力矩向量。

如图 2 所示, 双发尾座式垂直起降无人机的机翼可分成 6 个区域, 其中 L3、L1、R1 和 R3 不受螺旋桨尾流影响, 被称为“自由区”, 而 L2 和 R2 会受到螺旋桨尾流影响, 被称为“滑流区”。自由区的空速$\boldsymbol{V}_{\mathrm{a,out}}^{\mathrm{B}}$为

$$\boldsymbol{V}_{\mathrm{a,out}}^{\mathrm{B}}=\boldsymbol{R}_{\mathrm{I}}^{\mathrm{B}}(\boldsymbol{V}_{\mathrm{g}}^{\mathrm{I}}-\boldsymbol{V}_{\mathrm{w}}^{\mathrm{I}}) \tag{3}$$

式中: $\boldsymbol{R}_{\mathrm{I}}^{\mathrm{B}}$为大地坐标系到机体坐标系的旋转矩阵; $\boldsymbol{V}_{\mathrm{w}}^{\mathrm{I}}$为风速。滑流区的空速$\boldsymbol{V}_{\mathrm{a,in}}^{\mathrm{B}}$为

$$\boldsymbol{V}_{\mathrm{a,in}}^{\mathrm{B}}=\boldsymbol{R}_{\mathrm{I}}^{\mathrm{B}}(\boldsymbol{V}_{\mathrm{g}}^{\mathrm{I}}-\boldsymbol{V}_{\mathrm{w}}^{\mathrm{I}})+\boldsymbol{V}_{\mathrm{induce}}^{\mathrm{B}} \tag{4}$$

式中: $\boldsymbol{V}_{\mathrm{induce}}^{\mathrm{B}}$为螺旋桨尾流向量; $\boldsymbol{V}_{\mathrm{induce}}^{\mathrm{B}}=[-u_0\ \ 0\ \ 0]^{\mathrm{T}}$; u_0为螺旋桨尾流的速度大小。根据动量理论, u_0计算如下:

$$u_0=\frac{v_0+\sqrt{v_0^2+2T_{\mathrm{P}}/(\rho S)}}{2\sqrt{l^2+R^2}\Big/\left(l+\sqrt{l^2+R^2}\right)} \tag{5}$$

式中: l为滑流区气动焦点到螺旋桨的垂直距离; v_0为螺旋桨x^{P}轴方向的来流速度; S和R分别为螺旋桨的面积和半径; T_{P}为推力大小; ρ为空气密度。

在飞行中, 自由区和滑流区具有不同的迎角, 滑流区迎角α_{in}和自由区迎角α_{out}分别为

$$\begin{cases} \alpha_{\mathrm{in}}=\arctan\left(V_{\mathrm{a,in},z}^{\mathrm{B}}/V_{\mathrm{a,in},x}^{\mathrm{B}}\right) \\ \alpha_{\mathrm{out}}=\arctan\left(V_{\mathrm{a,out},z}^{\mathrm{B}}/V_{\mathrm{a,out},x}^{\mathrm{B}}\right) \end{cases} \tag{6}$$

式中: $V_{\mathrm{a,in},z}^{\mathrm{B}}$和$V_{\mathrm{a,in},x}^{\mathrm{B}}$分别为滑流区的空速在$z^{\mathrm{B}}$和$x^{\mathrm{B}}$轴上的分量; $V_{\mathrm{a,out},z}^{\mathrm{B}}$和$V_{\mathrm{a,out},x}^{\mathrm{B}}$分别为自由区的空速在$z^{\mathrm{B}}$和$x^{\mathrm{B}}$轴上的分量。对机翼每个区域分别计算升力、阻力和俯仰力矩, 可得到飞机总的升力 L、阻力 D 以及俯仰力矩 M 为

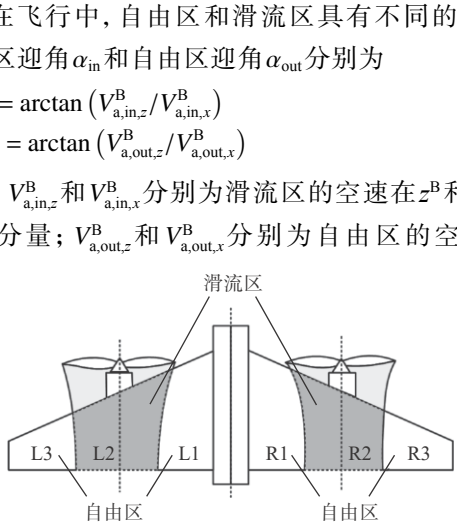


图 2　机翼分解示意图

Fig. 2　Wind components breakdown

$$\begin{cases} L=\displaystyle\sum_{n=1}^{6}\rho v_n^2\left(S_{\mathrm{w},n}C_{L_n}+S_{\mathrm{w},n}C_{Lq}\frac{\bar{c}}{2v_n}q+S_{\mathrm{e},n}C_{L\delta}\delta_{\mathrm{e}}\right)/2 \\ D=\displaystyle\sum_{n=1}^{6}\rho v_n^2\left(S_{\mathrm{w},n}C_{D_n}+S_{\mathrm{w},n}C_{Dq}\frac{\bar{c}}{2v_n}q+S_{\mathrm{e},n}C_{D\delta}\delta_{\mathrm{e}}\right)/2 \\ M=\displaystyle\sum_{n=1}^{6}\rho v_n^2\bar{c}\left(S_{\mathrm{w},n}C_{M_n}+S_{\mathrm{w},n}C_{Mq}\frac{\bar{c}}{2v_n}q+S_{\mathrm{e},n}C_{M\delta}\delta_{\mathrm{e}}\right)/2 \end{cases} \tag{7}$$

式中: n代表第n个区域; $S_{\mathrm{w},n}$为第n个区域的机翼面积; $S_{\mathrm{e},n}$为第n个区域的舵面面积; v_n为第n个区域的空速; $\bar{c}$为平均气动弦长; δ_{e}为舵偏角; q为俯仰角速度; C_{L_n}、C_{Lq}、$C_{L\delta}$、C_{D_n}、C_{Dq}、$C_{D\delta}$、C_{M_n}、C_{Mq}和$C_{M\delta}$为气动系数。

将式(7)的计算结果转换到机体坐标系下, 得到升力和阻力在 3 个机体坐标轴上的分量如下:

$$\begin{cases} \boldsymbol{L}^{\mathrm{B}}=[L\sin\alpha_{\mathrm{out}}\ \ \ 0\ \ \ -L\cos\alpha_{\mathrm{out}}]^{\mathrm{T}} \\ \boldsymbol{D}^{\mathrm{B}}=[-D\cos\alpha_{\mathrm{out}}\ \ \ 0\ \ \ -D\sin\alpha_{\mathrm{out}}]^{\mathrm{T}} \end{cases} \tag{8}$$

式中: $\boldsymbol{L}^{\mathrm{B}}$和$\boldsymbol{D}^{\mathrm{B}}$分别为机体坐标系下的升力向量和阻力向量。纵向的总气动力和气动力矩为

$$\begin{cases} \boldsymbol{F}_{\mathrm{aero}}=\boldsymbol{L}^{\mathrm{B}}+\boldsymbol{D}^{\mathrm{B}} \\ \boldsymbol{M}_{\mathrm{aero}}=\displaystyle\sum_{n=1}^{6}\begin{bmatrix} F_{\mathrm{aero},y}^{n}l_y^n \\ M_n+F_{\mathrm{aero},z}^{n}l_x^n \\ F_{\mathrm{aero},x}^{n}l_y^n \end{bmatrix} \end{cases} \tag{9}$$

式中: F_{aero}^{n}和M_n分别为第n个区域的气动力和气动力矩; l_x^n和l_y^n分别为第n个区域的气动中心到飞机重心的连线在x^{B}轴和y^{B}轴上的投影距离。单个螺旋桨产生的推力大小为

$$T_{\mathrm{P}}=C_{\mathrm{T}}\rho\left(\frac{\omega}{60}\right)^2S \tag{10}$$

式中: C_{T}为推力系数; ω为螺旋桨转速。因此, 推力向量和推力力矩向量为

$$\begin{cases} \boldsymbol{F}_{\mathrm{T}}=[F_{\mathrm{T1}}+F_{\mathrm{T2}}\ \ \ 0\ \ \ 0]^{\mathrm{T}} \\ \boldsymbol{M}_{\mathrm{T}}=[M_{\mathrm{T1}}+M_{\mathrm{T2}}\ \ \ 0\ \ \ (F_{\mathrm{T1}}-F_{\mathrm{T2}})l_{\mathrm{r}}]^{\mathrm{T}} \end{cases} \tag{11}$$

式中: F_{T1}和M_{T1}分别为左螺旋桨产生的推力和力矩; F_{T2}和M_{T2}分别为右螺旋桨产生的推力和力矩; l_{r}为左右螺旋桨到飞机对称平面的距离。将式(9)和式(11)代入式(2)可得到飞机的合外力和合外力矩, 进而得到理想情况下的纵向运动方程:

$$\begin{cases} \dot{u}=-g\sin\theta-wq+\left(F_{\mathrm{e},x}^{\mathrm{B}}+F_{\mathrm{T1}}+F_{\mathrm{T2}}\right)/m \\ \dot{w}=g\cos\theta+qu+F_{\mathrm{e},z}^{\mathrm{B}}/m \\ \dot{\theta}=q \\ \dot{q}=M_{\mathrm{e},y}^{\mathrm{B}}/I_{yy} \\ \dot{X}^{\mathrm{I}}=u\cos\theta+w\sin\theta \\ \dot{Z}^{\mathrm{I}}=-u\sin\theta+w\cos\theta \end{cases} \tag{12}$$

式中：g为重力加速度；$F^{B}_{e,x}$和$F^{B}_{e,z}$分别为合力在x^{B}和z^{B}轴上的分量；$M^{B}_{e,y}$为合力矩在y^{B}轴上的分量；I_{yy}为对y^{B}轴的转动惯量；$\dot{Z}^{I}$和$\dot{X}^{I}$分别为无人机的高度变化率和前向位置变化率；θ为俯仰角；u和w分别为机体坐标系下沿x^{B}和z^{B}轴的速度。

1.3 模型验证

采用Arkward尾座式飞机为试验样机，如图3所示，其基本参数如表1所示。为验证1.2节中模型的准确性，利用图4(a)的小型风洞设备进行试验。该风洞设备能够在2～25 m/s的速度范围内提供稳定的低雷诺气流。图4(b)为风洞试验原理，在试验过程中，制造了一个经过改装的Arkward样机以实现内部平衡的刚性安装，并为气动测试提供一个不可变形的翼面。改装样机的尺寸与原样机相同。将样机安装在测试台上，通过五分量测力天平测量样机的力和力矩，当力和力矩实现平衡时，样机的舵偏角、总功率及俯仰角即为配平结果。样机的舵偏角通过安装在舵面上的电位计测量。螺旋桨功率借助直流速度控制器测量的螺旋桨转动速度间接测量。风洞设备使用Xbee无线电模块将测量数据传送到控制站。将计算得到的理论平飞配平结果与风洞试验的配平结果进行对比，如图5所示。可知，计算配平结果与风洞试验数据相符，因此，本文构建的动力学模型准确，可用于研究过渡轨迹优化问题。

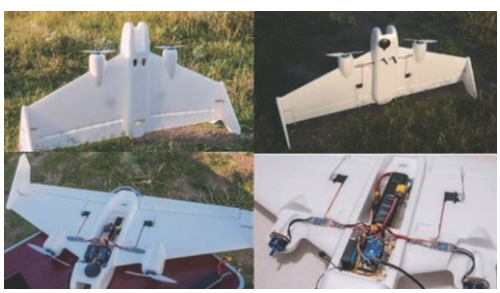


图3 试验尾座式飞机

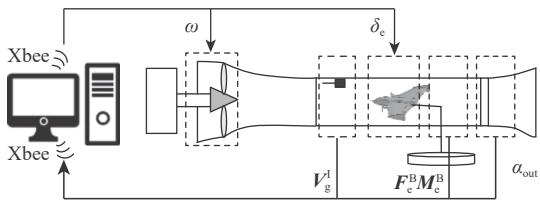
Fig. 3 Tested tail-sitter vehicle

表1 Arkward尾座式飞机参数

Table 1 Parameters of Arkward tail-sitter vehicle

物理参数	数值
起飞质量/kg	0.93
翼展/m	0.89
机翼表面积/m^2	0.13
螺旋桨桨盘半径/m	0.16
平均气动弦长/m	0.14
单个升降舵面积/m^2	0.05
单个螺旋桨最大推力/N	6.90
升降舵最大偏转角度/(°)	30
螺旋桨最大推力变化速率/(N·s^{-1})	20
升降舵最大偏转速率/((°)·s^{-1})	110

(a) 试验设备



(b) 试验原理

图4 风洞试验设备及原理

Fig. 4 Wind-tunnel facility and principle

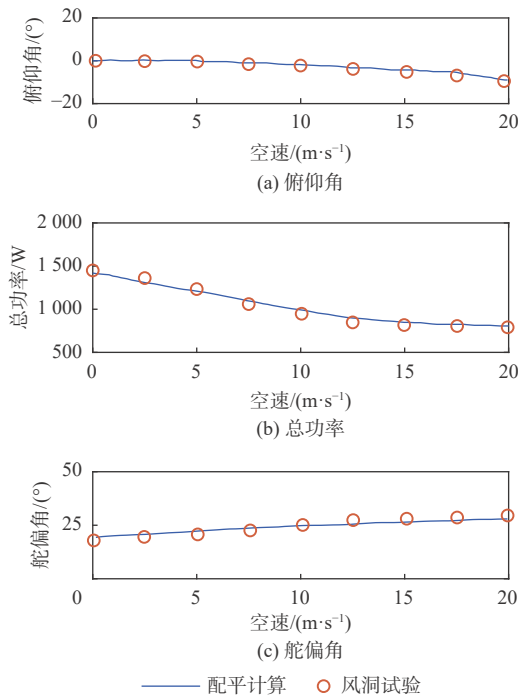


图5 模型计算配平结果和风洞试验数据对比

Fig. 5 Calculated trim vs wind tunnel experimental data

2 过渡走廊设计

2.1 过渡走廊概念和形式

对于倾转旋翼机，过渡走廊是指从直升机模式过渡到固定翼模式时，在不同倾转角下最小和最大速度的边界形成的包线[21-22]。在求解过渡走廊时，通常首先确定飞机的机翼升力特性、功率特性及失速特性，然后使用遍历的计算方法对所有状态点进行配平计算，得到所有的可行状态点后，最终进行边界插值确定过渡走廊边界。传统倾转旋翼机的过渡走廊通常以倾转角-速度曲线表示，如图6所

示。本文类比倾转旋翼过渡走廊的定义, 将尾座式垂直起降无人机的过渡走廊定义为: 在给定的俯仰角条件下, 会存在一个飞行速度范围, 在该范围内飞机可以通过配平实现稳定的飞行, 而尾座式垂直起降无人机过渡走廊就是不同俯仰角条件下所有飞行速度范围的集合。

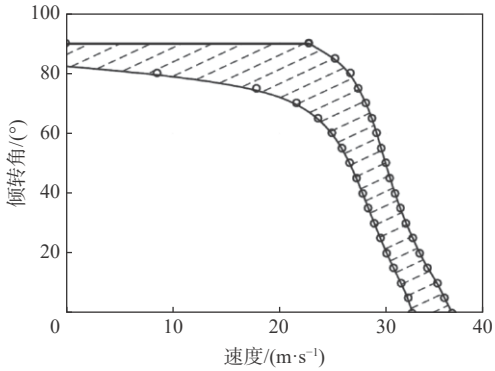


图 6　传统倾转旋翼机过渡走廊示意图

Fig. 6　Schematic diagram of the transition corridor for a conventional tiltrotor

2.2　过渡走廊具体设计过程

在进行前向过渡时, 尾座式垂直起降无人机会受到多重约束条件的限制, 这些限制反映了无人机在前向过渡模式下的可行状态空间。

1) 状态约束。前向过渡是加速、低头的过程, 因此, 合外力分量应满足 $F^{\mathrm{B}}_{\mathrm{e},x}\geqslant 0$, 俯仰角速度应满足 $q\leqslant 0$。此外, 由于样机的最大速度为 20 m/s, 舵面最大偏转角为 30°, 因此, 前向过渡走廊速度和舵偏角的约束范围分别为 $V\in[0,20\ \mathrm{m/s}]$ 和 $\delta_{\mathrm{e}}\in[-30°,30°]$。

2) 迎角约束。迎角对于无人机的气动力具有决定性的影响。为了确保前向过渡的安全, 设定前向过渡时的自由区机翼迎角约束条件为 $\alpha_{\mathrm{out}}\in[0,\alpha_{\max}]$, $\alpha_{\max}$ 为允许的最大迎角, 早期试验中通常将 $\alpha_{\max}$ 取为失速迎角值的 80%。

3) 推力约束。本文样机单个螺旋桨的最大推力为 6.9 N, 因此, 总推力约束范围为: $0\leqslant F_{\mathrm{T1}}+F_{\mathrm{T2}}\leqslant 13.8$ N。

此外, 为确保前向过渡的安全, 需要保证飞机不能掉高, 同时, 不能出现高度增加过大的情况。因此, 设定高度变化约束条件如下:

$$\begin{cases} F^{\mathrm{B}}_{\mathrm{e},z}\leqslant 0 \\ -3\ \mathrm{m/s}\leqslant \dot{Z}^{\mathrm{I}}\leqslant 0 \end{cases} \tag{13}$$

综上, 尾座式垂直起降无人机完整的前向过渡走廊求解方程组如下:

$$\begin{cases} F^{\mathrm{B}}_{\mathrm{e},x}\geqslant 0 \\ F^{\mathrm{B}}_{\mathrm{e},z}\leqslant 0 \\ M^{\mathrm{B}}_{\mathrm{e},y}=0 \\ q\leqslant 0 \\ -3\ \mathrm{m/s}\leqslant \dot{Z}^{\mathrm{I}}\leqslant 0 \\ 0\leqslant F_{\mathrm{T1}}+F_{\mathrm{T2}}\leqslant 13.8\ \mathrm{N} \\ V\in[0,20\ \mathrm{m/s}] \\ \delta_{\mathrm{e}}\in[-30°,30°] \\ \alpha_{\mathrm{out}}\in[0,\alpha_{\max}] \end{cases} \tag{14}$$

式中: V 为自由区空速 $\boldsymbol{V}^{\mathrm{B}}_{\mathrm{a,out}}$ 的大小。

将速度作为 x 轴, 俯仰角作为 y 轴, 计算符合式 (14) 的所有可行飞行点, 最终得到关于速度和俯仰角的尾座式垂直起降无人机前向过渡走廊, 如图 7(a) 中的灰色区域所示。同理, 得到后向过渡走廊求解方程组如下:

$$\begin{cases} F^{\mathrm{B}}_{\mathrm{e},x}\leqslant 0 \\ F^{\mathrm{B}}_{\mathrm{e},z}\leqslant 0 \\ M^{\mathrm{B}}_{\mathrm{e},y}=0 \\ q\geqslant 0 \\ -3\ \mathrm{m/s}\leqslant \dot{Z}^{\mathrm{I}}\leqslant 0 \\ 0\leqslant F_{\mathrm{T1}}+F_{\mathrm{T2}}\leqslant 13.8\ \mathrm{N} \\ V\in[0,20\ \mathrm{m/s}] \\ \delta_{\mathrm{e}}\in[-30°,30°] \\ \alpha_{\mathrm{out}}\in[0,\alpha_{\max}] \end{cases} \tag{15}$$

通过求解式 (15) 得到后向过渡走廊, 如图 7(b) 灰色区域所示。

为了说明设计的过渡走廊与无人机实际过渡

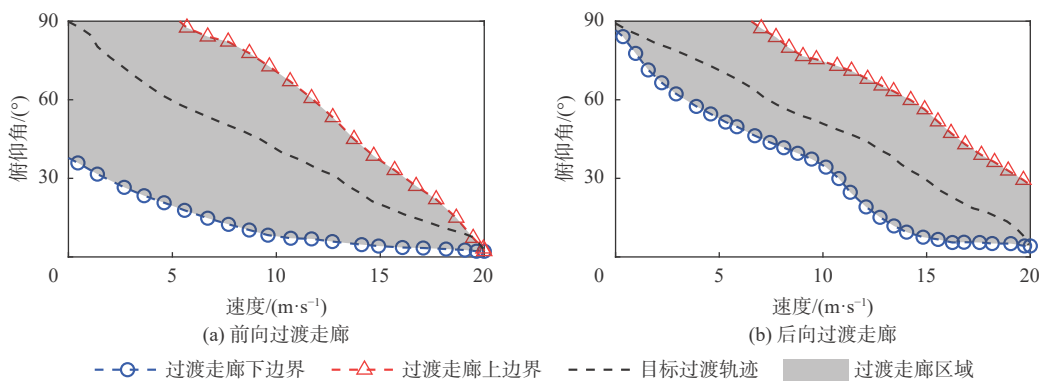


图 7　前后向过渡走廊

Fig. 7　Front and back transition corridors

模态转换能力的匹配程度，进行20组过渡飞行试验（10组前向过渡试验，10组后向过渡试验），并将该20组试验数据标注到构建的过渡走廊中，如图8所示。可见，大部分试飞轨迹位于构建的过渡走廊内部，说明本文设计的过渡走廊与该样机实际过渡转换能力的匹配程度良好。

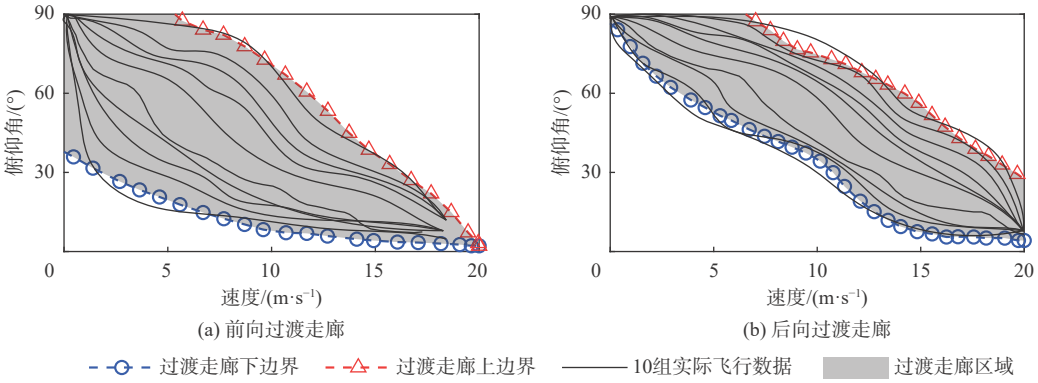


图8 尾座式垂直起降无人机过渡走廊计算结果与实际飞行数据对比

Fig. 8 Comparison of the calculation results of the tail-sitter vertical take-off and landing UAV transition corridor with the actual flight data

2.3 模型误差对前/后向过渡走廊的影响

在实际应用中，获取准确的气动参数十分困难。因此，基于理想模型所构建的过渡走廊不可避免地存在偏差。以升力系数偏差和俯仰力矩系数偏差为例，图9和图10分别为升力系数 C_L 和俯仰力矩系数 C_M 对过渡走廊大小的影响。可知，升力系数绝对值的减小会导致前/后向过渡走廊变窄，且任意速度对应的最小俯仰角均出现增大的情况。俯仰力矩系数的变化对后向过渡走廊的影响较大，随着俯仰力矩系数绝对值的增大，任意俯仰角所对应的最大飞行速度减小。即如果实际俯仰力矩系数绝对值比理论俯仰力矩系数绝对值大，则实际的

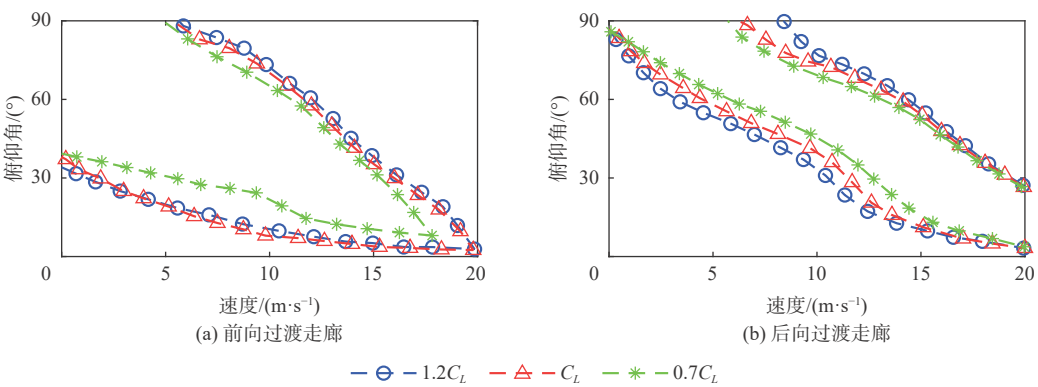


图9 升力系数对前后向过渡走廊的影响

Fig. 9 Effect of lift coefficient on the front and back transition corridors

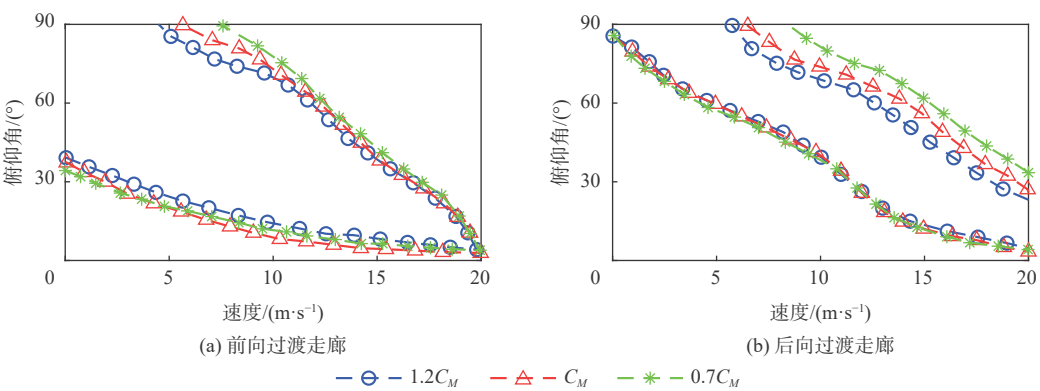


图10 俯仰力矩系数对前后向过渡走廊的影响

Fig. 10 Effect of pitching moment coefficient on the front and back transition corridors

后向过渡走廊会更窄。

3　目标过渡轨迹设计

理论上, 过渡走廊内部的任意一条轨迹都可以作为过渡轨迹。然而, 由于实际过渡走廊的偏差, 有必要在过渡走廊内部求解一条具有最大安全裕度的过渡轨迹作为目标过渡轨迹。本文在同时考虑过渡走廊下边界和上边界的情况下, 定义安全裕度函数为

$$\theta^*(V)=\frac{|\theta(V)-\theta_{\min}(V)|}{\theta_{\max}(V)-\theta_{\min}(V)}\cdot\frac{|\theta(V)-\theta_{\max}(V)|}{\theta_{\max}(V)-\theta_{\min}(V)} \tag{16}$$

式中: $\theta^*(V)$ 为目标过渡轨迹俯仰角; $\theta(V)$ 为实际过渡轨迹俯仰角; $\theta_{\min}(V)$为过渡走廊下边界; $\theta_{\max}(V)$为过渡走廊上边界。则求解前/后向过渡走廊目标过渡轨迹的最优化问题可描述为

$$\begin{cases}\text{find } \theta(V)\\ \max\ J=\int_{V_0}^{V_1}\theta^*(V)\mathrm{d}V\\ \quad\text{s.t.}\begin{cases}\theta_{\min}(V)\leqslant\theta(V)\leqslant\theta_{\max}(V)\\ \theta(V_0)=\pi/2\\ \theta(V_1)=\pi/36\end{cases}\end{cases} \tag{17}$$

$$\begin{cases}\text{find } \theta(V)\\ \max\ J=\int_{V_0}^{V_1}\theta^*(V)\mathrm{d}V\\ \quad\text{s.t.}\begin{cases}\theta_{\min}(V)\leqslant\theta(V)\leqslant\theta_{\max}(V)\\ \theta(V_0)=\pi/36\\ \theta(V_1)=\pi/2\end{cases}\end{cases} \tag{18}$$

式中: V_0 为过渡开始时的飞机速度; V_1 为过渡结束时的飞机速度。求解式 (17) 和式 (18), 得到前/后向过渡走廊目标过渡轨迹, 如图 7 中的黑色虚线所示。可知, 求得的目标过渡轨迹总体上位于 2 个过渡走廊居中的位置。

4　最优过渡轨迹

本节在目标过渡轨迹基础上进一步优化, 寻找一条最优过渡轨迹。该过渡轨迹能够确保尾座式垂直起降无人机实现快速且安全的过渡, 并能留有充足的作动器余量以应对未知的干扰。将最优过渡轨迹的求解过程视为最优控制问题[15,23-24], 通过寻找可行的最优控制序列 $\boldsymbol{U}^*(t)$, 使飞机从初始状态 $\boldsymbol{x}(t_0)$转移到目标最终状态 $\boldsymbol{x}(t_\mathrm{n})$的过程中, 尽可能接近目标过渡轨迹, 并且满足预设的代价函数值最小的条件。

4.1　过渡最优化目标

在保证过渡安全的前提下, 从高度变化最优、过渡轨迹安全裕度最优、作动器裕度最优及控制效果最优 4 个方面设置优化目标。

1) 高度变化最优。在过渡过程中, 高度变化是一个关键指标。设置过渡高度变化最优化目标函数为

$$j_1=\int_{t_0}^{t_\mathrm{n}}(h(\tau)-h_{\text{init}})^2\mathrm{d}\tau \tag{19}$$

式中: $h(\tau)$为飞机实时高度; h_{init}为飞机初始高度; t_0 和t_n分别为过渡开始和结束时间。

2) 过渡轨迹安全裕度最优。目标过渡轨迹能使无人机在过渡时具有最大的安全裕度, 因此, 在设计最优过渡轨迹时, 应考虑最优过渡轨迹与目标过渡轨迹之间的偏差。设定过渡轨迹安全裕度最优化目标函数为

$$j_2=\int_{V_0}^{V_1}(\theta(V)-\theta^*(V))^2\mathrm{d}V \tag{20}$$

3) 作动器裕度最优。无人机在过渡时所需作动器的输出越小, 则用于机动飞行和姿态调整的剩余作动器输出越多。因此, 设定作动器裕度最优化目标为

$$j_3=\int_{t_0}^{t_\mathrm{n}}\left[\left(\frac{\omega(\tau)}{\omega_{\max}}\right)^2+\left(\frac{\delta_\mathrm{e}(\tau)}{\delta_{\mathrm{e},\max}}\right)^2\right]\mathrm{d}\tau \tag{21}$$

式中: $\omega(\tau)$为螺旋桨转速; $\omega_{\max}$为螺旋桨的最大转速; $\delta_\mathrm{e}(\tau)$为舵偏角; $\delta_{\mathrm{e},\max}$为最大舵偏角。

4) 控制效果最优。无人机在进行过渡时, 需要考虑过渡轨迹可行性。为了减小实际状态与期望状态之间的偏差, 设置过渡控制效果最优化目标函数为

$$j_4=\beta_u(u(t_\mathrm{n})-u)^2+\beta_w(w(t_\mathrm{n})-w)^2+\beta_\theta(\theta(t_\mathrm{n})-\theta)^2 \tag{22}$$

式中: β_u、β_w、β_θ为权重系数。

4.2　过渡最优化求解模型

综合以上分析, 最终的过渡最优化目标代价函数 j 为

$$\begin{aligned}j=&k_1j_1+k_2j_2+k_3j_3+j_4=\\ &k_1j_1+k_2j_2+k_3j_3+\beta_u(u(t_\mathrm{n})-u)^2+\\ &\beta_w(w(t_\mathrm{n})-w)^2+\beta_\theta(\theta(t_\mathrm{n})-\theta)^2\end{aligned} \tag{23}$$

式中: k_1、k_2、k_3为权重系数, 权重系数的值越大, 对应项越重要。通过大量的仿真和试飞得出适合本文样机的权重系数为: k_1=1, k_2=1, k_3 = 2, $\beta_u=\beta_w=\beta_\theta=1$。过渡最优化求解模型可描述为

$$\begin{cases}\text{find } \left[U_\mathrm{p},U_\delta\right]\\ \min j & \cdots\text{式 (23)}\\ \quad\text{s.t. } \dot{\boldsymbol{x}}(t)=f(\boldsymbol{x},u) & \cdots\text{式 (12)}\\ \quad\boldsymbol{x}(t_0)=\boldsymbol{x}_0 & \cdots\text{初始边界条件}\\ \quad\boldsymbol{x}(t_\mathrm{n})=\boldsymbol{x}_\mathrm{n} & \cdots\text{终止边界条件}\\ \quad p(\boldsymbol{x},u)\leqslant0 & \cdots\quad\text{表 2}\end{cases} \tag{24}$$

式中：p 为约束函数；$\boldsymbol{x}_{\mathrm{n}}$ 为终止边界条件下的系统状态向量。第 1 条约束条件是尾座式垂直起降无人机的动力学约束。第 2 条和第 3 条约束条件是初始边界条件和和终止边界条件，即过渡开始时的无人机飞行状态$\boldsymbol{x}(t_0)$和过渡结束时的无人机飞行状态$\boldsymbol{x}(t_{\mathrm{n}})$。设定的边界条件包括：

$$\begin{cases}\boldsymbol{x}(t_0)=\left[h(t_0)\quad \dot{h}(t_0)\quad \theta(t_0)\quad q(t_0)\quad u(t_0)\right]^{\mathrm{T}}\\ \boldsymbol{x}(t_{\mathrm{n}})=\left[h(t_{\mathrm{n}})\quad \dot{h}(t_{\mathrm{n}})\quad \theta(t_{\mathrm{n}})\quad q(t_{\mathrm{n}})\quad u(t_{\mathrm{n}})\right]^{\mathrm{T}}\end{cases} \tag{25}$$

式中：h 为高度；$\dot{h}$为爬升速度。值得注意的是，边界条件的值并不固定，应根据任务的需要和飞机性能进行决定。以本文的实际飞行测试为例，设定的前向过渡的边界条件值为：$h(t_0)=h(t_{\mathrm{n}})=30$ m，$\dot{h}(t_0)=\dot{h}(t_{\mathrm{n}})=0$ m/s，$\theta(t_0)=85°$，$\theta(t_{\mathrm{n}})=5°$，$q(t_0)=q(t_{\mathrm{n}})=0(°)$/s，$u(t_0)=0$ m/s，$u(t_{\mathrm{n}})=16$ m/s；后向过渡的边界条件值为：$h(t_0)=h(t_{\mathrm{n}})=30$ m，$\dot{h}(t_0)=\dot{h}(t_{\mathrm{n}})=0$ m/s，$\theta(t_0)=5°$，$\theta(t_{\mathrm{n}})=85°$，$q(t_0)=q(t_{\mathrm{n}})=0(°)$/s，$u(t_0)=16$m/s，$u(t_{\mathrm{n}})=0$m/s。式 (24) 的第 4 条约束条件为状态约束条件，为确保轨迹优化结果的可行性，需要考虑过渡时间、飞机状态及作动器状态约束条件，表 2 为需要考虑的约束条件及各自的约束范围。

表 2　轨迹优化问题的约束条件

Table 2　Constraints of trajectory optimization problems

约束对象	约束范围
螺旋桨转速ω/(r·min^{-1})	$[0, \omega_{\max}]$
舵偏角δ_e/(°)	$[-\delta_{e,\max}, \delta_{e,\max}]$
过渡开始时间t_0/s	[0, 0]
过渡结束时间t_n/s	[5, 5]
水平速度$\dot{X}^{\mathrm{I}}(t)$/(m·s^{-1})	[0, 20]
垂直速度$\dot{Z}^{\mathrm{I}}(t)$/(m·s^{-1})	[−1, 1]

求解该过渡最优化模型时，使用文献 [25] 提出的一种数值求解方法，该方法的特点为将无穷维决策变量离散到有限维，使优化问题能用带约束的多元函数极值问题逼近。求解完成后可得到最优控制序列（螺旋桨的最优控制输入序列 U_{p} 和舵机的最优控制输入序列U_δ）及飞机的各状态量随时间的变化轨迹。最优控制输入序列和各状态量随时间的变化轨迹共同组成了飞机的最优过渡轨迹。

5　过渡控制框架

在文献 [26-28] 中，最优控制序列直接被用作开环控制输入来控制无人机完成过渡。虽然文献中的实验结果表明了这种做法的可行性，但单靠开环控制结构不能有效地抑制过渡过程中可能遇到的各种干扰。因此，本文设计了一个过渡控制框架，采用前馈控制和反馈控制的组合方式（见图 11），其中，前馈控制部分直接使用最优控制输入来提高系统的动态响应，而反馈控制部分则借助简单的反馈控制器来减小跟踪误差。需要说明的是，本文的研究重点是过渡轨迹的优化过程，而非过渡控制器的设计。为了突出本文所提出的过渡轨迹优化方法对过渡性能的提升效果，在过渡框架中仅采用简单的 P 控制器和 PD 控制器作为反馈控制器。各控制器的具体表达式详见附录 A。

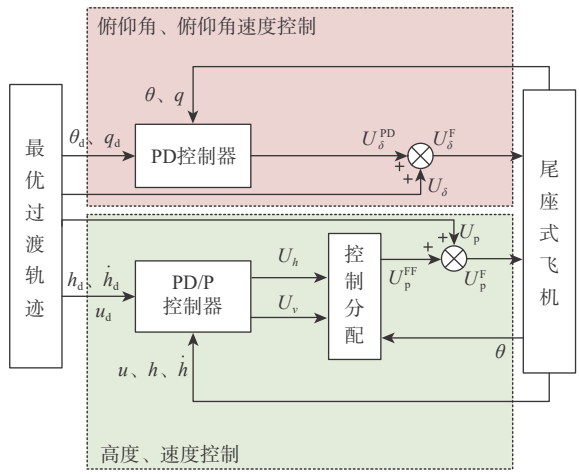


图 11　过渡控制框架

Fig. 11　The transition control structure

6　硬件在环仿真

6.1　硬件在环仿真平台

为了验证所提出的过渡优化方法的有效性并确保后续实际飞行测试的安全，搭建了一个硬件在环仿真平台，如图 12 所示。该平台由 5 个部分组成：自驾仪、NVIDIA 板卡、无线路由器、Lipo 电池和实时动态仿真器。其中，桌面计算机负责运行实时动态仿真模型，根据所建立的飞机数学模型实时计算飞机的姿态和位置等状态信息。Pixhawk 自驾仪和英伟达板卡组成飞行控制单元，根据飞机状态信息实时计算控制指令并将其回传给动态仿真模型。

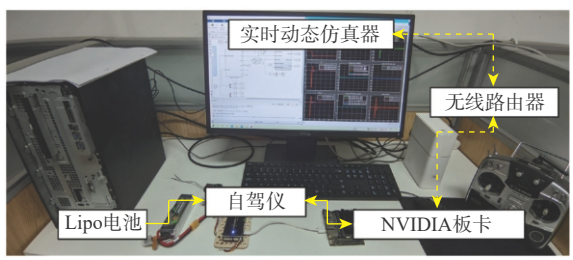


图 12　硬件在环仿真平台

Fig. 12　Hardware-in-the-loop (HITL) simulation platform

6.2　硬件在环仿真任务设置

使用上述仿真平台，进行包含完整飞行过程的

仿真实验: 飞机首先以垂直姿态开始爬升, 到达高度 30 m 后, 依次进行前向过渡、水平飞行和后向过渡, 最后进行垂直降落。在前向过渡轨迹优化过程中, 无人机的初始状态为: 速度 0 m/s, 高度 30 m, 俯仰角 85°, 终止条件为: 速度 20 m/s, 高度 30 m, 俯仰角 5°。在后向过渡轨迹优化过程中, 无人机的初始状态为: 速度 20 m/s, 高度 30 m, 俯仰角 5°, 终止条件为: 速度 0 m/s, 高度 30 m, 俯仰角 85°。

在仿真过程中, 选用另外 3 种过渡轨迹优化方法作比较, 分别为文献 [14] 中的定高过渡方法、文献 [15] 中的能量最优过渡方法及 PX4 原生固件中的过渡方法。选择上述 3 种过渡方法是因为: 文献 [14] 中的定高过渡方法具有出色的定高效果; 文献 [15] 中的能量最优过渡方法在兼顾无人机高度变化要求和能量消耗要求方面表现突出, 是比较实用的一种过渡方法; PX4 原生固件中的过渡方法被广泛使用且十分简单易懂, 可以作为其他更复杂过渡方法的对照组。具体而言, 文献 [14] 方法的优异定高效果得益于将飞机高度变化作为过渡轨迹优化问题的唯一指标, 并在最大高度变化值的基础上, 引入飞机的高度变化均值, 从而形成如下代价函数:

$$j=\frac{1}{K}\sum_{b=1}^{K}\left(h_b-h_{\mathrm{d},b}\right)^2+\max\left(h-h_{\mathrm{d}}\right) \tag{26}$$

式中: K 为高度数据采样点个数; h_{d} 为期望高度; h_b 为第 b 个采样点高度; $h_{\mathrm{d},b}$ 为第 b 个采样点期望高度。文献 [15] 能兼顾无人机高度变化要求和能量消耗要求, 是因为同时考虑了能量消耗指标和俯仰角跟踪指标, 形成了如下代价函数:

$$j=\int_{t_0}^{t_{\mathrm{F}}}\left[\left(\frac{F_{\mathrm{t}}(t)}{F_{\mathrm{t,max}}}\right)^2+\eta\left(\frac{\theta(t)-\theta_{\mathrm{c}}(t)}{\pi}\right)^2\right]\mathrm{d}t \tag{27}$$

式中: F_{t} 为实际推力; $F_{\mathrm{t,max}}$ 为飞机的理论最大推力; θ_{c} 为飞机期望俯仰角; η 为俯仰角跟踪权重系数; t_{F} 为过渡结束时间。虽然文献 [15] 没有直接将高度变化指标引入代价函数中, 但其通过将高度变化和俯仰角变化进行映射, 得到了能够实现较小高度变化的期望俯仰角 θ_{c}。PX4 原生固件中的过渡方法通过设置一个固定的推力输入和一个由经验确定的线性变化的俯仰角设定值引导飞机完成过渡。

6.3　仿真结果与分析

图 13 和图 14 分别为飞机在前向和后向过渡时的高度、俯仰角、空速、舵偏角及推力变化情况。表 3 总结了用于分析过渡性能的关键性指标: 过渡耗时 t_{trans}、高度变化最大值 $\Delta h_{\max}$、推力最大百分比 $T_{\max}$ 及舵面偏转最大百分比 $\delta_{\max}$。

前向过渡的仿真结果表明, 4 种过渡方法均能确保飞机安全完成前向过渡。PX4 原生固件中的过渡方法能引导飞机以最短时间完成前向过渡 (t_{trans}=2.15 s), 但会导致最大的高度变化 ($\Delta h_{\max}$=7.60 m) 和最大的作动器输出 ($\delta_{\max}$=99.8%)。文献 [14] 过渡方法高度变化最小 ($\Delta h_{\max}$=1.63 m), 且作动器输出最小 ($\delta_{\max}$=29.1%), 但完成前向过渡耗时最长 (t_{trans}=3.31 s)。经对比, 文献 [15] 过渡方法和本文过渡方法表现一致, 无论是最大爬升高度还是过渡耗时, 均介于另外 2 种过渡方法之间, 且产生的作动器输出也较小。虽然本文过渡方法在高度控制上不及文献 [15] 过渡方法, 但过渡时间略有缩短。

后向过渡的仿真结果表明了 4 种过渡方法在后向过渡过程中的有效性。其中, PX4 原生固件中的过渡方法高度变化最大 ($\Delta h_{\max}$=7.51 m), 过渡耗时最长 (t_{trans}=3.51 s)。文献 [14] 过渡方法高度变化最小 ($\Delta h_{\max}$=2.13 m), 且过渡耗时最短 (t_{trans}=0.80 s), 但作动器出现饱和现象 ($\delta_{\max}$=99.8%)。这是由于在后向过渡初始阶段, 飞机的速度较快, 为了减小高度变化, 文献 [14] 过渡方法需要快速增大迎角来减

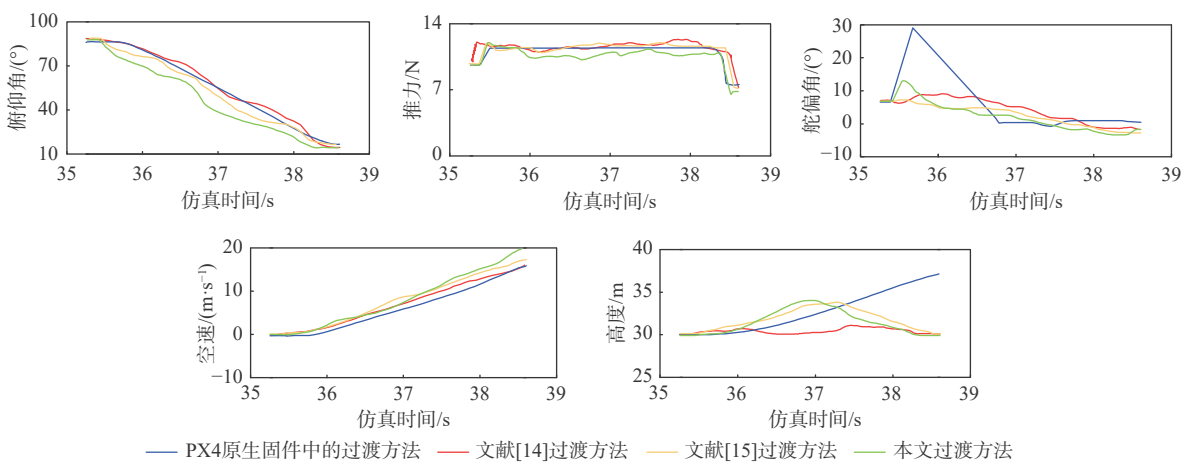


图 13　4 种过渡方法在前向过渡硬件在环仿真时的表现

Fig. 13　Performance of four different transition methods during front transition simulation

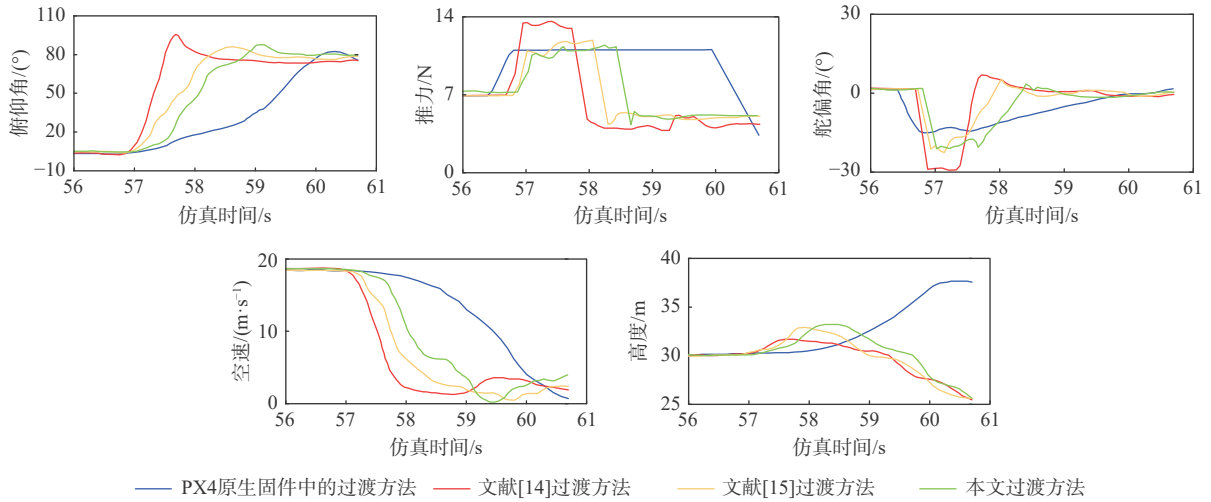


图 14 4种过渡方法在后向过渡硬件在环仿真时的表现

Fig. 14 Performance of four different transition methods during back transition simulation

表 3 4种过渡方法的过渡指标对比

Table 3 Comparison of transition indicators for four transition methods

过渡阶段	过渡方法	t_{trans}/s		Δh_{max}/m		T_{max}/%		δ_{max}/%	
		硬件在环仿真	实际飞行	硬件在环仿真	实际飞行	硬件在环仿真	实际飞行	硬件在环仿真	实际飞行
前向过渡	PX4原生固件中的过渡方法	2.15	1.80	7.60	7.21	85.7	96.4	99.8	93.4
	文献[14]过渡方法	3.31	3.31	1.63	1.33	80.2	82.1	29.1	13.6
	文献[15]过渡方法	3.09	2.46	3.87	3.38	82.6	84.2	32.6	14.5
	本文过渡方法	2.88	3.18	3.93	4.09	83.9	87.1	46.7	15.0
后向过渡	PX4原生固件中的过渡方法	3.51	5.08	7.51	13.80	78.3	97.1	48.9	66.7
	文献[14]过渡方法	0.80	1.32	2.13	2.19	99.6	99.8	99.8	99.7
	文献[15]过渡方法	1.29	2.57	2.69	4.04	84.5	92.8	87.1	60.3
	本文过渡方法	1.81	2.96	2.71	5.01	82.1	91.1	83.6	65.9

小空速在x^B轴上的分量，但快速增大迎角会导致飞机的升力迅速减小，故需要快速增大推力来平衡重力。本文过渡方法和文献[15]过渡方法在后向过渡中同样具有相似的表现，既不会出现高度增加过大的情况，也不会出现电机和作动器出现饱和的情况。

综上所述，PX4原生固件中的过渡方法在前/后向过渡时会导致最大的飞行高度变化和作动器输出，文献[14]过渡方法在前向过渡时过渡耗时最长，在后向过渡时出现作动器饱和现象，本文过渡方法既不会出现高度增加过大的情况，也不会出现作动器饱和的情况，且过渡耗时适中。由于较大的高度变化不利于飞行安全，会导致更多的能量消耗；较长的过渡时间会增加飞机失速的风险，不利于控制器性能的发挥；较大的作动器输出会削弱了飞机抵抗外部干扰的能力。因此，从飞行安全、能耗、失速风险及抗干扰能力等角度分析可知，PX4原生固件中的过渡方法和文献[14]过渡方法在实际飞行过程的可行性较低，而本文过渡方法的可行性最大。

图15为4种过渡方法下的过渡轨迹在过渡走廊中的位置。在前向过渡时4条过渡轨迹均位于前向过渡走廊内部，但PX4原生固件中的过渡方法的过渡轨迹有一部分接近走廊的上边界。在后向过渡时，PX4原生固件中的过渡方法的部分过渡轨迹超出了后向过渡走廊。为了更直观地反映过渡轨迹与目标过渡轨迹的误差，定义误差因子如下：

$$Q_e = \frac{1}{100}\sum_{r=0}^{20}(\theta_r(V)-\theta_r^*(V))^2 \tag{28}$$

式中：r为采样点；$\theta_r(V)$为实际过渡轨迹在采样点下的俯仰角；$\theta_r^*(V)$为目标过渡轨迹在采样点下的俯仰角。表4为4种过渡方法的误差因子。可知，无论是前向过渡还是后向过渡，本文过渡方法的误差因子均最小，说明本文过渡方法产生的实际过渡轨迹的安全裕度最大，即提高了过渡轨迹的鲁棒性。

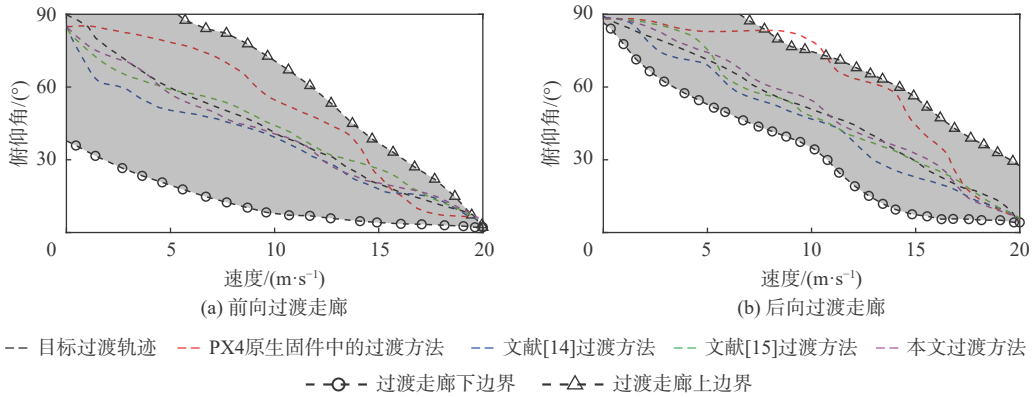


图 15　4 种过渡方法的过渡轨迹在过渡走廊中的位置

Fig. 15　Trajectories for four transition methods shown in the transition corridors

表 4　4 种过渡方法下的误差因子对比

Table 4　Comparison of error factors for four transition methods

过渡阶段	过渡方法	误差因子	
		硬件在环仿真	实际飞行
前向过渡	PX4原生固件中的过渡方法	34.287 0	39.632 0
	文献[14]过渡方法	15.039 0	18.662 0
	文献[15]过渡方法	9.474 4	10.006 0
	本文过渡方法	2.227 5	3.178 0
后向过渡	PX4原生固件中的过渡方法	40.602 0	48.776 0
	文献[14]过渡方法	18.773 0	22.003 0
	文献[15]过渡方法	9.949 0	13.356 0
	本文过渡方法	4.003 1	5.227 1

注: 粗体表示最优值。

7　飞行试验

进行实际飞行测试以进一步验证所提过渡方法的有效性, 如图 16 所示。实际飞行演示视频见网址 https://b23.tv/E3agJwl。实际飞行结果如图 17 和图 18 所示, 表 3 为过渡指标的数值。实际飞行结果与仿真结果一致, 在前向过渡过程中, PX4 原生固件中的过渡方法的过渡耗时最短(t_{trans}=1.80 s), 但高度变化最大(Δh_{max}=7.21 m), 作动器输出最大(δ_{max}=93.4%)。文献 [14] 过渡方法高度变化最小(Δh_{max}=1.33 m), 作动器输出最小(δ_{max}=13.6%), 但过渡耗时最长(t_{trans}=3.31 s)。文献 [15] 过渡方法和本文过渡方法无论是最大爬升高度还是过渡耗时, 均介于另外 2 种方法之间, 并且产生的作动器输出较小。在后向过渡过程中, PX4 原生固件中的过渡方法过渡耗时最长(t_{trans}=5.08 s), 高度变化最大(Δh_{max}=13.80 m)。文献 [14] 过渡方法过渡耗时最短(t_{trans}=1.32 s), 高度变化最小(Δh_{max}=2.19 m), 但出现了作动器饱和的现象(δ_{max}=99.7%)。文献 [15] 过渡方法和本文过渡方法在后向过渡时既不会出现高度增加过大的情况, 也不会出现作动器饱和的情况。在进行后向过渡实际飞行时, 仅文献 [14] 过渡方法在飞行最后阶段出现了反向倾斜的现象(俯仰角接近 100°)。这是由于在后向过渡结束时, 飞机仍有较大的水平速度, 此时飞机需要依靠反向倾斜产生反向力, 从而减小水平飞行速度。相比之下, 采用另外 3 种过渡方法的飞机可以通过爬升高度将多余的动能转化为势能, 需求的反向力也相应降低, 导致反向倾斜角度更小。

综上所述, PX4 原生固件中的过渡方法在前/后向过渡时会导致最大的飞行高度变化和作动器输出, 文献 [14] 过渡方法在前向过渡时过渡耗时最长, 在后向过渡时出现作动器饱和现象, 本文过渡方法既不会出现高度增加过大的情况, 也不会出现作动器饱和的情况, 且过渡耗时适中。由于较大的高度变化不利于飞行安全, 会导致更多的能量消耗, 较长的过渡耗时会增加飞机失速的风险, 不利

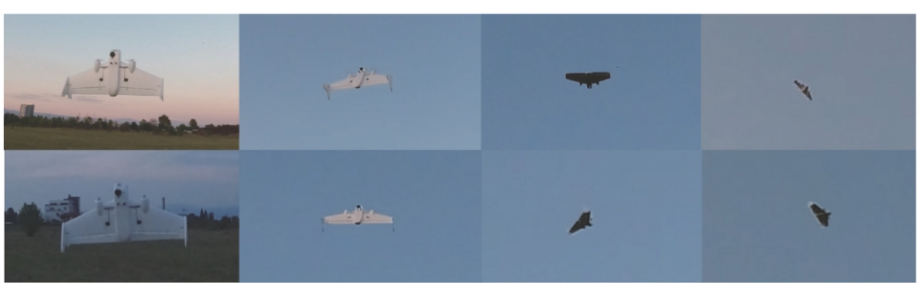


图 16　试飞中的尾座式垂直起降无人机

Fig. 16　Photographs of the test tail-sitter vertical take-off and landing UAV in flight tests

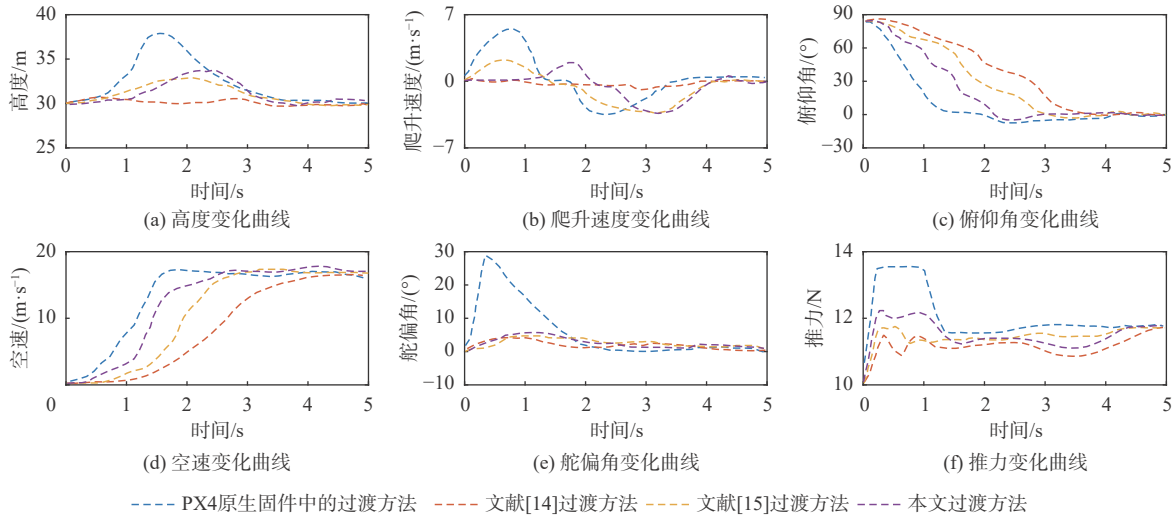


图 17 4 种过渡方法在前向过渡时的表现

Fig. 17 Performance of four different transition methods during front transition

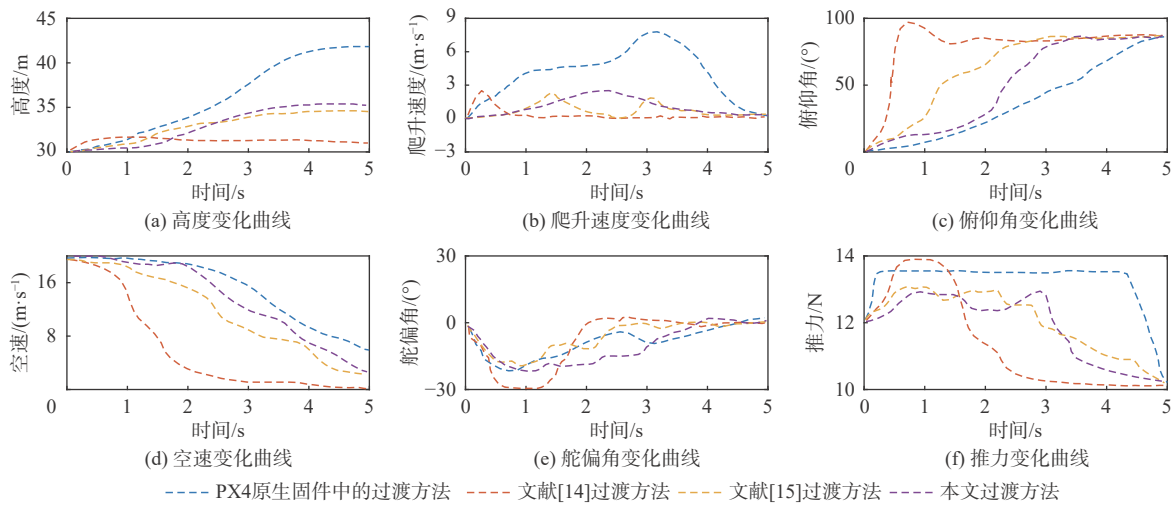


图 18 4 种过渡方法在后向过渡时的表现

Fig. 18 Performance of four different transition methods during back transition

于控制器性能的发挥，较大的作动器输出会削弱飞机抵抗外部干扰的能力，因此，从飞行安全、能耗、失速风险及抗干扰能力等角度分析可知，PX4 原生固件中的过渡方法和文献 [14] 过渡方法在实际飞行过程的可行性较低，而本文过渡方法的可行性最大。

根据 4 种过渡方法的实际前/后向过渡轨迹在前/后向过渡走廊中的位置（见图 19）可知，除 PX4 原生固件中的过渡方法外，其他 3 种过渡方法的前/后向过渡轨迹均位于过渡走廊的内部。表 4 为 4 种过渡方法的实际飞行过渡轨迹误差因子，可知，

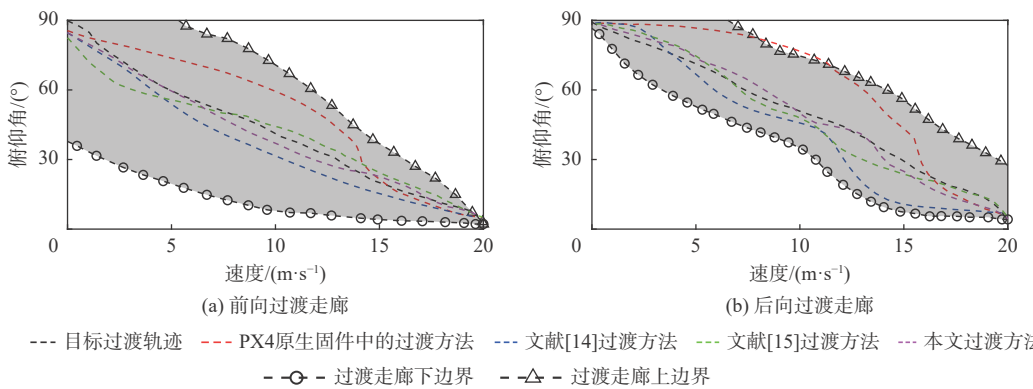


图 19 4 种过渡方法的实际过渡轨迹在过渡走廊中的位置

Fig. 19 Practical trajectories for four different transition methods shown in the transition corridor

无论是前向过渡还是后向过渡, 本文过渡方法的误差因子值均最小, 说明本文过渡方法产生的实际过渡轨迹的安全裕度最大, 即提高了过渡轨迹的鲁棒性。

8 结　论

1) PX4 原生固件中的过渡方法在前/后向过渡时会导致最大的飞行高度变化和作动器输出, 文献 [14] 过渡方法在前向过渡时过渡耗时最长, 在后向过渡时出现作动器饱和现象, 本文过渡方法既不会出现高度增加过大的情况, 也不会出现作动器饱和的情况, 且过渡耗时适中。由于较大的高度变化不利于飞行安全, 会导致更多的能量消耗, 较长的过渡时间会增加飞机失速的风险, 不利于控制器性能的发挥, 较大的作动器输出会削弱飞机抵抗外部干扰的能力, 因此, 从飞行安全、能耗、失速风险及抗干扰能力等角度分析可知, PX4 原生固件中的过渡方法和文献 [14] 过渡方法在实际飞行过程的可行性较低, 而本文过渡方法的可行性最大。

2) 无论是前向过渡还是后向过渡, 本文过渡方法相比其他 3 种方法, 减小了实际过渡轨迹与具有最大安全裕度的目标过渡轨迹之间的误差因子, 说明本文过渡方法产生的实际过渡轨迹的安全裕度最大, 即相比其他 3 种方法, 本文过渡方法提高了过渡轨迹的鲁棒性。

参考文献 (References)

[1] 刘志豪, 闵荣, 方成, 等. 多飞行模式垂直起降无人机过渡飞行控制策略[J]. 上海交通大学学报, 2019, 53(10): 1173-1181.
LIU Z H, MIN R, FANG C, et al. Transition flight control strategy of multiple flight mode vertical take-off and landing unmanned aerial vehicle[J]. Journal of Shanghai Jiao Tong University, 2019, 53(10): 1173-1181(in Chinese).

[2] 王子安, 龚正, 陈永亮, 等. 混合动力复合翼应急迫降在线航迹规划与制导[J]. 航空学报, 2019, 40(10): 323105.
WANG Z A, GONG Z, CHEN Y L, et al. Online path planning and guidance for hybrid VTOL UAVs forced landing[J]. Acta Aeronautica et Astronautica Sinica, 2019, 40(10): 323105(in Chinese).

[3] TAYYAB M, XIA Y Q, ZHAI D H, et al. Trajectory tracking control of a VTOL unmanned aerial vehicle using offset-free tracking MPC[J]. Chinese Journal of Aeronautics, 2020, 33(7): 2024-2042.

[4] 张飞, 王云, 谭锟. 复合式垂直起降固定翼无人机旋翼和机翼的干扰分析[J]. 航空工程进展, 2019, 10(6): 810-816.
ZHANG F, WANG Y, TAN K. Interference analysis of rotor and wing of compound vertical takeoff and landing fixed wing UAV[J]. Advances in Aeronautical Science and Engineering, 2019, 10(6): 810-816(in Chinese).

[5] 刘泽宇, 招启军, 张夏阳, 等. 倾转机翼无人倾转旋翼机飞行动力学稳定性分析[J]. 飞行力学, 2021, 39(3): 1-7.
LIU Z Y, ZHAO Q J, ZHANG X Y, et al. Analysis of flight dynamics stability of unmanned tilt-rotor aircraft with tilting wings[J]. Flight Dynamics, 2021, 39(3): 1-7(in Chinese).

[6] 刘德元, 刘昊, LEWIS F L. 尾座式无人飞行器鲁棒容错编队控制[J]. 航空学报, 2021, 42(2): 324296.
LIU D Y, LIU H, LEWIS F L. Robust fault-tolerant formation control for tail-sitters[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(2): 324296(in Chinese).

[7] KUBO D, SUZUKI S. Robust optimal autopilot design for a tail-sitter unmanned aerial vehicle[J]. Journal of Aerospace Computing, Information, and Communication, 2008, 5(5): 135-154.

[8] LIU D Y, LIU H, LI Z Y, et al. Robust attitude control for tail-sitter unmanned aerial vehicles in flight mode transitions[J]. International Journal of Robust and Nonlinear Control, 2019, 29(4): 1132-1149.

[9] BENREZKI R R, TAYEBI A, TADJINE M. Adaptive trajectory tracking control for VTOL-UAVs with unknown inertia, gyro-bias, and actuator LOE[J]. International Journal of Robust and Nonlinear Control, 2018, 28(17): 5247-5261.

[10] JOHNSON E N, WU A, NEIDHOEFER J C, et al. Flight-test results of autonomous airplane transitions between steady-level and hovering flight[J]. Journal of Guidance, Control, and Dynamics, 2008, 31(2): 358-370.

[11] ITASSE M, MOSCHETTA J M, AMEHO Y, et al. Equilibrium transition study for a hybrid MAV[J]. International Journal of Micro Air Vehicles, 2011, 3(4): 229-245.

[12] ÇAKICI F, LEBLEBICIOĞLU M K. Design and analysis of a mode-switching micro unmanned aerial vehicle[J]. International Journal of Micro Air Vehicles, 2016, 8(4): 221-229.

[13] SAGHAFI F, BANAZADEH A. Investigation on the flight characteristics of a conceptual fluidic thrust-vectored aerial tail-sitter[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2007, 221(5): 741-755.

[14] YANG Y J, ZHU J H, WANG X Y, et al. Dynamic transition corridors and control strategy of a rotor-blown-wing tail-sitter[J]. Journal of Guidance, Control, and Dynamics, 2021, 44(10): 1836-1852.

[15] LI B Y, SUN J X, ZHOU W F, et al. Transition optimization for a VTOL tail-sitter UAV[J]. IEEE/ASME Transactions on Mechatronics, 2020, 25(5): 2534-2545.

[16] SUN J X, LI B Y, WEN C Y, et al. Design and implementation of a real-time hardware-in-the-loop testing platform for a dual-rotor tail-sitter unmanned aerial vehicle[J]. Mechatronics, 2018, 56: 1-15.

[17] 夏济宇, 周洲, 王正平, 等. 倾转动力无人机三维过渡走廊[J]. 北京航空航天大学学报, 2024, 50(3): 886-895.
XIA J Y, ZHOU Z, WANG Z P, et al. The research on three-dimensional transition corridor of tilt-propulsion UAV[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(3): 886-895(in Chinese).

[18] 曹芸芸. 倾转旋翼飞行器飞行动力学数学建模方法研究[D]. 南京: 南京航空航天大学, 2012.
CAO Y Y. Study on mathematical modeling method of flight dynamics of tilt-rotor aircraft[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012(in Chinese).

[19] 俞志明, 陈仁良, 孔卫红. 倾转四旋翼飞行器倾转过渡走廊分析

方法[J]. 北京航空航天大学学报, 2020, 46(11): 2106-2113.
YU Z M, CHEN R L, KONG W H. Analysis method for conversion corridor of quad tilt rotor aircraft[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(11): 2106-2113 (in Chinese).

[20] 樊有容, 王皓, 杨乐, 等. 基于过渡走廊的倾转旋翼无人机纵向控制研究[J]. 飞行力学, 2022, 40(2): 61-66.
FAN Y R, WANG H, YANG L, et al. Research on longitudinal control of tilt-rotor UAV based on transition corridor[J]. Flight Dynamics, 2022, 40(2): 61-66(in Chinese).

[21] KANG Y, PARK B, YOO C, et al. Flight test results of automatic tilt control for small scaled tilt rotor aircraft[C]//Proceedings of the International Conference on Control, Automation and Systems. Piscataway: IEEE Press, 2008: 47-51.

[22] MURAOKA K, OKADA N, KUBO D, et al. Transition flight of quad tilt wing VTOL UAV[C]//Proceedings of the 28th Congress of the International Council of the Aeronautical Sciences. Bonn: ICAS, 2012, 4: 3242-3251.

[23] 严旭飞, 陈仁良. 倾转旋翼机动态倾转过渡过程的操纵策略优化[J]. 航空学报, 2017, 38(7): 520865.
YAN X F, CHEN R L. Control strategy optimization of dynamic conversion procedure of tilt-rotor aircraft[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(7): 520865(in Chinese).

[24] KIM C J, SUNG S, PARK S H, et al. Numerical time-scale separation for rotorcraft nonlinear optimal control analyses[J]. Journal of Guidance, Control, and Dynamics, 2014, 37(2): 658-673.

[25] KELLY M. An introduction to trajectory optimization: how to do your own direct collocation[J]. SIAM Review, 2017, 59(4): 849-904.

[26] BANAZADEH A, TAYMOURTASH N. Optimal control of an aerial tail sitter in transition flight phases[J]. Journal of Aircraft, 2016, 53(4): 914-921.

[27] MAQSOOD A, GO T H. Optimization of hover-to-cruise transition maneuver using variable-incidence wing[J]. Journal of Aircraft, 2010, 47(3): 1060-1064.

[28] NALDI R, MARCONI L. Optimal transition maneuvers for a class of V/STOL aircraft[J]. Automatica, 2011, 47(5): 870-879.

附录 A

俯仰角和角速度控制器表达形式为

$$U_{\delta}^{\mathrm{PD}} = -I_{yy}\left(k_{\theta}e_{\theta} + k_{q}e_{q}\right)/a_{M\delta} \tag{A1}$$

式中：e_{θ}和e_{q}分别为俯仰角和角速度的跟踪误差；k_{θ}和k_{q}为正增益系数；$a_{M\delta}$为

$$a_{M\delta} = \sum_{n=1}^{6}\frac{1}{2}\rho v_{n}^{2}\bar{c}S_{\mathrm{e},n}C_{M\delta} \tag{A2}$$

为了提高系统的动态响应，使用舵面最优控制输入U_{δ}作为前馈控制输入，则舵面控制输入U_{δ}^{F}为

$$U_{\delta}^{\mathrm{F}} = U_{\delta} + U_{\delta}^{\mathrm{PD}} \tag{A3}$$

高度控制器表达形式为

$$U_{h} = -(k_{h}e_{h} + k_{\dot{h}}e_{\dot{h}})/a_{zT} \tag{A4}$$

式中：e_{h}和$e_{\dot{h}}$分别为高度和高度变化速度的跟踪误差；k_{h}和$k_{\dot{h}}$为正增益系数；a_{zT}为

$$a_{zT} = -\frac{2\sin\theta}{m} \tag{A5}$$

速度控制器的表达形式为

$$U_{v} = -a_{VT}^{-1}k_{u}e_{u} \tag{A6}$$

式中：e_{u}为速度跟踪误差；k_{u}为正增益系数；a_{VT}=2/m。

螺旋桨的控制输入由U_{h}和U_{v}经过控制分配得到。在俯仰角较大时，螺旋桨产生的推力主要用于高度控制，在俯仰角较小时，螺旋桨产生的推力主要用于速度控制，因此，控制分配设置为

$$U_{\mathrm{p}}^{\mathrm{FF}} = U_{h}\sin^{2}\theta + U_{v}(1-\sin^{2}\theta) \tag{A7}$$

使用螺旋桨最优控制输入U_{p}作为前馈控制输入，则螺旋桨的最终控制输入$U_{\mathrm{p}}^{\mathrm{F}}$为

$$U_{\mathrm{p}}^{\mathrm{F}} = U_{\mathrm{p}} + U_{\mathrm{p}}^{\mathrm{FF}} \tag{A8}$$

Optimization method of transition trajectory for tail-sitter unmanned aerial vehicles

ZOU Xu[1], LIU Zhenbao[1, 2], ZHAO Wen[1, *], WANG Lina[1]

(1. School of Civilaviation, Northwestern Polytechnical University, Xi'an 710072, China;
2. Shenzhen Research Institute, Northwestern Polytechnical University, Shenzhen 518057, China)

Abstract: Most transition trajectories obtained by traditional optimization methods are not feasible in the actual flight process. Moreover, the robustness of these transition trajectories is poor due to not considering the retention of sufficient actuator margin. In order to solve this problem, a transition corridor-based transition method for tail-sitter unmanned aerial vehicle (UAV) is proposed. Considering a type of dual-rotor tail-sitter UAV, a nonlinear dynamic model is constructed by distinguishing the aerodynamic characteristics of the wing in the propeller airflow region from those outside the region. Inspired by the research on transition corridors for tilt-rotor aircraft, a transition corridor is designed. The feasibility of the transition corridor is improved by limiting the range of climb rate and pitch angle rate. A trajectory located inside the transition corridor with the maximum safety margin is chosen as the target transition trajectory by the optimization method. Then, the transition process of the tail-sitter UAV is regarded as a trajectory optimization problem. By solving the trajectory optimization problem, the optimal transition trajectory which is closest to the target transition trajectory and retains sufficient actuator margin is obtained. Finally, based on the designed transition control framework, hardware-in-the-loop simulation experiments and real flight tests are conducted. Simulation results are consistent with the real flight results, which prove that the designed transition method can guide the aircraft to complete the transition quickly and safely while retaining a certain actuator margin, avoiding such unfavorable conditions for the transition flight as excessive altitude increase, long transition time and actuator saturation.

Keywords: tail-sitter unmanned aerial vehicles; dynamic model; optimization method; transition corridor; transition trajectory

Received: 2023-07-12; **Accepted**: 2023-10-13; **Published Online**: 2023-11-30 17:04

URL: link.cnki.net/urlid/11.2625.V.20231130.1113.001

Foundation items: National Natural Science Foundation of China (52072309); Key Research and Development Program of Shaanxi (2019ZDLGY14-02-01); Natural Science Basic Research Program of Shaanxi (2021JQ-086); Shenzhen Fundamental Research Program (JCYJ20190806152203506); Aeronautical Science Foundation of China (ASFC-2018ZC53026)

*** Corresponding author**. E-mail: zhaowen@nwpu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0486

长时间气动力/热作用下累积热变形演化规律

刘深深[1,2]，蒋波[2]，韩青华[2]，余婧[2]，杨肖峰[1,2]，魏东[2]，朱言旦[2,*]，桂业伟[2]

(1. 中国空气动力研究与发展中心 空天飞行空气动力科学与技术全国重点实验室，绵阳 621000；2. 中国空气动力研究与发展中心计算空气动力研究所，绵阳 621000)

摘 要：长时间气动力/热作用下产生的累积热变形及其在整个飞行历程中引发的不利影响是未来高超声速巡航飞行器设计不可忽视的问题。针对高超声速飞行器典型翼面结构，以自研热环境/热响应耦合计算分析平台为基础，结合自适应时间步的双向耦合计算策略建立全数值的气动力/热/结构多场耦合累积热变形预测方法；在此基础上开展翼面结构在长时间气动力/热耦合作用下的累积热变形演化规律及形成机理研究，并分析其在时间变化历程中对气动特性的影响。研究结果表明，由于翼面结构所受气动力、气动加热、结构传热、变形响应等物理过程的时空尺度差异，累积热变形呈现第1发展、第2发展、充分发展等非线性特征突出的3个演化阶段，并且各种因素需要经过较长的时间才能充分发展并实现累积变形的稳定。上述累积热变形演化行为引发了伴随整个飞行历程的气动特性的非线性变化，进一步带来升力下降、升阻比下降及俯仰力矩偏差等不利影响。相关不利影响亟需在未来长航时高超声速飞行器设计中加以考虑并主动应对。

关 键 词：长航时；机翼结构；累积热变形；气动力/热耦合作用；热变形影响

中图分类号：V221+.3; TB553

文献标志码：A **文章编号**：1001-5965(2025)09-3086-12

当前高超声速飞行器前沿探索研究向着“更快、更远、更经济”的方向发展，以高超声速飞机、两级入轨飞行器、空天飞机等为代表的长航时可重复使用的飞行器概念不断涌现，正在成为新一代高超声速飞行器的研究热点。这些飞行器通常需要采用吸气式发动机为动力，并且以马赫数大于6的速度在稠密大气层中进行小时级的长时间飞行，因此，相较于以往飞行器的一个典型特征是气动力/热严酷作用时间更长，同时，该类飞行器构型多采用大尺度细长的升力体构型[1]，且由于结构质量限制，机身及各翼/舵面通常广泛采用轻质材料和大型薄壁结构设计，具备更强的柔性特征[2-4]。大尺度柔性薄壁结构在长时间气动加热作用下会产生巨大的非均匀温升，进而引发材料的“软化”及热应力[5-6]，在气动力载荷叠加作用下将会带来更加明显的热变形。区别于气动力带来的弹性变形，力/热共同作用下的变形将会随着持续的气动加热和结构传热而缓慢变化，进而与飞行过程相耦合，因而呈现时间上的累积效应，本文将其称之为累积热变形。预测较大的累积热变形的实现将会显著改变飞行器的气动外形及其特性，且该变化将持续整个飞行时间过程，并对飞行器的最终效能和任务执行产生影响[4,7]。由于高超声速飞行器的最优飞行区间较狭窄且对飞行状态和外形变化高度敏感，因此，对于变形问题不仅要关心最终的稳定状态，还需要关心力/热作用下变形的变化过程，以及整个变化过程对气动特性的影响。高超声速飞行器在飞行中既需要最终稳定状态的高效飞行，也需要整个变化过程

收稿日期：2023-07-27；**录用日期**：2023-12-01；**网络出版时间**：2024-04-03 12:28

网络出版地址：link.cnki.net/urlid/11.2625.V.20240402.1828.002

基金项目：高超声速冲压发动机技术重点实验室基金(2021-JCJQ-LB-020-11)；国家自然科学基金(12002361)；国家重点研发计划(2019YFA0405202)；智强基金

* **通信作者**. E-mail：zhupai@mail.ustc.edu.cn

引用格式：*刘深深，蒋波，韩青华，等. 长时间气动力/热作用下累积热变形演化规律[J]. 北京航空航天大学学报，2025，51(9)：3086-3097.*
LIU S S，JIANG B，HAN Q H，et al. Cumulative thermal deformation evolution under prelonged aerodynamic and thermal loads[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3086-3097 (in Chinese).

中飞行姿态的稳定可控, 因此, 有必要研究气动力/热长时间耦合作用下结构变形随时间的演化规律及其对气动特性的影响。

气动力/热作用下的累积热变形是典型的静热气动弹性问题。热气动弹性问题最早由 Roger[5] 提出, 当时 X-15 验证机在初期飞行中曾发生垂尾颤振, 这使人们意识到高超声速飞行下的热气动弹性问题研究的重要性, 随后国内外的研究多集中于壁板颤振、翼/舵翼面等结构的动热气动弹性问题, 重点关注气动加热效应及产生的热应力对结构动力学稳定性和颤振边界的影响[2,8-10], 对于静热气动弹性问题的研究相对较少。近些年来, 随着新一代高超声速飞行器的研制需求, 静热气动弹性问题开始逐渐受到关注。叶正寅团队[2] 综述文章指出, 在动热气动弹性研究之外, 对于大长细比飞行器, 静热气动弹性问题将成为影响飞行器配平等整体气动特性不可忽视的因素。李国曙等[11] 针对高超声速翼面开展的研究表明, 热应力导致的结构静变形与气动力引起的变形相比不可忽略。Zhang 等[12] 采用工程快速气动力/热计算方法针对类 X-20、X-43 飞行器开展的研究发现, 考虑热效应后飞行器最大变形量增大了约 1 倍, 引起了较大的升力系数和舵效偏差。刘磊等[13] 基于工程快速气动力/热预测方法和有限元方法对高超声速机翼的研究分析表明, 巡航状态下机翼热变形会引起较大的升阻比下降。罗金玲[14] 和刘凯[15] 等针对细长体轴对称类飞行器结构开展了静热气动弹性分析, 发现如果忽略热的影响可能会导致变形规律的偏差。受限于计算资源制约和问题本身的复杂性, 上述研究大多进行了简化, 多采用工程气动力/热计算方法开展, 并且多未关注气动加热和变形随时间的变化历程。

对于考虑时间历程的热变形问题研究, 叶坤[16] 基于数值求解器和松耦合的时间策略对高超声速翼面结构开展了 50 s 过程的静热变形研究, 表明力/热共同作用下的热变形会导致机翼外形及扭转角随时间演化, 对气动特性产生较大的影响。受限于计算资源的限制, 上述研究累积热变形尚未达到稳态, 但已经揭示了变形在短时间内存在的演化过程。而对于这种演化过程在更长的数千秒的飞行过程中将如何变化、在多大的时间尺度上达到稳定及产生何种影响尚缺乏研究。

得益于高性能计算集群、动网格技术和高效数据传递方法的不断发展, 采用全数值的气动力/热/结构多场耦合计算方法开展长时间过程的累积热变形分析正在逐步成为可能。针对上述问题, 本文选取典型的高超声速飞行器翼面结构, 以高精度数值 CFD 气动力/热环境、CTD 瞬态热传导计算及 CSD 结构静力学分析求解器为基础, 结合自适应时间步耦合计算策略, 建立长航时静热气动弹性全数值气动力/热/结构多场耦合分析方法, 并对该方法进行考核验证; 在此基础上开展翼面结构在长时间气动力/热共同作用下的累积热变形演化规律研究, 并对力/热在时间特征尺度下各自的作用机理进行探讨分析, 同时, 分析长时飞行过程中累积变形对升力、阻力、升阻比等气动力特性和力矩特性的影响, 并给出相关的规律性结论。

1　耦合分析方法及验证

1.1　累积热变形各物理场分析方法

累积热变形分析中, 各物理场的求解器采用团队自研的热环境/热响应耦合计算分析平台 FL-CAPTER[17]。该分析平台具备数值气动力/热模拟、结构传热模拟、结构应力/应变模拟等功能模块, 可开展单独模块的计算或不同物理场间的耦合计算, 软件各求解器经过考核验证及实际的工程应用[18-20]。各物理场具体计算方法中, 气动力/热环境通过求解三维非定常可压缩 Navier-Stokes 方程获得, 计算方法为有限体积法; 结构温度场通过求解三维瞬态热传导方程获得, 计算方法为有限体积法; 结构应力/应变通过有限元数值求解结构静力学方程得到, 具体见文献 [17]。

1.2　耦合计算策略

耦合计算分析采用双向耦合的计算策略, 计算过程中既考虑气动力/热环境对结构的影响, 也考虑结构变形及结构温升对气动力/热环境带来的影响。耦合方法中, 计算框架和各物理场数据传递流程如图 1 所示。t_0为起始时刻, $t_0+\Delta t$为下一时刻。其具体过程为:

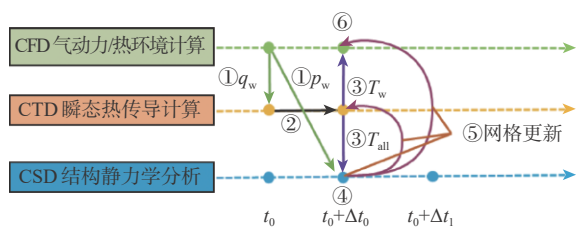


图 1　耦合计算框架[18]

Fig. 1　Framework of coupling simulation[18]

1) 开始t_0时刻的流场计算。首先, 通过求解准定常三维可压缩 Navier-Stokes 方程计算得到冷壁条件下的气动力/热环境, 获得结构表面的热流数据q_w与压力数据p_w, 将q_w传递至瞬态热传导计算, p_w传递到结构静力学分析, 流场/固体结构交界面采用基于物理量梯度修正的紧支径向基函数方法[21]

进行数据交互，其他物理量采用同样的数据传递方法进行传递。

2）结构温度场采用无内热源的三维热传导方程进行数值求解分析，得到$t_0+\Delta t_0$时刻的结构温度场分布。

3）将$t_0+\Delta t_0$时刻的结构表面温度T_w传递至流场网格，更新壁面温度条件。同时，将整个结构场的温度T_{all}传递给CSD求解器，根据结构温度场分布进行$t_0+\Delta t_0$时刻的应力/应变计算。

4）基于固体结构热应力/应变场，求解考虑热膨胀效应的三维弹性力学方程，得到结构$t_0+\Delta t_0$时刻的热变形。

5）采用基于径向基函数的动网格技术[22]，根据结构表面位移更新流场计算网格，得到$t_0+\Delta t_0$时刻的流场计算网格。同时根据结构每一点的变形，将其与原始CSD和CTD计算网格点坐标进行叠加，同步更新结构传热和应力/应变计算网格。

6）根据新的流场网格和壁面温度条件，求解准定常三维可压缩Navier-Stokes方程，开始新一轮计算。

在整个耦合计算中，采用自适应的耦合时间步策略[23]，Δt_0初始时刻取各物理场最小的时间尺度，将其设置为1×10^{-4} s量级，在计算过程中随着温度升高和热变形逐步趋于稳定，耦合步长逐步变大，其每轮迭代的时间步长Δt_n为

$$\Delta t_n=\Delta t_0\times(1+b)^n \tag{1}$$

式中：b取值为0.1。同时对Δt_n的最大值进行限制（令$\max(\Delta t_n)\leqslant\Delta T_{max}$），取$\Delta T_{max}$=2 s。图2为不同耦合时间步数下的时间步长变化情况。该自适应时间步将逐步从结构的特征时间尺度过渡到热传导的时间尺度，以保证多场耦合计算早期以更小的时间步获得更精准的计算结果，同时在气动加热过程逐步稳定的后期获得更高的计算效率。

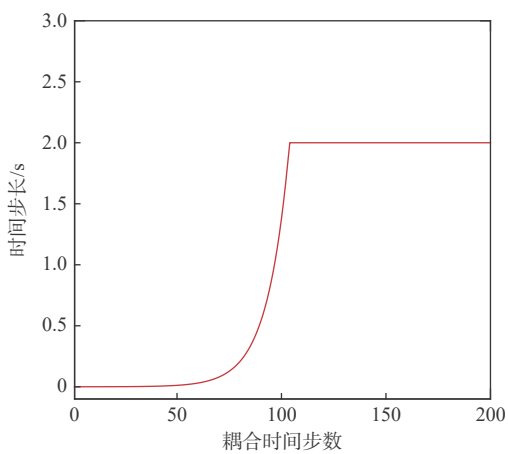


图2 自适应时间步耦合计算策略

Fig. 2 Adaptive time step coupling computing strategy

1.3 计算方法的验证

由于气动力/热共同作用下的变形试验数据较为匮乏，通过对累积热变形计算中涉及的静气动弹性变形及热应力预测、气动加热-传热-结构耦合分别进行验证的方法对计算方法的可靠性进行验证。其中悬臂梁算例用来验证在力载荷作用下单场结构变形计算的可靠性；球形热应力算例用来验证传热过程及热应力计算的可靠性；二维圆柱算例用来验证耦合计算方法在进行气动加热-传热-结构三场耦合计算的可靠性。

1.3.1 悬臂梁力载荷变形算例

悬臂梁在均匀受力载荷时，其变形存在解析解。假定悬臂梁长度为200 mm，高度为10 mm，上表面均匀压力为0.5 MPa，一端固支，材料选择标准钢，弹性模量E为209 GPa，泊松比为0.3，根据材料力学理论分析其挠曲线方程可以得到梁变形的解析解[24]。图3为梁的变形值与解析解的对比。可以看出，本文计算方法与理论解析解的结果一致性非常好，位移计算最大误差为2.8%，验证了本文方法在求解力载荷作用下变形的可靠性。

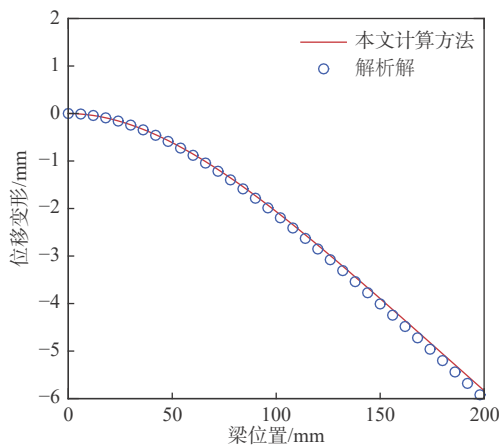


图3 本文计算变形与解析解对比

Fig. 3 Comparison of present displacement with analytical solution

1.3.2 球形热应力算例

热应力计算的准确性是热变形计算的基础，文献[25]对各向同性的半径为r_0均质球体施加恒温边界从初始温度T_0到终止温度T_e，给出了不同时刻球体各部分的温度$T(r,t)$、径向热应力$\sigma_r(r,t)$及周向热应力$\sigma_c(r,t)$的解析解，本文采用此算例验证随传热过程进行的热应力计算的可靠性。假定材料密度ρ为4 000 kg/m^3，比热容c为0.25 kJ/(kg·K)，导热率κ为0.02 kW/(m·K)，球半径r_0为0.3 m，弹性模量E为4 000 MPa，泊松比为0.2，线膨胀系数α为2.5×10^{-5}/K，假定初始温度T_0为1 273 K，恒温边界终止温度T_e为273 K，计算时间为40 s。

计算获得的不同时刻径向热应力和周向热应力分布与解析解的对比分别如图 4 和图 5 所示。可以看出, 本文方法热应力计算最大误差在 5% 以内, 表明了本文方法计算热应力的可靠性。

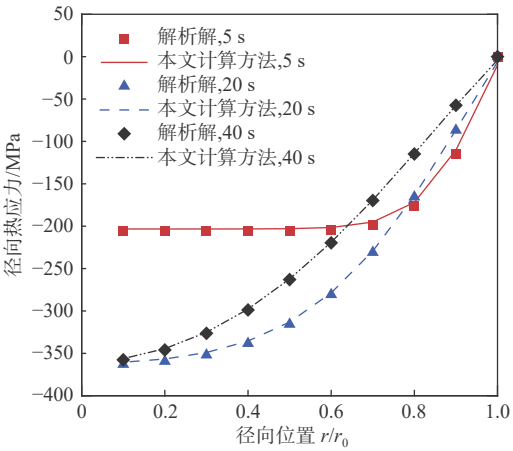


图 4　本文计算径向热应力与解析解对比

Fig. 4　Comparison of present radial thermal stress with analytical solution

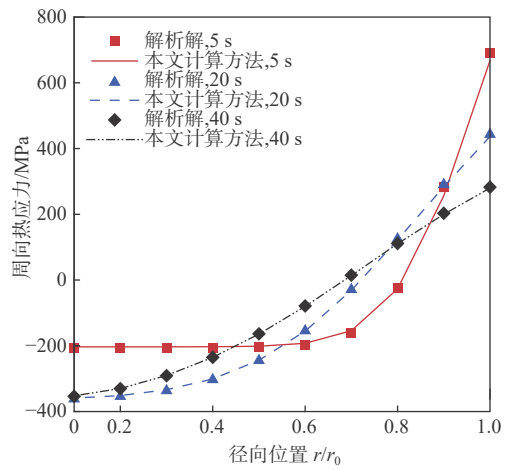


图 5　本文计算周向热应力与解析解对比

Fig. 5　Comparison of present circumferential thermal stress with analytical solution

1.3.3　二维圆柱气动加热-传热-结构耦合算例

Wieting[26] 在 NASA 兰利研究中心开展的高超声速空气气流绕不锈钢圆管的气动加热实验被很多研究用于气动对结构加热的验证算例, 本文选择该算例并结合文献 [26] 的试验数据和文献 [27] 的数值模拟结果进行气动加热-结构热传导耦合计算的验证。

圆管内、外直径分别为 50.8 mm、76.2 mm。圆管展向相对较长, 可近似认为无限长。来流马赫数为 6.47, 来流温度为 241.5 K, 来流压力为 648.1 Pa, 结构初始温度为 294.4 K。以圆管外直径为参考量的来流雷诺数 Re=1×10^5, 故计算状态为层流。计算网格如图 6 所示。流场域和结构域均采用结构网格, 流场域网格数量为 121×61, 结构域网格数量为 61×30, 并向第 3 个方向延拓生成三维网格进行计算。计算采用标准不锈钢的热力学参数, 发射率为 0.6。在初始结构温度 294.4 K 的条件下计算得到的驻点热流密度为 490.5 kW/m^2, 文献 [27] 数值计算给出的驻点热流密度为 483 kW/m^2, 两者相差 1.5%, 吻合较好, 而试验中驻点热流密度附近数据散布较大, 计算热流密度位于试验数据散布区间中。图 7 和图 8 分别为以驻点压力和热流密度进行无量纲化后的表面压力、热流密度与试验结果的对比曲线。可以看出, 初始时刻压力和热流密度的分布都与试验数据吻合较好。

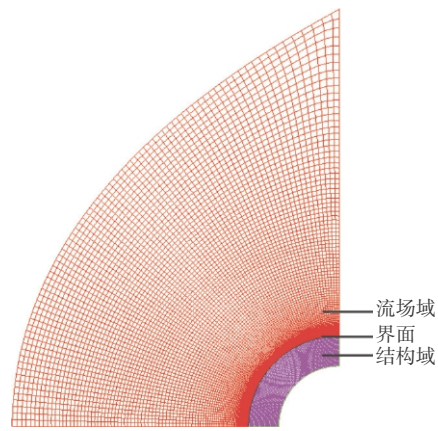


图 6　计算网格示意图

Fig. 6　The grid mesh of calculation

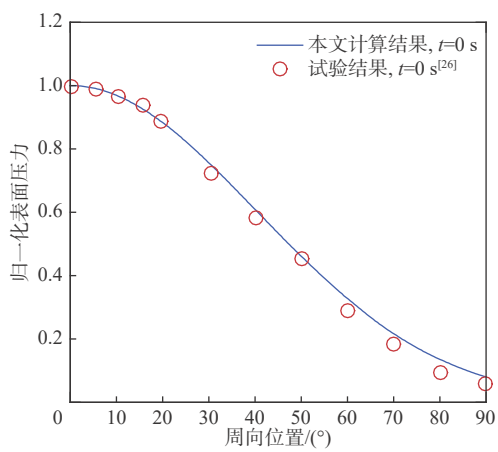


图 7　表面压力数据对比

Fig. 7　Comparison of wall pressure data

图 9(a)为 t=2 s 时本文计算的结构内部温度分布与文献 [27] 数值模拟得到的温度分布对比(图中°R 为兰氏温度单位)。可以看出, 本文计算结果无论定性或定量, 均与文献结果吻合较好。图 9(b)为 t=2 s 时本文计算得到的周向热应力分布。2 s 时驻点的周向热应力大小为 220.96 MPa, 比文献 [27] 的 220.8 MPa 偏高 0.07%, 两者吻合良好, 表明本文所采用的耦合计算方法进行气动加热-热传导-结构

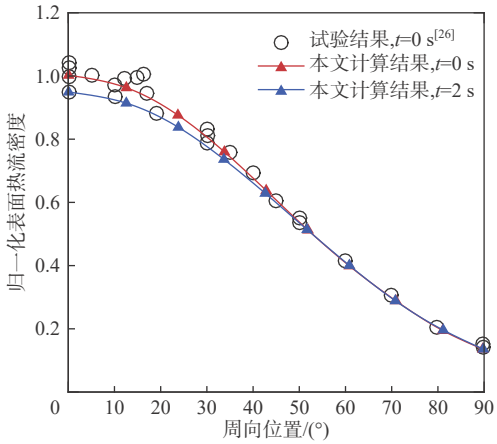


图 8 表面热流密度数据对比

Fig. 8 Comparison of wall heat flux data

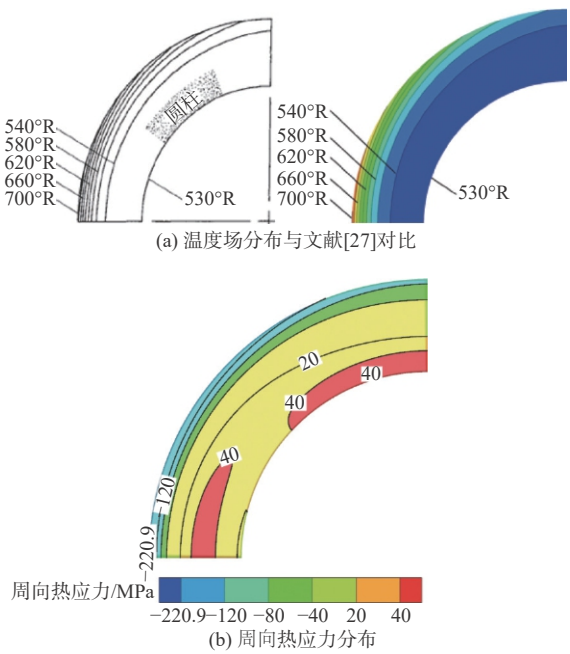


图 9 t=2 s 本文计算结果

Fig. 9 The present calculation result at t=2 s

三场耦合计算时的有效性。

2 计算模型及设置

选取典型的高超声速飞行器翼面构型开展结构在长时间气动力/热共同作用下的累积热变形相关问题研究。该模型具有大后掠角和小展弦比的基本特点。模型具体几何参数如图 10 所示。翼根长为 3.55 m, 翼展为 1.25 m, 前缘后掠角为 70°, 板厚 15 mm, 前缘和后缘半径均为 7.5 mm。计算材料使用镍基高温合金 GH1015, 具体材料属性如表 1 和表 2 所示。考虑到未来的长航时飞行器多以马赫数大于 6 的速度飞行, 本文选取的飞行速度为马赫数等于 6, 高度为 20 km, 飞行迎角选择翼面结构最大升阻比对应的 4°迎角。

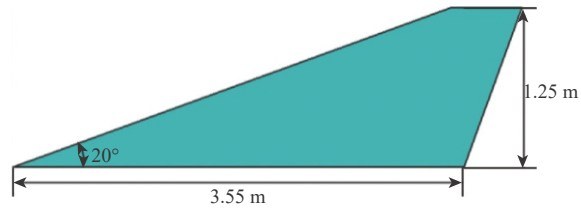


图 10 高超声速飞行器机翼模型

Fig. 10 Hypersonic wing model

表 1 高温合金 GH1015 材料物理性能

Table 1 Material physical properties of high temperature alloy GH1015

性能	数值
密度/(kg·m^{-3})	8 320
泊松比	0.3
表面辐射系数	0.8

表 2 高温合金 GH1015 材料热物理性能

Table 2 Thermal-physical properties of high temperature alloy GH1015

温度/K	导热系数/(W·(m·K)$^{-1}$)	比热容/(J·(kg·K)$^{-1}$)	弹性模量/GPa	热膨胀系数/10^{-6} K^{-1}
273	9.9	440	170	14.0
373	12.0	450	166	14.4
473	13.5	475	158	14.7
573	15.5	495	152	15.0
673	17.3	505	142	15.5
773	19.0	530	135	15.8
873	21.0	555	128	16.1
973	22.8	575	115	16.4
1 073	24.9	615	108	16.7
1 173	26.7	650	94	17.0
1 273	28.6	710	78	17.3

在 CFD 数值模拟中, 采用同一套网格进行气动力/热的数值计算, 同时, 结构传热和结构应力/应变分析也采用同一套网格进行计算, 流场域网格(见图 11)总量约为 603 392, 结构域网格(见图 12)总量为 49 020, 在边界层内进行法向加密处理, 网格第 1 层间距y^+=8×10^{-7} m, 网格第 1 层间距满足y^+<1, 上述网格已开展网格无关性验证。采用的边界条件设置: 远场为自由来流边界, 初始壁面温度为 300 K, 考虑到数值计算来流雷诺数为 1.11×10^7/m, 计算中采用 SST 湍流模型。温度场计算时考虑热辐射影响, 外形有气流流过的面为对流加热边界, 机翼根部设定固支边界, 其他部分均为绝热边界, 整个累积热变形耦合计算过程采用 1.2 节计算方法, 耦合计算时长为 1 000 s。计算采用 Intel Xeon E5 2692 处理器, 主频为 2.2 GHz, 单处理器 12 核 24 线程, 计算核数为 64 核, 计算耗时约 25 天。

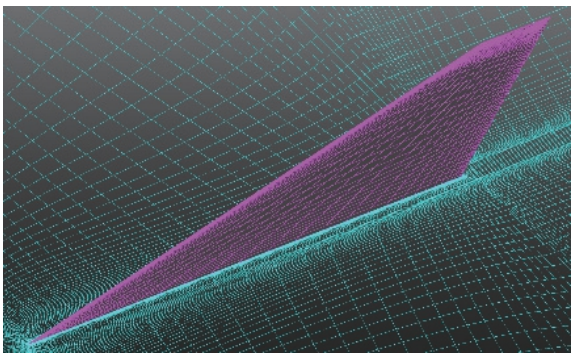


图 11　高超声速飞行器机翼气动力/热计算网格

Fig. 11　The aerodynamic force and heating calculation grid for wing of hypersonic aircraft

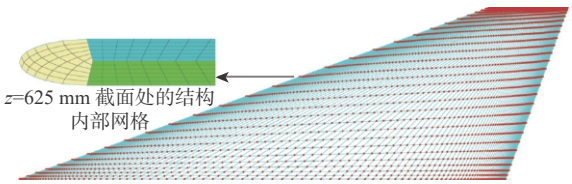


图 12　翼面结构模型有限元网格

Fig. 12　The finite element grid for wing structure

3　计算结果分析

3.1　初始流场特征

图 13 和图 14 为初始时刻的表面气动力/热情况。在当前飞行状态下, 从翼根到翼稍前缘热流密度 Q 由 1 800 kW/m^2 逐步过渡到 1 200 kW/m^2, 迎风面大面积区域热流密度约为 128～512 kW/m^2, 背风面热流密度基本在 100 kW/m^2 以内。迎风面压力系数 C_p 方面, 翼前缘在 0.2～0.24 之间, 大面积区域范围在 0.02～0.08 之间, 背风面压力系数多数小于 0。因此, 初始时刻迎背风气动力/热环境均差异较大。

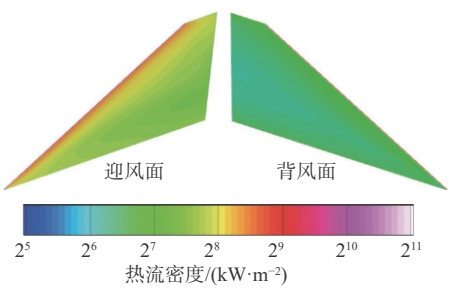


图 13　初始时刻热流分布云图

Fig. 13　Contour of heating rate distribution at initial moment

3.2　结构气动加热过程分析

图 15 为机翼结构温度随时间的变化情况。可以看出, 随时间增加, 迎风面和背风面温度均逐步升高, 结构迎风面温度稍高于背风面, 500 s 后温度趋于稳定。

图 16 为 3 个监测点的温度 T 和热壁热流密度 Q 的变化情况。可以看出, 温度整体呈现早期快速升高后逐步变缓的宏观特征。随着结构温度的升高, 热壁热流密度逐步降低, 迎背风温差同步呈现

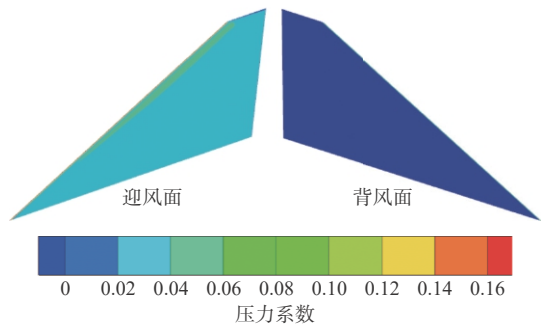


图 14　初始时刻表面压力分布云图

Fig. 14　Contour of surface pressure distribution at initial moment

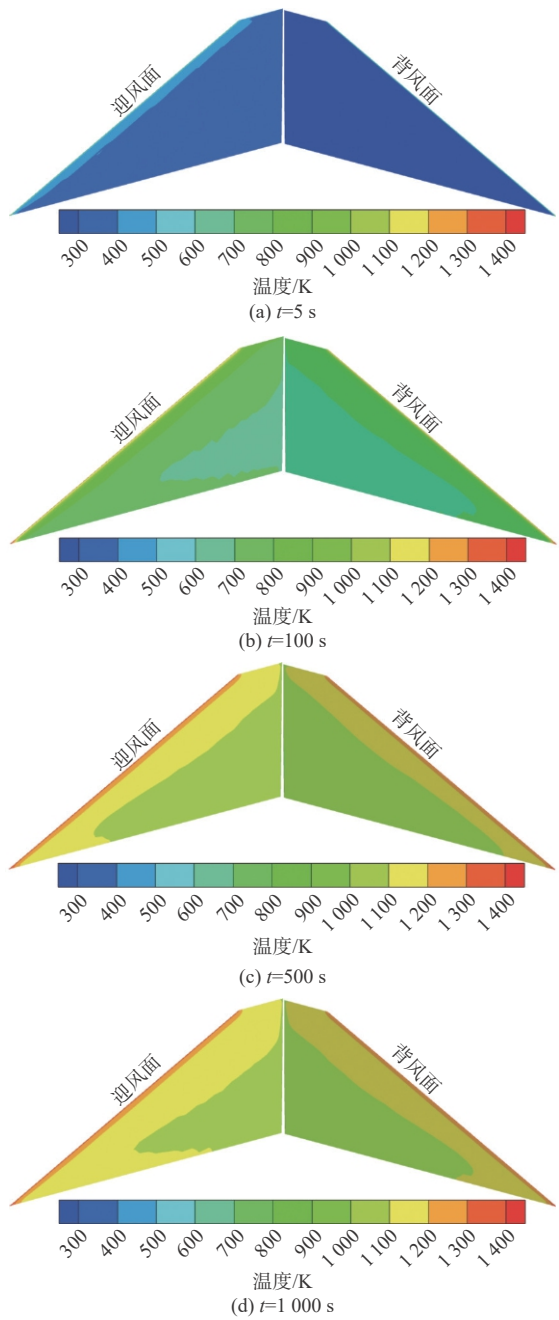


图 15　温度场随时间变化

Fig. 15　Variation of the temperature field with time

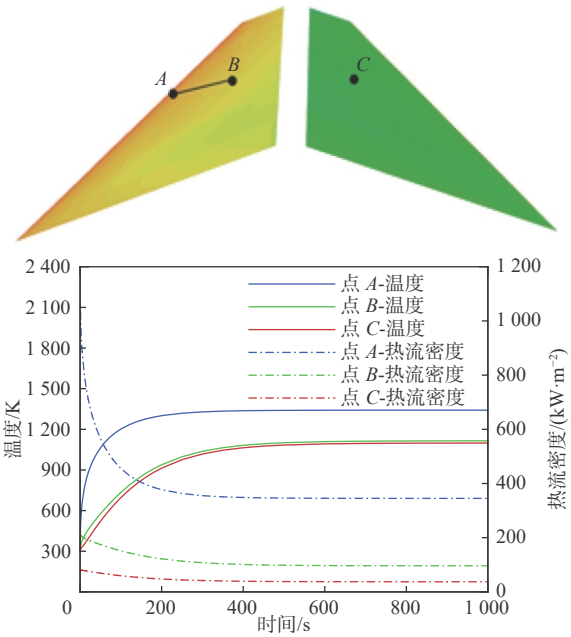


图 16 监测点位置以及温度和热流密度随时间的变化

Fig. 16 Monitor points and variation of the temperature and heating rate for monitor points with time

先迅速增大再逐渐减小后趋于稳定的趋势。

3.3 累积热变形量值随时间变化规律

在气动力/热共同作用下，翼面结构将发生随时间变化的累积热变形，图 17 为 1 000 s 时累积热变形不同方向分量的分布云图。可以看出，翼面结构

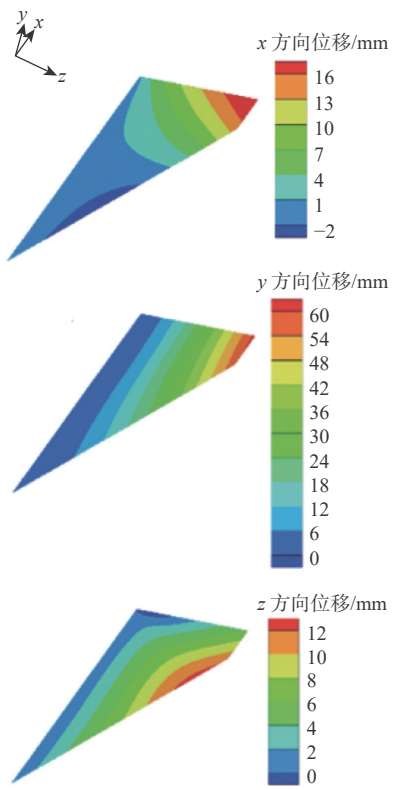


图 17 1 000 s 时的累积热变形分布

Fig. 17 The accumulated thermal deformation distribution of 1 000 s

主要发生 x 和 z 方向的膨胀以及 y 方向的上翘 3 个维度的变形。其中，y 方向的变形显著大于 x 和 z 方向的膨胀变形量，因此，重点分析 y 方向变形的演化规律。在分布特征中，y 方向变形量沿展向逐步增加，而 x 方向的膨胀集中于翼面后缘和翼梢，z 方向膨胀集中于翼面前缘和翼梢。

图 18 为不同时刻翼面变形与原始结构（灰色）的对比（y 方向变形进行了 10 倍放大）。可以看出，随着时间历程的进行，累积热变形呈现较为复杂的动态变化过程，在 t=100 s 之前变形持续增加且增速较快，而后在 t=300～1 000 s 过程中变形缓慢增加。从 t=100 s 和 300 s 的对比可看出，翼梢前后缘的变形增量存在大小量值的交替变化。

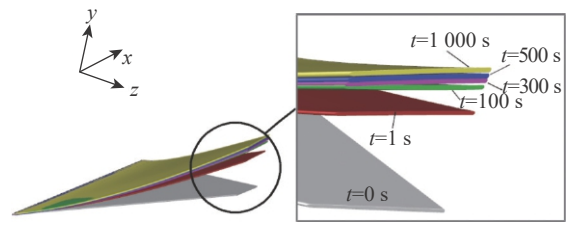


图 18 不同时刻的累积热变形对比

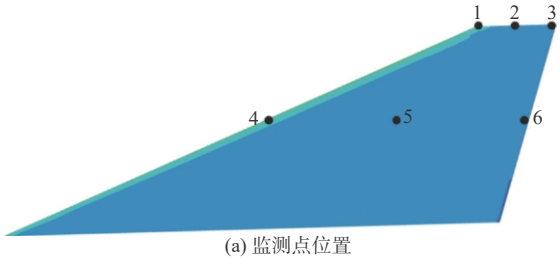
Fig. 18 The accumulated thermal deformation at different time

图 19 为机翼迎风面不同监测点处变形量随时间的变化情况。可以看出，x 和 z 方向的膨胀变形随时间增加逐步增大并趋于稳定。而从图 19(c)可以看出，y 方向累积热变形的变化存在快速增大并回落和缓慢增大 2 个大的时间尺度特征，为了更好地对不同时间尺度内变形量值的变化规律进行分析，图 20 给出了 y 方向变形在对数坐标时间轴下的变化规律。

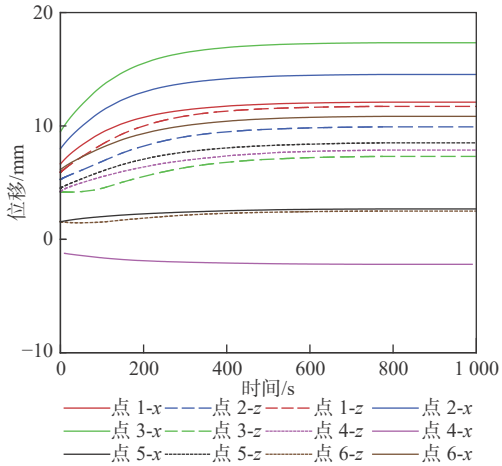
从图 20 可进一步看出，整体累积热变形在时间上呈现 3 个阶段的变化。计算开始时，气动力开始加载导致翼面变形，随后，变形导致气动力环境发生变化，变形小幅回落收敛，在 10^{-3} s 时基本稳定，这种冲击响应在真实飞行中并不存在，主要是由于耦合计算的虚拟起始效应导致；随着气动加热的作用，在 10^{-3}～2×10^{1} s 的时间尺度变形迅速增大；在 10^{1}～3×10^{2} s 时间尺度上变形量增长变缓，翼前缘监测点 1 和 4 甚至出现变形量回落；随后在 3×10^{2}～10^{3} s 时间尺度上变形量再次缓慢增大直至稳定。值得注意的是，点 1 和点 4 气动加热较为严重，其累积热变形增大和回落程度比其他点更明显，同时，还可以看到点 1 和点 3 及点 4 和点 5 的变形量大小存在交替变化，因此，热变形的量值变化除引起整体上翘之外，还导致翼面扭转角随时间发生变化。

3.4 弯曲和扭转随时间的变化规律

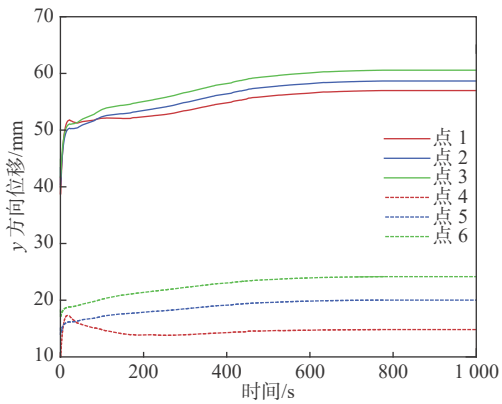
定义监测点 2 与翼根中点连线和 xz 平面的夹角为翼面的整体弯曲角度，则可以得到翼面整体弯



(a) 监测点位置



(b) 监测点 x 和 z 方向膨胀变形随时间变化规律



(c) 监测点 y 方向变形随时间变化规律

图 19　监测点累积热变形随时间的变化规律

Fig. 19　Changes of accumulated heat deformation with time at monitoring points

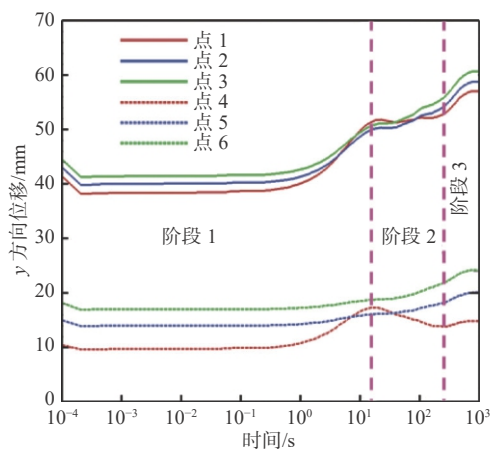


图 20　监测点y方向热变形随时间变化规律(对数坐标)

Fig. 20　Variation of thermal deformation in y direction for monitor points with time (in logarithmic coordinate system)

曲角度随时间的变化规律, 如图 21 所示。可以看出, 翼面整体弯曲角度在飞行早期迅速增大, 随后增加缓慢, 600 s 后趋于稳定, 最终稳定在 2.68° 左右。

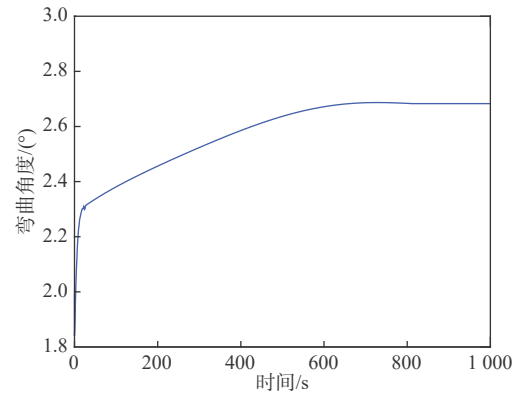


图 21　翼面整体弯曲角度随时间的变化规律

Fig. 21　Variation of bending angle of wing structure with time

图 22 为展长 z=1 250 mm 和 z=625 mm 处扭转角随时间的变化情况。可以明显看出, 在变形早期, 翼面主要产生负的扭转角, 随着累积热变形的快速增长, 扭转角逐步减小且由负变正, 在 20 s 后达到正的极值, 此后随时间的增加, 扭转角由正变负且逐步增大, 在 600 s 后逐步稳定。

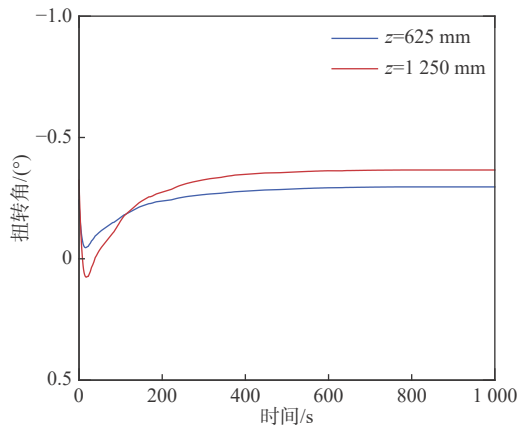


图 22　典型展长处扭转角随时间的变化

Fig. 22　Variation of torsion angle with time at typical span

3.5　累积热变形规律成因分析

分别对仅有气动力加载、仅有气动热加载的翼面结构变形规律进行计算分析, 并将其同两者共同作用下的变形规律进行对比分析(取图 19(a)中监测点 1、点 2 和点 3 为代表进行分析), 如图 23 所示。可以看出, 在单独的气动力作用下, 其变形量具备瞬时加载特征, 经过初始的几个时间步振荡后不再随时间变化, 且单独气动力作用会带来恒定的负扭转角。而单独气动热作用产生的变形呈现先迅速变大后缓慢回落的特征, 且单独气动热作用会导致正的扭转角, 因此, 力/热变形下变形的动态演

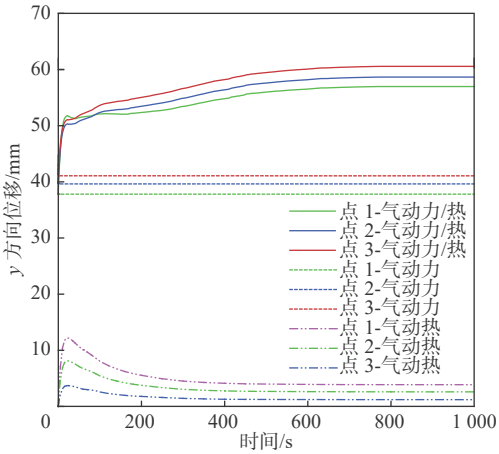


图 23 不同分析方法下的y方向位移

Fig. 23 The deformations in y direction for different analysis methods

化特征主要是由于气动加热引起的热应力的变化所导致。还可以看到，稳定状态力/热共同作用下导致的变形量大于气动力和气动热单独导致的变形量之和。

y方向热变形主要取决于上下表面的温差产生的热应力，图 24 为迎背风面对应点的温差随时间的变化。可以看出，上下表面温差的走势与热应力引起的热变形规律基本一致，表明上下表面温差达到局部峰值又回落的传热物理特征导致了力/热共同作用下热变形第 1 个局部峰值及增长率拐点的出现。同时考虑到由于温度的升高，长时间气动加热下弹性模量降低了约 50%，这种软化效应将会导致变形量进一步加大，这也解释了最终平衡状态的热变形大于单独气动力作用和气动热作用之和的原因。

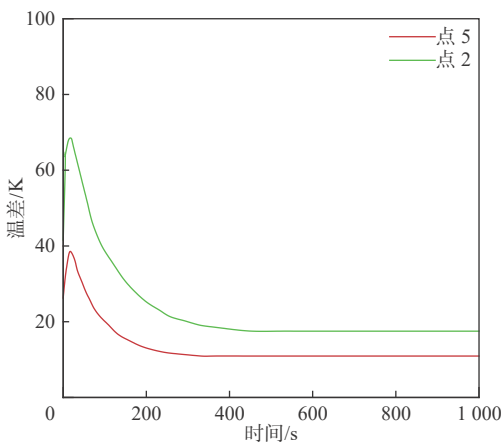


图 24 温差变化情况

Fig. 24 Variation of temperature difference

综上，在力/热的共同作用下，翼面结构整体累积变形主要体现为 3 个阶段的变化特征：

1）第 1 发展阶段（冷结构气动加热作用为主）：在热传导过程的早期，背风面升温较慢，迎风面升温较快，结构温度梯度快速增大达到极值，随之热应力和变形达到极值，此时，弯曲和扭转角变化剧烈，翼梢处扭转角由负变正，该特征时间尺度约为 10^1 s 量级，和热传导扩散过程时间一致。

2）第 2 发展阶段（热结构气动加热作用为主）：随着热传导继续进行，背风面温度持续升高，温度梯度快速回落，导致热应力产生的变形增长变缓并回落，尤其是翼前缘变形回落幅度较大，导致翼稍部分扭转角由正逐步变负，而整体弯曲度持续增加，该特征时间尺度约为 10^2 s 量级，和热传导稳定时间一致。

3）充分发展阶段（力/热作用趋于平衡）：温度梯度趋于稳定，气动加热继续进行，材料的弹性模量进一步减小，“软化”效应进一步增强，因而气动力载荷导致的变形继续增大，最终结构温度和温度梯度均达到平衡，累积热变形趋于稳定，该作用具备最大的时间尺度，大于热传导稳定的时间尺度。

3.6 变形的演化对气动特性影响

图 25 为翼面压力分布随时间的变化情况。可以明显看出，在长时间的累积历程中，由于上翘变形，翼梢附近压力逐渐减小，在扭转和弯曲变化较为剧烈的早期，压力还存在一定的非线性变化，400 s 之后压力逐渐稳定。

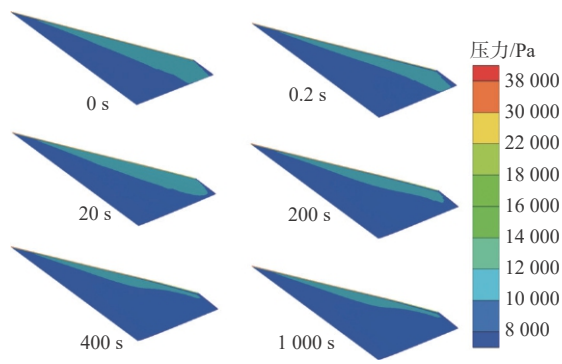
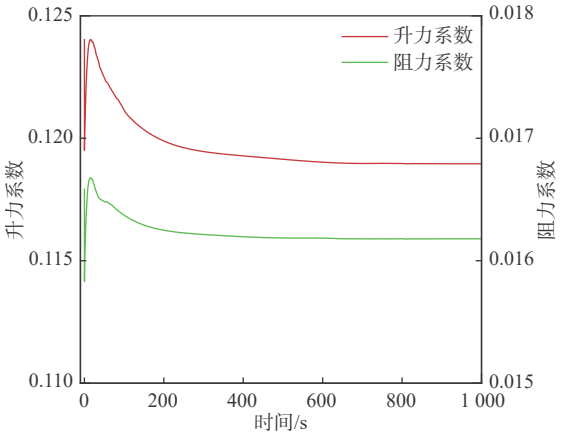


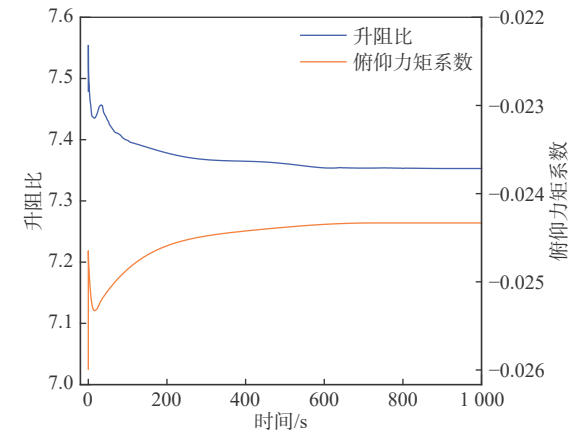
图 25 不同时刻压力分布变化

Fig. 25 Variation of pressure distribution at different time

图 26 为 1 000 s 时间飞行过程中，伴随着累积热变形的发生，翼面结构气动特性随时间的变化情况（参考面积为 1 m^2，参考长度为 3.5 m，俯仰力矩轴取为 1/2 弦线）。可以看出，由于累积热变形，整个气动特性随时间增加发生了非线性变化。升力、阻力和俯仰力矩随着初期翼面的上翘快速减小，且随着热变形引起的正扭转角的增加而快速增大，而后回落并快速减小，随后，随着热变形均匀地缓慢上翘而继续缓慢减小，最终稳定后，升力系数相比原始构型减小 4.09%，阻力系数减小 2.44%，俯仰力矩系数减小 6.39%，升阻比降低 1.7%，导致气动效



(a) 升力系数和阻力系数



(b) 升阻比和俯仰力矩系数

图 26　气动特性变化情况

Fig. 26　Variation of aerodynamic characteristics

率的明显下降和气动特性的显著偏离。除翼面上翘导致的等效迎角下降、整体升阻比及气动效率均降低之外, 在翼面早期快速扭转过程中, 升/阻力及俯仰力矩的变化也十分复杂且显著, 这将给初期巡航状态的维持和控制系统带来较大的压力。

从数据变化规律可以看出, 气动特性的变化会与飞行状态发生快/慢 2 种尺度的耦合影响, 机翼热变形快速扭转区对升阻特性和力矩的影响在短期内影响较大, 对飞行状态维持和控制有短期快尺度的影响; 而气动特性的缓慢变化及最终稳态的气动效率下降将会通过慢尺度的时间积分效应持续影响最终的航程。同时考虑到真实的结构多为空芯加筋结构, 相较于本文理想化的实心结构, 热变形和气动特性影响量将会进一步扩大, 达到热变形稳态的时间将会更长, 由此对控制和飞行效率的不利影响将会更进一步扩大, 需要在设计中提早进行考虑。

4　结　论

1) 典型翼面结构长时间气动加热作用下会持续发生变形, 达到稳定的时间通常需要数百秒, 这种变形效应在长时间的飞行过程中不可忽略。

2) 累积热变形的演化呈现快慢尺度叠加的 2 个特征, 具体可以分为以冷结构气动加热为主的第 1 发展阶段、以热结构气动加热为主的第 2 发展阶段和各因素共同作用的充分发展阶段。第 1 发展阶段热变形迅速增大, 第 2 发展阶段热变形增速缓慢乃至回落, 充分发展阶段变形缓慢增大直至稳定。

3) 累积热变形以上翘变形为主。上翘变形同时带来弯曲角度和扭转角随时间的演化, 弯曲角度随时间增加主要呈增加趋势, 而扭转角的变化则存在随累积热变形而变化的正负扭转角交替的现象。

4) 累积热变形造成了伴随整个时间历程的气动特性非线性变化, 气动特性的变化会与飞行状态发生快/慢 2 种尺度的耦合影响, 会给控制和长时间巡航性能造成较大影响, 需要在设计中予以考虑并主动应对。

本文基于全数值的多场耦合计算方法开展了长航时力/热作用下的翼面结构的累积热变形规律研究, 揭示了累积热变形随时间历程的非线性特性及其原因和影响。真实翼面结构通常会采用空芯加筋条的结构, 这将会导致热传导过程及非均匀热应力的进一步复杂化及更大的变形量, 进而产生更大更复杂的影响, 对于更真实结构下的机理规律揭示还需进一步研究。

参考文献（References）

[1] FRIEDMANN P, THURUTHIMATTAM B, MCNAMARA J, et al. Hypersonic aeroelasticity and aerothermoelasticity with application to reuseable launch vehicles[C]//Proceedings of the 12th AIAA International Space Planes and Hypersonic Systems and Technologies. Reston: AIAA, 2003: 7014.

[2] 叶正寅, 孟宪宗, 刘成, 等. 高超声速飞行器气动弹性的近期进展与发展展望[J]. 空气动力学学报, 2018, 36(6): 984-994.
YE Z Y, MENG X Z, LIU C, et al. Progress and prospects on aeroelasticity of hypersonic vehicles[J]. Acta Aerodynamica Sinica, 2018, 36(6): 984-994(in Chinese).

[3] 邹学锋, 潘凯, 燕群, 等. 多场耦合环境下高超声速飞行器结构动强度问题综述[J]. 航空科学技术, 2020, 31(12): 3-15.
ZOU X F, PAN K, YAN Q, et al. Overview of dynamic strength of hypersonic vehicle structure in multi-field coupling environment[J]. Aeronautical Science & Technology, 2020, 31(12): 3-15(in Chinese).

[4] 华如豪, 叶正寅. 吸气式高超声速飞行器多学科动力学建模[J]. 航空学报, 2015, 36(1): 346-356.
HUA R H, YE Z Y. Multidisciplinary dynamics modeling and analysis of a generic hypersonic vehicle[J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(1): 346-356(in Chinese).

[5] ROGER M. Aerothermoelasticity[J]. Aerospace Engineering, 1958, 17(10): 34-43.

[6] CARRICK I E. A survey of aerothermoelasticity[J]. Aerospace En-

gineering, 1963, 22(1): 140-147.

[7] 桂业伟. 高超声速飞行器综合热效应问题[J]. 中国科学: 物理学 力学 天文学, 2019, 49(11): 139-153.
GUI Y W. Combined thermal phenomena of hypersonic vehicle[J]. Scientia Sinica (Physica, Mechanica & Astronomica), 2019, 49(11): 139-153(in Chinese).

[8] MEI C, ABDEL-MOTAGALY K, CHEN R. Review of nonlinear panel flutter at supersonic and hypersonic speeds[J]. Applied Mechanics Reviews, 1999, 52(10): 321.

[9] 杨智春, 夏巍. 壁板颤振的分析模型、数值求解方法和研究进展[J]. 力学进展, 2010, 40(1): 81-98.
YANG Z C, XIA W. Analytical models, numerical solutions and advances in the study of panel flutter[J]. Advances in Mechanics, 2010, 40(1): 81-98(in Chinese).

[10] MCNAMARA J, FRIEDMANN P. Aeroelastic and aerothermoelastic analysis of hypersonic vehicles: current status and future trends[C]//Proceedings of the 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston: AIAA, 2007: 2013.

[11] 李国曙, 万志强, 杨超. 高超声速翼面气动热与静气动弹性综合分析[J]. 北京航空航天大学学报, 2012, 38(1): 53-58.
LI G S, WAN Z Q, YANG C. Integrated analysis of aerothermal-aeroelastic wings in hypersonic flow[J]. Journal of Beijing University of Aeronautics and Astronautics, 2012, 38(1): 53-58(in Chinese).

[12] ZHANG X, WANG Y, LI F, et al. Static aerothermoelasticity of hypersonic vehicles[C]//International Space Planes and Hypersonic Systems and Technologies Conferences. Reston: AIAA, 2017: 2367.

[13] 刘磊, 桂业伟, 耿湘人, 等. 高超声速飞行器热气弹静态问题研究[J]. 空气动力学学报, 2013, 31(5): 559-563.
LIU L, GUI Y W, GENG X R, et al. Study on static aerothermoelasticity for hypersonic vehicle[J]. Acta Aerodynamica Sinica, 2013, 31(5): 559-563(in Chinese).

[14] 罗金玲, 康宏琳, 操小龙. 吸气式高超声速飞行器空气动力学[M]. 北京: 科学出版社, 2021: 313-325.
LUO J L, KANG H L, CAO X L. Aerodynamics of air-breathing hypersonic vehicle[M]. Beijing: Science Press, 2021: 313-325(in Chinese).

[15] 刘凯, 许云涛. CFD 在高速飞行器热气弹问题中的应用[J]. 气体物理, 2019, 4(4): 50-55.
LIU K, XU Y T. Application of CFD in high speed aircraft thermal aeroelastic problems[J]. Physics of Gases, 2019, 4(4): 50-55(in Chinese).

[16] 叶坤. 吸气式高速飞行器关键热气动弹性问题研究[D]. 西安: 西北工业大学, 2018.
YE K. Study on key thermoelastic problems of air-breathing high-speed aircraft[D]. Xi'an: Northwestern Polytechnical University, 2018(in Chinese).

[17] 桂业伟, 刘磊, 代光月, 等. 高超声速飞行器流-热-固耦合研究现状与软件开发[J]. 航空学报, 2017, 38(7): 020844.
GUI Y W, LIU L, DAI G Y, et al. Research status of hypersonic vehicle fluid-thermal-solid coupling and software development[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(7): 020844(in Chinese).

[18] 桂业伟, 刘磊, 耿湘人, 等. 气动力/热与结构多场耦合计算策略与方法研究[J]. 工程热物理学报, 2015, 36(5): 1047-1051.
GUI Y W, LIU L, GENG X R, et al. Study on the computation strategy and method of aero-dynamic-thermal-structural coupling problem[J]. Journal of Engineering Thermophysics, 2015, 36(5): 1047-1051(in Chinese).

[19] 桂业伟, 刘磊, 杜雁霞. 热防护系统耦合分析方法与应用[J]. 现代防御技术, 2014, 42(4): 9-14.
GUI Y W, LIU L, DU Y X. Coupled analysis methods and applications of thermal protection system[J]. Modern Defence Technology, 2014, 42(4): 9-14(in Chinese).

[20] 刘磊, 桂业伟, 耿湘人, 等. 热气动弹性变形对飞行器结构温度场的影响研究[J]. 空气动力学学报, 2015, 33(1): 31-35.
LIU L, GUI Y W, GENG X R, et al. Study on the temperature field of hypersonic vehicle structure with aerothermoelasticity deformation[J]. Acta Aerodynamica Sinica, 2015, 33(1): 31-35(in Chinese).

[21] 刘智侃, 刘深深, 刘骁, 等. 基于物理量梯度修正的 RBF 数据传递方法[J]. 航空学报, 2021, 42(7): 124506.
LIU Z K, LIU S S, LIU X, et al. RBF data transfer based on physical gradient modification[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(7): 124506(in Chinese).

[22] 代光月. 高超声速气动/热/结构多场耦合方法及应用研究[D]. 绵阳: 中国空气动力研究与发展中心, 2016.
DAI G Y. Method and application research on the fluid-thermal-structural interaction problems in hypersonic flow[D]. Mianyang: China Aerodynamics Research and Development Center, 2016(in Chinese) .

[23] YANG X F, GUI Y W, TANG W, et al. Surface thermochemical effects on TPS-coupled aerothermodynamics in hypersonic Martian gas flow[J]. Acta Astronautica, 2018, 147: 445-453.

[24] 张少实. 新编材料力学[M]. 北京: 机械工业出版社, 2002: 124-127.
ZHANG S S. New material mechanics[M]. Beijing: China Machine Press, 2002: 124-127(in Chinese).

[25] 王小增, 杨久红, 曾辉. 大型轴承钢球热处理过程中瞬态温度和应力分布[J]. 热加工工艺, 2012, 41(20): 199-203.
WANG X Z, YANG J H, ZENG H. Transient temperature and stress distribution of large bearing steel ball during heat treatment process[J]. Hot Working Technology, 2012, 41(20): 199-203(in Chinese).

[26] WIETING A. Experimental study of shock wave interference heating on a cylindrical leading edge: NASA-TM-100484[R]. Washington, D. C.: NASA, 1987: 60-160.

[27] DECHAUMPHAI P, WIETING A R, THORNTON E A. Flow-thermal-structural study of aerodynamically heated leading edges [J]. Journal of Spacecraft and Rockets, 1989, 26(4): 201-209.

Cumulative thermal deformation evolution under prelonged aerodynamic and thermal loads

LIU Shenshen[1, 2], JIANG Bo[2], HAN Qinghua[2], YU Jing[2], YANG Xiaofeng[1, 2],
WEI Dong[2], ZHU Yandan[2, *], GUI Yewei[2]

(1. State Key Laboratory of Aerodynamics, China Aerodynamics Research and Development Center, Mianyang 621000, China;
2. Computational Aerodynamics Institute, China Aerodynamics Research and Development Center, Mianyang 621000, China)

Abstract: Cumulative thermal deformations during long periods of aerodynamic thermal coupling and their associated adverse effects are issues that cannot be ignored for future hypersonic aircrafts with long endurance. For typical structures of hypersonic wing structure, a fully numerical aerodynamic/thermal/structural multi-field coupling cumulative thermal deformation prediction method is established, based on in-house numerical solvers, and in conjunction with adaptive time-stepping and two-way coupling strategy. Based on this method, an analysis of the cumulative thermal deformation characteristics and their causes of the wing structure under long-term aerodynamic/thermal co-activity was carried out, as well as the the deformation impacts on the aerodynamic characteristics were performed. The results of the study indicate that, due to the temporal and spatial scale differences of physical processes such as aerodynamic forces, aerodynamic heating, aerodynamic heat transfer and deformation response in wing structures, the cumulative thermal deformation exhibits three distinct nonlinear evolutionary stages: first development, second development, and full development. Moreover, it requires a considerable amount of time for various factors to fully develop and achieve stable accumulated deformation. The aforementioned evolution of accumulated thermal deformation behavior triggers the nonlinear changes in aerodynamic characteristics throughout the entire flight journey, resulting in adverse effects such as decrease of lift, decrease of lift-to-drag ratio and the deviation of pitching moment. Therefore, in the design of long-duration hypersonic aircraft in the future, these associated adverse effects need to be duly considered and actively addressed.

Keywords: long endurance; wing structure; cumulative thermal deformation; aerodynamic/thermal coupling; thermal deformation effects

Received: 2023-07-27; **Accepted**: 2023-12-01; **Published Online**: 2024-04-03 12:28

URL: link.cnki.net/urlid/11.2625.V.20240402.1828.002

Foundation items: Science and Technology on Scramjet Laboratory Found (2021-JCJQ-LB-020-11); National Natural Science Foundation of China (12002361); National Key Research and Development Program of China (2019YFA0405202); Zhiqiang Found

* **Corresponding author**. E-mail: zhupai@mail.ustc.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0427

线束拓扑图纸快速签收方法

雷璐[1]，吴明宇[1]，郭智博[1]，乔晓利[1,2,*]

(1. 长春理工大学 机电工程学院，长春 130022；2. 长春理工大学重庆研究院，重庆 401135)

摘 要：针对人工校验签收线束拓扑图纸效率低、失误率高且不能适应智能化生产的问题，提出了基于图文法和谓词逻辑的线束拓扑图纸快速签收方法。该方法主要由图重构和图解析组成。图重构过程中，为使图结构能更好反映图纸数据，引入谓词逻辑对图资源进行提取，将拓扑图重构为标准图。图解析包含图语法分析器和图语义分析器。图语法分析器使用图文法归约算法对图结构进行识别解析，判断设计过程中是否出现图纸结构性错误，实现制图逻辑检查；图语义分析器应用归约判断重构后的图信息描述是否正确，并参照信息唯一性准则，将路径信息解析后与电线表比对，完成轨迹信息匹配性检测。在此基础上开展了12组实物实验，结果表明：所提方法可以将线束拓扑图纸签收耗时从3 h降低为9 s，且在此过程中未出现漏识别现象。

关 键 词：线束拓扑图纸；签收；图文法；谓词逻辑；归约

中图分类号：TP391.7

文献标志码：A **文章编号**：1001-5965(2025)09-3098-10

拓扑图纸[1]作为线束拓扑设计阶段和工艺设计阶段相互衔接的数据传递媒介，承担了将线束电路原理设计、实际零部件选型及线束相关图纸绘制等信息输送至工艺设计人员的任务，方便对方结合工艺规范绘制工装图版[2]指导实际生产。设计单位下发拓扑图纸到工艺设计部门，对该图纸的审查和接收称为拓扑数据签收[3]。此过程中，能否及时全面地发现拓扑图纸中潜藏的问题，对于保证工艺设计，乃至生产和装配过程顺利进行至关重要。

根据文献[4]总结的问题表征对于问题求解的重要性所述“问题表征说明了思维中问题的产生及呈现过程，是问题解决的必要条件”，在线束拓扑图纸分析审签过程中同样如此。线束拓扑图纸根据机载设备和飞机结构进行绘制，表征的合理性毋庸置疑，但其答案却不是显而易见的，仅靠人眼定位线束与连接器，并发现图纸中的全部问题[5]十分困难。例如，结点未连接、线束号重复等隐蔽性较强的问题通过目视审查难以发现，一旦流入工艺设计环节，会严重影响产品的质量和可靠性。据调查发现，电缆线路引发的故障占飞机空中故障的20%[2]，且因线束对飞机各模块的重要性[6]，易导致故障升级，从而发生重大空难事故。

目前，国内外对线束领域的研究主要集中在电气设计、电器件物理设计、产品数据管理、流程监控和质量控制等领域[7-10]，对于线束拓扑图纸签收技术的研究极为薄弱，大多使用模式识别的方法，基于深度优先[11]或广度优先搜索[12]对图纸路径解析，极少开展以发现设计缺陷为目标的相关研究工作。

鉴于上述原因，本文提出基于谓词逻辑[13]和图文法[14]的线束拓扑图纸快速签收方法，将拓扑图纸抽象为图数据类型[15]，采用谓词逻辑实现图重构，结合图文法对拓扑数据进行描述，并通过图文法归约算法[16]实现图纸规范性检查与轨迹信息匹配性

收稿日期：2023-06-29；**录用日期**：2023-11-10；**网络出版时间**：2023-12-09 11:45
网络出版地址：link.cnki.net/urlid/11.2625.V.20231208.1517.001
基金项目：吉林省科技发展计划(20200401128GX)；重庆英才计划包干制项目(cstc2022ycjh-bgzxm0187)
*通信作者. E-mail：qiaoxiaoli@cust.edu.cn

引用格式：雷璐，吴明宇，郭智博，等. 线束拓扑图纸快速签收方法[J]. 北京航空航天大学学报，2025，51(9)：3098-3107.
LEI L，WU M Y，GUO Z B，et al. Fast sign-off method for wiring harness topology drawings[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3098-3107 (in Chinese).

检测的目的, 使线束拓扑图纸签收效率和准确率大幅度提升。

1　问题分析

图 1 为航空线束拓扑示意图, 主要由电连接器、线束段、结点等构成。由于线束拓扑图具有图数据结构的典型特征, 通常使用图对线束拓扑图进行描述。

线束拓扑图[17]中包含的信息比图的基本形式更多, 故使用图扩展定义——超图[18]进行描述: $\boldsymbol{G}(V, E, V_{\mathrm{A}}, V_{\mathrm{DI}}, E_{\mathrm{A}})$。其中, V 表示图中的结点, E 表示边, V_{A} 表示结点属性, V_{DI} 表示与结点存在映射关系的设备信息, E_{A} 表示边属性。拓扑图中结点分为 2 类: $\mathrm{Dim}(V_i)=1$ 称为外结点; $\mathrm{Dim}(V_i)>1$ 称为内结点。Dim 为结点的度运算符, i 为节点编号。通常情况下, 拓扑图中出现的设备信息只能够与外结点绑定, 故相较于内结点, 外结点除了自身基本属性外, 还包含了设备信息 V_{DI}, 常以表格形式出现。

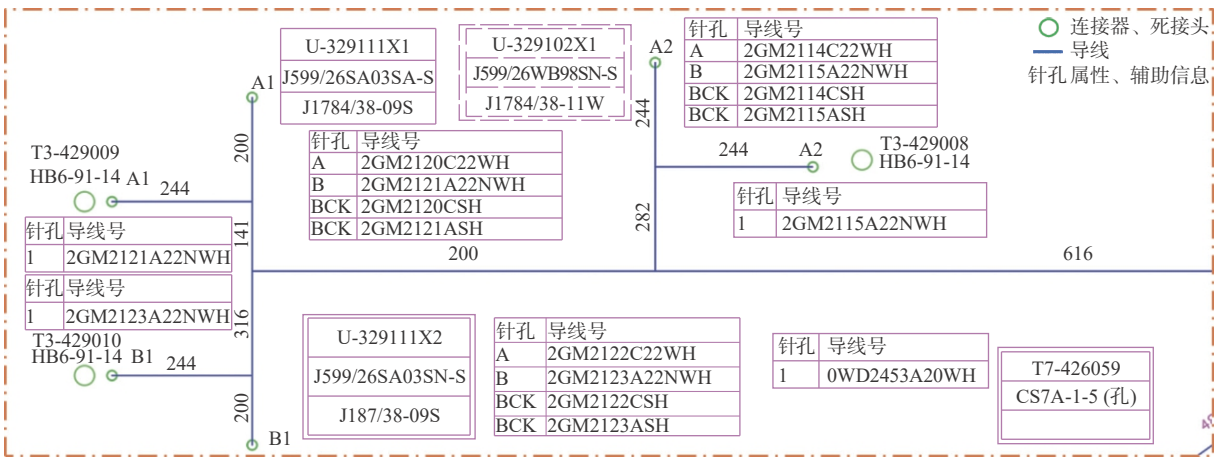


图 1　航空线束拓扑示意图

Fig. 1　Aviation wiring harness topology

线束拓扑图设计阶段出现的错误主要包含 2 类: 结构错误和信息错误。如何能够在较短时间内准确判定图纸是否正确成为难点问题。本文提出基于谓词逻辑和图文法的线束拓扑图纸快速签收方法。

首先, 以线束拓扑图纸绘制标准为依据, 提出基于谓词逻辑的图重构方法。线束拓扑图纸固有的图属性使其适用于使用图文法进行信息表述和操作, 应用的前提是目标图为标准图, 需先将超图进行重构。但目前没有标准方法用于图重构过程, 因此, 提出基于谓词逻辑的图重构技术, 设计对应谓词逻辑, 完成图纸资源提取, 进而将超图重构成标准图。

其次, 基于图文法的图解析包含图语法分析[19]和图语义分析[20]2 个解析审查过程, 以完成图纸数据审签。图语法分析实现制图规范性检查, 主要负责图结构检测; 图语义分析负责图属性检查, 实现轨迹信息匹配性检测。

最后, 根据本文方案构建应用系统, 设计实验分析数据, 对所提方法进行验证。

2　基于谓词逻辑的图重构

图重构[21]需要经过特征图形提取、属性提取和信息匹配 3 个阶段。首先, 提取相关图元, 根据线束拓扑图绘制方法总结特征图形提取规则, 将图元进行提取识别; 其次, 提取属性信息并进行总结, 根据拓扑图纸文本信息提取规则识别文本; 最后, 将识别出的图元分别与其相关信息建立绑定关系。根据线束拓扑图纸绘制规则, 设计谓词逻辑与相关函数, 具体见表 1 和表 2。

2.1　特征图形提取

图纸数据被读取后转化为有意义的图形信息, 依据特征提取规则, 将线束拓扑图中对应图形分别转换为结点、端头、拐点和边, 完成图形数据解读。

1) 结点提取规则。

$\mathrm{Circle}(V_i) \wedge \mathrm{Radius}(V_i, R_0)$

释义: 若点 V_i 为圆形, V_i 的半径为指定长度, 则 V_i 为结点。其中, "∧" 为逻辑 "与" 运算符号。

2) 端头提取规则。

$\mathrm{Node}(V_i) \wedge \mathrm{EqualsNumber}(\mathrm{ConnectedCount}(V_i), 1)$

释义: 若点 V_i 为结点, 与 V_i 相连边的数量为 1, 则 V_i 为端头。

3) 拐点提取规则。

$\mathrm{Node}(V_i) \wedge \mathrm{GreatThanNumber}(\mathrm{ConnectedCount}(V_i), 1)$

释义: 若点 V_i 为结点, 与 V_i 相连边的数量大于 1, 则 V_i 为拐点。

表1 谓词逻辑设计

Table 1 Predicate logic design

函数	参数	释义
Circle(V_i)	V_i为点对象	V_i为圆
Radius(V_i, R_0)	V_i为点对象, R_0为固定半径	V_i半径为R_0
Node(V_i)	V_i为点对象	V_i为结点
Line(E_i)	E_i为边对象	边为直线
SamePoint(O_{V_i}, O_{V_j})	V_i、V_j为点对象	两点重合
Outer(V_i)	V_i为点对象	V_i为外结点
Knee(V_i)	V_i为点对象	V_i为拐点
EquipInfoTable(g)	g为图元	元素为设备信息表成员
PinHoleTable(g)	g为图元	元素为针孔表成员
Text(g)	g为图元	元素为文本
Table(g)	g为图元	元素为表格
GreatThanNumber(m, N_0)	m为数值, N_0为常数	m大于N_0
EqualsNumber(m, N_0)	m为数值, N_0为常数	m等于N_0
NaturalNumber(t)	t为文本对象	文本为自然数
SameChar(C_0, C_1)	C_0、C_1为字符	两字符相同
Unique(t)	t为文本对象	文本内容唯一存在

表2 过程函数设计

Table 2 Process function design

函数	参数	释义
ConnectedCount(V_i)	V_i为点对象	获取点V_i相连边的数量
Dis(V_j, V_i)	V_j、V_i为点对象	获取两者距离
Sub(g, 0, 1)	g为图元	获取首字母
Length(g)	g为图元	获取元素长度
Min(d)	d为两者距离	匹配距离最短元素
Identifier(g)	g为图元	获取对应元素唯一标识码

4）边提取规则。

$\text{Line}(E_i) \wedge \text{SamePoint}(P_{\text{s_}E_i}, V_{j0}) \wedge \text{SamePoint}(P_{\text{E_}E_i}, V_{j1})$

释义：若边E_i为直线，起点$P_{\text{s_}E_i}$与结点V_{j0}重合，终点$P_{\text{E_}E_i}$与结点V_{j1}重合，则E_i为边。

2.2 属性提取

线束拓扑图中属性信息以文本的方式展现，对其提取需要依据文本排列规则与对应长度设定，完成设备属性、拐点属性、边属性和区号属性的解读。

1）设备属性提取规则。

$\text{Text}(g) \wedge (\text{EquipInfoTable}(g) \vee \text{PinHoleTable}(g))$

释义：若元素g为文本信息，且为设备信息表或针孔信息表，则g为设备属性。其中，“$\vee$”为逻辑“或”运算符号。

2）拐点属性提取规则。

$\text{Text}(g) \wedge \text{Table}(g) \wedge \text{SameChar}(\text{Sub}(g, 0, 1), C_0) \wedge \text{NaturalNumber}(\text{Sub}(g, 1, m-1)) \wedge \text{GreatThanNumber}(\text{Length}(g), 3)$

释义：若元素g为文本信息，长度为m，不为表格成员，以固定首字母C_0开头，其余部分为自然数，长度大于3，则g为拐点属性。

3）边属性提取规则。

$\text{Text}(g) \wedge \text{Table}(g) \wedge \text{NaturalNumber}(g)$

释义：若元素g为文本信息，不为表格成员，整体为自然数，则g为边属性。

4）区号属性提取规则。

$\text{Text}(g) \wedge \text{Table}(g) \wedge \text{SameChar}(\text{Sub}(g, 0, 1), C_0) \wedge \text{NaturalNumber}(\text{Sub}(g, 1, m-1)) \wedge \text{EqualsNumber}(\text{Length}(g), 3)$

释义：若元素g为文本信息，不为表格成员，以固定首字母C_0开头，其余部分为自然数，长度等于3，则g为区号属性。

2.3 信息匹配

信息匹配阶段通过建立图纸资源间内在逻辑关系，完成图资源与属性一体化存储，将线束拓扑示意图形式结构映射为标准图结构。图属一体化是指将特征图形与其对应的文本信息相互绑定。具体可分为设备属性与端头、拐点属性与拐点、边属性与边。

在该过程中，遍历各结点、轨迹与文本信息，依照匹配规则完成不同类型文本信息与特征图形的匹配。匹配完成后的文本信息以属性方式存储在特征图形的数据类中。

1）设备属性匹配端头。

$\text{Outer}(V_i) \wedge \text{Min}(\text{Dis}(T_{\text{EI}j}, V_i))$

释义：若结点V_i为外结点，将V_i与距离最近的设备属性表$T_{\text{EI}j}$匹配。

2）拐点属性匹配拐点。

$\text{Knee}(V_i) \wedge \text{Min}(\text{Dis}(V_j, V_i))$

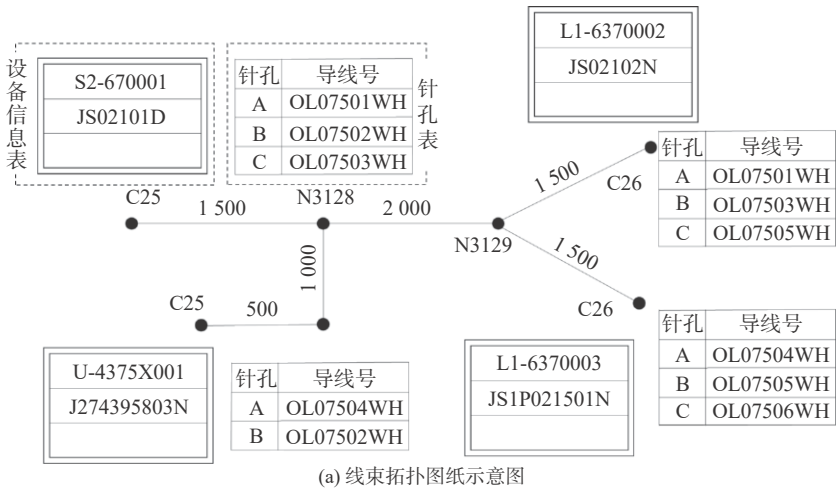
释义：若结点V_i为拐点，将V_i与距离最近的拐点V_j属性匹配。

3）边属性匹配边。

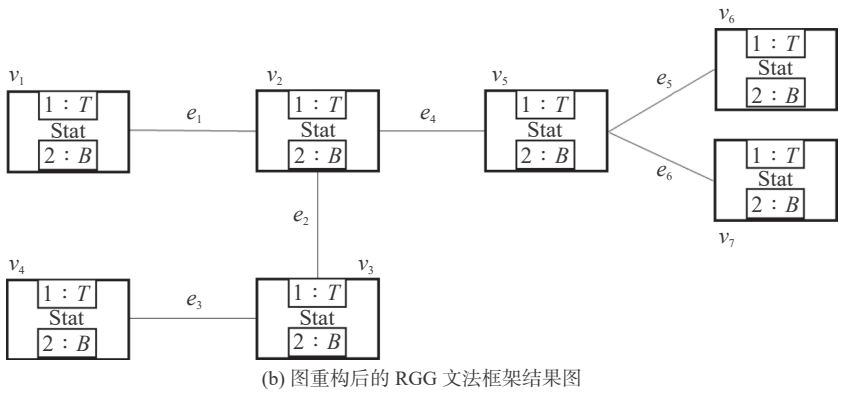
$\text{Min}(\text{Dis}(O_{E_j}, L_{\text{T}}))$

释义：将E_j与距离轨迹中心O_{E_j}最近的边属性文本L_{T}匹配。

以图2(a)为例，在图重构操作完成后，线束拓扑图纸转换为如图2(b)所示的关系生长文法(relation growth grammar, RGG)框架[22]结构，图内结点和边的特征标识分别用v_1、v_2和e_1、e_2等表示。下阶段图分析需依据其进行图结构和信息分析。



(a) 线束拓扑图纸示意图



(b) 图重构后的 RGG 文法框架结果图

图 2　线束拓扑图纸图重构起始图-结果图

Fig. 2　Start diagram-result diagram of graph reconstruction of wiring harness topology drawing

3　基于图文法的图解析

线束拓扑图重构阶段主要关注特征图形分析和属性完备性分析, 未对图纸整体结构合法性与信息表述准确性进行检测。图解析负责图纸解析工作, 在图重构基础上实现线束拓扑图结构的进一步描述, 利用图文法归约算法针对图结构进行制图逻辑检查; 依据电线表中给出的设备导通数据检测图纸反映的拓扑关系, 对图结点中绑定的属性信息及电线表中提供的轨迹信息进行匹配性检测, 确定图纸设计的正确性。

3.1　图语法分析

图语法分析使用归约算法对图进行制图规范性检查, 可对结点悬空、电线悬空和电线成环 3 种结构性错误进行判断。

图 3 为一组 RGG 产生式集合, 产生式①左端 $\boldsymbol{\lambda}$ 为文法初始图, 右端为 2 个度为 1 的结点和 1 条连接它们的边; 产生式②左端为 2 个有边相连的结点, 右端将有边相连的右边结点又连接 1 个叶子结点, 即度为 1 的结点。

以图 2 为例, 利用图 3 的 RGG 产生式能够得到其归约过程。设该图用 $\boldsymbol{G}$ 表示, 按照式 (1) 对图 $\boldsymbol{G}$ 循环使用产生式②进行图归约, 式 (1) 中“*”表

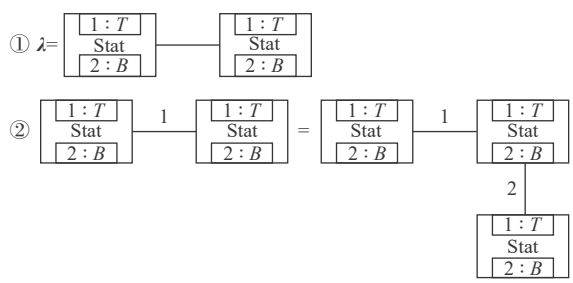


图 3　图文法 RGG 产生式集合

Fig. 3　Graph grammar RGG production set

示多次使用, 可得到产生式①右端所示图结构 $\boldsymbol{G}'$。最后, 执行式 (2) 将整张图归约为一个初始图 $\boldsymbol{\lambda}$。

$$\boldsymbol{G} \Rightarrow *\boldsymbol{G}' \tag{1}$$

$$\boldsymbol{G}' \Rightarrow \boldsymbol{\lambda} \tag{2}$$

该过程将图中叶子结点(即度为 1 的结点)和其连接的边全部消除, 具体如图 4 所示, 经 RGG 产生式集合作用将整张图归约完成。

图归约处理完成后, 上述 3 种错误情况分别对应如图 5 所示 3 种结果。

1) 生成初始图中有单个结点不与其他结点相连, 即图内包含悬结点。

2) 没有生成初始图 $\boldsymbol{\lambda}$, 图内存在未连接结点的

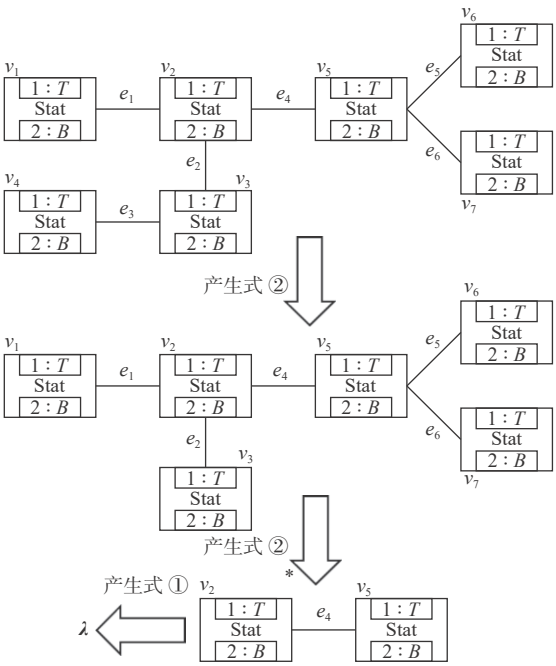


图 4 图归约过程

Fig. 4 Graph reduction process

电线，即图内包含悬边。

3）没有生成初始图 $\boldsymbol{\lambda}$，而生成了由多点多线连接而成的环，说明线束拓扑图纸中存在环。

3.2 图语义分析

完成语法分析的结果图已经在结构上符合校验标准，但仍需通过语义分析进行轨迹信息匹配性检测，过程中须对以下 3 个制图准则进行校验：

1）电线唯一准则：图中每根电线以线束号唯一标识身份，谓词逻辑描述为 Unique(Identifier(wire))。

2）设备唯一准则：图中每台设备以设备号唯一标识身份，谓词逻辑描述为 Unique(Identifier (device))。

3）针孔唯一准则：图中每个针孔以针孔号唯一标识身份，谓词逻辑描述为 Unique(Identifier (pinhole))。

线束路径解析是对图纸绘制语义的解析。线束拓扑图纸描述了电连接器之间的连接关系，若在图的角度表示，则为两结点之间相互连通。在该阶段需要在电线表中取得每条电线两端电连接器设备信息标识，在线束拓扑图中找出其对应的电连接器位置，判断图纸中它们之间能否获得一组线段相互连通，如图 6 加粗部分所示。

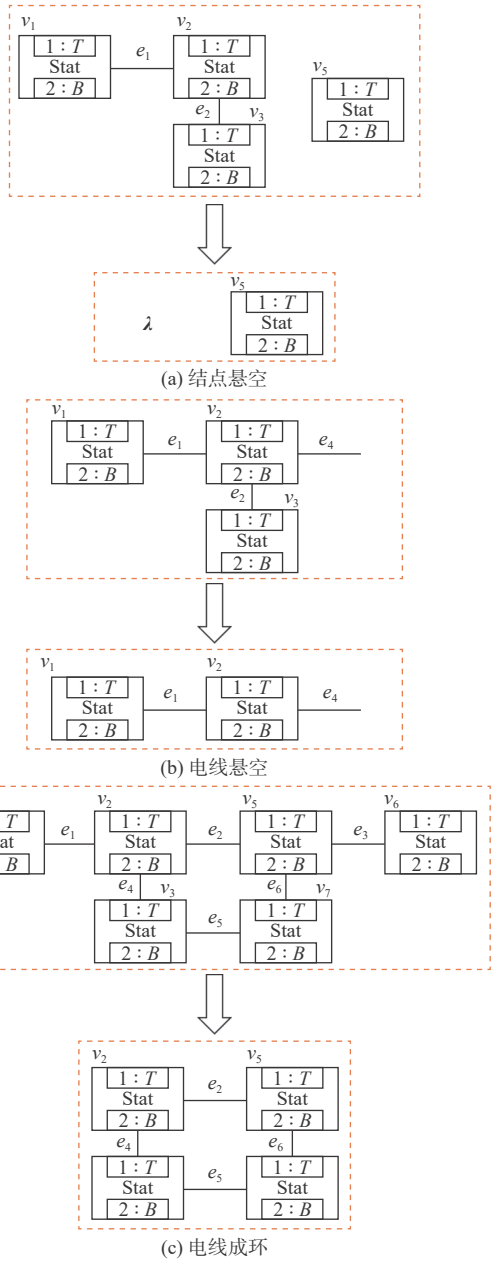
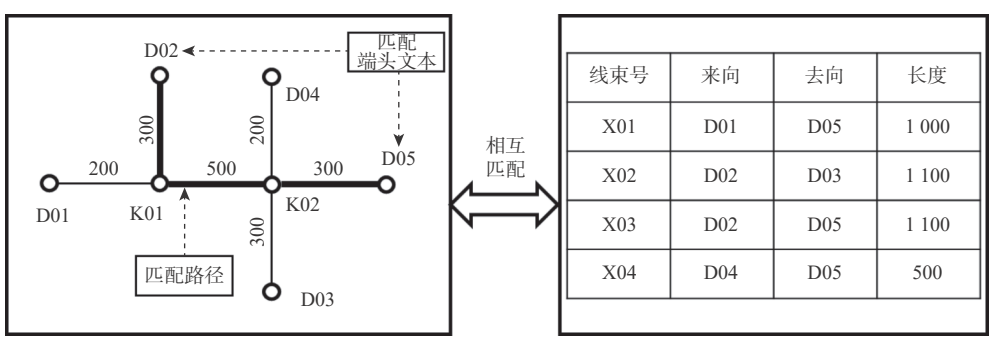


图 5 归约完成的 3 种错误情况

Fig. 5 Three error situations for completion of reduction



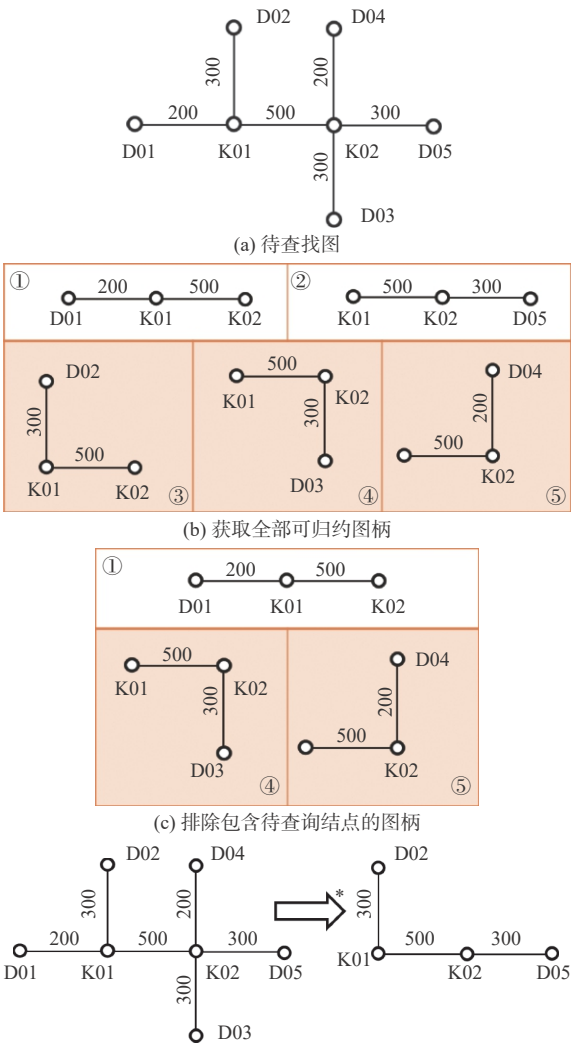
线束号	来向	去向	长度
X01	D01	D05	1 000
X02	D02	D03	1 100
X03	D02	D05	1 100
X04	D04	D05	500

图 6 线束路径解析

Fig. 6 Wiring harness path analysis

在进行图路线寻找中, 广度优先搜索和深度优先搜索为比较常用的方法, 但在线束路径搜索中均表现不佳, 主要是由于递归算法占用了过多的时间和空间资源, 导致线束路径识别缓慢。本文利用归约算法对线束正确性进行判定, 并在此方法基础上以结点属性信息为参考建立归约应用条件, 对正在操作的图柄进行判断, 若此图柄包含被选结点, 则不进行归约。归约完成后, 边集 E 中剩余的边即为线束两结点间的路径。

以图 6 为例, 若选中来向为 D02、去向为 D05 的路径, 其归约过程如图 7 所示。查找全部可归约图柄, 从中排除包含 D02 和 D05 的图柄, 应用式 (1) 进行归约。重复上述操作, 直至仅剩余包含所求结点的可归约图柄。进行归约操作得到所求结果图, 该图内边集 E 即为从结点 D02 到 D05 的路径。



(a) 待查找图

(b) 获取全部可归约图柄

(c) 排除包含待查询结点的图柄

(d) 使用图柄进行归约

图 7　路径查找过程

Fig. 7　Path lookup process

图文法归约算法的时间复杂度和空间复杂度取决于算法的实现和问题的规模。n 为图的节点数, e 为边数, 时间复杂度与图的大小和规则复杂性相关, 线束拓扑图纸可通过提取与简化规则, 尽可能降低其时间复杂度, 本文方法时间复杂度为 $O(n\log_2 n)$。同时, 空间复杂度受算法中使用的数据结构影响, 考虑到存储图结构、规则和中间结果等信息, 空间复杂度为 $O(n+e)$。

4　实　验

实验基于 Windows7(64 bit)操作系统, 利用 Teigha 控件进行 CAD 图纸图元数据提取及显示, 处理器为 i5-2430M@2.40 GHz×2, 内存为 8 GB, 固态硬盘容量为 512 GB。

采用某企业提供的线束拓扑图纸与电线表原始文件进行测试, 线束拓扑图纸内分别描述了 10～300 条不等的线束轨迹。依据图纸中边数, 划分为 4 种图纸: 小型、中型、大型和超大型, 边数< 50 为小型; 50≤边数 <100 为中型; 100≤边数 < 200 为大型; 边数≥200 为超大型。

4.1　方案可行性分析

依据图重构设计内容进行实验, 记录数据如表 3 所示。

表 3　图重构过程耗时

Table 3　Time consumption of graph reconstruction process

实验编号	线束拓扑图纸边数/条	图重构时间/s
1	17	1.97
2	31	2.23
3	45	2.87
4	62	2.95
5	80	3.18
6	89	3.20
7	116	3.28
8	137	3.82
9	169	4.03
10	203	4.29
11	217	4.59
12	262	6.77

重构过程中, 对于拓扑图纸资源提取的方案可行性为 100%。提取耗时随图纸规模的增大而增加, 小型图纸平均时长为 2.36 s, 中型图纸平均时长为 3.11 s, 大型图纸平均时长为 3.71 s, 超大型图纸平均时长为 5.22 s。

依据图语法分析器设计内容, 记录数据如表 4 所示。

图语法分析过程中, 图语法分析器对于拓扑图纸制图逻辑检查的方案可行性为 100%。检查耗时随图纸规模的增大而增加, 小型图纸平均时长为 1.09 s, 中型图纸平均时长为 1.49 s, 大型图纸平均

表4 图语法分析结果

Table 4 Graph grammar analysis results

实验编号	线束拓扑图纸边数/条	制图逻辑检查时间/s
1	17	0.85
2	31	1.09
3	45	1.32
4	62	1.39
5	80	1.53
6	89	1.54
7	116	1.76
8	137	2.17
9	169	2.61
10	203	3.25
11	217	3.37
12	262	4.28

时长为2.18 s, 超大型图纸平均时长为3.63 s。

依据图语义分析器设计内容, 记录数据如表5所示。

表5 图语义分析结果

Table 5 Graph semantic analysis results

实验编号	线束拓扑图纸边数/条	轨迹信息匹配性检查时间/s
1	17	0.19
2	31	2.21
3	45	2.76
4	62	2.94
5	80	3.17
6	89	3.26
7	116	3.34
8	137	3.76
9	169	3.89
10	203	4.28
11	217	4.91
12	262	6.55

图语义分析过程中, 图语义分析器对于拓扑图纸轨迹信息匹配性检查的方案可行性为100%。检查耗时随图纸规模的增大而增加, 小型图纸平均时长为1.72 s, 中型图纸平均时长为3.12 s, 大型图纸平均时长为3.66 s, 超大型图纸平均时长为5.25 s。

4.2 签收优势对比

对于不同规模图纸, 记录计算机判图总时长与人工判图总时长, 如表6所示。根据表6数据绘制效率对比图, 如图8所示。

图纸签收效率平均提升率计算公式为

$$\begin{cases} \eta = t/e \\ P = (\eta_r - \eta_c)/\eta_c \end{cases} \tag{3}$$

式中: t 为所用时间; e 为图纸边数; η 为签收效率;

表6 签收效率

Table 6 Sign-off efficiency

实验编号	线束拓扑图纸边数/条	计算机判图总时长/s	人工判图总时长/h
1	17	3.01	1
2	31	5.53	1.2
3	45	6.95	1.5
4	62	7.28	2.1
5	80	7.88	2.7
6	89	8.00	2.8
7	116	8.38	3.1
8	137	9.75	3.9
9	169	10.53	4.5
10	203	11.82	6.1
11	217	12.87	6.2
12	262	17.60	6.7

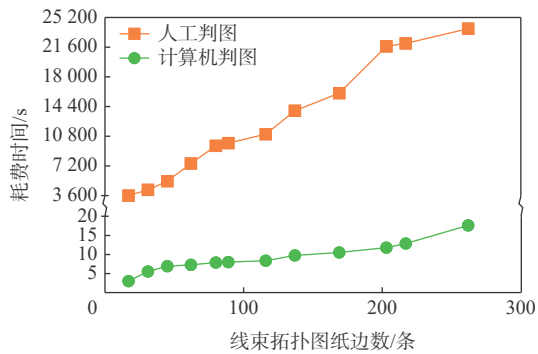


图8 签收效率对比

Fig. 8 Comparison of sign-off efficiency

η_r、η_c 分别为人工与计算机签收效率; P 为判图效率提升率。代入数据可得各种规模图纸判图效率提升率。小型图纸判图效率提升率 P_s=91 706.17%, 中型图纸判图效率提升率 P_m=117 632.14%, 大型图纸判图效率提升率 P_l=143 573.46%, 超大型图纸判图效率提升率 P_{xl}=165 319.61%, 总体平均判图效率提升率 P=129 557.8%。应用此方案, 成功将签收时长量级从小时降到秒。

依据实验, 记录签收准确性数据, 如表7所示。根据表7数据绘制对比图, 如图9所示。比较计算机与人工判图失效次数, 小型图纸平均人工失效次数为0.33次, 中型图纸平均人工失效次数为2次, 大型图纸平均人工失效次数为4次, 超大型图纸平均人工失效次数为8.33次, 总体平均人工失效次数为3.67次。计算机判图签收准确率为100%, 未出现漏判、错判现象。

4.3 数据分析

整体分析数据可知, 随着线束拓扑图纸规模增大, 轨迹逐渐增多, 签收方案中图重构、图语法分析与图语义分析耗时随之增加, 人工判图时间亦是如

表 7　签收准确性

Table 7　Sign-off accuracy

实验编号	线束拓扑图纸边数/条	人工判图失效次数	计算机判图失效次数
1	17	0	0
2	31	1	0
3	45	0	0
4	62	2	0
5	80	0	0
6	89	4	0
7	116	3	0
8	137	5	0
9	169	4	0
10	203	7	0
11	217	6	0
12	262	12	0

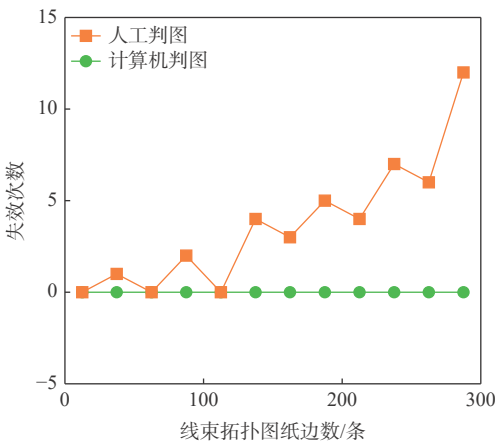


图 9　签收准确性对比

Fig. 9　Comparison of sign-off accuracy

此。且随着线束复杂程度不断上升, 人工判图漏判、误判等现象较为严重, 而计算机可以有效减小失误率, 提升签收数据准确性。进一步与现有的图纸自动识别和校验方法对比, 其较多关注拓扑图纸在传输过程中的改变, 通过提取原版图纸拉普拉斯矩阵特征值序列与现版对比得出结论, 其理论较为复杂, 工程背景不完全匹配, 致使复杂度较高; 本文签收方法基于校验规则, 流程相对简单, 易于实现, 对已知的线束设计错误识别准确且高效, 在时间复杂度和空间复杂度上也有独特的优势。

5　结　论

1) 设计了基于谓词逻辑和 RGG 图文法的线束拓扑图纸快速签收方案。首先, 依据线束拓扑图纸资源表征规则, 使用谓词逻辑设计图重构, 以图元为基本单位进行图形、信息的提取与匹配; 然后, 基于图文法能够对二维图结构进行有效分析的特性, 设计了图语法分析器, 利用归约算法对图纸设计结构性错误进行判断, 完成制图逻辑检查; 最后, 根据线束拓扑图纸设计准则构造语义分析器, 进行线束路径解析, 与电线表中数据进行匹配性校验, 从而准确、快速判断图纸设计的正确性, 完成图版签收任务。

2) 通过实验数据验证说明了设计方案的可行性与可靠性, 展现出本文方法在应用中的高效率与高准确度。

目前, 方法设计实现主要基于线束设计规则, 对其他基础设计逻辑不同的拓扑图纸或新型的、未录入对应图纸设计准则的设计错误, 尚不能直接使用, 需消耗一定的时间成本将对应的设计逻辑和校验标准进行补充和完善, 未来将重点研究方法的通用性设计, 以提高其普适性。

参考文献 (References)

[1] 李伟, 闫静. 基于 CATIA 二次开发的线束设计数据提取和结构识别[J]. 电子元器件与信息技术, 2020, 4(1): 71-72.
LI W, YAN J. Data extraction and structure identification of wire harness design based on CATIA secondary development[J]. Electronic Component and Information Technology, 2020, 4(1): 71-72 (in Chinese).

[2] 王萍, 张晓茜, 罗苑琪. 面向工艺的飞机线束精准设计技术探索[J]. 航空计算技术, 2022, 52(6): 16-20.
WANG P, ZHANG X X, LUO Y Q. Process oriented precision design technology for aircraft harness[J]. Aeronautical Computing Technique, 2022, 52(6): 16-20(in Chinese).

[3] 王春阳. 航空线束产品三维全信息模型定义与应用研究[D]. 北京: 北京理工大学, 2017: 73-82.
WANG C Y. Research on the definition and application of three-dimensional full information model of aviation wire harness products[D]. Beijing: Beijing Institute of Technology, 2017: 73-82 (in Chinese).

[4] LARKIN J H, SIMON H A. Why a diagram is (sometimes) worth ten thousand words[J]. Cognitive Science, 1987, 11(1): 65-100.

[5] 武鹏, 张一哲, 闻敬谦, 等. 面向航空线束的数字化制造关键问题探讨[J]. 航空制造技术, 2017, 60(14): 97-100.
WU P, ZHANG Y Z, WEN J Q, et al. Discussion on key issues of digital manufacturing for aircraft wire harness[J]. Aeronautical Manufacturing Technology, 2017, 60(14): 97-100(in Chinese).

[6] 王小辉, 朱丽, 车程, 等. 电气线路互联系统安全性设计与分析方法研究[J]. 西北工业大学学报, 2022, 40(3): 690-698.
WANG X H, ZHU L, CHE C, et al. Research on safety design and analysis method of electrical line interconnection system[J]. Journal of Northwestern Polytechnical University, 2022, 40(3): 690-698(in Chinese).

[7] 胡波, 赵猛, 卢山, 等. EWIS 数字化设计技术分析[J]. 飞机设计, 2019, 39(2): 56-59.
HU B, ZHAO M, LU S, et al. Analysis of EWIS digital design technology[J]. Aircraft Design, 2019, 39(2): 56-59(in Chinese).

[8] YANG S, TONG W, TIAN W. Lilium joins forces with GKN

aerospace for lilium jet's electrical wiring interconnection system (EWIS)[EB/OL]. (2023-01-31)[2023-06-20]. https://www.proquest.com/magazines/lilium-joins-forces-with-gkn-aerospace-jets/docview/2771207620/se-2.

[9] 石旭东, 张和茂, 赵宏旭, 等. 复杂线束在双 BCI 耦合下的终端响应机理[J]. 北京航空航天大学学报, 2021, 47(9): 1739-1747.
SHI X D, ZHANG H M, ZHAO H X, et al. Terminal response mechanism of complex wiring harness under double BCI coupling[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(9): 1739-1747(in Chinese).

[10] SOHAIL M, PFEIFFER H, WEVERS M. Addressing safety concerns in hybrid electric aircrafts: in-flight icing detection, moisture detection in fuselage and electrical wiring and interconnect system (EWIS)[J]. IOP Conference Series: Materials Science and Engineering, 2022, 1226(1): 012060.

[11] NIKOLAEV S, ROMANOV O, NYSHCHUK A. Method of modified depth-first search in a graph for constructing all possible gray codes of a specified length[J]. Cybernetics and Systems Analysis, 2023, 59(3): 359-364.

[12] PANAFIEU É. Analytic combinatorics of connected graphs[J]. Random Structures & Algorithms, 2019, 55(2): 427-495.

[13] 乔晓利, 李林, 刘贡平, 等. 基于谓词逻辑的飞机线束工装图版设计[J]. 北京航空航天大学学报, 2021, 47(9): 1756-1764.
QIAO X L, LI L, LIU G P, et al. Predicate logic based tooling drawing design of aircraft harness[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(9): 1756-1764(in Chinese).

[14] LIU Y F, YANG F. EGG+: a graph grammar formalism with uncertain structure processing mechanism[J]. Journal of Logic and Computation, 2021, 31(7): 1800-1819.

[15] 郝国舜, 刘静华, 李士才. 基于对象的 P&ID 设计软件无向图模型[J]. 北京航空航天大学学报, 2004, 30(2): 122-126.
HAO G S, LIU J H, LI S C. Object-oriented undirected graph model in P & ID design system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2004, 30(2): 122-126(in Chinese).

[16] 邹阳, 吕建, 曹春, 等. 上下文相关图文法的表达能力分析[J]. 软件学报, 2012, 23(7): 1635-1655.
ZOU Y, LV J, CAO C, et al. Analysis of expressive ability of context-correlated grammar[J]. Journal of Software, 2012, 23(7): 1635-1655(in Chinese).

[17] YEH C H, SMITH R E, DHOLLANDER T, et al. Connectomes from streamlines tractography: assigning streamlines to brain parcellations is not trivial but highly consequential[J]. NeuroImage, 2019, 199: 160-171.

[18] CLARK G J, THOMAZ F, STEPHEN A T. Comparing the principal eigenvector of a hypergraph and its shadows[J]. Linear Algebra and Its Applications, 2023, 673: 46-68.

[19] REKERS J, SCHÜRR A. Defining and parsing visual languages with layered graph grammars[J]. Journal of Visual Languages & Computing, 1997, 8(1): 27-55.

[20] KONG J, ZHANG K, DONG J, et al. Specifying behavioral semantics of UML diagrams through graph transformations[J]. Journal of Systems and Software, 2009, 82(2): 292-306.

[21] 陈晋音, 黄国瀚, 张敦杰, 等. 一种面向图神经网络的图重构防御方法[J]. 计算机研究与发展, 2021, 58(5): 1075-1091.
CHEN J Y, HUANG G H, ZHANG D J, et al. A graph reconfiguration defense method for graph neural networks[J]. Computer Research and Development, 2021, 58(5): 1075-1091(in Chinese).

[22] SHI Z Y, SUN Y Q, GU Y W, et al. The study about parsing of RGG grammar[J]. Advanced Materials Research, 2011, 204-210: 255-258.

Fast sign-off method for wiring harness topology drawings

LEI Lu[1]，WU Mingyu[1]，GUO Zhibo[1]，QIAO Xiaoli[1, 2, *]

(1. School of Mechanical and Electrical Engineering，Changchun University of Science and Technology，Changchun 130022，China；
2. Chongqing Research Institute of Changchun University of Science and Technology，Chongqing 401135，China)

Abstract：To address the low efficiency, the high error rate, and inability to adapt to intelligent production when wiring harness drawings are manually verified and signed off, this paper proposed a fast sign-off method for wiring harness topology drawings based on graph grammar and predicate logic. The method mainly consisted of two parts: graph reconstruction and graph parsing. In the process of graph reconstruction, to make the graph structure better reflect drawing data, predicate logic was introduced to extract graph resources, and the topology drawing was reconstructed into a standard graph. Graph parsing involved a graph grammar analyzer and a graph semantic analyzer. In the graph grammar analyzer, a graph grammar reduction algorithm was used to identify and analyze the graph structure, so as to judge whether there were structural errors in the drawing during the design process and to realize the logic check of the drawing. The graph semantic analyzer applied the reduction judgment to reconstructed graph information description and refers to the information uniqueness criterion for comparing the path information with the wire table after parsing to complete the matching detection of the trajectory information. On this basis, 12 groups of physical experiments were carried out. The experimental results show that this method can reduce the time required for signing off the wiring harness drawings from 3 hours to 9 seconds, and there is no missing recognition phenomenon in the process.

Keywords：wiring harness topology drawings; sign-off; graph grammar; predicate logic; reduction

Received：2023-06-29；**Accepted**：2023-11-10；**Published Online**：2023-12-09 11：45

URL：link.cnki.net/urlid/11.2625.V.20231208.1517.001

Foundation items：Science and Technology Development Plan of Jilin Province (20200401128GX); Chongqing Talent Plan Contract System Project (cstc2022ycjh-bgzxm0187)

* **Corresponding author**. E-mail：qiaoxiaoli@cust.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0428

基于资源供需差与情境理解度的情境意识模型

王一行，刘双，完颜笑如*，冯传宴，周孙夏，钱春颖
(北京航空航天大学 航空科学与工程学院，北京 100191)

摘　　要：为进一步满足航空器人机界面的研制需求，开展航空器任务和显控界面设计紧密结合特点下作业人员情境意识(SA)水平的理论预测，提出一种基于资源供需差与情境理解度的情境意识量化分析模型。该模型以经典且广泛适用的情境意识评估技术(SART)量表对情境意识的定义方式为基础，将人员情境意识表征为资源供需差与情境理解度的加和，并引入多维度人机界面显示格式对情境理解度进行量化，结合任务网络分析与多资源理论对各任务单元的资源供需差进行衡量。为对模型的有效性进行验证，采用30名被试开展不同情境评估难度下的5边飞行模拟实验，通过调整改变仿真气象条件、仪表显示模式、飞行导航地图显示模式以实现对情境评估难度的调控，并结合主观量表、飞行绩效及眼动指标对作业人员在不同情境评估难度下的情境意识水平进行测评。数据分析结果表明，所提情境意识理论模型的计算结果与实验主观（SART量表、机组意识评定量表(CARS)、美国国家航空航天局-任务负荷指数量表(NASA-TLX)）、客观（操纵绩效、注视熵、扫视次数）测量结果均呈现出良好的相关性，对模型的效度进行了初步验证。所提情境意识理论模型可结合航空器任务操控流程设计和显控界面设计特征对人员情境意识分析预测提供一定的理论依据和技术方法。

关　键　词：情境意识；资源供需差；情境理解度；绩效测量；眼动测量
中图分类号：V7; R857.1
文献标志码：A　　**文章编号**：1001-5965(2025)09-3108-09

情境意识是对情境元素的感知、意义的理解和未来状况的预测。飞行员的情境意识与航空安全密切相关，根据飞行事故数据调查统计结果，62%的事故原因涉及飞行员情境意识[1]。在大多数严重事故和灾难中，飞行员对于确定正确的行动及执行行动并没有困难，但对自身所处情境缺乏充分了解。相关研究表明，较低的情境意识可能会导致更多的认知与操作差错，在复杂的认知情境下还可能导致灾难性的后果[1-2]，因此，对飞行员的情境意识进行准确预测或评估具有重要意义。

在航空工效学研究领域，对人员情境意识的预测评估途径主要包括心理生理测量与理论建模分析2种典型方法[3-4]。目前，心理生理测量方法相对比较成熟，一般包括直接测量方法与间接测量方法。直接测量方法指通过针对任务情境的提问或观察获得作业人员对任务情境感知和理解状况的方法，具体包括记忆探查测量法和主观评价法；间接测量方法通过在线或离线测定作业人员执行任务过程中的各项行为操作或生理参数来反映作业人员的情境意识，主要包括绩效测量法和生理测量法。心理生理测量方法通常仅能对人员先前的或当前的情境意识状态进行实时监测或事后评估，而要实现对人员情境意识的事前预测分析，则需构建情境意识量化分析模型。

收稿日期：2023-06-30；**录用日期**：2023-10-05；**网络出版时间**：2023-11-06 12:09
网络出版地址：link.cnki.net/urlid/11.2625.V.20231103.1743.002
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(U1733118)；航空科学基金(201813300002)；国家自然科学基金(71301005)
* **通信作者**. E-mail：wanyanxiaoru@buaa.edu.cn

引用格式：王一行，刘双，完颜笑如，等. 基于资源供需差与情境理解度的情境意识模型[J]. 北京航空航天大学学报，2025，51(9)：3108-3116. WANG Y H，LIU S，WANYAN X R，et al. Situation awareness model based on resource supply-demand difference and understanding[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3108-3116 (in Chinese).

构建情境意识量化分析模型一般可从实测数据驱动或认知机理驱动 2 种视角出发开展建模研究。其中, 实测数据驱动方法主要基于传感器输出的人员多维度生理参数或操作绩效数据进行数据学习与模型训练; 而认知机理驱动方法则主要基于信息加工过程与情境意识的形成机理来构建模型。基于机器学习等方法构建的实测数据驱动模型能较为准确地实现人员情境意识状态的判别[5], 但模型的非显性特征难以提供有关人员内部心理表征方面的解释[6], 也难以用于情境意识的事前分析。在认知机理驱动建模方面, 研究人员基于情境意识形成的三层次理论、注意与情境意识关系理论、工作记忆与情境意识关系理论等提出了一系列有价值的情境意识量化表征模型[7-10]。上述模型可从事前角度对人员情境意识水平进行一定程度的诊断性分析及预测。

考虑到现有情境意识理论分析模型较少面向航空任务情境, 且多从人员作业状态角度进行建模评估, 与航空器任务设计或显控界面设计结合尚不够紧密, 为进一步面向航空器人机界面研制需求提升情境意识理论分析模型的工程应用价值, 本文提出一种基于资源供需差与情境理解度的情境意识量化分析模型, 并通过设计不同任务情境评估难度下的飞行模拟实验, 对模型的有效性进行初步验证。该模型可在早期设计阶段对航空器人机界面可支撑的情境意识水平进行一定程度的量化分析, 从而为优化航空器人机界面设计与任务流程设计、保障飞行安全提供理论分析依据与量化评估工具。

1　情境意识理论建模

Taylor 提出了经典的情境意识评估技术(situation awareness rating technique, SART), 将情境意识定义为资源供需差(资源供应量与资源需求量的差值)与情境理解度之和[11]。SART 在面向航空、轨道交通、核电等领域的诸多任务情境下获得广泛应用[12], 具有较高的测量效度, 能够很好地对信息资源的质量和数量进行判断, 并对任务相关的支持情境意识形成的认知过程进行有效评估。

在前人所提模型[10] 基础上, 本文基于 SART 对于情境意识的定义方式, 从作业人员的资源供需差与情境理解度 2 个维度出发进行情境意识建模。对于资源供需差或情境理解度的计算, 需结合任务网络建模开展。利用任务网络分析方法对任务进行逐级分解至任务单元, 并将任务单元按照一定的任务逻辑组合即可生成任务网络。常见的任务单元组合逻辑包括固定序列、偶然序列、条件选择序列、并发序列等[13], 对于每个任务单元, 包含一系列任务属性, 如任务名称、任务序号、任务时长及分布、任务触发条件、任务触发效应、任务结束效应等。

通过任务网络分析, 将完整任务阶段 X 按时间段划分为 m 个时长为 T 的时间切片, 其中, $[jT,(j+1)T]\ (j\in\{0,1,\cdots,m\})$ 为完整任务阶段的一个时长为 T 的时间切片, 基于资源供需差与情境理解度分别计算各时间切片内可支持的情境意识水平 A_j, 则完整任务阶段的情境意识水平 A 可由式 (1) 表示:

$$A=\frac{1}{m}\sum_{j=0}^{m}A_j \tag{1}$$

构建的情境意识模型理论框架如图 1 所示。

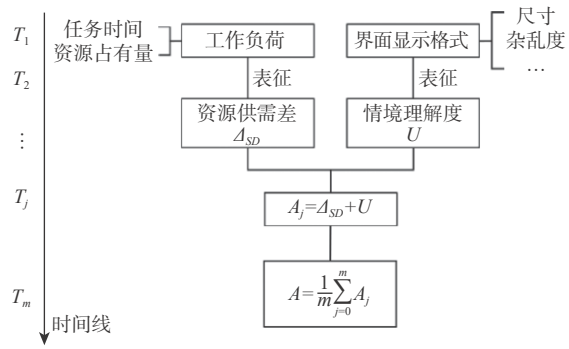


图 1　情境意识模型理论框架

Fig. 1　Situation awareness model theoretical diagram

对于资源供需差, 采用工作负荷进行表征。一方面, 工作负荷是情境意识的重要影响因素之一[10], 大量研究表明, 在飞行操作较为复杂的任务情境下, 人员工作负荷与情境意识存在负相关关系[14]; 另一方面, 工作负荷也常被描述为由任务决定的资源需求量与由操作者能力决定的资源供应量之间的相对关系[15]。

通过在传统视觉-听觉-认知-运动(visual auditory cognitive psychomotor, VACP)的多资源理论模型基础上引入时间压力[16-17], 从内在需求和外在任务压力 2 个方面对作业人员的工作负荷进行度量, 同时与任务网络分析结合, 对于串行任务, 可将完整任务阶段划分为一系列串行任务单元, 通过计算每个任务单元的工作负荷进而得到完整任务阶段的工作负荷; 对于并行任务, 同样划分不同的任务单元, 并将完整任务阶段分解为时间切片, 通过计算每个时间切片内的任务单元的工作负荷进而得到完整任务阶段的工作负荷。对于某一任务单元序列 $X=\{X_1,\cdots,X_i,\cdots,X_n\}\ (i\in\{1,2,\cdots,n\})$, X 的工作负荷 W 可表示为

$$W=\sum_{j=0}^{m}\sum_{i=1}^{n}\frac{1}{T_{X_i}}\sum_{k=1}^{4}\lambda_{k_{X_i}} \tag{2}$$

$$T_{X_i}=L\left(\left[jT,(j+1)T\right]\cap\left[T_{0X_i},T_{\mathrm{EX}_i}\right]\right) \tag{3}$$

式中：T_{0X_i}、T_{EX_i}分别为任务单元X_i的起止时间；$L(\cdot)$为区间的持续时长；$\lambda_{k_{X_i}}$为任务单元X_i的VACP通道资源占有量。对于串行任务的工作负荷计算，T_{X_i}为每个任务单元X_i的持续时长。

进一步基于工作负荷评定结果建立资源供需差Δ_{SD}的阈值函数[10]，表示为式(4)，其反映了工作负荷W与资源供需差Δ_{SD}之间的关系。

$$\Delta_{SD}=\begin{cases}0 & W>60\\ 0.5 & 40<W\leqslant 60\\ 1 & W\leqslant 40\end{cases} \tag{4}$$

相关研究表明，工作负荷W超出一定阈值时，可能损害作业人员的情境意识。当$W\leqslant 40$时，Δ_{SD}取值为1，此时资源供应量高于资源需求量，剩余资源有助于形成较好的人员情境意识水平；当$40<W\leqslant 60$时，Δ_{SD}取值为0.5，此时人员的情境意识水平因W的增加而受到损害；当$W>60$时，Δ_{SD}取值为0，此时人员处于高负荷状态，剩余资源难以维持有效的情境意识[18]。

合理的人机界面设计有助于促进人机系统中的信息传递效率，保障人员情境理解度，从而提高人员情境意识水平[19]。因此，本文基于人机界面设计特征来表征其可支撑的人员情境理解度。Wickens等[9]针对航空界面设计所提出的显示格式与态势感知关系模型(display formatting and situation awareness model, DFSAM)可被有效用于度量人员基于不同界面设计特征条件下的情境理解度。其中，界面设计关键特征包括界面尺寸、界面显示杂乱度、界面显示覆盖程度、界面显示突显性、界面显示维度等[20]，结合考虑不同界面设计特征的权重，情境理解度U可表示为

$$U=\sum_{b=1}^{g}l_b u_b \tag{5}$$

式中：b为不同界面设计特征指标；g为不同界面设计特征的数量；l_b为不同界面设计特征的相对重要程度；u_b为不同界面设计特征指标可支撑的情境理解度。

根据SART对人员情境意识的定义，情境意识可用情境理解度U、资源供应量S和资源需求量D这3个维度进行表征，即

$$A=U+(S-D) \tag{6}$$

式中：$(S-D)$为资源供需差Δ_{SD}。进一步考虑不同情境下资源供需差与情境理解度的相对重要程度，则人员情境意识可用式(7)表示，其中，q为资源供需差的权重。

$$A=q\Delta_{SD}+(1-q)U \tag{7}$$

结合任务网络分析，基于资源供需差与情境理解度的情境意识模型可对特定任务单元或完整任务阶段的情境意识进行量化分析。

需要说明的是，本文基于SART评估原理构建的情境意识模型中，资源供需差与情境理解度之间存在一定程度的相互影响关系，考虑到其主要适用于对不同工况条件(包括任务操控流程设计与显控界面设计)下人员的情境意识水平进行对比分析，情境意识模型输出的理论预测值是相对值而非绝对值，因此，资源供需差与情境理解度2个因素之间的耦合关联关系不影响情境意识模型的可用性。

2 实验验证

通过设计不同任务情境评估难度下的飞行模拟实验，对所构建的情境意识模型的有效性进行初步验证。

2.1 被 试

实验共招募有效被试30名(其中，男性22名，女性8名)，年龄范围为21～25岁(平均年龄为23岁，标准差为0.9)，视力或矫正视力正常，无色盲色弱，心肺功能正常，听力正常。所有被试均为北京航空航天大学航空科学与工程学院学生，实验前接受充分的飞行模拟培训，能够顺利完成预设的飞行模拟任务。

2.2 实验仪器设备

实验采用基于Xplane软件系统的飞行模拟仿真平台开展，系统组成如图2所示，主要包括飞行仪表系统、视景仿真系统、操纵杆、油门杆、脚蹬等。飞行界面模型参考塞斯纳172 G1000航电系统，并划分为4个兴趣区，包括外部视景区、主飞行仪表显示区、地图显示区和机械仪表显示区。

实验采用Tobii Glass 2可穿戴式眼动追踪系统实时记录被试的眼动数据，采样频率为50 Hz。同时，采用光电脉搏传感器采集被试的心电数据，采样频率为64 Hz。此外，采用绩效记录系统TACVIEW进行关键飞行数据的记录与输出，包括飞机位置(经纬度、离地高度、海拔高度等)及系统状态(指示空速、俯仰角、发动机功率等)等参数，采样频率为1 Hz。

2.3 实验设计

采用单因素被试内设计，自变量为基于任务情境下的情境评估难度，包括高、中、低3个水平。

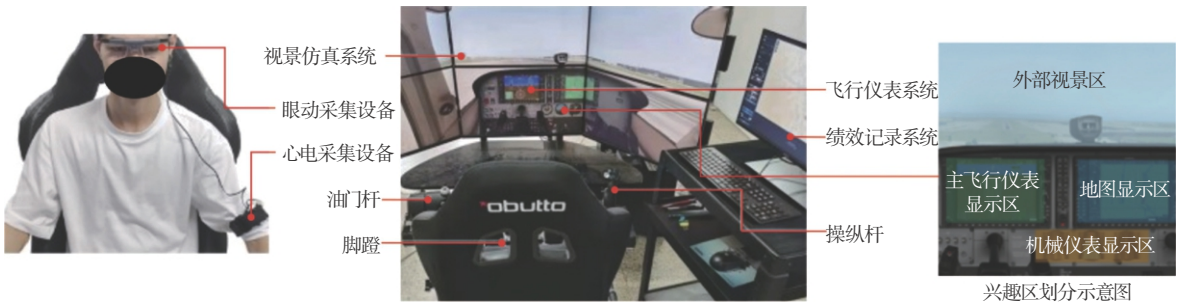


图 2　飞行模拟仿真平台

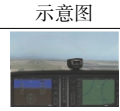
Fig. 2　Flight simulation platform

情境评估难度通过调整改变仿真气象条件、仪表显示模式、飞行导航地图显示模式实现[21-22]。其中, 每个情境评估难度下的完整任务分为起飞、巡航和着陆 3 个任务阶段, 不同任务阶段包含操纵油门、操纵襟翼等不同的任务单元; 各情境评估难度下的界面显示特征在杂乱度、信息布局及突显性等维度也有所差异, 总体而言, 杂乱度越低, 信息布局越优, 突显性越高, 则情境评估难度越低。

任务情境设计如表 1 所示。具体而言, 在低情境评估难度下, 仪表显示模式正常, 被试从布局位置较优的电子仪表读取飞行速度和高度信息; 地图显示模式正常, 表现为信息组织有序且突显性高; 天气设置为目视飞行条件(visual flight rules, VFR), 被试在着陆过程可以清晰看见跑道和下滑道指示灯。在中情境评估难度下, 速度电子仪表显示正常, 高度电子仪表显示故障, 被试需要从布局位置较差的备份机械仪表读取飞行高度信息; 地图信息显示杂乱且突显性高; 天气设置为临界目视飞行条件(marginal visual flight rules, MVFR), 被试在着陆过程能看见跑道和指示灯, 有云层干扰, 视野较为模糊。在高情境评估难度下, 速度和高度电子仪表均显示故障, 被试需要从布局位置均较差的备份机械仪表读取飞行速度和高度信息; 地图信息显示杂乱且突显性较低; 天气设置为仪表飞行条件(instrument flight rules category1, IFR CAT1), 即天气能见度差, 着陆阶段难以识别跑道。

表 1　不同情境评估难度条件下的任务情境设计

Table 1　Task situation design under different situation difficulties

情境评估难度	气象条件	仪表显示模式	地图显示模式	界面示意图
低	VFR	正常显示	正常模式	
中	MVFR	混合显示	信息杂乱	
高	IFR CAT1	备份显示	信息杂乱 突显性低	

2.4　实验任务与流程

实验任务为 5 边模拟飞行任务, 被试需要在 3 种不同任务情境评估难度条件下各完成一次模拟飞行任务。完整任务分为起飞、巡航和着陆 3 个阶段。其中, 起飞阶段被试手动驾驶飞机从 36R 跑道起飞, 爬升至 700 ft (1 ft=0.304 8 m)后左转 90°进入第 2 边开始巡航, 继续向前飞行与 36L 跑道相切时, 高度爬升至 1 000 ft, 向左转 90°进入第 3 边开始巡航直飞, 巡航直飞阶段高度保持 1 000 ft, 飞行至地标时左转 90°进入第 4 边, 最后转入第 5 边开始着陆阶段进近降落在跑道上。具体任务参数如图 3 所示。

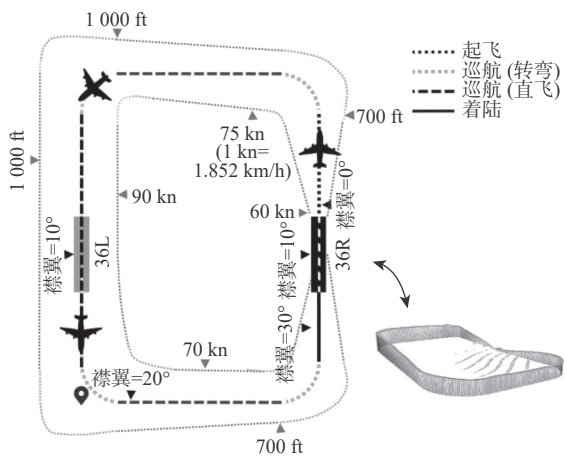


图 3　5 边模拟飞行任务轨迹示意图

Fig. 3　Trajectory of the simulated flight task in airfield traffic pattern

实验包括培训和正式实验 2 个阶段。每名被试的培训时长约为 3～4 h, 培训内容包括熟悉飞行界面仪表参数, 适应模拟飞行的操作特性, 掌握起飞、爬升、转弯、进近、降落等基本技能。经过充分培训后, 被试需达到以下绩效标准: ①能够对飞行界面信息的呈现位置及仪表读数、任务操作流程、关键任务参数进行完整准确复述; ②能够在 10 min 内按照预设飞行轨迹完成单次 5 边飞行模拟任务; ③能够成功进行跑道着陆。正式实验前, 被试签订知情同意书、佩戴生理设备并进行校标。在正式实验阶段, 单次模拟飞行任务时长约 10 min,

每2次模拟飞行任务之间休息5 min。为避免疲劳效应和练习效应，将3次模拟飞行任务的实验顺序进行交叉平衡。模拟飞行任务结束后，被试填写主观评价量表。

2.5 测量方法与指标

测量方法包括主观测评法、绩效测量法和眼动测量法。主观测评法用于测量被试的情境意识与工作负荷，包括SART量表、机组意识评定量表(crew awareness rating scale, CARS)及美国国家航空航天局-任务负荷指数量表(national aeronautics and space administration task load index, NASA-TLX)。

结合本次模拟飞行任务特征，采用绩效测量法实时采集被试在模拟飞行过程中的航向角、俯仰角、滚转角等关键飞行数据，通过计算其平均值$\bar{x}$、方差σ及变异系数c_v以建立不同任务阶段的综合绩效评价指标，从而对任务操纵绩效进行量化评价。本文中，完整飞行任务的操纵绩效e由起飞、巡航和着陆任务阶段的绩效e_T、e_C、e_L综合度量，各阶段的绩效由操纵变异度(即俯仰、航向及滚转姿态的波动程度)及操纵偏离度(即俯仰、航向及滚转姿态的偏差程度)的综合评价结果进行度量[23]，计算如式(8)所示：

$$e=(e_T+e_C+e_L)/3 \tag{8}$$

不同任务阶段的操纵特点如表2所示。结合各任务阶段的操纵要求和操纵特点，提取操纵绩效评价指标，如表3所示，其中，下标p、r、R、y、t分别表示俯仰角、滚转角、滚转率航向角和转弯率。

表2 不同任务阶段的操纵特点

Table 2 Manipulation characteristics of different task stages

任务阶段	操纵特点
起飞	保持10°～20°俯仰角稳定爬升高度，航向保持不变
巡航(直飞)	保持1 000 ft高度平稳飞行，航向保持不变，俯仰角和滚转角保持0°
巡航(转弯)	保持10°～20°滚转角进行转弯，高度保持不变，俯仰角保持0°
着陆	保持0°～10°俯仰角稳定降低高度，稳定调整航向和滚转角以对准跑道

表3 不同任务阶段的操纵绩效评价指标

Table 3 Manipulation performance evaluation indicators for different task stages

任务阶段	操纵绩效评价指标
起飞	σ_p、$\bar{x}_t$、c_{vR}
巡航(直飞)	σ_p、σ_y、$\bar{x}_R$、$\bar{x}_t$
巡航(转弯)	σ_p、c_{vr}、c_{vt}
着陆	σ_p、σ_r、σ_y、$\bar{x}_R$、$\bar{x}_t$

采用眼动测量法对被试在模拟飞行过程中的注视行为和扫视行为进行实时监测。相关研究表明，注视熵是指通过量化注视空间的分散程度来评估注视行为的一种方法，其可以反映信息搜索幅度，对于需要专注于特定操作的任务，注视熵越小，通常反映出作业人员更为稳定集中的注意力及更好的情境意识水平[24]。注视熵s的计算方法如下：

$$s=-\sum_{i=1}^{n}P_i\log_2 P_i \tag{9}$$

$$P_i=x_i/(x_1+\cdots x_i+\cdots+x_n) \tag{10}$$

式中：x_i为某兴趣区的注视点数目；P_i为某兴趣区的注视概率。

视觉扫视行为主要反映作业人员对特定视觉信息的捕获，可用于揭示眼球运动和注意转移之间的关系[25]，相关研究表明，在作业人员情境意识较低时，可能会出现扫视速度增加、扫视次数减少等特征[24,26-27]。

3 实验结果和分析

3.1 模型理论计算结果

首先，将5边飞行任务流程分解为多个任务单元，并依据VACP多资源理论模型对各任务单元的通道资源占有量和占用时长进行评定，根据式(2)～式(4)可计算出各任务单元的注意资源供需差Δ_{SD}。同时，依据DFSAM方法，对不同任务情境评估难度条件下的情境理解度因素进行分级量化。具体考量的人机界面设计特征包括界面显示的杂乱度、信息布局优劣程度及突显性，经专家评定，不同人机界面设计特征可支撑的情境理解度如表4所示。最后，将综合评定结果进行标准化得到任务单元的情境理解度U。

表4 界面显示格式评估维度对情境意识的支撑程度

Table 4 Support for situation awareness by interface display format evaluation dimension

人机界面设计特征	程度	情境理解度
界面显示的杂乱度	低	0
	中	−1
	高	−2
界面显示的信息布局优劣程度	优	−0.75
	良	−1.5
	差	−2
界面显示的突显性	低	0
	中	1.2
	高	1.5

不同任务阶段的任务单元及操作流程不同，根据式(7)可得出各任务阶段的情境意识理论值，其中，q值依据任务的资源供需差与情境理解度的相

对重要程度进行评定, 本文结合任务情境分析, 认为资源供需差与情境理解度对于维持人员情境意识的贡献度等同, 故 q 取值为 0.5。进而可根据式 (1) 得出面向完整任务及各任务阶段的情境意识理论计算结果, 如表 5 所示。可见, 在模拟飞行的各任务阶段, 随着情境评估难度的增大, 作业人员的情境意识不断下降。

表 5　情境意识模型的理论计算结果

Table 5　Theoretical calculation results of situation awareness model

情境评估难度	情景意识水平			
	起飞	巡航	着陆	完整任务
低	1.142 9	1.384 6	1.250 0	1.285 7
中	0.883 6	1.125 4	0.990 7	1.026 5
高	0.625 4	0.880 3	0.668 5	0.756 1

3.2　实验测量结果

完整任务阶段下的实验测量结果如表 6 所示。对于完整任务阶段的主观测量结果, 随着情境评估难度的增加, CARS 量表(检验统计量 F=30.53, 显著性 p<0.001, 组间差异效应量 η^2=0.412、Cohen's f=0.838)和 SART 量表(F=41.45, p<0.001, η^2=0.488, Cohen's f=0.976)得分显著降低, NASA-TLX 量表(F=42.11, p<0.001, η^2=0.492, Cohen's f=0.984)得分显著增加, 且组间差异较大(η^2>0.14, Cohen's f>0.4)。

表 6　完整任务阶段实验测量结果

Table 6　Measurement result of the complete task phase

测量方法与指标	平均值(标准差)			F	p
	低情境评估难度	中情境评估难度	高情境评估难度		
CARS	13.23(1.38)	12.53(1.53)	10.20(1.79)	30.53	<0.001
SART	7.27(2.03)	5.43(2.08)	2.23(2.37)	41.45	<0.001
NASA-TLX	40.81(10.27)	47.80(11.42)	66.50(11.89)	42.11	<0.001
注视熵	1.33(0.23)	1.53(0.21)	1.62(0.18)	15.52	<0.001
扫视次数	297.86(92.24)	260.97(82.55)	224.58(76.49)	5.71	0.005
操纵绩效	3.08(0.76)	3.34(0.53)	4.14(0.65)	21.75	<0.001

对于完整任务阶段的眼动测量结果, 随着情境评估难度的增加, 注视熵(F=15.52, p<0.001, η^2=0.263, Cohen's f=0.597)显著增加, 且组间差异较大(η^2>0.14, Cohen's f>0.4); 扫视次数(F=5.71, p=0.005, η^2=0.116, Cohen's f=0.362)显著减少, 组间差异中等(η^2>0.06, Cohen's f>0.25)。

对于完整任务阶段的操纵绩效测量结果, 随着情境评估难度的增加, 操纵绩效(F=21.75, p <0.001, η^2=0.333, Cohen's f=0.707)显著增加, 且组间差异较大(η^2>0.14, Cohen's f>0.4)。

3.3　相关性分析结果

将完整任务阶段下情境意识模型的理论计算结果与实验测量结果进行相关性分析。对于主观测评结果, 模型理论计算结果与 SART 量表得分显著正相关(相关系数 r=0.72, 显著性 p<0.001), 与 CARS 量表得分显著正相关(r=0.609, p<0.001), 与 NASA-TLX 量表得分显著负相关(r=−0.67, p<0.001); 对于眼动测量结果, 模型理论计算结果与注视熵显著负相关(r=−0.503, p<0.001), 与扫视次数显著正相关(r=0.351, p=0.003); 对于操纵绩效结果, 模型理论计算结果与操纵绩效显著负相关(r=−0.611, p<0.001)。

进一步, 分别对起飞、巡航和着陆阶段的不同情境评估难度条件下的情境意识理论计算结果 A_T、A_C、A_L 与实验测量结果做相关性分析, 结果如图 4 所示, 其中: *表示 p<0.05; **表示 p<0.01; ***表示 p<0.001。

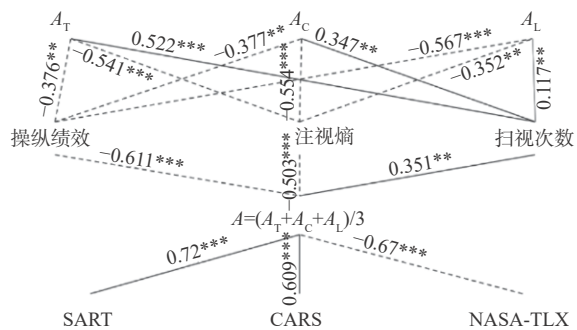


图 4　起飞、巡航、着陆阶段不同情境评估难度条件下相关性分析结果

Fig. 4　Results of correlation analysis under different situation difficulties in take-off, cruise, and landing stages

起飞阶段模型理论计算结果与操纵绩效显著负相关(r=−0.376, p=0.001), 与注视熵显著负相关(r=−0.541, p<0.001), 与扫视次数显著正相关(r=0.522, p<0.001); 巡航阶段模型理论计算结果与操纵绩效显著负相关(r=−0.377, p=0.001), 与注视熵显著负相关(r=−0.554, p<0.001), 与扫视次数显著正相关(r=0.347, p=0.003); 着陆阶段模型理论计算结果与操纵绩效显著负相关(r=−0.567, p<0.001), 与注视熵显著负相关(r=−0.352, p=0.002), 与扫视

次数正相关(r=0.117, p>0.05)。

4 数据结果分析

为了进一步满足航空器人机界面的研制需求,开展航空器任务和显控界面设计紧密结合特点下的作业人员情境意识水平的理论预测,提出一种基于资源供需差与情境理解度的情境意识量化分析模型。为对该模型进行验证,首先,构造了不同情境评估难度下的飞行模拟任务以诱发人员不同的情境意识状态;进一步,结合5边飞行模拟任务开展多维度的情境意识测量并筛选敏感性测量指标;最后,获得情境意识模型的理论计算结果,并将其与敏感指标进行相关性分析,以完成对模型的有效性验证。

本文主观测量结果在完整任务阶段下对情境意识测量敏感。如表6所示,NASA-TLX量表得分随情境评估难度增加而显著增加,而CARS量表得分和SART量表得分均随情境评估难度的增加而显著下降,说明在高情境评估难度任务条件下,作业人员的工作负荷更高,而情境意识水平更低。上述结果与前人研究发现[4,14,28]一致。随着情境评估难度的增加,作业人员的工作负荷上升,注意资源的消耗迅速增加[29],从而导致人员对任务情境元素的感知、理解和预测均更为困难[18],表现为情境意识水平的不断削弱。这一现象可为模型假设“工作负荷与情境意识存在负相关关系”提供实验验证依据。

操纵绩效和眼动测量具有实时和客观的优点,可以间接反映作业人员的情境意识,但它们在不同情境意识水平下的变化趋势尚缺乏一致性研究结论[27,30-31]。面向这一问题,探讨绩效指标和眼动指标对于情境意识评估的可用性。在完整任务阶段下,操纵绩效、眼动注视熵和扫视次数均对不同情境评估难度下的情境意识测量敏感,表现为随着情境评估难度的增加,作业人员的操纵绩效(操纵精度与操纵稳定性)逐渐恶化,注视熵逐渐增加,而扫视次数逐渐减少。考虑到注视熵增加通常与注意力分散相关[24],而扫视次数减少通常与信息搜索困难相关[27],因此,这一现象提示作业人员在情境意识较低时表现为飞行绩效显著下降[32],注意力更加分散,且信息搜索效率低下[33]。验证了绩效指标与眼动指标对于从多个维度测量与描述作业人员情境意识的必要性与有效性。

5 结 论

提出了基于资源供需差与情境理解度的情境意识量化分析模型,并开展不同情境评估难度下的飞行模拟实验对模型进行了初步验证。具体结论如下:

1)本文提高了模型效度验证的样本量(即采用被试全集数据而非均值数据),从而降低随机因素对预测结果造成的影响,使模型更为客观地反映总体情况。情境意识模型的理论计算结果与主观测量结果(SART量表、CARS量表、NASA-TLX量表)和客观测量结果(操纵绩效、注视熵、扫视次数)均呈现出良好的相关性,从而验证了模型的有效性。

2)主观测量结果展现出对情境意识测量良好的敏感性,同时发现工作负荷与情境意识存在负相关关系。客观测量指标中操纵绩效与眼动注视熵对情境意识测量的敏感性较高。

3)建立的情境意识理论模型与航空器任务设计及显控界面设计结合紧密,适用于由个体操作者、不同显控界面和任务操作流程构成的常规飞行任务情境下人员情境意识的理论分析与量化表征。

本文仍存在一定局限性,首先,理论模型中需要结合专家经验确定的模型参数可能会引入主观判断偏差;其次,理论模型目前主要适用于常规飞行情境,对于异常飞行情境(如告警界面设计、紧急处置操作流程等),则需要获得更多经验数据的支持;最后,模拟飞行任务情境和真实飞行任务情境间仍存在不可忽视的差异,需在后续工作中对理论模型进一步优化和验证,以提高理论模型的有效性。

参考文献(References)

[1] ENDSLEY M R. Situation awareness global assessment technique [C]//Proceedings of the IEEE 1988 National Aerospace and Electronics Conference. Piscataway: IEEE Press, 2002: 789-795.

[2] 靳慧斌, 刘亚威, 朱国蕾. 基于眼动和绩效分析的管制员情境意识测量[J]. 中国安全科学学报, 2017, 27(7): 65-70.
JIN H B, LIU Y W, ZHU G L. Measuring situation awareness of tower controllers based on eye movement and performance analysis [J]. China Safety Science Journal, 2017, 27(7): 65-70(in Chinese).

[3] NGUYEN T, LIM C P, NGUYEN N D, et al. A review of situation awareness assessment approaches in aviation environments[J]. IEEE Systems Journal, 2019, 13(3): 3590-3603.

[4] 冯传宴, 完颜笑如, 陈浩, 等. 基于多资源负荷理论的情境意识模型与应用[J]. 北京航空航天大学学报, 2018, 44(7): 1438-1446.
FENG C Y, WANYAN X R, CHEN H, et al. Situation awareness model based on multi-resource load theory and its application[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(7): 1438-1446(in Chinese).

[5] ZHOU F, YANG X J, DE WINTER J C F. Using eye-tracking data to predict situation awareness in real time during takeover transitions in conditionally automated driving[J]. IEEE Transactions on

Intelligent Transportation Systems, 2022, 23(3): 2284-2295.

[6] YEO L G, SUN H Q, LIU Y S, et al. Mobile EEG-based situation awareness recognition for air traffic controllers[C]//Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. Piscataway: IEEE Press, 2017: 3030-3035.

[7] KIM J, MYUNG R. A predictive model of situation awareness with ACT-R[J]. Journal of the Ergonomics Society of Korea, 2016, 35(4): 225-235.

[8] GORE B F, HOOEY B L , HAAN N, et al. A validated set of MIDAS V5 task network model scenarios to evaluate nextgen closely spaced parallel operations concepts: HCSL-13-03[R]. Moffett Field: NASA Ames Research Center, 2013: 1-35.

[9] WICKENS C D, GOH J, HELLEBERG J, et al. Attentional models of multitask pilot performance using advanced display technology [J]. Human Factors, 2003, 45(3): 360-380.

[10] MEYER D R. Effects of automation on aircrew workload and situation awareness in tactical airlift missions[D]. Fort Belvoir: Air Force Institute of Technology, 2015.

[11] TAYLOR R M. Situational awareness rating technique (SART): the development of a tool for aircrew systems design[M]//TAYLOR R M. Situational awareness. London: Routledge, 2017: 111-128.

[12] LEE Y J, JUNG K T, LEE H C. Use of gaze entropy to evaluate situation awareness in emergency accident situations of nuclear power plant[J]. Nuclear Engineering and Technology, 2022, 54(4): 1261-1270.

[13] SHEPHERD A. HTA as a framework for task analysis[J]. Ergonomics, 1998, 41(11): 1537-1552.

[14] FRIEDRICH M, BIERMANN M, GONTAR P, et al. The influence of task load on situation awareness and control strategy in the ATC tower environment[J]. Cognition, Technology & Work, 2018, 20(2): 205-217.

[15] PARASURAMAN R, SHERIDAN T B, WICKENS C D. Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs[J]. Journal of Cognitive Engineering and Decision Making, 2008, 2(2): 140-160.

[16] 田书婕, 王波, 王丽, 等. 基于时间-多资源占用的工作负荷评估模型[J]. 北京航空航天大学学报, 2017, 43(12): 2497-2504.
TIAN S J, WANG B, WANG L, et al. Workload evaluation model based on occupation of time and multi-resource[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(12): 2497-2504(in Chinese).

[17] MCCRACKEN J H, ALDRICH T B. Analyses of selected LHX mission functions: implications for operator workload and system automation goals: ASI 479-024[R]. Fort Rucker: ASI, 1984.

[18] GRIER R, WICKENS C, KABER D, et al. The red-line of workload: theory, research, and design[J]. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2008, 52(18): 1204-1208.

[19] KIM S K, SUH S M, JANG G S, et al. Empirical research on an ecological interface design for improving situation awareness of operators in an advanced control room[J]. Nuclear Engineering and Design, 2012, 253: 226-237.

[20] 肖旭, 完颜笑如, 庄达民. 显示界面多维视觉编码综合评价模型[J]. 北京航空航天大学学报, 2015, 41(6): 1012-1018.
XIAO X, WANYAN X R, ZHUANG D M. Comprehensive evaluation model of multidimensional visual coding on display interface [J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(6): 1012-1018(in Chinese).

[21] BLUNDELL J, SCOTT S, HARRIS D, et al. With flying colours: pilot performance with colour-coded head-up flight symbology[J]. Displays, 2020, 61: 101932.

[22] LIN C J, LIN P H, CHEN H J, et al. Effects of controller-pilot communication medium, flight phase and the role in the cockpit on pilots' workload and situation awareness[J]. Safety Science, 2012, 50(9): 1722-1731.

[23] 许卫宝. 飞行员着舰的多层次模糊综合评价方法[J]. 中国舰船研究, 2013, 8(2): 17-21.
XU W B. A comprehensive evaluation method for carrier landing based on fuzzy analytical hierarchy process[J]. Chinese Journal of Ship Research, 2013, 8(2): 17-21(in Chinese).

[24] 靳慧斌, 吕川, 朱国蕾. 塔台管制中情境意识与视觉搜索策略关系研究[J]. 重庆交通大学学报 (自然科学版), 2019, 38(1): 108-112.
JIN H B, LV C, ZHU G L. Relation between visual search strategy and situation awareness in tower control[J]. Journal of Chongqing Jiaotong University (Natural Science), 2019, 38(1): 108-112(in Chinese).

[25] STANTON N A, LI W C, HARRIS D. Editorial: ergonomics and human factors in aviation[J]. Ergonomics, 2019, 62(2): 131-137.

[26] KOEN V D M, HENK V D, ROLF Z. Eye movements as an indicator of situation awareness in a flight simulator experiment[J]. The International Journal of Aviation Psychology, 2012, 22(1): 78-95.

[27] FENG C Y, WANYAN X R, LIU S, et al. An approach to situation awareness (SA) assessment in flight simulation: SA dynamic circulation (SADC) model[J]. Human Factors and Ergonomics in Manufacturing & Service Industries, 2021, 31(5): 559-569.

[28] 冯传宴, 完颜笑如, 刘双, 等. 负荷条件下注意力分配策略对情境意识的影响[J]. 航空学报, 2020, 41(3): 123307.
FENG C Y, WANYAN X R, LIU S, et al. Influence of different attention allocation strategies under workloads on situation awareness[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(3): 123307(in Chinese).

[29] WICKENS C D, MCCARLEY J S, ALEXANDER A L, et al. Attention-situation awareness (A-SA) model of pilot error: AHFD-04-15/NASA-04-5[R]. Mountain View: NASA Ames Research Center, 2005.

[30] LIU S, WANYAN X R, ZHUANG D M. Modeling the situation awareness by the analysis of cognitive process[J]. Bio-Medical Materials and Engineering, 2014, 24(6): 2311-2318.

[31] JIANG S Q, CHEN W J, KANG Y T. Correlation evaluation of pilots' situation awareness in bridge simulations via eye-tracking technology[J]. Computational Intelligence and Neuroscience, 2021, 2021: 7122437.

[32] DE WINTER J C F, EISMA Y B, CABRALL C D D, et al. Situation awareness based on eye movements in relation to the task environment[J]. Cognition, Technology & Work, 2019, 21(1): 99-111.

[33] IVETA Š, PECHO P, AŽALTOVIČ V, et al. Number of saccades and fixation duration as indicators of pilot workload[J]. Transportation Research Procedia, 2020, 51: 67-74.

Situation awareness model based on resource supply-demand difference and understanding

WANG Yihang，LIU Shuang，WANYAN Xiaoru*，FENG Chuanyan，ZHOU Sunxia，QIAN Chunying

(School of Aeronautic Science and Engineering，Beihang University，Beijing 100191，China)

Abstract：To further meet the development needs for human-machine interface of aviation equipment and to theoretically predict pilots' situation awareness (SA) under the close combination of task with display and control interface design, this paper proposed a model for quantitative analysis of SA based on resource supply-demand difference and understanding. The model was based on the way the classical situation awareness rating technique (SART) scale defined SA, characterizing SA as the combination of resource supply-demand difference and understanding. The understanding was measured by multidimensional human-machine interface display format, while the resource supply-demand difference of task units was measured by combining task network analysis with multiple resource theory. To verify model validity, 30 subjects were selected to complete the simulated flight task in the airfield traffic pattern under different situation difficulties. The situation difficulty was adjusted by changing the simulated weather conditions, instrument display modes, and navigation map display modes. The SA was assessed under different situation difficulties by using subjective scales, flight performance, and eye movement indicators. The data analysis results show that the SA model results have good correlations with subjective (SART, crew awareness rating scale (CARS), and national aeronautics and space administration task load index (NASA-TLX) scales) and objective (manipulation performance, gaze entropy, and number of saccades) measurements. This preliminarily verifies the validity of the model. The proposed model can provide a theoretical basis and technical methods for the analysis and prediction of SA in combination with task flow design and display and control interface design features.

Keywords：situation awareness；resource supply-demand difference；understanding；performance measurement；eye movement measurement

Received：2023-06-30；**Accepted**：2023-10-05；**Published Online**：2023-11-06 12：09

URL：link.cnki.net/urlid/11.2625.V.20231103.1743.002

Foundation items：Joint Fund of National Natural Science Foundation of China and Civil Aviation Administration of China (U1733118); Aeronautical Science Foundation of China (201813300002); National Natural Science Foundation of China (71301005)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0435

冷云辐射对超压气球热动力学特性影响

贺澜[1,2]，刘强[1,*]，杨燕初[1,2]，祝榕辰[1]，周江华[1,2]

(1. 中国科学院空天信息创新研究院，北京 100094；2. 中国科学院大学，北京 101408)

摘　　要：针对冷云辐射对超压气球热动力特性影响的问题，提出融合冷云辐射特性的超压气球热动力学模型。研究不同相对高度条件下，冷云对超压气球热环境的影响。分析冷云高度、厚度、云粒子有效半径、温度、含水量等特征，建立冷云的辐射吸收、透过和反射特性模型。在动力学、运动学和热力学分析的基础上，建立包含冷云辐射特性的超压气球热动力学模型，开发仿真计算程序，利用实测数据验证热动力学模型和仿真程序的准确性。利用风云卫星冷云数据，仿真分析冷云辐射影响下的超压气球热动力学特性，结果显示：冷云会降低超压气球温度，影响超压气球升空速度和高度，当云层厚度或云层温度到达一定条件时，超压气球将无法抵达预期驻留高度。

关 键 词：超压气球；热力学特性；动力学特性；冷云环境；辐射特性

中图分类号：V273

文献标志码：A　　**文章编号**：1001-5965(2025)09-3117-13

浮空器具有飞行高度范围广、飞行时间长、运行成本低、搭载设备分辨率高、安全性好等优势，在对地观测、中继通信等方面具有广阔前景，成为当前研究热点[1]。相关领域的巨大应用需求，要求浮空器具备全天候升空作业的能力，因此，在升空-平飞-下降过程中遇到云层的可能性极大。

云层对浮空器热环境有较大的影响。云层可以吸收和反射外部辐射，同时也向外发射红外辐射，浮空器穿过低温云层时也会与云层进行对流换热，这些因素会改变浮空器的热环境，影响浮空器热动力学特性。

在云层对浮空器热动力学特性影响方面，研究者大多通过判断云量，引入遮云系数简化计算云层对浮空器热特性的影响。1981年，Bird等[2]建立了遮云系数与云量和太阳天顶角的关系式。同年，Nielsen等[3]建立了遮云系数与云量、太阳高度角和太阳辐射总量的关系式。2005年，Farley[4]研究了不同相对高度条件下，云层对气球的热环境及热动力学特性的影响。2019年，程晨[5]利用云层覆盖率计算云层反射辐射特性，并分析了无云、少云和多云情况下浮空器的热特性。

本文分析冷云的高度、厚度、云粒子有效半径、温度、含水量等特征，引入冷云的光学厚度作为表征参数，建立冷云的辐射吸收、透过和反射特性模型，研究云层对浮空器热动力学特性的影响，为浮空器设计研制和任务规划提供参考。

1 浮空器热环境

1.1 大气环境

根据标准大气模型，海拔高度H对应的大气温度、压强和密度计算公式如下[6]：

$$T_{\text{air}}=\begin{cases}288.15-0.006\,5H & 0<H\leqslant 11\,000\\ 216.65 & 11\,000<H\leqslant 20\,000\\ 216.65+0.001\,0(H-20\,000) & 20\,000<H\leqslant 32\,000\end{cases} \tag{1}$$

收稿日期：2023-07-04；**录用日期**：2023-09-14；**网络出版时间**：2023-10-10 14:47
网络出版地址：link.cnki.net/urlid/11.2625.V.20231009.1320.003
基金项目：国家重点研发计划 (2022YFB3903005-01)；中国科学院青年创新促进会 (2020133)
*通信作者. E-mail：liuqiang1001@aoe.ac.cn

引用格式：贺澜，刘强，杨燕初，等. 冷云辐射对超压气球热动力学特性影响 [J]. 北京航空航天大学学报，2025，51（9）：3117-3129.
HE L，LIU Q，YANG Y C，et al. Influence of cold cloud radiation on thermal-dynamic characteristics of super-pressure balloon [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（9）：3117-3129（in Chinese）.

$$P_{\text{air}}=\begin{cases}101\ 325[(288.15-0.006\ 5H)/288.15]^{5.255\ 77} & 0<H\leqslant 11\ 000\\ 22\ 632\mathrm{e}^{[-(H-11\ 000)/6\ 341.62]} & 11\ 000<H\leqslant 20\ 000\\ 5\ 474.87\{[216.65+0.001\ 0(H-20\ 000)]/216.65\}^{-34.163} & 20\ 000<H\leqslant 32\ 000\end{cases} \tag{2}$$

$$\rho_{\text{air}}=\frac{P_{\text{air}}}{R_{\text{air}}T_{\text{air}}} \tag{3}$$

式中：T_{air}为大气温度；P_{air}为大气压强；ρ_{air}为大气密度；R_{air}为空气气体常数。

1.2 热辐射环境

浮空器是一种热力学与动力学特性高度耦合的飞行器，在飞行过程中面临复杂的热环境。热环境主要因素有太阳短波辐射、长波辐射和对流换热3类。太阳短波辐射包括太阳直射辐射、天空散射辐射、地表及云层反射辐射；长波辐射包括大气红外辐射、地面红外辐射、云红外辐射及浮空器内表面单元之间的辐射换热；对流换热包括浮空器外表面与外部大气环境的对流换热、浮空器内表面与内部填充气体的对流换热[5]。

当浮空器处于云层之下时，其热环境如图1所示。

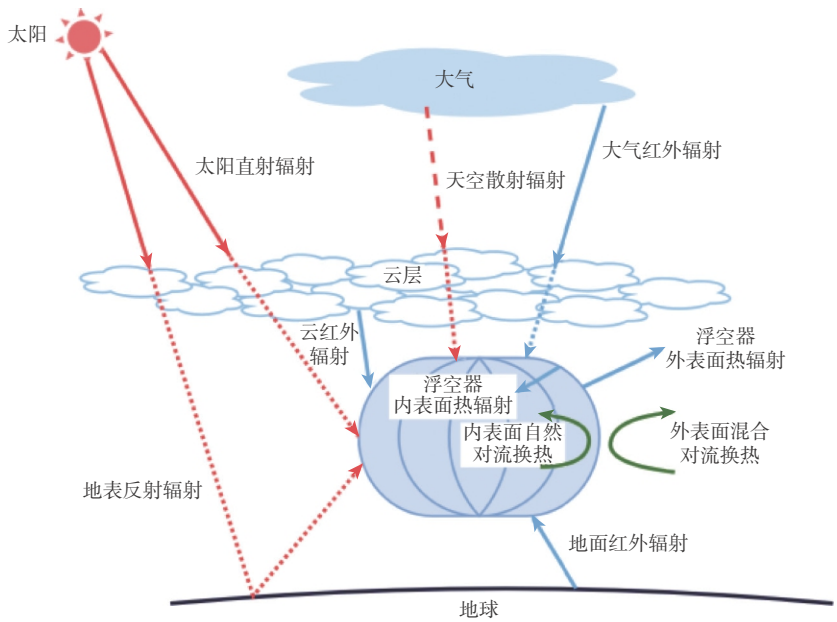


图1 浮空器在云层之下的热环境

Fig. 1 Thermal environment of aerostat under cloud

云层对太阳直射辐射、天空散射辐射、地表反射辐射及大气红外辐射均有1层透射衰减作用，同时，浮空器上表面会接收云层发射的红外辐射。

当浮空器处于云层之中时，其热环境如图2所示。

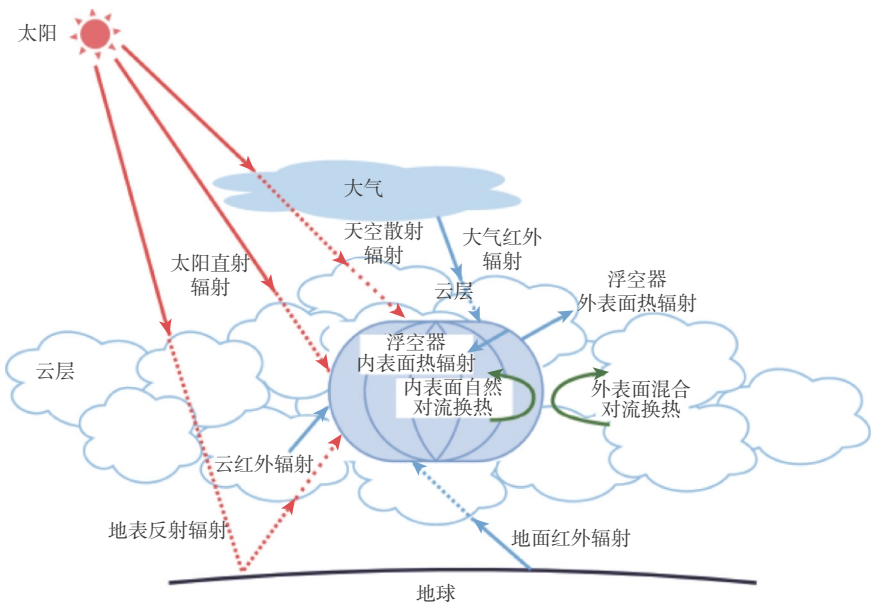


图2 浮空器在云层之中的热环境

Fig. 2 Thermal environment of aerostat in cloud

云层对太阳直射辐射、天空散射辐射、地面红外辐射、大气红外辐射均有 1 层透射衰减作用，对地表反射的太阳辐射有 2 层透射衰减作用，浮空器受到云自身发射的红外辐射，同时，云层与浮空器之间会基于温差引发对流换热。

当浮空器处于云层之上时，其热环境如图 3 所示。

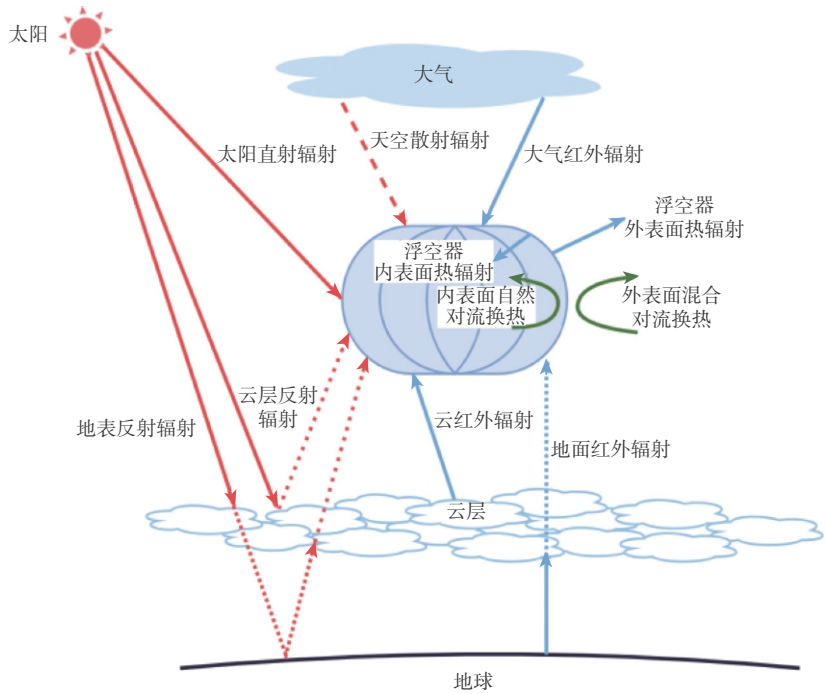


图 3　浮空器在云层之上的热环境

Fig. 3　Thermal environment of aerostat above cloud

云层对地表反射辐射有 2 层透射衰减作用，对地面红外辐射有 1 层透射衰减作用，浮空器下表面会接收云层反射辐射及云自身发射的红外辐射。

2　冷云辐射模型

2.1　冷云基本情况

冷云指有冰相过程参与的云，温度低于大气温度，对浮空器热辐射环境有较大影响，主要体现在云层的辐射透过率、反射率和发射率，其计算所需的云层高度、厚度、云粒子有效半径、温度、含水量等特征参数取值又因云的类型或分布高度不同而有所差异。

根据云底高度，可将云大致分成 3 类：低云、中云和高云，同时，根据云状等特征又细分为 10 种，如图 4 所示[7]。

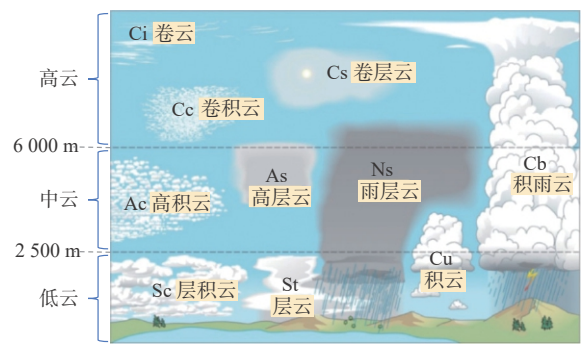


图 4　云的分类[7]

Fig. 4　Classification of cloud[7]

2.2　冷云热辐射特性模型

2.2.1　冷云的光学厚度、单次散射比和不对称因子

当辐射波长 $\lambda<4$ μm 时，云的光学厚度 ζ_λ、单次散射比 ω_0 和不对称因子 ς_0 计算如下[8]：

$$\zeta_\lambda = w_{\mathrm{LWP}}\left(a_1+\frac{a_2}{r_\mathrm{e}}\right) \tag{4}$$

$$w_{\mathrm{LWP}} = \int_{H_{\mathrm{cloudbase}}}^{H_{\mathrm{cloudtop}}} w_{\mathrm{LWC}}\mathrm{d}z \tag{5}$$

$$1-\omega_0 = a_3+a_4 r_\mathrm{e} \tag{6}$$

$$\varsigma_0 = a_5+a_6 r_\mathrm{e} \tag{7}$$

式中：a_i 为拟合系数；r_e 为云粒子有效半径；w_{LWC} 为云的含水量；$H_{\mathrm{cloudbase}}$ 为云底高度；H_{cloudtop} 为云顶高度；z 为云层厚度。

当辐射波长 $\lambda>4$ μm 时，云的光学厚度 ζ_λ、单次散射比 ω_0 和不对称因子 ς_0 计算如下[9]：

$$\zeta_\lambda = w_{\mathrm{LWC}} z \varphi_\lambda \tag{8}$$

$$\varphi_\lambda = b_1+b_2 r_\mathrm{e}+b_3/r_\mathrm{e}+b_4/r_\mathrm{e}^2+b_5/r_\mathrm{e}^3 \tag{9}$$

$$1-\omega_0 = c_1+c_2/r_\mathrm{e}+c_3 r_\mathrm{e}+c_4 r_\mathrm{e}^2 \tag{10}$$

$$\varsigma_0 = d_1+d_2/r_\mathrm{e}+d_3 r_\mathrm{e}+d_4 r_\mathrm{e}^2 \tag{11}$$

式中：b_i、c_i、d_i 为拟合系数。

2.2.2 冷云的透过率、反射率和发射率

云的透过率τ_c计算如下[10]:

$$\tau_c=\frac{4u}{B} \tag{12}$$

$$u^2=\frac{1-\omega_0+2\beta(\chi_0)\omega_0}{1-\omega_0} \tag{13}$$

$$B=(u+1)^2\exp(\zeta_{\text{eff}})-(u-1)^2\exp(-\zeta_{\text{eff}}) \tag{14}$$

$$\zeta_{\text{eff}}=\frac{\left[(1-\omega_0)(1-\omega_0+2\beta(\chi_0)\omega_0)\right]^{1/2}\zeta_\lambda}{\chi_0} \tag{15}$$

式中: χ_0为辐射入射角余弦; $\beta(\chi_0)$为反向散射比, 可由下式计算[11]:

$$\beta(\chi_0)=\frac{\int_0^1 p\mathrm{d}\chi}{\int_{-1}^1 p\mathrm{d}\chi} \tag{16}$$

式中: χ为散射角余弦; p为相函数, 表示散射光在各方向的分布, 云的米散射相函数近似为 Henyey-Greenstein 函数[12]:

$$p=\frac{1-\varsigma_0^2}{(1+\varsigma_0^2-2\varsigma_0\chi)^{3/2}} \tag{17}$$

云的反射率ϑ_c计算如下[13]:

$$\vartheta_c=\frac{\vartheta^\infty\left[\exp(\delta\zeta_\lambda)-\exp(-\delta\zeta_\lambda)\right]}{\exp(\delta\zeta_\lambda)-\vartheta^\infty\exp(-\delta\zeta_\lambda)} \tag{18}$$

$$\delta=2\sqrt{(1-\omega_0)(1-\omega_0\varsigma_0)} \tag{19}$$

$$\vartheta^\infty=\frac{\sqrt{1-\omega_0\varsigma_0}-\sqrt{1-\omega_0}}{\sqrt{1-\omega_0\varsigma_0}+\sqrt{1-\omega_0}} \tag{20}$$

式中: ϑ^∞为云反射率的上限。

将云自身发射的红外热辐射近似成发射率为0.97～0.99的灰体辐射, 反射比为0[14], 则云的发射率ε_c计算如下[15]:

$$\varepsilon_c=1-a_0\exp(-b_0w_{\text{LWC}}z) \tag{21}$$

式中: a_0、b_0为系数, 层云取a_0=0.92, b_0=0.27, 高层云取a_0=0.95, b_0=0.20, 积云取a_0=0.96, b_0=0.16。

3 超压气球热动力学特性模型

3.1 球体几何模型

超压气球体积计算如下[4]:

$$V=m_{\text{He}}R_{\text{He}}T_{\text{He}}/P_{\text{He}} \tag{22}$$

式中: V为球体体积; m_{He}为球体内部氦气质量; R_{He}为氦气气体常数; T_{He}为氦气温度; P_{He}为氦气压强。

在俯视平面上气球直径为[4]

$$d=1.424V^{1/3} \tag{23}$$

膨胀成型时球体表面积为[4]

$$A_{\text{surf}}\approx 5.3V^{2/3} \tag{24}$$

顶部投影面积为[4]

$$A_{\text{top}}=\pi d^2/4 \tag{25}$$

被太阳直接辐射照射的投影区域面积为[4]

$$A_{\text{projected}}=A_{\text{top}}(0.821\,9+0.178\,1\cos(\pi-2l)) \tag{26}$$

式中: l为太阳高度角。

3.2 动力学与运动学模型

3.2.1 受力特性

超压气球在飞行过程中主要受浮力F、系统重力G和气动阻力D影响, 如图5所示。

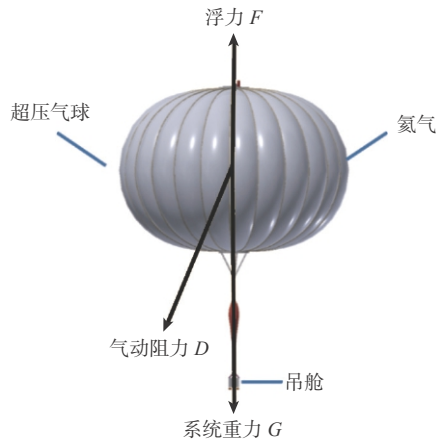


图5 超压气球受力分析

Fig. 5 Force analysis of super-pressure balloon

浮力、系统重力和气动阻力计算如下[16]:

$$F=\rho_{\text{air}}Vg \tag{27}$$

$$G=(m_{\text{film}}+m_{\text{He}}+m_{\text{payload}})g \tag{28}$$

$$D=0.5\rho_{\text{air}}v_H^2C_dA_{\text{top}} \tag{29}$$

式中: m_{film}为球膜质量; m_{payload}为有效载荷质量; g为重力加速度; v_H为气球上升速度; C_d为阻力系数。

3.2.2 动力学模型

忽略球体姿态影响, 将超压气球系统视为质点, 超压气球升空过程的动力学微分方程为[16]

$$M_V\ddot{H}=F-D-G \tag{30}$$

式中: $\ddot{H}$为气球飞行加速度; M_V为气球总质量, 计算如下:

$$M_V=m_{\text{film}}+m_{\text{He}}+m_{\text{payload}}+C_{\text{vir}}\rho_{\text{air}}V \tag{31}$$

其中: $C_{\text{vir}}\rho_{\text{air}}V$为被气球拖拽的空气引起的虚拟附加质量, C_{vir}为虚拟质量系数。

3.2.3 运动学模型

气球在高度方向上的运动学微分方程为[16]

$$\ddot{H}=\mathrm{d}v_H/\mathrm{d}t \tag{32}$$

式中: t为时间。

3.3 热力学模型

3.3.1 球膜热力学模型

因球膜厚度在20～30 μm量级, 沿厚度方向的

热传导可忽略不计，球膜平均温度的变化率由以下瞬态能量平衡公式计算：

$$\frac{\mathrm{d}T_{\mathrm{film}}}{\mathrm{d}t}=(Q_{\text{s-direct}}+Q_{\text{s-scatter}}+Q_{\text{s-earth}}+Q_{\text{s-cloud}}+Q_{\text{IR-earth}}+Q_{\text{IR-atm}}+Q_{\text{IR-cloud}}+Q_{\text{IR-filmabsorb}}+Q_{\text{IR-filmout}}+Q_{\text{Conv-ext}}+Q_{\text{Conv-int}})/(c_{\mathrm{f}}m_{\mathrm{film}}) \tag{33}$$

式中：T_{film}为球膜平均温度；$Q_{\text{s-direct}}$为吸收的太阳直射辐射热量；$Q_{\text{s-scatter}}$为吸收大气散射辐射热量；$Q_{\text{s-earth}}$为吸收的地表反射辐射热量；$Q_{\text{s-cloud}}$为吸收的云层反射辐射热量；$Q_{\text{IR-earth}}$为吸收的地面红外辐射热量；$Q_{\text{IR-atm}}$为吸收的大气红外辐射热量；$Q_{\text{IR-cloud}}$为吸收的云层红外辐射热量；$Q_{\text{IR-filmabsorb}}$为吸收的球体内部辐射热量；$Q_{\text{IR-filmout}}$为球膜表面发射的长波辐射热量；$Q_{\text{Conv-ext}}$为球膜与外部环境的对流换热热量；$Q_{\text{Conv-int}}$为球膜与内部氦气的对流换热热量；c_{f}为球膜比热容。

各分量计算公式如下：

1）球膜吸收的太阳直射辐射热量计算公式为

$$Q_{\text{s-direct}}=\begin{cases}I_{\text{s-direct}}A_{\text{projected}}\alpha_{\text{film}}[1+\tau_{\text{film}}(1+\vartheta_{\text{eff}})]\tau_{\text{c}} & H<H_{\text{cloudbase}}\\ I_{\text{s-direct}}A_{\text{projected}}\alpha_{\text{film}}[1+\tau_{\text{film}}(1+\vartheta_{\text{eff}})]\tau_{\text{c-up}} & H_{\text{cloudbase}}\leqslant H\leqslant H_{\text{cloudtop}}\\ I_{\text{s-direct}}A_{\text{projected}}\alpha_{\text{film}}[1+\tau_{\text{film}}(1+\vartheta_{\text{eff}})] & H>H_{\text{cloudtop}}\end{cases} \tag{34}$$

式中：α_{film}为球膜对可见光的吸收率；τ_{film}为球膜对可见光的透过率；$\tau_{\text{c-up}}$为超压气球穿过云层时，随超压气球飞行高度与云顶高度之差不断变化的云层上部透过率；$I_{\text{s-direct}}$为抵达球膜的太阳直接辐射强度，计算公式如下[4]：

$$I_{\text{s-direct}}=I_0\tau_{\text{atm}} \tag{35}$$

其中：I_0为地球大气层上边界处的太阳直射辐射强度，因一年中日期变化有所不同，具体计算详见文献 [17]；τ_{atm}为大气透过率，随高度与太阳辐射倾角等参数而变化，具体计算详见文献 [4]。

ϑ_{eff}为球膜内部有效反射率，计算公式为[4]

$$\vartheta_{\text{eff}}=\vartheta+\vartheta^2+\vartheta^3+\vartheta^4+\vartheta^5+\ldots \tag{36}$$

式中：ϑ为球膜的反射率。

2）球膜表面吸收的大气散射辐射热量计算公式为

$$Q_{\text{s-scatter}}=\begin{cases}I_{\text{s-scatter}}A_{\text{surf}}\alpha_{\text{film}}[1+\tau_{\text{film}}(1+\vartheta_{\text{eff}})]\tau_{\text{c}} & H<H_{\text{cloudbase}}\\ I_{\text{s-scatter}}A_{\text{surf}}\alpha_{\text{film}}[1+\tau_{\text{film}}(1+\vartheta_{\text{eff}})]\tau_{\text{c-up}} & H_{\text{cloudbase}}\leqslant H\leqslant H_{\text{cloudtop}}\\ I_{\text{s-scatter}}A_{\text{surf}}\alpha_{\text{film}}[1+\tau_{\text{film}}(1+\vartheta_{\text{eff}})] & H>H_{\text{cloudtop}}\end{cases} \tag{37}$$

式中：$I_{\text{s-scatter}}$为大气散射辐射强度，计算公式为[18]

$$I_{\text{s-scatter}}=I_0\tau_{\text{atm}}\kappa \tag{38}$$

其中：κ为大气散射系数。

3）球膜表面吸收的地表反射辐射热量计算公式为

$$Q_{\text{s-earth}}=\begin{cases}I_{\text{albedo}}A_{\text{surf}}\alpha_{\text{film}}X_{\text{e-b}}[1+\tau_{\text{film}}(1+\vartheta_{\text{eff}})]\tau_{\text{c}} & H<H_{\text{cloudbase}}\\ I_{\text{albedo}}A_{\text{surf}}\alpha_{\text{film}}X_{\text{e-b}}[1+\tau_{\text{film}}(1+\vartheta_{\text{eff}})]\tau_{\text{c}}\tau_{\text{c-below}} & H_{\text{cloudbase}}\leqslant H\leqslant H_{\text{cloudtop}}\\ I_{\text{albedo}}A_{\text{surf}}\alpha_{\text{film}}X_{\text{e-b}}[1+\tau_{\text{film}}(1+\vartheta_{\text{eff}})]\tau_{\text{c}}^2 & H>H_{\text{cloudtop}}\end{cases} \tag{39}$$

式中：$X_{\text{e-b}}$为球体表面与地球表面进行辐射换热的辐射角系数，具体计算详见文献 [4]；$\tau_{\text{c-below}}$为超压气球穿过云层时，随超压气球飞行高度与云底高度之差不断变化的云层下部透过率；I_{albedo}为地表反射辐射强度，计算公式如下[4]：

$$I_{\text{albedo}}=\vartheta_{\text{albedo}}I_{\text{ground}} \tag{40}$$

其中：$\vartheta_{\text{albedo}}$为地球表面的反射率，取平均值 0.3；$I_{\text{ground}}$为地表太阳辐射强度。

4）当超压气球在云层上方时，球膜吸收的云层反射辐射热量计算公式为

$$Q_{\text{s-cloud}}=\vartheta_{\text{c}}\left(I_{\text{s-direct}}\sin l\cdot A_{\text{projected}}+\frac{1}{2}I_{\text{s-scatter}}A_{\text{surf}}\right)\alpha_{\text{film}}\cdot[1+\tau_{\text{film}}(1+\vartheta_{\text{eff}})] \tag{41}$$

5）球膜表面吸收的地面红外辐射热量计算公式为

$$Q_{\text{IR-earth}}=\begin{cases}I_{\text{IR-earth}}A_{\text{surf}}\alpha_{\text{film-IR}}X_{\text{e-b}}[1+\tau_{\text{film-IR}}(1+\vartheta_{\text{eff}})] & H<H_{\text{cloudbase}}\\ I_{\text{IR-earth}}A_{\text{surf}}\alpha_{\text{film-IR}}X_{\text{e-b}}[1+\tau_{\text{film-IR}}\cdot(1+\vartheta_{\text{eff}})]\tau_{\text{c-below}} & H_{\text{cloudbase}}\leqslant H\leqslant H_{\text{cloudtop}}\\ I_{\text{IR-earth}}A_{\text{surf}}\alpha_{\text{film-IR}}X_{\text{e-b}}[1+\tau_{\text{film-IR}}(1+\vartheta_{\text{eff}})]\tau_{\text{c}} & H>H_{\text{cloudtop}}\end{cases} \tag{42}$$

式中：$\alpha_{\text{film-IR}}$为球膜对红外辐射的吸收率；$\tau_{\text{film-IR}}$为球膜对红外辐射的透过率；$I_{\text{IR-earth}}$为地面红外辐射强度，计算公式如下[4]：

$$I_{\text{IR-earth}}=\sigma\varepsilon_{\text{earth}}T_{\text{earth}}^4\tau_{\text{atm-IR}} \tag{43}$$

其中：σ为斯蒂芬-玻尔兹曼常数；$\varepsilon_{\text{earth}}$为地球表面平均发射率，取值为 0.95；$T_{\text{earth}}$为地球表面温度；$\tau_{\text{atm-IR}}$为地球表面红外辐射透射率。

6）球膜表面吸收的天空辐射热量计算公式为

$$Q_{\text{IR-atm}}=\begin{cases} I_{\text{IR-atm}}A_{\text{surf}}\alpha_{\text{film-IR}}(1-X_{\text{e-b}})[1+\tau_{\text{film-IR}}\cdot \\ \qquad (1+\vartheta_{\text{eff}})]\tau_{\text{c}} & H<H_{\text{cloudbase}} \\ I_{\text{IR-atm}}A_{\text{surf}}\alpha_{\text{film-IR}}(1-X_{\text{e-b}})[1+\tau_{\text{film-IR}}\cdot \\ \qquad (1+\vartheta_{\text{eff}})]\tau_{\text{c-up}} & H_{\text{cloudbase}}\leqslant H\leqslant H_{\text{cloudtop}} \\ I_{\text{IR-atm}}A_{\text{surf}}\alpha_{\text{film-IR}}(1-X_{\text{e-b}})[1+\tau_{\text{film-IR}}(1+\vartheta_{\text{eff}})] & \\ \qquad & H>H_{\text{cloudtop}} \end{cases} \tag{44}$$

式中：$I_{\text{IR-atm}}$为天空辐射强度，计算公式如下[5]：

$$I_{\text{IR-atm}}=\sigma\varepsilon_{\text{atm}}T_{\text{air}}^4 \tag{45}$$

其中：ε_{atm}为天空等效发射率。

7）球膜表面吸收的云层辐射热量计算公式为

$$Q_{\text{IR-cloud}}=\sigma\varepsilon_{\text{c}}T_{\text{cloud}}^4\cdot\frac{1}{2}A_{\text{surf}}\alpha_{\text{film-IR}}[1+\tau_{\text{film-IR}}(1+\vartheta_{\text{eff}})] \tag{46}$$

式中：T_{cloud}为云层温度。

8）球膜表面吸收的内部辐射热量计算公式为[4]

$$Q_{\text{IR-filmabsorb}}=\sigma\varepsilon_{\text{film}}T_{\text{film}}^4A_{\text{surf}}\alpha_{\text{film-IR}}(1+\vartheta_{\text{eff}}) \tag{47}$$

式中：$\varepsilon_{\text{film}}$为球膜材料平均红外发射率，假设满足基尔霍夫定理，即红外发射率等于红外吸收率[1]。

9）球膜表面发射的辐射热量计算公式为[4]

$$Q_{\text{IR-filmout}}=-2\sigma\varepsilon_{\text{film}}T_{\text{film}}^4A_{\text{surf}} \tag{48}$$

10）球膜外表面与外部环境的混合对流换热热量计算公式为

$$Q_{\text{Conv-ext}}=\begin{cases} h_{\text{ext}}A_{\text{surf}}(T_{\text{air}}-T_{\text{film}}) & \\ \qquad H\leqslant H_{\text{cloudbase}},H\geqslant H_{\text{cloudtop}} & \\ h_{\text{ext}}A_{\text{surf}}(T_{\text{cloud}}-T_{\text{film}}) & \\ \qquad H_{\text{cloudbase}}<H<H_{\text{cloudtop}} & \end{cases} \tag{49}$$

式中：h_{ext}为外部混合对流换热系数，计算公式如下：

$$h_{\text{ext}}=(h_{\text{ext-force}}^3+h_{\text{ext-free}}^3)^{1/3} \tag{50}$$

$$h_{\text{ext-force}}=\frac{k_{\text{air}}}{d}(2+0.41Re^{0.55}) \tag{51}$$

$$h_{\text{ext-free}}=\frac{k_{\text{air}}Nu_{\text{air}}}{d} \tag{52}$$

其中：k_{air}为空气的导热系数；Re为雷诺数；Nu_{air}为无量纲参数努塞尔数，具体计算详见文献[4]。

11）球膜内表面与氦气自然对流换热热量计算公式为[4]

$$Q_{\text{Conv-int}}=h_{\text{int}}A_{\text{surf}}(T_{\text{He}}-T_{\text{film}}) \tag{53}$$

式中：h_{int}为内部自然对流换热系数，计算公式如下：

$$h_{\text{int}}=0.13k_{\text{He}}\left(\frac{\rho_{\text{He}}^2g|T_{\text{film}}-T_{\text{He}}|Pr_{\text{He}}}{T_{\text{He}}\mu_{\text{He}}^2}\right)^{1/3} \tag{54}$$

其中：μ_{He}为氦气的动力黏度；k_{He}为氦气的导热系数；Pr_{He}为氦气的普朗特数。

3.3.2 氦气热力学模型

气球内部氦气平均温度的变化率可由以下瞬态能量平衡公式计算[4]：

$$\frac{\mathrm{d}T_{\text{He}}}{\mathrm{d}t}=-\frac{Q_{\text{Conv-int}}}{c_vm_{\text{He}}}+(\gamma-1)\frac{T_{\text{He}}}{\rho_{\text{He}}}\cdot\frac{\mathrm{d}\rho_{\text{He}}}{\mathrm{d}t} \tag{55}$$

式中：$\gamma=c_p/c_v$，$R_{\text{He}}/c_v=\gamma-1$，$c_v$为定体积比热容，$c_p$为定压比热容；$\rho_{\text{He}}$为氦气密度。

4 计算分析

4.1 仿真程序开发

根据超压气球热动力学特性数学模型，开发超压气球热动力学特性仿真计算程序，程序结构如图6所示。

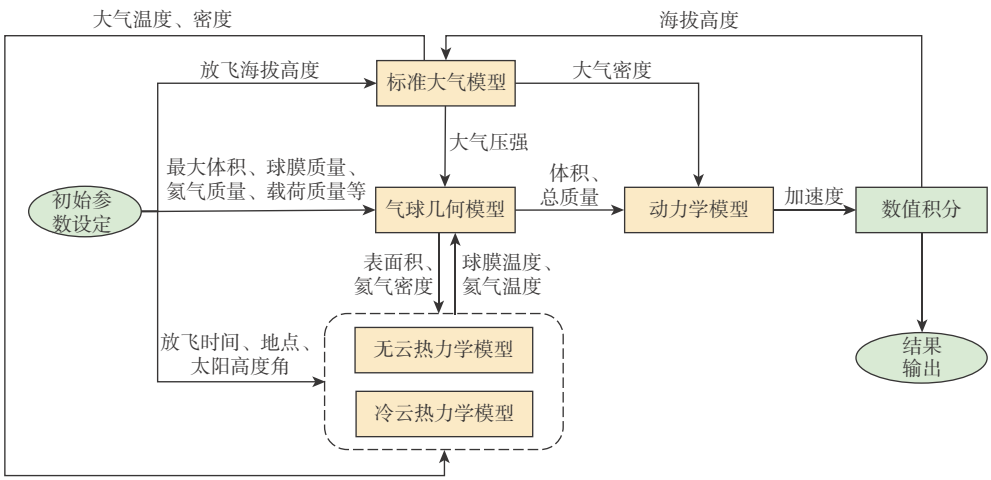


图6 仿真程序结构

Fig. 6 Structure of simulation program

该程序由主程序和5个子程序组成：主程序可求解球膜温度、氦气温度、升空速度和飞行高度；子程序分别求解标准大气模型、气球几何模型、动力学模型、无云环境下热力学模型及冷云环境下热力学模型。

4.2　模型验证

利用某次超压气球飞行试验数据，验证本文开发的超压气球热动力学特性计算模型和仿真程序的准确性。该超压气球系统设计参数详见表 1，球膜材料热辐射特性参数详见表 2。

表 1　某超压气球系统设计参数

Table 1　Design parameters of a super-pressure balloon system

参数	数值
浮升气体	氦气
放飞时间	8月下旬上午6:45
驻留高度/km	25
最大体积/m^3	7 240
总质量/kg	280
球膜质量/kg	120
载荷质量/kg	124

表 2　球膜材料热辐射特性参数

Table 2　Thermal radiation properties of balloon film

参数	数值
球膜比热容/($J\cdot kg^{-1}\cdot K^{-1}$)	2 302.7
球膜对可见光吸收率	0.02
球膜对可见光透过率	0.9
球膜对红外辐射吸收率	0.078
球膜对红外辐射透过率	0.9

超压气球内部氦气温度由 3 个温度传感器采集，其在球体内部沿中心线垂直分布，如图 7 所示。

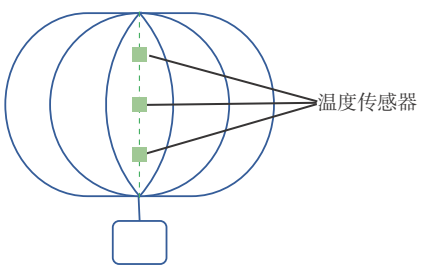


图 7　温度传感器分布示意图

Fig. 7　Distribution of temperature sensors

超压气球飞行试验当天，天空晴朗无云。超压气球内部氦气温度的仿真计算数据与实测数据对比如图 8 所示。

由图 8 可知，氦气温度仿真计算数据与实测数据变化趋势一致，数值接近。氦气初始温度为 280 K，升空过程前期，球体急剧膨胀，大气温度快速下降，球膜与环境之间的外部对流换热及球膜与氦气之间的内部对流换热加强，氦气急剧膨胀对外做功，导致氦气温度下降至最低点，仿真计算温度最低值为 209.3 K，实测温度最低值为 212 K，差值为 2.7 K。抵达设计平飞高度后，球体不再膨胀，太阳辐射等

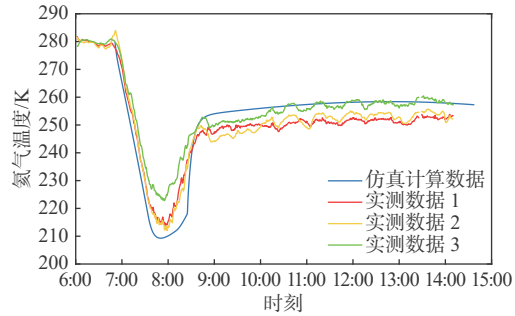


图 8　无云环境飞行过程氦气温度

Fig. 8　Helium temperature during flight process in environment without cloud

增强，氦气温度回升，仿真计算温度最大值为 258.5 K，实测温度最大值为 260.5 K，差值为 2 K。

超压气球飞行过程中，升空速度的仿真计算数据与实测数据对比如图 9 所示。

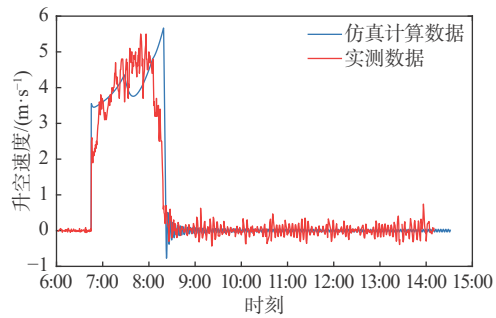


图 9　无云环境升空速度

Fig. 9　Ascending velocity in environment without cloud

由图 9 可知，气球升空速度仿真计算数据与实测数据趋势一致，数值接近。在放飞初始时刻，球内氦气提供的浮力大于系统重力，气球系统升空速度迅速攀升，仿真计算数值为 3.5 m/s，实测数值为 2.5 m/s；升空过程中，受大气环境温度变化和氦气膨胀做功双重影响，仿真计算数据出现双“V”变化趋势，实测数据因环境影响因素较多表现为不规则振荡。其中，最大速度仿真计算数据为 5.7 m/s，实测数据为 5.5 m/s，误差为 3.5%；平均速度仿真计算数据为 2.8 m/s，实测数据为 2.4 m/s，误差为 14.3%。

超压气球飞行过程中，飞行高度的仿真计算数据与实测数据对比如图 10 所示。

由图 10 可知，气球飞行高度仿真计算数据与实测数据变化趋势相近，稳定状态下的预测驻留高度为 24.95 km，实测驻留高度为 24.9 km，相差 0.2%。

某型超压气球升空飞行氦气温度、升空速度和飞行高度数据对比结果显示，本文开发的超压气球热动力学特性仿真模型和计算程序较为准确，为进一步研究冷云辐射特性奠定了基础。

4.3　仿真结果与分析

利用风云四号卫星应用平台数据，研究冷云辐

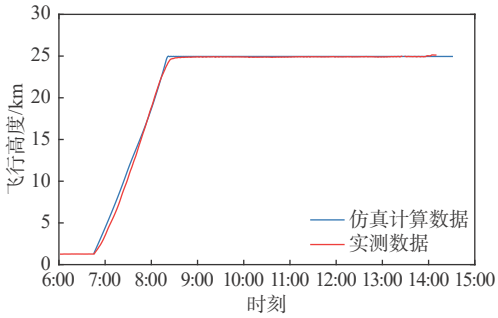


图 10 无云环境飞行高度

Fig. 10 Flight altitude in environment without cloud

射对超压气球热动力学特性影响。

某卷云覆盖区域的云底高度为 12.5 km, 云顶高度为 13.7 km, 以云顶温度 205 K 作为云层平均温度。仿真分析超压气球在冷云区域内升空和平飞经过冷云区域上方 2 种过程中的热动力学特性。

4.3.1 冷云区域内升空过程仿真

超压气球在冷云区域内的升空过程如图 11 所示。

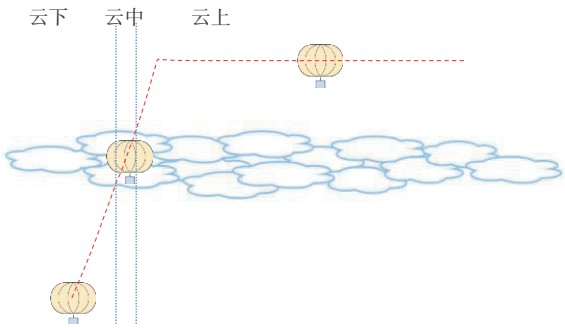


图 11 冷云区域内升空过程示意图

Fig. 11 Ascending process in environment with cold cloud

如图 11 所示, 超压气球从冷云下方区域放飞升空, 并穿过云层进入冷云上方区域, 其升空全过程可分为“云下-云中-云上”3 个阶段。

仿真计算超压气球升空过程中在无云环境和冷云环境下氦气温度, 如图 12 所示。

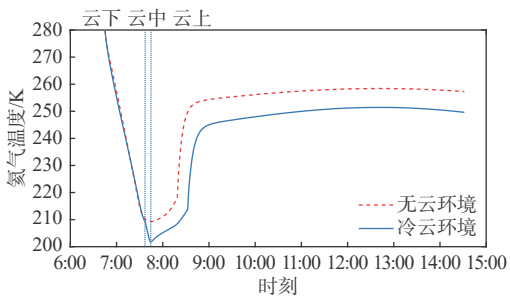


图 12 升空过程氦气温度对比

Fig. 12 Comparison of helium temperature during ascending process

由图 12 可知, 在“云下”阶段, 氦气温度变化与无云环境较为贴合, 对比差值最大约为 2 K; 在“云中”阶段, 超压气球穿越云层时有较大的热量损失, 温度曲线出现低温峰值 201.8 K, 较无云环境降低 7.5 K; 在“云上”阶段, 氦气温度回升, 最大值为 251.4 K, 低于无云环境 7.1 K。

冷云环境与无云环境之间的温度差主要取决于超压气球与云层之间的相对位置。“云下”和“云上”阶段, 冷云的主要影响是对热环境辐射的衰减, “云中”阶段, 冷云影响体现为热环境辐射衰减和强对流换热的双重作用, 因此, 该阶段温度差值较大。

仿真计算超压气球升空过程中在无云环境和冷云环境下球体内外压差, 如图 13 所示。

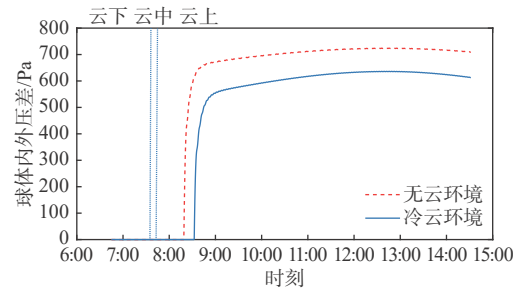


图 13 升空过程球体内外压差对比

Fig. 13 Comparison of pressure difference during ascending process

由图 13 可知, 在“云下”和“云中”的升空阶段, 此时超压气球还未充满, 超压气球内外压差为 0 Pa。

抵达驻留高度后, 体积膨胀至最大值, 随着温度升高, 超压气球内外压差随之增大。冷云环境下的最大压差值为 636 Pa, 无云环境下最大压差为 723 Pa, 相差 87 Pa。因冷云环境下的氦气温度低于无云环境, 故冷云环境下内外压差也低于无云环境。

仿真计算超压气球升空过程中在无云环境和冷云环境下升空速度, 如图 14 所示。

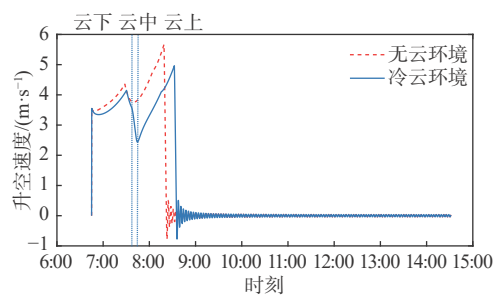


图 14 升空过程速度对比

Fig. 14 Comparison of velocity during ascending process

由图 14 可知, 无云环境下平均升空速度为 4.1 m/s, 冷云环境下平均升空速度为 3.6 m/s。在“云下”阶段, 升空速度对比差值最大约为 0.3 m/s; 在“云中”阶段, 升空速度快速降低至最小值 2.4 m/s, 低于无云环境升空速度最小值 1.4 m/s; 在

“云上”阶段, 升空速度先上升至最大值 5 m/s, 低于无云环境升空速度最大值 0.7 m/s, 随后, 快速下降至 0 m/s 振荡。冷云环境下的超压气球氦气温度比无云环境下低, 球体体积膨胀程度不足, 升力随之减小, 导致冷云环境下的升空速度整体减小, 在“云中”阶段, 氦气温度最低, 速度随之降至最低点。

仿真计算超压气球升空过程中在无云环境和冷云环境下飞行高度, 如图 15 所示。

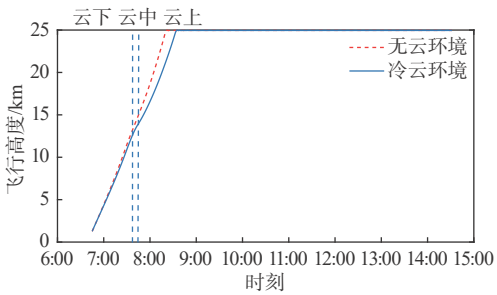


图 15　升空过程飞行高度对比

Fig. 15　Comparison of flight altitude during ascending process

由图 15 可知, 在“云下”阶段, 冷云环境飞行高度与无云环境相差不大; 在“云中”和“云上”阶段, 因速度降低导致冷云环境高度曲线比无云环境平缓, 同一时刻下的飞行高度低于无云环境; 最终, 冷云环境下的超压气球较无云环境延迟抵达相同的驻留高度。

4.3.2　平飞经过冷云区域上方过程仿真

超压气球平飞经过冷云区域上方的过程如图 16 所示。

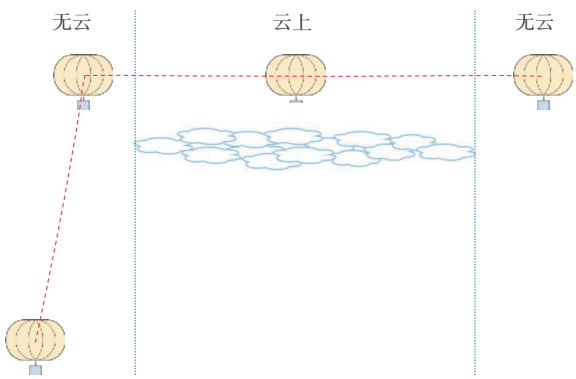


图 16　平飞经过冷云区域上方过程示意图

Fig. 16　Horizontal flight process above cold cloud

如图 16 所示, 超压气球在无云环境下升空并抵达驻留高度, 在平飞过程中从无云环境进入冷云上方区域, 并最终离开冷云区域, 其飞行过程可分为“无云-云上-无云”3 个阶段。

仿真计算超压气球在飞行过程中, 不经过冷云上空和经过冷云上空条件下氦气温度, 如图 17 所示。

由图 17 可知, 超压气球平飞进入“云上”阶段后, 氦气温度整体低于无云环境下的氦气温度。白天, 冷云环境下氦气温度在 12:45 到达峰值 251.5 K, 低于无云环境氦气温度峰值 7 K; 夜间, 冷云环境下氦气温度为 218.5 K, 低于无云环境 15 K。原因是冷云衰减了抵达超压气球表面的热辐射强度。

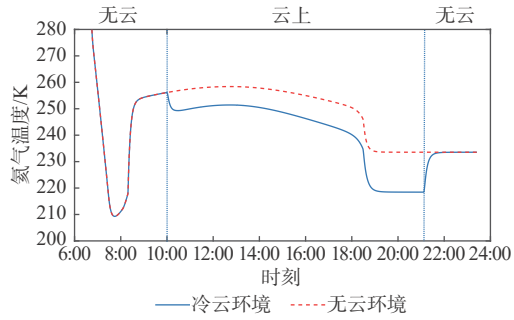


图 17　平飞过程氦气温度对比

Fig. 17　Comparison of helium temperature during horizontal flight process

仿真计算超压气球在飞行过程中, 不经过冷云上空和经过冷云上空条件下内外压差, 如图 18 所示。

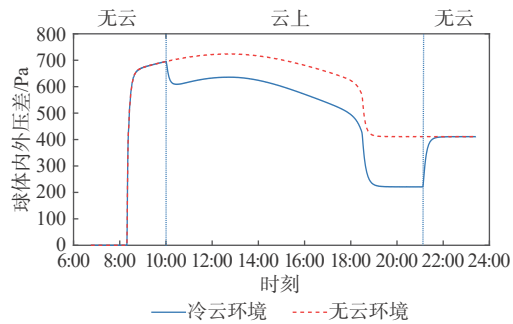


图 18　平飞过程球体内外压差对比

Fig. 18　Comparison of pressure difference during horizontal flight process

由图 18 可知, 超压气球平飞进入“云上”阶段后, 内外压差整体低于无云环境。白天, 超压气球内外压差曲线随太阳高度角变化而变化; 夜间, 无云环境下内外压差为 411 Pa, 冷云环境下内外压差为 220 Pa, 相差 191 Pa。“云上”阶段的氦气温度较无云环境降低, 导致内外压差减小, 但气球仍在“超压”状态。

仿真计算超压气球在飞行过程中, 不经过冷云上空和经过冷云上空条件下升空速度和飞行高度, 分别如图 19 和图 20 所示。

由图 19 和图 20 可知, 超压气球平飞进入“云上”阶段后, 升空速度、飞行高度曲线与无云环境下的曲线重合。因气球仍处于“超压”状态, 气球体积等于最大体积, 对应的受力情况无变化, 动力学特性保持一致。

4.3.3　冷云环境对超压气球影响参数分析

通过冷云区域内升空过程和冷云区域上方平飞过程的仿真分析可知, 云层温度越低, 发射的红

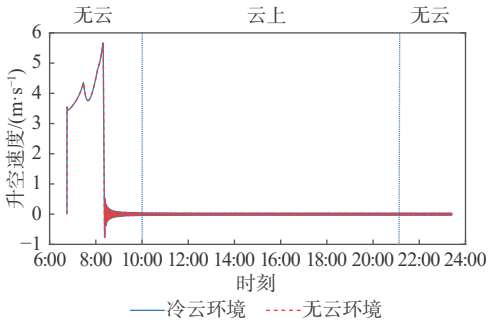


图 19 平飞过程升空速度对比

Fig. 19 Comparison of flight velocity during horizontal flight process

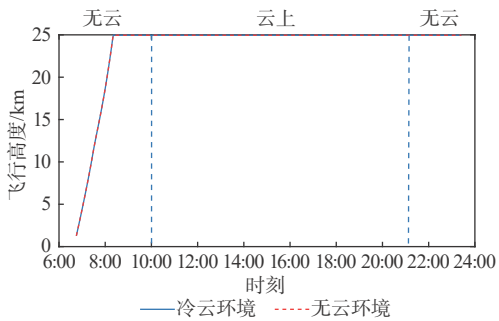


图 20 平飞过程飞行高度对比

Fig. 20 Comparison of flight altitude during horizontal flight process

外辐射越少，超压气球在穿云过程中的对流换热损失越强；云层越厚，对环境辐射的衰减作用越强，超压气球穿过云层所需的时间越长。

当云层温度和云层厚度到达临界值时，超压气球将无法抵达预期驻留高度，进一步研究云层温度和厚度对超压气球热动力学特性的影响。

1）云层温度的影响分析。当云底高度为 12.5 km、云顶高度为 13.7 km 时，超压气球在飞行过程中穿越云层。

仿真计算不同云层温度条件下，超压气球的氦气温度，如图 21 所示。

由图 21 可知，当云层厚度为定值时，云层温度

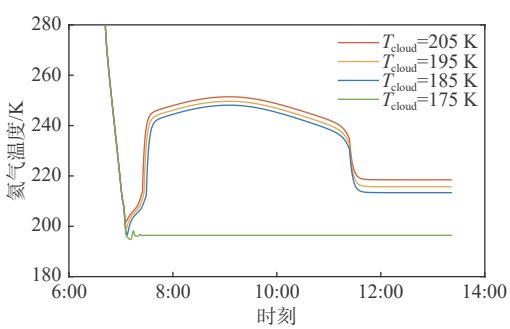


图 21 不同云层温度条件下飞行过程氦气温度

Fig. 21 Helium temperature during flight process at different cloud temperatures

越低，氦气温度越低。当云层温度为 175 K 时，氦气温度最终停滞在 196.4 K，不再回升。

仿真计算不同云层温度条件下超压气球的升空速度，如图 22 所示。由图 22 可知，当云层厚度为定值时，云层温度越低，整体升空速度越小。当云层温度为 175 K 时，速度曲线的双“V”变化趋势未形成。对比同时刻的氦气温度，温度越低，球体膨胀程度不足，速度相应减小。

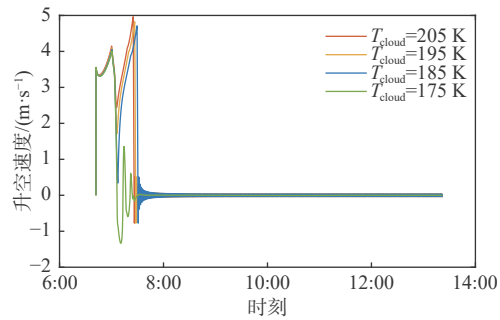


图 22 不同云层温度条件下升空速度

Fig. 22 Ascending velocity at different cloud temperatures

仿真计算不同云层温度条件下超压气球的飞行高度，如图 23 所示。由图 23 可知，当云层厚度为定值时，云层温度越低，抵达驻留高度的时间越晚。当云层温度为 175 K 时，到达最大高度 13.5 km，随后稳定在 12.6 km，即滞留在云层中，未抵达驻留高度。

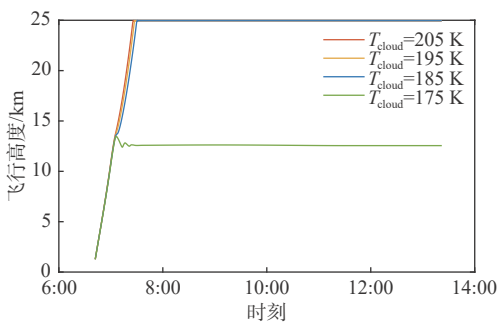


图 23 不同云层温度条件下飞行高度

Fig. 23 Flight altitude at different cloud temperatures

2）云层厚度的影响分析。当云层温度为 205 K、云底高度为 12.5 km 时，超压气球在飞行过程中穿越云层。

仿真计算不同云顶高度（即云层厚度）条件下超压气球的氦气温度，如图 24 所示。

由图 24 可知，当云层温度为定值时，云层越厚，氦气温度在最小温度值停滞的时间越长。当云层厚度为 4 200 m 时，氦气温度稳定在 196.4 K，不再回升。

仿真计算不同云顶高度（即云层厚度）条件下超压气球的升空速度，如图 25 所示。

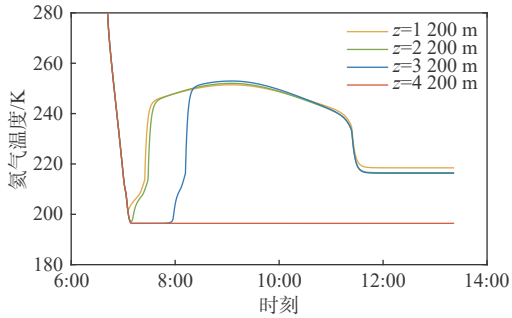


图 24　不同云层厚度条件下飞行过程氦气温度

Fig. 24　Helium temperature during flight process at different cloud thicknesses

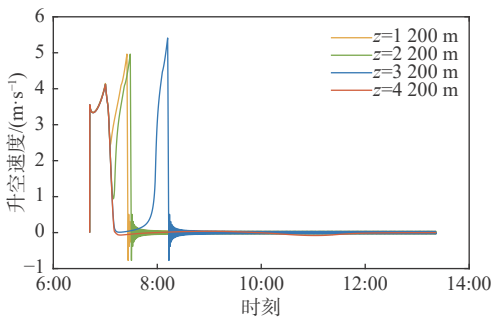


图 25　不同云层厚度条件下升空速度

Fig. 25　Ascending velocity at different cloud thicknesses

由图 25 可知，当云层温度为定值时，云层越厚，升空速度在穿云过程中的低速爬升状态时间越长。当云层厚度为 4 200 m 时，升空速度在云层中降至 0 m/s 振荡。穿云过程中氦气温度降低，使得超压气球体积不再膨胀，提供的升力有限，爬升速度较小。

仿真计算不同云顶高度（即云层厚度）条件下超压气球的飞行高度，如图 26 所示。

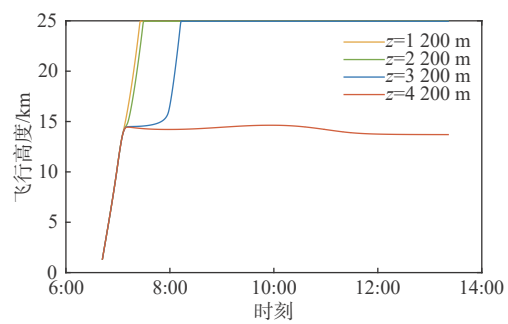


图 26　不同云层厚度条件下飞行高度

Fig. 26　Flight altitude at different cloud thicknesses

由图 26 可知，当云层温度为定值时，云层越厚，超压气球在穿云过程中耗费的时间越长，抵达驻留高度的时间越晚。当云层厚度为 4 200 m 时，与云层形成热平衡的超压气球升力有限，超压气球稳定在 13.6 km 左右的高度，未抵达预期驻留高度。

4.3.4　云层温度与厚度耦合特性分析

由 4.3.3 节可知，云层温度或云层厚度到达临界值时，超压气球将无法抵达预期驻留高度。实际上，云层温度与云层厚度对于超压气球的影响是耦合关系，进一步探究云层温度、云层厚度和超压气球能否抵达预期驻留高度三者之间的联系，为超压气球飞行试验提供参考。

当云底高度为 12.5 km 时，超压气球在飞行过程中穿越云层，仿真计算不同云顶高度（即云层厚度）和云层温度的耦合条件下，超压气球可以抵达预期驻留高度的临界点，并对临界点进行数值拟合，结果如图 27 所示。

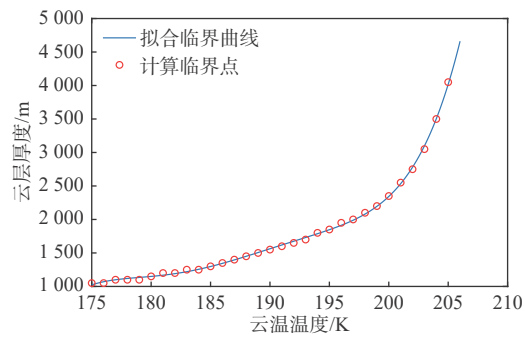


图 27　云层温度与厚度耦合特性分析

Fig. 27　Analysis of coupling characteristics of cold cloud temperature and thickness

由图 27 可知，红色圆圈为云层温度与云层厚度参数耦合下的临界点，即超压气球恰好滞留在云层中。

在拟合临界曲线下方区域，超压气球可以穿过云层，最终抵达预期高度；在曲线上方区域，超压气球将滞留云层，无法抵达预期高度。临界曲线计算公式如下：

$$\begin{aligned} z_{\mathrm{P}} = {} & 0.001\,288\,021\,269\,012\,68T_{\mathrm{cloud}}^{5} - \\ & 1.204\,268\,349\,301\,46T_{\mathrm{cloud}}^{4} + \\ & 450.170\,608\,429\,497T_{\mathrm{cloud}}^{3} - \\ & 84\,098.198\,488\,116\,3T_{\mathrm{cloud}}^{2} + \\ & 7\,851\,390.547\,529\,49T_{\mathrm{cloud}} - \\ & 293\,050\,141.372\,838 \end{aligned} \tag{56}$$

式中：z_{P}为根据云层温度仿真计算的临界云层厚度。当实际云层厚度大于临界云层厚度时，超压气球将滞留云层，无法抵达预期驻留高度；当实际云层厚度小于临界云层厚度时，超压气球可以升空抵达预期驻留高度。

5　结　论

本文分析云的宏观和微物理特性，建立了冷云辐射影响下的超压气球热动力学模型。以某型超

压气球为研究对象，仿真分析冷云影响下的热动力学特性，主要结论如下：

1）当超压气球在云层下方时，云层对太阳直射辐射、天空散射辐射、地表反射辐射和大气红外辐射起衰减作用，同时，云层自身向周围环境发射长波辐射。最终，超压气球吸收的辐射总量比无云环境条件下更少，氦气温度、升空速度和飞行高度更低。

2）当超压气球穿过云层时，云层对环境中的辐射均具有衰减作用，同时，冷云与超压气球的强对流换热导致氦气温度迅速下降，升空速度随之降低，推迟抵达驻留高度的时间。

3）当超压气球在云层上方时，云层对地表反射辐射和地面红外辐射有衰减作用，同时，超压气球底部会吸收云层反射的太阳辐射及云自身发射的红外辐射。白天，吸收的辐射总量随太阳高度角变化，氦气峰值温度低于无云环境 7 K；夜间，氦气温度低于无云环境 15 K。

4）当云层温度恒定时，云层越厚，对超压气球上升过程影响越大，到达临界值后，超压气球将滞留在云层中，无法抵达设计高度；当云层厚度恒定时，云层温度越低，对超压气球上升过程影响越大，到达临界值后，超压气球将滞留在云层中，无法抵达设计高度。

5）分析了云层厚度与温度的耦合特性，计算了超压气球可以抵达设计高度的临界条件，并拟合出临界曲线，在曲线下方区域条件下，超压气球可以升空抵达预期驻留高度，可为超压气球飞行试验提供参考。

参考文献（References）

[1] 吕明云, 巫资春. 高空气球热力学模型与上升过程仿真分析[J]. 北京航空航天大学学报, 2011, 37(5): 505-509.
LÜ M Y, WU Z C. Thermodynamic model and numerical simulation of high altitude balloon ascending process[J]. Journal of Beijing University of Aeronautics and Astronautics, 2011, 37(5): 505-509 (in Chinese).

[2] BIRD R E, HULSTROM R L. Simplified clear sky model for direct and diffuse insolation on horizontal surfaces: SERI/TR-642-761[R]. Golden: Solar Energy Research Institute, 1981.

[3] NIELSEN L B, PRAHM L P, BERKOWICZ R, et al. Net incoming radiation estimated from hourly global radiation and/or cloud observations[J]. Journal of Climatology, 1981, 1(3): 255-272.

[4] FARLEY R. BalloonAscent: 3-D simulation tool for the ascent and float of high-altitude balloons[C]//Proceedings of the AIAA 5th ATIO and 16th Lighter-Than-Air Sys Tech. and Balloon Systems Conferences. Reston: AIAA, 2005: 7412.

[5] 程晨. 平流层浮空器瞬态热模型及热特性研究[D]. 上海: 上海交通大学, 2019.
CHENG C. Study on transient thermal model and thermal characteristics of stratospheric aerostat[D]. Shanghai: Shanghai Jiao Tong University, 2019(in Chinese).

[6] U. S. Committee on Extension to the Standard Atmosphere. US standard atmosphere, 1976[S]. Washington, DC: U. S. Government Printing Office, 1976.

[7] ORGANIZATION W M. International cloud atlas[M]. Geneva: World Meteorological Organization, 2017.

[8] SLINGO A. A GCM parameterization for the shortwave radiative properties of water clouds[J]. Journal of the Atmospheric Sciences, 1989, 46(10): 1419-1427.

[9] LINDNER T H, LI J. Parameterization of the optical properties for water clouds in the infrared[J]. Journal of Climate, 2000, 13(10): 1797-1805.

[10] STEPHENS G L, ACKERMAN S, SMITH E A. A shortwave parameterization revised to improve cloud absorption[J]. Journal of the Atmospheric Sciences, 1984, 41(4): 687-690.

[11] 娄树理, 周晓东. 基于光学厚度的云红外辐射计算[J]. 应用光学, 2011, 32(2): 343-347.
LOU S L, ZHOU X D. Calculation of cloud infrared radiation based on optical depth[J]. Journal of Applied Optics, 2011, 32(2): 343-347 (in Chinese).

[12] 施帅红. 近 43 年太阳辐射的变化特征以及云的光学特性的研究[D]. 南京: 南京信息工程大学, 2007.
SHI S H. Variation characteristics of solar radiation and optical characteristics of clouds in recent 43 years[D]. Nanjing: Nanjing University of Information Science & Technology, 2007(in Chinese).

[13] 孙治安, 刘晶淼, 曾宪宁, 等. 云天地表总辐射和净辐射瞬时值的计算方法[J]. 气象与环境学报, 2014, 30(3): 1-9.
SUN Z A, LIU J M, ZENG X N, et al. Estimation of global and net solar radiation at the Earth surface under cloudy-sky condition[J]. Journal of Meteorology and Environment, 2014, 30(3): 1-9(in Chinese).

[14] 郭新军, 金伟其, 王霞, 等. 云层红外表观辐射模型的研究[J]. 北京理工大学学报, 2003, 23(4): 414-418.
GUO X J, JIN W Q, WANG X, et al. An infrared apparent radiation model for clouds[J]. Journal of Beijing Institute of Technology, 2003, 23(4): 414-418(in Chinese).

[15] CHYLEK P, RAMASWAMY V. Simple approximation for infrared emissivity of water clouds[J]. Journal of the Atmospheric Sciences, 1982, 39(1): 171-177.

[16] 刘强, 武哲, 祝明, 等. 平流层气球热动力学仿真[J]. 北京航空航天大学学报, 2013, 39(12): 1578-1583.
LIU Q, WU Z, ZHU M, et al. Thermal-dynamic simulation of stratospheric balloon[J]. Journal of Beijing University of Aeronautics and Astronautics, 2013, 39(12): 1578-1583(in Chinese).

[17] RAN H J, THOMAS R, MAVRIS D. A comprehensive global model of broadband direct solar radiation for solar cell simulation[C]//Proceedings of the 45th AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2007: 33.

[18] 夏新林, 李德富, 杨小川. 平流层浮空器的热特性与研究现状[J]. 航空学报, 2009, 30(4): 577-583.
XIA X L, LI D F, YANG X C. Thermal characteristics of stratospheric aerostats and their research[J]. Acta Aeronautica et Astronautica Sinica, 2009, 30(4): 577-583(in Chinese).

Influence of cold cloud radiation on thermal-dynamic characteristics of super-pressure balloon

HE Lan[1, 2]，LIU Qiang[1, *]，YANG Yanchu[1, 2]，ZHU Rongchen[1]，ZHOU Jianghua[1, 2]

(1. Aerospace Information Research Institute，Chinese Academy of Sciences，Beijing 100094，China；
2. University of Chinese Academy of Sciences，Beijing 101408，China)

Abstract：To investigate the influence of cold cloud radiation on the thermal-dynamic characteristics of super-pressure balloons, this paper proposed a thermal-dynamic model that incorporated cold cloud radiation characteristics for super-pressure balloons. First, the paper probed into the influence of cold cloud on the thermal environment of a super-pressure balloon at different relative heights. Then, it analyzed cold cloud characteristics including height, thickness, effective radius of cloud droplet, temperature, and water content and further described the radiation absorption, transmission, and reflection equations of cold cloud. On the basis of dynamic, kinetic, and thermal analyses, a thermal-dynamic model that incorporated cold cloud radiation characteristics was established for the super-pressure balloon, and a computer program was developed. Finally, measured data was used to verify the accuracy of the thermal-dynamic model and the program. Cold cloud data from the Fengyun satellite was introduced to simulate the thermal-dynamic characteristics of a super-pressure balloon under the influence of cold cloud radiation. The results show that cold cloud may lower the temperature of the super-pressure balloon and therefore exert influence on its flight velocity and altitude. When the cloud thickness and temperature reach a critical condition, the super-pressure balloon could not reach its designed flight altitude.

Keywords：super-pressure balloon；thermal characteristics；dynamic characteristics；cold cloud environment；radiation characteristic

Received：2023-07-04；**Accepted**：2023-09-14；**Published Online**：2023-10-10 14：47

URL：link.cnki.net/urlid/11.2625.V.20231009.1320.003

Foundation items：National Key Research and Development Program of China (2022YFB3903005-01); Youth Innovation Promotion Association of Chinese Academy of Sciences (2020133)

* **Corresponding author**. E-mail：liuqiang1001@aoe.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0439

轨控发动机含冰晶多相羽流辐射特性

罗吴迪[1]，任军学[1,*]，李志辉[2,3]，汤海滨[1]

(1. 北京航空航天大学 宇航学院，北京 100191；2. 中国空气动力研究与发展中心，绵阳 621000；
3. 国家计算流体力学实验室，北京 100191)

摘 要：为获得高空轨控发动机含冰晶的多相羽流的辐射特性，分析其对遥测效果的影响，基于爱因斯坦辐射理论，推导了气体辐射发射系数，结合微粒散射Mie理论，获得综合考虑气体热辐射和冰晶辐射与冰晶散射的辐射传输模型。基于最新的HITRAN2020谱线库，仿真得到了高空轨控发动机含冰晶多相羽流辐射特性参数以及近红外、中长波红外等波段辐射强度分布，并分析了辐射的主要影响因素。结果表明：该型发动机羽流辐射主要包括近红外波段的冰晶散射和中长波红外波段的气体热辐射。在近红外波段，考虑太阳背景辐射，计算域内总辐射量约10^{-6} W/sr量级，轴向距离0.5 m以内，气体热辐射占主导，轴向距离0.5 m以外，冰晶散射占主导；在中长波红外波段，计算域内总辐射量约10 W/sr量级，轴向距离1 m以内，气体热辐射占主导，轴向距离1 m以外，冰晶热辐射占主导。

关 键 词：辐射；轨控发动机；多相羽流；热辐射；散射
中图分类号：V435
文献标志码：A **文章编号**：1001-5965(2025)09-3130-11

发动机尾焰辐射是飞行器的重要特征信息。研究目标的辐射特性在发动机性能诊断、目标隐身性能研究以及目标识别跟踪等方面有着重要意义[1]。从20世纪中叶开始，对火箭发动机尾焰辐射特性的研究就一直没有间断过[2-3]，这方面的研究主要有数值计算和实验测量2种手段，早期由于理论和数据库尚不完善，计算机技术尚不成熟，对辐射特性的研究主要是实验手段的研究，如采用红外辐射计、红外干涉光谱仪和红外光谱辐射计等仪器直接对固体或液体火箭发动机尾焰进行测量[4-5]。随着计算机技术的发展和辐射理论、计算方法的完善，采用数值方法研究尾焰不同波段辐射特性变得越来越常见[6-8]。

当前，国内外对介质辐射特性计算建立了许多方法，一种是纯气相尾焰流场的辐射计算[7,9]，以及考虑气固两相辐射的模型[6,10]，采用的辐射传输方程计算方法主要包括蒙特卡罗法[11-12]、有限体积法[13-15]和路径积分法等[7,16]。大部分文献主要研究的是低空高温尾焰，其中主要考虑的微粒是固体火箭发动机尾焰中的金属氧化物颗粒或碳烟颗粒[6,10,17]。这些发动机燃烧室温度高达数千K，羽流温度也有上千K。但是，一些高空小型轨控发动机燃烧室温度约3 000 K，由于大膨胀比喷管的作用，尾焰羽流温度仅数百K，热辐射集中于中长波红外波段。同时，由于羽流温度较低，其中的水会凝结成大量微小的冰晶颗粒，而这些冰晶颗粒的存在会对羽流自身辐射以及对阳光的散射产生显著的影响。为此，本文基于HITRAN2020谱线库，结合气体辐射理论与微粒散射Mie理论，综合考虑流场中的气体和微粒对辐射的贡献，研究了某型轨控发动机多相羽流的热辐射和散射2种辐射机制，采用图形处理器(graphics processing unit, GPU)并行计算，获得了高空轨控发动机含冰晶多相羽流辐射特性参数以及红外波段辐射强度分布、光谱特性，并分析羽流辐

收稿日期：2023-07-04；录用日期：2023-08-15；网络出版时间：2023-09-11 11:28
网络出版地址：link.cnki.net/urlid/11.2625.V.20230908.1404.003
*通信作者. E-mail：rjx_buaa@163.com

引用格式：罗吴迪，任军学，李志辉，等. 轨控发动机含冰晶多相羽流辐射特性[J]. 北京航空航天大学学报，2025，51(9)：3130-3140.
LUO W D，REN J X，LI Z H，et al. Radiation characteristics of multi-phase plumes containing ice crystals from orbit-control engines[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3130-3140 (in Chinese).

射的主要来源。

1　尾焰多相羽流辐射模型建立

1.1　气体辐射特性参数的计算

气体的辐射特性参数包括气体辐射吸收系数和发射系数，本节基于分子辐射爱因斯坦理论，推导计算气体辐射特性参数的表达式。

根据分子辐射爱因斯坦理论，用波数ν（单位 cm^{-1}）表示的爱因斯坦系数A_{lu}和B_{lu}关系如下[18]（下标 l 和 u 表示分子能级跃迁的低能级和高能级）：

$$A_{\mathrm{lu}} = 8\pi hc\nu^3 B_{\mathrm{lu}} \tag{1}$$

式中：c 为光速；h 为普朗克常数。

根据爱因斯坦理论，气体分子辐射分为自发辐射和受激辐射。在 dt时间里，同时考虑 2 种辐射 dV体积到单位立体角的能量为[18]

$$j_\nu \mathrm{d}V\mathrm{d}\nu\mathrm{d}\Omega\mathrm{d}t = f(\nu)n_{\mathrm{u}}\mathrm{d}V\left(\frac{A_{\mathrm{lu}}}{4\pi}+\frac{B_{\mathrm{lu}}I_\nu(\Omega)}{c}\right)hc\nu\mathrm{d}\nu\mathrm{d}\Omega\mathrm{d}t \tag{2}$$

式中：n_{u}为处在高能级上的粒子数密度；$f(\nu)$为归一化谱线线形函数；$I_\nu(\Omega)$为辐射亮度；j_ν为发射系数。

可得

$$j_\nu = n_{\mathrm{u}}\left(\frac{A_{\mathrm{lu}}}{4\pi}+\frac{B_{\mathrm{lu}}I_\nu(\Omega)}{c}\right)hc\nu f(\nu) \tag{3}$$

对背景辐射的吸收用辐射亮度I_ν表示：

$$-\mathrm{d}I_\nu = -\frac{\mathrm{d}E_{\mathrm{abs}}}{\mathrm{d}S\,\mathrm{d}t\mathrm{d}\Omega\mathrm{d}\nu} = n_{\mathrm{l}}hc\nu B_{\mathrm{ul}}I_\nu c^{-1}f(\nu)\mathrm{d}s \tag{4}$$

式中：E_{abs}为吸收能量；n_{l}为低能级粒子数密度。

吸收系数k_ν为在x方向上单位长度的辐射亮度减小率：

$$k_\nu = \frac{-\mathrm{d}I_\nu}{I_\nu \mathrm{d}x} \tag{5}$$

对比式(4)和式(5)可得

$$k_\nu = n_{\mathrm{l}}B_{\mathrm{ul}}h\nu f(\nu) \tag{6}$$

由$B_{\mathrm{ul}} = B_{\mathrm{lu}}$，联立式 (1)、式 (3)、式 (6) 可以得出发射系数为

$$j_\nu = n_{\mathrm{u}}\left(2c\nu^2\frac{k_\nu}{n_{\mathrm{l}}}+\frac{k_\nu}{n_{\mathrm{l}}hc\nu}I_\nu(\Omega)\right)hc\nu \tag{7}$$

根据局部热平衡条件，考虑能级简并条件：

$$\frac{n_{\mathrm{l}}}{n_{\mathrm{u}}} = \frac{g_{\mathrm{l}}}{g_{\mathrm{u}}}\mathrm{e}^{hc\nu/(k_bT)} \tag{8}$$

和普朗克函数：

$$B_\nu(T) = 2hc^2\nu^3\frac{1}{\mathrm{e}^{hc\nu/(k_bT)}-1} \tag{9}$$

式中：g_{u}和g_{l}分别为上、下能级的统计权重；k_b为玻尔兹曼常数；T为当地气体温度。

令$k'_\nu = k_\nu\frac{g_{\mathrm{u}}}{g_{\mathrm{l}}}\left[1-\mathrm{e}^{-hc\nu/(k_bT)}\right]$，联立式 (7)～式 (9) 得

$$j_\nu = \left\{k_\nu B_\nu(T)\left[\mathrm{e}^{hc\nu/(k_bT)}-1\right]+k_\nu I_\nu(\Omega)\right\}\frac{g_{\mathrm{l}}}{g_{\mathrm{u}}}\mathrm{e}^{-hc\nu/(k_bT)} = k'_\nu B_\nu(T)+\frac{g_{\mathrm{l}}}{g_{\mathrm{u}}}\mathrm{e}^{-hc\nu/(k_bT)}k_\nu I_\nu(\Omega) \tag{10}$$

采用k'_ν代替k_ν相当于对吸收系数k_ν进行了修正，考虑了辐射上下能级的能级统计权重 g 对谱线强度的影响，计算实际谱线强度更加可靠。该修正系数在低频段有比较明显的修正作用，能提高本文对羽流长波红外辐射特性的计算准确性。此处计算只考虑了一种成分一条谱线的辐射，若要计算所有成分所有谱线的发射系数，只需将它们线性叠加即可。

1.2　Mie 理论与微粒辐射参数

多相羽流中，由于存在固体微粒，它们的辐射特性不同于气体。1908 年，Mie 通过求解麦克斯韦方程组，得到了均匀介质球远场弹性散射的严格解，可以计算任意大小、任意材料的球形颗粒在光场中的吸收、发射和散射情况，且在微粒足够小的情况下对于形状不规则的微粒同样适用[19]。

根据 Mie 理论，均匀球形介质颗粒散射的消光系数k_{ext}、吸收系数k_{abs}、散射系数k_{sca}分别表示为[19]

$$k_{\mathrm{ext}} = n\pi r^2\frac{2}{\alpha^2}\sum_{i=1}^{\infty}(2i+1)\mathrm{Re}(a_i+b_i) \tag{11}$$

$$k_{\mathrm{sca}} = n\pi r^2\frac{2}{\alpha^2}\sum_{i=1}^{\infty}(2i+1)\left(|a_i|^2+|b_i|^2\right) \tag{12}$$

$$k_{\mathrm{abs}} = k_{\mathrm{ext}}-k_{\mathrm{sca}} \tag{13}$$

式中：r为颗粒半径，用于表征颗粒大小；n为颗粒数密度；无量纲量α为$\pi r/\lambda$；λ为电磁辐射波长；a_i、b_i为 Mie 系数，用 Ricatti-Bessel 函数ψ和ξ表示为

$$a_i = \frac{\psi_i(\alpha)\psi'_i(m\alpha)-m\psi'_i(\alpha)\psi_i(m\alpha)}{\xi_i(\alpha)\psi'_i(m\alpha)-m\xi'_i(\alpha)\psi_i(m\alpha)} \tag{14}$$

$$b_i = \frac{m\psi_i(\alpha)\psi'_i(m\alpha)-\psi'_i(\alpha)\psi_i(m\alpha)}{m\xi_i(\alpha)\psi'_i(m\alpha)-\xi'_i(\alpha)\psi_i(m\alpha)} \tag{15}$$

其中：m 为介质的复折射率。

经过估算，本文讨论的尾焰羽流中冰晶尺度$\alpha \ll 0.1$，式 (11)～式 (13) 取第 1 项即可，泰勒展开后取非零的第一阶近似可以得到瑞利估计表达式[19]：

$$k_{\mathrm{ext}} = 4n\pi r^2\alpha\mathrm{Im}\left(\frac{m^2-1}{m^2+2}\right) \tag{16}$$

$$k_{\mathrm{sca}} = \frac{8}{3}n\pi r^2\alpha^4\left|\frac{m^2-1}{m^2+2}\right| \tag{17}$$

由于$\alpha \ll 0.1$，$k_{\mathrm{sca}} \ll k_{\mathrm{ext}}$，从而$k_{\mathrm{abs}} \approx k_{\mathrm{ext}}$。

1.3　辐射传输方程

同时考虑介质对辐射的吸收和发射，辐射传输

方程可以写成[18]

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s}=-k_\nu I_\nu+j_\nu \tag{18}$$

式中：I_ν为辐射亮度。

将式(10)代入并考虑微粒热辐射，整理得气体辐射传输方程：

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s}=-k'_\nu I_\nu+k'_\nu B_\nu(T)+k_{\mathrm{abs}}B_\nu(T_{\mathrm{ice}}) \tag{19}$$

式中：$B_\nu(T)$为温度T下的黑体辐射亮度；T_{ice}为冰晶温度。

而微粒辐射的辐射传输方程为[20]

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s}=-k_{\mathrm{ext}}I_\nu+\frac{k_{\mathrm{sca}}}{4\pi}\int I_\nu(\Omega_{\mathrm{in}})\Phi(\Omega_{\mathrm{in}},\Omega)\mathrm{d}\Omega_{\mathrm{in}} \tag{20}$$

式中：Φ为散射相函数；Ω_{in}为入射方位角；Ω为辐射方位角。

2种辐射叠加整理后可写为如下一般形式：

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s}=-\beta(s)I_\nu+J(s) \tag{21}$$

式中：综合吸收系数β和综合发射系数J表示为

$$\beta=k'_\nu+k_{\mathrm{ext}} \tag{22}$$

$$J=k'_\nu B_\nu\left(T_{\mathrm{gas}}\right)+k_{\mathrm{abs}}B_\nu(T_{\mathrm{ice}})+k_{\mathrm{sca}}\int I_{\mathrm{SE}}(\Omega_{\mathrm{in}})\Phi(\Omega_{\mathrm{in}},\Omega)\mathrm{d}\Omega_{\mathrm{in}} \tag{23}$$

求解该方程得解析解：

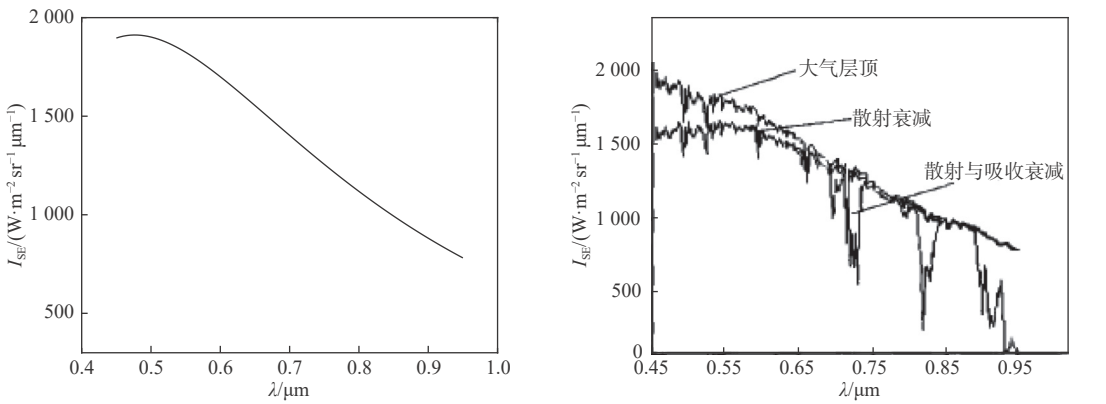
$$I_\nu(s)=I_0\mathrm{e}^{-\int_0^s\beta(s')\mathrm{d}s'}+\int_0^s J(s')\mathrm{e}^{-\int_{s'}^s\beta(s'')\mathrm{d}s''}\mathrm{d}s' \tag{24}$$

式中：I_0为背景辐射亮度；右端第1项表示辐射透射亮度，第2项表示总发射亮度，直接对辐射光路利用传输方程路径积分法求解。

1.4 背景辐射的处理

辐射传输方程的边界条件即为背景辐射，本文选取宇宙3 K背景辐射，即温度为3 K的黑体辐射。

考虑到可能存在的强背景光源，如太阳或地球表面的辐射在羽流区产生的散射光会对辐射强度提供较大贡献，必须将其考虑在内。对于太阳辐射，本文采用的模型是5 900 K的黑体辐射模型，地球附近接收到的太阳辐射亮度可以表示为$I_{\mathrm{SE}}=R_{\mathrm{S}}^2I_{\mathrm{S}}/L_{\mathrm{SE}}^2$，其中，$R_{\mathrm{S}}$和$L_{\mathrm{SE}}$分别为太阳半径和日地平均距离，$I_{\mathrm{S}}$和$I_{\mathrm{SE}}$分别为太阳表面处和地球附近的太阳辐射亮度。根据该模型求得的地球附近的太阳辐射亮度与文献[21]给出的一致，如图1所示。



(a) 本文背景辐射模型地球附近太阳辐射亮度　　(b) 文献[21]给出的地球附近太阳辐射亮度(大气内外)

图1　地球附近太阳辐射亮度

Fig. 1　The Sun's radiance near the Earth

2 羽流多波段辐射计算

2.1 羽流流场参数

本文研究基于某轨控发动机多相羽流仿真的流场结果。流场计算采用Bird的DSMC方法[22]。两相流场计算采用Boyd双向耦合算法，计算固体颗粒和气相之间的相互作用[23]。基于稀薄两相羽流双向耦合DSMC方法，采用经典成核理论，建立了真空羽流的相变过程的数值仿真方法[24]。基于上述方法构建了一套包含相变过程的DSMC方法，对某轨控发动机羽流开展了仿真。

仿真条件为：喷管入口压强2.5 MPa，燃烧室总温3 000 K，羽流辐射计算域大小及流场计算结果云图如图2和图3所示。羽流流场数据包含了各场点处的温度、压强、气体分子以及冰晶粒子密度、各组分分压等参数。羽流主要气体成分及其占比如表1所示。

从图3可以看出，由于燃烧室燃气经过大面积比喷管膨胀后羽流内部温度较低，因此羽流中不仅仅含有气体分子，还含有部分冰晶微粒。并且羽流中2种成分的辐射特性并不相同，故计算辐射时不仅要考虑气体辐射，也要考虑冰晶颗粒对辐射的吸收、发射和散射。从表1中还可以看出，该轨控发动机羽流中除冰晶外几乎不产生其他固体微粒，因此研究微粒辐射时仅考虑冰晶微粒的影响。

羽流中冰晶微粒粒度d_{ice}如图4所示，约为

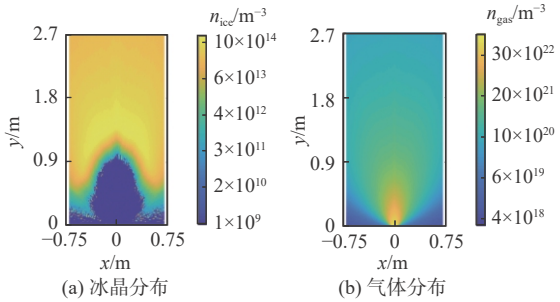


图 2　冰晶和气体的微粒数密度分布

Fig. 2　Number density distributions of particles for ice crystals and gases

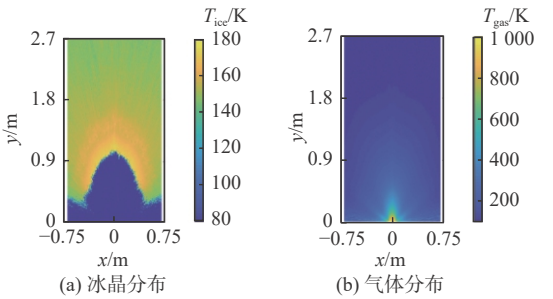


图 3　冰晶和气体的温度分布

Fig. 3　Temperature distributions of ice crystals and gases

表 1　羽流气体分压占比

Table 1　Partial pressure proportions of gases in plume

组分	占比/%
N_2	30.53
H_2O	31.13
CO	14.21
H_2	20.99
CO_2	3.14

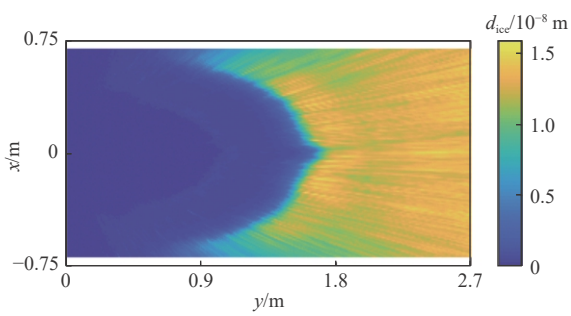


图 4　冰晶粒度分布

Fig. 4　Ice crystal size distribution

10^{-8} m 量级，满足远小于所研究辐射波长的条件。从图中可以看到，冰晶颗粒从喷口下游 1 m 左右的位置开始生成，并逐渐生长，在 1.5～2 m 处逐渐生长到最大。

2.2　气体辐射吸收系数计算

气体辐射吸收系数 k_ν 可以利用谱线库进行计算，常见的谱线库有美国的 HITRAN 谱线库、HITEMP 谱线库以及 JPL 谱线库等，本文采用的是最新的 HITRAN2020 数据库。HITEMP 谱线库对高温气体的一些谱线进行了修正，但是本文使用的算例气体温度较低，并不需要使用 HITEMP 库。

计算辐射吸收系数的方法有多种，如 LBL 方法、谱带模型、总体模型。本文采用分辨率较高的 LBL 方法。谱线数据经由 HITRAN 官方提供的 python 第三方库 HAPI 下载[25]。

谱线由线强参数和线形函数 2 部分相乘得到，谱线库中的线强参数采用的是 p_{ref}=1.013×10^5 Pa，T_{ref} = 296 K 条件下的数值，可用式(25)化为工作温度下的线强：

$$S(T)=S(T_{ref})\frac{Q(T_{ref})}{Q(T)}\cdot\frac{e^{-c_2E''/T}}{e^{-c_2E''/T_{ref}}}\cdot\frac{1-e^{-c_2\nu'/T}}{1-e^{-c_2\nu'/T_{ref}}} \tag{25}$$

式中：$c_2=hc/k_b$；Q 为总内配分函数；E'' 为低能态能量对应的波数；ν' 为谱线真空中心波数。

谱线线型本文采用的是洛伦兹线形函数，其展宽和线形函数由式(26)决定：

$$\gamma(p,T)=\sum_k\left(\frac{T_{ref}}{T}\right)^{n_k}\gamma_k p x_k \tag{26}$$

$$f(\nu;\nu',T,p)=\frac{1}{\pi}\cdot\frac{\gamma(p,T)}{\gamma(p,T)^2+\left[\nu-\left(\nu'+\sum_k\delta_k p\right)\right]^2} \tag{27}$$

式中：γ_k、n_k、δ_k 和 x_k 分别为背景 k 物质的展宽半宽、展宽半宽的温度系数、谱线波数的压致位移系数和组分占比。

进而辐射吸收系数 k_ν (cm^{-1}) 可表示为

$$k_\nu(\nu,T,p,x_k)=N(x_k)S(T)f(\nu;\nu',T,p) \tag{28}$$

式中：N 为分子数密度。

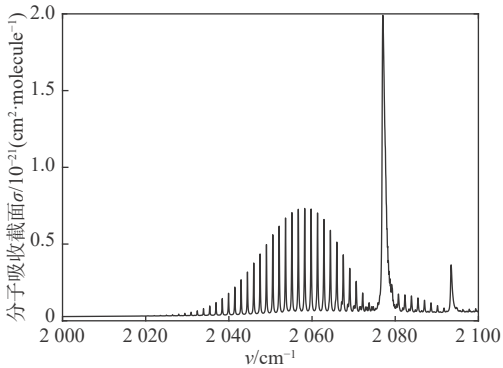
图 5 是用上述方法计算的 296 K、1.013×10^5 Pa 下的 CO_2 分子的吸收系数频谱，与文献 [25] 进行对比，可见两者的结果是一致的。

2.3　微粒辐射与散射计算

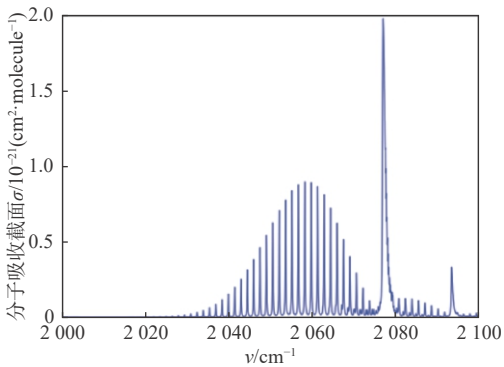
对 Mie 理论微粒吸收系数和散射系数的计算，采用 1.2 节得到的模型，利用多物理场仿真软件 COMSOL 材料库中冰晶的复折射率数据，绘制冰晶微粒在波数 1 000～4 000 cm^{-1} 频段吸收系数与散射系数谱如图 6 所示。可以看出，冰晶微粒的吸收峰与 H_2O 分子的吸收系数都在 1 500～2000 cm^{-1} 和 3 000～3 500 cm^{-1} 波段存在峰值，具有一定的一致性，符合一般的认知，也说明了本文微粒辐射特性计算模型是可靠的。

2.4　计算结果

根据第 1 章中介绍的理论和计算方法，计算了高空轨控发动机含冰晶多相羽流在吸收厚度为



(a) 本文计算的 CO_2 频谱



(b) 文献[25]给出的 CO_2 频谱

图 5 CO_2 吸收系数谱比较

Fig. 5 Comparison of CO_2 absorption coefficient spectra

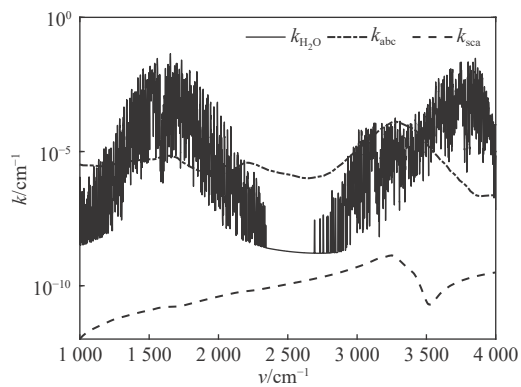


图 6 冰晶吸收系数与散射系数($d = 3\times10^{-8}$ m, $n = 1\times10^{14}$ m^{-3})

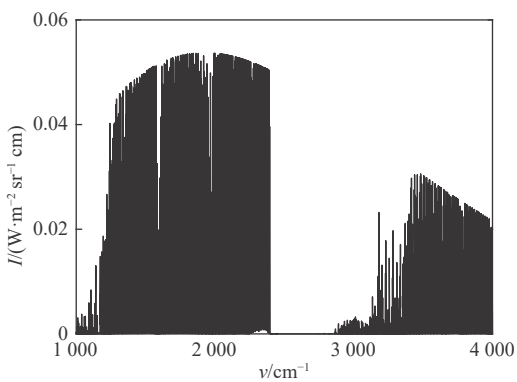
Fig. 6 Absorption coefficient and scattering coefficient of ice crystals ($d = 3 \times 10^{-8}$ m, $n = 1 \times 10^{14}$ m^{-3})

射发射谱,可见由于高空羽流压强较低,辐射谱线的展宽明显较小,且随着离喷口距离越远,由于流场温度迅速降低,辐射功率也迅速降低,辐射能量更加集中在低频段。

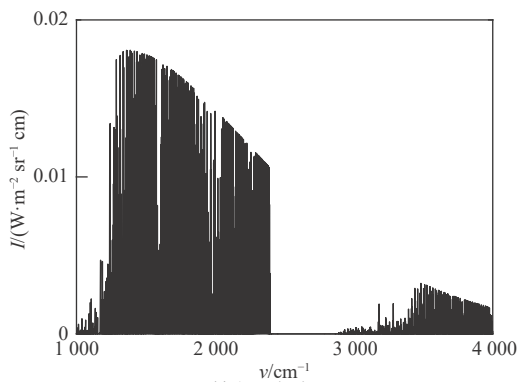
羽流大部分区域的热辐射集中于中长波红外波段,而由图 1 太阳光辐射在近红外到可见光波段达到峰值,冰晶散射强度的峰值也将在近红外到可见光波段,因此这几个波段是主要研究的范围。计算域内总辐射计算结果总结于表 2。

本文辐射亮度结果与研究的尾焰羽流成分相似的文献 [15] 中的仿真结果在温度相近的区域量

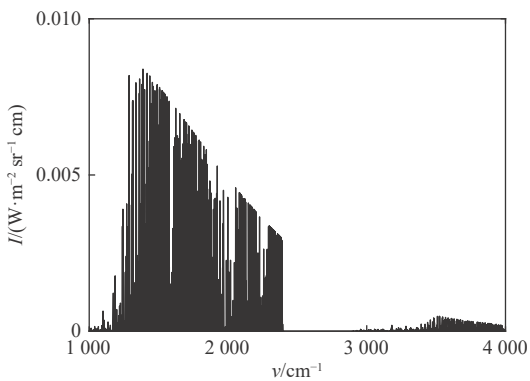
0.3 m时轴向距离喷口 2 cm、12 cm、22 cm、32 cm 处的综合热辐射谱如图 7 所示。根据图 7 的热辐



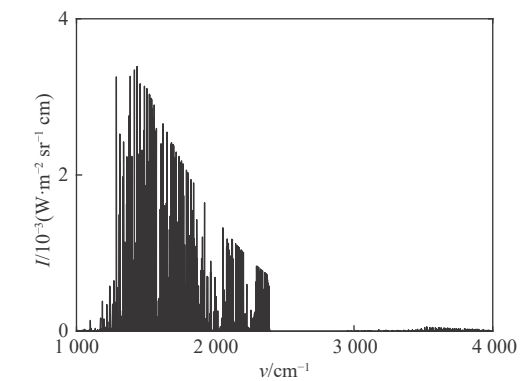
(a) 轴向距离喷口 2 cm



(b) 轴向距离喷口 12 cm



(c) 轴向距离喷口 22 cm



(d) 轴向距离喷口 32 cm

图 7 尾焰核心区域热辐射发射谱

Fig. 7 Thermal emission spectra in core region of plume

表 2　计算域内各波段羽流辐射与散射计算结果

Table 2　Calculation results of plume radiation and scattering in each band in calculation domain

辐射波段	波长/μm	总辐射量/(W·sr^{-1})	辐射在该波段占比/%
近红外波段	0.77～1.25	3.7×10^{-6}	100
中长波红外波段	2.5～3.3	1.83×10^{-4}	18.9
	3.3～5	4.71×10^{-4}	48.5
	5～10	3.17×10^{-4}	32.6

级上一致，因此表中在计算域内积分得到的总辐射量具有一定的参考意义。如表 2 所示，可见，总体来说由于羽流大部分区域辐射强度较低，近红外至光学波段的热辐射弱，中长波红外波段辐射较强，其中总辐射量最强的波段在 3.3～5 μm 的中波红外波段。本节将对各波段辐射的具体构成展开分析。

2.4.1　近红外波段

近红外波段的热辐射主要由 H_2O 分子贡献。图 8 为采用本文研究的轨控发动机羽流核心处的组分数据，在 0.77～1.25 μm 波段光谱计算结果。可见，该波段主要的辐射光带包含了水分子的 0.77～0.87 μm、0.87～1.04 μm、1.04～1.25 μm 3 个辐射光带和 CO_2 分子 1.17～1.25 μm 的辐射光带。羽流中其他成分的气体在该波段辐射吸收能力较弱，几乎是透明的。

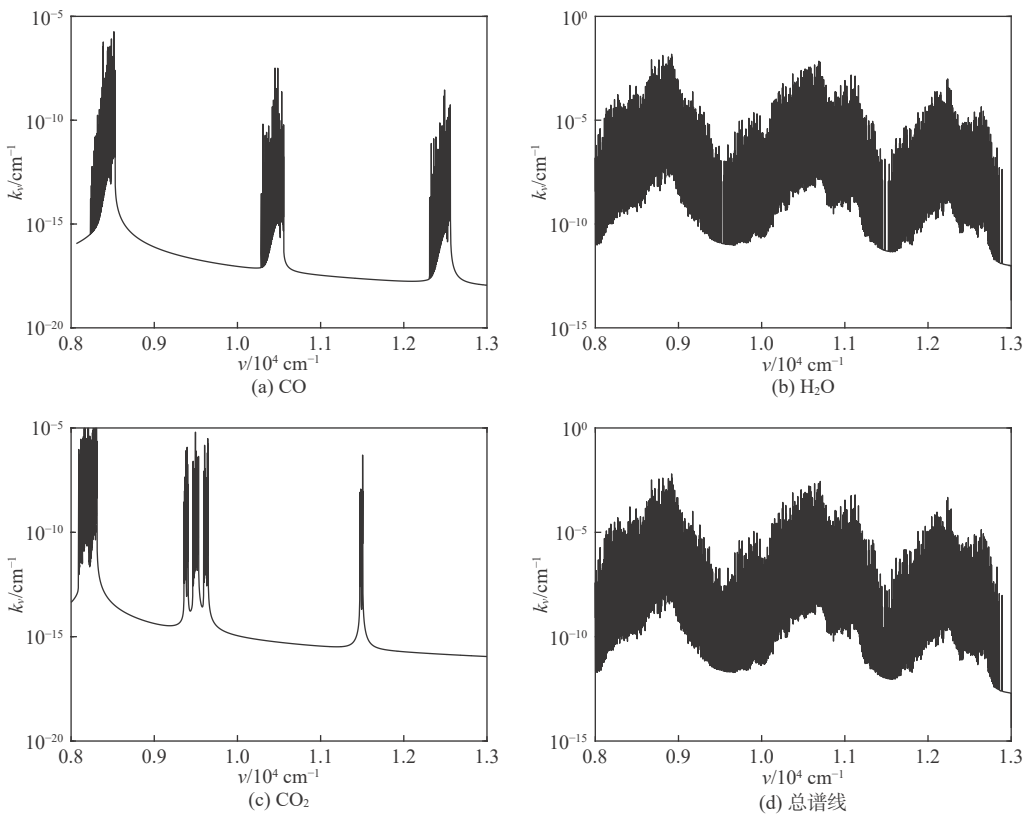


图 8　0.77～1.25 μm 波段辐射吸收系数(p=500 Pa, T=296 K)

Fig. 8　Radiation absorption coefficient of 0.77–1.25 μm band (p = 500 Pa, T = 296 K)

图 9 为 0.77～1.25 μm 波段尾焰流场的辐射亮度分布云图。可见，羽流辐射亮度分布存在明显的分段现象。以喷口下游 0.5 m 为界，辐射主要区域分成集中于喷口周围数厘米范围的高温气体辐射区域以及下游弥散的冰晶辐射区域。出口高温气体辐射区域的辐射亮度大但面积小，冰晶弥散区域辐射亮度小，但可以延伸到较远的距离。

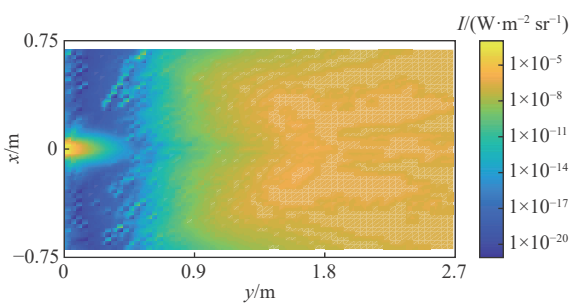


图 9　0.77～1.25 μm 波段的尾焰辐射亮度分布

Fig. 9　Radiance of plume in 0.77–1.25 μm band

2.4.2　中长波红外波段

随着辐射波长增大，冰晶的散射效应也逐渐减弱，因此热辐射逐渐占据主导地位。采用本文研究的轨控发动机羽流核心处的组分数据，在波数 1 000～4 500 cm^{-1} 频段，即波长为 2.4～10 μm 波段绘制 H_2O 分子和 CO_2 分子的吸收系数谱图如图 10 所示。

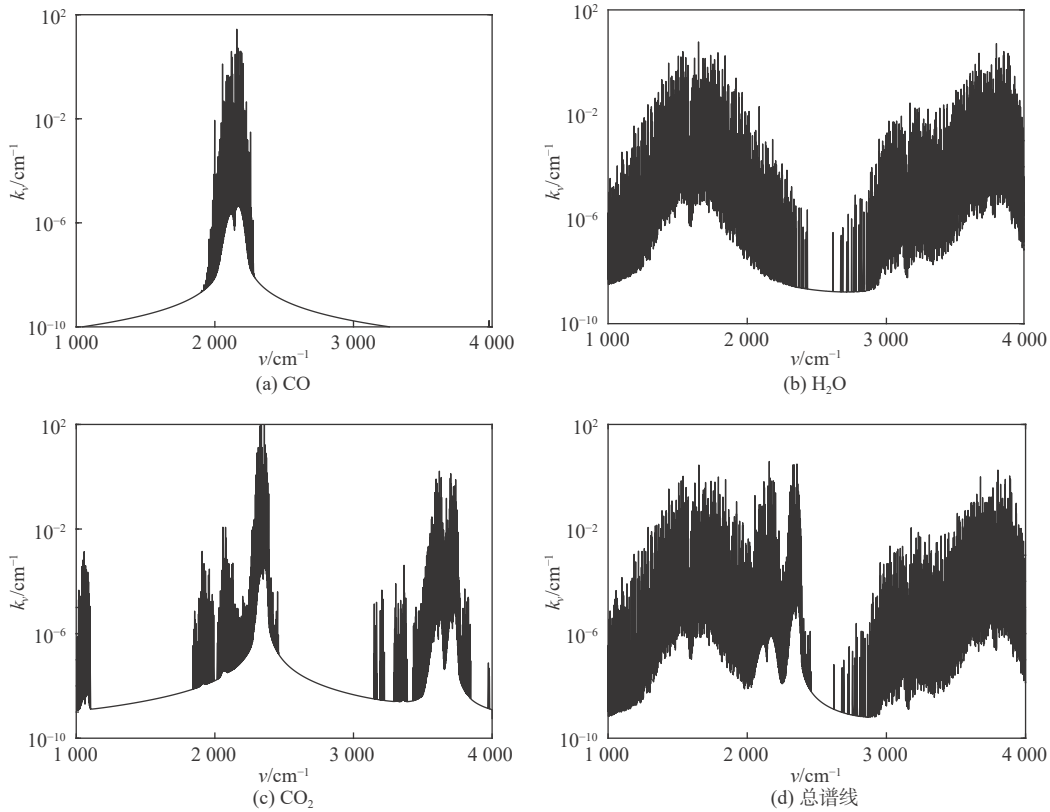


图 10 2.4～10 μm 波段辐射吸收系数(p=500 Pa, T=296 K)

Fig. 10 Radiation absorption coefficient of 2.4−10 μm band (p = 500 Pa, T = 296 K)

在 1 000 K 左右的气体温度下，尾焰气体的主要辐射光带都在中长波红外波段，从图中可以清晰地看出该波段的谱线包含了 H_2O 的第 1 辐射光带(2.4～3.3 μm)与第 2 辐射光带(4.5～8 μm)和 CO_2 的第 1 辐射光带(2.64～2.80 μm)与第 2 辐射光带(4.11～4.80 μm)，可见，不同于近红外波段主要的谱线都是由 H_2O 分子贡献，在中长波红外波段，CO 与 CO_2 对辐射谱线均有较强的贡献。

图 11 为波长 2.5～10 μm 波段尾焰流场的辐射亮度分布云图。可见，羽流辐射同样存在分段现象。以喷口下游 1 m 左右为界，辐射主要集中于喷口周围数十厘米范围的高温气体辐射区域，在辐射亮度上远大于近红外波段，以及冰晶弥散区域的亮度小面积大的辐射区域。

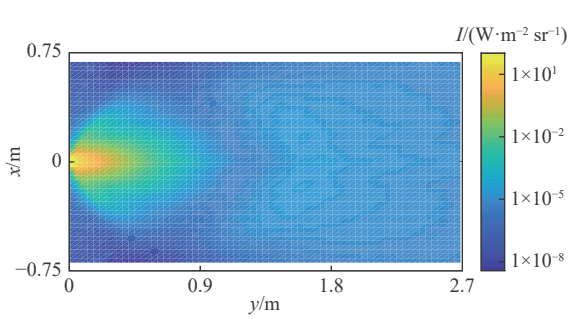


图 11 2.5～10 μm 波段的尾焰辐射亮度分布

Fig. 11 Radiance of plume in 2.5−10 μm band

2.5 不同波段尾焰辐射对比分析

对比图 9 与图 11 可以发现，在近红外波段和中长波红外波段，一方面，尾焰辐射分布结构有显著不同，同样是前后两段结构，分界线位置、两段辐射强弱关系均不同，另一方面，2 个波段羽流的总体辐射强度也存在十几个数量级的显著差异。经分析，这些差异与该发动机多相羽流组分的分布特征和辐射源特征有关。

2.5.1 多相羽流组分分布对辐射特性的影响

多相羽流组分分布对辐射特性的影响主要体现为空间分布的差异。为分析羽流中冰晶与气体各组分分布情况对辐射的贡献，对近轴沿轴向的气体、冰晶组分的密度、温度分布以及羽流辐射分布绘制线图如图 12～图 14 所示。需要说明，图 13、图 14 中的辐射亮度表示在近轴区域沿观察方向路径积分得到的辐射亮度，因此尽管图 12 喷口附近没有冰晶微粒，但是由于径向较远处的积分路径上存在冰晶颗粒，从而在图 13、图 14 中距喷口小于 1 m 的区域内能观察到冰晶辐射和散射。

对比 2 个波段的冰晶散射、冰晶热辐射、气体热辐射 3 种辐射的贡献可以看到，在距离喷口较近的区域，由于冰晶微粒尚未形成，近红外波段辐射是以气体热辐射为主，而在较远的区域，气体温度

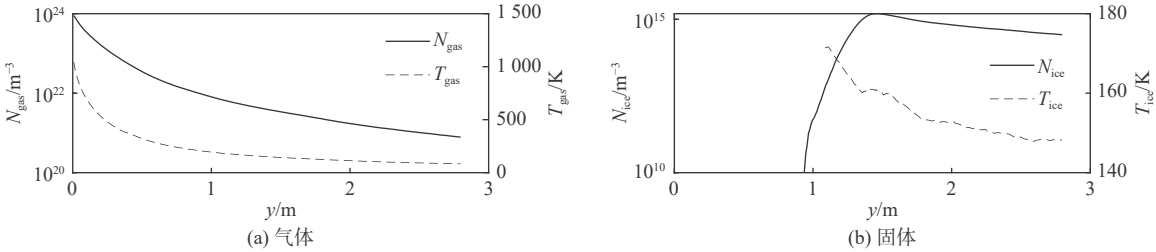


图 12　近轴区域气固两相密度与温度分布

Fig. 12　Density and temperature distributions of gas and ice crystals in paraxial region

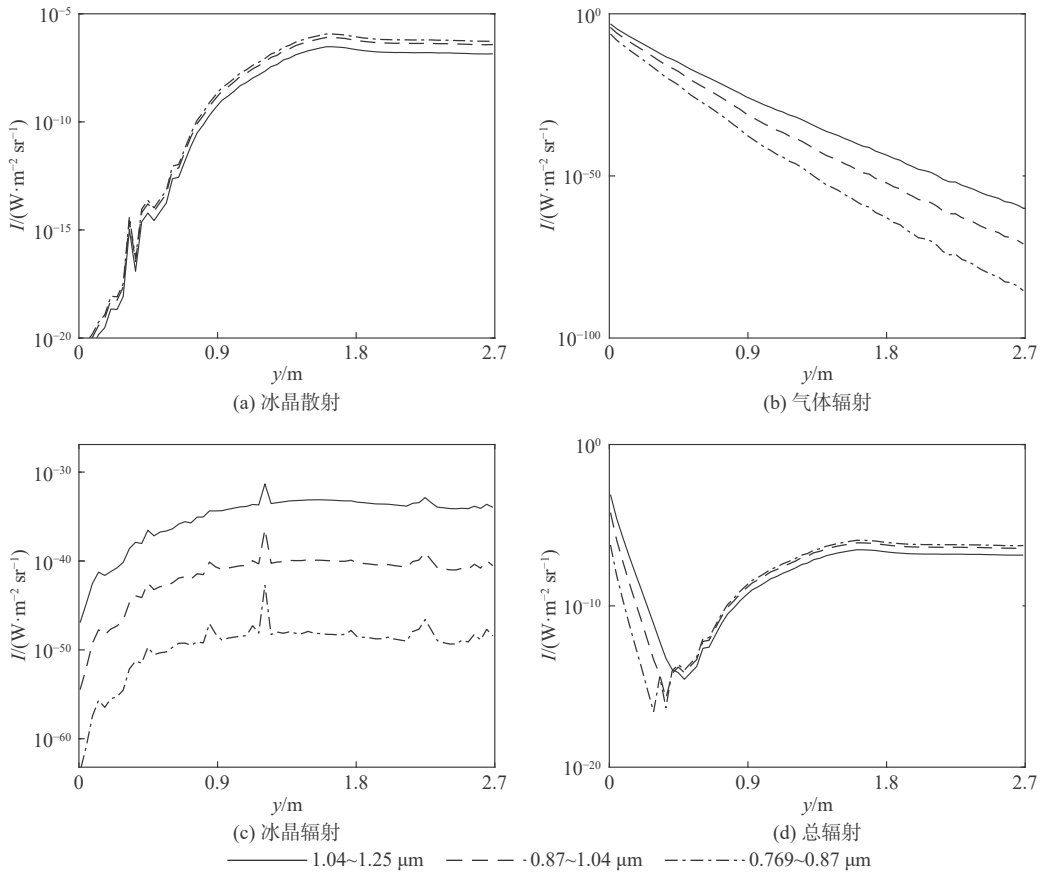


图 13　沿轴线方向羽流各组分近红外波段辐射贡献

Fig. 13　Near-infrared radiation contribution of plume components along axis

压强骤降，冰晶逐渐增多且温度密度较为稳定，辐射转而以冰晶的散射和热辐射为主，这就形成了如图 9 与图 11 的独特辐射分布特征。

2.5.2　辐射源对辐射特性的影响

辐射源对辐射特性的影响主要体现为频谱分布的差异。对比图 13 与图 14，可以发现随波长减小，冰晶散射强度增强，到近红外波段，下游的冰晶区散射成为辐射主导成分；而随波长增大，羽流气固两相的热辐射强度以数十个量级的速度上升，到中长波红外波段，不论是上游的高温气体区还是下游的冰晶区，热辐射逐渐成为辐射主导成分。这是由于冰晶散射的辐射源是对应温度约 5 900 K 的太阳，而羽流气固两相的热辐射辐射源则是温度几百 K 的羽流本身。

为研究 2 种辐射源的区别，采用对应温度的黑体辐射模型来定性分析辐射强度关于温度和波长的变化关系。在本文选定的 2 个波段，对比太阳 5 900 K 和羽流 300 K 2 种特征温度下的黑体辐射谱，如图 15 所示。可见，随波长增大，5 900 K 的黑体辐射强度降低约 2～3 个数量级，而 300 K 的黑体辐射强度显著升高了十几个数量级，这就解释了图 13 与图 14 中羽流气固两相的热辐射在不同波段下差异巨大，而冰晶散射在 2 个波段的差异相对较小，也解释了为什么随波长增大冰晶散射逐渐减弱而多相羽流本身的热辐射迅速增强占据主导地位。

另一方面，从图 15 中还可见，近红外波段辐射强度随温度变化远比中长波红外波段更敏感，表现在图 13 与图 14 中，即近红外波段气体辐射随轴向

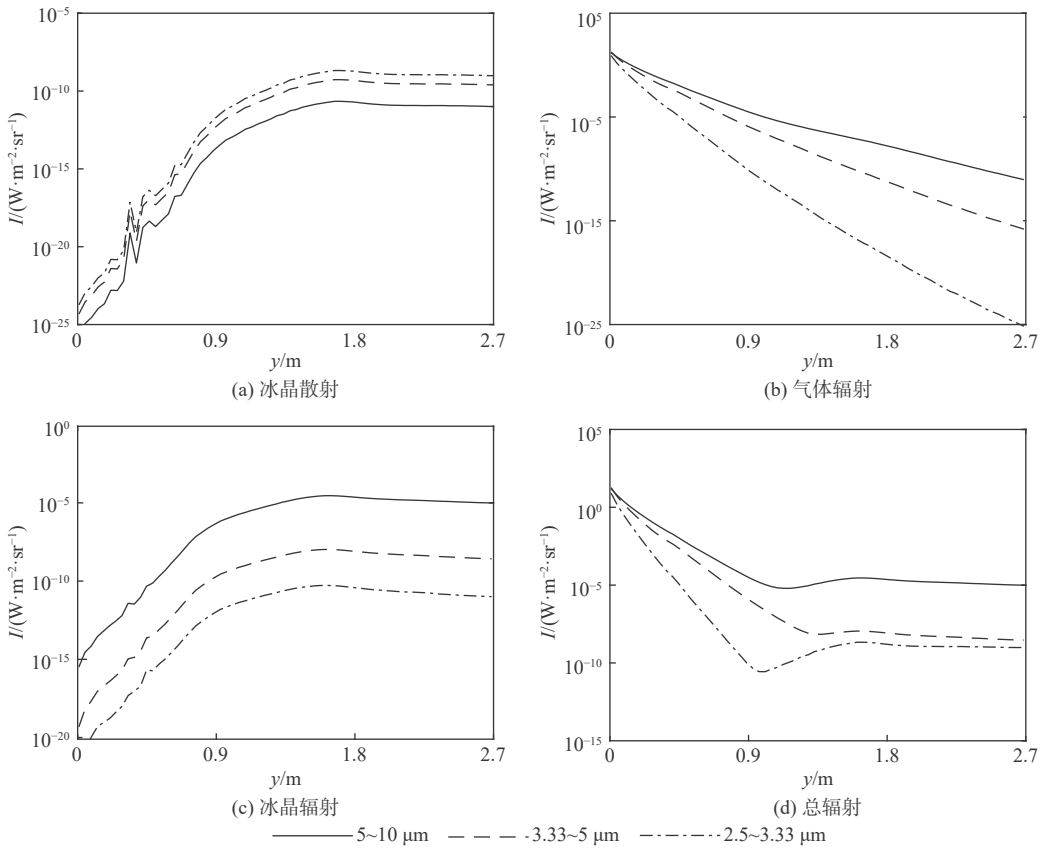


图 14 沿轴线方向羽流各组分中长波红外波段辐射贡献

Fig. 14 Mid-wave and long-wave infrared radiation contribution of plume components along axis

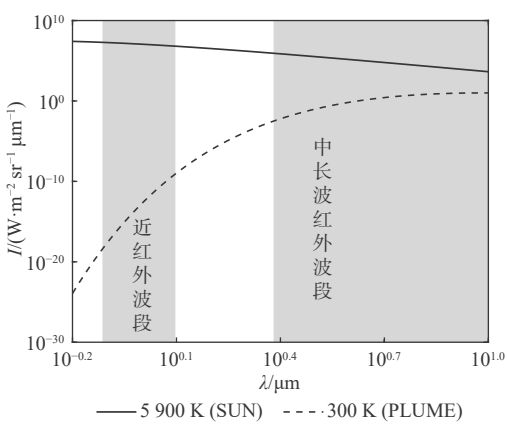


图 15 5 900 K 和 300 K 下黑体辐射强度对比

Fig. 15 Comparison of blackbody radiation intensity at 5 900 K and 300 K

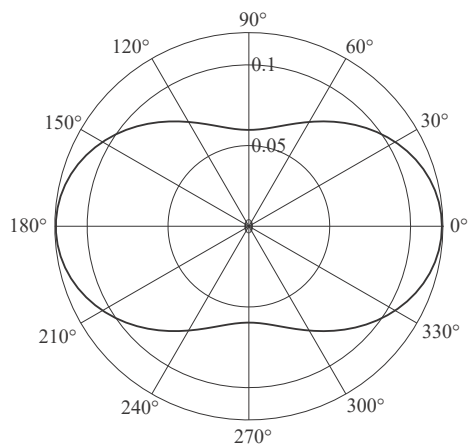


图 16 瑞利散射模型散射相函数

Fig. 16 Scattering phase function of Rayleigh scattering model

距离下降的斜率比中长波红外波段更大，这就导致与冰晶辐射或散射曲线的上升沿交点提前，从而图 9 相较于图 11 中辐射分布的分段点更为靠近喷口。

此外，随着波长增大，散射强度不断减小，除了由于前面提到的阳光辐射强度本身随波长增大而减小外，还有一个原因，即散射辐射波长增大导致了尺度因子α减小。值得注意的是，由于瑞利散射模型的散射光强与散射角满足图 16 所示的散射相函数[19]，因此随着太阳光入射角的变化，散射光强会有最大约 50% 的变化。本文计算散射光的默认太阳光入射方向与视线方向夹角为 30°。

3 结 论

1）本文研究的某型轨控发动机多相羽流中对辐射产生贡献的主要成分包括 CO、CO_2 和 H_2O 气体以及冰晶微粒，羽流温度约 200～1 000 K，同时考虑到高空中存在 5 900 K 太阳光背景辐射，羽流的主要的辐射包括近红外波段的散射光和中长波红

外波段的气体热辐射。

2）从总辐射量来看，在近红外波段，计算域内总辐射量大约 10^{-6} W/sr 量级，在中长波红外波段计算域内总辐射量大约 10 W/sr 量级。

3）从羽流辐射的空间和频谱分布特征来看，羽流辐射空间上主要分为集中于喷口周围数厘米范围的高温气体辐射区域以及下游弥散的冰晶辐射区域。频谱分布上，在近红外波段，气体辐射区内热辐射占主导，且辐射弱，冰晶辐射区内散射占主导；而在中长波红外波段，气体辐射区内内热辐射占主导，冰晶辐射区内也是热辐射占主导，这些差异与该型发动机多相羽流组分的分布特征和辐射源特征有关。

4）由于冰晶尾迹的存在，羽流的近红外辐射中散射光占主导地位。由于瑞利散射的特性，随着太阳光入射角的变化，散射光强会有最大约 50% 的变化。

综上，对于该型轨控发动机尾焰，其在中长波红外波段的热辐射辐射最强，其次是在近红外波段的冰晶散射。由于羽流中存在大量冰晶微粒，且能弥散至羽流下游很远的位置，因此在近红外至光学波段，由冰晶散射太阳光产生的辐射可能形成一条面积较大的辐射尾迹，对于相关目标探测有一定意义。

参考文献（References）

[1] 聂万胜, 蔡红华. 火箭发动机尾焰红外辐射特性研究综述[J]. 装备学院学报, 2017, 28(1): 47-53.
NIE W S, CAI H H. Overview of infrared characteristics study of rocket engine plume[J]. Journal of Equipment Academy, 2017, 28(1): 47-53(in Chinese).

[2] ROCHELLE W C. Review of thermal radiation from liquid and solid propellant rocket exhausts[R]. Washington, D.C.: NASA, 1967: 7-201.

[3] NELSON H F. Infrared radiation signature of tactical rocket exhausts: AIAA 1982-913[R]. Reston: AIAA, 1982.

[4] SNAZA C J. Investigation of the effects of solid rocket motor propellant composition on plume signature[D]. Monterey: Naval Postgraduate School, 1994: 17-25.

[5] HARWELL K, HYMAN W H, JACKSON J J, et al. Effects of external flow velocity on the spatial distribution of infrared radiation from a rocket exhaust plume[C]//Proceedings of the 11th Thermophysics Conference. Reston: AIAA, 1976: 443.

[6] 王平. 固体火箭发动机尾焰毫米波辐射特性研究[D]. 武汉: 华中科技大学, 2015.
WANG P. Study on millimeter wave radiation characteristics of solid rocket motor tail flame[D]. Wuhan: Huazhong University of Science and Technology, 2015(in Chinese).

[7] 张小英, 朱定强, 向红军, 等. 液体火箭喷焰红外特性的数值仿真[J]. 北京航空航天大学学报, 2005, 31(11): 1250-1253.
ZHANG X Y, ZHU D Q, XIANG H J, et al. Numerical simulation of infrared characteristics of liquid rocket plume[J]. Journal of Beijing University of Aeronautics and Astronautics, 2005, 31(11): 1250-1253(in Chinese).

[8] 郑海晶. 火箭发动机尾焰多波段辐射特性研究[D]. 北京: 北京理工大学, 2018.
ZHENG H J. Study on multi-band radiation characteristics of rocket engine tail flame[D]. Beijing: Beijing Institute of Technology, 2018(in Chinese).

[9] 乔野, 聂万胜, 丰松江, 等. 复燃对氢氧火箭发动机尾焰流场及辐射特性影响数值研究[J]. 导弹与航天运载技术, 2016(2): 22-25.
QIAO Y, NIE W S, FENG S J, et al. Numerical research on influence exerted by afterburning on flow field and radiation characteristics of LH2/LOX rocket engine[J]. Missiles and Space Vehicles, 2016(2): 22-25(in Chinese).

[10] BAO X D, YU X L, WANG Z H, et al. Numerical investigation on flow and radiation characteristics of solid rocket motor plume near the ground[J]. Procedia Computer Science, 2020, 174: 645-650.

[11] BURT J, BOYD I. A Monte Carlo radiation model for simulating rarefied multiphase plume flows[C]//Proceedings of 38th AIAA Thermophysics Conference. Reston: AIAA, 2005: 4691.

[12] YU Q Z, LIU L H, PAN Y C, et al. Monte Carlo method for simulating the radiative characteristics of an anisotropic medium[J]. Heat Transfer—Asian Research, 1999, 28(3): 201-210.

[13] 郑才浪, 朱定强, 乔要宾. 固液混合火箭发动机尾焰的红外特性研究[J]. 红外, 2013, 34(10): 26.
ZHENG C L, ZHU D Q, QIAO Y B. Study of infrared radiation characteristics of hybrid rocket engine exhaust plume[J]. Infrared, 2013, 34(10): 26(in Chinese).

[14] CAI H H, NIE W S, SU L Y, et al. Infrared radiation characteristics of liquid oxygen/kerosene rocket engine plume with different number of nozzles[J]. Spectroscopy Letters, 2019, 52(3-4): 159-167.

[15] 郑才浪, 朱定强, 乔要宾. 轨控发动机羽流红外辐射的数值仿真[J]. 宇航学报, 2014, 35(5): 521-527.
ZHENG C L, ZHU D Q, QIAO Y B. Numerical simulation of the infrared radiation of orbit control thruster exhaust plume[J]. Journal of Astronautics, 2014, 35(5): 521-527(in Chinese).

[16] NIU Q L, HE Z H, DONG S K. IR radiation characteristics of rocket exhaust plumes under varying motor operating conditions[J]. Chinese Journal of Aeronautics, 2017, 30(3): 1101-1114.

[17] ZHANG D M, BAI L, WANG Y K, et al. An improved SHDOM coupled with CFD for simulating infrared radiation signatures of rocket plumes[J]. Infrared Physics & Technology, 2022, 122: 104054.

[18] 王福恒, 王嵩薇. 近代科学技术中的原子分子辐射理论[M]. 成都: 成都科技大学出版社, 1991: 33-35.
WANG F H, WANG S W. The theory of atomic and molecular radiation in modern science and technology[M]. Chengdu: Chengdu University of Science and Technology Press, 1991: 33-35(in Chinese).

[19] MIE G. Beiträge zur optik trüber medien, speziell kolloidaler Metallösungen[J]. Annalen der Physik, 1908, 330: 377-445.

[20] FRISVAD J R, CHRISTENSEN N J, JENSEN H W. Computing the scattering properties of participating media using Lorenz-Mie theory[C]//ACM SIGGRAPH 2007. New York: ACM, 2007: 60-es.

[21] 邱荣, 吴健, 杨春平, 等. 0.45-0.95μm 太阳辐射在大气中的传输计算[J]. 大气与环境光学学报, 2006, 1(1): 21-26.

QIU R, WU J, YANG C P, et al. Solar radiation transfer in clear atmosphere at 0.45～0.95 μm[J]. Journal of Atmospheric and Environmental Optics, 2006, 1(1): 21-26(in Chinese).

[22] BIRD G A. Molecular gas dynamics and the direct simulation of gas flows[M]. Oxford: Clarendon Press, 1994.

[23] BURT J M, BOYD I D. Monte Carlo simulation of a rarefied multiphase plume flow[C]//Proceedings of the 43rd AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2005.

[24] LOTHE J, POUND G M. Reconsideration of nucleation theory[J]. The Journal of Chemical Physics, 1962, 36(8): 2080-2085.

[25] KOCHANOV R V, GORDON I E, ROTHMAN L S, et al. HITRAN application programming interface (HAPI): a comprehensive approach to working with spectroscopic data[J]. Journal of Quantitative Spectroscopy and Radiative Transfer, 2016, 177: 15-30.

Radiation characteristics of multi-phase plumes containing ice crystals from orbit-control engines

LUO Wudi[1], REN Junxue[1,*], LI Zhihui[2,3], TANG Haibin[1]

(1. School of Astronautics, Beihang University, Beijing 100191, China;
2. China Aerodynamics Research and Development Center, Mianyang 621000, China;
3. National Laboratory for Computational Fluid Dynamics, Beijing 100191, China)

Abstract: To obtain the radiation characteristics of multi-phase plumes containing ice crystals from high-altitude orbit-control engines and analyze their influence on the telemetry effect, this paper deduced the gas radiation emission coefficient based on Einstein's theory of radiation. In the light of the Mie theory of particle scattering, the study developed a radiation transfer model comprehensively considering gas and ice crystal radiation and ice crystal scattering. Based on the latest HITRAN2020 spectral line library, the study obtained through simulation the radiation characteristic parameters of multi-phase plumes containing ice crystals from high-altitude orbit-control engines and the radiation intensity distributions in the near-infrared and mid-wave and long-wave infrared bands and analyzed the main influencing factors of the radiation. The results show that the radiation of the engine plume researched in this study primarily includes scattered light in the near-infrared band and the thermal radiation of gas in the mid-wave and long-wave infrared band. In the near-infrared band, considering solar background radiation, the total radiation amount in the calculation domain is about 10^{-6} W/sr, and the thermal radiation is dominant within the axial distance of 0.5 m, while the ice crystal scattering is dominant beyond the axial distance of 0.5 m. In the mid-wave and long-wave infrared band, the total radiation amount in the calculation domain is about 10 W/sr, and the thermal radiation of gas dominates within the axial distance of 1 m, while the thermal radiation of ice crystals dominates beyond the axial distance of 1 m.

Keywords: radiation; orbit-control engine; multi-phase plumes; heat radiation; scattering

Received: 2023-07-04; **Accepted**: 2023-08-15; **Published Online**: 2023-09-11 11:28

URL: link.cnki.net/urlid/11.2625.V.20230908.1404.003

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0440

基于双天线 TDCP 的 GNSS/IMU 组合导航航向角增强算法

吉莉[1,2]，孙蕊[1,2,*]，王媛媛[1,2]，戴晔莹[1,2]

(1. 南京航空航天大学 海上智能网信技术教育部重点实验室，南京 211106；2. 南京航空航天大学 空中交通管理系统全国重点实验室，南京 211106)

摘 要：在全球导航卫星系统 (GNSS) 和惯性测量单元件 (IMU) 组合导航系统中，姿态估计尤其是航向角的准确估计对于车辆行驶状态的实时监控尤为重要。但由于 IMU 在高度通道上是发散的，若不加以准确约束其误差会逐渐累积，因此，在航向角变化频繁的车载应用中其航向角的估计精度不足。为解决传统 GNSS 与 IMU 松组合方式导航姿态估计精度较差的问题，提出一种基于双天线历元间载波相位差分 (TDCP) 的 GNSS/IMU 组合导航航向角增强算法。该算法通过双天线 TDCP 求解车辆航向角来增加组合导航滤波融合观测值输入维度，并利用 Hatch 滤波和抗差自适应滤波分别实现观测域伪距精度提升和 GNSS/IMU 组合导航定位、定姿性能的提升。实验结果表明：所提算法相对于传统 GNSS/IMU 组合导航方法在三维方向上的定位、测速精度分别提高了 22.12%、41.27%，车辆航向角精度提高了 46.29%。

关 键 词：GNSS/IMU 组合导航；多天线；双天线历元间载波相位差分；抗差卡尔曼滤波；姿态精度

中图分类号：P228; V324; V249.3

文献标志码：A **文章编号**：1001-5965(2025)09-3141-09

全球导航卫星系统 (global navigation satellite system, GNSS) 与惯性测量单元 (inertial measurement unit, IMU) 在各自的定位性能上具有很强的互补性，所构建的 GNSS/IMU 组合导航系统广泛应用于车载系统中，实现车辆动态连续导航功能。车辆的姿态信息，尤其是航向角信息的准确获取对于地面载具的实时导航和异常行为监控十分重要。在低成本组合导航中，姿态信息往往仅由 IMU 获取，IMU 的高度通道发散会导致航向角估计不准确，此外，GNSS 部分常用伪距进行 GNSS 解算，由于卫星信号受反射导致伪距观测值出现较大的噪声及多径误差的问题一定程度上影响 GNSS/IMU 组合导航性能，其航向角的估计精度难以保证。为实现高精度的 GNSS 组合导航，通常使用载波相位观测量，包括基于精密单点定位 (precise point positioning, PPP) 和实时动态差分定位 (real time kinematic, RTK) 的 GNSS/IMU 组合导航方法，但载波相位观测在高动态环境及复杂环境下存在模糊度固定困难的问题，难以满足车辆实时动态航向角估计需求。

历元间载波相位差分 (time-differenced carrier phase, TDCP) 通过对相邻历元间的载波相位观测量作差，可以获取车辆在相邻历元间的位移量，具有噪声小、精度高、硬件需求低且无需进行模糊度固定的优点，可用于车辆姿态信息的求解[1-2]。但 TDCP 并不能直接提供车辆姿态信息，需要进行额外的推算[2]。文献 [2] 提出一种基于单天线 TDCP

收稿日期：2023-07-04；**录用日期**：2023-08-11；**网络出版时间**：2023-09-18 14:27

网络出版地址：link.cnki.net/urlid/11.2625.V.20230915.1408.002

基金项目：国家自然科学基金 (42222401,42174025,41974033)；工信部专项科研项目 (TC220A04A-79)；江苏省“六大人才高峰”项目 (KTHY-014)；江苏省自然科学基金 (BK20211569)；中央高校基本科研业务费专项资金 (xcxjh20220726)

* **通信作者**. E-mail：rui.sun@nuaa.edu.cn

引用格式：吉莉，孙蕊，王媛媛，等. 基于双天线 TDCP 的 GNSS/IMU 组合导航航向角增强算法 [J]. 北京航空航天大学学报，2025，51（9）：3141-3149. JI L，SUN R，WANG Y Y，et al. Heading enhancement algorithm of GNSS/IMU integrated navigation based on dual-antenna TDCP[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（9）：3141-3149（in Chinese）.

的车辆姿态近似估计方法，通过TDCP求解的车辆位移解算车辆航向角，但该方法无法在车辆速度较低时进行航向角估计，航向角估计的连续性无法保证。因此，有学者将TDCP与IMU进行组合来应对上述问题[3-9]。文献[4]提出一种基于噪声相关性改进卡尔曼滤波的TDCP/惯性导航系统（inertial navigation system, INS）紧组合算法，提高了城市环境中的车辆状态估计精度。文献[5]提出一种基于抗差扩展卡尔曼滤波的TDCP/INS组合导航算法，较传统组合导航算法实现了11.84%的航向角估计精度提升。但由于上述算法中载体的姿态信息仅由IMU观测得出，算法的航向角估计精度提升有限。此外，TDCP与IMU提供的观测量均为位移量，上述方法用于长时解算时无法避免误差累积导致结果发散的问题，影响车辆航向角估计精度。

针对上述问题，本文提出基于双天线TDCP的GNSS/IMU组合导航航向角增强算法。该算法主要分为2个部分，第1部分是基于双天线TDCP和观测域Hatch滤波的GNSS初始状态解算，由GNSS观测值直接解算车辆位置及航向角，构建包含绝对位置及姿态观测信息在内的多维观测量，实现航向角在滤波系统中的直接输入；第2部分是基于状态不符值和新息的抗差自适应卡尔曼滤波融合策略，实现GNSS/IMU组合导航参数的动态调节，使得组合导航系统在观测环境变化时能对GNSS观测值进行及时的修正。

1 本文算法框架

本文算法框架如图1所示，具体步骤如下。

1）根据GNSS观测数据分别对双天线位置参数进行解算。对前后天线分别通过如下4个阶段进行求解：第1阶段，由GNSS接收机得到的星历文件获取卫星位置，从而计算历元间卫星位置增量；第2阶段，由载波相位观测值在历元间作差得TDCP值，并利用多普勒频移对TDCP进行周跳检测与修复；第3阶段，使用Hatch滤波对初始伪距观测值进行滤波平滑，对平滑后的伪距基于最小二乘法（least square method, LSM）进行初始定位解算；第4阶段，根据历元间卫星位置增量、接收机与卫星的方向向量及TDCP对双天线分别解算历元间位置增量。本步骤输出结果为前天线的初始定位结果、前后天线的历元间位置变化量。

2）根据双天线位置参数对车辆的GNSS初始状态进行估计。将主天线的初始定位解作为GNSS估计的车辆初始位置，并根据双天线的历元间位置增量求解车辆初始速度与初始航向角。本步骤输出结果为车辆的GNSS初始位置、速度、姿态估计结果。

3）滤波融合与最终状态估计。IMU的原始观测量比力与角速度经力学编排后，与第2步得到的GNSS初始状态估计结果通过抗差自适应卡尔曼滤波进行融合，得出最终状态估计结果。本步骤的输出结果为车辆的最终位置、速度、姿态。

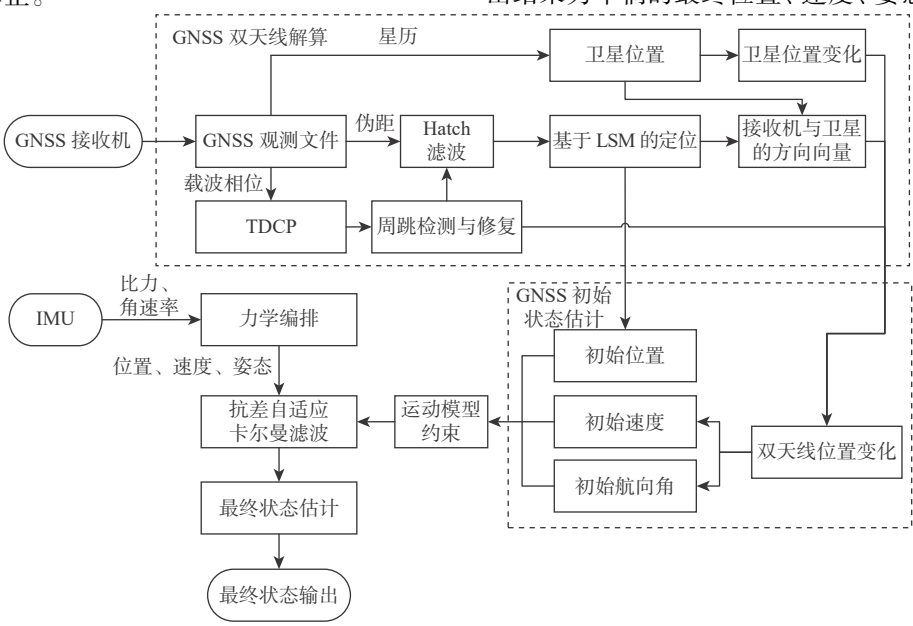


图1 本文算法框架

Fig. 1 The proposed algorithm framework

2 GNSS初始状态估计

2.1 TDCP解算位置增量

载波相位观测模型可以表示为

$$\Phi=\frac{1}{\lambda}(r-I_{\Phi}+T_{\Phi})+\frac{c}{\lambda}(\mathrm{d}t_{\mathrm{R}}-\mathrm{d}t_{\mathrm{S}})+N+\varepsilon_{\Phi} \tag{1}$$

式中：Φ为以周数表示的载波相位观测值；r为接收机与卫星间的几何距离；c为真空中的光速；λ为载

波波长; I_Φ为电离层延迟; T_Φ为对流层延迟; $\mathrm{d}t_\mathrm{R}$为接收机钟差; $\mathrm{d}t_\mathrm{S}$为卫星钟差; N为整周模糊度; ε_Φ为其余所有的未建模误差及观测噪声。上述所提到的部分误差在较短的采样间隔内变化十分微小, 因此可以忽略其在历元间的变化。此外, 当载波跟踪环路保持锁定时, 整周模糊度N也能够保持不变, 通过对观测方程在历元间作差分可得

$$\Delta\Phi = \Phi_{t(k+1)} - \Phi_{t(k)} = \frac{1}{\lambda}\Delta r + \frac{c}{\lambda}\Delta\mathrm{d}t_\mathrm{R} + \Delta\varepsilon_\Phi \tag{2}$$

式中: $\Delta\Phi$为历元间载波相位变化量; Δr为历元间卫星与接收机之间的几何距离的变化量; $\Delta\mathrm{d}t_\mathrm{R}$为接收机的钟差变化量。

卫星与接收机之间的几何关系如图 2 所示, $\boldsymbol{u}_{t(k+1)}$和$\boldsymbol{u}_{t(k)}$分别为历元$t(k+1)$及$t(k)$由卫星指向接收机的单位向量。考虑到卫星的速度远大于车辆运动速度, $\boldsymbol{u}_{t(k+1)}$与$\boldsymbol{u}_{t(k)}$可以近似相等, 因此, 根据几何关系可得

$$\begin{aligned}\Delta r = &\boldsymbol{u}_{t(k+1)}\cdot\left(\boldsymbol{X}^\mathrm{S}_{t(k+1)} - \boldsymbol{X}^\mathrm{R}_{t(k+1)}\right) - \boldsymbol{u}_{t(k+1)}\cdot\left(\boldsymbol{X}^\mathrm{S}_{t(k)} - \boldsymbol{X}^\mathrm{R}_{t(k)}\right) = \\ &\boldsymbol{u}_{t(k+1)}\cdot\left(\Delta\boldsymbol{r}^\mathrm{S} - \Delta\boldsymbol{r}^\mathrm{R}\right)\end{aligned} \tag{3}$$

式中: $\boldsymbol{X}^\mathrm{S}_{t(k+1)}$和$\boldsymbol{X}^\mathrm{S}_{t(k)}$分别为历元$t(k+1)$及$t(k)$的卫星位置; $\boldsymbol{X}^\mathrm{R}_{t(k+1)}$和$\boldsymbol{X}^\mathrm{R}_{t(k)}$分别为对应的接收机位置; $\Delta\boldsymbol{r}^\mathrm{S}$和$\Delta\boldsymbol{r}^\mathrm{R}$分别为两历元间卫星和接收机在地心地固 (earth-centered earth-fixed, ECEF) 坐标系中的几何位置增量。因此, 可以得出历元间载波相位的差分表达式为

$$\Delta\Phi = \frac{1}{\lambda}\boldsymbol{u}_{t(k+1)}\cdot\left(\Delta\boldsymbol{r}^\mathrm{S} - \Delta\boldsymbol{r}^\mathrm{R}\right) + \frac{c}{\lambda}\Delta\mathrm{d}t_\mathrm{R} + \Delta\varepsilon_\Phi \tag{4}$$

采用文献 [6] 中提出的多普勒积分法对周跳进行探测和修复后, 若可见卫星颗数满足定位条件时, 可以通过最小二乘法解算$\Delta\boldsymbol{r}^\mathrm{R}$。进一步推算可得到当前历元的速度$\boldsymbol{v} = \Delta\boldsymbol{r}^\mathrm{R}/\Delta t$, 其中, Δt为采样间隔。

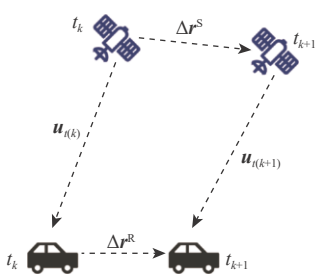


图 2　历元间卫星与车辆的相对几何关系

Fig. 2　Relative satellite vehicle geometry between two epochs

2.2　基于双天线位置增量的航向角估计

如图 3 所示, 车载双天线沿车身方向纵向放置, 天线 A 与天线 B 构成基线向量$\boldsymbol{L}$, 基线长度为l。历元$t(k)$的基线向量为$\boldsymbol{L}_{t(k)}$, 经车辆运动后历元$t(k+1)$的基线向量为$\boldsymbol{L}_{t(k+1)}$, 天线 A 与天线 B 在历元

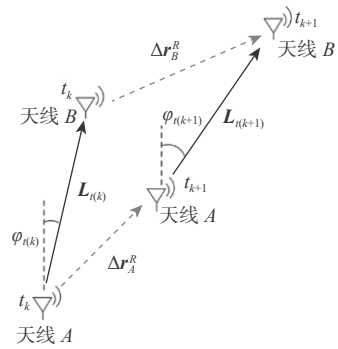


图 3　历元间双天线相对几何关系

Fig. 3　Relative geometry of two antennas between two epochs

$t(k)$与$t(k+1)$间的位移由 ECEF 坐标系转到北东地 (north east down, NED) 坐标系后分别为$\Delta\boldsymbol{r}_A^\mathrm{R}$和$\Delta\boldsymbol{r}_B^\mathrm{R}$。因此, 基线向量$\boldsymbol{L}_{t(k+1)}$可表示为

$$\boldsymbol{L}_{t(k+1)} = \boldsymbol{L}_{t(k)} + \Delta\boldsymbol{r}_B^\mathrm{R} - \Delta\boldsymbol{r}_A^\mathrm{R} \tag{5}$$

$\boldsymbol{L}_{t(k)}$可由历元$t(k)$的车辆姿态求得:

$$\boldsymbol{L}_{t(k)} = \begin{bmatrix} l\cos\theta_{t(k)}\cos\varphi_{t(k)} \\ l\cos\theta_{t(k)}\sin\varphi_{t(k)} \\ -l\sin\theta_{t(k)} \end{bmatrix} \tag{6}$$

式中: $\theta_{t(k)}$为历元$t(k)$的车辆俯仰角; $\varphi_{t(k)}$为历元$t(k)$的车辆航向角。由历元$t(k+1)$的基线向量$\boldsymbol{L}_{t(k+1)} = \left[L^\mathrm{N}_{t(k+1)} \quad L^\mathrm{E}_{t(k+1)} \quad L^\mathrm{D}_{t(k+1)}\right]^\mathrm{T}$可求出当前航向角$\varphi_{t(k+1)}$为

$$\varphi_{t(k+1)} = \begin{cases} \arctan\left(L^\mathrm{E}_{t(k+1)}/L^\mathrm{N}_{t(k+1)}\right) & L^\mathrm{N}_{t(k+1)} \geqslant 0 \\ \arctan\left(L^\mathrm{E}_{t(k+1)}/L^\mathrm{N}_{t(k+1)}\right) + \pi & L^\mathrm{N}_{t(k+1)} < 0 \text{且} L^\mathrm{E}_{t(k+1)} \geqslant 0 \\ \arctan\left(L^\mathrm{E}_{t(k+1)}/L^\mathrm{N}_{t(k+1)}\right) - \pi & L^\mathrm{N}_{t(k+1)} < 0 \text{且} L^\mathrm{E}_{t(k+1)} < 0 \end{cases} \tag{7}$$

由于车辆在道路上行驶的物理性能限制, 其俯仰角与横滚角通常接近于 0, 因此, 本节仅对车辆航向角进行讨论。基于本节算法进行航向角求解避免了传统双天线定姿算法的模糊度求解过程, 简化了算法复杂度。

2.3　基于 Hatch 滤波的位置估计

通常在进行 GNSS 定位解算时, 伪距的精度不如载波相位高, 但伪距观测可以避免载波相位观测时的模糊度解算过程, 更适应于车辆实时导航的快速单点定位需求。因此, 在接收机同时进行伪距观测与载波相位观测的条件下, 通过单频载波相位观测值对伪距观测值进行滤波平滑, 可以结合 2 种观测值的优势, 基于观测域滤波实现在不依赖外部数据信息的前提下削弱观测噪声对伪距观测值的影响, 提高伪距观测值精度, 从而实现更高精度的 GNSS 初始位置估计[10-15]。但由于电离层延迟变化的影响, 长时间平滑会带来累计偏差, 导致定位结果发散, 因此, 本文采用了基于滑动窗的 Hatch 滤波, 超出滑动窗的历史数据会被丢弃, 以此消除窗

口内电离层变化趋势项的影响，降低电离层延迟积累误差。文献[16]给出了基于滑动窗的Hatch滤波的递推公式：

$$\begin{cases}\bar{\rho}_{t(k)}=\dfrac{1}{k}\rho_{t(k)}+\left(1-\dfrac{1}{k}\right)\left[\bar{\rho}_{t(k-1)}+\lambda\left(\phi_{t(k)}-\phi_{t(k-1)}\right)\right] & k\leqslant m\\ \bar{\rho}_{t(k)}=\bar{\rho}_{t(k-1)}+\lambda\left(\phi_{t(k)}-\phi_{t(k-1)}\right)+\\ \qquad\dfrac{1}{m}\left[\left(\rho_{t(k)}-\rho_{t(k-m)}\right)-\lambda\left(\phi_{t(k)}-\phi_{t(k-m)}\right)\right] & k>m\end{cases} \tag{8}$$

式中：m为滑动窗大小；$\bar{\rho}_{t(k)}$和$\bar{\rho}_{t(k-1)}$分别为历元$t(k)$和$t(k+1)$的平滑伪距；$\rho_{t(k)}$和$\rho_{t(k-m)}$分别为历元$t(k)$和$t(k-m)$的伪距观测值；$\phi_{t(k)}$和$\phi_{t(k-m)}$分别为历元$t(k)$和$t(k-m)$的载波相位观测值。

经Hatch滤波平滑后的平滑伪距$\bar{\rho}_{t(k)}$进一步削弱了观测噪声对伪距的影响，用平滑伪距代替原始伪距观测值基于LSM得出当前历元更精确的初始位置估计结果$\boldsymbol{P}=[P_{\mathrm{N}}\quad P_{\mathrm{E}}\quad P_{\mathrm{D}}]^{\mathrm{T}}$。

3 GNSS/IMU松组合滤波算法设计

3.1 运动模型约束

根据载体运动特性的限制，在输入滤波前采用非完整约束（non-holonomic constraint, NHC）及零速修正（zero velocity update, ZUPT）对GNSS初始状态估计结果进行限制。

非完整性约束作用在载体坐标系中，对车辆运动状态进行假设，当车辆在路面上平稳行驶时不会发生侧滑、漂移或跳跃，可认为车辆的位移仅跟随车辆前向进行运动，因此，车辆基于载体坐标系的侧向和垂向速度可以近似为0，由此可对GNSS解算的载体速度进行运动约束。历元$t(k)$根据GNSS观测数据所求得的车辆速度在ECEF坐标系中表示为$\boldsymbol{v}^{\mathrm{e}}_{t(k)}$，将其转换为载体坐标系下的速度为

$$\boldsymbol{v}^{\mathrm{b}}_{t(k)}=\begin{bmatrix}v^{\mathrm{bx}}_{t(k)}\\ v^{\mathrm{by}}_{t(k)}\\ v^{\mathrm{bz}}_{t(k)}\end{bmatrix}=\boldsymbol{C}^{\mathrm{b}}_{\mathrm{e}}\boldsymbol{v}^{\mathrm{e}}_{t(k)} \tag{9}$$

式中：$v^{\mathrm{bx}}_{t(k)}$、$v^{\mathrm{by}}_{t(k)}$和$v^{\mathrm{bz}}_{t(k)}$分别为历元$t(k)$车辆在载体坐标系中的侧向、前向及垂向速度；$\boldsymbol{C}^{\mathrm{b}}_{\mathrm{e}}$为地心地固坐标系到载体系的旋转矩阵。经非完整约束的车辆在载体系中的速度$\boldsymbol{v}^{\mathrm{b}}_{\mathrm{NHC},t(k)}$为

$$\boldsymbol{v}^{\mathrm{b}}_{\mathrm{NHC},t(k)}=\begin{bmatrix}0\\ v^{\mathrm{by}}_{t(k)}\\ 0\end{bmatrix} \tag{10}$$

将$\boldsymbol{v}^{\mathrm{b}}_{\mathrm{NHC},t(k)}$重新转换到地心地固坐标系，可得到经非完整约束后的载体在地心地固坐标系下的速度$\boldsymbol{v}^{\mathrm{e}}_{\mathrm{NHC},t(k)}$：

$$\boldsymbol{v}^{\mathrm{e}}_{\mathrm{NHC},t(k)}=\boldsymbol{C}^{\mathrm{e}}_{\mathrm{b}}\boldsymbol{v}^{\mathrm{b}}_{\mathrm{NHC},t(k)} \tag{11}$$

式中：$\boldsymbol{C}^{\mathrm{e}}_{\mathrm{b}}$为由载体系到地心地固坐标系的旋转矩阵。

零速修正可应用于车辆行驶过程中的短暂停车过程，将停车时载体的速度真实值为0作为观测值进行滤波输入，对系统导航误差进行修正和补偿。设定零速判定滑动窗长度为n，n根据经验进行取值，对窗口内的IMU数据分别进行加速度幅值检测、加速度滑动方差检测、角速度幅值检测及角速度滑动方差检测，低于相应阈值即可认定当前处于零速状态，此时的速度观测值可表示为

$$\boldsymbol{v}=\begin{bmatrix}0 & 0 & 0\end{bmatrix}^{\mathrm{T}} \tag{12}$$

3.2 滤波算法设计

本文采用松组合导航模型对GNSS/IMU进行滤波融合，输入卡尔曼滤波器的21维系统误差表示如下：

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{\delta}_r & \boldsymbol{\delta}_v & \boldsymbol{\delta}_\psi & \boldsymbol{b}_{\mathrm{a}} & \boldsymbol{b}_{\mathrm{g}} & \boldsymbol{s}_{\mathrm{a}} & \boldsymbol{s}_{\mathrm{g}}\end{bmatrix}^{\mathrm{T}}_{21\times1} \tag{13}$$

式中：$\boldsymbol{\delta}_r$、$\boldsymbol{\delta}_v$、$\boldsymbol{\delta}_\psi$分别为地心地固坐标系下3维位置、速度和姿态误差向量；$\boldsymbol{b}_{\mathrm{g}}$为陀螺仪3轴零偏；$\boldsymbol{b}_{\mathrm{a}}$为加速度计三轴零偏；$\boldsymbol{s}_{\mathrm{g}}$为陀螺仪3轴刻度误差；$\boldsymbol{s}_{\mathrm{a}}$为加速度计刻度误差。抗差自适应卡尔曼滤波的递推公式为

$$\boldsymbol{X}_{k,k-1}=\boldsymbol{\Phi}_{k,k-1}\boldsymbol{X}_{k-1} \tag{14}$$

$$\boldsymbol{\Sigma}_{k,k-1}=\boldsymbol{\Phi}_{k,k-1}\boldsymbol{\Sigma}_{k-1}\boldsymbol{\Phi}^{\mathrm{T}}_{k,k-1}+\boldsymbol{Q}_{k-1} \tag{15}$$

$$\boldsymbol{V}_k=\boldsymbol{Z}_k-\boldsymbol{H}_k\boldsymbol{X}_{k,k-1} \tag{16}$$

$$\boldsymbol{K}_k=\frac{1}{\alpha_k}\boldsymbol{\Sigma}_{k,k-1}\boldsymbol{H}^{\mathrm{T}}_k\left[\boldsymbol{H}_k\left(\frac{1}{\alpha_k}\boldsymbol{\Sigma}_{k,k-1}\right)\boldsymbol{H}^{\mathrm{T}}_k+\frac{\boldsymbol{R}_k}{\boldsymbol{\omega}_k}\right]^{-1} \tag{17}$$

$$\boldsymbol{X}_k=\boldsymbol{X}_{k-1}+\boldsymbol{K}_k\boldsymbol{V}_k \tag{18}$$

$$\boldsymbol{\Sigma}_k=[\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k]\boldsymbol{\Sigma}_{k,k-1}[\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k]^{\mathrm{T}}+\boldsymbol{K}_k\frac{\boldsymbol{R}_k}{\boldsymbol{\omega}_k}\boldsymbol{K}^{\mathrm{T}}_k \tag{19}$$

式中：$\boldsymbol{X}_{k,k-1}$为一步预测状态向量；$\boldsymbol{\Phi}_{k,k-1}$为状态转移矩阵；$\boldsymbol{\Sigma}_{k,k-1}$为一步预测状态协方差矩阵；$\boldsymbol{Q}_{k-1}$为运动模型过程误差的方差矩阵；$\boldsymbol{Z}_k$为观测向量。本文采用组合导航间接法，所输入的观测向量为GNSS的输出状态与IMU的推算状态之差，$\boldsymbol{H}_k$为观测矩阵，$\boldsymbol{V}_k$为预测残差，$\boldsymbol{K}_k$为增益矩阵，$\boldsymbol{R}_k$为观测噪声矩阵，$\boldsymbol{X}_k$为估计状态向量，$\boldsymbol{\Sigma}_k$为估计状态的协方差矩阵，α_k为自适应因子，$\boldsymbol{\omega}_k$为抗差因子矩阵。

传统算法中观测向量$\boldsymbol{Z}_k$通常为3维向量，即GNSS输出3维位置与IMU推算位置之差。为了充分利用GNSS观测信息，本文将$\boldsymbol{Z}_k$拓展为7维，即GNSS输出3维位置、3维速度及航向角与IMU推算之差。$\boldsymbol{Z}_k$表示如下：

$$\boldsymbol{Z}_k=\begin{bmatrix}\boldsymbol{P}_{\mathrm{IMU}}-\boldsymbol{P} & \boldsymbol{v}_{\mathrm{IMU}}-\boldsymbol{v} & \varphi_{\mathrm{IMU}}-\varphi\end{bmatrix}^{\mathrm{T}} \tag{20}$$

式中: $\boldsymbol{P}_{\mathrm{IMU}}$、$\boldsymbol{v}_{\mathrm{IMU}}$和$\varphi_{\mathrm{IMU}}$分别为 IMU 推算的三维位置、三维速度和航向角; $\boldsymbol{P}$、$\boldsymbol{v}$、φ分别为对应的经运动状态约束后的 GNSS 状态估计参数。

根据速度构造状态不符值统计量$\Delta\tilde{T}_k$, 从而构建自适应因子α_k, 计算式如下[17-20]:

$$\Delta\tilde{T}_k=\frac{\|\boldsymbol{v}-\boldsymbol{v}_{\mathrm{IMU}}\|}{\sqrt{\mathrm{tr}\left(\boldsymbol{\Sigma}_{\tilde{\boldsymbol{V}}_k}\right)}} \tag{21}$$

式中: $\tilde{\boldsymbol{V}}_k$为一步预测的速度误差协方差; $\mathrm{tr}(\cdot)$为矩阵的迹。自适应因子α_k由 IGGⅢ权函数得出[21]:

$$\alpha_k=\begin{cases}1 & \Delta\tilde{T}_k\leqslant c_0\\ \dfrac{c_0}{\Delta\tilde{T}_k}\left(\dfrac{c_1-\Delta\tilde{T}_k}{c_1-c_0}\right) & c_0<\Delta\tilde{T}_k\leqslant c_1\\ 0 & \Delta\tilde{T}_k>c_1\end{cases} \tag{22}$$

式中: c_0和c_1为 2 个特定的常数[22], 本文根据经验取值$c_0=4$, $c_1=8$。

抗差因子矩阵计算式如下:

$$\omega_k^j=\begin{cases}1 & \left|\tilde{V}_k^j\right|\leqslant d_j\\ d_j/\left|\tilde{V}_k^j\right| & \left|\tilde{V}_k^j\right|>d_j\end{cases} \tag{23}$$

式中: ω_k^j为$\boldsymbol{\omega}_k$的第j个对角线元素; d_j为常数向量$\boldsymbol{d}$的第j个元素, 取值范围通常为 1.0～1.5。$\tilde{V}_k^j$的计算式如下:

$$\tilde{V}_k^j=V_k^j/\sqrt{\Sigma_{V_k}^{jj}} \tag{24}$$

式中: V_k^j为$\boldsymbol{V}_k$的第j个元素; $\boldsymbol{\Sigma}_{V_k}$为残差向量$\boldsymbol{V}_k$的协方差矩阵; $\Sigma_{V_k}^{jj}$为$\boldsymbol{\Sigma}_{V_k}$的第$j$个对角线元素。

经卡尔曼滤波解算的 IMU 误差状态量用于反馈校正 IMU, 解算得出的运动误差状态向量用于修正惯性导航力学编排所解算得出的载体三维位置、速度及姿态, 输出最终运动状态参数。

4　实验分析

为了验证本文算法的性能, 在南京市区道路进行了车载实验, 参考行驶轨迹如图 4 所示。本次实验采用浙江中裕通信的 ZYACF-S806 天线搭配北斗星通 C520-AT 接收机采集 GNSS 数据, 其中, 包括 GPS 和 BDS 这 2 个星座数据, 采集时长为 877 s, 采样频率为 10 Hz; 使用 STIM300 IMU 采集惯导数据。参考数据由霍尼韦尔的高精度组合导航系统 HGuide N580 进行采集, 后经商业软件 Inertial Explorer 进行后处理解算, 获取参考导航参数。实验车及所搭载的实验平台如图 5 所示, 使用文献 [23] 中的方法进行安装误差的标定, 杆臂测量误差控制在厘米级, 安装角误差在 0.001°量级, 经修正后基本可忽略安装误差对实验结果的影响。

实验数据采集时长为 877 s, 可见卫星数量如图 6 所示, 部分历元存在 GNSS 信号丢失或可见卫星数量不满足定位条件的情况。由于本文算法基于 GNSS 观测值实现车辆状态估计, 在不满足 GNSS 定位条件时组合导航系统仅由 IMU 进行推算定位, 无法通过算法实现精度的提升, 因此, 本文仅对满足定位条件的历元进行统计。

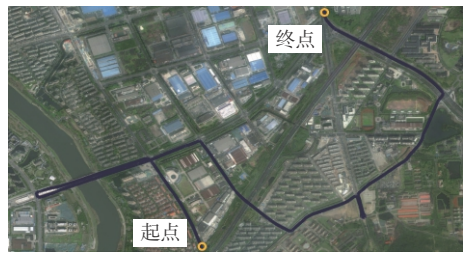
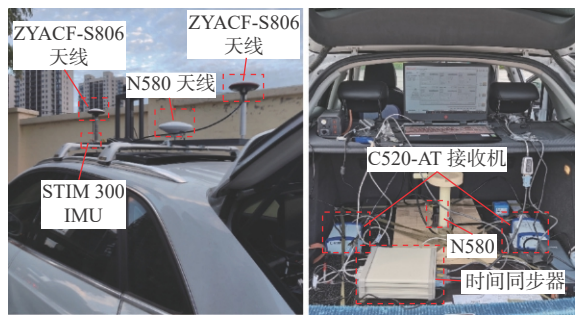


图 4　车载实验参考行驶轨迹

Fig. 4　Reference driving track of onboard experiment



(a) GNSS 天线及惯导设备　　(b) 数据接收平台

图 5　实验车及搭载的实验设备

Fig. 5　The experimental vehicle and the experimental equipment onboard

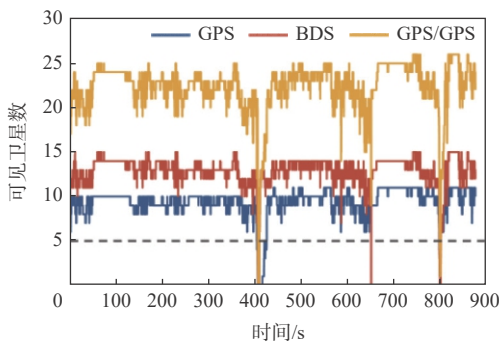


图 6　可见卫星数

Fig. 6　Number of visible satellites

图 7 和图 8 分别为组合导航算法的定位和测速误差图, 图中对比了传统 GNSS/IMU 松组合导航算法与本文算法的定位、测速性能。由图可知, 本文算法的定位及测速精度都比传统的 GNSS/IMU 松组合导航更高, 误差峰值也得到抑制。表 1 和表 2 分别统计了传统 GNSS/IMU 松组合算法和本文算法的定位、测速的均方根误差(root mean square error, RMSE)。由表可知, 本文算法的北向、东向和地向的定位精度相比传统 GNSS/IMU 松组合导航算法分别提高了 8.23%、38.73% 和 27.95%, 水平方

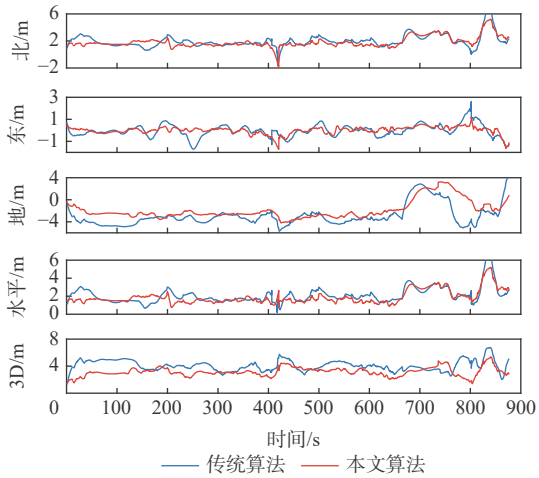


图7 算法定位误差对比

Fig. 7 Position errors comparison of the algorithms

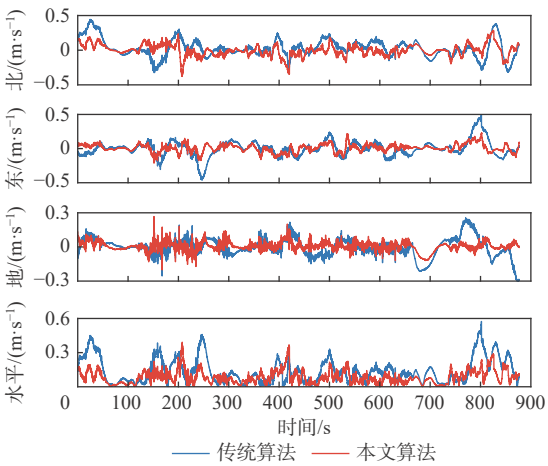


图8 算法测速误差对比

Fig. 8 Velocity errors comparison of the algorithms

向的提升率为9.78%，3维方向的总提升率为22.12%；测速精度在各方向上分别提升了36.19%、48.66%、52.11%，水平方向提升了41.27%，本文算法实现了定位、测速精度地有效提升。

图9显示了组合导航算法的定姿误差，表3统计了传统GNSS/IMU松组合导航算法与本文算法的姿态RMSE，对传统GNSS/IMU松组合导航算法与本文算法的车辆滚转角、俯仰角及航向角估计精度进行了对比。由图9和表3可知，本文算法所估计的车辆滚转角和俯仰角较传统算法分别提升了8.48%和6.90%，说明本文算法能够一定程度上提升车辆的滚转角和俯仰角估计精度。但由于车辆本身性能限制及实际道路的物理性质，通常情况下其滚转角和俯仰角基本不变且保持在0°左右，因此，算法对于滚转角及俯仰角精度的提升有限。

图10显示了车辆直线行驶及转弯时航向角估计结果，场景1为车辆航向角保持在64°～72°范围内的直线行驶，由图可知，场景1直线行驶过程中

表1 算法定位误差均方根对比

Table 1 Position RMSE comparison of the algorithms

对比算法	北向RMSE/m	东向RMSE/m	地向RMSE/m	水平RMSE/m	3D RMSE/m
传统算法	2.182 3	0.554 4	3.467 8	2.251 6	4.134 6
本文算法	2.002 7	0.339 7	2.498 4	2.031 3	3.220 0

表2 算法测速误差均方根对比

Table 2 Velocity RMSE comparison of the algorithms

对比算法	北向RMSE/m	东向RMSE/m	地向RMSE/m	水平RMSE/m
传统算法	0.131 8	0.115 3	0.080 4	0.175 1
本文算法	0.084 1	0.059 2	0.038 5	0.102 8

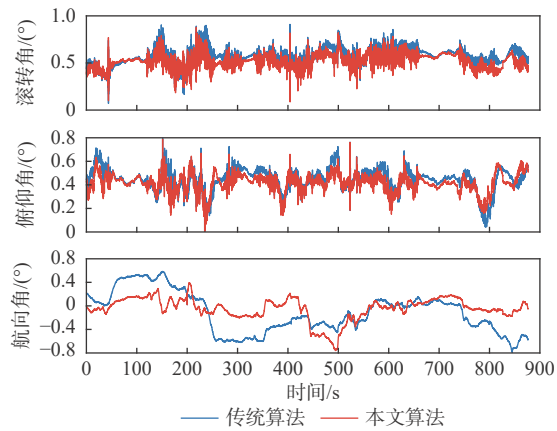


图9 算法定姿误差

Fig. 9 Attitude errors comparison of the algorithms

表3 算法定姿误差均方根对比

Table 3 Attitude RMSE comparison of the algorithms

对比算法	滚转RMSE/(°)	俯仰RMSE/(°)	航向RMSE/(°)
传统算法	0.574 0	0.453 8	0.360 3
本文算法	0.525 3	0.422 5	0.193 5

本文算法的航向角曲线更加贴近于参考值，表明本文算法能够在航向角变化不大的匀状态运动中保持更高精度的航向角估计；场景2车辆航向角大幅变化时本文算法的航向角基本与参考值相同，而传统GNSS/IMU松组合导航算法的航向角与参考曲线存在一定的偏移，表明本文算法在车辆航向角变化的高动态状态下能够实现更高精度的实时航向角估计。表3中显示通过本文算法航向角RMSE由0.360 3降低到0.193 5，精度提升率为46.29%，表明本文算法能够实现车辆航向角估计的精度大幅提升。这是由于传统算法中车辆姿态仅由IMU获取，而本文算法基于双天线TDCP可以连续获得可靠的航向角观测值，在组合导航系统中作为由GNSS提供的额外观测量与IMU观测值进行组合，进一步提高了测速的精度及可靠性。

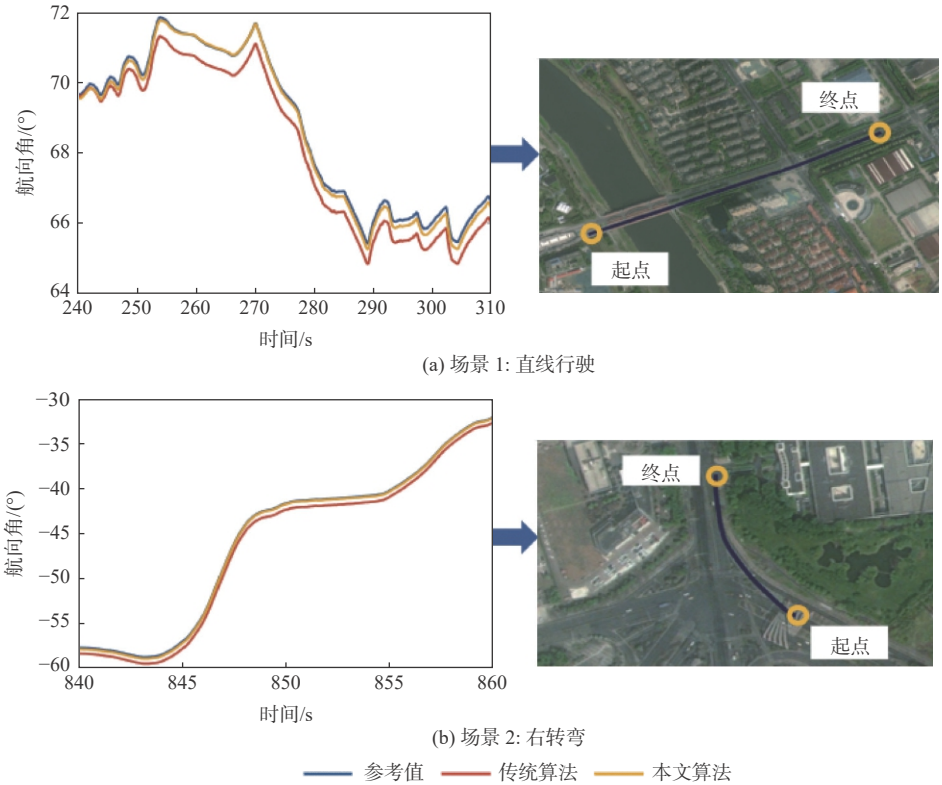


图 10　不同行驶状态下的车辆航向角估计结果

Fig. 10　Vehicle heading estimation results under different driving state

5　结　论

1）本文使用的基于双天线 TDCP 的姿态解算方法能够实现车辆姿态的快速解算。由于 TDCP 算法避免了整周模糊度的求解和固定过程，该方法能够利用单频载波观测值实现车辆姿态的测定。

2）在 GNSS/IMU 组合导航中引入双滤波可提升组合导航系统的状态估计精度。利用观测域的 Hatch 滤波可以通过载波观测值辅助伪距定位解算，提高 GNSS 初始状态估计精度；利用位置域的抗差自适应卡尔曼滤波可以合理调整 GNSS 观测值及 IMU 运动学推算值对定位结果的贡献权重。

3）实测实验结果表明，相较于传统 GNSS/IMU 松组合导航算法，利用基于双天线 TDCP 的 GNSS/IMU 双滤波组合导航算法的航向角精度可提升 46.29%，3 维定位精度提升了 22.12%，水平测速精度提升了 41.27%，说明本文算法可实现车辆航向角估计性能的增强，同时在定位、测速性能方面也有提升。

参考文献（References）

[1] FREDA P, ANGRISANO A, GAGLIONE S, et al. Time-differenced carrier phases technique for precise GNSS velocity estimation[J]. GPS Solutions, 2015, 19(2): 335-341.

[2] JI L, SUN R, CHENG Q, et al. Evaluation of the performance of GNSS-based velocity estimation algorithms[J]. Satellite Navigation, 2022, 3(1): 18.

[3] SUN R, CHENG Q, WANG J H. Precise vehicle dynamic heading and pitch angle estimation using time-differenced measurements from a single GNSS antenna[J]. GPS Solutions, 2020, 24(3): 84.

[4] KIM J, PARK M, BAE Y, et al. A low-cost, high-precision vehicle navigation system for deep urban multipath environment using TDCP measurements[J]. Sensors, 2020, 20(11): 3254.

[5] CHEN K, CHANG G, CHEN C, et al. An improved TDCP-GNSS/INS integration scheme considering small cycle slip for low-cost land vehicular applications[J]. Measurement Science and Technology, 2021, 32(5): 055006.

[6] JIAN W L, DONG Y, WANG D J, et al. A modified TDCP/INS tightly coupled navigation[C]//Proceedings of the China Satellite Navigation Conference. Berlin: Springer, 2019: 503-512.

[7] XU G C. GPS: theory, algorithms and applications[M]. Berlin: Springer, 2007.

[8] 甘雨, 隋立芬, 张清华, 等. 利用时间差分载波相位的 GNSS/INS 紧组合导航[J]. 测绘通报, 2015(6): 5-8.
GAN Y, SUI L F, ZHANG Q H, et al. GNSS/INS integrated navigation using time-differenced carrier phase measurements[J]. Bulletin of Surveying and Mapping, 2015(6): 5-8(in Chinese).

[9] QIAN N J, CHANG G B, GAO J X. GNSS pseudorange and time-differenced carrier phase measurements least-squares fusion algorithm and steady performance theoretical analysis[J]. Electronics Letters, 2019, 55(23): 1238-1241.

[10] LI F C, GAO J X, ZHENG N S, et al. A novel dual-domain filtering method to improve GNSS performance based on a dynamic model constructed by TDCP[J]. IEEE Access, 2020, 8: 79716-79723.

[11] ZHANG Q H, CHEN Z S, RONG F J, et al. An improved Hatch fil-

ter and its application in kinematic positioning with single-frequency GPS[J]. Measurement, 2019, 146: 868-878.

[12] LEE H K, RIZOS C. Position-domain Hatch filter for kinematic differential GPS/GNSS[J]. IEEE Transactions on Aerospace and Electronic Systems, 2008, 44(1): 30-40.

[13] PARK B, LIM C, YUN Y, et al. Optimal divergence-free hatch filter for GNSS single-frequency measurement[J]. Sensors, 2017, 17(3): 448.

[14] 陈正生, 张清华, 李林阳, 等. 电离层延迟变化自模型化的载波相位平滑伪距算法[J]. 测绘学报, 2019, 48(9): 1107-1118.
CHEN Z S, ZHANG Q H, LI L Y, et al. An improved carrier phase smoothing pseudorange algorithm with self-modeling of ionospheric delay variation[J]. Acta Geodaetica et Cartographica Sinica, 2019, 48(9): 1107-1118(in Chinese).

[15] 范士杰, 孔祥元. 基于 Hatch 滤波的 GPS 伪距相位平滑及其在单点定位中的应用[J]. 勘察科学技术, 2007(4): 40-42.
FAN S J, KONG X Y. GPS carrier phase smoothing on code pseudorange based on the hatch filter and its application in point positioning[J]. Site Investigation Science and Technology, 2007(4): 40-42 (in Chinese).

[16] 陈正生, 张清华, 崔阳, 等. 单频载波相位移动开窗平滑伪距算法及精度分析[J]. 武汉大学学报 (信息科学版), 2020, 45(7): 964-973.
CHEN Z S, ZHANG Q H, CUI Y, et al. Single frequency carrier smoothing pseudorange algorithm and accuracy analysis based on moving window[J]. Geomatics and Information Science of Wuhan University, 2020, 45(7): 964-973(in Chinese).

[17] 杨元喜, 何海波, 徐天河. 论动态自适应滤波[J]. 测绘学报, 2001, 30(4): 293-298.
YANG Y X, HE H B, XU T H. Adaptive robust filtering for kinematic GPS positioning[J]. Acta Geodaetica et Cartographic Sinica, 2001, 30(4): 293-298(in Chinese).

[18] 杨元喜, 任夏, 许艳. 自适应抗差滤波理论及应用的主要进展[J]. 导航定位学报, 2013, 1(1): 9-15.
YANG Y X, REN X, XU Y. Main progress of adaptively robust filter with applications in navigation[J]. Journal of Navigation and Positioning, 2013, 1(1): 9-15(in Chinese).

[19] GAGLIONE S, ANGRISANO A, CROCETTO N. Robust Kalman filter applied to GNSS positioning in harsh environment[C]//Proceedings of the European Navigation Conference. Piscataway: IEEE Press, 2019: 1-6.

[20] HE W, LIAN B W, TANG C K. GNSS/INS integrated navigation system based on adaptive robust Kalman filter restraining outliers[C]//Proceedings of the IEEE/CIC International Conference on Communications in China-Workshops. Piscataway: IEEE Press, 2014: 32-35.

[21] KIM L, LEE Y, LEE H K. Kalman–Hatch dual-filter integrating global navigation satellite system/inertial navigation system/onboard diagnostics/altimeter for precise positioning in urban canyons[J]. IET Radar, Sonar & Navigation, 2022, 16(2): 379-397.

[22] 王坚, 刘超, 高井祥, 等. 基于抗差 EKF 的 GNSS/INS 紧组合算法研究[J]. 武汉大学学报 (信息科学版), 2011, 36(5): 596-600.
WANG J, LIU C, GAO J X, et al. GNSS/INS tightly coupled navigation model based on robust EKF[J]. Geomatics and Information Science of Wuhan University, 2011, 36(5): 596-600(in Chinese).

[23] CHEN Q J, ZHANG Q, NIU X J. Estimate the pitch and heading mounting angles of the IMU for land vehicular GNSS/INS integrated system[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(10): 6503-6515.

Heading enhancement algorithm of GNSS/IMU integrated navigation based on dual-antenna TDCP

JI Li[1, 2], SUN Rui[1, 2, *], WANG Yuanyuan[1, 2], DAI Yeying[1, 2]

(1. Key Laboratory of Maritime Intelligent Cyberspace Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
2. State Key Laboratory of Air Traffic Management System, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: In global navigation satellite system (GNSS) and inertial measurement units (IMU) integrated navigation systems, attitude estimation, especially the accurate heading estimation, is very important for the real-time monitoring of vehicle state. However, due to the divergence of IMU on the altitude channel, its errors will gradually accumulate if the IMU cannot be accurately constrained, so the estimation accuracy of the heading is insufficient in the vehicle application where the heading changes frequently. In order to solve the problem of poor attitude estimation accuracy in GNSS and IMU loosely coupled navigation, a heading enhancement algorithm of GNSS/IMU integrated navigation based on dual-antenna time-differenced carrier phase (TDCP) is proposed. The vehicle heading is solved by dual-antenna TDCP to increase the input dimension of observations in integrated navigation filtering, and the Hatch filter and robust adaptive filter are used to improve the pseudorange accuracy of the observation domain and the GNSS/IMU integrated navigation positioning and attitude estimation performance, respectively. The algorithm evaluation results based on the measured data show that compared with the traditional GNSS/IMU integrated navigation method, the proposed algorithm improves the positioning and velocity measurement accuracy by 22.12% and 41.27%, respectively, and the heading accuracy by 46.29%.

Keywords: GNSS/IMU integrated navigation; multiple antennas; TDCP; robust Kalman filtering; attitude accuracy

Received: 2023-07-04; **Accepted**: 2023-08-11; **Published Online**: 2023-09-18 14:27

URL: link.cnki.net/urlid/11.2625.V.20230915.1408.002

Foundation items: National Natural Science Foundation of China (42222401,42174025,41974033); Ministry of Industry and Information Technology Special Project of China (TC220A04A-79); Jiangsu Provincial Six Talent Peaks Project (KTHY-014); Natural Science Foundation of Jiangsu Province (BK20211569); The Fundamental Research Funds for the Central Universities (xcxjh20220726)

*** Corresponding author**. E-mail: rui.sun@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0467

基于 STF-Net 的信号调制波形识别方法

哈晖[1,2]，高翔[1,*]，姚秀娟[1]，付降寅[1]，李伟[3]，张晓燕[3]

(1. 中国科学院国家空间科学中心，北京 100190；2. 中国科学院大学，北京 100049；3. 国家无线电监测中心，北京 100037)

摘 要：信号调制波形识别是空间频谱认知领域的关键技术之一，是实现低轨卫星频谱资源监测与管控的重要手段。针对现阶段基于深度学习的调制波形识别方法存在的参数量多、计算复杂度高等问题，提出一种基于空时融合网络（STF-Net）的轻量级信号调制波形识别方法。将信号预处理为时域-频域形式的双通道数据，通过卷积神经网络（CNN）提取信号空间特征并减少特征冗余，进而利用长短时记忆网络（LSTM）提取时序信息，输出识别结果。实验结果表明：所提方法在信噪比大于0 dB时，调制波形的平均识别准确率达到91.79%；与同等方法相比，所提方法参数量降低了96%，效率提升了2.7倍。

关 键 词：调制波形识别；深度学习；STF-Net；CNN；LSTM
中图分类号：V47; TN927+.21
文献标志码：A **文章编号**：1001-5965(2025)09-3150-11

低轨卫星通信系统较同步轨道卫星通信系统研发建设成本低、传输时延短、路径损耗小，通过多颗卫星组成的星座可实现全球覆盖。天地一体通信网络趋势下，低轨卫星通信系统成为地面通信系统的重要补充，在支持移动通信、边远地区基本通信和高速率用户接入等方面发挥重要作用[1]。但随着低轨星座系统爆发式发展，有限的频谱资源不能满足世界各国旺盛需求。为实现频谱资源可重复高效利用，如何构建面向低轨卫星的频谱认知系统是目前的研究热点。调制波形识别是在完全未知或具备较少先验信息的条件下，判别信号的调制方式[2]，是构建低轨卫星频谱监测系统，开展频谱资源管控的关键技术之一。准确识别和分类调制波形有助于实现接收机自动配置、干扰信号有效规避、频谱资源高效分配。

经典调制识别方法分为基于假设检验的最大似然识别方法和基于特征提取的统计模式识别方法[3]。基于假设检验的最大似然识别方法计算复杂度高且需要大量信道参数作为先验知识才能正确识别信号的调制波形[4]；基于特征提取的统计模式识别方法是依据专家手动设计的特征设计分类器进行分类，常用特征包括小波域特征[5-6]、循环谱[7-8]和高阶统计量[9-10]等。然而，基于特征提取的方法通用性较差且耗费专业人员大量的时间和精力。

深度学习能够通过多层神经网络实现复杂的特征组合和非线性变换，提取高级别的特征表示，在计算机视觉[11]、语音识别[12]和自然语言处理[13]等众多领域取得良好的表现。近年来，利用深度学习方法解决信号调制波形识别问题的研究和应用也日益增多。基于深度学习的调制波形识别方法主要分为2类，一是将待识别信号转化成图像，如眼图[14]、星座图[15-16]、累计极坐标图[17-18]、时频图[19]等，继而利用深度学习网络对图像进行分类，但是在该类方法中，信号至图像的转换过程带来额外的计算量并且可能丢失原始信息。二是利用深度学习网络直接对原始信号进行识别，如文献[20]设计

收稿日期：2023-07-14；**录用日期**：2023-09-14；**网络出版时间**：2023-10-13 10:12
网络出版地址：link.cnki.net/urlid/11.2625.V.20231011.1343.003
基金项目：国家重点研发计划(2020YFB1807900)；北京市自然科学基金(L222003)
* **通信作者**. E-mail：gaoxiang@nssc.ac.cn

引用格式：哈晖，高翔，姚秀娟，等. 基于STF-Net的信号调制波形识别方法[J]. 北京航空航天大学学报，2025，51(9)：3150-3160.
HA H，GAO X，YAO X J，et al. Signal modulation waveform recognition method based on STF-Net[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3150-3160 (in Chinese).

一种卷积神经网络(convolutional neural network, CNN)模型, 该模型直接利用接收信号的同相/正交数据(in-phase-quadrature data, IQ_data)进行训练, 在信噪比(signal-to-noise ratio, SNR)大于 4 dB 时的最高识别准确率可达 87.4%。文献 [21] 引入了信号畸变校正模块, 并与 CNN 模型协同训练, 解决由频率偏移和相位偏移引起的识别结果不理想问题。文献 [22-23] 利用 CNN 和长短时记忆(long short-term memory, LSTM)网络结合的网络模型识别信号调制波形, 较仅利用信号空间相关性的 CNN 具有更高的分类精度。文献 [24] 将接收信号 IQ_data、幅度/相位数据(amplitude-phase data, AP_data)输入基于 CNN-LSTM 的双流结构网络, 完成信号调制波形识别任务, 当信噪比大于 4 dB 时平均识别率达到 91% 左右。文献 [25] 提出一种用于信号调制波形识别的新型复值并行卷积递归神经网络, 该模型直接将原始复值信号输入到复值 CNN 获取复值特征图, 然后将其转换为幅度和相位并输入至 LSTM 以获取识别结果, 最高识别准确率可达 87.98%。文献 [26] 提出基于自编码器的同相/正交数据增强算法, 并将其应用至文献 [27] 所提的基于多通道的时空识别模型中, 改进后模型的最高识别准确率可达 93.11%。

基于 CNN-LSTM 的网络模型在多种调制信号的识别方面表现出良好的效果, 但以增加计算复杂度为代价提高识别准确率, 这对计算机的计算和存储能力要求较高。本文针对基于深度学习的信号调制波形识别方法存在的识别准确率与计算复杂度之间的矛盾问题, 提出一种基于 STF-Net 的轻量级信号调制波形识别方法, 主要创新点如下:

1) 提出基于空时融合特征的信号调制波形识别模型, 优化信号空时特征提取机制。分别利用 CNN 和 LSTM 提取调制信号局部空间特征及全局时序信息, 提高模型对不同调制方式的特征表示能力及低信噪比下的泛化能力。

2) 设计基于逐点卷积的特征降维方法, 减小模型计算复杂度。通过调整 1×1 卷积核的数量降低特征维度, 减小特征冗余度, 从而降低模型参数量、提升算法时效性。

3) 与现阶段基于深度学习网络的方法相比, 本文方法在保持较高识别精度的同时, 有效地减少了模型参数、降低了计算量。

1　相关技术

1.1　CNN

CNN 是一种深度学习模型[28], 核心思想是通过多层卷积网络学习输入数据高层次、抽象特征, 并利用反向传播算法优化网络参数, 使网络能自动学习并提取输入数据蕴含的有效特征。每层卷积网络数据处理过程为: 卷积计算、非线性激活、池化操作。

二维卷积计算式为

$$\boldsymbol{S}(i,j)=(\boldsymbol{I}*\boldsymbol{K})(i,j)=\sum_{m}\sum_{n}\boldsymbol{I}(i-m,j-n)\boldsymbol{K}(m,n) \quad (1)$$

式中: $\boldsymbol{I}$和$\boldsymbol{S}$分别为输入数据和输出数据; i和j为输入数据索引; $\boldsymbol{K}$为卷积核; m和n分别为卷积核的宽和高。

二维卷积计算示意如图 1 所示。卷积操作通过滑动卷积核获取输入数据不同位置的局部特征, 通过参数共享机制, 同一个卷积核可以在输入数据的不同位置上提取相同的特征, 实现平移不变性。

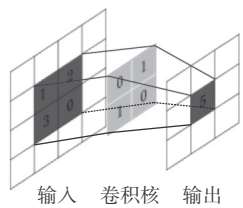


图 1　二维卷积示意图

Fig. 1　Illustration of 2D convolution

非线性激活函数是卷积神经网络的重要组成部分, 在前向计算及反向传播过程均发挥关键作用。在前向计算过程中, 非线性激活函数可以提升网络非线性表达能力; 在反向传播过程中, 非线性激活函数的可微性保证损失函数优化的可靠性。ReLU 函数是卷积神经网络常用的一种非线性激活函数, 定义式为

$$R_{\mathrm{eLU}}(x)=\begin{cases}x & x\geqslant 0\\ 0 & x<0\end{cases} \quad (2)$$

池化操作是指使用某位置相邻输出的总体统计特征作为该位置的输出, 在减小网络参数计算量的同时保证特征图平移不变性, 常用池化操作包括最大池化和均值池化, 示意图如图 2 所示。

深层卷积神经网络通过多次特征提取及非线性变换逐步提取输入数据多层次抽象特征: 低层次特征可以捕捉到边缘、纹理等局部特征; 高层次特征可以捕捉到更抽象的语义信息。但 CNN 的局部感知性、参数共享性、平移不变性等性质使其无法捕捉特征间的时序关系。假设 CNN 提取输入数据 $\boldsymbol{D}$的n个特征: $[f_1,f_2,\cdots,f_n]$, 依据:

$$\begin{cases}P(f_1)P(f_2)\cdots P(f_n)>\eta & \boldsymbol{D}\in C_1\\ P(f_1)P(f_2)\cdots P(f_n)\leqslant\eta & \boldsymbol{D}\in C_2\end{cases} \quad (3)$$

判断 $\boldsymbol{D}$所属类别: 若n个特征同时出现的概率大于门限值, 则 $\boldsymbol{D}\in C_1$, 反之 $\boldsymbol{D}\in C_2$。式 (3) 中, $P(f_i)$表示特征 f_i出现的概率, η表示判决门限。由

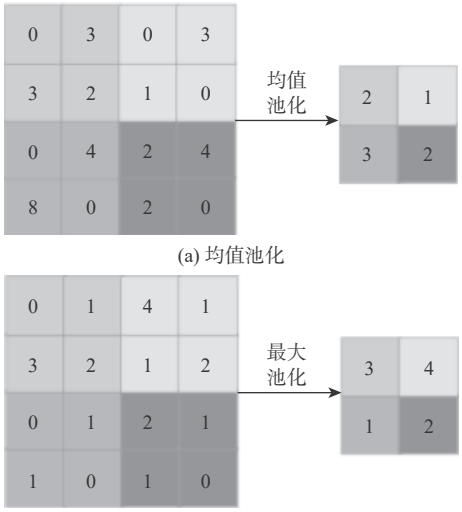


图2 池化操作示意图

Fig. 2 Illustration of pooling operation

式(3)可知,特征的排列顺序不会对判决结果产生影响,这表明CNN对特征间的时间相关性不敏感。由于调制信号在时间维度存在相关性,特征的不同时间排列可能表示不同调制波形,因此,基于CNN的调制波形识别模型的缺陷为损失调制信号的时序信息。

1.2 LSTM

LSTM是一种循环神经网络[29],核心思想是通过引入记忆单元和门控机制,有效地捕捉和管理输入序列中的长期依赖关系,解决传统循环神经网络在处理长序列数据时的梯度消失和梯度爆炸问题。LSTM具有3个门控机制,分别是输入门、遗忘门和输出门,内部结构如图3所示。

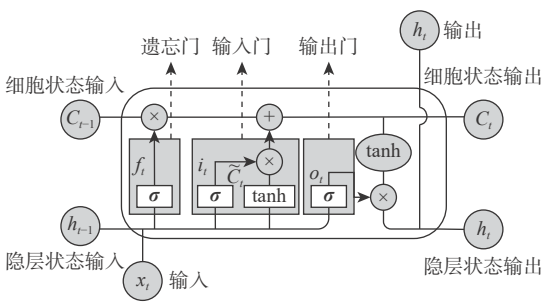


图3 LSTM基本结构

Fig. 3 Basic structure of LSTM

遗忘门控制细胞状态信息遗忘程度,输出当前状态的遗忘权重,计算式为

$$\boldsymbol{f}_t=\sigma(\boldsymbol{W}_{\mathrm{f}}\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_{\mathrm{f}}) \tag{4}$$

式中:$\boldsymbol{h}_{t-1}$为前一时刻隐藏层状态输入;$\boldsymbol{x}_t$为当前时刻输入数据;$\boldsymbol{W}_{\mathrm{f}}$为权重参数;$\boldsymbol{b}_{\mathrm{f}}$为偏置;$\sigma(*)$表示Sigmoid函数;$\boldsymbol{f}_t$为当前遗忘门控单元输出。

输入门控制细胞状态输入接收程度,输出当前输入信息的接收权重,计算式为

$$\begin{cases}\boldsymbol{i}_t=\sigma(\boldsymbol{W}_i\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_i)\\ \tilde{\boldsymbol{C}}_t=\tanh(\boldsymbol{W}_{\mathrm{c}}\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_{\mathrm{c}})\end{cases} \tag{5}$$

式中:tanh(*)为tanh函数;$\boldsymbol{i}_t$为输入门控状态;$\tilde{\boldsymbol{C}}_t$为将加入细胞状态的候选信息。细胞状态更新表达式为

$$\boldsymbol{C}_t=\boldsymbol{f}_t\cdot\boldsymbol{C}_{t-1}+\boldsymbol{i}_t\cdot\tilde{\boldsymbol{C}}_t \tag{6}$$

式中:C_t为更新的细胞状态;C_{t-1}为前一时刻细胞状态。

输出门控制细胞状态输出认可程度,输出当前输出信息的认可权重,计算式为

$$\begin{cases}\boldsymbol{o}_t=\sigma(\boldsymbol{W}_{\mathrm{o}}\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_{\mathrm{o}})\\ \boldsymbol{h}_t=\boldsymbol{o}_t\cdot\tanh(\boldsymbol{C}_t)\end{cases} \tag{7}$$

式中:$\boldsymbol{o}_t$为输出门控状态;$\boldsymbol{h}_t$为当前时刻隐藏层输出。

在信号调制波形识别任务中,调制信号前后样本之间存在相关性和依赖性,LSTM则可以通过记忆单元和门控机制获取调制信号的时序特征。但由式(4)~式(7)可知,LSTM无法获取调制信号的局部空间特征。此外,记忆单元和门控机制的引入导致LSTM计算复杂度和内存消耗增加。因此,基于LSTM的调制波形识别模型的缺陷为丢失调制信号局部特征且参数计算量大。

2 方法优化

在无线通信系统中,信号调制波形识别本质上是一个多分类的模式识别问题。本文从输入数据预处理机制、识别网络架构两方面优化现阶段调制波形识别方法。

2.1 数据预处理

接收信号$r(t)$一般可以表示为

$$r(t)=a(t)\mathrm{e}^{\mathrm{j}2\pi\Delta f_{\mathrm{c}}(t)+\Delta\varphi(t)}\int_0^{\tau_0}s(\Delta_{\mathrm{co}}(t-\tau))h(t,\tau)\mathrm{d}\tau+n(t) \tag{8}$$

式中:$s(t)$为发送信号;$a(t)$、$\Delta f_{\mathrm{c}}(t)$、$\Delta\varphi(t)$分别为发送信号的幅度变化、频率偏移和相位偏移;$\Delta_{\mathrm{co}}(t-\tau)$为发射机与接收机之间的定时偏移,最大的延迟时间为$\tau_0$;$h(t)$为信道函数;$n(t)$为均值0、方差$\sigma_n^2$的复加性高斯白噪声。

接收信号经采样后表示为:$\boldsymbol{r}=\boldsymbol{I}+\mathrm{j}\boldsymbol{Q}$,其中,$\boldsymbol{I}$和$\boldsymbol{Q}$分别表示离散接收信号$\boldsymbol{r}$的同相分量和正交分量。假设在一个采样周期内可以采集N个样本点,即$\boldsymbol{r}[n]\in\mathbf{C},n=0,1,\cdots,N-1$,那么第$k$个接收信号可以表示为

$$\boldsymbol{r}_k=[\boldsymbol{r}_k[0],\boldsymbol{r}_k[1],\cdots,\boldsymbol{r}_k[N-1]]^{\mathrm{T}} \tag{9}$$

调制识别的目标就是根据$\boldsymbol{r}_k$判断信号的调制波形。

本文考虑 3 种数据预处理方法: ①将接收信号表示为与原始复数样本相同的 IQ_data; ②通过获取信号幅度和相位信息将接收信号表示为 AP_data; ③使用快速傅里叶变换将接收信号表示为 FD_data。

1) IQ_data

IQ_data 中的 $\boldsymbol{I}$ 分量和 $\boldsymbol{Q}$ 分量分别是接收信号$\boldsymbol{r}_k$的同相分量和正交分量, 即

$$\begin{cases} \boldsymbol{x}_k^{\boldsymbol{I}} = [\mathrm{Re}(\boldsymbol{r}_k[0]), \mathrm{Re}(\boldsymbol{r}_k[1]), \cdots, \mathrm{Re}(\boldsymbol{r}_k[N-1])]^{\mathrm{T}} \\ \boldsymbol{x}_k^{\boldsymbol{Q}} = [\mathrm{Im}(\boldsymbol{r}_k[0]), \mathrm{Im}(\boldsymbol{r}_k[1]), \cdots, \mathrm{Im}(\boldsymbol{r}_k[N-1])]^{\mathrm{T}} \end{cases} \tag{10}$$

式中: Re(∗)表示复数的实部; Im(∗)表示复数的虚部。因此, 由第k个待识别信号$\boldsymbol{r}_k$经预处理得到的 IQ_data 为

$$\boldsymbol{r}_k \to \boldsymbol{x}_k^{\boldsymbol{IQ}} = \begin{bmatrix} (\boldsymbol{x}_k^{\boldsymbol{I}})^{\mathrm{T}} \\ (\boldsymbol{x}_k^{\boldsymbol{Q}})^{\mathrm{T}} \end{bmatrix} \tag{11}$$

2) AP_data

第k个接收信号的第n个样本点的幅度为

$$\boldsymbol{r}_k^A[n] = \sqrt{(\mathrm{Re}(\boldsymbol{r}_k[n]))^2 + (\mathrm{Im}(\boldsymbol{r}_k[n]))^2} \tag{12}$$

相位$\boldsymbol{r}_k^P[n]$为

$$\boldsymbol{r}_k^P[n] = \phi_k[n] + p_{r_k}[n] \tag{13}$$

式中: $\phi_k[n]$为相位测量值; $p_{r_k}[n]$为真实值与测量值之间的差值; n为不小于 2 的整数。$\phi_k[n]$、$p_{r_k}[n]$表达式分别为

$$\begin{cases} \phi_k[n] = \arctan\left(\dfrac{\mathrm{Im}(\boldsymbol{r}_k[n])}{\mathrm{Re}(\boldsymbol{r}_k[n])}\right) \\ p_{r_k}[n] = \begin{cases} p_{r_k}[n-1] - 2\pi & \phi_k[n] - \phi_k[n-1] > \pi \\ p_{r_k}[n-1] + 2\pi & \phi_k[n-1] - \phi_k[n] > \pi \\ p_{r_k}[n-1] & |\phi_k[n] - \phi_k[n-1]| < \pi \end{cases} \end{cases} \tag{14}$$

可得, 第k个接收信号$\boldsymbol{r}_k$经预处理得到的 AP_data 为

$$\begin{cases} \boldsymbol{r}_k \to \boldsymbol{x}_k^{AP} = \begin{bmatrix} (\boldsymbol{x}_k^A)^{\mathrm{T}} \\ (\boldsymbol{x}_k^P)^{\mathrm{T}} \end{bmatrix} \\ (\boldsymbol{x}_k^A)^{\mathrm{T}} = [\boldsymbol{r}_k^A[0], \boldsymbol{r}_k^A[1], \cdots, \boldsymbol{r}_k^A[N-1]] \\ (\boldsymbol{x}_k^P)^{\mathrm{T}} = [\boldsymbol{r}_k^P[0], \boldsymbol{r}_k^P[1], \cdots, \boldsymbol{r}_k^P[N-1]] \end{cases} \tag{15}$$

3) FD_data

频域表示形式是利用快速傅里叶变换实现原始时域复数据到其频域复数据的映射。第k个接收信号$\boldsymbol{r}_k$经预处理得到的频域表示形式为

$$\boldsymbol{r}_k \to \boldsymbol{x}_k^{FD} = \begin{bmatrix} \mathrm{Re}(F(r_k))^{\mathrm{T}} \\ \mathrm{Im}(F(r_k))^{\mathrm{T}} \end{bmatrix} \tag{16}$$

式中: $F(*)$表示信号的快速傅里叶变换。

2.2 STF-Net

2.2.1 网络模型设计思路

本文提出的基于 STF-Net 的调制波形识别网络通过将 CNN 与 LSTM 级联处理, 同时提取并利用接收信号空时特征。网络模型的泛化误差可分解为偏差、方差和噪声之和, 即

$$E(f, \boldsymbol{D}) = b_{\mathrm{ias}}^2(\boldsymbol{x}) + v_{\mathrm{ar}}(\boldsymbol{x}) + \varepsilon^2 \tag{17}$$

式中: $E(*)$为泛化误差; $b_{\mathrm{ias}}^2(*)$为偏差, 表示期望输出与真实标签的差别; $v_{\mathrm{ar}}(*)$为方差, 度量数据扰动对模型性能的影响; ε^2为噪声; f为网络模型; $\boldsymbol{D}$和$\boldsymbol{x}$分别表示训练数据集和测试样本。一般来讲, STF-Net 级联的卷积模块与 LSTM 模块数量越多, 网络模型f的表达能力越强, 偏差越低, 但方差也会随之升高。因此, 盲目增加 CNN 和 LSTM 的数量将导致模型泛化能力下降。此外, CNN 卷积层的参数量n_{c}为

$$n_{\mathrm{c}} = c_{\mathrm{i}} c_{\mathrm{o}} k \tag{18}$$

式中: c_{i}和c_{o}分别为输入和输出数据通道数; k表示卷积核的大小。LSTM 参数量n_{L}为

$$n_{\mathrm{L}} = 4[(d_{\mathrm{i}} + d_{\mathrm{h}}) d_{\mathrm{h}} + d_{\mathrm{h}}] \tag{19}$$

式中: d_{i}和d_{h}分别为输入维度和隐藏层维度。若假设$c_{\mathrm{i}} = c_{\mathrm{o}} = d_{\mathrm{i}} = d_{\mathrm{h}} = 50$、$k = 3$, 那么同时增加一个卷积模块和 LSTM 模块, 模型参数至少增加 27 700。

本文综合考虑网络模型性能与复杂度, 选取 2 个卷积模块和 2 个 LSTM 模块以提取信号空时融合特征。为解决由于数据维度过高而导致的 LSTM 模块参数量和计算量大幅度增加的问题, STF-Net 对卷积模块提取的不同尺度特征进行特征降维, 减少冗余特征。连接层拼接两降维模块输出, 融合不同尺度空间特征, 增强模型表达能力。特征融合数据依次输入至 LSTM 模块, 获取调制信号时序特征。最后, 数据经过全连接分类层输出待识别信号的调制方式。STF-Net 网络整体结构如图 4 所示, 各网络层输入数据、输出数据结构如表 1 所示。

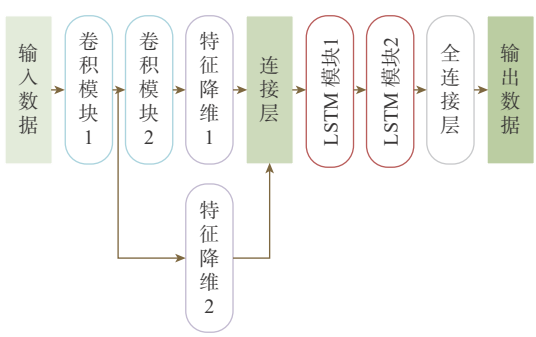


图 4　STF-Net 网络结构

Fig. 4　STF-Net network architecture

2.2.2 空间特征提取

输入数据经预处理后表示为

表1 STF-Net网络层数据结构

Table 1 Data structures of STF-Net network layers

网络层	输入数据结构	输出数据结构
卷积模块1	(1,2,128)	(N_1,1,130)
卷积模块2	(N_1,1,130)	(N_2,1,130)
特征降维1	(N_1,1,130)	(6,1,130)
特征降维2	(N_2,1,130)	(14,1,130)
连接层	(6,1,130), (14,1,130)	(20,130)
LSTM模块1	(20,130)	(50,130)
LSTM模块2	(50,130)	(20,130)
全连接层	20	11

$$\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_k,\boldsymbol{x}_{k+1},\cdots,\boldsymbol{x}_m\}^{\mathrm{T}} \tag{20}$$

式中：$\boldsymbol{x}_k$为第k个待识别信号三维向量表示；m为待识别信号数量。基于CNN的空间特征提取网络由2个卷积模块构成，每个卷积模块均包括卷积层、批量归一化层、Dropout层及非线性激活层，结构如图5所示。

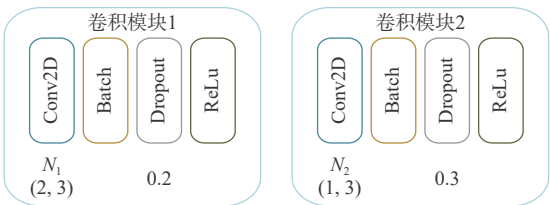


图5 卷积模块结构

Fig. 5 Convolutional module structure

在卷积模块1中，卷积核数量为N_1，大小为2×3，待识别信号$\boldsymbol{x}_k$经卷积层输出形状为(N_1,1,130)的特征图：$\boldsymbol{c}_k=f_{\mathrm{c}}(\boldsymbol{x}_k)$，$f_{\mathrm{c}}(*)$表示卷积函数，计算过程如式(1)所示。添加批量归一化层，减少训练时间并避免梯度爆炸或消失，输出：

$$\boldsymbol{bn}_k=f_{\mathrm{b}}(\boldsymbol{c}_k)=(\boldsymbol{c}_k-\mu)/\sqrt{\sigma^2} \tag{21}$$

式中：μ、σ^2分别为$\boldsymbol{c}_k$均值和方差。Dropout层以一定的概率将神经元输出置0，缓解模型过拟合问题。非线性激活层采用R_{eLU}函数作为非线性激活函数。待识别信号$\boldsymbol{x}_k$经卷积模块1输出：

$$\boldsymbol{o}_{\mathrm{c1}}=f_{\mathrm{C1}}(\boldsymbol{x}_k)=R_{\mathrm{eLU}}(f_{\mathrm{d}}(f_{\mathrm{b}}(f_{\mathrm{c}}(\boldsymbol{x}_k)))) \tag{22}$$

式中：$f_{\mathrm{C1}}(*)$表示卷积模块1的系统函数。

卷积模块1与卷积模块2的级联结构可以增加网络深度，提高网络特征表示能力。两卷积模块的结构相似，区别在于卷积模块2采用的卷积核数量为N_2，大小为1×3。卷积模块2的输入为卷积模块1的输出，输出为

$$\boldsymbol{o}_{\mathrm{c2}}=f_{\mathrm{C2}}(\boldsymbol{o}_{\mathrm{c1}})=R_{\mathrm{eLU}}(f_{\mathrm{d}}(f_{\mathrm{b}}(f_{\mathrm{c}}(\boldsymbol{o}_{\mathrm{c1}})))) \tag{23}$$

式中：f_{C2}表示卷积模块2的系统函数。

2.2.3 特征降维

LSTM参数计算量与其输入数据维度相关，数据维度越高，计算复杂度越大。直接将CNN模块提取的特征输入至LSTM模块，将导致模型参数量和计算量大幅度增加。为解决上述问题，本文设计基于逐点卷积的特征降维方法，结构如图6所示。

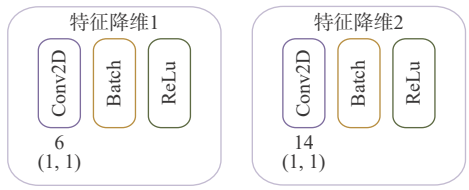


图6 特征降维结构

Fig. 6 Feature reduction architecture

卷积模块的级联结构使卷积模块2提取到更高级别的空间特征，本文为获取不同空间尺度下更抽象的特征表示，分别将逐点卷积层使用的卷积核数设置为6和14，卷积核大小均为1×1。特征降维1的输入为卷积模块1的输出，输出为

$$\boldsymbol{o}_{\mathrm{d1}}=f_{\mathrm{D1}}(\boldsymbol{o}_{\mathrm{c1}})=R_{\mathrm{eLU}}(f_{\mathrm{b}}(f_{\mathrm{c}}(\boldsymbol{o}_{\mathrm{c1}}))) \tag{24}$$

式中：f_{D1}为特征降维1的系统函数。特征降维2的输入为卷积模块2的输出，输出为

$$\boldsymbol{o}_{\mathrm{d2}}=f_{\mathrm{D2}}(\boldsymbol{o}_{\mathrm{c2}})=R_{\mathrm{eLU}}(f_{\mathrm{b}}(f_{\mathrm{c}}(\boldsymbol{o}_{\mathrm{c2}}))) \tag{25}$$

式中：f_{D2}为特征降维2的系统函数。$\boldsymbol{o}_{\mathrm{d1}}$和$\boldsymbol{o}_{\mathrm{d2}}$形状分别为(6,1,130)和(14,1,130)。

连接层拼接特征降维输出：$\boldsymbol{o}_{\mathrm{f}}=[\boldsymbol{o}_{\mathrm{d1}},\boldsymbol{o}_{\mathrm{d2}}]$，将不同尺度空间特征融合起来，从而提高模型泛化能力。连接层输出数据$\boldsymbol{o}_{\mathrm{f}}$形状为(20,1,130)，为符合LSTM网络对输入数据格式的要求，修改$\boldsymbol{o}_{\mathrm{f}}$形状为(20,130)。

2.2.4 时域特征提取

基于LSTM的时序特征提取模块由2个级联的LSTM模块构成，模块隐藏层数目分别为50、20。LSTM模块1输入数据为$\boldsymbol{o}_{\mathrm{f}}$，形状为(20,130)，其中，130表示时间序列观测长度，20表示每个时间步的特征维度。LSTM模块1的输出数据形状为(50,130)，该数据即为LSTM模块2的输入数据。保留LSTM模块2最后一个单元的输出，得到20×1的特征向量。该向量经过展平、全连接操作后得到模型的输出，具体可以表示为

$$\hat{\boldsymbol{y}}=[\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_i,\cdots,\hat{y}_K] \tag{26}$$

式中：K为调制池中调制波形的个数；$\hat{y}_i$为将信号的调制方式预测为第i种调制方式的概率。

3 实验验证

3.1 实验设置

3.1.1 实验数据集

实验使用公开数据集RadioML2016.10a[20]验证

本文方法的有效性。RadioML2016.10a 数据集由 GNU Radio 软件平台仿真生成, 具有 11 种不同的调制信号, 具体包括 8 种数字调制方式: BPSK、QPSK、8PSK、16QAM、64QAM、GFSK、CPFSK、PAM4, 3 种模拟调制方式: WB-FM、AM-SSB、AM-DSB。信噪比(信号功率与噪声功率的比值)范围为−20～18 dB, 增量为 2 dB。每种调制方式在每种信噪比下具有 1 000 个信号, 共有 220 000 组信号。数据集包含实际信道缺陷, 如频率偏移、采样率偏移和噪声等, 表 2 中列出了用于信号生成的详细参数。实验按照 8∶1∶1 的比例将数据集拆分成训练集、验证集和测试集。

表 2　数据集参数

Table 2　Dataset parameters

采样频率/kHz	采样频率标准差/Hz	最大采样频率偏移量/Hz	载波频率偏移标准差/Hz	最大载波频率偏移/Hz	频率选择性衰落中使用的正弦波个数	衰落中使用的最大多普勒频率/Hz	衰落模型	莱斯因子	功率时延谱的部分样本延迟	每个延迟时间对应的大小	插入功率时延谱的滤波器长度
200	0.01	50	0.01	500	8	1	Rician	4	[0, 0.9, 1.7]	[1, 0.8, 1.3]	8

实验在 Windows10 64 位操作系统上, 使用 Pytorch 深度学习框架完成网络的搭建、训练、验证和测试。网络训练数据的批处理大小为 64, 损失函数采用的是交叉熵损失函数, 使用 Adam 作为优化器, 学习率设置为 0.001, 其余参数为 Pytorch 中的默认参数。硬件设备配置部分: CPU 为 Intel Core i9-10900X; GPU 为 NVIDIA RTX 2080Ti; 内存为 64 GB。

3.1.2　模型结构最优化处理

为分析卷积模块中卷积核数目对调制波形识别性能的影响, 本文固定 STF-Net 其他结构, 选取不同数目的卷积核进行对比实验, 优化模型结构。实验设置卷积模块 1 与卷积模块 2 的卷积核数 N_1、N_2 均为 N, 采用平均识别准确率度量网络模型的泛化能力, 平均识别准确率为正确识别的调制信号个数与所有待识别信号个数的比值。

图 7 为输入单通道数据 IQ_data 时, 不同卷积核数目 N 在不同信噪比下的识别准确率的比较。由图可知: ①不同卷积核数的识别准确率均随信噪比的增加而增加, 在信噪比低于−6 dB 时, 识别准确率较低, 识别结果不具备参考意义; ②在信噪比大于 0 dB 时, 识别准确率趋于稳定, 并且 N=N_1=N_2=80 的识别准确率较高。表 3 列出了不同卷积核数在信噪比分别位于−20～0 dB、0～18 dB、−20～18 dB 下的平均识别准确率。由表可知, 在 N_1=N_2=80, 即卷积核数目和为 160 时, STF-Net 在 3 种信噪比范围内均取得最佳识别效果(见表 3 加黑数值)。

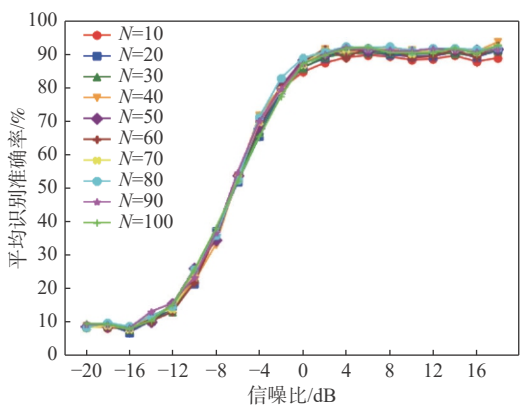


图 7　不同卷积核数目识别准确率比较

Fig. 7　Comparison of recognition accuracy with different numbers of convolutional kernels

表 3　不同卷积核数的平均识别准确率

Table 3　Average recognition accuracy for different numbers of convolutional kernels

卷积核数目N	平均识别准确率/%		
	[−20,0) dB	[0,18] dB	[−20,18] dB
10	30.99	88.03	59.51
20	30.35	89.36	59.85
30	30.80	89.45	60.12
40	31.19	90.78	60.98
50	31.31	90.51	60.91
60	30.07	89.96	60.52
70	31.30	90.62	60.96
80	**32.08**	**91.04**	**61.56**
90	31.90	90.43	61.16
100	31.02	90.35	60.68

在一定范围内, 增加模型参数量可以提高网络模型的泛化性。本文通过最大化卷积模块参数量来进一步调整卷积核数量、优化识别网络。输入单通道数据、双通道数据及三通道数据时, 卷积模块的参数量分别为

$$\begin{cases} Y_1 = 6N_1 + 3N_1N_2 \\ Y_2 = 12N_1 + 3N_1N_2 \\ Y_3 = 18N_1 + 3N_1N_2 \end{cases} \tag{27}$$

将 N_2=160-N_1 代入式 (27) 得

$$\begin{cases} Y_1 = -3N_1(N_1-162) \\ Y_2 = -3N_1(N_1-164) \\ Y_3 = -3N_1(N_1-166) \end{cases} \tag{28}$$

由此可得, 在 N_1+N_2=160 限制条件下, 当 N_1=81、N_2=79 时, 单通道输入数据对应的卷积模块参数量

最大；当 N_1=82、N_2=78 时，双通道输入数据对应的卷积模块参数量最大；当 N_1=83、N_2=77 时，三通道输入数据对应的卷积模块参数量最大。

表 4 列出了 N_1=N_2=80 及上述 3 种卷积核数在不同信噪比范围内的平均识别准确率。可得，当 N_1=83、N_2=77 时，模型识别效果最佳(见表 4 加黑数值)，因此，本文采用取得最佳识别效果 N_1=83、N_2=77 作为卷积模块 1 和卷积模块 2 的卷积核数。

表 4 不同数目卷积核的平均识别准确率

Table 4 Average recognition accuracy with different numbers of convolutional kernels

卷积核数目(N_1,N_2)	平均识别准确率/%		
	[−20,0) dB	[0,18] dB	[−20,18] dB
(80,80)	32.08	91.04	61.56
(81,79)	32.30	91.35	61.83
(82,78)	32.41	91.16	61.79
(83,77)	**32.65**	**91.35**	**61.95**

3.1.3 输入数据最优化处理

本文选取 IQ_data、AP_data、FD_data 这 3 种单通道输入数据，IQ-AP_data、IQ-FD_data、AP-FD_data 这 3 种双通道输入数据及 IQ-AP-FD_data 三通道输入数据进行对比实验，验证输入数据表示形式对识别结果的影响。

图 8 分别为单通道、双通道及三通道输入数据在不同信噪比下识别准确率的比较，可得：①采用不同输入数据表示形式获得的识别准确率均随信噪比的提高而增加，且当信噪比超过 4 dB 时，识别准确率趋于稳定；②相较于其他格式输入数据，FD_data 识别准确率最低；③在信噪比低于−8 dB 时，AP_data 的识别准确率要略高于 IQ_data 及 FD_data 的识别准确率；在信噪比高于−8 dB 时，IQ_data 的识别准确率要略高于 AP_data 的识别准确率。

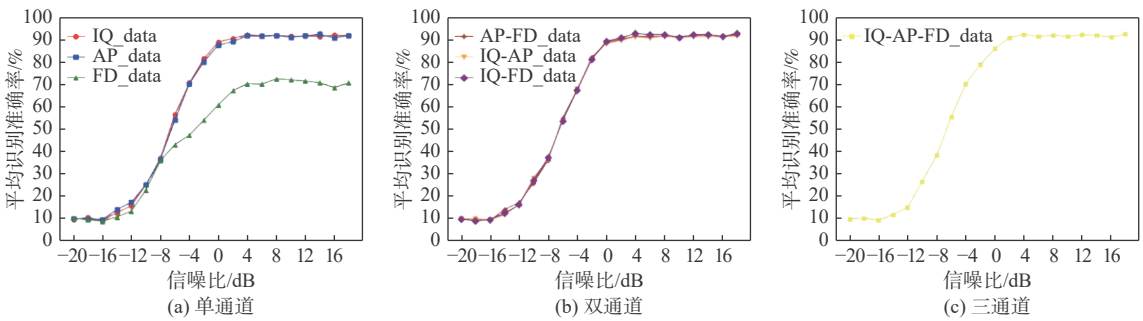


图 8 数据表示形式识别准确率比较

Fig. 8 Comparison of data format recognition accuracy

由图 8 可知，不同数据表示形式在极低信噪比下识别性能均非常差，识别结果不具备参考意义。因此，为更准确地了解数据表示形式与识别结果之间的关系，表 5 分别列出了信噪比在−20～0 dB、0～18 dB、−20～18 dB 范围内时，7 种输入数据的平均识别准确率。实验结果表明，IQ-FD_data 形式的双通道数据在信噪比位于 0～18 dB 和−20～18 dB 范围内的平均识别准确率均为最高，原因是 IQ_data 表示接收信号无损的时域信息，而 FD_data 数据表示接收信号的频域信息，2 种数据表示形式互相补充，丰富信号的有效特征。此外，由于 AP_data 与 IQ_data 均表示接收信号时域信息，因此 AP-FD_data 和 IQ-FD_data 的平均识别准确率相近：在信噪比为−20～18 dB 范围时，AP-FD_data 平均识别准确率比 IQ-FD_data 平均识别准确率仅低 0.08%；在信噪比为 0～18 dB 范围时，AP-FD_data 平均识别准确率比 IQ-FD_data 平均识别准确率低 0.53%。AP-FD_data 的平均识别准确率略低于 IQ-FD_data 的原因可能是接收数据至 AP_data 形式数据的转换过程丢失部分原始信息。综上，本文采用取得最佳识别效果的 IQ-FD_data 双通道数据作为识别模型的输入数据。

表 5 不同表示形式数据的平均识别准确率

Table 5 Average recognition accuracy of different data representations

输入数据形式	平均识别准确率/%		
	[−20,0) dB	[0,18] dB	[−20,18] dB
IQ_data	32.65	91.35	61.95
AP_data	32.46	90.89	61.68
FD_data	25.33	69.32	47.33
IQ-AP_data	32.57	90.79	61.68
IQ-FD_data	32.33	91.79	62.06
AP-FD_data	32.70	91.26	61.98
IQ-AP-FD_data	32.32	91.27	61.80

3.2 实验结果

实验选取文献 [22-26] 中的方法分别进行实现并与本文方法在识别准确率与算法复杂度方面进行比较。文献 [22-26] 提出的方法与本文方法均采用 CNN 和 LSTM 网络作为识别模型，但模型结构都各不相同。对比方法与本文方法的网络模型参数如表 6 所示。

3.2.1 算法准确率比对

图 9 显示了对比方法与本文方法在不同信噪

表 6　不同方法的网络结构

Table 6　Network structures of different methods

算法	CNN层数	卷积核尺寸	通道数	LSTM层数	隐藏单元数
文献[22]	3	(1,8),(1,8),(1,8)	50,50,50	1	50
文献[23]	2	(1,3),(2,3)	128,32	1	128
文献[24]	3	(1,3),(2,3)(1,3)	256,256,80	2	100,50
文献[25]	3	(1,5),(1,7)(1,9)	8,16,32	2	256,256
文献[26]	3	(2,8),(1,8),(2,5)	50,50,100	2	128,128
本文算法	2	(2,3),(1,3)	83,77	2	50,20

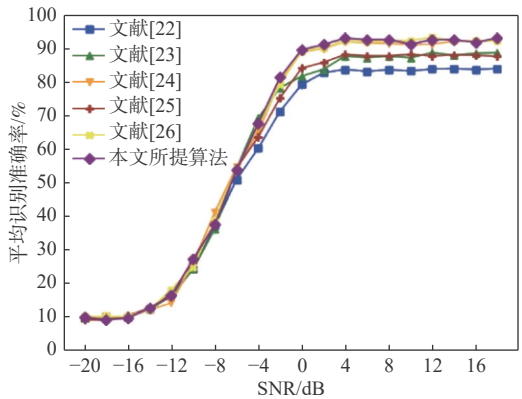


图 9　不同方法识别准确率比较

Fig. 9　Comparison of recognition accuracy among different methods

比下的识别准确率: ①对比方法与本文方法的识别准确率均随信噪比的增加而增加, 并且当信噪比大于 4 dB 时识别准确率趋于稳定。原因是随着信噪比的提升, 接收数据中有用信息占比增加, 模型可以更准确的学习特征并进行分类。②图中所有方法在信噪比低于−6 dB 时, 识别准确率均不足 50%, 识别性能较差。原因是随着信噪比的降低, 接收信号含有的有效信息减少, 模型难以学习到有效特征并进行准确分类。③在信噪比为−4～18 dB 范围时, 本文方法的识别准确率最高, 最高可到 92.82%; 在信噪比为−20～−6 dB 范围时, 对比方法与本文方法的平均识别准确率相近。

表 7 列出了对比方法及本文方法在信噪比为 0～18 dB 和−20～18 dB 范围内的平均识别准确率,

表 7　不同方法的平均识别准确率

Table 7　Average recognition accuracy of different methods

算法	平均识别准确率/%	
	[0,18] dB	[−20,18] dB
文献[22]	82.88	56.46
文献[23]	86.70	59.96
文献[24]	91.50	61.98
文献[25]	87.09	59.20
文献[26]	91.58	61.68
本文算法	**91.79**	**62.06**

由表可知, 本文方法的平均识别准确率与文献 [24]、文献 [26] 算法的平均识别准确率相近, 均高于文献 [22]、文献 [23]、文献 [25] 所提方法的平均识别准确率。

为更好地理解限制平均识别准确率的因素, 图 10 中分别显示了本文方法在信噪比为 6 dB、0 dB

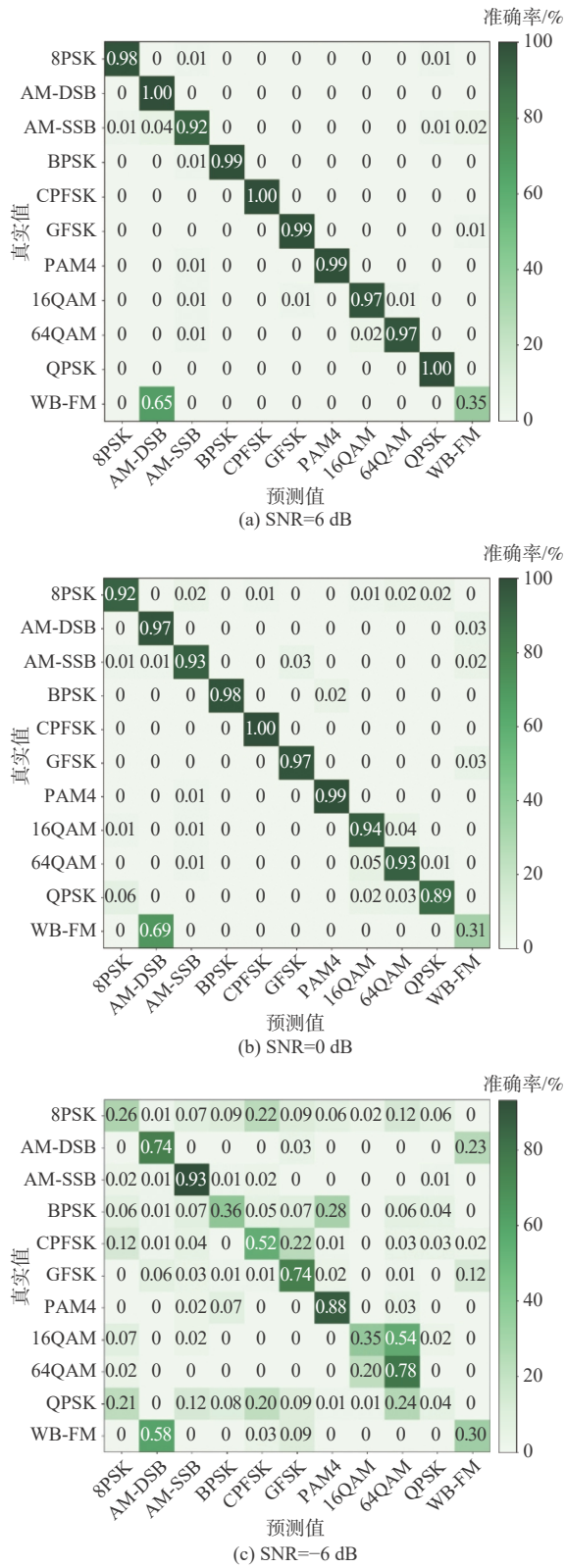


图 10　不同信噪比下的混淆矩阵

Fig. 10　Confusion matrix at different signal-to-noise ratios

和−6 dB下的混淆矩阵。由图10(a)和图10(b)可知,在信噪比为6 dB、0 dB的情况下,除WB-FM外的识别精度均处于较高水平。主要问题在于WB-FM被错误地识别为AM-DSB,原因可能是在模拟信号静默期中只有一个载波频率,模型无法正确区分这2种调制方式。如图10(c)所示,随着信噪比的降低,混淆矩阵对角线变得不清晰,混淆程度严重。但模型仍以较高的准确率识别出AM-SSB和PAM4。总体而言,在信噪比大于0 dB时,本文方法可以以较高的精度识别出大多数调制方式。

3.2.2 方法效率比对

表8对比了本文方法与对比方法的模型参数量与训练时间,其中,训练时间代表训练一轮全数据集花费的时间,文献[26]的网络参数和训练时间分别为数据增强网络与识别网络参数量及训练时间的总和。由表可知,文献[22-26]的模型参数量分别是本文模型参数量的1.69, 2.58, 28.17, 21.43, 14.34倍,所需训练时间分别是本文方法所需训练时间的4.59, 0.95, 2.76, 1.42, 1.06倍。

表8 不同方法的模型参数量及训练时间

Table 8 Model parameters and training time of different methods

方法	网络参数/个	训练时间/s
文献[22]	71 311	209.70
文献[23]	108 971	43.36
文献[24]	1 187 943	126.11
文献[25]	903 664	64.71
文献[26]	604 505	48.45
本文方法	42 165	45.71

综合表7和表8分析,性能方面,本文方法平均识别准确率最高,当信噪比在−20～18 dB范围时,本文方法平均识别准确率为62.06%,比文献[24]高0.08%;当信噪比在0～18 dB范围时,本文方法平均识别准确率为91.79%,比文献[24]高0.29%。方法效率方面,本文方法网络参数量最少,数目为42 165,但训练时间比文献[23]多2.35 s。综合方法性能与方法效率:本文方法识别准确率与文献[24]相近,但网络参数量降低约96%,训练时间缩短近2.7倍;本文方法训练时间虽比文献[23]多2.35 s,但在信噪比为0～18 dB范围内的平均识别准确率比文献[23]高5.09%。

4 结 论

针对现阶段基于深度学习的信号调制波形识别方法在实现高识别精度的同时也带来了高计算复杂度的问题,本文提出一种基于STF-Net的轻量化调制波形识别方法。采用RadioML2016.10a数据集对该方法进行验证,得到:

1)信号表示形式影响调制波形识别结果,采用IQ-FD_data形式的双通道数据作为识别模型的输入获得最佳识别效果,具体表现为:在−20～18 dB信噪比范围内,平均识别准确率为62.06%;在0～18 dB信噪比范围内,平均识别准确率为91.79%。

2)与现阶段基于深度学习的调制波形识别方法相比,本文方法不仅将0～18 dB信噪比下的平均识别准确率提升了0.29%,而且仅使用不到其4%的参数量。

综合方法性能与方法效率,本文方法在减小算法复杂度的同时,保证较高的识别准确率。

参考文献(References)

[1] 李高, 王威, 吴启晖. 面向低轨卫星的频谱认知智能管控[J]. 中兴通讯技术, 2021, 27(5): 7-11.
LI G, WANG W, WU Q H. Cognitive intelligent spectrum management and control for low earth orbit satellite system[J]. ZTE Technology Journal, 2021, 27(5): 7-11(in Chinese).

[2] 古相平. 浅谈通信信号调制样式自动识别方法[J]. 无线互联科技, 2011, 8(2): 22-23.
GU X P. Discussion on automatic identification method of communication signal modulation pattern[J]. Wireless Internet Technology, 2011, 8(2): 22-23(in Chinese).

[3] 张海燕, 闫文君, 张立民, 等. 通信信号调制识别综述[J]. 海军航空大学学报, 2022, 37(1): 126-132.
ZHANG H Y, YAN W J, ZHANG L M, et al. An overview of communication signal modulation recognition[J]. Journal of Naval Aviation University, 2022, 37(1): 126-132(in Chinese).

[4] 林冲, 闫文君, 张立民, 等. 通信信号调制识别综述[J]. 中国电子科学研究院学报, 2021, 16(11): 1074-1085.
LIN C, YAN W J, ZHANG L M, et al. An overview of communication signals modulation recognition[J]. Journal of China Academy of Electronics and Information Technology, 2021, 16(11): 1074-1085 (in Chinese).

[5] HASSAN K, DAYOUB I, HAMOUDA W, et al. Automatic modulation recognition using wavelet transform and neural network[C]// Proceedings of the 9th International Conference on Intelligent Transport Systems Telecommunications. Piscataway: IEEE Press, 2009: 234-238.

[6] 闫文康, 闫毅, 范亚楠, 等. 基于小波变换熵值及高阶累积量联合的卫星信号调制识别算法[J]. 空间科学学报, 2021, 41(6): 968-975.
YAN W K, YAN Y, FAN Y N, et al. A modulation recognition algorithm based on wavelet transform entropy and high-order cumulant for satellite signal modulation[J]. Chinese Journal of Space Science, 2021, 41(6): 968-975(in Chinese).

[7] 李世平, 陈方超, 王隆, 等. 多径信道下基于循环谱特征的调制识别算法[J]. 计算机应用, 2012, 32(8): 2123-2127.
LI S P, CHEN F C, WANG L, et al. Modulation identification algorithm based on cyclic spectrum characteristics in multipath channel[J]. Journal of Computer Applications, 2012, 32(8): 2123-2127

(in Chinese).

[8] DONG S L, LI Z P, ZHAO L F. A modulation recognition algorithm based on cyclic spectrum and SVM classification[C]// Proceeding of the IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference. Piscataway: IEEE Press, 2020: 2123-2127.

[9] FLOHBERGER M, GAPPMAIR W, KOUDELKA O. Modulation classifier for signals used in satellite communications[C]// Proceeding of the 5th Advanced Satellite Multimedia Systems Conference and the 11th Signal Processing for Space Communications Workshop. Piscataway: IEEE Press, 2010: 198-202.

[10] 李跃, 郭兴吉, 赵欣. 基于高阶累积量的调制方式识别研究[J]. 西南科技大学学报, 2018, 33(3): 64-68.
LI Y, GUO X J, ZHAO X. Study on modulation recognition based on higher-order cumulants[J]. Journal of Southwest University of Science and Technology, 2018, 33(3): 64-68(in Chinese).

[11] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[12] TOKOZUME Y, HARADA T. Learning environmental sounds with end-to-end convolutional neural network[C]// Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2017: 2721-2725.

[13] GAN C Q, WANG L, ZHANG Z F, et al. Sparse attention based separable dilated convolutional neural network for targeted sentiment analysis[J]. Knowledge-Based Systems, 2020, 188: 104827.

[14] WANG D S, ZHANG M, LI Z, et al. Modulation format recognition and OSNR estimation using CNN-based deep learning[J]. IEEE Photonics Technology Letters, 2017, 29(19): 1667-1670.

[15] PENG S L, JIANG H Y, WANG H X, et al. Modulation classification based on signal constellation diagrams and deep learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(3): 718-727.

[16] DOAN V S, HUYNH-THE T, HUA C H, et al. Learning constellation map with deep CNN for accurate modulation recognition[C]// Proceeding of the IEEE Global Communications Conference. Piscataway: IEEE Press, 2020: 1-6.

[17] TENG C F, LIAO C C, CHEN C H, et al. Polar feature based deep architectures for automatic modulation classification considering channel fading[C]// Proceeding of the IEEE Global Conference on Signal and Information Processing. Piscataway: IEEE Press, 2018: 554-558.

[18] TENG C F, CHOU C Y, CHEN C H, et al. Accumulated polar feature-based deep learning for efficient and lightweight automatic modulation classification with channel compensation mechanism[J]. IEEE Transactions on Vehicular Technology, 2020, 69(12): 15472-15485.

[19] KARRA K, KUZDEBA S, PETERSEN J. Modulation recognition using hierarchical deep neural networks[C]// Proceeding of the IEEE International Symposium on Dynamic Spectrum Access Networks. Piscataway: IEEE Press, 2017: 1-3.

[20] O'SHEA T J, CORGAN J, CLANCY T C. Convolutional radio modulation recognition networks[C]// Proceeding of the Engineering Applications of Neural Networks. Berlin: Springer, 2016: 213-226.

[21] YASHASHWI K, SETHI A, CHAPORKAR P. A learnable distortion correction module for modulation recognition[J]. IEEE Wireless Communications Letters, 2019, 8(1): 77-80.

[22] WEST N E, O'SHEA T. Deep architectures for modulation recognition[C]// Proceeding of the IEEE International Symposium on Dynamic Spectrum Access Networks. Piscataway: IEEE Press, 2017: 1-6.

[23] WU Y L, LI X J, FANG J. A deep learning approach for modulation recognition *via* exploiting temporal correlations[C]// Proceeding of the IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications. Piscataway: IEEE Press, 2018: 1-5.

[24] ZHANG Z F, LUO H, WANG C, et al. Automatic modulation classification using CNN-LSTM based dual-stream structure[J]. IEEE Transactions on Vehicular Technology, 2020, 69(11): 13521-13531.

[25] REN Y, JIANG W, LIU Y. Complex-valued parallel convolutional recurrent neural networks for automatic modulation classification [C]// Proceeding of the IEEE 25th International Conference on Computer Supported Cooperative Work in Design. Piscataway: IEEE Press, 2022: 804-809.

[26] ZHANG F X, LUO C B, XU J L, et al. An autoencoder-based I/Q channel interaction enhancement method for automatic modulation recognition[J]. IEEE Transactions on Vehicular Technology, 2023, 72(7): 9620-9625.

[27] XU J L, LUO C B, PARR G, et al. A spatiotemporal multi-channel learning framework for automatic modulation recognition[J]. IEEE Wireless Communications Letters, 2020, 9(10): 1629-1632.

[28] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541-551.

[29] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.

Signal modulation waveform recognition method based on STF-Net

HA Hui[1,2], GAO Xiang[1,*], YAO Xiujuan[1], FU Jiangyin[1], LI Wei[3], ZHANG Xiaoyan[3]

(1. National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China;
3. State Ratio Monitoring Center, Beijing 100037, China)

Abstract: Signal modulation waveform recognition is one of the key technologies in the field of spectrum cognition and an important means to achieve monitoring and control of spectrum resources for low-orbit satellites. To address the issues of high parameter count and computational complexity in existing deep learning-based modulation waveform recognition methods, a lightweight signal modulation waveform recognition method based on space-time fusion network (STF-Net) is proposed. The method first preprocesses the signals into dual-channel data in the time-frequency domain. It then utilizes convolutional neural network (CNN) to extract signal spatial features and reduce feature redundancy. Long short-term memory (LSTM) is employed to capture temporal information and output recognition results. Experimental results show that the proposed method achieves an average recognition accuracy of 91.79% for modulation waveforms when the signal-to-noise ratio is greater than 0dB. Compared with equivalent methods, the proposed method reduces the parameter count by 96% and improves efficiency by 2.7 times.

Keywords: modulation waveform recognition; deep learning; STF-Net; CNN; LSTM

Received: 2023-07-14; **Accepted**: 2023-09-14; **Published Online**: 2023-10-13 10:12

URL: link.cnki.net/urlid/11.2625.V.20231011.1343.003

Foundation items: National Key Research and Development Program of China (2020YFB1807900); Beijing Municipal Natural Science Foundation (L222003)

* **Corresponding author**. E-mail: gaoxiang@nssc.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0469

民机货舱氮气-细水雾协同灭火数值仿真研究

黄苏平[1]，朱培[1,2,*]，刘全义[3]，邵荃[1]

(1. 南京航空航天大学 民航学院，南京 211106；2. 中国民航大学民航热灾害防控与应急重点实验室，天津 300300；
3. 民机火灾科学与安全工程四川省重点实验室，广汉 618307)

摘　　要：货舱火灾是民机运输安全的重要威胁之一。考虑不同火源位置和环境压力的影响，基于火灾动力学模拟(FDS)代码开展了全尺度飞机货舱氮气-细水雾协同熄灭航空煤油池火有效性评估研究。结果表明：货舱火源位置影响灭火效果，角落火最难扑灭；以角落火为例，氮气灭火初期对火源难以有效抑制，而细水雾启动后可快速衰减火源热释放率；在氮气和细水雾单喷头流量分别为0.222 kg/s和15 L/min时，与单独氮气和细水雾灭火相比，协同灭火时冷却降温和窒息效应均显著增强，灭火效率分别提高46.20%和48.80%，其中氮气的加入可使水消耗量减少48.78%；当灭火参数相同时，灭火剂释放速率增加可同时降低灭火时间和灭火剂消耗量，实际中应采用大流量快速释放的喷头结构设计；环境压力降低可显著降低灭火时间，环境压力从101 kPa降低至46 kPa，协同灭火降温速率可提高约48.60%。

关 键 词：民机货舱；氮气；细水雾；协同灭火；环境压力
中图分类号：V244.1+2
文献标志码：A　　**文章编号**：1001-5965(2025)09-3161-13

民机货舱作为货物运输的重要载体，是飞机火灾事故的高发区域之一。飞机巡航阶段货舱环境压力一般维持在76 kPa左右，其中部分货舱无增压系统可降低至46 kPa左右，导致低压低氧环境的存在[1]；同时，货舱空间受限、燃烧物载荷大、防火控制薄弱等。因此，低压受限特殊环境给货舱火灾防治带来了严峻挑战。货舱灭火主要采用全淹没方式，保证货舱内部区域快速达到灭火体积分数是抑灭火的关键。传统货舱灭火剂哈龙1301因破坏臭氧层被限制生产和使用，到目前为止货舱哈龙替代难题尚未有效解决[2]。而细水雾和氮气由于其环境友好性，是理想的哈龙替代品。因此，探究货舱低压受限环境下氮气-细水雾协同灭火有效性具有重要现实意义。

对于低环境压力下火灾燃烧行为特性，贺元骅等[3]分别在海拔470 m和4 290 m下探究了航空电缆燃烧特性，发现随着环境压力降低，点燃温度和时间增加，而燃烧速率降低。Wang等[4]、刘全义等[5]探究了货舱正庚烷池火燃烧特性，结果表明环境压力降低导致火焰高度增加和燃烧速率降低，而火焰高度增加会进一步增大货舱顶棚内衬层等结构材料受热失稳破坏的风险。针对氮气、细水雾及其协同灭火性能，氮气灭火机理主要是氧稀释、冷却吸热等，如Dinesh等[6]分别通过杯式燃烧器和大尺度池火数值模拟测试，获得氮气灭火最小设计体积分数为0.34，而同等条件下要求重量为哈龙1301的7.5倍多；Hu等[7]探讨了灭火惰性气体体积分数与排放速率、通风量的关系，表明所需体积分数受火灾环境的影响。细水雾灭火机理主要是冷却吸热、氧稀释、热辐射衰减及动力学作用等，如Lee等[8]、

收稿日期：2023-07-17；**录用日期**：2023-09-22；**网络出版时间**：2023-10-23 15:26
网络出版地址：link.cnki.net/urlid/11.2625.V.20231023.1039.002
基金项目：国家自然科学基金(52202416)；中国民航大学民航热灾害防控与应急重点实验室开放基金(RZH2021-KF-02)
*通信作者. E-mail：zpei@nuaa.edu.cn

引用格式：黄苏平，朱培，刘全义，等. 民机货舱氮气-细水雾协同灭火数值仿真研究[J]. 北京航空航天大学学报，2025，51(9)：3161-3173. HUANG S P，ZHU P，LIU Q Y，et al. Numerical simulation study on synergetic fire suppression of nitrogen/water-mist in aircraft cargo compartment[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3161-3173 (in Chinese).

Wang等[9]研究了雾滴粒径、流量、雾锥角和喷头与火源相对位置对灭火效率的影响。Zhang等[10]探究了氮气-细水雾协同抑灭锂离子电池火灾性能，发现协同作用可有效防止复燃，且具有较好的冷却降温效果。

而环境压力对灭火介质的扩散雾化特性及其灭火效能均产生影响，现有研究的开展主要集中在常压环境与高海拔低压环境。Zhang等[11]探究了环境压力对气体射流扩散特性的影响，得出环境压力降低引起气体扩散速率和扩散角度增大。Wang等[12]、Zhu等[13]基于小尺度低压舱探究了环境压力对细水雾特性的影响，发现环境压力降低，喷雾中心区域雾滴粒径和轴向速度减小，而边缘处粒径和径向速度有所增加。对于低压环境灭火机理，戴经天[14]通过小尺寸低压舱室研究了环境压力对细水雾灭火性能的影响，发现环境压力降低，雾滴最佳粒径减小，火焰冷却与燃油表面冷却为主导灭火机理。Ma等[15]基于中尺度低压舱探究了氮气-水双流体细水雾熄灭正庚烷油池火的有效性，发现环境压力降低导致平均雾滴粒径和灭火时间减小。但上述研究结果均基于中小尺度模拟舱，难以有效分析和直接用于揭示实际全尺度货舱火灾的灭火性能及作用机制。

飞机货舱哈龙替代灭火系统必须通过美国联邦航空管理局(Federal Aviation Administration, FAA)制定的货舱灭火最低性能测试标准(minimum performance standard, MPS)[16]，而FAA-MPS灭火测试标准规定了4类火灾场景，包括散装火(纸箱火)、集装箱火、表面燃烧火(航空煤油池火)和气溶胶爆炸等。Xiong等[17]依据FAA-MPS标准，采用数值模拟方法评估了氮气抑灭第3类表面燃烧火的有效性，结果表明：与哈龙1301相比，氮气作为灭火介质，能够实现灭火后更低的火场温度；Marker等[18]、Reinhardt[19]开展了一系列全尺寸货舱细水雾灭火实验测试，结果表明：①单流体细水雾可有效长时间抑制A类和B类火灾，但耗水量较多；②氮气和细水雾混合灭火系统能够通过MPS灭火测试，且相对单独细水雾，耗水量可以减少50%以上；吕志豪[20]基于全尺度实验舱开展了空气-水双流体细水雾熄灭航空煤油池火的实验研究，得出细水雾在该场景下主导机理为火焰冷却。但上述研究受测量手段限制，尚未系统细致揭示氮气-细水雾协同灭火主导机制及微观演化机理，同时较少涉及高空巡航环境下货舱抑灭火性能。朱培等[21]基于全尺寸低压舱初步开展了环境压力影响下细水雾抑灭FAA-MPS标准表面燃烧火的有效性研究，发现随着环境压力降低，火焰熄灭时间明显缩短，但对环境压力、灭火系统参数、受限空间等耦合作用下协同灭火动力学特征及其灭火机制尚未深入探究。

综上，本文采用FDS数值模拟方法，选择FAA-MPS测试标准中第3类表面燃烧火灾场景，进一步系统探究高空巡航环境下氮气-细水雾协同熄灭货舱火灾的有效性及作用机理，评估特殊货舱环境下应用其协同灭火方式的可行性并获得相关灭火性能参数，为实际民机货舱哈龙替代灭火系统设计提供理论依据。

1 理论基础

FDS数值模型主要应用于低速、热浮力驱动流动问题模拟研究，在消防工程领域中广泛应用，相关燃烧、湍流、传热传质等理论模型参见FDS手册[22-23]。本文研究中主要理论模型设置如下：燃烧模型采用混合分数模型；火焰熄灭模型采用基于氧消耗方法，即通过估算维持燃烧所需的极限氧质量分数进行求解，而极限氧质量分数$Y_{O_{2,lim}}$由局部火焰温度和临界火焰温度决定，如下所示[23]：

$$Y_{O_{2,lim}} = Y_{OI}\left(\frac{T_{OI}-T}{T_{OI}-T_{\infty}}\right) \tag{1}$$

式中：Y_{OI}为氧指数，即火焰熄灭时氧气体积分数；T_{OI}为临界火焰温度，℃；T为局部火焰温度，℃；T_{∞}为环境温度，℃；当火焰区域网格单元内氧质量分数Y_{O_2}低于$Y_{O_{2,lim}}$时，即认为火焰熄灭。环境压力改变将影响火焰临界温度和局部氧体积分数。

细水雾模拟采用拉格朗日粒子喷雾流模型，雾滴粒径大小分布采用罗辛-拉姆勒-对数正态分布(Rosin-Rammler-Lognormal)分布模型；离散雾滴在连续气相中运动方程如下[23]：

$$\frac{\mathrm{d}(m_{\mathrm{d}}u_{\mathrm{d}})}{\mathrm{d}t} = m_{\mathrm{d}}g - \frac{1}{2}\rho C_{\mathrm{D}}\pi r_{\mathrm{d}}^2(u_{\mathrm{d}}-u)|u_{\mathrm{d}}-u| \tag{2}$$

式中：m_{d}为液滴质量，kg；ρ为气体密度，kg/m^3；C_{D}为水滴在气体中运动的阻力系数；r_{d}为液滴半径，m；u_{d}为液滴运动速度，m/s；u为气流速度，m/s。环境压力降低会导致ρ和C_{D}减小，同时由于真空蒸发效应，r_{d}也会降低，而u_{d}有所增加[7]，进而影响雾滴动力学性能。

液滴与气相间的质量和能量传输决定其吸热降温性能，液滴蒸发速率由液相平衡蒸汽质量分数、局部气相蒸汽质量分数、液体温度和局部气相温度共同决定，本文采用默认传热传质模型[23]，质量和温度传输方程如下所示：

$$\frac{\mathrm{d}m_{\mathrm{d}}}{\mathrm{d}t} = -A_{\mathrm{d}}h_{\mathrm{m}}\rho_{\mathrm{g}}\left(Y_{\mathrm{d}}-Y_{\mathrm{g}}\right) \tag{3}$$

$$m_{\mathrm{d}}c_{\mathrm{p}}\frac{\mathrm{d}T_{\mathrm{d}}}{\mathrm{d}t}=A_{\mathrm{d}}h_{\mathrm{d}}\left(T_{\mathrm{g}}-T_{\mathrm{d}}\right)\frac{\mathrm{d}m_{\mathrm{d}}}{\mathrm{d}t}h_{\mathrm{v}} \tag{4}$$

式中: m_{d}为液滴质量, kg; A_{d}为液滴表面积, m^2; h_{m}为传质系数; ρ_{g}为气体密度, kg/m^3; c_{p}为水比热容; h_{d}为液滴和气体之间的传热系数; h_{v}为液体蒸发的潜热; Y_{g}可从气相质量传输方程中获得; Y_{d}为液体平衡蒸汽质量分数。环境压力将影响雾滴粒径和气体密度, 即式 (3) 和式 (4) 中ρ_{g}、A_{d}和m_{d}等参数改变, 进而影响雾滴冷却降温效果。

在实际中雾滴不仅起到直接吸热降温作用, 同时降低可燃物热解速率, 进而影响燃烧速率。因此, 在模拟中需要对燃烧速率进行修正, 此时局部质量燃烧速率通过下式计算[23]:

$$\dot{m}''_{\mathrm{f}}(t)=\dot{m}''_{\mathrm{f,p}}(t)\cdot\exp\left[-\int k(t)\mathrm{d}t\right] \tag{5}$$

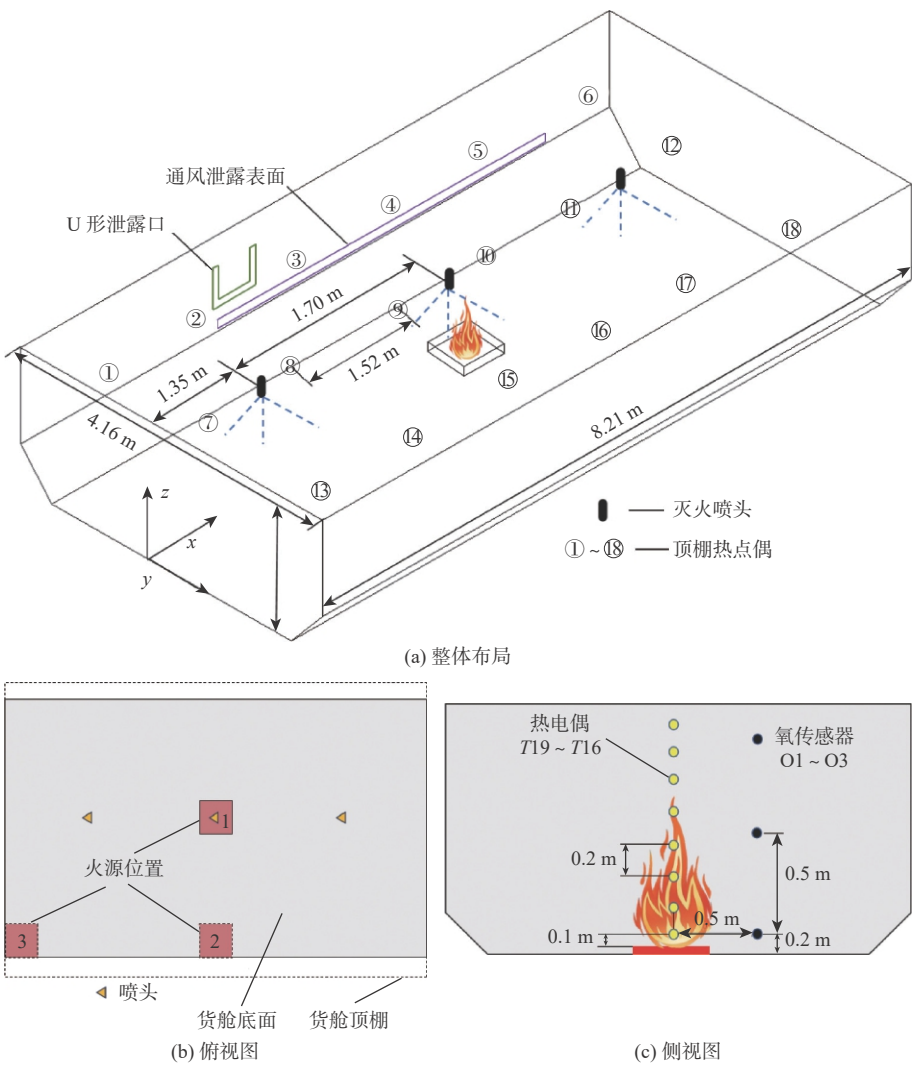
式中: $\dot{m}''_{f}(t)$为液滴作用时局部质量损失率; $\dot{m}''_{\mathrm{f,p}}(t)$为无液滴作用时燃烧速率; $k(t)$为局部单位面积水质量的线性函数, $k(t)=E_{\mathrm{coefficient}}\cdot\dot{m}''_{\mathrm{w}}(t)$, 其中, $\dot{m}''_{w}(t)$为单位面积水质量, $E_{\mathrm{coefficient}}$为灭火系数, 基于前人研究设置灭火系数为 16.4 m^2/(kg·s)。

对于氮气灭火系统, 主要依据其氧稀释原理, 其火焰熄灭作用主要基于氧消耗方法, 而低环境压力主要影响环境氧体积分数和氮气释放扩散动力学过程, 相关理论模型采用默认设置[23]。

2　模型和参数设置

2.1　货舱物理模型构建及测量系统布置

依据 FAA-MPS 标准[16], 构建宽体飞机前货舱全尺寸数值模型, 如图 1 所示, 内部尺寸为 8.11 m×4.16 m×1.67 m; 为模拟实际货舱气体泄露, 在侧壁设置固定流量 (23 L/s) 的 U 形泄漏口和泄漏表面 (0.001 5 m^2)[16]。舱内环境温度设置为 20 ℃, 环境压力考虑为 46 kPa、76 kPa、101 kPa, 分别对应巡航阶段外部环境压力和舱内压力及地面常压环境。顶棚中心线上 x = 1.35 m、x = 4.05 m 和 x = 6.75 m 处布置 3 个喷头安装点 (细水雾和氮气喷头)。顶棚热电偶布置如图 1(a) 所示, 编号依次为 T1～T18, 测点间



(a) 整体布局

(b) 俯视图　　(c) 侧视图

图 1　货舱物理模型和测量系统布局示意

Fig. 1　Physical model of the cargo compartment and schematic layout of the measurement system

距为1.52 m;采用表面池火火源,油盘尺寸为0.609 m×0.609 m×0.102 m,为分析不同火源位置对灭火效果的影响,考虑舱内中心、侧壁和角落3个火源位置,如图1(b)所示;如图1(c)所示,火源上方从下往上依次布置一列热电偶树T19-26测量火焰温度,第1个测点距离油盘表面0.1 m,其余间隔0.2 m;沿y方向距火源0.5 m处竖直方向布置3个间隔0.5 m的氧传感器,从下往上依次编号为O1～O3,用于测量舱内氧体积分数。

2.2 火源参数设置

选择第3类表面燃烧火灾,采用航空煤油(Jet-A)燃料,其物性参数依据参考文献[16]。火源热释放速率(heat release rate, HRR)采用$\tanh$函数增长模型[23],60 s达到峰值,单位面积热释放率可由式(6)计算:

$$\dot{Q}=\dot{m}''\Delta H_{\mathrm{c,eff}}A_{\mathrm{f}}\left(1-\mathrm{e}^{-k\beta D}\right) \tag{6}$$

式中:$\dot{m}''$为单位面积质量损失速率,kg/(m^2·s);$\Delta H_{\mathrm{c,eff}}$为航空煤油有效燃烧热,kJ/kg;$A_{\mathrm{f}}$为油盘水平燃烧面积,m^2;$D$为油盘直径,m;$k\beta$为经验常数,3.5 m^{-1}[17]。

为获得数值模拟中最大火源热释放率输入参数,以角落火为例,式(4)中$\dot{m}''$值参考文献数据[21],分别计算得到环境压力46 kPa、76 kPa和101 kPa时最大火源热释放率分别为258 kW/m、455 kW/m、759 kW/m;而细水雾作用下火源热释放率计算则采用式(5)进行修正。

2.3 网格敏感性分析

数值计算中网格尺寸直接影响计算精度。FDS中火源附近的网格尺寸由无量纲表达式$D^*/\delta_x=\left(Q/\rho_\infty C_{\mathrm{p}}T_\infty\sqrt{g}\right)^{2/5}/\delta_x$确定,式中$D^*$为特征火灾直径,m,$\delta_x$为单元网格尺寸,m;$Q$为火源总热释放速率,kW;$\rho_\infty$为空气密度,kg/m^3;$C_{\mathrm{p}}$为空气比热容,kJ/(kg·K);$T_\infty$为环境温度,K;$g$为重力加速度,m/s^2。$D^*/\delta_x$的合理范围确定在4～16[9]。以常压环境火源功率参数为例,计算D^*为0.65 m,即网格尺寸δ_x合理范围为0.035～0.14 m;因此选择0.06 m、0.08 m、0.10 m、0.12 m、0.14 m这5种网格尺寸进行敏感性分析,以火焰上方温度测点T19的温度数据进行对比分析,不同网格尺寸下温度数据结果如图2所示。从图中可以看出,随着单元网格尺寸减小,稳定阶段火焰温度增加,当网格尺寸从0.14 m减小至0.10 m时,平均温度由690 ℃增加至900 ℃,继续降低网格尺寸大小,平均温度变化不大,考虑到计算成本,网格尺度0.10 m即可满足数值计算精度要求。

2.4 喷头参数设置及工况设置

由于仿真模型边界条件限制,难以直接模拟双流体喷头,本文研究中将氮气和细水雾分别释放形成近似等效双流体细水雾,其中喷头输入参数均代表实际中双流体稳定雾场特征参数。对于氮气喷射模型,采用气体粒子模型设置,并通过速度补偿方法设置竖直向下的喷射方向和速度[23],速度值设置为20 m/s;考虑到受限空间内氮气熄灭B类火灾一般要求灭火剂施放时间为60 s[24],依据文献[17]氮气总流量设置为0.667 kg/s,单个喷头流量约为0.222 kg/s;对于细水雾喷雾模型,雾滴初始喷射速度同样设置为20 m/s,平均粒径、喷雾锥角和流量等参数通过敏感性分析确定;本文分别考虑氮气、细水雾及其协同灭火3种模式,其中协同灭火通过将氮气和细水雾喷头同时开启来实现。为保证灭火启动时火灾发展处于稳定状态,将灭火启动时间设置在火源燃烧70 s时,详细工况见下文,同时定义当火灾热释放率降为0时,即为火焰熄灭。

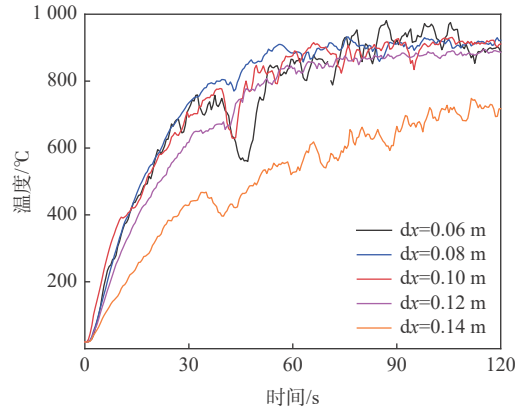


图2 不同单元网格尺寸热电偶T19温度变化结果

Fig. 2 The variation results of temperature for thermocouple T19 with different cell mesh size

3 结果讨论与分析

3.1 单一灭火剂灭火效果

3.1.1 氮 气

图3为常压环境下氮气熄灭货舱不同位置火灾时HRR变化结果。从图中可以看出,灭火启动后,中心火2 s时熄灭,而侧壁火和角落火灭火时间分别为36 s和39 s,其中角落火灭火时间预测结果与文献[17]相近。这是因为中心火位于中间喷头正下方,氮气释放后快速撞击并压制火焰,在动力学作用和氧隔绝效应耦合作用下快速达到熄灭体积分数;而侧壁和角落火时,火源与喷头距离相对较远,气体灭火剂扩散速度衰减迅速,导致火焰区达到熄灭体积分数时间较长,且有较长一段时间对HRR基本无抑制作用,当对火焰的窒息效应明显增强时,即灭火启动28 s和33 s时开始逐渐熄灭,故灭火时间较长。

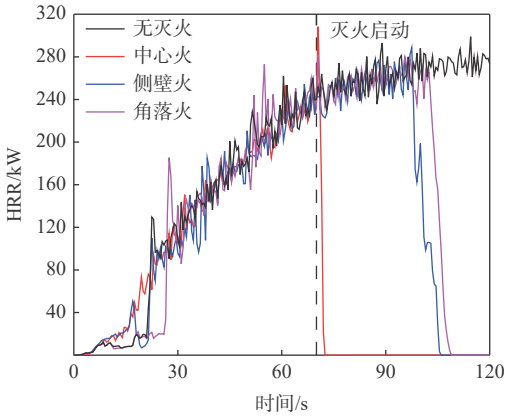


图 3　氮气熄灭不同位置火时 HRR 变化结果

Fig. 3　Results of HRR for the fire at different positions suppressed by N_2

3.1.2　细水雾

为探究雾特性对灭火性能的影响, 进而确定合适的灭火参数, 首先以角落火为例, 分析细水雾平均粒径、流量和喷雾锥角等参数对灭火的影响, 分别考虑平均粒径范围为 100～400 μm, 流量范围为 5～30 L/min, 2 种空心喷雾锥角 (半锥角如图 4 所示, 对应 30°～90°和 60°～120°)。图 5 为灭火时间随平均粒径、流量、喷雾角度变化结果。从图 5(a) 可以发现, 随着平均粒径增加, 灭火时间先增加后减小再趋于稳定; 当平均粒径较小时, 雾滴动量小, 不足以穿透火焰, 此时主要以水蒸气窒息效应参与灭火; 当平均粒径增大到 200 μm 时, 窒息效应减弱但火焰冷却效应仍不足, 灭火时间增加, 此时达到临界值; 当继续增大雾滴粒径时, 雾动量增加, 穿透火焰参与灭火的雾滴量增加, 其冷却降温效应增强, 灭火时间开始减小; 当平均粒径达到 300 μm 左右时, 灭火时间趋于稳定, 2 种效应达到平衡。在灭火时间相近时, 大粒径对火焰热释放率的衰减较为显著, 可有效降低火灾危险性, 故后续研究平均粒径采用 300 μm。

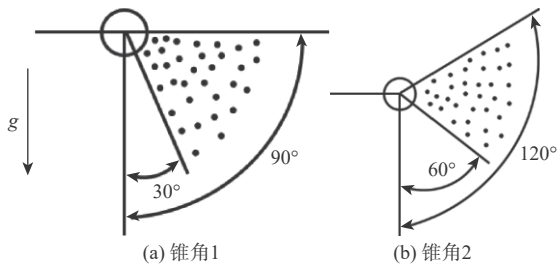


图 4　半雾锥角示意

Fig. 4　Half spray cone angle diagram

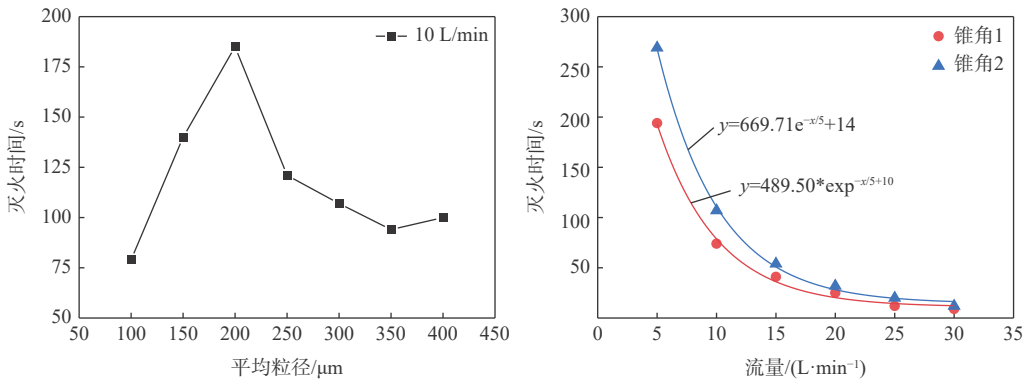


图 5　不同雾特性参数下灭火时间变化结果

Fig. 5　The variation of fire extinguishing time with water mist parameters

图 5(b) 给出了平均粒径 300 μm 时, 不同喷雾流量和不同喷雾锥角时灭火时间变化结果。从图中可以看出, 随着喷雾流量增加, 灭火时间减小, 当流量从 5 L/min 增加到 30 L/min, 喷雾锥角 1 和 2 对应的灭火时间分别由 194 s 和 269 s 减小到 9 s 和 12 s, 并呈现幂指数衰减; 对比 2 种喷雾锥角灭火时间, 发现喷雾锥角 1 灭火时间短, 主要因为火源位于货舱底部, 半锥角为 30°～90°时水雾有效作用量较大, 后续研究喷雾锥角采用锥角 1。为了更好地对比协同灭火效果, 流量选择与氮气灭火时间接近的 15 L/min。

图 6 为常压环境下细水雾熄灭货舱不同位置火灾时 HRR 变化结果。从图中可以看出, 侧壁火最易熄灭, 而角落火灭火时间最长, 同样由于与喷头距离越远, 水雾有效作用量越少, 而灭火阶段中心火 HRR 衰减速率较小, 这主要是空心喷雾导致中心区域雾滴量较少, 对正下方火源直接抑制作用受限。

3.2　氮气-细水雾协同灭火效果

以灭火难度最高的角落火为例分析氮气-细水雾协同灭火效果, 其中灭火剂参数选取 3.1 节中单一灭火介质相关参数。

3.2.1　灭火时间

图 7 给出了 3 种灭火模式下的 HRR 变化结果;

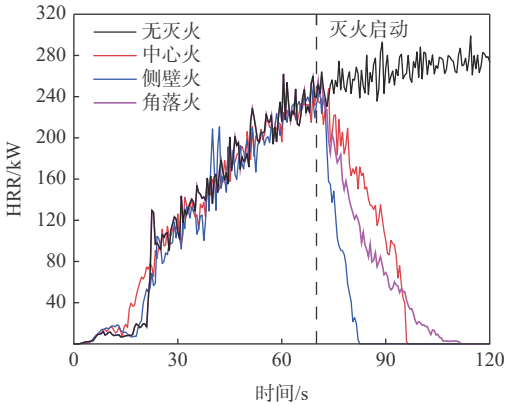


图 6 细水雾熄灭不同位置火时 HRR 变化结果

Fig. 6 Results of HRR for the fire at different positions suppressed by water mist

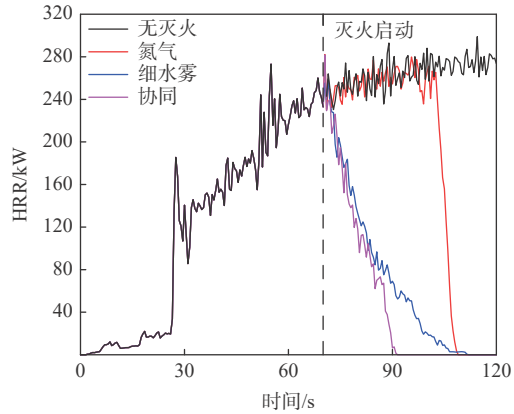


图 7 不同灭火模式下 HRR 变化结果

Fig. 7 The variation of HRR under different fire extinguishing modes

可以看出，当灭火时间相近时，细水雾对 HRR 衰减作用明显强于氮气，这主要由于雾滴优异的冷却降温效应；而协同灭火时灭火时间显著降低，灭火前期以冷却降温为主，后期窒息效应明显增强。如表 1 工况 1 所示，当同样的氮气和细水雾流量时，与单一氮气和细水雾相比，协同灭火时间分别缩短 46.2% 和 48.8%；同时协同灭火时，氮气的加入使得水消耗量降低了 48.78%。为进一步探究协同灭火时灭火剂流量改变对灭火效果的影响，将氮气和细水雾流量分别降低至 0.10 kg/s 和 10 L/min，对比结果如表 1 工况 2 所示，灭火剂流量降低后，灭火时间明显增加，且灭火剂的消耗量均增加。因此，实际中灭火喷头应考虑大流量快速释放的设计方案，以实现快速灭火同时降低消耗量。

表 1 不同灭火剂流量下灭火对比结果

Table 1 Comparative results of fire extinguishing under different flow rates of extinguishing agents

工况序号	灭火剂	灭火时间/s	消耗量/kg	
			氮气	细水雾
1	氮气(0.222 kg/s)	39	26.01	0
	细水雾(15 L/min)	41	0	30.75
	协同(氮气: 0.222 kg/s, 水:15 L/min)	21	14.01	15.75
2	氮气(0.10 kg/s)	65	19.50	0
	细水雾(10 L/min)	70	0	35.00
	协同(氮气: 0.10 kg/s, 水:10 L/min)	54	16.20	27.00

3.2.2 温度结果

图 8 给出了 3 种灭火模式下火焰上方温度变化结果。由图可知，灭火剂施加后温度迅速下降，降温速率依次为：协同>细水雾>氮气。对于氮气，其

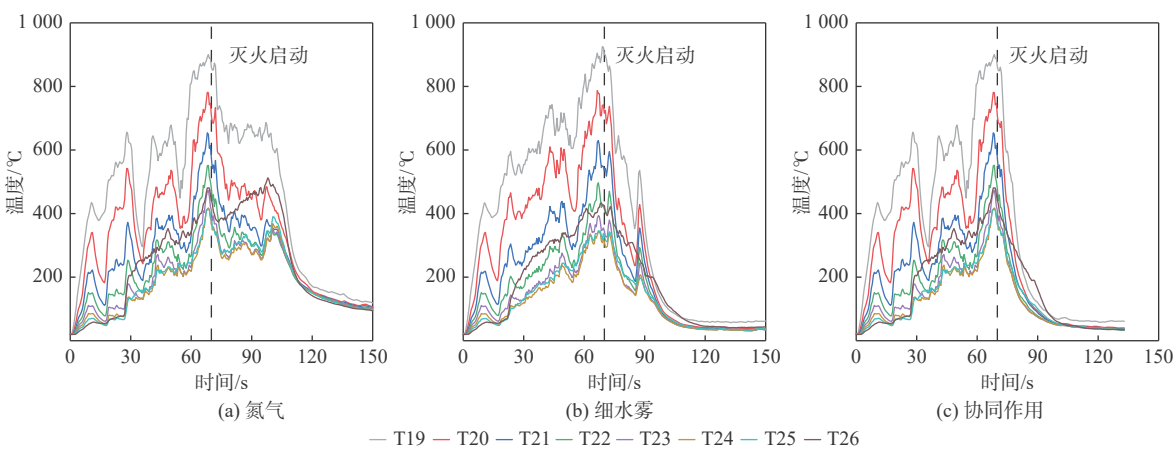


图 8 火焰上方热电偶温度变化结果

Fig. 8 The variation of temperature above the flame

灭火作用主要是氧稀释, 启动后火焰温度有一定程度的降低, 随后保持较高温度稳定直至火焰熄灭; 而细水雾参与可以快速降低火场温度, 二者协同时可以起到正向协同效应。由表 1 工况 1 可知, 协同作用时灭火时间为 21 s, 选择灭火启动 15 s 时不同灭火模式下油盘中心截面 (x=7.8 m) 温度场分布结果进一步对比分析, 如图 9 所示, 发现氮气作用时火焰区温度基本未降低且火焰上方出现一个明显高温区, 这说明氮气冷却降温作用弱, 且在灭火早期阶段会引起一定火焰强化现象, 带来一定的危险性; 而单独细水雾和协同作用时火焰温度和高温区域范围分别依次降低, 同时货舱顶棚和空间区域温度降低明显。

货舱顶棚温度是评估灭火剂灭火效果的关键参数。图 10 为灭火启动 15 s 时顶棚温度分布结果。从图中可以看出, 启动前, 顶棚温度从货舱角落区域往四周逐步衰减, 火焰上方区域最高温度达到 442 ℃; 结合图 10(b), 氮气作用下, 火焰附近区域顶棚温度明显增加, 火焰上方区域最高温度达到

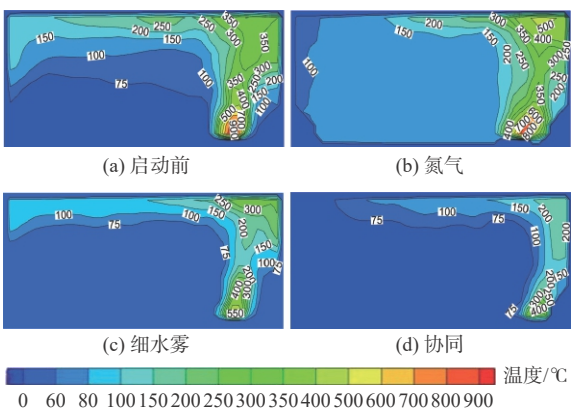


图 9　灭火启动 15 s 时油盘中心截面 (x=7.8 m) 温度分布结果

Fig. 9　Temperature distribution of the central section of the oil pan (x=7.8 m) for 15 s after the activation of extinguishing agent

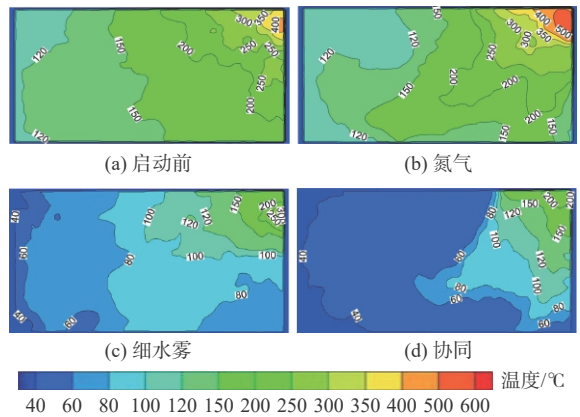


图 10　灭火启动 15 s 时顶棚温度分布结果

Fig. 10　Temperature distribution of the ceiling for 15 s after the activation of extinguishing agent

590 ℃, 这主要由于气体释放引起火焰周围气流湍流强度增强, 空气卷吸效应引起火焰高度增加, 进而导致火焰上方出现高温区; 而细水雾及协同灭火时由于水雾冷却降温效应, 顶棚温度明显降低, 如图 10(c) 和 10(d) 所示, 火焰上方区域最高温度分别 310 ℃ 和 240 ℃, 且顶棚高温区范围依次减小, 这说明协同灭火更具有优势。

3.2.3　氧体积分数结果

火焰附近氧体积分数参数反映灭火剂释放后的窒息效应。图 11 为不同灭火模式下从燃烧开始至火焰熄灭阶段的氧体积分数变化结果。灭火启动前, 由于烟气分层现象, 顶部烟气层内氧体积分数下降较快, 灭火启动时降低至 0.17 左右, 而下部区域下降不明显; 灭火启动后, 由于灭火剂释放, 3 个测点氧体积分数均快速降低, 但单独氮气作用下氧体积分数下降最快, 火焰熄灭时氧体积分数为 0.11 左右, 主要由于氮气窒息效应较强; 协同时氧体积分数下降程度次之, 火焰熄灭时氧体积分数为 0.139 左右, 这主要是细水雾冷却降温效应导致火

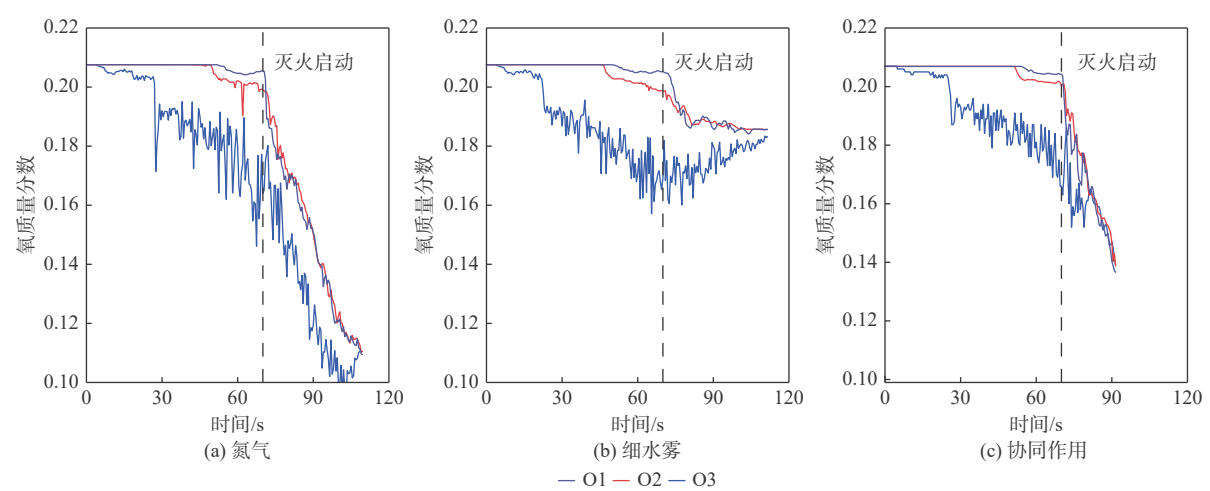


图 11　测点 O1～O3 氧体积分数变化结果

Fig. 11　Results of oxygen volume fraction change in the measurement point O1～O3

焰熄灭时间减小，进而氮气释放量减少；而单独细水雾作用时中部和下部测点氧体积分数下降到0.19左右时，逐步趋于稳定，直至火焰熄灭时氧体积分数0.185左右，这主要由于细水雾以冷却降温为主，窒息效应相对较弱，上部测点氧浓度升高，这主要由于喷雾动力学效应下引起烟气分层失稳破裂，空气卷吸进入氧气。

灭火启动15 s时不同灭火模式下油盘中心截面(x=7.8 m)氧体积分数分布结果进一步对比分析，如图12所示。从图中可以看出，氮气和协同作用下火场氧体积分数降低明显，且单独氮气作用时火焰区域氧体积分数最低，而细水雾作用下氧体积分数下降不明显，这再次验证了氮气的窒息效应较强。

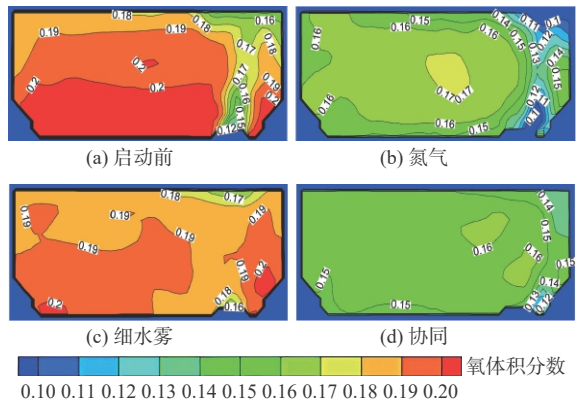


图12 灭火剂启动15 s时油盘中心截面(x=7.8 m)氧体积分数分布结果

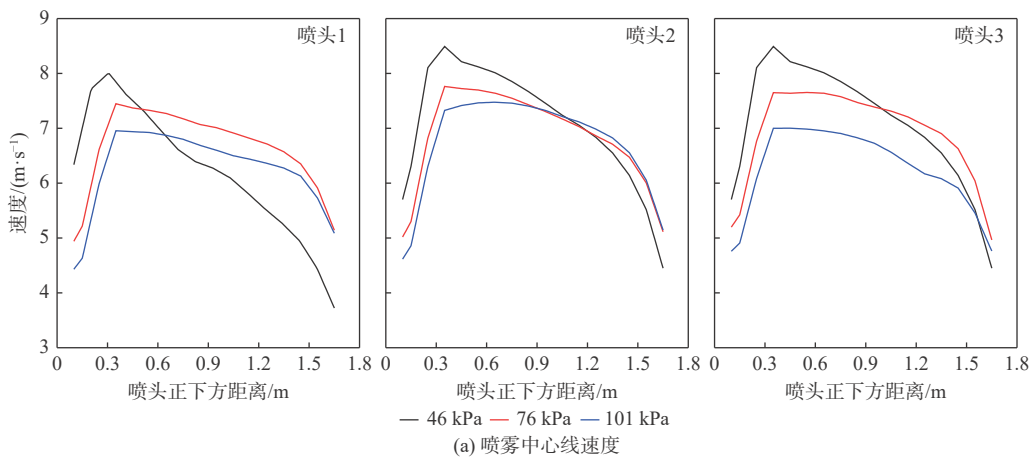
Fig. 12 Oxygen concentration distribution of the central section of the oil pan (x=7.8 m) for 15 s after the activation of extinguishing agent

3.3 巡航货舱环境下灭火效果

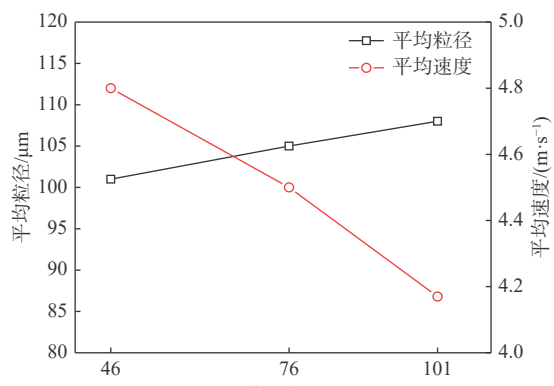
为评估飞机巡航环境下氮气-细水雾协同灭火有效性，考虑不同环境压力对灭火效果的影响，其中灭火剂参数同样选取3.1节中单一灭火介质相关参数。

3.3.1 灭火剂释放扩散特性

灭火剂扩散特性直接影响火源区域灭火剂浓度能否快速达到灭火要求，而环境压力对于流体扩散过程产生一定影响。对于氮气，主要通过稳定阶段气体射流中心线速度来表征，如图13(a)所示，环境压力46 kPa时，喷头附近气体初始速度较大，但随着中心线喷射距离增加速度衰减变快，而环境压力增加时，气体扩散速度有所降低；此外，中心喷头2整体扩散速度较大。对于细水雾，主要通过喷头中心下方0.8 m处雾滴平均粒径和速度表征，以喷头2，流量为15 L/min时为例，如图13(b)所示，随着



(a) 喷雾中心线速度



(b) 雾滴平均粒径和速度

图13 灭火剂扩散参数随环境压力变化结果

Fig. 13 Results of diffusion parameters of extinguishing agent varying with ambient pressure

环境压力降低, 雾滴平均粒径减小, 而平均速度增大, 这与前人小尺度实验研究结果[12] 相一致。

3.3.2　灭火效果

图 14 为不同环境压力和灭火模式下 HRR 变化结果。无灭火时, 根据 2.2 节不同环境压力下最大火源热释放率设置, 不同环境压力下火源热释放速率曲线如图所示; 灭火启动后, 同一环境压力下不同灭火剂的灭火效果与常压保持一致, 而随环境压力降低, 灭火时间均减小, 详细结果如表 2 所示。低环境压力下灭火效果好, 主要原因为: ①低压下火源热释放率较低, 且货舱空间内绝对氧含量较低, 导致灭火所需的灭火剂量减少; ②环境压力降低, 氮气初始扩散速度相对增加, 窒息效应有所增加; 雾滴粒径减小且速度增加, 冷却效应增强。在巡航阶段货舱发生火灾时, 建议打开货舱通风阀减压, 以达到快速灭火的目的。

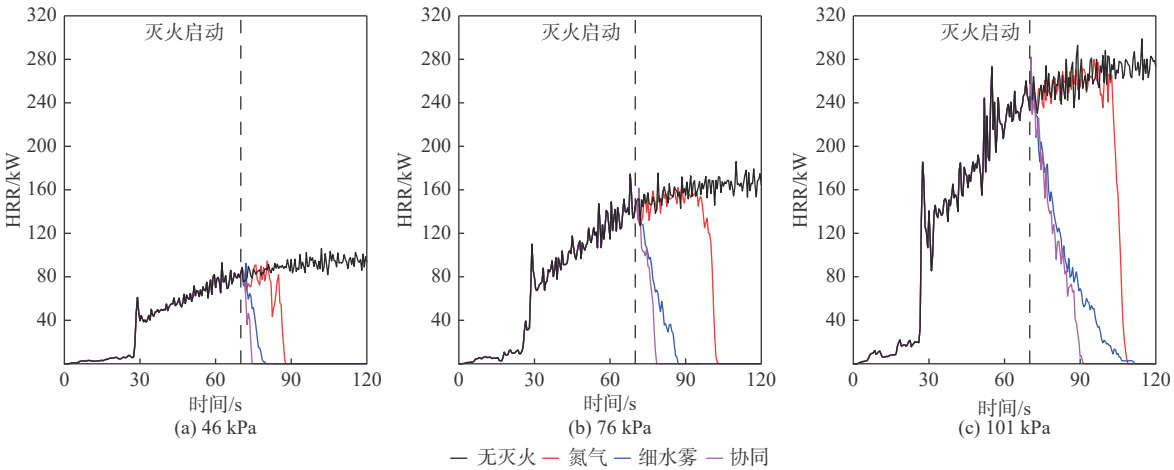


图 14　不同环境压力与灭火模式下 HRR 变化结果

Fig. 14　HRR variation results of fire source under different ambient pressure and fire extinguishing mode

表 2　不同环境压力与灭火模式下灭火时间结果

Table 2　Fire extinguishing time data results under different ambient pressure and fire extinguishing mode

灭火剂	灭火时间/s		
	46 kPa	76 kPa	101 kPa
氮气	18	33	39
细水雾	10	17	41
协同	5	9	21

图 15 为不同环境压力与灭火模式下火焰上方热电偶 T19 温度变化结果。随着环境压力降低, 降温速率均明显增加, 其中协同灭火降温效果最好; 对于氮气, 较高环境压力时, 灭火启动后火焰温度会首先快速降低, 然后维持高温一段时间直至火焰熄灭, 这是因为受气体射流压制, 火焰温度有所降低, 同时氮气冷却效应较弱而出现高温阶段; 但环境压力 46 kPa 时则基本无高温阶段, 这是因为灭火时间较短; 而细水雾和协同灭火时不同环境压力下火焰温度均快速降低直至熄灭; 不同环境压力下协同灭火时降温速率数据如表 3 所示, 环境压力从 101 kPa 降低至 46 kPa, 降温速率提高 48.60%。

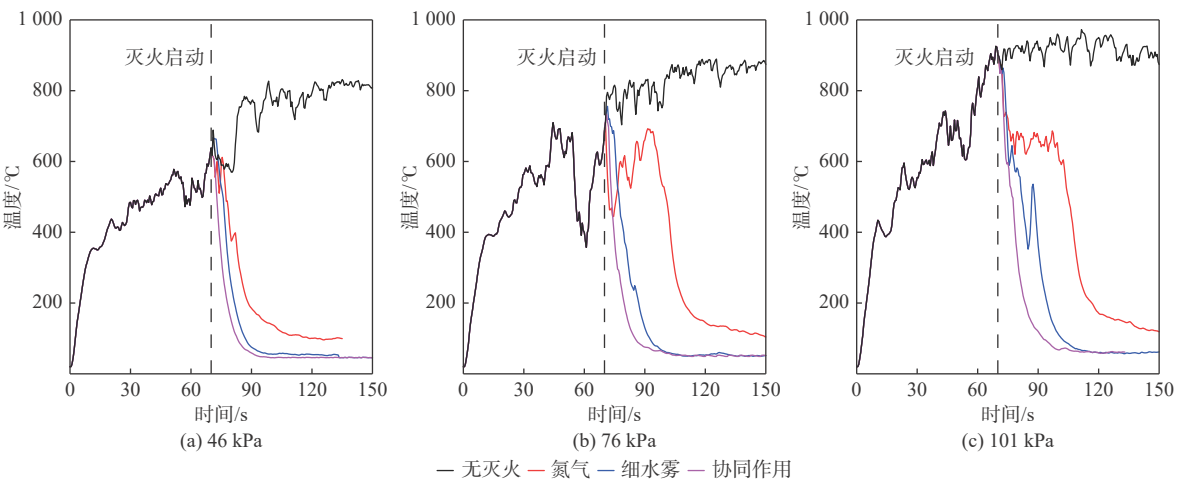


图 15　不同环境压力与灭火模式下热电偶 T19 温度变化结果

Fig. 15　Temperature changes of thermocouple T19 under different ambient pressures and fire extinguishing mode

表 3 不同环境压力下协同灭火时降温速率数据

Table 3 Data of synergetic cooling rate under different ambient pressures

环境压力/ kPa	最高温度/℃	达到常温时间/s	降温速率/(℃·s^{-1})
46	638	23	26.0
76	769	43	18.0
101	900	49	17.5

以协同灭火工况为例，不同环境压力下启动 5 s 时油盘中心截面 (x=7.8 m) 和顶棚温度分布结果如图 16 和图 17 所示。从图中可以看出，无灭火剂时，随着环境压力降低，火焰区和货舱顶棚附近温度均有所降低；环境压力从 101 kPa 降低至 46 kPa，火源上方顶棚最高温度从 442 ℃ 降低至 387 ℃，这

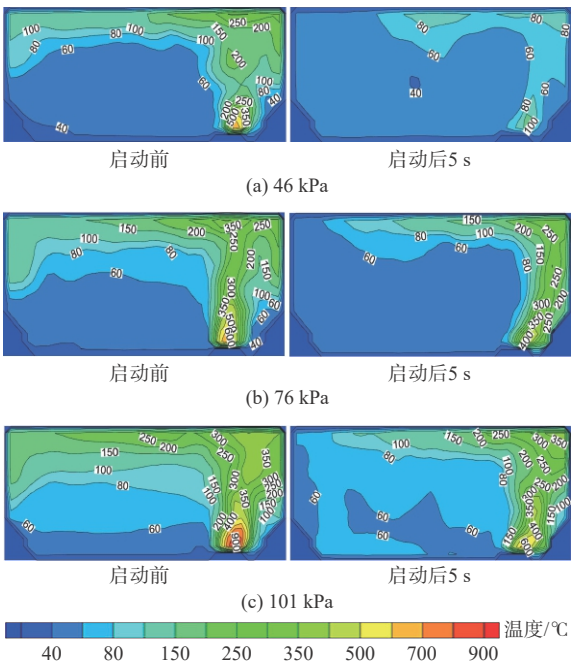


图 16 灭火启动 5 s 时油盘中心截面 (x=7.8 m) 温度分布结果

Fig. 16 Temperature distribution results of the central section of oil pan (x=7.8 m) for 5 s after the activation of extinguishing

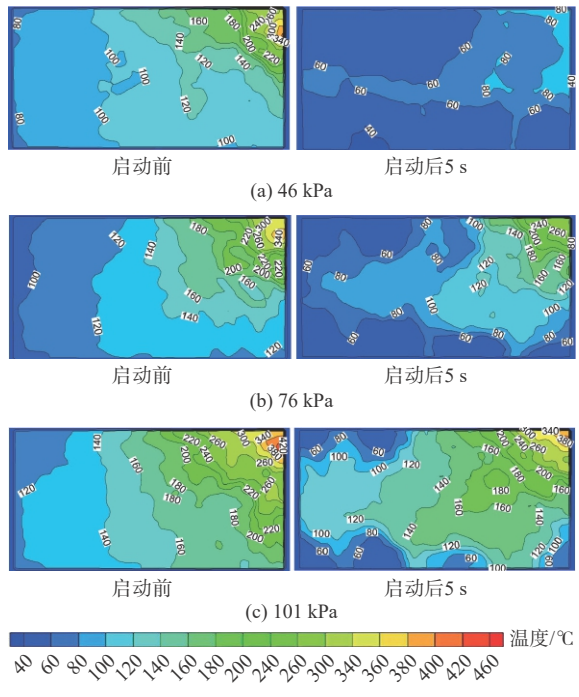


图 17 不同环境压力下协同灭火时顶棚温度变化结果

Fig. 17 Results of ceiling temperature variation under different ambient pressure

主要由于低环境压力下火源热释放率低，其燃烧后释放的热量较低。灭火启动 5 s 时，随着环境压力降低，火焰温度和高温区域范围同样分别依次降低，在同一流量下，环境压力从 101 kPa 降低至 46 kPa，火源上方顶棚最高温度从 384 ℃ 降低至 101 ℃。

为区分不同环境压力下实际氧体积分数，本文采用绝对氧体积分数，即相对平原环境的货舱空间单位体积内氧体积分数进行对比分析。环境压力 101 kPa, 76 kPa 和 46 kPa 对应的货舱空间环境绝对氧体积分数分别为 0.21、0.17 和 0.14。图 18 为不同环境压力和灭火剂模式下测点 O1 绝对氧含量变化结果。从图中可知，随着环境压力下降低，不同灭火剂的灭火时间均明显减小，且火焰熄灭时对应

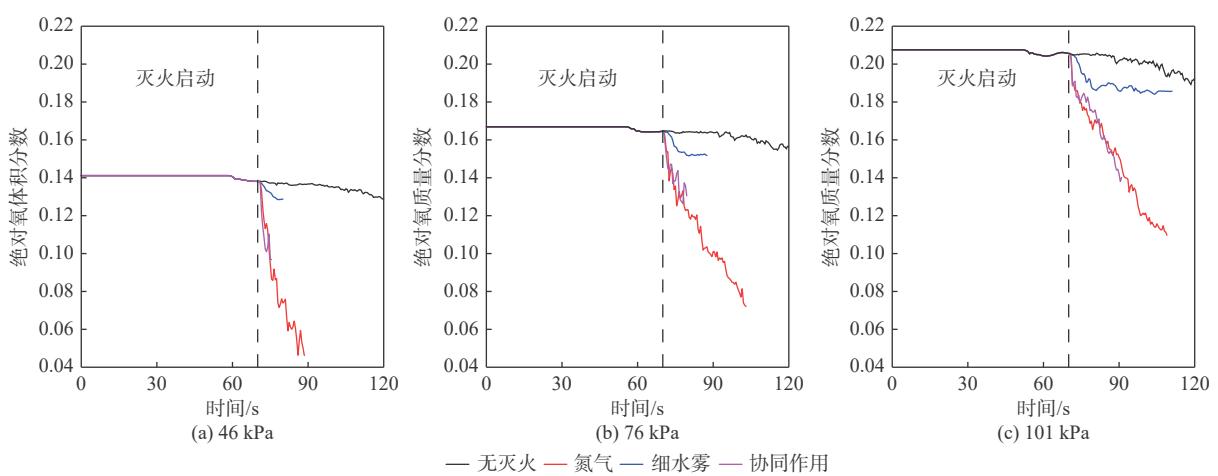


图 18 不同环境压力和灭火模式下测点 O1 氧体积分数变化结果

Fig. 18 Results of O1 oxygen volume fraction under different ambient pressure and extinguishing mode

的绝对氧体积分数下降。以协同灭火为例, 不同环境压力下油盘中心截面 (x=7.8 m) 绝对氧体积分数分布结果如图 19 所示, 从图中可知, 灭火启动前, 不同环境压力下仅火焰区和顶棚附近区域氧体积分数下降; 灭火启动后, 火源附近区域和货舱内绝对氧体积分数均进一步降低, 而低环境压力下氧体积分数降低更显著。

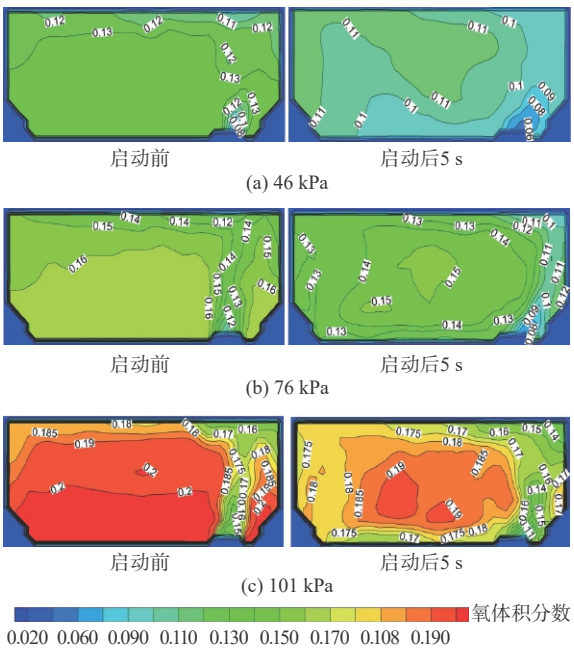


图 19　不同环境压力下协同灭火时油盘中心截面 (x=7.8 m) 氧体积分数变化结果

Fig. 19　Results of oxygen concentration in the central section of the oil pan (x=7.8 m) under different ambient pressure

4　结　论

1) 火源与喷头相对距离增加时, 作用于火源区域的灭火剂有效量减少, 灭火时间增加, 角落火最难扑灭。

2) 对于角落火, 氮气灭火启动后一段时间内对火源功率基本无抑制作用, 导致其应用于货舱灭火时存在一定的安全问题; 细水雾启动后火源热释放快速衰减, 灭火效果则由进入火源区的雾滴有效作用量决定, 其灭火主控机制受细水雾参数影响, 当喷雾流量一定时, 随着雾滴平均粒径增加, 灭火时间先增加后减小再趋于稳定, 平均粒径 200 μm 为主控灭火机制由窒息效应转为冷却降温的临界值; 随着水流量增加, 灭火时间呈现幂指数衰减; 固定喷雾锥角时细水雾喷射方向直接影响灭火效果。

3) 对于不同灭火方式, 在氮气和细水雾灭火时间相近时, 细水雾的热释放率衰减作用较好; 相较于单一灭火剂, 协同时冷却降温和窒息效应均显著增强; 在氮气和细水雾单喷头流量分别为 0.222 kg/s 和 15 L/min 时, 与单独氮气和细水雾相比, 协同灭火效率分别提高 46.20% 和 48.80%, 其中氮气的加入可使水耗量减少 48.78%; 灭火剂释放速率增加可显著降低灭火时间, 大流量快速释放喷头更有利于灭火。

4) 环境压力降低, 氮气扩散速度加快, 细水雾雾滴粒径减小且速度增加。随着环境压力降低, 各灭火剂作用下灭火时间均减小, 而协同灭火时, 环境压力从 101 kPa 降低至 46 kPa, 灭火降温速率提高 48.60%。

未来研究中需要进一步考虑不同可燃物种类和荷载量等参数, 并进行大尺度实验验证和模型普适性扩展研究。

参考文献（References）

[1] CRAIG P L. “Environmental Control Systems” in the encyclopedia of aerospace engineering[M]. Hoboken: John Wiley & Sons, Ltd, 2010.

[2] PROTOCOL M, DEPLETE O. Halons technical options committee assessment report[R]. Geneva: UN Environment, 2018.

[3] 贺元骅, 张政, 伍毅. 低压环境下航空电缆材料燃烧特性的研究[J]. 塑料科技, 2020, 48(1): 71-74.
HE Y H, ZHANG Z, WU Y. Experimental study on combustion characteristics of aviation cable materials under low pressure environment[J]. Plastics Science and Technology, 2020, 48(1): 71-74(in Chinese).

[4] WANG W, WANG L, YANG R, et al. Investigation of the effect of low pressure on fire hazard in cargo compartment[J]. Applied Thermal Engineering, 2019, 158: 113775.

[5] 刘全义, 朱博, 邓力, 等. 低压环境下不同直径油池火特征参量研究[J]. 中国安全科学学报, 2021, 31(4): 105-110.
LIU Q Y, ZHU B, DENG L, et al. Study on characteristic parameters of oil pool fire of different diameters in low pressure environment[J]. China Safety Science Journal, 2021, 31(4): 105-110(in Chinese).

[6] DINESH A, BENSON C M, HOLBORN P G, et al. Performance evaluation of nitrogen for fire safety application in aircraft[J]. Process Safety and Environmental Protection, 2020, 141: 110-122.

[7] HU X Q, KRAAIJEVELD A, LOG T. Numerical investigation of the required quantity of inert gas agents in fire suppression systems[J]. Energies, 2020, 13: 2536.

[8] LEE J, MOON J. Numerical analysis of the effect of horizontal distance between a water mist nozzle and ignition source on reduction in heat release rate[J]. Annals of Nuclear Energy, 2020, 144: 107560.

[9] WANG Z, Wang W H, WANG Q S. Optimization of water mist droplet size by using CFD modeling for fire suppressions[J]. Journal of Loss Prevention in the Process Industries, 2016, 44: 626-632.

[10] ZHANG T W, LIU H, SONG J W, et al. Synergistic inhibition ef-

fect on lithium-ion batteries during thermal runaway by N_2-twin-fluid liquid mist[J]. Case Studies in Thermal Engineering, 2022, 37: 102269.

[11] ZHANG B, ZHU Q, XU Y, et al. The effect of ambient pressure on gas jet diffusion in air: experimental and numerical study[J]. International Journal of Heat and Mass Transfer, 2019, 134: 392-403.

[12] WANG X S, ZHU P, LI Y, et al. Effect of low ambient air pressure on spray characteristics of water mist[J]. Experimental Thermal and Fluid Science, 2015, 66: 7-12.

[13] ZHU P, WANG X S. Numerical study on the effects of ambient air pressure on water mist characteristics[J]. Energy Procedia, 2015, 66: 169-172.

[14] 戴经天. 低压环境细水雾灭火有效性实验研究[D]. 广汉: 中国民用航空飞行学院, 2019: 13-48.
DAI J T. Experimental study on fire extinguishing effectiveness of water mist in low pressure environment[D]. Guanghan: Civil Aviation Flight College of China, 2019: 13-48(in Chinese).

[15] MA Q J, WAN M S, SHAO J C, et al. Pool fire suppression performance by twin-fluid water mist under low pressures in an altitude chamber[J]. Process Safety Progress, 2021, 40(1): e12155.

[16] REINHARDT J. Minimum performance standard for aircraft cargo compartment halon replacement fire suppression systems (2012 update[R]. Washington, D.C.: Federal Aviation Administration, 2012.

[17] XIONG Y F, DIAKOSTEFANIS M, DINESH A, et al. Numerical assessment for aircraft cargo compartment fire suppression system safety[J]. Journal of Fire Sciences, 2021, 39(3): 240-261.

[18] MARKER T R, REINHARDT J. Water spray as a fire suppression agent for aircraft cargo compartment Fires: DOT/FAA/AR-TN01/1 [R]. Virginia: National Technical Information Service, 2001.

[19] REINHARDT J. The evaluation of water mist with and without nitrogen as an aircraft cargo compartment fire suppression system: DOT/FAA/AR-01/121[R]. Virginia: National Technical Information Service, 2002.

[20] 吕志豪. 民机货舱细水雾雾场特性及抑灭火有效性研究[D]. 广汉: 中国民用航空飞行学院, 2020: 37-79.
LV Z H. Study on characteristics of water mist field and effectiveness of fire suppression in Civil aircraft cargo Hold[D]. Guanghan: Civil Aviation Flight College of China, 2020: 37-79(in Chinese).

[21] 朱培, 罗圣峰, 刘全义, 等. 货舱低压环境下细水雾抑灭航空煤油池火有效性[J]. 清华大学学报 (自然科学版), 2022, 62(1): 21-32.
ZHU P, LUO S F, LIU Q Y, et al. Effectiveness of aviation kerosene pool fire suppression by water mist in a cargo compartment with low-pressure environment[J]. Journal of Tsinghua University (Science and Technology), 2022, 62(1): 21-32(in Chinese).

[22] KEVIN M G, SIMO H, RANDALL M D, et al. Fire dynamics simulator user's guide[M]. Gaithersburg: National Institute of Standards and Technology, 2019.

[23] KEVIN M G, SIMO H, RANDALL M D, et al. Fire dynamics simulator technical reference guide volume 1: mathematical mode[M]. Gaithersburg: National Institute of Standards and Technology, 2019.

[24] Klein RA. SFPE handbook of fire protection engineering[M]. 5th ed. New York; Springer, 2015.

Numerical simulation study on synergetic fire suppression of nitrogen/water-mist in aircraft cargo compartment

HUANG Suping[1], ZHU Pei[1, 2, *], LIU Quanyi[3], SHAO Quan[1]

(1. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
2. Key Laboratory of Civil Aviation Thermal Disaster Prevention and Emergency, Civil Aviation University of China, Tianjin 300300, China;
3. Sichuan Provincial Key Laboratory of Civil Aircraft Fire Science and Safety Engineering, Guanghan 618307, China)

Abstract: The threat of fire in the cargo compartments of civil aircraft poses a considerable safety risk. The synergetic fire suppression effectiveness of nitrogen-water mist to extinguish aviation kerosene pool fire in a full-scale aircraft cargo compartment was studied using the Fire Dynamics Simulator (FDS) code, considering different fire source locations and ambient pressures. The results showed that the location of fire source affects the effectiveness of fire suppression, and the corner fire is the most difficult to extinguish. For corner fire, it is difficult to suppress by nitrogen at the initial stage, whereas water mist can rapidly reduce the heat release rate of fire. When the flow rates of nitrogen and water mist were set at 0.222 kg/s and 15 L/min, respectively, the synergetic suppression demonstrated significantly enhanced cooling and suffocation effects compared to using nitrogen or water mist alone. The suppression efficiency increased by 46.2% and 48.8% respectively, while nitrogen reduced water consumption by 48.87%. Increasing the release rate of extinguishing agents with constant suppression parameters could reduce both the suppression time and consumption. Therefore, it is recommended to design nozzles with high flow rates and rapid release. Furthermore, decreasing ambient pressure significantly reduced the suppression time. A decrease for ambient pressure from 101 kPa to 46 kPa increased the cooling rate by 48.60% during synergistic suppression.

Keywords: civil aircraft cargo compartment; N_2; water mist; synergetic fire suppression; ambient pressure

Received: 2023-07-17; **Accepted**: 2023-09-22; **Published Online**: 2023-10-23 15:26

URL: link.cnki.net/urlid/11.2625.V.20231023.1039.002

Foundation items: National Natural Science Foundation of China (52202416); Opening Fund of Key Laboratory of Civil Aviation Thermal Disaster Prevention and Emergency, Civil Aviation University of China (RZH2021-KF-02)

* **Corresponding author**. E-mail: zpei@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0495

格网误差建模辅助的GNSS/IMU组合导航综合质量控制算法

戴晔莹，孙蕊*，邓思瑜，吉莉，王媛媛，黄薛东

(南京航空航天大学 民航学院，南京 211106)

摘　　要：在城市复杂环境中，全球导航卫星系统（GNSS）由于高楼等障碍物的遮挡和反射，易产生非视距接收（NLOS）与多径干扰（MI），定位精度和可靠性严重下降，无法满足用户高精度和高可靠性的定位、导航和授时（PNT）服务需求。基于GNSS原始观测量的信号分类和多径建模方法对于缓解城市环境中GNSS多径效应，提高定位精度有重要意义，但这类模型的精度、效率和检验方法往往存在欠缺。提出了一种格网误差建模辅助的GNSS/IMU组合导航综合质量控制算法，通过对城市区域进行格网伪距误差建模，并且实现基于格网拟合精度和卫星故障检验构建精细化的综合质量控制，有效地优化了城市复杂环境下的GNSS/IMU组合导航定位性能。城市环境车载实验表明，所提算法相较传统GNSS/IMU组合导航算法在水平和3D方向上分别提升了50.23%和66.77%，相较基于格网建模算法提升了11.56%和40.53%。

关　键　词：GNSS/IMU组合导航；伪距误差建模；质量控制；加权最小二乘；故障检验
中图分类号：V324; P228
文献标志码：A　　**文章编号**：1001-5965(2025)09-3174-09

智能交通系统是未来城市交通系统重要的发展方向，位置等实时导航信息的精确感知是智能交通系统高效、智能运作的关键。全球导航卫星系统(global navigation satellite system, GNSS)已广泛应用于定位、导航和授时(positioning, navigation, timing, PNT)服务，但在城市环境中，GNSS信号往往易被建筑物或树木遮挡或反射[1]，从而导致严重的非视距接收(non-light-of-sight, NLOS)和多径干扰(multi-path interference, MI)，无法满足智能交通车辆定位精度的需求。

国内外多位学者围绕城市多径环境中的GNSS定位问题展开研究，主要分为天线设计、信号处理和数据处理3个方向。射频端可采用扼流圈天线等抗多径天线抑制低仰角多径信号的入射[2]。但扼流圈天线和双极化天线成本较高，不利于大量推广[3]。基带信号处理端可采用窄相关、多径估计延迟锁定环路等技术抑制多径误差[4]。高精度的GNSS接收机往往会同时采用抗多径天线和基带信号处理方法，但对于多径影响的消除仍有一定的局限性[5]。

在后端的GNSS数据处理和算法设计层面，国内外学者对利用卫星观测量和相关信息消除NLOS和多径影响展开研究。通过对接收信号进行分类和NLOS排除可以提高定位精度，传统的基于单一C/N0，卫星仰角和方位角阈值比较的方法在复杂的城市环境中分类精度较低[6]，如NLOS信号也可能具有较高的C/N0。基于聚类算法和机器学习算法的多径/NLOS检测方法被广泛提出[7-8]，能够综合处

收稿日期：2023-07-27；**录用日期**：2023-09-01；**网络出版时间**：2023-12-04 16:02
网络出版地址：link.cnki.net/urlid/11.2625.V.20231204.0907.001
基金项目：国家自然科学基金(42222401,42174025,41974033)；工信部专项科研项目(TC220A04A-79)；江苏省“六大人才高峰”项目(KTHY-014)；江苏省自然科学基金(BK20211569)；中央高校基本科研业务费专项资金(xcxjh20220722)
* **通信作者**. E-mail：rui.sun@nuaa.edu.cn

引用格式：*戴晔莹，孙蕊，邓思瑜，等. 格网误差建模辅助的GNSS/IMU组合导航综合质量控制算法[J]. 北京航空航天大学学报，2025，51(9)：3174-3182. DAI Y Y，SUN R，DENG S Y，et al. Grid error modeling aided GNSS/IMU integrated navigation comprehensive quality control algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3174-3182 (in Chinese).*

理 C/N0 等多种特征, 相应的输入和输出变量中提取的规则应用于测试的 GNSS 数据, 预测信号分类并排除 NLOS/多径信号进行定位[6]。然而, 通过这种方法实现的定位精度受到限制, 因为分类精度受到机器学习算法所需的离线标记过程引入的错误的影响。Sun 等[9] 提出了一种基于梯度增强决策树的信号分类方法, 采用的是 C/N0、伪距残差和卫星仰角 3 个特征, 增强了信号接收分类的准确性, 在静态数据上验证了对 NLOS、多径和 LOS 3 种信号的分类准确率分别达到 85.7%、81.5% 和 99.8%, 优于其他算法。但卫星剔除可能导致卫星数较少或几何构型差导致定位精度反而下降。不同于直接剔除卫星, 使用加权最小二乘的方法降低观测质量较差的卫星在定位解算中的权重, 常用的有基于高度角的正弦、余弦模型, 基于信噪比或载噪比的 Brunner 模型或综合加权模型[10], 其中综合加权模型效果最优, 在城市复杂环境中定位精度平均提高了 35.11%[11]。但加权方法对定位精度的影响较小, 准确地直接修正伪距则更为直接地大幅度提升定位精度。Sun 等[12] 使用梯度决策树(gradient boosting decision tree, GBDT)通过考虑信号强度、卫星仰角和伪距残差等多个变量来预测伪距误差并进行修正, 在城市复杂环境中的静态试验中 3D 定位精度为 23.3 m, 比伪距误差修正前的定位精度提升了 70%。除了直接修正之外, 基于预测的伪距误差动态调整 GNSS/IMU(inertial measurement unit)紧组合卡尔曼滤波中的观测噪声协方差阵, 3D 定位精度达到 9.21 m, 相比传统扩展卡尔曼滤波(extended Kalman filter, EKF)和直接修正伪距的组合导航定位精度分别提升了 55% 和 15%[13]。为了进一步提升城市环境中多径误差预测模型的分辨率和计算效率, 该团队继续提出了基于格网的伪距误差建模方法, 伪距误差修正后 GNSS 水平和 3D 定位精度相比未修正分别提升了 40.8% 和 63.3%[14]。GNSS 城市环境中存在动态障碍物且观测条件恶劣多变, 因此基于伪距预测误差模型直接进行全部卫星的修正会导致定位精度提升效果有限甚至精度进一步恶化, 但以上模型均对每一个历元的每一颗卫星进行了统一的修正或加权等处理, 自适应性不足, 缺少检测和自适应策略的切换。

本文提出了一种基于格网误差建模辅助的 GNSS/IMU 组合导航综合质量控制算法。对城市环境区域进行了基于六边形格网的伪距误差建模, 并设计综合质量控制策略:

1) 根据格网拟合精度, 切换该历元 GNSS 数据为基于预测的伪距误差直接进行修正或加权最小二乘定位, 实现每个历元的自适应优化策略。

2) 在伪距修正的历元, 基于每颗卫星故障检测标记结果自适应选择该颗卫星的处置策略, 包括剔除、不修正或修正伪距误差, 实现每颗卫星的自适应质量控制策略。

1　算法框架

本文算法如图 1 所示。离线部分: 先在建模区域内进行多日 GNSS 众包数据采集与处理。选取方位角、仰角和C/N_0作为特征输入, 在多径误差标定后根据星座和格网进行数据集划分, 基于贝叶斯优化集成回归模型分别训练每个格网的回归模型用于在线伪距误差预测。在线测试部分: 利用当前历元惯导测量的位置计算当前每颗卫星的仰角和方位角, 以及根据该位置搜索确定当前历元所在格网并调用该格网的全球卫星定位系统(global positioning system, GPS)和北斗卫星导航系统(BeiDou navigation system, BDS)多径误差预测模型。输入每颗卫星的仰角、方位角可得到相应卫星的伪距

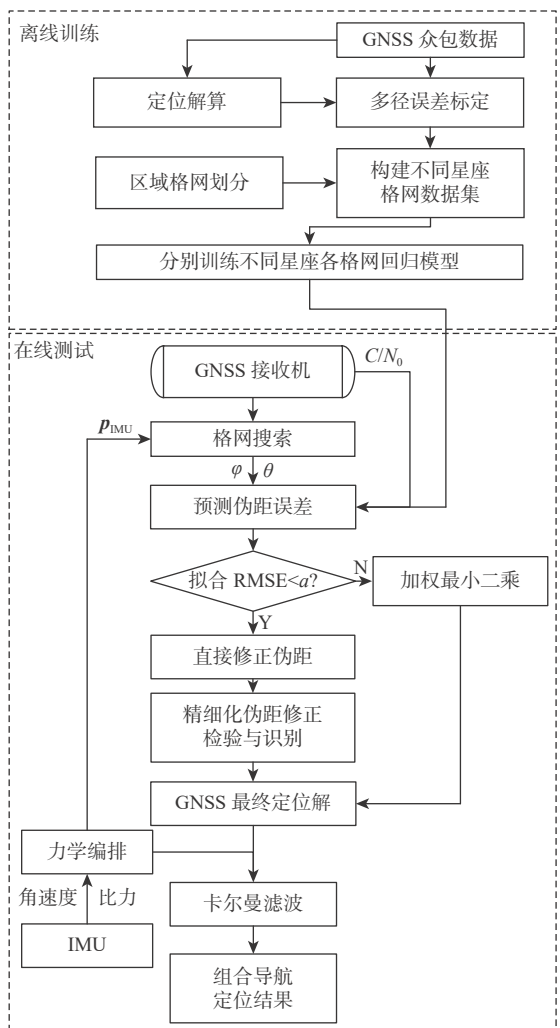


图 1　本文算法结构

Fig. 1　Scheme of the proposed algorithm

误差预测值。为解决城市恶劣多变环境中多径误差预测精度不稳定的问题，在基于格网建模的GNSS修正优化定位算法中，对预测伪距误差的应用加以精细化的自适应策略，以提升算法的可用性与定位性能。由于每个格网基于集成回归模型分别进行训练，因此每个格网选用的回归模型不同，训练集数量不同，模型的拟合精度和训练效果也并不相同，需要根据格网自身的拟合精度进行不同策略的自适应。当前历元落在拟合精度较低的格网内时，预测精度不足以进行直接伪距修正，只利用预测的伪距误差作为权重计算依据进行加权最小二乘定位。当所在格网精度较高时，对所有卫星直接进行伪距修正，为避免个别卫星修正前后出现故障情况，基于伪距残差进行故障识别，构建精细化修正策略，实现不同卫星的模型自适应质量控制。最终的GNSS定位解输入EKF，与IMU进行组合导航得到组合导航定位结果。

2 基于格网的伪距误差预测模型

2.1 基于特征变量的格网数据集

模型采用正六边形（蜂窝）格网形状对城市建模区域进行分割划分，每个格网内都有按既定路线重复多日采集到的GNSS数据集。卫星周期性经过特定位置时，与同一位置的接收机将呈现相似的几何形状[15]。因为在一周期的同一时间，接收机接近原始位置，且周围环境没有太大变化的情况下，多径误差也是相似的。因此，基于划分的道路格网和原始GNSS测量，可以使用先验数据建立准确的伪距误差预测模型。当用户位置落在某个格网内，输入某颗卫星的仰角、方位角和C/N0，此格网能够将对应的伪距误差预测值提供给用户。由于GPS卫星轨道周期为24 h，北斗轨道周期为7 d，因此GPS训练数据集为连续多日数据，BDS训练数据集为每隔7 d的数据。时空匹配越准确，信号传播路径的相似度就越高，伪距误差预测精度越高。时间和空间划分过于精细会导致模型数量过多，运算效率降低。因此，需要将格网时空分辨率调整到合适的大小，对于空间层面，六边形网格不严格区分路段、车道、方向等细节，其大小可以根据GNSS数据质量和应用要求进行更改，权衡匹配精准度和计算复杂度，本文建立的区域格网模型边长为10 m，时间分辨率为2 h。由于GPS和北斗的卫星轨道重复周期不同，本文训练集和模型根据星座类型进行划分，对于每个格网，2种星座分别训练和执行对应的规则。图2为格网模型示意。

基于单一特征变量的信号分类或多径误差预测在城市动态环境中的准确率难以保证，因此该模型特征变量选取为卫星载噪比、仰角和方位角。载噪比C/N_0表征信号强度，NLOS/MI信号的载噪比由于反射等原因通常低于LOS信号，因此C/N_0可用来作为输入变量之一预测伪距误差。而卫星仰角θ越大，信号越不容易被遮挡，因此θ能够作为特征变量之一。在某个格网内，不同的卫星方位上可能有不同的建筑物遮挡或反射信号，因此，卫星方位角φ与伪距误差有相关性。

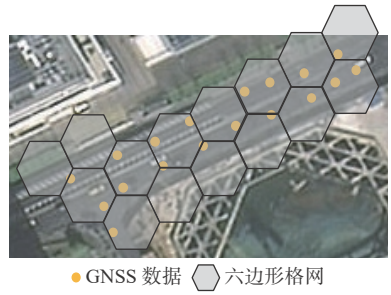


图2 格网模型示意

Fig. 2 Grid model diagram

伪距误差的值作为离线训练阶段的标签，伪距中的卫星钟差，接收机钟差，电离层误差和对流层误差都可以通过模型进行修正，修正后的伪距如下：

$$\rho^{\mathrm{c}} = r + c(\Delta t_{\mathrm{u}} - \Delta t_{\mathrm{s}}) + \Delta D^{\mathrm{trop}} + \Delta D^{\mathrm{iono}} + \varepsilon \tag{1}$$

$$r = \sqrt{(x_{\mathrm{s}} - x_{\mathrm{u}})^2 + (y_{\mathrm{s}} - y_{\mathrm{u}})^2 + (z_{\mathrm{s}} - z_{\mathrm{u}})^2} \tag{2}$$

式中：r为卫星和接收机间的几何距离；$(x_{\mathrm{s}},y_{\mathrm{s}},z_{\mathrm{s}})$和$(x_{\mathrm{u}},y_{\mathrm{u}},z_{\mathrm{u}})$分别为卫星和接收机在地心地固坐标系（earth-centered earth-fixed, ECEF）下的坐标；c为光速；Δt_{s}、Δt_{u}、ΔD^{trop}和ΔD^{iono}分别为修正后的卫星钟差、接收机钟差、对流层误差和电离层误差的残差；ε包含多径误差和观测噪声。

$$\Delta\rho = \rho^{\mathrm{c}} - r = c(\Delta t_{\mathrm{u}} - \Delta t_{\mathrm{s}}) + \Delta D^{\mathrm{trop}} + \Delta D^{\mathrm{iono}} + \varepsilon \tag{3}$$

式中：$\Delta\rho$为修正后的伪距误差，即为模型要预测的伪距误差，其他残差项和观测噪声非常小，因此主要为多径误差。

2.2 基于贝叶斯优化集成回归的多径误差预测模型

由于每个格网都要训练一个回归模型，为避免模型固定和手动调参，提高模型精度并减少时间成本，选用自动优化的贝叶斯优化集成回归模型进行拟合，更高效地对每个模型进行自动模型选择和参数调节。所有的格网模型均为一种参数最优的袋装树模型或参数最优的最小二乘提升模型（least-squares boosting, LSBoost），袋装树模型参数包括学习周期数和叶子节点含有的最少样本，最小二乘提升模型参数包括学习周期数、学习率和叶子节点含

有的最少样本。

在伪距误差建模中, 对每个格网数据集而言, 训练集中的一个样本为

$$x_i = [C/N_{0i}, \theta_i, \varphi_i] \tag{4}$$

$$y_i = \Delta\rho_i \tag{5}$$

式中: i 为每个格网训练集的样本序号, $i = 1,2,\cdots,I$, I为训练集的样本数量; $\Delta\rho_i$为标定的伪距误差。训练数据集为

$$T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_I, y_I)\} \tag{6}$$

集成回归模型为集合袋装树模型时, 对训练样本集进行随机抽样, 为训练单个回归树生成一个训练样本子集, 共训练M棵回归树, 最终输出M棵回归树预测值的平均值$y^{\text{bag}}(x)$:

$$y^{\text{bag}}(x) = \frac{1}{M}\sum_{m=1}^{M} y_m^{\text{bag}}(x) \tag{7}$$

式中: y_m^{bag}为第m棵回归树的伪距误差预测值。

集成回归模型为 LSBoost 方法时, 基于最小二乘构建损失函数, 共进行L次迭代, 每一次迭代计算残差为

$$r_i = y_i - F_{l-1}(x_i) \tag{8}$$

式中: $F_{l-1}(x)$为第$l-1$次迭代的伪距误差预测模型。

最小化损失函数:

$$(\alpha_l, a_l) = \arg\min_{a,\alpha}\sum_{i=1}^{N}[r_i - \alpha h(x_i; a)]^2 \tag{9}$$

式中: α_l和a_l能使损失函数达到最小值。

$$F_l(x) = F_{l-1}(x) + \eta\alpha_l h(x; a_l) \tag{10}$$

式中: η为学习率, 能够防止过拟合, $0 < \eta \leqslant 1$; $F_l(x)$为第l次迭代的伪距误差预测模型。

迭代L次后, 输出模型$F_L(x)$, 为最终训练生成的伪距误差预测模型。

3　格网模型辅助的综合质量控制算法

根据上一历元组合导航定位结果与这一历元惯导的测量值可以推算出当前历元惯导的定位结果, 将该位置作为该历元初始定位结果, 判断该历元所在的格网编号。设μ_j为第j个格网的拟合精度, 即拟合均方根误差(root mean square error, RMSE), 当μ_j过大时, 该格网内的伪距误差预测精度较差, 不能对预测得到的结果直接在伪距中进行修正, 而选用相对不敏感的加权最小二乘定位, 基于预测得到的伪距误差进行权重矩阵的构造, 通过加权最小二乘得到 GNSS 定位解。当格网精度较高时, 则是对伪距误差直接进行修正, 但针对个别卫星存在预测不准确及故障问题, 构建基于伪距残差的故障识别及精细化修正策略, 对不同的卫星分别采用修正、不修正和剔除这 3 种方法。

3.1　格网建模辅助的加权最小二乘定位

最小二乘在实际定位过程中, 由于不同卫星信号的传播路径和受到的干扰不同, 每颗卫星的测量误差也不同, 因此在定位解算时根据卫星测量的精度对不同卫星赋予不同的权重, 提升观测质量较高的卫星在定位中的权重, 降低精度较差的卫星信号的权重, 从而提高 GNSS 的定位精度[16]。

最小二乘观测方程为

$$\boldsymbol{G}\Delta\boldsymbol{x} = \boldsymbol{b} \tag{11}$$

式中: $\Delta\boldsymbol{x}$为方程解得的修正量向量, 包括三维位置修正值和接收机钟差; $\boldsymbol{b}$为伪距观测值残差向量, 每个元素为对应卫星的伪距观测值与接收机定位位置反演得到的伪距之差; $\boldsymbol{G}$为卫星方向余弦矩阵。该方程解为

$$\Delta\boldsymbol{x} = (\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{b} \tag{12}$$

设权重矩阵为$\boldsymbol{W}$, $\boldsymbol{W}$是一个$n \times n$的对角阵, 对角线元素为第k颗卫星的权重w_k, 表示为

$$\boldsymbol{W} = \mathrm{diag}(w_1, \cdots, w_n) \tag{13}$$

通过基于格网的伪距误差预测模型得到第 k 颗卫星的伪距误差e_k, 伪距误差越大, 卫星的权重应该越低, 提出基于预测伪距误差分段加权模型如下:

$$w_k = \begin{cases} 1 & e_k \leqslant c_0 \\ 1/e_k^2 & e_k > c_0 \end{cases} \tag{14}$$

当预测伪距误差不大于阈值c_0时, 认为该颗卫星的伪距不需要修正, 权重设为 1; 当预测伪距误差超过阈值时, 权重设为伪距误差平方的倒数。则观测方程变为

$$\boldsymbol{W}\boldsymbol{G}\Delta\boldsymbol{x} = \boldsymbol{W}\boldsymbol{b} \tag{15}$$

加权最小二乘解为

$$\Delta\boldsymbol{x} = (\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{b} \tag{16}$$

式中: $\boldsymbol{C}$为最小二乘解中的权重矩阵, $\boldsymbol{C} = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}$。

3.2　基于伪距残差的卫星伪距修正检验与识别

对于直接进行伪距修正的历元, 为避免个别卫星的伪距误差预测误差较大造成定位精度恶化, 对卫星伪距修正前后进行故障检验与识别。伪距残差向量可以反映出卫星伪距测量精度, 当存在故障时, 某颗卫星含有的粗差会使伪距残差值变大。

由最小二乘解可得最小二乘残差矢量$\boldsymbol{v}$可表示为[17]

$$\begin{aligned}\boldsymbol{v} &= \left[\boldsymbol{I} - \boldsymbol{G}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\right]\boldsymbol{b} = \left[\boldsymbol{I} - \boldsymbol{G}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\right](\boldsymbol{G}\Delta\boldsymbol{x} + \boldsymbol{\varepsilon}) = \\ &\left[\boldsymbol{I} - \boldsymbol{G}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\right]\boldsymbol{\varepsilon}\end{aligned} \tag{17}$$

$$\boldsymbol{Q}=[\boldsymbol{I}-\boldsymbol{G}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}] \tag{18}$$

式中：$\boldsymbol{\varepsilon}$为考虑噪声情况下的观测误差向量；$\boldsymbol{Q}$为伪距残差向量的协因数阵。

利用巴尔达数据探测法，采用最小二乘残差矢量构建检验统计量为[18]

$$\boldsymbol{d}_k=\frac{|\boldsymbol{v}_k|}{\sigma_0\sqrt{\boldsymbol{Q}_{kk}}} \tag{19}$$

式中：$\boldsymbol{d}_k$为第k颗卫星的检验统计量；σ_0为伪距残差的标准差。将每颗卫星的检验统计量与检测门限T_d比较，如果大于该门限，则判定该颗卫星故障，标记f_{flag}=1。

分别对伪距修正前后的GNSS观测值进行最小二乘定位解算，并分别计算每颗卫星的故障识别统计量。设伪距修正前第k颗卫星的故障检验统计量为$\boldsymbol{d}_k$，故障标记为f_{flag}。伪距修正后第k颗卫星的故障检验统计量为$\boldsymbol{d}'_k$，故障标记为f'_{flag}。根据故障检验识别结果，对每颗卫星均采取修正策略如图3所示，当修正后卫星无故障时，用修正定位解作为最终定位解。若修正前后卫星均故障，则将该颗卫星剔除，不参与定位。若修正前卫星为正常卫星，伪距修正后发生故障，则说明错误的伪距误差预测结果污染了卫星质量，不应用修正结果作为定位结果，直接采用伪距单点定位结果为最终定位解。

	f_{flag}=1	f_{flag}=0
f_{flag}=1	剔除	修正
f_{flag}=0	不修正	修正

图3 基于故障检验与识别的精细伪距修正策略

Fig. 3 Refined pseudorange correction strategy based on fault detection and identification

4 实 验

4.1 实验设置

为验证本文算法的有效性，进行实际数据车载实验。实验环境为南京市南京南站附近城市环境区域，部分路段为严重城市峡谷，选取实验建模区域及路径如图4所示，实验环境如图5所示。在算例数据集中，在16:00—18:00这一时间段进行数据采集，GPS星座训练数据集为连续多日数据，BDS训练数据集为每隔7 d的数据，测试数据则选在北斗轨道重复周期日。最终该实验区域共训练了240个GPS格网模型，240个BDS格网模型。

图4 实验路线

Fig. 4 Experiment region

图5 实验环境

Fig. 5 Experiment environment

为验证本文算法的性能，本文实验分别对2种对比算法和本文算法进行测试，实验算法分别为算法1：基于EKF的GNSS/IMU组合导航算法；算法2：基于格网伪距误差预测的GNSS/IMU组合导航算法；本文算法：格网误差建模辅助的GNSS/IMU组合导航综合质量控制算法。3种对照算法及本文算法内容如表1所示。

表1 实验算法描述

Table 1 Experiment algorithm description

算法	算法内容及步骤
算法1	基于传统卡尔曼滤波的GNSS/IMU松组合
算法2	1. 基于格网模型预测值对全部卫星进行伪距修正后定位； 2. 与IMU进行松组合
本文算法	1. 根据格网拟合精度选择当前历元进行伪距修正或加权最小二乘； 2. 根据卫星故障检验与识别结果选择每颗卫星是否进行剔除或修正； 3. 与IMU进行松组合

实验车平台及其设备如图6所示，GNSS接收机的型号为BDStar Navigation C520-AT，采样率为10 Hz。IMU的型号为STIM300，采样率为125 Hz，其参数如表2所示。参考轨迹是通过使用Novatel

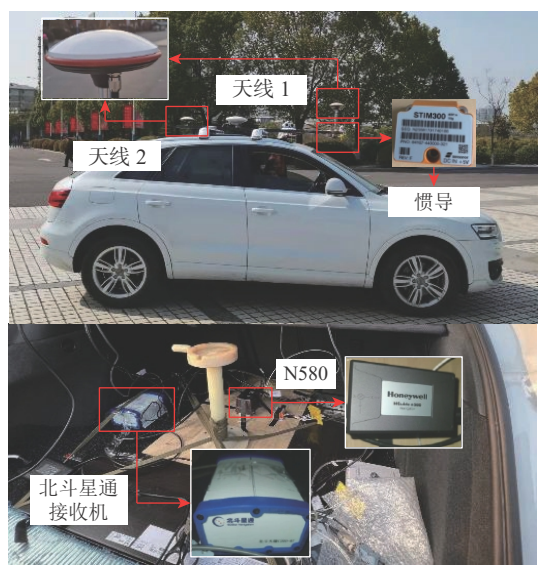


图6 车载实验平台

Fig. 6 Vehicle test platform

Inertial Explorer 软件对高精度的 HGuideN580inertial sensor 的数据进行后处理得到的, N580 相关性能参数如表 3 所示。N580 连接天线 1, BDStar Navigation C520-AT 接收机连接天线 2, 天线型号均为 ZYACF-S806。

表 2　STIM300 参数

Table 2　Parameters of STIM300

参数	数值
加速度计零偏不稳定性/mg	0.05
加速度计随机游走/(m·s^{-1}·h$^{-1/2}$)	0.06
陀螺仪零偏不稳定性/((°)·h^{-1})	0.5
陀螺仪随机游走/((°)·h$^{-1/2}$)	0.15

表 3　HGuideN580 参数

Table 3　Parameters of HGuideN580

参数	数值
加速度计零偏不稳定性/mg	0.025
加速度计随机游走/(m·s^{-1}·h$^{-1/2}$)	0.03
陀螺仪零偏不稳定性/((°)·h^{-1})	0.25
陀螺仪随机游走/((°)·h$^{-1/2}$)	0.04

4.2　实验验证与分析

测试数据的 GPS 和 BDS 卫星数目如图 7 所示, 位置精度因子(position dilution of precision, PDOP)值如图 8 所示, 在 400～500 s, 卫星数量较少, 且出现一些历元 PDOP 值过大的问题。

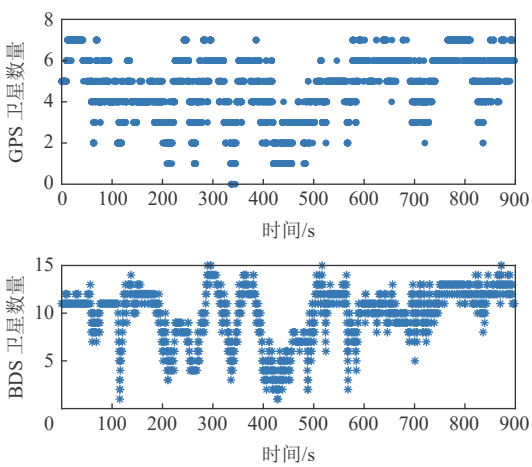


图 7　卫星数量

Fig. 7　Number of satellites

图 9 为 3 种算法定位结果在地图上的轨迹, 图 9(a) 为整体轨迹图, 可以看出, 与其他组合导航算法相比, 本文算法更接近参考值, 尤其在图 9(d) 放大部分, 由于高架的遮挡, 算法 1 中 GNSS 的严重多径使得组合导航结果产生了几十米的误差, 经过伪距误差修正, 算法 2 和算法 3 均大幅修正了定位误差。在图 9(b) 部分定位误差较小的部分, 本文

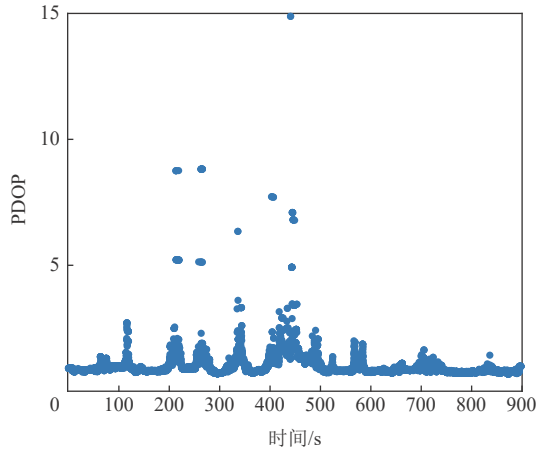


图 8　PDOP 值

Fig. 8　PDOP value

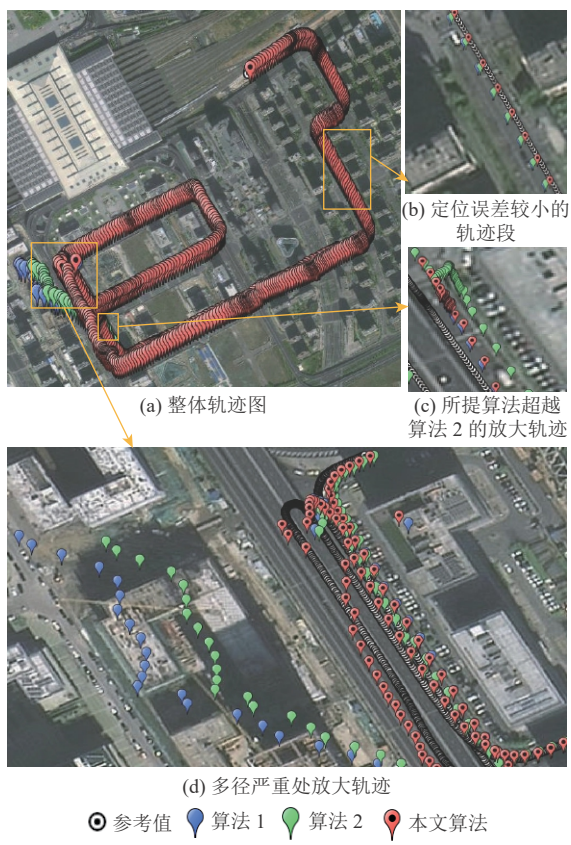


图 9　3 种算法定位轨迹

Fig. 9　Positioning trajectories of the three algorithms

算法依然能够提升定位精度。而在图 9(c) 部分, 算法 2 由于模型错误的伪距误差预测值污染了原本良好的伪距观测, 恶化了定位精度, 而本文算法解决了这一问题。

实验中 3 种算法的定位精度及提升比例如表 4 所示。算法 1 在东方向和地方向定位误差较大, 水平位置误差和三维位置误差分别高达 10.61 m 和 22.88 m。基于格网伪距误差建模与修正的算法 2 的水平和 3D 位置误差分别提升了 43.71% 和 44.16%, 这充分说明了格网伪距误差预测模型对于缓解多

表 4 3 种算法定位精度对比

Table 4 Positioning accuracy comparison of the three algorithms

类型	RMSE/m			提升率1/%	提升率2/%
	算法1	算法2	本文算法		
北	2.93	2.82	2.37	3.75	19.11
东	10.20	5.27	4.72	48.33	53.75
地	20.27	11.30	5.47	44.25	64.92
水平	10.61	5.97	5.28	43.71	50.23
3D	22.88	12.78	7.60	44.16	66.77

注：提升率1是算法2相对算法1的提升比例、提升率2是本文算法相对算法1的提升比例。

径误差，提高城市环境定位精度的有效性。而本文算法在北东地 3 个方向上的定位精度分别达到了 2.37 m、4.72 m 和 5.47 m，水平和三维位置误差分别为 5.28 m 和 7.60 m，相较算法 2 定位精度进一步提升，垂直方向尤为显著。这是因为本文算法实现了算法对格网质量和卫星故障的自适应，对于模型的应用增加了精细化策略和约束检验。本文算法相较传统 GNSS/IMU 组合导航算法，在水平方向上定位精度提升了 50.23%，3D 定位精度提升了 66.77%。

图 10 为北东地 3 个方向上 3 种算法的定位误差对比，图 11 为 3 种算法的水平和 3D 定位误差。

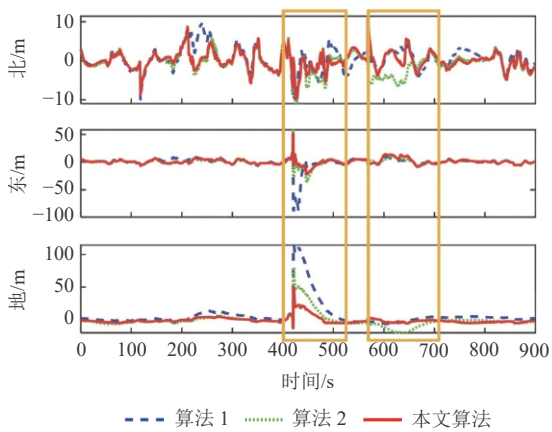


图 10 北东地方向定位误差

Fig. 10 Positioning error in NED

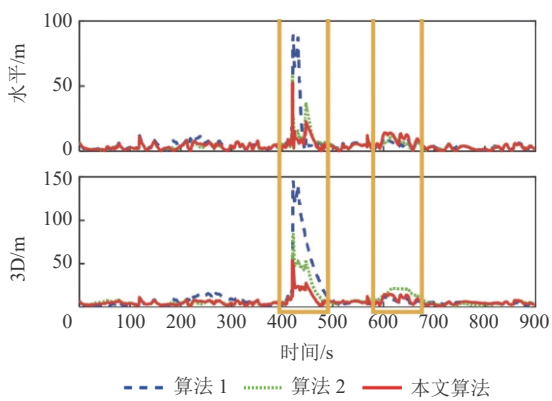


图 11 水平和 3D 定位误差

Fig. 11 Horizontal and 3D positioning error

在第 1 个方框 400～500 s 内算法 2 由于有效的伪距误差预测模型降低了定位误差，而本文算法由于针对每颗卫星修正前后故障情况进行了自适应策略，进一步提高了定位精度。在第 2 个方框 600～700 s，由于格网模型预测不准确，算法 2 由于错误的修正值污染了伪距测量值，降低了定位精度，而本文算法由于其检验和自适应策略能够解决这一问题。

5 结 论

1）基于格网伪距误差建模的 GNSS 伪距修正模型能够一定程度上缓解多径效应，修正伪距中的多径误差，提高定位精度。

2）提出的格网辅助的综合质量控制算法能够进一步提升组合导航定位精度，根据格网拟合精度的不同进行伪距修正和加权最小二乘 2 种策略的切换。在伪距修正过程中，基于伪距残差故障检验与识别进行精细伪距修正，降低故障卫星和错误的修正值对定位结果的影响。

3）本文算法在城市环境中实地测试中具备有效性和优越性，相较传统 GNSS/IMU 组合导航 3D 定位精度提高了 66.77%，相较单纯基于格网伪距建模修正的 GNSS/IMU 组合导航算法 3D 定位精度提升了 30.53%。

4）本文提出的基于格网伪距误差建模综合质量控制的 GNSS/IMU 组合导航算法可以迁移至载波相位观测量的误差建模，以及面向无人机的立体格网模型构建。

5）现有的格网模型时空局限性较大，后续研究中，可以进行格网内不同方位的遮蔽角信息或者遮蔽物高程信息提取，遮蔽角信息或者遮蔽物高程信息属于空间信息模型或环境建图，可以通过使用 3D 城市模型或激光雷达、相机建图等手段获得。将遮蔽角信息、建筑物高度信息等空间信息作为输入参与到模型构建，模型的应用将不受区域、时间的影响，在任意时间场景都能够使用，提升模型适用性和效率。

参考文献（References）

[1] ZHU N, MARAIS J, BÉTAILLE D, et al. GNSS position integrity in urban environments: a review of literature[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(9): 2762-2778.

[2] 冯晓超，程晓滨，赵珂. GNSS 接收机抗多径技术[J]. 电讯技术，2010, 50(8): 180-184.
FENG X C, CHENG X B, ZHAO K. Anti-multipath technology for GNSS receivers[J]. Telecommunication Engineering, 2010, 50(8): 180-184(in Chinese).

[3] Kumar S, Moore K B. The evolution of global positioning system (GPS) technology[J]. Journal of Science Education and Technology,

2002, 11: 59-80.

[4] 邵晨, 曾庆化, 邱文旗, 等. GNSS 多径抑制基带处理算法综述[J]. 导航定位与授时, 2022, 9(5): 1-14.
SHAO C, ZENG Q H, QIU W Q, et al. Overview of GNSS multipath mitigation baseband processing algorithm[J]. Navigation Positioning and Timing, 2022, 9(5): 1-14(in Chinese).

[5] 李岩, 董玮, 徐鹏程, 等. GNSS 抑制多径技术研究进展[J]. 汽车电器, 2022(7): 55-57.
LI Y, DONG W, XU P C, et al. Research status and prospect of GNSS multi-path signal suppression technology[J]. Auto Electric Parts, 2022(7): 55-57(in Chinese).

[6] YOZEVITCH R, MOSHE B B, WEISSMAN A. A robust GNSS LOS/NLOS signal classifier[J]. Navigation-Journal of the Institute of Navigation, 2016, 63(4): 429-442.

[7] QUAN Y M, LAU L, ROBERTS G W, et al. Convolutional neural network based multipath detection method for static and kinematic GPS high precision positioning[J]. Remote Sensing, 2018, 10(12): 2052.

[8] GUERMAH B, EL GHAZI H, SADIKI T, et al. A robust GNSS LOS/multipath signal classifier based on the fusion of information and machine learning for intelligent transportation systems[C]// Proceedings of the 2018 IEEE International Conference on Technology Management, Operations and Decisions. Piscataway: IEEE Press, 2018: 94-100.

[9] SUN R, WANG G Y, ZHANG W Y, et al. A gradient boosting decision tree based GPS signal reception classification algorithm[J]. Applied Soft Computing, 2020, 86: 105942.

[10] 刘成, 李芳. 不同卫星定位加权方法的比较与分析[J]. 测绘科学, 2018, 43(8): 39-44.
LIU C, LI F. Comparison and analysis of different GNSS weighting methods[J]. Science of Surveying and Mapping, 2018, 43(8): 39-44(in Chinese).

[11] 付红波. 城市复杂环境下不同加权模型对动态 SPP 定位性能影响分析[J]. 测绘通报, 2023(3): 139-143.
FU H B. Analysis of the influence of different weighted models on dynamic SPP positioning performance in urban complex environment[J]. Bulletin of Surveying and Mapping, 2023(3): 139-143(in Chinese).

[12] SUN R, WANG G Y, CHENG Q, et al. Improving GPS code phase positioning accuracy in urban environments using machine learning[J]. IEEE Internet of Things Journal, 2021, 8(8): 7065-7078.

[13] SUN R, ZHANG Z X, CHENG Q, et al. Pseudorange error prediction for adaptive tightly coupled GNSS/IMU navigation in urban areas[J]. GPS Solutions, 2021, 26(1): 28.

[14] SUN R, FU L X, CHENG Q, et al. Resilient pseudorange error prediction and correction for GNSS positioning in urban areas[J]. IEEE Internet of Things Journal, 2023, 10(11): 9979-9988.

[15] SONG B, MA X D, ZHANG S B. Analysis of BDS satellite orbit periodicity[J]. GNSS World of China, 2016, 41(5): 47-50.

[16] 程琦, 胡杰, 王均晖, 等. 基于数据驱动的无人机加权最小二乘定位算法[J]. 指挥信息系统与技术, 2020, 11(6): 76-80.
CHENG Q, HU J, WANG J H, et al. Weighted least square positioning algorithm of UAV based on data drive[J]. Command Information System and Technology, 2020, 11(6): 76-80(in Chinese).

[17] 韩清清. GNSS 接收机自主完好性监测算法优化及性能分析[D]. 西安: 长安大学, 2020.
HAN Q Q. Optimization and performance analysis of autonomous integrity monitoring algorithm for GNSS receiver[D]. Xi'an: Changan University, 2020(in Chinese).

[18] 施颖. 面向非高斯噪声条件的列车卫星定位完好性监测方法研究[D]. 北京: 北京交通大学, 2019.
SHI Y. Research on integrity monitoring method for satellite based train positioning under non-gaussian noise conditions[D]. Beijing: Beijing Jiaotong University, 2019(in Chinese).

Grid error modeling aided GNSS/IMU integrated navigation comprehensive quality control algorithm

DAI Yeying，SUN Rui*，DENG Siyu，JI Li，WANG Yuanyuan，HUANG Xuedong

(College of Civil Aviation，Nanjing University of Aeronautics and Astronautics，Nanjing 211106，China)

Abstract：In the complex urban environment, Global navigation satellite system (GNSS) signals are prone to non-line-of-sight reception (NLOS) and multipath interference (MI) due to the occlusion and reflection of obstacles such as tall buildings. The positioning accuracy and reliability are seriously reduced, which cannot meet the user's high precision and reliability positioning, navigation and timing (PNT) service requirements. Signal classification and multipath modeling methods based on machine learning and data-driven are of great significance for alleviating GNSS multipath effect and improving positioning accuracy in urban areas. But the accuracy, efficiency and adaptivity of such models still need to be improved. In this paper, a grid error modeling aided GNSS/Inertial measurement unit (IMU) integrated navigation comprehensive quality control algorithm is proposed. Besides the grid pseudorange error modeling, it proposes a refined comprehensive quality control strategy based on grid fitting accuracy and satellite fault detection, thus optimizes the performance of GNSS/IMU integrated navigation in complex urban environments. The field test in urban environment shows that compared with the traditional GNSS/IMU integrated navigation algorithm, the horizontal and 3D positioning accuracy of the proposed algorithm are improved by 50.23% and 66.77%, respectively, and compared with the grid pseudorange error modeling algorithm, they are improved by 11.56% and 40.53%, respectively.

Keywords：GNSS/IMU integrated navigation；pseudorange error modelling；quality control；weighted least square；fault detection

Received：2023-07-27；**Accepted**：2023-09-01；**Published Online**：2023-12-04 16：02

URL：link.cnki.net/urlid/11.2625.V.20231204.0907.001

Foundation items：National Natural Science Foundation of China (42222401,42174025,41974033); Ministry of Industry and Information Technology Special Project of China (TC220A04A-79); Jiangsu provincial Six Talent Peaks Project (KTHY-014); Natural Science Foundation of Jiangsu Province (BK20211569); the Fundamental Research Funds for the Central Universities (xcxjh20220722)

* **Corresponding author**. E-mail：rui.sun@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0498

基于直接概率积分的混合可靠性分析方法

朱启滔，李洪双*
(南京航空航天大学 航空学院，南京 210016)

摘　　要：针对证据变量和随机变量混合不确定性以及极限状态函数复杂并存的情况，提出了一种基于直接概率积分法的证据-概率混合可靠性分析方法。该方法从证据理论的概率解释入手，采用证据变量均匀化方法，将证据变量转换为随机概率变量，达到混合变量归一化目的，进而通过直接概率积分法进行结构可靠性分析。最后通过3个算例对所提方法的性能进行验证，结果表明：与现有方法相比，所提方法在计算精度相近的情况下，极大地提高了计算效率。

关 键 词：证据理论；直接概率积分法；结构可靠性分析；混合不确定性；随机概率变量
中图分类号：V221+.3; TB553
文献标志码：A　　**文章编号**：1001-5965(2025)09-3183-10

实际工程设计中，结构可靠性分析起着举足轻重的作用。传统的随机可靠性分析是在概率论和数理统计的基础上得以发展的。然而当极限状态函数过于复杂时，采用传统的随机抽样方法进行结构可靠性分析成本过于昂贵，效率低下。Chen 等[1-2]提出了直接概率积分法(direct probability integral method, DPIM)，该方法高效地计算静力和动力作用下的线性和非线性结构的响应概率密度函数，进而计算得出结构的可靠性分析结果。

近年来，直接概率积分法得到了极大发展。刘思琦[3]将直接概率积分法应用到车-(轨)-桥耦合系统中，对其进行随机动力学分析。李小兰[4]在直接概率积分法的基础上，建立了求解含多设计点的结构静/动力可靠度优化设计问题的统一框架。李石[5]将直接概率积分法和一种新的明确时间集成算法结合，进行了大型结构的随机振动分析。高天骁等[6]采用直接概率积分法，对重载铁路梁进行了可靠性分析。Meng 等[7]将直接概率积分法和指数凸模型结合，进行了混合不确定性传播和可靠性研究。

1977年，Dempster 等[8]最先提出了证据理论，其学生 Shafer[9]在其基础上进行了研究和发展，进而逐步形成了较为完善的证据理论，又称 Dempster-Shafer 理论或者 D-S 证据理论。证据理论是认知不确定性理论的重要组成部分，可以较好地针对复杂系统进行不确定性建模，近年来，证据理论已被引入结构可靠性分析领域，成为一个热点研究方向[10]。

Oberkampf 等[11]总结了证据理论应用于可靠性分析中的优点和缺点。Bae 等[12]提出了一种基于证据理论的结构静态和动态可靠性分析模型，并且构建了有关的代理模型。Bai 等[13]将证据理论代理模型相关技术引入可靠性应用。Zhang 等[14]提高了证据理论可靠性分析代理模型的精度。Wang 等[15]将证据理论和逆高斯退化结合，提出了一种可靠性分析方法。

当概率变量和证据变量同时存在时，就演变出证据-概率混合可靠性分析问题，将其应用到具有复杂结构和变量的实际工程问题，具有极高的实用价值。目前，关于这方面的研究，国内外成果尚少，还处于起步阶段。最初，Du[16]建立了一种统一框架下的概率-证据混合问题的可靠性灵敏度分析。姜潮等[17]将证据理论与一次二阶矩法融合，提出了一种高效的概率-证据混合可靠性分析方法。Tao 等[18]

收稿日期：2023-07-27；录用日期：2023-09-04；网络出版时间：2023-09-15 15:01
网络出版地址：link.cnki.net/urlid/11.2625.V.20230914.2020.001
*通信作者. E-mail：hongshuangli@nuaa.edu.cn

引用格式：朱启滔，李洪双. 基于直接概率积分的混合可靠性分析方法 [J]. 北京航空航天大学学报，2025，51（9）：3183-3192.
ZHU Q T，LI H S. A mixed reliability analysis method based on direct probability integral [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（9）：3183-3192 （in Chinese）.

提出了和样本点法有关的一种包含概率变量和证据变量的不确定性多目标可靠性分析方法。梁云飞等[19]基于Kriging代理模型，提出了一种概率-证据混合可靠性分析方法，并将其应用到工业机器人定位精度可靠性领域中。曹亮等[20]将概率-证据混合可靠性问题与子集模拟方法相结合，有效解决了小概率可靠性问题。张梦[21]将概率-证据混合可靠性与最基本的蒙特卡洛方法联系，对桥梁进行了可靠性评估分析。孟增等[22]提出了一种适用于概率-证据混合可靠性的自适应共轭分析方法，一定程度上丰富了其发展。

综上所述，已有研究虽然在一定程度上解决随机不确定性和认知不确定性混合问题，但难以解决极限状态函数过于复杂的可靠性问题。本文针对上述问题，提出了一种基于直接概率积分法的证据-概率混合可靠性分析方法，旨在解决具有复杂结构极限状态函数的证据-概率变量并存的实际工程问题。该方法在保持计算精度的同时，显著提高了计算效率。

1 基于证据理论的可靠性分析

证据理论是基于辨识框架所建立的。辨识框架是所有可能发生的相互独立事件的集合，用符号Θ表示。

基本概率分配(basic probability assignment, BPA)是指确定了辨识框架后，通过专家估计或者测量计算等方法对辨识框架中的每一个子集进行概率赋值。对于一个集合A的基本概率分配，记作$m(A)$，表示对集合A的信任支持程度，m为识别框架上的基本概率分配函数。当$m(A)>0$时，A为焦元。BPA具有以下基本性质：

$$m(A)\geqslant 0 \qquad A\in 2^{\Theta} \tag{1}$$

$$m(\varnothing)=0 \tag{2}$$

$$\sum_{A\subseteq 2^{\Theta}} m(A)=1 \tag{3}$$

对于一个命题的信任程度用一个区间上下界形式进行表示：[Bel(A),Pl(A)]。Bel(∗)和Pl(∗)分别为信任函数(belief function, Bel)和似然函数(plausibility function, Pl)：

$$\mathrm{Bel}(A)=\sum_{B\subseteq A} m(B) \tag{4}$$

$$\mathrm{Pl}(A)=\sum_{B\cap A\neq\varnothing} m(B) \tag{5}$$

证据理论在工程可靠性领域主要用于认知不确定性的建模。收集原始信息，通过专家估计或者实际测量等手段来构建证据变量识别框架以及每个焦元对应的BPA。

根据单个变量的识别框架，采用笛卡儿积D来定义联合识别框架Θ_x以及联合焦元D_x：

$$D=X_1\times X_2\times\cdots\times X_n=\{D_k=[X_{1i},X_{2i},\cdots,X_{ni}]\} \tag{6}$$

式中：$X_1,X_2,\cdots,X_n$分别为进行证据合成之后的证据变量的焦元；D_k为D的焦元；$X_{1i},X_{2i},\cdots,X_{ni}$分别为$X_1,X_2,\cdots,X_n$的焦元。考虑到证据变量之间的独立性，根据独立性假设，可以构建出多维联合基本可信度分配函数：

$$m(D)=\prod_{i=1}^{n} m(X_i) \tag{7}$$

对每个联合焦元进行极值分析获得可靠性分析中极限状态函数的极大值和极小值。分析联合焦元与极限状态面的位置关系来判断焦元是否失效。

根据给定的不同阈值，不断重复前面几个步骤，得到不同阈值下的可信度函数数值和似真度函数数值。将其进行连线，进而得到累积可信度函数(complementary cumulative belief function, CCBF)曲线和累积似真度函数(complementary cumulative plausibility function, CCPF)曲线。不同阈值下可靠度位于CCBF曲线和CCPF曲线之间。

2 证据-概率混合可靠性分析

当结构中同时存在证据变量和随机变量，原始输入变量表示为$\boldsymbol{Z}=(\boldsymbol{X},\boldsymbol{Y})$，其极限状态函数可以表示为$g(\boldsymbol{X},\boldsymbol{Y})$。其中$\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$表示$n$维输入证据变量，联合证据空间中的各焦元可计算其联合BPA；$\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_m)$表示$m$维随机输入变量。值得注意的是：各变量之间相互独立。

$g(\boldsymbol{X},\boldsymbol{Y})<c$时，结构失效，$c$为阈值。则失效概率表示为

$$P_{\mathrm{f}}=P_{\mathrm{r}}\{g(\boldsymbol{X},\boldsymbol{Y})<c\} \tag{8}$$

式中：P_{f}为对应结构的失效概率；P_{r}为对应结构的可靠度。

根据全概率公式，P_{f}展开为

$$P_{\mathrm{f}}=\sum_{i=1}^{n} P_{\mathrm{r}}\{g(\boldsymbol{X},\boldsymbol{Y})<c\}P_{\mathrm{r}}\{\boldsymbol{X}\} \tag{9}$$

其中焦元对应的联合BPA为

$$P_{\mathrm{r}}\{\boldsymbol{X}\}=m_{\boldsymbol{X}}(X_i) \tag{10}$$

由于无法得到$P_{\mathrm{r}}\{\boldsymbol{X}\}$的精确值，证据-概率混和可靠性通过其最大值和最小值进行计算：

$$p_{\mathrm{f}}^{\max}=P_{\mathrm{r}}\{g_{\min}(\boldsymbol{X},\boldsymbol{Y})<c\} \tag{11}$$

$$p_{\mathrm{f}}^{\min}=P_{\mathrm{r}}\{g_{\max}(\boldsymbol{X},\boldsymbol{Y})<c\} \tag{12}$$

最终失效概率的上下边界为似真度(Pl)和可信度(Bel), 表示为

$$\mathrm{Pl}(X,Y)=\sum_{i=1}^{n}P_{\mathrm{r}}\{g_{\min}(\boldsymbol{X},\boldsymbol{Y})<c\}\cdot m_{\boldsymbol{X}}(X_i) \tag{13}$$

$$\mathrm{Bel}(X,Y)=\sum_{i=1}^{n}P_{\mathrm{r}}\{g_{\max}(\boldsymbol{X},\boldsymbol{Y})<c\}\cdot m_{\boldsymbol{X}}(X_i) \tag{14}$$

3　基于直接概率积分法的混合可靠性分析

本文从证据理论均匀化入手, 将证据变量转换为随机变量, 结合直接概率积分法, 高效地开展结构可靠性工作。

3.1　证据变量均匀化

图 1 为包含互不重叠的 5 个焦元的证据变量均匀化过程。对于存在证据变量区间重叠的情况, 通过简单的累加即可, 如图 2 所示。

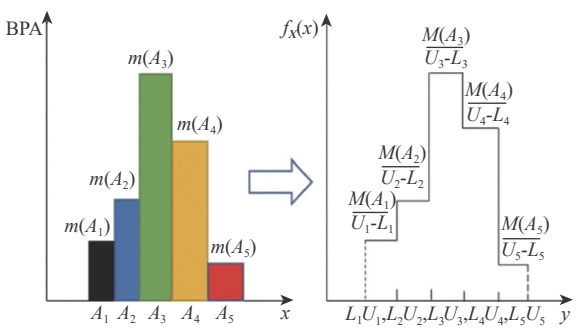


图 1　不重叠证据变量随机化过程

Fig. 1　Randomization process of non overlapping evidence variables

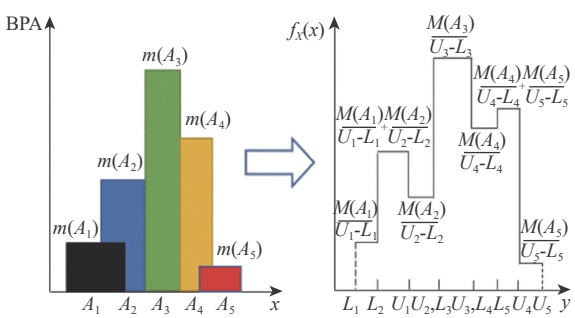


图 2　重叠证据变量随机化过程

Fig. 2　Randomization process of overlapping evidence variables

考虑到证据变量具有严重的离散性, 导致计算量巨大, 在实际工程中运用较为困难。本文采用均匀化方法[23], 将证据变量转换为具有概率密度函数的随机变量。均匀化公式如下:

$$f_X(x)=\sum_{j=1}^{n}\delta_j(x)m\left(A_j\right)/\left(U_j-L_j\right) \tag{15}$$

式中: f_X表示证据变量均匀化后的随机变量对应的概率密度函数; $\delta_j(x)$的取值与焦元的包含关系有关, 当$x\in A_j$时, $\delta_j(x)=1$, 否则, $\delta_j(x)=0$; A_j为证据变量X的焦元; n为焦元数; $m(A_j)$为每个焦元对应的 BPA; U_j和L_j分别为对应焦元基本概率分配的上界和下界。

3.2　概率密度积分方程的构建

经过证据变量均匀化后, 证据-概率混合变量可靠性分析问题仅包含随机变量。本文使用直接概率积分法进行结构可靠性分析。直接概率积分法源于概率守恒原理: 由于保守系统不引入其他新的随机因素, 原有的随机因素也没消失, 因此在随机系统的状态演化过程中, 概率保持不变。静力系统和动力系统中的概率守恒原理分别表示为

$$\int_{\Omega_z}P_{\boldsymbol{Z}}(z)\mathrm{d}z=\int_{\Omega_{\boldsymbol{\theta}}}P_{\boldsymbol{\theta}}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \tag{16}$$

$$\int_{\Omega_z}P_{\boldsymbol{Z}}(z,t)\mathrm{d}z=\int_{\Omega_{\boldsymbol{\theta}}}P_{\boldsymbol{\theta}}(\boldsymbol{\theta},t)\mathrm{d}\boldsymbol{\theta} \tag{17}$$

式中: $\boldsymbol{\theta}$和$\Omega_{\boldsymbol{\theta}}$分别表示输入随机向量和其对应的样本空间; $\boldsymbol{Z}$和$\Omega_{\boldsymbol{Z}}$分别表示输出随机向量和其对应的样本空间; $P_{\boldsymbol{\theta}}(\boldsymbol{\theta})$和$P_{\boldsymbol{\theta}}(\boldsymbol{\theta},t)$分别表示静力系统和动力系统中的原始输入采样点的概率密度函数; $P_{\boldsymbol{Z}}(z)$和$P_{\boldsymbol{Z}}(z,t)$表示静力系统和动力系统中响应量的概率密度函数。

静力系统或动力系统可以通过具体映射关系表示。实际工程中的映射G一般是隐式的, 很难求解。由于狄拉克δ函数具有筛选性质, 引入狄拉克函数, n维输入随机向量的概率密度函数变为

$$P_{\boldsymbol{\theta}}(\boldsymbol{\theta})=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}P_{\boldsymbol{\theta}}(\boldsymbol{s})\delta(\boldsymbol{s}-\boldsymbol{\theta})\mathrm{d}\boldsymbol{s} \tag{18}$$

式中: $\boldsymbol{\theta}$和$\boldsymbol{s}$为输入空间中的相互对称向量。

结合狄拉克δ函数与映射G, 有

$$p_{\boldsymbol{Z}}(z)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p_{\boldsymbol{\theta}}(\boldsymbol{\theta})\delta\left[z-\boldsymbol{g}(\boldsymbol{\theta})\right]\mathrm{d}\boldsymbol{\theta} \tag{19}$$

同理可得, 动力系统输出的概率密度函数为

$$p_{\boldsymbol{Z}}(z,t)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p_{\boldsymbol{\theta}}(\boldsymbol{\theta})\delta\left[z-\boldsymbol{g}(\boldsymbol{\theta},t)\right]\mathrm{d}\boldsymbol{\theta} \tag{20}$$

式 (19) 和式 (20) 分别表示静力系统和动力系统中的概率转换的积分形式, 称为概率密度积分方程 (probability density integral equation, PDIE)[3]。

3.3　直接概率积分法的数值求解

直接求解概率密度积分方程很困难, 可以采用一些如概率转换方法等间接数值计算方法。DPIM 可以高效地解决这一问题。这一方法包含 2 个重要部分: ①基于广义 F(generalized F, GF)选点技术进行概率空间剖分[24]; ②狄拉克δ函数平滑化。

采用 Voronoi 单元剖分技术对概率空间进行剖

分生成代表性区域$\Omega_{\boldsymbol{\theta},q}$:

$$\Omega_{\boldsymbol{\theta},q}=\left\{\boldsymbol{x}\in R^n:\|\boldsymbol{x}-\boldsymbol{\theta}_q\|\leqslant\|\boldsymbol{x}-\boldsymbol{\theta}_j\|\ \text{for all}\ j\right\} \tag{21}$$

式中: $q=1,2,\cdots,N$, $\boldsymbol{\theta}_q$表示子域$\Omega_{\boldsymbol{\theta},q}$的代表点; $\boldsymbol{\theta}_j$表示子域$\Omega_{\boldsymbol{\theta},j}$的代表点; N为代表点总数。

代表区域的概率测度被分配到对应的代表点, 每个代表点的赋得概率为

$$P_q=\int_{\Omega_{\boldsymbol{\theta},q}}P_{\boldsymbol{\theta}}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \tag{22}$$

实现了结构的概率密度积分方程和物理方程的解耦以及离散化求和。静力系统和动力系统的概率密度函数演变为

$$p_{\boldsymbol{Z}}(z)=\sum_{q=1}^{N}\left\{\delta\left[z-\boldsymbol{g}\left(\boldsymbol{\theta}_q\right)\right]P_q\right\} \tag{23}$$

$$p_{\boldsymbol{Z}}(z,t)=\sum_{q=1}^{N}\left\{\delta\left[z-\boldsymbol{g}\left(\boldsymbol{\theta}_q,t\right)\right]P_q\right\} \tag{24}$$

为了提高数值积分的计算精度, 本文选择使用高斯函数进行狄拉克δ函数光滑化。高斯函数有

$$\delta(z,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-z^2/2\sigma^2} \tag{25}$$

式中: σ为平滑参数。静力系统和动力系统最佳平滑参数表达式分别为

$$\sigma=\frac{A}{N^{1/5}}\min_{q=1,2,\cdots,N}\left\{\mathrm{std}\left[\boldsymbol{g}\left(\boldsymbol{\theta}_q\right)\right],\frac{\mathrm{iqr}\left[\boldsymbol{g}\left(\boldsymbol{\theta}_q\right)\right]}{1.34}\right\} \tag{26}$$

$$\sigma=\frac{A}{N^{1/5}}\min_{q=1,\cdots,N}\left\{\mathrm{std}\left[\boldsymbol{g}\left(\boldsymbol{\theta}_q,t\right)\right],\frac{\mathrm{iqr}\left[\boldsymbol{g}\left(\boldsymbol{\theta}_q,t\right)\right]}{1.34}\right\} \tag{27}$$

式中: A为光滑化因子, 一般$A\in(0,1]$; iqr(∗)表示四分位距; std(∗)表示标准差。

静力系统和动力系统的概率密度函数进一步演变为

$$p_{\boldsymbol{Z}}(z)=\sum_{q=1}^{N}\left\{\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\left[z-g\left(\boldsymbol{\theta}_q\right)\right]^2/2\sigma^2}P_q\right\} \tag{28}$$

$$p_{\boldsymbol{Z}}(z,t)=\sum_{q=1}^{N}\left\{\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\left[z-g\left(\boldsymbol{\theta}_q,t\right)\right]^2/2\sigma^2}P_q\right\} \tag{29}$$

3.4 基于直接概率积分法的可靠性分析流程

对于随机静力系统, DPIM可以直接对结构响应在失效域上的概率密度函数进行积分来得到失效概率, 即

$$P_{\mathrm{f}}=\int_{\Omega_{\mathrm{f},Z}}p_{\boldsymbol{Z}}(z)\mathrm{d}z=\int_{\Omega_{\mathrm{f},Z}}\sum_{q=1}^{N}\left\{\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\left[z-g\left(\boldsymbol{\theta}_q\right)^2/2\sigma^2\right]}P_q\right\}\mathrm{d}z=$$
$$\sum_{q=1}^{N}\left\{\int_{\Omega_{\mathrm{f},Z}}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\left[z-g\left(\boldsymbol{\theta}_q\right)^2/2\sigma^2\right]}\mathrm{d}z\right\}\cdot P_q \tag{30}$$

对于动力系统, 结合首次超越准则, 将可靠性分析转换为时间$(0,t]$内的等价极值映射。最后结合直接概率积分法求解动力系统极值的概率密度函数, 有

$$P_{\boldsymbol{Z}_{\mathrm{ext}}}(z_{\mathrm{ext}},t)=\int_{\Omega_{\boldsymbol{\theta}}}\delta[z_{\mathrm{ext}}-\boldsymbol{Z}_{\mathrm{ext}}(\boldsymbol{\theta},t)]p_{\boldsymbol{\theta}}(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}=$$
$$\sum_{q=1}^{N}\left\{\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\left[z_{\mathrm{ext}}-\boldsymbol{Z}_{\mathrm{ext}}\left(\boldsymbol{\theta}_q,t\right)\right]^2/2\sigma^2}P_q\right\} \tag{31}$$

式中: $\boldsymbol{Z}_{\mathrm{ext}}$表示狄拉克$\delta$函数的一个变量。最后, t时刻的动力可靠度可以由安全域内对时间$(0,t]$内动力响应极值的概率密度函数积分, 即

$$R(t)=\int_{\Omega_s}P_{\boldsymbol{Z}_{\mathrm{ext}}}(z_{\mathrm{ext}},t)\mathrm{d}z_{\mathrm{ext}} \tag{32}$$

3.5 混合可靠性分析流程

将本文提出的随机变量和证据变量融合的可靠性分析方法步骤概括如图3所示:

步骤1 确定原始输入信息, 包括n维证据模型的不确定性输入以及相对应的BPA; m维随机输入变量以及相对应的概率密度函数。

步骤2 通过证据变量均匀化方法, 将证据变量转化为随机变量。

步骤3 采用GF偏差选点技术, 获得代表点; 并且根据Voronoi剖分技术产生代表区域; 计算代表区域的赋得概率。

步骤4 根据映射$G:\Theta\to Z$或者$Z(t)$在每个代表点处求解静态响应$z_{1,\mathrm{g}}=\boldsymbol{g}_1\left(\boldsymbol{\theta}_q\right)$或者动态响应$z_{1,\mathrm{g}}(t)=\boldsymbol{g}_1\left(\boldsymbol{\theta}_q,t\right)$。

步骤5 对于结构静力系统, 直接对结构响应在失效域上的概率密度函数进行积分来得到失效概率区间。对于结构动力系统则结合首次超越准则和等价极值映射对相应的概率密度函数进行积分, 计算某一时刻的动力可靠度区间, 或者相对应

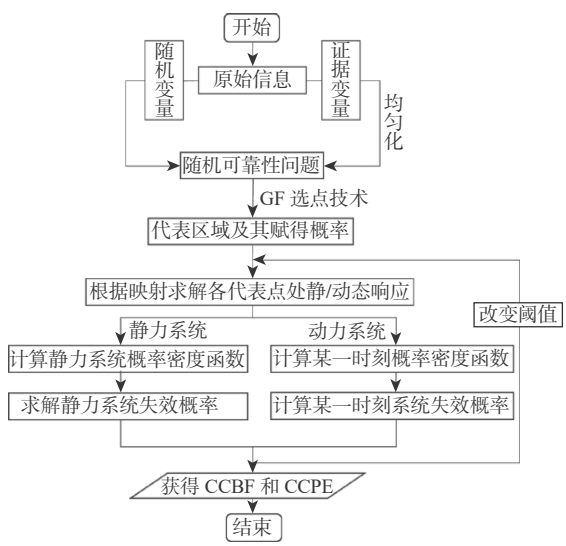


图3 混合可靠性分析流程

Fig. 3 Hybrid reliability analysis flowchart

的动力失效概率区间。

步骤 6　改变阈值, 重复步骤 4 和步骤 5, 计算可得到不同阈值下的可信度函数数值和似真度函数数值。通过曲线拟合得到 CCBF 曲线和 CCPF 曲线。不同阈值下可靠度结果位于 CCBF 曲线和 CCPF 曲线之间。

4　算例分析

4.1　解析函数

考虑以下数值算例的极限状态函数:

$$g(X_1,Y_1,Y_2)=X_1{}^2+Y_1+Y_2 \tag{33}$$

式中: Y_1和Y_2为随机变量, 具体分布参数见表 1。X_1为证据变量, 具体不确定性参数见表 2。该极限函数可靠域为$g(X_1,Y_1,Y_2)<c$, c为阈值。

表 1　算例 1 随机变量分布参数

Table 1　Random variable distribution parameters in Example 1

变量	分布类型	均值	标准差
Y_1	正态	3	0.1
Y_2	正态	2	0.3

表 2　算例 1 证据变量

Table 2　Evidence variable in Example 1

变量	取值范围	BPA
X_1	[1.0,1.1]	0.3
	[1.1,1.2]	0.3
	[1.2,1.3]	0.4

采用本文方法计算不同阈值下的该极限状态函数可靠度, 计算结果见表 3。相应的 CCBF 和 CCPF 曲线见图 4。

表 3　算例一可靠性分析结果

Table 3　Reliability analysis results of example 1

阈值	本文方法	IS	与IS相对误差/%	MCS	与MCS相对误差/%
10.0	[1.371 8×10^{-4},0.000 9]	[1.387 2×10^{-4},0.000 8]	[−1.12,11.1]	[1.362 0×10^{-4},0.000 8]	[0.72,11.1]
10.2	[0.000 9,0.003 0]	[0.000 8,0.003 2]	[11.1,−6.67]	[0.000 9,0.002 9]	[0.00,6.66]
10.4	[0.001 3,0.007 9]	[0.001 5,0.007 8]	[−15.4,1.27]	[0.001 3,0.007 7]	[0.00,2.53]
10.6	[0.005 6,0.016 8]	[0.005 2,0.016 7]	[7.14,0.60]	[0.005 5,0.016 6]	[1.79,1.20]
10.8	[0.011 1,0.034 8]	[0.010 5,0.034 6]	[5.41,0.57]	[0.010 9,0.034 2]	[1.80,1.72]
11.0	[0.026 1,0.066 7]	[0.025 8,0.064 8]	[1.15,2.85]	[0.024 5,0.061 7]	[6.13,7.50]
11.2	[0.050 1,0.115 6]	[0.048 7,0.115 1]	[2.79,0.43]	[0.048 1,0.114 6]	[3.99,0.87]
11.4	[0.092 5,0.182 3]	[0.090 1,0.180 8]	[2.59,0.82]	[0.090 2,0.181 0]	[2.49,0.71]
11.6	[0.150 2,0.279 2]	[0.152 0,0.276 6]	[−1.19,0.93]	[0.147 3,0.274 1]	[1.97,1.86]
11.8	[0.237 4,0.381 7]	[0.236 8,0.381 1]	[0.25,0.16]	[0.237 2,0.380 9]	[0.08,0.21]
12.0	[0.335 0,0.499 4]	[0.334 1,0.499 3]	[0.27,0.02]	[0.330 1,0.498 0]	[1.48,0.28]
12.2	[0.442 7,0.611 3]	[0.441 0,0.611 7]	[0.38,−0.07]	[0.442 7,0.611 0]	[0.00,0.05]
12.4	[0.564 8,0.722 0]	[0.563 2,0.721 9]	[0.28,0.01]	[0.561 5,0.720 5]	[0.59,0.21]
12.6	[0.676 0,0.817 5]	[0.676 2,0.816 0]	[−0.03,0.18]	[0.675 6,0.814 2]	[0.06,0.49]
12.8	[0.771 0,0.880 2]	[0.771 0,0.878 8]	[0.00,0.16]	[0.770 8,0.878 8]	[0.02,0.16]
13.0	[0.849 0,0.929 7]	[0.847 0,0.928 9]	[0.24,0.09]	[0.848 4,0.928 1]	[0.23,0.17]
13.2	[0.908 1,0.968 5]	[0.907 2,0.962 7]	[0.10,0.60]	[0.907 2,0.961 0]	[0.10,0.78]
13.4	[0.948 4,0.978 7]	[0.947 9,0.978 0]	[0.05,0.07]	[0.947 8,0.977 8]	[0.06,0.09]
13.6	[0.971 2,0.989 8]	[0.971 1,0.984 0]	[0.01,0.59]	[0.970 1,0.980 6]	[0.11,0.94]
13.8	[0.986 5,0.995 5]	[0.985 5,0.992 7]	[0.10,0.28]	[0.986 0,0.995 0]	[0.05,0.05]
14.0	[0.995 9,0.998 7]	[0.994 0,0.999 1]	[0.19,−0.04]	[0.993 9,0.997 8]	[0.20,0.09]

表 3 中还展示了基于蒙特卡罗模拟法(Monte Carlo simulation, MCS)的传统证据-概率混和可靠性方法计算结果, 以及文献 [17] 中的方法结合重要抽样方法(importance samplings, IS)的计算结果。可以看出: 本文方法与上述 2 种方法的计算精度基本吻合。在计算效率方面, 本文方法调用极限状态函数10^3次, IS 调用极限函数5×10^5, MCS 调用极限状态函数10^7次, 本文方法显著提升了计算效率。

4.2　五层弹性框架静载可靠性分析

考虑如图 5 所示五层弹性框架静载结构[2]。每个层间刚度$k_i(i=1,2,\cdots,5)$是正态分布。

此框架结构的极限状态函数为

$$G=g(\boldsymbol{\Theta})=\bar{Y}-\max_{1\leqslant i\leqslant 5}\{Y_i(\boldsymbol{\Theta})\} \tag{34}$$

式中: $\bar{Y}$表示地板能承受的位移的最大阈值。

$Y_i(\boldsymbol{\Theta})$表示框架结构的第i层位移, 可由平衡方

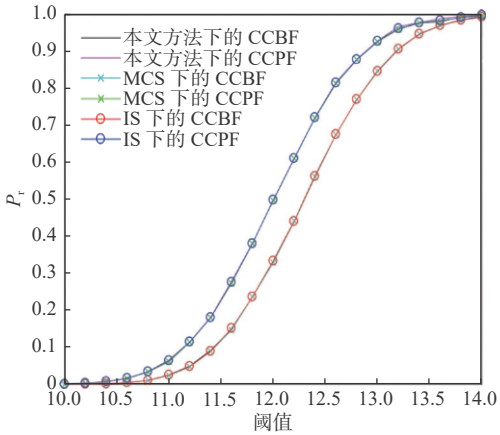


图 4 算例 1 的 CCPF 和 CCBF 曲线

Fig. 4 CCPF and CCBF curves of Example 1

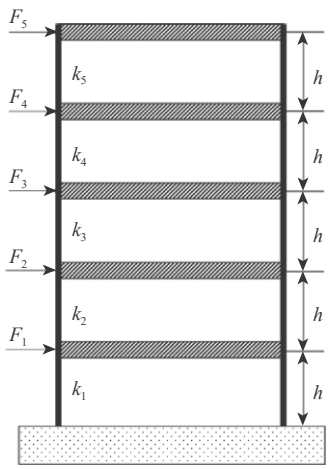


图 5 静载作用下的五层剪力式框架结构

Fig. 5 Five story shear frame structure under static load

程求解：

$$\boldsymbol{K}(\Theta)\boldsymbol{Y}=\boldsymbol{F} \tag{35}$$

式中：$\boldsymbol{Y}=[\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_5]^{\mathrm{T}}$表示地板的位移矢量；$\boldsymbol{F}$表示载荷矢量；$\boldsymbol{K}(\Theta)$表示层间刚度，随机刚度矩阵由式 (36) 构成：

$$k_i=\mu_i(1+\Theta_i C_i),i=1,2,\cdots,5 \tag{36}$$

式中：$\Theta_i,i=1,2,\cdots,5$表示基本输入随机变量；C_i表示第i个随机变量的变异系数，本文选取$C=0.2$。

横向作用在楼板上的静载荷$F_1\sim F_5$具体数值为：$F_1=0.5\times10^6$ N，$F_2=F_3=F_4=2F_1$，$F_5=F_1$。

将五层框架结构中前三层的基本随机变量考虑为概率分布并服从正态分布，具体分布参数见表 4。将第 4 层和第 5 层的基本变量定义为证据变量，具体数值见表 5。

5 层框架结构可靠性分析结果见表 6，CCPF 和 CCBF 曲线见图 6。

本文方法和 MCS 以及 IS 这 2 种对比方法的计算精度基本吻合，但是在计算效率方面，本文方法

表 4 前 3 层结构层间刚度概率分布以及统计特性

Table 4 Probability distribution and statistical characteristics of interlayer stiffness of the first three story frame structure

层数	均值/10^{-5} (kN·m^{-1})	变异系数
Θ_1	3.8	0.2
Θ_2	3.6	0.2
Θ_3	3.4	0.2

表 5 第 4、5 层框架结构层间刚度不确定参数

Table 5 Uncertain parameters of interlayer stiffness of the fourth and fifth story frame structure

层数	均值/10^{-5} (kN·m^{-1})	基本概率分配
Θ_4	[3.0, 3.2]	0.4
	[3.2, 3.4]	0.6
Θ_5	[2.8, 3.0]	0.4
	[3.0, 3.2]	0.6

调用极限状态函数10^3次，MCS 调用极限状态函数10^7次，IS 调用极限状态函数5×10^5，本文方法显著提升了计算效率。

4.3 受非平稳随机激励的弹性薄板可靠性分析

考虑一个矩形薄板[2]受非平稳随机激励时可靠性分析。矩形薄板的边界条件具有简支 (S)、固支 (C) 以及自由 (F)，具体见图 7。

该问题的物理映射Γ如下：

$$\Gamma:\rho h\frac{\partial^2 w(x,y,t)}{\partial t^2}+c\frac{\partial w(x,y,t)}{\partial t}+D\nabla^4 w(x,y,t)=-\rho hY(\Theta_{\mathrm{f}},t) \tag{37}$$

式中：双调和算子$\nabla^4=\dfrac{\partial^4}{\partial x^4}+2\dfrac{\partial^4}{\partial x^2\partial y^2}+\dfrac{\partial^4}{\partial y^4}$；$\rho$为板的体积密度；$H$为板的厚度；横向挠度为$w(x,y,t)$；板的黏滞阻尼系数$C=2\rho h\zeta\omega$，其中，$\zeta=5\%$表示板的最大模态阻尼比；$\omega$表示板的最大固有频率；$Y(\Theta_{\mathrm{f}},t)$为非平稳随机过程。

施加完全非平稳随机加速度激励可采用以下时间调制函数：

$$f(t)=A\left(e^{-\alpha t}-e^{-\beta t}\right) \tag{38}$$

式中：参数$A=4$、α=0.099 5 和$\beta=2\alpha$。

非平稳随机加速度激励的功率谱密度公式为

$$S_Y(t,\omega)=|f(t)|^2\frac{\omega^4}{\omega^4+\omega_h^4}\cdot\frac{1+4\omega^2\zeta_g^2(t)/\omega_g^2(t)}{\left(1-\omega^2/\omega_g^2(t)\right)^2+4\omega^2\zeta_g^2(t)/\omega_g^2(t)}S_0 \tag{39}$$

式中：强度因子$S_0=1$ m^2/s^3，截止频率$\omega_n=2$ rad/s；地面阻尼比$\zeta_g(t)=0.64$。

谱的非平稳性由下式表示：

表 6　5 层框架结构可靠性分析结果

Table 6　Reliability analysis results of five story frame structure

最大位移阈值/m	本文方法	IS	与IS相对误差/%	MCS	与MCS相对误差/%
0.022	[2.493 7×10^{-7},2.760 8×10^{-7}]	[2.378 4×10^{-7}×10^{-7},2.723 1×10^{-7}]	[4.62,1.37]	[2.359 6×10^{-7},2.680 8×10^{-7}]	[5.68,2.98]
0.024	[5.224 5×10^{-6},3.164 6×10^{-5}]	[5.222 6×10^{-6},3.124 6×10^{-5}]	[0.04,1.26]	[5.211 6×10^{-6},3.095 6×10^{-5}]	[0.25,2.23]
0.026	[6.282 9×10^{-4},0.001 8]	[6.235 8×10^{-4},0.001 8]	[0.85,0.00]	[6.126 8×10^{-4},0.001 6]	[0.55,11.1]
0.028	[0.008 9,0.017 6]	[0.007 9,0.016 9]	[11.2,3.98]	[0.008 1,0.016 6]	[8.99,6.02]
0.030	[0.049 7,0.085 3]	[0.047 8,0.084 6]	[3.82,0.82]	[0.045 8,0.083 1]	[8.52,2.65]
0.032	[0.164 9,0.234 2]	[0.162 3,0.228 1]	[1.58,2.60]	[0.162 7,0.222 2]	[1.60,5.40]
0.034	[0.348 8,0.433 9]	[0.346 4,0.415 6]	[0.69,4.22]	[0.343 7,0.433 7]	[1.48,0.05]
0.036	[0.551 6,0.632 7]	[0.554 2,0.636 6]	[−0.47,−0.62]	[0.556 3,0.638 5]	[0.84,0.91]
0.038	[0.723 0,0.780 2]	[0.723 1,0.790 1]	[−0.01,−1.27]	[0.731 9,0.791 4]	[1.21,1.42]
0.040	[0.835 9,0.873 9]	[0.856 0,0.888 8]	[−2.40,−1.71]	[0.850 4,0.886 8]	[−1.73,−1.48]
0.042	[0.911 6,0.934 8]	[0.910 5,0.941 2]	[0.12,−0.68]	[0.920 3,0.940 8]	[−0.95,0.64]
0.044	[0.953 9,0.966 0]	[0.957 2,0.966 6]	[−0.35,−0.06]	[0.958 3,0.969 3]	[0.46,0.34]
0.046	[0.977 9,0.985 1]	[0.978 5,0.984 6]	[−0.06,0.05]	[0.978 1,0.983 8]	[0.06,0.13]
0.048	[0.991 8,0.994 4]	[0.988 9,0.993 8]	[0.29,0.06]	[0.988 3,0.991 2]	[0.35,0.32]
0.050	[0.996 2,0.996 9]	[0.995 8,0.996 5]	[0.04,0.04]	[0.993 5,0.995 1]	[0.27,0.18]
0.052	[0.997 7,0.998 3]	[0.996 8,0.997 9]	[0.09,0.04]	[0.996 3,0.997 1]	[0.14,0.12]
0.054	[0.999 0,0.999 2]	[0.997 2,0.998 0]	[0.18,0.12]	[0.997 8,0.998 2]	[0.12,0.10]
0.056	[0.999 3,0.999 4]	[0.999 0,0.999 1]	[0.03,0.03]	[0.998 6,0.998 9]	[0.10,0.05]
0.058	[0.999 5,0.999 6]	[0.999 2,0.999 5]	[0.03,0.01]	[0.999 1,0.999 2]	[0.04,0.04]
0.060	[0.999 7,0.999 8]	[0.999 4,0.999 7]	[0.03,0.01]	[0.999 4,0.999 5]	[0.03,0.03]
0.062	[0.999 8,0.999 9]	[0.999 8,0.999 9]	[0.00,0.00]	[0.999 6,0.999 8]	[0.02,0.01]

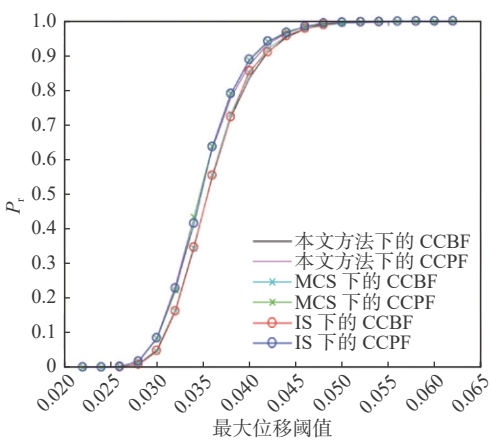


图 6　5 层框架结构的 CCPF 和 CCBF 曲线

Fig. 6　CCPF and CCBF curves of five story shear frame structure

$$\omega_g(t)=\omega_0-(\omega_0-\omega_T)\frac{t}{T} \tag{40}$$

式中: T为随机激励的总持续时间, T = 25 s; ω_0为初始时刻的地面频率, ω_0 = 39.4 rad/s; ω_T为持续时间结束时地面频率, ω_T = 4.86 rad/s。

完全非平稳激励通过使用基于随机函数的谱产生:

$$Y(\Theta_{\mathrm{f}},t)=f(t)\sum_{K=0}^{N}\sqrt{S(\omega_K)\Delta\omega}\cdot [\cos(\omega_K t)R_K+\sin(\omega_k t)Q_K] \tag{41}$$

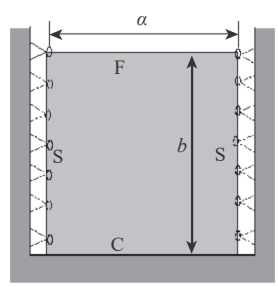


图 7　试验矩形薄板

Fig. 7　Rectangular thin plate

式中: R_k和Q_k表示一对正交随机变量, 将其定义为关于基本随机变量$\Theta_{\mathrm{f}}=[\Theta_1,\Theta_2]$:

$$R_k=\cos(k\Theta_1)+\sin(k\Theta_1) \tag{42}$$

$$Q_k=\cos(k\Theta_2)+\sin(k\Theta_2) \tag{43}$$

失效准则为板的横向挠度首次超过特定阈值, 极限状态函数可表示为

$$Z=g(\Theta,t)=\overline{\omega}-\max\{\omega(x,y,t)\} \tag{44}$$

式中: $x\in[0,a]$, $y\in[0,b]$, a为板长, b为板宽; $\overline{\omega}$为横向挠度$\omega(x,y,t)$的阈值。

将薄板的几何尺寸考虑为随机不确定性变量, 具体数据见表 7。

将薄板的弹性模量、泊松比和质量密度考虑为

表7 薄板的几何尺寸

Table 7 Geometric dimensions of the thin plates

变量	分布形式	均值/m	标准差
板宽a	正态分布	6	0.03
板长b	正态分布	8	0.02
板高h	正态分布	0.12	0.01

表8 薄板的证据不确定性变量

Table 8 Evidence uncertainty variables for the thin plates

变量	取值范围	BPA
弹性模量E/GPa	[35.8,36.2]	0.3
	[36.2,36.6]	0.7
泊松比v	[0.15,0.20]	0.3
	[0.20,0.25]	0.7
质量密度ρ/(kg·m^{-3})	[2 200,2 400]	0.3
	[2 400,2 600]	0.7

证据不确定性变量，具体数据见表8。

本文方法和2种对比方法计算结果见表9，计算所得的CCBF和CCPF曲线如图8所示。在计算效率方面，本文方法极限状态函数调用次数为10^3次，而基于MCS的传统方法为10^6次，IS为2×10^5，本文方法显著提升了计算效率。

表9 薄板结构可靠性分析结果

Table 9 Reliability analysis results of the thin plate

横向挠度阈值/m	本文方法	IS	与IS相对误差/%	MCS	与MCS相对误差/%
0.000 60	[0.002 4,0.003 6]	[0.002 4,0.003 5]	[0.00,2.78]	[0.002 3,0.003 5]	[4.17,2.78]
0.000 64	[0.023 0,0.026 0]	[0.022 4,0.025 8]	[2.61,0.77]	[0.022 1,0.025 3]	[3.91,2.70]
0.000 68	[0.091 0,0.117 0]	[0.089 6,0.116 2]	[1.54,0.68]	[0.089 4,0.115 1]	[1.79,1.62]
0.000 72	[0.227 0,0.278 1]	[0.225 2,0.277 9]	[0.79,0.07]	[0.224 8,0.275 2]	[0.97,1.04]
0.000 76	[0.415 8,0.455 9]	[0.414 4,0.455 0]	[0.34,0.20]	[0.413 7,0.454 4]	[0.51,0.33]
0.000 80	[0.617 9,0.649 0]	[0.615 9,0.647 8]	[0.32,0.18]	[0.616 1,0.648 2]	[0.17,0.12]
0.000 84	[0.771 0,0.786 0]	[0.768 8,0.780 2]	[0.29,0.75]	[0.769 7,0.781 5]	[0.17,0.57]
0.000 88	[0.863 9,0.879 1]	[0.861 5,0.876 6]	[0.28,0.28]	[0.859 4,0.874 3]	[0.47,0.42]
0.000 92	[0.926 0,0.939 0]	[0.920 8,0.936 8]	[0.56,0.23]	[0.921 5,0.935 4]	[0.49,0.38]
0.000 96	[0.968 0,0.974 0]	[0.966 6,0.970 2]	[0.14,0.39]	[0.965 2,0.971 5]	[0.29,0.35]
0.001 00	[0.985 0,0.989 9]	[0.981 0,0.988 6]	[0.41,0.13]	[0.981 4,0.988 9]	[0.37,0.10]
0.001 04	[0.992 0,0.994 8]	[0.991 5,0.993 2]	[0.05,0.16]	[0.991 0,0.991 9]	[0.10,0.29]
0.001 08	[0.993 5,0.996 4]	[0.992 6,0.995 8]	[0.09,0.06]	[0.992 1,0.995 0]	[0.14,0.14]
0.001 12	[0.994 6,0.998 0]	[0.993 8,0.997 6]	[0.08,0.04]	[0.994 2,0.997 3]	[0.04,0.07]
0.001 16	[0.997 9,0.999 9]	[0.997 8,0.998 8]	[0.01,0.11]	[0.997 7,0.998 4]	[0.02,0.15]
0.001 20	[0.999 2,0.999 9]	[0.999 1,0.999 9]	[0.01,0.00]	[0.999 0,0.999 6]	[0.02,0.03]

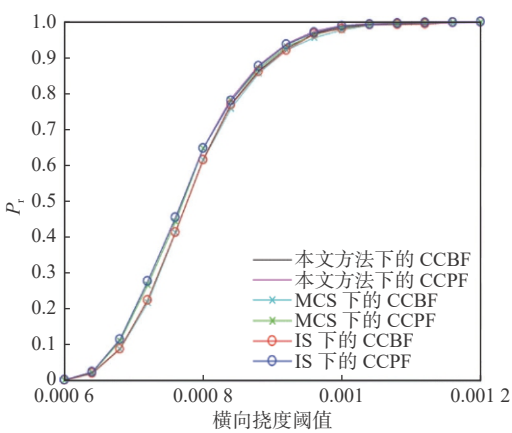


图8 薄板结构的CCPF和CCBF曲线

Fig. 8 CCPF and CCBF curves of the thin plate

5 结 论

针对证据变量和随机变量同时存在的混合不确定性问题，同时考虑复杂极限状态函数，提出了一种新的证据-概率混合变量的结构可靠性分析方法。首先采用证据理论的均匀化方法，将证据变量转换成随机变量，进而采用直接概率积分法进行结构可靠度分析。高效地解决了极限状态函数复杂的混合变量可靠性分析问题。文中利用3个算例对所提方法进行了性能分析，可知：本文方法在满足计算精度的同时，避免了直接计算可靠度的复杂性，将嵌套计算进行解耦，很大程度上提高了计算效率。本文方法为工程结构可靠性分析提供了一种新的证据-概率混合可靠性分析方法。

参考文献（References）

[1] CHEN G H, YANG D X. Direct probability integral method for stochastic response analysis of static and dynamic structural systems[J]. Computer Methods in Applied Mechanics and Engineering, 2019, 357(C): 112612.

[2] CHEN G H, YANG D X. A unified analysis framework of static and dynamic structural reliabilities based on direct probability integral method[J]. Mechanical Systems and Signal Processing, 2021, 158: 107783.

[3] 刘思琦. 基于直接概率积分法的车—(轨)—桥耦合系统随机动力学分析[D]. 大连: 大连理工大学, 2021.
LIU S Q. Stochastic dynamic analysis of vehicle-(rail)-bridge coupling system based on direct probability integral method[D]. Dalian: Dalian University of Technology, 2021(in Chinese).

[4] 李小兰. 结构静/动力可靠度分析与优化设计的高效算法研究[D]. 大连: 大连理工大学, 2021.
LI X L. Study on efficient algorithms for static/dynamic reliability analysis and reliability-based design optimization of structures[D]. Dalian: Dalian University of Technology, 2021(in Chinese).

[5] 李石. 结构特性相关的显式时程积分算法性能改进及工程应用[D]. 大连: 大连理工大学, 2021.
LI S. Performance improvement and engineering application of explicit structure-dependent time integration algorithms[D]. Dalian: Dalian University of Technology, 2021(in Chinese).

[6] 高天骁, 李铸珅, 张佳旺, 等. 基于直接概率积分法的重载铁路 RC 梁疲劳可靠度分析[J]. 铁道科学与工程学报, 2023, 20(4): 1416-1424.
GAO T X, LI Z S, ZHANG J W, et al. Fatigue reliability analysis of RC beams in heavy-haul railways based on direct probability integral method[J]. Journal of Railway Science and Engineering, 2023, 20(4): 1416-1424(in Chinese).

[7] MENG Z, ZHAO J Y, CHEN G H, et al. Hybrid uncertainty propagation and reliability analysis using direct probability integral method and exponential convex model[J]. Reliability Engineering & System Safety, 2022, 228: 108803.

[8] DEMPSTER A P, LAIRD N M, RUBIN D B. Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1977, 39(1): 1-22.

[9] SHAFER G. A mathematical theory of evidence[M]. Princeton: Princeton University Press, 1976.

[10] 范松. 基于证据理论的机械可靠性分析及优化设计方法[D]. 长沙: 湖南大学, 2015.
FAN S. An evidence-theory based mechanical reliability analysis method and optimization design method[D]. Changsha: Hunan University, 2015(in Chinese).

[11] OBERKAMPF W L, HELTON J C. Investigation of evidence theory for engineering application, non-deterinistic approaches forum[C]//Proceedings of the 4th Non-Deterministic Approaches Forum. Reston: AIAA, 2002, 43: 8-11.

[12] BAE H R, GRANDHI R V, CANFIELD R A. Sensitivity analysis of structural response uncertainty propagation using evidence theory[J]. Structural and Multidisciplinary Optimization, 2006, 31(4): 270-279.

[13] BAI Y C, HAN X, JIANG C, et al. Comparative study of metamodeling techniques for reliability analysis using evidence theory[J]. Advances in Engineering Software, 2012, 53: 61-71.

[14] ZHANG Z, JIANG C, HAN X, et al. A response surface approach for structural reliability analysis using evidence theory[J]. Advances in Engineering Software, 2014, 69: 37-45.

[15] WANG Y W, FENG H L. Reliability analysis based on inverse Gauss degradation process and evidence theory[J]. International Journal of Performability Engineering, 2019, 15(2): 353-361.

[16] DU X P. Uncertainty analysis with probability and evidence theories[C]// Proceedings of the ASME 2006 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Philadelphia: ASME, 2006: 1025-1038.

[17] 姜潮, 范松, 张哲, 等. 一种高效的概率-证据混合可靠性分析方法[J]. 计算力学学报, 2016, 33(2): 135-143.
JIANG C, FAN S, ZHANG Z, et al. An efficient structural reliability anslysis method with both probability variable and evidence variable[J]. Chinese Journal of Computational Mechanics, 2016, 33(2): 135-143(in Chinese).

[18] TAO Y R, HAN X, DUAN S Y, et al. Reliability analysis for multidisciplinary systems with the mixture of epistemic and aleatory uncertainties[J]. International Journal for Numerical Methods in Engineering, 2014, 97(1): 68-78.

[19] 梁云飞, 张德权, 彭周源. 一种高效的概率-证据混合工业机器人定位精度可靠性分析方法[J]. 燕山大学学报, 2022, 46(4): 309-318.
LIANG Y F, ZHANG D Q, PENG Z Y. An efficient reliability analysis method for positioning accuracy of industrial robot through probability and evidence theory[J]. Journal of Yanshan University, 2022, 46(4): 309-318(in Chinese).

[20] 曹亮, 陈国强, 董丽君, 等. 基于概率论和证据理论的小概率问题可靠性分析[J]. 湖南工程学院学报 (自然科学版), 2019, 29(2): 30-35.
CAO L, CHEN G Q, DONG L J, et al. Reliability analysis of small probability problem based on probality and evidence theory[J]. Journal of Hunan Institute of Engineering (Natural Science Edition), 2019, 29(2): 30-35(in Chinese).

[21] 张梦. 基于区间-概率混合模型的在役桥梁可靠性评估[J]. 黑龙江交通科技, 2023, 46(3): 80-82,86.
ZHANG M. Reliability evaluation of in service bridges based on interval stochastic hybrid model[J]. Communications Science and Technology Heilongjiang, 2023, 46(3): 80-82,86(in Chinese).

[22] 孟增, 汪栋梁, 郝鹏. 面向概率-区间混合可靠性分析的自适应共轭控制方法与应用[J]. 应用力学学报, 2022, 39(6): 1076-1085.
MENG Z, WANG D L, HAO P. Adaptive conjugate control method for probability-interval hybrid reliability analysis and its engineering application[J]. Chinese Journal of Applied Mechanics, 2022, 39(6): 1076-1085(in Chinese).

[23] JIANG C, ZHANG Z, HAN X, et al. A novel evidence-theory-based reliability analysis method for structures with epistemic uncertainty[J]. Computers & Structures, 2013, 129: 1-12.

[24] CHEN J B, YANG J Y, LI J. A GF-discrepancy for point selection in stochastic seismic response analysis of structures with uncertain parameters[J]. Structural Safety, 2016, 59: 20-31.

A mixed reliability analysis method based on direct probability integral

ZHU Qitao，LI Hongshuang*

(College of Aerospace Engineering，Nanjing University of Aeronautics and Astronautics，Nanjing，210016，China)

Abstract：Based on the idea of direct probability integral, a mixed reliability analysis method is proposed for the mixed uncertainty analysis of evidence variables and random variables with a complex limit state function. The proposed method starts with the probability interpretation of evidence theory, adopts the method of unifying evidence variables for converting evidence variables into random variables, and achieve the goal of unifying two types of uncertainty. Then, the structural reliability analysis is carried out through the direct probability integral method. Finally, the performance of the proposed method is validated through three numerical examples. The results show that compared with existing methods, the proposed method greatly improves computational efficiency with similar computational accuracy.

Keywords：evidence theory；direct probability integral method；structure reliability analysis；mixed uncertainty；randow variables

Received：2023-07-27；**Accepted**：2023-09-04；**Published Online**：2023-09-15 15：01

URL：link.cnki.net/urlid/11.2625.V.20230914.2020.001

* **Corresponding author**. E-mail：hongshuangli@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0415

繁忙终端空域连续爬升与下降运行的调度优化

杜卓铭，张军峰*，归旭豪
(南京航空航天大学 民航学院，南京 210016)

摘　　要：基于现行终端空域结构，提出了一种融合航迹优化、冲突探测与多目标优化的进离场航空器调度方法，助力繁忙终端空域实施连续爬升与下降运行。基于多阶段最优控制理论，采用高斯伪谱法，提出了连续爬升与下降运行的垂直剖面优化方法，实现了基于成本指数的连续爬升与下降运行的航迹优化。依据跑道使用的尾流间隔与放行间隔、空中运行的水平与垂直间隔，采用马氏距离建立了航空器冲突探测模型。考虑空管、航司、机场等运行单位的诉求，提出了优化结果可达的进离场航空器多目标调度模型和方法。选取繁忙时段广州白云国际机场两组进离场数据，设置多种间隔参数，引入备选路径，开展实例分析与对比研究。结果表明：在离场占优的繁忙时段，广州白云国际机场终端空域能够实现繁忙时段下的连续爬升与下降运行；在进场占优的繁忙时段，出现2架航空器无法调度的情况，引入备选路径的方法可以减少无法调度航空器的数量。

关 键 词：空中交通管理；连续下降运行；航迹优化；航空器调度；多目标优化
中图分类号：V355
文献标志码：A　　**文章编号**：1001-5965(2025)09-3193-10

连续爬升与下降运行作为一种新型运行理念，由于其显著的节能减排降噪特点，已在国内外诸多机场开展验证与实施。早期研究主要聚焦于垂直剖面的设计与优化[1]，以及相应节能减排降噪效果的评估[2]。结果表明，连续爬升与下降运行确实达到了节能减排降噪的效果，然而实际运行往往受限于夜间空闲时刻，无法适应繁忙时段的终端空域运行。毕竟，在繁忙时段，连续爬升与下降运行的实施无疑增加了管制员的工作负荷[3]。

为了实现连续爬升与下降运行在繁忙终端空域实施，并进一步节能减排降噪的目标，国内外诸多学者进行了不同领域的尝试。

一方面，部分学者通过设计新型的进场程序以实现繁忙终端空域内进场航空器连续下降运行。张军峰等[4]提出了多目标帝国竞争算法，研究了基于点融合系统的多目标进场排序与调度问题。Liang等[5]提出了考虑离场运行的点融合程序，研究了平行跑道进离场协同调度问题。杨磊等[6]设计了一种倒皇冠型的柔性进场路径，相较于点融合系统，柔性进场路径能够更好地融合各方向的进场交通流，增强了进场延误的吸收能力。上述新型的进场程序确实能够助力繁忙终端空域的连续下降运行，然而终端空域资源受限的现状[7]，导致了点融合系统或倒皇冠型结构难以顺利推广。

另一方面，部分学者基于现行空域结构和飞行程序，通过改变飞行路径以调整航空器到达时隙，从而提升繁忙终端空域内进场航空器连续下降运行的成功率。Sáez等[8]通过外扩终端空域与引入“延长三边”(parallel-offset)，为进场调度获得更强的冲突解脱与延误吸收能力。进一步，Sáez等[9]利用最优控制与混合整数规划，基于扩展进场管理实现繁忙机场的连续下降运行。Gui等[10]基于现行

收稿日期：2023-06-28；录用日期：2023-08-25；网络出版时间：2023-09-01 18:01
网络出版地址：link.cnki.net/urlid/11.2625.V.20230901.1703.005
基金项目：国家自然科学基金(52372315,U1933117)；南京航空航天大学科研与实践创新计划(xcxjh20220714)
*通信作者. E-mail：zhangjunfeng@nuaa.edu.cn

引用格式：杜卓铭，张军峰，归旭豪. 繁忙终端空域连续爬升与下降运行的调度优化[J]. 北京航空航天大学学报，2025，51(9)：3193-3202. DU Z M，ZHANG J F，GUI X H. Scheduling optimization for continuous climb and descend operations in busy terminal area[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3193-3202 (in Chinese).

空域结构，使用标准进场程序，辅以“延长三边”和“进场偏离”（dog-leg）等冗余路径，达到冲突解脱与进场调度的目的。上述飞行路径改变的方式[8-10]更符合实际运行，也更容易推广，然而目前的研究并未将离场运行纳入考量，因此繁忙终端空域内同时实现连续爬升与下降运行仍然是亟待解决的关键问题。

此外，为了增强管制员对连续下降运行航空器的态势感知能力，Hong 等[11]在点融合的框架下，提出了动态鲁棒的航空器调度算法，考虑飞行时间的不确定性，最小化连续航空器额外缓冲间隔；Kamo 等[12]将气象不确定性纳入考量，构建了鲁棒的连续下降运行航空器航迹优化与调度模型；Toratani 等[13]则通过设置固定航径角，增强了管制员对繁忙空域内连续下降航空器到达时间的控制手段。

分析上述研究进展，不难发现：繁忙终端空域实现连续爬升与下降运行的核心离不开进离场航空器调度优化，而进离场调度优化研究经历了从进场[14-15]到进离场[16-17]，从静态到动态[18]，从单目标到多目标[19]，从确定性到不确定性[20]，从模型驱动到数据驱动[21]的漫长历程。对于繁忙终端空域连续爬升与下降运行调度优化问题，如何构建融合进离场航迹优化与排序调度模型是核心。

鉴于此，本文首先基于高斯伪谱法，提出了单航空器连续爬升与下降运行的航迹优化方法，实现基于成本指数的连续爬升与下降运行的垂直剖面优化。其次，提出了一种基于马氏距离（Mahalanobis distance）的四维航迹冲突探测方法，能够准确识别潜在冲突，进而构建冲突集合。接着，充分考虑空管、航司、机场等不同运行主体的诉求，设计了多种目标函数，基于冲突集合完善了约束限制，从而构建了繁忙终端空域连续爬升与下降运行调度优化模型，并提出了多目标的优化方法。最后，以广州白云国际机场的实际运行为例展开案例分析，验证了模型和方法的有效性。

1 连续爬升与下降运行调度优化方法

本文提出的繁忙终端空域连续爬升与下降运行调度优化的结构框架，如图 1 所示，主要由 3 个模块组成。

1）航迹优化模块，基于现行空域结构和飞行程序，对于进场航空器，通过采用不同成本指数（cost

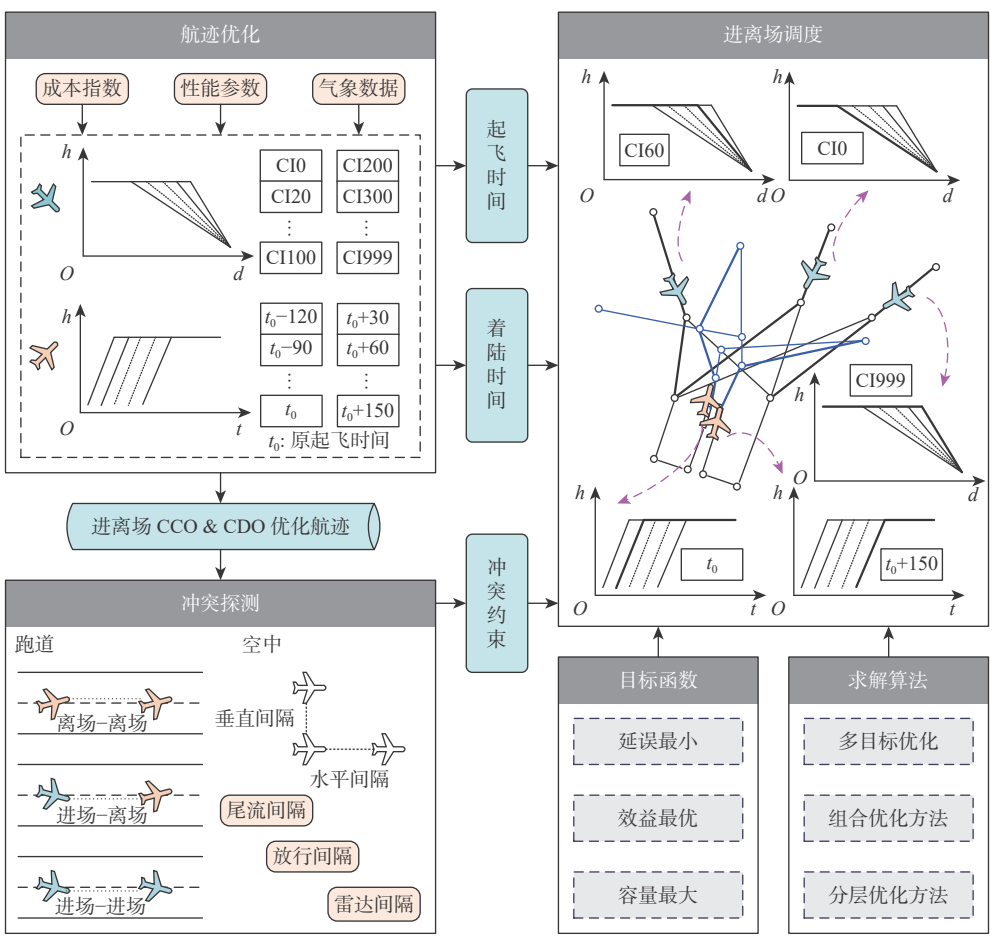


图 1 连续爬升与下降运行调度优化方法的框架示意

Fig. 1 Schematic diagram of the scheduling optimization method for the continuous climb and descent operation

index, CI), 实现航空器连续下降运行的航迹优化; 对于离场航空器, 通过采用不同离场时间, 实现航空器连续爬升运行的航迹优化。

2) 冲突探测模块, 基于进离场航空器之间在跑道和空中的间隔需求, 设计快速冲突探测算法, 识别不同航空器连续爬升与下降运行航迹间的冲突, 构建约束限制。

3) 进离场调度模块, 基于空管、航司、机场的不同诉求, 设计多目标函数, 考虑连续爬升与下降运行航迹间的冲突约束, 构建多目标进离场调度模型, 采用多目标优化算法进行求解。

本文提出的繁忙终端空域连续爬升与下降运行调度优化方法, 可以简化为: 在满足进场连续下降运行、离场连续爬升运行的前提下, 考虑终端空域运行安全间隔、跑道放行和尾流间隔等约束, 以期达到减少延误、提升效益、增强容量的目的。

2　连续爬升与下降运行调度优化模型

2.1　航迹优化模型

鉴于本文提出的方法基于现行空域结构和飞行程序, 因此航迹优化主要涉及垂直剖面的优化。连续爬升与下降运行的垂直剖面优化, 可以描述为多阶段 Bolza 型的最优控制问题:

$$\begin{cases} J=\sum_{p=1}^{N_p}\left[\Phi^{(p)}\left(\boldsymbol{x}_0^{(p)},t_0^{(p)},\boldsymbol{x}_{\mathrm{f}}^{(p)},t_{\mathrm{f}}^{(p)}\right)+\int_{t_0^{(p)}}^{t_{\mathrm{f}}^{(p)}}G^{(p)}\left(\boldsymbol{x}^{(p)},\boldsymbol{u}^{(p)},t\right)\mathrm{d}t\right] \\ \quad \text{s.t.}\quad \dot{\boldsymbol{x}}^{(p)}=f^{(p)}\left(\boldsymbol{x}^{(p)},\boldsymbol{u}^{(p)},t\right) \\ \quad \phi_{\min}^{(p)}\leqslant\phi^{(p)}\left(\boldsymbol{x}_0^{(p)},t_0^{(p)},\boldsymbol{x}_{\mathrm{f}}^{(p)},t_{\mathrm{f}}^{(p)}\right)\leqslant\phi_{\max}^{(p)} \\ \quad C_{\min}^{(p)}\leqslant C^{(p)}\left(\boldsymbol{x}^{(p)},\boldsymbol{u}^{(p)},t\right)\leqslant C_{\max}^{(p)} \\ \quad P^{(s)}\left(\boldsymbol{x}_{\mathrm{f}}^{(p-1)},t_{\mathrm{f}}^{(p-1)},\boldsymbol{x}_0^{(p)},t_0^{(p)}\right) \end{cases} \tag{1}$$

式中: $\boldsymbol{x}_0^{(p)}$、$t_0^{(p)}$、$\boldsymbol{x}_{\mathrm{f}}^{(p)}$、$t_{\mathrm{f}}^{(p)}$分别表示在最优控制阶段$p$的初始与终端状态变量与时间; N_p表示阶段总数; $\boldsymbol{x}^{(p)}$、$\boldsymbol{u}^{(p)}$表示阶段p关于自变量时间t的状态变量与控制变量; $\Phi^{(p)}(\cdot)$、$G^{(p)}(\cdot)$分别表示p阶段 Mayer 型与拉格朗日(Lagrangian)型指标; 式(1)中, $f^{(p)}$、$\phi^{(p)}$、$C^{(p)}$分别表示阶段p的动态、边界、路径和连接约束函数, $P^{(s)}$表示第s对左右待连接阶段的连接约束函数, 分别对应着系统的动力学微分方程、阶段始末状态、阶段过程中状态的约束。

对于连续爬升与下降运行而言, 其动态约束将影响垂直剖面优化的结果。采用下式所示的质点运动模型作为动态约束。

$$\frac{\mathrm{d}V_{\mathrm{TAS}}}{\mathrm{d}t}=\frac{1}{m}(T-D)-g\sin(\gamma) \tag{2a}$$

$$\frac{\mathrm{d}h}{\mathrm{d}t}=V_{\mathrm{TAS}}\sin(\gamma) \tag{2b}$$

$$\frac{\mathrm{d}s}{\mathrm{d}t}=V_{\mathrm{TAS}}\cos(\gamma)+W_{\mathrm{h}} \tag{2c}$$

$$\frac{\mathrm{d}m}{\mathrm{d}t}=-f_{\mathrm{Fuel}} \tag{2d}$$

式中: T、D、m、g分别表示航空器推力、阻力、质量与重力加速度; γ表示飞行航径角; h、s、V_{TAS}分别表示航空器高度、距离与真空速; W_{h}表示风速的沿航迹分量; f_{Fuel}为燃油流量。式(2)中推力和阻力可以通过 BADA 提供的性能参数获得[22]。边界约束、路径约束分别是用于约束优化过程始末端点处和约束优化过程中的状态、控制变量, 而连接约束则用于优化过程中不同阶段状态变量的衔接。

鉴于高斯伪谱法具有结构简单、精度较高和收敛速度较快的特点, 因此本文采用高斯伪谱法求解多阶段最优控制问题(1)。具体步骤包括: 时域变换、离散化、变量逼近与问题转化[23]。如此多阶段最优控制问题则可以转化为代数约束下的非线性规划问题, 并通过软件包 SNOPT[24] 对上述非线性规划问题进行求解, 从而获得不同成本指数对应的进场连续下降运行四维航迹与不同起飞时间的离场连续爬升运行四维航迹。

2.2　冲突探测模型

获取进离场航空器所有的连续爬升与下降运行四维航迹后, 本文通过遍历起降航空器的进离场四维航迹组合, 得到航空器发生冲突的剖面组合, 作为后续进离场调度的约束限制。进离场航空器发生的冲突可以分为空域和跑道 2 个场景。

在空域中, 航空器间需满足符合中国民用航空规章规定的最小垂直和水平间隔标准。为了准确、快速识别运行中的潜在冲突, 本文使用马氏距离确定航空器空中冲突。利用马氏距离判断空中冲突, 可以很好地适应仪表着陆中航空器截取下滑道后横向间隔改变的需要[25]。

在跑道上, 航空器间隔主要依赖于多跑道的运行模式。同一跑道, 进场与进场航空器间需满足尾流间隔, 离场与离场航空器间需满足放行间隔, 进场与离场航空器间需满足尾流间隔。不同跑道, 相关平行进近与隔离平行运行需要考虑雷达间隔; 独立平行进近与独立平行离场, 本文不考虑起降航空器在跑道上的间隔, 毕竟空中的冲突探测能够弥补此类情形的潜在冲突。

2.3　进离场调度模型

2.3.1　决策变量

模型的决策变量为$x_{i,k}$,

$$x_{i,k}=\begin{cases}1 & \text{如果航空器}i\text{使用剖面}k\\ 0 & \text{其他}\end{cases}$$

式中: $i\in A$, A为待调度的航空器集合; $k\in P_i$, P_i为

航空器i可使用的连续下降或爬升的剖面集合。

2.3.2 目标函数

1）最小化进场平均延误。对于进场运行，其延误时间往往是空管、航司、机场等各方关注的重点指标，在满足约束条件的前提下，优化进场飞行延误，使得终端空域内进场平均延误最小，即

$$\min \frac{1}{N}\sum_{i\in A}\sum_{k\in P_i} x_{i,k}\left(t_{i,k}^{\mathrm{TTOA}}-t_i^{\mathrm{ETOA}}\right) \tag{3}$$

式中：$i\in A$、$k\in P_i$，N表示航空器集合A中的航空器数量；$t_{i,k}^{\mathrm{TTOA}}$为航空器i使用剖面k的到达跑道目标着陆时间（target time of arrival, TTOA），t_i^{ETOA}为航空器i在性能约束允许的最早到达跑道时间（earliest time of arrival, ETOA），$t_{i,k}^{\mathrm{TTOA}}-t_i^{\mathrm{ETOA}}$即为航空器$i$使用剖面$k$到达跑道所产生的延误。以往研究表明[26]，最小化进场平均延误与最小化进场交通流的最晚着陆时间等价，也即最小化该目标能够确保跑道进场容量的提升。

2）最小化平均离场偏离。对于离场运行，每架航空器在起飞前需要历经机位推出、滑行、等待排队、进入跑道等过程。本文不涉及机坪的运行优化，因此对于起飞航班，将调度后的目标起飞时间与实际起飞时间的平均偏离作为目标，即

$$\min \frac{1}{N}\sum_{i\in A}\sum_{k\in P_i} x_{i,k}\cdot\left|t_{i,k}^{\mathrm{TTOT}}-t_i^{\mathrm{ATOT}}\right| \tag{4}$$

式中：$t_{i,k}^{\mathrm{TTOT}}$为航空器i使用剖面k的目标起飞时间，$t_{i,k}^{\mathrm{ATOT}}$为航空器i的实际起飞时间（actual take off time, ATOT），$\left|t_{i,k}^{\mathrm{TTOT}}-t_i^{\mathrm{ATOT}}\right|$即为航空器$i$使用剖面$k$与实际起飞时间的偏差。

3）最小化最大进场飞行时间。前述2个目标函数可以从总体上提升终端空域内进离场运行效率，其结果往往会通过优化某一架次或某几个特定架次的起降航航班实现运行效率的提升，这显然有失公允。因此，本文引入最小化最大进场飞行时间的目标，以期保障进离场调度的公平性，即

$$\min \quad \max\left(x_{i,k}\cdot\left(t_{i,k}^{\mathrm{TTOA}}-t_i^{\mathrm{TOE}}\right)\right) \tag{5}$$

式中：t_i^{TOE}为航空器器i进入终端空域的时间（time of entry, TOE），$t_{i,k}^{\mathrm{TTOA}}-t_i^{\mathrm{TOE}}$即为航空器器$i$使用剖面$k$到达跑道的飞行时间。

2.4 多目标优化方法

本文主要探究进场航空器连续下降运行在考虑离场航空器时的调度情况，而航空器的调度往往需要同时考虑多个目标，如公平性、效率等。为使得模型求解时能够在多个目标之间进行权衡，得到更加全面的优化结果，因此，考虑对前述2.3节目标函数赋予一定的优先级和权重系数，研究在不同优先级和权重条件下的调度情况。

考虑以下3种优化方式：

1）分层优化。将多个目标函数按照优先级顺序进行优化，先优化优先级较高的目标函数，在不降低较高优先级目标函数的前提下，优化次优先目标函数。

2）组合优化。将多个目标按照一定权重系数加和，模型会自然地偏向权重较高的目标函数。

3）混合优化。在分层优化中嵌套了组合优化，即在分层优化的某一层中以组合优化方法处理多目标函数。

3 连续爬升与下降运行调度仿真验证

3.1 场景与参数

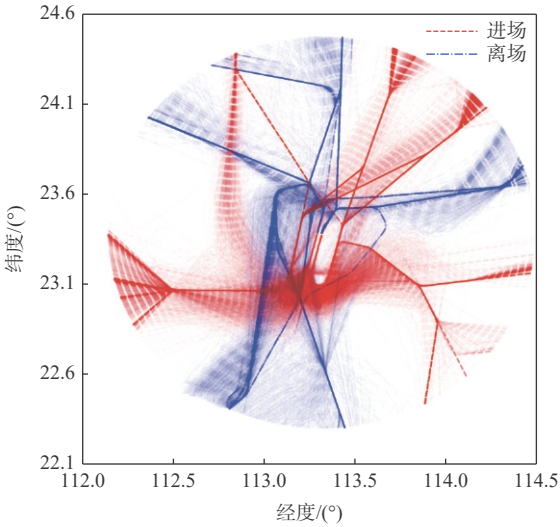
3.1.1 验证场景

以广州白云国际机场（ZGGG）为研究对象，对连续下降和爬升运行的进离场航空器进行调度。实例数据取自2019年12月3至4日某2个繁忙时段的进离场航空器雷达综合航迹数据（如图2(a)所示），包括航班号、机型、起飞机场、着陆机场、进离港点、计划起飞时间、实际起飞时间、计划着陆时间、实际着陆时间等字段。12月3日繁忙时段共80架次起降航班，其中进场36架，离场44架；12月4日繁忙时段共82架次起降航班，其中进场47架，离场35架。运行场景的进离场程序来源于广州白云国际机场的航行资料汇编（aeronautical information publication, AIP），如图2(b)所示。

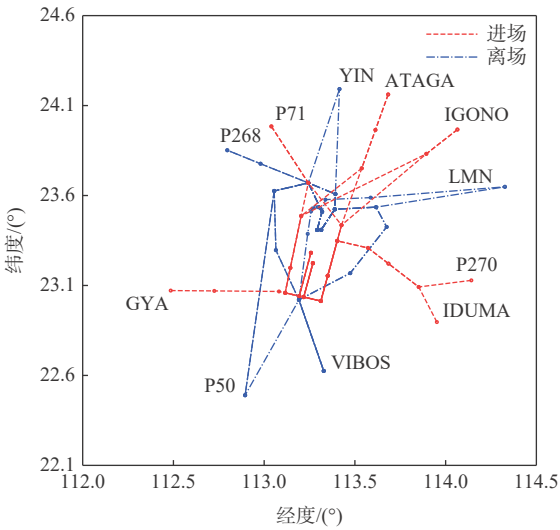
本文涉及的模型与算法应用Python 3.9编程实现，运行环境：Windows 11操作系统，八核AMD R7-5700H CPU、3.2 GHz主频和16 GB内存的微机。航迹优化部分使用的求解器为内置SNOPT的GPOPS 5.0工具箱，进离场调度部分使用的求解器为Gurobi 10.0。

3.1.2 参数设置

对于进场航空器的航迹优化，其航迹起始于以机场参考点为圆心、固定距离为半径的截取边界，终止于机场跑道。其初始状态沿用真实历史数据中的记录，即出现时间、高度、速度与历史数据一致；其终止位置为进场程序中的最后进近定位点（final approach fix, FAF），状态为高度2 000 ft、校正空速160节。鉴于航空器由FAF点至着陆跑道往往以3°下滑方式进近着陆，本文估算耗时约220 s。该参数的设置，方便推算航空器着陆时间，有助于后续进离场航空器跑道上间隔的计算。在航迹优化过程中，最小航径角的约束限制为最小−7°，以保



(a) 雷达综合航迹示意



(b) 进离场程序示意

图 2　广州白云国际机场验证场景示意

Fig. 2　Schematic diagram of ZGGG validation scenario

障所优化航迹实际可飞。

对于离场航空器的航迹优化, 其航迹起始于机场跑道, 终止于离港点。其初始状态沿用真实历史数据中的记录; 其终止高度根据离港点管制移交协议而定, 终止速度不做额外限制。起飞爬升推力需小于最大可用推力, 大于三档减推力(即 80% 最大可用推力), 具体参数由离场航空器的所属机型决定。

离场偏离时间是指航空器可以选择在实际起飞时间前后一定时间内择机起飞, 本文分别考虑以 1 min 为间隔、提前 4 min、延后 5 min 和以 0.5 min 为间隔、提前 2 min、延后 2.5 min 这 2 种离场偏离策略。如此, 2 种策略均可生成 10 个时刻作为离场备选起飞时间集合。

对于航空器间隔配备, 同样考虑 2 组参数, 具体如表 1 所示。其中, 起飞间隔配备 90 s, 着陆间隔及起降间隔配备 120 s 和 150 s; 空中间隔配备: 垂直 300 m 和 400 m, 水平 6 km 和 10 km。联合离场偏离时间, 共构建 4 组参数组合, 目的在于考察不同调配裕度和间隔裕度的进离场调度效果。

表 1　参数设置

Table 1　Parameters setting table

参数组合	离场偏离/min	跑道运行间隔			空中运行间隔	
		离场/s	进场/s	进离/s	垂直/m	水平/km
1	[−4:1:5]	90	120	120	300	6
2	[−4:1:5]	90	150	150	400*	10*
3	[−2:0.5:2.5]	90	120	120	300	6
4	[−2: 0.5:2.5]	90	150	150	400	10

注: *垂直间隔400 m与水平间隔10 km为本文仿真设置的比实际运行更严格的间隔标准。

由于本节主要验证本文方法的有效性, 因此对于间隔配备采取了简化处理。例如, 忽略航空器尾流等级对于起降运行的影响, 统一采取宽泛的间隔配备。再如, 忽略不同跑道之间起降运行的影响, 其依据在于: ①广州白云国际机场采取独立平行进近与独立平行离场的运行模式; ②后续终端空域进离场的冲突探测能够识别潜在的冲突航班。

3.2　航迹优化结果分析

以 B737-800 机型为例, 本节展示航迹优化的结果与分析, 图 3 为从终端区东北部进港点 ATAGA 进港的一架 B737-800 航班优化的最早和最晚到达的连续下降运行的垂直剖面示意, 该进港点对应 2 条距离不同的标准进场航路。

最早到达剖面(图 3(a))中, 航空器选择距离较短的进场航路, 先是等速平飞, 到达下降点后开始下降, 随后保持等校正空速(CAS)下降, 最后减速下降直至 160 kn 到达 FAF 点高度 2 000 ft。最早到达剖面(图 3(b))中, 航空器使用距离较长的进场航路进场, 到达下降点后减速下降至某一校正空速, 以该速度缓慢下降一段时间后, 再减速至 FAF 点(以满足优化的边界条件)。该示例中, 最早到达耗时 1 014 s, 最晚到达耗时 1 304 s。

图 4 是成本指数变化下, 该航空器飞行时间与燃油消耗的变化。

由图 4 可见, 成本指数在 0～200 kg/min 时, 随着成本指数的增加, 时间成本占目标函数的比重逐渐上升, 优化得出的燃油消耗逐渐增加, 飞行时间逐渐降低。当成本指数大于 200 kg/min 时, 所优化航迹已达最早到达剖面, 燃油和飞行时间不再随成本指数的增加而变化。

3.3　进离场调度结果分析

3.3.1　引入离场后的调度结果对比

本节对比仅进场调度[7-8]与进离场协同调度的结果, 数据采取 12 月 3 日繁忙时段 80 架次起降航

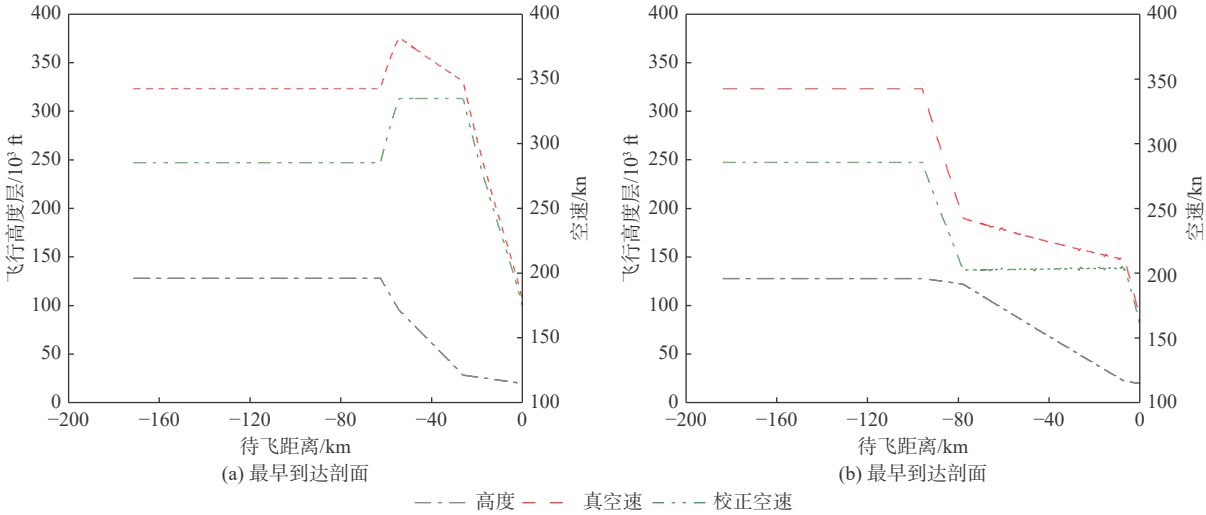


图 3 连续下降运行垂直剖面示意

Fig. 3 Diagram of the vertical profiles of CDO

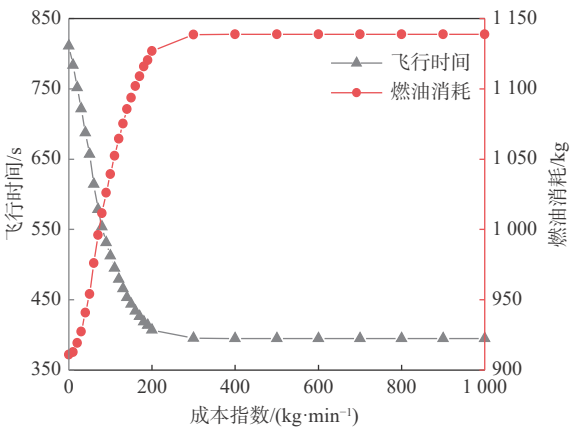


图 4 飞行时间与燃油消耗随成本指数的变化

Fig. 4 Variation of flight time and fuel consumption with different cost index

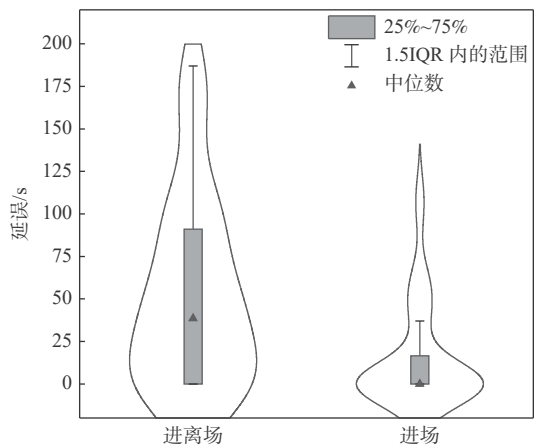


图 5 进离场调度与进场调度延误分布对比

Fig. 5 Variation of flight time and fuel consumption with different cost index

班，间隔标准采用表 1 中的参数组合 1，具体结果如表 2 所示。进一步，通过小提琴图展示进场延误分布的对比，如图 5 所示。

表 2 进场调度与进离场调度结果对比

Table 2 Arrival scheduling versus arrival and departure scheduling results s

调度类型	平均进场延误	平均离场偏离	最大飞行时间	平均飞行时间
进场	15.5		1 359.0	1 120.0
进离场	53.5	47.7	1 407.0	1 158.0

由表 2 可知，仅考虑进场调度，平均延误、最大飞行时间和进场平均飞行时间等指标都显著优于进离场协同调度的结果。其中，仅考虑进场调度的平均延误仅为 15.5 s，而考虑进离场的调度平均延误会高达 53.5 s。由图 5 可知，进离场协同调度最大延误达 187 s，而进场调度最大延误仅 114 s。因此，引入离场运行会对原有的进场调度产生较大影响，原本的进场调度为了满足离场航空器的约束做出了时间空间上的调整，导致进场飞行时间增加，进场延误提高。

实际运行中，航空器进离场运行同时发生，只将进场连续下降纳入考量，而不顾离场航空器对进场运行的影响，会使优化结果过于理想，难以客观反映现实环境中的调度表现。因此，引入离场运行，实现连续爬升与下降运行的协同调度优化，能够提升繁忙终端空域内同时实施连续爬升与下降运行的可行性。

3.3.2 不同优化方法的调度结果对比

本节考察不同优化方法如表 3 所示的调度结果的对比，数据依旧采取 12 月 3 日繁忙时段 80 架次起降航班，间隔标准采用表 1 中参数组合 1。具体结果如图 6 所示。

首先，考虑平均进场延误和与平均离场偏离的优化。a、b、c、d 这 4 种优化方法对进场延误权重

表 3　多目标优化的权重系数与优先级

Table 3　Weights & priorities of multi-objective optimization

优化方案	最小化平均进场延误式(3)		最小化平均离场偏离式(4)		最小化最大进场飞行时间式(5)	
	权重	优先级	权重	优先级	权重	优先级
a		首先		其次		
b	2.0		1.0			
c	1.0		1.0			
d	1.0		2.0			
e		其次		再次		首先
f	1.0		1.0			首先
g		再次		其次		首先
h				其次		首先

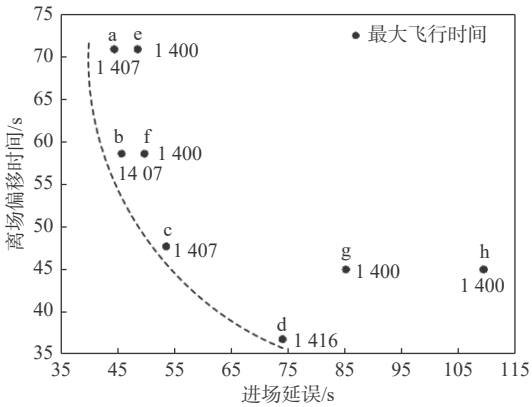


图 6　不同优化方法优化结果对比

Fig. 6　Comparison of the optimization results with

降低、离场偏离权重升高, 由图 6 可知, 这 4 种方法构成帕累托前沿, 随着二者权重和优先级的变化, 进场延误逐渐变大, 离场偏移逐渐变小。

其次, 将最大飞行时间纳入优化考量。e、f、g、h 这 4 种优化方法均优先最小化最大飞行时间, 在此基础上进行进场延误和离场偏离的优化。

由图 6 可知, 4 种方法的结果构成第二级的帕累托前沿。本案例中最小的最大飞行时间为 1 400 s, 与不考虑最大飞行时间的优化方法相比, 最大飞行时间减少 7～16 s。考察其对进场延误和离场偏离的优化情况, 可以发现, 最大飞行时间的降低往往是以更多的延误或偏离为代价的。例如, 使用方法 #e, 相较于方法 a 减少最大飞行时间(7 s), 然而会导致平均延误增加超过 4 s。该现象表明, 若降低最大飞行时间, 会造成终端空域内所有进离场航空器受到额外延误或偏离的影响。

最后, 对优化前后航空器所使用的跑道进行分析。原始数据与优化跑道分布如图 7 所示。

由图 7 可见, 2019 年 12 月 3 日的历史运行, 01 跑道作为进离场混合运行的跑道, 1 h 内共有 43 架次起降航空器。02L(仅离场)和 02R(仅进场)

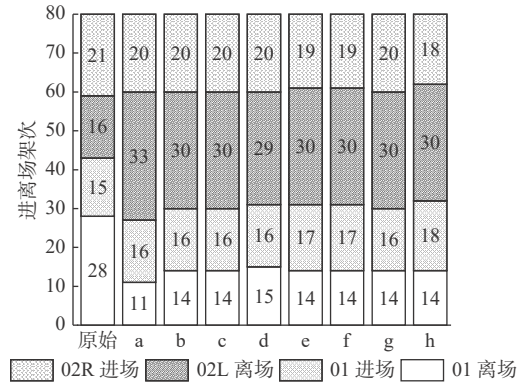


图 7　不同优化方法对应的跑道使用分布

Fig. 7　Distribution of runway use with different optimization methods

分别有 16 架次起飞和 21 架次降落。因此, 该时段的实际运行中管制员更倾向于使用 01 跑道完成航班起降运行。

本文提出的考虑连续爬升与下降运行的进离场调度模型, 在使用不同优化方法时, 跑道使用总体较为均衡。一方面, 将 01 跑道负责的离场航空器, 部分转移至负载较低的 02L 跑道, 充分利用了 02L 的跑道容量, 有利于实现连续爬升与下降运行, 进一步减少延误, 提高运行水平。另一方面, 02R 跑道在优化前后, 均保持了较为一致的着陆航空器数量, 这说明 02R 跑道的实际运行已经达到较为理想的运行状态。

3.3.3　不同间隔参数的调度结果对比

本节考察不同间隔参数(如表 1 所示)的调度结果的对比, 如表 4 所示, 数据采取 12 月 3 日繁忙时段 80 架次起降航班, 优化方法采用方案 c。

表 4　进场调度与进离场调度结果对比

Table 4　Arrival scheduling versus arrival and departure scheduling results　　s

参数组合	平均进场延误	平均离场偏离	最大飞行时间	平均飞行时间
1	53.5	47.7	1 407.0	1 158.0
2	115.6	100.9	1 512.0	1 220.1
3	65.1	27.3	1 462.0	1 169.6
4	110.9	64.1	1 512.0	1 211.3

分析表 4, 对比标准间隔(参数 1, 参数 3)和大间隔(参数 2, 参数 4)的优化结果可以发现, 空中和地面间隔的扩大, 导致进离场调度各项指标的降低。

对比疏(参数 1, 参数 2)密(参数 3, 参数 4)起飞偏移时间的优化结果可以发现, 离场偏移窗口的压缩和偏移间隔的缩小, 可以使得离场航空器找到更优的起飞时间, 即平均离场偏移时间更短。因为离场可选时间更密集, 离场航空器前后

调节的阶跃就越小，离场时间变化更平滑，有利于减少对进场航空器的影响，进而减少因为离场偏移间隔过大而导致的时间窗浪费。然而，进场延误随离场偏离参数的改变并不固定，较密的偏移可能会使进场延误增加。究其根本，在于进离场偏离与延误同权重优化，当离场时隙参数的改变使得离场偏离能够为相应目标函数带去更多的提升时，模型会通过增加进场延误来达到整体优化。对于造成进场变差的情况，可以考虑赋予进场更高的优先级或权重解决。

3.3.4 路径变化的调度结果对比

前述分析结果表明，本文提出的面向连续爬升与下降运行的进离场调度模型均能在标准进离场程序下完成优化调度，本节选取更为繁忙时段（12月4日，82架次起降航班，其中进场47架，离场35架），考察标准进离场程序下无法完成连续运行调度的解决方法及其效果。具体结果如表5所示，其中优化方法选择方案c。

表5 路径变化的进场调度与进离场调度结果对比

Table 5 Arrival scheduling versus arrival and departure scheduling results under the change of arrival routes

参数组合	进场路径	平均进场延误/s	平均离场偏离/s	最大飞行时间/s	平均飞行时间/s	无法调度/架
1	STAR	66.8	21.8	1 512.0	1 200.9	2
2	STAR	67.7	40.0	1 444.0	1 203.9	7
1	GYA-DL	86.3	21.8	1 549.0	1 223.9	0
2	GYA-DL	112.6	36.4	1 602.0	1 241.5	5
2	GYA-DL IDUMA-SC	117.3	36.4	1 602.0	1 241.5	5

由表5可知，采用标准进离场程序，针对更为繁忙时段的数据，基于标准间隔（参数1）与大间隔（参数2），均无法实现连续爬升与下降运行的进离场调度。具体而言，标准间隔下2架（GYA）航空器无法调度，大间隔下7架（来自GYA、IGONO、IDUMA）航空器无法调度。与此同时，随着间隔的加大，调度优化的指标也相应下降。鉴于此，为GYA进场程序设计一条进场偏离路径（dog-leg, DL），为IDUMA进场程序设计一条直飞路径（short-cut, SC），考察路径变化下的进离场调度优化结果，具体路径如图8所示。

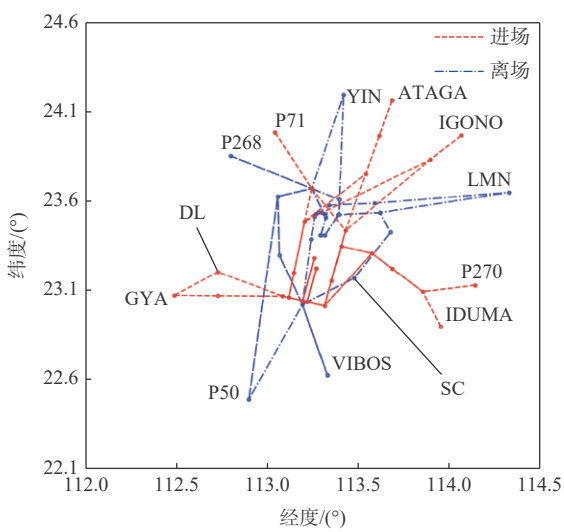


图8 进场路径设计示意

Fig. 8 Diagram of the arrival routes design

由表5可知，在标准进离场程序的基础上新增GYA的进场偏离路径，针对更为繁忙时段的数据，基于标准间隔（参数1），可以实现连续爬升与下降运行的进离场调度。当然，不可避免地会导致调度优化指标的下降。进一步，修改标准间隔（参数1）为大间隔（参数2），此时将会出现5架进场航空器无法调度。究其原因，GYA方向的进场偏离路径在大间隔下，只能解决机场西侧航空器调度问题，而对于东侧（IDUMA、IGONO）并没有帮助。

此外，本节引入IDUMA方向的直飞路径。由表5可知，并不能进一步降低无法调度的航空器数量，反而使得进场延误进一步恶化。这是因为本文对于进场延误的定义是实际到达时间减去最早到达时间，直飞路径的加入使得最早到达时间进一步提前，而由于空域十分繁忙，调度模型无法为

航空器安排更早的到达时间，航空器实际到达时间几乎未发生变化，因而使得延误有所提升。因此，对于解决无法调度的进场航空器，需要更多考虑“延长三边”和“进场偏离”等，而非允许航空器以捷径更快着陆。

4 结 论

1）融合航迹优化、冲突探测与进离场调度，构建了面向繁忙终端空域连续爬升与下降运行的调度优化模型。此外，以最小化进场平均延误、平均离场偏离、最大进场飞行时间为目标，同时满足了空管、航司、机场以及民众的不同诉求。

2）与仅考虑连续下降运行的进场调度相比，同时考虑连续爬升与下降运行的进离场调度的进场延误和飞行时间将会增大，表明连续爬升与下降运

行的进离场调度会相互影响, 本文方法更贴近繁忙终端空域的实际运行。

3) 多目标优化的权重选择可以得出不同偏好的调度结果。赋予进场延误较大的权重, 调度模型可以获得较低的进场延误; 赋予离场偏离时间较大的权重, 则可以获得较低的离场偏离时间。实际运行过程中可以根据具体的进离场交通需求进行权重设置。

4) 现行标准进离场程序难以满足特别繁忙时段下的进离场连续爬升与下降运行的实施, 通过增加进场备选路径(如“延长三边”“进场偏离”“直飞”等), 能够有助于连续爬升与下降运行在繁忙终端空域的实施。

5) 本文主要解决了单机场繁忙终端空域连续爬升与下降运行的调度优化, 未来的研究会扩展到多机场繁忙终端空域的运行场景。

参考文献（References）

[1] LIU J, ZHANG J F, DAI X M, et al. Research on trajectory generation and optimization in continuous descent operations[C]//Proceedings of the 2018 Aviation Technology, Integration, and Operations Conference. AIAA, 2018: 2018-3986.

[2] JIN L, CAO Y, SUN D F. Investigation of potential fuel savings due to continuous-descent approach[J]. Journal of Aircraft, 2013, 50(3): 807-816.

[3] DALMAU R. Optimal trajectory management for aircraft descent operations subject to time constraints[D]. Barcelona: Technical University of Catalonia, 2019: 5-6.

[4] 张军峰, 游录宝, 杨春苇, 等. 基于多目标帝国竞争算法的进场排序与调度[J]. 航空学报, 2021, 42(6): 324439.
ZHANG J F, YOU L B, YANG C W, et al. Arrival sequencing and scheduling based on multi-objective Imperialist competitive algorithm[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(6): 324439(in Chinese).

[5] LIANG M, DELAHAYE D, MARECHAL P. Conflict-free arrival and departure trajectory planning for parallel runway with advanced point-merge system[J]. Transportation Research Part C: Emerging Technologies, 2018, 95: 207-227.

[6] 杨磊, 李文博, 刘芳子, 等. 柔性空域结构下连续下降航迹多目标优化[J]. 航空学报, 2021, 42(2): 324157.
YANG L, LI W B, LIU F Z, et al. Multi-objective optimization of continuous descending trajectories in flexible airspace[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(2): 324157(in Chinese).

[7] HU R, FENG H L, WITLOX F, et al. Airport capacity constraints and air traffic demand in China[J]. Journal of Air Transport Management, 2022, 103: 102251.

[8] SÁEZ R, PRATS X, POLISHCHUK T, et al. Traffic synchronization in terminal airspace to enable continuous descent operations in trombone sequencing and merging procedures: An implementation study for Frankfurt airport[J]. Transportation Research Part C: Emerging Technologies, 2020, 121: 102875.

[9] SÁEZ R, POLISHCHUK T, SCHMIDT C, et al. Automated sequencing and merging with dynamic aircraft arrival routes and speed management for continuous descent operations[J]. Transportation Research Part C: Emerging Technologies, 2021, 132: 103402.

[10] GUI D D, LE M L, HUANG Z C, et al. Optimal aircraft arrival scheduling with continuous descent operations in busy terminal maneuvering areas[J]. Journal of Air Transport Management, 2023, 107: 102344.

[11] HONG Y, CHOI B, LEE K, et al. Dynamic robust sequencing and scheduling under uncertainty for the point merge system in terminal airspace[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(9): 2933-2943.

[12] KAMO S, ROSENOW J, FRICKE H, et al. Robust optimization integrating aircraft trajectory and sequence under weather forecast uncertainty[J]. Transportation Research Part C: Emerging Technologies, 2023, 152: 104187.

[13] TORATANI D, WICKRAMASINGHE N K, WESTPHAL J, et al. Feasibility study on applying continuous descent operations in congested airspace with speed control functionality: fixed flight-path angle descent[J]. Aerospace Science and Technology, 2020, 107: 106236.

[14] LIEDER A, BRISKORN D, STOLLETZ R. A dynamic programming approach for the aircraft landing problem with aircraft classes[J]. European Journal of Operational Research, 2015, 243(1): 61-69.

[15] 张军峰, 游录宝, 周铭, 等. 基于点融合系统的多目标进场排序与调度[J]. 北京航空航天大学学报, 2023, 49(1): 66-73.
ZHANG J F, YOU L B, ZHOU M, et al. Multi-objective arrival sequencing and scheduling based on point merge system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(1): 66-73(in Chinese).

[16] BENNELL J A, MESGARPOUR M, POTTS C N. Airport runway scheduling[J]. Annals of Operations Research, 2013, 204(1): 249-270.

[17] IKLI S, MANCEL C, MONGEAU M, et al. The aircraft runway scheduling problem: a survey[J]. Computers & Operations Research, 2021, 132: 105336.

[18] BENNELL J A, MESGARPOUR M, POTTS C N. Dynamic scheduling of aircraft landings[J]. European Journal of Operational Research, 2017, 258(1): 315-327.

[19] SAMÀ M, D'ARIANO A, D'ARIANO P, et al. Scheduling models for optimal aircraft traffic control at busy airports: tardiness, priorities, equity and violations considerations[J]. Omega, 2017, 67: 81-98.

[20] YANG Y C, GAO Z C, HE C. Stochastic terminal flight arrival and departure scheduling problem under performance-based navigation environment[J]. Transportation Research Part C: Emerging Technologies, 2020, 119: 102735.

[21] DU Z M, ZHANG J F, KANG B. A data-driven method for arrival sequencing and scheduling problem[J]. Aerospace, 2023, 10(1): 62.

[22] ZHANG J F, LIU J, HU R, et al. Online four-dimensional trajectory prediction method based on aircraft intent updating[J]. Aerospace Science and Technology, 2018, 77: 774-787.

[23] RAO A V, BENSON D A, DARBY C, et al. Algorithm 902: GPOPS, a MATLAB software for solving multiple-phase optimal

control problems using the Gauss Pseudo-spectral method[J]. ACM Transactions on Mathematical Software, 2010, 37(2): 1-39.

[24] GILL P E, WONG E, MURRAY W, et al. User's guide for SNOPT version 7.5: software for large-scale nonlinear programming[R]. San Diego: University of California, 2015.

[25] GUI X H, ZHANG J F, TANG X M, et al. A data-driven trajectory optimization framework for terminal maneuvering area operations[J]. Aerospace Science and Technology, 2022, 131(1): 108010.

[26] ZHANG J F, ZHAO P L, ZHANG Y, et al. Criteria selection and multi-objective optimization of aircraft landing problem[J]. Journal of Air Transport Management, 2020, 82: 101734.

Scheduling optimization for continuous climb and descend operations in busy terminal area

DU Zhuoming，ZHANG Junfeng*，GUI Xuhao

(College of Civil Aviation，Nanjing University of Aeronautics and Astronautics，Nanjing 210016，China)

Abstract：Based on the current terminal airspace structure, a novel scheduling method for arrival and departure aircraft based on trajectory optimization, conflict detection, and multi-objective optimization is proposed to assist in continuous climb and descent operations in busy terminal airspace. Firstly, a vertical profile optimization method for continuous climb and descent operations is proposed based on multi-stage optimal control theory and the Gaussian pseudo-spectral method, achieving trajectory optimization for continuous climb and descent operations by cost index. Secondly, according to the wake separation and clearance separation used by the runway operations and the horizontal and vertical separation of air operations, an aircraft conflict detection model is established using Mahalanobis distance. Subsequently, considering the demands of operational units such as air traffic control, airlines, and airports, a multi-objective scheduling model and method with achievable optimization results were proposed for arrival and departure aircraft. Finally, two sets of arrival and departure data from Guangzhou Baiyun Airport during peak hours were selected, multiple interval parameters were set, alternative paths were designed, and case analysis and comparative research were conducted. The results indicate that during peak departure hours, the terminal airspace of Guangzhou Baiyun Airport can achieve continuous climb and descent operations during peak hours. Furthermore, during peak arrival hours, two aircraft cannot be scheduled. Introducing alternative paths can reduce the number of aircraft that cannot be scheduled.

Keywords：air traffic management; continuous descent operation; trajectory optimization; aircraft scheduling; multi-objective optimization

Received：2023-06-28；**Accepted**：2023-08-25；**Published Online**：2023-09-01 18：01

URL：link.cnki.net/urlid/11.2625.V.20230901.1703.005

Foundation items：National Natural Science Foundation of China (52372315, U1933117); Postgraduate Research & Practice Innovation Program of NUAA (xcxjh20220714)

* **Corresponding author**. E-mail：zhangjunfeng@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0429

基于改进 TCN-Elman 神经网络的电离层杂波抑制方法

刘强，尚尚*，乔铁柱，祝健，石依山
(江苏科技大学 海洋学院，镇江 212003)

摘 要：高频地波雷达因其卓越的海面目标探测能力，被世界各国应用于海上工程领域，而提升其目标探测能力的关键要素之一在于回波信号中电离层杂波的抑制，针对这一现象，提出一种基于瓶颈膨胀卷积模块改进时序卷积 (ITCN)-Elman 神经网络结合混合注意力机制的电离层杂波预测抑制模型 (Mixatt-ITCN-Elman)。对电离层杂波时间序列进行相空间重构和乱序归一化，利用 ITCN 提取高维相空间内的空间特征，依据自注意力机制突出其中关键的空间特征，将空间特征与原时间序列组合输入 Elman 神经网络，结合注意力机制突显序列的空时特征，通过空时特征与 Elman 神经网络输出序列组合输出，得到最终预测结果。所提模型与 Elman、TCN、Att-CNN-Elman 和 TCN-Elman 模型相对比，具有较好的预测性能和稳定性，对于电离层杂波的抑制具有较高应用价值。

关 键 词：高频地波雷达；电离层杂波；时序卷积网络；Elman 神经网络；注意力机制
中图分类号：TN957.54
文献标志码：A **文章编号**：1001-5965(2025)09-3203-09

高频地波雷达 (high frequency surface wave radar, HFSWR) 作为现代海洋探测监控遥感领域的主要新兴技术之一，利用高频频段 (3～30 MHz) 无线电波实现导电海平面绕射传播，具备一系列工程优势 (如传播衰减小速度快、探测超视距大范围、检测全天候全时段、造价成本低以及频段反隐身等)，在军民应用领域有着巨大的实用价值[1-2]。实际工程应用中，由于海上情况的时变性和突发性，其发射的垂直极化电磁波受多种客观因素 (如大气、同频段无线电等) 的影响而产生不可控散射，使得雷达目标回波中掺杂着多种干扰，其中电离层杂波干扰问题尤为显著，因其较高的反射能量和广泛的距离单元覆盖范围，对雷达弱目标检测性能产生了至关重要的影响。因此，如何有效抑制电离层杂波干扰成为热点方向并持续开展研究。

纵观高频地波雷达电离层杂波抑制方法研究，根据抑制处理方法不同，分为电离层杂波基础特性层面抑制和电离层杂波建模层面抑制。电离层杂波基础特性层面抑制分为单通道杂波处理和多通道杂波处理。单通道杂波处理适用于某些简单情况以及特定需求，当杂波分布更为复杂或与目标特性相近时，则达不到预期效果，造成抑制不彻底或目标丢失，如子空间分解类[3]、时频特性类[4]和小波分解类[5]。子空间分解类通过特征值划分噪声和信号子空间，实现特定范围特征值的杂波抑制，与目标特征值相近的电离层杂波抑制效果不佳；时频特性类基于时域和频域相关联，适用于单一情况，复杂情况下易造成目标丢失；小波分解类通过多尺度分析实现抑制，对于平稳信号抑制效果好，反之抑制能力受限，目标信号受损。多通道杂波处理解决

收稿日期：2023-06-30；**录用日期**：2023-11-03；**网络出版时间**：2023-11-13 16:44
网络出版地址：link.cnki.net/urlid/11.2625.V.20231110.1522.002
基金项目：国家自然科学基金 (61801196)；江苏省研究生科研与实践创新计划资助项目 (SJCX23_2138)
* **通信作者**. E-mail：shangshang@just.edu.cn

引用格式：刘强，尚尚，乔铁柱，等. 基于改进 TCN-Elman 神经网络的电离层杂波抑制方法 [J]. 北京航空航天大学学报，2025，51(9)：3203-3211. LIU Q，SHANG S，QIAO T Z，et al. Ionospheric clutter suppression method based on improved TCN-Elman neural network[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(9)：3203-3211 (in Chinese).

了多数单通道杂波处理的问题，但主要依赖于系统阵列的大小，对初始系统的要求高，排列复杂，成本昂贵，如旁瓣对消类[6]、自适应波束形成类[7]、空时自适应类[8]、极化滤波类[9]和偏振滤波类[10]。旁瓣对消类利用辅助通道加权对消，抑制效果受目标方向性影响；自适应波束形成类多用于多通道杂波处理，调整波束实现抑制，需要高算法和计算资源；空时自适应类通过多通道信息调整参数，提高信号质量，但计算维度高且耗时，不利于实时处理；极化滤波类利用垂直和水平极化实现抑制，但系统复杂且成本高；偏振滤波类通过偏振信息实现抑制，但硬件要求高且处理时间长，工程实现困难。

在电离层杂波建模层面抑制，通过对电离层杂波信号建模，实现电离层杂波抑制。依据电波传播理论和光线追踪结合多普勒光谱分析搭建模型，实现自适应识别处理，但模型的准确性取决于数据集的多少，计算复杂度较高，如电离层杂波模型[11]、电离层杂波功率模型[12]；依据电离层不同层高所产生电离层杂波统计特性的不同，结合智能杂波分类，实现点对点分类抑制，效果高于单一方法处理，但不同统计特性需使用不同抑制方法，系统较为复杂，如电离层杂波分类建模[13]、马尔可夫跳跃电离层高度偏差建模[14]；依据电离层杂波的混沌特性，采用非线性预测类方法建立混沌动力学模型，从时间序列上进行建模，实现电离层杂波的抑制，预测精度高，抑制效果好，具有强大的计算能力以及时变适应能力，如电离层杂波混沌模型[15]。

近年来，随着深度学习和电子硬件的爆发式增长，深度学习成功应用于混沌时间序列预测，相较于传统神经网络，具备更高的预测精度、更强的计算能力以及更优的时变适应能力等优势。文献[16]提出一种改进后时域卷积模型，在保证精度的前提下，提高了训练速度，网络更加稳定，但由于模型使用了大量的卷积模块，适用于数据量大的预测情景；文献[17]提出一种注意力机制优化混合神经网络的时序预测模型，在高维和低维的问题处理上，都具有不错的表现；文献[18]利用混沌理论与Elman神经网络组合，提高海水冷却系统参数的短期预测精度，为航运智能化提供了新的研究方向。

相较于现有电离层杂波抑制方法，受到电离层空时变化的限制，可能影响雷达弱目标检测的性能。伴随着深度学习的爆发式增长，电离层杂波混沌模型对非线性数据具有更好的拟合逼近能力、计算能力以及稳定性。为了进一步提高电离层杂波抑制效果，深入研究电离层杂波特性，本文提出一种基于改进时域卷积(improved temporal convolutional network, ITCN)-Elman神经网络结合混合注意力机制(mixed attention mechanism)的电离层杂波预测抑制模型(Mixatt-ITCN-Elman)，其中引入了瓶颈膨胀卷积残差模块，自注意力机制(self-attentive mechanism)和注意力机制(attention mechanism)，运用ITCN和Elman神经网络实现电离层杂波复杂时间序列预测抑制。该方法首先将雷达回波信号中电离层杂波时间序列进行数据预处理，即相空间重构和乱序归一化；其次，利用自注意力机制自迭代突显卷积放大后的空间特征；最后，利用注意力机制提取混合模型的空时特征，得到最终预测结果。实验结果表明，Mixatt-ITCN-Elman模型能有效实现电离层杂波时间序列的预测抑制，较其他模型具有更好的预测精度、网络梯度结构和时变适应能力，优化了电离层杂波抑制效果。

1 数据预处理

1.1 电离层杂波混沌特性验证

本文采用主成分分析法(principal component analysis, PCA)和最大李雅普诺夫(Lyapunov)指数法[19]分别从定性和定量2个方面对电离层杂波时间序列的混沌特性进行验证。主成分分析法首先通过计算得电离层杂波时间序列相空间矩阵$\boldsymbol{X}_{l\times m}$($l$为轨道维数，$m$为嵌入维数)的协方差矩阵$\boldsymbol{C}$，其次计算得$\boldsymbol{C}$的特征值$\lambda$，按从大到小排序并对其求和得$\gamma$。以嵌入维数$m$为$x$轴，$\ln(\lambda_i/\gamma)$为$y$轴绘出主成分分析图如图1所示，电离层杂波主成分在图上呈一条斜率为负的直线，斜率$k=-0.15$，从而定性地判断电离层杂波时间序列具有混沌特性。

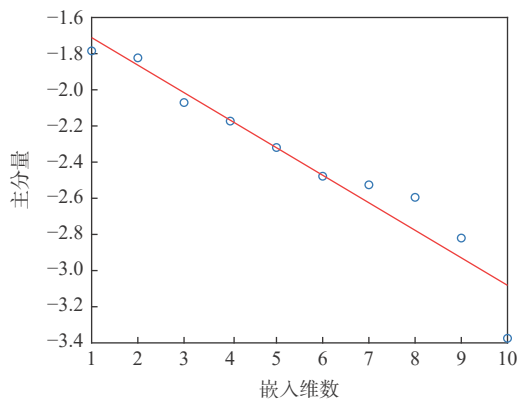


图1 电离层杂波主成分分析图

Fig. 1 Principal component analysis of ionospheric clutter

Lyapunov指数法通过Wolf方法计算最大Lyapunov指数，意为系统相邻轨线在相空间中随时间演化是收敛或是发散的平均指数率，定义为

$$\lambda_1=\frac{1}{t_m-t_0}\sum_{k=1}^{m}\log_2\frac{L(t_k)}{L(t_{k-1})} \tag{1}$$

式中：m为最大迭代次数；$L(t_k)$为邻近两点在t_k时刻

间距。计算得到电离层杂波最大 Lyapunov 指数 $\lambda = 0.379\,5 > 0$, 从而定量地判断电离层杂波时序序列具有混沌特性。综上所得, 判断电离层杂波时序序列具有混沌特性。

1.2　电离层杂波相空间重构

为有效提取时间序列动力系统模型内包含的信息, 通过相空间重构将一维时间序列拓展至高维相空间, 近似恢复出反映原始混沌系统演化特征的混沌吸引子。本文采用 C-C 方法[20] 得到相空间重构所需要的嵌入维数 m 和时间延迟 τ。

假设, 电离层杂波一维时间序列 $\boldsymbol{x}(n)$ 为

$$\{\boldsymbol{x}(n)|n = 1,2,\cdots,N\} \tag{2}$$

经相空间重构, 将一维时间序列扩展至 m 维相空间, 得高维相空间序列 $\boldsymbol{X}$:

$$\boldsymbol{X}_i = \{\boldsymbol{x}_i, \boldsymbol{x}_{i+\tau}, \cdots, \boldsymbol{x}_{i+(m-1)\tau} | i = 1,2,\cdots,N-(m-1)\tau\} \tag{3}$$

根据 Haykin[21] 所提出的预测方程观点, 电离层杂波预测方程为

$$\boldsymbol{x}_{i+m\tau} = F(\boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \cdots, \boldsymbol{x}_{i+m\tau-1}) \tag{4}$$

1.3　数据乱序归一化

在深度学习领域中, 不同评价指标具有不同的量纲和量纲单位, 数据归一化处理可有效消除奇异样本数据所导致的影响, 在一定程度上提高模型的训练速度和收敛精度, 而乱序处理能有效提高模型的泛化能力, 防止模型产生“过拟合”现象。公式如下:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{5}$$

$$x'' = \{x'(i) | i = \text{randperm}(N)\} \tag{6}$$

1.4　电离层杂波抑制

高频地波雷达回波信号 R-D 谱中, 同一片电离层杂波云团内部非线性动力系统相似, 利用相邻距离单元中不含目标的电离层杂波时间序列作为训练样本建立预测模型, 因电离层杂波的动力系统与目标信号不同, 通过将原始距离单元雷达回波与预测值对消实现电离层杂波的抑制, 实现目标信号的保留。电离层杂波预测抑制模型的建立过程如图 2 所示。

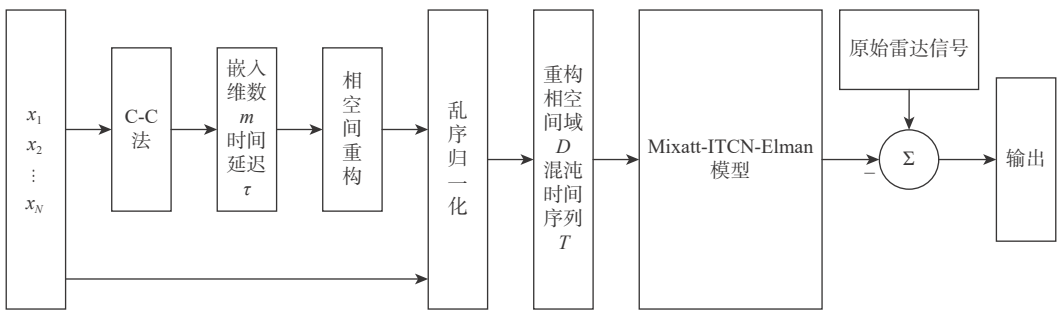


图 2　电离层杂波抑制流程

Fig. 2　Flow chart of ionospheric clutter suppression

步骤 1　选取距离单元中不含目标的电离层杂波时间序列作为训练样本输入, 并对其进行数据预处理。

步骤 2　设置训练误差容限和最大训练迭代次数, 将重构相空间 D 和时间序列 T 输入 Mixatt-ITCN-Elman 模型中迭代训练, 得到电离层杂波预测模型。

步骤 3　将原始距离单元雷达回波数据与预测值相对消, 实现雷达回波中电离层杂波的抑制。

2　Mixatt-ITCN-Elman 模型

Mixatt-ITCN-Elman 模型如图 3 所示, 不同网络处理侧重点不同, 即最后产生的作用也不同。预处理后的高维相空间序列既包含原始混沌系统的空间特征, 也蕴含着大量的时间特征。将一维时间序列重构至高维相空间, 放大序列的空时特征, 通过乱序归一化消除尺度效应, 利用瓶颈膨胀卷积残差模块和残差连接以提高模型的非线性映射能力, 更好地适应数据中的复杂特征, 用于捕捉电离层杂波数据中的长期和短期变化, 通过 ITCN 模型提取高维相空间内的空间特征, 运用自注意力机制自迭代突出序列中关键的空间特征, 运用 Elman 模型的短时记忆功能提取空间特征下的时间特性和高维相空间的关联性, 并采用注意力机制捕获突出模型训练后的关键信息, 使得模型更具有适应性和泛化能力, 能够更好地解决电离层杂波的复杂性问题。模型详细描述如下。

2.1　TCN 模块

针对传统神经网络无法大规模并行处理的弊端, 时序卷积网络(temporal convolutional network, TCN)模型以卷积神经网络为基础, 通过扩展感受野和堆叠参残差块来增强其性能, 具备多项优点, 其中包括有效捕捉时间序列之间的依赖关系、避免梯度弥散以及并行计算输出等。其核心组成部分

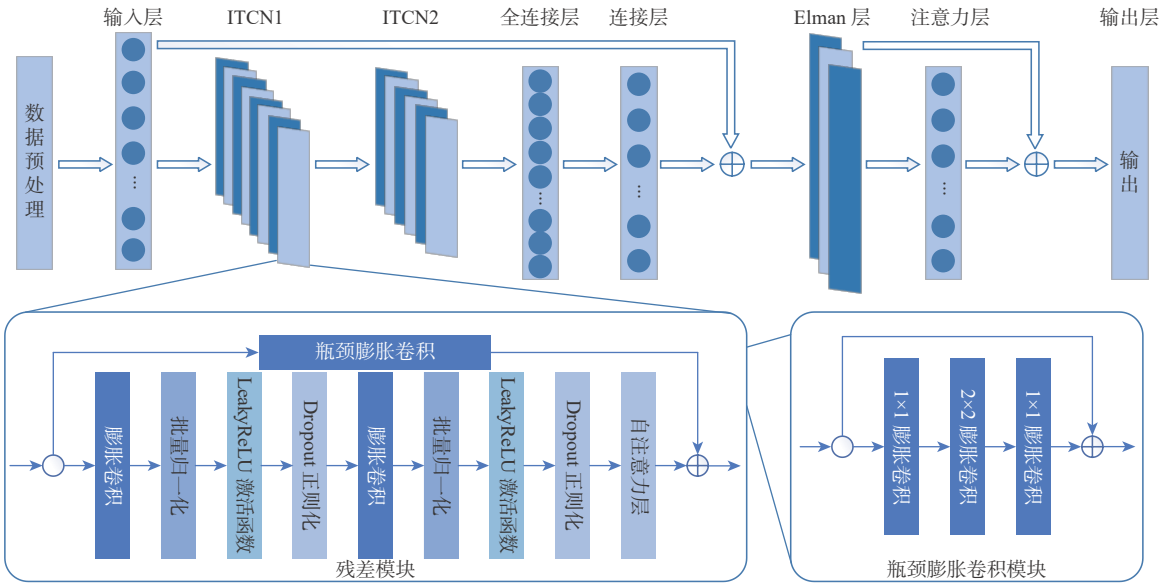


图 3 Mixatt-ITCN-Elman 模型

Fig. 3 Mixatt-ITCN-Elman model

包括因果卷积、膨胀卷积和残差模块，模块介绍如下。

2.1.1 因果卷积

相较于卷积神经网络 (convolutional neural networks, CNN) 的前后时间信息同时提取，因果卷积与循环神经网络 (recurrent neural networks, RNN) 相类似，为单向结构，如图 4 所示。为确保未来时间序列信息不会泄露到过去，且输入序列长度与输出序列长度一致，对输入序列进行填充，以遵循因果关系规则。因需要感知视野联系前向时间序列，在处理长序列问题时，不可避免地要求网络具有更深的结构或更大的滤波器，网络层数的不断累积，使得训练更加复杂。

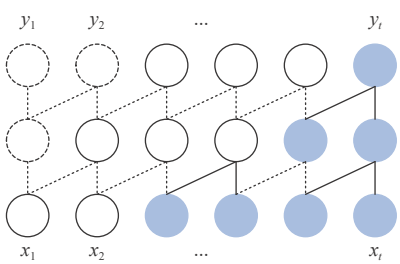


图 4 因果卷积结构

Fig. 4 Causal convolution structure

2.1.2 膨胀卷积

为适应各种类型的序列，得到更大的感知视野，在因果卷积的基础上加入膨胀因子，以较少的层数获得较大范围的序列信息。对于一维时间序列 $\{x_1, x_2, \cdots, x_t\}$，滤波器 $\{f_1, f_2, \cdots, f_k\}$，在 x_m 处的膨胀卷积公式为

$$F(m) = (x *_d f)(m) = \sum_{n=0}^{k-1} f(n) \cdot x_{m-d \cdot n} \tag{7}$$

式中：$*$ 为卷积运算；d 为膨胀因子，呈指数增长 ($d = 2^k$)；k 为卷积核大小；$x_{m-d \cdot n}$ 为前向联系序列，由于膨胀卷积输入输出特性，每一层都得在序列最左端补值，数量为 $(k-1)d$。如图 5 所示，卷积核 $k = 2$，膨胀因子 $d = [1, 2]$。

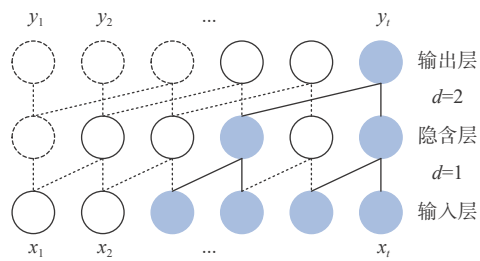


图 5 膨胀卷积结构

Fig. 5 Expansion convolution structure

2.1.3 残差模块

随着网络层数的不断叠加，带来种种好处的同时，问题也随之而来，如梯度范数过大、梯度弥散以及训练集精度下降等，残差连接便可有效解决这些问题。如图 3 所示，其结构类似于长短期记忆网络 (long short-term memory, LSTM) 中控制门，网络权重的链式反向传播学习由 $x \to f(x)$ 映射减少为 $x \to (f(x) - x)$ 映射，使得特征学习更为简单，优化了模型的泛化能力。

2.2 注意力机制

在深度学习领域，注意力机制模仿人类认知注意力被提出，用于提取模型特征，提高模型精度和训练速度，广泛应用于计算机视觉领域和自然语言处理领域等[22]。本文将注意力机制应用于提取 Mixatt-ITCN-Elman 模型中的关键空时特征，优化模型的训练精度，以提高模型的精确性和稳定性。如

图 6 所示, 注意力机制可以理解为一种对于序列权重赋值求和的全连接网络, 主要负责对序列特征向量的加权求和, 从而提取关键特征向量$\boldsymbol{c}$, 计算公式如下:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{n}\exp\left(e_j\right)} \tag{8}$$

$$\boldsymbol{c} = \sum_{i=1}^{n}\alpha_i \boldsymbol{V}_i \tag{9}$$

式中: α为$\boldsymbol{V}$的注意力权重; e为个体序列相似性, $e_i = \mathrm{MLP}(\boldsymbol{K}, \boldsymbol{L})$, 其全连接层激活函数为 tanh, $\boldsymbol{K}$为参量 Query, $\boldsymbol{L}$为参量 Value; n为 Elman 网络输入步长; $\boldsymbol{v}$为 Elman 网络输出向量。

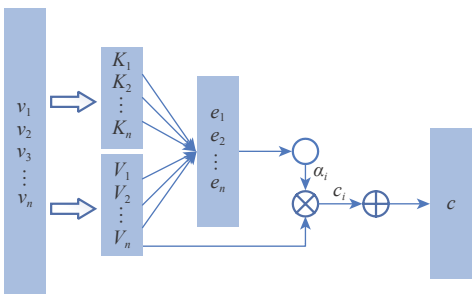


图 6　注意力模型

Fig. 6　Attention model

2.3　自注意力机制

自注意力机制为注意力机制中变体分支, 其不仅包含个体序列的主体信息, 也包含和其他序列的关系信息, 为目标序列内部元素互相做注意力机制, 因其权重是动态变化的, 具有适应度高以及计算速度快等优势[23]。本文将自注意力机制应用于残差模块中, 用于捕获个体序列之间的关系信息, 有效优化了空间特征提取的准确性和全面性。如图 7 所示, 自注意力机制与注意力大体流程类似, 但内部个体序列相似性e计算方式不同, 计算公式如下:

$$e_{ij} = \frac{Q_i K_j}{\sqrt{d}} \tag{10}$$

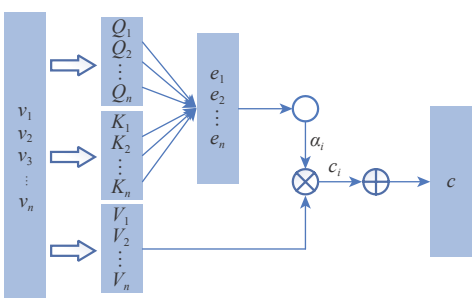


图 7　自注意力模型

Fig. 7　Self-attention model

式中: d为Q和K的维度, K为参量 Key。

2.4　ITCN 模块

图 3 列出了 ITCN 模块的结构框架, 在 TCN 模块的基础上引入自注意力机制, 通过自注意力机制自迭代捕获高维相空间序列的空间特征, 提升模型的自适应能力和特征提取能力。同时, 将残差模块中一维卷积模块结合瓶颈残差[24]和膨胀卷积生成瓶颈膨胀卷积模块, 用于灵活改变序列维度, 提升感知视野, 提高模型的计算能力和预测精度, 从而提高模型的稳定性和精确性。

2.5　Elman 模块

Elman 神经网络又称局部反馈神经网络, 如今已成功应用于各个领域[25-26]。相对于传统反向传播神经网络增加了一个具有短期记忆功能的特殊结构层, 即承接层, 有效提高了网络的非线性处理能力、时变适应能力以及全局稳定性。结构如图 8 所示, 数学模型如下:

$$\boldsymbol{x}(k) = \boldsymbol{f}(\boldsymbol{w}_2\boldsymbol{x}_c(k) + \boldsymbol{w}_1\boldsymbol{u}(k) + \boldsymbol{b}_1) \tag{11}$$

$$\boldsymbol{x}_c(k) = \boldsymbol{x}(k-1) \tag{12}$$

$$\boldsymbol{y}(k) = g(\boldsymbol{w}_3\boldsymbol{x}(k) + \boldsymbol{b}_2) \tag{13}$$

式中: k为当前时刻状态; $\boldsymbol{y}$、$\boldsymbol{x}$、$\boldsymbol{u}$、$\boldsymbol{x}_c$分别为输出向量、隐含层节点向量、输入向量和承接层节点向量; $\boldsymbol{w}$为各层之间权重矩阵; $\boldsymbol{b}$为各层阈值; f为 tanh 函数; g为 purelin 函数。

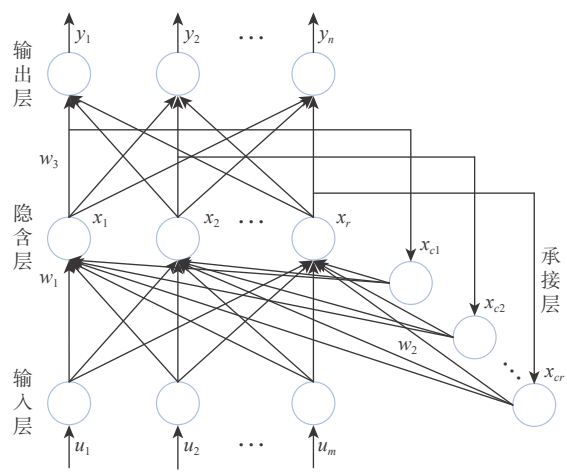


图 8　Elman 模块

Fig. 8　Elman module

3　实验分析与验证

本文中, 用于实验的系统配置包括 AMD Ryzen7 5800H with Radeon Graphics GPU、Nvidia GeForce RTX 3050 Ti GPU 和 Python(3.8 版本)。实验选用工作频率 5.1 MHz, 调制周期 49.77 ms, 脉冲宽度 0.42 ms, 脉冲重复周期 2.37 ms, 基带信号采样频率 0.1 MHz

的高频地波雷达(哈尔滨工业大学威海实验站)实测数据进行结果验证。由于雷达回波信号具有相位信息,需分别构建实部和虚部模型。下面实验以实部数据模型为例,虚部模型同理。根据C-C方法计算获得电离层杂波时间序列嵌入维数$m=3$,时间延迟$\tau=5$,故而可得Elman神经网络的输入层节点为15,隐含层节点为5,承接层节点为5,输出层节点为1。

电离层杂波预测模型的预测精度评估计算公式如下:

$$\sigma=1-\frac{\mathrm{var}(\boldsymbol{y}_i-\hat{\boldsymbol{y}}_i)}{\mathrm{var}(\hat{\boldsymbol{y}}_i)} \tag{14}$$

式中:var()为方差;$\hat{\boldsymbol{y}}_i$为时间序列预测值;$\boldsymbol{y}_i$为时间序列真实值。

为验证Mixatt-ITCN-Elman模型对于电离层杂波相邻距离单元的预测能力,利用Elman模型、TCN模型、TCN-Elman模型以及加入注意力机制的CNN-Elman模型形成对比实验,为评价模型的预测性能,本实验采用4个评价指标进行具象化评估,分别为均方根误差(RMSE, R)、平均绝对误差(MAE, M)、平均绝对百分比误差(SMAPE, S)和斯皮尔曼相关系数(ρ),公式分别如下:

$$R=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\boldsymbol{y}}_i-\boldsymbol{y}_i)} \tag{15}$$

$$M=\frac{1}{n}\sum_{i=1}^{n}|\hat{\boldsymbol{y}}_i-\boldsymbol{y}_i| \tag{16}$$

$$S=\frac{100\%}{n}\sum_{i=1}^{n}\frac{|\hat{\boldsymbol{y}}_i-\boldsymbol{y}_i|}{(|\hat{\boldsymbol{y}}_i|-|\boldsymbol{y}_i|)/2} \tag{17}$$

$$\rho=1-\frac{6\sum\boldsymbol{d}_i^2}{n(n^2-1)} \tag{18}$$

式中:n为预测样本长度;$\boldsymbol{d}_i$为时间序列位次差。

选取雷达回波数据第75距离单元中不含目标信号的电离层杂波构建1 024组实验训练样本集,将第66距离单元构建实验测试样本集,每个模型独立运行30次取平均值,各个模型训练误差指标结果如表1所示。

如表1所示,Mixatt-ITCN-Elman与Elman、TCN、Att-CNN-Elman和TCN-Elman模型对比,表中各项误差指标均为最优,即本文所提出的Mixatt-ITCN-Elman模型预测精度准确性高且稳定性较强。值得注意的是,Mixatt-ITCN-Elman与TCN-Elman模型相比,得到了进一步的提升,说明瓶颈膨胀卷积模块能有效提高模型的计算能力,同时自注意力机制能进一步突出序列的空间特征,注意力机制能捕获突出模型训练后的关键信息,从而提高模型预测值。单一模型TCN优于Elman,说明时间序列相空间重构到高维空间后,从空间特征方面进行预测要胜于时间特征。TCN-Elman模型相对于单一模型TCN和Elman来说,各项误差指标得到了一定程度的提升,说明Elman神经网络能有效提取时间序列空时特征下的时间特征,优化网络的预测性能。

表1 模型训练误差指标对比

Table 1 Comparison of model training error metrics

模型	R	M	S/%	ρ
Elman	0.053 0	0.042 6	7.203 8	0.968 2
TCN	0.030 5	0.020 3	3.895 8	0.993 2
Att-CNN-Elman	0.020 8	0.015 7	2.543 2	0.997 2
TCN-Elman	0.020 1	0.013 8	2.155 1	0.997 4
Mixatt-ITCN-Elman	0.010 0	0.007 6	1.145 5	0.999 2

如表2所示,在100次epochs迭代训练下,单一模型Elman和TCN训练时间较短,为6.28 s和14.74 s,但精度与混合模型Att-CNN-Elman、TCN-Elman和Mixatt-ITCN-Elman相比预测精度较低。混合模型Att-CNN-Elman、TCN-Elman和Mixatt-ITCN-Elman中TCN-Elman模型训练时间最长,为20.42 s,Mixatt-ITCN-Elman模型预测精度最高,训练时间相较于TCN-Elman模型短了1.72 s,相对于Att-CNN-Elman模型长了2.74 s,预测精度提高为99.66%,综合各个模型的训练误差指标、训练时间和训练精度对比,Mixatt-ITCN-Elman模型综合优势较高,具有较好的预测性能。

表2 模型训练时间和预测精度对比

Table 2 Comparison of model training time and prediction accuracy

模型	训练时间/s	预测精度/%
Elman	6.28	90.56
TCN	14.74	96.88
Att-CNN-Elman	15.96	98.55
TCN-Elman	20.42	98.64
Mixatt-ITCN-Elman	18.70	99.66

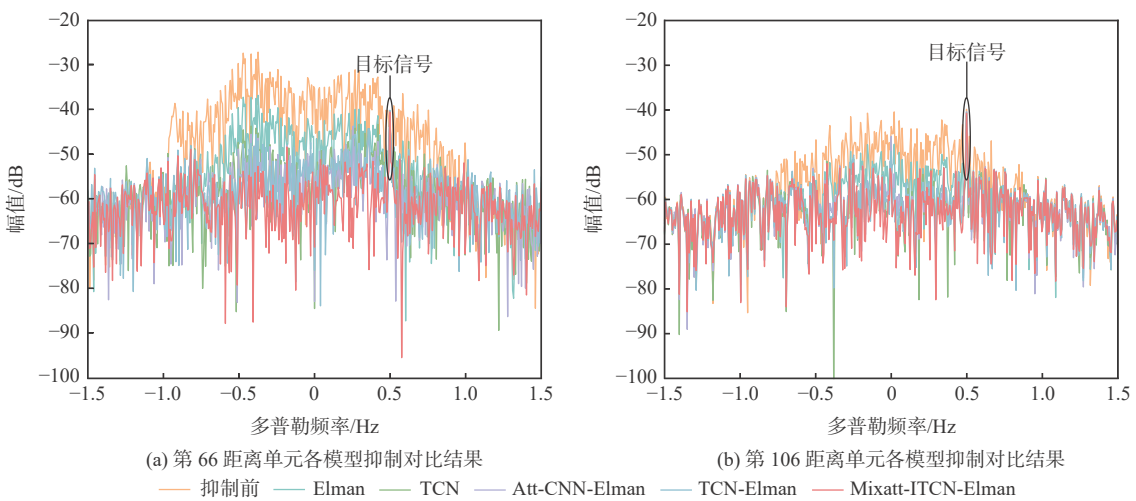
为进一步验证模型对于电离层杂波的抑制效果和抑制过程中对于目标信号是否产生影响,分别在第66、86和106距离单元中加入一个多普勒频率为0.5 Hz的目标信号。由表3和图9所示,第66距离单元中,目标信号被完全淹没,极大地影响了后续雷达目标检测的性能,经单一模型Elman和TCN抑制后,电离层杂波得到一定程度的抑制,但预测精度较低,抑制效果较差,信杂比较处理之前分别提升了8.78 dB和14.26 dB,然而,电离层杂波幅值仍与目标信号幅值相近,无法有效识别出目标

表 3　模型不同距离单元信杂比对比

Table 3　Comparison of signal-to-clutter ratio for different distance units of the model

模型	信杂比/dB		
	66距离单元	86距离单元	106距离单元
原始	−2.96	1.14	9.27
Elman	5.82	10.22	17.65
TCN	11.30	19.21	25.19
Att-CNN-Elman	12.53	17.15	24.72
TCN-Elman	12.58	19.70	24.83
Mixatt-ITCN-Elman	15.55	25.81	30.51

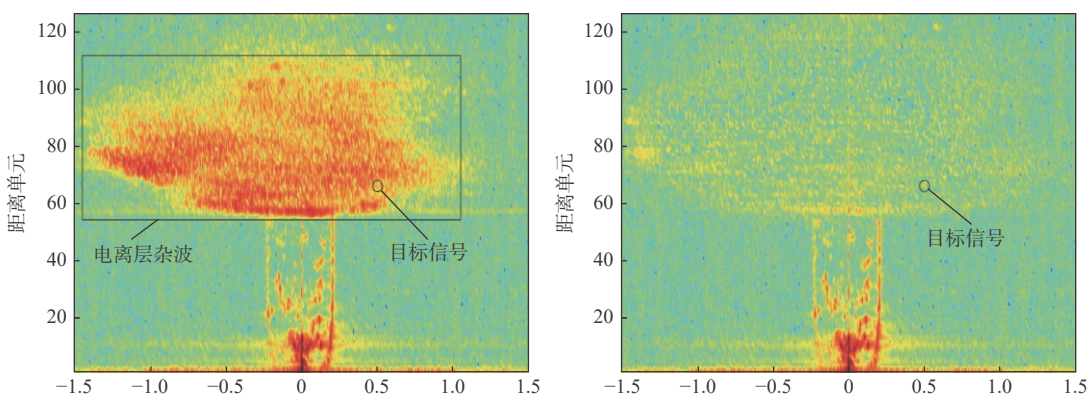
信号; 相比于单一模型较快的训练速度, 混合模型的训练时长有所增加, 但对于电离层杂波的抑制效果得到了极大的提升, 经 Att-CNN-Elman 和 TCN-Elman 模型预测抑制后, 信杂比较处理之前分别提升 15.49 dB 和 15.54 dB, 目标信号已基本显现, 但目标周围的电离层杂波幅值较高, 为高频地波雷达目标检测带来了一定难度; 经 Mixatt-ITCN-Elman 模型处理后, 将电离层杂波抑制至背景噪声附近的同时, 还有效解决了目标信号周围电离层杂波幅值相近的问题, 目标信号得到较为完整的保存, 较处理之前提升 18.51 dB, 相较于混合模型 Att-CNN-Elman 和 TCN-Elman, 信杂比分别进一步提升 3.02 dB 和 2.97 dB。且在目标信号处于未淹没于电离层杂波的第 86 距离单元和较远距离的第 106 距离单元时, Att-CNN-Elman 模型处理效果低于 TCN 和 TCN-Elman 模型, 抑制效果产生了波动, 而 Mixatt-ITCN-Elman 模型的信杂比依旧优于其他模型, 且抑制效果更为稳定。虽训练时长较 Att-CNN-Elman 模型有所提高, 但时长增加幅度不大, 较原始 TCN-Elman 模型训练时长有所降低, 综上所得, Mixatt-ITCN-Elman 模型综合优势较大, 对于电离层杂波具有较好的预测和抑制效果表现。图 10 为同一时刻 Mixatt-ITCN-Elman 模型处理不同距离单元电离层杂波群距离-多普勒对比, 经 Mixatt-ITCN-Elman 模型处理后, 群距离-多普勒图上电离层杂波云团



(a) 第 66 距离单元各模型抑制对比结果　　(b) 第 106 距离单元各模型抑制对比结果

图 9　各模型抑制结果对比

Fig. 9　Comparison of suppression results among different models



(a) 抑制前群距离-多普勒谱　　(b) 抑制后群距离-多普勒谱

图 10　实测数据电离层杂波抑制前后群距离-多普勒谱对比

Fig. 10　Comparison of group distance-Doppler spectra before and after ionospheric clutter suppression of measured data

处理效果显著，目标信号可有效识别。

4 结 论

针对高频地波雷达回波信号中电离层杂波云团干扰目标检测问题，结合瓶颈膨胀卷积模块优化残差模块，利用自注意力机制突出ITCN模型的空间特征，注意力机制捕获混合模型ITCN-Elman的空时特征，提出了一种基于Mixatt-ITCN-Elman模型的电离层杂波云团预测抑制模型。结合相空间理论，将一维时间序列拓展至高维相空间，采用ITCN模型提取高维相空间的空间特征，自注意力机制突显空间特征，Elman模型依据空间特征提取出序列内的时间特征。最后，通过注意力机制对序列的特征进行提取，提高预测精度，构建电离层杂波云团预测抑制模型，实现电离层杂波云团的预测抑制。实验结果表明，本文提出的模型相较于模型Elman、TCN、Att-CNN-Elman和TCN-Elman效果更好，极大程度上提高了电离层杂波预测的精确性和稳定性，在抑制电离层杂波云团的基础上，有效保留了目标信号，降低了目标信号周围的电离层杂波幅值，在电离层杂波抑制中具有良好的研究前景，为进一步提升雷达目标探测性能提供了可靠的理论依据。

参考文献（References）

[1] COOK T M, TERRILL E J, GARCIA-MORENO C, et al. Observations of ionospheric clutter at near equatorial high frequency radar stations[J]. Remote Sensing, 2023, 15(3): 603.

[2] YANG X G, WANG M J, HUANG W M, et al. Experimental observation and analysis of ionosphere echoes in the mid-latitude region of China using high-frequency surface wave radar and ionosonde[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 4599-4606.

[3] CHEN Z Z, XIE F, ZHAO C, et al. An orthogonal projection algorithm to suppress interference in high-frequency surface wave radar[J]. Remote Sensing, 2018, 10(3): 403.

[4] WANG Z Q, LI Y J, SHI J N, et al. Spread sea clutter suppression in HF hybrid sky-surface wave radars based on general parameterized time-frequency analysis[J]. International Journal of Antennas and Propagation, 2020, 2020: 1-12.

[5] JANGAL F, SAILLANT S, HELIER M. Ionospheric clutter mitigation using one-dimensional or two-dimensional wavelet processing [J]. IET Radar, Sonar & Navigation, 2009, 3(2): 112-121.

[6] DING M K, WEI Y S, YU L, et al. Adaptive suppression of mainlobe spread Doppler clutter with high directivity for HFSSWR using oblique projection[J]. Electronics Letters, 2019, 55(23): 1245-1247.

[7] LI J M, YANG Q, ZHANG X. Space-time adaptive processing algorithm based on hyper beamforming for ionospheric clutter suppression in small-array high-frequency surface wave radar[J]. IET Radar, Sonar & Navigation, 2023, 17(4): 545-555.

[8] LI Y J, WANG Z Q, XU L, et al. Spread sea clutter suppression via prior knowledge-based space time adaptive processing in high frequency hybrid sky-surface wave radar[J]. IET Radar, Sonar & Navigation, 2023, 17(5): 830-844.

[9] YANG Y L, MAO X P, HOU Y G, et al. A two-step method for ionospheric clutter mitigation for HFSWR with two-dimensional dual-polarized received array[J]. IEEE Access, 2020, 8: 105903-105913.

[10] GENG H, LIANG Y, YANG F, et al. Joint estimation of target state and ionospheric height bias in over-the-horizon radar target tracking[J]. IET Radar, Sonar & Navigation, 2016, 10(7): 1153-1167.

[11] CHEN S Y, HUANG W M, GILL E W. A vertical reflection ionospheric clutter model for HF radar used in coastal remote sensing[J]. IEEE Antennas and Wireless Propagation Letters, 2015, 14: 1689-1693.

[12] CHEN S Y, GILL E W, HUANG W M. A high-frequency surface wave radar ionospheric clutter model for mixed-path propagation with the second-order sea scattering[J]. IEEE Transactions on Antennas and Propagation, 2016, 64(12): 5373-5381.

[13] 位寅生, 周建宇, 许荣庆. 基于杂波聚类与贪婪策略的电离层杂波智能处理方法[J]. 雷达学报, 2020, 9(4): 589-607.
WEI Y S, ZHOU J Y, XU R Q. Intelligent suppression method for ionospheric clutter based on clustering and greedy strategy[J]. Journal of Radars, 2020, 9(4): 589-607(in Chinese).

[14] GENG H, LIANG Y, CHENG Y H. Target state and Markovian jump ionospheric height bias estimation for OTHR tracking systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(7): 2599-2611.

[15] LYU Z, YU C, LIU A . Chaotic dynamics of ionospheric clutter from high frequency surface wave radar[C]//IET International Radar Conference. London: IET, 2020: 1640-1646.

[16] CHENG W, WANG Y, PENG Z, et al. High-efficiency chaotic time series prediction based on time convolution neural network[J]. Chaos, Solitons & Fractals, 2021, 152: 111304.

[17] HUANG W J, LI Y T, HUANG Y. Prediction of chaotic time series using hybrid neural network and attention mechanism[J]. Acta Physica Sinica, 2021, 70(1): 010501.

[18] WEI L, CAO H, MA Z H. Parameter prediction of marine seawater cooling system based on chaos-Elman combined model[J]. IEEE Access, 2022, 10: 77272-77283.

[19] WOLF A, SWIFT J B, SWINNEY H L, et al. Determining Lyapunov exponents from a time series[J]. Physica D: Nonlinear Phenomena, 1985, 16(3): 285-317.

[20] KIM H S, EYKHOLT R, SALAS J D. Nonlinear dynamics, delay times, and embedding windows[J]. Physica D: Nonlinear Phenomena, 1999, 127(1-2): 48-60.

[21] HAYKIN S. Radar clutter attractor: implications for physics, signal processing and control[J]. IEE Proceedings-Radar, Sonar and Navigation, 1999, 146(4): 177-188.

[22] 陈永, 陈锦, 陶美风. 多尺度特征和注意力融合的生成对抗壁画修复[J]. 北京航空航天大学学报, 2021, 49(2): 254-264.
CHEN Y, CHEN J, TAO M F. Mural inpainting with generative adversarial networks based on multi-scale feature and attention fusion[J]. Journal of Beijing University of Aeronautics and Astro-

nautics, 2023, 49(2): 254-264(in Chinese).

[23] ZHU W Q, WANG Z Y, WANG X C, et al. A Dual self-attention mechanism for vehicle re-Identification[J]. Pattern Recognition, 2023, 137: 109258.

[24] KIM D, KIM Y. Understanding effects of architecture design to invariance and complexity in deep neural networks[J]. IEEE Access, 2021, 9: 9670-9681.

[25] XING W Y, BAI Y L, DING L, et al. Application of a hybrid model based on GA–ELMAN neural networks and VMD double processing in water level prediction[J]. Journal of Hydroinformatics, 2022, 24(4): 818-837.

[26] SUN Y Z, ZHANG J H, YU Z J, et al. WOA (whale optimization algorithm) optimizes Elman neural network model to predict porosity value in well logging curve[J]. Energies, 2022, 15(12): 4456.

Ionospheric clutter suppression method based on improved TCN-Elman neural network

LIU Qiang，SHANG Shang*，QIAO Tiezhu，ZHU Jian，SHI Yishan

(Ocean College，Jiangsu University of Science and Technology，Zhenjiang 212003，China)

Abstract：High-frequency surface wave radar is used in the field of offshore engineering worldwide because of its excellent sea surface target detection capability. One of the key elements to improve its target detection capability lies in the suppression of ionospheric clutter in the echo signal. In response, this paper proposed an ionospheric clutter prediction and suppression model (Mixatt-ITCN-Elman) based on bottleneck expansion convolution module improved temporal convolution (ITCN)-Elman neural network combined with hybrid attention mechanism. First, the ionospheric clutter time series was reconstructed in phase space and subjected to disordered normalization. The spatial features within the high-dimensional phase space were extracted by using ITCN, and the key spatial features were highlighted by combining the self-attention mechanism. Then, the spatial features were combined with the original time series and input into the Elman neural network. The spatial-temporal features of the sequences were highlighted by combining the attention mechanism. Finally, the spatial-temporal features combined with the Elman neural network output sequence were output to obtain the final prediction result. In comparison with Elman, TCN, Att-CNN-Elman, and TCN-Elman models, the proposed model has better prediction performance and stability, having high application value for the suppression of ionospheric clutter.

Keywords：high-frequency surface wave radar；ionospheric clutter；temporal convolutional network；Elman neural network；attention mechanism

Received：2023-06-30；**Accepted**：2023-11-03；**Published Online**：2023-11-13 16：44

URL：link.cnki.net/urlid/11.2625.V.20231110.1522.002

Foundation items：National Natural Science Foundation of China (61801196); Jiangsu Postgraduate Research and Practice Innovation Program (SJCX23_2138)

* **Corresponding author**. E-mail：shangshang@just.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0450

基于子弹击穿理论的直升机机身弹伤快速分析方法

邢肖兵[1]，谭剑锋[1,*]，杨宇霄[1]，雍天[1]，于领军[2]，武瑞[3]

(1. 南京工业大学 机械与动力工程学院，南京 211816；2. 陆军航空兵学院 航空机械工程系，北京 101123；
3. 陆军航空兵学院 第三飞行训练旅，北京 041000)

摘　　要：直升机机身外形与结构复杂，基于有限元方法的直升机机身弹伤数值仿真分析计算量大、周期长，难于适合复杂多弹环境的直升机机身弹伤快速分析评估。为此，耦合子弹击穿理论和弹击碰撞计算模型，建立直升机机身弹伤快速分析方法，并通过与试验结果、有限元数值仿真结果对比，验证快速分析方法的准确性。随后研究多弹打击的直升机机身弹击坐标、剩余速度以及机身损伤面积。研究表明：快速分析方法计算的剩余速度与试验结果相吻合；相比有限元弹伤数值仿真，快速分析方法计算的弹体剩余速度最大误差为4.7%，机身损伤面积最大误差为9.56%，计算时间减少92.1%，计算效率显著提高。随着弹体入射角增大，剩余速度减小，损伤面积增大；当弹体入射角过大，弹体无法穿透机身表面，产生水槽状滑擦损伤，造成较大面积损伤。

关 键 词：直升机机身；多弹体；弹伤快速分析方法；剩余速度；损伤面积
中图分类号：V221$^+$.3; TB553
文献标志码：A　　**文章编号**：1001-5965(2025)09-3212-10

飞机战伤评估 (aircrafts battle damage assessment, ABDA) 是指对飞机战伤程度、修理所需时间和资源、要完成的修理工作以及修理后作战能力等方面做出的评估。战伤评估是战伤修理的前提和必要条件。现代局部战争中，直升机低空作战时机身易受舰炮弹击导致损伤，而弹伤的快速性评估是直升机战伤修复的前提。为此，直升机机身弹伤破坏程度的分析速度直接影响战场快速抢修效率。

近年来，国内外学者已开展了弹击研究。Wadagbalkar 等[1]、Petrudi 等[2]、惠旭龙等[3]开展了基于有限元的弹伤建模与仿真。Wadagbalkar 等[1]通过有限元仿真，训练神经网络和决策树回归模型，并应用于预测弹丸对层压板的穿透力分析。Petrudi 等[2]采用有限元数值仿真方法计算弹体高速冲击，并研究质量、直径、弹头形状和性质对陶瓷材料的冲击损伤影响特性。惠旭龙等[3]采用有限元数值仿真方法分析了平板厚度、撞击角度，以及弹头形状对TC4钛合金平板撞击损伤及弹道极限特性的影响。研究结果表明：平板厚度越大，弹丸撞击角度越小，平板的弹道极限速度越大，而弹丸直径相同时，弹头形状对薄板的弹道极限速度有一定影响。此外，平板厚度、撞击角度及弹头形状对平板的撞击损伤形式有较大影响。然而，上述有限元数值仿真方法主要针对简单平板模型的残余速度及损伤形式，并未针对直升机机体等复杂模型。随后，Singh 等[4]、程玉等[5]、刘铁磊等[6]、高润芳等[7]、赵丽俊等[8]等开展基于理论的弹伤建模与研究。Singh 等[4]采用约翰逊-库克材料和破坏模型以及米-格鲁奈森状态方程，模拟空间碎片对航天器结构的超高速冲击影响。程玉等[5]根据破片初速及剩余速度建立回归方程，得到破片的弹道极限速度，并研究无量纲弹长及无量纲靶厚对弹道极限速度的影响。刘铁

收稿日期：2023-07-10；**录用日期**：2023-09-08；**网络出版时间**：2023-10-25 11:49
网络出版地址：link.cnki.net/urlid/11.2625.V.20231024.1120.001
基金项目：国家自然科学基金 (12172165)；江苏省自然科学基金 (BK20211259)；江苏省高校“青蓝工程”优秀青年骨干教师项目
*通信作者. E-mail：Jianfengtan@njtech.edu.cn

引用格式：*邢肖兵，谭剑锋，杨宇霄，等. 基于子弹击穿理论的直升机机身弹伤快速分析方法 [J]. 北京航空航天大学学报，2025，51（9）：3212-3221.*
XING X B，TAN J F，YANG Y X，et al. An efficient analysis method of helicopter fuselage bullet damage based on bullet penetration theory [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（9）：3212-3221（in Chinese）.

磊等[6]应用信赖域方法获得模型系数, 建立弹道极限速度计算模型。高润芳等[7]基于理论和实验研究战斗部预制破片钨球、钨柱和钨块对装甲钢板的侵彻性能, 并分析着靶姿态对极限穿透速度的影响, 得到不同着靶姿态侵彻阻力面的理论模型和极限穿透速度的公式, 以及 3 种典型预制破片的极限穿透速度。赵丽俊等[8]分析预制破片侵彻均质装甲钢的受力特点, 推导预制破片极限穿透速度公式, 并结合实弹驱动试验和弹道枪模拟复核试验, 得到极限穿透速度计算公式修正系数。上述研究均基于试验数据推导得到弹体穿透材料的弹道极限速度, 但仍然缺少剩余速度及损伤特性。随后, 胡静等[9]开展单层铝合金薄板的弹体冲击试验, 结果表明: 弹体及靶板尺寸形状相同情况下, 随着撞击角度增大, 靶板弹道极限先减小后增大; 撞击角度为 15°时, 铝合金靶板的弹道极限最低, 靶板的失效模式逐渐由花瓣开裂过渡为拉伸撕裂破坏。然而, 上述试验局限于试验结构件, 材料特性和结构特性均与直升机机身存在较大差异。

鉴于直升机机身外形与结构复杂特性, 试验方法与有限元弹伤数值仿真方法存在结构单一、计算量大、周期长问题, 难于对复杂多弹环境下直升机机身结构弹伤快速分析评估。为此, 本文基于子弹击穿理论和弹击碰撞计算模型, 建立直升机机身弹伤快速分析方法, 并与实验数据对比验证, 研究多弹打击下的直升机机身弹击坐标、剩余速度以及机身损伤面积。

1　直升机机身弹伤快速分析方法

1.1　子弹击穿理论

目前弹伤易损性定量计算主要针对常规威胁终端产物 (如弹丸、碎片等) 造成的损伤, 子弹击穿理论研究子弹打击各种材料靶板后的实际弹道及运动状态变化, 涉及弹道极限、子弹剩余速度和损伤面积等计算。

1.1.1　弹道极限

弹道极限 V_{50} 指子弹完全穿过靶板所需的最低速度, 对弹道极限的计算可以预先判读子弹能否穿透被打击结构。设子弹入射速度为 V, 如果 $V>V_{50}$, 子弹会穿透结构, 造成穿孔; 如果 $V<V_{50}$, 子弹会弹开, 对结构表面造成划痕。为此, 采用 JTCG/ME 方程[10]表示弹道极限方程, 即

$$V_{50}=C_{\mathrm{b}f}\left(\frac{\rho_f T A_{\mathrm{pf}}}{W}\right)^{b_f}\sec\theta^{h}\left(\frac{\rho_f T A_{\mathrm{pf}}}{W_0}\right)^{f} \tag{1}$$

式中: ρ_f 为子弹密度; T 为被打击结构表面法向厚度; A_{pf} 为子弹平均暴露面积, $A_{\mathrm{pf}}=C_f wd$, C_f 为子弹形状系数, w 为子弹宽度, d 为子弹长度; W 为子弹初始质量; θ 为子弹的攻击方向与面元法线的夹角, 即入射角; W_0 为参考质量, 大小为 6.48g; $C_{\mathrm{b}f}$、b_f、h、f 为弹道极限常数。

1.1.2　子弹剩余速度

假定穿孔以钝子弹打击的形式发生, 当 $V>V_{50}$, 子弹穿透结构后的剩余速度 V_{r} 为[11]

$$V_{\mathrm{r}}=\sqrt{V^2-\frac{V_{50}^2}{1+\dfrac{\rho_f T A_{\mathrm{pf}}}{W\cos\theta}}} \tag{2}$$

若子弹发生跳弹, 剩余速度 V_{r} 为

$$V_{\mathrm{r}}=V\sin\theta\sin^2\theta_{\mathrm{r}} \tag{3}$$

式中: θ_{r} 为子弹打击跳弹后的出射角。

1.1.3　结构损伤面积

计算损伤面积[12]需要先计算结构暴露面积, 被打击结构暴露面积表达式为

$$A_{\mathrm{p}i}(A_{\mathrm{p}})=\sum_{j=1}^{W}\omega h \tag{4}$$

式中: ω 为结构表面被击中的个数; A_{p_i} 为编号 i 的结构暴露面积; h 为结构垂直于威胁攻击方向平面内的投影面积。此时结构的易损面积表达式为

$$A_{\mathrm{v}i}=\sum_{j=1}^{W}P_{k/h_{ij}}\omega h \tag{5}$$

式中: A_{v_i} 为编号 i 的结构易损面积; $P_{k/h_{ij}}$ 为结构编号 i 在击中轨迹编号 j 打击时的杀伤概率, 由打击形式及被打击位置特性决定。对于击穿损伤, 杀伤概率表达式如下:

$$P_{k/h_{ij}}=\begin{cases}0 & E_b\leqslant 4.5\\ 1+2.65\mathrm{e}^{-0.34E_b}-2.96\mathrm{e}^{-0.14E_b} & E_b>4.5\end{cases} \tag{6}$$

式中: E_b 为被打击结构单位面积法向厚度承受比动能, kg·m/(cm^2·mm)。

$$E_b=\frac{m_f v_f^2}{2Sh}=\frac{m_f^{1/3}v_f^2}{2C_f h}=\frac{m_f^{1/3}v_f^2}{2C_f\delta}\cos\theta \tag{7}$$

式中: m_f 为子弹质量; v_f 为子弹打击速度; S 为子弹和被打击结构接触面积期望值, 且 $S=C_f m_f^{2/3}$; h 为结构击穿厚度, $h=\delta/\cos\theta$, δ 为结构厚度。

1.2　弹击碰撞计算模型

1.2.1　碰撞点计算

弹击碰撞中的三维模型为 STL 格式直升机模型, 即直升机模型由多个三角形面片组成。因此, 弹体与直升机机身的碰撞转化为三维空间内点与三角形面片的交点判断。图 1 为弹击碰撞计算模型流程。

首先, 通过读取函数将三维直升机模型导入并

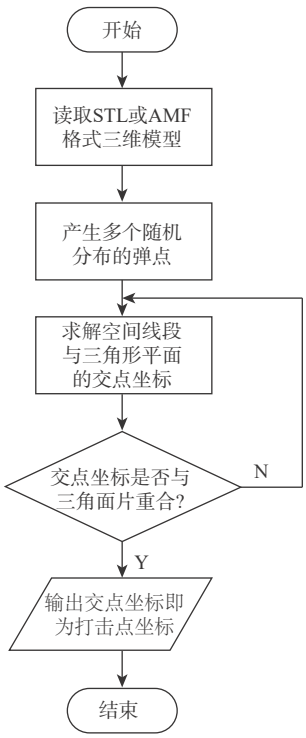


图1 弹击碰撞计算模型流程

Fig. 1 Flow chart of the calculation model of projectile impact

可视化；然后，通过正态分布随机函数生成多个随机弹点；其次，根据空间线段与三角形交点坐标求解方式得到每个弹点与三角形平面的交点；最后，循环判断交点是否与三角形面片重合，若重合则输出交点坐标，若未重合则回到上一步判断下一个交点。

假设三维空间有3个点 $P_1(x_1,y_1,z_1)$、$P_2(x_2,y_2,z_2)$、$P_3(x_3,y_3,z_3)$。3个点构成1个三角形，另外有2个点 $Q(x_q,y_q,z_q)$、$R(x_r,y_r,z_r)$ 构成1条空间线段。

假设线段坐标由参数方程表示：

$$\begin{cases} x = x_q + (x_r - x_q)\,t \\ y = y_q + (y_r - y_q)\,t \\ z = z_q + (z_r - z_q)\,t \end{cases} \qquad 0 \leqslant t \leqslant 1 \tag{8}$$

则三角形所在平面方程由行列式表示：

$$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix} = 0 \tag{9}$$

联立方程(8)和方程(9)可求解出线段参数方程中的 t 值，判断其是否落在区间[0,1]。如果落在其中，则必有交点，且交点可由方程(8)得到，从而得到空间线段与三角形平面的交点坐标。

1.2.2 碰撞点与机身重合判断

利用面积法判断交点是否在三角形内：若交点在三角形面片内，则该交点将三角形面片分成3个子三角形，3个子三角形面积之和等于三角形面片的面积，如果交点在三角形面片外，则子三角形面积之和不等于三角形面片面积，从而得到子弹打击机身坐标点。

本文弹击碰撞计算模型用于计算弹体打击直升机机身的坐标位置；子弹击穿理论用于计算弹体撞击结构后的剩余速度和损伤面积。两者均无法单独实现多弹环境下的直升机机身机构弹伤分析。为实现快速分析多弹体打击直升机机身后的损伤状况，辅助直升机弹伤抢修，本文快速分析方法耦合了弹击碰撞计算模型与子弹击穿理论，获得多弹体打击下直升机机身损伤点与损伤特性。随后耦合有限元数值仿真模型，快速分析多弹环境下直升机弹伤损伤过程、损伤形式。

2 直升机机身弹伤快速分析方法验证

2.1 实验验证

为验证直升机机身弹伤快速分析方法的合理有效性，选取文献[13]的实验数据做对比。实验选取结构钢作为靶板材料[14-16]，靶板截面为150 mm×150 mm，厚度分别为1.5 mm、3 mm和4.5 mm。基于实际弹击情况[17-19]，子弹模型的弹径为5.8 mm，弹速为930 m/s。子弹材料选取钢，材料参数如表1所示。

表1 子弹材料参数[13]

Table 1 Bullet material parameter[13]

材料	密度ρ/(kg·m^{-2})	弹性模量E/MPa	泊松比μ
钢	7 800	200 000	0.3

弹伤快速分析方法计算得到的高速打击后弹体剩余速度如表2和图2所示，同时，图表中给出实验数据，并用于对比验证。

表2 弹体剩余速度实验与弹伤快速分析数值计算对比

Table 2 Comparison of residual velocity experiment on projectile body and efficient analysis method of bullet damage

金属板	序号	板厚/mm	射入角度/(°)	剩余速度/(m·s^{-1})		误差/%
				实验	本文	
钢板	1	1.5	0	878.73	875.65	0.3
钢板	2	3	0	830.37	829.35	0.12
钢板	3	4.5	0	816.99	785.67	3.8

相同入射角下，弹伤快速分析方法计算得到不同靶板厚度的弹体剩余速度与实验测量所得结果吻合较好，4.5 mm板厚时误差最大，为3.8%。由此表明，所建立的直升机机身弹伤快速分析方法能准确计算不同靶板厚度下的弹体剩余速度。

2.2 有限元数值仿真验证

为确保所建立的直升机机身弹伤快速分析方法的可靠性，再将有限元仿真计算结果与快速分析

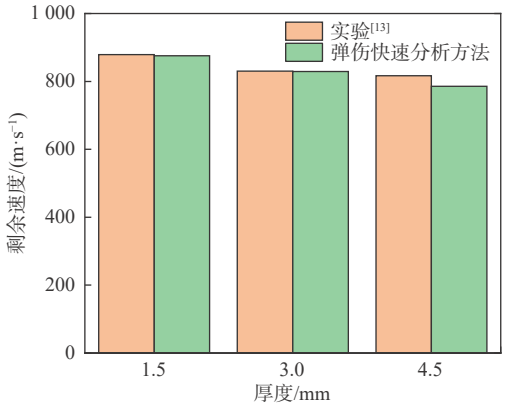


图 2　实验与弹伤快速分析数值计算对比

Fig. 2　Comparison of results of experiment and efficient analysis method of bullet damage

方法计算结果对比验证。

基于 ABAQUS 软件平台建立直径为 5.8 mm 的弹体和截面为 150 mm×150 mm 的矩形靶板。弹速为 930 m/s, 靶板厚度分别为 1.5 mm、3 mm 和 4.5 mm, 入射角度为 0°。三维模型装配如图 3 所示。

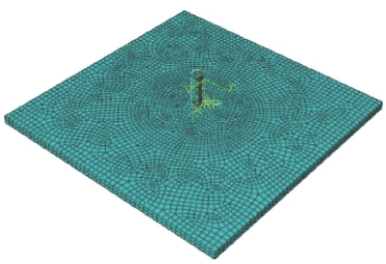


图 3　三维模型装配

Fig. 3　Three-dimensional model assembly

子弹材料选用钢, 材料参数如表 1 所示, 靶板材料选用结构钢。由于结构钢具有渐进性损伤退化性能, 因此选用 Johnson-Cook 失效模型作为损伤准则以及断裂能作为损伤演化。针对撞击接触相互作用问题中切向行为的摩擦因素选用罚函数法, 添加摩擦系数并且添加法向接触行为。靶板材料常用参数及失效参数如表 3 所示。

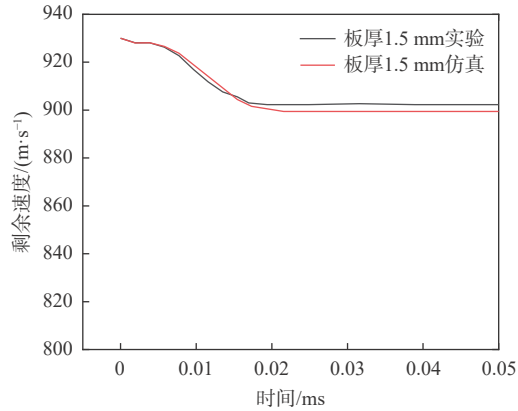
表 3　靶板材料常用参数及失效参数[14]

Table 3　Common parameters and failure parameters of target plate materials[14]

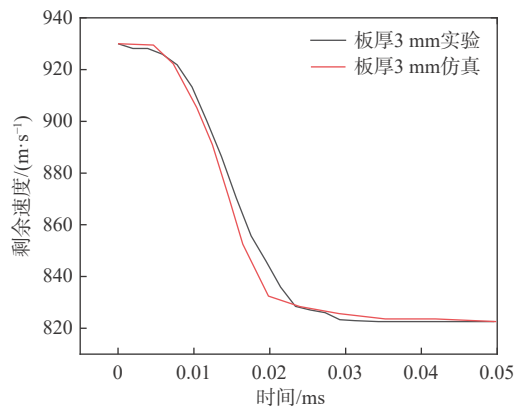
材料	密度ρ/(g·m^{-2})	弹性模量 E/MPa	泊松比μ	屈服强度 σ/MPa	A/Pa	B/Pa	n
钢	7 800	200 000	0.3	235	506	320	0.28
材料	m	C	D_1	D_2	D_3	D_4	D_5
钢	1.06	0.064	0.1	0.76	1.57	0.005	−0.84

注: A为初始屈服应力; B为硬化常数; n为硬化指数; m为热软化指数; C为应变速率常数; D_1-D_5为材料参数。

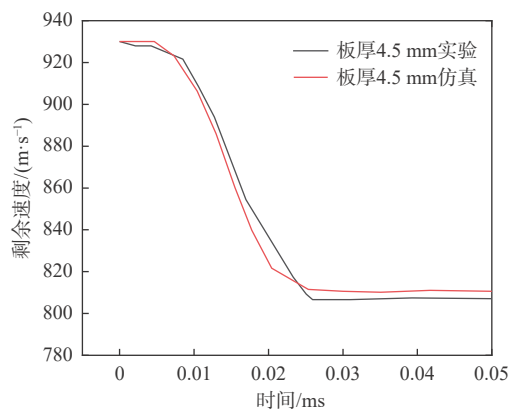
有限元仿真能够得到弹体剩余速度随时间的变化趋势, 将实验和有限元仿真得到的弹体剩余速度数据作对比曲线如图 4 所示。



(a) 1.5 mm板厚剩余速度



(b) 3 mm板厚剩余速度



(c) 4.5 mm板厚剩余速度

图 4　有限元仿真弹体剩余速度与实验对比曲线

Fig. 4　Curves of experiment and finite element simulation of residual velocity of projectile body

基于有限元计算得到的不同靶板厚度弹体剩余速度时间历程曲线与实验结果吻合较好, 由此表明有限元计算的弹体剩余速度随时间变化的精度较高, 可用于快速分析方法的对比验证。随着时间的增加, 子弹剩余速度迅速减小, 减小速率先增大后减小。随着厚度的增加, 剩余速度减小速率也明显增大, 穿透 3 mm 靶板剩余速度下降速率约为 1.5 mm 靶板的 3.6 倍, 穿透 4.5 mm 靶板剩余速度下降速率约为 1.5 mm 的 4.4 倍。

基于有限元和弹伤快速分析方法计算得到弹体剩余速度结果对比如表 4 和图 5 所示。

表 4 剩余速度对比

Table 4 Residual velocity comparison

金属板	板厚/mm	剩余速度/(m·s^{-1})			误差/%		时间/s	
		实验	仿真	本文	仿真	本文	仿真	本文
钢板	1.5	878.73	893.09	875.65	1.63	0.3	45	0.032
	3	830.37	828.27	829.35	0.78	0.12	55	0.025
	4.5	816.99	810.17	785.67	0.83	3.8	62	0.021

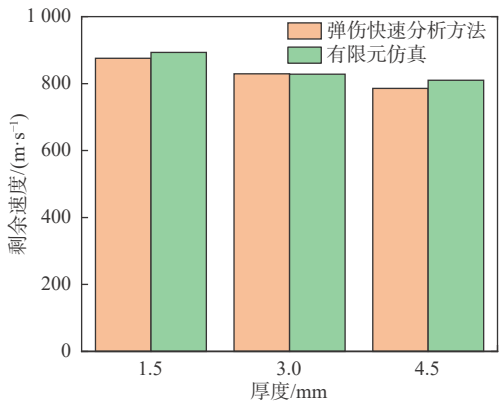


图 5 剩余速度对比

Fig. 5 Residual velocity comparison

在相同入射角下，弹伤快速分析方法与有限元仿真计算的不同靶板厚度弹体剩余速度吻合较好。弹伤快速分析方法计算与有限元仿真计算得到的剩余速度误差均小于 5%。此外，有限元仿真的总时间为 162 s,弹伤快速分方法的计算总时间约为 0.1 s。由此表明，弹伤快速分析方法计算所得到的结果精度与有限元仿真所得到的结果相近，但快速分析方法计算所用的时间远小于有限元仿真。因此，相比于有限元数值仿真，本文快速分析方法计算的弹伤剩余速度精度相近，计算周期显著减小，可用于多弹打击的机身弹伤快速分析。

3 多弹打击的直升机机身弹伤特性研究

3.1 打击点与损伤面积特性

直升机机身弹伤快速分析方法包含 2 部分：①模拟弹体对三维模型的打击数值模型，入射轨迹可固定或者随机，由弹击模型计算碰撞点坐标及入射角；②耦合基于弹击理论的弹伤快速分析方法实现对打击点的损伤情况快速分析。根据三维直升机模型计算弹击碰撞点，然后基于碰撞点处直升机结构参数 (包括厚度、材料等) 计算弹击后的剩余速度、损伤面积等，因此能够体现机体结构变化的影响。

首先，模拟对某型直升机弹伤打击，弹体弹径 10.6 mm、弹长 4.8 mm×10 mm，弹速为 1 000 m/s。根据高射炮弹点正态分布的散布规律，采用 normrnd 正态分布函数随机生成围绕直升机模型的 50 个空间弹点，*y*-*z* 平面的投影如图 6 所示，打击模拟如图 7 所示。

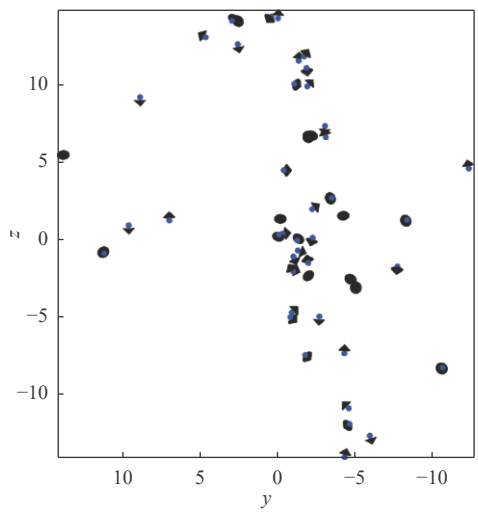


图 6 随机弹点在 *y*-*z* 平面投影

Fig. 6 Projection of random bullet points on the *y*-*z* plane

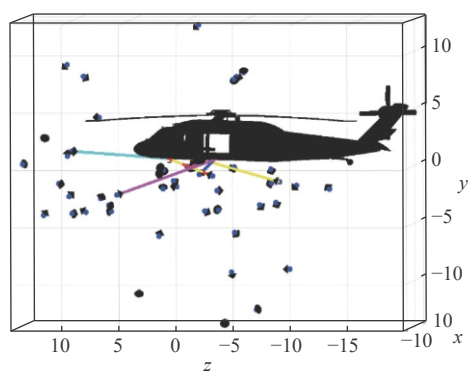


图 7 打击模拟图

Fig. 7 Strike simulation diagram

经弹伤快速分析方法计算得到随机生成的 50 个弹点中，与直升机机身共有 6 个打击点，三维坐标为 (215.03,46.74,−23.26)、(113.67,47.93)、(248.88,54.16,188.23)、(308.52,50.20,−242.55)、(48.52,64.92,−74.00)、(236.57,54.80,183.41)，入射线与碰撞点面法线的夹角分别为 43.51°、28.99°、77.01°、50.53°、8.16°、87.71°。

随后，采用本文弹伤快速分析方法分析 6 个碰

撞点的弹伤影响程度。总的计算时间包含碰撞模拟的时间和耦合弹伤快速分析计算的时间, 结果如表 5 和图 8 所示。

表 5　直升机机身弹伤快速分析方法结果

Table 5　Results of efficient analysis method of helicopter fuselage bullet damage

打击点	材料	入射角/(°)	初速度/(m·s^{-1})	剩余速度/(m·s^{-1})	损伤面积/m^2	计算时间/min
1	铝合金	8.16	1 000	858.8	6.21×10^{-4}	25
2	铝合金	28.99	1 000	840.1	9.41×10^{-4}	25
3	铝合金	43.51	1 000	808.6	1.08×10^{-3}	25
4	铝合金	50.53	1 000	764.6	1.26×10^{-3}	25
5	铝合金	77.01	1 000	607.5	2.95×10^{-3}	25
6	铝合金	87.71	1 000	460.2	4.13×10^{-3}	25

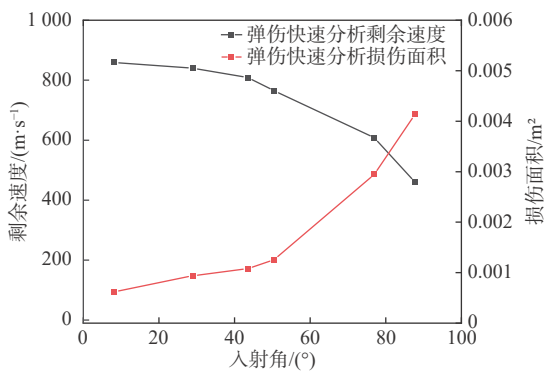


图 8　弹伤快速分析方法不同入射角的剩余速度和损伤面积

Fig. 8　Residual velocity and damage area calculated by efficient analysis method of bullet damage at different incidence angles

随着入射角的增大, 弹伤快速分析方法计算得出的弹体剩余速度随之减小, 损伤面积也随入射角的增大而增大。另外, 当入射角从 50.53°到 77.01°, 剩余速度下降速率增大, 损伤面积上升速率增大, 原因为当入射角较大时, 弹体逐渐无法穿透机身, 产生了跳弹现象, 弹体在机身表面滑擦导致速度快速下降并且滑擦导致较大的损伤面积。因此, 当打击点入射角较大时, 机身表面结构的强度降低较大, 危险程度较高; 当打击点入射角较小时, 打击点损伤面积较小, 危险程度较低。

3.2　损伤特性

基于直升机机身弹伤快速分析方法计算得到的 50 个弹点的打击状况和损伤面积, 结合基于 ABAQUS 软件的有限元数值仿真, 进一步分析 6 个打击点直升机机身弹伤过程和损伤形式, 从而快速分析弹伤特性。

基于 SOLIDWORKS 建立某型直升机三维模型, 如图 9 所示。由于弹伤中弹体受损较小, 变形轻微, 因此假设弹体为刚体, 针对倾角和材料因素分析, 选取弹径 10.6 mm、弹长 4.8 mm×10 mm, 弹速选择 1 000 m/s, 如图 10 所示。

图 9　直升机三维模型

Fig. 9　3D model of helicopter

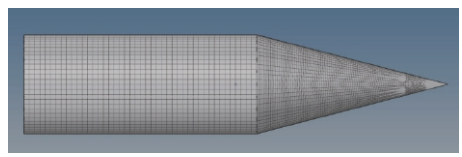


图 10　弹体模型

Fig. 10　Projectile body model

选择机身材料为铝合金 2A50[20], 材料参数如表 6 所示。

表 6　铝合金 2A50 材料参数及 Johnson-Cook 模型失效参数[20]

Table 6　Aluminum alloy 2A50 material parameters and failure parameters of Johnson-Cook model[20]

材料	密度/(kg·m^{-3})	弹性模 E/MPa	泊松比μ	A/Pa	B/Pa			
铝合金	2 750	70 000	0.3	324.1	113.8			
材料	n	m	C	D_1	D_2	D_3	D_4	D_5
铝合金	0.42	1.7	0.008	0.77	1.45	0.47	0.007	0

基于 HYPERMESH 建立直升机机身网格, 由于直升机模型为复杂模型, 选取本文主要研究的机身划分网格, 网格划分类型为六面体网格, 单元网格约为 4 mm, 网格数量为 2.2×10^5, 划分结果如图 11 所示。

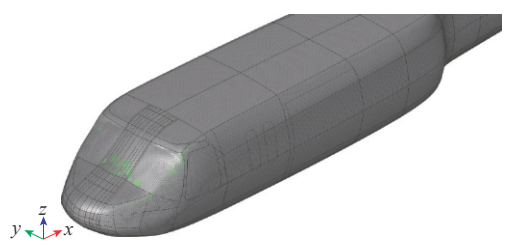


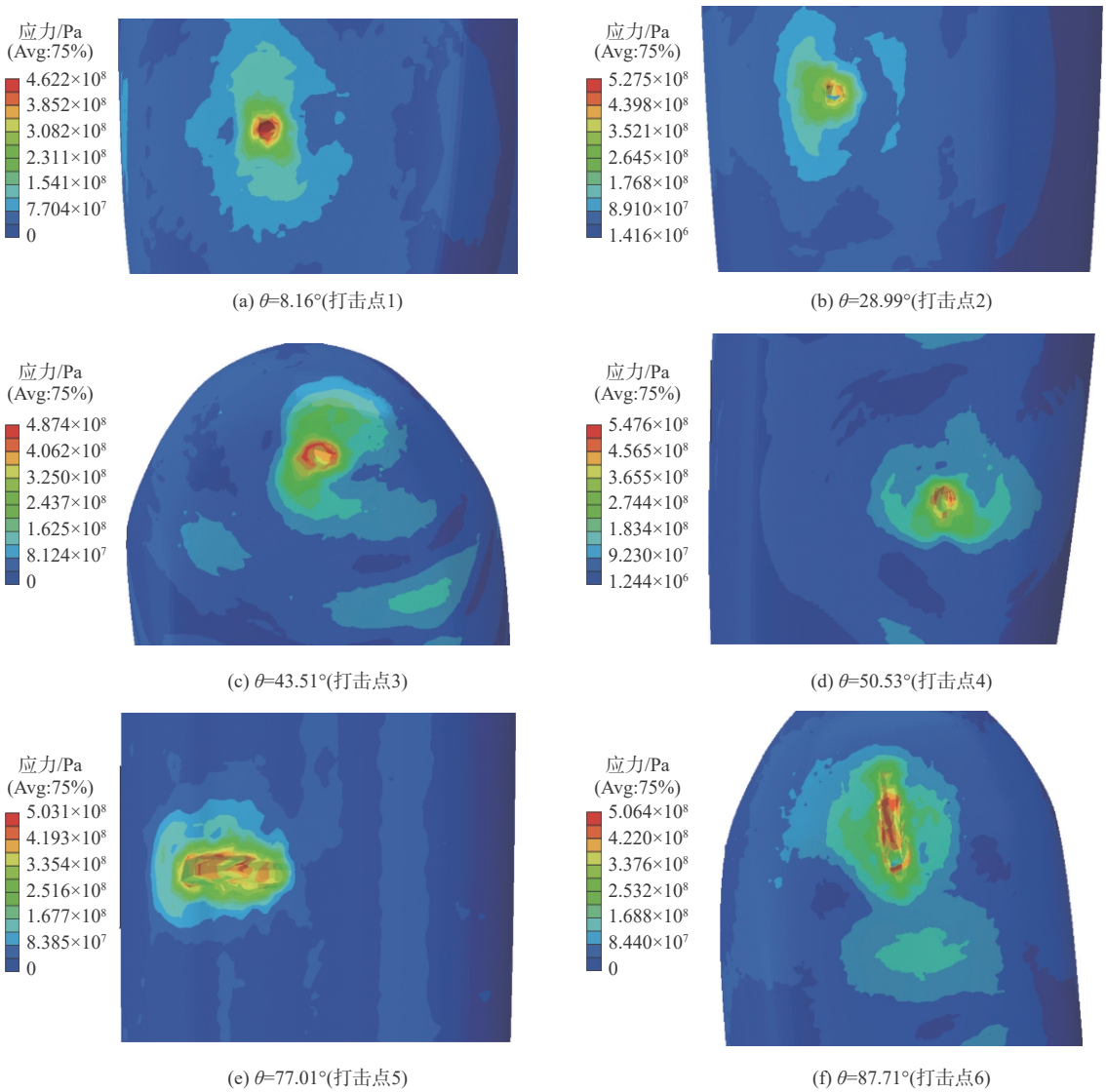
图 11　直升机机身划分网格

Fig. 11　Helicopter fuselage grid

根据本文弹伤快速分析方法计算得到的 6 个碰撞点坐标、入射角、弹伤特性, 在装配模块中分别将弹体与直升机机身按照相应的入射点与入射角装配。选择 ABAQUS/Explicit 计算冲击, 边界条件为机身后部固支, 弹体初速度为 1 000 m/s。

与弹伤快速分析方法相对应的打击点应力云图如图 12 所示。6 个打击点入射角分别为 8.16°、28.99°、43.51°、50.53°、77.01°和 87.71°时, 相对应

的最大应力为462.2、527.5、487.4、547.6、503.1和506.4 MPa。随着入射角的增大，最大应力的变化趋势为先增大后减小，然后再增大再减小，当入射角较大时，最大应力变化幅度减小，最大应力值是入射角为50.53°时的547.6 MPa。另外，随着入射角的增大，损伤面积随之增大，当入射角达到77.01°时，弹体与机身表面的夹角微小导致跳弹，弹体在机身表面滑擦，产生了较大的损伤面积。因此，当打击点入射角较大时，弹体产生的滑擦损伤会使机身表面结构强度降低较多，危险程度较高。



(a) θ=8.16°(打击点1)　(b) θ=28.99°(打击点2)

(c) θ=43.51°(打击点3)　(d) θ=50.53°(打击点4)

(e) θ=77.01°(打击点5)　(f) θ=87.71°(打击点6)

图12　6个打击点的应力云图

Fig. 12　Stress nephogram for six strike points

图13为不同入射角下机身表面最大应力以及弹体剩余速度曲线；图14为不同打击点的剩余速度随时间变化的历程曲线。弹体打击直升机机身后的剩余速度随打击点入射角的增大而减小；随着入射角的增大，弹体剩余速度趋于平稳的时间也随之增大，当入射角从50.53°到77.01°，剩余速度下降速率增大，而入射角为77.01°和87.71°时的剩余速度下降速率基本相同。由此表明当入射角较大时，弹体无法完全穿透机身，而是在机身表面滑擦，弹体与机身表面的接触面积增大，阻力增大，导致剩余速度下降较快。此外，入射角为8.16°、28.99°、43.51°、50.53°、77.01°和87.71°的剩余速度为886.4、823.6、790.8、747.7、580.3和445.1 m/s，当入射角较小时，弹体能够击穿直升机机身后速度趋于稳定，而当入射角过大时，弹体与机身表面为挤压滑擦形式，无法完全穿透机身，因此产生了跳弹现象，此时弹体经过滑擦后也趋于平稳速度。

当入射角度较小时，弹体完全穿透机身表面形成孔洞，孔径较小，损伤面积较小，在直升机机身结构可修复范围。当入射角度增大到无法穿透机身表面后，弹体在表面滑擦，损伤形式转变成槽状破坏并且损伤面积较大，修复性随之降低。

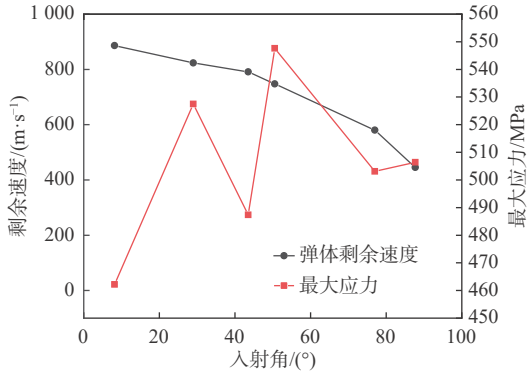


图 13　不同入射角最大应力和剩余速度

Fig. 13　Maximum stress and residual velocity at different incidence angles

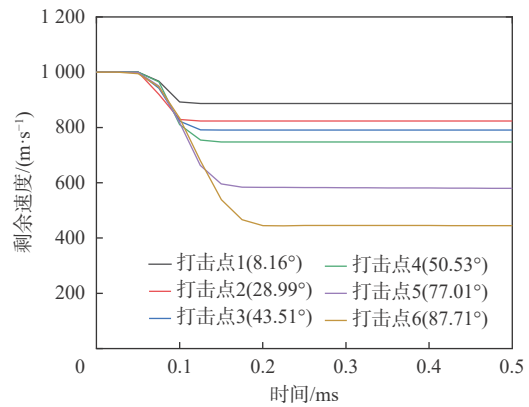


图 14　打击点的剩余速度时间历程

Fig. 14　Residual velocity time history of the strike point

3.3　直升机机身弹伤快速分析方法与有限元方法的比较

直升机机身弹伤快速分析方法的耗时包含仿真计算时间包含碰撞模拟时间和耦合弹伤快速分析计算时间, 有限元数值仿真耗时包含仿真计算时间。将六种不同打击点下直升机机身弹伤快速分析方法计算结果与有限元数值仿真计算结果作对比如表 7 和图 15 所示, 其中损伤面积为机身表面受弹体打击后损伤破坏单元的最大包络面积。有限元分析中, 可根据材料的损伤数值, 确定损伤单元, 并计算损伤区域的包络面积获得损伤面积。弹伤快速分析方法与有限元仿真计算时间随打击点个数增加的变化如图 16 所示。

表 7　直升机机身弹伤快速分析计算与有限元仿真计算数据对比

Table 7　Comparison of data between efficient analysis and finite element simulation of helicopter fuselage bullet damage

打击点入射角/(°)	本文剩余速度/(m·s^{-1})	有限元剩余速度/(m·s^{-1})	本文损伤面积/m^2	有限元损伤面积/m^2
8.16	858.8	886.4	6.21×10^{-4}	5.75×10^{-4}
28.99	840.1	823.6	9.41×10^{-4}	8.96×10^{-4}
43.51	808.6	790.8	1.08×10^{-3}	1.02×10^{-3}
50.53	764.6	747.7	1.26×10^{-3}	1.15×10^{-3}
77.01	607.5	580.3	2.95×10^{-3}	2.82×10^{-3}
87.71	460.2	445.1	4.13×10^{-3}	3.84×10^{-3}

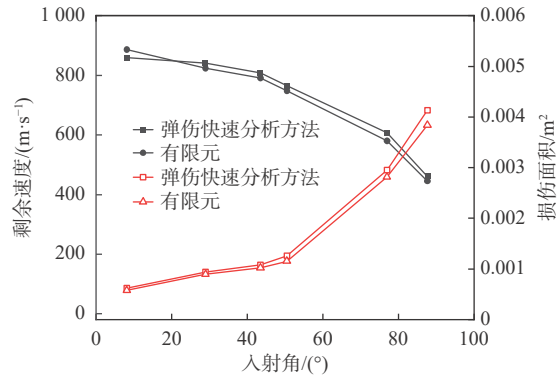


图 15　剩余速度和损伤面积对比曲线

Fig. 15　Curves of residual velocity and damage area

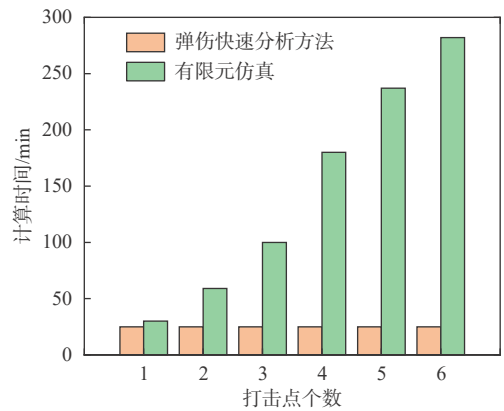


图 16　计算时间对比

Fig. 16　Comparison of computational time

本文弹伤快速分析方法与有限元仿真计算的弹体剩余速度与损伤面积吻合程度良好。6 种打击点下的弹伤快速分析方法剩余速度相对于有限元仿真剩余速度百分比差分别为 3.1%、2.0%、2.20%、2.26%、4.7% 和 3.4%; 弹伤快速分析方法损伤面积相对于有限元仿真损伤面积百分比差分别为 8%、5.02%、5.88%、9.56%、4.88% 和 7.97%。在 6 个不同打击点的情况下, 弹伤快速分析方法的弹体剩余速度相对于有限元数值仿真的弹体剩余速度误差小于 4.7%, 弹伤快速分析方法的损伤面积相对于有限元数值仿真的损伤面积误差小于 9.56%, 表明弹伤快速分析方法能够得到与有限元数值仿真近似结果。

此外, 弹伤快速分析方法能够同时实现多弹打击的损伤分析, 随着打击数量的增加, 计算时间基本相同; 而有限元仿真的计算时间随着打击数量的增大基本呈现线性增加。在 6 个不同打击点的情况下, 弹伤快速分析方法相对于有限元数值仿真所消耗的时间缩减了 92.1%, 因此, 本文建立的直升机

机身弹伤快速分析方法效率更高。

4 结 论

1）基于子弹击穿理论和弹击碰撞计算模型，建立多弹体的直升机机身弹伤快速分析方法，计算的剩余速度与实验结果相吻合，打击3种厚度钢靶板的最大误差为3.8%。

2）相比于有限元数值仿真，本文快速分析方法计算的弹体剩余速度最大误差为4.7%，机身损伤面积最大误差为9.56%，计算时间缩减92.1%，计算效率显著提高。

3）随着入射角的增大，弹体剩余速度随之减小。当入射角较小时，弹体能够完全穿透机身造成花瓣状弹孔，损伤面积变化较小；当入射角过大将会导致跳弹现象，弹体在机身表面产生滑擦，从而形成水槽状损伤面积且损伤面积变化较大。

参考文献（References）

[1] WADAGBALKAR P, LIU G R. Real-time prediction of projectile penetration to laminates by training machine learning models with finite element solver as the trainer[J]. Defence Technology, 2021, 17(1): 147-160.

[2] PETRUDI A M, VAHEDI K, RAHMANI M. Numerical and analytical simulation of ballistic projectile penetration due to high velocity impact on ceramic target[J]. Fracture and Structural Integrity, 2020, 14(54): 226-248.

[3] 惠旭龙, 刘小川, 王计真, 等. TC4钛合金平板高速撞击损伤及弹道极限特性[J]. 科学技术与工程, 2017, 17(11): 1-8.
HUI X L, LIU X C, WANG J Z, et al. High velocity impact failure and ballistic performance of TC4 plate[J]. Science Technology and Engineering, 2017, 17(11): 1-8(in Chinese).

[4] SINGH P K, LUMAR M. Hypervelocity impact behavior of projectile penetration on spacecraft structure: a review[J]. Materials Today: Proceedings, 2022, 62: 3167-3171.

[5] 程玉, 陈智刚, 杨芮, 等. 无量纲参数对破片弹道极限速度的影响[J]. 弹道学报, 2022, 34(4): 8-14,51.
CHENG Y, CHEN Z G, YANG R, et al. Influence of dimensionless parameters on ballistic limit velocity of fragments[J]. Journal of Ballistics, 2022, 34(4): 8-14,51(in Chinese).

[6] 刘铁磊, 徐豫新, 王晓锋, 等. 钨合金球形破片侵彻低碳钢的弹道极限速度计算模型[J]. 兵工学报, 2022, 43(4): 768-779.
LIU T L, XU Y X, WANG X F, et al. Ballistic limit calculation model of tungsten alloy spherical fragments penetrating into low carbon steel plate[J]. Acta Armamentarii, 2022, 43(4): 768-779(in Chinese).

[7] 高润芳, 韩峰, 马晓青, 等. 几种钨合金破片垂直侵彻装甲钢板极限穿透速度研究[J]. 弹箭与制导学报, 2005, 25(4): 57-59,62.
GAO R F, HAN F, MA X Q, et al. Investigation of tungsten fragments of different shape penetrating armour plate[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2005, 25(4): 57-59,62(in Chinese).

[8] 赵丽俊, 焦志刚, 李晓婕, 等. 预制破片侵彻均质装甲钢的极限穿透速度[J]. 爆炸与冲击, 2018, 38(1): 183-190.
ZHAO L J, JIAO Z G, LI X J, et al. Critical penetration velocity of prefabricated fragment in penetrating homogeneous armor steel plate[J]. Explosion and Shock Waves, 2018, 38(1): 183-190(in Chinese).

[9] 胡静, 吕瑜晗. 斜撞击角度对铝合金薄板失效特性影响研究[J]. 机械强度, 2022, 44(5): 1119-1127.
HU J, LÜ Y H. Study on the effect of oblique impact angle on the failure characteristics of aluminum alloy sheet[J]. Journal of Mechanical Strength, 2022, 44(5): 1119-1127(in Chinese).

[10] Ball R E. The fundamentals of aircraft combat survivability analysis and design[M]. Second edition. Monterey: American Institute of Aeronautics and Astronautics, Inc, 2003: 739-778.

[11] PEI Y, SONG B F, HAN Q. A generic calculation model for aircraft single-hit vulnerability assessment based on equivalent target[J]. Chinese Journal of Aeronautics, 2006, 19(3): 183-189.

[12] 宋笔锋, 韩庆. 飞行器易损性指标计算的基本理论与方法（Ⅰ）[J]. 现代防御技术, 1999, 27(4): 39-47.
SONG B F, HAN Q. Basic theory and method of aircraft vulnerability index calculation (Ⅰ)[J]. Modern Defence Technology, 1999, 27(4): 39-47(in Chinese).

[13] 陈刚, 陈小伟, 陈忠富, 等. A3钢钝头弹撞击45钢板破坏模式的数值分析[J]. 爆炸与冲击, 2007, 27(5): 390-397.
CHEN G, CHEN X W, CHEN Z F, et al. Simulations of A3 steel blunt projectiles impacting 45 steel plates[J]. Explosion and Shock Waves, 2007, 27(5): 390-397(in Chinese).

[14] 张铁纯, 田璐, 邓云飞, 等. 2A12铝合金薄板抗叶片形弹体撞击的数值仿真研究[J]. 机械强度, 2019, 41(3): 734-738.
ZHANG T C, TIAN L, DENG Y F, et al. Numerical simulation research on the aluminum alloy thin plate impacted by blade projectile[J]. Journal of Mechanical Strength, 2019, 41(3): 734-738(in Chinese).

[15] 李金福. Q235叠层靶抗破片侵彻性能的研究[D]. 太原: 中北大学, 2021: 10-12.
LI J F. Study on resistance of Q235 laminated target to fragment penetration[D]. Taiyuan: North University of China, 2021: 10-12(in Chinese).

[16] 周卿, 黄松, 赵鹏铎, 等. 平头破片侵彻中厚Q235靶板的破坏模式研究[J]. 兵器装备工程学报, 2020, 41(7): 69-74.
ZHOU Q, HUANG S, ZHAO P D, et al. Study on failure mode of flat-headed projectile penetrating medium-thick Q235 target[J]. Journal of Ordnance Equipment Engineering, 2020, 41(7): 69-74(in Chinese).

[17] 何降润. 穿燃含能破片侵彻靶板后的毁伤效能[D]. 太原: 中北大学, 2021: 12-20.
HE J R. Study on the damage effectiveness of energy-containi-ng fragments penetrating a target plate through combustion[D]. Taiyuan: North University of China, 2021: 12-20(in Chinese).

[18] 张钰龙, 郑宾, 郭华玲, 等. 球形钨破片侵彻钢靶毁伤效应研究[J]. 兵器装备工程学报, 2020, 41(5): 32-36.
ZHANG Y L, ZHENG B, GUO H L, et al. Study on damage effect of spherical tungsten fragments penetrating steel targets[J]. Journal of Ordnance Equipment Engineering, 2020, 41(5): 32-36(in Chinese).

[19] 舒张忆南, 梁争峰, 阮喜军, 等. 多种因素影响下钨合金破片的穿甲深度研究[J]. 兵器装备工程学报, 2022, 43(4): 25-30.
SHU Z Y N, LIANG Z F, RUAN X J, et al. Research on penetration depth of tungsten alloy fragments under influence of many factors[J]. Journal of Ordnance Equipment Engineering, 2022, 43(4): 25-30(in Chinese).

[20] 李亚, 任杰, 马大为, 等. 多角度高速侵彻金属靶的有限元模拟和实验研究[J]. 兵器材料科学与工程, 2017, 40(5): 72-77.
LI Y, REN J, MA D W, et al. FEM simulation and experimental study of multiangular high velocity penetration of metallic targets[J]. Ordnance Material Science and Engineering, 2017, 40(5): 72-77(in Chinese).

An efficient analysis method of helicopter fuselage bullet damage based on bullet penetration theory

XING Xiaobing[1], TAN Jianfeng[1,*], YANG Yuxiao[1], YONG Tian[1], YU Lingjun[2], WU Rui[3]

(1. School of Mechanical and Power Engineering, Nanjing Tech University, Nanjing 211816, China;
2. Department of Aeronautical Mechanical Engineering, Army Aviation Institute, Beijing 101123, China;
3. The Third Training Brigade, Army Aviation Institute, Beijing 041000, China)

Abstract: The shape and structure of helicopter fuselage are complicated. Finite element simulation of helicopter fuselage bullet damage has a high computing burden and is time-consuming, which makes it difficult to be applied in rapid analysis and evaluation of helicopter fuselage bullet damage in complex multi-bullet environments. In view of this, an efficient analysis method of helicopter fuselage bullet damage was proposed by coupling the bullet penetration theory and the model of projectile impact. The results obtained from the method were compared with those of experiment and finite element simulation to validate its accuracy. Then, the helicopter fuselage impact coordinate, residual velocity, and fuselage damage area under multi-bullet attack were studied. The results show that the residual velocity calculated by the efficient analysis method is in agreement with the experimental result. The maximum error of the residual velocity of the projectile body and the maximum error of the fuselage damage area are 4.7% and 9.56%, respectively for the proposed method compared with the finite element simulation of bullet damage. Additionally, the efficiency of the present method is obviously improved as the computational time is reduced by 92.1%. With the increase of the incidence angle of the projectile body, the residual velocity decreases, and the damage area increases. When the incidence angle is too large, the projectile body cannot penetrate the surface of the fuselage, which results in flume-like skid damage and a large area of damage.

Keywords: helicopter fuselage; multi-bullet; efficient analysis method of bullet damage; residual velocity; damage area

Received: 2023-07-10; **Accepted**: 2023-09-08; **Published Online**: 2023-10-25 11:49
URL: link.cnki.net/urlid/11.2625.V.20231024.1120.001
Foundation items: National Natural Science Foundation of China (12172165); Natural Science Foundation of Jiangsu Province(BK20211259); Outstanding Young Backbone Teacher Project of Jiangsu Qinglan Project

* **Corresponding author**. E-mail: Jianfengtan@njtech.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0500

基于混合注意力的全天候机场跑道异物检测

张敬博，任杰*，王美琪
(石家庄铁道大学 机械工程学院，石家庄 050043)

摘　　要：机场跑道的异物检测对于飞机安全起降起着至关重要的作用，然而不同光照、天气跑道环境下现有的检测算法存在漏检、误检等现象，为此提出了一种适用于全天候机场跑道环境的YOLOv5异物检测算法。针对原有网络中池化过程存在一定程度的特征损失问题，设计了跨阶段局部空间金字塔池化模块，可以自适应提取深层特征语义信息，增强网络多尺度表征能力；在特征融合部分引入混合注意力模块，利用通道和空间特征权值再分配，强化异物和无关背景要素的特征差异；针对小目标异物难以识别和定位从而导致漏检现象，设计了多尺度定位损失函数，通过加入相似性度量提高小目标的检测能力；使用优化后的训练策略训练MS-FOD数据集。实验结果表明：改进的算法达到了95.83%的均值平均精度和94.31%的召回率，相比原始的YOLOv5分别提高了3.68%和15.69%，同时检测速度为68帧/s，满足实时性异物检测的需求，有效验证了所提算法对机场跑道异物检测的有效性。

关　键　词：机场跑道异物；YOLOv5；跨阶段局部结构；混合注意力机制；多尺度损失函数
中图分类号：TP391.41; V328
文献标志码：A　　**文章编号**：1001-5965(2025)09-3222-11

机场跑道异物 (foreign object debris, FOD) 的检测一直以来都是航空领域的关键问题，是维护机场安全运营的重要一环。FOD是指出现在机场跑道上的任何物体，主要包括金属零件、机械工具、碎石块、塑料品、其他杂物等[1]。其很大可能导致飞机轮胎被划破或者航空发动机强力吸入致使关键部件损坏而发生重大安全事故，造成无法想象的生命威胁和财产损失。目前，中国大多数机场仍然采用人工巡检的方式来检测跑道上的FOD，受工作人员主观意识、时间成本较高、环境多变等因素制约，无法严格保障机场的安全运营。因此，研究FOD的智能化检测对于航空业的发展具有重要的现实意义。

近年来，FOD检测识别方法[2]大致分为三大类：基于雷达探测技术的自动识别方法、基于光学成像探测技术的自动识别方法、基于多传感器数据融合的自动识别方法。雷达探测技术[3-5]具有高精度定位、不受环境变化所影响等特点，可以全天候对机场跑道进行检测。但其本身造价和维护成本高，容易受复杂外界因素干扰导致虚警，并且缺少目标语义描述，不利于异物的清除。光学成像探测技术又分为传统数字图像处理技术[6-7]和深度学习目标检测[8]技术。传统数字图像处理技术根据任务中的要求，人为设计基于颜色、形状等特征提取器，然后利用机器学习等方法对其分类识别。在光线良好的环境下，可以达到很好的效果，对于光线昏暗或背景噪声干扰大的情况下检测效果有待提高。随着卷积神经网络 (convolutional neural network, CNN) 等深度学习技术的发展，越来越多的研究者将其应用到各自的领域。深度学习可以从给定数据中自动学习到更高层次的抽象特征，具有强大的

收稿日期：2023-07-29；**录用日期**：2023-09-22；**网络出版时间**：2023-10-18 09:51
网络出版地址：link.cnki.net/urlid/11.2625.V.20231017.0921.001
基金项目：国家自然科学基金 (12102273)
* **通信作者**. E-mail：renjiebeijing@163.com

引用格式：*张敬博，任杰，王美琪. 基于混合注意力的全天候机场跑道异物检测 [J]. 北京航空航天大学学报，2025，51（9）：3222-3232.*
ZHANG J B，REN J，WANG M Q. All-weather airport runway foreign object debris detection based on mixed attention [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（9）：3222-3232 （in Chinese）.

鲁棒性。FOD 检测属于目标检测领域的范畴, 其主要分为两大类: 一类是基于候选框的两阶段目标检测算法, 另一类是基于回归的一阶段目标检测算法。多传感器数据融合的探测技术[9] 即结合雷达和光学成像探测技术, 可以充分发挥每个技术模块的优势, 但是现有技术难以实现多源信息的高效融合以及机场的全方位部署。

本文重点研究基于深度学习目标检测的光学成像探测技术。其中基于候选框检测算法的研究中, Cao 等[10] 首次使用深度学习的方法用于机场 FOD 的检测, 通过区域候选网络 (region proposal network, RPN) 提供的候选框传给 CNN 分类器, 达到了不错的检测效果; 于晨[11] 在 Faster R-CNN[12] 的基础上, 通过在基础主干网络添加特征融合模块来提取更有效的特征, 然后利用先验知识对 RPN 网络重新获取锚点, 最后设计了多属性语义识别决策层, 可以实现对材料和危险等级的识别。基于回归检测算法的研究中, 郭晓静等[13] 在 YOLOv3[14] 的基础上, 在网络结构、聚类方法和损失函数上进行改进, 达到了平均精度为 86.5% 的检测效果; 何自芬等[15] 针对夜间 FOD 入侵提出了 CSPTNet 模型, 通过将融合自注意力特征替换残差模块和采用多头自注意力机制, 实现了对夜间跑道的检测。由于机场跑道具有全天候开放的特殊性以及 FOD 多为中小目标, 对 FOD 的检测也提出了更高的要求。现有的 FOD 检测算法归结起来存在以下问题: ①候选框类检测算法在网络结构设计上注重高精度, 导致检测速度较慢; 回归类检测算法采用端到端的结构设计, 检测速度更快, 但识别精度相对较低, 尤其对于小目标; ②目前, 机场跑道均为针对某一种环境下 (如白天或夜间) 的算法研究, 缺少应对多种跑道场景下的 FOD 检测; ③对于 FOD 的检测需要满足高精度和高实时性, 现有算法在平衡二者上有一定程度上的漏检、误检等现象。

基于上述, 针对复杂机场跑道下的 FOD 检测中所出现的漏检误检、小目标精度低以及跑道环境多变等问题, 本文提出了一种应对全天候机场跑道环境的 YOLOv5[16] 异物检测算法。主要的工作如下:

1) 网络结构部分在池化模块引入跨阶段局部结构[17], 加强网络特征提取的能力。同时, 在特征融合结构中加入混合注意力机制模块, 提升网络复杂环境下区分目标和背景的能力。

2) 采用回归性能更好的 CIoU 损失函数, 同时引入 Wasserstein 距离[18] 共同计算回归框的定位损失, 提高小目标识别精度。

3) 对训练策略进行优化, 采用改进的 Mosaic 数据增强方式, 丰富恶劣天气下机场跑道场景的变化性, 提高模型应对多变天气的鲁棒性, 并在训练前重新聚类初始锚框的大小, 加速模型收敛。

4) 最后在自建 MS-FOD 数据集上验证了所提算法模型的有效性, 为智能化 FOD 检测系统的搭建提供有效的技术支撑。

1　本文算法

YOLOv5 作为基于回归的一阶段网络模型结合了之前各类算法的多种优势, 无论是精度还是速度上都可以达到不错的识别效果, 因此在目标检测任务中受到广泛使用。其中提供了网络复杂度由浅到深的 s、m、l、x 这 4 个版本, 考虑到保证精度的同时也要兼顾检测速度, 故本文选择 YOLOv5s 模型作为 FOD 检测的基础网络。针对 YOLOv5 在跑道环境中 FOD 检测出现的漏检、误检等问题, 对网络结构重新构建。改进后本文算法的整体框架如图 1 所示。

整体框架由 4 部分组成: ①输入端: 数据预处理操作, 包括固定图片的输入尺寸、自适应先验框尺寸和图像数据增强处理; ②主干网络: 使用 CSP-Darknet53 提取图像中的浅层位置信息和深层语义信息, 并使用本文的 P-SPPCSPC 池化操作自适应的强化提取特征; ③颈部网络: 结合 FPN 和 PAN 结构, 加强浅层网络的语义信息和深层网络的位置信息。同时引入混合注意力机制, 充分融合不同层级的特征表达, 输出 3 个信息丰富的特征图; ④检测头: 对每个特征图生成一组先验框, 训练时使用分类损失、置信度损失和本文设计的多尺度定位损失, 对其进行分类和回归。推理时利用非极大值抑制得出最后的预测结果。

1.1　跨阶段局部空间金字塔池化模块

YOLOv5 主干网络中的空间金字塔池化模块 (spatial pyramid pooling, SPP), 能够增大感受野, 使得算法适应不同的分辨率图像, 其是通过最大池化来获得不同感受野, 同时也可以提取不同尺度的空间特征信息, 提升特征图中空间整体特征的表达。但模块中的最大池化操作, 会带来一定程度特征信息的损失, 这对于跑道场景下的中小异物检测是十分不利的。为此, 设计了一种融合跨阶段局部结构和部分卷积的空间金字塔池化模块 P-SPPCSPC, 结构如图 2 所示, 该模块能够有效改善 FOD 目标特征损失问题。

图 2 所示的 P-SPPCSPC 是在 SPP 模块的基础上, 引入跨阶段局部网络 (cross stage partial network, CSPNet) 结构对池化操作进行了改进, 将 SPP 模块

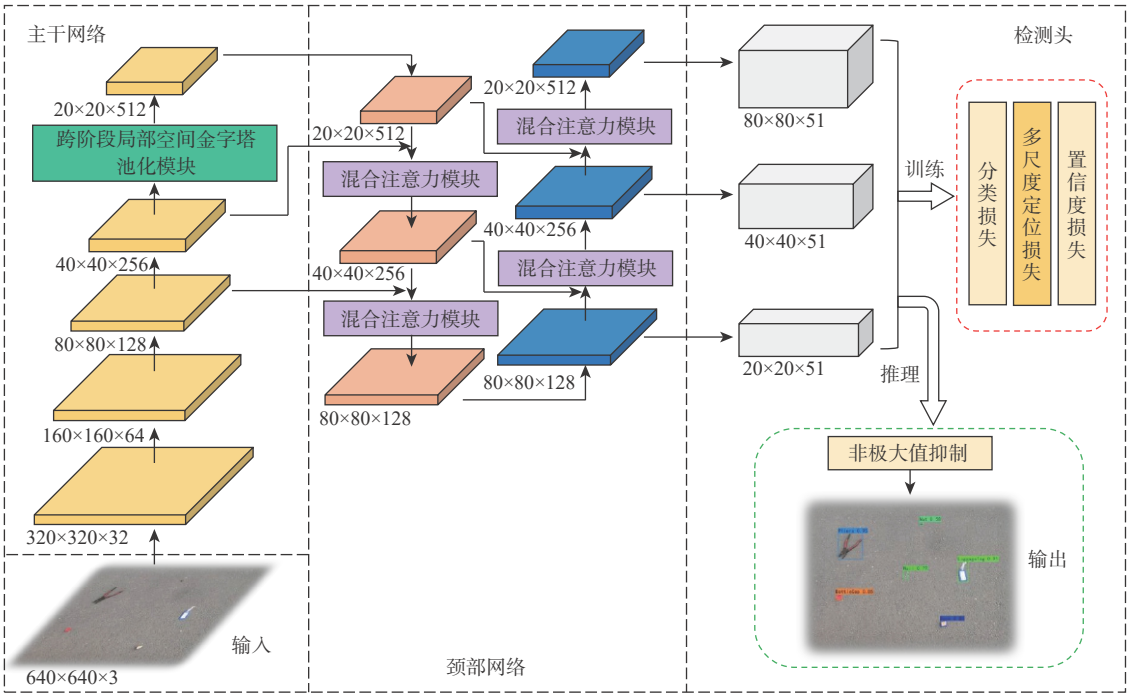


图 1 本文算法的整体框架

Fig. 1 The overall framework of our algorithm

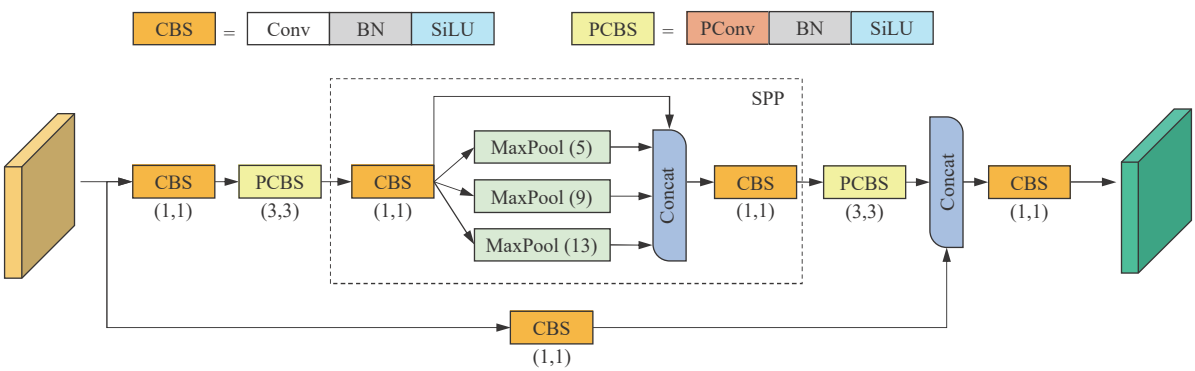


图 2 P-SPPCSPC 模块

Fig. 2 P-SPPCSPC module

嵌入到跨阶段层次结构中，主分支上池化前后增加不同大小卷积进行特征提取，另外从输入端增加一条 short cut 分支，来减少网络传播过程中梯度信息的损失，最后通过拼接操作将 2 条分支提取的特征融合。此模块特征提取的过程如下：

1）由主干网络提取到的特征分别送入 2 个分支，主分支首先对输入使用 1×1 卷积，然后经过 3×3 卷积提取特征，经过 SPP 池化操作增大网络感受野后，再次通过 3×3 卷积提取特征。short cut 分支则只进行 1×1 卷积。

2）最后将 2 条分支进行通道拼接融合后，经过 1×1 卷积调整输出通道，传出融合不同尺度的有效特征层。

其中，1×1 卷积是为了调整通道数，便于后面的操作。3×3 卷积是为了减少池化操作引起的信息损失，增强更多的语义信息，本文使用 3×3 的部分卷积 (partial convolution, PConv) 作为特征提取卷积。部分卷积是 FasterNet[19] 网络模型中提出的新算子，其与标准卷积的卷积过程对比如图 3 所示。PConv 只需在输入通道的一部分上应用标准卷积进行空间特征提取，并保持其余通道不变，可以在保证特征有效提取的同时，降低参数量和冗余的内存访问。需要注意的是，保留剩下的通道而不是简单地删除可能包含冗余特征信息的通道，是因为后续的 1×1 卷积允许特征信息通过所有通道进行交流，加强了特征信息的融合。short cut 分支则是避免提取到过多的重复梯度信息，提高网络的检测精度。

此外，可以看到经过 MaxPool 的有 4 条分支，分别是 1、5、9、13 这 4 个不同的最大池化就代表着能够处理不同的对象。也就是说，这 4 个不同尺

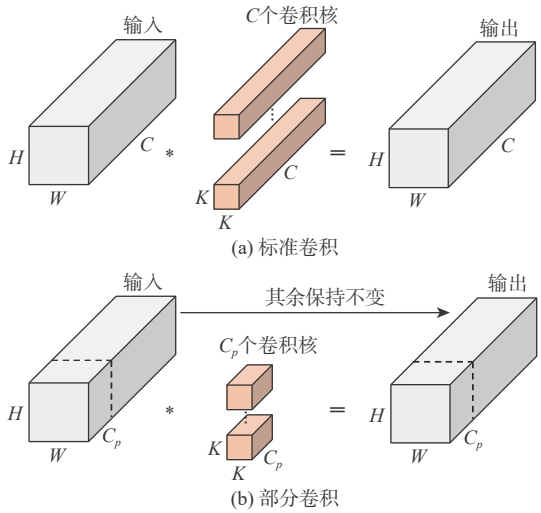


图 3　部分卷积和标准卷积的卷积过程

Fig. 3　The convolutional process of partial convolution and convolution

度的最大池化有 4 种感受野, 用来区别不同尺度的 FOD。通过借鉴 CSPNet 结构的思想构建的 P-SPPCSPC 池化模块, 可以实现局部特征和全局特征的多尺度信息融合来获取更丰富的梯度融合信息, 加强不同分辨率下的表征, 扩大感受野, 提高网络的语义表征能力。

1.2　混合注意力机制模块

针对夜晚或光线较弱环境下目标特征不显著引起的异物漏检, 以及道面背景要素被误检等现象, 在特征融合部分加入卷积注意力模块 (convolutional block attention module, CBAM)[20] 混合注意力机制, 使网络更加关注目标感兴趣区域, 从而提高检测精度。CBAM 注意力机制由通道注意力和空间注意力混合构成, 可以很容易地嵌入到网络中提升网络性能, CBAM 模块结构如图 4 所示。

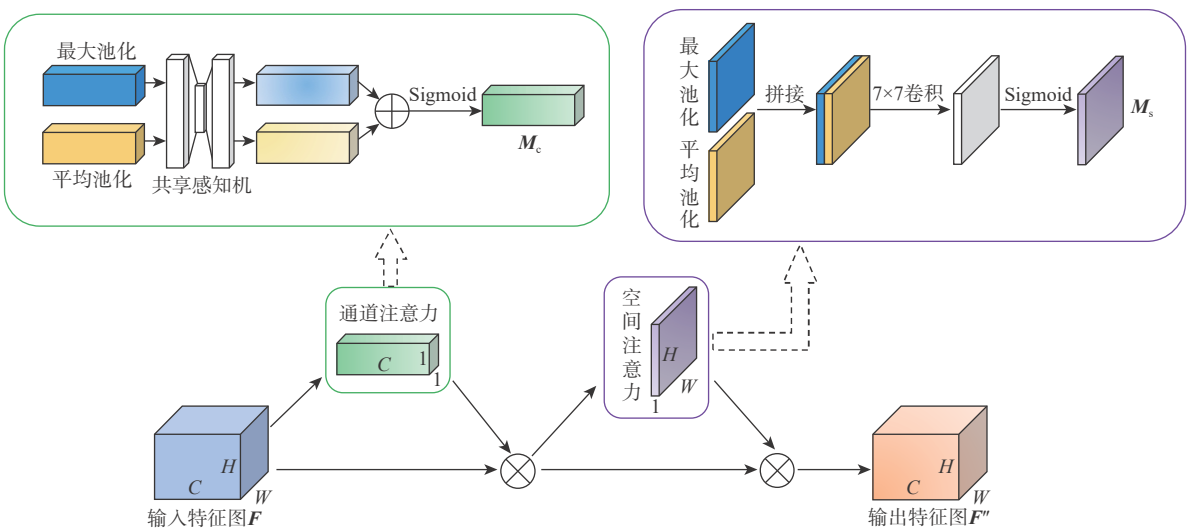


图 4　CBAM 注意力模块

Fig. 4　CBAM attention module

一方面, 通道注意力可以根据不同通道的语义信息增加场景中 FOD 的关注度, 减弱无关背景信息 (裂缝、坑洼以及路面标志线等) 的干扰。图 4 中, 输入特征图记为 $\boldsymbol{F}\in\mathbf{R}^{C\times H\times W}$, H和W分别为特征图的高和宽, C为特征图的通道。首先, 对输入特征图进行全局最大池化和全局平均池化得到$C\times1\times1$的特征图, 再分别经过由全连接层和 Relu 激活函数组成的共享多层感知机得到 2 个一维特征图, 然后将其相加后通过 Sigmoid 归一化函数得出通道注意力权重 $\boldsymbol{M}_c\in\mathbf{R}^{C\times1\times1}$。

另一方面, 空间注意力关注每个像素的重要程度来区分物体的边缘细节信息, 以强化 FOD 的定位识别能力。图 4 中, 以经过通道注意力重标定后的特征图 $\boldsymbol{F}'$ 作为输入, 首先在通道维度上对输入特征图作全局最大池化和全局平均池化, 利用通道拼接操作得到$2\times H\times W$的特征图, 然后采用 7×7 卷积和 Sigmoid 激活函数标准化处理得到最终的空间注意力权重 $\boldsymbol{M}_s\in\mathbf{R}^{1\times H\times W}$。

综上所述, CBAM 混合注意力机制从通道和空间两个维度上对原特征图重新标定。首先对输入特征图进行空间压缩得到通道注意力权值$\boldsymbol{M}_c$, 与输入特征图$\boldsymbol{F}$逐元素相乘后得到包含通道信息的加权特征响应$\boldsymbol{F}'$。然后在此基础上, 通过强化空间维度上的关键特征信息得到空间注意力权值$\boldsymbol{M}_s$, 与特征图$\boldsymbol{F}'$进行逐元素相乘后输出同时包含通道和空间信息的精细化特征图$\boldsymbol{F}''$。经过 CBAM 注意力模块, 使得网络可以更容易地捕获到目标的通道和空间位置信息, 削弱跑道中与 FOD 无关的背景干扰。其过程可表示为

$$\begin{cases}\boldsymbol{F}'=\boldsymbol{M}_c(\boldsymbol{F})\otimes\boldsymbol{F}\\ \boldsymbol{F}''=\boldsymbol{M}_s(\boldsymbol{F}')\otimes\boldsymbol{F}'\end{cases}\tag{1}$$

式中：“⊗”表示逐元素相乘。

1.3 多尺度定位损失函数

YOLOv5网络的损失函数包括分类损失、边框定位损失和置信度损失。在进行机场跑道FOD检测中，使用基于最小包围矩阵的GIoU计算定位损失，经常会遇到预测框与真实框处于包含关系时，GIoU会退化为传统IoU而导致无法衡量相对位置关系，并且在同一最小包围矩阵的情况下，在水平垂直方向上因距离相同导致无法准确表示当前位置的损失值，不利于预测框的回归。此外，FOD检测中多小尺度目标，使用IoU尺度计算回归损失比其他尺度目标更加敏感，容易引起小目标异物出现漏检现象。针对以上问题，设计了一种多尺度定位损失函数NC-Loss，其中“多尺度”指NWD (normalized Gaussian Wasserstein distance)[21]尺度和CIoU[22]尺度。

CIoU损失函数综合考虑了预测框和真实框中心点的距离以及宽高的因素，相较于GIoU损失，回归更加稳定，在表示二者距离上可以有效减小目标定位误差，加速模型收敛。对于预测框$B=(x,y,w,h)$和真实框$B_{gt}=(x_{gt},y_{gt},w_{gt},h_{gt})$，$(x,y)$为中心点坐标，$w$、$h$分别为宽和高。有关CIoU的计算公式为

$$L_{CIoU}=L_{IoU}-\frac{\rho^2(B,B_{gt})}{c^2}-\alpha v \tag{2}$$

$$L_{IoU}=\frac{B\cap B_{gt}}{B\cup B_{gt}} \tag{3}$$

$$\alpha=\frac{v}{(1-L_{IoU})+v} \tag{4}$$

$$v=\frac{4}{\pi^2}\left(\arctan\frac{w_{gt}}{h_{gt}}-\arctan\frac{w}{h}\right)^2 \tag{5}$$

式中：L_{IoU}为预测框与真实框的交并比；$\rho^2(B,B_{gt})$为预测框与真实框之间中心点的欧氏距离；c为包围真实框和预测框最小包围矩阵的对角线长度；α和v分别为权衡参数、衡量预测框与真实框宽高比一致性的参数。

引入NWD尺度可以有效减轻IoU尺度对于小目标回归度量的敏感性，无论边框重叠与否均能够度量其相似性。首先，对预测框B和真实框B_{gt}分别建模为二维高斯分布N和N_{gt}，然后，利用Wasserstein距离尺度计算2个高斯分布的距离$W_2^2(N,N_{gt})$，最后，将其归一化处理后得到$\Phi_{NWD}(N,N_{gt})$来表示预测框和真实框的相似性。有关NWD的计算公式为

$$W_2^2(N,N_{gt})=\left\|\left(\left[x,y,\frac{w}{2},\frac{h}{2}\right]^{\mathrm{T}},\left[x_{gt},y_{gt},\frac{w_{gt}}{2},\frac{h_{gt}}{2}\right]^{\mathrm{T}}\right)\right\|_2^2 \tag{6}$$

$$\Phi_{NWD}(N,N_{gt})=\exp\left(-\frac{\sqrt{w_2^2(N,N_{gt})}}{C}\right) \tag{7}$$

式中：C为根据经验设置的常数。

最后，将NWD损失尺度和CIoU损失尺度同时对目标边界框进行回归损失的计算，提高模型收敛速度和小目标FOD的检测精度。得到本文的NC-Loss定位损失函数为

$$L_{NC}=1-\mu\Phi_{NWD}(N,N_{gt})-(1-\mu)L_{CIoU} \tag{8}$$

式中：μ为NWD尺度在NC-Loss损失函数中所占比例的权重大小，范围为[0,1]，取值由权重参数优化实验中得出。

2 实验结果与分析

2.1 机场跑道MS-FOD数据集

由于目前并没有公开有效的机场跑道异物数据集，基于此，本文依据美国联邦航空局的通告手册[1]对FOD类型的定义以及出现频次的统计，针对全天候机场跑道环境FOD检测任务，模拟了不同背景(沥青和混凝土)、光照(阴天、晴天、黄昏、夜晚)下的场景，拍摄并构建了符合实际机场跑道的多场景异物数据集(multiple scenes foreign object debris, MS-FOD)。

该数据集包含螺母、螺栓、钉子、扳手、钳子、螺丝刀、碎石块、笔、行李牌、电线、瓶盖、树枝12类常见的FOD目标的多场景图像共5 180幅，共27 513个目标实例，各类别实例数量分布如图5所示。其中，金属异物占56.8%，橡胶、塑料异物占25.5%，并且涵盖了不同尺寸和形状特征的异物，基本符合实际机场异物情况。采用LabelImg标注工具统一标注为PASCAL VOC格式，并按照8∶2比例划分为训练集和测试集。其中，MS-FOD的部分图像标注如图6所示。

2.2 实验配置及训练策略优化

2.2.1 实验环境与参数设置

本文实验环境为Windows10操作系统，Intel Core i5-12400F(CPU)，16 GB运行内存(RAM)，NVIDIA RTX 3060 12 GB(GPU)。深度学习框架为pytorch 1.7.1，并采用CUDA11.0、cuDNN8.0加速模型训练。另外，采用表1所示的初始训练参数对MS-FOD数据集中的训练集进行训练。为了提高模型训练效率，减少训练收敛时间，本文所有实验使用基于COCO数据集的预训练权重作为模型初始权重参数进行训练。

2.2.2 数据增强策略优化

机场跑道环境的特殊性，在秋冬时期难免会遇

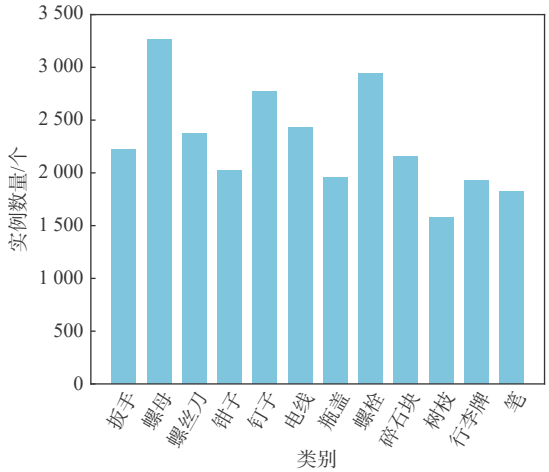


图 5　各类别的实例数量

Fig. 5　Number of instances for each category

图 6　MS-FOD 数据集部分标注示意

Fig. 6　Partial annotation diagram of MS-FOD dataset

表 1　初始训练参数

Table 1　Initial training parameters

参数	数值
输入图片尺寸/像素	640×640
批处理大小	16
优化器	SGD
初始学习率	0.01
动量	0.937
权重衰减	0.000 5
训练轮数	300

到雾天的干扰, 导致成像出现模糊, 细节特征信息缺失等情况。因此本文在原有的 Mosaic 数据增强的基础上, 根据大气散射模型[23] 的成雾原理随机加入不同浓度的雾, 达到丰富目标场景的目的, 从而提高网络模型对雾天的检测, 增强模型的鲁棒性和泛化能力。大气散射模型的描述为

$$\boldsymbol{I}(x)=\boldsymbol{J}(x)t(x)+\boldsymbol{A}(1-t(x)) \tag{9}$$

$$t(x)=\mathrm{e}^{-\beta d(x)} \tag{10}$$

式中: x为像素坐标; $\boldsymbol{I}(x)$为合成的有雾图像; $\boldsymbol{J}(x)$为经过 Mosaic 数据增强后的无雾图像; $\boldsymbol{A}$为大气光向量; $t(x)$为透射率; 衰减系数β控制雾的浓度, 值越大, 雾越浓, 本文将其取值为 [0.02,0.05], 来表示不同浓度的雾; $d(x)$为相机到物体之间的距离, 本文设置为常数 10。

图 7 为优化后的 Mosaic 数据增强示意。将选出的 4 张图片经过随机裁剪、缩放等变换后拼接为一张图片, 然后利用大气散射模型生成有雾图像, 模拟雾天场景下的机场跑道环境, 提高应对恶劣天气下的异物检测能力。

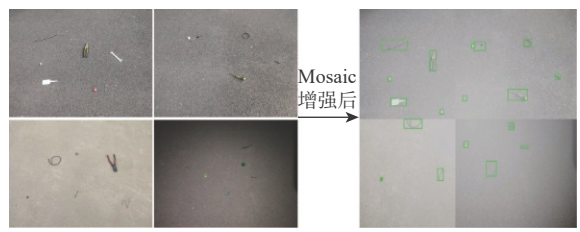


图 7　优化后的 Mosaic 数据增强示意

Fig. 7　Optimized Mosaic data enhancement diagram

2.2.3　重聚类先验框策略

YOLOv5 原网络下的默认先验框是根据 COCO 数据集聚类得来, 包含不同尺度大小的先验框, 与机场 FOD 目标尺度差距较大, 容易导致检测结果定位不精确, 收敛速度较慢。应用 K-means 聚类算法对 FOD 数据集标签真实框进行聚类, 获得贴合异物大小的初始先验框, 有利于网络模型的训练, 提高边界框的回归精度和收敛速度。表 2 为 MS-FOD 数据集真实框聚类前后的先验框对比。可以看到, 聚类后的 P_5 大检测层先验框尺寸明显减小, 整体来看更符合跑道异物的尺度大小。并且按照目标检测中尺度大小的定义: 小于 32 像素×32 像素为小目标, 介于 32 像素×32 像素和 96 像素×96 像素之间为中等目标, 大于 96 像素×96 像素为大目标。因此, MS-FOD 数据集均为中小尺度异物。

2.3　评价指标

本文使用以下评价指标评估模型性能: ①精确率 P, 表示预测正确的目标占所有预测目标的比率; ②召回率 R, 表示预测正确的目标占所有实际目标的比率; ③均值平均精度 mAP, 表示 IoU 为 0.5 时的检测精度; ④检测速度, 单位为帧/s, 表示每秒检测图片数; ⑤参数量为模型中所包含的参数数量。

表2 聚类前后先验框尺寸

Table 2 Prior box size before and after clustering

检测层	默认先验框	聚类后的先验框
P_3	(10,13), (16,30), (33,23)	(11,11), (17,17), (26,14)
P_4	(30,61), (62,45), (59,119)	(23,24), (37,35), (78,28)
P_5	(116,90), (156,198), (373,326)	(31,74), (66,56), (92,95)

2.4 空间金字塔池化模块对比实验

为了验证本文所提P-SPPCSPC模块在检测任务中的有效性,通过保持实验参数的一致性,在MS-FOD数据集上进行了实验。具体地,在YOLOv5s网络的基础上对SPP模块引入跨阶段局部结构,并分别使用标准卷积、分组卷积(GConv, 4groups)、深度可分离卷积(DWConv)代替跨阶段局部结构中的3×3特征提取卷积,从而构成3个新模型。然后与本文的P-SPPCSPC模块进行检测精度mAP和参数量的实验对比。

表3为使用不同空间金字塔池化模块的实验结果对比。整体来看,在池化模块引入跨阶段局部结构虽然给模型带来了一定的参数量,但是相较于原网络模型检测精度均提升了,证明该结构可以有效减少特征信息损失,提高模型的检测能力。在跨阶段局部结构中分别采用不同卷积提取特征,模型1采用的是基于标准卷积的SPPCSPC模块,检测精度mAP提升了1.28%,但是模型参数量也增加了6.3×10^6。模型2采用基于GConv的SPPCSPC模块,相较于模型1参数量减轻了5.3×10^6,但由于特征图缺少通道间的信息交流导致出现一定的信息损失,mAP下降了0.55%。模型3采用基于DWConv的SPPCSPC模块,通过1×1卷积加强通道间的信息融合,相较于模型2略微增加了一些参数量,但mAP提升了0.74%。模型4采用本文所提的基于PConv的P-SPPCSPC模块,相较于模型3参数量相差不大的情况下,mAP提升了0.25%。综上所述,本文提出的P-SPPCSPC池化模块在检测任务中表现最佳,引入基于部分卷积的CSP结构加强了池化模块对于不同尺度目标的特征提取能力,其和原网络相比在增加少量参数的同时mAP提高了1.72%。

表3 不同池化模块效果对比

Table 3 Effect comparison of different pooling modules

模型	空间金字塔池化模块	mAP/%	参数量
基线模型	SPP	92.15	7.2×10^6
模型1	SPPCSPC(Conv)	93.43(+1.28)	13.5×10^6
模型2	SPPCSPC(GConv,4groups)	92.88(+0.73)	8.2×10^6
模型3	SPPCSPC(DWConv)	93.62(+1.47)	9.3×10^6
模型4	P-SPPCSPC	93.87(+1.72)	9.1×10^6

2.5 损失函数权重参数优化实验

本文设计的NC-Loss多尺度定位损失函数由NWD尺度和CIoU尺度联合计算,为了验证所提损失函数的有效性以及确定两者最优权重的比例,采用目前主流的GIoU、DIoU、CIoU损失函数以及权重比例μ为0.3、0.5、0.7、1的NC-Loss损失函数进行对比实验。实验对比结果如表4所示。

表4 不同损失函数效果对比

Table 4 Effect comparison of different loss functions

损失函数	P/%	R/%	mAP/%
GIoU	95.66	78.62	92.15
DIoU	95.27	78.83	92.24(+0.09)
CIoU	95.04	79.29	92.39(+0.24)
NC-Loss(0.3)	94.53	81.69	92.21(+0.07)
NC-Loss(0.5)	93.47	83.99	92.58(+0.43)
NC-Loss(0.7)	93.30	88.83	93.57(+1.42)
NWD	89.41	90.27	91.69(-0.46)

由表4可知,在IoU损失尺度中采用CIoU尺度提升最大,mAP和召回率相较于GIoU分别提升了0.24%和0.67%,验证了考虑边界框中心点距离和长宽比的CIoU作为定位损失函数是有效的。然后,随着权重大小μ从0.3增加到0.7,NWD尺度占比的增加使得网络模型能够获取更多小尺度异物的位置信息,以提高预测框对小目标的定位能力,减少了IoU尺度因损失计算敏感导致的小尺度异物漏检问题,因此mAP和召回率逐渐上升。而当NWD尺度占比增加到1,即完全使用NWD尺度计算损失时,简单的增加权重大小会使模型过分关注其他无关的背景要素,大大降低模型的识别效果,相较于原网络精确率从95.66%下降到89.41%。综合对比实验结果,权重比例为0.7时的检测效果最好,mAP和召回率分别达到了93.57%和88.83%,因此将$\mu=0.7$作为本文所提NC-Loss多尺度损失函数的权重大小。

2.6 消融实验

为了评估本文改进的各个模块对模型性能的贡献程度,在相同的实验环境和训练参数下,设计了消融实验以定量地衡量训练策略优化、P-SPPCSPC池化模块、CBAM注意力机制模块和多尺度NC-Loss损失函数带来的效果。表5为5组消融实验

表 5　消融实验结果

Table 5　Ablation experimental results

模型	训练策略优化	P_SPPCSPC	CBAM	NC-Loss	mAP/%	mAP_S/%	mAP_M/%	mAP_L/%	检测速度/(帧·s^{-1})	参数量
基线模型					92.15	85.78	90.59	95.52	80	7.2×10^6
G1	√				92.81(+0.66)	87.39	92.35	95.89	80	7.2×10^6
G2	√	√			94.34(+1.53)	91.42	94.77	96.18	75	9.2×10^6
G3	√	√	√		94.93(+0.59)	92.56	95.51	96.58	68	9.3×10^6
G4	√	√	√	√	95.83(+0.90)	94.44	95.98	96.46	68	9.3×10^6

的结果。其中, mAPS、mAPM 和 mAPL 分别表示小、中和大目标尺度的检测精度。

由表 5 可知, G1 在采用数据增强优化策略和重聚类初始先验框策略后, 不同尺度异物的检测效果都有一定的提升, 整体检测精度 mAP 较基线模型提升了 0.66%, 因此优化训练策略对 FOD 的检测是有效的。G2 在 G1 的基础上采用了改进的 P-SPPCSPC 模块进行空间自适应特征提取, 整体 mAP 相较于 G1 增加了 1.53%, 其中小、中尺度异物的检测精度 mAPS、mAPM 分别提高了 4.03% 和 2.42%, 提升效果显著。由此可以看出本文设计的金字塔池化模块可以自适应提取深层特征语义信息, 增强网络多尺度表征能力。G3 中 mAP 进一步增加到 94.93%, 证明加入 CBAM 注意力机制能够提高模型对目标的关注度。在此基础上, G4 使用了本文所提 NC-Loss 多尺度损失函数, 能够更精确地度量预测框和真实框之间的位置关系, 并且更适用于小尺度目标的检测。最终改进模型的 mAP 达到了 95.83%, 尤其小尺度异物的检测精度 mAPS 较原模型提高了 8.66%, 验证了本文改进算法的有效性。由于改进增加了一定的参数量, 使得检测速度下降至 68 帧/s, 但基本满足巡检需求。

2.7　不同模型对比实验

通过将本文改进的 YOLOv5 算法模型与当前主流的目标检测算法作对比, 进一步验证本文模型的检测性能, 每个实验均在同一环境下对 MS-FOD 数据集进行训练和测试。Faster R-CNN、SSD[24]、YOLOv3、YOLOv5s、YOLOv7[25] 和本文算法训练过程中的 mAP 对比如图 8 所示, 可以看出, 本文算法改进的模块中权重参数是随机初始化的, 导致模型的收敛速度较慢, 虽然在训练前期的检测精度低于 YOLOv5s 算法, 但是在 150 个 Epoch 之后 mAP 超过其他算法, 最后收敛至 95.83%。

由表 6 的定量分析可知, 对于衡量模型整体算法性能的检测精度 mAP 而言, 本文模型在 YOLOv5s 的基础上设计了跨阶段局部池化结构, 在 Neck 部分引入注意力机制关注重点区域, 并使用多尺度定位损失函数提高模型定位能力, 使得模型在其他网

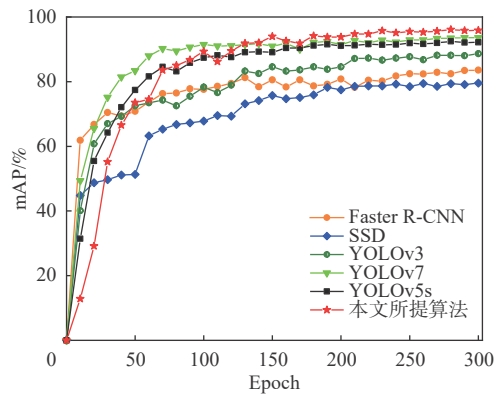


图 8　训练过程中各算法 mAP 对比

Fig. 8　The mAP comparison of each algorithm in the training process

络中检测精度最高, 模型 mAP 达到了 95.83%, 召回率也较原网络提升了 15.69%。检测精度 mAP 相较于双阶段的 Faster R-CNN 提高了 12.29%, 比单阶段的 SSD 和 YOLOv3 分别高出 16.28% 和 7.19%, 相比于最新的 YOLOv7 提高了 2.24%。

在模型参数量方面, 跨阶段局部结构和注意力模块的引入, 网络结构发生了改变, 使得模型相较于原网络参数量略微增加了 2.1×10^6, 但是参数量仅为 Faster R-CNN 模型参数量的 6.8%, 相对于 YOLOv3 缩减了 52.3×10^6, 是 YOLOv7 的四分之一。因此, 本文模型在机场跑道 FOD 检测部署方面更具优势。

在模型检测速度方面, YOLOv5s 得益于 CSPNet 架构的骨干网络以及精简化的模块数量, 在对照组中检测速度最快。由于本文算法的改进增加了一定的参数量, 检测速度相较于 YOLOv5s 基础网络下降了 12 帧/s, 达到 68 帧/s, 但仍快于其他大多数模型, 满足机场跑道巡检的实时性需求。

2.8　检测效果与定性分析

为了更直观地展示改进后检测效果的提升, 将原 YOLOv5s 模型与本文改进模型的多场景检测结果进行可视化。由图 9 的检测结果可知, 在不同跑道场景下, 本文模型可以检测到原模型没有检测到的钉子、螺母等小尺度高危异物, 并且目标置信度均得到了提高, 证实了本文模型在小目标检测和有效特征提取方面的优越性。另外, 在图 9 前 3 行的

表 6 不同算法模型的检测结果对比

Table 6 Comparison of detection results of different algorithm models

模型	主干网络	P/%	R/%	mAP/%	检测速速/(帧·s^{-1})	参数量/M
Faster R-CNN	ResNet50	69.67	91.90	83.54	11	136.9×10^6
SSD	VGG-16	89.62	68.27	79.55	69	25.0×10^6
YOLOv3	Darknet53	91.28	79.12	88.64	36	61.6×10^6
YOLOv5s	CSPDarknet53	95.66	78.62	92.15	80	7.2×10^6
YOLOv7	CBS+ELAN+MP+SPPCSPC	93.48	89.19	93.59	28	36.9×10^6
本文所提算法	CBS+C3+P-SPPCSPC	95.80	94.31	95.83	68	9.3×10^6

混凝土复杂背景下，原模型存在一定的漏检、误检，而本文模型不仅准确识别出所有 FOD，并且没有受道面坑洼、裂缝、轮胎印、标志线等背景因素的干扰，说明模型更加关注目标本身，具有抑制无关背景噪声的能力。对于图 9 第 7 行的雾天场景，本文模型相较于原模型在恶劣天气下具有更好的识别效果，说明使用改进的数据增强策略训练的模型具有更强的泛化能力，鲁棒性更强。因此，本文模型可以应用到多个场景下的 FOD 检测。

(a) 原图 (b) YOLOv5s模型 (c) 本文模型

图 9 改进前后多种场景下的效果对比

Fig. 9 Comparison of effects in various environments before and after improvement

3 结 论

本文从影响飞机安全起降的关键问题“机场 FOD 自动检测技术”出发，针对全天候、多场景复杂环境下异物漏检、错检问题，提出了一种基于混合注意力的机场跑道异物检测算法。

1）在主干网络的池化模块中引入跨阶段局部结构，同时使用轻量化 PConv 提取有效特征，使得模型整体性能提升；此外，在特征融合部分加入 CBAM 注意力机制，抑制了复杂背景中的无关噪声。

2）设计了多尺度定位损失函数 NC-Loss，解决了在对跑道上 FOD 检测时使用 GIoU 损失导致预测框难以收敛以及 IoU 损失尺度对小目标异物敏感的问题，大大提升了小异物的检测能力。并且当权重$\mu=0.7$时，检测效果最好。

3）采用优化的训练策略训练 MS-FOD 数据集，本文所提算法相较于原网络有了较大提升，mAP 和召回率分别提高了 3.68% 和 15.69%，并优于现有的主流检测算法，能够满足多场景机场跑道 FOD 的巡检需求。

虽然本文在数据集上进行了验证，但与实际机场跑道环境还存在一定的差异，后续工作会对真实机场跑道图像进行测试，同时会将算法部署到巡检无人机中进行实际异物探测实验，代替或辅助人工为机场 FOD 检测提供智能化检测方案。

参考文献（References）

[1] FAA. Airport foreign object debris detection equipment: AC 150/5220-24[R]. Washington. D.C.: Federal Aviation Administration, 2009: 1-13.

[2] 陈唯实, 李敬. 基于视频数据的机场跑道外来物检测[J]. 北京航空航天大学学报, 2014, 40(12): 1678-1684.
CHEN W S, LI J. Foreign object debris detection for airport runway with video data[J]. Journal of Beijing University of Aeronautics and Astronautics, 2014, 40(12): 1678-1684(in Chinese).

[3] WANG Y, HUANG H, WANG J, et al. An image denoising method for arc-scanning SAR for airport runway foreign object debris detection[J]. Electronics, 2023, 12(4): 984.

[4] FUTATSUMORI S, SHIBAGAKI N. 96 GHz millimeter-wave

radar system for airport surface detection purpose[C]//Proceedings of the 2022 IEEE Conference on Antenna Measurements and Applications. Piscataway: IEEE Press, 2022: 1-2.

[5] REN S J, HAN S Y, WANG B S. Stationary and small target detection for millimeter-wave radar[C]//Proceedings of the 2022 IEEE 22nd International Conference on Communication Technology. Piscataway: IEEE Press, 2022: 1698-1702.

[6] HAO X J, ZHAN Y H. Research on foreign object debris detection algorithm[C]//Proceedings of the 2022 International Conference on Mechanical and Electronics Engineering. Piscataway: IEEE Press, 2022: 312-318.

[7] 王国屹, 孙永荣, 张怡, 等. 背景对齐差分的机场跑道异物分块检测与跟踪算法[J]. 计算机辅助设计与图形学学报, 2021, 33(3): 413-423.
WANG G Y, SUN Y R, ZHANG Y, et al. Block detection and tracking algorithm of foreign objects debris in airport runway based on background alignment and difference[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(3): 413-423(in Chinese).

[8] ADI K, WIDODO C E, WIDODO A P, et al. Detection of foreign object debris (Fod) using convolutional neural network (CNN)[J]. Journal of Theoretical and Applied Information Technology, 2022, 100(1): 184-191.

[9] 李沙, 李春娟. 机场跑道异物检测系统设计与算法研究[J]. 现代雷达, 2021, 43(6): 80-85.
LI S, LI C J. A study on design and algorithm of FOD detection and surveillance system for airport runway[J]. Modern Radar, 2021, 43(6): 80-85(in Chinese).

[10] CAO X G, GONG G P, LIU M M, et al. Foreign object debris detection on airfield pavement using region based convolution neural network[C]//Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications. Piscataway: IEEE Press, 2016: 1-6.

[11] 于晨. 基于深度学习的机场跑道异物检测与识别技术研究[D]. 北京: 北京工业大学, 2019: 39-52.
YU C. Research of deep learning based airport runway foreign object debris detection and recognition[D]. Beijing: Beijing University of Technology, 2019: 39-52(in Chinese).

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

[13] 郭晓静, 隋昊达. 改进 YOLOv3 在机场跑道异物目标检测中的应用[J]. 计算机工程与应用, 2021, 57(8): 249-255.
GUO X J, SUI H D. Application of improved YOLOv3 in foreign object debris target detection on airfield pavement[J]. Computer Engineering and Applications, 2021, 57(8): 249-255(in Chinese).

[14] FARHADI A, REDMON J. Yolov3: an incremental improvement [C]//Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018, 1804: 1-6.

[15] 何自芬, 陈光晨, 王森, 等. 融合自注意力特征嵌入的夜间机场跑道异物入侵检测[J]. 光学精密工程, 2022, 30(13): 1591-1605.
HE Z F, CHEN G C, WANG S, et al. Detection of foreign object debris on night airport runway fusion with self-attentional feature embedding[J]. Optics and Precision Engineering, 2022, 30(13): 1591-1605(in Chinese).

[16] WU W T, LIU H, LI L L, et al. Application of local fully convolutional neural network combined with YOLO v5 algorithm in small target detection of remote sensing image[J]. PLoS One, 2021, 16(10): e0259283.

[17] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2020: 1571-1580.

[18] VALLENDER S S. Calculation of the Wasserstein distance between probability distributions on the line[J]. Theory of Probability & Its Applications, 1974, 18(4): 784-786.

[19] CHEN J R, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 12021-12031.

[20] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the Computer Vision – ECCV 2018. Berlin: Springer International Publishing, 2018: 3-19.

[21] WANG J W, XU C, YANG W, et al. A normalized Gaussian Wasserstein distance for tiny object detection[EB/OL]. (2021-10-26)[2023-07-05]. https://doi.org/10.485501/arXiv.2110.13389.

[22] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[23] 杨勇, 邱根莹, 黄淑英, 等. 基于改进大气散射模型的单幅图像去雾方法[J]. 北京航空航天大学学报, 2022, 48(8): 1364-1375.
YANG Y, QIU G Y, HUANG S Y, et al. Single image dehazing method based on improved atmospheric scattering model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(8): 1364-1375(in Chinese).

[24] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Proceedings of the Computer Vision – ECCV 2016. Berlin: Springer, 2016: 21-37.

[25] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the 2023IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 7464-7475.

All-weather airport runway foreign object debris detection based on mixed attention

ZHANG Jingbo，REN Jie*，WANG Meiqi

(School of Mechanical Engineering，Shijiazhuang Tiedao University，Shijiazhuang 050043，China)

Abstract：The foreign object debris (FOD) detection of airport runway plays an important role in the safe take-off and landing of aircraft. However, the existing detection algorithms in different light and weather runway environments have the phenomenon of missed detection and false detection. Therefore, a YOLOv5 FOD detection algorithm suitable for all-weather airport runway environments is proposed. Firstly, aiming at the problem of feature loss in the pooling process of the original network, a cross stage partial spatial pyramid pooling module is designed, which can adaptively extract deep feature semantic information and enhance the multiscale representation ability of the network. Secondly, the mixed attention module is introduced in the feature fusion part, and the channel and spatial feature weights are redistributed to strengthen the feature differences between FOD and unrelated background elements. Then, a multiscale positioning loss function is designed to improve the detection ability of small targets by adding similarity measures, aiming at the phenomenon that small target FOD are difficult to identify and locate, which leads to missed detection. Finally, the optimized training strategy is used to train the MS-FOD dataset. The experimental results show that the improved algorithm achieves an average accuracy of 95.83%and a recall rate of 94.31%, which is 3.68 and 15.69 percentage points higher than the original YOLOv5, respectively. At the same time, the detection speed FPS is 68 frames per second, which meets the needs of real-time FOD detection. The effectiveness of the proposed algorithm for airport runway FOD detection is effectively verified.

Keywords：foreign object debris on airport runways；YOLOv5；cross stage partial structure；mixed attention mechanism；multiscale loss function

Received：2023-07-29；**Accepted**：2023-09-22；**Published Online**：2023-10-18 09：51

URL：link.cnki.net/urlid/11.2625.V.20231017.0921.001

Foundation item：National Natural Science Foundation of China (12102273)

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

《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

IISTJA 工信科技期刊集群 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9期　2025年9月

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)
(Monthly, Started in 1956)
Vol.51 No.9　September 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

国内定价：50.00元／期

微信公众号　超星电子书