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



电子机械制动系统关键技术研究进展

赵立金　杨世春　曲婧瑶

4 | 2025
Vol.51

北京航空航天大学学报

第 51 卷　第 4 期　（总第 386 期）　2025 年 4 月

目　　次

电子机械制动系统关键技术研究进展 …… 赵立金，杨世春，曲婧瑶（1037）
基于残差全局上下文注意和跨层特征融合的去雾网络 …… 杨燕，陈飞（1048）
自适应不完备多视角聚类 …… 陈梅，马学艳，张弛，张锦宏，钱罗雄（1059）
基于 RISC-V 向量扩展的图像预处理加速方法 …… 刘强，尹蔚，李凯（1074）
基于 DSGIoU 损失与双分支坐标注意力的目标检测算法 ……
…… 马素刚，李宁博，侯志强，余旺盛，杨小宝（1085）
颜色一致性的无人机航拍低光照图像增强算法 …… 王殿伟，刘旺，房杰，许志杰（1096）
基于新型趋近律的 AABS 积分滑模控制系统设计 …… 徐萌，李艳，高洁，徐海，高冰（1107）
基于融合卷积 Transformer 的航空发动机故障诊断 …… 赵洪利，杨佳强（1117）
基于混合整数规划的终端区柔性进场航迹优化 …… 赵向领，周志亮（1127）
基于用户调度的多无人机辅助通信系统资源优化 …… 唐菁敏，黄嘉琪，王炳文，宋耀莲，虞贵财（1143）
基于 SSA-IWT-EMD 的滚动轴承故障诊断方法 ……
…… 雷春丽，焦孟萱，樊高峰，刘世超，薛林林，李建华（1152）
直驱泵阀协同式线控制动单元压力串级控制 …… 谭草，于鹏，李波，陆佳瑜，任云云（1163）
基于 MDAM-GhostCNN 的滚动轴承故障诊断方法 …… 郭俊锋，谭宝宏，王智明（1172）
基于多层域适应的无标签数据故障诊断方法 …… 王进花，刘瑞，曹洁（1185）
基于结构参数分析的 Gyroid 换热器性能 …… 杨晓军，张雪丽，李沛燃（1195）
基于相邻特征融合与特征解耦的一阶段目标检测 …… 郑剑，贺朝辉，于祥春（1205）
基于 Trans-Attention 的飞行区航空器监视数据融合方法 …… 王兴隆，尹昊，丁俊峰（1215）
面向卫星星座的多地面站系统链路分配策略 ……
…… 姜逸飞，何婉霞，刘文正，吴树范，魏骁，莫乾坤（1224）
TBO 模式下基于复杂网络的空中交通复杂性分析 …… 彭娅婷，温祥西，吴明功，杨志伟，王楠（1234）
小腔对排气活门快速调压能力的影响 …… 吴豪，刘猛，王浚（1245）
翼伞翼型非定常气动特性数值模拟研究 …… 王震，钟伟，王同光，李旭东，张红英（1255）
基于不确定理论的网络时间可靠性评估方法 …… 马骥，李瑞莹，张清源，康锐（1267）
不同气体模型与滑移边界模型下升力体摩阻热流预测 ……
…… 严攀，李沁，黄潇，郭晓明，翁谊辉，尤延铖（1277）
3 种六自由度动力下降凸优化制导方法 …… 王驰，刘伟，高扬（1292）
运载火箭发射过程中的最大气动载荷预测方法 …… 程胡华，武帅，姜祝辉，张入财（1304）
基于卷积神经网络的星载粒子鉴别平台及应用 ……
…… 白超平，张珅毅，张鑫，孙越强，张帅，王子婷（1313）
基于时频重排算法的 LPI 雷达信号识别 …… 姬丽彬，朱岩，崔天舒，王栋，黄永辉（1324）
基于奇异值方法的推力矢量/气动舵飞机复合控制律参数优化方法 ……
…… 阮仕龙，董哲，孙尧，曲晓雷，霍少泽（1332）

基于鲁棒最优滑模的非充气轮胎驱动车辆稳定性控制 …… 李田，赵又群，徐涛，沈亚伟，林棻（1342）
月面应急直接上升交会轨迹快速规划方法 …… 班焕恒，周聪，闫晓东（1352）
多智能体仿射编队机动控制的有向交互拓扑优化设计 ……
…… 马苏慧，张栋，王孟阳，王庭晖，刘送丹（1367）
基于异常点重构的调频引信抗扫频干扰方法 …… 段乐帆，郝新红，陈齐乐（1377）
基于深度强化学习的离散状态转移算法求解柔性作业车间调度问题 ……
…… 朱家政，王聪，李新凯，董颖超，张宏立（1385）
岸基 GNSS-R 码测高反演模型 …… 宁保蛟，王娜子，荆丽丽，高凡，孔亚慧，贺匀峤（1395）
融合 LBP 与并行注意力机制的微表情识别方法 …… 李帅超，李明泽，孙嘉傲，卢树华（1404）

期刊基本参数：CN 11-2625/V * 1956 * m * A4 * 378 * zh * P * ¥50.00 * 400 * 35 * 2025-04

（**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. 4 （Sum 386） April 2025

CONTENTS

Research progress of EMB systems key technology
…… ZHAO Lijin，YANG Shichun，QU Jingyao （1037）
Dehazing network based on residual global contextual attention and cross-layer feature fusion
…… YANG Yan，CHEN Fei （1048）
Adaptive incomplete multi-view clustering
…… CHEN Mei，MA Xueyan，ZHANG Chi，ZHANG Jinhong，QIAN Luoxiong （1059）
Image preprocessing acceleration method based on RISC-V vector extension
…… LIU Qiang，YIN Wei，LI Kai （1074）
Object detection algorithm based on DSGIoU loss and dual branch coordinate attention
…… MA Sugang，LI Ningbo，HOU Zhiqiang，YU Wangsheng，YANG Xiaobao （1085）
Low illumination image enhancement algorithm for UAV aerial photography with color consistency
…… WANG Dianwei，LIU Wang，FANG Jie，XU Zhijie （1096）
Design of aircraft anti-skid braking system integral sliding mode control system based on novel reaching law
…… XU Meng，LI Yan，GAO Jie，XU Hai，GAO Bing （1107）
Aero-engine fault diagnosis based on fusion convolutional Transformer
…… ZHAO Hongli，YANG Jiaqiang （1117）
Flexible arrival trajectory optimization at terminal airspace based on mixed integer programming
…… ZHAO Xiangling，ZHOU Zhiliang （1127）
Resource optimization of multi UAV assisted communication system based on user scheduling
…… TANG Jingmin，HUANG Jiaqi，WANG Bingwen，SONG Yaolian，YU Guicai （1143）
Fault diagnosis method of rolling bearings based on SSA-IWT-EMD
…… LEI Chunli，JIAO Mengxuan，FAN Gaofeng，LIU Shichao，XUE Linlin，LI Jianhua （1152）
Pressure cascade control of brake-by-wire unit based on direct drive pump-valve cooperative
…… TAN Cao，YU Peng，LI Bo，LU Jiayu，REN Yunyun （1163）
Fault diagnosis method of rolling bearing based on MDAM-GhostCNN
…… GUO Junfeng，TAN Baohong，WANG Zhiming （1172）
Unlabeled data fault diagnosis method based on multi-domain adaptation
…… WANG Jinhua，LIU Rui，CAO Jie （1185）
Performance of Gyroid heat exchanger based on structural parameters analysis
…… YANG Xiaojun，ZHANG Xueli，LI Peiran （1195）
One-stage object detection based on adjacent feature fusion and feature decoupling
…… ZHENG Jian，HE Zhaohui，YU Xiangchun （1205）
Aircraft surveillance data fusion method in flight area based on Trans-Attention
…… WANG Xinglong，YIN Hao，DING Junfeng （1215）
Multi-ground station system-based link allocation strategy for satellite constellation
…… JIANG Yifei，HE Wanxia，LIU Wenzheng，WU Shufan，WEI Xiao，MO Qiankun （1224）
Complex network-based air traffic complexity analysis in TBO
…… PENG Yating，WEN Xiangxi，WU Minggong，YANG Zhiwei，WANG Nan （1234）
Effect of balance chamber on rapid pressure regulation ability of outflow valve
…… WU Hao，LIU Meng，WANG Jun （1245）
Numerical simulation of unsteady aerodynamic characteristics of parafoil airfoil
…… WANG Zhen，ZHONG Wei，WANG Tongguang，LI Xudong，ZHANG Hongying （1255）
Network time reliability evaluation method based on uncertainty theory
…… MA Ji，LI Ruiying，ZHANG Qingyuan，KANG Rui （1267）
Friction and heat flux prediction of lift body under different gas models and slip boundary models
…… YAN Pan，LI Qin，HUANG Xiao，GUO Xiaoming，WENG Yihui，YOU Yancheng （1277）
Three convexification-based methods for six-degree-of-freedom powered descent guidance
…… WANG Chi，LIU Wei，GAO Yang （1292）

Maximum aerodynamic load prediction method during launching of vehicle launch
…… CHENG Huhua，WU Shuai，JIANG Zhuhui，ZHANG Rucai （1304）
Spaceborne particle identification platform and application based on convolutional neural network
…… BAI Chaoping，ZHANG Shenyi，ZHANG Xin，SUN Yueqiang，ZHANG Shuai，WANG Ziting （1313）
LPI radar signal recognition based on time-frequency reassignment algorithm
…… JI Libin，ZHU Yan，CUI Tianshu，WANG Dong，HUANG Yonghui （1324）
Parameter optimization method of thrust vector/pneumatic rudder composite control law for aircraft based on singular value method …… RUAN Shilong，DONG Zhe，SUN Yao，QU Xiaolei，HUO Shaoze （1332）
Stability control of vehicles powered by non-pneumatic wheels based on robust optimal sliding mode
…… LI Tian，ZHAO Youqun，XU Tao，SHEN Yawei，LIN Fen （1342）
Rapid planning method for lunar direct ascent and rendezvous trajectory in emergency cases
…… BAN Huanheng，ZHOU Cong，YAN Xiaodong （1352）
Directed interactive topology optimization design for multi-agent affine formation maneuver control
…… MA Suhui，ZHANG Dong，WANG Mengyang，WANG Tinghui，LIU Songdan （1367）
Anti-sweep jamming method for FM fuze based on outlier reconstruction
…… DUAN Lefan，HAO Xinhong，CHEN Qile （1377）
A deep reinforcement learning based on discrete state transition algorithm for solving fuzzy flexible job shop scheduling problem
…… ZHU Jiazheng，WANG Cong，LI Xinkai，DONG Yingchao，ZHANG Hongli （1385）
The retrieval model of shore-based GNSS-R code altimetry
…… NING Baojiao，WANG Nazi，JING Lili，GAO Fan，KONG Yahui，HE Yunqiao （1395）
A micro expression recognition method integrating LBP and parallel attention mechanism
…… LI Shuaichao，LI Mingze，SUN Jiaao，LU Shuhua （1404）

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0210

电子机械制动系统关键技术研究进展

赵立金[1,2]，杨世春[1,*]，曲婧瑶[2]

(1. 北京航空航天大学 交通科学与工程学院，北京 100191；2. 中国汽车工程学会，北京 102607)

摘 要：电子机械制动系统(electro-mechanical brake, EMB)是真正意义上的纯线控制动，完全摒弃液压/气压装置，实现人车解耦、更高的响应速度、更精准的制动力控制，是高级别自动驾驶技术的理想制动执行机构，然而由于系统可靠性、功能安全及成本等原因仍未规模量产应用。分析EMB的技术发展现状及典型系统结构，比较线性自增力式和非线性增力式两大技术路线，介绍基于楔形自增力结构的电子楔式制动器(electronic wedge brake，EWB)及混合EMB系统，并探讨其在自动驾驶环境下的适用性。针对EMB系统关键技术，重点研究了系统架构、电控单元、通信架构、失效模式等冗余设计方案，以及无传感器夹紧力估算方法的夹紧力控制技术，并对其后续产业化应用所面临的挑战进行了分析和后续研究展望。

关 键 词：电子机械制动；系统构型；冗余设计；夹紧力控制；功能降级

中图分类号：U463.5

文献标志码：A **文章编号**：1001-5965(2025)04-1037-11

制动系统是汽车安全行驶最重要的保障，汽车电动化、智能化、网联化驱动制动系统变革，对制动系统提出新的要求：自动驾驶功能需要高动态建压、低延迟、高控制精度要求；低碳化、低能耗需要高能量回收率、低拖滞甚至“0”拖滞力卡钳。满足自动驾驶需求的安全冗余设计有：电源系统、通信系统、传感系统、计算单元、执行单元等各个环节具备软硬件冗余；运动控制算法集成，适应整车域控电子电气架构的演变；驾驶性能具备自学习功能；驱动和制动系统需要满足不同驾驶习惯的踏板感，通过驾驶风格自学习实现踏板感可调，满足用户体验持续迭代，覆盖产品量产后的应用及功能升级。

线控制动以电子液压制动(electro-hydraulic brake, EHB)系统和电子机械制动(electro-mechanical brake, EMB)系统为主要形式[1]。EHB系统替代了真空助力器，根据制动踏板与制动轮缸之间的耦合程度可以采用非解耦、半解耦或全解耦3种模式，但其以液压为制动能量源，液压控制管路相对复杂，液压的产生和电控化相对比较困难，不易与其他电控系统整合[2-3]。而EMB即电子机械式线控制动，主要由踏板模拟器、EMB执行器及控制器等组成，通过控制执行器的夹紧力来实现对每个车轮制动力的独立控制[1,4]，避免了液压系统 P-V 特性对响应速度的影响，提高了响应速度，结构简单、控制灵活、效率更高，实现完全人车解耦，是真正意义上的纯线控制动。EHB线控制动已处于量产应用及完善阶段，EMB线控制动技术渗透尚处于起步发展阶段，大规模应用仍需技术突破和成本下降。此外，EMB可以有效集成电子稳定程序(electronic stability program, ESP)、自动紧急制动(autonomous emergency brake, AEB)、自适应巡航(adaptive cruise control, ACC)和自动泊车(automatic parking, AP)等技术，能够大幅提高车辆行驶的稳定性和安全性[5]。EMB是高级别自动驾驶技术的理想制动执行机

收稿日期：2023-04-25；**录用日期**：2023-05-19；**网络出版时间**：2023-07-05 17:37
网络出版地址：link.cnki.net/urlid/11.2625.V.20230705.1529.001
基金项目：国家重点研发计划(2022YFB2503105,2021YFB2501700)
*通信作者. E-mail：yangshichun@buaa.edu.cn

引用格式：赵立金，杨世春，曲婧瑶. 电子机械制动系统关键技术研究进展[J]. 北京航空航天大学学报，2025，51(4)：1037-1047.
ZHAO L J，YANG S C，QU J Y. Research progress of EMB systems key technology[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1037-1047 (in Chinese).

构，符合自动驾驶和低碳化汽车发展进程需求[6]。首先，本文分析EMB系统发展现状与典型系统结构；然后，对EMB系统的冗余设计方案、夹紧力控制等关键技术进行深入研究。

1 EMB系统发展历程与典型结构

1.1 EMB系统发展现状

1899年，Sperry电气铁路公司提出安装有电磁铁的电磁盘式制动器。在此基础上，对结构进行改进提出电磁摩擦离合器式的盘式制动器结构，并应用在汽车上[2]。瑞典斯凯孚集团在2001年日内瓦汽车展览会上亮相的Bertone-SKFFILO概念车第一次公开展示了其EMB技术[7]；2004年，德尔福公司推出了混合线控制动系统，用电动制动钳代替传统液压制动钳，并应用于后轮制动[8]。随着电机驱动技术和减速器的发展及汽车的电动化，从20世纪90年代开始，国外众多著名的零部件厂商纷纷开始对EMB系统进行研究，如德国博世、德国西门子、德国大陆、韩国摩比斯、韩国万都、美国天合、瑞典Halde等公司均已申请相关专利并陆续开发了EMB系统测试产品。2013年，舍弗勒开发无人物流车搭载轮毂电机及EMB系统；2016年，奥迪公司开发的R8 e-tron车型采用了EHCB电液制动和电子机械复合制动系统；2021年，意大利布雷博也研发出EHB和EMB混合线控制动系统，并进行了实车试验。由于商用车的空间优势，对其进行EMB系统开发应用较乘用车有一定优势，最接近量产的是瀚德万安的产品。瀚德万安[9]的商用车EMB系统组成采用桥控单元控制执行整车策略，建立桥控互联机制，并以桥控单元为核心节点，在脚刹模组建立自身控制机制，构建多回路冗余安全网络。

1.2 典型构型

执行机构是EMB与EHB系统区别最大的部分，其设计也是EMB系统的关键技术之一，从传动原理上通常有两大技术路线，分别为线性自增力式和非线性增力式[10]。

线性增力式EMB系统结构简单，是目前国内外研究与专利较多的类型，电机输出的旋转力矩经过行星或定轴齿轮机构减速增扭、再通过滚珠丝杠将旋转运动转化为平动位移传递给摩擦制动器，制动力与电机输出力矩为线性关系。典型的线性非自增力式EMB系统构型以德国大陆、德国博世直驱构型为代表，电机可采用横向磁通电机或永磁同步电机，是目前应用最广泛的一种结构[10-11]。德国博世EMB[12]方案采用电机外置匹配行星齿轮机构与滚珠丝杠副，利用电磁离合器接合断开可以实现4种工况，如图1所示。德国大陆集团EMB[13]方案采用电机内置匹配行星齿轮机构与滚珠丝杠副，间隙调整方式为智能控制，有棘轮机构实现制动力保持和驻车，如图2所示。以上2种方案轴向空间占用较大，且需要大直径空心电机，对电机的性能要求高，实际制动响应偏慢。

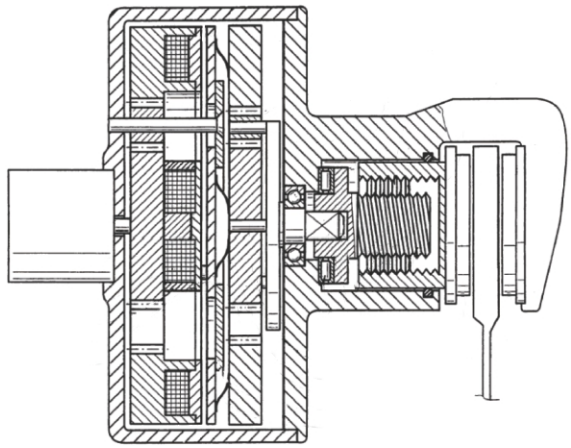


图1 德国博世EMB结构[12]

Fig. 1 Structure of Bosch EMB[12]

非线性自增力式EMB系统利用楔形块的自增力原理，取消齿轮减速增扭机构，以Siemens VDO的电子楔式制动器（electronic wedge brake, EWB）为代表[1,10]，楔形自增力结构可以匹配低功率小型电机配合滚珠丝杠驱动楔形块。装有西门子EWB系统的宝马样车在结冰路面上测试，初始速度为80 km/h，制动距离为64.5 m，比装有防抱死制动系统（anti-lock brake system, ABS）液压制动系统缩短了15%[14]，其采用2个电机对称布置，占用空间小，结构紧凑。楔形块增力原理如图3所示。在进入制动状态后，利用制动盘旋转动能作为自增力制动动力源，将整车的部分动能转化为制动能量，2部分能量共同作用，使制动器达到越夹越紧的自增力效果，当达到目标制动力或初始车轮转速较高时，电机提供反向扭矩给楔形块，避免越夹越紧的状态[10]。

杨岩松[10]针对楔形自增力式“死点”可控性差的问题，提出单浮动楔形传动增力构型（见图4）和双浮动楔形传动增力构型2种改进方案，将电机、减速机构、电机锁紧装置横置，同时引入滑动滚针排提高传动效率，并提出一种最优楔形传动角优化方法，通过对活塞增加径向约束，避免楔形传动机构出现“死点”的非稳态状态，但该构型存在克服间隙第2阶段加载较慢等问题。Xu和Cho[15]设计

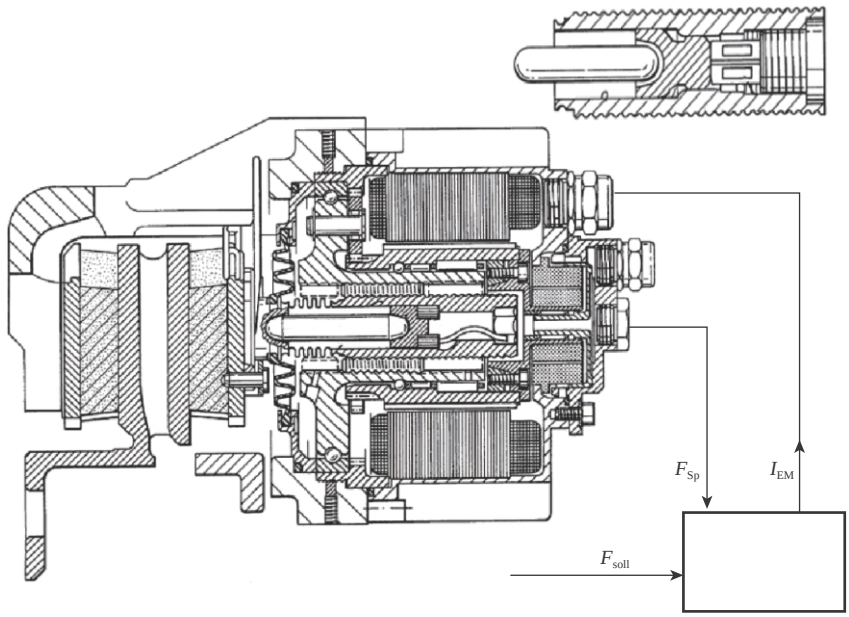


图 2　德国大陆 EMB 结构[13]

Fig. 2　Structure of Continental EMB[13]

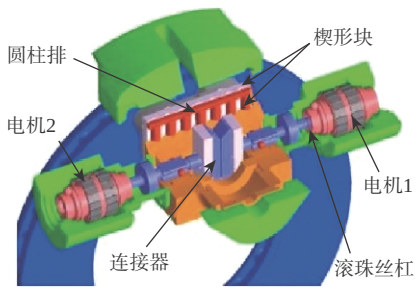
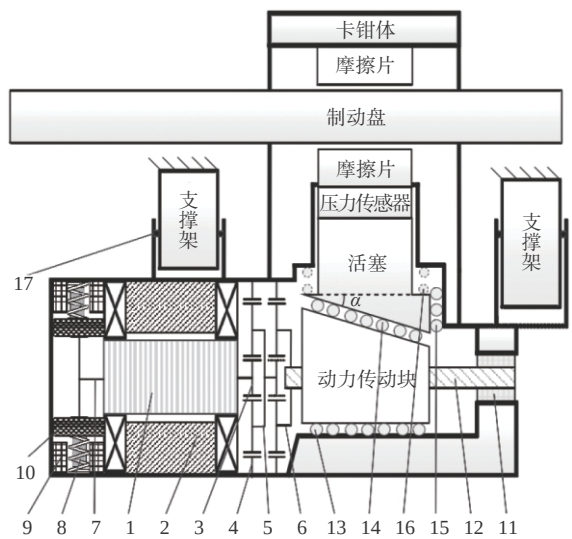


图 3　西门子楔形自增力 EWB 系统

Fig. 3　Siemens wedge self EWB system

了一种单电机无行星齿轮和滚珠丝杠的 EWB 系统，如图 5 所示，通过步进电机驱动无滚珠丝杠推动楔形块，进而促动内制动摩擦片，简化机械传动结构，降低系统能耗，在 EWB 模型中引入自抗扰控制 (active disturbance rejection control, ADRC) 算法，相较于 PID 和滑模变结构控制（sliding mode control, SMC）算法，提高了 EWB 系统的抗干扰能力和响应时间。

以上 2 种系统构型方案各有利弊，前者的机构原理简单易控制，但是减速增扭的效果在一定程度上取决于电机，为应对夹紧力需求，往往希望升压到 24 V 或 48 V；而后者的执行机构放宽了对电机的要求，可以兼容采用现有整车 12 V 低压电气平台，但制造工艺相对复杂，且不易进行转矩的精确控制[12]。EWB 和 EMB 系统实物图如图 6 所示。Cheon 等[16-17] 提出一种前 EWB 后 EMB 系统构型方案，兼容 12 V 电压平台满足前轴制动力需求，该方案为前后轴异构方案，有利于软件系统冗余设计。Cheon 等[16-17] 在完成样件设计和台架性能验证后，



1. 转子；2. 定子；3. 一级太阳轮；4. 一级齿圈；5. 一级行星架；6. 二级行星架；7. 转子制动轴；8. 电磁阀芯及复位弹簧；9. 电磁线圈；10. 转子制动摩擦片；11. 丝杠轴承；12. 丝杠；13、14、15. 滑动滚针排；16. 活塞复位弹簧；17. 弹性密封环。

图 4　单浮动楔形传动增力构型[10]

Fig. 4　Single floating wedge-shaped force augmentation[10]

进行了整车性能验证，整车制动减速度可以达到 1.0g，如图 7 所示。

国内高校和制动零部件企业针对 EMB 系统进行了开发研究，清华大学、浙江大学、吉林大学、同济大学、重庆大学等高校均对 EMB 系统开展了工程设计及系统特性研究[9]；此外，菲格智能科技有限公司、浙江亚太机电股份有限公司[18]、芜湖伯特利汽车安全系统股份有限公司[19-20]、重庆长安汽车股

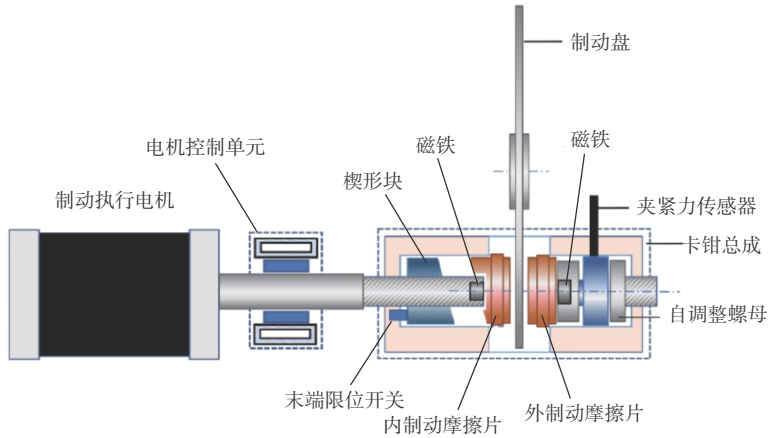
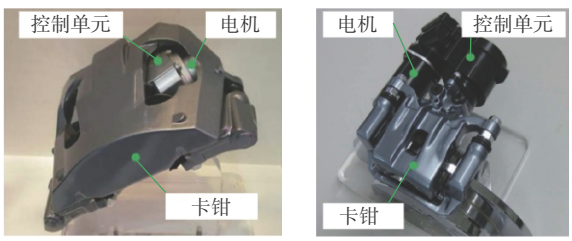


图5 单电机EWB系统[15]

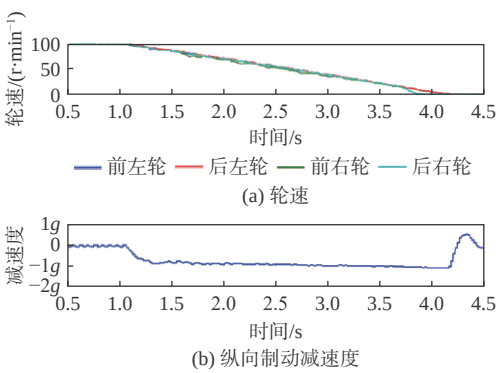
Fig. 5 Single motor EWB system[15]



(a) EWB (b) EMB

图6 EWB和EMB系统实物图

Fig. 6 Pictures of EWB and EMB system



(a) 轮速

(b) 纵向制动减速度

图7 整车制动性能测试结果

Fig. 7 Vehicle brake performance test results

份有限公司[21]等企业也开发了EMB产品，申请了相关专利。研究结构紧凑、有安全备份系统、安全可靠的EMB系统是目前国内外重点研究的问题。

2 EMB系统关键技术

EMB系统冗余设计与容错控制、夹紧力控制技术、复杂环境下无刷电机可靠性、减速增力机构创新化小型化是制约其量产应用的关键技术。基于EMB系统的整车动力学控制系统，不同于基于液压的闭环控制，需要开发全新动力学模型及整车协同控制算法。

2.1 系统冗余设计方案

与传统制动系统和EHB线控制动系统相比，驾驶员无法成为EMB系统的冗余备份，这也是如今制约EMB系统量产化面临的主要挑战[22-23]。联合国欧洲经济委员会(The United Nations Economic Commission for Europe, UNECE)制定的UNECE-R13H[24]和加拿大运输署制定的《液压和电动制动系统》[25]关于制动系统的技术条件和功能安全要求，增加了EMB制动系统在系统设计、性能指标、单点失效后的性能衰退等要求，从系统设计上要求独立供电装置、独立储能装置及任何一项电气系统单点失效引起的性能降级。

2.1.1 执行机构冗余

目前，常见的执行机构冗余设计包含双电机单制动卡钳冗余构型(见图8和图9)和双电机双减速机构冗余构型(见图10和图11)。现有冗余执行机构主要通过增加中间齿轮传动[26-27](见图8和图10)和带传动[21,28](见图9和图11)实现双动力耦合传递，且双电机布局通过控制力矩匹配可以有效降低摩擦片的不均衡磨损。双电机EMB可用于需要制动扭矩较大的主制动轮(即车辆前轮)，而对于车辆后轮，EMB可通过配置一个制动电机实现。博世

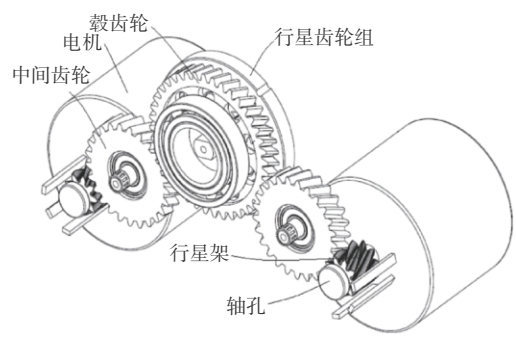


图8 博世双电机冗余构型[26]

Fig. 8 Bosch dual motor redundant structure[26]

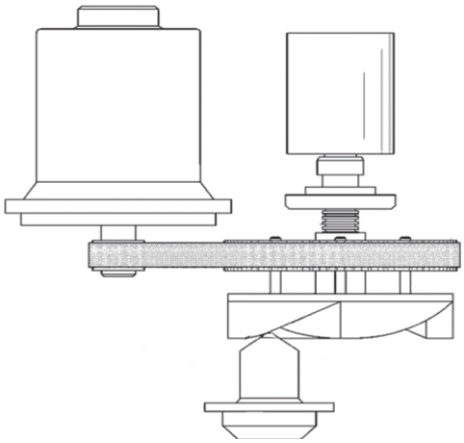


图 9　万都双电机冗余构型[27]

Fig. 9　Wando dual motor redundant structure[27]

2022 年公布的 EMB 专利方案采用双电机单减速机构[26]，执行电机和制动执行装置位于传动机构的同一侧，可以获得较小的轴向尺寸以便布置在车轮轮毂内。EMB 多集成 EPB 功能，万都 EMB[27] 通过冗余小电机旋转推动导向机构产生直线位移，推动制动盘夹紧实现驻车功能，可通过 EMB 电子控制单元（electronic control unit, ECU）上的 EPB 硬件开关或整车上的 EPB 软开关，接受请求电子手刹夹紧或释放，EMB 总成在夹紧卡钳后，按照上述过程进行锁止实现驻车。

2.1.2　电控系统冗余

1）电源系统冗余。根据系统功能安全 ASIL D 要求，电源系统需要进行冗余设计，具备电源分配和系统诊断等功能[23,29]。应用在电动汽车的 EMB 供电系统主要通过高压动力电池（400 V/800 V）和直流电压转换器（direct current to direct current converter, DCDC）进行高低压转化，获得 12 V/24 V/48 V 不同电压等级。传统液压制动系统通常采用 X 型或 H 型供电拓扑来进行电源冗余设计。文献 [23] 所述电源冗余方案可分为双低压蓄电池 X 型供电、H 型供电、全冗余供电拓扑和高压低压动力源同时提供 X 型供电拓扑，如图 12 所示。目前，很多 EMB 系统设计方案都采用 X 型双电源供电冗余拓扑。Reinhard[30] 采用支持电容局部能量存储器用于短时备份供电系统失效；刘强等[31] 发明了一种基于高压动力电池和低压蓄电池的异构电源冗余备份方案，主供电电路包含高压动力电池和 DCDC，份供电电路包含低压蓄电池、2 个超级电容、DCDC，冗余电源回路 DCDC 用于将低压蓄电池升压至 48 V 与超级电容充电。

2）电控单元冗余。电控单元冗余通常分为双控单元、三控单元和四控单元。双控单元系统的冗余可以分为主备方式、冗余互备方式和冗余双工方式[32]。主备方式是指当某套失效时，完全切换到另一套；冗余互备方式是指备份控制器仅接受信息不计算，主控制器故障时备份控制器接管；冗余双工方式是指 2 个控制器各自进行计算，不进行交互，自检故障时自主切断，另一路工作。另一种实现双工方式的是锁步核[33]，通过一个芯片上的 2 个处理器处理命令和监控。Lee 等[34] 通过 EMB 采用非对称双核 ECU 控制器，结合外部看门狗装置为主处理器提供电源监控，通过系统硬件在环仿真验证，能有效减少微控制器单元时钟系统的共因失效，当 ECU 出现故障或软件执行故障时，该控制系统硬件设计方案能够满足功能安全性要求。根据底盘域集成控制的发展趋势，EMB 控制系统朝着轮端执行控制器和中央控制单元混合拓扑架构方向发

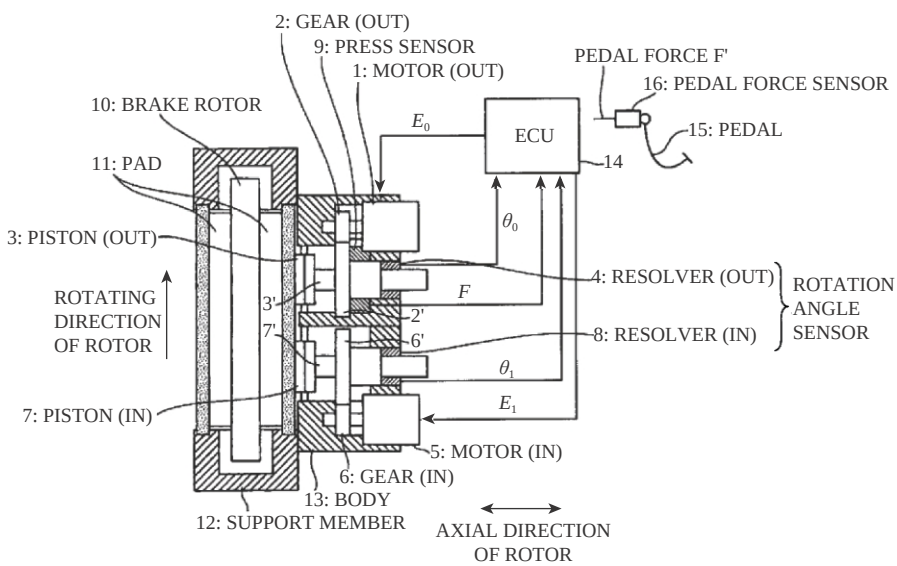
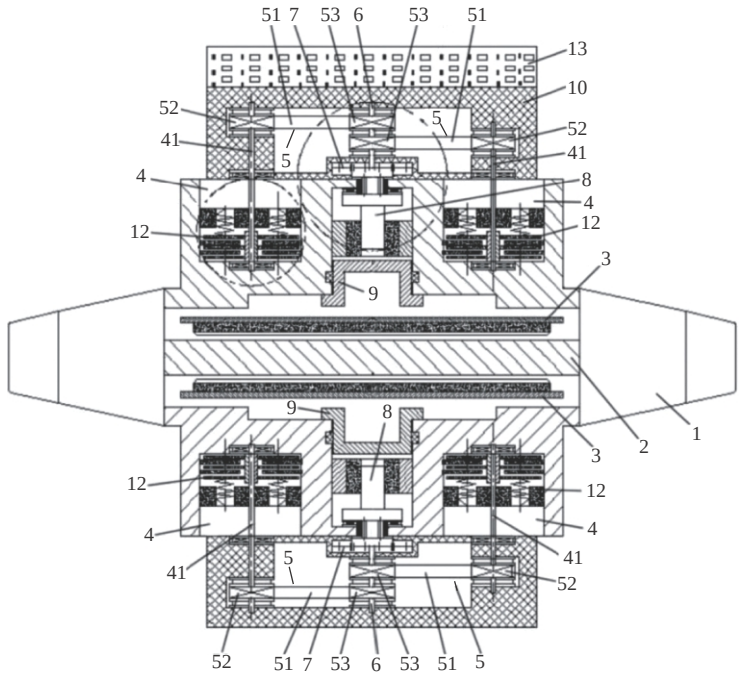


图 10　曙光制动 EMB 冗余构型[28]

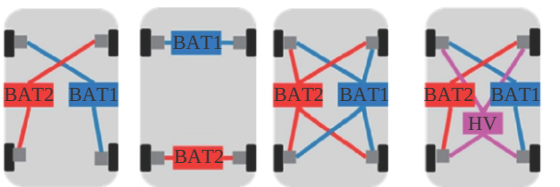
Fig. 10　Akebono brake EMB redundant structure[28]



1. 制动钳壳体; 2. 制动盘; 3. 制动片; 4. 电机; 5. 带传动机构; 6. 传动轴; 7. 减速机构; 8. 行星滚柱丝杠; 9. 活塞; 10. 减速箱壳体; 12. 驻车机构; 13. 制动控制器; 41. 电机轴; 51. 传动带; 52. 第 1 带轮; 53. 第 2 带轮。

图 11 重庆长安双电机双减速机构冗余构型[21]

Fig. 11 Chongqing Changan dual motor dual gearbox redundant structure[21]



(a) X型冗余供电拓扑 (b) H型冗余供电拓扑 (c) 全冗余供电拓扑 (d) X型+动力电池冗余供电拓扑

图 12 电源系统冗余拓扑方案[23]

Fig. 12 Power system redundancy topology solutions[23]

展[6], 如图 13 所示。Anders 等[35] 发明了一种通过增加原始信号和执行端独立的通信网络, 应对双工系统(制动集成控制器或底盘域集成控制器)信息不一致时的系统控制方法, 如图 14 所示。

3) 通信冗余。随着汽车智能网联化发展, 增加了车内网络通信性能的要求, 车载 Ethernet 满足低成本、高可靠性、带宽分配、低延迟及同步实时性等方面的要求, 被认为是解决智能网联汽车车内网

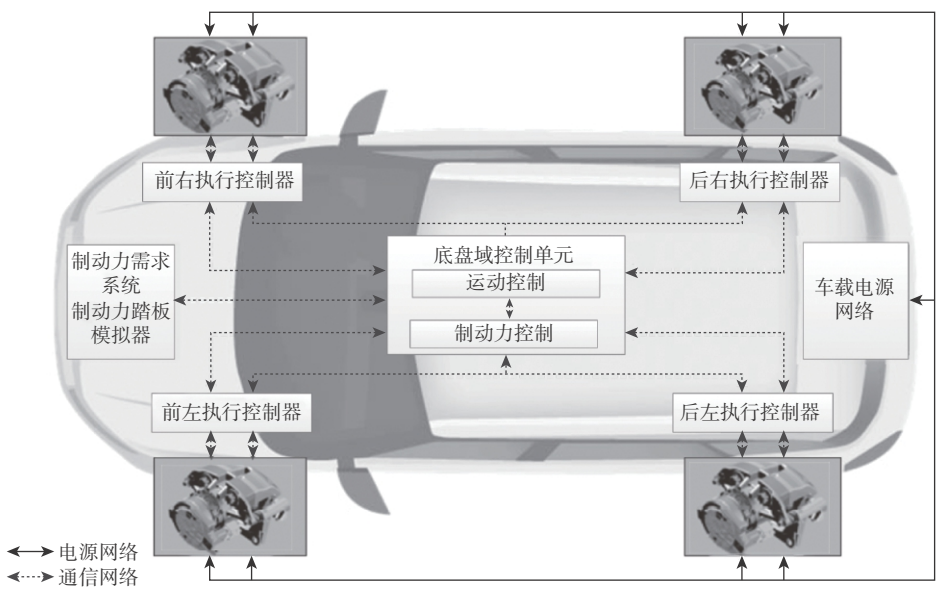


图 13 日立智能制动系统架构[6]

Fig. 13 Hitachi EMB control architecture[6]

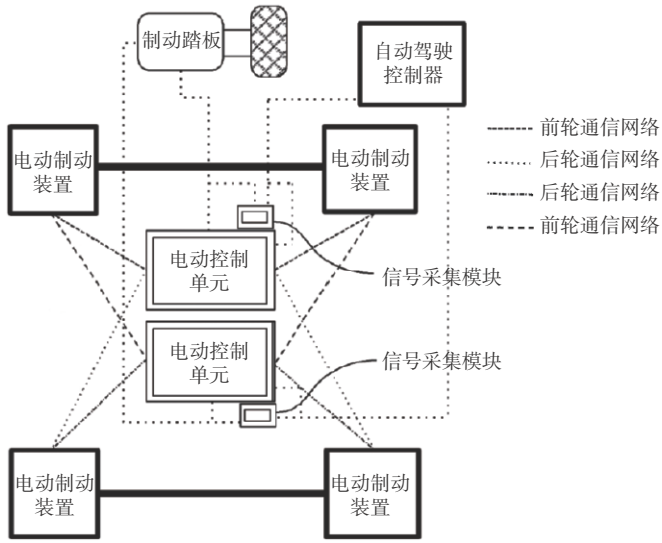


图 14　瀚德万安 EMB 系统控制架构[35]

Fig. 14　Haldex EMB system control architecture[35]

络带宽需求的有效技术[36]。常见的车载通信网络拓扑形式主要为总线拓扑、星形拓扑、环形拓扑、网状拓扑及混合型拓扑，混合型拓扑包含星形总线拓扑（见图 15）和环形总线拓扑（见图 16）[37]。分布式星形拓扑冗余总线，中央网关仍然存在单点失效风险，需要对中央网关备份。环形通信网络拓扑提供了非常高的鲁棒性，可以防止局部机械、电气故障单点失效，研究结果表明，环形总线拓扑可以满足未来 EMB 系统安全冗余设计要求[23]。

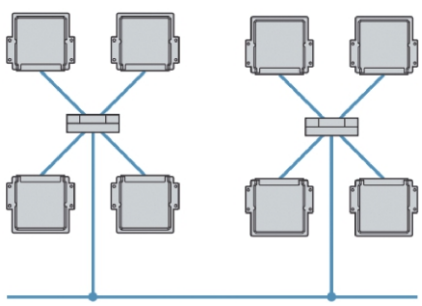


图 15　星形总线拓扑[37]

Fig. 15　Star bus topology[37]

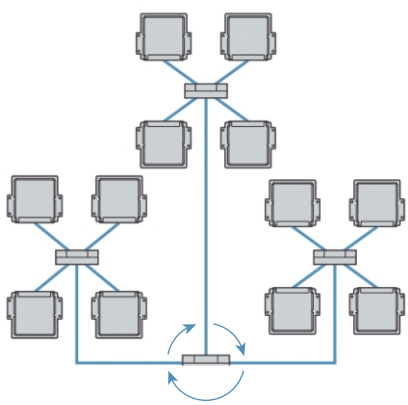


图 16　环形总线拓扑[37]

Fig. 16　Ring bus topology[37]

Cheon[16] 采用两路踏板行程传感器和一路踏板力传感器构建三重冗余系统，通过仲裁系统监测故障信号。文献 [38] 的 EMB 控制系统冗余方案包含踏板模拟器、主控 ECU、冗余 ECU、EMB 总成、轮速传感器、主电源及冗余电源；踏板行程传感器包括 1 路踏板行程传感器和 2 路踏板行程传感器，主控 ECU 与冗余 ECU 通过私有 CAN 连接，相互进行信号校验，校验主控 ECU 和冗余 ECU 的有效性，2 路踏板行程传感器信号和 EMB 执行情况，如图 17 所示。

2.2　EMB 系统夹紧力控制

现阶段对 EMB 控制算法的研究主要集中在精确输出所需制动夹紧力、提升制动系统反应速度、相关临界点识别等[39]。夹紧力的控制效果直接关系到车辆的制动性能，如何精确快速地控制各个车轮的夹紧力是目前 EMB 系统的研究重点和难点[5]。EMB 系统中的夹紧力传感器较贵，且安装在制动垫片附近存在高温工艺设计和参数稳定性问题[39-40]，其使用寿命和夹紧力测量精度面临挑战，目前，各车企和高校正在开展无传感器夹紧力控制技术研究。

应用于 EMB 系统的制动力控制算法主要基于传统 PID，其核心是基于执行系统的模型及目标制动力对电机转矩、转速、转角三闭环控制[5,40-41]。但是，实际应用中存在电机电流环、转速环、夹紧力控制回路的强耦合关系[5,40]，参数整定工作量大，且系统控制精度鲁棒性不足，难以处理系统摩擦和刚度等引起的非线性扰动问题。通过增加库仑力、迟滞摩擦力、Stribeck 摩擦的摩擦力模型对负载进行前馈补偿控制[20,40]，可以显著提高夹紧力控制的精确性和鲁棒性。Line[42] 研究执行器饱和、负载扰动

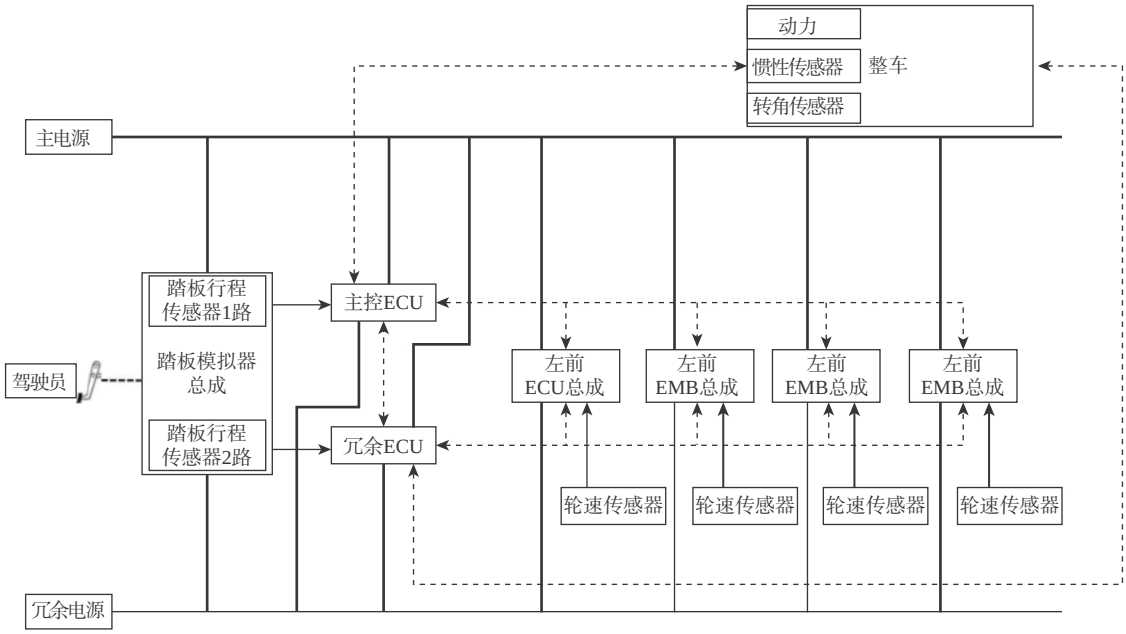


图 17 重庆长安 EMB 控制系统方案[38]

Fig. 17 Chongqing changan EMB control system solution[38]

和系统非线性刚度模型，通过前馈逆增益补偿隔离系统因温度、磨损等引起的非线性刚度，叠加摩擦力和电机扰动反馈线性化补偿和模型预测控制（model predictive control, MPC）控制优化电机转矩控制，可显著提升低制动负荷需求下的夹紧力控制精度和响应时间，但是系统性能对建模误差较为敏感。De Castro 等[43]研究了鲁棒性 H_∞ 最优控制器应对执行单元最差应用场景，通过与串级 PI 控制效果进行对比分析，鲁棒最优控制对执行器给系统带来的扰度不敏感，但是系统控制性能过于保守，难以兼顾高响应控制性能和系统不确定性。Lee 和 Manzie[44]研究了基于线性变参数系统基于线性变参数（linear parameter varying, LPV）和自适应前馈补偿，以车轮角速度和加速度作为调度依据，实验研究表明，LPV 控制结构对制动抖动具有良好的控制效果，但是采样率要求高。

目前，常用的 EMB 系统无压力传感器制动力伺服控制主要分为 3 类：转矩动态方程法、电机转角与夹紧力的特性曲线法、实验方法[45-46]。针对电机转角与夹紧力的特性曲线法，Kwon 等[46]采用三阶多项式特性曲线进行夹紧力估算可以满足制动和释放工况的最优精度要求；文献 [41] 重点研究转矩夹紧力平衡方程中力矩放大系数的修正方法，特别是在低扭矩区间通过查表获得转矩和夹紧力对应的非线性区；Xu 和 Gerada[47]采用超螺旋观测器（super-twisting observer, STO）进行基于动态扭矩平衡方程的夹紧力预测，通过试验证明 STO 在高频动态跟踪过程性能优异；Fu 等[48]研制了一种磨损间隙自动调整方案，可以满足 UNECE-R13H[24]法规要求，提高 EMB 系统安全；张俊峰等[49]在考虑系统传动间隙的基础上，基于转角夹紧力特性曲线进行夹紧力估算，应用滑模控制方法降低建模和扰动系统误差，提高夹紧力估计精度和控制精度；文献 [40,50] 通过最大行程夹紧力曲线与任意点释放夹紧力曲线的相似推演特性，并通过不同电机电流控制斜率试验进行模型修订，采用台架试验标定摩擦片磨损量的接触点位移、夹紧和释放过程夹紧力与电机转角曲线组、夹紧和释放过渡区曲线组进行查表计算，对夹紧力进行估算；Riva 等[51]研究了一种纯数据驱动的夹紧力估计模型，包含执行器与环境，基于非线性静态曲线和线性动态系统串联而成的 Hammerstein 模型进行训练和辨识，并通过台架试验对比该数据驱动算法模型和特性曲线法对夹紧力估计的误差，结果表明，数据驱动模型具有更小的估计误差，但所需训练数据量巨大，目前研究大多在常温和未磨损状态下。

2.3 EMB 系统功能降级

文献 [24-25,52] 制动系统技术要求中，对于系统失效后应急制动性能的要求是基于传统真空助力的制动系统提出的。但是，针对全 EMB 或 EHB+EMB 混合制动系统，可以结合新能源汽车驱动电机能量回收、EPB 系统等多层跨系统冗余方案，通过线控制动系统的诊断策略识别不同失效模式匹配不同冗余备份策略，达到优化制动系统失效时整车制动性能的目的。

为满足 L4\L5 高级别自动驾驶功能的实施，自动驾驶对制动系统备份的功能和性能要求进一步提高，可以提供制动减速度大于 6.4 m/s^2，且达到

6.4 m/s^2 减速的时间不超过 1.5 s[53]；具备防抱死性能，使车辆维持转向能力。根据 EMB 系统冗余方案，进行不同等级失效降级模式设计。采用双电机自身冗余备份系统或多绕组电机的 EMB 系统，当 EMB 系统单轮失效时，由于双电机或双绕组存在，通过制动力重新分配，协调转向、悬架配合控制车身姿态，结合驱动电机再生制动，可以提供大于 6.4 m/s^2 的减速度[4]；如果 EMB 系统出现多轮失效，单纯采用驱动电机能量回收制动，仍然可以提供大于 2.9 m/s^2 的减速度[54]，同时需要保留 ABS 控制功能，AEB 功能根据失效模式可以进行性能降级。

3　结 束 语

EMB 系统相比电子液压助力制动系统和电气助力制动系统具有结构精简、响应迅速、控制精度高、易于实现与整车纵向动力学其他控制功能相集成等优势，符合智能车辆模块化、集成化、机电一体化、智能化、低碳化的发展需求，成为行业未来发展的重要趋势。但是，EMB 系统小型化、可靠性、功能安全设计、夹紧力控制仍然是影响其量产应用的重点和难点，且控制算法的优劣影响其与自动驾驶系统的功能集成。

本文重点分析了 EMB 系统结构设计、功能安全设计和夹紧力控制的研究进展，并就进一步的研究作出以下展望：

1）乘用车受限于空间问题，EMB 系统传动机构创新和小型化设计仍然是今后研究的重点内容，继续向结构简单、性能可靠稳定、整体空间紧凑等方向发展；且 EMB 系统易于与轮毂电机集成在轮端形成独立的模块，对于整车底盘系统架构、减少线束、系统冗余备份有重要的意义，需进一步研究高集成化轮端模块。

2）虽然国内外有相关文献研究 EMB 系统功能安全设计，包括需求分析、系统架构和拓扑设计，但是对冗余备份的降级机制和各种故障安全临界性的文献不多，且鲜有明确的评价方法，有待进一步深入研究。

3）系统的动刚度非线性特性、系统温度和耐久后参数变化，是 EMB 无传感器夹紧力控制面临的难题，未来可以进一步研究如何解决数据驱动等算法的训练时间，提高系统精度和鲁棒性，降低对模型精度和温度等参数的敏感性。随着对 EMB 系统控制研究的逐渐深入，相关学者进一步展开基于四轮 EMB 的纵向横向动力学协同控制研究。

4）法规适应性是 EMB 线控制动产品落地的基本保障，需要加紧研究其应急制动系统和各类测试验证条件和性能要求。

参考文献（References）

[1] 邓美俊, 孙仁云, 潘湘芸, 等. 汽车电子机械制动系统技术发展分析[J]. 汽车零部件, 2021(9): 103-109.
DENG M J, SUN R Y, PAN X Y, et al. Research status analysis of automotive electro-mechanical brake system[J]. Automobile Parts, 2021(9): 103-109(in Chinese).

[2] 雍健羽, 王洪亮, 董金聪, 等. 轻型车辆线控制动技术研究现状及发展趋势综述[J]. 中国汽车, 2022, 32(11): 16-22.
YONG J Y, WANG H L, DONG J C, et al. Review on the research status and development trend of wire control technology[J]. China Auto, 2022, 32(11): 16-22(in Chinese).

[3] 牟希东. 面向动力学控制的电子机械制动系统控制策略研究[D]. 长春: 吉林大学, 2021.
MU X D. Research on control strategy of electro-mechanical braking system oriented to dynamic control[D]. Changchun: Jilin University, 2021(in Chinese).

[4] 夏利红. 基于 EMB 的分布式复合制动系统研究[D]. 重庆: 重庆大学, 2019.
XIA L H. The study on distributed hybrid braking system based on EMB[D]. Chongqing: Chongqing University, 2019(in Chinese).

[5] 张奇祥, 靳立强, 靳博豪, 等. EMB 夹紧力控制与传感器故障诊断研究进展[J]. 汽车工程, 2022, 44(5): 736-746.
ZHANG Q X, JIN L Q, JIN B H, et al. Research progress of EMB clamping force control and sensor fault diagnosis[J]. Automotive Engineering, 2022, 44(5): 736-746(in Chinese).

[6] NEZU T, INOUE Y, USUI T. Electronic control technology for chassis subsystems[J]. Hitachi Review, 2022, 71(1): 76-77.

[7] BROWN S, HOLWEG E. Smart electro-mechanical actuation for drive-by-wire applications[C]//Proceedings of the SAE Technical Paper Series. Warrendale: SAE International, 2004.

[8] 德尔福集团. 德尔福混合线控制动系统[J]. 汽车与配件, 2004(23): 48.
Delphi. Delphi hybrid brake-by-wire[J]. Automobile & Parts, 2004(23): 48(in Chinese).

[9] 高国伟. 面向线控底盘平台的弯道制动力分配控制策略研究[D]. 天津: 天津职业技术师范大学, 2022.
GAO G W. Research on bend braking force distribution control strategy for wire controlled chassis platform[D]. Tianjin: Tianjin University of Technology and Education , 2022(in Chinese).

[10] 杨岩松. 基于楔形传动的电子机械线控制动系统的研究与设计[D]. 广州: 华南理工大学, 2020.
YANG Y S. Research and design of the electro-mechanical brake-by-wire system based on wedge transmission[D]. Guangzhou: South China University of Technology, 2020(in Chinese).

[11] 宋瑞琦. 商用车电子机械制动系统设计与控制策略研究 [D]. 青岛: 青岛理工大学, 2022.
SONG R Q. Research on design and control strategy of electro-mechanical braking system of commercial vehicle[D]. Qingdao: Qingdao University of Technology, 2022(in Chinese).

[12] KELLER F. Electromagnetic wheel brake device: US6536561[P]. 2003-03-25.

[13] CAO C T, BAUMANN D, HOFMANN D, et al. Self-reinforcing electromechanical disc brake: US8439170[P]. 2013-05-14.

[14] NEWTON M. Siemens' wedge brake hits the ice[J]. Automotive Engineering International, 2007, 115(7): 26-30.

[15] XU F, CHO C. A novel electronic wedge brake based on active disturbance rejection control[J]. Energies, 2022, 15(14): 5096.

[16] CHEON J S. Brake by wire system configuration and functions using front EWB (electric wedge brake) and rear EMB (electro-mechanical brake) actuators[C]//Proceedings of the SAE Technical Paper Series. Warrendale: SAE International, 2010.

[17] CHEON J S, KIM J, JEON J. New brake by wire concept with mechanical backup[J]. SAE International Journal of Passenger Cars-Mechanical Systems, 2012, 5(4): 1194-1198.

[18] 姚兴娟, 李学佳, 张水清, 等. 一种电子机械制动器: CN218294260U[P]. 2023-01-13.
YAO X J, LI X J, ZHANG S Q, et al. An electronic mechanical brake: CN218294260U[P]. 2023-01-13(in Chinese).

[19] 郭锋亮, 叶乙兴, 朱元澄, 等. 一种制动钳总成: CN218063174U[P]. 2022-12-16.
GUO F L, YE Y X, ZHU Y C, et al. Brake caliper assembly: CN218063174U[P]. 2022-12-16(in Chinese).

[20] DE MORAIS A E, 许志, 李运动, 等. 多缸同步建压电控制动卡钳: CN217926862U[P]. 2022-11-29.
DE MORAIS A E, XU Z, LI Y D, et al. Multi-cylinder synchronous built-in piezoelectric control caliper: CN217926862U[P]. 2022-11-29(in Chinese).

[21] 郑祖雄, 刘磊, 邹佳航. 一种带传动制动钳、制动系统、汽车和设计方法 : CN115507141A[P]. 2022-12-23.
ZHENG Z X, LIU L, ZOU J H. Belt transmission brake caliper, brake system, automobile and design method: CN115507141A[P]. 2022-12-23(in Chinese).

[22] LIU H Z, DENG W W, HE R, et al. Fault-tolerant control of brake-by-wire systems based on control allocation[C]//Proceedings of the SAE Technical Paper Series. Warrendale: SAE International, 2016: 923-936.

[23] SCHRADE S, NOWAK X, VERHAGEN A, et al. Short review of EMB systems related to safety concepts[J]. Actuators, 2022, 11(8): 214.

[24] The United Nations Economic Commission for Europe. Uniform provisions concerning the approval of passenger cars with regard to braking: UNECE-R13H[S]. Geneva: The United Nations Economic Commission for Europe, 2015.

[25] Transport Canada. Hydraulic and electric brake systems: technical standards document No. 105, revision 5[S]. Ottawa: Transport Canada, 2015.

[26] 张晓坤, 于飞. 电子机械制动器和车辆: CN218177792U[P]. 2022-12-30.
ZHANG X K, YU F. Electromechanical brake and vehicle: CN218177792U[P]. 2022-12-30(in Chinese).

[27] 沈暟勋, 张在勋. 电动制动系统及其操作方法: CN115087814A[P]. 2022-09-20.
SHEN J X, ZHANG Z X. Electric brake system and operating method thereof: CN115087814A[P]. 2022-09-20(in Chinese).

[28] TAKAHASHI H, TAKAHASHI K. Electric brake: US11523500[P]. 2010-10-05.

[29] KÖHLER A, BERTSCHE B. An approach of fail operational power supply for next generation vehicle powernet architectures[C]//Proceedings of the 30th European Safety and Reliability Conference and 15th Probabilistic Safety Assessment and Management Conference. Singapore: Research Publishing Services, 2020: 60-67.

[30] REINHARD W. Electric brake system, particularly electro-mechanical brake system for motor vehicle, has brake circuit, and control unit for implementation of driver braking demand: DE102009046231[P]. 2011-05-11.

[31] 刘强, 秦宬, 付永泉, 等. 车辆电子机械制动系统和具有其的车辆: CN112550189B[P]. 2021-02-10.
LIU Q, QIN C, FU Y Q, et al. Vehicle electromechanical brake system and vehicle with same: CN112550189B[P]. 2021-02-10(in Chinese).

[32] 孙冰. 线控转向系统冗余安全方案设计与测试研究[D]. 长春: 吉林大学, 2022.
SUN B. Research on design and testing scheme for redundant security of steer-by-wire system[D]. Changchun: Jilin University, 2022 (in Chinese).

[33] PUTZ M H, SEIFERT H, ZACH M, et al. Functional safety (ASIL-D) for an electro mechanical brake[C]//Proceedings of the SAE Technical Paper Series. Warrendale: SAE International, 2016.

[34] LEE K J, KI Y H, CHEON J S, et al. Approach to functional safety-compliant ECU design for electro-mechanical brake systems[J]. International Journal of Automotive Technology, 2014, 15(2): 325-332.

[35] ANDERS N, ANDERS L, DONG X B. A brake system for a vehicle: WO2021139954A1[P]. 2021-07-01.

[36] 尹荣彬, 陈博, 孔祥明. 智能网联汽车车载以太网技术应用研究[J]. 汽车文摘, 2019(11): 1-5.
YIN R B, CHEN B, KONG X M. Research and application on Ethernet technologies for connected and intelligent vehicles[J]. Automotive Digest, 2019(11): 1-5(in Chinese).

[37] Robert Bosch GmbH. Bosch automotive electrics and automotive electronics[M]. Wiesbaden: Springer Fachmedien Wiesbaden, 2014: 70-81.

[38] 卢天义, 胡浩, 刘佳勇, 等. 一种电子机械制动 EMB 控制系统及汽车: CN115107722A[P]. 2022-09-27.
LU T Y, HU H, LIU J Y, et al. Electromechanical braking (EMB) control system and automobile: CN115107722A[P]. 2022-09-27(in Chinese).

[39] 张帅. 电子机械制动 EMB 控制策略研究 [D]. 西安: 长安大学, 2021.
ZHANG S. Research on control strategy of electro-mechanical braking (EMB)[D]. Xi'an: Chang'an University, 2021(in Chinese).

[40] LI Y J, SHIM T, SHIN D H, et al. Control system design for electromechanical brake system using novel clamping force model and estimator[J]. IEEE Transactions on Vehicular Technology, 2021, 70(9): 8653-8668.

[41] WEI Z, XU J, HALIM D. Clamping force control of Sensor-less electro-mechanical brake actuator[C]//Proceedings of the IEEE International Conference on Mechatronics and Automation. Piscataway: IEEE Press, 2017: 764-769.

[42] LINE C, MANZIE C, GOOD M C. Electromechanical brake modeling and control: from PI to MPC[J]. IEEE Transactions on Control Systems Technology, 2008, 16(3): 446-457.

[43] DE CASTRO R, TODESCHINI F, ARAÚJO R E, et al. Adaptive-robust friction compensation in a hybrid brake-by-wire actuator[J]. Proceedings of the Institution of Mechanical Engineers, Part I:

Journal of Systems and Control Engineering, 2014, 228(10): 769-786.

[44] LEE C F, MANZIE C. Active brake judder attenuation using an electromechanical brake-by-wire system[J]. IEEE/ASME Transactions on Mechatronics, 2016, 21(6): 2964-2976.

[45] 赵逸云, 林辉, 李兵强. 电子机械制动系统无压力传感器控制策略[J]. 北京航空航天大学学报, 2023, 49(10): 2711-2720.
ZHAO Y Y, LIN H, LI B Q. Clamping force sensorless control strategies for electromechanical brake systems[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(10): 2711-2720 (in Chinese).

[46] KWON S, LEE S, LEE J, et al. Accurate state estimation for electro-mechanical brake systems[J]. Journal of Electrical Engineering & Technology, 2019, 14(2): 889-896.

[47] XU Z, GERADA C. Enhanced force estimation for electromechanical brake actuators in transportation vehicles[J]. IEEE Transactions on Power Electronics, 2021, 36(12): 14329-14339.

[48] FU Y F, HU X H, WANG W R, et al. Simulation and experimental study of a new electromechanical brake with automatic wear adjustment function[J]. International Journal of Automotive Technology, 2020, 21(1): 227-238.

[49] 张峻峰, 张俊智, 季园, 等. 一种夹紧力估计系统、夹紧力控制方法及汽车制动系统: CN115431938A[P]. 2022-12-06.
ZHANG J F, ZAHNG J Z, JI Y, et al. Clamping force estimation system, clamping force control method and automobile braking system: CN115431938A[P]. 2022-12-06(in Chinese).

[50] 林东焕. 电子机械制动器: CN115771488A[P]. 2023-03-10.
LIN D H. Electro-hydraulic brake: CN115771488A[P]. 2023-03-10 (in Chinese).

[51] RIVA G, FORMENTIN S, SAVARESI S M. Force estimation in electro-mechanical systems: theory and experiments[C]//Proceedings of the IEEE Conference on Control Technology and Applications. Piscataway: IEEE Press, 2021: 992-997.

[52] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 乘用车制动系统技术要求及试验方法: GB 21670—2008[S]. 北京: 中国标准出版社, 2008.
General Administration of Quality Supervision, Inspectionand Quarantine, Standardization Administration of China. Technical requirements and testing methods for passenger car braking systems: GB 21670—2008[S]. Beijing: Standards Press of China, 2008(in Chinese).

[53] 高惠民. 新能源汽车制动系统解析 (三)[J]. 汽车维修与保养, 2022(5): 53-56.
GAO H M. Analysis of new energy vehicle braking system (Ⅲ)[J]. Motor China, 2022(5): 53-56(in Chinese).

[54] 陈韦纲. 线控液压制动系统制动力分配与控制策略研究[D]. 长春: 吉林大学, 2022.
CHEN W G. Research on braking force distribution and control of hydraulic brake-by-wire system[D]. Changchun: Jilin University, 2022(in Chinese).

Research progress of EMB systems key technology

ZHAO Lijin[1, 2]，YANG Shichun[1, *]，QU Jingyao[2]

(1. School of Transportation Science and Engineering，Beihang University，Beijing 100191，China；
2. China Society of Automotive Engineers，Beijing 102607，China)

Abstract: The electro-mechanical brake (EMB) is a true brake-by-wire system that completely eliminates hydraulic/pneumatic components, enabling decoupling between driver and vehicle, achieving higher response speed, and providing more precise braking force control. As an ideal braking actuator for high-level autonomous driving technology, EMB has not yet to be widely mass-produced due to challenges related to system reliability, functional safety, and cost.This paper analyzes the current technological development and typical system structures of EMB, comparing the two major technological approaches: linear self-amplifying and nonlinear force-amplifying methods. It introduces the electronic wedge brake (EWB) , as well as hybrid EMB systems, and discusses their applicability in autonomous driving environments. Second, this study focuses on the redundancy design solutions for system architecture, electronic control units (ECU), communication architecture, and failure modes. Additionally, it explores clamping force control technology based on a sensorless clamping force estimation method. Finally, the paper analyzes the challenges faced by EMB for future industrial applications and presents prospects for further research.

Keywords: electro-mechanical brake; system configuration; redundant design; clamping force control; function degradation

Received：2023-04-25；**Accepted**：2023-05-19；**Published Online**：2023-07-05 17：37

URL：link.cnki.net/urlid/11.2625.V.20230705.1529.001

Foundation items：National Key Research and Development Program of China (2022YFB2503105,2021YFB2501700)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0194

基于残差全局上下文注意和跨层特征融合的去雾网络

杨燕*，陈飞

(兰州交通大学 电子与信息工程学院，兰州 730070)

摘 要：基于深度学习的图像去雾算法通常在提取特征时使用传统的卷积层，容易造成图像的细节和边缘等信息丢失，提取特征时忽略图像的位置信息，融合特征时忽略图像原始信息，不能恢复出结构完整、清晰的高质量无雾图像。针对该问题，提出了一种基于残差全局上下文注意和跨层特征融合的去雾算法。对提出的残差全局上下文注意块串行得到残差组结构，并对网络的前2层(即浅层)进行特征提取，得到浅层丰富的上下文信息；引入坐标注意力，建立具有位置信息的注意力图，并将其应用于残差上下文特征提取，放置在网络的第3层(即深层)，提取更深层次的语义信息；在网络中间层，通过跨层融合来自不同分辨率流的特征信息，增强深浅层的信息交换，达到特征增强的目的；聚合网络得到具有丰富语义信息的特征与原始输入特征，提升复原效果。在RESIDE和Haze4K数据集上的实验结果表明：所提算法在视觉效果与客观指标上都取得了较好的效果。

关 键 词：图像去雾；深度学习；残差结构；注意力机制；特征融合

中图分类号：TP391.4

文献标志码：A **文章编号**：1001-5965(2025)04-1048-11

单幅图像去雾旨在从雾、霾等影响下的退化图像中重建出令人在视觉上感到清晰的图像。雾、霾的存在导致视觉能见系统清晰度降低，在这种天气条件下，使采集到的图像、视频通常呈现出对比度、图像细节等方面的视觉质量退化问题，严重影响到高级计算机视觉相关任务的性能，也因此备受关注。

目前，图像复原类算法是主要的去雾算法，这类算法从图像退化的物理本质出发，基于先验知识对大气散射模型进行约束，通过估计模型中的参数反推求解清晰无雾的图像。例如，文献[1]提出基于暗通道先验(dark channel prior, DCP)的去雾算法，通过对大量无雾图像进行研究，发现大多数图像至少在一个颜色通道有最小的像素值，但对含天空区域的图像处理效果不理想，使得复原结果颜色失真；文献[2]提出基于颜色衰减先验的去雾算法，利用图像亮度和饱和度之间的关系，建立与景深相关的线性关系，但易使透射率估计偏大，出现去雾不彻底的现象；文献[3]提出利用基于最小通道区间估计与透射率自适应约束模型的去雾算法，有效解决了透射率估计不足和偏色等问题。基于先验的图像去雾算法均是基于传统大气散射模型进行复原，虽然能取得一定程度的效果，但由于先验具有统计特性，估计参数会存在误差，导致去雾效果不理想。

近年来，随着深度学习的快速发展，一些学者将神经网络模型应用到了图像去雾领域。例如，文献[4]提出基于卷积神经网络(convolutional neural

收稿日期：2023-04-23；**录用日期**：2023-08-04；**网络出版时间**：2023-09-21 18:04

网络出版地址：link.cnki.net/urlid/11.2625.V.20230921.1219.002

基金项目：国家自然科学基金(61561030,62063014)；甘肃省高等学校产业支撑计划(2021CYZC-04)；兰州交通大学研究生教改项目(JG201928)；甘肃省教育厅优秀研究生“创新之星”项目(2023CXZX-547)

* **通信作者**. E-mail：yangyantd@mail.lzjtu.cn

引用格式：杨燕，陈飞. 基于残差全局上下文注意和跨层特征融合的去雾网络[J]. 北京航空航天大学学报，2025，51(4)：1048-1058.

YANG Y，CHEN F. Dehazing network based on residual global contextual attention and cross-layer feature fusion[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1048-1058 (in Chinese).

network, CNN) 的去雾算法 DehazeNet, 通过学习降质雾图中的透射率进行去雾; 文献 [5] 通过改进大气散射模型, 将透射率和大气光合并进行训练, 提出一体化去雾网络 AOD-Net; 文献 [6] 建立了多尺度密集特征融合的网络模型, 利用增强策略和反投影技术来加强特征融合; 文献 [7] 提出了一种端到端的特征融合注意网络 (feature fusion attention network, FFA-Net), 学习重要特征的权重, 对特征自适应融合, 增强网络映射能力, 进行去雾; 文献 [8] 提出了一种处理高分辨率图像的网络模型, 通过仿射双边网格引导学习高清图像, 生成高质量图像特征, 以提升去雾效果; 文献 [9] 提出一种对比正则化的去雾网络, 对比学习模糊图像和清晰图像的信息, 分别从正负样本中获益, 有效生成无雾图像; 文献 [10] 提出了基于自引导的图像去雾模型 SGID, 利用输入的雾图引导帮助去雾, 通过渐进式特征融合策略, 在迭代融合过程中逐步去除残留雾气的影响。

虽然 CNN 在图像去雾方面取得了重大进展, 但当前基于深度学习的去雾算法仍然存在以下问题:

1) 更少关注网络组成结构。一些去雾算法通常致力于通过增加卷积层的深度或宽度来改善模型性能, 缺失对 CNN 结构学习能力的探索[7,11]。

2) 忽视图像全局和特征位置信息。一些去雾网络在局部特征上赋予更多的权重, 而忽视了全局上下文信息、特征位置信息, 导致图像恢复细节缺失和去雾不彻底[6,12]。

3) 图像原始信息不能有效利用。在融合特征时忽略图像原始信息 (如颜色), 缺失了各层之间的信息传递, 不能很好地利用深浅层的特征信息, 导致不能恢复出结构完整、清晰的高质量无雾图像[10,13]。

针对上述分析, 本文提出了一种新的融合全局上下文信息及图像深层特征位置信息的网络结构, 整体目标是: 在整个网络中保持图像的原始特征信息, 并从低分辨率表示中接收补充的上下文信息, 以及深层特征的空间位置信息, 增强深浅层信息的流动, 保证得到清晰生动的图像, 同时避免因全局信息、特征位置信息缺失导致的去雾不彻底等问题。通过主观和客观 2 个方面与其他主流算法进行对比, 实验结果表明, 本文模型取得了更好的去雾效果。

综上, 本文主要贡献如下:

1) 提出了基于全局上下文注意 (global contextual attention, GCA) 的浅层残差特征提取模块, 不仅能够捕获丰富的细粒度特征信息, 还能够维持通道间的依赖关系, 有效实现初步去雾。

2) 设计了基于坐标注意力 (coordinate attention, CA) 的深层残差特征提取模块, 使网络深入挖掘并去除影响图像退化的因素, 同时保留图像所需的精细的空间细节信息, 进一步提升去雾效果。

3) 构建了基于混合注意力的跨层特征融合方式, 充分利用不同层的重要特征, 增强网络信息利用及表达能力, 保证复原图像的质量。

1　本文算法

本文算法的网络结构如图 1 所示。为实现特征的深浅层融合, 采用编解码器结构, 设计多级去雾网络结构增强信息流动, 有效实现特征保留与融合。在网络的浅层部分, 设计了基于残差全局上下文位置的残差特征提取模块, 通过串行 2 个浅层特征提取模块, 并使用残差连接方式避免上下采样的细节信息缺失, 同时提供了更精确和丰富的上下文特征表示; 在中间部分, 设计跨层融合方式, 有效融合来自不同分辨率流的特征, 用于增强特征信息; 在网络的深层部分, 设计了基于残差全局上下文位置的残差特征提取模块, 在删除影响图像退化内容的同时, 保留所需的精细空间细节信息 (如边缘和纹理)。总体流程为: 网络上半部分, 通过特征提取模块和下采样提取特征, 中间部分, 将第 1、4 模块和第 2、3 模块进行特征融合, 通过下半部分逐步上采样进行图像重建。

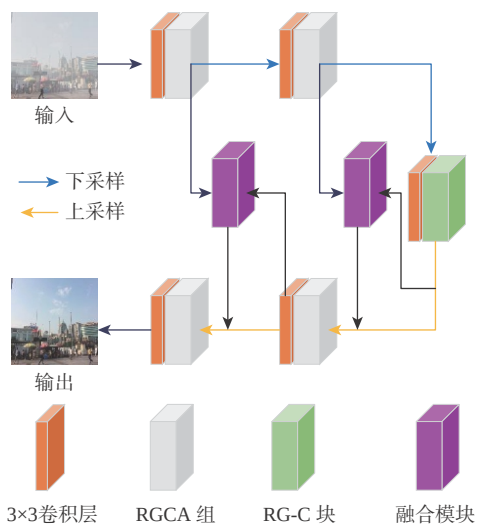


图 1　网络总体结构

Fig. 1　Overall network structure

1.1　浅层残差特征提取模块

浅层特征提取部分采用 2 层编码器结构。经过卷积层提取初步特征, 浅层特征主要为细粒度的信息, 与输入较接近, 蕴含更多的像素点信息, 通常包括图像的色彩、纹理等细节信息。网络浅层的感

受野较小，重叠部分也较小，能够保证网络捕获更多细节特征。同时，为挖掘更多信息，提出了残差全局上下文注意组 (residual global contextual attention group, RGCA group)，如图 2 所示，其能够通过 2 个残差全局上下文注意块 (residual global contextual attention block, RGCA block) 及残差连接保持原始分辨率生成空间精确的输出，提取浅层丰富的上下文信息。对第 2 层进行 2 倍下采样操作，对不同分辨率的特征映射进行特征提取操作。浅层特征通过来自低分辨率流的上下文信息，传送到深层次特征提取部分以巩固高分辨率特征。

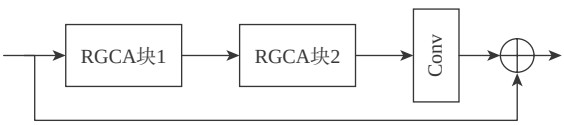


图 2 残差全局上下文注意组

Fig. 2 Residual global contextual attention group

1.1.1 残差全局上下文注意块

RGCA 块是基于全局上下文注意的残差特征提取块，如图 3 所示，能够抑制用处较小的特征，只允许信息更丰富的特征通过。首先，输入一幅图像，经过普通卷积得到初步特征 $\boldsymbol{F}_1$；其次，通过 2 个卷积激活层得到进一步的特征映射 $\boldsymbol{F}_2$；然后，利用全局上下文注意得到通道间的依赖关系 $\boldsymbol{F}_3$，同时通过逐元素相加将这些依赖关系融合到原始特征 $\boldsymbol{F}_2$ 上，得到具有丰富的上下文信息的特征 $\boldsymbol{F}_4$；最后，与 $\boldsymbol{F}_1$ 相加得到最终特征 $\boldsymbol{F}_5$。整体流程可表示为

$$\boldsymbol{F}_5 = \boldsymbol{F}_1 + N(\mathrm{GCA}(\boldsymbol{F}_2)) \tag{1}$$

式中：$\boldsymbol{F}_2 \in \mathbf{R}^{H\times W\times C}$ 表示通过 2 个卷积层得到的特征映射，H 为特征图的高，W 为特征图的宽，C 为特征图的通道数；GCA(·)表示全局上下文注意；N 表示 1×1 卷积层；$\boldsymbol{F}_5$ 表示最终得到的具有全局上下文信息的特征。

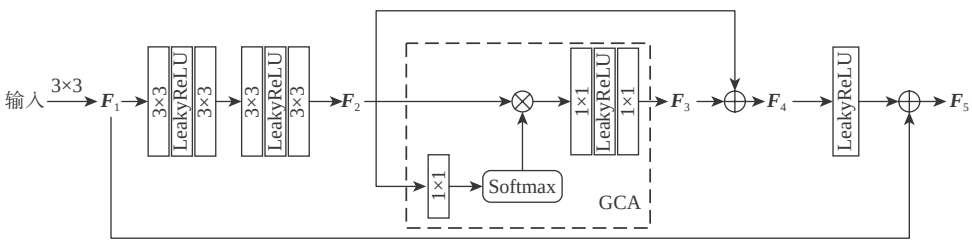


图 3 残差全局上下文注意块

Fig. 3 Residual global contextual attention block

1.1.2 全局上下文注意

图 4 为全局上下文注意[14]。图中：r 为通道缩减因子。先通过上下文建模部分 W_k 得到全局特征描述，再通过卷积层 N_{v1} 和 N_{v2} 捕获通道间依赖，最终进行融合得到上下文特征信息。流程可表示为

$$\boldsymbol{z}_i = \boldsymbol{x}_i + N_{v2}(\alpha(N_{v1}(\gamma(N_k(\boldsymbol{x}_i))\otimes\boldsymbol{x}_i))) \tag{2}$$

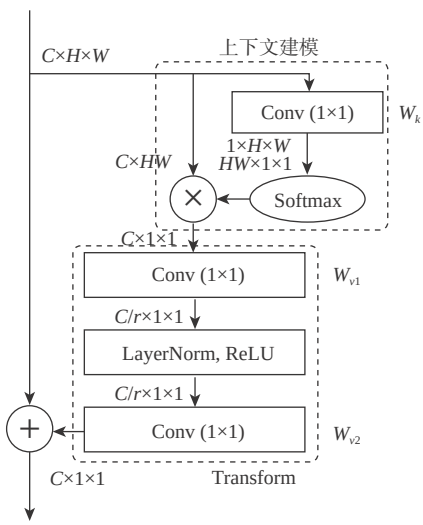


图 4 全局上下文注意示意图

Fig. 4 Global contextual attention

式中：$\boldsymbol{z}_i$ 为得到的残差上下文特征；$\boldsymbol{x}_i$ 为初始特征；N_k、N_{v1}、N_{v2} 分别为具有不同参数的卷积层；α 为 LeakyReLU 激活函数；γ 为 Softmax 激活函数；“⊗”表示对应元素相乘。

1.2 深层残差特征提取模块

深层特征提取部分为浅层特征 2 次下采样后进行的特征提取操作。网络深层提取的特征与输出较接近，主要包含粗粒度的特征信息，其中蕴含更抽象的信息，即语义信息。随着感受野变大，感受野之间重叠部分增加，图像内部信息进行压缩，得到的是图像整体性信息。同时，为防止跨通道信息缺失和特征空间位置信息缺失，提出一种能够同时提取图像上下文特征信息和特征位置信息的残差结构，放置在网络的深层部分，用于捕获深层特征方向感知和位置感知的信息，还能捕捉到沿着一个空间方向上的长距离依赖关系，总体通过强化信息表示的方法来增强特征表达。

1.2.1 残差全局上下文位置块

残差全局上下文位置块 (residual global context location block, RG-C block) 是融合全局上下文注意和坐标注意的残差特征提取块，如图 5 所示。首

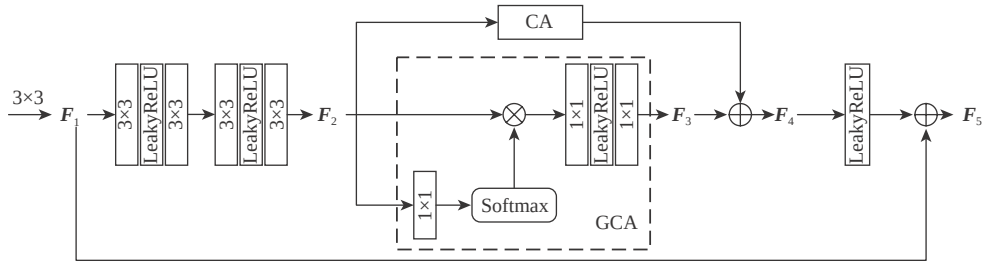


图 5　残差全局上下文位置块

Fig. 5　Residual global contextual location block

先, 输入一幅图像, 经过普通卷积得到初步特征 $\boldsymbol{F}_1$; 其次, 通过 2 个卷积激活层得到进一步的特征映射 $\boldsymbol{F}_2$; 然后, 利用全局上下文注意得到通道间的依赖关系 $\boldsymbol{F}_3$, 同时, 与经过坐标注意得到的位置信息逐元素相加, 将这些依赖关系融合到原始特征 $\boldsymbol{F}_2$ 上, 得到具有位置信息和丰富上下文信息的特征 $\boldsymbol{F}_4$; 最后, 与 $\boldsymbol{F}_1$ 相加得到最终特征 $\boldsymbol{F}_5$。整体流程可表示为

$$\boldsymbol{F}_{\text{RG-C}} = \boldsymbol{F}_1 + (N(\text{GCA}(\boldsymbol{F}_2)) + \text{CA}(\boldsymbol{F}_2)) \tag{3}$$

式中: CA(·)表示坐标注意力; $\boldsymbol{F}_{\text{RG-C}}$为最终得到的具有全局上下文信息和位置信息的特征。

1.2.2　坐标注意力

坐标注意力[15] 是高效的注意力机制, 如图 6 所示。通过将位置信息嵌入到通道注意力, 使网络获取更大区域的信息, 避免引入大的开销。同时, 为避免 2D 全局池化引入导致位置信息损失, 坐标注意力分解通道注意为 2 个并行的 1D 特征编码, 高效地整合空间坐标信息。首先, 使用尺寸为 (H,1) 和 (1,W) 的池化核分别沿着水平坐标和垂直坐标对每个通道进行编码, 沿 2 个空间方向聚合特征, 得到特征图, 保存特征的位置信息, 如式 (4) 所示。然后, 通过卷积变换操作, 如式 (5)所示。最后, 坐标注意力的输出如式 (6) 所示, 其能捕获跨通道信息, 还能捕获位置信息, 从而有助于模型更准确地定位和识别感兴趣的特征。

$$\begin{cases} z_c^h(h) = \dfrac{1}{W}\displaystyle\sum_{0\leqslant i\leqslant W} x_c(h,i) \\ z_c^w(w) = \dfrac{1}{H}\displaystyle\sum_{0\leqslant j\leqslant H} x_c(j,w) \end{cases} \tag{4}$$

$$\begin{cases} f = \delta\left(N\left(\left[z^h, z^w\right]\right)\right) \\ g^h = \sigma\left(N_h\left(f^h\right)\right) \\ g^w = \sigma\left(N_w\left(f^w\right)\right) \end{cases} \tag{5}$$

$$y_c(i,j) = x_c(i,j)\, g_c^h(i)\, g_c^w(j) \tag{6}$$

式中: z_c^h和z_c^w分别表示高度为 h 和宽度为 w 的第 c 通道的输出; [·,·]表示沿空间维度的拼接运算; δ为非线性激活函数; f为在水平方向和垂直方向上的中间特征; σ为 Sigmoid 激活函数; N_h和N_w表示卷积变换; g_c^h和g_c^w为注意力权重; y_c为注意力模块的输出。

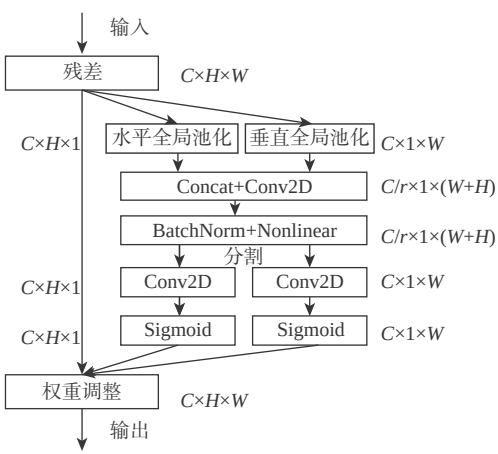


图 6　坐标注意力示意图

Fig. 6　Coordinate attention

1.3　跨层特征融合

本文总体结构采用了类似于编码器-解码器的网络架构。编码器-解码器结构会在某些方面受到限制, 如在编码器的下采样过程中缺少空间信息, 并且在非相邻层级的特征之间缺少足够的连接[8]。因此, 在中间层引入跨层特征融合模块, 将编码器中部分特征与解码器中部分特征融合, 避免网络忽视低级特征。低级特征 (如边缘和轮廓等) 对于恢复清晰图像同样具有不可忽视的作用, 防止在经过中间层后逐渐失去对细节信息的关注。

跨层特征融合在增强信息从浅层向深层流动的同时, 能够保证原始特征的保留。基于上述讨论, 本文提出一种基于 CBAM[16] 和像素注意力 (pixel attention, PA)[7] 融合方式深浅层特征融合 (low and high features fusion, LHFF), 如图 7 所示, 其中, CBAM 和 PA 分别如图 8 和图 9 所示。利用 CBAM

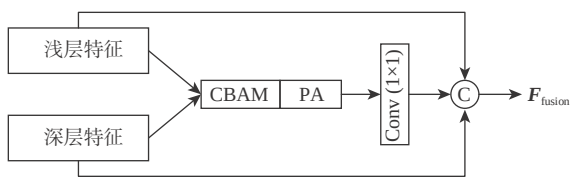


图 7　跨层特征融合

Fig. 7　Cross layers feature fusion

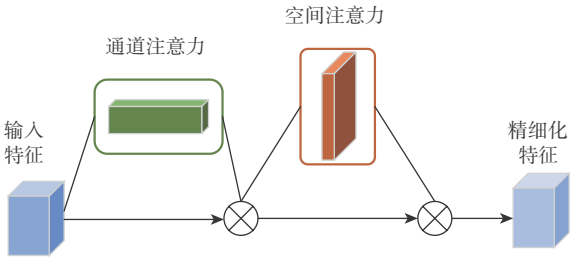


图8 CBAM示意图

Fig. 8 Convolutional block attention module

能够序列化地在通道和空间2个维度上产生注意力特征图信息，2种特征图信息通过PA进行相乘，得到自适应特征修正，再与原输入的深浅层特征图进行拼接，产生最终的特征图。流程可表示为

$$\boldsymbol{F}_{\text{fusion}}=\boldsymbol{F}_{\text{low}}+\boldsymbol{F}_{\text{high}}+\text{PA}(\text{CBAM}(\boldsymbol{F}_{\text{mid}})) \tag{7}$$

式中：$\boldsymbol{F}_{\text{fusion}}$为深层特征和浅层特征融合后特征；$\boldsymbol{F}_{\text{low}}$为浅层特征；$\boldsymbol{F}_{\text{high}}$为深层特征；$\boldsymbol{F}_{\text{mid}}$为中间融合特征。

1.4 图像复原

图像复原通过对深度特征应用卷积层，得到残差图像$\boldsymbol{R}\in\mathbf{R}^{H\times W\times 3}$。最终得到的复原图像表达式为

$$\hat{\boldsymbol{I}}=\boldsymbol{I}+\boldsymbol{R} \tag{8}$$

式中：$\hat{\boldsymbol{I}}$为最终复原图像；$\boldsymbol{I}$为给定的输入图像；$\boldsymbol{R}$为网络输出的残差图像。

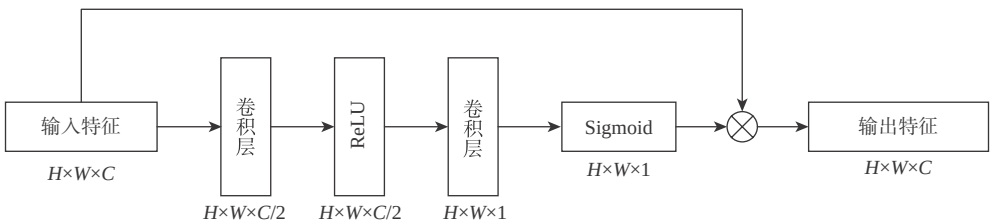


图9 像素注意力示意图

Fig. 9 Pixel attention

1.5 损失函数

1.5.1 平滑L_1损失

平滑L_1损失[17]能够直观计算出网络输入和输出之间的差距，通过逐像素比较输出与真实图像之间的像素差值，避免了L_1损失在零点处不可导而不收敛的情况。函数可表示为

$$L_{\text{SL}_1}=\frac{1}{X}\sum_{i=1}^{X}G(O_i-J_i) \tag{9}$$

$$G(x)=\begin{cases}0.5x^2 & |x|<1\\ |x|-0.5 & \text{其他}\end{cases} \tag{10}$$

式中：X为总像素数；i为第i个像素；$\boldsymbol{O}$为输出图像；$\boldsymbol{J}$为真实图像；x为输出图像与真实图像之间的差距。

1.5.2 结构相似性损失

结构相似性(structural similarity, SSIM)损失[18]用于评价图像质量，捕获图像结构信息，反映场景中物体的结构信息，评估2幅图像是否在结构上相似。结构相似性函数可表示为

$$\text{SSIM}(\boldsymbol{O},\boldsymbol{J})=\frac{(2\mu_O\mu_J+c_1)(2\sigma_{OJ}+c_2)}{(\mu_J^2+\mu_O^2+c_1)(\sigma_J^2+\sigma_O^2+c_2)} \tag{11}$$

式中：μ_O、μ_J分别为输出图像、真实图像的均值；σ_{OJ}为输出图像和真实图像的协方差；σ_J^2、σ_O^2分别为输出图像、真实图像的方差；c_1、c_2表示防止分母为0的常数。

结构相似性损失可表示为

$$L_{\text{SSIM}}=1-\text{SSIM}(\boldsymbol{O},\boldsymbol{J}) \tag{12}$$

1.5.3 总损失函数

将平滑L_1损失与结构相似性损失结合，可有效避免单一损失函数带来的误差。总损失函数可表示为

$$L=L_{SL1}+L_{\text{SSIM}} \tag{13}$$

2 实验结果及分析

2.1 数据集及训练设置

本文算法在RESIDE[19]和Haze4K[20]数据集上训练网络，训练集包括RESIDE数据集室内真实场景ITS(1 399张真实图像和13 990张雾图)和室外真实场景OTS(选取1 000张真实图像和10 000张雾图)；测试集为SOTS，包括500张室内测试图像和500张室外测试图像。Haze4K数据集包含3 000个合成训练图像和1 000张合成测试图像。

算法基于Pytorch框架实现，在Ubuntu环境下采用NVIDIA V100 GPU进行训练，使用Adam优化器进行优化，参数为默认值。初始学习率设置为10^{-4}，批处理大小设置为12，采用余弦退火策略[18]，将学习率从初始值调整到10^{-6}。为训练模型，从原始图像中随机裁剪尺寸为256×256的图像作为输入，采用2种数据增强技术：90°或180°或270°旋转、垂直或水平翻转。

2.2 去雾质量定性分析

为验证本文算法的有效性，本文从主观效果上与先进算法进行实验对比，包括DCP去雾算法[1]、

DehazeNet 去雾算法[4]、AOD-Net 去雾算法[5]、门控上下文聚合网络 (gated context aggregation network, GCANet) 算法[13]、FFA-Net[7]、UHD 去雾算法[8]、SGID 去雾算法[10]。此外, 另选取 3 幅真实雾图与不同算法进行实验对比, 进一步验证了本文算法在真实场景上的有效性。

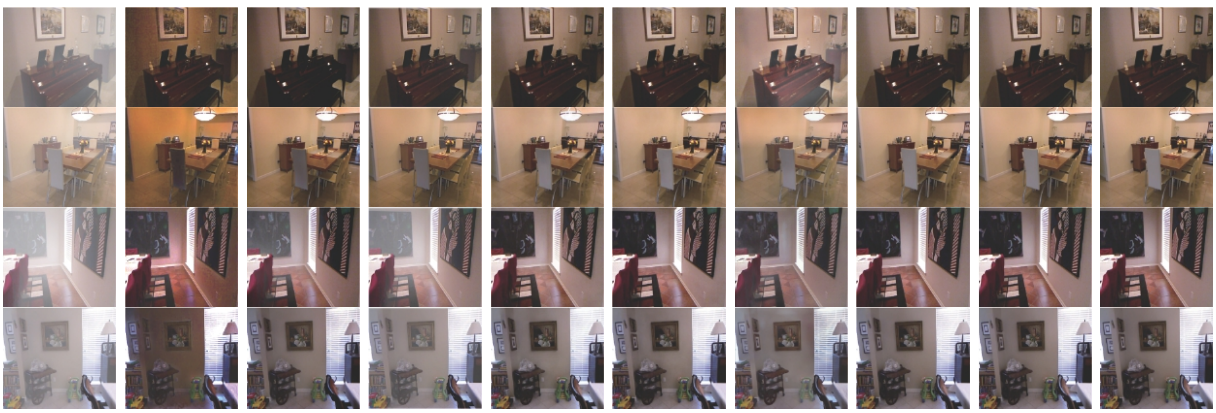
图 10 为不同算法在合成雾图数据集 SOTS 室内测试集上选取的雾图视觉效果对比。可以看出, DCP 的复原效果较彻底, 但由于暗通道先验存在局限性, 与其他算法相比复原效果较暗; DehazeNet、AOD-Net 和 UHD 出现了不同程度的去雾不彻底现象, 如图 10(c) 的第 3 幅图, 图 10(d) 的第 3、4 幅图, 图 10(g) 的第 1 幅图; 与真实图像 GT 相比, GCANet、FFA-Net 和 SGID 去雾结果与本文算法的结果较接近, 但 GCANet 和 FFA-Net 在边缘细节上恢复效果略差。

图 11 为不同算法在 Haze4K 数据集测试集上选取的雾图视觉效果对比。可以看出, DCP、DehazeNet、GCANet 和 UHD 存在不同程度的颜色失真、光晕伪影现象, 如图 11 的第 1 幅图较为明显; AOD-Net、FFA-Net 和 SGID 存在不同程度的去雾不彻底现象, 如图 11 的第 4 幅图较为明显; 本文算法不存在上述问题, 恢复效果较好。

图 12 为不同算法在合成雾图数据集 SOTS 室外测试集上选取的雾图视觉效果对比。DCP 存在严重的颜色失真现象; AOD-Net 和 UHD 也存在一定的颜色偏移问题, 如图 12(d) 的第 1 幅图和图 12(g) 的第 2 幅图; 相比真实图像 GT, 其他算法与本文效果接近, 去雾较为彻底, 色彩自然。

图 13 为真实世界的雾图视觉效果对比。可以看出, 其他算法在处理结果上保持雾霾, 或产生颜色偏差和伪影, 相反, 本文算法可以得到视觉效果更好的去雾结果。

图 14 为 Haze4K 测试集(第 1 行和第 2 行图像)和 SOTS 室外测试集(第 3 行和第 4 行图像)上的视觉主观局部放大效果比较。可以看出, DCP、DehazeNet、GCANet 和 SGID 细节信息丢失, 出现了明显光晕伪影, 同时存在不同程度的颜色失真现象, 如第 1 幅图红框区域树枝边缘与天空交界局部放大处, 均存在不同程度的光晕; 第 5 幅图红框区域天空部



(a) 雾图　(b) DCP[1]　(c) DehazeNet[4]　(d) AOD-Net[5]　(e) GCANet[13]　(f) FFA-Net[7]　(g) UHD[8]　(h) SGID[10]　(i) 本文　(j) GT

图 10　不同算法在 SOTS 室内测试集上的复原结果

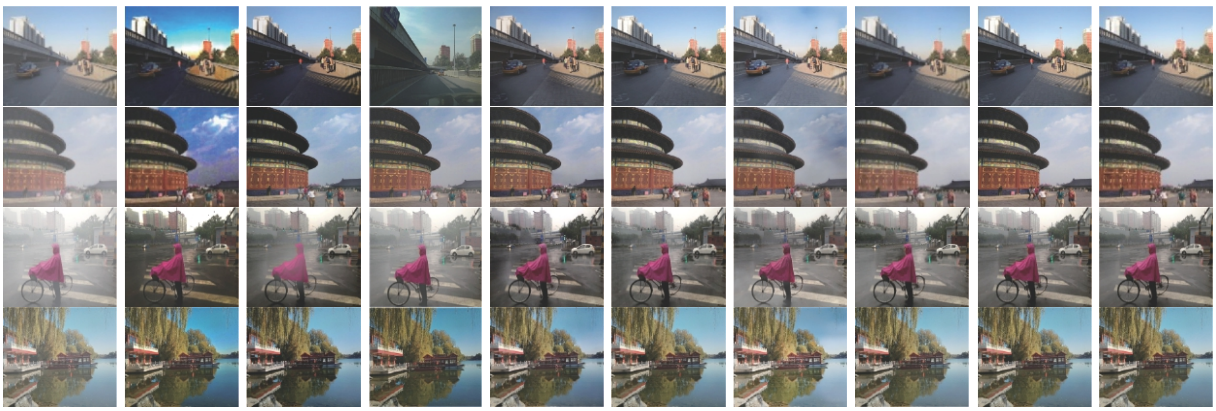
Fig. 10　Recovery results of different algorithms on SOTS-indoor test set



(a) 雾图　(b) DCP[1]　(c) DehazeNet[4]　(d) AOD-Net[5]　(e) GCANet[13]　(f) FFA-Net[7]　(g) UHD[8]　(h) SGID[10]　(i) 本文　(j) GT

图 11　不同算法在 Haze4K 测试集上的复原结果

Fig. 11　Recovery results of different algorithms on Haze4K test set



(a) 雾图 (b) DCP[1] (c) DehazeNet[4] (d) AOD-Net[5] (e) GCANet[13] (f) FFA-Net[7] (g) UHD[8] (h) SGID[10] (i) 本文 (j) GT

图 12 不同算法在 SOTS 室外测试集上的复原结果

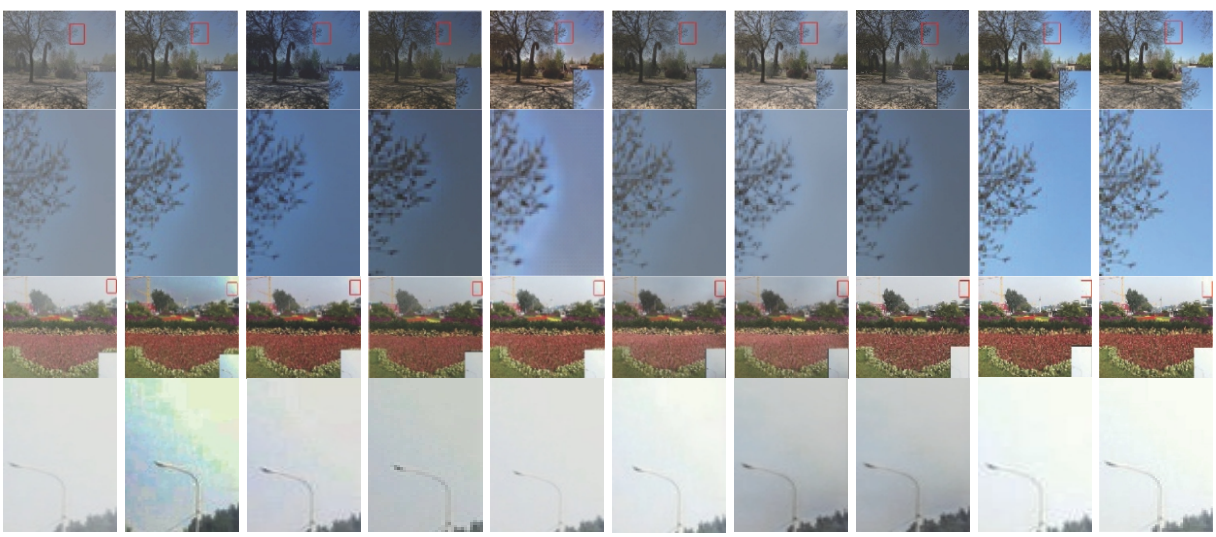
Fig. 12 Recovery results of different algorithms on SOTS-outdoor test set



(a) 雾图 (b) DCP[1] (c) DehazeNet[4] (d) AOD-Net[5] (e) GCANet[13] (f) FFA-Net[7] (g) UHD[8] (h) SGID[10] (i) 本文

图 13 不同算法下真实图像的复原结果

Fig. 13 Recovery results of real-world images by different algorithms



(a) 雾图 (b) DCP[1] (c) DehazeNet[4] (d) AOD-Net[5] (e) GCANet[13] (f) FFA-Net[7] (g) UHD[8] (h) SGID[10] (i) 本文 (j) GT

图 14 不同算法下数据集上图像的复原结果及局部放大展示

Fig. 14 Recovery results and partial amplification display of images in datasets by different algorithms

分局部放大处，天空部分复原较偏深蓝色。

图 15 为真实图像上的视觉主观局部放大效果比较，如第 2 幅图的天空处和第 3 幅图红框区域红墙局部放大部分。

综上，本文算法的处理结果不存在上述问题，对颜色失真、细节缺失、去雾不彻底等方面有一定的改善，在色彩细节方面与真实图像最为接近，取得了最佳的视觉体验。

2.3 去雾质量定量分析

本文在 3 个测试集上从定量的客观角度进行评价。采用峰值信噪比 (peak signal-to-noise ratio, PSNR) 和 SSIM 来测试不同算法的图像复原效果[21]，对比如表 1 所示（表中数据为所有测试结果均值）。



(a) 雾图　(b) DCP[1]　(c) DehazeNet[4]　(d) AOD-Net[5]　(e) GCANet[13]　(f) FFA-Net[7]　(g) UHD[8]　(h) SGID[10]　(i) 本文

图 15　不同算法下真实图像的复原结果及局部放大展示

Fig. 15　Recovery results and partial amplification display of real-world images by different algorithms

表 1　不同算法在不同测试集上的评测指标

Table 1　Evaluation metrics for different algorithms on different test sets

算法	PSNR/dB			SSIM		
	SOTS室内	Haze4K	SOTS室外	SOTS室内	Haze4K	SOTS室外
DCP[1]	16.52	14.21	19.03	0.818 6	0.724 6	0.821 6
DehazeNet[4]	20.13	19.24	22.16	0.845 7	0.842 8	0.823 3
AOD-Net[5]	19.08	17.18	19.15	0.850 3	0.834 9	0.845 8
GCANet[13]	30.20	23.57	30.14	0.980 6	0.949 2	0.952 1
FFA-Net[7]	33.29	24.97	30.79	0.976 4	0.951 4	0.934 2
UHD[8]	21.59	13.28	25.80	0.867 9	0.745 7	0.963 9
SGID[10]	33.45	21.25	29.98	0.985 6	0.929 1	0.978 0
本文	34.14	25.53	31.29	0.983 0	0.962 4	0.980 8

通过实验数据可以看出, 本文算法的 PSNR 值在所有的比较算法中最高, 而 SSIM 值仅比最新的 SGID 低 0.002 6。结果进一步表明, 本文算法具有一定的可行性及优越性。

为更全面地与其他算法对比, 体现各算法的复杂度, 表 2 展示了各算法的速度数据。可以看出, 运行时间最短为 AOD-Net, 本文算法与排名前几位算法运行时间差距相对较小。传统算法 DCP 运行时间最长, 深度学习算法中 AOD-Net、DehazeNet、UHD、GCANet 虽然运行时间较短, 但从主观实验对比中可以看出, 复原结果均存在一定的问题。综合定性定量分析来看, 本文能够平衡算法效率和图像复原质量之间的关系。

表 2　平均运行时间对比

Table 2　Average runtime comparison

算法	平均运行时间/s
DCP[1]	1.907
DehazeNet[4]	0.812
AOD-Net[5]	0.574
GCANet[13]	0.914
FFA-Net[7]	1.578
UHD[8]	0.896
SGID[10]	1.102
本文	1.079

2.4　消融实验

为证明本文算法中每个成分的有效性, 对使用的 3 个模块 (RGCA、RG-C、LHFF) 进行消融分析。首先, 构建 Base 网络作为基线, 使用基本残差块 RB[22] 作为特征提取, 同时使用基于通道注意力 SE[23] 的融合作为跨层融合模块构建基线模型。然后, 将不同的模块替换到 Base 网络中, 构造了 4 个不同的变体: ①Base+RGCA: 将 Base 中 RB 替换为 RGCA 组。②Base+RGCA+RG-C: 用 RG-C 块代替 Base+RGCA 中深层的 RGCA 组。③Base+RG-C+LHFF: 将 Base+RGCA+RG-C 中基于 SE 的跨层特征融合模块替换为 LHFF 模块, 同时替换浅层 RB 为 RGCA 组。④Base+RGCA+LHFF:将 Base+RG-C+LHFF 中的深层 RB 替换为 RG-C 块, 同时替换浅层 RB 为 RGCA 组。本文用所有提出的模块替换 Base 中基本模块。这些模型在 ITS 数据集上进行训练, 并在 SOTS 室内数据集上进行评估。表 3 总结了模型的性能指标, 可以证明本文提出模块的有效性。

为进一步表明本文所提模块的有效性, 对逐步添加每个模块的去雾结果进行主观比较, 如图 16 所示。Base 模块: 仅使用 RB 进行特征提取, 缺乏提取雾浓度信息和处理特征的能力, 如图 16(b) 所示, 基本没有去除雾霾。Base+RGCA 模块: 添加了

表3 消融实验

Table 3 Ablation experiment

模块	PSNR/dB	SSIM
Base	27.25	0.956 2
Base+RGCA	28.89	0.961 9
Base+RGCA+RG-C	28.85	0.976 1
Base+RG-C+LHFF	31.17	0.979 2
Base+RGCA+LHFF	30.61	0.978 0
本文	33.08	0.981 4

RGCA组，具有很好的雾浓度感知提取和特征处理能力，感知雾浓度上下文信息，因此，能够较好地去除雾霾，如图16(c)所示，大部分雾霾已经去除，但在细节方面处理得不够好。Base+RGCA+RG-C模块：增加了RG-C块，用于感知深层雾霾特征位置并提取，利用全局上下文注意和坐标注意能够进一步提升复原的细节，如图16(d)所示，在细节恢复上有着明显提升。本文模型添加了LHFF模块，用于不同层之间的信息交流，跨层融合多尺度特征，增强了前3组实验在雾霾去除、细节保留和原始特征信息利用(如颜色)的能力，如图16(e)所示，能够复原出与真实图像更接近的结果。

综上，消融实验表明，RGCA组能够提升图像特征提取能力，RG-C块能够提升图像细节恢复，LHFF模块能够增强图像复原的质量，每个模块都对提升图像复原结果的质量起着促进的作用，使网络达到最佳的去雾效果。



(a) 雾图 (b) Base模块 (c) Base+RGCA模块 (d) Base+RGCA+RG-C模块 (e) 本文 (f) GT

图16 消融实验主观对比

Fig. 16 Subjective comparison of ablation experiment

3 结 论

1）本文针对深度学习去雾算法提取特征忽视图像细节信息和图像原始特征信息的问题，提出了一种基于残差全局上下文注意和跨层特征融合的去雾算法。

2）提出浅层残差特征提取模块，获得了图像中丰富的上下文信息，嵌入坐标注意力提取了网络更深层次的语义信息，能够达到良好的特征提取效果。

3）提出跨层特征融合模块增强了特征信息的表达能力，使图像特征信息更加完备。

4）通过大量实验证明了本文算法的有效性及在图像去雾中的效果。

本文在算法效率上仍有提升空间，与先进算法相比提升差距较小，本文网络模块及结构还可继续优化，后续研究将更加重视算法的效率和实时性，应用于对复杂环境的工程研究中。

参考文献（References）

[1] 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.

[2] ZHU Q S, MAI J M, SHAO L. A fast single image haze removal algorithm using color attenuation prior[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3522-3533.

[3] 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) .

[4] CAI B L, XU X M, JIA K, et al. DehazeNet: an end-to-end system for single image haze removal[J]. IEEE Transactions on Image Processing, 2016, 25(11): 5187-5198.

[5] LI B Y, PENG X L, WANG Z Y, et al. AOD-Net: all-in-one dehazing network[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 4780-4788.

[6] DONG H, PAN J S, XIANG L, et al. Multi-scale boosted dehazing network with dense feature fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 2154-2164.

[7] QIN X, WANG Z L, BAI Y C, et al. FFA-Net: feature fusion attention network for single image dehazing[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 11908-11915.

[8] 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.

[9] WU H Y, QU Y Y, LIN S H, et al. Contrastive learning for compact single image dehazing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 10546-10555.

[10] BAI H R, PAN J S, XIANG X G, et al. Self-guided image dehazing using progressive feature fusion[J]. IEEE Transactions on Image Processing, 2022, 31: 1217-1229.

[11] LIU X H, MA Y R, SHI Z H, et al. GridDehazeNet: attention-based multi-scale network for image dehazing[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 7313-7322.

[12] LI R D, PAN J S, LI Z C, et al. Single image dehazing via conditional generative adversarial network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8202-8211.

[13] CHEN D D, HE M M, FAN Q N, et al. Gated context aggregation network for image dehazing and deraining[C]//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2019: 1375-1383.

[14] CAO Y, XU J R, LIN S, et al. GCNet: non-local networks meet squeeze-excitation networks and beyond[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop. Piscataway: IEEE Press, 2019: 1971-1980.

[15] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 13708-13717.

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

[17] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 1440-1448.

[18] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[19] 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.

[20] LIU Y, ZHU L, PEI S D, et al. From synthetic to real: image dehazing collaborating with unlabeled real data[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 50-58.

[21] HE T, ZHANG Z, ZHANG H, et al. Bag of tricks for image classification with convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 558-567.

[22] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

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

Dehazing network based on residual global contextual attention and cross-layer feature fusion

YANG Yan*, CHEN Fei

(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: Current deep learning-based image dehazing algorithms usually use traditional convolutional layers when extracting features, which easily cause loss of information, such as details and edges of the image, ignore the location information of the image in feature extraction, and neglect the original information of the image in feature fusion, and they thus fail to recover a high-quality dehazing image with complete and clear structure. To address this problem, a dehazing algorithm based on residual contextual attention and cross-layer feature fusion was proposed. Firstly, the residual group structure was obtained by serializing the proposed residual contextual blocks, and feature extraction was performed on the first two layers of the network, i.e. the shallow layers, to obtain rich contextual information in the shallow layers; secondly, coordinate attention was introduced to build an attention graph with location information and apply it to the residual contextual feature extraction, which was placed in the third layer of the network, i.e. the deep layer, to extract deeper semantic information; then, by fusing feature information from different resolution streams across layers in the middle layer of the network, the information exchange between the deep and shallow layers was enhanced to achieve feature enhancement; finally, the semantic information-rich features obtained from the network were combined with the original input, thus enhancing the recovery effect. Experimental results on the RESIDE dataset and the Haze4K dataset show that the proposed algorithm achieves better results in terms of visual effects and objective metrics.

Keywords: image dehazing; deep learning; residual structure; attention mechanism; feature fusion

Received: 2023-04-23; **Accepted**: 2023-08-04; **Published Online**: 2023-09-21 18:04

URL: link.cnki.net/urlid/11.2625.V.20230921.1219.002

Foundation items: National Natural Science Foundation of China (61561030,62063014); Industrial Support Program for Higher Education Institutions in Gansu Province (2021CYZC-04); Lanzhou Jiaotong University Graduate Teaching and Research Project (JG201928); Outstanding Graduate "Innovation Star" Project of Gansu Provincial Department of Education (2023CXZX-547)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0178

自适应不完备多视角聚类

陈梅*，马学艳，张弛，张锦宏，钱罗雄
(兰州交通大学 电子与信息工程学院，兰州 730070)

摘 要：高质量的完备初始图能够有效提高不完备多视角聚类的性能，缺失值填充不恰当会导致初始图丢失数据潜在的结构，同时，各视角的仿射图融合不完全会造成学习到的一致表征缺失视角间的互补信息。为此，提出自适应不完备多视角聚类(AIM)模型。在初始图构建中，AIM模型采用有效视角的相似度均值来填充对应位置的缺失值，以获取数据更加完整的潜在结构，同时引入稀疏约束来提高模型对噪声的鲁棒性；在图优化过程中，引入低秩约束捕获数据的全局结构，通过谱约束增强类内数据间的紧密性，使仿射图具有更清晰的块对角结构，并引入一致性约束最小化各视角的仿射图与一致表征之间的差异来捕获视角间的互补信息，得到具有高鉴别特征的一致鲁棒表征图。与9种不完备多视角聚类方法在真实和多种缺失率下仿真的不完备多视角数据集中进行实验对比，结果表明：AIM模型均获得了最好的聚类性能。

关 键 词：自适应构图；低秩表示；图融合；图算法；不完备多视角聚类
中图分类号：TP311
文献标志码：A **文章编号**：1001-5965(2025)04-1059-15

现实中的事物往往具有多面性，如某个新闻经过不同的媒体处理可分别成为文字、语言和视频；某一物体可以从不同角度进行拍摄；一个故事可被制作为小说和电影等。事物的每种表达形式可被视为单视角数据，通过对单视角数据的组合，可得到对事物描述更加完善的多视角数据。如何有效地从多视角数据集中获取事物的高鉴别特征是当前图学习聚类领域亟待解决的问题[1]。

谱聚类[2-3]是经典的图聚类方法，可以有效处理面向高维复杂数据的聚类任务，因此，许多学者在多视角图聚类方法中引入了谱聚类思想[4-6]。然而，谱聚类方法的聚类效果往往依赖于初始图的质量，低质量的初始图通常难以取得较好的聚类结果。因此，学习到高质量的初始图对基于图的方法来说是至关重要的。在现有的众多图构造方法中，最常用的是基于距离度量的图构造方法。Nishom[7]利用数据-点之间的距离来计算相似度值；Yin和Sun[8]利用余弦函数构造各视角初始图；文献[9]提出一种随机紧凑高斯核模型，通过高斯核度量数据点间的相似度值。这类利用距离度量数据点间相似度的方法尽管保留了原始空间的流形结构，但无法自动获取适合于图聚类的初始图。为更好地保留数据的局部几何结构，通过全局自表示得到数据点间相似度的方法被提出。例如，Shang等[10]提出了一种基于对偶图正则化特征选择的自表示方法；Weng等[11]基于数据的自表示性及拉普拉斯平滑性来构建初始图。但这类方法针对高维数据集的复杂度较高。Wen等[6]提出了通过数据点的近邻得到初始图的方法，利用kNN-Graph得到各视角的邻接矩阵。这类方法对邻域参数k较为敏感，不合适的参数k会极大地降低方法的聚类效果。

在多视角图聚类领域中，高质量的初始图能够

收稿日期：2023-04-14；**录用日期**：2023-09-14；**网络出版时间**：2023-10-20 09:01
网络出版地址：link.cnki.net/urlid/11.2625.V.20231019.1646.004
基金项目：国家自然科学基金(62266029)；甘肃省重点研发计划(21YF5GA053)；甘肃省高等学校产业支撑计划(2022CYZC-36)
* **通信作者**. E-mail：chenmeilz@mail.lzjtu.cn

引用格式：陈梅，马学艳，张弛，等. 自适应不完备多视角聚类[J]. 北京航空航天大学学报，2025，51(4)：1059-1073.
CHEN M，MA X Y，ZHANG C，et al. Adaptive incomplete multi-view clustering[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1059-1073 (in Chinese).

反映每个视角内部完整的结构特征，同时蕴含各视角间的互补信息，如何充分挖掘这些信息是当前领域的突出问题。Wang 等[12] 分别对每个视角施加低秩约束，通过视角结构一致性约束得到统一低维表征。这种方法虽然可以有效保证各视角之间的局部流形结构，但对具有高度异构特点的多视角数据聚类性能较差。文献 [13] 提出一种基于结构化低秩矩阵分解的方法，通过最小化各视角潜在于数据簇表示之间的分歧目标，迭代地协调所有视角使其一致。Gao 等[5] 提出多视角子空间聚类 (multi-view subspace clustering, MVSC)，采用同一个指标保证聚类结构的一致性，使模型能够学习到用于聚类的一致指示矩阵。Zhang 等[14] 结合自表示和共识表示项提出了一种新的多视角一致性学习方法，利用学习到的一致表征来探索数据点之间潜在交叉视角的关系。上述方法在特定场景下能够取得不错的效果，但在实际应用中，由于一些不可控的环境和设备性能因素，所收集到的数据可能会出现某些视角信息缺失的现象，从而极大地降低上述方法的聚类性能。

为避免出现视角信息缺失影响聚类性能的情况，学者们提出众多解决方案。文献 [15-18] 使用特征平均值填充各视角缺失实例，这种填充策略将缺失实例视为同一个类，直接影响聚类结果，并且在视角融合时还会因为填充值对各视角的一致表征产生负面影响。文献 [6,14,19-20] 使用 0 值填充策略处理缺失值，在缺失比例较大的情况下，这种粗略的填充方式会对聚类结果造成较大的偏差值。文献 [21] 提出利用完备视角数据的拉普拉斯矩阵来补全不完全图，但必须至少有一个完整视角。

为构建能够反映数据全局结构的仿射图，更加精准地填补缺失值并最大程度地挖掘视角间的互补信息来得到高质量的一致低维表征，同时避免视角融合时视角间权重失衡，本文提出自适应不完备多视角聚类 (adaptive incomplete multi-view clustering, AIM) 方法。首先，通过自适应初始图学习模型得到不完备初始图，其次，利用所有有效视角相似度向量的平均值对缺失值进行填充，最后，通过低秩约束、$l_{2,1}$ 范数约束、谱约束及一致性约束获得各视角的一致低维表征。本文方法能够同时处理完备和不完备 2 种多视角数据，对于不完备多视角数据，不需要任何视角的完备性且不限制不完备视角数量；同时，克服了全连接图和 k 近邻图这 2 种图结构的缺点，学习出适合高维复杂数据的更加鲁棒的初始图。在图表示过程中，AIM 模型通过 $l_{2,1}$ 范数弥补隐藏在多视角子空间中的重建误差，提高模型对噪声的鲁棒性，更加准确地解释数据的本质子空间结构。

1 相关理论

1.1 图构造方法

图常以矩阵的形式参与聚类计算，现有的转化方法大多是将数据 $\boldsymbol{X}$ 转化为邻接矩阵，常有以下 2 种图结构：

1）全连接图。数据集中任意 2 个数据对象 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 之间都存在边连接。

2）k 近邻图。数据集中的某个数据对象 $\boldsymbol{x}_i$ 仅与其前 k 个近邻对象之间存在边连接。

令 $\boldsymbol{S}\in\mathbf{R}^{n\times n}$ 为构造出的相似矩阵，如果数据点 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 之间存在边连接，通常，邻接矩阵中对应的元素 s_{ij} 有以下 3 种计算方式：

① 二值 (0-1) 相似度：$s_{ij}=1$。

② 余弦相似度：$s_{ij}=\dfrac{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|}$。

③ 高斯核相似度：$s_{ij}=\exp\left(\dfrac{-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|}{2\sigma}\right)$，其中，$\sigma$ 为缩放因子。

这些方法简单易用，但均存在一定的局限性。例如，二值相似度仅用 0 或 1 表示，不能反映复杂数据之间的相似度；余弦相似度没有考虑数据的局部几何结构；而高斯核相似度基于距离的度量方式使其对噪声和异常点非常敏感。

1.2 谱聚类

设 $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n]\in\mathbf{R}^{m\times n}$ 为含有 n 个样本的数据集，m 为特征维数。谱聚类先根据数据 $\boldsymbol{X}$ 构建相似矩阵 $\boldsymbol{S}$，再根据相似矩阵 $\boldsymbol{S}$ 构建邻接矩阵 $\boldsymbol{W}=[w_{ij}],\forall i,j\in 1,\cdots,n$，以及度矩阵 $\boldsymbol{D}=\left[d_{ii}=\sum_{j=1}^{n}w_{ij}\right]$，计算拉普拉斯矩阵 $\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{W}$，最终利用目标函数得到低维表征，如下：

$$\min_{\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}=\boldsymbol{I}}\mathrm{Tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{F}) \tag{1}$$

式中：$\mathrm{Tr}(\cdot)$ 表示矩阵迹运算；$\boldsymbol{I}$ 为单位矩阵；$\boldsymbol{F}$ 为特征映射矩阵。

模型 (1) 为经典的特征值分解问题，其最优解 $\boldsymbol{F}\in\mathbf{R}^{m\times n}$ 为拉普拉斯矩阵 $\boldsymbol{L}$ 的前 c_1 个最小特征值对应的特征向量的集合。谱聚类仅需要数据之间的相似度矩阵，因此，其能够有效处理稀疏数据的聚类，但聚类效果非常依赖于初始图的构建，不同初始图得到的聚类结果差异较大。

1.3 低秩表示

基于数据通常近似分布于联合子空间的假设，低秩表示 (low-rank representation, LRR) 方法[22] 在低

维表征矩阵上施加低秩约束, 得到能够反映数据全局结构的表示图。LRR 方法的目标函数如下:

$$\begin{cases}\min\limits_{\boldsymbol{Z}}\|\boldsymbol{Z}\|_* \\ \text{s.t. } \boldsymbol{X}=\boldsymbol{X}\boldsymbol{Z}\end{cases} \tag{2}$$

式中: $\boldsymbol{Z}$为表示矩阵; $\|\boldsymbol{Z}\|_*$为$\boldsymbol{Z}$的核范数约束, 计算方式为$\|\boldsymbol{Z}\|_*=\sum\limits_{i=1}^{n}\delta_i$, δ_i为矩阵$\boldsymbol{Z}$的第i个奇异值。

由于实际应用中的数据集通常包含噪声和异常点, 为消除噪声和异常点对聚类结果产生的负面影响, LRR 方法改进为如下模型:

$$\begin{cases}\min\limits_{\boldsymbol{Z},\boldsymbol{E}}\{\|\boldsymbol{Z}\|_*+\lambda\|\boldsymbol{E}\|_1\} \\ \text{s.t. } \boldsymbol{X}=\boldsymbol{X}\boldsymbol{Z}+\boldsymbol{E}\end{cases} \tag{3}$$

式中: $\boldsymbol{E}$为混入数据集中的噪声和异常点; λ为惩罚项参数。

通过对模型 (3) 的求解, 可以得到表示矩阵$\boldsymbol{Z}$, 采用谱聚类方法得到聚类结果。

1.4　多视角子空间聚类

由于多视角数据各视角的特征分布不同, 所构造的仿射图也有较大差异。通过各视角学习一致低维表征, 可以挖掘视角间的互补信息, 从而减少差异。Gao 等[5] 提出 MVSC, 从各视角仿射图中学习一致低维表征, 目标函数如下:

$$\begin{cases}\min\limits_{\boldsymbol{Z}^{(v)},\boldsymbol{F}}\left\{\sum\limits_v\left\|\boldsymbol{X}^{(v)}-\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}\right\|_F^2+\right. \\ \qquad\left.\lambda_1\sum\limits_v\mathrm{Tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{L}^{(v)}\boldsymbol{F})+\lambda_2\sum\limits_v\left\|\boldsymbol{E}^{(v)}\right\|_1\right\} \\ \text{s.t. } \boldsymbol{Z}^{(v)\mathrm{T}}\boldsymbol{J}=\boldsymbol{J},z_{ii}^{(v)}=0,\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}=\boldsymbol{I}\end{cases} \tag{4}$$

式中: $\boldsymbol{X}^{(v)}=\left[\boldsymbol{x}_1^{(v)},\boldsymbol{x}_2^{(v)},\cdots,\boldsymbol{x}_n^{(v)}\right]\in\mathbf{R}^{m\times n}$为第$v$个视角的样本集合; $\boldsymbol{Z}^{(v)}$为第v个视角的子空间表示; $z_{ii}^{(v)}=0(i\in1,2,\cdots,n)$表示矩阵$\boldsymbol{Z}^{(v)}$的对角元素为 0; $\boldsymbol{J}$为所有元素都为 1 的列向量; λ_1和λ_2为惩罚项参数; $\boldsymbol{L}^{(v)}$为第v个视角的拉普拉斯矩阵, $\boldsymbol{L}^{(v)}=\boldsymbol{D}^{(v)}-\boldsymbol{W}^{(v)},\boldsymbol{W}^{(v)}=\left(\left|\boldsymbol{Z}^{(v)}\right|+\left|\boldsymbol{Z}^{(v)}\right|^{\mathrm{T}}\right)/2,d_{ii}^{(v)}=\sum\limits_j w_{ij}^{(v)},w_{ij}^{(v)}$为矩阵$\boldsymbol{W}^{(v)}$的第$(i,j)$个元素; $\boldsymbol{E}^{(v)}$为第v个视角的误差矩阵; $\|\cdot\|_F^2$和$\|\cdot\|_1$分别为 Frobenius 范数和l_1范数。

MVSC 通过融合低维表征学习和仿射图学习, 得到更合理的低维聚类表征结果。针对不完备的数据, 这种将各视角低维表征简单累加为统一表征的融合方法, 可能会因为各视角的缺失率 (partial example ratio, PER) 不同导致视角间的权重失衡, 从而降低聚类效果。

2　本文方法

AIM 方法框架如图 1 所示。

2.1　自适应初始图学习模型

由第 1 节相关理论可知, 基于距离度量构造的相似度矩阵对噪声和异常点敏感。因此, 通过对初始图矩阵$\boldsymbol{S}$施加稀疏化约束, 增加模型对噪声和异常点的鲁棒性, 引入基于距离信息的流形正则项, 使数据点间的距离和相似度值成反比。故本文提出如下自适应初始图学习模型:

$$\begin{cases}\min\limits_{\boldsymbol{S}}\sum\limits_v\left(\sum\limits_{i,j=1}^{n}\left\|\boldsymbol{x}_i^{(v)}-\boldsymbol{x}_j^{(v)}\right\|_2^2s_{ij}^{(v)}+\beta\left\|\boldsymbol{S}^{(v)}\right\|_F^2\right) \\ \text{s.t. } \forall v,s_{ii}^{(v)}=0,0\leqslant s_{ij}^{(v)}\leqslant1,\boldsymbol{J}^{\mathrm{T}}\boldsymbol{s}_i^{(v)}=1\end{cases} \tag{5}$$

式中: $\boldsymbol{s}_i^{(v)}$为第v个视角中对象$\boldsymbol{x}_i$与其他对象的相似度值所组成的向量; $\beta\left\|\boldsymbol{S}^{(v)}\right\|_F^2$为诱导因子, 可使每个数据点与$\boldsymbol{x}_i$的相似度为$1/n$, 从而避免模型出现无用解; $\boldsymbol{J}^{\mathrm{T}}\boldsymbol{s}_i^{(v)}=1$为$\boldsymbol{S}$的稀疏约束项, 可增强模型对噪声和离群点的鲁棒性。

为解决多视角聚类方法无法从不完备的数据中构造初始图的问题, 本文假设每个数据对象至少在一个视角下是存在的, 利用符号标识缺失值生成各视角的不完备初始图。

$$\Gamma(\boldsymbol{x}_i^{(v)})=\begin{cases}\boldsymbol{x}_i^{(v)} & \text{第}v\text{个视角存在 }\boldsymbol{x}_i^{(v)} \\ \text{NaN} & \text{其他}\end{cases} \tag{6}$$

式中: NaN 为无效数字标识符, 代替缺失实例。

据此, 可将式 (5) 改为

$$\begin{cases}\min\limits_{\boldsymbol{S}}\sum\limits_v\left(\sum\limits_{i,j=1}^{n}\left\|\Gamma(\boldsymbol{x}_i^{(v)})-\Gamma(\boldsymbol{x}_j^{(v)})\right\|_2^2s_{ij}^{(v)}+\beta\left\|\boldsymbol{s}_i^{(v)}\right\|_F^2\right) \\ \text{s.t.}\forall v,s_{ii}^{(v)}=0,0\leqslant s_{ij}^{(v)}\leqslant1,\boldsymbol{J}^{\mathrm{T}}\boldsymbol{s}_i^{(v)}=1\end{cases} \tag{7}$$

式 (7) 将数据缺失转换为相似度缺失, 若第v个视角下数据对象$\boldsymbol{x}_i^{(v)}$或$\boldsymbol{x}_j^{(v)}$缺失, 式 (7) 给变量$s_{ij}^{(v)}$分配 NaN, 否则分配相应的相似度值。为解决相似度缺失的问题, 本文采用所有有效视角相似度平均值填充策略进行填充, 如下:

$$\Omega(\boldsymbol{s}_i^{(v)})=\begin{cases}\boldsymbol{s}_i^{(v)} & s_{ij}^{(v)}\neq\text{NaN} \\ \sum\limits_j\dfrac{\boldsymbol{s}_i^{(j)}}{N_v} & \text{其他}\end{cases} \tag{8}$$

式中: $\boldsymbol{s}_i^{(j)}$为第j个有效视角的第i个相似度向量; N_v为所有有效视角的个数。

通过模型 (7) 和式 (8) 可获得各视角的完备初始图, 这种构造方式克服了对噪声敏感的缺点, 同时补充了缺失实例, 并保留了部分特征结构。

2.2　不完备多视角聚类模型

由于高维数据一般分布在具有低秩结构的子空间中, 对仿射图施加低秩约束可以捕获数据的全局结构。据此, AIM 模型在模型 (4) 的基础上对每

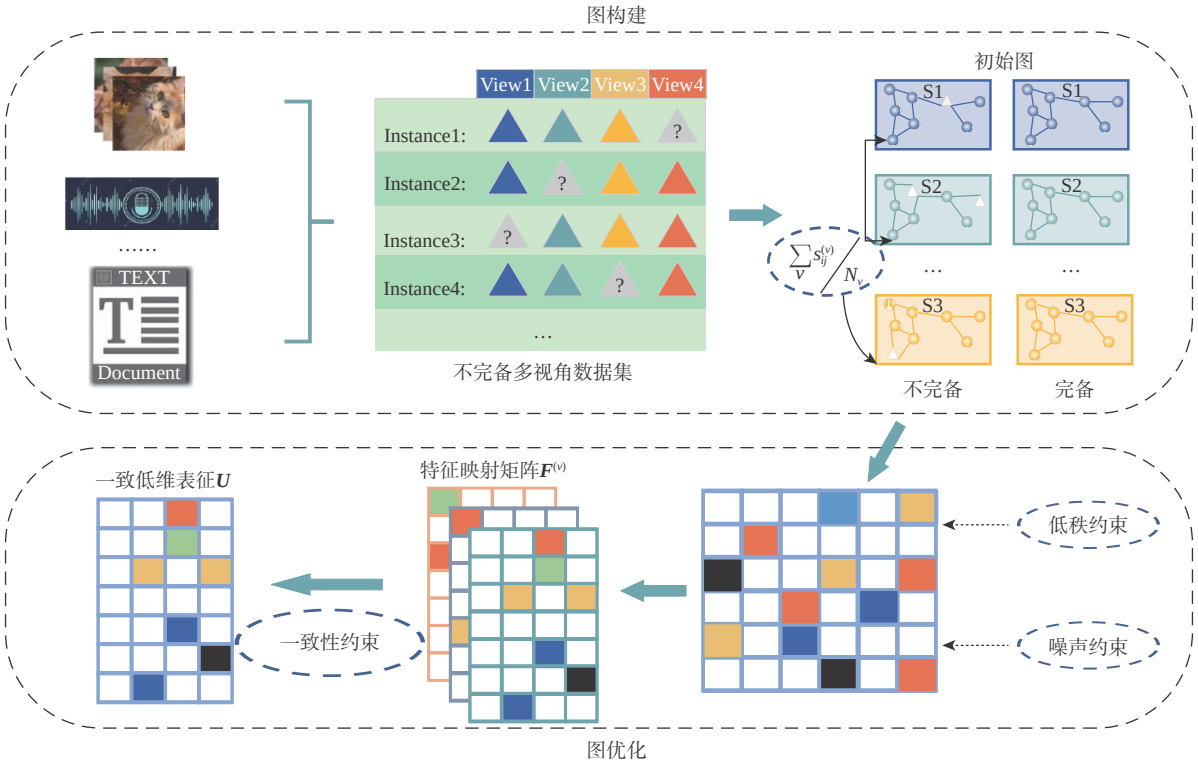


图1 AIM框架

Fig. 1 Framework of AIM

个视角的仿射图施加低秩约束，得到如下模型：

$$\begin{cases}\min\limits_{\boldsymbol{Z}^{(v)},\boldsymbol{F}^{(v)}}\left\{\sum\limits_{v}\operatorname{rank}(\boldsymbol{Z}^{(v)})+\lambda_1\sum\limits_{v}\operatorname{Tr}(\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{L}^{(v)}\boldsymbol{F}^{(v)})+\right.\\ \left.\lambda_2\sum\limits_{v}\left\|\boldsymbol{E}^{(v)}\right\|_1\right\}\\ \text{s.t.}\begin{cases}\boldsymbol{X}^{(v)}=\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}+\boldsymbol{E}^{(v)},\boldsymbol{Z}^{(v)\mathrm{T}}\boldsymbol{J}=\boldsymbol{J}\\ z_{ii}^{(v)}=0,\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{F}^{(v)}=\boldsymbol{I}\end{cases}\end{cases}\tag{9}$$

式中：$\operatorname{rank}(\boldsymbol{Z}^{(v)})$表示矩阵$\boldsymbol{Z}^{(v)}$的秩。

为解决秩极小化中NP难的问题，本文采用核范数代替秩函数。同时，假设数据中只有特定样本被严重破坏，而其他样本是干净的，即误差矩阵$\boldsymbol{E}^{(v)}$具有稀疏的列支持。因此，将模型(9)修改为

$$\begin{cases}\min\limits_{\boldsymbol{Z}^{(v)},\boldsymbol{E}^{(v)},\boldsymbol{F}^{(v)}}\left\{\sum\limits_{v}(\left\|\boldsymbol{Z}^{(v)}\right\|_*+\lambda_1\operatorname{Tr}(\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{L}^{(v)}\boldsymbol{F}^{(v)}))+\right.\\ \left.\sum\limits_{v}\lambda_2\left\|\boldsymbol{E}^{(v)}\right\|_{2,1}\right\}\\ \text{s.t.}\begin{cases}\boldsymbol{X}^{(v)}=\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}+\boldsymbol{E}^{(v)},\boldsymbol{Z}^{(v)}\boldsymbol{J}=\boldsymbol{J},\boldsymbol{0}\leqslant\boldsymbol{Z}^{(v)}\leqslant\boldsymbol{O}\\ z_{ii}^{(v)}=0,\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{F}^{(v)}=\boldsymbol{I}\end{cases}\end{cases}\tag{10}$$

式中：$\left\|\boldsymbol{Z}^{(v)}\right\|_*$表示矩阵$\boldsymbol{Z}^{(v)}$的核范数约束，其值为矩阵$\boldsymbol{Z}^{(v)}$的所有奇异值之和；$\boldsymbol{O}$为全1矩阵。

为判断所有样本的类别，模型(10)引入约束$\boldsymbol{Z}^{(v)}\boldsymbol{J}=\boldsymbol{J}$，使所有样本至少与其余样本中的一个样本存在某种关联。矩阵$\boldsymbol{F}^{(v)}$为指示矩阵，其中每一行可以视为原始样本具有高鉴别性的低维表征。同时，模型(10)对所有视角仿射图进行累加得到一致低维表征。对于不完备数据来说，该融合方式会造成缺失视角和未缺失视角的权重失衡，进而增加不同类样本的累加约束权重大于同类样本的概率。为解决该问题，本文在模型(10)中引入各视角表征矩阵差异约束，得到如下模型：

$$\begin{cases}\min\limits_{\boldsymbol{Z}^{(v)},\boldsymbol{E}^{(v)},\boldsymbol{F}^{(v)},\boldsymbol{U}}\sum\limits_{v}\left(\left\|\boldsymbol{Z}^{(v)}\right\|_*+\lambda_1\operatorname{Tr}(\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{L}^{(v)}\boldsymbol{F}^{(v)})+\right.\\ \left.\lambda_2\left\|\boldsymbol{E}^{(v)}\right\|_{2,1}+\dfrac{\lambda_3}{2}\Upsilon(\boldsymbol{F}^{(v)},\boldsymbol{U})\right)\\ \text{s.t.}\begin{cases}\boldsymbol{X}^{(v)}=\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}+\boldsymbol{E}^{(v)},\boldsymbol{Z}^{(v)}\boldsymbol{J}=\boldsymbol{J},\boldsymbol{0}\leqslant\boldsymbol{Z}^{(v)}\leqslant\boldsymbol{O}\\ z_{ii}^{(v)}=0,\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{F}^{(v)}=\boldsymbol{I},\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}\end{cases}\end{cases}\tag{11}$$

式中：λ_3为惩罚项参数；矩阵$\boldsymbol{U}\in\mathbf{R}^{n\times c}$为所有视角一致低维表征矩阵，$n$为样本个数，$c$为簇个数；函数$\Upsilon(\boldsymbol{F}^{(v)},\boldsymbol{U})$用来衡量各视角的低维表征$\boldsymbol{F}^{(v)}$与一致低维表征$\boldsymbol{U}$之间的差距，具体如下：

$$\Upsilon(\boldsymbol{F}^{(v)},\boldsymbol{U})=\left\|\frac{\boldsymbol{K}_{\boldsymbol{U}}}{\|\boldsymbol{K}_{\boldsymbol{U}}\|_F^2}-\frac{\boldsymbol{K}_{\boldsymbol{F}^{(v)}}}{\|\boldsymbol{K}_{\boldsymbol{F}^{(v)}}\|_F^2}\right\|_F^2\tag{12}$$

其中：$\boldsymbol{K}_{\boldsymbol{U}}$为一致低维表征$\boldsymbol{U}$的核矩阵；$\boldsymbol{K}_{\boldsymbol{F}^{(v)}}$为各视角的表示矩阵$\boldsymbol{F}^{(v)}$的核矩阵；$\boldsymbol{K}_{\boldsymbol{U}}=\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}$，$\boldsymbol{K}_{\boldsymbol{F}^{(v)}}=\boldsymbol{F}^{(v)}\boldsymbol{F}^{(v)\mathrm{T}}$。

根据$\|\boldsymbol{K}_{\boldsymbol{U}}\|_F^2=c_4$和$\|\boldsymbol{K}_{\boldsymbol{F}^{(v)}}\|_F^2=c_4$，式(12)可简化为

$$\Upsilon(\boldsymbol{F}^{(v)},\boldsymbol{U})=2(c_4-\operatorname{Tr}(\boldsymbol{F}^{(v)}\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}))/c_4^2\tag{13}$$

忽略常数c_4, 根据模型 (11) 和式 (13) 可得到最终目标模型:

$$\begin{cases}\min\limits_{\boldsymbol{Z}^{(v)},\boldsymbol{E}^{(v)},\boldsymbol{F}^{(v)},\boldsymbol{U}}\sum\limits_{v}(\left\|\boldsymbol{Z}^{(v)}\right\|_{*}+\lambda_1\mathrm{Tr}(\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{L}^{(v)}\boldsymbol{F}^{(v)})+\\ \qquad\lambda_2\left\|\boldsymbol{E}^{(v)}\right\|_{2,1}-\lambda_3\mathrm{Tr}(\boldsymbol{F}^{(v)}\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}))\\ \text{s.t.}\begin{cases}\boldsymbol{X}^{(v)}=\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}+\boldsymbol{E}^{(v)},\boldsymbol{Z}^{(v)}\boldsymbol{J}=\boldsymbol{J},\boldsymbol{0}\leqslant\boldsymbol{Z}^{(v)}\leqslant\boldsymbol{O}\\ z_{ii}^{(v)}=0,\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{F}^{(v)}=\boldsymbol{I},\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}\end{cases}\end{cases}\tag{14}$$

AIM 方法通过模型 (14) 在完备初始图上弥补隐藏在多视角子空间中的重建误差, 获得具有数据全局结构信息和视角间互补信息的一致低维表征, 通过对其进行 k-Means 得到聚类结果。

2.3　模型优化

2.3.1　自适应初始图学习模型优化过程

由于模型 (7) 需要单独计算每个视角的初始图, 本节给出第v个视角的自适应初始图模型优化过程。令$d_{ij}^{(v)}=\left\|\Gamma(\boldsymbol{x}_i^{(v)})-\Gamma(\boldsymbol{x}_j^{(v)})\right\|_2^2$, 同时令$\boldsymbol{d}_i^{(v)}$为第$j$个元素是$d_{ij}^{(v)}$的向量, 则模型 (7) 用向量的形式可以表示为

$$\begin{cases}\min\limits_{\boldsymbol{s}_i^{(v)}}\left\|\boldsymbol{s}_i^{(v)}+\dfrac{\boldsymbol{d}_i^{(v)}}{2\beta}\right\|_2^2\\ \text{s.t.}\,s_{ii}^{(v)}=0,0\leqslant s_{ij}^{(v)}\leqslant 1,\boldsymbol{J}^{\mathrm{T}}\boldsymbol{s}_i^{(v)}=1\end{cases}\tag{15}$$

模型 (15) 的增广拉格朗日函数为

$$\ell(\boldsymbol{s}_i^{(v)},\eta,\boldsymbol{\zeta})=\frac{1}{2}\left\|\boldsymbol{s}_i^{(v)}+\frac{\boldsymbol{d}_i^{(v)}}{2\beta}\right\|-\eta(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{s}_i^{(v)}-1)-\boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{s}_i^{(v)}\tag{16}$$

式中: η和$\boldsymbol{\zeta}$为拉格朗日乘子。

对式 (16) 求变量$\boldsymbol{s}_i^{(v)}$的偏导, 并令偏导数为 0, 可以得到

$$\boldsymbol{s}_i^{(v)}+\frac{\boldsymbol{d}_i^{(v)}}{2\beta}-\eta\boldsymbol{J}-\boldsymbol{\zeta}=\boldsymbol{0}\tag{17}$$

将式 (17) 表示为单个元素:

$$s_{ij}^{(v)}+\frac{d_{ij}^{(v)}}{2\beta}-\eta-\zeta_j=0\tag{18}$$

根据 Karush-Kuhn-Tucker (KKT) 条件: $s_{ij}^{(v)}\zeta_j=0$。将式 (18) 两边同时乘$s_{ij}^{(v)}$并做简单推导, 得到$s_{ij}^{(v)}$的解 (记为$\hat{s}_{ij}^{(v)}$) 如下:

$$\hat{s}_{ij}^{(v)}=\left(-\frac{d_{ij}^{(v)}}{2\beta}+\eta\right)_{+}\tag{19}$$

本文假设$d_{i1}^{(v)},d_{i2}^{(v)},\cdots,d_{in}^{(v)}$由大至小进行排列, 设定每个元素$x_{ij}$仅有$k$个邻域, 即$\boldsymbol{s}_i^{(v)}$中最多有$k$个非 0 值, 因此, $\boldsymbol{s}_i^{(v)}$的学习需要满足$\hat{s}_{ik}^{(v)}>0$且$\hat{s}_{i,k+1}^{(v)}=0$, 可以得出

$$\begin{cases}-\dfrac{d_{ik}^{(v)}}{2\beta}+\eta>0\\ -\dfrac{d_{i,k+1}^{(v)}}{2\beta}+\eta\leqslant 0\end{cases}\tag{20}$$

同时, 将式 (19) 代入约束$\boldsymbol{J}^{\mathrm{T}}\boldsymbol{s}_i^{(v)}=1$, 可以推出最优的$\eta$为

$$\eta=\frac{1}{k}+\frac{1}{2k\beta}\sum_{j=1}^{k}d_{ij}^{(v)}\tag{21}$$

将式 (21) 代入式 (20) 中, 可以得到如下不等式:

$$\frac{k}{2}d_{ik}^{(v)}-\frac{1}{2}\sum_{j=1}^{k}d_{ij}^{(v)}<\beta\leqslant\frac{k}{2}d_{i,k+1}^{(v)}-\frac{1}{2}\sum_{j=1}^{k}d_{ij}^{(v)}\tag{22}$$

为使学习到的$\hat{s}_{ij}^{(v)}$最多只有k个值, 令参数β的值为

$$\beta=\frac{k}{2}d_{i,k+1}^{(v)}-\frac{1}{2}\sum_{j=1}^{k}d_{ij}^{(v)}\tag{23}$$

根据式 (20)～式 (23), 可以求得向量$\boldsymbol{s}_i^{(v)}$中元素$\hat{s}_{ij}^{(v)}$如下:

$$\hat{s}_{ij}^{(v)}=\begin{cases}\dfrac{d_{i,k+1}^{(v)}-d_{ij}^{(v)}}{kd_{i,k+1}^{(v)}-\sum\limits_{h=1}^{k}d_{ih}^{(v)}} & j\leqslant k\\ 0 & j>k\end{cases}\tag{24}$$

上述为模型 (7) 的详细优化过程, 通过式 (24) 可得到各视角的不完备初始图, 再根据式 (8) 填充缺失值得到完备初始图。

方法 1　自适应初始图学习。

输入: 数据矩阵$\boldsymbol{X}$和近邻个数k。

输出: 初始图$\boldsymbol{S}^{(v)}$。

```
begin
    //依次计算每个视角的初始图
    for i = 1 to n do
        for j = 1 to n do
            计算式 (16), 即拉格朗日函数;
            将d_in^(v)进行降序排序;
            通过式 (21) 计算拉格朗日乘子η;
            通过式 (23) 计算拉格朗日乘子β;
            通过式 (24) 计算向量s_i^(v)中元素ŝ_ij^(v)。
        end
    end
end
```

2.3.2　不完备多视角聚类模型优化过程

本节采用交替方向乘子法对 AIM 模型进行优化。为使模型中各变量可以单独求解, 引入辅助变量$\boldsymbol{P}$和$\boldsymbol{Q}$, 如下:

$$\begin{cases}\min\limits_{\boldsymbol{Z}^{(v)},\boldsymbol{E}^{(v)},\boldsymbol{F}^{(v)},\boldsymbol{U},\boldsymbol{P}^{(v)},\boldsymbol{Q}^{(v)}}\sum\limits_{v}(\left\|\boldsymbol{P}^{(v)}\right\|_{*}+\lambda_1\mathrm{Tr}(\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{L}_w^{(v)}\boldsymbol{F}^{(v)})+\\ \qquad\lambda_2\left\|\boldsymbol{E}^{(v)}\right\|_{2,1}-\lambda_3\mathrm{Tr}(\boldsymbol{F}^{(v)}\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}))\\ \text{s.t.}\begin{cases}\boldsymbol{X}^{(v)}=\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}+\boldsymbol{E}^{(v)},\boldsymbol{Z}^{(v)}=\boldsymbol{P}^{(v)},\boldsymbol{Z}^{(v)}=\boldsymbol{Q}^{(v)}\\ \boldsymbol{Q}^{(v)}\boldsymbol{J}=1,\boldsymbol{0}\leqslant\boldsymbol{Q}^{(v)}\leqslant\boldsymbol{O},q_{ii}^{(v)}=0,\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{F}^{(v)}=\boldsymbol{I}\\ \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}\end{cases}\end{cases}\tag{25}$$

式中：$\boldsymbol{L}_w^{(v)}$为拉普拉斯矩阵。

式(25)的增广拉格朗日函数为

$$\begin{aligned}L(a)=&\sum_v\left(\left\|\boldsymbol{P}^{(v)}\right\|_*+\lambda_1\mathrm{Tr}(\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{L}_w^{(v)}\boldsymbol{F}^{(v)})\right)+\\&\sum_v\left(\lambda_2\left\|\boldsymbol{E}^{(v)}\right\|_{2,1}-\lambda_3\mathrm{Tr}(\boldsymbol{F}^{(v)}\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}})\right)+\\&\frac{\mu}{2}\sum_v\left(\left\|\boldsymbol{Z}^{(v)}-\boldsymbol{P}^{(v)}+\boldsymbol{C}_2^{(v)}/2\right\|_F^2+\right.\\&\left.\left\|\boldsymbol{Z}^{(v)}-\boldsymbol{Q}^{(v)}+\boldsymbol{C}_3^{(v)}/2\right\|_F^2\right)+\\&\sum_v\left(\frac{\mu}{2}\left\|\boldsymbol{X}^{(v)}-\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}-\boldsymbol{E}^{(v)}+\boldsymbol{C}_1^{(v)}/2\right\|_F^2-\right.\\&\left.\boldsymbol{\psi}^{(v)\mathrm{T}}(\boldsymbol{Q}^{(v)}\boldsymbol{J}-\boldsymbol{J})\right)\end{aligned}\tag{26}$$

式中：$\boldsymbol{C}_1^{(v)}$、$\boldsymbol{C}_2^{(v)}$、$\boldsymbol{C}_3^{(v)}$和$\boldsymbol{\psi}^{(v)}$为拉格朗日乘子；μ为惩罚项参数。

通过固定相关变量，对问题(26)中的剩余变量逐个优化得到最优解，以下为详细优化过程。

步骤1 更新变量$\boldsymbol{Z}^{(v)}$。固定所有与$\boldsymbol{Z}^{(v)}$无关的变量，则变量$\boldsymbol{Z}^{(v)}$的求解变为如下形式：

$$\begin{aligned}L(\boldsymbol{Z}^{(v)})=&\left\|\boldsymbol{Z}^{(v)}-\boldsymbol{P}^{(v)}+\boldsymbol{C}_2^{(v)}/2\right\|_F^2+\left\|\boldsymbol{Z}^{(v)}-\boldsymbol{Q}^{(v)}+\right.\\&\left.\boldsymbol{C}_3^{(v)}/2\right\|_F^2+\left\|\boldsymbol{X}^{(v)}-\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}-\boldsymbol{E}^{(v)}+\boldsymbol{C}_1^{(v)}/2\right\|_F^2\end{aligned}\tag{27}$$

对式(27)求变量$\boldsymbol{Z}^{(v)}$的偏导数，如下：

$$\begin{aligned}\frac{\partial L(\boldsymbol{Z}^{(v)})}{\partial\boldsymbol{Z}^{(v)}}=&2\boldsymbol{Z}^{(v)}-2\boldsymbol{M}_1^{(v)}+2\boldsymbol{Z}^{(v)}-\\&2\boldsymbol{M}_2^{(v)}+2\boldsymbol{X}^{(v)\mathrm{T}}(\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}-\boldsymbol{M}_3^{(v)})\end{aligned}\tag{28}$$

式中：$\boldsymbol{M}_1^{(v)}=\boldsymbol{P}^{(v)}-\boldsymbol{C}_2^{(v)}/\mu$；$\boldsymbol{M}_2^{(v)}=\boldsymbol{Q}^{(v)}-\boldsymbol{C}_3^{(v)}/\mu$；$\boldsymbol{M}_3^{(v)}=\boldsymbol{X}^{(v)}-\boldsymbol{E}^{(v)}+\boldsymbol{C}_3^{(v)}/\mu$。

令$\frac{\partial L(\boldsymbol{Z}^{(v)})}{\partial\boldsymbol{Z}^{(v)}}=0$，可得变量$\boldsymbol{Z}^{(v)}$的最优解，如下：

$$\boldsymbol{Z}^{(v)}=(\boldsymbol{X}^{(v)\mathrm{T}}\boldsymbol{X}^{(v)}+2\boldsymbol{I})^{-1}(\boldsymbol{X}^{(v)\mathrm{T}}\boldsymbol{M}_3^{(v)}+\boldsymbol{M}_1^{(v)}+\boldsymbol{M}_2^{(v)})\tag{29}$$

步骤2 更新变量$\boldsymbol{P}^{(v)}$。固定与$\boldsymbol{P}^{(v)}$无关的其他变量，可得到如下优化子问题：

$$\min_{\boldsymbol{P}^{(v)}}\left\{\left\|\boldsymbol{P}^{(v)}\right\|_*+\frac{\mu}{2}\left\|\boldsymbol{Z}^{(v)}-\boldsymbol{P}^{(v)}+\boldsymbol{C}_2^{(v)}/2\right\|_F^2\right\}\tag{30}$$

式(30)为典型的核范数优化问题，本文利用奇异值阈值收缩(singular value decomposition, SVT)算子[23]得到如下最优解：

$$\boldsymbol{P}^{(v)}=\Theta_{1/\mu}(\boldsymbol{Z}^{(v)}+\boldsymbol{C}_2^{(v)}/\mu)\tag{31}$$

式中：Θ表示SVT收缩算子。

步骤3 更新变量$\boldsymbol{Q}^{(v)}$和$\boldsymbol{\psi}^{(v)}$。固定与$\boldsymbol{Q}^{(v)}$和$\boldsymbol{\psi}^{(v)}$无关的其他变量，可得到如下优化子问题：

$$\begin{aligned}\min_{\boldsymbol{Q}^{(v)}\geqslant\boldsymbol{0},q_{ii}=0}&\left\{\lambda_1\mathrm{Tr}(\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{L}_w^{(v)}\boldsymbol{F}^{(v)})+\frac{\mu}{2}\left\|\boldsymbol{Z}^{(v)}-\boldsymbol{Q}^{(v)}+\right.\right.\\&\left.\left.\boldsymbol{C}_3^{(v)}/2\right\|_F^2-\boldsymbol{\psi}^{(v)\mathrm{T}}(\boldsymbol{Q}^{(v)}\boldsymbol{J}-\boldsymbol{J})\right\}\end{aligned}\tag{32}$$

令$\boldsymbol{R}^{(v)}=\boldsymbol{Z}^{(v)}+\boldsymbol{C}_3^{(v)}/2$，可得到如下优化问题：

$$\begin{aligned}\min_{\boldsymbol{Q}^{(v)}\geqslant\boldsymbol{0},q_{ii}=0}&\left\{\frac{\lambda_1}{2}\sum_{i,j=1}^{n_v}\left\|\boldsymbol{f}_{i,:}^{(v)}-\boldsymbol{f}_{j,:}^{(v)}\right\|_2^2q_{ij}^{(v)}+\right.\\&\left.\frac{\mu}{2}\left\|\boldsymbol{Q}^{(v)}-\boldsymbol{R}^{(v)}\right\|_F^2-\boldsymbol{\psi}^{(v)\mathrm{T}}(\boldsymbol{Q}^{(v)}\boldsymbol{J}-\boldsymbol{J})\right\}\end{aligned}\tag{33}$$

式中：n_v为视角个数；$\boldsymbol{f}_i^{(v)}$和$\boldsymbol{f}_j^{(v)}$分别为指示矩阵$\boldsymbol{F}^{(v)}$的第i行和第j行向量。

从式(33)可以看出，该优化问题的各行相互独立，令$h_{ij}^{(v)}=\left\|\boldsymbol{f}_{i,:}^{(v)}-\boldsymbol{f}_{j,:}^{(v)}\right\|_2^2$，则问题(33)可简化为如下优化问题：

$$\min_{\boldsymbol{Q}^{(v)}\geqslant\boldsymbol{0},q_{ii}^{(v)}=0}\left\{\frac{\mu}{2}\left\|\boldsymbol{q}_{i,:}^{(v)}-(\boldsymbol{r}_{i,:}^{(v)}-\frac{\lambda_1}{2\mu}\boldsymbol{h}_{i,:}^{(v)})\right\|_2^2-\boldsymbol{\psi}_i^{(v)\mathrm{T}}(\boldsymbol{q}_{i,:}^{(v)}\boldsymbol{J}-\boldsymbol{J})\right\}\tag{34}$$

显然，问题(34)有如下最优解：

$$q_{ij}^{(v)}=\begin{cases}0 & j=i\\ r_{ij}^{(v)}-\dfrac{\lambda_1}{2\mu}h_{ij}^{(v)}+\dfrac{\psi_{ij}^{(v)}}{\mu} & \text{其他}\end{cases}\tag{35}$$

令$\boldsymbol{Q}^{(v)}=\max(\boldsymbol{Q}^{(v)},\boldsymbol{0})$，将式(35)代入约束条件$\boldsymbol{Q}^{(v)}\boldsymbol{J}=\boldsymbol{J}$中，得到拉格朗日乘子$\psi_{ij}^{(v)}$的最优值如下：

$$\psi_{ij}^{(v)}=\mu\left[1-\sum_{j=1,j\neq i}^{n_v}(r_{ij}^{(v)}-\frac{\lambda_1}{2\mu}h_{ij}^{(v)})\right]\Big/(n_v-1)\tag{36}$$

式中：n_v为视角个数。

步骤4 更新变量$\boldsymbol{E}^{(v)}$。固定与$\boldsymbol{E}^{(v)}$无关的其他变量，可得到如下优化子问题：

$$\min_{\boldsymbol{E}^{(v)}}\left\{\lambda_2\left\|\boldsymbol{E}^{(v)}\right\|_{2,1}+\frac{\mu}{2}\left\|\boldsymbol{X}^{(v)}-\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}-\boldsymbol{E}^{(v)}+\boldsymbol{C}_1^{(v)}/2\right\|_F^2\right\}\tag{37}$$

令$\boldsymbol{B}^{(v)}=\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}-\boldsymbol{X}^{(v)}-\boldsymbol{C}_1^{(v)}/2$，可将式(37)简化为

$$\min_{\boldsymbol{E}^{(v)}}\frac{\lambda_2}{\mu}\left(\left\|\boldsymbol{E}^{(v)}\right\|_{2,1}+\frac{1}{2}\left\|\boldsymbol{E}^{(v)}-\boldsymbol{B}^{(v)}\right\|_F^2\right)\tag{38}$$

最小化式(38)，可以得到$\boldsymbol{E}^{(v)}$的最优值如下：

$$\boldsymbol{E}^{(v)}(:,i)=\begin{cases}\dfrac{\|\boldsymbol{b}_i\|_2-\lambda_2/\mu}{\|\boldsymbol{b}_i\|_2}\boldsymbol{b}_i & \dfrac{\lambda_2}{\mu}\leqslant\|\boldsymbol{b}_i\|_2\\ \boldsymbol{0} & \text{其他}\end{cases}\tag{39}$$

步骤5 更新变量$\boldsymbol{F}^{(v)}$。固定与$\boldsymbol{F}^{(v)}$无关的其他变量，可得到如下优化子问题：

$$\min_{\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{F}^{(v)}=\boldsymbol{I}}\left\{\lambda_1\mathrm{Tr}(\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{L}_w^{(v)}\boldsymbol{F}^{(v)})-\lambda_3\mathrm{Tr}(\boldsymbol{F}^{(v)}\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}})\right\}\tag{40}$$

问题(40)可以转化为

$$\max_{\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{F}^{(v)}=\boldsymbol{I}}\mathrm{Tr}(\boldsymbol{F}^{(v)\mathrm{T}}(\lambda_3\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}-\lambda_1\boldsymbol{L}_w^{(v)})\boldsymbol{F}^{(v)})\tag{41}$$

问题(41)为典型的特征值分解问题，最优解为矩阵$(\lambda_3\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}-\lambda_1\boldsymbol{L}_w^{(v)})$的前$c_2$个最大特征值组成的特征向量。

步骤6 更新变量$\boldsymbol{U}$。固定与$\boldsymbol{U}$无关的其他变

量, 可得到如下优化子问题:

$$\min_{\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}} -\lambda_3 \mathrm{Tr}(\boldsymbol{F}^{(v)}\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}) \tag{42}$$

问题 (42) 可以转化为

$$\max_{\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}} -\lambda_3 \mathrm{Tr}(\boldsymbol{U}^{\mathrm{T}}(\boldsymbol{F}^{(v)}\boldsymbol{F}^{(v)\mathrm{T}})\boldsymbol{U}^{\mathrm{T}}) \tag{43}$$

问题 (43) 也是特征值分解问题, 最优解为矩阵 $\boldsymbol{F}^{(v)}\boldsymbol{F}^{(v)\mathrm{T}}$ 的前 c_3 个最大特征值组成的特征向量。

步骤 7　更新变量 $\boldsymbol{C}_1^{(v)}$、$\boldsymbol{C}_2^{(v)}$、$\boldsymbol{C}_3^{(v)}$ 和 μ。与这些变量相关的优化式如下:

$$\begin{cases} \boldsymbol{C}_1^{(v)} = \boldsymbol{C}_1^{(v)} + \mu(\boldsymbol{X}^{(v)} - \boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)} - \boldsymbol{E}^{(v)}) \\ \boldsymbol{C}_2^{(v)} = \boldsymbol{C}_2^{(v)} + \mu(\boldsymbol{Z}^{(v)} - \boldsymbol{P}^{(v)}) \\ \boldsymbol{C}_3^{(v)} = \boldsymbol{C}_3^{(v)} + \mu(\boldsymbol{Z}^{(v)} - \boldsymbol{Q}^{(v)}) \end{cases} \tag{44}$$

$$\mu = \min(\rho\mu, \mu_0) \tag{45}$$

式中: ρ 和 μ_0 为预定义的正常数。

上述为目标模型 (14) 的详细优化过程, 通过迭代逐一更新优化以上变量, 可得到这些变量的局部最优解。不完备多视角聚类模型的方法过程如方法 2 所示。AIM 利用方法 2 得到各视角的统一低维表征矩阵后, 采用 k-Means 算法得到最终的聚类结果。

方法 2　不完备多视角聚类模型。

输入: 数据矩阵 $\boldsymbol{X}$、完备初始图 $\boldsymbol{Z}$、参数 λ_1、λ_2 和 λ_3。

输出: 表征矩阵 $\boldsymbol{U}$。

1. 初始化: 通过对各视角的拉普拉斯矩阵进行特征值分解初始化 $\boldsymbol{F}^{(v)}$, 使用式 (39) 初始化矩阵 $\boldsymbol{U}$, 令 $\mu_0 = 10^8$, $\rho = 1.1$, $\mu = 0.01$。
2. while not converge do
3. for $v = 1$ to n_v do
4. 通过式 (29) 更新变量 $\boldsymbol{Z}^{(v)}$。
5. 通过式 (31) 更新变量 $\boldsymbol{P}^{(v)}$。
6. 通过式 (35) 和式 (36) 更新变量 $\boldsymbol{Q}^{(v)}$ 和 $\boldsymbol{\psi}^{(v)}$。
7. 计算 $\boldsymbol{Q}^{(v)} = \max(\boldsymbol{Q}^{(v)}, \boldsymbol{0})$。
8. 通过式 (39) 更新变量 $\boldsymbol{E}^{(v)}$。
9. 通过式 (41) 更新变量 $\boldsymbol{F}^{(v)}$。
10. 通过式 (44) 更新变量 $\boldsymbol{C}_1^{(v)}$、$\boldsymbol{C}_2^{(v)}$ 和 $\boldsymbol{C}_3^{(v)}$。
11. end
12. 通过式 (43) 更新变量 $\boldsymbol{U}$。
13. 通过式 (45) 更新变量 μ。
14. end

2.4　时间复杂度分析

本节对 AIM 方法进行时间复杂度分析。方法 1 主要的复杂度在于计算各视角的距离矩阵, 该项的复杂度为 $O(n^2)$, 因此, 方法 1 的时间复杂度为 $O(n_v n^2)$, 其中, n_v 为视角个数, n 为数据集的点数。在忽略矩阵基本运算开销的前提下, 方法 2 的时间复杂度主要包括步骤 1 中更新变量 $\boldsymbol{Z}^{(v)}$ 时矩阵的逆运算、步骤 2 中更新变量 $\boldsymbol{P}^{(v)}$ 时的奇异值分解、步骤 5 中更新变量 $\boldsymbol{F}^{(v)}$ 和步骤 6 中更新变量 $\boldsymbol{U}$ 时的特征值分解。对于大小为 $n \times n$ 的矩阵, 后 3 项的时间复杂度均为 $O(n^3)$。由于步骤 1 中矩阵的逆运算 $(\boldsymbol{X}^{(v)\mathrm{T}}\boldsymbol{X}^{(v)} + 2\boldsymbol{I})^{-1}$ 在迭代循环外, 忽略其时间复杂度。因此, 方法 2 的时间复杂度为 $O(\tau(3n_v n^3))$, 其中, τ 为迭代次数。因此, AIM 方法的时间复杂度为 $O(\tau(3n_v n^3) + O(n_v n^2))$。由于方法 1 的时间复杂度远小于方法 2, 故 AIM 方法的总时间复杂度为 $O(\tau(3n_v n^3))$。

3　实验与分析

3.1　实验设置

本文实验环境为 AMD Ryzen 7 5800H with Radeon Graphics@ 3.20 GHz, 编程环境为 MATLAB 2019a, 操作系统为 Windows11。

选取 6 个多视角数据集为基准数据集, 包括 2 个真实不完备多视角文本数据集 BBCSport 和 3Sources, 1 个合成二维数据集 Two-Moon, 以及 3 个完备多视角图像数据集 ORL、100Leaves 和 Scene-15[24]。对 ORL、100Leaves 和 Scene-15 这 3 个完备数据集按缺失率进行随机删除, 得到仿真不完备多视角数据集, 缺失率分别为 0、0.1、0.3、0.5、0.7、0.9。数据集详细信息如表 1 所示。表中: d_v 表示各视角特征维数。

表 1　实验数据集描述

Table 1　Description of experimental dataset

数据集	n_v	c	n	d_v	类型
BBCSport	2	5	737	3 183/3 203	运动文档
3Sources	3	6	416	3 560/3 631/3 068	新闻
Two-Moon	2	2	200	200/200	
ORL	4	40	400	256/256/256/256	人脸图像
100Leaves	3	100	1 600	64/64/64	叶片图像
Scene-15	3	15	4 485	20/59/40	场景图像

本文选取以下 9 个方法与 AIM 方法进行对比实验: 最佳单视角聚类(best single view, BSV)[25]、部分多视角聚类(partial multi-view clustering, PVC)[15]、不完全多视角聚类(multi-incomplete-view clustering, MIC)[16]、双对齐不完全多视角聚类(doubly aligned incomplete multi-view clustering, DAIMC)[26]、在线多视角聚类(online multi-view clustering, OMVC)[18]、基于张量补全的不完全多视图聚类(tensor completion-

based incomplete multiview clustering, TCIMC)[27]、自适应图学习的不完全多视图谱聚类(incomplete multiview spectral clustering with adaptive graph learning, IMSC_AGL)[6]、张量化拓扑图学习用于广义不完全多视图聚类(tensorized topological graph learning for generalized incomplete multi-view clustering, TTGL)[28]和基于高置信局部结构的一致性图学习用于不完全多视图聚类(highly confident local structure based consensus graph learning for incomplete multi-view clustering, HCLS_CGL)[29]。其中,BSV为最佳视角上的谱聚类;DAIMC利用加权半非负矩阵分解(non-negative matrix factorization, NMF)学习各视角非负一致低维表示;MIC利用加权NMF学习一个$l_{2,1}$正则化的潜在子空间;PVC采用NMF技术学习潜在表示矩阵;OMVC给缺失实例动态赋予较低的权重,最小化缺失数据的负面影;TCIMC采用基于张量Schatten-p范数补全缺失值;IMSC_AGL通过捕获低秩结构学习一致低维表征;TTGL引入张量结构捕获视角间的高阶相似性信息,构造用于不完备多视角聚类的拓扑相似一致图;HCLS_CGL提出新的置信图结构,帮助模型获得高质量的共识图。对所有方法,由于k-Means对初始值有一定依赖性,将实验运行30次k-Means,并记录平均值。

3.2 评价指标

本文使用精度及归一化互信息衡量聚类结果。2个指标数值越大,聚类效果越好。

$$A_{\mathrm{acc}}=\frac{T_{\mathrm{P}}+T_{\mathrm{N}}}{T_{\mathrm{P}}+T_{\mathrm{N}}+F_{\mathrm{P}}+F_{\mathrm{N}}} \tag{46}$$

式中:T_{P}表示真阳性;F_{P}表示假阳性;T_{N}表示真阴性;F_{N}表示假阴性。

$$N_{\mathrm{nmi}}(X,Y)=2\frac{I(X,Y)}{H(X)+H(Y)} \tag{47}$$

式中:互信息$I(X,Y)$定义为$\sum_{x}\sum_{y}p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)}$,$p(x,y)$为$X$与$Y$的联合概率分布,$p(x)$为$x$的边际概率分布;$H(X)=-\sum_{i}p(x_i)\log_2 p(x_i)$为信息熵。

3.3 实验结果分析

3.3.1 合成数据集Two-Moon上的结果

本节使用合成的Two-Moon数据集评估每个方法的性能,进行一致性表示的可视化对比实验。该数据集包含2个视角,由2个簇构成,每个簇包含200个数据点,数据分布呈现为2个弯月形状。为验证方法对噪声的鲁棒性及缺失值填充策略的有效性,在数据集中随机添加12%的高斯噪声,并随机删除30%的实例。图2为可视化后的Two-Moon原始数据集和各方法得到的一致表征。为更清晰地进行可视化,本文将属于同一个簇的数据点标记为相同的符号和颜色,实心表示可用实例,空心表示缺失实例。

图2(a)和图2(b)分别表示各视角的原始数据点。图2(c)~图2(k)为不同方法从所有视角的原始数据中学习到的一致表征的可视化展示。如图2(c)、图2(e)、图2(f)、图2(h)、和图2(j)所示,DAIMC、TCIMC、PVC、IMSC_AGL和HCLS_CGL方法所学习到的一致表征分别呈现椭圆、直线、扇形、环形和折线形状,因此,这些方法不能有效划分数据点,表2也显示这几种方法聚类精度均不超过80%。图2(d)展示的可视化图为空,这是因为MIC方法产生的一致低维表征在迭代中被全部置为空。OMVC方法受限于NMF方法无法处理负向数据的特点,所得到的一致低维表征有很多数据点位于坐标轴上,如图2(g)所示,并且表2也显示其聚类精度仅为65.50%。如图2(i)所示,TTGL方法通过张量结构将数据点分散在坐标轴附近,难以达到好的聚类效果,表2展示其聚类精度仅为84.00%。

如图2(k)所示,本文提出的AIM方法将数据点划分为上下2个组,且大多数数据点都位于正确的集群上。这些现象表明,AIM方法可以充分挖掘视角间的互补信息,进而从原始的不完备多视角数据中学习到判别性更强且更紧凑的一致表征。从表2也可以看出,AIM方法取得了最佳的精度和归一化互信息。

3.3.2 不完备多视角数据集上的结果

表2展示了AIM方法及对比方法在2个真实不完备多视角数据集BBCSport和3Sourse上的实验结果,表3和图3展示了AIM方法及对比方法在3个仿真不完备图像多视角数据集ORL、100Leaves和Scene-15上的实验结果。通过对实验结果进行分析,可以得出如下结论:

1)从表2、表3和图3的实验结果可以看出,BSV方法在大多数数据集中取得相对最差的结果,PVC方法次之。这是因为BSV方法仅使用一个视角,放弃了其他视角的特有信息及视角间的互补信息,PVC方法则是只能学习2个视角的一致表征,对于超过2个视角的数据集,也只选用最好的2个视角。而DAIMC、MIC、OMVC、TCIMC、IMSC_AGL、TTGL、HCLS_CGL及AIM方法均在所有视角中寻求一致表征,通常能获得更好的性能。这说明利用更多视角的聚类技术可以提高多视角数据的聚类准确性,同时学习一致表征图的方法能够充分利用不完备多视角数据的互补信息。

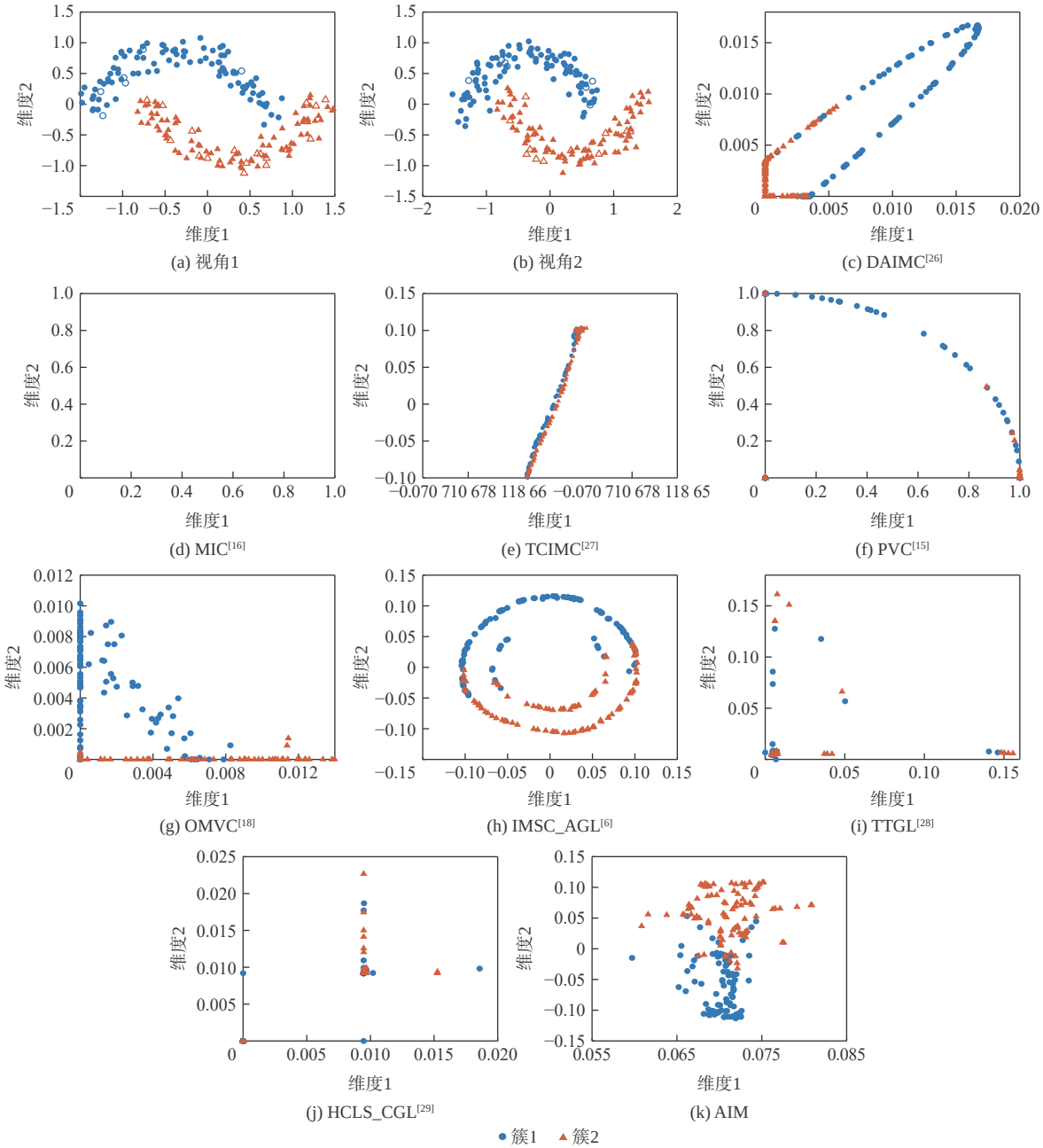


图 2　不同方法在 Two-Moon 数据集上可视化的一致表示

Fig. 2　Consistent representation of different methods for Two-Moon dataset visualization

表 2　不同方法在 BBCSport、3Sources 和 Two-Moon 数据集上的精度和归一化互信息

Table 2　Accuracy and NMI of different algorithms on BBCSport, 3Sources, and Two-Moon datasets

方法	精度/%			归一化互信息		
	Two-Moon	BBCSport	3Sources	Two-Moon	BBCSport	3Sources
BSV[25]	79.50	40.99	22.50	0.270 1	0.266 0	0.053 0
DAIMC[26]	73.00	75.16	58.68	0.220 3	0.578 0	0.488 0
MIC[16]	50.50	58.66	43.73	0.005 0	0.462 6	0.389 4
PVC[15]	76.50	44.33	26.45	0.260 6	0.137 7	0.137 7
TCIMC[27]	56.00	67.44	77.16	0.010 4	0.594 9	0.676 3
OMVC[18]	65.50	57.94	49.76	0.093 5	0.663 5	0.663 5
IMSC_AGL[6]	75.50	88.05	65.86	0.180 9	0.537 6	0.619 8
TTGL[28]	84.00	84.93	77.64	0.387 9	0.635 7	0.601 1
HCLS_CGL[29]	72.50	74.76	65.63	0.151 6	0.688 9	0.572 4
AIM	91.00	89.55	86.15	0.563 5	0.709 5	0.711 6

表3 缺失率为0,0.1,0.3,0.5,0.7,0.9时，不同方法在100Leaves、ORL、Scene-15数据集上的归一化互信息

Table 3 NMI of different algorithms with PER of 0, 0.1, 0.3, 0.5, 0.7, 0.9 on 100Leaves、ORL、Scene-15 datasets

数据集	缺失率	归一化互信息									
		BSV[25]	DAIMC[26]	MIC[16]	OMVC[18]	TCIMC[27]	PVC[15]	IMSC_AGL[6]	TTGL[28]	HCLS_CGL[29]	AIM
100Leaves	0	0.541 7	0.875 4	0.848 8	0.235 0	0.893 7	0.660 0	0.935 5	0.785 6	0.883 0	0.953 6
	0.1	0.536 3	0.840 6	0.817 0	0.565 9	0.886 0	0.582 8	0.924 7	0.750 3	0.873 3	0.930 3
	0.3	0.474 7	0.762 8	0.741 7	0.495 2	0.690 1	0.529 4	0.883 4	0.732 0	0.810 4	0.903 5
	0.5	0.435 9	0.683 3	0.673 5	0.461 5	0.698 2	0.443 1	0.838 1	0.676 8	0.798 6	0.862 5
	0.7	0.396 1	0.630 2	0.631 7	0.435 4	0.708 4	0.398 1	0.821 5	0.648 0	0.776 0	0.840 5
	0.9	0.355 2	0.581 3	0.594 4	0.470 0	0.782 8	0.360 6	0.766 3	0.601 2	0.754 0	0.787 6
ORL	0	0.134 6	0.779 3	0.786 9	0.736 4	0.860 0	0.655 0	0.835 1	0.715 1	0.838 2	0.880 3
	0.1	0.430 7	0.769 1	0.751 5	0.691 0	0.859 5	0.570 0	0.791 3	0.690 2	0.824 4	0.873 9
	0.3	0.394 6	0.734 4	0.633 1	0.648 3	0.823 6	0.460 0	0.770 6	0.650 8	0.793 2	0.857 2
	0.5	0.445 5	0.709 6	0.576 1	0.604 9	0.825 2	0.355 0	0.752 4	0.578 7	0.761 0	0.845 0
	0.7	0.464 3	0.656 5	0.491 0	0.568 1	0.790 8	0.322 5	0.722 7	0.522 0	0.738 4	0.827 1
	0.9	0.470 3	0.616 2	0.423 3	0.556 3	0.770 5	0.205 0	0.691 1	0.505 4	0.700 3	0.782 3
Scene-15	0	0.007 4	0.304 8	0.344 0	0.326 2	0.215 8	0.333 6	0.420 3	0.391 4	0.399 9	0.463 4
	0.1	0.005 8	0.305 3	0.324 1	0.277 2	0.404 5	0.346 3	0.399 4	0.376 3	0.389 1	0.448 2
	0.3	0.007 6	0.279 1	0.287 4	0.254 0	0.410 2	0.275 6	0.385 8	0.333 6	0.362 6	0.431 5
	0.5	0.007 5	0.246 8	0.250 7	0.218 0	0.399 4	0.253 7	0.372 0	0.311 0	0.295 9	0.402 0
	0.7	0.007 9	0.209 2	0.211 6	0.235 6	0.366 0	0.260 4	0.353 3	0.264 7	0.321 0	0.405 3
	0.9	0.008 8	0.183 6	0.192 1	0.219 8	0.347 1	0.212 0	0.338 2	0.250 0	0.307 4	0.367 7

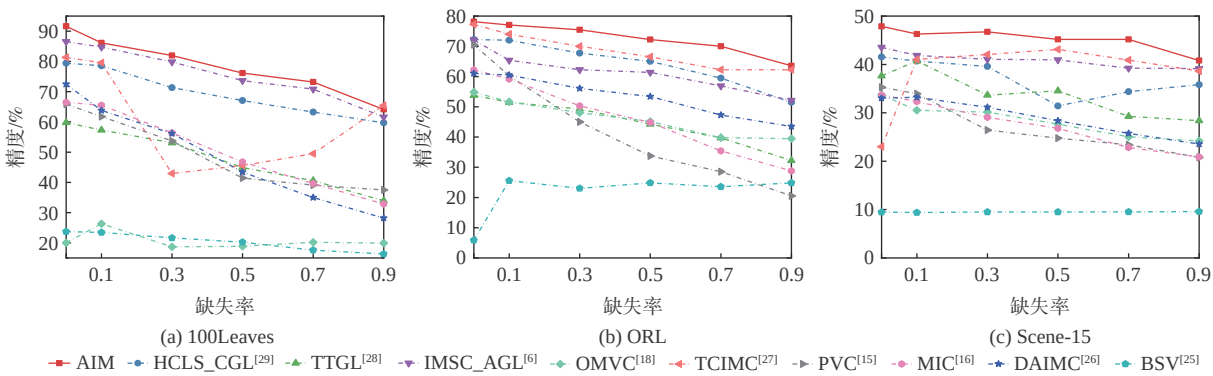


图3 缺失率为0,0.1,0.3,0.5,0.7,0.9时，不同方法在100Leaves、ORL 、Scene-15数据集上的精度

Fig. 3 Accuracy of different algorithms at 100 Leaves, ORL, Scene-15 datasets with PER of 0, 0.1, 0.3, 0.5, 0.7, 0.9

2）如表3所示，随着缺失率的增加，所有方法的聚类性能均显著下降，表明特征缺失对聚类结果有一定的影响。实验结果表明，DAIMC和IMSC_AGL方法的表现在大多数情况下优于MIC和OMVC方法，这是因为MIC和OMVC方法使用的均值填充策略带来的噪声随着缺失率的提高而增加，导致聚类性能降低。DAIMC和IMSC_AGL方法直接将缺失值填充为0，HCLS_CGL方法仅从可用实例中的某些相似信息中学习共识图，虽然这相对于均值填充策略缓解了缺失实例对一致表征的负面影响，但这种填充方法直接放弃了缺失值的结构信息。TTGL方法通过训练已知样本之间的关系逼近预定义的结构信息，估算缺失值之间的相似度。这种迭代补全缺失值的方法在一定程度上保证了缺失样本的特征，但实验结果表明，这种方法带来的负面影响也不可忽视。实验结果还显示，AIM方法的表现优于以上所有方法，表明本文的补全方案在处理缺失实例时表现优异，既避免了缺失实例的负面影响，又补充了其结构信息。

3）为更加直观地展现各聚类方法的有效性，图4展示了不同方法在缺失率0.3时的ORL人脸数据集上的混淆矩阵，其中，行和列分别表示每个

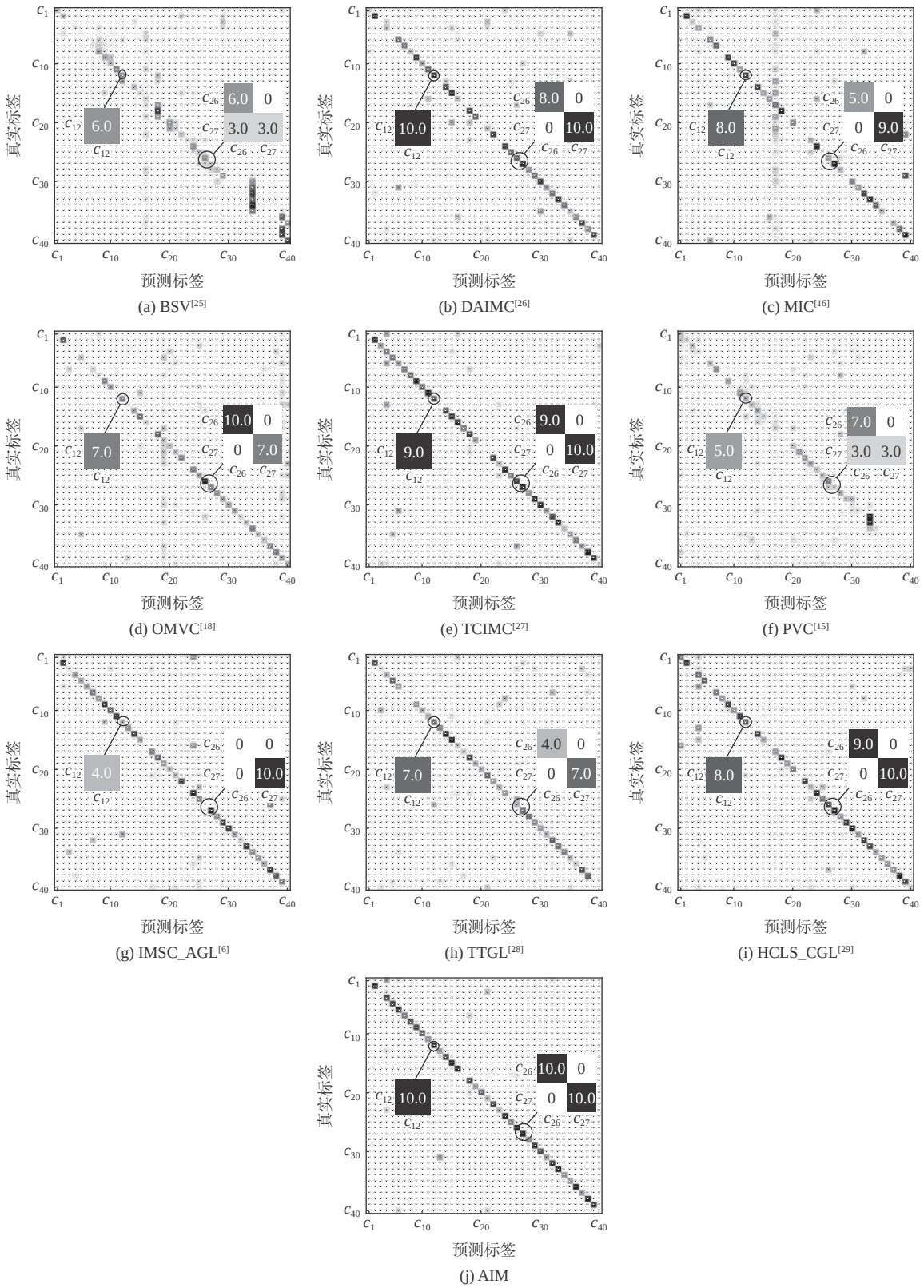


图 4　缺失率为 0.3 时, 不同方法在 ORL 数据集上的混淆矩阵

Fig. 4　Confusion matrix of different methods on ORL dataset with PER = 0.3

真实簇中数据点的真实标签集合和预测标签集合。实验使用精度的最优排列映射函数来计算矩阵中的每个元素。如果混淆矩阵对角线上真实标签集合和预测标签集合所对应的块颜色深, 并且数字接近 10, 说明对应方法对该簇的识别效果越好, 非对角线上的块颜色浅, 并且数字接近 0, 说明对应方法对该簇的识别效果越差。如图 4 中放大的 3 个对角块所示, AIM 方法对应的混淆矩阵中, 该 3 个对角块中的数字均达到 10, 而所有对比方法相应的数字均低于 10, 说明 AIM 方法对这 3 个簇的

识别优于对比方法。同时，从图4还可以看出，AIM方法的混淆矩阵中，非对角线区域内的浅色块数量明显少于其余混淆矩阵，故AIM方法对大多数簇的识别效果都优于对比方法。

4）AIM方法在这些数据集上取得最好的聚类结果，验证了其在不完备多视角聚类上的有效性。

3.4 参数敏感性分析

实验以缺失率为0.3的ORL人脸数据集和缺失率为0.5的100Leaves叶片数据集为例，分析AIM方法中近邻数k及3个惩罚项参数λ_1、λ_2和λ_3对聚类性能的影响。由于篇幅所限，本文只基于归一化互信息进行分析。

图5展示了固定参数λ_1、λ_2和λ_3时，近邻数k的变化对聚类效果的影响。可以看出，当k取不同值时，归一化互信息波动较小，即AIM方法对近邻数k鲁棒。

固定k研究惩罚项参数λ_1、λ_2和λ_3对聚类性能的影响。这3个参数在模型中具有重要作用，其中，λ_1用于调节低秩约束项，λ_2用于调节误差项，λ_3用于调节视角融合约束项。本文使用的参数λ_1、λ_2和λ_3范围均为$\{10^{-5},10^{-4},10^{-3},10^{-2},10^{1},10^{2},10^{3},10^{4}\}$。首先，固定参数$\lambda_2$，得到AIM方法在$\lambda_1$和$\lambda_3$取不同值时的归一化互信息变化曲线。如图6(a)和图7(a)所示，可以发现，当$\lambda_1 \leqslant 10^{-2}$时，$\lambda_3$取任何值，2个数据集上的聚类结果都较为稳定，因此，AIM方法对参数λ_3不敏感，对λ_1敏感。这是因为参数λ_1为调节低秩约束项的惩罚项参数，在模型学习中占据主导地位，从而导致参数λ_1的取值对聚类效果的影响较大。固定λ_1和λ_3，得到AIM方法在λ_2取不同值时的归一化互信息变化曲线。从图6(b)和图7(b)可以看出，当$\lambda_2 \geqslant 10^{1}$时，在ORL人脸数据集上的结果趋向平稳。而$\lambda_2$的取值对100Leaves叶片数据集没有较大影响。这是因为λ_2为误差约束项，而人脸数据结构复杂，此时较小的λ_2会导致重构误差增大，不利于得到好的聚类结果。综上所述，AIM方法对近邻数k和参数λ_3不敏感，对参数λ_1和λ_2敏感。

综上，AIM方法对近邻数k和参数λ_3不敏感，对参数λ_1和λ_2敏感。

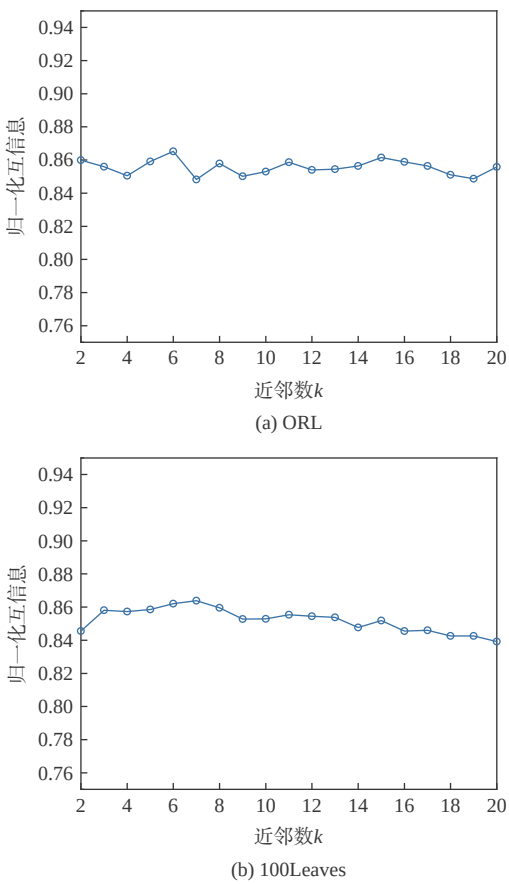


图5 在缺失率为0.3的ORL数据集和缺失率为0.5的100Leaves数据集上，AIM方法的归一化互信息与近邻数k的关系

Fig. 5 Relationship between NMI and nearest neighbor k of AIM algorithm on ORL dataset with PER = 0.3 and 100Leaves dataset with PER = 0.5

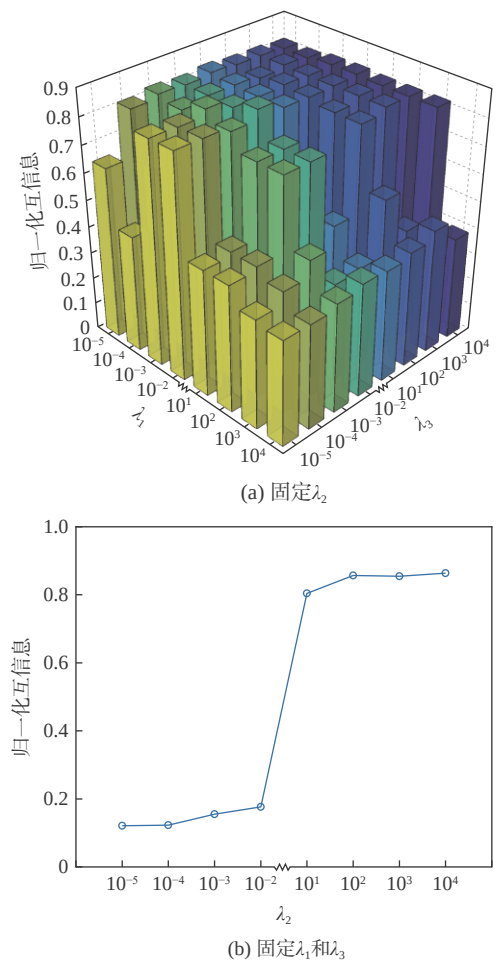
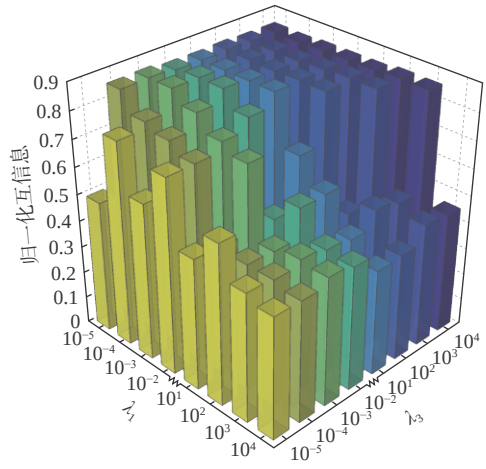
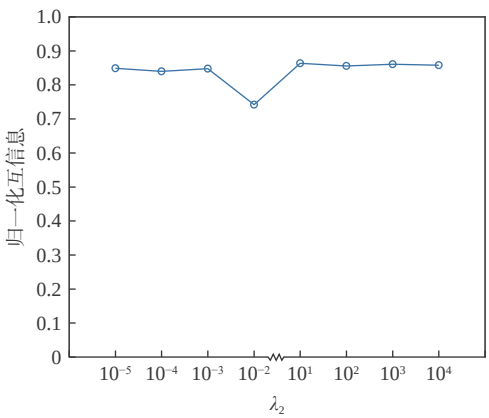


图6 缺失率为0.3时，ORL数据集上归一化互信息与参数λ_1、λ_2和λ_3的变化关系

Fig. 6 Relationship between NMI and parameters λ_1, λ_2, and λ_3 on ORL dataset with PER = 0.3



(a) 固定λ_2



(b) 固定λ_1和λ_3

图 7　缺失率为 0.5 时, 100Leaves 数据集上归一化互信息与参数λ_1、λ_2和λ_3的变化关系

Fig. 7　Relationship between NMI and parameters λ_1, λ_2, and λ_3 on 100Leaves dataset with PER = 0.5

3.5　收敛性分析

实验以缺失率为 0.3 的 ORL 人脸数据集和缺失率为 0.5 的 100Leaves 叶片数据集为例, 验证 AIM 方法的收敛性。图 8(a) 和图 8(b) 分别展示了在 ORL 和 100Leaves 数据集上目标函数值与迭代步数之间的关系。其中, 目标函数值的计算式为

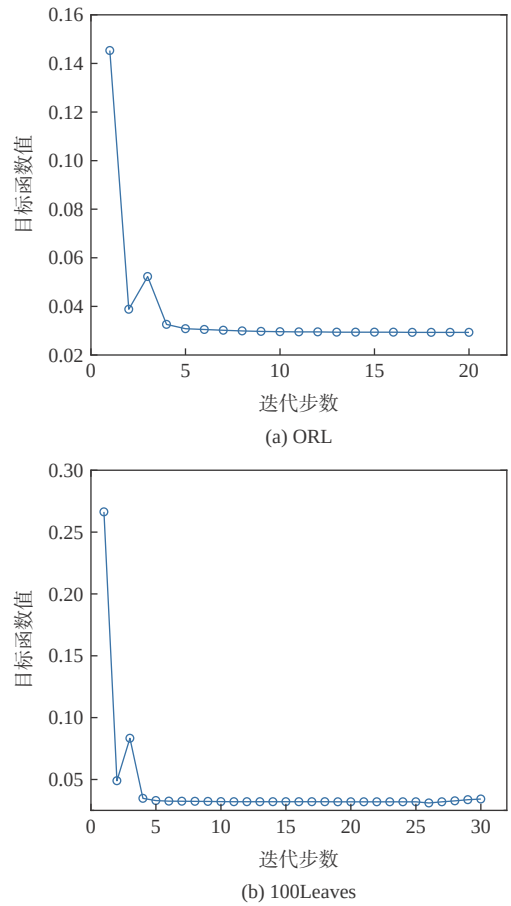
$$o=\frac{\sum_{v=1}^{k}\left(\gamma_1^{(v)}+\gamma_2^{(v)}\right)}{\sum_{v=1}^{k}\left\|\boldsymbol{X}^{(v)}\right\|_F^2} \tag{48}$$

式中:

$$\gamma_1^{(v)}=\left\|\boldsymbol{Z}^{(v)}\right\|_*+\lambda_1\mathrm{Tr}(\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{L}^{(v)}\boldsymbol{F}^{(v)})+\lambda_2\left\|\boldsymbol{E}^{(v)}\right\|_{2,1}+\lambda_3(c_4-\mathrm{Tr}(\boldsymbol{F}^{(v)}\boldsymbol{F}^{(v)\mathrm{T}}\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}))$$

$$\gamma_2^{(v)}=\left\|\boldsymbol{X}^{(v)}-\boldsymbol{X}^{(v)}\boldsymbol{Z}^{(v)}-\boldsymbol{E}^{(v)}\right\|_F^2+\left\|\boldsymbol{Z}^{(v)}-\boldsymbol{P}^{(v)}\right\|_F^2+\left\|\boldsymbol{Z}^{(v)}-\boldsymbol{Q}^{(v)}\right\|_F^2$$

从图 8 可以看出, 目标函数值随着迭代步数增加先急剧下降后马上趋于平稳, 表明 AIM 方法能很快达到收敛, 具有良好的收敛性。



(a) ORL

(b) 100Leaves

图 8　在缺失率为 0.3 的 ORL 数据集和缺失率为 0.5 的 100Leaves 数据集上, 目标函数值与迭代步数之间的变化关系

Fig. 8　Relationship between object function value and iteration steps on ORL dataset with PER = 0.3 and 100Leaves dataset with PER = 0.5

4　结　论

1) 本文为自适应构图模型引入稀疏约束项, 再利用数据驱动模型, 使 2 个样本间的相似度值与其距离成反比, 由此得到适合高维数据集的高质量初始图。

2) AIM 方法利用所有有效视角相似度平均值填充策略来补充缺失值, 在避免填充值对一致表征产生负面影响的同时, 补充了缺失实例的部分结构信息。

3) AIM 方法通过引入低秩约束得到数据的全局结构信息, 再通过 $l_{2,1}$ 范数约束重构无噪声数据, 通过一致性约束避免视角间由于缺失率不同而出现权重失衡的情况。

4) 通过 9 个对比方法在 6 种不同的数据集上与 AIM 方法进行对比, 证明其能够获得具有高鉴

别性的一致表征，同时能够取得良好的划分结果。

本文方法在处理大规模不完备多视角数据方面具有一定的优势，未来将继续研究，在挖掘数据完整结构信息的同时减少参数的设定。

参考文献（References）

[1] ZHAO J, XIE X J, XU X, et al. Multi-view learning overview: recent progress and new challenges[J]. Information Fusion, 2017, 38: 43-54.

[2] 谢娟英, 丁丽娟. 完全自适应的谱聚类算法[J]. 电子学报, 2019, 47(5): 1000-1008.
XIE J Y, DING L J. The true self-adaptive spectral clustering algorithms[J]. Acta Electronica Sinica, 2019, 47(5): 1000-1008(in Chinese).

[3] XIA K J, GU X Q, ZHANG Y D. Oriented grouping-constrained spectral clustering for medical imaging segmentation[J]. Multimedia Systems, 2020, 26(1): 27-36.

[4] SHARMA K K, SEAL A. Multi-view spectral clustering for uncertain objects[J]. Information Sciences, 2021, 547: 723-745.

[5] GAO H C, NIE F P, LI X L, et al. Multi-view subspace clustering [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 4238-4246.

[6] WEN J, XU Y, LIU H. Incomplete multiview spectral clustering with adaptive graph learning[J]. IEEE Transactions on Cybernetics, 2020, 50(4): 1418-1429.

[7] NISHOM M. Perbandingan akurasi Euclidean distance, Minkowski distance, Dan Manhattan distance pada algoritma K-means clustering berbasis Chi-square[J]. Jurnal Informatika: Jurnal Pengembangan IT, 2019, 4(1): 20-24.

[8] YIN J, SUN S L. Incomplete multi-view clustering with cosine similarity[J]. Pattern Recognition, 2022, 123: 108371.

[9] DING X J, LIU J, YANG F, et al. Random compact Gaussian kernel: application to ELM classification and regression[J]. Knowledge-Based Systems, 2021, 217: 106848.

[10] SHANG R H, ZHANG Z, JIAO L C, et al. Self-representation based dual-graph regularized feature selection clustering[J]. Neurocomputing, 2016, 171: 1242-1253.

[11] WENG L B, DORNAIKA F, JIN Z. Graph construction based on data self-representativeness and Laplacian smoothness[J]. Neurocomputing, 2016, 207: 476-487.

[12] WANG Y, ZHANG W J, WU L, et al. Iterative views agreement: an iterative low-rank based structured optimization method to multi-view spectral clustering[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York: ACM, 2016: 2153-2159.

[13] WANG Y, WU L, LIN X M, et al. Multiview spectral clustering via structured low-rank matrix factorization[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(10): 4833-4843.

[14] ZHANG P, LIU X W, XIONG J, et al. Consensus one-step multi-view subspace clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(10): 4676-4689.

[15] LI S Y, JIANG Y, ZHOU Z H. Partial multi-view clustering[C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2014: 1968-1974.

[16] SHAO W, HE L. Multiple incomplete views clustering via weighted nonnegative matrix factorization with $l_{2,1}$ regularization[C]//Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Berlin: Springer, 2017: 318-334.

[17] GAO H, PENG Y X, JIAN S L. Incomplete multi-view clustering [C]//Proceedings of the 9th IFIP TC 12 International Conference on Intelligent Information Processing VIII. Berlin: Springer, 2016: 245-255.

[18] SHAO W X, HE L F, LU C T, et al. Online multi-view clustering with incomplete views[C]//Proceedings of the IEEE International Conference on Big Data. Piscataway: IEEE Press, 2016: 1012-1017.

[19] CHEN J, YANG S X, PENG X, et al. Augmented sparse representation for incomplete multiview clustering[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(3): 4058-4071.

[20] LIU X W, LI M M, TANG C, et al. Efficient and effective regularized incomplete multi-view clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(8): 2634-2646.

[21] TRIVEDI A, RAI P, DAUMÉ H. Multiview clustering with incomplete views[C]//Proceedings of the Conference and Workshop on Neural Information Processing Systems workshop. Cambridge: MIT Press, 2010: 1-8.

[22] LIU G C, LIN Z C, YAN S C, et al. Robust recovery of subspace structures by low-rank representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 171-184.

[23] CAI J F, CANDÈS E J, SHEN Z W. A singular value thresholding algorithm for matrix completion[J]. SIAM Journal on Optimization, 2010, 20(4): 1956-1982.

[24] XIAO X L, GONG Y J, HUA Z Y, et al. On reliable multi-view affinity learning for subspace clustering[J]. IEEE Transactions on Multimedia, 2020, 23: 4555-4566.

[25] ZHAO H D, LIU H F, FU Y. Incomplete multi-modal visual data grouping[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York: ACM, 2016: 2392-2398.

[26] HU M L, CHEN S C. Doubly aligned incomplete multi-view clustering[EB/OL]. (2019-03-07)[2023-04-01]. https://arxiv.org/abs/1903.02785v1.

[27] XIA W, GAO Q X, WANG Q Q, et al. Tensor completion-based incomplete multiview clustering[J]. IEEE Transactions on Cybernetics, 2022, 52(12): 13635-13644.

[28] ZHANG Z, HE W J. Tensorized topological graph learning for generalized incomplete multi-view clustering[J]. Information Fusion, 2023, 100: 101914.

[29] WEN J, LIU C L, XU G H, et al. Highly confident local structure based consensus graph learning for incomplete multi-view clustering[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 15712-15721.

Adaptive incomplete multi-view clustering

CHEN Mei*，MA Xueyan，ZHANG Chi，ZHANG Jinhong，QIAN Luoxiong

(School of Electronic and Information Engineering，Lanzhou Jiaotong University，Lanzhou 730070，China)

Abstract：A high-quality complete initial graph can effectively improve the performance of incomplete multi-view clustering. However, inappropriate filling of missing values will lead to the initial graph losing the underlying structure of the data, and incomplete fusion of affine graphs of each view will make the unified learned representations miss the complementary information among the views. To address the aforementioned problems, an adaptive incomplete multi-view clustering (AIM) method was proposed in this paper. In the initial graph construction, AIM used the average value of similarity of valid views to fill the missing values at corresponding positions to obtain a complete potential structure of the data and introduced sparsity constraints to improve the robustness of the model to noise. In the graph optimization process, initially, low-rank constraints were introduced to capture the global structure of the data, followed by spectral constraints to enhance the closeness between data within classes to make the affine graph have a clearer block diagonal structure. The consistency constraints were introduced to minimize the differences between the affine graph and the unified representation of each view to capture the complementary information between the views. Ultimately, a unified robust representation graph with high discriminative features was obtained. The experimental comparisons with nine kinds of incomplete multi-view clustering in real and incomplete multi-view datasets simulated under multiple missing rates demonstrate that AIM obtains the best clustering performance.

Keywords：adaptive graph construction; low-rank representation; graph fusion; graph algorithm; incomplete multi-view clustering

Received：2023-04-14；**Accepted**：2023-09-14；**Published Online**：2023-10-20 09：01

URL：link.cnki.net/urlid/11.2625.V.20231019.1646.004

Foundation items：National Natural Science Foundation of China (62266029); Gansu Key Research and Development Program (21YF5GA053); Gansu Higher Education Industry Support Program (2022CYZC-36)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0208

基于 RISC-V 向量扩展的图像预处理加速方法

刘强[1, 2, *]，尹蔚[1, 2]，李凯[1, 2]

(1. 天津大学 微电子学院，天津 300072；2. 天津市成像与感知微电子技术重点实验室，天津 300072)

摘 要：作为卷积神经网络 (CNN) 计算的前序步骤，图像预处理不可或缺又非常耗时。为加速图像预处理，提出一种基于 RISC-V 向量扩展的加速方法，对灰度化、标准化、高斯滤波等 11 种图像预处理算法进行加速。从计算模式上将 11 种图像预处理算法归为 4 类，并基于 RISC-V 向量扩展对各类图像预处理算法设计了加速方案；为进一步提高性能，新增 6 条自定义的向量指令，并通过修改编译器和设计硬件模块实现了 6 条自定义向量指令；使用现场可编程门阵列 (FPGA) 进行测试，并分析了向量处理器配置对性能和资源消耗的影响。结果显示：所提方法相比标量处理器实现了 3.13～9.97 倍的加速效果，可有效解决图像预处理在深度学习过程中的性能瓶颈问题。

关 键 词：卷积神经网络；预处理；RISC-V；向量扩展；算法加速
中图分类号：TN402
文献标志码：A **文章编号**：1001-5965(2025)04-1074-11

当前，卷积神经网络 (convolutional neural network, CNN) 广泛应用于图像分类、人脸识别等领域。作为 CNN 计算的前序步骤，图像预处理在扩充训练数据集、降低 CNN 推理或训练误差等方面均有重要意义。Pal 等[1] 调研了归一化、标准化和零分量分析 (zero component analysis, ZCA) 对 3 种不同结构的 CNN 识别准确率的影响，结果表明，经过预处理的数据识别准确率相比原始数据提升约 10%。Şaban 和 Akdemir[2] 研究了归一化、中值滤波和维纳滤波等图像预处理算法对病理学图像识别的影响，结果显示，预处理使 CNN 训练收敛速度更快，并使推理准确率提升约 1.7%。多项研究表明，高效的图像预处理算法可有效改善 CNN 的性能[3-4]。

CNN 常用的图像预处理算法从应用角度可以分为 3 类：①统一图像或数据格式，以满足 CNN 的输入要求，如标准化、归一化、图像缩放、通道转换等；②扩充 CNN 训练数据集，在许多应用场景中，供 CNN 训练的数据集往往是不充分的，预处理能够有效增加 CNN 训练的样本量，从而提高 CNN 的泛化能力，常见的扩充数据集的图像预处理算法有旋转、镜像、平移、裁剪、亮度调整、对比度调整、添加随机噪声等；③图像增强，目的是提高推理速度和准确率，常见的方法包括灰度化、二值化、各类滤波算法、图像分割、边缘检测、锐化 (边缘增强) 等。

在嵌入式应用环境中，CPU+CNN 加速器的异构组合是常见的深度学习计算平台。在该异构平台中，CPU 负责图像的预处理和加速器的控制，加速器负责 CNN 的计算。在先前的研究中，大量的工作集中在 CNN 推理或训练部分的优化[5-6]。然而，在嵌入式平台中，由于 CPU 并行度低，性能不佳，图像预处理也会成为瓶颈。在多项研究[7-9] 中，图像预处理占据了 60% 以上的计算时间。因此，加速图像预处理成为了进一步提升性能的关键。

为解决这一问题，NVIDIA 发布了图像预处理计算库 NVIDIA Data Loading Library (DALI)[10]，该

收稿日期：2023-04-24；**录用日期**：2023-05-26；**网络出版时间**：2023-06-09 10:13
网络出版地址：link.cnki.net/urlid/11.2625.V.20230608.1155.002
基金项目：国家自然科学基金 (U21B2031)
* **通信作者**. E-mail：qiangliu@tju.edu.cn

引用格式：刘强，尹蔚，李凯. 基于 RISC-V 向量扩展的图像预处理加速方法 [J]. 北京航空航天大学学报，2025，51（4）：1074-1084.
LIU Q，YIN W，LI K. Image preprocessing acceleration method based on RISC-V vector extension[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1074-1084（in Chinese）.

计算库能够高效地进行 CNN 计算中的图像解码和预处理等操作, 其加速原理是将在 CPU 上进行的图像预处理工作部署到图形处理单元 (graphics processing unit, GPU) 上, 但是, DALI 仅能被应用在 CPU+GPU 的深度学习平台, 不适用于 CPU+CNN 加速器的嵌入式深度学习环境。Ma[11] 通过高层次综合 (high level synthesis, HLS) 的方法将灰度化、图像缩放和显著性分析算法固化成 IP 核, 相比 ARM CPU 实现了约 20 倍的加速效果。

RISC-V 是一种开源的 CPU 指令集架构 (instruction set architecture, ISA)。除了标准的整数指令集外, RISC-V 还针对不同的应用需求定义了扩展指令集, 如嵌入式扩展 (embedded extension)、位操作扩展 (bit manipulation extension)、向量扩展 (vector extension)、加密扩展 (cryptographic extension) 等。此外, RISC-V 还支持用户自定义指令集。RISC-V 指令集扩展在算法加速、硬件安全等领域表现出优秀的性能和良好的灵活性。Kuo 等[12] 扩展了一组迦瓦罗域的 RISC-V 计算指令, 对非二进制纠错码和后量子密码算法进行加速, 实现了约 5 倍的加速效果, 同时硬件逻辑资源消耗仅增加 1.27%; Razilov 等[13] 使用高性能 RISC-V 向量处理器 Ara[14] 和 RISC-V 向量扩展 (RISC-V vector extension, RVV) 指令来加速广义频分复用 (generalized frequency division multiplexing, GFDM) 算法, 实现了约 60 倍的加速效果; 刘强和李一可[15] 基于 RISC-V 指令扩展提出了一种可配置的故障注入检测方法, 支持时间冗余和信息冗余 2 种检测模式, 相比单一的信息冗余检测方法, 故障检测率提高了 13.34%, 仅引入 4.4% 的资源开销。

RVV 是一种单指令多数据流 (single instruction multiple data, SIMD) 的技术。RVV1.0[16] 是 RVV 的技术规范, 在 2021 年被批准。支持 RVV 的 RISC-V 处理器有用于实时和嵌入式计算的向量处理器 Vicuna[17]、64 位高性能向量处理器 Ara[14] 等。

由于不同网络结构和应用场景所需的图像预处理算法不同, 图像预处理难以被设计成专用集成电路 (application specific integrated circuit, ASIC) 进行加速。同时, 图像预处理算法对并行度要求高, 通用的嵌入式 CPU 往往并行计算能力较差。RVV 能够增强 RISC-V CPU 的并行计算能力, 还可以通过自定义向量指令实现对特定应用的定制化加速。因此, 本文采用 RVV 方案对图像预处理进行加速。

使用 RISC-V 向量处理器加速图像预处理算法面临以下 2 个问题: ①图像预处理算法种类多, 难以设计通用加速方案; ②现有标准 RISC-V 指令集不能完全满足图像预处理加速的需求。

为解决上述问题, 本文面向常用的 CNN 图像预处理算法提出了一种基于 RVV 的加速方案。

1) 针对标准化、缩放、旋转、镜像、平移、亮度调整、灰度化、二值化、高斯滤波、拉普拉斯滤波、索贝尔边缘检测 11 种常用的 CNN 图像预处理算法, 按照计算模式分为 4 类, 对每类算法设计了加速方案。

2) 在标准向量扩展的基础上新增了 6 条自定义向量指令, 用于进一步加速图像预处理算法中的操作。为实现这些指令, 对编译器进行修改, 并设计了相应的译码和计算模块。

3) 使用现场可编程门阵列 (field programmable gate array, FPGA) 对本文方法和硬件架构进行验证和评估, 实验结果显示, 与标量处理器相比, 本文设计的加速方法能够实现 3.13～9.97 倍的加速。

1　硬件架构

在嵌入式深度学习应用中, 硬件资源和功耗往往是受限的。因此, 本文采用轻量级处理器 ibex[18]+向量协处理器 Vicuna[17] 的组合来开展研究, 如图 1 所示。其中, ibex 是 32 位的标量主处理器, 设有 2 级流水线, 支持 RV32IEMCB 指令集, 该处理器核的优点是可配置且轻量化, 适合嵌入式控制应用。Vicuna 是 32 位整数向量协处理器, 支持 8、16、32 位的向量元素宽度, 其遵循 RVV1.0 规范, 实现了 Zve32x 指令集 (向量扩展的一个子集)。

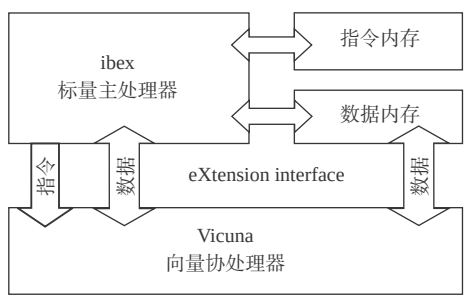


图 1　硬件平台架构

Fig. 1　Hardware architecture

ibex 和 Vicuna 通过 eXtension interface[19] 相互连接。当取到标量指令时, 由主处理器 ibex 执行; 当取到向量指令时, ibex 会将该指令通过 eXtension interface 发送至 Vicuna 执行。Vicuna 在执行向量指令时, 能够直接从数据内存中读取数据到向量寄存器文件 (vector register file, VRF) 中, 并在计算完成后将结果写回数据内存。在此期间, 如果存在数据依赖, 则 ibex 会停滞流水线直到 Vicuna 完成写

回，如果不存在数据依赖，则 ibex 可以和 Vicuna 同时运行。

2 图像预处理算法加速方案设计

由于卷积运算的独特性，本文按照是否涉及卷积运算，将 11 种图像预处理算法分为非卷积类和卷积类。卷积类算法包括高斯滤波、拉普拉斯滤波和索贝尔边缘检测。非卷积类算法按照特点又可以分为 3 类，分别为像素位置变化类、像素数值变化类及求全局平均值类。其中，像素位置变化类包括旋转、镜像、平移和缩放，像素数值变化类包括灰度化、二值化和亮度调整，求全局平均值类包括标准化。

2.1 非卷积类算法

2.1.1 像素位置变化类算法

涉及像素位置变化的算法包括旋转、镜像、平移和缩放。其中，旋转和缩放采用最近邻插值法，其他插值方法（如线性插值法）由于需要单独计算每个像素点数值的变化，不适合进行向量化。

这类算法的特点是：输出图像与输入图像之间，仅像素值的存储地址发生了变化，像素值本身并未发生变化。本文方法使用 RVV 中的索引加载 (indexload) 指令对这类算法实现定制化加速。

RVV 中的加载指令用于实现从数据内存中读取数据并加载到向量处理器的向量寄存器文件中。按照加载方式的不同，加载指令可以分为以下 3 类：①unit-strideload，该指令能够从内存中加载一组相邻的数据元素到向量寄存器中，是最常用的加载指令；②stridedload，该指令能按照一定的步长 (stride) 间隔加载一组固定间隔的数据到向量寄存器中；③indexload，即索引加载指令，该指令能够按照索引矩阵中描述的顺序加载一组数据元素到向量寄存器中。

索引加载指令能够按照既定的顺序加载数据。而涉及像素位置变化的图像预处理算法本质上是数据存储位置的变化。当使用索引加载指令实现这些算法时，原有的图像数据被按照索引矩阵描述的顺序读取到向量寄存器中重新排列，再写回数据内存，以此完成像素数据位置的变化。

如图 2 所示，以一张 3×3 大小的图像顺时针旋转 90°为例，说明索引加载指令如何实现像素位置的变化。

图 2 最上方是索引矩阵，左侧是旋转前后的图像，右侧是旋转前后的图像像素值在一维数组中存储的变化。在索引加载指令操作过程中，输入像素值按照式（1）进行重排序：

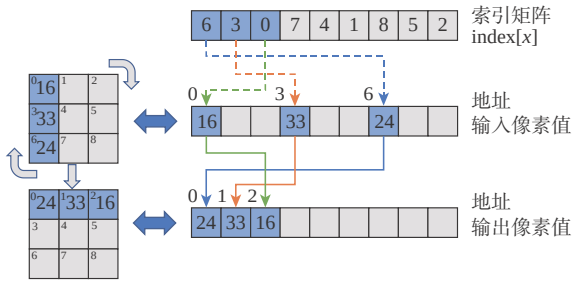


图 2 索引加载指令实现像素位置变化

Fig. 2 Pixel location change by index loading instruction

$$\text{output}[x] = \text{input}\left[\text{index}[x]\right] \tag{1}$$

例如，索引矩阵中第一个元素为 6，则向量处理器会加载输入数组地址为 6 的元素到向量寄存器地址 0 的位置，后续元素同理。加载完毕后，将向量寄存器中的数据写回数据内存，以此来实现像素位置的变化。标量处理器完成同样的操作需要对每个数据进行单独读写，这需要大量的指令开销，而一条索引加载指令能够完成一批数据的位置变化，显著提高效率。

另外，索引矩阵决定了算法中像素位置变化的规则，如果多张图像都采用同一种变换方式，则索引矩阵可以被提前计算好并作为常量存储，以减少计算量。

2.1.2 像素数值变化类算法

涉及像素数值变化的算法包括灰度化、二值化和亮度调整。这类算法的特点是：输出图像和输入图像之间仅像素值发生变化，像素位置并未发生变化。在实现过程中，可以使用常规的 unit-strideload 加载指令加载数据，按照算法原理使用对应的向量算数运算指令进行计算，将计算结果写回数据内存。

本节以灰度化为例说明加速这类算法的方法。灰度化有多种实现方法，如分量法、最大值法、均值法和加权平均法。本文采用图像处理库 OpenCV 中的加权平均来保证通用性。

$$\text{Gray}[x] = 0.072B[x] + 0.715G[x] + 0.213R[x] \tag{2}$$

式中：Gray[x] 表示像素 x 的灰度值；B、G、R 分别表示蓝、绿、红 3 个颜色通道。

计算灰度值的加权平均公式有浮点乘法，需要消耗大量的计算资源。因此，本文将式（2）中等号右侧所有系数都乘以 256 再右移 8 位，如式（3）所示。这种方式使用 8 位的整数乘法和移位来代替浮点乘法，易于硬件实现。经过评估，由此带来的精度损失不超过 1%。

$$\text{Gray}[x] = (18B[x] + 183G[x] + 55R[x]) \gg 8 \tag{3}$$

使用 RVV 指令对算法进行实现，伪代码如算法 1 所示。其中，输入为 B、G、R 这 3 个通道的像

素值, 输出结果 RES 为灰度值; 循环次数 n 取决于需要处理的数据量和向量寄存器的长度; src 表示该变量存储在数据内存中, vec 表示该变量存储在向量处理器的寄存器中。

使用向量指令实现的灰度化算法可以分为以下 4 步: ①将数据从内存加载到向量寄存器中 (算法 1 的第 2～4 行)。②在 3 个通道上分别进行乘累加操作 (第 5～7 行), 其中, 括号内的第 1 个参数代表需要累加的数据, 后 2 个参数代表指令中的操作数 1 和操作数 2。在该算法使用的乘累加指令中, 操作数 1 的数据类型是整形, 操作数 2 的数据类型是向量寄存器的地址。在实际运算中, 会根据地址调取该向量寄存器中的所有元素进行计算, 实现 SIMD 的效果。③将乘累加后的结果右移 8 位 (第 8 行)。④将结果从向量寄存器写回数据内存 (第 9 行)。

算法 1　灰度化算法的向量实现。

输入: src_B, src_G, src_R。

输出: src_RES。

1. for i = 1 to n do
2. vec_B_i = VLOAD(src_B_i);
3. vec_G_i = VLOAD(src_G_i);
4. vec_R_i = VLOAD(src_R_i);
5. vec_RES$_i$ = VMAC(vec_RES$_i$,18,vec_B_i);
6. vec_RES$_i$ = VMAC(vec_RES$_i$,183,vec_G_i);
7. vec_RES$_i$ = VMAC(vec_RES$_i$,55,vec_R_i);
8. vec_RES$_i$ = VSHIFT(vec_RES$_i$,8);
9. src_RES$_i$ = VSTORE(vec_RES$_i$);
10. end for

2.1.3　求全局平均值类算法

许多 CNN 算法在图像数据输入卷积层前需要进行数据标准化的操作。标准化涉及求图像全局的平均值。求平均值要先求全局像素值的和, 本文使用规约求和指令实现该类算法的加速。

规约指令是 RVV 指令中的一类特殊指令, 能够对指定向量寄存器中的元素两两运算, 将其规约为一个标量值。RVV 中的规约指令有规约求和、规约求最大值、规约求与等。

首先, 使用规约求和指令求得所有像素的和; 然后, 使用移位代替除法求得图像的全局平均值; 最后, 在当前像素值的基础上调用向量减法指令减去全局平均值, 即可实现数据标准化。

2.2　卷积类算法

高斯滤波、拉普拉斯滤波和索贝尔边缘检测等算法的共同特点是需要对比或计算像素点与其周围像素点的关系, 类似卷积运算。因此, 本文设计了一种基于 RVV 的卷积加速方案。

2.2.1　卷积加速方案

使用 RVV 指令实现卷积操作的难点在于向量指令易于实现地址连续的操作, 但是卷积运算处理的二维数据的地址是不连续的。如图 3 所示, 左侧为原始图像, 中间为卷积核, 右侧为计算结果。在原始图像中, 像素点 a_i、b_i、c_i 的存储地址是连续的, 而像素点 c_i 和像素点 d_i 由于处在不同行, 其地址不是连续的, 这不利于向量指令进行算法实现。

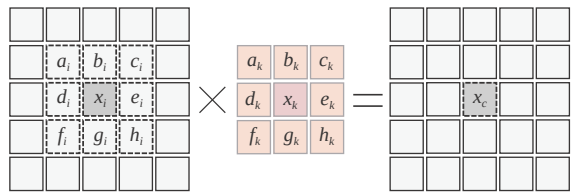


图 3　卷积计算过程

Fig. 3　Convolution computation process

在卷积运算过程中, 卷积核按照一定的步长遍历整个图像。以图 3 为例, 当卷积的步长为 1 时, 卷积核会均匀地滑过图像中的所有像素点。图像中的每个元素都要与卷积核中每个权重相乘。

基于以上分析, 本文方案第 1 步为计算该图像和卷积核中所有元素的乘积。这一步操作是地址连续的向量与标量的相乘, 易于向量实现。以图 3 中的卷积核为例, a_k 与图像所有像素点相乘得到第 1 张图像, b_k 与图像所有像素点相乘得到第 2 张图像, 后续同理, 一共产生 9 张图像 img1～img9, 如图 4 左侧所示。这 9 张图像称为临时卷积图像。

卷积运算的过程实际上是乘累加的过程, 第 1 步相当于完成了“乘累加”中的乘法操作, 下一步需要从 9 张临时卷积图像中找到需要的值并累加。

图 4 的最右侧是计算结果 RES, 其中, x_c 对应的 9 个临时乘积分别位于临时卷积图像 img1～img9 中 x_1～x_9 的位置。因此, $x_c = x_1 + x_2 + \cdots + x_9$。假设 x_c 在数组中的位置为 m, 在计算 x_c 时, 需采用以下公式:

$$\begin{aligned} x_c = \mathrm{RES}[m] = {} & \mathrm{img1}[m-r-1] + \mathrm{img2}[m-r] + \\ & \mathrm{img3}[m-r+1] + \mathrm{img4}[m-1] + \mathrm{img5}[m] + \\ & \mathrm{img6}[m+1] + \mathrm{img7}[m+r-1] + \\ & \mathrm{img8}[m+r] + \mathrm{img9}[m+r+1] \end{aligned} \tag{4}$$

式中: r 为临时卷积图像一行的元素数量。然而, 由于 x_1～x_9 在数组中的位置不统一, 这种计算方式也难以进行向量实现。

2.1.1 节中提到的索引加载指令能够改变像素的存储位置。为使操作易于向量化, 本文方法使用索引加载指令索引 img1～img9, 以此统一 x_1～x_9 在数组中的位置。索引矩阵决定了像素位置变换的

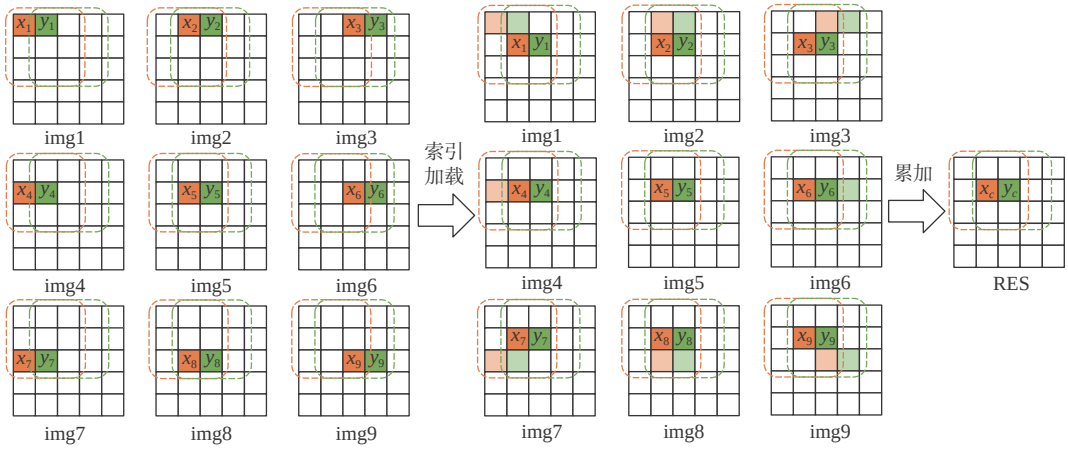


图4 索引加载和累加计算过程

Fig. 4 Computation process of index loading and accumulation

规则，如img1对应的索引矩阵负责将所有像素往右下移动一格，img2对应的索引矩阵负责将所有像素往下移动一格，其余索引矩阵同理，共9个。索引加载过程如图4所示。

索引加载完成后，计算 x_c 采用的公式为

$$\begin{aligned} x_c = \mathrm{RES}[m] = {} & \mathrm{img1}[m] + \mathrm{img2}[m] + \mathrm{img3}[m] + \\ & \mathrm{img4}[m] + \mathrm{img5}[m] + \mathrm{img6}[m] + \\ & \mathrm{img7}[m] + \mathrm{img8}[m] + \mathrm{img9}[m] \end{aligned} \tag{5}$$

这种计算方式易于进行向量实现。此外，不仅计算 x_c 需要的 x_1～x_9 位置进行了统一，图像中的其他点，如计算 y_c 需要的 y_1～y_9 位置也得到了统一。调用向量指令将索引后的9张临时卷积图像累加，便得到了整张图像的卷积计算结果。至此，完成了“乘累加”中的累加。

本文提出的卷积加速方法可以概括为以下4步：①将图像与卷积核中的每个权重相乘，得到9张临时卷积图像；②计算9个索引矩阵；③用索引矩阵索引9张临时卷积图像；④累加，得到实际卷积结果。值得注意的是，如果需要处理大量图像的卷积，9个索引矩阵实际上只需要被计算一次。

2.2.2 卷积加速方案实例——高斯滤波

高斯滤波算法通过计算像素值的加权和来平滑图像并减少噪声，提高CNN推理的准确率。其算法原理是：取图像中每个像素点相邻的矩形窗口，按照权重矩阵计算窗口内所有像素点的加权和。计算式如式(6)所示。高斯滤波的计算过程本质上是一种卷积。权重矩阵的权重值遵循二元正态分布，越接近矩阵中心，权重值越大。

$$C(x,y) = \sum_{j=y-n}^{y+n} \sum_{i=x-n}^{x+n} W(i,j) * P(i,j) \tag{6}$$

式中：C 为计算结果；W 和 P 分别为权重和该点周围的像素。

本文提出的卷积加速方法还有一个常规卷积计算方法不具备的优势，即当卷积核中有重复的权值时，可以避免重复计算。图5展示了图像中像素点 a、b、c 的高斯滤波计算过程。高斯滤波权重矩阵中，距离卷积核中心距离相等的点权值相同。根据2.2.1节的分析，图像中每个像素点在卷积计算的过程中会与权重矩阵中的所有元素相乘，即乘4次 W_1，4次 W_2，1次 W_3。重复的计算会导致资源和功耗的浪费，实际上，每个点只需要分别与 W_1、W_2、W_3 相乘1次。因此，本文提出的卷积加速方法能够规避卷积核中有重复权值产生的额外乘法计算。高斯滤波算法向量实现伪代码如算法2所示。

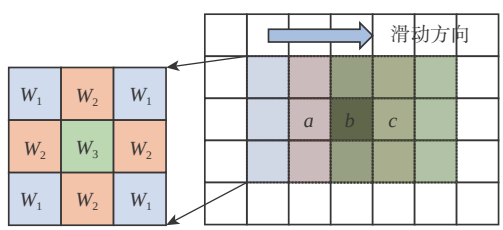


图5 高斯滤波计算过程

Fig. 5 Gaussian filter computation process

将输入图像从数据内存加载到向量寄存器(算法2的第2行)后，先计算图像的所有像素点与 W_1、W_2、W_3 这3个权重的乘积，由于权重均为浮点数据，采取2.1.2节中用到的方法，用整数乘法和移位来代替浮点乘法，得到3张临时卷积图像(第4～6行)。计算完毕后，利用索引加载指令，索引在上一步中计算完毕的临时卷积图像。直到9张临时卷积图像都被索引并累加完毕(第9～11行)。最后，将计算结果写回数据内存(第13行)。

算法2 高斯滤波算法的向量实现。

输入：src_input，W_1，W_2，W_3，src_index$_1$～src_index$_9$。

输出：src_RES。

1. for i = 1 to n do
2. vec_input$_i$ = VLOAD(src_input$_i$);
3. for j = 1 to 3 do
4. vec_W_j = VMUL(vec_input$_i$, W_j);
5. vec_W_j = VSHIFT(vec_W_i, 8);
6. src_W_j = VSTORE(vec_W_j);
7. end for
8. for k = 1 to 9 do
9. vec_index$_k$ = VLOAD(src_index$_k$);
10. vec_RES$_k$ = VINDEX(src_W, vec_index$_k$);
11. vec_RES$_k$ = VADD(vec_RES$_k$, vec_RES$_k$);
12. end for
13. src_RES = VSTORE(vec_RES);
14. end for

3　自定义向量指令

在使用 RVV 指令对图像预处理算法进行加速时, 标准的 RVV 并不能完全满足算法向量化的要求。例如, 索贝尔边缘检测中需要进行阈值判断并置位, 这在标准 RVV 中不被支持。

因此, 为实现标准 RVV 中不支持的功能和进一步对算法进行加速, 本文新增了 6 条自定义向量指令, 并通过修改编译器、设计硬件电路的译码和执行模块实现了这 6 条指令。

3.1　指令定义与结构

新增的 6 条自定义向量指令如图 6 所示。指令的长度为 32 位, 其中, 不同的编码段代表的意义不同。OP-V major code 表示该指令属于 RVV 指令; funct6 用于和该指令集内的其他指令进行区分。因此, 在自定义向量指令时, 该编码段不能与已有的标准向量指令重复。考虑到自定义向量指令的需求, RVV1.0 预留了一些没有被使用的编码空间。每条指令都有 2 个操作数, 使用 funct3 来区分 2 个操作数的类型, 其中, funct3 = 110 和 100 分别表示该指令的 2 个操作数分别为向量和标量; vs2 和 rs1 分别表示向量源寄存器和标量源寄存器的地址, vd 表示目的寄存器的地址; vm 表示掩码位。

funct6				funct3		OP-V major code	
111001	vm	vs2	rs1	100	vd	1010111	vset255.vx
010110	vm	vs2	rs1	100	vd	1010111	vset0.vx
111001	vm	vs2	rs1	110	vd	1010111	vabs.vx
010110	vm	vs2	rs1	110	vd	1010111	vcmpset.vx
010101	vm	vs2	rs1	110	vd	1010111	vwmsau.vx
001101	vm	vs2	rs1	110	vd	1010111	vmuls.vx

图 6　自定义向量指令格式

Fig. 6　Customized vector instruction format

在本文实现的图像预处理算法中, 需要扩展指令的操作有: 在拉普拉斯滤波和亮度调整算法中, 需要判断像素值改变后是否越界 (0～255); 在索贝尔边缘检测和二值化算法中, 需要判断像素值是否大于阈值并进行置位; 在索贝尔边缘检测算法中, 需要计算数据的绝对值。

针对上述操作, 本文新增了 4 条自定义指令, 其逻辑功能如表 1 所示。vset255 和 vset0 这 2 条指令能够将越界的数据调整为上限或下限数据; vabs 指令判断输入操作数的正负, 如果操作数为负则取反, 否则返回原值; vcmpset 指令先判断 2 个操作数的大小, 如果操作数 1 大于操作数 2, 则返回一个置位值, 否则返回另一个置位值。

此外, 本文使用乘法和移位代替浮点运算, 为进一步减少指令调度, 提高加速效果, 本文将这 2 步操作进行合并, 新增了无符号加宽型向量乘移位累加 vwmsau 和向量乘移位 vmuls, 其逻辑功能如表 1 所示, 2 条指令的主要区别在于是否累加。

表 1　自定义向量指令功能

Table 1　Function of customized vector instructions

指令	功能
vset255	vd = vs1 >rs2 ? 255: vs1
vset0	vd = vs1 >rs2 ?vs1:0
vabs	vd = vs1 > rs2 ? vs1: −vs1
vcmpset	vd = vs1 > rs2 ?255: 0
vwmsau	vd = vd + (vs1·rs2 >> 8)
vmuls	vd = vs1·rs2 >>8

3.2　自定义向量指令的实现

自定义指令的实现分为 2 部分: 对编译器的修改和对处理器本身的修改。

以 GNU 编译套件 (GNU compiler collection, GCC) 为例, 在修改编译器前, 按照 RVV 指令规范定义向量编码。将自定义的指令变量添加到 GCC 的二进制工具 (RISC-V binutils) 中并重新编译, 使 GCC 能够成功识别自定义向量指令并生成对应的机器码。

自定义向量指令的硬件实现如图 7 所示。左侧是标量处理器 ibex 和数据内存, 右侧是向量处理器 Vicuna。Vicuna 中的比较置位模块实现了 vset255、vset0、vabs 和 vcmpset 指令, 乘移位模块实现了 vwmsau 和 vmuls 指令。

实现自定义向量指令需要先对指令译码器进行修改, 译码器根据指令的 funct6 和 funct3 编码段来辨别自定义向量指令, 按照指令的操作类型生成控制信号及操作数据, 并发送到对应的执行模块。

为实现 vset255、vset0、vabs 和 vcmpset 指令, 本

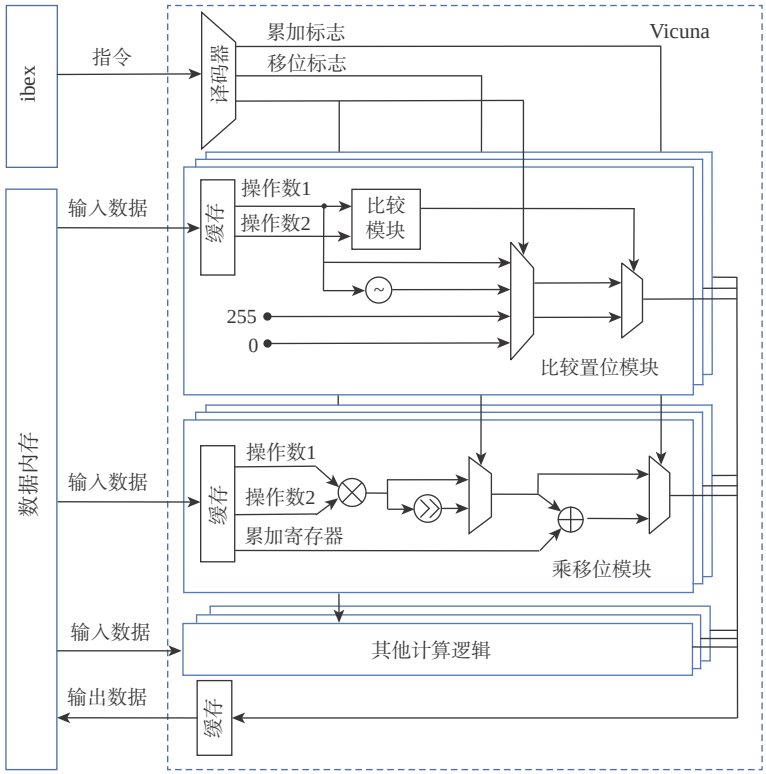


图7 自定义向量指令的硬件实现

Fig. 7 Hardware implementation of customized vector instructions

文设计了比较置位模块。数据进入该模块时，会先进入比较置位模块进行大小比较。在该模块中，根据指令的功能定义设置了4种不同的置位值：vs1(操作数1)、−vs1、255和0，根据译码阶段的控制信号及比较置位模块的输出信号，通过两级多路复用器(multiplexer, MUX)选择置位值输出。这4条指令复用了比较置位模块。

为实现vwmsau和vmuls指令，本文设计了乘移位模块。该模块复用了Vicuna原有的乘法模块逻辑，仅增加移位标志、移位器和MUX，以减少资源开销。如果译码模块遇到这2条指令，移位标志就会被置为1。由于没有影响原乘法器的功能，乘移位模块根据不同指令能够实现4种操作，分别为乘法、乘累加、乘移位和乘移位累加。

4 实验结果与分析

4.1 仿真及实验环境

本文选择riscv32-unknown-elf-gcc v12.0.1作为编译工具，选择Verilator v4.210作为仿真工具来测试自定义模块的功能正确性和图像预处理算法的加速效果。使用Vivado v2019.2进行硬件电路的综合和实现，并部署到Xilinx Artix-7 FPGA开发板上进行性能验证，评估面积和功耗，工作频率为50 MHz。用于测试的图像数据来自数据集CIFAR-10[20]，图像大小为32×32×3。

4.2 不同处理器配置下的性能与资源消耗

本文以周期数为指标评价性能。由于时间=周期数/频率，经评估，本文设计并未影响处理器的关键路径，其依然能按照原有的最大频率工作。因此，周期数的变化能准确反映计算时间的变化。

不同的处理器配置会对性能和面积产生影响。本节就2种参数的配置对图像预处理加速实验结果的影响展开讨论，分别为向量寄存器长度VLEN和向量处理单元位宽PIPE_W。

4.2.1 向量寄存器长度

RISC-V向量处理器设有32个向量寄存器v0～v31，每个向量寄存器的长度为VLEN，单位为bit。RVV中，向量寄存器长度不是固定的，Vicuna的向量寄存器长度可配置的范围为64～2 048 bit。向量寄存器长度可配置是指在硬件电路中，向量寄存器有不同长度的实现，而不是软件层面的配置。

图8展示了随向量寄存器长度的变化，单张图像灰度化算法向量实现周期数的变化(向量处理单元位宽固定为32 bit)。可以发现，随着向量寄存器长度增长，算法需要的周期数显著减少，向量寄存器长度从64 bit增加到1 024 bit，周期数减少了60.2%。当向量寄存器长度增大时，单条向量指令(如加载指令、运算指令和存储指令)处理的数据量得到提升。

不同向量寄存器长度配置下，FPGA查找表

(look up table, LUT)、触发器 (flip flop, FF)、块随机存取存储器 (block random access memory, BRAM) 和数字信号处理单元 (digital signal processor, DSP) 的消耗量如表 2所示。向量寄存器长度的变化不会影响 BRAM 和 DSP 的用量, 但 LUT 和 FF 的消耗会增加。向量寄存器长度的变化对性能和资源消耗均有明显影响, 因此, 应针对不同的应用环境要求选择合适的向量寄存器长度配置。

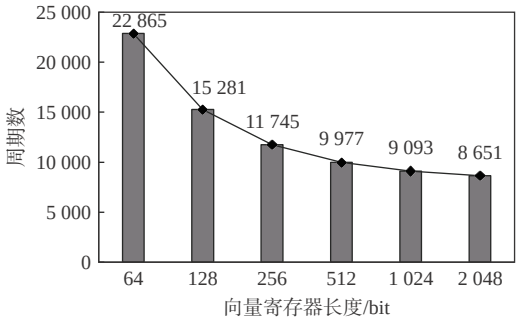


图 8　不同向量寄存器长度配置下的灰度化周期数

Fig. 8　Cycles of grayscale processing under different vector register length configurations

表 2　不同向量寄存器长度配置下资源消耗及功耗

Table 2　Resource and power consumption under different vector register length configurations

向量寄存器长度/bit	LUT	FF	BRAM	DSP	功耗/W
64	9 880	5 976	64	7	0.336
128	10 373	6 964	64	7	0.346
256	11 312	8 716	64	7	0.353
512	13 262	12 168	64	7	0.373
1 024	17 374	19 079	64	7	0.424
2 048	41 836	33 068	64	7	0.533

4.2.2　向量处理单元位宽

向量处理单元位宽是向量处理单元的位宽, 即算术逻辑单元 (arithmetic logical unit, ALU)、乘法单元等运算模块能同时处理的数据量, 代表了向量处理单元的计算性能。不同向量处理单元位宽对灰度化算法周期数的影响如图 9 所示。资源消耗如表 3 所示 (向量寄存器长度固定为 2 048 bit)。向量处理单元位宽取值范围为 32～VLEN/2, 单位为 bit。

向量处理单元位宽影响了 LUT、FF 和 DSP 资源的使用量, 但其增加对周期数的影响并不明显。向量处理单元位宽从 32 bit 提升到 512 bit, 周期数仅减少了 2.5%。由于图像预处理算法是数据密集型的应用, 限制算法效率的往往不是向量处理单元的计算性能, 而是可用的内存带宽。Vicuna 的内存读取端口宽度为 32 bit, 限制了算法的效率, 因此, 向量处理单元位宽的变化并不会显著影响性能。此外, 只涉及像素位置变化的算法, 如旋转、镜像、

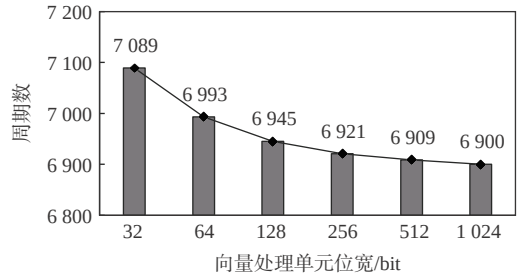


图 9　不同向量处理单元位宽配置下的灰度化周期数

Fig. 9　Cycles of grayscale processing under different vector processing pipe width configurations

表 3　不同向量处理单元位宽配置下资源消耗及功耗

Table 3　Resource and power consumption under different vector processing pipe width configurations

向量处理单元位宽/bit	LUT	FF	BRAM	DSP	功耗/W
32	41 386	33 068	64	7	0.533
64	42 693	33 642	64	11	0.539
128	44 133	34 772	64	19	0.545
256	47 470	37 058	64	35	0.577
512	52 061	41 584	64	67	0.607
1 024	81 279	48 705	64	131	0.738

平移等, 不会用到逻辑运算单元, 向量处理单元位宽的增加对这些算法的效率没有影响。因此, 在图像预处理类应用中, 本文选择尽量小的向量处理单元位宽来减少资源消耗。

4.3　加速效果及分析

结合 4.2 节中对向量处理器关键参数配置的分析, 本节的实验结果均基于向量寄存器长度为 2 048 bit、向量处理单元位宽为 32 bit 的处理器配置。以标量处理器 ibex 的执行周期数为基准, 列举了文中方法对各类图像预处理算法的加速效果。

使用索引加载指令对像素位置变化类算法进行加速, 效果如表 4 所示。实验结果显示, 本文方法对像素位置变化类算法相比标量处理器实现了 6.93～9.75 倍的加速效果。索引加载指令能够对这类算法实现高效的加速。

对像素数值变化类算法进行加速, 效果如表 5 所示。实验结果显示, 本文方法对 3 种像素数值变化类算法相比标量处理器实现了 6.69～9.97 倍的加速。标准向量扩展仅对灰度化算法有较好的加

表 4　像素位置变化类算法加速效果

Table 4　Acceleration results of pixel location change algorithms

算法	标量基准周期数	RVV周期数	加速倍数
旋转/镜像	46 728	4 791	9.75
平移	47 439	5 679	8.35
缩放	40 681	5 868	6.93

速效果，对二值化和亮度调整算法没有明显的加速效果。其中，二值化算法无法使用标准向量扩展指令加速，因此，其周期数与标量基准一致；由于标准向量指令集不支持亮度调整算法中的比较置位运算，这部分运算需要在标量处理器上执行，额外的标量处理器读写和运算操作导致加速效果被严重稀释。

高效的自定义RVV指令实现了标准RVV指令集不支持的功能，解决了加速效果被标量处理器稀释的问题，使得二值化和亮度调整算法的效率大幅提升。合并乘法与移位的扩展指令也使得灰度化算法得到进一步的加速。

求全局平均值类算法为标准化算法。本文使用规约求和指令对这类算法进行加速，效果如表6所示。实验结果显示，本文方法对求全局平均值类算法实现了7.52倍的加速效果。

表5 像素数值变化类算法加速效果

Table 5 Acceleration results of pixel value change algorithms

算法	标量基准周期数	标准RVV周期数	标准RVV加速倍数	自定义RVV周期数	自定义RVV加速倍数
灰度化	60 488	10 786	5.61	8 651	6.99
二值化	42 576	42 576	1.00	4 272	9.97
亮度调整	116 627	80 730	1.44	17 422	6.69

表6 求全局平均值类算法加速效果

Table 6 Acceleration results of global average computation algorithms

算法	标量基准周期数	RVV周期数	加速倍数
标准化	65 620	8 724	7.52

卷积类算法包括高斯滤波、拉普拉斯滤波和索贝尔边缘检测，应用本文提出的卷积加速方案对这类算法进行加速，结果如表7所示。实验结果显示，本文方法对3种卷积类算法相比标量处理器实现了3.13～4.26倍的加速。应用本文提出的卷积加速方案，在使用标准RVV指令时，能取得一定的加速效果，但仍存在加速效果被稀释的现象。而结合卷积加速方案和自定义RVV指令能够解决这一问题，运算效率得到进一步提升。

表7 卷积类算法加速效果

Table 7 Acceleration results of convolution algorithms

算法	标量基准周期数	标准RVV周期数	标准RVV加速倍数	自定义RVV周期数	自定义RVV加速倍数
高斯滤波	268 604	66 178	4.06	63 037	4.26
拉普拉斯滤波	214 456	105 235	2.04	57 650	3.72
索贝尔边缘检测	214 783	197 202	1.09	68 542	3.13

4.4 CNN应用中的资源消耗与功耗

以文献[21]中的MobileNet加速器为例，预估本文硬件平台在实际CNN加速应用中的资源消耗，如表8所示。Vicuna的配置为：向量寄存器长度为2 048 bit，向量处理单元位宽为32 bit。应用于CPU+CNN加速器的深度学习平台中，向量处理器导致的资源消耗增长仅约10%，自定义指令导致的额外硬件资源开销小于0.1%。

表8 资源消耗及功耗

Table 8 Resource and power consumption

硬件架构	LUT	FF	BRAM	DSP	功耗/W
ibex+CNN	310 840	279 836	957	2 161	11.65
ibex+Vicuna+CNN	349 618	311 699	1 005	2 168	11.88
ibex+自定义Vicuna+CNN	349 835	311 964	1 005	2 168	11.88

4.5 在CNN中的应用实例

本节选取嵌入式图像分类应用[9]和车牌识别应用为例，预估本文方法在深度学习应用场景中的效果。

如图10所示，在图像分类应用中，CPU预处理阶段先进行图像的缩放、灰度化和归一化，再由二值神经网络(binary neural network, BNN)加速器进行推理分类。整个计算过程中，加速器占用时间为24%(137 ms)，CPU图像预处理占用时间为76%(433 ms)，其中，缩放占用30%(171 ms)，灰度化占用32%(182 ms)，归一化占用14%(80 ms)。应用本文方法，CPU图像预处理的占用时间大幅缩短，整体计算时间相比原来减少54.2%(309 ms)。

如图11所示，在车牌识别应用中，预处理阶段进行灰度化、高斯滤波和索贝尔边缘检测，再由3层的反向传播神经网络(back propagation neural

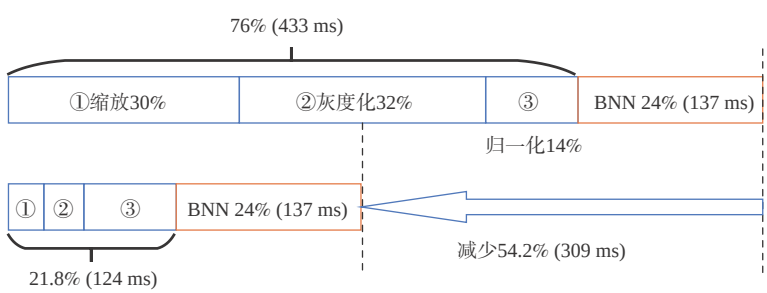


图10 图像分类应用中的效果

Fig. 10 Effect in image classification application

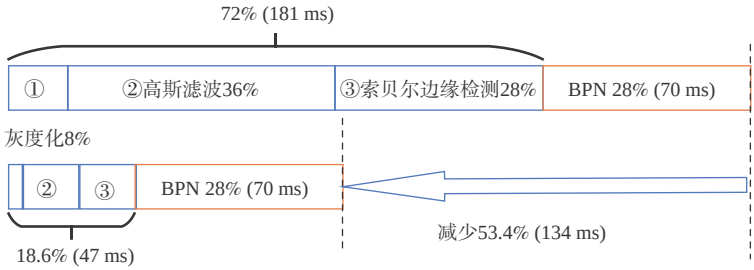


图 11　车牌识别应用中的效果

Fig. 11　Effect in license plate recognition application

network, BPN) 进行识别。BPN 占用时间为 28% (70 ms), CPU 图像预处理占用时间为 72%(181 ms), 其中, 灰度化占用 8%(20 ms), 高斯滤波占用 36% (91 ms), 索贝尔边缘检测占用 28%(70 ms), 应用本文方法, 整体计算时间相比原来减少 53.4%(134 ms), 有效解决了 CPU 图像预处理在 CNN 计算中的性能瓶颈问题。

5　结　论

1) 针对 11 种常见的图像预处理算法, 按照算法特点分为 4 类, 并针对每种分类设计了基于 RISC-V 向量扩展的加速方案。

2) 为解决标准向量扩展不能满足图像预处理加速需求的问题, 自定义了 4 条向量指令; 为进一步提高算法运算效率, 自定义了 2 条向量指令以合并乘法与移位运算。通过设计 RISC-V 处理器微架构实现了这 6 条指令。

3) 实验结果表明, 本文提出的方案与标量处理器相比实现了 3.13～9.97 倍的加速效果。在 2 种实际应用中, 预估能减少 50% 左右的整体计算时间, 可有效解决 CPU 图像预处理在 CNN 整体计算中的性能瓶颈问题。

参考文献 (References)

[1] PAL K K, SUDEEP K S. Preprocessing for image classification by convolutional neural networks[C]//Proceedings of the IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology. Piscataway: IEEE Press, 2016: 1778-1781.

[2] ŞABAN Ö, AKDEMIR B. Effects of histopathological image preprocessing on convolutional neural networks[J]. Procedia Computer Science, 2018, 132: 396-403.

[3] PITALOKA D A, WULANDARI A, BASARUDDIN T, et al. Enhancing CNN with preprocessing stage in automatic emotion recognition[J]. Procedia Computer Science, 2017, 116: 523-529.

[4] TABIK S, PERALTA D, HERRERA-POYATOS A, et al. A snapshot of image pre-processing for convolutional neural networks: case study of MNIST[J]. International Journal of Computational Intelligence Systems, 2017, 10(1): 555-568.

[5] DU C Y, TSAI C F, CHEN W C, et al. A 28 nm 11.2 TOPS/W hardware-utilization-aware neural-network accelerator with dynamic dataflow[C]//Proceedings of the IEEE International Solid-State Circuits Conference. Piscataway: IEEE Press, 2023: 1-3.

[6] KELLER B, VENKATESAN R, DAI S, et al. A 95.6-TOPS/W deep learning inference accelerator with per-vector scaled 4-bit quantization in 5 nm[J]. IEEE Journal of Solid-State Circuits, 2023, 58(4): 1129-1141.

[7] KARNIK T, KURIAN D, ASERON P, et al. A cm-scale self-powered intelligent and secure IoT edge mote featuring an ultra-low-power SoC in 14 nm tri-gate CMOS[C]//Proceedings of the IEEE International Solid-State Circuits Conference. Piscataway: IEEE Press, 2018: 46-48.

[8] NARAYANAN D, SANTHANAM K, PHANISHAYEE A, et al. Accelerating deep learning workloads through efficient multi-model execution[C]//Proceedings of the NeurIPS Workshop on Systems for Machine Learning. [S. l.]: NeurIPS, 2018: 20-27.

[9] JIA T Y, JU Y H, JOSEPH R, et al. NCPU: an embedded neural CPU architecture on resource-constrained low power devices for real-time end-to-end performance[C]//Proceedings of the 53rd Annual IEEE/ACM International Symposium on Microarchitecture. Piscataway: IEEE Press, 2020: 1097-1109.

[10] NVIDIA. NVIDIA data loading library (DALI) [EB/OL]. [2023-04-01]. https://github.com/NVIDIA/DALI.

[11] MA N. A SoC-based acceleration method for UAV runway detection image pre-processing algorithm[C]//Proceedings of the 25th International Conference on Automation and Computing. Piscataway: IEEE Press, 2019: 1-6.

[12] KUO Y M, GARCÍA-HERRERO F, RUANO O, et al. RISC-V Galois field ISA extension for non-binary error-correction codes and classical and post-quantum cryptography[J]. IEEE Transactions on Computers, 2023, 72(3): 682-692.

[13] RAZILOV V, MATÚŠ E, FETTWEIS G. Communications signal processing using RISC-V vector extension[C]//Proceedings of the International Wireless Communications and Mobile Computing. Piscataway: IEEE Press, 2022: 690-695.

[14] CAVALCANTE M, SCHUIKI F, ZARUBA F, et al. Ara: a 1-GHz scalable and energy-efficient RISC-V vector processor with multi-precision floating-point support in 22-nm FD-SOI[J]. IEEE Transactions on Very Large Scale Integration Systems, 2020, 28(2): 530-543.

[15] 刘强, 李一可. 基于指令扩展的 RISC-V 可配置故障注入检测方法[J/OL]. 北京航空航天大学学报, 2023(2023-03-10)[2023-04-01]. https://bhxb.buaa.edu.cn/bhzk/cn/article/doi/10.13700/j.bh.1001-5965.2022.0995.

LIU Q, LI Y K. Configurable fault detection method for RISC-V processors based on instruction extension[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2023(2023-03-10)[2023-04-01]. https://bhxb.buaa.edu.cn/bhzk/cn/article/doi/10.13700/j.bh.1001-5965.2022.0995(in Chinese).

[16] ASANOVIC K. Vector extension 1.0[EB/OL]. (2021-09-20)[2023-04-01]. https://github.com/riscv/riscv-v-spec/releases/tag/v1.0.

[17] PLATZER M, PUSCHNER P. Vicuna: a timing-predictable RISC-V vector coprocessor for scalable parallel computation[C]//Proceedings of the 33rd Euromicro Conference on Real-Time Systems. Porto: [s. n.], 2021: 1-18.

[18] SCHIAVONE P D, CONTI F, ROSSI D, et al. Slow and steady wins the race? A comparison of ultra-low-power RISC-V cores for Internet-of-things applications[C]//Proceedings of the 27th International Symposium on Power and Timing Modeling, Optimization and Simulation. Piscataway: IEEE Press, 2017: 1-8.

[19] OpenHW Group. OpenHW group eXtension interface[EB/OL]. [2023-04-01].https://docs.openhwgroup.org/projects/openhw-group-core-v-xif/en/latest/index.html.

[20] ALEX K, VINOD N. The CIFAR-10 dataset[EB/OL]. [2023-04-01]. http://www.cs.toronto.edu/~kriz/cifar.html.

[21] YAN S, LIU Z Y, WANG Y, et al. An FPGA-based MobileNet accelerator considering network structure characteristics[C]//Proceedings of the 31st International Conference on Field-Programmable Logic and Applications. Piscataway: IEEE Press, 2021: 17-23.

Image preprocessing acceleration method based on RISC-V vector extension

LIU Qiang[1, 2, *], YIN Wei[1, 2], LI Kai[1, 2]

(1. School of Microelectronics, Tianjin University, Tianjin 300072, China;
2. Tianjin Key Laboratory of Imaging and Sensing Microelectronic Technology, Tianjin 300072, China)

Abstract: As the pre-order step of convolutional neural network (CNN) computing, image preprocessing is indispensable but time-consuming. To accelerate image preprocessing, a method based on RISC-V vector extension was proposed to accelerate eleven image preprocessing algorithms such as gray scale processing, standardization, and Gaussian filtering. Firstly, eleven image preprocessing algorithms were classified into four categories according to the computing mode, and acceleration schemes for the preprocessing algorithms were designed based on RISC-V vector extension. In order to further improve the performance, six customized vector instructions were added. The customized instructions were implemented by modifying the compiler and designing the hardware module. Finally, a field programmable gate array (FPGA) was used for testing, and the impact of vector processor configuration on performance and resource consumption was analyzed. The results showed that the proposed method achieves 3.13–9.97 times speedup compared with scalar processors, which effectively solves the performance bottleneck problem of image preprocessing in deep learning.

Keywords: convolutional neural network; preprocessing; RISC-V; vector extension; algorithm acceleration

Received: 2023-04-24; **Accepted**: 2023-05-26; **Published Online**: 2023-06-09 10:13

URL: link.cnki.net/urlid/11.2625.V.20230608.1155.002

Foundation item: National Natural Science Foundation of China (U21B2031)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0192

基于DSGIoU损失与双分支坐标注意力的目标检测算法

马素刚[1,2,*]，李宁博[1]，侯志强[1,2]，余旺盛[3]，杨小宝[1]

(1. 西安邮电大学 计算机学院，西安 710121；2. 西安邮电大学 陕西省网络数据分析与智能处理重点实验室，西安 710121；
3. 空军工程大学 信息与导航学院，西安 710077)

摘 要：针对YOLOX算法中边界框回归损失效果有限和多尺度特征表示能力不足，导致检测结果不准确的问题，提出一种基于距离形状广义交并比(DSGIoU)损失与双分支坐标注意力的目标检测算法。在交并比(IoU)损失项的基础上，通过添加真实框与预测框之间的非重叠面积、中心点距离及宽高比3个惩罚项，优化边界框的回归收敛效果；通过平均池化和最大池化沿着2个方向对特征进行编码，获取方向感知信息和位置信息，从而对特征进行增强。为验证所提算法的检测性能，分别以网络大小为Tiny、S、M的YOLOX为基准，在PASCAL VOC和KITTI数据集上进行测试。实验结果表明：所提算法在PASCAL VOC数据集上的检测精度分别达到80.0%、82.6%、85.8%，相比基准算法YOLOX提升了1.5%、1.6%、2.0%；在KITTI数据集上的检测精度分别达到87.7%、89.7%、90.7%，相比基准算法YOLOX提升了1.7%、2.9%、1.3%。所提算法能够优化网络收敛性，提高多尺度特征的表示能力，有效提高检测精度。

关 键 词：目标检测；损失函数；边界框回归；坐标注意力；YOLOX
中图分类号：TP391.4
文献标志码：A **文章编号**：1001-5965(2025)04-1085-11

目标检测是计算机视觉领域中研究的热点问题[1]，在文本检测[2]、自动驾驶[3]、医学影像[4]等多个场景有着重要的研究意义。

近年来，基于卷积神经网络(convolutional neural network, CNN)的目标检测算法得到了迅速发展，主要分为2类。一类是以YOLO系列[5-8]、SSD系列[9-10]等为代表的单阶段目标检测算法。YOLOv1[5]将输入图像划分为多个互不重叠的网格，并在每个网格上设定长宽比不同的锚框以确保预测的完整性和准确性，当目标物体的中心点落入某个网格中，则由该网格负责对此目标进行预测；SSD[9]采用分而治之的思想，将大小不同的目标物体映射到不同尺度的特征图上进行检测。另一类是以R-CNN系列[11-13]为代表的双阶段目标检测算法。该类算法采用选择性区域搜索或区域候选网络生成目标候选框，对候选框进行分类和定位从而得到最终的检测结果。与单阶段目标检测算法相比，双阶段目标检测算法的检测精度较高，但检测速度相对较低。随着Transformer[14]在自然语言处理领域的不断应用，研究者们将其引入了计算机视觉领域中。在目标检测任务中，为实现端到端的自动训练和学习，减少计算量，提出了DETR[15]、UP-DETR[16]等基于Transformer的目标检测算法。

边界框的回归收敛效果极大地影响着目标检测算法的性能。早期，目标检测算法多数采用基于L_n-norm的边界框回归损失函数，可表示为$L_n(\boldsymbol{A},\boldsymbol{B})=$

收稿日期：2023-04-21；**录用日期**：2023-05-15；**网络出版时间**：2023-06-19 15:58
网络出版地址：link.cnki.net/urlid/11.2625.V.20230619.1244.001
基金项目：国家自然科学基金(62072370)；陕西省自然科学基金(2023-JC-YB-598)；西安市科技计划(22GXFW0125)
*通信作者. E-mail：msg@xupt.edu.cn

引用格式：马素刚，李宁博，侯志强，等. 基于DSGIoU损失与双分支坐标注意力的目标检测算法[J]. 北京航空航天大学学报，2025，51(4)：1085-1095. MA S G，LI N B，HOU Z Q，et al. Object detection algorithm based on DSGIoU loss and dual branch coordinate attention[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1085-1095 (in Chinese).

$\|\boldsymbol{A}-\boldsymbol{B}\|_n$。YOLOv1采用均方误差(mean square error, MSE),即L_2-norm损失函数优化边界框回归;SSD和Faster R-CNN[13]算法采用L_1-smooth损失函数计算边界框位置信息的回归损失。上述2种损失函数将边界框的坐标信息作为4个独立的变量进行优化,在网络训练过程中存在1个或2个变量得到很好的优化,但整体效果与真实框相差较大的情况。另外,L_n-norm具有尺度敏感性,使得算法对相同的损失值回归出许多形状大小不同的边界框,这些边界框与真实框相差较大,不能准确反映两者之间的交并比(intersection over union, IoU)指标,导致网络回归收敛效果较差。为使边界框回归与IoU评估指标之间能够等价优化,Yu等[17]将IoU指标作为边界框回归损失函数对网络模型进行优化,通过将4个独立的坐标变量作为一个整体进行回归,很好地解决了训练过程中边界框回归对尺寸敏感的问题。然而对于IoU损失而言,当预测框与真实框不相交时,损失函数L_{IoU}恒为常量1,梯度无法下降,从而导致边界框回归分支的参数无法及时更新;IoU无法准确反映预测框与真实框的重合度,当多个形状不同的预测框与真实框IoU值相等时,无法分辨哪个预测框的回归效果更好。Rezatofighi等[18]提出一种广义交并比(generalized intersection over union, GIoU)及其损失函数L_{GIoU},解决了预测框与真实框不相交时梯度无法下降的问题,并且还能准确反映两者的重合程度。Zheng等[19]提出了距离交并比(distance intersection over union, DIoU)及其损失函数L_{DIoU},在IoU损失项的基础上,引入欧氏距离公式计算预测框与真实框的中心点距离,与L_{IoU}和L_{GIoU}相比,网络收敛速度更快。基于DIoU,Zheng等[19]提出了新的评价指标完全交并比(complete intersection over union, CIoU)及其损失函数L_{CIoU},将重叠面积、中心点距离及纵横比同时作为判断预测框回归效果的评价指标,并且采用L_{CIoU}优化算法的边界框回归分支。上述损失函数虽然不断提高边界框回归的收敛性,但在某些情况下依然存在收敛效果不佳、检测结果不准确的问题。本文提出了一种距离形状广义交并比(distance shape of generalized intersection over union, DSGIoU),基于IoU损失项,将真实框与预测框的非重叠面积、中心点距离、宽高比同时作为惩罚项优化边界框回归,准确反映预测框与真实框之间的差异。

注意力机制以权重的生成和分配重新调整网络参数,并通过训练最终使网络聚焦于关键特征,起到增强识别特征、抑制无关信息的作用,从而优化CNN的学习机制[20-21]。SENet[22]通过2D全局池化计算网络对不同特征通道的关注度,牺牲较小的运算成本以提高网络性能,然而,其忽略了空间位置信息对网络性能的影响。CBAM[23]在通道和空间2个维度通过卷积操作学习得到2组不同的权重系数矩阵,分别与特征图点乘后优化网络参数。GCT[24]通过高斯函数映射捕获网络对不同通道的关注度,在不增加任何参数量的情况下提高网络性能。坐标注意力(coordinate attention, CA)[25]将位置信息嵌入到通道注意力,对输入特征沿着水平和垂直方向进行1D全局池化,减轻因2D全局池化造成位置信息的丢失。为了能够充分获取特征的位置信息,捕获输入特征的长距离相关性从而优化网络参数,本文提出了一种双分支坐标注意力(dual branch coordinate attention, DBCA),并将其引入目标检测网络中。对输入特征分别沿着水平和垂直2个方向进行1D全局最大池化和1D全局平均池化,将输入特征与最终生成的2组不同的注意力激活权重进行点乘,调整网络参数,使得多尺度特征获取丰富的位置信息,从而增强特征表示能力。

本文主要贡献如下:

1)提出一种边界框回归损失函数DSGIoU损失,优化网络的收敛效果,使边界框定位更准确。

2)提出一种双分支坐标注意力机制,充分获取特征的方向感知信息和位置信息,增强特征的表示能力。

3)以YOLOX[26]为基准,在公开数据集PASCAL VOC和KITTI上进行大量实验,验证本文算法的有效性。

1 本文算法

本文算法框架如图1所示,主要包括3部分:改进的主干网络、路径聚合网络(path aggregation network, PANet)和解耦检测头。其中,主干网络部分,在CSP残差块中添加双分支坐标注意力,沿着水平方向与垂直方向对特征分别进行平均池化和最大池化,保留平均响应值与最大响应值,通过捕获长距离位置信息增强特征表示能力。

1.1 DSGIoU损失

IoU即为Jaccard系数,常被用于度量2个任意检测框之间的相似性,是目标检测中常用的评价指标。IoU作为损失项优化边界框回归,旨在消除尺度敏感性。IoU损失的计算公式为

$$L_{\mathrm{IoU}}(B^{\mathrm{gt}},B)=1-\frac{\left|B^{\mathrm{gt}}\cap B\right|}{\left|B^{\mathrm{gt}}\cup B\right|} \tag{1}$$

式中:B^{gt}和B分别为真实框和预测框。

GIoU损失解决了预测框与真实框无交集时

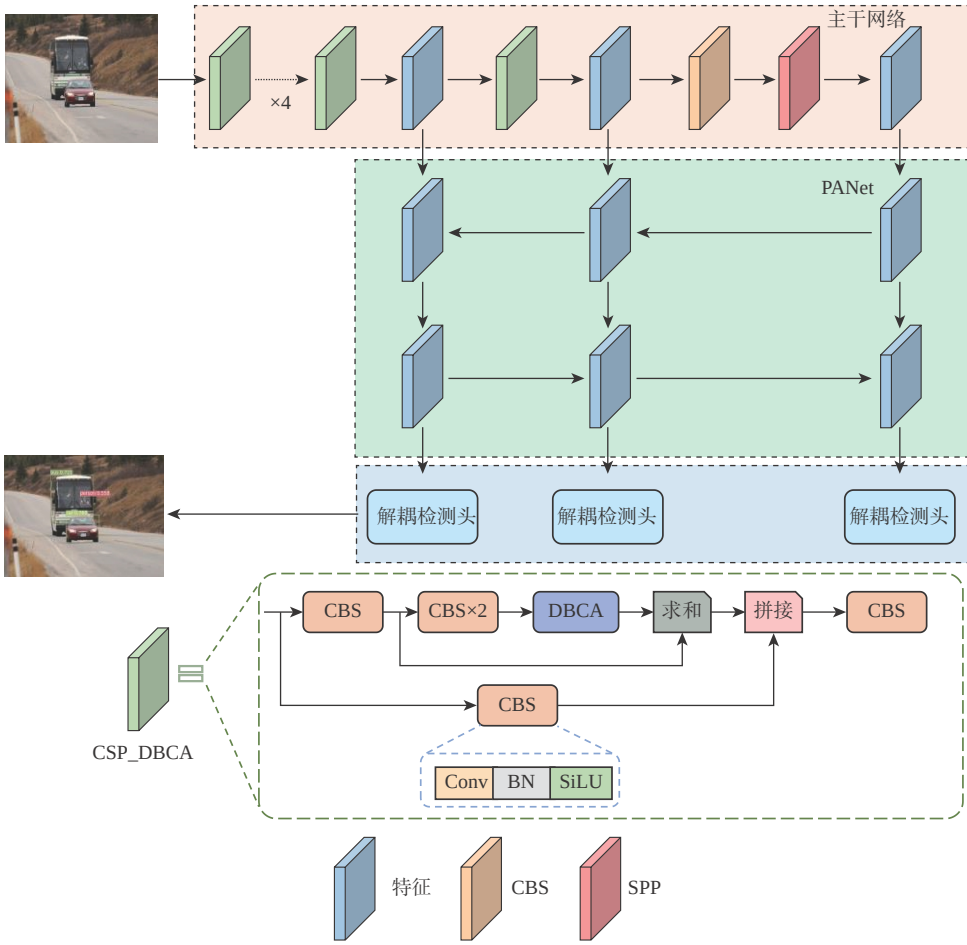


图 1　本文算法框架

Fig. 1　Framework of the proposed algorithm

$L_{\mathrm{IoU}}(B^{\mathrm{gt}},B)$恒等于 1 的问题, 其计算公式为

$$L_{\mathrm{GIOU}}(B^{\mathrm{gt}},B)=1-I+\frac{\left|G-(B^{\mathrm{gt}}\cup B)\right|}{|G|} \tag{2}$$

式中: $I=\left|B^{\mathrm{gt}}\cap B\right|/\left|B^{\mathrm{gt}}\cup B\right|$; G为包含B^{gt}和B的最小矩形框。

GIoU 的不足之处在于: 当真实框与预测框处于包含关系时, GIoU 损失中$\left|G-(B^{\mathrm{gt}}\cup B)\right|$恒为 0, GIoU 退化为 IoU。

DIoU 损失的计算公式为

$$L_{\mathrm{DIoU}}(B^{\mathrm{gt}},B)=1-I+\frac{p^2(B^{\mathrm{gt}},B)}{g^2} \tag{3}$$

式中: $p(B^{\mathrm{gt}},B)$为真实框与预测框之间的中心点距离; g为包含B^{gt}和B的最小矩形框的对角距离。

DIoU 的不足之处在于: 当真实框与预测框中心点重合时, DIoU 损失中$p^2(B^{\mathrm{gt}},B)/g^2$恒为 0, DIoU 也会退化为 IoU。

CIoU 损失的计算公式为

$$L_{\mathrm{CIoU}}(B^{\mathrm{gt}},B)=1-I+\frac{p^2(B^{\mathrm{gt}},B)}{g^2}+av \tag{4}$$

式中: 边界框的纵横比$v=\frac{4}{\pi^2}(\arctan(w^{\mathrm{gt}}/h^{\mathrm{gt}})-\arctan(w/h))^2$, w^{gt}、h^{gt}和w、h分别为真实框和预测框的宽、高; 权衡参数$a=v/(1-I+v)$。

v对w、h的偏导计算公式为

$$\begin{cases}\dfrac{\partial v}{\partial w}=\dfrac{8}{\pi^2}\left(\arctan\left(\dfrac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}}\right)-\arctan\left(\dfrac{w}{h}\right)\right)\dfrac{h}{w^2+h^2}\\ \dfrac{\partial v}{\partial h}=-\dfrac{8}{\pi^2}\left(\arctan\left(\dfrac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}}\right)-\arctan\left(\dfrac{w}{h}\right)\right)\dfrac{w}{w^2+h^2}\end{cases} \tag{5}$$

将$\frac{\partial v}{\partial w}$与$\frac{\partial v}{\partial h}$求和可得

$$\frac{\partial v}{\partial w}+\frac{\partial v}{\partial h}=\frac{8}{\pi^2}\left(\arctan\left(\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}}\right)-\arctan\left(\frac{w}{h}\right)\right)\frac{h-w}{w^2+h^2} \tag{6}$$

从式 (6) 可知, 当预测框的h和w相等时, $\partial v/\partial w+\partial v/\partial h=0$, 此时, 纵横比损失项失效, 并不能提高网络的收敛效果, 反之, 还会增加网络的计算量, 减缓网络的训练速度。纵横比反映的是真实框与预测框宽高比之间的差异, 并不能准确衡量w^{gt}与w或h^{gt}与h之间的差异; 同时, 纵横比涉及反三角函数的计算, 复杂的反向传播增加了网络的计算量。

为解决上述问题, 本文设计了一种高效的边界框回归损失函数, 即 DSGIoU 损失, 在 IoU 的基础

上，添加了非重叠面积、中心点距离及宽高比3个损失项进一步加快网络的收敛性。DSGIoU如图2所示。其中，G表示包含B^{gt}和B的最小矩形框，即灰色虚线矩形框；D表示矩形框G中不含B^{gt}和B并集的部分，即2个黄色矩形框；d为B^{gt}和B的中心点距离，即蓝色线段长度，g为矩形框G的对角距离，即绿色线段长度；w^g和h^g分别表示矩形框G的宽和高。

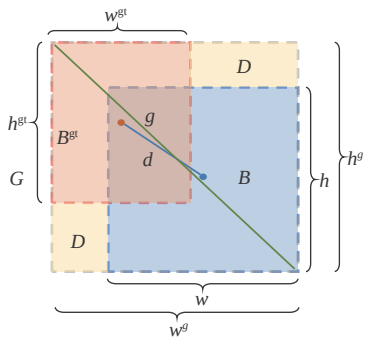


图2 DSGIoU示意图

Fig. 2 Schematic of DSGIoU

DSGIoU损失被定义为

$$\begin{aligned}L_{\mathrm{DSGIoU}} &= L_{\mathrm{GIoU}} + L_{\mathrm{Distance}} + L_{\mathrm{Shape}} = \\ &1 - I + \frac{|G-(B^{gt}\cup B)|}{|G|} + \\ &\frac{p^2(B^{gt},B)}{g^2} + \left(\frac{w^{gt}-w}{w^g}\right)^2 + \left(\frac{h^{gt}-h}{h^g}\right)^2 \end{aligned} \tag{7}$$

式中：$0\leqslant I\leqslant 1$，$0\leqslant |G-(B^{gt}\cup B)|/|G|\leqslant 1$，$0\leqslant p^2(B^{gt},B)/g^2\leqslant 1$，$0\leqslant [(w^{gt}-w)/w^g]^2\leqslant 1$，$0\leqslant [(h^{gt}-h)/h^g]^2\leqslant 1$。因此，$0\leqslant L_{\mathrm{DSGIoU}}\leqslant 5$，$L_{\mathrm{Shape}}$对$w$和$h$的偏导计算公式为

$$\begin{cases}\dfrac{\partial L_{\mathrm{Shape}}}{\partial w} = \dfrac{2w-2w^{gt}}{w^g} \\ \dfrac{\partial L_{\mathrm{Shape}}}{\partial h} = \dfrac{2h-2h^{gt}}{h^g}\end{cases} \tag{8}$$

DSGIoU损失相比于GIoU与DIoU损失的优势在于：如图3所示，检测框出现在真实框内部，并且中心点重合，从式(2)可知，当$B\subseteq B^{gt}$时，$G=B^{gt}$，$B^{gt}\cup B=B^{gt}$，则$\left|G-(B^{gt}\cup B)\right|=0$，此时GIoU便退化成IoU；式(3)中的中心点距离$p^2(B^{gt},B)/g^2=0$，则DIoU也会退化成IoU。而DSGIoU损失通过计算宽高比能够较好地解决此类问题。

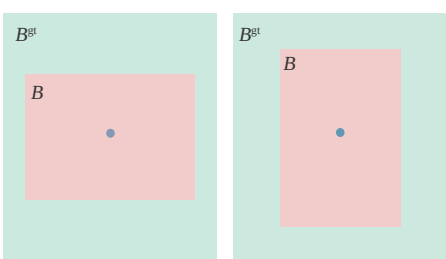


图3 GIoU、DIoU退化为IoU

Fig. 3 GIoU and DIoU degenerating into IoU

与CIoU损失相比较，DSGIoU损失添加了真实框与预测框的非重叠面积，更有利于网络对边界框回归分支的优化；同时，相比于CIoU损失中的纵横比，DSGIoU损失中宽高比L_{Shape}在反向传播过程中仅涉及简单的线性求导过程，并不会带来复杂的计算量；并且L_{Shape}能够准确衡量w^{gt}与w、h^{gt}与h之间的差异，解决了CIoU损失中预测框的w与h相等时反向传播值为0的问题。

1.2 双分支坐标注意力

Hou等[25]将通道注意力分解为2个1D特征编码过程，对输入特征沿着水平和垂直方向进行1D全局池化，沿着2个空间方向聚合特征，捕获长距离依赖并且保留准确的位置信息，将位置信息嵌入通道注意力中。文献[27-28]证明了池化操作通过合并相邻区域神经元信息不仅能够降低计算量，而且能够扩大像素点的感受野，使缩小后的特征图中的像素点获得来自原图像更大范围的信息，提取更高级的语义信息。平均池化对区域内特征点求均值，能够体现出对应区域的平均响应情况；最大池化对区域内特征点求最大值，保留对应区域的最大响应，防止其被周围信息弱化。

基于上述分析，本文提出一种双分支坐标注意力。首先，对输入特征沿着2个方向进行1D全局最大池化和平均池化，得到2组不同的激活权重；其次，将输入分别与2组权重进行点乘操作，充分获取特征的位置坐标信息；最后，对2个分支得到的特征进行相加。

双分支坐标注意力如图4所示，主要包括平均池化和最大池化2个分支。与SE中采用的平均池化操作不同，本文将全局池化转化为1D特征编码操作，对输入特征$\boldsymbol{X}\in\mathbf{R}^{c\times h\times w}$沿着水平方向和垂直方向分别进行最大池化和平均池化操作，特征$\boldsymbol{X}$在2个方向上的均值$\boldsymbol{X}^h_{\mathrm{avg}}$和$\boldsymbol{X}^w_{\mathrm{avg}}$、最大值$\boldsymbol{X}^h_{\mathrm{max}}$和$\boldsymbol{X}^w_{\mathrm{max}}$的计算公式为

$$\begin{cases}\boldsymbol{X}^h_{\mathrm{avg}} = \left[\dfrac{1}{W}\displaystyle\sum_{0\leqslant j<W} x_c(h,j)\right] \\ \boldsymbol{X}^h_{\mathrm{max}} = \left[\max\limits_{0\leqslant j<W}(x_c(h,j))\right] \\ \boldsymbol{X}^w_{\mathrm{avg}} = \left[\dfrac{1}{H}\displaystyle\sum_{0\leqslant i<H} x_c(i,w)\right] \\ \boldsymbol{X}^w_{\mathrm{max}} = \left[\max\limits_{0\leqslant i<H}(x_c(i,w))\right]\end{cases} \tag{9}$$

式中：$0\leqslant w\leqslant W$，$0\leqslant h\leqslant H$，$0\leqslant c\leqslant C$，W、H、C分别表示输入特征$\boldsymbol{X}\in\mathbf{R}^{c\times h\times w}$的宽、高、通道数。

通过对输入特征$\boldsymbol{X}$在水平方向和垂直方向进行聚合信息，不仅能够降低计算量，还能够提取更高级的语义信息，均值$\boldsymbol{X}^h_{\mathrm{avg}}$和$\boldsymbol{X}^w_{\mathrm{avg}}$能够表示特征$\boldsymbol{X}$在某

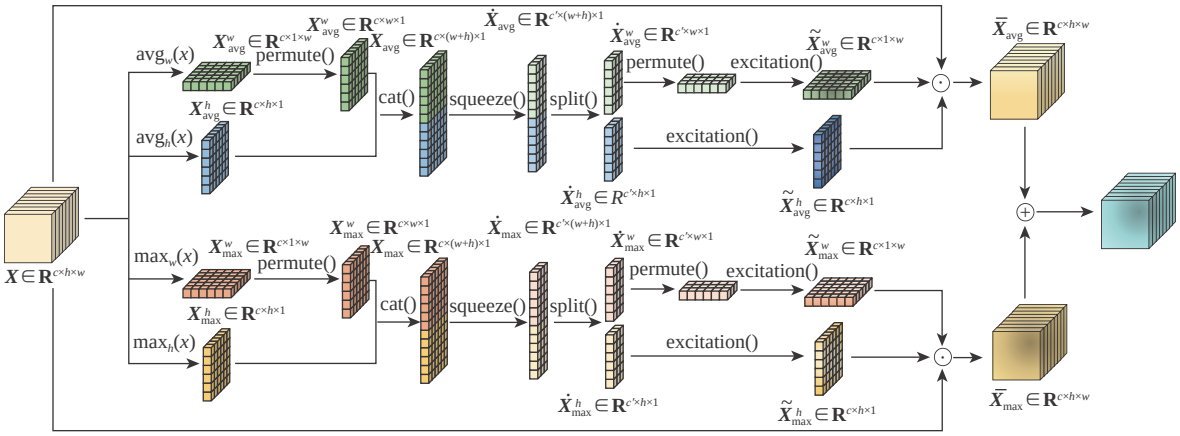


图 4　双分支坐标注意力

Fig. 4　Dual branch coordinate attention

一通道处沿着水平和垂直方向神经元的平均响应, 最大值$\boldsymbol{X}_{\max}^{h}$和$\boldsymbol{X}_{\max}^{w}$则能够保留细节信息, 反映出特征在某一方向维度神经元的最大响应, 防止其被弱化而丢失信息。

由于水平和垂直方向获取的响应图维度不同, 对$\boldsymbol{X}_{\text{avg}}^{h}$和$\boldsymbol{X}_{\max}^{h}$进行维度变换, 使其与$\boldsymbol{X}_{\text{avg}}^{w}$和$\boldsymbol{X}_{\max}^{w}$维度相同。通过式 (10) 对$\boldsymbol{X}_{\text{avg}}^{h}$和$\boldsymbol{X}_{\text{avg}}^{w}$、$\boldsymbol{X}_{\max}^{h}$和$\boldsymbol{X}_{\max}^{w}$沿着垂直方向进行拼接得到$\boldsymbol{X}_{\text{avg}}$和$\boldsymbol{X}_{\max}$。

$$\begin{cases}\boldsymbol{X}_{\text{avg}}=[\boldsymbol{X}_{\text{avg}}^{h},\boldsymbol{X}_{\text{avg}}^{w}]\\ \boldsymbol{X}_{\max}=[\boldsymbol{X}_{\max}^{h},\boldsymbol{X}_{\max}^{w}]\end{cases}\tag{10}$$

式 (11) 能够获得 2 组位置编码信息矩阵$\dot{\boldsymbol{X}}_{\text{avg}}$和$\dot{\boldsymbol{X}}_{\max}$, 其中, δ表示$h_$swish激活函数, BN 表示批量归一化。通过式 (12) 对$\dot{\boldsymbol{X}}_{\text{avg}}$和$\dot{\boldsymbol{X}}_{\max}$在水平和垂直方向拆分为 2 对具有位置信息独立的编码矩阵$\dot{\boldsymbol{X}}_{\text{avg}}^{h}$和$\dot{\boldsymbol{X}}_{\text{avg}}^{w}$、$\dot{\boldsymbol{X}}_{\max}^{h}$和$\dot{\boldsymbol{X}}_{\max}^{w}$, 对这 2 对编码矩阵通过式 (13) 变换为与输入特征$\boldsymbol{X}$具有相同通道数量的矩阵$\tilde{\boldsymbol{X}}_{\text{avg}}^{h}$和$\tilde{\boldsymbol{X}}_{\text{avg}}^{w}$、$\tilde{\boldsymbol{X}}_{\max}^{h}$和$\tilde{\boldsymbol{X}}_{\max}^{w}$, 其中, σ表示Sigmoid激活函数。

$$\begin{cases}\dot{\boldsymbol{X}}_{\text{avg}}=\delta(\text{BN}(\text{Conv}_{\text{avg}}^{1\times1}(\boldsymbol{X}_{\text{avg}})))\\ \dot{\boldsymbol{X}}_{\max}=\delta(\text{BN}(\text{Conv}_{\max}^{1\times1}(\boldsymbol{X}_{\max})))\end{cases}\tag{11}$$

$$\begin{cases}(\dot{\boldsymbol{X}}_{\text{avg}}^{h},\dot{\boldsymbol{X}}_{\text{avg}}^{w})=\text{Split}(\dot{\boldsymbol{X}}_{\text{avg}})\\ (\dot{\boldsymbol{X}}_{\max}^{h},\dot{\boldsymbol{X}}_{\max}^{w})=\text{Split}(\dot{\boldsymbol{X}}_{\max})\end{cases}\tag{12}$$

$$\begin{cases}(\tilde{\boldsymbol{X}}_{\text{avg}}^{h},\tilde{\boldsymbol{X}}_{\text{avg}}^{w})=\sigma(\text{Conv}_{\text{avg_}h}^{1\times1}(\dot{\boldsymbol{X}}_{\text{avg}}^{h}),\text{Conv}_{\text{avg_}w}^{1\times1}(\dot{\boldsymbol{X}}_{\text{avg}}^{w}))\\ (\tilde{\boldsymbol{X}}_{\max}^{h},\tilde{\boldsymbol{X}}_{\max}^{w})=\sigma(\text{Conv}_{\max_h}^{1\times1}(\dot{\boldsymbol{X}}_{\max}^{h}),\text{Conv}_{\max_w}^{1\times1}(\dot{\boldsymbol{X}}_{\max}^{w}))\end{cases}\tag{13}$$

通过式 (14) 将得到的 2 对具有位置信息的编码矩阵分别与输入特征$\boldsymbol{X}$进行点乘, 通过式 (15) 得到具有丰富的位置信息的特征$\boldsymbol{Y}$。

$$\begin{cases}\bar{\boldsymbol{X}}_{\text{avg}}=\boldsymbol{X}\tilde{\boldsymbol{X}}_{\text{avg}}^{h}\tilde{\boldsymbol{X}}_{\text{avg}}^{w}\\ \bar{\boldsymbol{X}}_{\max}=\boldsymbol{X}\tilde{\boldsymbol{X}}_{\max}^{h}\tilde{\boldsymbol{X}}_{\max}^{w}\end{cases}\tag{14}$$

$$\boldsymbol{Y}=\bar{\boldsymbol{X}}_{\text{avg}}+\bar{\boldsymbol{X}}_{\max}\tag{15}$$

2　实验结果与分析

2.1　实验设置

1) 实验环境。操作系统为 Ubuntu18.04, CPU 为 Interl(R) Xeon(R) Gold 6136, GPU 为单张 NVIDIA TITAN V, 显存大小为 12 GB, 深度学习框架为 PyTorch 1.7.0, CUDA 版本为 10.2。

2) 数据集。本文采用 2 种数据集: ①公开基准数据集 PASCAL VOC 2007 和 PASCAL VOC 2012 的组合, 其中, 包含训练图像 16 551 张, 测试图像 4 952 张, 目标类别数为 20。②自动驾驶数据集 KITTI, 其中, 包括 7 481 张图像, 本文对其进行划分, 训练图像 5 984 张, 测试图像 1 497 张, 目标类别数为 3。

3) 训练细节。初始学习率为 3×10^{-5}, 采用 Adam 优化器调整学习率, 学习率下降方式采用余弦退火算法, 动量和衰减系数分别为 0.937 和 0, 学习率最小为 1×10^{-5}; 输入图像采用 mosaic 和 mixup 增强策略, 图像大小为 640×640; Batchsize 设置为 16, 在 PASCAL VOC 数据集上训练 100 个 epoch, 在 KITTI 数据集上训练 150 个 epoch。

4) 评价指标。本文利用平均精度均值 (mean average precision, mAP) 衡量多个类别目标的平均检测精度, 值越高, 表明算法的综合检测性能越好; 同时, 利用每秒帧数 (frame per second, FPS) 评估算法的检测速度。

精确度 P、召回率 R 和平均精度 (average precision, AP) 的计算公式分别为

$$P=\frac{T_{\text{p}}}{T_{\text{p}}+F_{\text{p}}}\tag{16}$$

$$R=\frac{T_{\text{p}}}{T_{\text{p}}+F_{\text{n}}}\tag{17}$$

$$A_P=\int_0^1 P\text{d}R\tag{18}$$

式中：T_p为正确检测框，即预测框与某个真实框标签匹配且IoU大于阈值0.5；F_p为误检框，即将图像背景检测为某个物体，IoU小于阈值0.5；F_n为漏检框，即没有预测框能够识别出该物体。

平均精度从精确度和召回率2个方面评估算法的检测性能，能够准确反映单个类别的检测精度。平均精度均值表示对所有类别的平均精度计算均值。

2.2 定性分析

本文算法与基准算法在PASCAL VOC数据集上的训练损失曲线如图5所示。网络经过90个epoch训练后，损失值趋于稳定，对比可知，本文算法的网络模型收敛性要优于基准算法。为进一步说明本文算法的检测性能，以YOLOX-S为基准，分别在PASCAL VOC和KITTI这2个数据集上进行实验，并对本文算法和基准算法的实验结果进行对比分析。

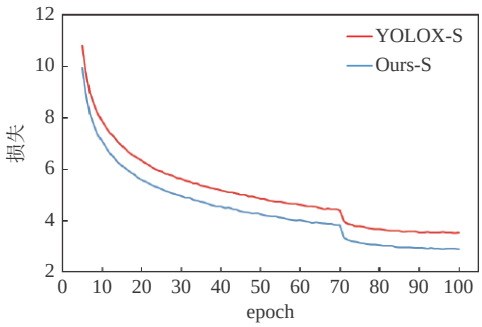


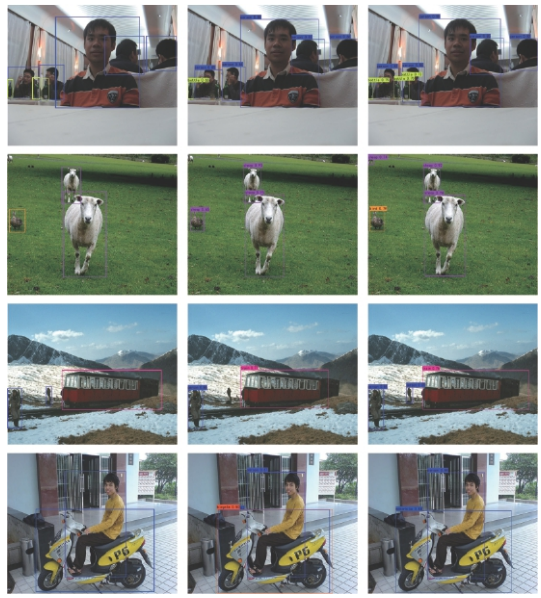
图5 损失曲线对比

Fig. 5 Comparison of loss curves

2.2.1 PASCAL VOC数据集上的定性分析

本文算法与基准算法的检测结果如图6所示。在第1行中，基准算法漏检了2个细小的瓶子，本文算法则能给出较好的检测结果，同时部分目标的分类得分也有所提高；在第2行中，基准算法将左边一个类别为鸟的目标误检为羊，本文算法解决了误检问题，给出该目标正确的分类标签，并且能够检测到图像左上角被遮挡的类别为羊的小目标，且该目标并未被Ground Truth标记；在第3行中，基准算法漏检了一个类别为人的小目标，并且类别为火车的检测框不准确，本文算法则能检测到漏检的人，并且给火车一个准确的检测框；在第4行中，基准算法存在一个类别为自行车的冗余检测框，本文算法能够去除该冗余框，保留更合理的检测结果，解决重复检测的问题。

实验结果表明，本文算法能够有效解决YOLOX在复杂场景中存在漏检、误检及重复检测的问题，保留更准确的检测结果，并且对部分目标能够提高其分类得分。



(a) Ground Truth (b) YOLOX-S (c) Ours-S

图6 PASCAL VOC数据集上的检测结果对比

Fig. 6 Comparison of test results on PASCAL VOC dataset

2.2.2 KITTI数据集上的定性分析

本文算法与基准算法的检测结果如图7所示。在第1行中，基准算法漏检了一个被遮挡的类别为小汽车的极小目标，本文算法能够检测到该极小目标；在第2行中，基准算法漏检了左边被遮挡的小汽车，本文算法则能成功检测到该目标；在第3行中，基准算法漏检了一个类别为骑自行车的人，本文算法则能检测到该目标。结果表明，本文算法对漏检问题提升明显，能够成功检测到小目标及被遮挡的目标。



(a) Ground Truth (b) YOLOX-S (c) Ours-S

图7 KITTI数据集上的检测结果对比

Fig. 7 Comparison of test results on KITTI dataset

2.3 定量分析

2.3.1 PASCAL VOC数据集上的定量分析

1）与基准算法的性能对比。为进一步说明本文算法的检测性能，以不同网络大小的YOLOX算法为基准在PASCAL VOC数据集上进行实验，具体结果如表1所示。在基本不增加网络参数的情况下，本文算法相比于不同网络大小的基准算法，检测性能均能有所提升。

2）与主流算法的整体性能对比。本文算法与

表 1　PASCAL VOC 数据集测试结果

Table 1　Test results on PASCAL VOC dataset

算法	模型大小/Mbit	mAP/%	检测速度/(帧·s^{-1})
YOLOX-Tiny	5.04	78.5	64.1
Ours-Tiny	5.09	80.0	35.6
YOLOX-S	8.95	81.0	60.5
Ours-S	9.02	82.6	31.5
YOLOX-M	25.3	83.8	45.8
Ours-M	25.6	85.8	19.2

当前检测性能优秀的算法进行对比, 其中, 包括单阶段目标检测算法、双阶段目标检测算法及无锚框目标检测算法。具体结果如表 2 所示。本文算法的检测精度为 82.6%, 相比于基准算法提升 1.6%。与其他主流算法对比, 本文算法相比于双阶段目标检测算法 Faster R-CNN、R-FCN 的检测精度分别提升 6.2% 和 3.1%, 检测速度分别提升 29.1 帧/s 和 25.7 帧/s; 与单阶段目标检测算法 SSD、DSSD、YOLOv3 相比, 检测精度分别提升 5.8%、1.1%、3.3%, 检测速度分别提升 12.5 帧/s、26 帧/s、5.5 帧/s; 与无锚框目标检测算法 FCOS 和 CenterNet 相比, 检测精度分别提升 2.4%、3.9%, 检测速度分别提升 15.5 帧/s、1.5 帧 /s; 与 CenterNet-DHRNet、CenterNet-Res101-FcaNet 相比, 检测精度分别提升 0.7%、0.3%, 检测速度提升 13.5 帧/s、3.9 帧/s; 与 YOLOv4、YOLOv7 相比, 检测精度分别提升 0.6%、0.3%。

表 2　各类算法在 PASCAL VOC 数据集上的测试结果

Table 2　Test results of various algorithms on PASCAL VOC dataset

算法	主干网络	图像分辨率	mAP/%	检测速度/(帧·s^{-1})
YOLOv3[7]	DarkNet-53	544×544	79.3	26.0
YOLOv4[8]	CSPDarkNet-53	448×448	82.0	43.0
SSD[9]	VGG-16	512×512	76.8	19.0
DSSD[10]	ResNet-101	513×513	81.5	5.5
Faster R-CNN[13]	ResNet-101	1 000×600	76.4	2.4
YOLOX-S[26]	Modified CSPv5	640×640	81.0	60.5
R-FCN[29]	ResNet-101	1 000×600	79.5	5.8
FCOS[30]	ResNet-50	800×800	80.2	16.0
CenterNet[31]	ResNet-101	512×512	78.7	30.0
CenterNet-DHRNet[32]	DHRNet	512×512	81.9	18.0
CenterNet-Res101-FcaNet[33]	Res101-FcaNet	512×512	82.3	27.6
YOLOv7[34]		640×640	82.3	45.0
Ours-S	Modified CSPv5	640×640	82.6	31.5

3) 与主流算法的单类性能对比。表 3 为各类检测算法在 PASCAL VOC 数据集上每个类别的检测精度。实验结果表明, Ours-S 在 14 个类别上能够达到最优水平, 说明本文算法性能较好, 其中, 类别为汽车、桌子、沙发、火车及电视机的检测精度提升 1.0% 以上; 在桌子、椅子等易受到遮挡的类别

表 3　各类算法在 PASCAL VOC 数据集上每个类别的平均精度比较

Table 3　AP comparison of various algorithms for each category on PASCAL VOC dataset　%

类别	Ours-S	YOLOX-S[26]	Faster R-CNN[13]	R-FCN[29]	SSD[9]	RetinaNet[35]	CenterNet-DHRNet[32]
aero	89.5	86.5	79.8	82.5	82.4	89.4	86.2
bike	90.4	89.5	80.7	83.7	84.7	86.6	88.6
bird	79.2	77.3	76.2	80.3	78.4	79.8	82.4
boat	74.6	73.9	68.3	69.0	73.8	67.8	72.8
bottle	73.5	71.6	55.9	69.2	53.2	70.8	73.4
bus	88.7	88.2	85.1	87.5	86.2	85.4	86.6
car	93.1	91.9	85.3	88.4	87.5	90.5	88.8
cat	87.5	87.4	89.8	88.4	86.0	88.8	87.3
chair	69.0	66.7	56.7	65.4	57.8	61.0	68.1
cow	83.4	82.0	87.8	87.3	83.1	75.6	86.9
table	80.9	79.6	69.4	72.1	70.2	65.8	78.4
dog	84.6	82.9	88.3	87.9	84.9	84.1	84.6
horse	89.3	89.1	88.9	88.3	85.2	84.4	88.5
mbike	87.6	86.7	80.9	81.3	83.9	84.9	86.5
person	89.1	88.7	78.4	79.8	79.7	85.7	86.0
plant	57.7	53.9	41.7	54.1	50.3	52.1	59.0
sheep	79.1	78.3	78.6	79.6	77.9	77.7	85.3
sofa	82.8	79.8	79.8	78.8	73.9	74.2	81.5
train	88.5	86.3	85.3	87.1	82.5	85.8	87.5
tv	83.1	79.0	72	79.5	75.3	79.6	80.2

上，本文算法也能够取得最佳检测结果，相较于其他算法能够准确检测到被遮挡的目标，具有良好的鲁棒性。

2.3.2 KITTI数据集上的定量分析

1）与基准算法的性能对比。本文算法与基准算法在KITTI数据集上的对比结果如表4所示，本文算法的检测精度均高于基准算法。

表4 KITTI数据集测试结果

Table 4 Test results on KITTI dataset

算法	模型大小/Mbit	mAP/%	检测速度/(帧·s^{-1})
YOLOX-Tiny	5.04	86.0	64.1
Ours-Tiny	5.09	87.7	35.6
YOLOX-S	8.95	86.8	60.5
Ours-S	9.02	89.7	31.5
YOLOX-M	25.3	89.4	45.8
Ours-M	25.6	90.7	19.2

2）与主流算法的整体性能对比。本文算法与其他经典的目标检测算法在KITTI数据集上的对比结果如表5所示。实验结果表明，本文算法与基准算法相比，检测精度提高了2.9%；相比于单阶段目标检测算法SSD、YOLOv3、AM-YOLOv3，检测精度分别提高了28.5%、4.8%、3.7%，检测速度分别提高了2.6帧/s、2.5帧/s、5.5帧/s；与无锚框目标检测算法CenterNet、CenterNet-DHRNet相比，检测精度分别提高了3.6%、2.6%，检测速度分别提高了1.5帧/s、13.5帧/s。

表5 各类算法在KITTI数据集上的测试结果

Table 5 Test results of various algorithms on KITTI dataset

算法	主干网络	图像分辨率	mAP/%	检测速度/(帧·s^{-1})
YOLOv3[7]	DarkNet-53	544×544	84.9	29.0
SSD[9]	VGG-16	512×512	61.2	28.9
YOLOX-S[26]	Modified CSPv5	640×640	86.8	60.5
CenterNet[31]		512×512	86.1	30.0
CenterNet-DHRNet[32]	DHRNet	512×512	87.1	18.0
AM-YOLOv3[36]	DarkNet-53	544×544	86.0	26.0
Ours-S	Modified CSPv5	640×640	89.7	31.5

3）与主流算法的单类性能对比。表6为各类检测算法在KITTI数据集上每个类别的检测精度。实验结果表明，本文算法每个类别的检测精度均高于SSD、YOLOv3、AM-YOLOv3、CenterNet、YOLOX-S，其中，在行人和骑自行车的人这2类上，本文算法较基准算法分别提高了4.4%和3.9%。

表6 各类算法在KITTI数据集上每个类别的平均精度比较

Table 6 AP comparison of various algorithms for each category on KITTI dataset %

类别	YOLOv3[7]	SSD[9]	YOLOX-S[26]	CenterNet[31]	CenterNet-DHRNet[32]	AM-YOLOv3[36]	Ours-S
Pedestrian	76.9	48.0	81.2	78.3	79.1	78.6	85.6
Car	94.1	85.1	95.1	95.4	96.7	94.5	95.5
Cyclist	83.8	50.6	84.0	84.6	85.5	84.9	87.9

2.4 消融实验

为验证本文所提2种方法对检测结果的有效性，以YOLOX-S为基准，在PASCAL VOC数据集上进行了不同模块的训练及测试。实验结果如表7所示。基准算法的检测精度为81.0%，在YOLOX-S中采用DSGIoU损失作为边界框回归损失函数，检测精度提升至82.0%，说明DSGIoU损失能够使目标定位更准确，网络收敛更好；将双分支坐标注意力嵌入到YOLOX-S中，检测精度提升至82.0%，表明双分支坐标注意力能够动态学习通道之间的相关性，增强特征表示能力，使网络聚焦于关键特征，抑制无关信息；在YOLOX-S中同时使用上述2种方法，检测精度提升至82.6%，验证了本文算法的有效性。

表7 在PASCAL VOC数据集上的消融实验

Table 7 Ablation experiment on PASCAL VOC dataset

YOLOX-S	DSGIoU损失	DBCA	mAP/%
√			81.0
√	√		82.0
√		√	82.0
√	√	√	82.6

为说明DSGIoU损失的有效性，与其他主流的边界框回归损失函数进行对比，实验结果如表8所示。GIoU损失、DIoU损失、CIoU损失对检测精度均有所提升，DSGIoU损失通过优化边界框宽高比及引入非重叠面积，综合考量边界框的重叠面

表8 不同边界框回归损失的测试结果

Table 8 Test results of different bounding box regression losses

方法	mAP/%	检测速度/(帧·s^{-1})
IoU损失	81.0	60.5
GIoU损失	81.4	63.0
DIoU损失	81.3	63.4
CIoU损失	81.4	64.5
DSGIoU损失	82.0	63.5

积、非重叠面积、中心点距离及宽高比, 从而达到优化边界框回归的效果。检测精度从 81.0% 提升到了 82.0%, 提升效果优于其他损失函数, 检测速度也较为稳定。

为验证双分支坐标注意力的有效性, 与其他主流的注意力模块进行对比, 实验结果如表 9 所示。双分支坐标注意力相比经典的 SENet 和 CBAM, 检测精度提高了 0.2% 和 0.4%; 与不增加网络参数的 GCT 相比, 检测精度提升 0.4%; 与 CA 相比, 检测精度提升 0.4%。结果表明, 双分支坐标注意力能够进一步增强特征表示能力, 提高检测精度。

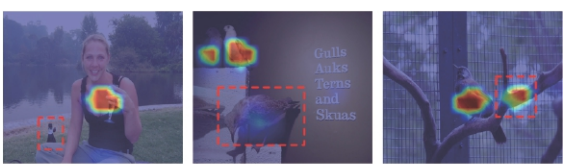
表 9　不同注意力模块的测试结果

Table 9　Test results of different attention modules

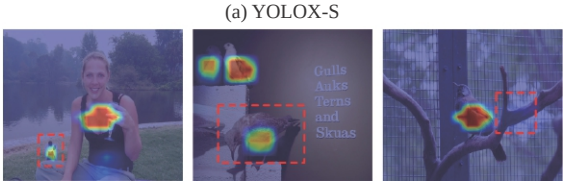
方法	mAP/%	检测速度/(帧·s^{-1})	模型大小/Mbit
YOLOX-S	81.0	60.5	8.95
SENet[22]	81.8	48.3	8.97
CBAM[23]	81.6	43.5	8.98
GCT[24]	81.6	53.7	8.95
CA[25]	81.6	40.8	8.98
DBCA	82.0	32.3	9.02

2.5　可视化热力图分析

可视化热力图如图 8 所示。其中, 图 8(a)和图 8(b)分别为 YOLOX-S 和本文算法在 PASCAL VOC 数据集上的可视化热力图, 在前 2 列图像中, 基准算法不能准确地聚焦于图像中的全部目标, 存在漏检的问题; 在第 3 列图像中, 基准算法关注到与目标无关的背景信息, 导致误检。反观本文算法, 采用双分支坐标注意力模块提高特征的语义信息, 通过 DSGIoU 损失提高边界框的定位准确度, 不仅使算法能够关注到图像中的所有目标, 而且能消除噪声, 不再关注背景信息。结果显示, 本文算法能够有效解决漏检及误检问题。



(a) YOLOX-S



(b) Ours-S

图 8　可视化热力图

Fig. 8　Visualized heatmap

3　结　论

1) 本文提出一种基于 DSGIoU 损失与双分支坐标注意力的目标检测算法, 在 PASCAL VOC 数据集和 KITTI 数据集上, mAP 分别达到 82.6% 和 89.7%, 相比于基准算法分别提高了 1.6% 和 2.9%。

2) DSGIoU 损失函数从交并比、非重叠面积、中心点距离及长宽比 4 个方面回归精准的检测框。在 PASCAL VOC 数据集上验证其先进性, 相比于 GIoU、DIoU 及 CIoU 损失, mAP 分别提升 0.6%、0.7%、0.6%, 且检测速度基本保持不变。

3) 双分支坐标注意力采用不同的池化方式沿着水平和垂直方向获取方向感知信息, 并与经典的注意力机制在 PASCAL VOC 数据集上进行比较, 检测精度能够达到最高, 但随着参数量的增加, 检测速度有所降低。

为提升算法的检测速度, 在后续工作中将进一步优化双分支坐标注意力, 降低其参数量。

参考文献（References）

[1] LIU L, OUYANG W L, WANG X G, et al. Deep learning for generic object detection: A survey[J]. International Journal of Computer Vision, 2020, 128(2): 261-318.

[2] DENG L J, GONG Y X, LIN Y, et al. Detecting multi-oriented text with corner-based region proposals[J]. Neurocomputing, 2019, 334: 134-142.

[3] ZABLOCKI É, BEN-YOUNES H, PÉREZ P, et al. Explainability of deep vision-based autonomous driving systems: review and challenges[J]. International Journal of Computer Vision, 2022, 130(10): 2425-2452.

[4] LIU Y H, ZHANG F D, ZHANG Q Y, et al. Cross-view correspondence reasoning based on bipartite graph convolutional network for mammogram mass detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 3811-3821.

[5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 779-788.

[6] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6517-6525.

[7] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL].(2018-04-08)[2023-04-17].https://arxiv.org/abs/1804.02767.

[8] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2023-04-17]. https://arxiv.org/abs/2004.10934.

[9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.

[10] FU C Y, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector[EB/OL]. (2017-01-23)[2023-04-17]. https://arxiv.org/abs/1701.06659.

[11] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hie-

rarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 580-587.

[12] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 1440-1448.

[13] 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.

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[15] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Proceedings of the 16th European Conference on Computer Vision. Berlin: Springer, 2020: 213-229.

[16] DAI Z G, CAI B L, LIN Y G, et al. UP-DETR: unsupervised pre-training for object detection with Transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 1601-1610.

[17] YU J H, JIANG Y N, WANG Z Y, et al. UnitBox: an advanced object detection network[C]//Proceedings of the 24th ACM International Conference on Multimedia. New York: ACM, 2016: 516-520.

[18] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 658-666.

[19] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 12993-13000.

[20] GUO M H, XU T X, LIU J J, et al. Attention mechanisms in computer vision: a survey[J]. Computational Visual Media, 2022, 8(3): 331-368.

[21] 康涛, 段蓉凯, 杨磊, 等. 融合多注意力机制的卷积神经网络轴承故障诊断方法[J]. 西安交通大学学报, 2022, 56(12): 68-77.
KANG T, DUAN R K, YANG L, et al. Bearing fault diagnosis using convolutional neural network based on a multi-attention mechanism[J]. Journal of Xi'an Jiaotong University, 2022, 56(12): 68-77(in Chinese).

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

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

[24] RUAN D S, WANG D Y, ZHENG Y, et al. Gaussian context Transformer[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 15124-15133.

[25] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 13708-13717.

[26] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL].(2021-08-06)[2023-04-17].https://arxiv.org/abs/2107.08430.

[27] BOUREAU Y L, PONCE J, LECUN Y. A theoretical analysis of feature pooling in visual recognition[C]//Proceedings of the 27th International Conference on Machine Learning. New York: ACM, 2010: 111-118.

[28] BOUREAU Y L, BACH F, LECUN Y, et al. Learning mid-level features for recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2010: 2559-2566.

[29] DAI J F, LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: ACM, 2016: 379-387.

[30] TIAN Z, SHEN C H, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 9626-9635.

[31] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as points[EB/OL]. (2019-04-25)[2023-04-17]. https://arxiv.org/abs/1904.07850.

[32] 王新, 李喆, 张宏立. 一种迭代聚合的高分辨率网络 Anchor-free 目标检测方法[J]. 北京航空航天大学学报, 2021, 47(12): 2533-2541.
WANG X, LI Z, ZHANG H L. High-resolution network Anchor-free object detection method based on iterative aggregation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(12): 2533-2541(in Chinese).

[33] 侯志强, 郭浩, 马素刚, 等. 基于双分支特征融合的无锚框目标检测算法[J]. 电子与信息学报, 2022, 44(6): 2175-2183.
HOU Z Q, GUO H, MA S G, et al. Anchor-free object detection algorithm based on double branch feature fusion[J]. Journal of Electronics & Information Technology, 2022, 44(6): 2175-2183(in Chinese).

[34] 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[EB/OL]. (2022-07-06)[2023-04-17]. https://arxiv.org/abs/2207.02696.

[35] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(2): 318-327.

[36] 鞠默然, 罗江宁, 王仲博, 等. 融合注意力机制的多尺度目标检测算法[J]. 光学学报, 2020, 40(13): 132-140.
JU M R, LUO J N, WANG Z B, et al. Multi-scale target detection algorithm based on attention mechanism[J]. Acta Optica Sinica, 2020, 40(13): 132-140(in Chinese).

Object detection algorithm based on DSGIoU loss and dual branch coordinate attention

MA Sugang[1, 2, *]，LI Ningbo[1]，HOU Zhiqiang[1, 2]，YU Wangsheng[3]，YANG Xiaobao[1]

(1. School of Computer Science & Technology，Xi'an University of Posts & Telecommunications，Xi'an 710121，China；
2. Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing，Xi'an University of Posts & Telecommunications，Xi'an 710121，China；
3. School of Information and Navigation，Air Force Engineering University，Xi'an 710077，China)

Abstract：The bounding box regression loss effect is limited, and the multi-scale feature representation ability is insufficient in the YOLOX algorithm, which leads to inaccurate detection results. To address this issue, an object detection algorithm based on distance shape of generalized intersection over union (DSGIoU) loss and dual branch coordinate attention was proposed. Based on the intersection over union (IoU) loss term, the regression convergence effect of the bounding box was optimized by adding three penalty terms: non-overlapping area, distance from the center, and aspect ratio between the true box and the predicted box. Meanwhile, the feature was encoded in two directions by using average pooling and max pooling to obtain directional perception information and position information, so as to enhance the feature. To demonstrate the detection performance of the proposed algorithm, YOLOX with network sizes of Tiny, S, and M was used as the benchmark to carry out tests on PASCAL VOC and KITTI datasets. The experimental results show that the detection accuracy of the proposed algorithm on the PASCAL VOC dataset reaches 80.0%, 82.6%, and 85.8%, respectively, which is 1.5%, 1.6%, and 2.0% higher than the YOLOX as the benchmark. On the KITTI dataset, the detection accuracy reaches 87.7%, 89.7%, and 90.7%, which is increased by 1.7%, 2.9%, and 1.3%, respectively. The proposed algorithm can optimize the network convergence, improve the representation ability of multi-scale features, and significantly boost the detection accuracy.

Keywords：object detection; loss function; bounding box regression; coordinate attention; YOLOX

Received：2023-04-21；**Accepted**：2023-05-15；**Published Online**：2023-06-19 15：58

URL：link.cnki.net/urlid/11.2625.V.20230619.1244.001

Foundation items：National Natural Science Foundation of China (62072370); Natural Science Foundation of Shaanxi Province (2023-JC-YB-598); Science and Technology Project of Xi'an City (22GXFW0125)

*** Corresponding author**. E-mail：msg@xupt.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0172

颜色一致性的无人机航拍低光照图像增强算法

王殿伟[1,*]，刘旺[1]，房杰[1]，许志杰[2]

(1. 西安邮电大学 通信与信息工程学院，西安 710121；2. 哈德斯菲尔德大学 计算机与工程学院，哈德斯菲尔德 HD1 3DH)

摘 要：针对无人机 (UAV) 在低光照条件下拍摄的图像存在亮度低、视觉效果差等问题，建立了无人机航拍低光照数据集，提出了一种颜色一致性的无人机航拍低光照图像增强算法。在亮度增强阶段，构建亮度增强网络 (BENet) 来增强图像的亮度，利用颜色学习网络 (CNet) 模块和金字塔颜色嵌入 (PCE) 模块将图像的颜色特征和内容特征相结合，避免增强后的图像出现颜色失真；在图像校正阶段，构建基于域转移的校正网络，使用自建的数据集训练网络，借助光照良好图像来校正亮度增强阶段增强后的图像，减少噪声对图像的影响，得到增强后的图像。实验结果表明：所提算法可提升图像亮度，避免出现噪声放大和颜色失真的问题，在客观指标上，整体优于对比算法，同时还能提升目标检测算法在夜间的检测性能。

关 键 词：低光照图像增强；无人机航拍图像；亮度增强；颜色一致；图像去噪

中图分类号：TP391.41

文献标志码：A **文章编号**：1001-5965(2025)04-1096-11

无人机 (unmanned aerial vehicle, UAV) 能广泛应用于各种领域，如侦查、公共安全、救灾等，但由于环境因素的影响，导致其在低光照环境下拍摄的图像质量低，不满足人们的要求，而且会影响后续图像算法的性能[1]。因此，无人机航拍低光照图像增强已经成为研究热点。

目前，常用的低光照图像增强算法可分为传统算法和基于深度学习算法两大类。Pizer 等[2] 提出了自适应的直方图均衡化方法，有效增强了图像的细节和亮度。Retinex 是低光照图像增强算法中经常使用的理论基础[3]，其将图像分解为光照分量和反射分量，再分别对 2 个分量做处理，得到增强后的结果。Guo 等[4] 提出了 LIME 的经典算法，选取像素中的最大值初始化光照图，再使用结构先验细化光照图，根据 Retinex 理论合成图像。数据驱动的图像增强算法通常指基于深度学习的增强算法，主要包括有监督学习和无监督学习。有监督学习通过成对匹配的训练集训练网络。Wei 等[5] 将深度学习和 Retinex 理论相结合，构建了分解网络和增强网络，实现了低光照图像增强。Zhang 等[6] 构建了 KinD 网络，使用分解网络将图像分解为光照分量和反射分量，再分别使用 2 个子网络对分量进行增强处理，同时，还通过多尺度注意模块解决了 KinD 视觉效果的问题，称为 KinD++[7]。Lv 等[8] 提出了一种端到端的多分支增强网络，通过特征提取模块、增强模块和融合模块协同来提升低光照图像质量。孙帮勇等[9] 提出了基于多头注意力的低光照图像增强算法，使用注意力机制获取图像全局特征，有效提升了图像的亮度，避免了噪声放大。Xu 等[10] 提出了基于频率分解的增强网络，在低频层中恢复图像内容，在高频层中增强图像细节，解决了噪声放大的问题。张雅荔等[11] 提出了一种融合注意力引导的多尺度低光照图像增强算法，增强了图像的亮度和细节信息。上述算法都是通过成对数据集

收稿日期：2023-04-10；**录用日期**：2023-05-19；**网络出版时间**：2023-06-27 16:19
网络出版地址：link.cnki.net/urlid/11.2625.V.20230627.1302.001
基金项目：国家自然科学基金青年基金 (62201454)；陕西省国际科技合作计划 (2023-GHYB-04)
* **通信作者**. E-mail：wangdianwei@xupt.edu.cn

引用格式：王殿伟，刘旺，房杰，等. 颜色一致性的无人机航拍低光照图像增强算法 [J]. 北京航空航天大学学报，2025，51（4）：1096-1106.
WANG D W，LIU W，FANG J，et al. Low illumination image enhancement algorithm for UAV aerial photography with color consistency [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1096-1106（in Chinese）.

训练网络, 有监督地学习图像之间的映射关系, 但成对数据集的建立存在很大困难, 因此, 学者们开始以无监督方式来训练网络。Jiang 等[12] 构建了 EnlightenGAN 网络模型, 使用注意力引导的 U-Net 网络作为生成器, 以全局-局部判别器来鉴别生成图像的真伪, 实现图像增强。向森等[13] 提出了一种双重迭代的低光照图像增强算法, 不需要任何训练数据集, 有效提升了图像的亮度。王殿伟等[14] 提出了一种受生物视觉启发的低光照图像增强算法, 将图像分解为结构通路和细节通路分别处理, 有效增强了图像的亮度和细节。Jin 等[15] 通过分解网络得到图像的光照层, 再使用光照层引导网络增强图像的亮度。上述算法都能有效提升图像的亮度, 但增强后的图像往往会导致颜色失真和噪声放大问题。

为解决上述问题, 本文提出颜色一致性的无人机航拍低光照图像增强算法, 将低光照图像增强分为 2 个阶段: 亮度增强阶段和图像校正阶段。在亮度增强阶段, 为避免亮度增强后的图像产生颜色失真, 本文构建了亮度增强网络(brightness enhancement network, BENet), 使用颜色学习网络(color network, CNet)模块学习图像中的颜色信息, 通过金字塔颜色嵌入(pyramid color embedding, PCE)模块将颜色信息和内容信息相结合, 在亮度增强的过程中结合了图像的颜色信息; 在图像校正阶段, 本文通过学习光照良好图像特征来校正亮度增强阶段增强后的图像, 使用噪声编码器和内容编码器来解耦图像的内容和噪声, 减少噪声对图像的影响, 同时使用自建的无人机数据集训练网络, 使校正后的图像接近于无人机拍摄的真实图像。

1　本文算法

为避免无人机航拍低光照图像增强后出现颜色失真和噪声放大等问题, 本文提出了一种两阶段的增强算法。第一阶段, 构建了颜色一致性的亮度增强网络, 避免在亮度增强的同时带来颜色失真。第二阶段, 使用域转移的思想来减少噪声对图像的影响。网络结构如图 1 所示。

1.1　亮度增强网络

无人机航拍低光照图像增强旨在提升图像的亮度信息, 但如果在增强的过程中只考虑图像的亮

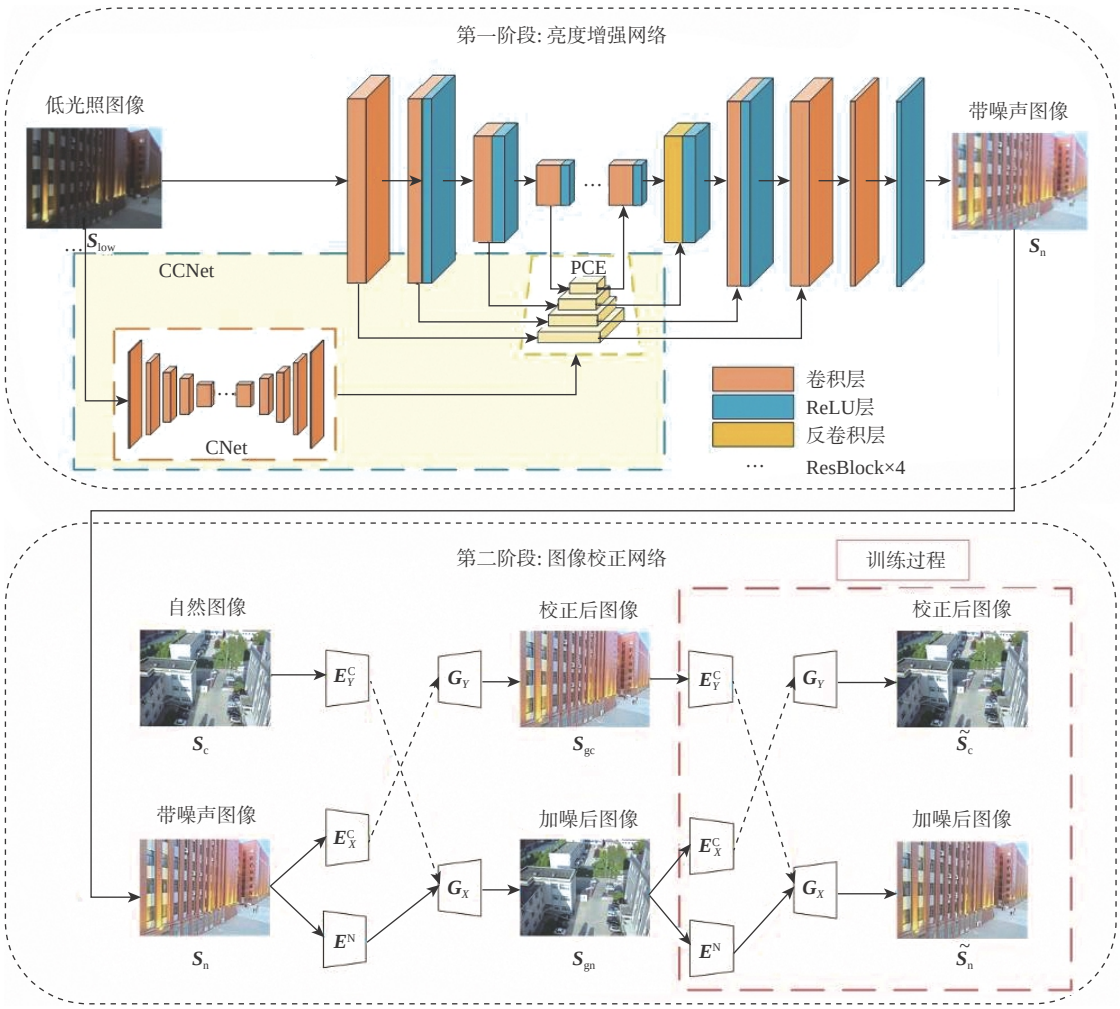


图 1　本文网络结构

Fig. 1　Structure of the proposed network

度信息，往往会导致增强后的图像出现颜色失真，不符合人眼的视觉感知。为此，本文构建了一种颜色一致性亮度增强网络，网络结构如图1所示。亮度增强网络是基于U-Net[16]网络改进的，额外添加颜色一致性网络（color consistency network, CCNet），CCNet能够保证增强前后图像颜色一致性，包含CNet模块和PCE模块。

CNet模块基于计算图像颜色直方图设计，目的是学习低光照图像颜色直方图到正常光照图像颜色直方图的映射关系，从而获得与正常光照图像一致的颜色特征。模块结构如图1所示。CNet是一种编解码的网络结构，使用有监督学习来训练网络，使网络能将输入的低光照图像颜色直方图转化为预测的颜色直方图，可表示为

$$\boldsymbol{C}_{\text{pre}} = \text{CNet}(\boldsymbol{S}_{\text{low}}) \tag{1}$$

式中：$\boldsymbol{C}_{\text{pre}}$为预测的颜色直方图；CNet表示网络计算；$\boldsymbol{S}_{\text{low}}$为输入的低光照图像。

为更好地重建颜色直方图，本文使用L_{c}损失来约束网络，使网络专注于学习图像颜色特征。L_{c}损失可表示为

$$L_{\text{c}} = \frac{1}{N\times 256}\|\boldsymbol{C}_{\text{pre}} - \boldsymbol{C}_{\text{high}}\|_1 \tag{2}$$

式中：$\boldsymbol{C}_{\text{high}}$为正常光照的颜色直方图；$N$=3，对应RGB的3个通道。

由于颜色直方图中不包含图像的空间信息，直接将CNet学习到的颜色特征输入到亮度增强网络中会影响图像增强效果。为将颜色特征和空间特征相结合，受文献[17]启发，采用PCE模块融合图像的颜色信息和空间信息，与原PCE不同的是：①调整了PCE的层次结构，从原始的6层结构变为更加适应本文亮度增强网络的4层结构，减少了参数量，这种渐进式结构也不会对网络整体性能具有较大影响；②在使用PCE模块进行每层颜色特征和内容特征匹配时，为改变颜色特征的分辨率大小，使用双线性插值卷积的上采样方式代替原有的反卷积上采样模式，有效避免了颜色特征上采样后出现棋盘效应。PCE模块结构如图2所示。

PCE模块中，颜色嵌入（color embedding, CE）结构用来解决颜色特征和内容特征不匹配问题，该结构使用双亲和矩阵（dual affinity matrix, DAM）来融合图像的颜色特征和特征内容。具体来说，先计算大小为$N\times H\times W$的颜色特征$\boldsymbol{C}$和内容特征$\boldsymbol{F}$的哈曼顿距离和内积，如下：

$$M(x,y) = -\|\boldsymbol{F}(x,y) - \boldsymbol{C}(x,y)\|_1 \tag{3}$$

$$P(x,y) = \boldsymbol{F}(x,y)\boldsymbol{C}(x,y) \tag{4}$$

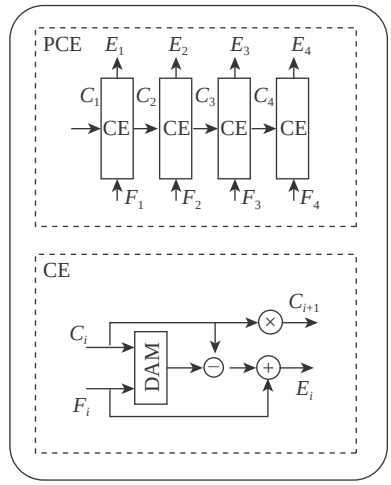


图2 PCE模块结构

Fig. 2 Structure of PCE module

式中：M和P分别为颜色特征和内容特征的哈曼顿距离和哈曼顿内积。

双亲和矩阵$\boldsymbol{A}$计算如下：

$$\boldsymbol{A} = 2\text{sigmoid}(M)\cdot\tanh(P) \tag{5}$$

式中：sigmoid和tanh为激活函数。

通过双亲和矩阵$\boldsymbol{A}$来匹配图像的内容信息和颜色信息，如下：

$$\boldsymbol{E} = \boldsymbol{A}\odot\boldsymbol{C} + \boldsymbol{F} \tag{6}$$

式中：$\boldsymbol{E}$为融合后的图像特征。

CE结构可以匹配每层的颜色和内容信息，使用PCE模块调整每层输入颜色特征和内容特征的大小，再通过每层的CE结构重新计算融合的图像特征。CE结构可以通过双亲和矩阵有效地将图像的颜色特征和内容特征相融合。

本文在亮度增强阶段使用重建损失L_{recon}、颜色损失L_{c}和结构相似损失L_{ssim}作为约束训练网络。L_{recon}损失能约束增强后图像和真实图像之间的差异，L_{c}损失用来恢复图像的颜色直方图，L_{ssim}损失用来约束2幅图像之间的结构相似度。L_{recon}损失函数表示为

$$L_{\text{recon}} = \|\boldsymbol{F}_{\text{net}}(\boldsymbol{S}_{\text{low}}) - \boldsymbol{S}_{\text{high}}\|_1 \tag{7}$$

式中：$\boldsymbol{S}_{\text{high}}$为正常光照图像；$\boldsymbol{F}_{\text{net}}(\boldsymbol{S}_{\text{low}})$为经过网络增强后的图像。

L_{ssim}损失如表示为

$$L_{\text{ssim}} = 1 - \text{SSIM}(\boldsymbol{S}_{\text{low}}, \boldsymbol{S}_{\text{high}}) \tag{8}$$

式中：相似性损失函数SSIM为

$$\text{SSIM}(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}\cdot\frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{9}$$

其中：μ_x和μ_y为图像的均值；σ_x和σ_y为图像的方差；c_1和c_2表示不为0的常数。

L_1损失表示为

$$L_1 = L_{\text{recon}} + L_{\text{c}} + L_{\text{ssim}} \tag{10}$$

式中: L_1为亮度增强阶段的整体损失函数。

1.2　图像校正网络

第一阶段的亮度增强网络可以有效提升图像的亮度信息, 还能避免出现颜色失真, 但亮度提升的同时也带来了噪声放大问题。受文献 [18] 启发, 本文使用两阶段的处理方式, 在图像校正阶段减少噪声对图像的影响。本文采用域转移的思想, 将图像从噪声域转移到清晰域中完成图像去噪, 并使用自建的无人机数据集训练校正网络, 使校正网络能够更加有效地针对无人机拍摄的低光照图像, 网络结构如图 1 所示。图像校正网络由 6 部分组成: ①内容编码器$\boldsymbol{E}_X^{\text{C}}$和$\boldsymbol{E}_Y^{\text{C}}$, 这 2 个编码器之间共享参数。②噪声编码器$\boldsymbol{E}^{\text{N}}$, 学习图像中的噪声分布。③噪声域图像生成器$\boldsymbol{G}_X$, 生成一个带噪声的图像。④清晰域图像生成器$\boldsymbol{G}_Y$, 生成一个清晰的图像。⑤噪声域图像鉴别器$\boldsymbol{D}_X$。⑥清晰域图像鉴别器$\boldsymbol{D}_Y$, 保证生成图像的可靠性。2 个生成器的网络结构与 U-Net[16] 中网络结构相同, 使用编解码网络。

图像校正网络的训练流程为: 首先, 将第一阶段增强后得到的带噪声图像$\boldsymbol{S}_{\text{n}}$和光照良好的图像$\boldsymbol{S}_{\text{c}}$输入到图像校正网络中, 使用内容编码器$\boldsymbol{E}_X^{\text{C}}$和噪声编码器$\boldsymbol{E}^{\text{N}}$将图像$\boldsymbol{S}_{\text{n}}$分解为噪声特征和内容特征, 使用内容编码器$\boldsymbol{E}_Y^{\text{C}}$将图像$\boldsymbol{S}_{\text{c}}$生成为内容特征。其次, 噪声域图像生成器$\boldsymbol{G}_X$利用内容特征和噪声特征生成噪声图像$\boldsymbol{S}_{\text{gn}}$, 清晰域图像生成器$\boldsymbol{G}_Y$利用内容特征生成清晰的图像$\boldsymbol{S}_{\text{gc}}$, 2 种生成器具有共同的网络结构, 但网络参数并不共享。最后, 将生成器$\boldsymbol{G}_X$生成的加噪图像$\boldsymbol{S}_{\text{gn}}$再次使用内容编码器和噪声编码器分离, 得到图像的内容特征和噪声特征, 将生成器$\boldsymbol{G}_Y$生成的清晰图像$\boldsymbol{S}_{\text{gc}}$使用内容编码器得到内容特征图, 再利用生成器$\boldsymbol{G}_X$和$\boldsymbol{G}_Y$得到校正后图像$\tilde{\boldsymbol{S}}_{\text{c}}$和加噪后图像$\tilde{\boldsymbol{S}}_{\text{n}}$。整个过程存在 2 种数据流: 一种是噪声图像转移为清晰图像再转移为噪声图像, 另一种是清晰图像转移为噪声图像再转移为清晰图像, 有利于使用循环一致损失来训练网络模型。

图像校正网络的工作流程为: 首先, 通过第一阶段亮度增强后得到带噪声图像, 将带噪声图像通过内容编码器和噪声编码器得到图像的内容特征和噪声特征。然后, 使用清晰域图像生成器$\boldsymbol{G}_Y$生成不带噪声的图像, 将生成的清晰图像作为处理图像, 实现图像从噪声域到清晰域的转移, 恢复出高质量的正常光照图像。

在图像校正阶段, 引入 KL 损失L_{KL}、对抗损失L_{adv}、循环一致损失L_{cc}、重建损失L_{rec}、内容感知损失L_{per}和背景一致损失L_{bc}作为约束训练网络。总的损失函数L_2表示为

$$L_2 = \lambda_{\text{KL}}L_{\text{KL}} + \lambda_{\text{adv}}L_{\text{adv}} + \lambda_{\text{cc}}L_{\text{cc}} + \lambda_{\text{rec}}L_{\text{rec}} + \lambda_{\text{per}}L_{\text{per}} + \lambda_{\text{bc}}L_{\text{bc}} \tag{11}$$

式中: λ为损失项权重。

KL 损失函数有助于网络从噪声图像中分离图像的内容信息, 如下:

$$L_{\text{KL}} = \frac{1}{2}\sum_{i=1}^{d}(-\ln\sigma_i^2 + \mu_i^2 + \sigma_i^2 - 1) \tag{12}$$

式中: d为噪声特征的维度; μ和σ分别为噪声的均值和方差特征。

对抗损失可以最小化真实图像和输出之间的分布距离, 使校正后的图像更接近于真实图像, 如下:

$$L_{\text{adv}} = \frac{1}{2}E_{x\sim p_r}[(D(x)-b)^2] + \frac{1}{2}E_{z\sim p_z}[(D(G(x))-a)^2] \tag{13}$$

式中: E为分布函数的期望值; $D(x)$为鉴别器对真实样本判别为真实样本的概率; $D(G(x))$为鉴别器对生成样本判别为真实样本的概率; a为生成样本的标签; b为真实样本的标签; z为潜在向量。

循环一致损失指导生成的噪声图像$\boldsymbol{S}_{\text{gn}}$转换回清晰图像, 并将去噪的图像$\boldsymbol{S}_{\text{gc}}$恢复成原来的图像, 如下:

$$L_{\text{cc}} = \|\boldsymbol{S} - \tilde{\boldsymbol{S}}\|_1 \tag{14}$$

式中: $\boldsymbol{S}$为输入图像; $\tilde{\boldsymbol{S}}$为域转换后重新生成的图像。

重建损失可保证生成的图像更加可靠, 如下:

$$L_{\text{rec}} = \|\boldsymbol{S}_{\text{g}} - \boldsymbol{S}\|_1 \tag{15}$$

式中: $\boldsymbol{S}_{\text{g}}$为校正后的图像。

内容感知损失使生成的图像具有更多的细节, 如下:

$$L_{\text{per}} = \|\boldsymbol{\phi}(\boldsymbol{S}_{\text{g}}) - \boldsymbol{\phi}(\boldsymbol{S})\|_2^2 \tag{16}$$

式中: $\boldsymbol{\phi}(\cdot)$为预训练的 VGG 网络提取的特征。

背景一致损失使生成后的图像具有更多纹理特征, 保证图像不会被过度平滑, 如下:

$$L_{\text{bc}} = \sum_{\varsigma=i,j,k}\lambda_{\varsigma}\|\boldsymbol{B}_{\varsigma}(\tilde{\boldsymbol{S}}) - \boldsymbol{B}_{\varsigma}(\boldsymbol{S})\|_1 \tag{17}$$

式中: λ_{ς}为用于平衡不同高斯模糊误差的超参数; $\boldsymbol{B}_{\varsigma}$为具有模糊核$\varsigma$大小的高斯算子。

受文献 [19]启发, 本文在实验中将σ'的大小设置为 5,9,15, 对应的$\lambda_{\sigma'}$大小为 0.25,0.5,1.0。实验结果表明, 设置 λ_{KL}=0.01, λ_{adv}=1, λ_{cc}=10, λ_{rec}=10, λ_{per}=0.1, λ_{bc}=5 时所训练的网络性能最佳。

2 实验结果与分析

2.1 无人机数据集

目前，常用的低光照图像增强网络使用的训练数据集通常是地面采集的，与无人机拍摄的图像存在较大差距，导致训练的网络模型不能很好地应用于无人机拍摄的图像。低光照数据集通常采用人工合成的方式，对光照良好的图像调整相机感光度和曝光时间等参数，获得与之对应的成对的低光照图像，无法反映真实的低光照场景特性，如光照不均匀、存在点光源等，导致训练出的网络模型不能很好地解决真实场景下的问题。但是，对于无人机来说，获取成对图像几乎是不可能的，通过人工合成的数据集不能反映真实的低光照场景。因此，本文建立了一个不成对的无人机航拍低光照数据集，使网络能更加有效地针对无人机拍摄的低光照图像。

数据集的建立应包含数量和场景，因此，本文采用拍摄和网络收集等方法，创建了一个多类场景下的无人机航拍低光照数据集，该数据集包含10 000张图像，其中，5 000张为低光照图像，5 000张为正常光照图像，场景包含街道、建筑和野外，场景的部分图像如图3所示。



(a) 街道 (b) 建筑 (c) 野外

图3 数据集图像

Fig. 3 Dataset images

本文建立的无人机航拍低光照数据集如表1所示。其中，街道和建筑场景的图像较多，比较符合现实情况，相比于普通视角下的数据集，本文建立的数据集包含的场景更多，同时也符合无人机拍摄的真实场景。

表1 无人机航拍低光照数据集

Table 1 Low illumination dataset of UAV photography

场景	图像数量/张
建筑	4 000
街道	4 000
野外	2 000

2.2 实验细节

本文算法配置在GPU为Nvidia Titan XP的PC平台上，使用PyTorch框架构建网络。使用渐进式的两阶段网络，第一阶段和第二阶段的网络模型可以分开单独训练。在第一阶段亮度增强网络训练过程中，本文使用GLADNet[20]中提出的数据集，共5 000对成对的图像，训练过程使用Adam优化器，初始学习率设为0.000 1，网络训练次数设为1 000次。在第二阶段图像校正网络训练过程中，使用创建的无人机航拍低光照数据集，Adam优化器，初始学习率设为0.000 1，网络训练次数设为10 000次。

本文使用大疆无人机拍摄的60多张真实场景的图像和搜集的图像组成100张图像作为实验测试集，该数据集包含了多类场景。

2.3 主观评价

2.3.1 自建测试数据集

为验证本文算法的有效性，与其他算法在自建的测试集上进行比较，算法包括RetinexNet[5]、DRBN[21]、Zero-DCE++[22]、EnlightenGAN[12]、KinD++[7]、URetinex-Net[23]、SCI[24]。实验选取3种不同场景进行测试，结果如图4～图6所示。

从图4可以看出，RetinexNet算法增强后，高楼的轮廓产生了严重的伪影，图像颜色整体失真，对物体暗处细节恢复不好；DRBN算法增强后，图像亮度偏暗，高楼细节丢失较多；Zero-DCE++算法增强后，图像整体泛白，不符合人眼视觉感知；EnlightenGAN算法增强后，高楼的颜色较暗，无法恢复更多的图像细节；KinD++算法增强后，高楼伪影严重，存在大量噪声；URetinex-Net算法增强后，天空区域出现了过曝光现象，但图像的整体亮度不高；SCI算法增强了图像的亮度，但无法处理图像中过亮的部分；本文算法增强后的图像在色彩上较为自然，也有效提高了图像的亮度。

从图5可以看出，对于黄色框中选中暗区，每种算法都能提升图像的亮度，基本满足人们的视觉感知。RetinexNet算法增强后的树木的颜色不真实，存在颜色失真；DRBN算法增强后的树木亮度



(a) 原图　(b) RetinexNet[5]　(c) DRBN[21]　(d) Zero-DCE++[22]　(e) EnlightenGAN[12]　(f) KinD++[7]　(g) URetinex-Net[23]　(h) SCI[24]　(i) 本文算法

图 4　不同算法在建筑场景下的增强结果

Fig. 4　Enhancement results of different algorithms in architectural scenarios



(a) 原图　(b) RetinexNet[5]　(c) DRBN[21]　(d) Zero-DCE++[22]　(e) EnlightenGAN[12]　(f) KinD++[7]　(g) URetinex-Net[23]　(h) SCI[24]　(i) 本文算法

图 5　不同算法在街道场景下的增强结果

Fig. 5　Enhancement results of different algorithms in street scenarios



(a) 原图　(b) RetinexNet[5]　(c) DRBN[21]　(d) Zero-DCE++[22]　(e) EnlightenGAN[12]　(f) KinD++[7]　(g) URetinex-Net[23]　(h) SCI[24]　(i) 本文算法

图 6　不同算法在野外场景下的增强结果

Fig. 6　Enhancement results of different algorithms in field scenarios

不足、细节不明显; Zero-DCE++算法增强后图像整体偏白; EnlightenGAN 算法增强后的地面和树木颜色过于偏黄; KinD++算法增强后树木颜色过于鲜艳, 且暗处细节模糊; URetinex-Net 算法对树木和点光源的增强比较符合人眼感知; SCI 增强后的图像亮度和纹理效果都很好, 但对车灯的处理出现了过曝光; 本文算法增强后的树木颜色真实, 也有效提升了图像的亮度。

从图 6 可以看出, 对于黄色框中选中暗区, 除 DRBN 算法外, 每种算法都能有效增强图像的亮度。RetinexNet 和 KinD++算法增强后的树林的颜色不真实, 细节模糊; Zero-DCE++算法增强后, 图像纹理清晰, 但会导致图像整体泛白; EnlightenGAN 算法增强后的图像会丢失山体的轮廓信息; URetinex-Net 算法增强后空中云的细节消失不见; SCI 算法增强后的图像会导致图像过曝光; 本文算法对天空区域的处理结果自然, 增强后图像符合人眼的视觉感知。

从以上 3 组不同场景的图像中可以看出, 本文算法能够有效增强无人机航拍低光照图像, 高楼增强后还保留了图像的边缘信息, 树木、云彩和烟雾增强后也没有出现颜色失真, 车灯和天空区域增强后也没有出现局部过曝光。

2.3.2　公开测试数据集

为进一步验证算法的有效性, 本文使用公开的真实环境下低光照数据集 LIME、NPE、MEF、DICM 进行测试, 随机挑选了若干组图像, 结果如图 7 所示。

从图 7 可以看出, 本文算法增强后的结果符合人眼的视觉感知, 对天空区域的处理不会带来过曝光问题, 能保留云彩的边缘轮廓, 对于点光源的处理没有产生局部过曝光问题, 树木等物体增强后的颜色也比较自然。总的来说, 本文算法有效增强了图像的亮度, 避免了颜色失真问题, 减少了噪声对图像的影响。



(a) 原图 (b) RetinexNet[5] (c) DRBN[21] (d) Zero-DCE++[22] (e) EnlightenGAN[12] (f) KinD++[7] (g) URetinex-Net[23] (h) SCI[24] (i) 本文算法

图 7 不同算法在公开数据集场景下的增强结果

Fig. 7 Enhancement results of different algorithms on public datasets

2.4 客观评价

2.4.1 自建测试数据集质量评价

本文算法与其他先进算法进行定量比较，采用收集的 100 张图像作为测试集，使用 4 种无参考图像评价指标定量评估算法性能：①亮度失真[25]（lightness order error, LOE）反映增强后图像自然度的亮度顺序误差，LOE 值越小，亮度顺序保持得越好；②无参考图像空间质量评估器[26]（blind reference-less image spatial quality evaluator, BRISQUE）是基于自然场景统计的无参考图像评价指标，值越小，图像表现的越自然；③自然图像质量评估器[27]（natural image quality evaluator, NIQE）的值越小，表示图像越符合人眼的自然感知；④处理一副图像所需要的时间越小，表示算法的速度越快。不同算法增强后的各项图像质量评价结果如表 2 所示。可以看出，本文算法的 NIQE 指标的值最小，LOE 和 BRISQUE 指标靠前，处理时间为 0.023 2 s，表示算法每秒处理约 43 张图像，能够达到实时效果。

表 2 不同算法在自建数据集上的客观评价指标

Table 2 Objective evaluation indicators for different algorithms on self-built dataset

算法	LOE	BRISQUE	NIQE	处理时间/s
RetinexNet[5]	702.80	31.47	3.663	0.027 0
DRBN[21]	462.32	28.69	3.250	0.024 1
Zero-DCE++[22]	433.86	15.37	3.357	0.001 2
EnlightenGAN[12]	408.01	17.63	3.247	0.015 0
KinD++[7]	609.83	23.05	3.813	0.022 0
URetinex-Net[23]	151.21	31.50	3.444	0.028 1
SCI[24]	222.67	37.73	3.568	0.006 7
本文算法	213.78	19.45	3.205	0.023 2

2.4.2 公开测试数据集质量评价

本文在公开真实环境的低光照数据集 LIME、NPE、MEF、DICM 上进行测试，用 NIQE 指标说明算法的有效性，实验结果如表 3 所示。可以看出，本文算法的 NIQE 值最低，增强后的图像更加符合人眼的自然感知。

表 3 不同算法在公开数据集上的 NIQE 值

Table 3 NIQE of different methods on public datasets

算法	LIME	NPE	MEF	DICM
RetinexNet[5]	4.907	4.080	4.904	4.674
DRBN[21]	4.133	4.112	3.356	3.431
Zero-DCE++[22]	3.762	4.253	3.283	3.362
EnlightenGAN[12]	3.608	3.425	3.121	2.846
KinD++[7]	4.206	3.551	3.374	2.965
URetinex-Net[23]	4.096	4.041	3.318	3.235
SCI[24]	4.138	4.160	3.433	3.709
本文算法	3.557	3.363	3.065	2.622

2.5 消融实验

为验证每个模块的有效性，在 LOL 数据集上对颜色一致性模块和图像校正模块进行了消融实验。

颜色一致性模块的消融实验结果如图 8 所示。图 8(a)为低光照图像，图 8(b)为没有添加颜色一致性模块得到的图像，图像整体比较鲜艳，局部出现了颜色失真；图 8(c)为添加了颜色一致性模块得到的图像，图像的颜色信息接近自然。结果表明，颜色一致性模块能够有效避免颜色失真。

图像校正模块的消融实验如图 9 所示，主要验证图像校正模块中的去噪功能，使用 BM3D[28] 代替本文中的校正模块。图 9(a)为原始的低光照图像；

图 9(b)为通过 BM3D 去噪后得到的图像, 图像整体模糊, 细节信息没有保留; 图 9(c)为通过本文图像校正模块增强后的图像, 减少了图像中存在的噪声, 并保留了物体的纹理信息。

本文采用峰值信噪比 PSNR 和结构相似性 SSIM 客观评价指标来定量评估模块。PSNR 表示信号最大可能功率和影响其表示精度的噪声功率的比值, 值越大, 图像质量越好。SSIM 表示 2 张图像的相似性, 值越大, 图像的质量越好。实验结果如表 4 所示。可以看出, 本文所提模块都能有效提升网络的性能。



(a) 低光照图像　(b) 无颜色一致性　(c) 本文算法

图 8　颜色一致性消融实验对比

Fig. 8　Comparison of color consistency ablation experiments



(a) 低光照图像　(b) 加入BM3D　(c) 本文算法

图 9　图像校正去噪消融实验对比

Fig. 9　Comparison of image correction ablation experiments

表 4　消融实验客观评价指标

Table 4　Objective evaluation indicators for ablation experiments

算法	PSNR/dB	SSIM
无颜色一致性	20.138	0.769
加入BM3D	19.893	0.778
本文算法	21.221	0.791

2.6　目标检测应用

使用 SSD[29] 和 Deformable DETR[30]2 种目标检测算法验证本文算法增强预处理后对目标检测算法性能的影响, 由于目前没有很好的夜间无人机航拍图像检测的公开数据集, 本文使用普通视角下的夜间目标检测 ExDark[31]。从数据集中随机选取 300 张低光照图像进行测试, 使用 COCO 预训练权重直接检测增强前后的图像, 部分可视化结果如图 10 所示。此外, 为说明本文算法对无人机航拍夜间图像检测性能的影响, 在图 10 中添加了 2 组从公开的 COCO 数据集中随机挑选的无人机航拍低光照图像。



(a) SSD300算法增强前目标检测结果　(b) SSD300算法增强后目标检测结果　(c) Deformable DETR算法增强前目标检测结果　(d) Deformable DETR算法增强后目标检测结果

图 10　检测算法结果

Fig. 10　Detection algorithm results

从图10中可以看出，2种检测算法对人和自行车的检测精度有所提升，但对自行车的检测出现了漏判。针对无人机航拍的低光照图像，2种检测算法的检测精度有一定提升，并且减少了漏检现象。本文采用平均精度（aaverage precision, AP）作为评价指标，定量分析图像增强前后对目标检测算法性能的影响，实验结果如表5所示。

SSD300算法中，Person和Boat类别的目标检测AP值有所提升，而Bicycle和Car类别有所下降，这是由于本文使用了COCO预训练权重直接检测增强后图像，增强后图像中的目标特征和增强前图像的目标特征存在差异，导致检测算法出现一定的漏检和误检，降低目标检测算法的性能。Deformable DETR算法的4种类别AP值都有所提升。总体来说，本文算法能提升目标检测算法在夜间的检测性能，对人和船等物体准确率增加0.01。

表5 图像增强前后算法检测结果平均精度对比

Table 5 Comparison of AP values of algorithm detection results before and after image enhancement

算法	主干网络	Person		Bicycle		Car		Boat	
		增强前	增强后	增强前	增强后	增强前	增强后	增强前	增强后
SSD300	VGG-16	0.355	0.361	0.344	0.343	0.351	0.346	0.364	0.375
Deformable DETR	ResNet-50	0.382	0.393	0.365	0.375	0.358	0.360	0.381	0.399

3 结 论

1）算法增强后的图像符合人眼的视觉感知，在多种场景下都有较为优异的表现，在亮度和颜色方面有着显著的突出。

2）算法增强后的图像在客观评价指标上也具有良好的表现，如与主流的低光照图像增强算法对比，NIQE指标最好，LOE指标也名列前茅。

3）算法增强后的图像可以用于图像检测，能有效提升夜间检测算法的准确性，如增强后的人、船等物体准确率都有0.01以上的提升。

为使算法具有更好的增强效果和应用性，后续一方面增强训练集，使训练出的模型具有更强的鲁棒性；另一方面设计更加轻量级的网络，减少处理图像的时间，做好预处理工作，更好地结合图像跟踪或目标检测领域。

参考文献（References）

[1] 李可夫, 钟汇才, 高兴宇, 等. 显著性引导的低光照人脸检测[J]. 北京航空航天大学学报, 2021, 47(3): 572-584.
LI K F, ZHONG H C, GAO X Y, et al. Saliency guided low-light face detection[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(3): 572-584(in Chinese).

[2] PIZER S M, AMBURN E P, AUSTIN J D, et al. Adaptive histogram equalization and its variations[J]. Computer Vision, Graphics, and Image Processing, 1987, 39(3): 355-368.

[3] LAND E H. The Retinex theory of color vision[J]. Scientific American, 1977, 237(6): 108-128.

[4] GUO X J, LI Y, LING H B. LIME: low-light image enhancement via illumination map estimation[J]. IEEE Transactions on Image Processing, 2017, 26(2): 982-993.

[5] WEI C, WANG W J, YANG W H, et al. Deep Retinex decomposition for low-light enhancement[EB/OL]. (2018-08-14)[2023-04-01]. https://arxiv.org/abs/1808.04560v1.

[6] ZHANG Y H, ZHANG J W, GUO X J. Kindling the darkness: a practical low-light image enhancer[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM, 2019: 1632-1640.

[7] ZHANG Y H, GUO X J, MA J Y, et al. Beyond brightening low-light images[J]. International Journal of Computer Vision, 2021, 129(4): 1013-1037.

[8] LV F F, LU F, WU J H, et al. MBLLEN: low-light image/video enhancement using CNNs[C]//Proceedings of the British Machine Vision Conference. [S. l.]: BMVC, 2018.

[9] 孙帮勇, 赵兴运, 吴思远, 等. 基于移位窗口多头自注意力U型网络的低照度图像增强方法[J]. 电子与信息学报, 2022, 44(10): 3399-3408.
SUN B Y, ZHAO X Y, WU S Y, et al. Low-light image enhancement method based on shifted window multi-head self-attention U-shaped network[J]. Journal of Electronics & Information Technology, 2022, 44(10): 3399-3408(in Chinese).

[10] XU K, YANG X, YIN B C, et al. Learning to restore low-light images via decomposition-and-enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 2278-2287.

[11] 张雅荔, 李文元, 李昌禄, 等. 融合注意力引导的多尺度低照度图像增强方法[J]. 西安电子科技大学学报, 2023, 50(1): 129-136.
ZHANG Y L, LI W Y, LI C L, et al. Method for enhancement of the multi-scale low-light image by combining an attention guidance[J]. Journal of Xidian University, 2023, 50(1): 129-136(in Chinese).

[12] JIANG Y F, GONG X Y, LIU D, et al. EnlightenGAN: deep light enhancement without paired supervision[J]. IEEE Transactions on Image Processing, 2021, 30: 2340-2349.

[13] 向森, 王应锋, 邓慧萍, 等. 基于双重迭代的零样本低照度图像增强[J]. 电子与信息学报, 2022, 44(10): 3379-3388.
XIANG S, WANG Y F, DENG H P, et al. Zero-shot learning for low-light image enhancement based on dual iteration[J]. Journal of Electronics & Information Technology, 2022, 44(10): 3379-3388(in Chinese).

[14] 王殿伟, 刘旺, 房杰, 等. 受生物视觉启发的无人机航拍低照度图

像增强算法[J]. 西北工业大学学报, 2023, 41(1): 144-152.
WANG D W, LIU W, FANG J, et al. Enhancement algorithm of low illumination image for UAV images inspired by biological vision[J]. Journal of Northwestern Polytechnical University, 2023, 41(1): 144-152(in Chinese).

[15] JIN Y Y, YANG W H, TAN R T. Unsupervised night image enhancement: when layer decomposition meets light-effects suppression[EB/OL].(2022-09-17)[2023-04-01].https://arxiv.org/abs/2207.10564v2.

[16] 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.

[17] ZHANG Z, ZHENG H, HONG R C, et al. Deep color consistent network for low-light image enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 1889-1898.

[18] ZHANG F, SHAO Y J, SUN Y S, et al. Unsupervised low-light image enhancement via histogram equalization prior[EB/OL]. (2021-11-03)[2023-04-01]. https://arxiv.org/abs/2112.01766v1.

[19] DU W C, CHEN H, YANG H Y. Learning invariant representation for unsupervised image restoration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 14471-14480.

[20] WANG W J, WEI C, YANG W H, et al. GLADNet: low-light enhancement network with global awareness[C]//Proceedings of the 13th IEEE International Conference on Automatic Face & Gesture Recognition. Piscataway: IEEE Press, 2018: 751-755.

[21] YANG W H, WANG S Q, FANG Y M, et al. From fidelity to perceptual quality: a semi-supervised approach for low-light image enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 3060-3069.

[22] LI C Y, GUO C L, LOY C C. Learning to enhance low-light image via zero-reference deep curve estimation[EB/OL]. (2021-03-01)[2023-04-01]. https://arxiv.org/abs/2103.00860v1.

[23] WU W H, WENG J, ZHANG P P, et al. URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 5891-5900.

[24] MA L, MA T Y, LIU R S, et al. Toward fast, flexible, and robust low-light image enhancement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 5627-5636.

[25] AHN H, KEUM B, KIM D, et al. Adaptive local tone mapping based on Retinex for high dynamic range images[C]//Proceedings of the IEEE International Conference on Consumer Electronics. Piscataway: IEEE Press, 2013: 153-156.

[26] MITTAL A, MOORTHY A K, BOVIK A C. No-reference image quality assessment in the spatial domain[J]. IEEE Transactions on Image Processing, 2012, 21(12): 4695-4708.

[27] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a “completely blind” image quality analyzer[J]. IEEE Signal Processing Letters, 2012, 20(3): 209-212.

[28] DABOV K, FOI A, KATKOVNIK V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. IEEE Transactions on Image Processing, 2007, 16(8): 2080-2095.

[29] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the Computer Vision. Berlin: Springer, 2016: 21-37.

[30] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable Transformers for end-to-end object detection[EB/OL]. (2021-03-18)[2023-04-01]. https://arxiv.org/abs/2010.04159v4.

[31] LOH Y P, CHAN C S. Getting to know low-light images with the exclusively dark dataset[J]. Computer Vision and Image Understanding, 2019, 178: 30-42.

Low illumination image enhancement algorithm for UAV aerial photography with color consistency

WANG Dianwei[1,*], LIU Wang[1], FANG Jie[1], XU Zhijie[2]

(1. School of Communications and Information Engineering, Xi'an University of Posts & Telecommunications, Xi'an 710121, China;
2. School of Computing and Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK)

Abstract: To address the issue of low brightness and poor visual effect of unmanned aerial vehicles (UAVs) in low illumination conditions, this paper established a low illumination dataset of UAV aerial photography and proposed a low illumination image enhancement algorithm for UAV aerial photography with color consistency. Firstly, in the brightness enhancement stage, this paper constructed a brightness enhancement network (BENet) to enhance the brightness of images and used the color network (CNet) module and the pyramid color embedding (PCE) module to combine the color features and content features of the images, which avoided color distortion in the enhanced image. In the image correction stage, this paper constructed a correction network based on domain transmission, trained the network with the self-built dataset, corrected the enhanced image after the first stage with the help of well-illuminated images, reduced the effect of noise on the image, and finally obtained the enhanced image. The experimental results show that the algorithm effectively avoids the problems of color distortion and noise amplification while enhancing the image brightness, and it outperforms other advanced algorithms in terms of objective indicators and improves the detection performance of the target detection algorithm at night.

Keywords: low illumination image enhancement; unmanned aerial vehicle aerial photography image; brightness enhancement; color consistency; image denoising

Received: 2023-04-10; **Accepted**: 2023-05-19; **Published Online**: 2023-06-27 16:19

URL: link.cnki.net/urlid/11.2625.V.20230627.1302.001

Foundation items: Youth Fund of the National Natural Science Foundation of China (62201454); Shaanxi Provincial International Science and Technology Cooperation Plan (2023-GHYB-04)

* **Corresponding author**. E-mail: wangdianwei@xupt.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0185

基于新型趋近律的AABS积分滑模控制系统设计

徐萌[1,2]，李艳[1]，高洁[1,2,*]，徐海[3]，高冰[4]

(1. 中国民航大学 电子信息与自动化学院，天津 300300；2. 天津市城市空中交通系统技术与装备重点实验室，天津 300300；
3. 中国民用航空沈阳航空器适航审定中心，沈阳 110043；4. 中国民航大学 工程技术训练中心，天津 300300)

摘 要：针对飞机防滑刹车系统(AABS)具有强时变性、强不可测性、内外干扰因素多的特点，提出一种变指数分数阶指数趋近律的AABS积分滑模控制(ISMC)方法。建立飞机着陆滑跑时系统的动力学模型；以提高系统快速性和鲁棒性为目标，引入变结构控制实现最佳滑移率和最大结合系数的跟踪问题；改进分数阶指数趋近律的积分滑模控制律，在抑制抖振现象的同时加快跟踪到最佳滑移率的速度；设计滑模观测器观测飞机速度，以减少存在的内外界非线性干扰；利用MATLAB仿真平台，对所提方法的可行性及有效性进行仿真验证。仿真结果表明：所设计的变指数分数阶指数趋近律的积分滑模控制器整体控制效果优于传统线性滑模控制器，滑模观测器可以精准估计飞机速度，所提方法提高了系统整体设计的鲁棒性，缩短了刹车时间和刹车距离，控制效果良好。

关 键 词：飞机防滑刹车系统；分数阶；积分滑模控制；滑移率；鲁棒性
中图分类号：V227+.5
文献标志码：A **文章编号**：1001-5965(2025)04-1107-10

飞机防滑刹车系统(aircraft anti-skid braking system, AABS)是飞机着陆过程中非常重要的安全保障机构，其不仅能使飞机迅速刹停，还能避免因车轮打滑或抱死而引起的安全事故[1]。随着现代飞机性能的大幅提升，对飞机制动系统的安全性和机轮刹车系统的刹车能力等提出了严格的要求，因此，需要建立相对精确的AABS数学模型。同时，具有高可靠性和优良动态性能的防滑刹车控制律是建立新型综合AABS模型的基础。

针对飞机刹车中的诸多非线性干扰因素，机轮-地面间的摩擦力由飞机速度、滑移率、跑道路面条件、侧风等不确定性和非线性干扰因素决定。为解决这类非线性和不确定性问题并建立精确的AABS数学模型，学者们提出了基于反馈的控制方法，包括PID控制、反馈线性化、滑模控制和智能控制等。例如，文献[2]针对偏心力矩和执行器故障提出了径向基函数和自适应估计器来估计未知非线性。文献[3]仅考虑纵向力建立AABS动力学模型，而AABS的动力学由纵向力矩、垂直力矩和俯仰力矩决定。此外，飞机前轮自由转动时产生动摩擦系数，而由制动装置产生的制动力矩对机轮产生影响的是静摩擦系数，这2个摩擦系数存在本质的区别。文献[4]建立的AABS模型考虑了飞机的纵向、垂直和俯仰力矩，但为了简化模型，其忽略了前轮摩擦和AABS动力学耦合的问题。要克服目前存在的局限性，有必要建立一种更适合实际应用的综合性AABS动力学模型。

综上所述，分析跑道对飞机主轮产生的摩擦力的过程，存在多个非线性和不确定性因素，这无疑给AABS控制方法的设计带来了巨大挑战。随着AABS智能控制技术的研究越来越多，自适应控制等被广泛应用于飞机的防滑制动方法中。然而，虽然智能控制技术给防滑制动方法带来了革新，但仍然有待进一步提高。由于主轮静摩擦系数受到飞

收稿日期：2023-04-18；**录用日期**：2023-08-02；**网络出版时间**：2023-08-22 17:41
网络出版地址：link.cnki.net/urlid/11.2625.V.20230822.0944.002
基金项目：国家自然科学基金(51707195,62173331)；中央高校基本科研业务费专项资金(3122023016)
*通信作者. E-mail：jiegao@cauc.edu.cn

引用格式：徐萌，李艳，高洁，等. 基于新型趋近律的AABS积分滑模控制系统设计[J]. 北京航空航天大学学报，2025，51(4)：1107-1116.
XU M，LI Y，GAO J，et al. Design of aircraft anti-skid braking system integral sliding mode control system based on novel reaching law[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1107-1116 (in Chinese).

机速度、滑移率、跑道条件等因素的影响而呈现时变的状态，摩擦系数的最大值对应滑移率的最优值，滑移率也是时变的。现有的控制方法大多用于跟踪恒定的参考滑移率信号值，而不是实时的最优滑移率值。因此，需要采用一些手段实时跟踪时变最佳滑移率，才能充分利用最大摩擦力来缩短制动距离和制动时间。目前，虽然已有一些策略能够实现滑移率寻优，但其实时跟踪性能有望进一步提高。文献 [5] 设计了一种分数阶滑模控制器用于跟踪车辆 ABS 的最佳滑移率；文献 [6] 针对任意阶非线性系统提出了一种动态滑模面的连续滑模控制律；文献 [7] 基于电动汽车设计了积分滑模控制器，提高了驾驶舒适性和控制质量；文献 [8] 针对多电系统提出了二阶自适应滑模控制方法；文献 [9] 提出了一种跑道最大摩擦力跟踪算法。除此之外，文献 [10-11] 针对飞机变跑道车轮制动提出了一种基于混合滑移减速和跑道非线性模型预测方法。以上方法均旨在跟踪最优目标值的同时提高控制系统的整体鲁棒性。

本文在建立 AABS 模型的基础上，分析了其特点及控制精度问题，介绍了新型分数阶指数趋近律的积分滑模控制（integral sliding mode control, ISMC）方法与传统线性滑模设计的区别，对所提出的新型控制策略进行仿真，并与线性滑模控制 (linear sliding mode control, LSMC) 律进行对比研究，变指数分数阶指数趋近律 (variable exponential fractional order exponential reaching law, VFERL) 的积分滑模控制律提高了 AABS 的鲁棒性和刹车效率，新型趋近律的提出在提高趋近速度的同时减小了超调，同时，通过仿真验证了本文方法的有效性和可行性。

1 系统模型建立和问题提出

1.1 系统描述

本文以某型飞机为研究对象，对 AABS 的动力学性能进行研究，建立的动力学模型包括飞机机体动力学、机轮受刹动力学、轮胎地面摩擦模型和刹车装置模型，并分析了主轮打滑情况。

在飞机制动过程中，假设机身和起落架为理想刚体，无垂直俯仰位移，且左右对称，不受横向力影响。

根据以上假设，建立机体受力分析图，如图 1 所示。

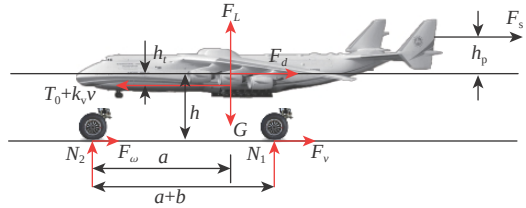


图 1 飞机机体受力分析

Fig. 1 Force analysis of aircraft body

考虑空气动力特性影响，飞机纵向动力学方程、垂直方向平衡方程和质心力矩平衡方程如下：

$$\begin{cases} T_0+k_v v-F_d-F_s-F_v-F_\omega=M\dot{v} \\ G-F_L-N_1-N_2=0 \\ N_2 a+(T_0+k_v v)h_t+F_s h_p=N_1 b+(F_v+F_\omega)h \end{cases} \tag{1}$$

式中：T_0、F_d、F_s、F_L、F_v和F_ω分别为发动机的剩余推力、空气阻力、阻力伞阻力、升力、主轮摩擦力和前轮摩擦力，$F_d=\rho C_d S v^2/2$，$F_s=\rho C_p S_p v^2/2$，$F_L=\rho C_L S v^2/2$，ρ为空气密度，C_d、C_p、C_L分别为飞机滑跑时的阻力系数、阻力伞阻力系数、飞机滑跑时的升力系数，S和S_p分别为机翼面积和阻力伞计算面积；N_1和N_2分别为主轮载荷和前轮载荷；h_p、h_t和h分别为阻力伞悬挂点距飞机水平轴上移距离、发动机推力线距飞机水平轴下移距离和飞机重心距地面的高度；a、b分别为飞机重心与前轮、主轮中心之间的距离；v为飞机速度；k_v为发动机的推力系数；M为飞机质量；G为飞机重力。

$$N_1=\frac{(G-F_L)(a-\mu_f h)+(T_0+k_v v)h_t+F_s h_p}{a+b+(\mu-\mu_f)h} \tag{2}$$

$$N_2=\frac{(G-F_L)(b+\mu_f h)+(T_0+k_v v)h_t+F_s h_p}{a+b+(\mu-\mu_f)h} \tag{3}$$

式中：μ、μ_f分别为主轮与跑道间的结合系数、前轮与跑道间的结合系数。

将式 (2)、式 (3) 代入式 (1) 可知，飞机加速度为

$$\begin{aligned} \dot{v}=&\frac{T_0+k_v v-F_d-F_s-F_v-F_\omega}{M}= \\ &\frac{T_0+k_v v}{M}-\frac{\rho C_d S v^2}{2M}-\frac{\rho C_p S_p v^2}{2M}- \\ &\mu\frac{(a-\mu_f h)\left(g-\dfrac{\rho C_L S v^2}{2M}\right)}{a+b+(\mu-\mu_f)h}+ \\ &\mu\frac{h_t\dfrac{T_0+k_v v}{M}+h_p\dfrac{\rho C_p S_p v^2}{2M}}{a+b+(\mu-\mu_f)h}- \\ &\mu_f\frac{(b+\mu h)\left(g-\dfrac{\rho C_L S v^2}{2M}\right)}{a+b+(\mu-\mu_f)h}- \\ &\mu_f\frac{h_t\dfrac{T_0+k_v v}{M}+h_p\dfrac{\rho C_p S_p v^2}{2M}}{a+b+(\mu-\mu_f)h} \end{aligned} \tag{4}$$

式中：g为重力加速度。

1.2 机轮模型

机轮的转动是由刹车力矩和结合力矩共同驱动的结果，依据机轮转动定律：

$$\dot{\omega}=\frac{M_j-M_g}{I} \tag{5}$$

$$v=R_{vb}\omega \tag{6}$$

式中: M_g 为刹车力矩; M_j 为地面结合力矩; I 为单个受刹机轮转动惯量; ω 为机轮角速度; R_{vb} 为机轮半径。

刹车力矩 M_g 由刹车装置的摩擦特性和刹车压力决定, 地面结合力矩 M_j 由力学公式表示。

因此, 机轮的数学模型为

$$\dot{\omega}=\frac{\mu N_1 R_{vb}-M_g}{I}=\mu R_{vb}\frac{(a-\mu_f h)\left(Mg-\dfrac{\rho C_L S v^2}{2}\right)}{nI\left[a+b+(\mu-\mu_f)h\right]}+\mu R_{vb}\frac{h_t(T_0+k_v v)+h_p\dfrac{\rho C_p S_p v^2}{2}}{nI\left[a+b+(\mu-\mu_f)h\right]}-\frac{k_b}{I}P_S \tag{7}$$

式中: n 为受刹机轮数目; k_b 为刹车机构的压力-力矩转换系数; P_S 为最大刹车压力。

机轮是飞机在着陆阶段与到跑道唯一接触的单元, 其特性对飞机刹车过程中的动态行为起到至关重要的作用[12]。基于轮胎-地面摩擦模型的研究, 从机理模型到经验、半经验模型, 以及可以精确描述轮胎与地面接触区域相互作用的有限元模型。由于纯机理模型的假设多样性与现实情况存在一定差距和有限元方法高昂的计算成本, 在常见研究中多采用经验和半经验模型。其中, 最常见的是用机轮滑移率与结合系数的代数函数表征轮胎与摩擦力的非线性关系。2 种经典的模型分别为魔术公式[13]和 Burckhardt 模型[14]。相较于魔术公式, Burckhardt 模型被广泛应用于道面识别中。根据 Burckhardt 提出的摩擦模型[15], 静摩擦系数表示为

$$\mu(\lambda,v)=\left[c_1\left(1-\mathrm{e}^{-c_2\lambda}\right)-c_3\lambda\right]\mathrm{e}^{-c_4\lambda v} \tag{8}$$

式中: μ 为 Burckhardt 摩擦模型中的最佳结合系数; $c_i\,(i=1,2,3,4)$ 取决于跑道条件。

表 1 列出了不同类型的跑道参数。图 2 为 v=25 m/s 时不同跑道的 μ-λ 曲线。图 3 为湿沥青跑道条件下 Burckhardt 摩擦模型的曲线。

表 1　不同类型的跑道参数

Table 1　Parameters of several types of runways

跑道类型	c_1	c_2	c_3	c_4
干沥青	1.280 1	23.99	0.52	0.03
干石灰	1.197 3	25.168	0.537 3	0.03
湿沥青	0.857	33.822	0.347	0.03
雪	0.194 6	94.129	0.064 6	0.03
冰	0.05	306.39	0	0.03

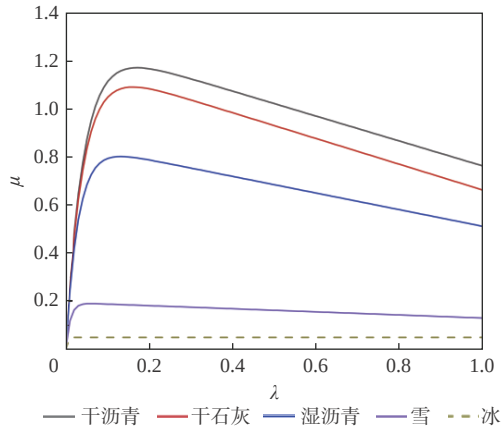


图 2　v=25 m/s 时不同跑道 μ-λ 曲线

Fig. 2　μ-λ curves of different runways when v=25 m/s

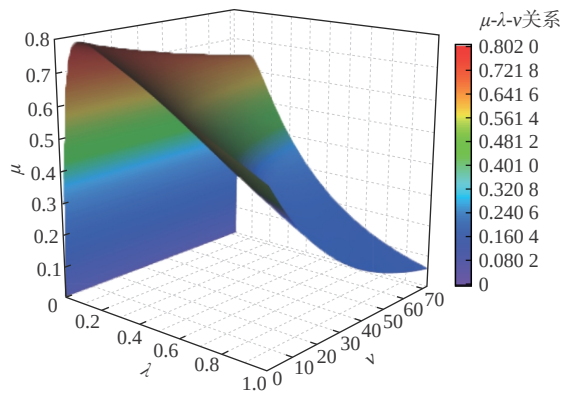


图 3　湿沥青跑道结合系数模型曲线

Fig. 3　Binding coefficient model curves of wet asphalt runway

2　新型趋近律积分滑模控制律设计

2.1　滑模面设计

飞机处于滑跑阶段时, 机轮受刹车力矩作用, 使飞机速度与机轮速度始终保持在 $v>R_{vb}\omega$。定义滑移率 λ 表示机轮相对跑道的滑动比:

$$\lambda=\frac{v-R_{vb}\omega}{v}\qquad \lambda\in[0,1) \tag{9}$$

在传统滑模控制中引入分数阶微积分算子, 对于存在非线性干扰和模型不确定性的系统鲁棒性更好。且与传统微积分比较, 分数阶微积分增加了微分、积分这 2 个运算自由度的可变性, 利用其遗传特性和记忆性可以进一步提高系统的控制品质和综合性能。因此, 引入分数阶切换控制代替上述 Ambrosino 切换函数, 以达到更好的消除抖振、消除稳态误差的效果。由于线性滑模面无法保证初始状态处于滑模面上, 本文提出一种积分滑模控制方法, 避免了线性滑模趋近阶段的缺点, 并引入分数阶指数趋近律, 旨在减小滑移率跟踪误差的同时, 加快响应速度, 抑制系统抖振。

首先, 滑移率跟踪误差及其一阶微分分别为

$$e=\lambda-\lambda_p \tag{10}$$

$$\dot{e}=\dot{\lambda}-\dot{\lambda}_{\mathrm{p}}=\frac{f_1(\boldsymbol{x})\omega-v\left(f_2(\boldsymbol{x})-\dfrac{k_{\mathrm{b}}}{I}u\right)}{v^2}R_{\mathrm{vb}}-\dot{\lambda}_{\mathrm{p}} \tag{11}$$

式中：λ_{p}为最佳滑移率；$f_1(\boldsymbol{x})$和$f_2(\boldsymbol{x})$分别为飞机加速度和机轮加速度模型；u为飞机防滑刹车控制器模型的输入，即控制律输入u。

定义含积分的滑模函数为

$$s=c\int e(t)\mathrm{d}t+e(t) \tag{12}$$

式中：c必须满足Hurwitz条件，即$c>0$，当$s=0$时，$e(t)+c\int e(t)\mathrm{d}t=0$，收敛结果为$e(t)=e(0)\mathrm{e}^{-t}$，当$t\to\infty$时，误差指数收敛于0，收敛速度取决于$c$值，本文中$c$为常值系数。对于初始状态$e(0)$，通过设计积分项$\int e(t)\mathrm{d}t=-e(t)/c$，当$t>0$时，轨线必须保持在滑模面$s$上。因此，跟踪最佳滑移率$\lambda_{\mathrm{p}}$的问题可转化为使标量$s\equiv0$的问题。

由式(12)可得

$$\dot{s}=ce+\dot{e}=c\left(\lambda-\lambda_{\mathrm{p}}\right)+\left(\dot{\lambda}-\dot{\lambda}_{\mathrm{p}}\right)=c\left(\lambda-\lambda_{\mathrm{p}}\right)+\frac{f_1(\boldsymbol{x})\omega-v\left(f_2(\boldsymbol{x})-\dfrac{k_{\mathrm{b}}}{I}u\right)}{v^2}R_{\mathrm{vb}}-\dot{\lambda}_{\mathrm{p}} \tag{13}$$

2.2 分数阶指数趋近律设计

分数阶滑模主要是通过在整数阶滑模控制中引入分数阶微分，使整数阶滑模控制增加一个额外的可调参数，从而改善传统整数阶滑模控制的收敛性能，抑制系统抖振。

引入分数阶指数微分方程，并建立一种类似于指数趋近律的新型分数阶指数趋近律：

$$s^{\beta}=-ks-\varepsilon\cdot\mathrm{sgn}(s) \tag{14}$$

式中：$k>0,\varepsilon>0,0<\beta<1$；$\mathrm{sgn}(s)$为符号函数。

$$\dot{s}=c\left(\lambda-\lambda_{\mathrm{p}}\right)+\frac{f_1(\boldsymbol{x})\omega-v\left(f_2(\boldsymbol{x})-\dfrac{k_{\mathrm{b}}}{I}u\right)}{v^2}R_{\mathrm{vb}}-\dot{\lambda}_{\mathrm{p}}=-D^{(\alpha)}\left(ks+\varepsilon\cdot\mathrm{sgn}(s)\right) \tag{15}$$

式中：$D^{(\alpha)}$为分数阶指数模型，即α阶导数，$0<\alpha<1$。

控制律为

$$u=\left[\frac{\dot{\lambda}_{\mathrm{p}}-c\left(\lambda-\lambda_{\mathrm{p}}\right)}{R_{\mathrm{vb}}}v-\frac{f_1(\boldsymbol{x})\omega}{v}+f_2(\boldsymbol{x})\right]\frac{I}{k_{\mathrm{b}}}-\frac{D^{(\alpha)}\left(ks+\varepsilon\cdot\mathrm{sgn}(s)\right)v}{R_{\mathrm{vb}}}\cdot\frac{I}{k_{\mathrm{b}}} \tag{16}$$

2.3 趋近律改进设计

趋近律的设计原则如下：

1）运动点在滑模面较远处时，趋近速度要大，加快响应速度，缩短趋近阶段的时间。

2）运动点在滑模面附近时，趋近速度要尽量小，以免产生较大的超调。

以趋近律设计原则出发，对分数阶指数趋近律进行改进，改进后的新型趋近律表达式如下：

$$\begin{cases}\dot{s}=-D^{(\alpha)}\left(ks+\varepsilon f(s)\cdot\mathrm{sgn}(s)\right)\\ f(s)=\dfrac{1}{p(2^{-|s|})^{q}+\dfrac{1}{|s|+1}}\end{cases} \tag{17}$$

式中：$p>0$，$q>0$。当运动点在无限远处时，$|s|\to\infty$，则$f(s)\to\infty$，趋近速度大；当运动点在滑模面附近时，$|s|\to0$，则$f(s)\approx1/(p+1)<1$，趋近速度较小。

因此，改进趋近律后的控制律为

$$u=\left[\frac{\dot{\lambda}_{\mathrm{p}}-c\left(\lambda-\lambda_{\mathrm{p}}\right)}{R_{\mathrm{vb}}}-\frac{f_1(\boldsymbol{x})\omega}{v}+f_2(\boldsymbol{x})\right]\frac{I}{k_{\mathrm{b}}}-\frac{D^{(\alpha)}\left(ks+\varepsilon f(s)\cdot\mathrm{sgn}(s)\right)}{R_{\mathrm{vb}}}v\frac{I}{k_{\mathrm{b}}} \tag{18}$$

切换面可达性证明如下：

设计Lyapunov函数为

$$V=\frac{1}{2}s^2 \tag{19}$$

若切换面可达，需要满足条件：$s\dot{s}\leqslant0$，即

$$\dot{V}=s\dot{s}=s\left[c\left(\lambda-\lambda_{\mathrm{p}}\right)+\frac{f_1(\boldsymbol{x})\omega-v\left(f_2(\boldsymbol{x})-\dfrac{k_{\mathrm{b}}}{I}u\right)}{v^2}R_{\mathrm{vb}}-\dot{\lambda}_{\mathrm{p}}\right]=s\left[-D^{(\alpha)}\left(ks+\varepsilon f(s)\cdot\mathrm{sgn}(s)\right)\right] \tag{20}$$

利用$\mathrm{sgn}\left(D^{(\alpha)}\left(-k\cdot\mathrm{sgn}(s)\right)\right)=-k\cdot\mathrm{sgn}(s)$[16]：

$$\begin{aligned}\mathrm{sgn}\left(\dot{V}\right)&=\mathrm{sgn}(s)\cdot\mathrm{sgn}\left(D^{(\alpha)}\left(-ks-\varepsilon f(s)\cdot\mathrm{sgn}(s)\right)\right)=\\&-\varepsilon f(s)\cdot\mathrm{sgn}(s)\cdot\mathrm{sgn}(s)-ks\cdot\mathrm{sgn}(s)=\\&-\varepsilon f(s)-ks\cdot\mathrm{sgn}(s)\end{aligned}$$

当$s>0$时，$\mathrm{sgn}(\dot{V})=-\varepsilon-k<0$；当$s<0$时，$\mathrm{sgn}(\dot{V})=-\varepsilon-k<0$。

因此，$\dot{V}\leqslant0\to D^{(\alpha)}V\leqslant0$，根据Lyapunov稳定定理可知，系统的平衡点是渐近稳定的，即切换面是可达的。

3 飞机速度滑模观测器设计

飞机速度v不能直接测量得到，因此，需要设计观测器对其进行观测。由飞机纵向受力：

$$M\dot{v}=T_0+k_{\mathrm{v}}v-F_d-F_{\mathrm{s}}-F_v-F_{\omega}$$

设有外界干扰改变了飞机滑跑时的初始速度，飞机速度观测值为$\hat{v}$，飞机加速度观测值为$\dot{\hat{v}}$，外界干扰$H(t)$为$\dot{\hat{v}}$与v差值的符号函数。

定义滑模控制量：

$$\sigma = \hat{v} - v \tag{21}$$

有

$$\dot{\sigma} = \dot{\hat{v}} - \dot{v} \tag{22}$$

则滑模观测器的方程为

$$M\dot{\hat{v}} = T_0 + k_v\hat{v} - F_d - F_s - F_v - F_\omega + H(t) \tag{23}$$

设 Lyapunov 函数为 $V = \frac{1}{2}\sigma^2$, 若要满足 $\dot{V} = \sigma\dot{\sigma} \leqslant 0$, 只需要满足:

$$K \leqslant -k_v + \frac{\rho C_d S(\hat{v}+v)}{2} + \frac{\rho C_p S_p(\hat{v}+v)}{2} + \frac{\mu\left[G - \frac{\rho C_L S(\hat{v}+v)}{2}\right]}{2} \tag{24}$$

令 $K \leqslant -k_v + K_1 + K_2 + K_3$, 其中, K_1、K_2、K_3 分别为 $\frac{\rho C_d S(\hat{v}+v)}{2}$、$\frac{\rho C_p S_p(\hat{v}+v)}{2}$、$\frac{\mu\left[G - \frac{\rho C_L S(\hat{v}+v)}{2}\right]}{2}$ 的上界。

飞机速度观测器和机轮角速度传感器可以分别获得飞机纵向速度和机轮角速度这 2 个变量, 为防滑刹车控制器的设计提供了数据基础。干跑道工况下飞机速度的观测效果如图 4 所示。假设飞机落地速度为 72 m/s, 滑模观测器的初始速度为 70 m/s, 结果显示, 所设计的滑模观测器可以实现对飞机速度的准确观测。

因此, 基于变指数分数阶指数趋近律的 AABS 积分滑模控制系统总体方法框图如图 5 所示, 各部分功能互补, 协同配合。

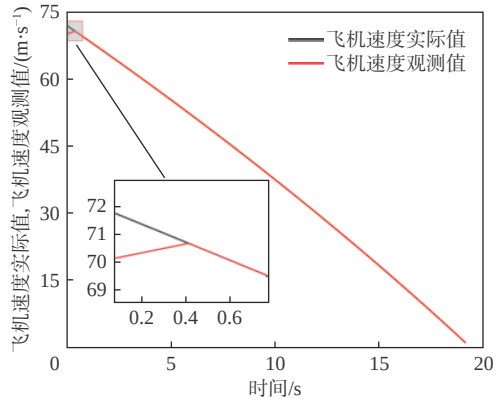


图 4　飞机速度实际值与飞机速度观测值

Fig. 4　Actual aircraft speed and observed aircraft speed

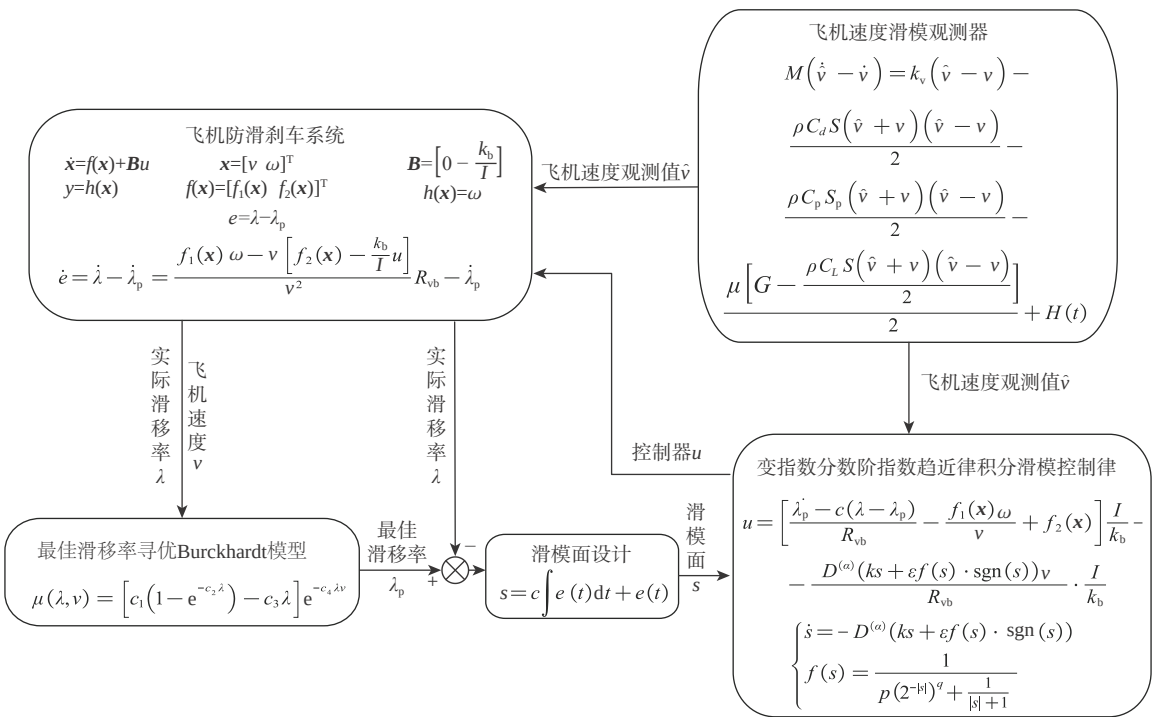


图 5　飞机防滑刹车控制系统闭环控制

Fig. 5　Closed-loop control diagram of AABS

4　仿真分析

为研究本文提出的变指数分数阶指数趋近律的积分滑模控制律是否适用于不同的跑道工况, 在 MATLAB 平台中利用 Simulink 进行仿真建模和分析, 并分别对干沥青、湿沥青、雪 3 种跑道工况进行仿真验证。

研究 ISMC-VFERL 与 LSMC、ISMC 在 3 种工况下对最佳滑移率跟踪的快速性和准确性, 并对飞机速度和机轮速度进行观测, 对滑移距离进行评估, 仿真对比结果如图 6～图 8 所示。其中, 图 6(a)、图 7(a)、图 8(a) 为 3 种控制律对最佳滑移率的跟踪

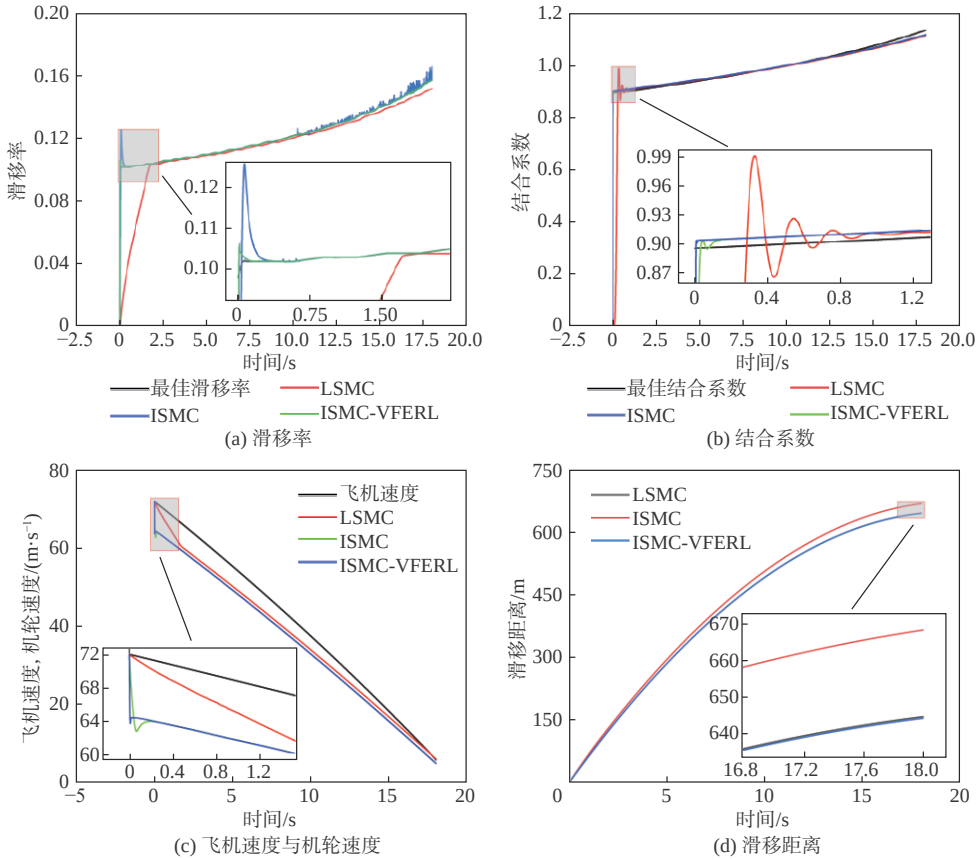


图 6 干沥青跑道下 3 种控制律对比

Fig. 6 Comparison of three control laws under dry asphalt runway

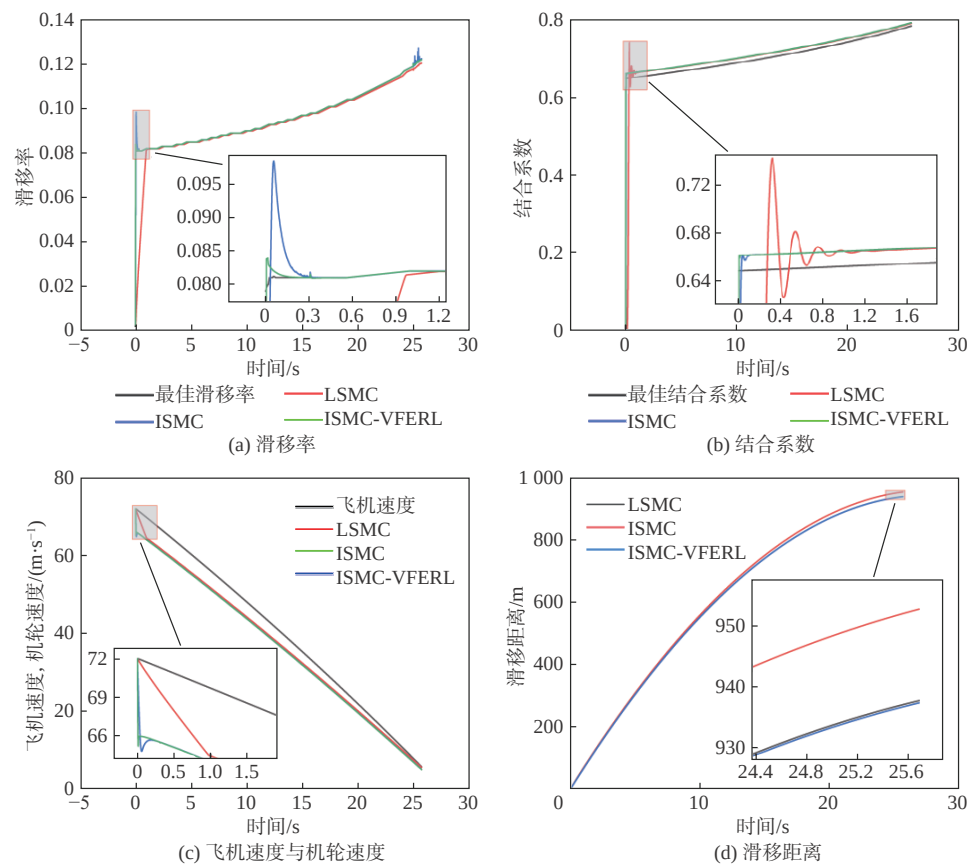


图 7 湿沥青跑道下 3 种控制律对比

Fig. 7 Comparison of three control laws under wet asphalt runway

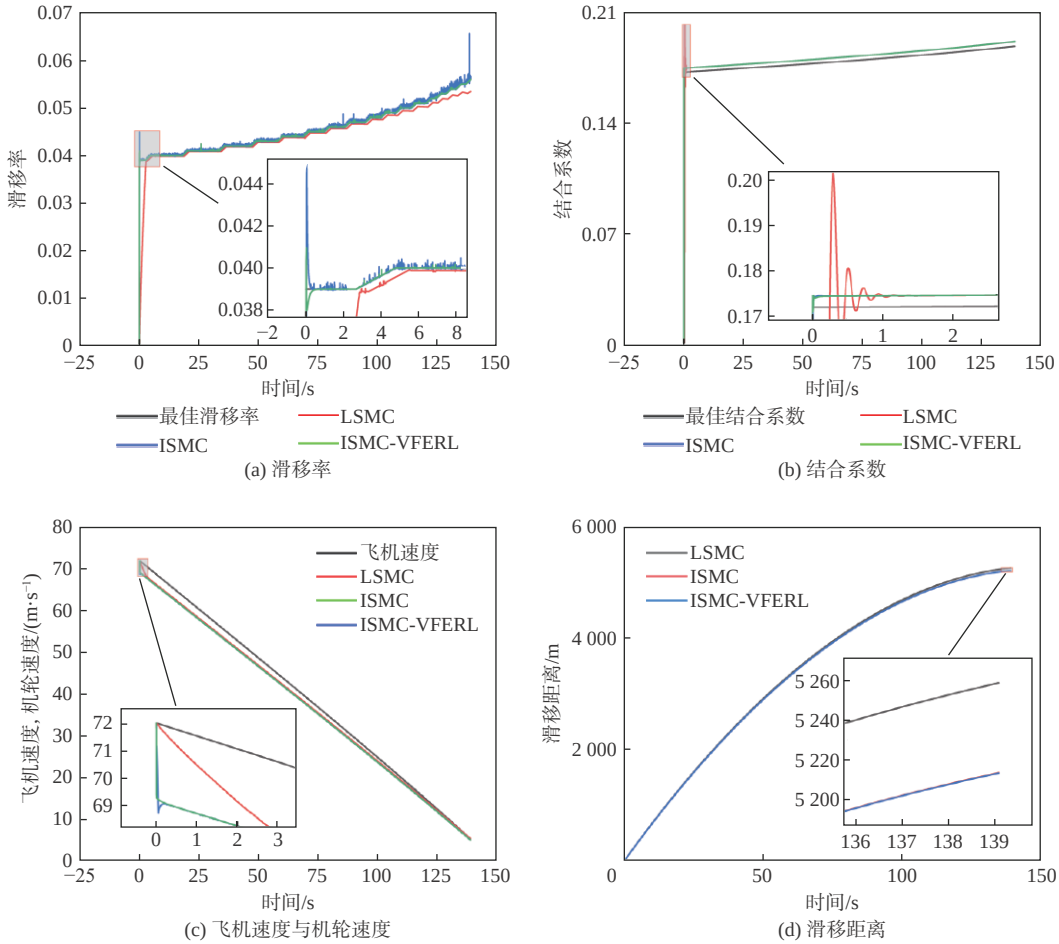


图 8　雪跑道下 3 种控制律对比

Fig. 8　Comparison of three control laws under snow-covered runway

效果对比, 图 6(b)、图 7(b)、图 8(b) 为对最大结合系数的跟踪效果对比, 图 6(c)、图 7(c)、图 8(c) 为飞机速度和机轮速度对比, 图 6(d)、图 7(d)、图 8(d) 为刹车滑移距离对比。

从仿真结果可见, 在干沥青跑道、湿沥青跑道和雪跑道 3 种工况下, LSMC、ISMC、ISMC-VFERL 在干沥青跑道的刹车时间分别为 18.339 s、18.006 s、17.992 s, 刹车滑移距离分别为 670.4 m、645.2 m、644.3 m; 在湿沥青跑道的刹车时间分别为 25.896 s、25.693 s、25.681 s, 刹车距离分别为 954.0 m、938.2 m、937.4 m; 在雪跑道的刹车时间分别为 139.146 s、139.052 s、139.043 s, 刹车距离分别为 5 221 m、5 214 m、5 213 m。可见, 在干沥青跑道、湿沥青跑道和雪跑道 3 种工况下, LSMC 难以适应刹车过程中的非线性因素, 对最佳滑移率的跟踪效果不佳, 刹车效率较低, 而采用积分滑模控制器后, 刹车过程中的稳态效果明显好转, 提出的 ISMC-VFERL 相比于 LSMC、ISMC, 同时兼顾了最佳滑移率跟踪的快速性和稳定性, 在整个刹车过程中, 飞机能够始终保持在最佳滑移率上, 且抑制了原有积分滑模的抖振现象。相比于 LSMC 和 ISMC, ISMC-VFERL 具有更好的鲁棒性, 并有效解决了 LSMC 方法中存在的趋近速度较慢和低速滑行时打滑的现象。

对顺逆风下的刹车距离和控制精度进一步仿真, 具体仿真验证及影响分析如下:

1) 假设顺风情况下的相对风速为 20 m/s, 则刹车的初始速度为 72+20 = 92 m/s。

2) 假设逆风情况下的相对风速为−20 m/s, 则刹车的初始速度为 72−20=52 m/s。

分别讨论干沥青跑道顺逆风情况下控制律的飞机防滑刹车距离和跟踪精度, 并对飞机速度和机轮速度进行观测, 仿真结果如图 9 和图 10 所示。图 9(a)、图 10(a) 为 3 种控制律分别在顺风和逆风工况下对最佳滑移率的跟踪效果对比, 图 9(b)、图 10(b) 为对最大结合系数的跟踪效果对比, 图 9(c)、图 10(c) 为飞机速度和机轮速度对比, 图 9(d)、图 10(d) 为飞机刹车滑移距离对比。

分析可知, ISMC-VFERL 顺风情况下的刹车时间为 24.471 s, 刹车距离为 1 157 m; 逆风情况下的刹车时间为 12.023 s, 刹车距离为 327.2 m; ISMC-VFERL 在无顺逆风情况下刹车时间为 17.992 s, 刹车距离为 644.3 m。顺风相对风速为 20 m/s 时, 刹车时间

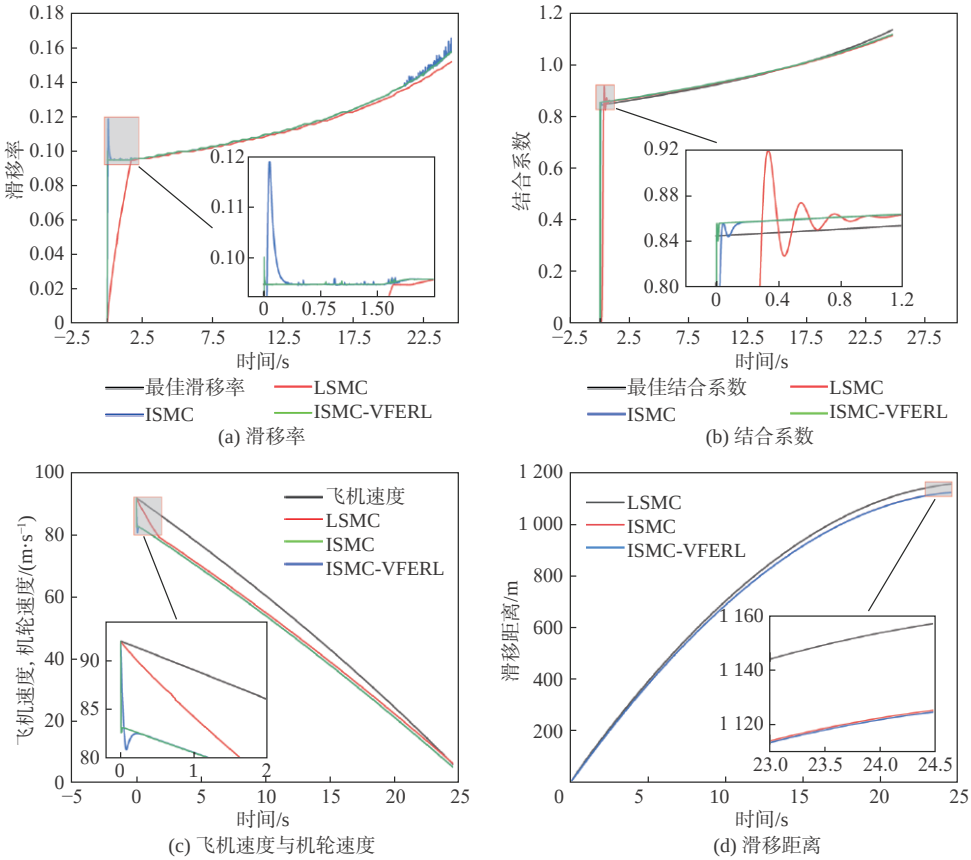


图 9 干沥青跑道顺风工况下 3 种控制律对比

Fig. 9 Comparison of three control laws under downwind condition of dry asphalt runway

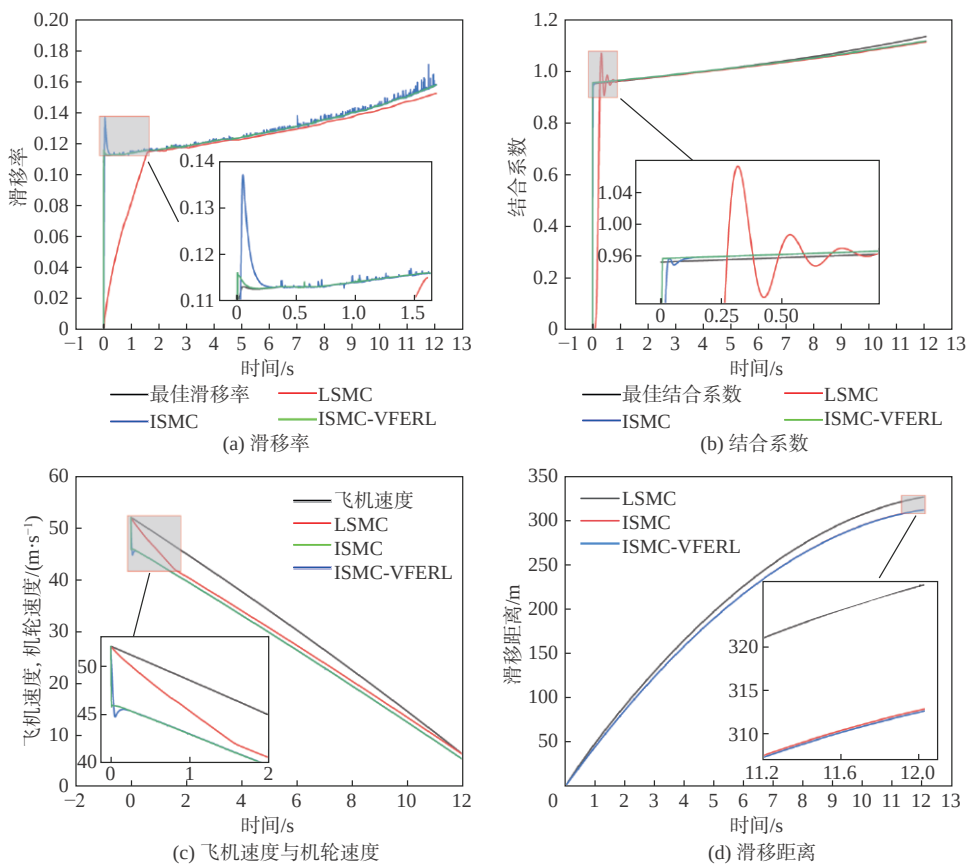


图 10 干沥青跑道逆风工况下 3 种控制律对比

Fig. 10 Comparison of three control laws under upwind condition of dry asphalt runway

增长了 6.479 s, 刹车距离增长了 492.7 m; 逆风相对风速为−20 m/s 时, 刹车时间缩短了 5.969 s, 刹车距离缩短了 317.1 m。

5 结　论

1) 实时跟踪最优滑移率信号以提高飞机防滑刹车系统制动性能。

2) 基于新型变指数分数阶指数趋近律的积分滑模控制律提高了趋近阶段的快速性和整体过程的鲁棒性, 该控制律不仅改善了系统鲁棒性, 同时也提升了刹车性能。

3) 设计滑模观测器以观测飞机速度值, 使其尽可能不受到外界干扰, 同时缩短了刹车时间和滑移距离, 提高了刹车效率。

4) 在顺风情况下, 刹车时间和刹车距离会增长, 在逆风情况下, 刹车时间和刹车距离会缩短, 但刹车精度几乎不受顺逆风影响。湿沥青跑道和雪跑道与干沥青跑道结论一致。

为使本文提出的控制律能应用于更多的综合复杂跑道工况, 后续还需要进一步设计自适应控制律。

参考文献（References）

[1] DU C L, LI F B, YANG C H, et al. Multiphase-based optimal slip ratio tracking control of aircraft antiskid braking system via second-order sliding-mode approach[J]. IEEE/ASME Transactions on Mechatronics, 2021, 27(2): 823-833.

[2] MENG Y Z, JIANG B, QI R Y. Adaptive fault-tolerant attitude tracking control of hypersonic vehicle subject to unexpected centroid-shift and state constraints[J]. Aerospace Science and Technology, 2019, 95: 105515.

[3] JIAO Z X, SUN D, SHANG Y X, et al. A high efficiency aircraft anti-skid brake control with runway identification[J]. Aerospace Science and Technology, 2019, 91: 82-95.

[4] QIU Y N, LIANG X G, DAI Z Y. Backstepping dynamic surface control for an anti-skid braking system[J]. Control Engineering Practice, 2015, 42: 140-152.

[5] TANG Y G, ZHANG X Y, ZHANG D L, et al. Fractional order sliding mode controller design for antilock braking systems[J]. Neurocomputing, 2013, 111: 122-130.

[6] HUŠEK P. Adaptive sliding mode control with moving sliding surface[J]. Applied Soft Computing, 2016, 42: 178-183.

[7] SAVITSKI D, IVANOV V, AUGSBURG K, et al. Wheel slip control for the electric vehicle with in-wheel motors: variable structure and sliding mode methods[J]. IEEE Transactions on Industrial Electronics, 2019, 67(10): 8535-8544.

[8] LI L B, SUN L L, ZHANG S Z. Mean deviation coupling synchronous control for multiple motors via second-order adaptive sliding mode control[J]. ISA Transactions, 2016, 62: 222-235.

[9] JIAO Z X, WANG Z Z, SUN D, et al. A novel aircraft anti-skid brake control method based on runway maximum friction tracking algorithm[J]. Aerospace Science and Technology, 2021, 110: 106482.

[10] D'AVICO L, TANELLI M, SAVARESI S M, et al. An anti-skid braking system for aircraft via mixed-slip-deceleration control and sliding mode observer[C]//Proceedings of the IEEE 56th Annual Conference on Decision and Control. Piscataway: IEEE Press, 2017: 4503-4508.

[11] CHEN M Q, XU F R, LIANG X L, et al. MSD-based NMPC aircraft anti-skid brake control method considering runway variation [J]. IEEE Access, 2021, 9: 51793-51804.

[12] 李繁飙, 黄培铭, 阳春华, 等. 基于非线性干扰观测器的飞机全电刹车系统滑模控制设计[J]. 自动化学报, 2021, 47(11): 2557-2569.
LI F B, HUANG P M, YANG C H, et al. Sliding mode control design of aircraft electric brake system based on nonlinear disturbance observer[J]. Acta Automatica Sinica, 2021, 47(11): 2557-2569 (in Chinese).

[13] BAKKER E, NYBORG L, PACEJKA H B. Tyre modelling for use in vehicle dynamics studies[R]. Warrendale: SAE International, 1987: 190-204.

[14] KIENCKE U, NIELSON L. Automotive control systems: forengine, driveline, and vehicle[J]. Measurement Science and Technology, 2000, 11(12): 1828.

[15] 林程, 梁晟, 宫新乐, 等. 面向极限工况的分布式驱动电动汽车动力学集成控制方法[J]. 汽车工程, 2022, 44(9): 1372-1385.
LIN C, LIANG S, GONG X L, et al. Integrated dynamic control strategy for extreme maneuvers of 4WIDEVs[J]. Automotive Engineering, 2022, 44(9): 1372-1385(in Chinese).

[16] EFE M O. Fractional fuzzy adaptive sliding-mode control of a 2-DOF direct-drive robot arm[J]. IEEE Transactions on Systems, Man, and Cybernetics Part B, Cybernetics, 2008, 38(6): 1561-1570.

Design of aircraft anti-skid braking system integral sliding mode control system based on novel reaching law

XU Meng[1, 2], LI Yan[1], GAO Jie[1, 2, *], XU Hai[3], GAO Bing[4]

(1. School of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China;
2. Key Laboratory of Technology and Equipment of Tianjin Urban Air Transportation System, Tianjin 300300, China;
3. Shenyang Aircraft Airworthiness Certification Center of CAAC, Shenyang 110043, China;
4. Engineering Techniques Training Center, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Aircraft anti-skid braking system (AABS) has the characteristics of strong time variability, strong unpredictability, and many internal and external interference factors. Therefore, an integral sliding mode control (ISMC) method of AABS with variable exponential fractional order exponential reaching law was proposed. Firstly, the dynamics model of the aircraft landing system was established. Secondly, in order to improve the rapidity and robustness of the system, a variable structure control method was introduced to realize the tracking of optimal slip rate and optimal binding coefficient. Then,by improving the ISMC law of the exponential fractional order reaching law, the chattering phenomenon was suppressed, and the tracking speed to the optimal slip rate was accelerated. In addition, a sliding mode observer was designed to observe the aircraft speed to reduce the internal and external nonlinear interference. Finally, the feasibility and effectiveness of the algorithm were verified by the MATLAB simulation platform. The simulation results show that the overall control effect of the designed integral sliding mode controller with variable exponential fractional order reaching law is better than the traditional linear sliding mode controller, and the sliding mode observer can accurately estimate the aircraft speed.The method improves the robustness of the overall system design and shortens the braking time and braking distance, and the control effect is great.

Keywords: aircraft anti-skid braking system; fractional order; integral sliding mode control; slip rate; robustness

Received: 2023-04-18; **Accepted**: 2023-08-02; **Published Online**: 2023-08-22 17:41

URL: link.cnki.net/urlid/11.2625.V.20230822.0944.002

Foundation items: National Natural Science Foundation of China (51707195,62173331); The Fundamental Research Funds for the Central Universities (3122023016)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0206

基于融合卷积 Transformer 的航空发动机故障诊断

赵洪利*，杨佳强

(中国民航大学 航空工程学院，天津 300300)

摘 要：航空发动机长期处于恶劣的气路环境下工作会面临腐蚀、侵蚀等问题，且故障参数特征不明显，因此，精准的航空发动机故障诊断方法对保证飞机安全运行具有重要意义。为提高预测准确性，提出了一种基于融合卷积 Transformer 的航空发动机故障诊断方法。利用自注意力机制提取有用特征，抑制冗余信息，并将最大池化层引入 Transformer 模型中，进一步降低模型内存消耗及参数量，缓解过拟合现象。采用基于 GasTurb 建模的涡扇发动机仿真数据集进行验证，结果与 Transformer 模型和反向传播 (BP) 神经网络、卷积神经网络 (CNN)、循环神经网络 (RNN) 等传统深度学习模型相比，准确率分别提高了 6.552% 和 28.117%、13.189%、10.29%，证明了所提方法的有效性，可为航空发动机故障诊断提供一定的参考。

关 键 词：航空发动机；故障诊断；自注意力机制；融合卷积 Transformer；深度神经网络
中图分类号：V263.6
文献标志码：A **文章编号**：1001-5965(2025)04-1117-10

安全是民航业的生命线，航空发动机作为飞机的动力装置，是保障飞行安全的重要系统。航空发动机是一种结构极为复杂的动力装置，并且长期处于高温、高压、高转速及振动等环境下工作。在多种气路条件的耦合作用下，航空发动机气路系统会产生叶片腐蚀、侵蚀、积垢等故障，造成发动机性能退化，燃油消耗率增加，工作稳定性变差[1]。因此，提出一种准确及时的航空发动机故障诊断方法，明确故障位置，对发动机在翼健康管理和维修规划有重要价值。

航空发动机发生气路故障通常体现在部件的效率和流量的变化上，但这些参数无法直接测量，需要根据航空发动机可测量的参数间接判断，如转速、温度、压力、流量及燃油流量等。当前，航空发动机故障诊断方法主要分为 2 种：①基于物理模型的方法，如卡尔曼滤波[2-3] 和粒子滤波[4]；②基于数据驱动的方法，如神经网络[5-6]、专家系统[7] 和支持向量机[8]。

基于物理模型的航空发动机故障诊断方法需要根据物理特性建立精确的航空发动机数字模型。然而，航空发动机是一个强非线性系统，气路故障过程复杂，对于用户来说，很难得到准确的发动机模型，因此，不能进行精准的诊断和预测。随着计算机技术和机器学习技术的不断发展，基于数据驱动的航空发动机故障诊断方法被越来越广泛地应用于工程和实验之中，其不依赖发动机的物理特性和热力循环，无需建立发动机物理模型，只依赖发动机的监控数据，就可以得到很好的预测效果。

深度学习是一种常用的数据驱动方法，在航空发动机故障诊断领域展示出强大的能力，也逐渐成为当前的研究重点。Malhi 等[9] 采用传统的循环神经网络（recurrent neural network, RNN）对时间序列进行预测，但由于 RNN 网络固有的位置关系，降低了模型的灵活性。为解决该问题，刘利军等[10] 提出

收稿日期：2023-04-24；**录用日期**：2023-05-26；**网络出版时间**：2023-06-09 13:09
网络出版地址：link.cnki.net/urlid/11.2625.V.20230608.1550.005
基金项目：中央高校基本科研业务费专项资金 (3122021049)；中国民航大学实验技术创新基金 (2021CXJJ90)；2022 年天津市研究生科研创新项目 (2022SKY156)；中国交通教育研究会 2022—2024 年度教育科学研究课题 (JT2022YB326)

* **通信作者**. E-mail：henleytrent@163.com

引用格式：赵洪利，杨佳强. 基于融合卷积 Transformer 的航空发动机故障诊断 [J]. 北京航空航天大学学报，2025，51（4）：1117-1126.
ZHAO H L，YANG J Q. Aero-engine fault diagnosis based on fusion convolutional Transformer[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1117-1126（in Chinese）.

了一种基于双向长短时记忆(bi-directional long short-term memory, Bi-LSTM)神经网络的航空发动机气路部件故障诊断模型,通过引入多个遗忘门来进行信息的控制。另外,王奉涛等[11]提出了一种基于卷积神经网络(convolutional neural network, CNN)的中介轴承故障诊断方法,与RNN网络和LSTM这种循环运行的神经网络不同,CNN网络通过卷积核采取局部连接的方式对原始数据进行学习,并逐步转换成更易识别的特征。但上述RNN网络和CNN网络都存在结构上的局限性。RNN网络不能进行并行运算,造成了长序列学习中的遗忘问题;CNN网络受限于卷积核的大小,会导致捕获远程信息的能力有限。Vaswani等[12]提出了一种基于注意力机制的Transformer模型,有效解决了以上问题,在图像识别[13]、语音识别[14]等领域都得到了广泛的应用,但Transformer模型仍存在模型参数量大、计算速度慢的问题。

基于此,本文提出了一种基于融合卷积的Transformer模型用于航空发动机的故障诊断。与传统Transformer不同,本文模型采用了池化层进一步降低网络的计算量,有效避免了过拟合的风险,提高了诊断准确率。

1 基于融合卷积Transformer故障诊断模型

Transformer模型是一个只依赖于自注意力机制来构建编码器(Encoder)和解码器(Decoder)架构的模型。本文用于航空发动机故障诊断的融合卷积Transformer模型主要由以下4个部分构成,分别为输入层、特征编码器、特征解码器及输出层。在特征解码器中,融入CNN网络中的池化层用来解决内部序列到序列训练速度较慢、内存消耗大的问题。融合卷积Transformer模型结构如图1所示。

1.1 输入层

输入层为位置编码模块,Transformer模型的核心是自注意力机制,但自注意力机制会导致其失去学习时序特征位置的能力,因此,需要对数据进行位置编码,来描述特征与特征之间的相互位置,使得模型的学习更加容易。Transformer模型中,位置编码计算过程为

$$\begin{cases} U_{(p,2i)} = \sin\left(\dfrac{p}{10\,000^{2i/d_{\text{model}}}}\right) \\ N_{(p,2i+1)} = \cos\left(\dfrac{p}{10\,000^{2i/d_{\text{model}}}}\right) \end{cases} \tag{1}$$

式中:p为输入数据的位置;i为对应的特征维度。每个输入数据都用周期不同的sin和cos函数计算,波长从2π逐渐增加到$10\,000\times2\pi$。

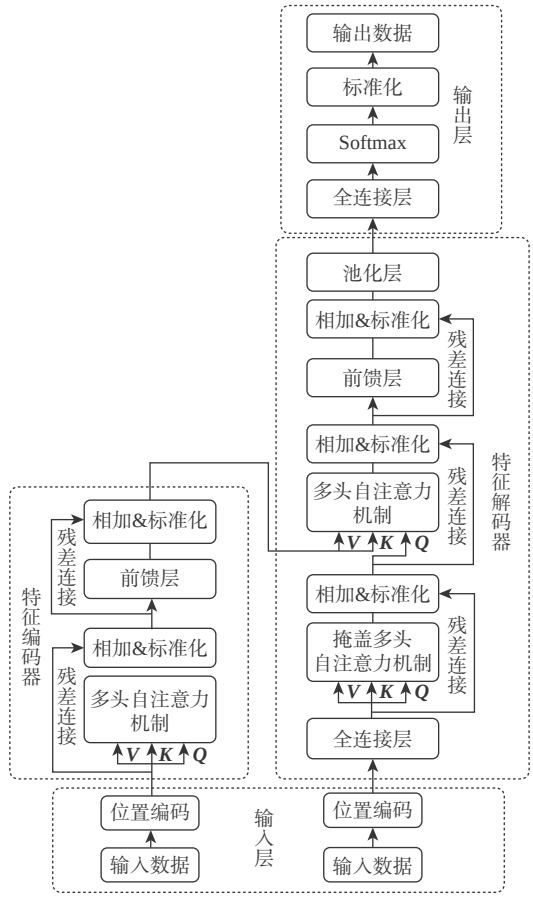


图1 融合卷积Transformer模型基本结构

Fig. 1 Basic structure of fusion convolutional Transformer model

1.2 特征编码器

特征编码器将经过位置编码的航空发动机退化参数进行特征编码,传输到特征解码器。特征编码器由多个相同的Encoder层组成,每个Encoder层由一个多头自注意力模块[15]和前馈神经网络模块组成,并且在模块之间采取残差连接[16]和层归一化[17]进行处理。

多头自注意力模块的核心原理为自注意力机制,其是注意力机制的变体。相对于注意力机制来说,多头自注意力模块减少了对外部信息的依赖,更善于捕捉信息层内部的相关性,主要通过计算权重系数来表征特征之间的相关性,进而达到传递重要信息的效果。自注意力机制模型采用缩放点积注意力来计算特征矩阵的注意力值,先通过缩放点积来计算查询矩阵$\boldsymbol{Q}$和键矩阵$\boldsymbol{K}$二者的近似相关性,再通过Softmax函数归一化处理,与值矩阵$\boldsymbol{V}$相乘,计算过程为

$$\begin{cases} \boldsymbol{Q} = \boldsymbol{W}^q \boldsymbol{I} \\ \boldsymbol{K} = \boldsymbol{W}^k \boldsymbol{I} \\ \boldsymbol{V} = \boldsymbol{W}^v \boldsymbol{I} \end{cases} \tag{2}$$

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{K}^{\mathrm{T}}\boldsymbol{Q}}{\sqrt{d}}\right)\boldsymbol{V} \tag{3}$$

式中: $\boldsymbol{I}$ 为特征矩阵; $\boldsymbol{W}^{q}$、$\boldsymbol{W}^{k}$、$\boldsymbol{W}^{v}$ 分别为 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 的权重矩阵; d 为 $\boldsymbol{K}$ 的维度。

自注意力机制是将注意力放在输入数据各个部分, 通过分配给关键信息更大的权重, 从而降低其他无关部分的影响。但单一的注意力机制只能在一个维度内关注航空发动机的退化信息, 为了更加有效地学习发动机的退化信息, 采用了多头自注意力机制, 如图 2 所示, 在多个不同维度中共同关注航空发动机的退化数据。多头自注意力机制本质是多个自注意力机制拼接在一起, 在不同维度内分别学习退化数据, 再将计算得到的输出矩阵拼接在一起后经过全连接层输出, 最终得到多头注意力值。计算过程为

$$\text{MulitiAttention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Concat}(H_1,\cdots,H_h)\boldsymbol{W} \tag{4}$$

$$H_h = \text{Attention}(\boldsymbol{I}\boldsymbol{W}_h^{\boldsymbol{Q}},\boldsymbol{I}\boldsymbol{W}_h^{\boldsymbol{K}},\boldsymbol{I}\boldsymbol{W}_h^{\boldsymbol{V}}) \tag{5}$$

式中: h 为多头自注意力头数; $\boldsymbol{W}_h^{\boldsymbol{Q}}$、$\boldsymbol{W}_h^{\boldsymbol{K}}$、$\boldsymbol{W}_h^{\boldsymbol{V}}$ 分别为第 h 个自注意力头 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 的权重矩阵; $\boldsymbol{W}$ 为多头自注意力权重矩阵。

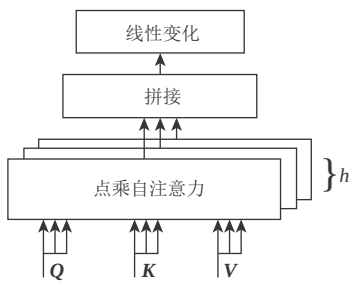


图 2　多头自注意力机制结构

Fig. 2　Structure of multi-headed self-attention mechanism

前馈神经网络模块是具有 2 层线性层的全连接网络, 由 2 层全连接层构成, 中间的隐含层使用 ReLU 函数激活, 所示前馈神经网络计算过程为

$$\text{FFN}(\boldsymbol{x}) = \max(\boldsymbol{0},\boldsymbol{x}\boldsymbol{B}_1+\boldsymbol{b}_1)\boldsymbol{B}_2+\boldsymbol{b}_2 \tag{6}$$

式中: $\boldsymbol{x}$ 为注意力层归一化后的输出矩阵; $\boldsymbol{B}$ 为权重向量; $\boldsymbol{b}$ 为偏置项。

残差连接结构位于特征编码器和特征解码器的每个子模块之间, 在引入残差网络后, 权重的变化对输出更加敏感,能够更好地优化权重矩阵。同时, 为了加快训练速度, 减少过拟合, 会将残差连接后的结果进行归一化处理, 计算过程为

$$\boldsymbol{X}_{\mathrm{B}} = \text{LN}(\text{MulitiAttention}(\boldsymbol{X}_{\mathrm{A}})+\boldsymbol{X}_{\mathrm{A}}) \tag{7}$$

$$\boldsymbol{X}_{\mathrm{C}} = \text{LN}(\text{FFN}(\boldsymbol{X}_{\mathrm{B}})+\boldsymbol{X}_{\mathrm{B}}) \tag{8}$$

式中: $\boldsymbol{X}_{\mathrm{A}}$ 为 Encoder 层的输入矩阵; $\boldsymbol{X}_{\mathrm{B}}$ 为前馈神经网络输入矩阵; $\boldsymbol{X}_{\mathrm{C}}$ 为 Encoder 层的输出矩阵; LN 表示层归一化处理函数。

1.3　特征解码器

特征解码器接收特征编码器的输出信号, 进行特征解码后传入输出层。特征解码器由多个相同的 Decoder 层组成, 在原始 Transformer 结构中, 每个 Decoder 层由一个掩盖多头自注意力模块、多头编码-解码注意力模块和前馈神经网络模块组成, 并且在模块之间采取残差连接和层归一化进行处理。与原始模型架构不同, 本文架构在 Decoder 层中增加了池化层, 位于前馈神经网络模块之后。

Transformer 模型最初常被用于机器翻译领域, 输入与输出的向量维度是一致的, 因此, 可以直接在多头编码-解码注意力模块中进行计算。然而, 在航空发动机故障诊断领域, 故障标签为一维向量, 会导致输入与输出的维度不一致, 进而无法在多头编码-解码注意力模块中计算。因此, 本文将航空发动机退化数据作为特征解码器的输入, 使模型更多地关注退化数据内部的关系。

特征解码器中, Decoder 层的堆叠方式为循环堆叠, 其运算量与 Decoder 层数成正比关系, 计算时会出现计算速度慢、参数冗余的问题。因此, 本文在特征解码器中引入 CNN 网络中常用的池化层, 位置位于标准化层后, 形成了融合卷积的 Decoder 模型, 该结构可以在保留原特征的基础上, 有效去除冗余信息, 减少内存消耗, 降低过拟合的风险。其中, 池化层采用最大池化, 最大池化是在神经元向前传播的过程中, 选取特定步长矩阵区域中值最大的点, 同时记录该点的位置信息, 计算过程为

$$y_{lij} = \max_{(m,n)\in R_{ij}} x_{lmn} \tag{9}$$

式中: y_{lij} 为与第 l 个特征有关的、在矩形区域 R_{ij} 的最大池化输出值; x_{lmn} 为矩形区域 R_{ij} 中位于 (m,n) 处的元素。

1.4　输出层

输出层接收特征解码器的输出信号, 通过全连接层特征信号进行加权求和, Softmax 层将输出值转化成概率分布形式, 经过正则化层, 有助于提高模型的泛化性能。输出层将数据展开为一维向量, 输出故障诊断预测结果。

1.5　融合卷积 Transformer 故障诊断流程

本文提出的航空发动机故障诊断训练与测试的流程如图 3 所示, 主要包括 2 部分: ①数据预处理。将发动机数据进行滤波降噪和标准化处理, 对数据添加故障标签, 并按照工作条件划分为训练集数据和测试集数据。②基于融合卷积 Transformer 航空发动机故障诊断过程。构建融合卷积 Transformer

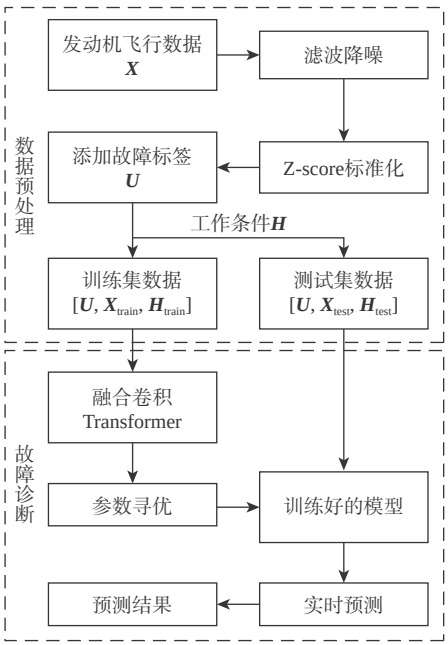


图3 航空发动机故障诊断流程

Fig. 3 Aero-engine fault diagnosis process

模型,使用训练集数据对模型进行离线训练及参数寻优,得到训练好的故障诊断模型,使用测试集数据进行在线预测,输出测试结果。

2 实验数据处理与模型构建

2.1 数据集

由于航空发动机故障数据难以获得,为验证本文提出的故障诊断方法的有效性,基于GasTurb发动机建模软件构建了涡扇发动机模型。该涡扇发动机模型主要结构包括风扇(FAN)、低压压气机(low pressure compressor, LPC)、高压压气机(high pressure compressor, HPC)、燃烧室、高压涡轮(high pressure turbine, HPT)、低压涡轮(low pressure turbine, LPT)和尾喷管。涡扇发动机结构和站位如图4所示。发动机在大气温度 T_0=15 ℃、标准大气压 P_0=10 135 Pa、相对湿度60%、低压转子修正 N_{1c}=92.25%、飞行高度AL=12 000 m、飞行马赫数 Ma=0.8时,具体属性如表1所示。

涡扇发动机的运行是通过各种部件相互作用的结果,气路部件的退化和外来物损伤都会影响气路性能,发动机气路故障占发动机故障的90%[18],

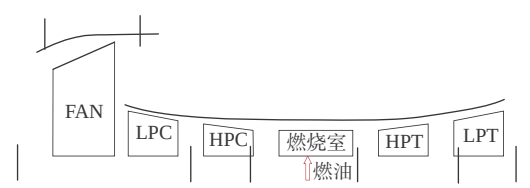


图4 航空发动机结构简图

Fig. 4 Schematic of aero-engine structure

表1 部件属性参数

Table 1 Property parameters of engine components

部件属性参数	设定值
进气道总压恢复系数	0.99
涵道比	8
风扇转速/(r·min^{-1})	4 000
风扇增压比	1.702
风扇效率/%	0.878 0
LPC转速/(r·min^{-1})	4 000
LPC增压比	4
LPC效率/%	0.870 0
HPC转速/(r·min^{-1})	18 000
HPC增压比	7
HPC效率/%	0.850 0
燃烧室燃烧效率/%	0.999 5
燃烧室总压恢复系数	0.95
HPT转速/(r·min^{-1})	18 000
HPT效率/%	0.88
HPT落压比	2.816
LPT转速/(r·min^{-1})	4 000
LPT效率/%	0.900 0
LPT落压比	7.166

因此,本文采用10个健康指数,即效率指数 E_i 和流量指数 W_i(i = 1,2,3,4,5)来表征涡扇发动机的气路性能。涡扇发动机的性能退化会导致发动机状态参数的变化,如温度、压力和转速。因此,本文选用涡扇发动机气路系统常用的传感器参数进行故障诊断。发动机的工作条件由飞行高度和飞行马赫数定义。

航空发动机模型的故障单元设为风扇故障、高压压气机故障、低压压气机故障、高压涡轮故障和低压涡轮故障。其中,对于压气机和风扇故障,部件的流量和效率指数在−0.5%～−5%之间变化。对于涡轮故障,部件的效率指数在−0.5%～−5%之间变化,流量指数在0.5%～5%之间变化,这是由于涡轮退化导致流量增加。故障部件的健康指数范围和样本数量如表2所示。

为使模拟数据更加贴近航空发动机实际工作时的条件,本文选择了若干典型工作点来验证飞行包线中的诊断准确性和稳定性性能,并以 N_{1c}=92.25%功率运行。考虑涡扇发动机实际工作条件,在飞行包线内选择了5种条件作为训练集,分别为Ⅰ(AL=0 km, Ma=0)、Ⅱ(AL=4 km, Ma=0.7)、Ⅲ(AL=6 km, Ma=0.6)、Ⅳ(AL=8 km, Ma=0.7)、Ⅴ(AL=12 km, Ma=0.8)。选择3种条件作为测试集,分别为Ⅵ(AL=1 km, Ma=0.1)、Ⅶ(AL=4 km, Ma=0.5)、Ⅷ(AL=10 km, Ma=0.8)。均为亚声速飞

行条件, 如图 5 所示。图中, 运行条件Ⅰ和Ⅵ在地面附近, 马赫数较低; 运行条件Ⅱ、Ⅲ和Ⅶ的飞行高度处于中间水平; 运行条件Ⅳ、Ⅴ和Ⅷ处在高高度、高飞行速度状况。

发动机在实际运行中, 传感器测量得到的状态参数会存在一定的误差, 因此, 本文对模拟数据添加了零平均值和单位标准偏差(1σ)的随机正态分布测量噪声[19], 噪声的具体值如表 3 所示。

表 2　航空发动机参数规范

Table 2　Parameter specification of aero-engine

故障部件	健康指数/%	样本数量
FAN	E_1, $W_1\in[-0.5, -5.0]$	5 152
LPC	E_2, $W_2\in[-0.5, -5.0]$	5 152
HPC	E_3, $W_3\in[-0.5, -5.0]$	5 152
HPT	$E_4\in[-0.5, -5.0]$, $W_4\in[0.5, 5.0]$	5 152
LPT	$E_5\in[-0.5, -5.0]$, $W_5\in[0.5, 5.0]$	5 152

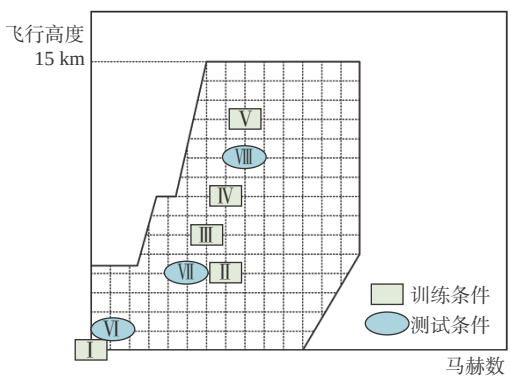


图 5　飞行包线中的典型工作点

Fig. 5　Typical engine running points in flight envelope

表 3　模拟参数的测量噪声范围

Table 3　Range of measurement noise for simulated parameters

状态参数	数值
燃油流量W_{FF}标准噪声值/(kg·s^{-1})	0.0176
进口总温T_2标准噪声值/K	± 2.6
进口总压P_2标准噪声值/%	± 0.5
风扇出口总温T_{13}标准噪声值/K	± 2.6
风扇出口总压P_{13}标准噪声值/%	± 0.5
HPC入口总温T_{25}标准噪声值/K	± 2.6
HPC入口总压P_{25}标准噪声值/%	± 0.5
HPC出口总温T_3标准噪声值/K	± 3.3
HPC出口总压P_3标准噪声值/%	± 0.5
LPT入口总温T_{45}标准噪声值/K	± 9.38
LPT入口总压P_{45}标准噪声值/%	± 0.5
LPT出口总温T_5标准噪声值/K	± 6.64
LPT出口总压P_5标准噪声值/%	± 0.5
高压转子转速N_2标准噪声值/%	± 0.1

使用 t 分布随机邻域嵌入(t-distributed stochastic neighbor embedding, t-SNE)方法对故障数据集进行降维可视化分析, 航空发动机数据的三维结果如图 6 所示。数据集中的不同故障在图中用不同颜色进行表示, 可以看出, 不同故障类型的数据点在空间中相互重叠, 说明几种故障状态所表现出的参数特征具有很强的耦合性。

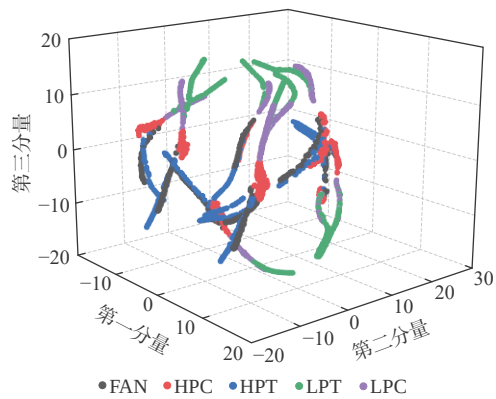


图 6　t-SNE 特征可视化

Fig. 6　t-SNE feature visualization

2.2　数据预处理

由于获得的故障数据具有较大的噪声干扰和随机波动, 会对预测效果产生不利影响。因此, 使用数字滤波器对故障数据进行降噪滤波处理。本文使用小波阈值滤波方法, 其基本原理是: 将信号经过小波变换后, 有效信号产生的小波系数较大, 噪声的小波系数较小, 经过阈值处理, 大于阈值的小波系数进行保留, 小于阈值的则认为是噪声, 从而达到降噪的目的[20]。具体流程如图 7 所示。

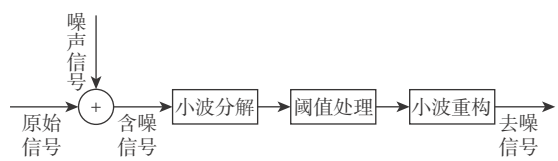


图 7　小波阈值去噪过程

Fig. 7　Wavelet threshold denoising process

航空发动机在不同工作条件下工作时, 气路参数之间数据差异大, 且不同参数的量纲也不同, 因此, 需要对数据进行标准化处理。为不改变数据的分布情况, 本文选用 Z-score 标准化的方法, 计算过程为

$$\sigma=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2} \tag{10}$$

$$z=\frac{x-\mu}{\sigma} \tag{11}$$

式中: μ 为总体数据的均值; σ 为标准差; z 为标准化之后的值。

将数据进行标准化处理后，根据数据的故障类型生成故障标签，各类标签如表4所示。

表4 航空发动机各类故障标签

Table 4 Various fault labels for aero-engine

故障种类	故障标签
FAN故障	0
HPC故障	1
HPT故障	2
LPT故障	3
LPC故障	4

2.3 模型性能评价标准

为验证融合卷积Transformer模型预测航空发动机故障诊断的性能，本文通过混淆矩阵来表示模型的预测结果，混淆矩阵示意图如图8所示。

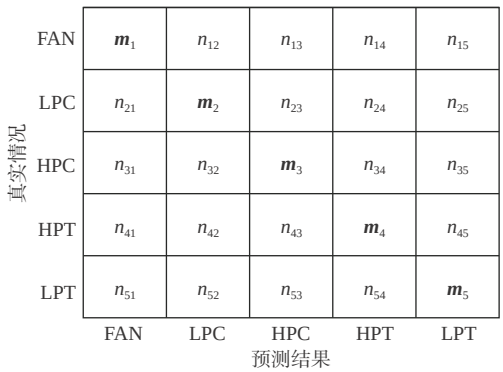


图8 混淆矩阵示意图

Fig. 8 Schematic of confusion matrix

通过混淆矩阵，本文选取4个参数来验证模型性能，分别为准确率、精准率、召回率及F_1值，计算过程为

$$\begin{cases} T_i = m_i \\ O_j = \sum_{i=1}^{5} n_{ij} \qquad i \neq j \\ V_i = \sum_{j=1}^{5} n_{ij} \qquad j \neq i \end{cases} \tag{12}$$

$$\begin{cases} A = \sum_{i=1}^{5} m_i \Big/ \sum_{i,j=1}^{5} n_{ij} \qquad i \neq j \\ P_y = \dfrac{T_y}{T_y + O_y} \\ R_y = \dfrac{T_y}{T_y + V_y} \\ F_{1y} = \dfrac{2P_yR_y}{P_y + R_y} \end{cases} \tag{13}$$

式中：T为真实类别和模型预测结果相同的数量；O为预测结果为此类，真实结果为其他类别的数量；V为真实结果为此类，预测结果为其他类别的数量；A为模型对所有数据的预测准确率；P表示模型预测为此类的结果中，真正为此类的数据所占的比例；R表示真实为本类的所有结果中，模型预测正确的数据比例；F_1值综合了准确率、精准率的结果，是综合评价指标。本文主要使用准确率和F_1值来表征模型性能。

2.4 实验环境和模型参数设置

本文所提网络模型在一台配置了Intel Core i7-11800H处理器、NVIDIA GeForce RTX 3060 Laptop GPU、16 GB RAM的计算机上搭建，语言环境为Python3.7，深度学习语言库版本为Pytorch1.11.0。模型使用的优化算法为随机梯度下降(stochastic gradient descent, SGD)算法，批训练大小设置为10，学习率设置为0.000 1，迭代次数设置为50次，Dropout系数为0.1，融合卷积模块中的最大池化(MaxPool)层的池化窗口大小和步长均设置为2。损失函数为交叉熵损失函数，计算过程为

$$L = \frac{1}{N}\sum_i L_i = -\frac{1}{N}\sum_i \sum_{c=1}^{M} y_{ic} \ln a_{ic} \tag{14}$$

式中：M为类别的数量；y为2个样本是否为同类型，当样本i的真实类别等于c时，y取1，否则取0；a为样本i的真实类别属于c时的概率。

3 实验设置与结果

3.1 融合卷积Transformer优化实验

在融合卷积Transformer模型中，编解码器的层数及多头自注意力机制的头数会对算法的诊断结果具有直接影响。编解码器的层数过多，造成模型收敛速度慢、梯度消失及过拟合。多头自注意力机制是在多个不同的投影空间中寻找输入向量之间的相互关系，但头数过多会影响模型的复杂度。为避免模型效率降低和过拟合问题，选择合适的自注意力头数和编解码器层数非常关键。

在模型中，特征维度大小必须可以整除多头自注意力的头数，因此，本文选择1、2、7、14这4种头数配置来进行实验，其他网络参数保持一致。经过50个Epoch的训练，4种方案的准确率如图9所示，评价指标如表5所示。

由实验结果可以看出，融合卷积Transformer模型的收敛速度快，自注意力头数为1和2时，模型的预测准确率要优于头数为7和14时，且头数为2时，模型的稳定度要优于头数为1时，综合考虑，模型自注意力头数选择为2。

在Transformer模型中，编解码器的层数为6，本文采用网格搜索方法，分别对编解码器层数为5、6、7的情况进行寻优。经过50个Epoch的训

练, 各方案的准确率及运行时间如表 6 所示。可以看出, 当编解码器层数为 5+5、5+6、5+7、6+5、6+6 时, 准确率均超过 98.5%, 得到了很好的预测效果, 综合考虑模型运行时间的影响, 得到层数结构为 5+5 时模型训练性能最好。

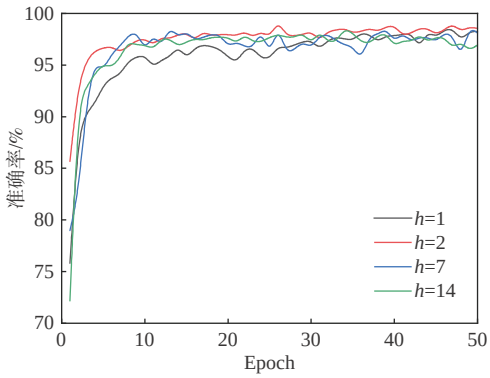


图 9　不同自注意力头数准确率对比

Fig. 9　Accuracy comparison of different self-attention heads

表 5　不同自注意力头数的对比结果

Table 5　Comparison results of different self-attention heads

h	F_1值					准确率/%
	FAN 故障	HPC 故障	LPC 故障	HPT 故障	LPT 故障	
1	0.978 42	0.971 83	0.977 78	0.989 01	0.996 36	98.261
2	0.982 08	0.978 72	0.981 55	0.989 01	0.996 36	98.551
7	0.978 57	0.989 24	0.989 01	0.978 41	0.970 37	98.116
14	0.975 09	0.948 91	0.953 41	0.977 78	0.992 75	96.957

表 6　不同编解码器层数的对比结果

Table 6　Comparison results of different encoder and decoder layers

编解码器层数结构	运行时间/s	准确率/%	损失函数值
5+5	140.02	98.986	3.543 4
5+6	170.99	98.986	3.768 6
5+7	180.64	98.551	5.013 1
6+5	163.97	98.696	4.096 3
6+6	184.38	98.551	3.825 9
6+7	197.45	97.826	5.535 2
7+5	173.84	98.401	5.324 3
7+6	188.31	97.826	7.516 8
7+7	207.74	98.116	6.769 4

在自注意力头数为 2、编解码器层数为 5+5 时, 数据未经过滤波处理时, 模型预测的准确率为 89.565%, 与数据经过降噪处理之后的准确率相比下降了 9.421%。同时, 本文采取的测试集数据处于不同的工作条件, 证明本文模型具有很好的泛化性。

3.2　算法对比实验

为验证本文融合卷积 Transformer 模型的性能, 构建了 Transformer 模型、BP 神经网络、CNN 和 RNN 网络作为对比模型。模型的具体参数设置如下:

1) Transformer 模型。综合考虑模型的复杂度和性能, 本文将多头自注意力头数设置为 2, 编解码器层数设置为 6 层, 采用 SGD 优化器更新参数, 学习率设置为 0.000 1。

2) BP 神经网络。激活函数选择 ReLU 激活函数, 神经元个数设置为 100, 采用 SGD 优化器更新参数, 学习率设置为 0.000 1。

3) CNN 网络。采用多层卷积结构, 第一层卷积层卷积核大小为 2, 池化层窗口大小为 2, 第二层卷积层卷积核大小为 2, 池化层窗口大小为 4。采用 Adam 优化器更新参数, 学习率设置为 0.000 1。

4) RNN 网络。采用多层 RNN 结构, 层数设置为 2, 采用 SGD 优化器更新参数, 学习率设置为 0.000 1。

对于对比模型, 均采用相同的数据和预处理方式, 与融合卷积 Transformer 进行比较研究, 训练过程如图 10 所示, 预测结果如图 11 所示。

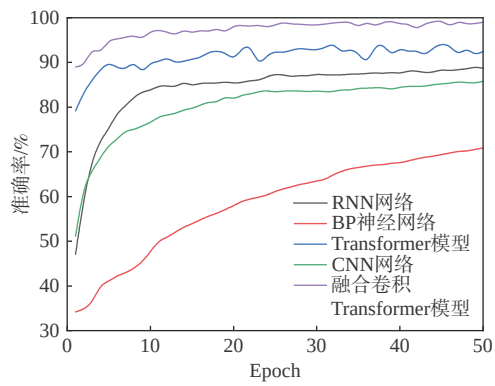


图 10　不同算法训练过程

Fig. 10　Training processes of different algorithms

从表 7 可以看出, 融合卷积 Transformer 模型相比于 BP 神经网络、CNN 和 RNN 网络, 准确率分别提升了 28.117%、13.189% 和 10.29%, 具有较高的预测可靠性。相比于 BP 神经网络、CNN 和 RNN 网络, 本文模型凭借独特的编解码器结构和自注意力机制, 对发动机中不同的特征进行区分, 对关键信息给予更高的关注, 有效增加了航空发动机故障诊断的精度。

从表 8 可以看出, 相比于 Transformer 模型, 本文模型预测的准确率提高了 6.522%, 在进行最大池化 (MaxPool) 操作后, 保留了特征矩阵中局部的关键信息的同时, 模型参数量减少了 75%, 总参数量减少了 121 657, 内存消耗降低了 35.84 MB, 运行速度提高了 84.57 s。

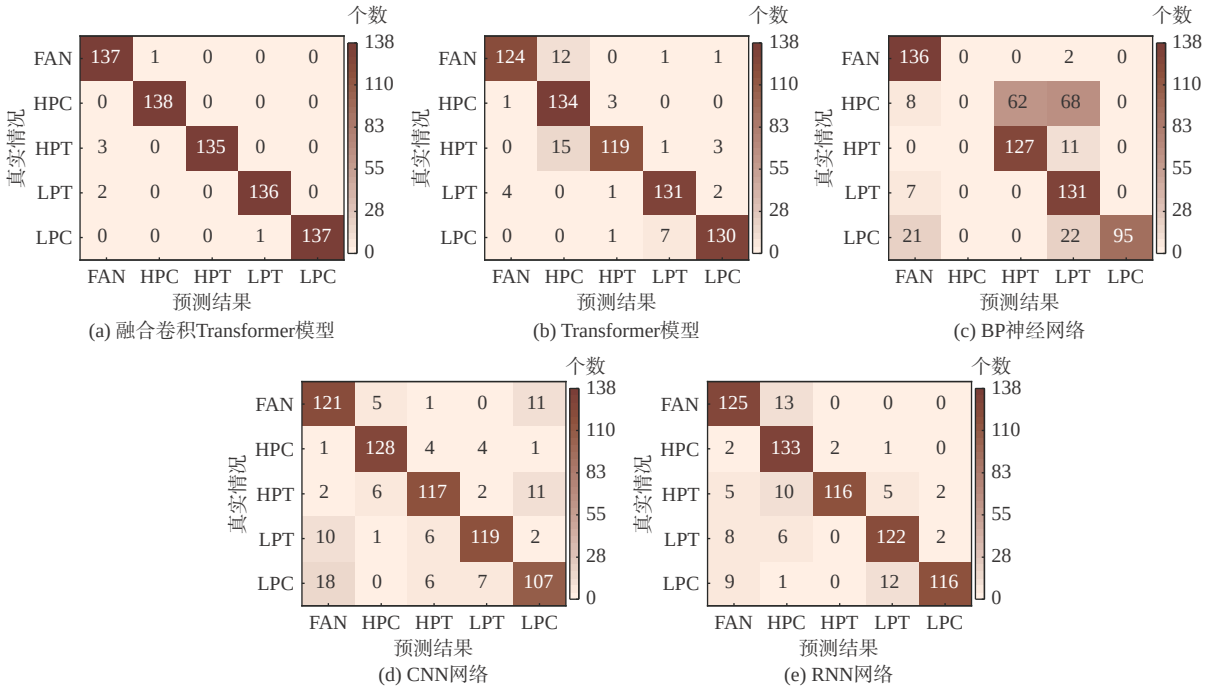


图 11 故障分类混淆矩阵

Fig. 11 Fault classification confusion matrix

表 7 不同算法的效果对比

Table 7 7Effect comparison of different algorithms

网络类型	准确率/%	损失函数值
融合卷积Transformer模型	98.986	3.543 4
Transformer模型	92.464	23.116 6
BP神经网络	70.869	94.608 4
CNN网络	85.797	38.677 5
RNN网络	88.696	42.261 4

表 8 本文模型与 Transformer 模型对比

Table 8 Comparison with the proposed model and Transformer model

模型	准确率/%	池化后参数量	总参数量	内存消耗/MB	运行时间/s
融合卷积Transformer	98.986	35	607 466	2 391.04	140.02
Transformer	92.464	140	729 123	2 426.88	224.59

4 结 论

本文提出了一种基于融合卷积 Transformer 的航空发动机故障诊断方法，并在仿真数据集上进行验证，结论如下：

1）融合卷积 Transformer 模型通过自注意力机制捕捉输入向量之间的相互关系，通过加入池化层降低了模型过拟合的风险，对不同高度的测试数据都有很好的诊断效果，具有很高的泛化性。

2）相比于 BP 神经网络、CNN 和 RNN 网络，本文模型的预测准确率分别提升了 28.117%、13.189% 和 10.29%，可有效提高航空发动机故障诊断的准确性，对及时发现潜在故障、保障飞机的安全飞行有重要的价值。

3）相比于 Transformer 模型，本文方法准确率提升了 6.552%，运行速度减少了 84.57 s，且在模型参数总量、内存消耗上有所降低，进一步证明了本文所提模型在航空发动机故障诊断领域的优越性。

参考文献（References）

[1] LI H H, GOU L F, ZHENG H, et al. Intelligent fault diagnosis of aeroengine sensors using improved pattern gradient spectrum entropy[J]. International Journal of Aerospace Engineering, 2021, 2021(1): 8868875.

[2] VOLPONI A J. Gas turbine engine health management: past, present, and future trends[J]. Journal of Engineering for Gas Turbines and Power, 2014, 136(5): 051201.

[3] BORGUET S, DEWALLEF P, LE'ONARD O. A way to deal with model-plant mismatch for a reliable diagnosis in transient operation [C]//Proceedings of the ASME Turbo Expo 2006: Power for Land, Sea, and Air. New York: ASME, 2008: 593-602.

[4] LIU X F, YUAN Y, SHI J, et al. Adaptive modeling of aircraft engine performance degradation model based on the equilibrium manifold and expansion form[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2014, 228(8): 1246-1272.

[5] ZHOU H W, HUANG J Q, LU F, et al. Echo state kernel recursive least squares algorithm for machine condition prediction[J]. Mechanical Systems and Signal Processing, 2018, 111: 68-86.

[6] 崔建国, 李国庆, 蒋丽英, 等. 基于深度自编码网络的航空发动机故障诊断[J]. 振动测试与诊断, 2021, 41(1): 85-89.
CUI J G, LI G Q, JIANG L Y, et al. Aero-engine fault diagnosis based on deep self-coding network[J]. Journal of Vibration, Mea-

surement & Diagnosis, 2021, 41(1): 85-89(in Chinese).

[7] PU X X, LIU S M, JIANG H D, et al. Sparse Bayesian learning for gas path diagnostics[J]. Journal of Engineering for Gas Turbines and Power, 2013, 135(7): 071601.

[8] 徐启华, 师军. 基于支持向量机的航空发动机故障诊断[J]. 航空动力学报, 2005, 20(2): 298-302.
XU Q H, SHI J. Aero-engine fault diagnosis based on support vector machine[J]. Journal of Aerospace Power, 2005, 20(2): 298-302 (in Chinese).

[9] MALHI A, YAN R Q, GAO R X. Prognosis of defect propagation based on recurrent neural networks[J]. IEEE Transactions on Instrumentation and Measurement, 2011, 60(3): 703-711.

[10] 刘利军, 雷宇, 余臻. 双向 LSTM 模型在航空发动机气路故障诊断的应用[J]. 航天控制, 2020, 38(5): 67-72.
LIU L J, LEI Y, YU Z. The application of bidirectional LSTM model in aero-engine gas path fault diagnosis[J]. Aerospace Control, 2020, 38(5): 67-72(in Chinese).

[11] 王奉涛, 薛宇航, 王洪涛, 等. GLT-CNN 方法及其在航空发动机中介轴承故障诊断中的应用[J]. 振动工程学报, 2019, 32(6): 1077-1083.
WANG F T, XUE Y H, WANG H T, et al. GLT-CNN and its application of aero-engine intermediary bearing fault diagnosis[J]. Journal of Vibration Engineering, 2019, 32(6): 1077-1083(in Chinese).

[12] VASWANI A, SHAZEER N, PARMER N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[13] YOO S, KWON O, LEE H. Image-to-graph Transformers for chemical structure recognition[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2022: 3393-3397.

[14] LU L, LIU C L, LI J Y, et al. Exploring Transformers for large-scale speech recognition[EB/OL]. (2020-08-11)[2023-04-01]. https://arxiv.org/abs/2005.09684v2.

[15] KIM Y, DENTON C, HOANG L, et al. Structured attention networks[EB/OL].(2017-02-16)[2023-04-01]https://arxiv.org/abs/1702.00887v3.

[16] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

[17] JIMMY L B, JAMIE R K, GENFFREY E H. Layer normalization[EB/OL]. (2016-07-21)[2023-04-01]. https://arxiv.org/abs/1607.06450.

[18] 黄金泉, 王启航, 鲁峰. 航空发动机气路故障诊断研究现状与展望[J]. 南京航空航天大学学报, 2020, 52(4): 507-522.
HUANG J Q, WANG Q H, LU F. Research status and prospect of gas path fault diagnosis for aeroengine[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(4): 507-522(in Chinese).

[19] ALOZIE O, LI Y G, PILIDIS P, et al. An integrated principal component analysis, artificial neural network and gas path analysis approach for multi-component fault diagnostics of gas turbine engines[C]//Proceedings of the ASME Turbo Expo 2020: Turbomachinery Technical Conference and Exposition. New York: ASME, 2021.

[20] ZHU K G, PENG C, JING C Y, et al. Suppression of motion noise based on a linear-homomorphic filtering algorithm in airborne electromagnetic survey[J]. Geophysical Prospecting, 2021, 69(5): 1102-1115.

Aero-engine fault diagnosis based on fusion convolutional Transformer

ZHAO Hongli*, YANG Jiaqiang

(School of Aeronautical Engineering, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Aero-engine faces the problems of corrosion and erosion when working in an atmospheric environment for a long time, and the features of fault parameters are not obvious. Therefore, an accurate aero-engine fault diagnosis method is of great significance to ensure the safe operation of aircraft. To improve the prediction accuracy, an aero-engine fault diagnosis method based on a fusion convolutional Transformer was proposed, which used the self-attention mechanism to extract useful features and restrain redundant information. In addition, the MaxPool was introduced into the Transformer model to reduce the model memory consumption further and the number of parameters and mitigate overfitting. The turbofan engine simulation dataset based on GasTurb modeling was used for verification. Compared with Transformer network and other traditional deep learning models, such as back propagation (BP) neural network, convolutional neural network (CNN), and recurrent neural network (RNN), the proposed model has improved prediction accuracy by 6.552%, 28.117%, 13.189%, and 10.29%, which proves its effectiveness, and it can provide a reference for aero-engine fault diagnosis.

Keywords: aero-engine; fault diagnosis; self-attention mechanism; fusion convolutional Transformer; deep neural network

Received: 2023-04-24; **Accepted**: 2023-05-26; **Published Online**: 2023-06-09 13:09

URL: link.cnki.net/urlid/11.2625.V.20230608.1550.005

Foundation items: The Fundamental Research Funds for the Central Universities (3122021049); Civil Aviation University of China Experimental Technology Innovation Fund (2021CXJJ90); 2022 Tianjin Postgraduate Research Innovation Project (2022SKY156); China Communications Education Research Association (CCERA) 2022—2024 Educational Research Projects (JT2022YB326)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0240

基于混合整数规划的终端区柔性进场航迹优化

赵向领*，周志亮
(中国民航大学 空中交通管理学院，天津 300300)

摘 要：随着航空运输量持续增长和空域容量日渐饱和，固定进场程序对航班调配缺乏灵活性，潜在冲突多、调配困难、等待时间长、碳排放量大等问题日趋凸显。综合考虑多航班在终端区的进场、进近、排序、汇聚等运行过程，提出和描述了基于柔性进场航迹优化解决方案。基于终端区和航空器运行特征，设计终端区航班进场时空离散网络栅格。以航迹和总飞行长度最短为目标，建立柔性进场航迹的混合整数规划模型，分别考虑网络栅格规则、航空器转弯特性、飞行冲突、安全间隔、航迹连续性等限制条件，以天津机场终端区为例，基于不同的交通密度等级和终端区入口时间窗，设置多组典型实验案例，使用商业求解器 Gurobi 对所建模型进行求解验证。结果表明：柔性进场航迹较标准仪表进场程序 (STAR)，航迹长度平均缩短 127.375 km，飞行总长度平均降低 534.75 km，航班进场时间平均减少 6.24 min；且在相同交通密度等级，入口时间窗设置合理的情况下，柔性进场航迹对航班量的小幅波动具有一定的鲁棒性。

关 键 词：空中交通管理；进场航班；四维航迹；进场排序；柔性空域；混合整数规划
中图分类号：V355
文献标志码：A **文章编号**：1001-5965(2025)04-1127-16

终端区空域作为空中交通网络关键节点，运行环境日益复杂。繁忙终端区管制员的航班调度，通常会出现航班额外等待、碳排放量大、噪声污染等问题，如何提高终端区运行效率成为解决空中交通问题的关键。

目前，航班进场通常采用标准仪表进场程序 (standard instrument arrival procedure, STAR)，通过管制指挥和雷达引导完成进场、进近和着陆任务。航班实施阶梯下降方式，管制员需频繁发出调度指令，导致航班时刻难以预测，管制员的工作负荷增加，并引发过多燃油消耗、碳排放等问题。中国部分机场实施基于性能的导航 (performance based navigation, PBN)程序，实现飞行的部分自主性，增加终端区容量，提高了运行效率。但是基于固定的飞行程序，不能充分发挥 PBN 程序简洁、灵活的导航优势与飞机性能潜力，造成空域资源浪费。在进场排序方面，先到先服务 (first come first served, FCFS) 进场方式虽然易于执行，但无法充分利用空域资源。为适应日益增长的航班量，考虑一定范围内的航班位置偏移及合理调整航班排序受到越来越多的重视。鉴于此，在终端区优化进场航迹和排序对经济和环境均有重要意义。

为实现管制自动化和智能化，民航“十四五”规划制定推进空管数字化系统建设，积极推广进场管理系统 (arrival management system, AMAN)。AMAN 作为一种自动化进港排序与调度系统，为管制员提供决策支持，为进场航班提供安全、高效的运行环境；并根据飞行计划和雷达数据等预测进场轨迹，计算航班预计到达时间，管制部门根据获取的信息向飞机发布相应指令，优化进场航班序列。但是，目前 AMAN 对进场排序与航迹优化是分开计算的，并未实现对航班序列与航迹的同时优化；另外，

收稿日期：2023-05-11；**录用日期**：2023-10-13；**网络出版时间**：2023-11-17 17:10
网络出版地址：link.cnki.net/urlid/11.2625.V.20231117.1428.003
基金项目：国家自然科学基金 (52272356)；中国民航大学研究生科研创新资助项目 (2022YJS087)
*通信作者. E-mail：zxl-llx@163.com

引用格式：赵向领，周志亮. 基于混合整数规划的终端区柔性进场航迹优化 [J]. 北京航空航天大学学报，2025，51（4）：1127-1142.
ZHAO X L，ZHOU Z L. Flexible arrival trajectory optimization at terminal airspace based on mixed integer programming[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1127-1142（in Chinese）.

其作用范围仅限于终端区，调度范围小，易造成机场附近等待的问题。

为确保进场航迹与进场交通密度相适应，充分发挥PBN技术与AMAN的优势，参考基于实际运行环境自主规划飞行航迹的自由航线空域(free route airspace, FRA)概念[1]，转变STAR的传统运行模式，本文设计柔性进场程序方案，合理利用空闲时隙，调整航班进场的时间与序列，协同优化四维航迹与进场排序。

目前，国内外从不同方面对进场航班运行效率进行研究，主要包括辅助进场程序设计、进场航迹优化、进场排序优化等。

在辅助进场程序的研究中，2006年欧控提出点融合程序，通过设计扇形排序弧，优化不同来向的航班融合和排序[2]，有效提升了航班的进场效率，增加了进场航迹的灵活性，但在交通密度较高或空域受限制的机场，可容纳的航班数量有限，依然会导致航班大面积延误[3-6]。

对于进场排序与调度的研究，戴喜妹以最小化偏离时间为目标，提出基于复合分配规则的两阶段优化方法[7]。刘继新等考虑不同交通密度下的进场航班，设计了带精英策略的快速非支配排序遗传算法[8]。乐美龙等基于多跑道、多路径选择，建立混合整数规划模型[9]。张军峰等综合考虑旅客、空管、机场和航空公司的诉求，提出了基于多目标帝国竞争算法的进场排序与调度方法[10]。Samà等引入变邻域搜索和禁忌搜索算法，有效提高进场效率[11]。Liang等基于滑动时间窗口机制，采用模拟退火算法进行动态优化[12]。Hong等针对飞机排序与调度问题提出两阶段随机规划方法，建立混合整数规划模型并采用粒子群算法进行决策[13]。Salehipour等建立混合整数规划模型，有效减少了着陆时间与目标时间总偏差[14]。

为进一步的提升航班进场效率，多数学者开始考虑进场航迹的选择。张洪海等提出基于航迹运行的截点直飞方式，建立混合整数规划模型，用商业求解器Gurobi求解[15]。杨磊等设计倒皇冠形的柔性进场程序，有效提升终端区运行的经济性和飞行效率[16]。陈昇等运用模拟退火算法和A*算法，生成进场程序的汇聚结构[17]。Cecen考虑3种冲突解脱方案，建立混合整数规划模型优化进场航迹；通过路径延伸与点融合方法，有效降低进场油耗[18-19]。Lee和Balakrishnan用动态规划方法将航班调度问题描述为网络最短路径问题[20]。Zhang等提出空中交通流诱导规则，引导进场航班在终端区自主选择路径[21]。Sáez等通过建立混合整数规划模型，安排航班沿最佳飞行剖面实现连续下降[22]。Chida等通过调整航向、速度或时隙等决策变量，分别规划不同场景下进场程序的最优拓扑结构[23]。目前对进场航班优化的研究大多体现在航班序列优化，或固定航迹优化方面，二者的组合研究仍有一定的提升空间。

在二者组合优化方面，Sáez等考虑到日益紧张的空域资源和交通动态性需求，通过分析飞机性能与运行环境，建立混合整数规划模型，规划适应交通需求的动态进场航迹，同时安排进场序列，在一定的时间窗口内高效安排航空器进场[24-25]。本文主要基于其研究，进一步完善冲突解脱模型，实现终端区的无冲突且高效运行。

原进场程序航路结构固定，当空中交通密度较低时，不能沿最优航迹进场，导致燃油消耗的增加；在交通密度较高时，终端区空域无法容纳全部进场航班，航空器在机场附近盘旋，导致延误、碳排放增多和噪声污染。

因此，为保证终端区的高效运行，将柔性进场航迹作为未来的研究方向。柔性进场航迹与传统的固定进场飞行程序相对应，是指根据航班密度和交通流量特征，实时规划和调整飞行程序。

本文在终端区空域，基于航空器性能，在保证无冲突运行条件下，规划给定时间段内航班的柔性进场航迹；同时，实现航班的排序融合，即终端区完整进近过程。

本文采用离散化时空网络模型对柔性进场航迹规划问题进行描述，以最小化进场航迹长度和航班飞行总长度为目标，考虑网络规则限制、转弯限制、飞行冲突、安全间隔、连续性等约束条件，建立柔性进场航迹的混合整数规划模型。以天津滨海国际机场日常运行数据为例进行算例验证，分析了进场航班密度、时间窗口偏移量等特征参数对航迹的影响。

1 问题描述

1.1 柔性进场航迹规划问题

在终端区，给定一段时间内的进场航班数量和预计进场时间窗，在满足飞行及汇聚安全的条件下，规划多组无冲突进场四维航迹，确保从不同方向进场的航班，在较短的时间内，沿航迹逐步排序融合，进近着陆。

柔性体现在打破传统的固定进场程序，基于不同时段进场交通密度与交通流特征，规划不同的进场航迹。该模式用于规划一个固定时间段内(如0.5 h、1 h、2 h等)的航班进场程序；不同时段，交通

密度不同, 进场程序不同。

图 1 为终端区水平方向上的进场航迹。终端区边界有 n 个入口(图中的 D_1～D_3), 在给定的时间段内, 来自不同方向上的 s 架航班从不同的入口进入终端区。需要规划航班从不同入口到达最终融合点(图中的 M_p)的四维航迹及飞行参数, 即各航班在终端区内任何时刻的经纬度、高度、航向和速度等; 并充分利用最终融合点的容量, 实现进场排序。在整个进场时间段内, 所有航空器遵循自身性能与管制间隔等限制。

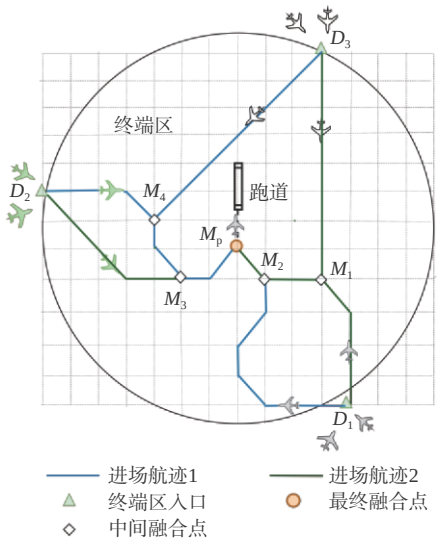


图 1　水平方向上的进场航迹

Fig. 1　Arrival trajectory form in horizontal direction

因此, 终端区柔性进场航迹优化问题可描述为: 在给定时间段内, 以实现所有航班的航迹总长度和到达融合点的时间最短为目标, 在遵循飞机性能和安全间隔条件下, 规划从不同入口进入终端区的每架航班到达融合点的航迹, 并对全部航班在融合点排序, 实现在融合点的安全汇聚。

1.2　时空网络栅格

如图 1 所示, 进场航班在自由飞行的终端区内可选择任意航迹, 在航班量较少、无冲突的条件下, 可以选择直飞融合点。但如果航班量较大, 就必须避免飞行冲突, 对航迹进行重新规划。由于在连续时间和空间内搜索无冲突的航迹计算成本较高, 本文将时间和终端区空域离散化处理。

航班的进场时间可以离散为一系列时刻序列 $T=\{t_1, t_2, \cdots, t_{|T|}\}$, 则每个航班的进场四维连续航迹可以描述为基于时刻序列 T 的离散航迹点。终端区空域离散为一系列的正方形栅格, 即网络图 $G(V, E)$, 其中, V 表示栅格顶点集合, 除边界顶点外, 每个顶点周边有 8 个相邻顶点; E 表示栅格边的集合, 任意相邻的 2 个顶点 i 和 j、边和对角线 (i, j) 或 (j, i) 都属于 E。正方形栅格的边长和对角线长度为进场航班时刻序列中前后相邻 2 个时刻点的时间间隔与航班进场速度的乘积。在进场过程中, 基于时刻序列的航班离散航迹点与终端区栅格上的顶点相匹配, 即每个时刻, 航班都在某一个栅格顶点上。多个航班同时进场过程中, 可能出现航班冲突, 则空闲的栅格顶点可以作为突发状况下的冲突解脱的改航点。

基于离散时空网络, 为便于预测与描述四维航迹, 定义如下。

1) 终端区入口。原标准仪表进场程序中的终端区入口, 供来自不同方向的航班进入终端区, 如图 1 中的 D_1～D_3。设 D 为终端区入口集合, $D=\{D_1, D_2, \cdots, D_{|d|}\}$, $D \subseteq V$。

2) 最终融合点。所有航班为进近着陆, 需要在该点进行汇聚(即图 1 中的 M_p, $M_p \in V$); 在无冲突的前提下, 依次通过该点降落至跑道。终端区入口与最终融合点属于特殊顶点。

3) 航迹点。其为被规划到进场航迹中的网络栅格顶点。在该点, 航班可以调整飞行、航向, 决策下一点。航迹点序列是航班进场航迹的组成部分。

4) 航迹段。其为相邻航迹点之间的航迹, 多个航迹段组成从入口至最终融合点的进场航迹。本文中, 航迹段是被选中的网络栅格的边与对角线。

1.3　基于时空网络的问题描述

通过时空网络图转化后, 将连续的航迹规划问题转化为离散的给定时间序列的不同航班的航迹点分配问题, 方便了问题描述和建模。

基于时空网络图, 终端区进场柔性航迹优化问题可描述为: 在给定的时刻序列内, 以实现所有航班经过的网络航迹段数量最少为目标, 在每个时刻, 满足飞机性能和安全间隔条件下, 为从不同终端区入口进入网络的每一架航班分配一个航迹点; 保证每架航班连续时刻航迹点的连续性, 并避免不同航班的栅格顶点的冲突; 最终实现所有航班按先后顺序在融合点汇聚、排队、着陆。

为实现最优进场航迹和排序, 同时便于应用到实际运行环境中, 在优化调度时考虑以下问题。

1) 航班终端区入口时间窗。以航班预计到达终端区入口点时刻为基准, 给定一个可浮动的微小时间段, 作为到达入口点的时间窗。规划航迹时, 在给定时间窗内为航班分配一个具体的进入终端区时刻。由于时间窗较短, 一般不超过 5 min, 分配时刻和预计到达时刻之间的差异不超过时间窗, 航班完全可以在巡航或下降阶段吸收掉。

2) 调度策略。在终端区空域内, 从同一个终端区入口进入的航班之间不允许相互超越(即同一条

进场航线)。在改变航向时,飞机的转弯角度需要满足动力学限制(飞机性能要求),选择合适的角度,确保运行安全。

3)下降方式假设。航班遵循连续下降运行(continuous descent operations, CDO)进近着陆,不考虑机型的差异,假设经过同一终端入口的全部航班沿相同的下降角度进近着陆。因航班下降梯度相同,则在给定入口高度、融合点高度及确定飞行距离之后,可以确定航班在任一航迹点的高度,不需要再决策。因此,连续进近假设,把三维轨迹规划简化为二维航迹问题。

4)航迹段航行时间相同假设。一般航班在终端区飞行速度相差不大,可假设每个航迹段航行所需时间相同。

5)融合点限制。规划的进场航线的结构主要与航线汇聚方式和融合点的位置有关,考虑到运行的安全性和管制工作负荷,规定在每个融合点只能有2条航线进行汇聚,鉴于此,中间融合点的位置与终端区入口的数量和进场航班特征相关联,一方面,需合理安排汇聚位置和航班到达融合点的时刻,确保无冲突运行;另一方面,2个融合点之间需确保一定的距离,并且与终端区入口保持一定距离,确保时间可预测性与安全运行。

2 模型建立

为实现进场航迹与排序的优化,本文考虑实际终端区运行环境、飞机性能、航行安全等方面的因素,建立基于栅格的混合整数规划模型。

2.1 决策变量

引入0-1整数决策变量x_{ij},如式(1)所示,用于判断栅格中的一个边是否被选为航迹段,如果是为1,否则为0。该边的前后2个顶点对为(i,j)。

引入变量f_{ij}表示沿航迹段(i,j)飞行的航班量,即从顶点i到顶点j流量,$(i,j)\in E$,流量具有非负特性,其表达式为

$$f_{ij}\geqslant 0 \quad \forall (i,j)\in E \tag{1}$$

若边(i,j)有正的流量,则该边为一个航迹段,约束如下:

$$x_{ij}\geqslant \frac{f_{ij}}{M} \quad \forall (i,j)\in V \tag{2}$$

式中:M为较大的正数,可取研究问题的整体时间段内的航班总量。

2.2 目标函数

1)最小化进场航迹长度。在满足各种约束条件下,缩短航班进场路径,使航迹从终端区入口到达最终融合点距离最短。该目标可表达为

$$\min \sum_{(i,j)\in E} l_{ij}x_{ij} \tag{3}$$

式中:l_{ij}为栅格边或对角线(i,j)的长度,$(i,j)\in E$。

2)最小化航班飞行总长度。在进场过程中,为体现柔性航迹的优越性,航迹段的航班量f_{ij}与该航迹段长度l_{ij}的乘积为该时段的飞行总长度,实现航班进场总长度最小。目标函数为

$$\min \sum_{(i,j)\in E} l_{ij}f_{ij} \tag{4}$$

两目标函数的组合如式(5)所示。考虑到进场过程中,尽可能节约整体航班的飞行时间和能耗,降低所有航班的飞行长度,故在该2个目标函数中,优先考虑最小化飞行总长度。

$$\min\left(a\frac{\sum\limits_{(i,j)\in E} l_{ij}x_{ij}}{|D|}+(1-a)\frac{\sum\limits_{(i,j)\in E} l_{ij}f_{ij}}{|A|}\right) \tag{5}$$

式中:a为权重系数,在案例测试中,设置a为0.1;A为该时间段内全部进场航班的集合,$|A|$为A中的航班数量。

2.3 网络栅格的基本规则限制

2.3.1 网络流平衡

入口是网络栅格的起始点,所有航班通过入口进入网络,因此,网络起始点没有进入边,只有出边。设经过入口b进入的航班集合为A_b,$A_b\subseteq A$,$|A_b|$表示经入口b进入的航班量,则:

$$\sum_{k:(k,b)\in E} f_{kb}-\sum_{j:(b,j)\in E} f_{bj}=-|A_b| \quad b\in D \tag{6}$$

所有通过终端区入口的航班,在最终融合点汇聚,排序进入跑道。模型中最终融合点为网络的结束点,只有进入边,没有出边。由于多数机场在某一时段为单跑道降落,通常只有1个最终融合点,本文只考虑单融合点情况。因此,通过最终融合点M_{p}的航班量为所有入口航班量之和:

$$\sum_{k:(k,M_{\mathrm{p}})\in E} f_{kM_{\mathrm{p}}}-\sum_{j:(M_{\mathrm{p}},j)\in E} f_{M_{\mathrm{p}}j}=|A| \tag{7}$$

除终端区入口和最终融合点之外,进出每个顶点的航班数量相同:

$$\sum_{k:(k,i)\in E} f_{ki}-\sum_{j:(i,j)\in E} f_{ij}=0 \quad i\in V/\{D\cup M_{\mathrm{p}}\} \tag{8}$$

2.3.2 顶点出度与入度

为避免冲突,降低汇聚压力,不是所有方向的航迹都首次在融合点M_{p}汇聚。部分航班可以在终端区的其他点(中间融合点,如图1中的M_1)先汇聚到同一条航迹中,然后再飞往最终融合点。

为避免航迹太复杂,不便于在实际运行环境中应用,规定最多允许有2条航迹在融合点汇聚,即

融合点的入度最大为 2, 如下:

$$\sum_{k:\,(k,\,i)\in E} x_{ki} \leqslant 2 \quad \forall i \in V/D \tag{9}$$

为保障航班在终端区逐渐趋于融合, 而不是分散, 对于非入口点与非最终融合点的网络顶点 i, 如果被选择作为航迹点, 则该点至多有 1 个出度, 即航班经过该点只能有 1 个航迹段, 其不等式为

$$\sum_{j:\,(i,\,j)\in E} x_{ij} \leqslant 1 \quad \forall i \in V/\left\{D \cup M_{\mathrm{p}}\right\} \tag{10}$$

最终融合点为结束点, 出度为 0(见式(11)), 经过该点, 航班沿固定航线着落到跑道, 不需要再规划航迹。

$$\sum_{j:\,(M_{\mathrm{p}},\,j)\in E} x_{M_{\mathrm{p}}j} = 0 \tag{11}$$

2.4　转弯限制

在四维航迹规划过程中, 航班主要通过改变航向搜索航迹。通过转弯角度 θ 来表示航向的改变。θ 指航班的航向相对于原有航向的改变大小, 如图 2 所示, 基于网络栅格特点, 航班航迹段 (i, j) 后, 在点 j 有 7 个可选转弯方向。由于受飞机性能和空气动力限制, 在每个航迹点和飞行方向上, 较短时间内转弯角度不能太大。根据这个特性, 规定航班单步转弯角度 θ 不超过 45°。因此, 航班的转弯只能在前方、右前、左前 3 个方向上选择。

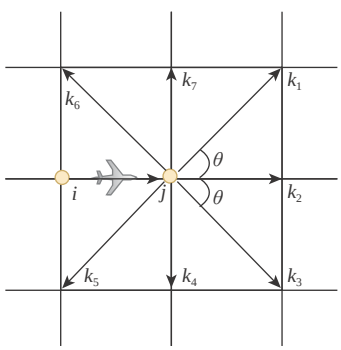


图 2　改变航向时的航迹段选择

Fig. 2　Trajectory selection during course change

如图 2 所示, 假设边 (i, j) 为进场航迹段, 航班从边 (i, j) 航行至下一航迹点时, 需要决策航向, 因此, 顶点 $k_1, k_2, \cdots, k_7$ 为潜在的航迹点。根据规定的转弯角度 θ 可知, 不可行的顶点集合为 $\{k_4, k_5, k_6, k_7\}$, 令 G_{ij} 为转弯角度大于 θ 的出边的集合, $|G_{ij}|$ 表示集合中个体数量。如果 j 不为 M_{p} 点, 则 $G_{ij}=\{k_4, k_5, k_6, k_7\}$, 即 $|G_{ij}|=4$; 如果 j 为 M_{p} 点, 则 $G_{ij}=\{k_1, k_2, \cdots, k_7\}$, 则转弯约束为

$$\left|G_{ij}\right| x_{ij} + \sum_{k\in G_{ij}} x_{jk} \leqslant \left|G_{ij}\right| \quad \forall (i,\ j) \in E \tag{12}$$

在航迹规划中, 用式 (12) 决策可行的航迹段。x_{jk} 表示航班选择的可行航迹段, 在式 (12) 不等号的右侧, 设置上限为 $|G_{ij}|$, 因此, 可以禁止使用集合 G_{ij} 中的任意边, 保证选择的航迹段符合转弯角度限制。

2.5　飞行冲突限制

在进场航迹规划中, 最重要的是避免航班之间发生飞行冲突。在航班进场过程中, 参考文献 [24-25] 的冲突限制, 考虑航迹段交叉冲突及尾流造成的冲突, 为避免冲突, 考虑对头冲突、顶点交叉冲突、航迹段交叉冲突。

以下将为 3 种冲突建立约束模型。其中, 对于交叉冲突, 基于网络特点, 可分为网络顶点交叉冲突和单元格内对角线航迹段交叉冲突。

由于安全间隔限制, 后机在同一航迹段不可以超越前机; 且由于航班在终端区飞行速度相差不大, 基于航迹段航行时间相同假设, 后机不能在同一航迹段超越前机。故安全间隔限制和假设已限制超越冲突的发生, 无需增加追赶冲突约束。

为表达航班的时空位置, 引入 0-1 整数决策变量 y_{pit}, 用于判断航班 p 在时刻 t 是否位于航迹点 i, 如果是为 1, 否则为 0。

2.5.1　对头冲突限制

对头冲突是指 2 架航班沿同一航迹段对头飞行产生的冲突。如图 3 所示, 航班 p 在航迹段 (i, j) 飞行, 同时, 航班 q 要经过航迹段 (j, i), 从而产生对头冲突。

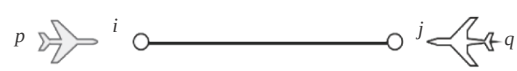


图 3　对头冲突限制示意图

Fig. 3　Schematic diagram of head-on conflict constraints

设航班在通过边 (i, j) 的航行时间为 t_{ij}, 由于本文模型假设所有航班以相同的恒定速度飞行, 2 个对头飞行航班在边 (i, j) 的航行时间相同, 即 $t_{ji}= t_{ij}$。设航班 p 当前时刻 t_p, 位于顶点 i, 即将沿航迹段 (i, j) 飞行到顶点 j。为避免对头冲突, 参考文献 [26] 的对头冲突解脱约束, 则航班 p 在航迹段 (i, j) 飞行的时间段 $[t_p, t_p+ t_{ij}]$ 内, 边 (i, j) 不能有其他航班进入。则对头进入航迹段 (j, i) 的航班只能在 t_p-t_{ij} 时刻之前或 $t_p+ t_{ij}$ 之后进入, 如图 4 所示。

因此, 为避免航班对头冲突, 需要保证只有 1 架航班在时间 $[t_p-t_{ij}, t_p+ t_{ij}]$ 之间通过航迹段 (i, j) 或 (j, i), 可表示为

$$y_{pit_p} + \sum_{t_q\in(t_p-t_{ij},\,t_p+t_{ij})} y_{qjt_q} \leqslant 1 \quad \forall p \in A_b, q \in A/A_b \tag{13}$$

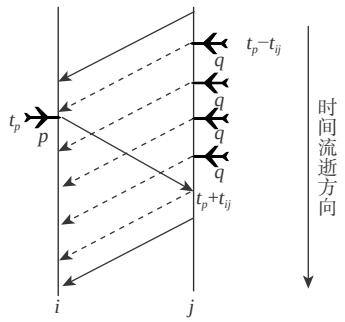


图4 对头冲突的时空网络

Fig. 4 Time-space network of head-on conflict

2.5.2 顶点交叉冲突限制

当多架航班从不同航迹点向同一航迹点飞行时，则在该点可能会发生冲突，即顶点交叉冲突。

如图5所示，设有多架航班要通过顶点j，顶点j相邻点的顶点集合为N_j，$N_j=\{i_1, i_2, \cdots, i_8\}$，则在某个时刻可能在顶点$j$处发生冲突。因此，需要在一定时刻限制航迹点处的航班流量。

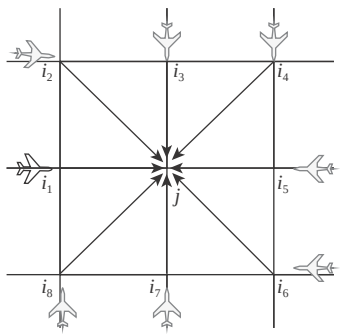


图5 顶点交叉冲突示意图

Fig. 5 Schematic diagram of vertex cross-conflict

引入0-1决策变量y_{pijt}用于判断航班p在时间段$[t_i, t_j]$是否在航迹段(i, j)，如果在为1，不在为0。

在给定的时间窗$[t_i, t_j]$内，在所有以j为到达顶点的航迹段$\{(i, j)|i\in N_j\}$中，最多只能有1个航迹段被占用，即

$$\sum_{i\in N_i} y_{pijt}\leqslant 1 \quad \forall j\in V/\{D\cup M_{\mathrm{p}}\}, \forall t\in t_{ij} \tag{14}$$

2.5.3 航迹段交叉冲突限制

由于在单元格的不同对角线存在交叉，则多条对角线不能同时做航迹段。可分为3种情况。

1）非入口与融合点单元格对角线航迹段冲突。在航迹规划过程中，从不同入口至最终融合点的航迹可能出现交叉，导致终端区空域复杂度增加，易造成航班之间的冲突。假设终端区网络有$m\times n$个单元格，每个单元格的对角线依据飞行方向不同可以产生4个顶点i、i+1、i+n、i+n+1，围成的单元格对应的4个可用对角线航迹段为$(i, i+n+1)$、$(i+1, i+n)$、$(i+n+1, i)$和$(i+n, i+1)$。则4个可用航迹段最多有1个组成进场航迹，如下：

$$x_{i(i+1+n)}+x_{(i+1+n)i}+x_{(i+n)(i+1)}+x_{(i+1)(i+n)}\leqslant 1$$
$$\forall i, i+1+n, i+n, i+1\in V/\{D\cup M_{\mathrm{p}}\} \tag{15}$$

2）入口单元格对角线航迹段冲突。由于终端区入口是该网络的航班起始点，没有到达该点的航迹段，因此，在终端区入口所处的单元格中，只需要考虑3条航迹段。如图6所示，如果航迹点i、i+1、i+n、i+n+1其中之一为终端区入口，则没有其他3点到达该点的对角线航迹段，则其他3个可用航迹段最多有1个航迹段组成进场航线。式(16)～式(19)分别表示终端区入口处于单元格不同位置的航迹段交叉冲突约束：

$$x_{i(i+1+n)}+x_{(i+n)(i+1)}+x_{(i+1)(i+n)}\leqslant 1 \quad \forall i\in D \tag{16}$$

$$x_{i(i+1+n)}+x_{(i+1+n)i}+x_{(i+1)(i+n)}\leqslant 1 \quad \forall i: i+1\in D \tag{17}$$

$$x_{i(i+1+n)}+x_{(i+1+n)i}+x_{(i+n)(i+1)}\leqslant 1 \quad \forall i: i+n\in D \tag{18}$$

$$x_{(i+1+n)i}+x_{(i+1)(i+n)}+x_{(i+n)(i+1)}\leqslant 1 \quad \forall i: i+n+1\in D \tag{19}$$

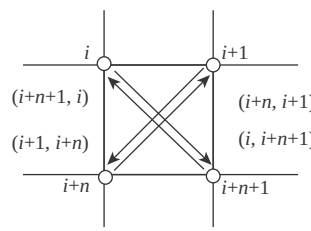


图6 单元格内的4个对角航迹段

Fig. 6 Four diagonal trajectory segments for grid

3）最终融合点单元格对角线航迹段冲突。由于最终融合点是所有航班的终点，全部进场航迹在该点汇聚，该点没有出边。因此，在最终融合点所在单元格内，只需考虑3条可用航迹段的交叉情况。如图7所示，蓝色标记的点代表最终融合点，则单元格内最多有1个航迹段组成进场航线。最终融合点邻近的4个单元格的对角线航迹段冲突约束分别为

$$x_{(M_{\mathrm{p}}-n-1)M_{\mathrm{p}}}+x_{(M_{\mathrm{p}}-1)(M_{\mathrm{p}}-n)}+x_{(M_{\mathrm{p}}-n)(M_{\mathrm{p}}-1)}\leqslant 1 \tag{20}$$

$$x_{(M_{\mathrm{p}}-n+1)M_{\mathrm{p}}}+x_{(M_{\mathrm{p}}-n)(M_{\mathrm{p}}+1)}+x_{(M_{\mathrm{p}}+1)(M_{\mathrm{p}}-n)}\leqslant 1 \tag{21}$$

$$x_{(M_{\mathrm{p}}+n-1)M_{\mathrm{p}}}+x_{(M_{\mathrm{p}}-1)(M_{\mathrm{p}}+n)}+x_{(M_{\mathrm{p}}+n)(M_{\mathrm{p}}-1)}\leqslant 1 \tag{22}$$

$$x_{(M_{\mathrm{p}}+n+1)M_{\mathrm{p}}}+x_{(M_{\mathrm{p}}+n)(M_{\mathrm{p}}+1)}+x_{(M_{\mathrm{p}}+1)(M_{\mathrm{p}}+n)}\leqslant 1 \tag{23}$$

2.6 安全间隔限制

航空器在空中飞行时，机翼上下表面产生的压力差在翼尖处产生涡流，并向后扩散，形成尾涡，俗称尾流。如果同一航线上的后机进入前机尾流，将会受到尾涡的滚转力矩影响，威胁飞行安全。因此，航班之间需要保持一定的间隔，称为尾流间

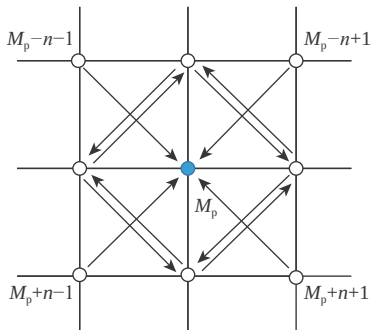


图 7　最终融合点栅格对应的 3 个可用对角航迹段

Fig. 7　Three available diagonal trajectory segments in each final merge point grid

隔。同时, 管制员为安全控制和方便指挥, 通常也会要求同一航线上的前后两架航班之间要保持一定的间隔, 称为管制间隔。尾流间隔和管制间隔必须同时满足, 因此, 取这 2 种间隔最大值作为最小限制间隔。

假设前后两架航空器之间所需保持的最小限制时间间隔为 σ, 令 A 为该时间段进入终端区的全部航空器的集合。考虑进场航迹前后两架航班之间的安全间隔, 若 t 时刻有航班 p 在航迹点 i, 则后续航班在时间段 $[t, t+\sigma]$ 内不能到达 i, 如下:

$$\sum_{\tau=t}^{t+\sigma}\sum_{p\in A} y_{pit} \leqslant 1 \quad \forall i \in V, \forall \tau \in T \tag{24}$$

2.7　进场航迹的连续性约束

进场航迹需要从终端区入口到最终融合点形成一条连续的航迹, 由一系列首尾相接的航迹段组成, 中间不能出现间断或跳跃。

图 8 为航班 p 规划进场航迹。航班 p 在相邻航迹点之间的飞行时间为 w, 航班 p 在当前时刻 t 位于航迹点 i, 则在下一时刻 $t+w$ 只能给该航班分配顶点 i 的邻接顶点, 如 k_1、k_2、k_3 等; 不能分配非邻接顶点, 如 g、h 等, 否则, 其进场航迹就不再连续, 出现跳跃。

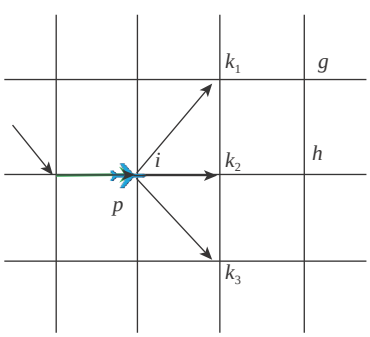


图 8　进场航迹结构示意图

Fig. 8　Schematic diagram of arrival trajectory structure

为清楚表达各终端区入口至最终融合点的进场航迹的连续特性, 引入 0-1 整数决策变量 x_{bij} 用于决策航迹段 (i, j) 是否为经入口 b 通往最终融合点的进场航迹的一部分, 如果是为 1, 否则为 0, 即

$$x_{bij} = \begin{cases} 1 & (i,j) \in E \text{且} b \in D \\ 0 & \text{其他} \end{cases} \tag{25}$$

则 x_{bij} 与决策变量 x_{ij} 关系为

$$x_{bij} \leqslant x_{ij} \quad \forall b \in D, \forall (i,\ j) \in E \tag{26}$$

进场航迹的连续性约束可表达为

$$\sum_{i:(i,j)\in E} x_{bij} \times y_{pit} = y_{pj(t+w)}$$
$$\forall b \in D, \forall p \in A_b, \forall j \notin D, \forall t \in \{t_1, t_2, \cdots, t_{|T|-1}\} \tag{27}$$

因为 x_{bij} 和 y_{pit} 都是决策变量, 2 个不同的决策变量相乘产生了非线性约束。为便于求解, 对其进行线性化处理。定义新 0-1 整数辅助决策变量 z_{pbijt} 作为 x_{bij} 和 y_{pit} 的乘积, 且该决策变量满足:

$$z_{pbijt} \leqslant x_{bij}$$
$$\forall p \in A_b, \forall i,\ j \in V, \forall b \in D, \forall t \in \{t_1, t_2, \cdots, t_{|T|-1}\} \tag{28}$$

$$z_{pbijt} \leqslant y_{pit}$$
$$\forall p \in A_b, \forall i,\ j \in V, \forall b \in D, \forall t \in \{t_1, t_2, \cdots, t_{|T|-1}\} \tag{29}$$

$$z_{pbijt} \geqslant x_{bij} - (1 - y_{pit})$$
$$\forall p \in A_b, \forall i,\ j \in V, \forall b \in D, \forall t \in \{t_1, t_2, \cdots, t_{|T|-1}\} \tag{30}$$

则式 (27) 可改写为线性表达式, 如下:

$$\sum_{(i,j)\in E} z_{pbijt} - y_{pj(t+w)} = 0$$
$$\forall b \in D, \forall p \in A_b, \forall j \notin D, \forall t \in \{t_1, t_2, \cdots, t_{|T|-1}\} \tag{31}$$

3　实例验证与分析

目前, 天津滨海国际机场(以下简称天津机场)采用双跑道隔离平行运行模式, 西跑道 (16R/34L) 用于离场, 东跑道 (16L/34R) 用于进场。天津机场距离北京首都和北京大兴机场较近, 多个终端区部分入口和进离场航段共用, 空域结构复杂, 因此, 对该区域的进场程序和航班调度进行优化, 具有重要意义。

测试以跑道 34R 为例, 建立终端区栅格图。使用 Python 语言调用优化求解器 Gurobi 9.5.2, 计算本文混合整数规划模型。参考天津机场历史航班信息, 选取不同进场交通密度场景, 对不同时间段的 1 h 内的进场航班, 基于进入终端区的不同时间窗, 分配入口时刻, 规划从终端区入口至最终融合点的柔性进场航迹, 并实现进场序列。

3.1 终端区栅格化与测试数据

天津机场34R跑道进场程序如图9所示，有3个终端区入口与4条进场程序，进场程序在中间进近定位点 (intermediate fix, IF) 完成融合后，截获下滑道完成着陆。4个进场程序的从终端区入口至IF点的航迹长度分别为140, 65, 146, 170 km。

保留终端区的入口和最终融合点，参考天津终端区运行环境和空中交通管制规定，在水平方向上将终端区栅格化。用边长为10 km的正方形覆盖该空域（根据航空器的进场速度，将安全间隔约束嵌入到离散时空网络中，简化整数规划模型），如图9所示，AVBOX、DUMAP和TJ867为3个终端区入口，其栅格坐标分别为 (0, 1), (14, 0) 和 (7, 9)；最终融合点 (TJ827) 的栅格坐标为 (9, 4)。

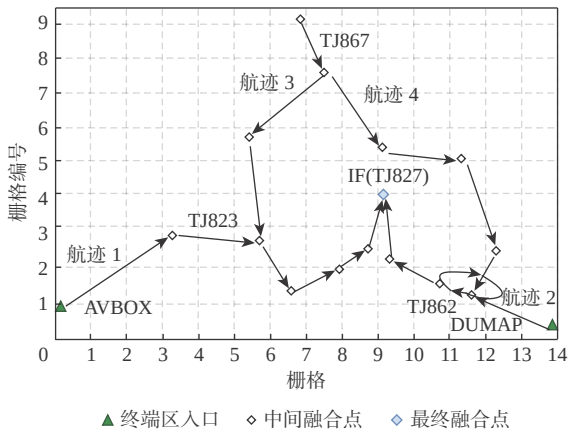


图9 栅格化的进场程序

Fig. 9 Rasterized arrival procedure

为便于计算，设置以下规定。

1) 安全间隔。由天津机场日常运行可知，进场航班由重型和中型飞机组成。根据民航规章，在进场阶段，前后重型与中型机之间的尾流间隔不少于2 min[27]。

2) 进场速度。参考航班进场过程中的航行速度，设置航班经过任意航迹段时间为2 min。

根据天津机场2019年的运行数据，非繁忙时段进场航班数量为10～20架次，繁忙时段进场航班数量为29～34架次。本文将小时进场航班量分为3种不同场景，低于15架次（含）称为低交通密度；小时进场航班量高于29架次（含）称为高交通密度；小时进场航班量为 (16, 28] 架次称为中等交通密度。针对不同交通密度，采集天津机场低、中、高3个时间段的进场航班量数据，在原进场航班时刻表基础上，通过增减航班量，或改变进入终端区不同入口的航班量，或者两者之间的组合，选择和设置几组代表性的航班架次，并手动设置入口时间窗，获得测试数据，如表1所示。其中“三入口航班量”表示分别从AVBOX-DUMAP-TJ867进入的航班量。

表1 不同测试场景的案例数据

Table 1 Case data in different test scenarios

测试案例	密度	航班量	时间窗/min	三入口航班量
L-1	低	13	0	4-5-4
L-2	低	13	1	4-5-4
M-1	中	22	0	8-8-6
M-2	中	22	1	8-8-6
M-3	中	22	2	8-8-6
H-1	高	32	2	10-10-12
H-2	高	32	3	10-10-12
H-3	高	32	4	10-10-12

3.2 航班飞行总长度分析

图10为不同场景的航班飞行总长度，表示在某时段，本文模型计算的航迹下的进场航班飞行总长度与STAR下飞行的总长度对比，其中，STAR(L)、STAR(M)、STAR(H)分别表示低交通密度、中交通密度、高交通密度。

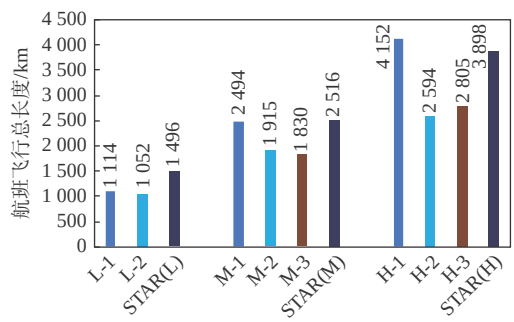


图10 不同场景的航班飞行总长度

Fig. 10 Total flight length in different scenarios

可见，本文模型明显减少某时段的航班在终端区飞行总距离，平均减少534.75 km，最大减少1 312 km。在低交通密度下，不设置时间窗 (L-1) 和1 min的时间窗 (L-2) 的总飞行长度分别减少382 km和444 km；在中交通密度下，设置入口时间窗可以有效优化航迹，1 min(M-2) 和2 min(M-3) 的时间窗下的总飞行长度分别减少601 km和686 km；在高交通密度下，H-1案例因为入口时间窗较小，所有航班在3条航线运行，反而需要更长的距离。若设置合理的入口时间窗，则飞行长度优化明显，3 min(H-2) 和4 min (H-3) 的时间窗的飞行总长度分别减少1 304 km和1 093 km。

综上可知，本文提出的进场模式可以明显地优化进场飞行总长度，但当交通流量较大时，需要设置合理的入口时间窗。

3.3 柔性进场航迹长度

为测试柔性进场航迹的效果，分析柔性进场航迹的长度特点，在低交通密度条件下，选取13架进

场航班的场景, 进场时间窗为 0 和 1 min, 进场航迹如图 11 所示。中等交通密度条件下, 选取 22 架进场航班的场景, 进场时间窗为 1 min 和 2 min 进场航迹如图 12 所示。高交通密度条件下, 选取 32 架进场航班的场景, 进场时间窗为 2, 3, 4 min, 进场航迹如图 13 所示。

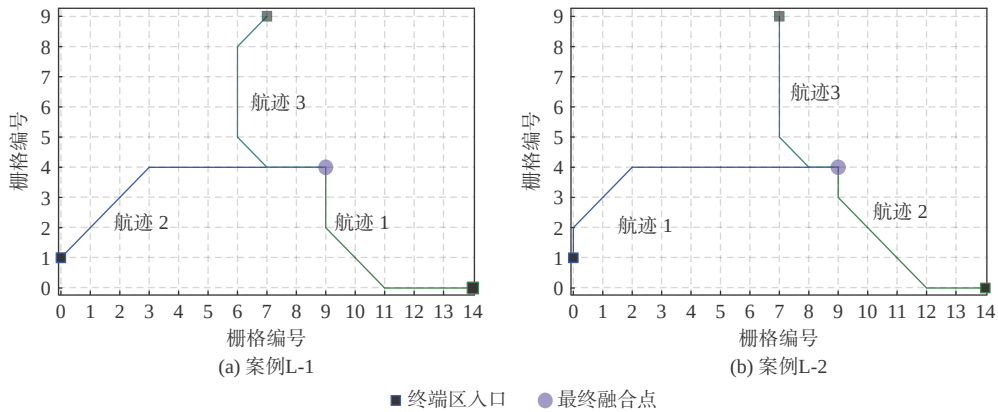


图 11　13 架航班场景的进场航迹

Fig. 11　Arrival trajectory of 13 aircraft scenario

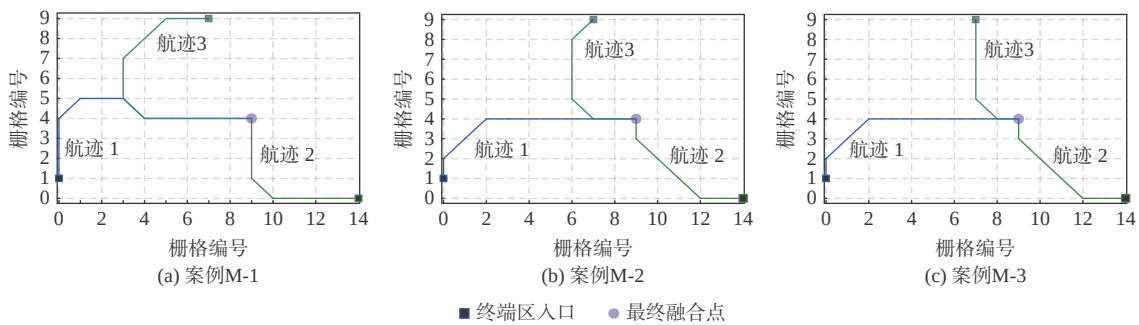


图 12　22 架航班场景的进场航迹

Fig. 12　Arrival trajectory of 22 aircraft scenario

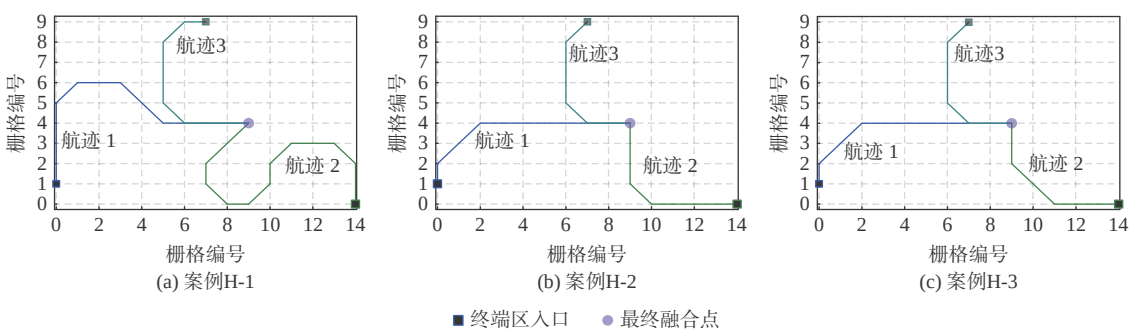


图 13　32 架航班场景的进场航迹

Fig. 13　Arrival trajectory of 32 aircraft scenario

与 STAR 相比, 航迹长度有明显变化。在以上场景中, 本文模型优化后的 3 条总航迹最短为 245 km, 最长为 395 km, 平均为 273 km, 与 STAR 相比平均缩短 127.375 km。航迹 1(AVBOX 入口到 IF) 与航迹 3(TJ867 入口左侧到 IF) 的长度明显减少, 与 STAR 相比分别平均缩短 25 km 和 61.5 km; 航迹 2(DUMAP 入口到 IF) 的长度有所增加, 平均增加长度为 22.875 km。

表 2 详细给出了每个场景中规划的进场航迹长度相对原进场程序长度的改变量, 其中, 航迹 4(TJ867 入口右侧到 IF) 在中等密度与低密度场景时规划的航迹中不存在, "-" 表示减少的航迹长度, "+" 表示增加的航迹长度。

由此可见, 在合理的时间窗口下, 柔性进场航迹可明显缩短航迹长度, 进而缩短飞行时间。

3.4　入口时间窗对航迹长度的影响

对比不同交通密度情况下的航迹改变情况, 如图 14 所示。在低交通密度, 本文模型相对于原 STAR 的进场航迹长度大幅度减少。在中等交通密度, 案例 M-1 入口时间窗设置为 0 min, 进场航迹长

表 2 优化航迹相对原进场程序的长度变化

Table 2 Changes in the length of arrival trajectory

案例	航迹1长度/km	航迹2长度/km	航迹3长度/km	航迹4长度/km	总长度/km
L-1	−37	+14	−67		−90
L-2	−31	+9	−82		−104
M-1	−11	+19	−13		−5
M-2	−31	+9	−67		−89
M-3	−31	+9	−82		−104
H-1	+3	+90	−47	−170	−124
H-2	−31	+19	−67	−170	−249
H-3	−31	+14	−67	−170	−254

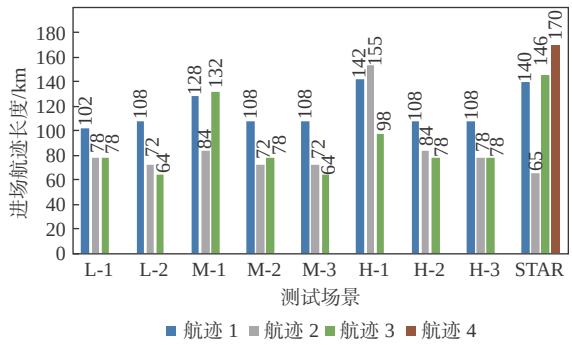


图 14 不同场景下的进场航迹长度

Fig. 14 Length of arrival trajectory in different scenarios

度无明显变化，而案例 M-2 和 M-3 设置入口时间窗后，进场航迹的长度明显减少。在高交通密度，由于终端区容量的限制，若设置较小的时间窗，需要更大空间对航班进行调度，进场航迹的长度明显大于 STAR，如案例 H-1 的图 13(a) 所示，飞行总长度为 395 km。如果合理设计入口时间窗，适当调配航班进入终端的时刻，航迹就会明显改善。如图 13(b) 和图 13(c) 所示，案例 H-2 和案例 H-3 的飞行总长度分别为 271 km 和 265 km，在案例 H-2 中比 STAR 缩短 80 km 和 104 km，在案例 H-3 中比 STAR 缩短 86 km 和 110 km。可见入口时间窗的调整对柔性航迹优化有一定的影响。

在低交通密度，由于航班量小，无需通过距离调整航班的冲突，时间窗口对进场航迹影响较小，如图 15(a) 所示。

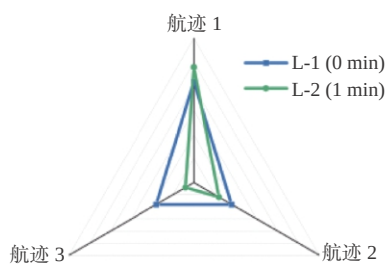
在中等交通密度，终端区内需合理调整航班冲突。若设置一定的入口时间窗，可以通过入口时间的偏移降低一定的冲突调配压力，进而较大程度上减少进场航迹长度。如图 15(b) 所示，入口时间窗从 0 min 变为 1 min，航迹 2 和航迹 3 明显缩短。

在高交通密度，终端区将存在较多航班冲突，无入口时间窗，甚至较小时间窗，都很难起到调配安全间隔的作用，故需要的较长的航迹长度来避免冲突。只有设置足够大的入口时间窗口，才能进一步缩短航班进场航迹，如图 15(c) 所示。

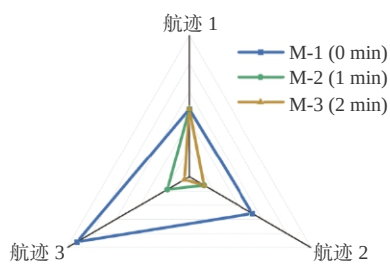
此外，并不是入口时间窗越大越好。如图 15(b) 中，时间窗从 1 min 变为 2 min，航迹 1 和航迹 2 都无变化，只有航迹 3 缩短了 14 km，相比时间窗从 0 min 变为 1 min，收益小了很多；同样如图 15(c) 中，时间窗从 3 min 增加到 4 min 时，航迹 1 和航迹 3 没有变化，航迹 2 缩短了 8 km，相比时间窗从 2 min 改为 3 min，收益太小。可见入口时间窗设置对柔性航迹长度的影响有一个最佳值。入口时间窗的优化将在后续工作中开展。

3.5 航班进场时间

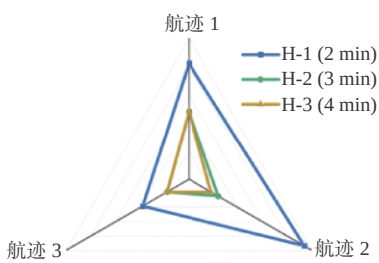
在航班进场时间方面，通过表 1 的 8 个案例可得，本文模型优化后的平均进场时间为 17.5 min，最大值为 26 min，最小值为 12 min，与 STAR 相比，平



(a) 13架航空器场景的进场航迹长度



(b) 22架航空器场景的进场航迹长度



(c) 32架航空器场景的进场航迹长度

图 15 不同时间窗的进场航迹长度

Fig. 15 Length of arrival trajectory based on different time window

均进近时间缩短 6.24 min, 平均提前 5.9 min 到达最终融合点。

图 16 为各案例下每架航班到最终融合点的时间与 STAR 的对比, 其中, 绿色点与蓝色点分别表示 STAR 与本文模型计算的到达最终融合点时刻, 黑色线表示二者相差的时间。由图 16 可知, 除案

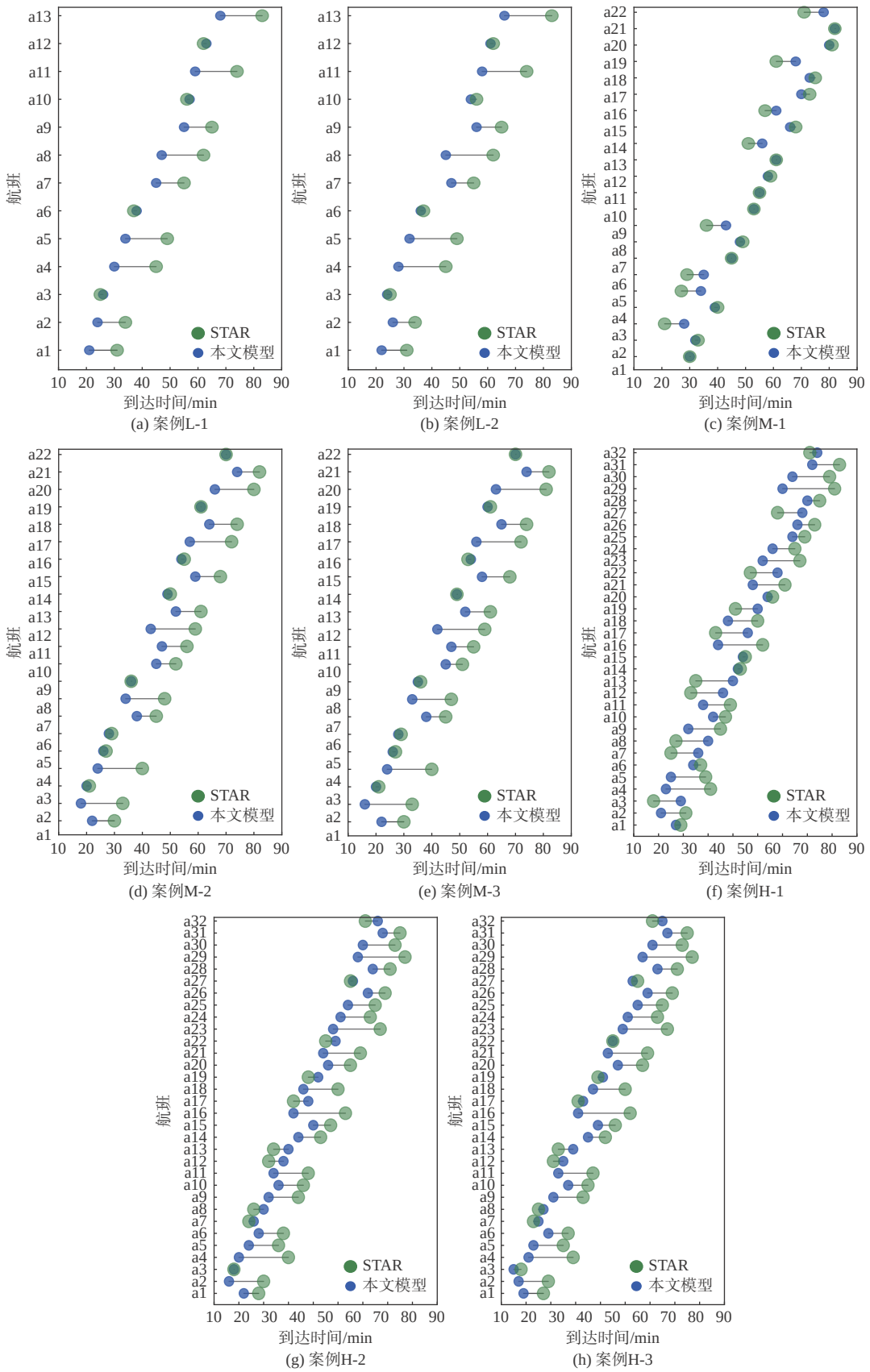


图 16　到达最终融合点的时间

Fig. 16　Time to reach final merge point

例 M-1(图 16(c))和 H-1(图 16(f))由于没用设置时间窗口或时间窗口较小,进场时间缩短不明显外,航班基本上可以较早地到达最终融合点,在整体上缩短了进场时间。

3.6 总运行时间

表 3 为 8 个案例优化航迹的最后一架航班到达最终融合点的时刻。该时刻表达了该案例的总运行时间。因为本文模型的开始时刻是基于给定时间段航班执行 STAR 入口时刻,时间的缩短仅基于航迹的变化,因此,与 STAR 相比,在一定程度上有所缩短。该缩短时间可用于后续时间段的柔性航迹规划和调整。

表 3 最后一架航班到达最终融合点的时间

Table 3 Time when the last flight arrived at the final merge point min

测试场景	本文模型	STAR
L-1	66	83
L-2	66	83
M-1	78	82
M-2	74	82
M-3	74	82
H-1	84	87
H-2	78	87
H-3	77	87

3.7 航班量对航迹的影响

为进一步验证本文模型的可行性,分析进场航迹与航班数量的关系,选取典型的进场时段与合理的时间窗口,做进一步测试,测试数据如表 4 所示。

表 4 不同测试案例的数据

Table 4 Data for different test cases

测试	密度	航班量	时间窗/min	三入口航班量
L-3	低	14	1	4-6-4
L-4	低	15	1	4-7-4
L-5	低	13	1	6-4-3
L-6	低	13	1	4-4-5
M-4	中	21	1	8-8-5
M-5	中	23	1	9-9-5
M-6	中	22	1	9-7-6
M-7	中	22	1	8-9-5
H-4	高	31	3	10-10-11
H-5	高	33	3	10-11-12
H-6	高	32	3	11-10-11
H-7	高	32	3	11-11-10

在低交通密度、中交通密度下,选取 1 min 的时间窗,航迹分别如图 17 和图 18 所示。在高交通密度下,选取 3 min 的时间窗,航迹如图 19 所示。

根据计算结果可知,相同等级的交通密度,在设置合理的进场时间窗口的情况下,当航班数量接近时或改变一定架次的航班进场入口时,进场航迹

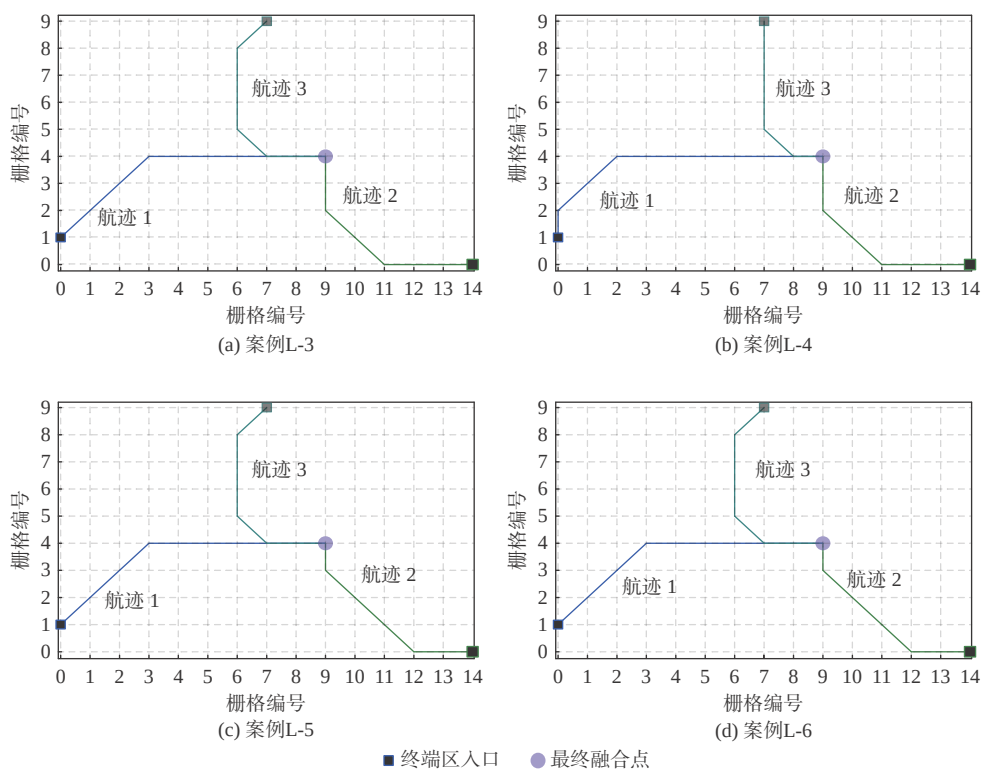


图 17 低交通密度下不同测试的航迹

Fig. 17 Trajectories for different tests in low traffic density

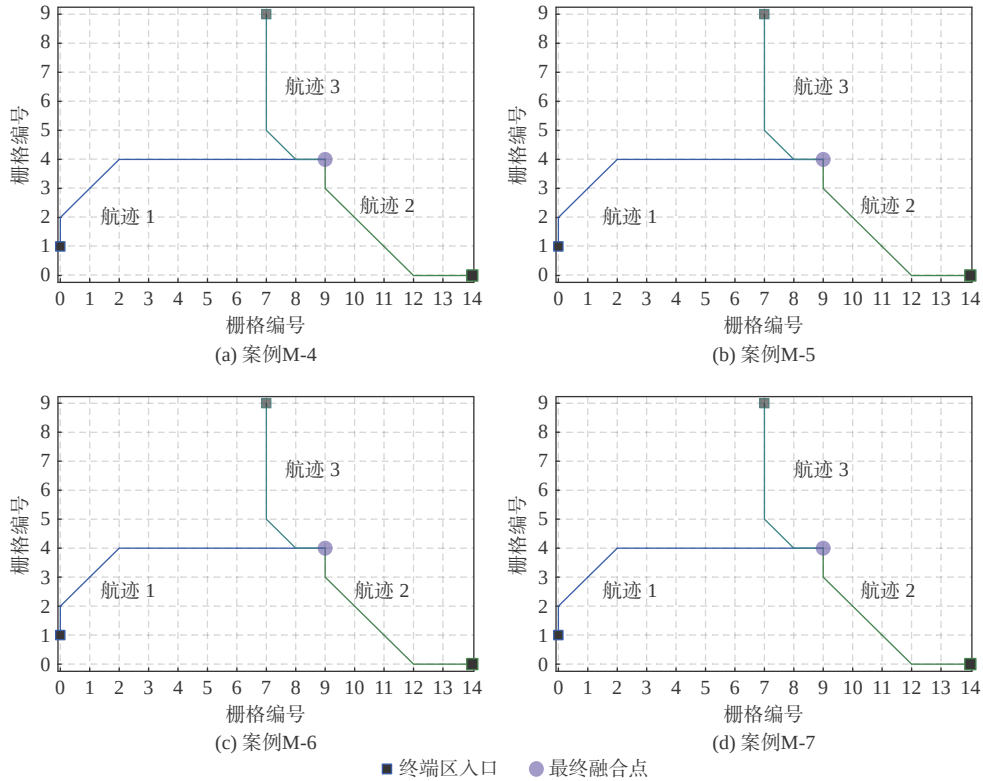


图 18　中交通密度下不同测试的航迹

Fig. 18　Trajectories for different tests in medium traffic density

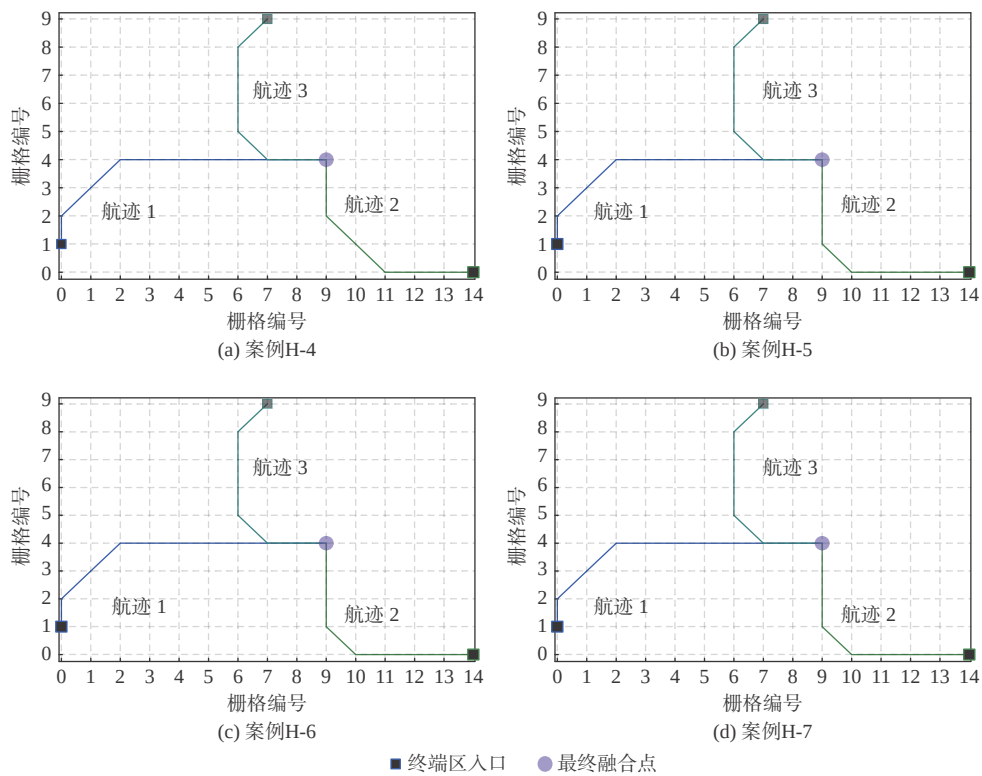


图 19　高交通密度下不同测试的航迹

Fig. 19　Trajectories for different tests in high traffic density

的长度与分布位置无明显变化。可见, 柔性航迹对航班量的小幅波动具有一定的鲁棒性。这可以降低少量航班变动带来大幅度的航迹调整, 造成管制指挥和监控的困难, 有利于管制员的调度和飞行员的操控。

3.8　不同目标函数权重系数对航迹的影响

选取案例 M-2 的航班数据, 设置不同的权重系数, 分析对进场航迹长度与飞行总长度的影响。权

重系数 a 分别设置为 0.1、0.3、0.5、0.7、0.9。

如图 20 所示，对目标函数归一化处理后，设置不同权重系数的求解结果。可见，随着权重系数的增大，进场航迹长度不断缩短，但飞行总长度不断增大，多目标帕累托前沿特征明显。本文模型设置权重系数 a 为 0.1，求解结果可以实现优先考虑最小化飞行总长度的目标。

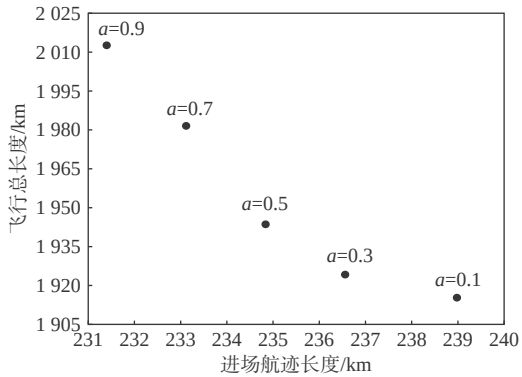


图 20 不同权重系数下的结果

Fig. 20 Results under different weight coefficients

4 结 论

1）针对当前终端区固定进场程序存在的航班冲突多、调配困难、等待时间长、碳排量大等问题，提出了基于柔性进场航迹优化的解决方案。设计了终端区时空离散网络栅格规划方法，建立柔性进场航迹规划的混合整数规划模型。

2）针对天津机场航班时刻和终端区，对本文模型进行了验证，实验表明，本文模型获得的进场航迹长度明显缩短，优化后的航迹长度与原 STAR 相比平均缩短 127.375 km；航班飞行总长度平均减少 534.75 km，最大减少 1 312 km；平均进近时间缩短 6.24 min。

3）入口时间窗对柔性航迹优化有一定的影响。低交通密度，入口时间窗对进场航迹影响较小；中等交通密度，入口时间窗为 1 min 时航迹较短，大于 1 min 后航迹无明显缩短趋势；高交通密度，入口时间窗为 3 min 时航迹较短，大于 3 min 后航迹缩短趋势较小。

4）柔性航迹对航班量的小幅波动具有一定的鲁棒性。相同等级的交通密度，在设置合理的进场时间窗口的情况下，当航班数量接近或改变少量航班架次时，进场航迹的长度与分布位置无明显变化。

目前本文的研究是基于航班的单一进场速度和安全间隔进行的。柔性进近航迹的进一步研究，需要讨论多航班多种变速进场和安全间隔条件下的规划，以及最优入口时间窗等问题。本文规划的航迹是基于完全的 CDO，因此，规划的是二维航迹，如果在下降过程中存在改变下降梯度问题，则需要进一步研究三维航迹。此外，由于主要考虑二维航迹规划问题，没有考虑进近过程中障碍物的避障或越障问题，因此，本文模型目前可用于障碍物较小或无障碍物的地区。

参考文献（References）

[1] BENTRUP L, HONFFMANN M. Free routing airspace in Europe[C]//Proceedings of the International Conference on Research in Air Transportation. Drexel: ICRAT, 2016.

[2] Eurocontrol. Point merge integration of arrival flows enabling extensive RNAV application and continuous descent - operational services and environment definition (V2.0) [EB/OL]. (2010-07-19) [2023-04-20]. https://www.eurocontrol.int/concept/point-merge.

[3] 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).

[4] 赵向领, 雷碧玉. 点汇聚进近飞行程序近地风险评估方法研究[J]. 中国安全科学学报, 2017, 27(3): 157-162.
ZHAO X L, LEI B Y. Research on method for assessing risk in ground approach under point merging procedure[J]. China Safety Science Journal, 2017, 27(3): 157-162(in Chinese).

[5] CECEN R K. Multi-objective TMA management optimization using the point merge system[J]. Aircraft Engineering and Aerospace Technology, 2021, 93(1): 15-24.

[6] LEE S, HONG Y, KIM Y. Optimal scheduling algorithm in point merge system including holding pattern based on mixed-integer linear programming[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2020, 234(10): 1638-1647.

[7] 戴喜妹. 基于分派规则的航班排序与调度研究[D]. 南京: 南京航空航天大学, 2019: 4-50.
DAI X M. Research on flight sequencing and scheduling based on assignment rules[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2019: 4-50(in Chinese) .

[8] 刘继新, 江灏, 董欣放, 等. 基于空中交通密度的进场航班动态协同排序方法[J]. 航空学报, 2020, 41(7): 279-294.
LIU J X, JIANG H, DONG X F, et al. Dynamic collaborative sequencing method for arrival flights based on air traffic density[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(7): 279-294(in Chinese).

[9] 乐美龙, 吴宪晟, 胡钰明. 基于滚动时域控制的多路径进场航班排序优化[J]. 北京航空航天大学学报, 2023, 49(12): 3222-3229.
LE M L, WU X S, HU Y M. Arrival flights optimal sequencing based on path selection and rolling horizon control[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(12): 3222-3229(in Chinese).

[10] 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): 464-476.

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): 464-476(in Chinese).

[11] SAMÀ M, D'ARIANO A, CORMAN F, et al. Metaheuristics for efficient aircraft scheduling and re-routing at busy terminal control areas[J]. Transportation Research Part C: Emerging Technologies, 2017, 80: 485-511.

[12] LIANG M, DELAHAYE D, MARÉCHAL P. Integrated sequencing and merging aircraft to parallel runways with automated conflict resolution and advanced avionics capabilities[J]. Transportation Research Part C: Emerging Technologies, 2017, 85: 268-291.

[13] HONG Y, CHOI B, KIM Y. Two-stage stochastic programming based on particle swarm optimization for aircraft sequencing and scheduling[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(4): 1365-1377.

[14] SALEHIPOUR A, MODARRES M, MOSLEMI NAENI L. An efficient hybrid meta-heuristic for aircraft landing problem[J]. Computers & Operations Research, 2013, 40(1): 207-213.

[15] 张洪海, 汤一文, 许炎. TBO 模式下终端区进场交通流优化模型与仿真分析[J]. 航空学报, 2020, 41(7): 319-332.
ZHANG H H, TANG Y W, XU Y. Optimizing arrival traffic flow in airport terminal airspace under trajectory based operations[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(7): 319-332(in Chinese).

[16] 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): 200-216.
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): 200-216(in Chinese).

[17] 陈昇, 周隽, 胡小兵, 等. 基于混合模拟退火算法的机场进场程序优化[J]. 计算机应用, 2022, 42(2): 606-615.
CHEN S, ZHOU J, HU X B, et al. Optimization of airport arrival procedures based on hybrid simulated annealing algorithm[J]. Journal of Computer Applications, 2022, 42(2): 606-615(in Chinese).

[18] CECEN R K. A stochastic programming model for the aircraft sequencing and scheduling problem considering flight duration uncertainties[J]. The Aeronautical Journal, 2022, 126(1304): 1736-1751.

[19] CECEN R K. A path stretching model for effective terminal airspace management[J]. International Journal of Aeronautical and Space Sciences, 2022, 23(5): 1043-1052.

[20] LEE H, BALAKRISHNAN H. A study of tradeoffs in scheduling terminal-area operations[J]. Proceedings of the IEEE, 2008, 96(12): 2081-2095.

[21] ZHANG H H, YANG X L, FAN W, et al. Guidance control strategy for air traffic flow in terminal areas[J]. Advances in Mechanical Engineering, 2016, 8(10): 1-12.

[22] 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.

[23] CHIDA H, ZUNIGA C, DELAHAYE D. Topology design for integrating and sequencing flows in terminal maneuvering area[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2016, 230(9): 1705-1720.

[24] 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.

[25] SÁEZ R, PRATS X, POLISHCHUK T, et al. Automation for separation with continuous descent operations: dynamic aircraft arrival routes[J]. Journal of Air Transportation, 2020, 28(4): 144-154.

[26] LIU J M, GUO Z, YU B. Optimising Gate assignment and taxiway path in a discrete time–space network: integrated model and state analysis[J]. Transportmetrica B: Transport Dynamics, 2023, 11(1): 1-23.

[27] 中国民用航空局. 民用航空空中交通管理规则: CCAR-93-R5[S]. 北京: 中国民用航空局, 2018: 86-91.
Civil Aviation Administration of China. Civil aviation air traffic management rules: CCAR-93-R5[S]. Beijing: Civil Aviation Administration of China, 2018: 86-91(in Chinese).

Flexible arrival trajectory optimization at terminal airspace based on mixed integer programming

ZHAO Xiangling*, ZHOU Zhiliang

(College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Problems with flight conflicts, difficult deployment, lengthy wait times, and high carbon emissions are becoming more frequent as a result of the continuous growth in air traffic, the saturation of available airspace, and the rigidity of the conventional fixed approach procedure used in the terminal area. A solution based on flexible approach trajectory optimization is developed and detailed, taking into account the operating processes of multiple flights such as descent, approach, sequencing, and margining in the terminal area. A spatiotemporal discrete network grid of flight approaches in the terminal area is created based on the characteristics of the terminal and aircraft operation. A mixed integer programming model is developed with the objective of minimizing the trajectory and overall flight length, as well as being constrained by factors including grid regulations, airplane turning characteristics, flying conflicts, safety separation, and trajectory continuity. The model is solved and validated using the commercial solver Gurobi, and multiple sets of typical experimental scenarios are put up depending on different traffic density levels and the entrance time frame, using the Tianjin Airport terminal area as an example. The results show that the flexible approach trajectory length is decreased by an average of 127.375 km and the total flight length is reduced by 534.75 km compared with the standard instrument approach procedure (STAR), with an average reduction of 6.24 min in flight arrival times. Additionally, with the same traffic density level and a fair entrance time window setting, the flexible approach trajectory is resilient enough to handle slight variations in the number of flights.

Keywords: air traffic management; arrival flights; four-dimensional trajectory; arrival sequencing; flexible airspace; mixed integer programming

Received: 2023-05-11; **Accepted**: 2023-10-13; **Published Online**: 2023-11-17 17:10

URL: link.cnki.net/urlid/11.2625.V.20231117.1428.003

Foundation items: National Natural Science Foundation of China (52272356); Graduate Research and Innovation Funding Project of Civil Aviation University of China (2022YJS087)

* **Corresponding author**. E-mail: zxl-llx@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0241

基于用户调度的多无人机辅助通信系统资源优化

唐菁敏[1]，黄嘉琪[1,*]，王炳文[1]，宋耀莲[1]，虞贵财[2]

(1. 昆明理工大学 信息工程与自动化学院，昆明 650500；2. 宜春学院 物理科学与工程技术学院，宜春 336000)

摘 要：为提高多用户移动通信下行无线传输系统的传输速率，提出一种基于用户调度和轨迹优化的多无人机(UAV)辅助通信系统资源优化算法。所提算法在满足用户调度、无人机总能耗和用户服务质量要求等约束条件下，以多用户总吞吐量最大化为准则建立优化问题。为解决该非凸问题，通过块坐标下降(BCD)法将原非凸问题分解成3个易于处理的非凸子问题，并通过引入松弛变量、一阶泰勒表达式、连续凸近似(SCA)等方法对子问题转换求解后交替迭代优化，得出原非凸问题的近似次优解。仿真结果表明：所提算法能有效地提高系统总吞吐量，并且在单、多无人机通信系统下均具有良好的收敛性。

关 键 词：无人机通信；资源分配；用户调度；波束成形；飞行速度

中图分类号：TN929.5

文献标志码：A **文章编号**：1001-5965(2025)04-1143-09

随着5G和6G网络的快速发展，未来的通信网络有望满足用户无处不在的通信服务需求[1-2]。但是当自然灾害发生时，传统通信网络一旦被破坏，很难在短时间内恢复通信连接。无人机(unmanned aerial vehicle, UAV)具有高机动性和灵活性等优势，其作为空中基站能有效快速地建立应急通信[3]，并且可以为用户提供成本低、效益较高的无线通信，但是UAV轨迹规划和资源分配等问题限制了UAV基站长时间续航和高数据传输速率的实现[4]。

目前，众多专家和学者对UAV通信的轨迹和资源优化展开了一系列的研究。文献[5-6]均以系统吞吐量最大化为目标，给出了UAV最佳悬停位置部署方案。文献[7]通过固定轨迹设计开发其机动性，提高无线通信系统的性能。文献[5-7]证明了合理的位置部署和轨迹设计是提高系统性能的关键因素之一，但是均未充分考虑UAV可移动性。而文献[8]设计优化UAV轨迹来最大化用户的最小吞吐量，利用优化UAV的轨迹所带来的信道状态的改善来显著提高系统吞吐量。由于UAV通信网络服务质量会受限于UAV能量消耗，所以通过合理的资源分配能进一步提升系统网络性能[9]。文献[10]考虑在满足用户最低速率传输要求下，通过联合优化UAV飞行轨迹和功率分配，实现系统的吞吐量最大化，但仅考虑无人机的通信能耗。而UAV辅助通信系统的总能耗通常主要由通信能耗和飞行能耗两部分组成，且通常飞行能耗占比更大。文献[11]研究了UAV飞行能耗的优化问题，提出了联合优化UAV的轨迹、飞行速度和加速度来提高系统的吞吐量，但是仅考虑了单天线UAV辅助通信。

随着毫米波、太赫兹等高频段通信技术的发展，天线尺寸进一步减小，在UAV上可配备多根天线，利用多天线技术可提供更多分集增益。文献[12]通过联合优化多天线UAV轨迹、发射功率和天线阵的方向，来实现最大化所有用户的最小吞吐量。文献[13]研究了多天线UAV辅助通信的波束成形和资源分配，提升了用户总传输速率。另外，随着用户数量不断增多并且分布越来越广，单个无人机

收稿日期：2023-05-11；**录用日期**：2023-07-28；**网络出版时间**：2023-09-07 16:13
网络出版地址：link.cnki.net/urlid/11.2625.V.20230907.1135.005
基金项目：国家自然科学基金(62261056)
*通信作者. E-mail：1540600368@qq.com

引用格式：唐菁敏，黄嘉琪，王炳文，等. 基于用户调度的多无人机辅助通信系统资源优化[J]. 北京航空航天大学学报，2025，51(4)：1143-1151. TANG J M，HUANG J Q，WANG B W，et al. Resource optimization of multi UAV assisted communication system based on user scheduling[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1143-1151 (in Chinese).

基站已经远远无法为所有用户提供更好的通信环境，甚至无法满足用户最基本的需求。而针对覆盖面积比较大的目标区域，多无人机可以进一步节省系统的飞行能耗，提高系统整体的通信质量，有效地提高系统能量利用率[7,14]。

目前大部分无人机基站的研究都是仅考虑了单个基站通信，而在实际场景中需要多个无人机基站为多个用户提供通信服务；同时，多天线无人机辅助通信的研究未能全面考虑用户调度、波束成形、无人机轨迹和飞行速度等因素优化对系统性能的影响。因此，本文在多UAV总能耗受限的情况下，研究了用户调度、UAV的轨迹和速度优化、飞行和功率分配等问题。

本文主要的研究工作如下：

1）建立了多无人机辅助地面用户通信的下行无线传输系统，结合固定翼无人机能耗模型，在用户调度和用户最低传输速率等约束条件下，设计一个联合优化用户调度、波束成形、无人机轨迹和飞行速度的资源优化算法来提高系统的总吞吐量的问题。

2）设计的问题是非凸的，难以直接求解。本文提出通过块坐标下降 (block coodinate descent, BCD) 法将非凸优化问题分解成3个子问题，引入松弛变量、一阶泰勒表达式、连续凸近似方法将3个非凸子问题转换成凸问题求解，最后对3个子问题交替迭代优化。

3）通过仿真对比实验，证明本文算法在满足用户的通信质量的前提下，能有效提高系统的总吞吐量并在有限次数迭代后收敛。

1 系统模型

考虑了固定翼无人机和用户组成的下行无线传输网络，如图1所示，将多架含N_t根天线的UAV作为基站，为k个单天线地面用户提供数据传输服务，其中，无人机$m\in\{1,2,\cdots,M\}$，$M\geqslant 1$和用户$k\in\{1,2,\cdots,K\}$，$K\geqslant 1$，并假设所有地面用户的位置是固定已知的。由于无人机是移动的，为便于描述无人机和用户的位置，将总服务时间T等分为N个时隙，每个时隙长度为$\delta_t=T/N$，假设时隙δ_t足够小以保证在每个时隙UAV和用户位置保持不变。每架UAV在飞行过程中高度固定为H，因此，UAVm在第n个时隙坐标可以用$\boldsymbol{q}_m^n=(x_m(n),y_m(n),H)$来表示，$n=1,2,\cdots,N$。此外，第$k$个用户的坐标位置表示为$\boldsymbol{g}_k=(x_k,y_k,0)$。考虑到UAV在飞行过程中一般存在飞行能力的限制，多架UAV在服务时间内的轨迹应满足如下约束：

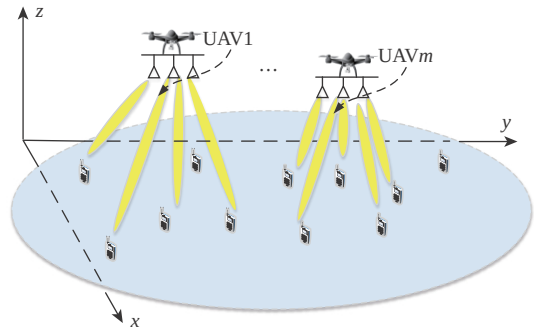


图1 多无人机通信系统

Fig. 1 Multiple UAV communication system

$$\left\|\boldsymbol{q}_m^{n+1}-\boldsymbol{q}_m^n\right\|^2\leqslant S_{\max}^2 \tag{1}$$

$$\left\|\boldsymbol{q}_m^n-\boldsymbol{q}_j^n\right\|^2\geqslant d_{\max}^2\quad\forall m,n \tag{2}$$

$$\boldsymbol{q}_m^1=\boldsymbol{q}_{m,0},\boldsymbol{q}_m^N=\boldsymbol{q}_{m,F}\quad\forall m \tag{3}$$

其中：$S_{\max}=v_{\max}\delta_t$为无人机在每个时间间隙中可以飞行的最大水平距离，$v_{\max}$为最大飞行速度；$d_{\max}$为任何2架无人机之间最小间距；$\boldsymbol{q}_{m,0}$和$\boldsymbol{q}_{m,F}$分别为UAV$m$的初始位置和终止位置。

假设UAV采用均匀线性阵阵列，则在第n个时隙，UAVm与用户k的天线阵列响应为

$$\boldsymbol{a}_{k,m}^n=[1,\mathrm{e}^{-\mathrm{j}\pi D\sin(\theta_{k,m}^n)},\cdots,\mathrm{e}^{-\mathrm{j}\pi D(N_t-1)\sin(\theta_{k,m}^n)}] \tag{4}$$

式中：$D=d/\lambda_c$；d为均匀阵列天线之间的距离；λ_c为载波中心频率的波长；$\theta_{k,m}^n$为用户k与阵列法向量之间的夹角：

$$\sin(\theta_{k,m}^n)=\frac{x_k-x_m^n}{\sqrt{\left\|\boldsymbol{q}_m^n-\boldsymbol{g}_k\right\|^2}} \tag{5}$$

由于UAV在一定高度与用户通信，未受障碍物的影响，大部分时间都存在视距链路传输路径，因此，下行链路考虑视距信道模型。在第n个时隙UAVm和用户k的信道增益可以表示为

$$\boldsymbol{h}_{k,m}^n=\sqrt{\frac{\beta_0}{\left\|\boldsymbol{q}_m^n-\boldsymbol{g}_k\right\|^2}}\boldsymbol{a}_{k,m}^n \tag{6}$$

式中：β_0为UAVm与用户间距离1 m时的功率增益。

定义二进制变量$s_{k,m}^n$表示UAVm在第n个时隙是否与地面用户建立通信连接。$s_{k,m}^n=1$表示UAVm在第n个时隙为用户k提供通信服务；$s_{k,m}^n=0$表示UAVm在第n个时隙不为用户k提供通信服务。为防止多无人机碰撞事件的发生，假设无人机之间的最小安全距离为$d_{\min}^2$，在每个时隙每个无人机最多只服务于一个用户。于是，用户调度约束如下：

$$\sum_{k=1}^{K}s_{k,m}^n\leqslant 1\quad\forall m,n \tag{7}$$

$$\sum_{m=1}^{M} s_{k,m}^{n} \leqslant 1 \quad \forall k,n \tag{8}$$

$$s_{k,m}^{n} \in \{0,1\} \quad \forall k,m,n \tag{9}$$

因此, 用户k在第n个时隙接收到的信号为

$$y_k^n = s_{k,m}^n \boldsymbol{h}_k^n \boldsymbol{w}_k^n x_k + \sum_{j,j\neq k}^{K} s_{k,m}^n \boldsymbol{h}_k^n \boldsymbol{w}_j^n x_j + n_i \tag{10}$$

式中: n_i服从正态分布$(0,\sigma_z^2)$, 表示加性高斯白噪声; $\boldsymbol{w}_{k,m}^n \in \mathbf{C}^{N_t\times 1}$表示在对用户$k$的波束成形矩阵, 每个时隙用户接收到的通信功率可以通过$\boldsymbol{w}_{k,m}^n$来调节控制。

则在第n个时隙, 用户k接收信号的干扰信噪比为

$$V_{\mathrm{SIR}k}^n = \frac{\|\boldsymbol{h}_{k,m}^n \boldsymbol{w}_{k,m}^n\|^2}{\sum\limits_{j,j\neq k}^{K} \|\boldsymbol{h}_{k,m}^n \boldsymbol{w}_{j,m}^n\|^2 + \sigma_z^2} \tag{11}$$

用户k的吞吐量可以表示为

$$R_k = \sum_{m=1}^{M}\sum_{n=1}^{N} s_{k,m}^n \log_2(1+V_{\mathrm{SIR}k}^n) \tag{12}$$

在整个飞行周期T内, 系统的总吞吐量可以表示为

$$R_{\mathrm{sum}} = \sum_{k=1}^{K} R_k \tag{13}$$

2　能耗模型

在整个飞行周期T内, UAV 的总能耗E_{sum}由通信能耗 E_{t} 和飞行能耗 E_{v} 组成, 可表示为

$$E_{\mathrm{sum}} = E_{\mathrm{t}} + E_{\mathrm{v}} \tag{14}$$

1) 通信能耗。UAV 提供通信传输数据消耗的能量, 可表示为

$$E_{\mathrm{t}} = \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{n=1}^{N} s_{k,m}^n \boldsymbol{w}_{k,m}^{n\mathrm{H}} \boldsymbol{w}_{k,m}^n \tag{15}$$

2) 飞行能耗。用来支持 UAV 飞行消耗的能量。根据文献 [15], 在第n个时隙第m个 UAV 飞行消耗的功率为

$$p_m(v_m^n) = c_1\left(1+\frac{3|v_m^n|^2}{U_{\mathrm{tip}}}\right) + c_2\left(\sqrt{1+\frac{|v_m^n|^4}{4v_m^4}} - \frac{|v_m^n|^2}{2v_m^2}\right)^{1/2} + \frac{1}{2}d_0\rho s A|v_m^n|^3 \tag{16}$$

式中: c_1和c_2为常数, 取决于 UAV 的质量、转子的速度、转子盘面积、叶片角速度和空气密度; U_{tip}为转子的叶尖速度; d_0为机身阻力比; v_m为平均转子速度; ρ为空气密度; s为转子密度; A为转子盘面积; v_m^n为 UAVm在第n个时隙的飞行速度。

因此, 在整个飞行周期T内, 所以多无人机飞行消耗的能量可表示为

$$E_{\mathrm{v}} = \sum_{m=1}^{M}\sum_{n=1}^{N} \delta_t p(v_m^n) \tag{17}$$

3　联合优化问题的建立

为提升整个系统数据传输的总吞吐量, 考虑基于阵列天线 UAV 基站用户调度、波束成形、UAV 轨迹和飞行速度规划联合优化问题, 该优化问题 P_1 可以表示为

$$\begin{cases} P_1: \max\limits_{\boldsymbol{s},\boldsymbol{w},\boldsymbol{q},\boldsymbol{v}} \quad R_{\mathrm{sum}} = \sum\limits_{k=1}^{K} R_k \\ \quad \text{s.t.} \quad C_1: R_k \geqslant \alpha \qquad \forall k \\ \qquad\quad C_2: \sum\limits_{k=1}^{K} s_{k,m}^n \leqslant 1 \qquad \forall m,n \\ \qquad\quad C_3: \sum\limits_{m=1}^{M} s_{k,m}^n \leqslant 1 \qquad \forall k,n \\ \qquad\quad C_4: \ s_{k,m}^n \in \{0,1\} \qquad \forall k,m,n \\ \qquad\quad C_5: E_{\mathrm{t}} + E_{\mathrm{v}} \leqslant p_{\max} T \qquad \forall m,n \\ \qquad\quad C_6: \left\|\boldsymbol{q}_m^{n+1} - \boldsymbol{q}_m^n\right\|^2 \leqslant S_{\max}^2 \qquad \forall m \\ \qquad\quad C_7: \left\|\boldsymbol{q}_m^n - \boldsymbol{q}_j^n\right\|^2 \geqslant d_{\min}^2 \qquad \forall m,n \\ \qquad\quad C_8: \boldsymbol{q}_m^1 = \boldsymbol{q}_{m,0}, \boldsymbol{q}_m^N = \boldsymbol{q}_{m,F} \qquad \forall m \\ \qquad\quad C_9: v_{\min} \leqslant v_{\mathrm{m}}^n \leqslant v_{\max} \qquad \forall m,n \end{cases} \tag{18}$$

式中: $\boldsymbol{s} = \left[s_{k,m}^n, \forall k,n\right]$; $\boldsymbol{w} = \left[\boldsymbol{w}_{k,m}^n, \forall k,n\right]$; $\boldsymbol{q} = \left[\boldsymbol{q}_m^n, \forall n\right]$; $\boldsymbol{v} = \left[v_m^n, \forall n\right]$; C_1 表示至少满足用户k最低的传输速率; C_2、C_3 和 C_4 表示无人机通信的用户调度约束; C_5 表示无人机系统通信能耗和飞行能耗受限于系统总能量; $p_{\max}$为无人机群系统最大功率, 并且 $p_{\max} = \sum\limits_{m=1}^{M} p_{m,\max} T$, $p_{m,\max}$为无人机m的最大发射功率; C_6、C_7 表示无人机的轨迹约束; C_8 表示无人机m的初始位置和终止位置; C_9 表示无人机m最大和最小飞行速度, 由于 P_1 中目标函数和约束条件存在多变量耦合、非凸问题, 难以直接求最优解。

4　UAV 资源分配优化方法

为解决问题 P_1, 提出了一种有效的基于块坐标下降次优迭代算法, 通过固定变量法依次优化各子问题中的变量, 并交替优化迭代直到收敛。该联合优化问题 P_1 可以分解 3 个子问题 P_2、P_3 和 P_4。

子问题 P_2。在固定波束成形、UAV 轨迹和飞行速度条件下, 对用户调度的进行优化。

子问题 P_3。在固定用户调度、UAV 的飞行轨迹和飞行速度条件下, 对波束成形优化。

子问题 P_4。在固定用户调度、波束成形矩阵条件下, 对 UAV 的飞行轨迹和速度优化。

4.1 用户调度优化

固定波束成形$\boldsymbol{w}$和飞行轨迹$\boldsymbol{q}$、速度$\boldsymbol{v}$，则子问题P_2可以表示为

$$\begin{cases} P_2:\max\limits_{\boldsymbol{s}} \quad R_{\text{sum}}=\sum\limits_{k=1}^{K}R_k \\ \text{s.t.} \quad C_1:R_k\geqslant\alpha \qquad \forall k \\ \qquad C_2:\sum\limits_{k=1}^{K}s_{k,m}^{n}\leqslant 1 \qquad \forall m,n \\ \qquad C_3:\sum\limits_{m=1}^{M}s_{k,m}^{n}\leqslant 1 \qquad \forall k,n \\ \qquad C_4:\ s_{k,m}^{n}\in\{0,1\} \qquad \forall k,m,n \end{cases} \tag{19}$$

由于子问题P_2是一个标准的线性规划问题，可以直接利用凸优化工具CVX求解。

4.2 波束成形优化

固定用户调度$\boldsymbol{s}$和飞行轨迹$\boldsymbol{q}$、速度$\boldsymbol{v}$，问题P_1转变为关于变量$\boldsymbol{w}$的子问题P_3，即

$$\begin{cases} P_3:\max\limits_{\boldsymbol{w}} \ R_{\text{sum}}=\sum\limits_{k=1}^{K}R_k \\ \text{s.t.} \quad C_1:R_k\geqslant\alpha \qquad \forall k \\ \qquad C_2:E_{\text{v}}\leqslant\sum\limits_{m=1}^{M}p_{m,\max}T-E_{\text{t}} \qquad \forall m,n \end{cases} \tag{20}$$

由于目标函数和约束条件C_1中的R_k^n都是关于$\boldsymbol{w}_{k,m}^n$非凸的，子问题P_3是非凸的，为解决这一问题，通过利用连续凸近似(successive convex approximation，SCA)的技术[16]来处理。首先，通过一阶泰勒展开式推导R_k的一个近似下界；然后，对该近似下界进行迭代计算逼近原来的非凸约束。

先将R_k重写为2个对数函数的差值，即

$$R_k=\sum_{m=1}^{\text{M}}\sum_{n=1}^{N}s_{k,m}^{n}\cdot \log_2\left(1+\frac{\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{k,m}^{n}\right\|^2}{\sum\limits_{j,j\neq k}^{K}\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n}\right\|^2+\sigma_z^2}\right)=R_{1k,m}-R_{2k,m} \tag{21}$$

其中：

$$\begin{cases} R_{1k,m}=\log_2\left(\sum\limits_{k=1}^{K}\|\boldsymbol{h}_{k,mm}^{n}\boldsymbol{w}_{k,m}^{n}\right\|^2+\sigma_z^2\right) \\ R_{2k,m}=\sum\limits_{j,j\neq k}^{K}\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n}\right\|^2+\sigma_z^2 \end{cases} \tag{22}$$

由于$R_{2k,m}$相对$\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n}\right\|^2$是凹函数，根据在任意点的一阶泰勒展开都是其全局上界，可得

$$R_{2k,m}\leqslant A_{k,m}^{n}\left(\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n}\right\|^2-\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n,l}\|^2\right)+\sum_{j,j\neq k}^{K}\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n,l}\right\|^2+\sigma_z^2=R_{2k,m}^{\text{up}} \tag{23}$$

式中：$\boldsymbol{w}_{j,m}^{n,l}$定义为第$l$次迭代中波束成形向量。

$A_{k,m}^{n}$可表示为

$$A_{k,m}^{n}=\frac{\log_2(\text{e})}{\sum\limits_{j,j\neq k}^{K}\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n,l}\right\|^2+\sigma_z^2} \tag{24}$$

$\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n}\right\|^2$是关于$\boldsymbol{w}_{j,m}^{n}$的凸函数，于是利用凸函数的性质，凸函数在任意点展开式都是其下界。因此，利用$\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n}\right\|^2$的一阶泰勒展开式作为其在给定点$\boldsymbol{w}_{j,m}^{n,l}$处的下界，可得

$$\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n}\right\|^2\geqslant\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n,l}\right\|^2+2R\{\boldsymbol{w}_{j,m}^{n,\text{H}}\boldsymbol{h}_{k,m}^{n\text{H}}\boldsymbol{h}_{k,m}^{n}(\boldsymbol{w}_{j,m}^{n}-\boldsymbol{w}_{j,m}^{n,u})\} \tag{25}$$

式中：R表示一阶泰勒展开。

因此，子问题P_3可以通过近似松弛非凸约束来解决，即在每次迭代中，使用一个更易于处理的函数来近似在给定点上的原始函数。因此，子问题P_3可以近似为问题$P_{3.1}$：

$$\begin{cases} P_{3.1}:\max\limits_{\boldsymbol{w}} \ \sum\limits_{k=1}^{K}\sum\limits_{m=1}^{\text{M}}\sum\limits_{n=1}^{N}s_{k,m}^{n}(R_{1k,m}-R_{2k,m}^{\text{up}}) \\ \text{s.t.} \quad C_1:\sum\limits_{m=1}^{\text{M}}\sum\limits_{n=1}^{N}s_{k,m}^{n}(R_{1k,m}-R_{2k,m}^{\text{up}})\geqslant\alpha \qquad \forall k \\ \qquad C_2:E_{\text{v}}\leqslant\sum\limits_{m=1}^{M}p_{m,\max}T-E_{\text{t}} \qquad \forall m,n \\ \qquad C_3:\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n}\right\|^2\geqslant\left\|\boldsymbol{h}_{k,m}^{n}\boldsymbol{w}_{j,m}^{n,l}\right\|^2+ \\ \qquad 2R\{\boldsymbol{w}_{j,m}^{n,\text{H}}\boldsymbol{h}_{k,m}^{n\text{H}}\boldsymbol{h}_{k,m}^{n}(\boldsymbol{w}_{j,m}^{n}-\boldsymbol{w}_{j,m}^{n,u})\} \end{cases} \tag{26}$$

由于问题$P_{3.1}$是一个凸优化问题，可以直接利用CVX工具有效地解决。

4.3 UAV飞行轨迹和速度优化

在固定用户调度$\boldsymbol{s}$和波束成形$\boldsymbol{w}$情况下，关于优化变量$\boldsymbol{q}$，$\boldsymbol{v}$的子问题P_4表述为

$$\begin{cases} P_4:\max\limits_{\boldsymbol{q},\boldsymbol{v}} \ R_{\text{sum}}=\sum\limits_{k=1}^{K}R_k \\ \text{s.t.} \quad C_1:R_k\geqslant\alpha \qquad \forall k \\ \qquad C_2:E_{\text{v}}\leqslant\sum\limits_{m=1}^{M}p_{m,\max}T\ -E_{\text{t}} \qquad \forall m,n \\ \qquad C_3:\ \left\|\boldsymbol{q}_m^{n+1}-\boldsymbol{q}_m^{n}\right\|^2\leqslant S_{\max}^2 \qquad \forall m,n \\ \qquad C_4:\left\|\boldsymbol{q}_m^{n}-\boldsymbol{q}_j^{n}\right\|^2\geqslant d_{\min}^2 \qquad \forall m,n \\ \qquad C_5:\boldsymbol{q}_m^{1}=\boldsymbol{q}_{m,0},\boldsymbol{q}_m^{N}=\boldsymbol{q}_{m,F} \qquad \forall m \\ \qquad C_6:v_{\min}\leqslant v_{\text{m}}^{n}\leqslant v_{\max} \qquad \forall m,n \end{cases} \tag{27}$$

目标函数和约束条件 C_1 和 C_3 都是非凸的, 为求解该问题,目标函数和约束条件 C_3 利用连续凸近似方法来处理; 对于约束条件 C_1, 通过引入松弛变量, 利用一个近似下界代替约束条件 C_1, 最后, 通过迭代计算逼近原来的约束条件。

C_1 中的 $\|\boldsymbol{h}_{k,m}^n \boldsymbol{w}_{k,m}^n\|^2$ 展开为

$$\left\|\boldsymbol{h}_{k,m}^n \boldsymbol{w}_{k,m}^n\right\|^2 = \frac{\beta_0 \left\|\boldsymbol{a}_{k,m}^n \boldsymbol{w}_{k,m}^n\right\|^2}{\left\|\boldsymbol{q}_m^n - \boldsymbol{g}_k\right\|^2} \tag{28}$$

于是 R_k 可以表示为

$$R_k = \sum_{m=1}^{\mathrm{M}} \sum_{n=1}^{N} s_{k,m}^n \log_2(1 + V_{\mathrm{SIR}k}^n) \tag{29}$$

$$V_{\mathrm{SIR}k}^n = \frac{\left\|\boldsymbol{h}_{k,m}^n \boldsymbol{w}_{k,m}^n\right\|^2}{\sum\limits_{j,j\neq k}^{K} \left\|\boldsymbol{h}_{k,m}^n \boldsymbol{w}_{j,m}^n\right\|^2 + \sigma_z^2} = \frac{\beta_0 \left\|\boldsymbol{a}_{k,m}^n \boldsymbol{w}_{k,m}^n\right\|^2}{\beta_0 \sum\limits_{j,j\neq k}^{K} \left\|\boldsymbol{a}_{k,m}^n \boldsymbol{w}_{j,m}^n\right\|^2 + (\|\boldsymbol{q}^n - \boldsymbol{g}_k\|^2)\sigma_z^2} \tag{30}$$

式中:

$$\begin{cases} P_{k,m}^n = \boldsymbol{w}_{k,m}^{n\mathrm{H}} \boldsymbol{a}_{k,m}^n \boldsymbol{a}_{k,m}^{n\mathrm{H}} \boldsymbol{w}_{k,m}^n \\ P_{j,m}^n = \boldsymbol{w}_{j,m}^{n\mathrm{H}} \boldsymbol{a}_{k,m}^n \boldsymbol{a}_{k,m}^{n\mathrm{H}} \boldsymbol{w}_{j,m}^n \end{cases} \tag{31}$$

因此, 可得

$$R_k = \sum_{k=1}^{K} \sum_{m=1}^{M} s_{k,m}^n \cdot \log_2\left(1 + \frac{P_{k,m}^n}{\sum\limits_{j,j\neq k}^{K} P_{j,m}^n + (\|\boldsymbol{q}^n - \boldsymbol{g}k\|^2)\sigma_z^2}\right) \tag{32}$$

R_k 是关于 $\boldsymbol{q}^n$ 的非凸函数, 但是相对 $\|\boldsymbol{q}^n - \boldsymbol{g}_k\|^2$ 是凸函数, 根据凸函数在任意点的一阶泰勒展开都是其全局的下界的性质, R_k 的全局下界为

$$R_k \geqslant \sum_{m=1}^{M} \sum_{n=1}^{N} s_{k,m}^n (A_{k,m}^{n,i} - B_{k,m}^{n,i} C^n) \tag{33}$$

式中:

$$A_{k,m}^{n,i} = \log_2\left(1 + \frac{\beta_0 P_{k,m}^n}{\beta_0 \sum\limits_{j,j\neq k}^{K} P_{j,m}^n + (\|\boldsymbol{q}^n - \boldsymbol{g}_k\|^2)\sigma_z^2}\right) \tag{34}$$

$$B_{k,m}^{n,i} = \log_2\left(1 + \frac{\beta_0 P_{k,m}^n + (\|\boldsymbol{q}^n - \boldsymbol{g}_k\|^2 + H^2)\sigma_z^2}{\mathrm{In}2\beta_0 \sum\limits_{j,j\neq k}^{K} P_{j,m}^n + \ln 2(\|\boldsymbol{q}^n - \boldsymbol{g}_k\|^2)\sigma_z^2}\right) \tag{35}$$

$$C^n = \|\boldsymbol{q}^n - \boldsymbol{g}_k\|^2 - \left\|\boldsymbol{q}^{n,i} - \boldsymbol{g}_k\right\|^2 \tag{36}$$

式(33)中 R_k 的右边定义为 R_k^{lb}, 根据海赛矩阵, R_k^{lb} 相对 $\boldsymbol{q}^n$ 是凸的。

由于约束 C_1 是非凸的, 引入一个松弛变量 $\varphi_{m,n} \geqslant \left|v_m^n\right|$, 于是 C_1 中的 $p(v_m^n)$ 可以改写为

$$p_{\mathrm{appro}}(\varphi_{m,n}) = c_1\left(1 + \frac{3\varphi_{m,n}^2}{U_{\mathrm{tip}}}\right) + c_2 u_{m,n} + \frac{1}{2} d_0 \rho s A \varphi_{m,n}^3 \tag{37}$$

式(37)仍然是非凸的, 通过引入另一个松弛变量 $u_{m,n}^2 \geqslant \sqrt{1 + \frac{\varphi_{m,n}^4}{4v_0^4}} - \frac{\varphi_{m,n}^2}{2v_0^2}$, 化简后得到, $\frac{1}{u_{m,n}^2} \leqslant u_{m,n}^2 + \frac{\varphi_{m,n}^2}{v_0^2}$, 然后任意局部点 $\{u_{m,n}^m, \varphi_{m,n}^m\}$, m 表示第 m 次迭代, $u_{m,n}^2 + \frac{\varphi_{m,n}^2}{v_0^2}$ 可以通过一阶泰勒展开式得到下界, 因为其相对于 $u_{m,n}$ 和 $\varphi_{m,n}$ 是共同凸的, 用 $X_{m,n}^{\mathrm{lb}}$ 表示下界函数:

$$X_n^{\mathrm{lb}} \triangleq (u_n^m)^2 + u_n^m(u_n - u_n^m) + \frac{(\varphi_n^m)^2}{v_0^2} + \frac{2\varphi_n^m}{v_0^2}(\varphi_n - \varphi_n^m) \tag{38}$$

UAV 飞行能耗计算式(33)近似为

$$p_{\mathrm{appro}}(\varphi_{m,n}) = c_1\left(1 + \frac{3\varphi_{m,n}^2}{U_{\mathrm{tip}}}\right) + c_2 u_{m,n} + \frac{1}{2} d_0 \rho s A \varphi_{m,n}^3 \tag{39}$$

综合上述分析, 子问题 P_4 可以转换成子问题 $P_{4.1}$ 如下:

$$\begin{cases} P_{4.1}: \max\limits_{\boldsymbol{q},\boldsymbol{v}} R_{\mathrm{sum}} = \sum\limits_{k=1}^{K} R_{k,m}^{\mathrm{lb}} \\ \quad \text{s.t.} \quad C_1: R_{k,m}^{\mathrm{lb}} \geqslant \alpha \qquad \forall k \\ \qquad C_2: \sum\limits_{m=1}^{M}\sum\limits_{n=1}^{N} \delta_t p_{\mathrm{appro}}(\varphi_{m,n}) \leqslant \sum\limits_{m=1}^{M} p_{m,\max} T - E_{\mathrm{t}} \\ \qquad\qquad \forall m,n \\ \qquad C_3: \varphi_{m,n}^2 \geqslant \left|v_m^n\right|^2 \\ \qquad C_4: X_{m,n}^{\mathrm{lb}} \geqslant \frac{1}{u_{m,n}^2} \\ \qquad C_5: \left\|\boldsymbol{q}_m^{n+1} - \boldsymbol{q}_m^n\right\|^2 \leqslant S_{\max}^2, S_{\max} = v_{\max}\delta_t \\ \qquad C_6: \left\|\boldsymbol{q}_m^n - \boldsymbol{q}_j^n\right\|^2 \geqslant d_{\max}^2 \qquad \forall m,n \\ \qquad C_7: \boldsymbol{q}_m^1 = \boldsymbol{q}_{m,0}, \boldsymbol{q}_m^N = \boldsymbol{q}_{m,F} \qquad \forall m \\ \qquad C_8: v_{\min} \leqslant v_m^n \leqslant v_{\max} \qquad \forall m,n \end{cases} \tag{40}$$

C_2、C_3 和 C_4 保证问题 P_4 可以转换成问题 $P_{4.1}$, 则问题 $P_{4.1}$ 是一个凸优化问题, 可以使用凸优化工具 CVX 求解。

4.4　本文算法设计与分析

综上所述, 本文提出的基于 BCD 的资源分配算法如下。

算法: 基于 BCD 的资源分配算法

初始化系统参数: s^0 $\boldsymbol{w}^0$, $\boldsymbol{q}^0$, $\boldsymbol{v}^0$; 设置最大迭代次数 $L\max$; 迭代次数 $l = 0$;

1: Repeat

2: $l=l+1$;

3: 根据给定的$\{\boldsymbol{w}^l, \boldsymbol{q}^l, \boldsymbol{v}^l\}$, 通过解决$P_2$, 得到用户调度解$s^{l+1}$;

4: 根据给定的$\{s^{l+1},\boldsymbol{q}^l,\boldsymbol{v}^l\}$, 通过解决$P_3$, 得到波束成形解$\boldsymbol{w}^{l+1}$;

5: 根据给定的$\{s^{l+1},\boldsymbol{w}^{l+1}\}$, 通过解决$P_4$, 得到UAV轨迹和飞行速度解$\boldsymbol{q}^{l+1},\boldsymbol{v}^{l+1}$;

6: Until $R_{\text{sum}}^{l}-R_{\text{sum}}^{l-1}\leqslant\varepsilon$或者$l\geqslant L_{\max}$;

7: 结束并输出结果。

定理 本文算法可得到P_1的次优解且收敛

证明 在第$l+1$次迭代过程中, 在给定波束成形$\boldsymbol{w}^l$、UAV飞行轨迹$\boldsymbol{q}^l$及速度$\boldsymbol{v}^l$下, 通过求解P_2可以得到最优的用户调度s^{l+1}, 此时有$C(s^{l+1},\boldsymbol{w}^l,\boldsymbol{q}^l,\boldsymbol{v}^l)\geqslant C(s^l,\boldsymbol{w}^l,\boldsymbol{q}^l,\boldsymbol{v}^l)$。其中, $C(\cdot)$表示相应参数下的系统总吞吐量。

在用户调度s^{l+1}、UAV飞行轨迹$\boldsymbol{q}^l$和速度$\boldsymbol{v}^l$下, 通过SCA方法求解P_3可以得到最优的波束成形$\boldsymbol{w}^{l+1}$, 此时有$C(s^{l+1},\boldsymbol{w}^{l+1},\boldsymbol{q}^l,\boldsymbol{v}^l)\geqslant C(s^{l+1},\boldsymbol{w}^l,\boldsymbol{q}^l,\boldsymbol{v}^l)$。

在求解UAV轨迹和速度子问题时, 在用户调度s^{l+1}和波束成形$\boldsymbol{w}^{l+1}$下, 通过一阶泰勒展开、SCA等方法求解得到UAV的飞行轨迹$\boldsymbol{q}^{l+1}$和速度$\boldsymbol{v}^{l+1}$。在此过程中, 一阶泰勒展开可以等得到相同的目标函数值, 引入松弛变量, 式(34)存在下界。此时有$C(s^{l+1},\boldsymbol{w}^{l+1},\boldsymbol{q}^{l+1},\boldsymbol{v}^{l+1})\geqslant C(s^{l+1},\boldsymbol{w}^{l+1},\boldsymbol{q}^l,\boldsymbol{v}^l)$。

因此, 在第$l+1$迭代中, 可以得到$C(s^{l+1},\boldsymbol{w}^{l+1},\boldsymbol{q}^{l+1},\boldsymbol{v}^{l+1})\geqslant C(s^l,\boldsymbol{w}^l,\boldsymbol{q}^l,\boldsymbol{v}^l)$, 由此可见, P_1的目标值在每次的迭代中是非递减的, 并且在给定精度阀值ε的条件下, P_1的优化目标函数具有有限的上界, 本文算法的收敛性可以得到保证。

证毕

本文算法的总计算复杂度主要取决于算法中的步骤3～步骤5。其中, 步骤3直接利用CVX求解, 其计算复杂度为$O\{(KN)^{3.5}\}$。步骤4和步骤5利用SCA和CVX进行求解, 其计算复杂度分别为$O\{L_1(KN)^{3.5}\}$和$O\{L_2(KN)^{3.5}+KN\}$[17], 其中, L_1和L_2分别为步骤3和步骤4的SCA逼近的次数。综上, 本文算法的总计算复杂度为$O\{((L_1+L_2+1)(KN)^{3.5}+KN)\ln(1/\varepsilon)\}$, 其中$\varepsilon$为收敛精度。

5 仿真分析

本节考虑了配备8根均匀天线阵列的UAV基站为随机分布在600 m×600 m的6个单天线用户提供通信服务。用户的位置通过随机产生得到如图2所示, 分别为$\boldsymbol{g}_1$=[360,111], $\boldsymbol{g}_2$=[26,403], $\boldsymbol{g}_3$=[128,194], $\boldsymbol{g}_4$=[187,38], $\boldsymbol{g}_5$=[171,338], $\boldsymbol{g}_6$=[307,429], $\boldsymbol{q}_I$=[200,25], $\boldsymbol{q}_F$=[250,400]。其中, H=100 m, N=50, δ_t=2 s, $V_{\max}$=30 m/s, α=1.8 bit/sHz, $\varepsilon=10^{-2}$, $d_{\min}$=10 m, D=0.5, $L_{\max}$=30; 接收信号的噪声功率σ_z^2=−110 dBm, 相对参考距离的信道增益β_0=−60 dB。另外仿真过程中其他的参数见表1。

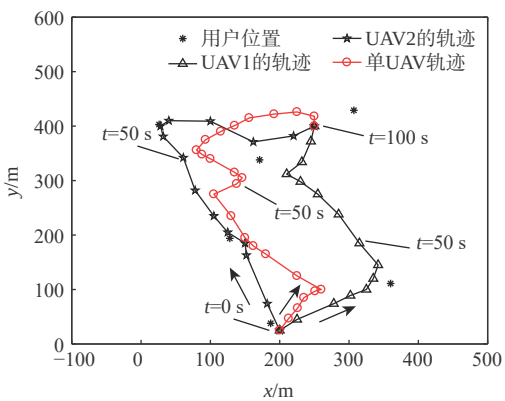


图2 UAV飞行轨迹

Fig. 2 UAV flight trajectory

表1 仿真参数

Table 1 Simulation parameters

c_1	c_2	转子的叶尖速度U_{tip}/(m·s^{-1})	机身阻力比d_0	平均转子速度v_{m}/(m·s^{-1})	空气密度ρ/(kg·m^{-3})	转子密度/10^3(kg·m^{-3})	转子盘面积A/m^2
60	81.5	120	0.6	4.03	1.225	1.5	0.603

为验证本文算法有效性, 对比分析了单、多无人机通信系统性能。单无人机通信系统, $m=1$, 总功率设置为$p_{\max}$=30 dBm; 双无人机通信系统, $m=2$, 且每架无人机的最大功率$p_{1,\max}=p_{2,\max}=p_{\max}/2$。无人机飞行周期均为$T$=100 s, 单、双通信系统的无人机的轨迹如图2所示。

图2分别为单、双无人机通信系统飞行轨迹优化的结果。由图可见, 单、双无人机系统远离用户时加速, 接近用户时逐渐减速, 这表明无人机可以停留在用户附近更长的时间, 从而获得更大的吞吐量。而在双无人机通信系统中, 两架无人机的飞行轨迹相互远离, 以保持安全距离且避免信号干扰。

图3为T=100 s, k=6时的3种波束成形方案下双无人机通信系统总吞吐量随飞行时间变化的情况。显然, 随着时间的推移, 3种方案的吞吐量都持续增加。其中, 本文的优化波束成形方案具有最优的吞吐量性能, 这是因为通过联合无人机轨迹设计和用户调度优化, 来获取到更多的自由度优化吞吐量。而随机波束成形方案吞吐量低于均等波束成形方案, 说明了均等波束成形方案, 能够适当减少

通信能耗, 提升系统的总吞吐量。

图 4 为 T=100 s, k=6 时的单、双无人机通信系统迭代次数的关系。可以看出, 2 种通信系统的总吞吐量随着迭代次数增加而增加, 最终在约 5 次迭代后均达到收敛, 说明本文算法均具有良好的收敛性, 而双无人机通信系统性能优于单无人机通信系统。

图 5 为 p_{max}=30 dBm, k=6 时, 单、双无人机系统在不同飞行周期下的通信能耗和飞行能耗占比关系。由图可见, 随着飞行周期增加, 2 种通信系统的通信能耗均先增后减, 而系统的飞行能耗均先减后增。这是因为当飞行周期设置较小时, 无人机很难到达最佳通信位置, 因此, 通信能耗占比较小; 而飞行周期设置过长, 将会消耗更多的飞行能量, 使得通信能耗减少。因此, 在该场景下飞行周期最佳的设置范围在 110～130 s 为宜。

图 6 为多无人机通信系统在不同飞行周期下的总吞吐量的对比关系。在多无人机 ($m = i$, i = 1、2、3)通信系统中, 假设每架无人机的最大功率为 p_{max}/i, 无人机飞行周期均为 T=100 s。由图可见, 随着飞行周期增加, 3 种系统的总吞吐量均先增后减, 这是由于无人机飞行周期设置长会造成大量的飞行能量消耗, 最终降低系统的通信总吞吐量。当飞行周期超过 130 s 后, 双无人通信系统的吞吐量反而低于单无人机通信系统; 当飞行周期超过 120 s 后, 3 无人通信系统的吞吐量也明显低于单无人机通信系统, 这是由于当飞行周期较长, 多无人机将消耗更多的飞行能耗。

图 7 为 T=100 s 时, 多无人机通信系统用户数

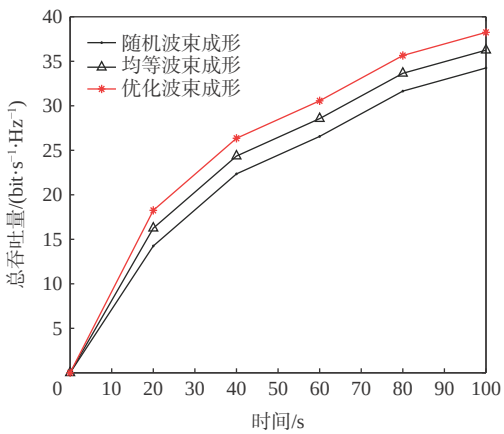


图 3　3 种波束成形方法的吞吐量对比

Fig. 3　Comparison of throughput of three beamforming methods

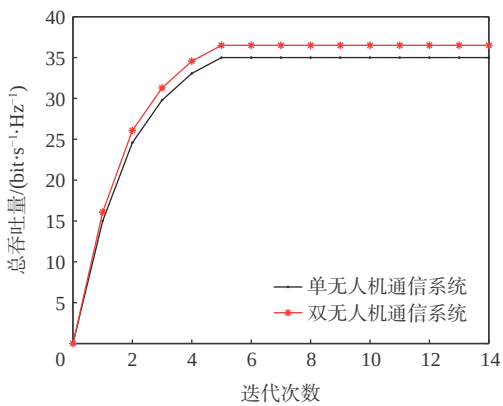


图 4　单、双无人机通信系统迭代次数与总吞吐量的关系

Fig. 4　Relationship between iteration times and total throughput of single and dual UAV communication systems

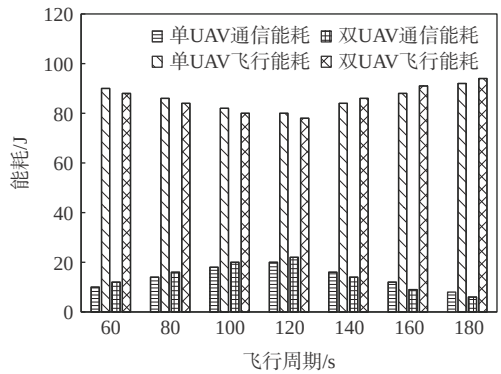


图 5　单、多无人机通信系统在不同飞行周期下的能耗关系

Fig. 5　Energy consumption relationship of single and dual unmanned aerial vehicle communication systems under different flight cycles

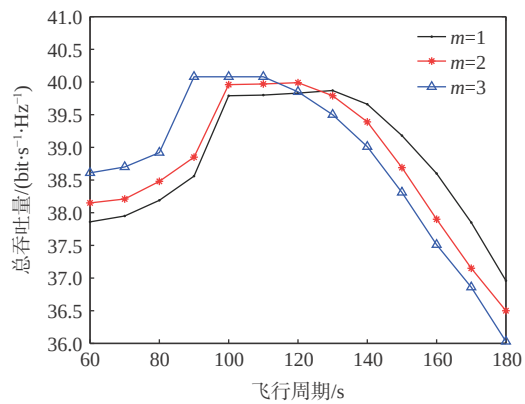


图 6　多无人机通信系统在不同飞行周期下的总吞吐量

Fig. 6　Total throughput of multiple unmanned aerial vehicle communication systems under different flight cycles

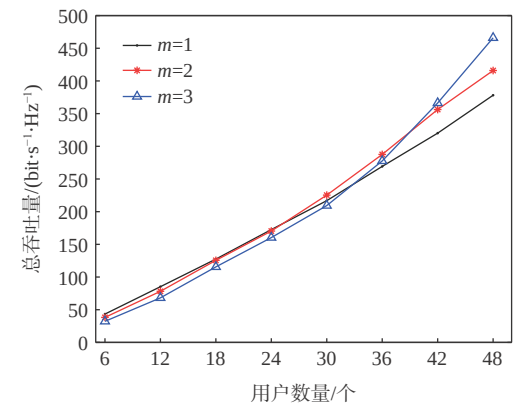


图 7　多无人机通信系统用户数量与系统吞吐量的关系

Fig. 7　Relationship between the number of users and system throughput in multi-UAV communication systems

量与系统吞吐量的关系，假设每次在目标区域（600 m×600 m）的随机生成新增6个用户位置坐标，无人机的最大总功率同步提升一倍。由图可见，当用户数量大于25个时，双无人机通信系统性能明显高于单无人机通信系统，且随着用户数的增加，两者差距将会越来越大；当用户数量大于32和39个后，3无人机通信系统性能分别高于单无人机和双无人机通信系统。因此，当用户数量较少时，应当选择单无人机通信系统，对于目标区域用户数量比较多时，尽可能选择多无人机通信系统。

6 结 论

1）本文研究了多UAV辅助无线通信系统下的资源分配问题，提出一种基于BCD资源分配算法。考虑了用户调度、UAV通信和飞行能耗、移动轨迹和用户最低传输速率等约束；建立了多变量耦合的系统最大化总吞吐量问题；通过BCD将优化问题分解成3个子问题，对3个子问题交替迭代优化，得出原优化问题的近似次优解。

2）仿真结果表明，本文算法在单、双无人机通信系统均具有良好的收敛性，并且在用户数量较多的情况下，多无人机通信系统性能明显优于单无人通信系统。

参考文献（References）

[1] 张海君, 陈安琪, 李亚博, 等. 6G移动网络关键技术[J]. 通信学报, 2022, 43(7): 189-202.
ZHANG H J, CHEN A Q , LI Y B, et al. Key technologies of 6G mobile network[J]. Journal on Communications, 2020, 43(7): 189-202(in Chinese).

[2] WANG M Y, LIN Y, TIAN Q, et al. Transfer learning promotes 6G wireless communications: recent advances and future challenges[J]. IEEE Transactions on Reliability, 2021, 70(2): 790-807.

[3] WANG B W, SUN Y J, SUN Z, et al. UAV-assisted emergency communications in social IoT: a dynamic hypergraph coloring approach[J]. IEEE Internet of Things Journal, 2020, 7(8): 7663-7677.

[4] ZENG F Z, HU Z Z, XIAO Z, et al. Resource allocation and trajectory optimization for QoE provisioning in energy-efficient UAV-enabled wireless networks[J]. IEEE Transactions on Vehicular Technology, 2020, 69(7): 7634-7647.

[5] ALZENAD M, EL-KEYI A, YANIKOMEROGLU H. 3-D placement of an unmanned aerial vehicle base station for maximum coverage of users with different QoS requirements[J]. IEEE Wireless Communications Letters, 2018, 7(1): 38-41.

[6] XIAO Z Y, DONG H, BAI L, et al. Unmanned aerial vehicle base station (UAV-BS) deployment with millimeter-wave beamforming [J]. IEEE Internet of Things Journal, 2020, 7(2): 1336-1349.

[7] XU J, ZENG Y, ZHANG R. UAV-enabled wireless power transfer: trajectory design and energy optimization[J]. IEEE Transactions on Wireless Communications, 2018, 17(8): 5092-5106.

[8] WU Q Q, ZENG Y, ZHANG R. Joint trajectory and communication design for multi-UAV enabled wireless networks[J]. IEEE Transactions on Wireless Communications, 2018, 17(3): 2109-2121.

[9] 张广驰, 陈娇, 崔苗, 等. 无人机交替中继通信及其轨迹优化和功率分配研究[J]. 电子与信息学报, 2021, 43(12): 3554-3562.
ZHANG G C, CHEN J, CUI M, et al. Trajectory optimization and power allocation for UAV alternate relay communications[J]. Journal of Electronics & Information Technology, 2021, 43(12): 3554-3562(in Chinese).

[10] WANG H C, REN G C, CHEN J, et al. Unmanned aerial vehicle-aided communications: joint transmit power and trajectory optimization[J]. IEEE Wireless Communications Letters, 2018, 7(4): 522-525.

[11] 张涛, 张君, 唐洪莹, 等. 无人机辅助无线供电通信网络节能通信[J]. 传感器与微系统, 2021, 40(12): 44-48.
ZHANG T, ZHANG J, TANG H Y, et al. Energy minimization communication in UAV-aided wireless powered communication network[J]. Transducer and Microsystem Technologies, 2021, 40(12): 44-48(in Chinese).

[12] GUO W, ZHANG W L, WANG Y C, et al. Joint attitude and power optimization for UAV-aided downlink communications[J]. IEEE Transactions on Vehicular Technology, 2019, 68(12): 12437-12442.

[13] SONG Q H, ZHENG F C, ZENG Y, et al. Joint beamforming and power allocation for UAV-enabled full-duplex relay[J]. IEEE Transactions on Vehicular Technology, 2019, 68(2): 1657-1671.

[14] LIN N, FAN Y B, ZHAO L, et al. GREEN: a global energy efficiency maximization strategy for multi-UAV enabled communication systems[C]//Proceedings of the IEEE Transactions on Mobile Computing. Piscataway: IEEE Press, 2023: 7104-7120.

[15] ZHANG T K, XU Y, LOO J, et al. Joint computation and communication design for UAV-assisted mobile edge computing in IoT[J]. IEEE Transactions on Industrial Informatics, 2020, 16(8): 5505-5516.

[16] RAZAVIYAYN M. Successive convex approximation: analysis and applications[D]. Minnesota: University of Minnesota, 2014.

[17] BOYD S P, VANDENBERGHE L. Convex optimization[M]. Cambridge: Cambridge University Press, 2004.

Resource optimization of multi UAV assisted communication system based on user scheduling

TANG Jingmin[1], HUANG Jiaqi[1,*], WANG Bingwen[1], SONG Yaolian[1], YU Guicai[2]

(1. School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China;
2. School of Physical Science and Engineering Technology, Yichun University, Yichun 336000, China)

Abstract: In order to improve the transmission rate of multi-user mobile communication downlink wireless transmission systems, a resource optimization method for multi-unmanned aerial vehicle (UAV) auxiliary communication systems based on user scheduling and trajectory optimization is proposed. The proposed method formulates an optimization problem based on maximizing the total throughput of multiple users while satisfying constraints such as user scheduling, total energy consumption of drones, and user service quality requirements. In order to solve the nonconvex problem, the original nonconvex problem is decomposed into three easy to deal with nonconvex subproblems by the block coordinate descent (BCD), and the approximate suboptimal solution of the original nonconvex problem is obtained by introducing relaxation variables, first-order Taylor expression, successive convex approximation (SCA) and other methods to transform the subproblems and solve them alternately. Simulation results show that the proposed method can effectively improve overall system throughput and has good convergence in single and multiple UAV communication systems.

Keywords: UAV communication; resource allocation; user scheduling; beamforming; flight speed

Received: 2023-05-11; **Accepted**: 2023-07-28; **Published Online**: 2023-09-07 16:13
URL: link.cnki.net/urlid/11.2625.V.20230907.1135.005
Foundation item: National Natural Science Foundation of China (62261056)
* **Corresponding author**. E-mail: 1540600368@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0174

基于 SSA-IWT-EMD 的滚动轴承故障诊断方法

雷春丽[1]，焦孟萱[2]，樊高峰[3]，刘世超[2]，薛林林[1]，李建华[1,*]

(1. 兰州理工大学 机电工程学院，兰州 730050；2. 中石化英力士（天津）石化有限公司，天津 300280；
3. 云南文山铝业有限公司，文山 663000)

摘　　要：针对小波阈值降噪不充分及经验模态分解 (EMD) 特征频率提取不明显的问题，提出一种基于麻雀搜索算法-改进小波阈值-EMD(SSA-IWT-EMD) 的滚动轴承故障诊断方法。引入2个调节因子，提出一种 IWT 函数，克服了传统软硬阈值的缺点，并运用 SSA 对其各参数进行全局寻优，实现滚动轴承信号降噪。提出一种综合指标 P 对 EMD 产生的分量进行选取重构，突出信号的故障特征信息。采用包络谱分析实现轴承的故障诊断。仿真和实测结果验证了所提方法的有效性；同时与单一指标选取分量的方法及文献方法进行对比，说明了综合指标 P 和所提方法具有更强的降噪能力及特征提取能力，包络谱幅值及倍频成分更明显，可以更好地实现对滚动轴承的故障诊断。

关 键 词：滚动轴承；改进阈值；综合指标；经验模态分解；故障诊断
中图分类号：TH133.33
文献标志码：A　　**文章编号**：1001-5965(2025)04-1152-11

旋转机械广泛应用于航空航天、数控机床、机器人等领域。对于旋转机械而言，轴承是至关重要的部件，其健康状态直接影响着系统的运行品质。据统计，有 45%～55% 的旋转机械故障是由滚动轴承引起的[1]。在电机领域，该故障占比高达 40% 以上[2]。所以对轴承进行有效的故障诊断是有必要的，其对于设备长期稳定地工作具有重要意义。

为能够准确地识别出轴承故障，学者提出了多种故障诊断方法。Shao 等[3]将小波阈值应用于经验模态分解（empirical mode decomposition, EMD）产生的本征模态函数（intrinsic mode function, IMF）分量上，有效实现了故障的特征提取。王普等[4]实现了小波阈值的分层自适应性，实现高效降噪，增强了信号的故障特征。由于轴承复杂的运行状态，其特征频率容易被掩盖，针对此现象，Cui 等[5]运用小波变化及稀疏分量分析的方式对信号进行特征提取，通过频谱分析确定故障所在位置，通过实验验证了该方法具有一定的优势。针对多数降噪方法会滤除部分重要信息的问题，Chen 等[6]首先将信号进行共振稀疏分解，然后，将处理后的信号进行小波分析，提取其中的故障特征，解决了强噪声干扰下的轴承故障诊断问题。崔玲丽等[7]将第 2 代小波用于信号分析，运用仿真和试验信号验证该方法可准确识别滚动轴承的早期故障类型。王志刚等[8]运用谐波小波分析凸显出信号中的奇异成分，有效应用于故障诊断领域。张晓峰和李功燕[9]运用小波分析提出特定频段的故障信号。邓飞跃等[10]提出一种自适应频率窗经验小波变换提取特征信息，实现了对轴承在强噪声条件下的故障诊断。阈值选取是小波阈值处理信号的一大难题，对此部分学者提出优化算法对阈值进行寻优进而确定阈值。Tuerxun[11]采用麻雀搜索算法（sparrow search algorithm, SSA）优化支持向量机，在收敛速度及优化能力上大大提高。Ren[12]运用麻雀搜索算法优化变分模态分解的参数，实现信号降噪，随后对降噪后的信号进行奇异值分解，降噪效果显著。经验模

收稿日期：2023-04-12；**录用日期**：2023-06-30；**网络出版时间**：2023-07-10 16:05
网络出版地址：link.cnki.net/urlid/11.2625.V.20230710.1537.002
基金项目：国家自然科学基金 (51465035)；甘肃省自然科学基金 (20JR5RA466)；兰州理工大学红柳一流学科建设项目
*通信作者. E-mail：li_jh@vip.sina.com

引用格式：雷春丽，焦孟萱，樊高峰，等. 基于 SSA-IWT-EMD 的滚动轴承故障诊断方法 [J]. 北京航空航天大学学报，2025，51（4）：1152-1162. LEI C L，JIAO M X，FAN G F，et al. Fault diagnosis method of rolling bearings based on SSA-IWT-EMD [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1152-1162（in Chinese）.

态分解由 Huang 等[13] 提出，该方法可以自适应地对非平稳信号进行分解，其在故障诊断方面被广泛应用。Keshtan 和 Nouri Khajavi[14] 将 EMD 运用到故障诊断中并运用希尔伯特边缘谱实现对滚轴轴承故障状态的识别。丛晓妍等[15] 将振动信号进行 EMD 分解，得到多个 IMF 分量，采用峭度准则对 IMF 分量进行选取，该方法极大地减少了干扰信号的影响。郭俊超等[16] 将调制信号双谱与加权平均集成经验模态分解相结合，准确获取滚动轴承故障特征。

上述方法对信号进行小波分析或小波阈值处理，将处理好的信号进行频谱分析或包络谱分析等实现故障诊断。对信号进行小波阈值处理时，阈值的选取是一个难点。而传统硬、软阈值函数分别存在不连续及失真现象，对试验结果会造成一定的误差。上述方法未考虑二者缺点；同时，对于阈值的选取是根据经验而定的，没有实现自适应选取。此外，上述方法中采用单一指标对 EMD 分量进行选取，在一定程度上会导致重要信息缺失，最终影响故障诊断效果。

本文提出一种基于 SSA-改进小波阈值（improved wavelet threshold, IWT）-EMD 的滚动轴承故障诊断方法。首先，提出了改进的小波阈值函数，并采用麻雀搜索算法对该函数中的调节因子及阈值进行寻优，同时进行降噪处理；然后，将降噪信号进行 EMD 分解，采用本文提出的综合指标 P 进行分量选取，将所选分量进行重构实现对故障特征的有效提取；最后，结合包络谱分析进行故障诊断。通过对仿真试验及实测试验结果进行分析，验证了本文方法的有效性和可靠性。

1　小波阈值降噪

1.1　小波阈值降噪原理

经过小波处理后信号会形成多个小波系数，这些系数含有重要的特征信息。以阈值作为区分噪声及信号的界限，保留系数大的重要信息，去除系数小的噪声，重组处理完后的小波系数，得到最终的信号。图 1 为阈值降噪的流程。

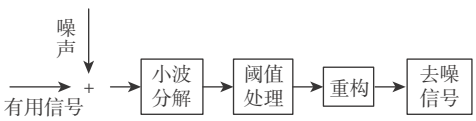


图 1　小波阈值去噪流程

Fig. 1　Flow chart of wavelet threshold denoising

1.2　改进小波阈值函数

小波阈值去噪是一种简单有效的去噪手段，其中，阈值函数的选取尤为重要[17]，硬阈值及软阈值函数较为常见。

硬阈值函数表达式为

$$\hat{\omega}=\begin{cases}\omega & |\omega|\geqslant\lambda\\ 0 & |\omega|<\lambda\end{cases} \tag{1}$$

式中：$\hat{\omega}$为处理过后的小波系数；ω为分解的原始小波系数；λ为阈值。

对于硬阈值函数来说，该方法保留信号的局部特征，具有较好的边界性[18]，但其不具连续性，小波系数在阈值处不连续，导致信号去噪效果粗糙，同时运用该方法形成的重构信号会产生附加震荡。

软阈值函数表达式为

$$\hat{\omega}=\begin{cases}\mathrm{sgn}(\omega)(|\omega|-\lambda) & |\omega|\geqslant\lambda\\ 0 & |\omega|<\lambda\end{cases} \tag{2}$$

式中：$\mathrm{sgn}(\omega)$为符号函数，当ω为正数时函数值为 1，当ω为负数时函数值为−1，当ω为 0 时函数值为 0。

相较硬阈值函数，软阈值函数在连续性上具有一定优势，具有较平稳的去噪效果[19]。但该方法会使小波系数产生偏差，导致信号产生失真与原信号产生差距。本文利用硬阈值函数具有保留局部特征及软阈值函数具有连续性的优势，引入 γ、β 这 2 个调节因子，提出了 IWT 函数，表达式为

$$\hat{\omega}=\begin{cases}\mu\omega+(1-\mu)\mathrm{sgn}(\omega)\left(|\omega|-\beta\lambda\mathrm{e}^{-\frac{\omega^2}{\lambda}}\right) & |\omega|\geqslant\lambda\\ 0 & |\omega|<\lambda\end{cases} \tag{3}$$

$$\mu=1-\mathrm{e}^{-\gamma(|\omega|-\lambda)^2} \tag{4}$$

式中：$\gamma>0$，$\beta\in(0,1)$，调节因子使得该函数处于硬阈值与软阈值之间，达到更好的降噪效果。

2 个调节因子在阈值函数中起着不同的作用，共同决定小波阈值降噪的过程与结果。当$\beta\to0$且$|\omega|\geqslant\lambda$，不论γ取何值，阈值函数均为硬阈值函数；当$\beta\to1$，$\gamma\to0$且$|\omega|\geqslant\lambda$时，使得新阈值函数接近软阈值函数，保留软阈值函数处理重构后信号更加圆滑的特点；当$\beta\to1$，$\gamma\to+\infty$且$|\omega|\geqslant\lambda$时，阈值函数逐渐向硬阈值函数靠近，保留硬阈值函数处理信号注重边缘细节的特点，从而使重构信号不容易发生失真。

由此可见，通过调节调节因子γ、β的取值可以决定阈值函数对小波系数的作用，其中，调节因子γ主要决定阈值函数的趋向，调节因子β决定阈值函数对小波系数的作用程度，通过γ、β这 2 个参数的共同作用，使得此阈值函数衍变为软阈值和硬阈值函数间一种广义函数，在保留 2 种阈值函数各自优点的同时，也克服了其在处理小波系数时的缺点。

1.3　IWT 函数数学理论分析

为说明 IWT 函数的连续性及偏差性，对 IWT

函数进行如下数学分析。目标函数为

$$f(x)=\left[1-\mathrm{e}^{-\gamma(|x|-\lambda)^2}\right]x+\mathrm{e}^{-\gamma(|x|-\lambda)^2}\mathrm{sgn}(x)\left(|x|-\beta\lambda\mathrm{e}^{-\frac{x^2}{\lambda}}\right) \tag{5}$$

IWT函数连续性分析：

当$x>0$时，

$$\lim_{x\to\infty}\frac{f(x)}{x}=\lim_{x\to\infty}\frac{\left[1-\mathrm{e}^{-\gamma(x-\lambda)^2}\right]x+\mathrm{e}^{-\gamma(x-\lambda)^2}\left(x-\beta\lambda\mathrm{e}^{-\frac{x^2}{\lambda}}\right)}{x}=1$$

当$x<0$时，

$$\lim_{x\to\infty}\frac{f(x)}{x}=\lim_{x\to\infty}\frac{\left[1-\mathrm{e}^{-\gamma(-x-\lambda)^2}\right]x-\mathrm{e}^{-\gamma(-x-\lambda)^2}\left(-x-\beta\lambda\mathrm{e}^{-\frac{x^2}{\lambda}}\right)}{x}=1$$

以上推论证明目标函数的连续性，有效解决了硬阈值函数不连续的问题，使去噪后的信号更加光滑，消除了伪吉布斯现象。

IWT函数偏差性分析：

$$\lim_{x\to\infty}f(x)-x=\lim_{x\to\infty}\left(1-\mathrm{e}^{-\gamma(|x|-\lambda)^2}\right)x+\mathrm{e}^{-\gamma(|x|-\lambda)^2}\mathrm{sgn}(x)\left(|x|-\beta\lambda\mathrm{e}^{-\frac{x^2}{\lambda}}\right)-x=0$$

新阈值函数下的小波系数与原始阈值函数下的小波系数不断靠近，解决了软阈值函数存在偏差的问题。根据本节数学分析可知目标函数逐渐接近$f(x)=x$，同时所得新的小波系数$\hat{\omega}$与原始小波系数ω不断靠近。由此推导可以看出，该阈值函数具有较好的连续性及偏差性，克服了传统硬、软阈值函数所存在的问题。

2 SSA和EMD方法

2.1 SSA

SSA在2020年被提出以来已经被应用于诸多领域，SSA是模仿麻雀觅食行为及反捕食衍化出来的算法，该算法收敛快、精度高、稳定性好，同时可以避免陷入局部最优[20]。麻雀主要扮演2种角色，一种是搜索者，另一种是跟随者。

搜索者位置更替可表示为

$$\boldsymbol{X}_{i,j}^{t+1}=\begin{cases}\boldsymbol{X}_{i,j}^{t}\exp\left(-\dfrac{i}{\alpha t_{\max}}\right) & R_2<S\\ \boldsymbol{X}_{i,j}^{t}+Q\boldsymbol{L} & R_2\geqslant S\end{cases} \tag{6}$$

式中：t为更新次数；$t_{\max}$最大更新次数；$\boldsymbol{X}_{i,j}^{t}$为第i个个体在j维上的位置；α为0～1中的任意数；R_2为警报值，R_2为0～1中的随机数；S为安全值，S为0.5～1中的随机数；Q为服从正态分布的随机数；$\boldsymbol{L}$为各个元素单元是1的$1\times d$维矩阵。

跟随者位置更替可表示为

$$\boldsymbol{X}_{i,j}^{t+1}=\begin{cases}Q\exp\left(\dfrac{\boldsymbol{X}_{\mathrm{worst}}^{t}-\boldsymbol{X}_{i,j}^{t}}{i^2}\right) & i>n/2\\ \boldsymbol{X}_{\mathrm{P}}^{t+1}+\left|\boldsymbol{X}_{i,j}^{t}+\boldsymbol{X}_{i,j}^{t+1}\right|\boldsymbol{A}^{+}\boldsymbol{L} & \text{其他}\end{cases} \tag{7}$$

式中：$\boldsymbol{X}_{\mathrm{p}}$为搜索者当前所占据的理想位置；$\boldsymbol{X}_{\mathrm{worst}}$为最差位置；$\boldsymbol{A}$矩阵中各个元素单元为1或者−1，维度是$1\times d$；$\boldsymbol{A}^{+}$为伪逆矩阵。

降低被捕食概率行为可表示为

$$\boldsymbol{X}_{i,j}^{i+1}=\begin{cases}\boldsymbol{X}_{\mathrm{best}}^{t}+\sigma\left(\boldsymbol{X}_{i,j}^{t}-\boldsymbol{X}_{\mathrm{best}}^{t}\right) & f_i>f_g\\ \boldsymbol{X}_{i,j}^{t}+W\left[\dfrac{\boldsymbol{X}_{i,j}^{t}-\boldsymbol{X}_{\mathrm{worst}}^{t}}{(f_{\mathrm{i}}-f_{\mathrm{w}})+\varepsilon}\right] & f_{\mathrm{i}}=f_{\mathrm{g}}\end{cases} \tag{8}$$

式中：$\boldsymbol{X}_{\mathrm{best}}$为全局最优位置；$\sigma$、$W$为随机数，$\sigma$服从方差为1、均值为0的正态分布，$W\in[-1,1]$；$f_{\mathrm{i}}$为个体适应度值；$f_{\mathrm{g}}$为全局最优适应度；$f_{\mathrm{w}}$为全局最差适应度；$\varepsilon$为最小常数，以避免零点整除错误。

SSA在寻求最优解时，需要计算当前的适应度函数值，通过对比选择更优位置，适应度函数的选取起着至关重要的作用。对于滚动轴承信号而言，峭度系数是衡量轴承在运转一周时产生冲击脉冲的参数，幅值越大表示故障现象越明显，冲击成分越多，如式(9)所示。同时考虑到峭度系数计算较为简单，所以本文将式(10)作为适应度函数，搜索最大峭度系数，运用SSA迭代更替实现位置的更新。

$$K=E\left((o-\theta)^4\right)/\delta^4 \tag{9}$$

$$F=1/K \tag{10}$$

式中：o为信号；θ为信号的均值；$E(\cdot)$为均值运算；δ为信号的标准差；K为峭度值。

通过SSA实现对新阈值函数中调节因子及阈值的选取，进而实现对滚动轴承信号的降噪。

2.2 EMD方法

经验模态分解是由Huang等[13]提出的一种时频域信号处理方法，其对信号按照频率大小进行分解，将信号分解成多个具有局部特征的IMF分量之和，频率越高，特征越明显，其可以高效提取信号中的主要分量。EMD分解过程如下：①寻找原始信号极值点；②用包络线对极值点进行拟合，对极大值进行上包络拟合，对极小值进行下包络拟合；③计算包络线均值；④计算中间信号，中间信号为原始信号与均值的差值；⑤判断中间信号是否满足IMF分量条件，若不满足则回到第1步；⑥满足分量条件的中间信号则作为原始信号重复以上步骤继续进行分解。

3　综合指标

样本熵是衡量时间序列复杂度的参数，其对数据长度依赖性小，其数值越大，代表时间复杂度越高。样本熵的计算过程如下：

1）对含有N个数据的$\boldsymbol{X}(i)$序列来说，重构维数为 m 的向量：

$$\boldsymbol{X}(i)=[x(i),x(i+1),\cdots,x(i+m-1)] \tag{11}$$

2）$\boldsymbol{X}(i)$和$\boldsymbol{X}(j)$之间的距离 d 为

$$d=\max|\boldsymbol{X}(i+k)-\boldsymbol{X}(j+k)| \tag{12}$$

3）统计$\boldsymbol{X}(i)$和$\boldsymbol{X}(j)$距离小于容限值 r 的数量为 B_i，其与距离总数φ比值记为 $B_i^m(r)$：

$$\varphi=N-m-1 \tag{13}$$

$$B_i^m(r)=B_i/\varphi \tag{14}$$

4）$A^m(r)$为$B_i^m(r)$的平均值，表示相似容限匹配 m点的概率：

$$A^m(r)=\sum_{i=1}^{N-m}B_i^m(r)\Big/(N-m) \tag{15}$$

5）重复以上步骤，维度增加为$m+1$，即可得到 $A^{m+1}(r)$，根据式(11)～式(15)可得样本熵：

$$\theta(m,r,N)=-\ln\left(A^m(r)/A^{m+1}(r)\right) \tag{16}$$

峭度指标对于瞬时冲击很敏感，但其只考虑冲击的强弱性而不关注冲击的周期性[21]，而滚动轴承故障信号一大特点就是冲击具有周期性，这就影响该指标进行故障诊断的有效性。相关系数可以表示分量与原始信号的相关性，但是区分程度不高，对处于边界处的信号不能准确判断，且容易受到噪声影响。样本熵可以描述时间复杂度，其较少依赖数据长度，对噪声及其他干扰的抵抗性较高，同时样本熵反映周期性冲击脉冲。样本熵越小，周期性脉冲越显著，其值越小，说明涵盖有效信息越多。由此可知，这 3 个参数单独做评价指标时较难完整地反映故障特征信息，本文综合利用三者各自的优点，提出一种综合指标 P，具体为

$$P_i=K_iq_i/\hat{\theta}_i \tag{17}$$

式中：q_i为相关系数；$\hat{\theta}_i$为样本熵；i为第i个分量。

综合指标 P 不仅保证了分量与原始信号的相关性，而且还可以反映信号的周期性脉冲及冲击成分程度，可以更好地保留故障信息。综合指标 P 不仅克服了峭度指标只关注冲击强弱而忽略周期性的问题，而且还解决了相关系数容易受噪声干扰的问题。P 值越大，代表故障特征越丰富，所以选取 P 值最大的分量作为重构信号。

4　基于 SSA-IWT-EMD 的故障诊断

本文提出基于 SSA-IWT-EMD 的滚动轴承故障诊断方法。该方法使用 SSA 对改进小波阈值函数中的多个参数进行寻优，寻求其全局最优解，设置种群规模，按照麻雀行为进行迭代，其中一个极值为个体适应度，另一个极值为全局适应度，通过对比二者大小，搜索出最优的调节因子及阈值。随后对降噪后的信号进行 EMD 分解，通过计算分解分量的综合指标 P，筛选重要分量进行组合，并对重组信号进行包络谱分析，对比理论故障频率，得出滚动轴承故障诊断结果。

基于 SSA-IWT-EMD 的滚动轴承故障诊断流程图如图 2 所示。

基于 SSA-IWT-EMD 的滚动轴承故障诊断方法具体步骤如下：

步骤 1　初始 SSA 中的各个参数，其中包括：迭代次数、麻雀规模及参数的上下限范围，为阈值及调节因子的寻优做准备。

步骤 2　按照式(10)计算初始适应度值，通过迭代寻找最优解及最优位置。

步骤 3　计算警报值，根据该指标与安全值的大小，引用式(6)更新搜索者位置。

步骤 4　应用式(7)更替跟随者的位置信息。

步骤 5　采用式(8)做出反捕捉行为，了解到周围存在捕食者，麻雀会向种群中其他个体靠近，从而更新自己的位置。

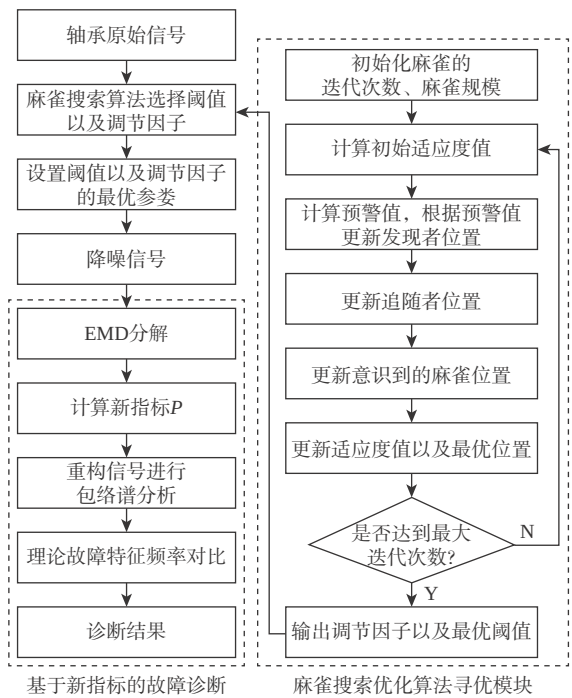


图 2　基于 SSA-IWT-EMD 的滚动轴承故障诊断流程

Fig. 2　Flow chart of fault diagnosis of rolling bearings based on SSA-IWT-EMD

步骤 6 计算个体适应度与步骤 5 最优解进行对比，选取最优解。

步骤 7 判断是否满足迭代结束条件。如果满足则输出最优解及最优位置；如不满足，则进行步骤 2，继续进行迭代。

步骤 8 将迭代寻优得到的调节因子及阈值代入改进小波阈值函数中，对原始信号进行降噪，得到处理后的滚动轴承信号。

步骤 9 对处理后的滚动轴承信号进行 EMD 分解，得到的 IMF 分量，通过综合指标进行重构，最终得到重构滚动轴承信号，观察最终的信号包络谱图得出故障诊断结果。

5 结果与分析

5.1 仿真信号

为验证本文方法的有效性，运用仿真信号 $x(l)$：

$$\begin{cases} x(l)=s(l)+n(l)=\sum_i A_i h(l-iT)+n(l) \\ h(l)=\exp(-Cl)\cos(2\pi f_n l) \\ A_i=1+A_0\cos(2\pi f_r l) \end{cases} \tag{18}$$

式中：$s(l)$为冲击响应信号；A_0为幅值，值为 0.3；f_r为转频，值为 30 Hz；C衰减系数，值为 700；f_n为共振频率，值为 4 kHz；$n(l)$为白噪声，信噪比为 4 dB；T为周期，值为 1/120；l为采样时刻。

冲击信号及仿真信号如图 3 和图 4 所示。通过理论计算可知该信号理论故障频率值为 120 Hz。图 4 显示为仿真信号，该信号掺杂噪声，图 5 为原始仿真信号包络谱，由图 5 可知，对应倍频被白噪声所掩盖，不能准确实现轴承的故障诊断，故需对信号进行处理。

图 6 为经过 SSA-IWT 的仿真信号时域图。由图可知，经过麻雀搜索算法处理后的信号冲击特征明显，滤除了噪声，能够对原始信号实现有效去噪。随后对处理信号进行 EMD 分解，如图 7 所示。

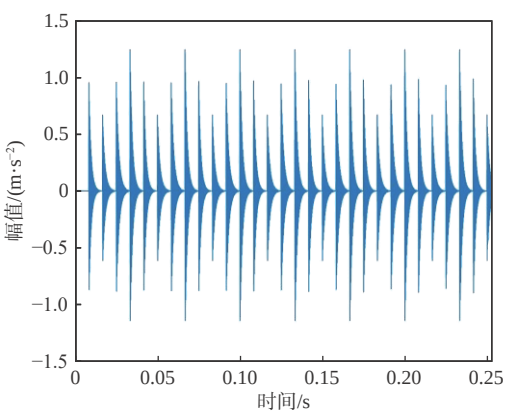


图 3 冲击信号

Fig. 3 Impact signal

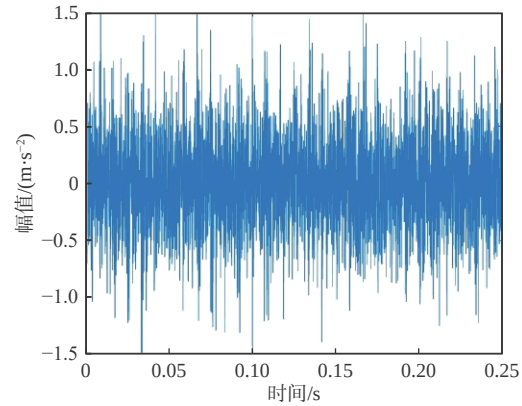


图 4 仿真信号

Fig. 4 Simulation signal

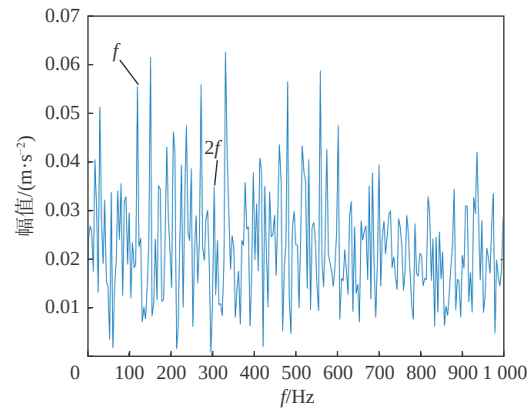


图 5 原始仿真信号包络谱

Fig. 5 Envelope spectrum of original simulation signal

由图 7 可以看出，信号被分解为 9 个 IMF 分量，对分量分别进行综合指标 P 的计算，计算结果如表 1 所示。选取综合指标 P 值最大的 IMF1 进行信号重构并进行包络谱分析，分析结果如图 8 所示。

故障频率幅值大小代表故障发生的概率，幅值越大说明故障发生概率越大。当滚动轴承发生故障时，频谱图上就会出现与之对应的峰值。但由于实际装配安装、转速变化等因素，导致实际故障特征频率与理论值有所差别。多数情况下，可对振动

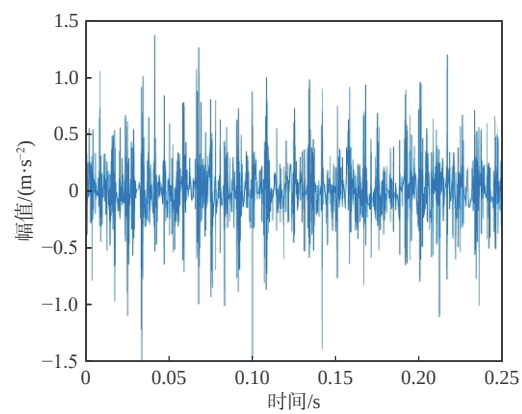


图 6 SSA-IWT 优化后仿真信号

Fig. 6 Simulation signal optimized by SSA-IWT

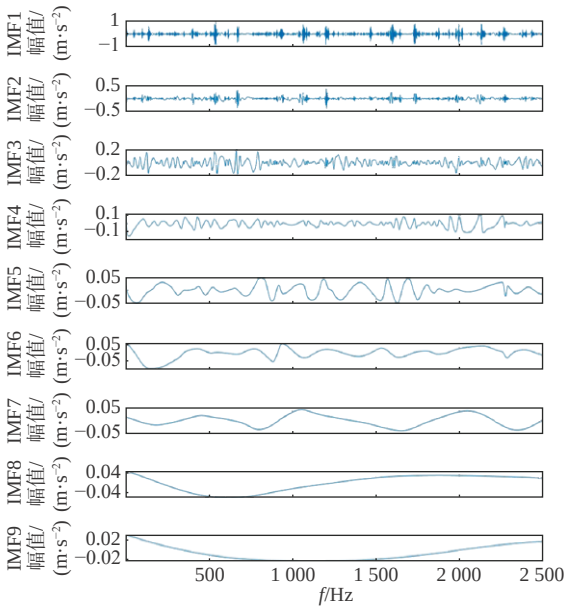


图 7　仿真信号 EMD 分解图

Fig. 7　EMD diagram of simulation signal

表 1　仿真信号 EMD 分解分量参数

Table 1　Parameters of EMD components of simulation signal

分量	综合指标P	分量	综合指标P	分量	综合指标P
IMF1	15.517	IMF4	4.091	IMF7	2.011
IMF2	7.913	IMF5	4.798	IMF8	2.067
IMF3	9.848	IMF6	1.888	IMF9	2.307

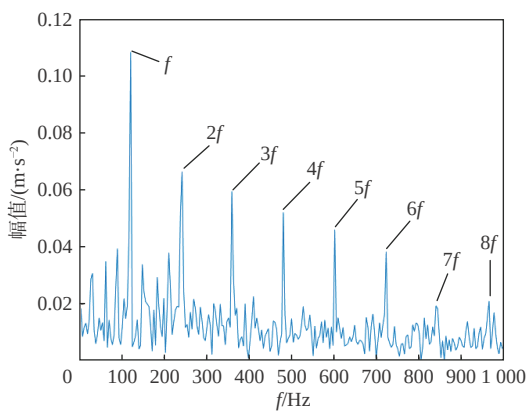


图 8　SSA-IWT-EMD 仿真信号包络谱

Fig. 8　Envelope spectrum of SSA-IWT-EMD simulation signal

信号进行包络谱解调，以特征频率作为基频，寻求二、三倍频等成分，倍频的出现同样可以判断滚动轴承的故障类型[22]。

由图 8 可看出，一阶频率峰值数值为 121.4 Hz，与故障特征频率理论值 120 Hz 相近。在图 8 中不仅能明显观察到 1～5 倍频的特征频率成分，而且可以在图中找到 6～8 倍频成分，故可判断该轴承发生了故障。通过观察处理后的滚动轴承信号包络谱图，可以发现本文方法能够实现阈值的最优选取，实现良好降噪，同时运用综合指标 P 对分量进行选取重构更能突出信号特征，结合包络谱中峰值频率和倍频分布可以有效实现轴承的故障诊断。

5.2　试验轴承信号

为验证本文方法的可靠性，本文采用 SQI 公司研发的机械故障模拟（mechanical fault simulation, MFS）试验平台对本文方法进行验证，试验平台如图 9 所示。主要由调速器、电机、加速度传感器、信号采集器及一对轴承组成，轴承分别固定在左、右两端轴承架上，其中一个为健康轴承，另一个为故障轴承。试验平台中间安装 5.1 kg 的转盘为系统提供 50 N 负载。

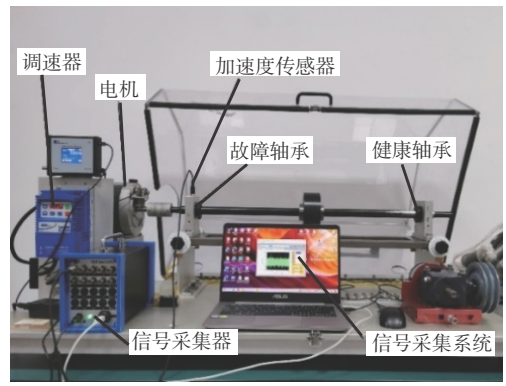


图 9　MFS 机械故障试验平台

Fig. 9　MFS test platform for mechanical fault

试验所选用的轴承型号为 ER-16k，该轴承具体参数如表 2 所示。轴承缺陷均是由激光刻蚀技术加工所得，在采样频率为 15.36 kHz 及电机转速 1 200 r/min 条件下分别采集内圈、外圈及滚动体单一类型故障信号，故障直径为 1.2 mm。通过表 2 中相应参数可计算出内圈、外圈及滚动体理论故障特征频率分别为 108.6, 71.4, 92.6 Hz。

表 2　ER-6k 轴承参数

Table 2　Parameters of ER-6k bearing

轴承节径/mm	滚珠直径/mm	滚动体个数/个	故障直径/mm
38.70	8	9	1.2

随机挑选 2 组试验信号用以故障诊断。图 10 为试验轴承原始信号 1，由图可知，原始振动信号 1 中包含大量干扰噪声。图 11 为原始轴承信号 1 的包络谱图，其中，f_b 为滚动体故障特征频率。可以看出，倍频成分被淹没在噪声之中，不能实现对滚动轴承的故障诊断。

图 12 为 SSA-IWT 优化后轴承信号 1 时域图。通过对比可知 SSA-IWT 优化后的振动信号 1 去噪效果明显同时冲击特性突出。对去噪后的信号进行 EMD 分解，EMD 分解图如图 13 所示。

图 13 可知，去噪信号被分解为 10 个 IMF 分量，分别计算各分量综合评价指标 P 的数值，计算结果如表 3 所示。由表 3 可知在各分量中 IMF3 分

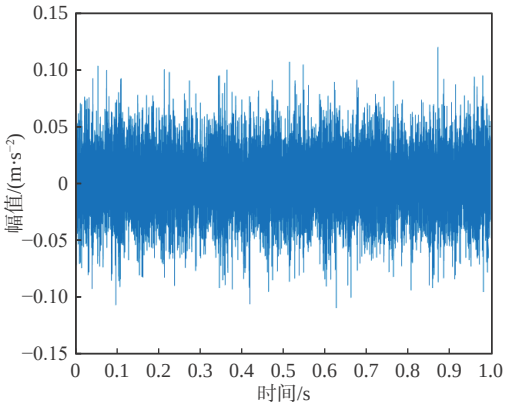


图 10 试验轴承原始信号 1

Fig. 10 Original signal 1 of test bearing

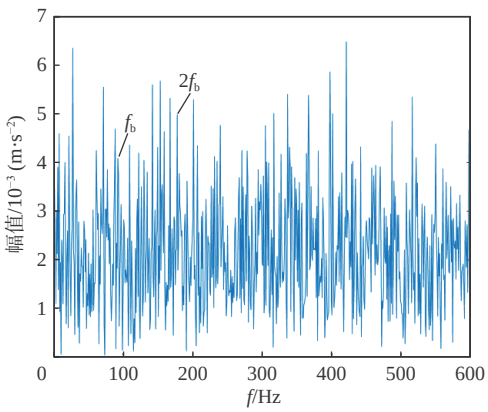


图 11 原始试验轴承信号 1 包络谱

Fig. 11 Envelope spectrum of original signal 1 of test bearing

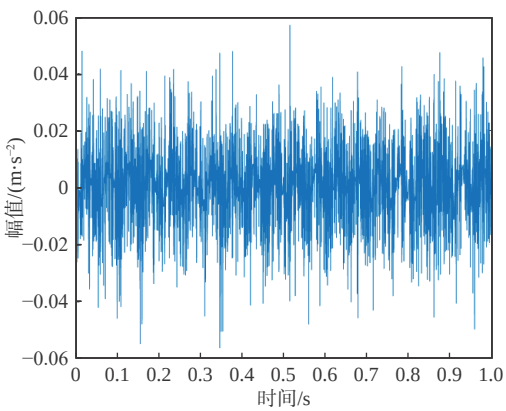


图 12 SSA-IWT 优化后轴承信号 1 时域图

Fig. 12 Time domain diagram of signal 1 of bearing optimized by SSA-IWT

量综合评价指标 P 最大，故选择 IMF3 作为重构信号进行包络谱分析，分析结果如图 14 所示。

由图 14 可以看出，一阶倍频成分大小为 93.7 Hz，该数值与滚动体理论故障特征频率92.6 Hz 相近，而且从包络谱中还可观察到 2～5 倍频成分，故可判断该轴承故障类型为滚动体故障，同时也验证本文方法的可靠性。

图 15 为试验轴承原始数据信号 2，可知该信号

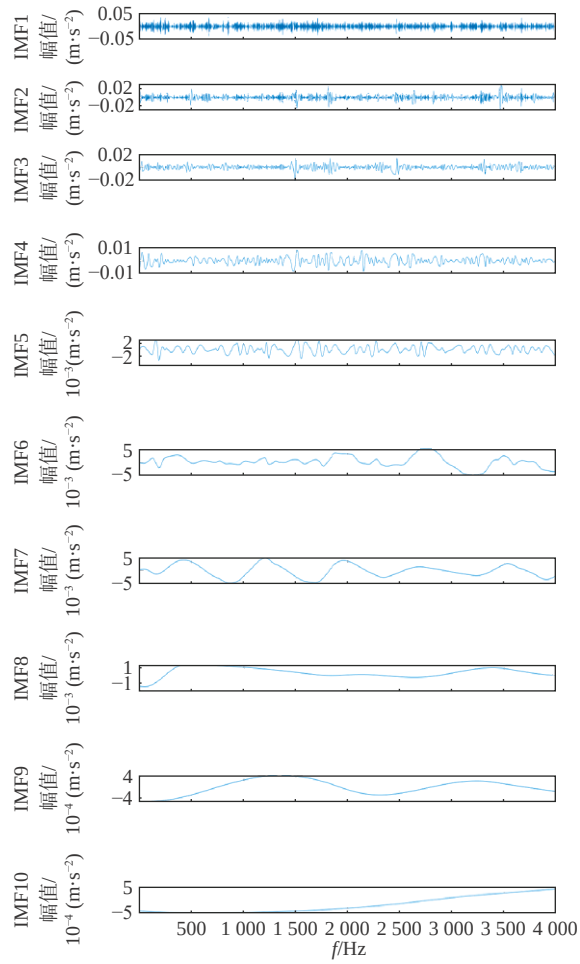


图 13 试验轴承信号 1 EMD 分解图

Fig. 13 EMD diagram of signal 1 of test bearing

表 3 试验轴承信号 1 EMD 分解分量参数

Table 3 Parameters of EMD components of signal 1 of test bearing

分量	综合指标P	分量	综合指标P
IMF1	3.410	IMF6	2.981
IMF2	3.614	IMF7	1.431
IMF3	4.319	IMF8	1.643
IMF4	2.644	IMF9	1.564
IMF5	2.245	IMF10	0.941

中掺杂噪声。图 16 为原始试验轴承信号 2 包络谱，其中，f_i 为内圈故障特征频率，可以看出，1 倍频较突出，而 2 倍频淹没在噪声中，从包络谱图中无法获得周期性故障特征，无法实现有效故障诊断。

图 17 为 SSA-IWT 优化后的轴承数据信号，对比图 15 可以看出，经麻雀搜索算法处理后的信号冲击特征更加明显。将处理后的信号进行 EMD 分解，图 18 为 EMD 分解图。

由图 18 可以看出，信号被分解为 9 个 IMF 分量，对这 9 个分量进行指标计算，计算结果如表 4 所示。由表 4 可知 IMF1 的综合指标 P 数值最高，数值越高代表其包含故障信息越多，所以选取该分

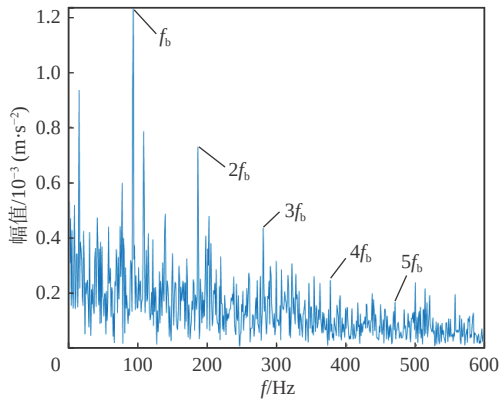


图 14　试验轴承信号 1 包络谱

Fig. 14　Envelope spectrum of signal 1 of test bearing

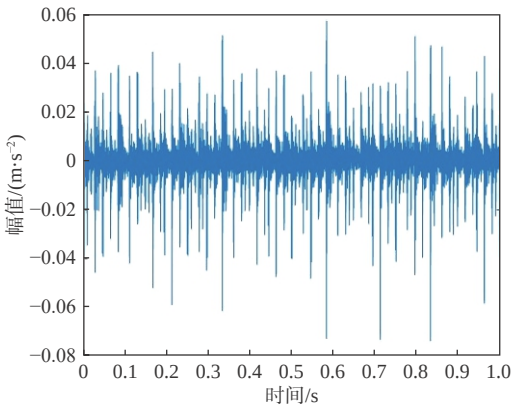


图 15　试验轴承原始信号 2

Fig. 15　Original signal 2 of test bearing

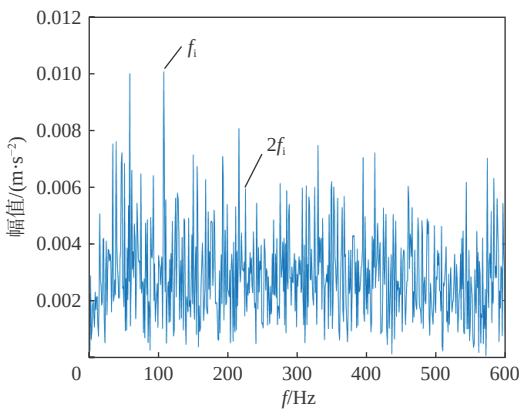


图 16　原始试验轴承信号 2 包络谱

Fig. 16　Envelope spectrum of signal 2 of test bearing

量进行重组。对重构试验轴承信号 2 进行包络谱分析, 图 19 为经过 SSA-IWT-EMD 方法处理后的试验轴承信号 2 包络谱图。

由图 19 可知, 1 倍频成分为 108.5 Hz, 与计算的轴承内圈故障特征频率理论值 108.6 Hz 相近, 同时在图中可寻找到 1～5 倍频成分, 其中 1～3 倍频幅值也较为突出, 故可判断该故障为内圈故障。通过图 19 可知, 本文方法更能突出信号的故障频率, 更可靠地实现对滚动轴承的故障诊断。

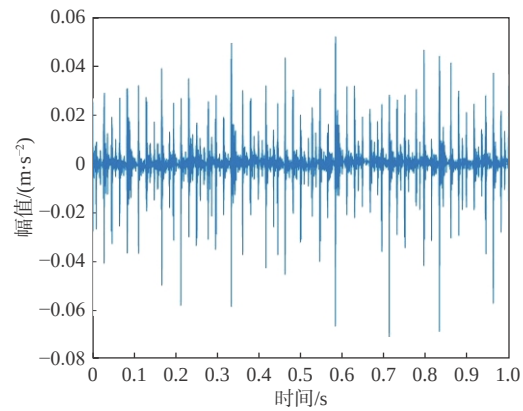


图 17　SSA-IWT 优化后信号 2

Fig. 17　Signal 2 optimized by SSA-IWT

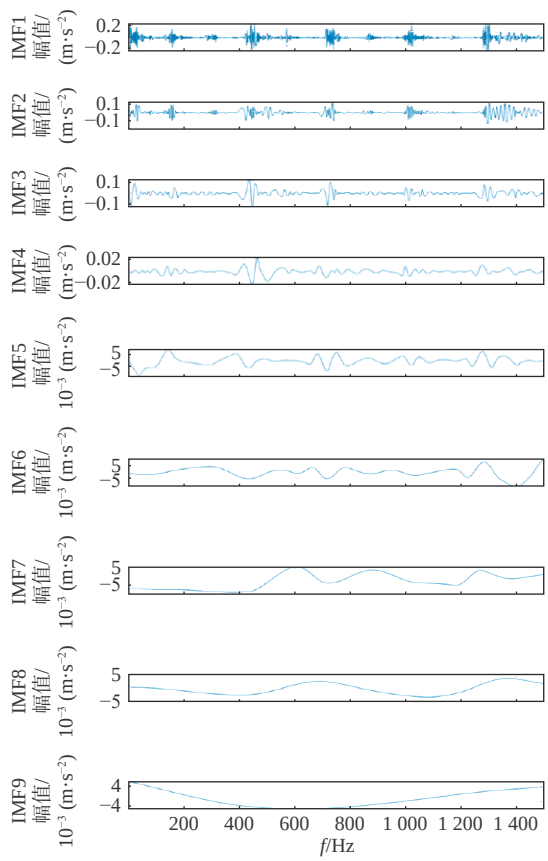


图 18　试验轴承信号 2 EMD 分解图

Fig. 18　EMD diagram of signal 2 of test bearing

表 4　试验轴承信号 2 EMD 分解分量参数

Table 4　Parameters of EMD components of signal 2 of test bearing

分量	综合指标P	分量	综合指标P	分量	综合指标P
IMF1	30.614	IMF4	4.906	IMF7	2.534
IMF2	21.787	IMF5	5.401	IMF8	2.407
IMF3	4.825	IMF6	3.149	IMF9	1.557

为进一步说明综合指标 P 的优越性, 采用传统指标—峭度对 EMD 分量进行分析, 所得分量峭度参数如表 5 所示。选取峭度值最大的 IMF4 对信号

进行重构，图 20 为重构信号的包络谱图。

与图 19 进行对比，可以看出本文方法处理过的试验轴承信号的内圈故障特征的 3 倍频清晰可见，而且还可以找出 4 倍频及 5 倍频的成分，采用峭度准则选

取分量的方法仅前 2 倍频清晰，后面倍频成分淹没在噪声之中，而且包络谱整体幅值较小，故无法根据包络谱中倍频特性实现精准故障诊断，同时可见综合指标 P 在分量选取上具有明显优势。

为进一步说明本文方法的优越性，选用遗传算法优化的 Morlet 小波滤波与经验模态分解相结合的故障诊断方法[23] 作为对比方案。首先，运用遗传算法优化 Morlet 小波滤波器相应参数实现对轴承

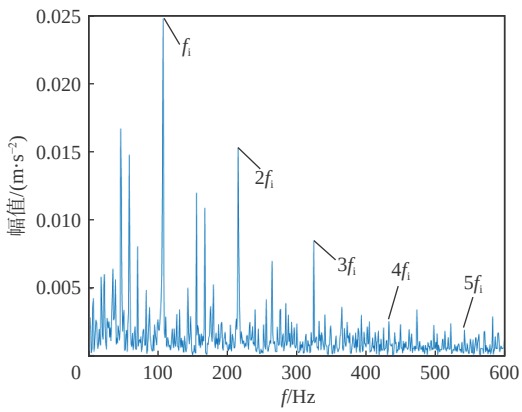


图 19 SSA-IWT-EMD 试验轴承信号 2 包络谱

Fig. 19 Envelope spectrum of signal 2 of bearing in SSA-IWT-EMD test

表 5 试验轴承信号 2EMD 分解分量峭度参数

Table 5 Kurtosis parameters of EMD components of signal 2 of test bearing

分量	峭度	分量	峭度	分量	峭度
IMF1	16.225	IMF4	16.772	IMF7	4.629
IMF2	14.114	IMF5	4.781	IMF8	2.686
IMF3	10.170	IMF6	5.954	IMF9	1.979

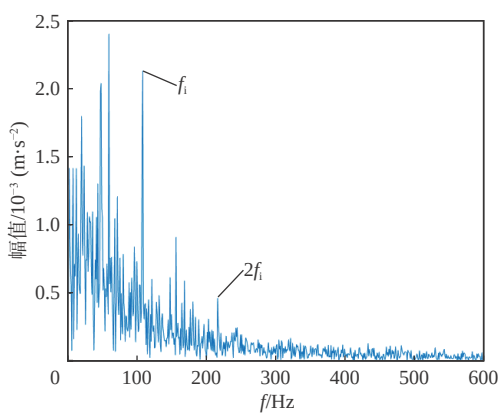


图 20 峭度指标试验轴承信号 2 包络谱

Fig. 20 Envelope spectrum of signal 2 of bearing in kurtosis indicator test

信号 2 的滤波，滤波后的振动信号如图 21 所示。通过对比图 21 与图 17，可以发现本文方法去噪效果更好，具有更好的冲击性。随后将滤波信号进行 EMD 分解，EMD 分解图如图 22 所示。

以峭度作为筛选分量的重要依据，认为峭度值大于 3 的 IMF 分量包含较多冲击成分，将其作为有效分量予以保留并重构，其余分量则进行舍弃，各分量峭度参数如表 6 所示。由表 6 可知，前 4 分量

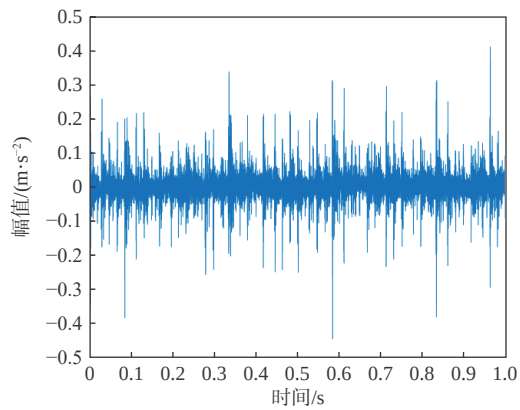


图 21 文献 [23] 方法滤波后的实验轴承信号 2

Fig. 21 Signal 2 of test bearing filtered by literature [23] method

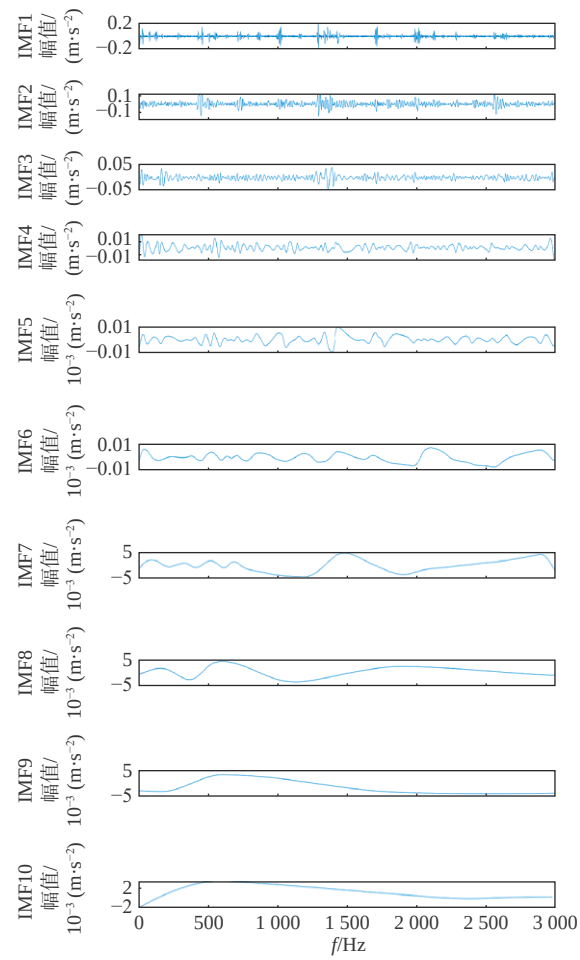


图 22 文献 [23] 方法 EMD 分解图

Fig. 22 EMD diagram of literature [23] method

峭度值大于 3，故将其进行重构并进行包络谱分析，分析结果如图 23 所示。由图 23 可看出，包络谱图中前 2 倍频较清晰，3 倍频已经被淹没在噪声中，与图 19 进行对比可知，本文方法降噪明显效果更好，同时包络谱整体幅值更大，倍频成分更明显，更能凸显信号的周期性和冲击特性，而幅值大小及倍频分布规律正是判别故障类别的重要标准，故验证出本文方法具有更强的故障诊断能力。

表 6　EMD 分解分量峭度参数[23]

Table 6　Kurtosis parameters of EMD components[23]

分量	峭度	分量	峭度
IMF1	18.152	IMF6	2.616
IMF2	7.099	IMF7	2.479
IMF3	6.743	IMF8	1.649
IMF4	3.680	IMF9	2.202
IMF5	2.808	IMF10	1.840

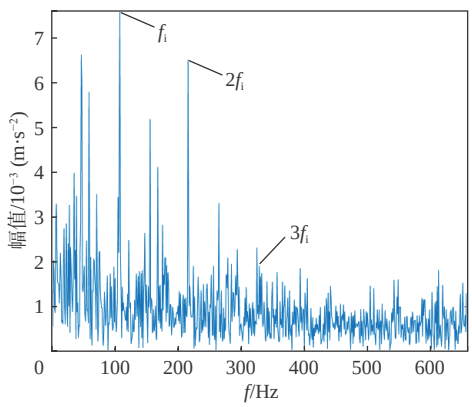


图 23　文献 [23] 方法实际轴承信号 2 包络谱

Fig. 23　Envelope spectrum of signal 2 of actual bearing by literature [23] method

6　结　论

针对滚动轴承信号中包含大量干扰噪声，严重影响后续故障诊断工作的问题，本文提出一种基于 SSA-IWT-EMD 的滚动轴承故障诊断方法。通过仿真信号及试验轴承信号验证所提方法的有效性并得出以下结论。

1）基于原始阈值提出了改进小波阈值函数，解决了求取最优阈值的问题，达到了良好的去噪效果，同时提高了轴承信号的冲击特性。

2）将结合峭度、相关系数及样本熵的综合指标 P 运用到 EMD 分量筛选上，保留了信号中的有用信息，克服单一指标可能造成信息缺失的问题，提取出更加充分的故障特征信息。

3）对比运用峭度单一指标进行分量选取重构后的信号包络谱图及遗传算法优化 Morlet 小波与 EMD 相结合的故障诊断方法，试验结果表明本文方法包络谱更加清晰，倍频成分更明显，幅值更高，说明综合指标 P 及本文方法的有效性。同时，包络谱中倍频及幅值的优越性，说明本文方法能够更好地实现滚动轴承故障诊断。

参考文献（References）

[1] RAI A, UPADHYAY S H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings[J]. Tribology International, 2016, 96: 289-306.

[2] 葛兴来, 邹丹. 多层降噪技术及 Hilbert 变换的轴承故障诊断方法[J]. 电机与控制学报, 2020, 24(8): 9-17.
GE X L, ZOU D. Bearing fault diagnosis method of multi-layer denoising technologies and Hilbert transformation[J]. Electric Machines and Control, 2020, 24(8): 9-17(in Chinese).

[3] SHAO R P, HU W T, LI J. Multi-fault feature extraction and diagnosis of gear transmission system using time-frequency analysis and wavelet threshold de-noising based on EMD[J]. Shock and Vibration, 2013, 20(4): 763-780.

[4] 王普, 李天垚, 高学金, 等. 分层自适应小波阈值轴承故障信号降噪方法[J]. 振动工程学报, 2019, 32(3): 548-556.
WANG P, LI T Y, GAO X J, et al. Bearing fault signal denoising method of hierarchical adaptive wavelet threshold function[J]. Journal of Vibration Engineering, 2019, 32(3): 548-556(in Chinese).

[5] CUI H Y, QIAO Y Y, YIN Y M, et al. An investigation on early bearing fault diagnosis based on wavelet transform and sparse component analysis[J]. Structural Health Monitoring, 2017, 16(1): 39-49.

[6] CHEN B J, SHEN B M, CHEN F F, et al. Fault diagnosis method based on integration of RSSD and wavelet transform to rolling bearing[J]. Measurement, 2019, 131: 400-411.

[7] 崔玲丽, 高立新, 殷海晨, 等. 基于第二代小波的复合故障诊断方法研究[J]. 中国机械工程, 2009, 20(4): 442-446.
CUI L L, GAO L X, YIN H C, et al. Research on composite fault diagnosis method based on the second generation wavelet[J]. China Mechanical Engineering, 2009, 20(4): 442-446(in Chinese).

[8] 王志刚, 李友荣, 李方. 基于谐波小波分析的故障诊断方法研究[J]. 振动与冲击, 2006, 25(2): 125-128.
WANG Z G, LI Y R, LI F. Fault diagnosis method based on harmonic wavelet analysis[J]. Journal of Vibration and Shock, 2006, 25(2): 125-128(in Chinese).

[9] 张晓峰, 李功燕. 应用小波分析提取故障诊断信号的特定频段[J]. 振动与冲击, 2004, 23(4): 47-50.
ZHANG X F, LI G Y. Extraction of fault diagnosis signal's frequency band with wavelet analysis[J]. Journal of Vibration and Shock, 2004, 23(4): 47-50(in Chinese).

[10] 邓飞跃, 强亚文, 杨绍普, 等. 一种自适应频率窗经验小波变换的滚动轴承故障诊断方法[J]. 西安交通大学学报, 2018, 52(8): 22-29.
DENG F Y, QIANG Y W, YANG S P, et al. A fault diagnosis method of rolling element bearings with adaptive frequency window empirical wavelet transform[J]. Journal of Xi'an Jiaotong University, 2018, 52(8): 22-29(in Chinese).

[11] TUERXUN W, XU C, GUO H Y, et al. Fault diagnosis of wind turbines based on a support vector machine optimized by the sparrow search algorithm[J]. IEEE Access, 2021, 9: 69307-69315.

[12] REN Y, ZHANG L L, CHEN J T, et al. Noise reduction study of pressure pulsation in pumped storage units based on sparrow optimi-

zation VMD combined with SVD[J]. Energies, 2022, 15(6): 2073.

[13] HUANG N E, SHEN Z, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proceedings of the Royal Society of London Series A, 1998, 454(1971): 903-998.

[14] KESHTAN M N, NOURI KHAJAVI M. Bearings fault diagnosis using vibrational signal analysis by EMD method[J]. Research in Nondestructive Evaluation, 2016, 27(3): 155-174.

[15] 丛晓妍, 王增才, 王保平, 等. 基于 EMD 与峭度滤波的煤岩界面识别[J]. 振动、测试与诊断, 2015, 35(5): 950-954.
CONG X Y, WANG Z C, WANG B P, et al. Identification of coal-rock interface based on EMD and kurtosis filtering[J]. Journal of Vibration, Measurement & Diagnosis, 2015, 35(5): 950-954(in Chinese).

[16] 郭俊超, 甄冬, 孟召宗, 等. 基于 WAEEMD 和 MSB 的滚动轴承故障特征提取[J]. 中国机械工程, 2021, 32(15): 1793-1800.
GUO J C, ZHEN D, MENG Z Z, et al. Feature extraction of rolling bearings based on WAEEMD and MSB[J]. China Mechanical Engineering, 2021, 32(15): 1793-1800(in Chinese).

[17] DENG C X, CHEN X X, LI S Q, et al. The improved wavelet threshold function and its application[J]. International Journal of Signal Processing, Image Processing and Pattern Recognition, 2016, 9(7): 79-92.

[18] ZHANG Y, DING W F, PAN Z F, et al. Improved wavelet threshold for image de-noising[J]. Frontiers in Neuroscience, 2019, 13: 39.

[19] MO F, MO Q, CHEN Y Y, et al. Wavelet quant, an improved quantification software based on wavelet signal threshold de-noising for labeled quantitative proteomic analysis[J]. BMC Bioinformatics, 2010, 11: 219.

[20] XUE J K, SHEN B. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. Systems Science & Control Engineering, 2020, 8(1): 22-34.

[21] 张龙, 熊国良, 黄文艺. 复小波共振解调频带优化方法和新指标[J]. 机械工程学报, 2015, 51(3): 129-138.
ZHANG L, XIONG G L, HUANG W Y. New procedure and index for the parameter optimization of complex wavelet based resonance demodulation[J]. Journal of Mechanical Engineering, 2015, 51(3): 129-138(in Chinese).

[22] 齐咏生, 樊佶, 李永亭, 等. 一种改进的解卷积算法及其在滚动轴承复合故障诊断中的应用[J]. 振动与冲击, 2020, 39(21): 140-150.
QI Y S, FAN J, LI Y T, et al. An improved deconvolution algorithm and its application in compound fault diagnosis of rolling bearings[J]. Journal of Vibration and Shock, 2020, 39(21): 140-150(in Chinese).

[23] HAN X M, XU J, SONG S N, et al. Crack fault diagnosis of vibration exciter rolling bearing based on genetic algorithm–optimized Morlet wavelet filter and empirical mode decomposition[J]. International Journal of Distributed Sensor Networks, 2022, 18(8): 155013292211145.

Fault diagnosis method of rolling bearings based on SSA-IWT-EMD

LEI Chunli[1], JIAO Mengxuan[2], FAN Gaofeng[3], LIU Shichao[2], XUE Linlin[1], LI Jianhua[1,*]

(1. School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, China;
2. Ineos Sinopec Tianjin Petrochemicals Limited, Tianjin 300280, China;
3. Yunnan Wenshan Aluminum Co., Ltd., Wenshan 663000, China)

Abstract: Wavelet threshold denoising is insufficient, and feature frequency extraction of empirical mode decomposition (EMD) is unclear. To address these issues, a fault diagnosis method of rolling bearings based on sparrow search algorithm - improved wavelet threshold-EMD (SSA-IWT-EMD) was proposed. Firstly, two adjustment factors were introduced, and an IWT function was presented to overcome the shortcomings of traditional soft and hard thresholds. The SSA was used to globally optimize the parameters of the IWT to reduce the noise of rolling bearing signals. Secondly, a comprehensive index *P* was put forward to select and reconstruct the components generated by EMD, so as to highlight the fault feature information of the signals. Finally, the fault diagnosis of bearings was realized by envelope spectrum analysis. The simulation and experimental results verified the effectiveness of the proposed method. At the same time, the comparison with the single index component selection method and the literature method indicated that the comprehensive index *P* and the method proposed in this paper had stronger denoising ability and feature extraction ability, and the envelope spectrum amplitude and frequency doubling component were more obvious, which could better realize the fault diagnosis of rolling bearings.

Keywords: rolling bearing; improved threshold; comprehensive index; empirical mode decomposition; fault diagnosis

Received: 2023-04-12; **Accepted**: 2023-06-30; **Published Online**: 2023-07-10 16:05

URL: link.cnki.net/urlid/11.2625.V.20230710.1537.002

Foundation items: National Natural Science Foundation of China (51465035); Natural Science Foundation of Gansu Province, China (20JR5RA466); Hongliu First-class Disciplines Development Program of Lanzhou University of Technology

* **Corresponding author**. E-mail: li_jh@vip.sina.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0216

直驱泵阀协同式线控制动单元压力串级控制

谭草[1]，于鹏[1]，李波[1,*]，陆佳瑜[1]，任云云[2]

(1. 山东理工大学 交通与车辆工程学院，淄博 255000；2. 舜泰汽车有限公司，淄博 255000)

摘　　要：针对分布式制动及高级别自动驾驶的需求，设计一种直驱泵阀协同式线控制动单元，通过电磁直线执行器直接驱动的液压泵与主动阀协调实现轮缸压力调节。为实现压力的精确控制，建立面向控制的系统动力学模型，提出一种基于外环轮缸压力控制和内环直驱泵活塞位置控制的串级控制方法。外环设计了滑模控制器提高响应速度，内环设计了自适应积分鲁棒控制器以减小参数不确定性和时变干扰的影响，证明了所提方法李雅普诺夫稳定。结果表明：直驱泵阀协同式线控制动单元为分布式制动提供一种可行的新方案，所提方法可以进一步提高制动压力的调节速度与控制精度。

关 键 词：线控制动；液压力控制；串级控制；滑模控制；自适应积分鲁棒控制

中图分类号：U463.5; U469.72

文献标志码：A　　**文章编号**：1001-5965(2025)04-1163-09

随着车辆底盘向着线控底盘、滑板底盘方向发展，对分布式制动系统的需求日益强烈；同时，高级别自动驾驶对线控系统的响应速度、控制精度提出了更高的需求[1-2]。

电子液压式线控制动系统凭借响应迅速、功率密度高、不改变制动器结构等优势得到了广泛的青睐[3-4]。Pan 等提出一种紧凑型电子液压式线控制动系统，不需要高动力电机，且易于实现备份[5]。余卓平等设计一种双动力源电子液压制动系统，该系统可对制动主缸液压力和踏板感觉进行独立主动控制，实现了踏板行程与液压力的解耦[6]。胡东海和何仁提出基于安全特性的电机泵式电子液压制动系统匹配设计方法[7]。刘兆勇等设计了一种集成式线控液压制动系统，实现该制动系统的高冗余性和4个轮缸的独立控制[8]。各具特点的电子液压制动系统被提出，推动了线控制动技术的发展，然而难以直接用于分布式制动系统[9]。于良耀等设计了一种分布式电液线控制动系统，通过运动转化机构将旋转电机的运动变为直线运动进而驱动液压轮缸[10]。Gong 等设计的电液制动单元采用直线电机驱动液压活塞，通过不等径的液压活塞与轮缸活塞放大电机驱动力[11]。由上可知，现有的电子液压制动系统往往基于成熟的液压制动系统改进，仍然保留了复杂的液压管路；分布式线控系统往往需要将整个制动单元集成到制动钳上，恶化了轮边空间与簧下质量。

线控制动系统制动压力快速调节与精确控制是实现高级别制动驾驶的重要基础，也是目前的研究热点。针对不同线控系统的特点，文献 [12] 基于模型设计了不同的控制方法。熊璐等设计基于LuGre 摩擦模型前馈补偿和死区补偿的联合控制方法，在考虑系统关键非线性特征的基础上，精确控制主缸液压力[13]。石琴等利用径向基网络逼近连续函数特性，对与系统状态量相关的非线性摩擦力进行估计，作为反步控制器的补偿，实现线控制动系统液压精确控制[14]。张俊智等针对线控系统中

收稿日期：2023-04-27；**录用日期**：2023-06-06；**网络出版时间**：2023-06-15 11:44

网络出版地址：link.cnki.net/urlid/11.2625.V.20230614.1457.001

基金项目：国家自然科学基金 (51975341)；山东省优秀青年人才基金 (ZR2022YQ51)；山东省的重大科技创新工程项目 (2021CXGC010703)；山东省高等学校“青创科技计划”团队项目 (2019KJB027,2022KJ232)；山东省青年科技人才托举工程项目 (SDAST2021qt20)

* **通信作者**. E-mail：libo@sdut.edu.cn

引用格式：谭草，于鹏，李波，等. 直驱泵阀协同式线控制动单元压力串级控制 [J]. 北京航空航天大学学报，2025，51（4）：1163-1171.

TAN C，YU P，LI B，et al. Pressure cascade control of brake-by-wire unit based on direct drive pump-valve cooperative [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1163-1171（in Chinese）.

的摩擦现象，建立了一种更适用的摩擦模型，并提出了一种针对制动系统主缸压力控制的高精度压力跟踪控制方法[15]。李波等建立了基于LuGre摩擦模型的直驱阀模型，通过串级控制实现了直驱阀控线控系统的压力精确控制[16]。对线控系统液压力控制方法的研究主要集中在针对系统摩擦等典型非线性补偿方面，而制动系统自身的参数不确定性及面临的时变干扰也需要进一步研究[17-19]。

综上所述，分布式制动系统需要一种降低液压管路复杂程度且不恶化轮边空间与簧下质量的制动单元；而高精度的制动液压力控制方法，则需要进一步考虑系统参数不确定性与时变干扰的影响。鉴于此，本文设计一种直驱泵阀协同式线控制动单元，通过电磁直线执行器直接驱动的液压泵与主动阀协调实现轮缸压力调节。进而设计一种基于外环轮缸压力控制和内环直驱泵柱塞位置控制的串级控制方法。通过实验验证了本文设计的制动单元和串级控制器的有效性。

1 方案设计与工作原理

针对分布式制动及高级别自动驾驶的需求，本文设计了一种直驱泵阀协同式线控制动单元，结构示意如图1所示，主要包括电磁直线执行器、液压泵、主动阀、不改变现有结构的制动轮缸、补油油箱。电磁直线执行器、液压泵、主动阀组及补油油箱集成安装在车架上，通过油管与不改变现有结构的制动轮缸相连，有效简化制动管路，实现分布式驱动的同时不恶化轮边空间与簧下质量。其中，电磁直线执行器直接驱动的液压泵活塞往复运动，结合主动阀组的开关控制，实现制动轮缸压力的迅速调节。主动阀组由2个高速电磁阀组成，补油阀为常关阀，控制补油油箱到直驱泵之间的油路通断；保压阀为常开阀，控制直驱泵到制动轮缸之间的油路通断。设计直驱泵活塞塞直径小于轮缸活塞直径，利用不等径液压放大原理来放大直线执行器的驱动力，从而使用较小执行器获得足够的制动力。

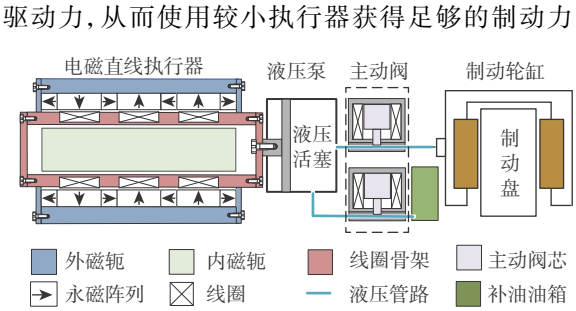


图1 直驱泵阀协同式线控制动单元原理图

Fig. 1 Principle diagram of brake-by-wire unit based on direct drive pump-valve cooperative

电磁直线执行器作为制动单元的动力来源，完成了电能到机械能的转换，采用高功率密度动圈式电磁直线执行器，其详细工作原理见文献[20]，省去旋转运动转换为直线运动的中间机械结构，提高了制动单元的响应速度。

直驱泵阀协同式线控制动单元典型的工作过程分为调压模式、保压模式与卸压模式。卸压模式中保压阀打开、补油阀关闭，在制动轮缸压力及制动轮缸复位弹簧作用下，制动轮缸内的液压油流回液压泵、直线执行器及液压泵活塞回到初始位置；若由于液压油泄漏等原因，直线执行器未回到初始位置时，保压阀关闭、补油阀打开，直线执行器驱动液压泵活塞回到初始位置。调压模式中，保压阀打开、补油阀关闭，电磁直线执行器驱动液压活塞快速排出液压泵头中液压油流入制动轮缸、推动制动轮缸活塞消除制动间隙，进而通过控制直线电机输出力快速调节制动轮缸压力。保压模式中，保压阀、补油阀关闭，直线执行器不工作，制动轮缸压力保持不变。直驱泵阀协同式线控制动单元相较于泵控制动系统，减轻直线执行器工作负担、减小保压工况下系统的工作能耗；而相较于阀控制动系统，可更加精确地控制轮缸压力，使制动更加准确和安全。

2 控制系统模型建立

2.1 电磁直线执行器建模

直驱泵阀协同式线控制动单元所采用的电磁直线执行器由内外磁轭、线圈骨架、线圈、永磁体组成[19]。电磁直线执行器与液压活塞直接相连，受到摩擦力、来自活塞的液压阻力等，其机械、磁路、电路子系统耦合的模型为

$$\begin{cases} M\dfrac{\mathrm{d}^2x}{\mathrm{d}t^2} = F_{\mathrm{m}} - F_{\mathrm{f}} - P_1S_1 - F_{\mathrm{dis}} \\ F_{\mathrm{m}} = NB_{\mathrm{e}}l_{\mathrm{e}}I = K_{\mathrm{m}}I \\ u = IR + \dfrac{\mathrm{d}I}{\mathrm{d}t}L + K_{\mathrm{e}}v \end{cases} \tag{1}$$

式中：F_{m}为通电线圈在磁场中产生的电磁力；S_1为柱塞横截面积；N为线圈总匝数；B_{e}为磁场强度；l_{e}为线圈单匝长度；I为线圈电流；K_{m}为电磁力系数；M为执行器的动质量；x为动子移动位移；F_{f}为执行器动子移动所受到的摩擦力；F_{dis}为不确定性误差和干扰；u为供电电压；R和L分别为线圈电阻和电感；K_{e}为反电势系数；v为动子移动速度；P_1为直驱泵内部压力。

为提高建模准确性，建立一种简单而有效的摩擦力表达形式，将摩擦力看为速度的静态非线性函数，具体表达式为

$$F_f = B_1\dot{x} + A_f \arctan(\beta\dot{x}) \tag{2}$$

式中: B_1 为黏性系数; A_f 为库伦摩擦系数, 通过光滑函数 arctan() 来表示传统的符号函数 sgn()。β 为常数, 设置 β 足够大, 使得该函数保持符号函数的特性并且使摩擦力的表达更加符合实际。

2.2　液压系统建模

液压泵需在高压工况下运行, 液体的可压缩性必须予以考虑。为简化直驱泵的数学模型, 假设液体在流动时无局部压力损失和沿程压力损失; 液压泵泵腔和液压活塞不发生形变; 液压泵泵腔内各处的压力相等; 整个制动单元密封良好, 无液压油泄漏。液压泵的液压力变化模型为

$$\dot{P}_1 = \beta_e(S_1\dot{x} - Q)/[S_1(l-x)] \tag{3}$$

式中: β_e 为液压油有效体积弹性模量; Q 为直驱泵流量; l 为直驱泵腔长度。

主动阀在泵阀协同制动单元中起到了配流的作用, 假设主动阀全开, 此时主动阀阀芯两端压差较小, 主动阀阀孔流量简化为

$$Q = k_0 c_0 A_0 \sqrt{\frac{2}{\rho}\Delta P} \tag{4}$$

式中: k_0 为流量线性化系数; c_0 为阀口流量系数; A_0 为孔口面积; ΔP 为阀口液压差; ρ 为液压油密度。

制动轮缸实现液压力到制动力的转化。体积模量定义为 $\beta_e S_2 \approx k_h(l_h + x_h)$, 由于活塞位移 x_h 较小, 且轮缸压力升高时, 活塞位移基本不变, $\dot{x}_h$ 可忽略, 轮缸的压力变化模型为

$$\dot{P}_2 = \beta_e \frac{Q}{S_2(l_h + x_h)} \tag{5}$$

式中: S_2 为轮缸活塞横截面积; x_h 为活塞位移; l_h 为轮缸腔内长度; P_2 为轮缸压力。为便于液压子系统建模, 设以下参数:

$$\begin{cases} \xi(x) = \dfrac{\beta_e}{l-x} \\ \phi = \dfrac{k_0 c_0 A_0}{S_1}\sqrt{\dfrac{2}{\rho}} \\ \psi = \dfrac{k_h k_0 c_0 A_0}{S_2^2}\sqrt{\dfrac{2}{\rho}} \end{cases} \tag{6}$$

通过所设参数, 直驱泵和轮缸内液压力为

$$\begin{cases} \dot{P}_1 = \xi(x)(\dot{x} - \phi\Delta P) \\ \dot{P}_2 = \psi\Delta P \end{cases} \tag{7}$$

由式(7)可以得到压力差表达式为

$$\Delta P = \frac{\xi(x)}{\xi(x)\phi + \psi}\dot{x} - \frac{1}{\xi(x)\phi + \psi}\Delta\dot{P} \tag{8}$$

根据式(8), 泵内压力进一步改写为

$$\begin{cases} \dot{P}_1 = \dfrac{\psi}{\phi}\dot{x} + D \\ D = \dfrac{\xi(x)\phi}{\xi(x)\phi + \psi}\Delta\dot{P} - \dfrac{\varphi^2}{\phi^2\xi(x) + \phi\psi}\dot{x} \end{cases} \tag{9}$$

2.3　制动单元状态空间模型

式 (1) 电路子系统中电感项 $\dfrac{dI}{dt}I$ 相对其他项较小, 为简化模型可忽略不计, 整合式 (1)、式 (7) 和式 (9) 得到的直驱泵阀协同式线控制动单元的系统动力学模型如式(10)所示, 其主要参数如表 1 所示。

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = \dfrac{K_m}{MR}u - \dfrac{K_m K_e}{MR}x_2 - \dfrac{S_1}{M}x_3 + \dfrac{1}{M}F_f + \tilde{d} \\ \dot{x}_3 = \dfrac{\psi}{\phi}x_2 + D \\ \dot{x}_4 = \psi(x_3 - x_4) \end{cases} \tag{10}$$

式中: $[x_1, x_2, x_3, x_4] = [x, \dot{x}, P_1, P_2]$; $\tilde{d}$ 为不确定性误差和干扰 F_{dis} 所产生的不确定项。

表 1　直驱泵阀协同式线控制动单元参数

Table 1　Parameters of brake-by-wire unit based on direct drive pump-valve cooperative

结构	参数	数值
电磁直线执行器	线圈电阻R/Ω	1.40
	线圈等效电感L/mH	0.91
	反电势系数K_e/(Vs·m^{-1})	24.61
	电磁力系数K_m/(N·A^{-1})	24.61
直驱泵	泵腔长度l/mm	16
	活塞面积S_1/mm^2	27.5
主动阀	阀芯直径/mm	8
	阀座直径/mm	4
制动液	弹性模量β_e/MPa	1 700
	密度ρ/(kg·m^{-3})	1 046

3　串级控制器设计

直驱泵阀协同式线控制动单元的液压力控制设计为压力环-位置环的串级控制器, 其中, 外环的压力环采用滑模控制, 利用其控制响应快的优点, 得到可以使压力快速到达目标液压力的理想时变位移。单元中电磁直线执行器与直驱泵柱塞固连, 由于缺少中间缓冲环节, 各种非线性和时变的内外部干扰直接作用在驱动单元上, 所以需要设计合适的控制器以降低未知非线性和时变干扰对控制性能的影响。内环设计自适应积分鲁棒控制减小执行器参数不确定性和未知干扰对执行器性能的影响。本文设计的串级控制器如图 2 所示, 其中, P_d 为目标液压力; x_d 为执行器目标位移。

3.1　外环控制器设计

3.1.1　滑模控制器

为满足控制器设计, 定义液压力误差为

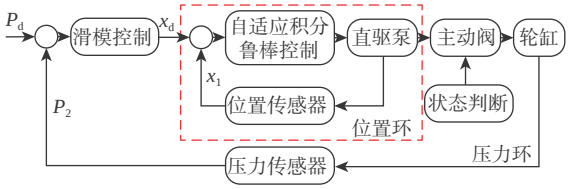


图2 控制系统总体框图

Fig. 2 Overall block diagram of control system

$$e = P_d - P_2 \tag{11}$$

式中：P_d为目标液压力。针对制动单元的液压子系统，滑模面s设计如下[21]：

$$s = c_a e + c_b \int e\mathrm{d}t \tag{12}$$

式中：c_a和c_b均为常数，为保证系统从任一点出发的状态都可到达滑模面，设计趋近律为

$$\dot{s} = -k_h s - q\mathrm{sat}(s) \tag{13}$$

式中：k_h、q均为常数且均大于0；sat(s)为削弱系统抖振所设计的饱和函数，具体形式为

$$\mathrm{sat}(s) = \begin{cases} 1 & s > \delta \\ \dfrac{s}{\delta} & |s| \leqslant \delta \\ -1 & s < -\delta \end{cases} \tag{14}$$

将式(12)代入式(13)得到执行器理想位移x_d：

$$x_d = \left[\psi\phi\left(c_a\dot{e} + c_b P_d + k_h s + q\mathrm{sat}(s) - c_b\int_0^t D\mathrm{d}t\right) + \phi c_b \dot{P}_2\right] \Big/ c_b\psi^2 \tag{15}$$

3.1.2 稳定性分析

为证明系统的稳定性，构造李亚普诺夫函数如下：

$$V_1 = \frac{1}{2}s^2 \tag{16}$$

对式(16)求导，并将式(13)代入得

$$\dot{V}_1 = s\dot{s} = s(-k_h s - q\mathrm{sat}(s)) \tag{17}$$

根据饱和函数sat(s)的定义可知，式(17)恒小于0，因此，外环控制器稳定有效。

3.2 内环控制器设计

3.2.1 自适应积分鲁棒控制器设计

根据系统动力学模型式(10)，将参数定义为：$\boldsymbol{\theta}$=$[\theta_1, \theta_2, \theta_3, \theta_4]^{\mathrm{T}}$，$\theta_1$=$MR/K_m$；$\theta_2$=$K_e$+$B_1R/K_m$；$\theta_3$=$A_fR/K_m$；$\theta_4$=$S_1R/K_m$，假设上述参数和不确定项范围已知，电磁直线执行器的动力学方程重新定义为

$$\theta_1\dot{x}_2 = u - \theta_2 x_2 - \theta_3\arctan(\beta x_2) - \theta_4 P_1 + \tilde{d} \tag{18}$$

$$\begin{cases} \left|\dot{\tilde{d}}\right| \leqslant \delta_1 \\ \left|\ddot{\tilde{d}}\right| \leqslant \delta_2 \end{cases} \tag{19}$$

$$\theta_i \in [\theta_{i\min}, \theta_{i\max}] \qquad i = 1,2,3,4 \tag{20}$$

式中：δ_1和δ_2为已知常数。

为满足控制器设计，定义执行器位置误差e_1、滑模变量e_2和辅助变量r如下：

$$\begin{cases} e_1 = x_1 - x_d \\ e_2 = \dot{e}_1 + k_1 e_1 \\ r = \dot{e}_2 + k_r e_2 \end{cases} \tag{21}$$

式中：k_1和k_r为正增益。将这些变量进一步整合得到

$$r = \ddot{x}_1 - \ddot{x}_d - k_1^2 e_1 + (k_1 + k_r)e_2 \tag{22}$$

将式(18)代入式(22)得

$$\theta_1 r = u - \theta_1\ddot{x}_d - \theta_2 x_2 - \theta_3\arctan(\beta x_2) - \theta_4 P_1 + \tilde{d} - \theta_1 k_1^2 e_1 + \theta_1(k_1 + k_r)e_2 \tag{23}$$

此外，通过不连续投影映射函数使参数估计值都处于有界范围内，因此，参数估计所需的投影映射函数定义为

$$\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \tag{24}$$

$$\dot{\hat{\boldsymbol{\theta}}} = \text{Proj-}\hat{\boldsymbol{\theta}}(\boldsymbol{\Gamma\tau}) \tag{25}$$

$$\mathrm{Proj}\hat{\theta}_i\{\zeta_i\} = \begin{cases} 0 & \hat{\theta}_i = \hat{\theta}_{i\max} \text{ 且 } \zeta_i \geqslant 0 \\ 0 & \hat{\theta}_i = \hat{\theta}_{i\min} \text{ 且 } \zeta_i < 0 \\ \zeta_i & \text{其他} \end{cases} \tag{26}$$

式中，$\hat{\boldsymbol{\theta}}$和$\tilde{\boldsymbol{\theta}}$分别为参数的估计值和误差矩阵；$\boldsymbol{\Gamma}$为系数矩阵；$\boldsymbol{\tau}$为参数估计变量矩阵；Proj为投影映射函数，上述自适应律有以下属性：

$$\begin{cases} \theta_{i\min} \leqslant \hat{\theta}_i \leqslant \theta_{i\max} \quad i = 1,2,\cdots,4 \\ \tilde{\boldsymbol{\theta}}^{\mathrm{T}}\left(\boldsymbol{\Gamma}^{-1}\text{Proj-}\hat{\boldsymbol{\theta}}(\boldsymbol{\Gamma\tau}) - \boldsymbol{\tau}\right) \leqslant 0 \end{cases} \tag{27}$$

根据动力学方程，控制器可设计为

$$\begin{cases} u = u_a + u_s \\ u_a = \boldsymbol{\varphi}^{\mathrm{T}}\hat{\boldsymbol{\theta}} \\ u_s = u_{s1} + u_{s2} \\ u_{s1} = -ke_2 \end{cases} \tag{28}$$

式中：k为正增益；u_a为自适应补偿项；u_s为鲁棒反馈项；u_{s1}为用于稳定系统的比例反馈项；u_{s2}为积分鲁棒项；$\boldsymbol{\varphi}$的具体表达式为

$$\boldsymbol{\varphi} = [\ddot{x}_d, x_2, \arctan(\beta x_2), P_1]^{\mathrm{T}} \tag{29}$$

由于系统不确定的非线性因素，仅依赖线性反馈项不能达到很好的控制效果，因此，利用积分鲁棒项进一步降低模型不确定性，具体表达式为

$$u_{s2} = -\int_0^t \left(kk_r e_2 + \mu\mathrm{sgn}(e_2)\right)\mathrm{d}w \tag{30}$$

式中：μ为常数，sgn()为符号函数。自适应积分鲁棒控制的具体控制框图如图3所示。

将控制器信号代入式(23)中得

$$\theta_1 r = \boldsymbol{\varphi}^{\mathrm{T}}\tilde{\boldsymbol{\theta}} - \int_0^t \left(kk_r e_2 + \mu\mathrm{sgn}(e_2)\right)\mathrm{d}w + \tilde{d} - ke_2 - \theta_1 k_1^2 e_1 + \theta_1(k_1 + k_r)e_2 \tag{31}$$

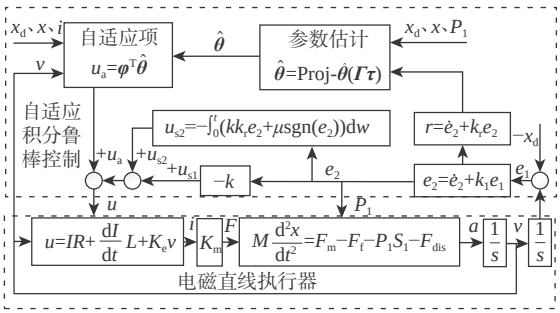


图 3　自适应积分鲁棒控制框图

Fig. 3　Adaptive integral robust control block diagram

为便于后续控制器稳定性分析验证, 对式 (31) 进行求导并整理得

$$\begin{aligned}\theta_1\dot{r}=&\boldsymbol{\varphi}^{\mathrm{T}}\text{Proj-}\hat{\boldsymbol{\theta}}(-\boldsymbol{\Gamma}\dot{\boldsymbol{\varphi}}r)+\dot{\boldsymbol{\varphi}}^{\mathrm{T}}\tilde{\boldsymbol{\theta}}-kr-\\&\mu\mathrm{sgn}(e_2)+\dot{\tilde{d}}+\theta_1(k_1+k_{\mathrm{r}})r+\\&\theta_1k_1^3e_1-\left[\theta_1k_1^2+\theta_1(k_1+k_{\mathrm{r}})k_{\mathrm{r}}\right]e_2\end{aligned}\tag{32}$$

3.2.2　稳定性分析

首先, 声明以下引理:

引理　定义辅助函数和其积分式如下:

$$L(t)=-r\mu\mathrm{sgn}(e_2)+r\dot{\tilde{d}}\tag{33}$$

$$P=\mu|e_2(0)|-e_2(0)\dot{\tilde{d}}(0)-\int_0^tL(w)\mathrm{d}w\tag{34}$$

式中: $\tilde{d}(0)$和 $e_2(0)$ 分别为$\tilde{d}$和e_2的初始值。如果控制增益满足 $\mu>\delta_1+\delta_2/k_{\mathrm{r}}$, 则函数 P 恒为正。

证明

对式 (33) 积分得

$$\begin{aligned}\int_0^tL(w)\mathrm{d}w=&\int_0^tk_re_2\left(\dot{\tilde{d}}-\mu\mathrm{sgn}(e_2)\right)\mathrm{d}w+\\&\int_0^t\dot{e}_2\left(\dot{\tilde{d}}-\mu\mathrm{sgn}(e_2)\right)\mathrm{d}w\end{aligned}\tag{35}$$

对式(35)进行分部积分得

$$\begin{aligned}\int_0^tL(w)\mathrm{d}w=&\int_0^tk_re_2\left(\dot{\tilde{d}}-\mu\mathrm{sgn}(e_2)\right)\mathrm{d}w+\\&e_2\dot{\tilde{d}}-\mu(|e_2(t)|-|e_2(0)|)-e_2(0)\dot{\tilde{d}}(0)-\int_0^t\dot{e}_2\ddot{\tilde{d}}\mathrm{d}w=\\&\int_0^tk_re_2\left(\dot{\tilde{d}}-\mu\mathrm{sgn}(e_2)\right)\mathrm{d}w+e_2\dot{\tilde{d}}-\\&e_2(0)\dot{\tilde{d}}(0)-\mu|e_2|+\mu|e_2(0)|-\int_0^t\dot{e}_2\ddot{\tilde{d}}\mathrm{d}w\end{aligned}\tag{36}$$

由式(36)可得

$$\begin{aligned}\int_0^tL(w)\mathrm{d}w\leqslant&\int_0^tk_r|e_2|\left(\delta_1-\mu+\frac{\delta_2}{k_{\mathrm{r}}}\right)\mathrm{d}w+\\&|e_2|(\delta_1-\mu)+\mu|e_2(0)|-e_2(0)\dot{\tilde{d}}(0)\end{aligned}\tag{37}$$

由式(37)可以看出, 若 $\mu>\delta_1+\delta_2/k_{\mathrm{r}}$, 函数 P 恒为正, 故引理得证。

证毕

定理　若通过调整参数 k、k_1、k_{r}, 对称矩阵 $\boldsymbol{\Lambda}$ 正定, 则系统 (18) 中所有信号均有界, 且控制器可获得渐近稳定性, 即 $t\to\infty$时 $\boldsymbol{z}\to\boldsymbol{0}$, 其中 $\boldsymbol{z}=[e_1, e_2, r]^{\mathrm{T}}$。

$$\boldsymbol{\Lambda}=\begin{bmatrix}k_1&-\frac{1}{2}&-\frac{1}{2}c_1\\-\frac{1}{2}&k_{\mathrm{r}}&-\frac{1-c_2}{2}\\-\frac{1}{2}c_1&-\frac{1-c_2}{2}&k_3\end{bmatrix}\tag{38}$$

定义参数:

$$\begin{cases}c_1=\theta_1k_1^3\\c_2=(k_1+k_{\mathrm{r}})k_{\mathrm{r}}\theta_1+k_1^2\theta_1\\k_3=k-\max\left\{\left|\boldsymbol{\varphi}^{\mathrm{T}}\text{Proj-}\hat{\boldsymbol{\theta}}\left(-\boldsymbol{\Gamma}\dot{\boldsymbol{\varphi}}\right)\right|\right\}-\theta_1(k_1+k_{\mathrm{r}})\end{cases}\tag{39}$$

证明　为证明系统稳定性, 定义非负函数如下:

$$V_2=\frac{1}{2}e_1^2+\frac{1}{2}e_2^2+\frac{1}{2}\theta_1r^2+\frac{1}{2}\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\tilde{\boldsymbol{\theta}}+P\tag{40}$$

对式 (40) 求导:

$$\dot{V}_2=e_1\dot{e}_1+e_2\dot{e}_2+\theta_1r\dot{r}+\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\dot{\tilde{\boldsymbol{\theta}}}+\dot{P}\tag{41}$$

将式 (21) 和式 (32) 代入式 (41) 得

$$\begin{aligned}\dot{V}_2=&e_1(e_2-k_1e_1)+e_2(r-k_re_2)+\\&r\{\boldsymbol{\varphi}^{\mathrm{T}}\text{Proj-}\hat{\boldsymbol{\theta}}\left(-\boldsymbol{\Gamma}\dot{\boldsymbol{\varphi}}r\right)+\dot{\boldsymbol{\varphi}}^{\mathrm{T}}\tilde{\boldsymbol{\theta}}-kr-\\&\mu\mathrm{sgn}(e_2)+\dot{\tilde{d}}+\theta_1(k_1+k_{\mathrm{r}})r+\theta_1k_1^3e_1-\\&\left[\theta_1k_1^2+\theta_1(k_1+k_{\mathrm{r}})k_{\mathrm{r}}\right]e_2\}+\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\dot{\tilde{\boldsymbol{\theta}}}+\dot{P}\end{aligned}\tag{42}$$

将式 (34) 中 P 的表达式代入式 (42) 并整理得

$$\begin{aligned}\dot{V}_2=&-k_1e_1^2-k_re_2^2+e_1e_2+e_2r+c_1e_1r-c_2e_2r-\\&kr^2+\left[\boldsymbol{\varphi}^{\mathrm{T}}\text{Proj-}\hat{\boldsymbol{\theta}}\left(-\boldsymbol{\Gamma}\dot{\boldsymbol{\varphi}}\right)+\theta_1(k_1+k_{\mathrm{r}})\right]r^2+\\&\dot{\boldsymbol{\varphi}}^{\mathrm{T}}\tilde{\boldsymbol{\theta}}r+\tilde{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}\dot{\tilde{\boldsymbol{\theta}}}\end{aligned}\tag{43}$$

根据自适应率属性和式 (39) 参数得

$$\dot{V}_2\leqslant-\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{z}\leqslant-\lambda_{\min}(\boldsymbol{\Lambda})\left(e_1^2+e_2^2+r^2\right)\triangleq-W\tag{44}$$

由于 $V\in L_\infty$和 $W\in L_2$, 并且对信号 $\boldsymbol{z}$ 和参数的估计是有界的。那么 x_1 始终有界, 同样控制输入 u 有界。根据 e_1、e_2 和 r 的计算式, 可知 W 的导数有界, 因此, W 是一致连续的。根据 Barbalat 引理[22], 当 $t\to\infty$时 $W\to0$, 故定理得证。

证毕

4　结果与分析

4.1　试验平台

本文搭建了如图 4 所示的直驱泵阀协同式线控制动单元试验平台。上位机的控制信号通过以太网传送给控制器, 控制器控制功率驱动模块控制电磁直线执行器的驱动电压、主动阀的开关, 从而控制制动轮缸液压力。控制器采用快速控制原型系统 RTU-BOX, 其数字控制器采用主频率为 300 MHz 的 32 位浮点数字信号处理器 TMS320C 28346。采用 0.01 mm 分辨率的位置传感器提供电磁直线执

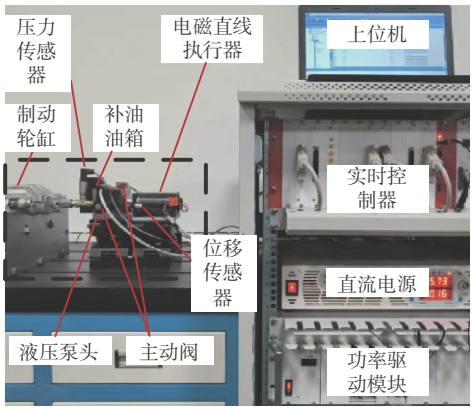


图 4 测试平台

Fig. 4 Experimental platform

行器的位置反馈，具有 0.01 MPa 分辨率的压力传感器提供轮缸液压力反馈。

将本文提出的串级控制与常用的双闭环 PID 控制进行比较。串级控制中，外环参数 c_a=0.02, c_b=3, k_h=5, q=0.002; 内环参数 k=500, k_l=130, k_r=4, μ=10, $\boldsymbol{\theta}_{max}$=[0.1, 50, 0.5, 0.01]^T, $\boldsymbol{\theta}_{min}$=[0, 0, 0, 0]^T, $\boldsymbol{\Gamma}$=diag[0.8, 3 600, 5.2, 0.02]。双闭环 PID 控制中外环 PID 参数 k_{p1}=0.06, k_{I1}=0.8, k_{d1}=0.05; 内环 PID 参数 k_{p2}=600, k_{I2}=5 000, k_{d2}=2 000。

4.2 响应性能测试

模拟目标小车紧急制动工况，设置阶跃响应目标为 4 MPa，测试结果如图 5 所示。在 2 种不同控制器下，本文所设计的直驱泵阀协同式线控制动单元均可以快速达到目标值。双闭环 PID 控制的响应时间为 0.028 s，而串级控制的响应时间为 0.023 s，相较于双闭环 PID 控制，响应时间提升了 17.8%，证明串级控制的快速响应性能优于双闭环 PID 控制。

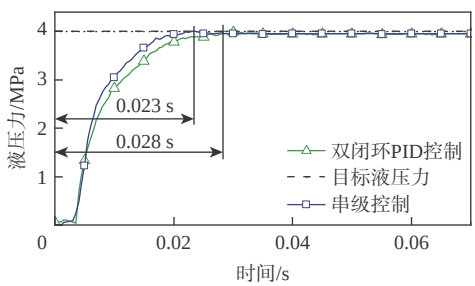


图 5 阶跃响应测试结果

Fig. 5 Experimental results of step response

串级控制器下的电磁直线执行器及主动阀的测试结果如图 6 所示。系统达到目标压力且工作模式切换为保压模式时，此时保压阀关闭导致液压力出现较小的波动，同时执行器控制信号为 0，电流值迅速下降，在泵腔内压力的作用下执行器向复位方向运动一段位移。

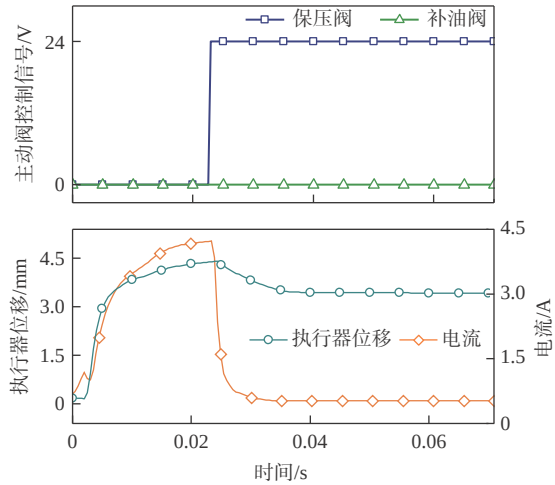


图 6 串级控制器时阶跃响应测试中直驱泵相关信号

Fig. 6 Relevant signals of direct-drive pump in step response experiment with cascade controller

4.3 跟踪性能测试

正弦制动压力工况是模拟驾驶员在某些特定工况下有规律地松踩制动踏板，设置目标正弦压力信号频率为 2.5 Hz、振幅为 2.5 MPa，液压力跟踪测试如图 7 所示。

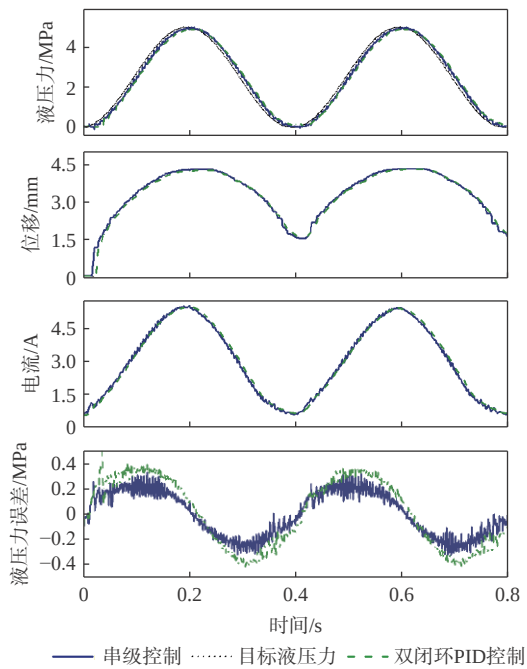


图 7 正弦压力目标信号试验结果

Fig. 7 Experimental results of sinusoidal pressure target signal

2 种控制均可以较好的跟踪正弦信号，在第 1 个峰值处，串级控制器和双闭环 PID 控制器的滞后时间分别为 0.007 s 和 0.009 s。由于制动单元内部液压油未充满，在 0.4 s 处执行器动子位置高于初始状态。为进一步比较 2 种控制器的跟踪性能，采用液压力平均误差 e_μ 和误差的标准偏差 e_σ 来比较 2 种控制器的液压力跟踪性能。e_μ 和 e_σ 的具体表

达式为

$$\begin{cases} e_{\mu}=\left(\sum_{i=1}^{n}|e_i|\right)\Big/n \\ e_{\sigma}=\sqrt{\left[\sum_{i=1}^{n}\left(|e_i|-e_{\mu}\right)^2\right]\Big/n} \end{cases} \tag{45}$$

在整个跟踪过程中, 双闭环 PID 控制的 e_{μ} 为 0.238 MPa, 串级控制的平均误差 e_{μ} 为 0.159 MPa,相比于双闭环 PID 控制减小了 33.2%。双闭环 PID 控制的 e_{σ} 为 0.107, 串级控制的标准偏差 e_{σ} 为 0.074 MPa, 标准偏差 e_{σ} 为 0.074 MPa, 其相比于双闭环 PID 控制减小了 30.8%。2 种控制的性能参数均较小, 说明 2 种控制均可较好的实现目标液压力的跟踪控制, 同时在正弦目标液压力下, 提出的串级控制器在跟踪性能上优于双闭环 PID 控制器。

为模拟较为缓和的制动过程, 设置三角波制动目标液压力频率为 2.5 Hz、振幅为 2.5 MPa, 2 种控制器的液压力跟踪结果如图 8 所示。由图可知, 在 0.2 s 和 0.4 s 处目标信号导数突变, 但 2 种控制器均能较好的跟随目标液压力。

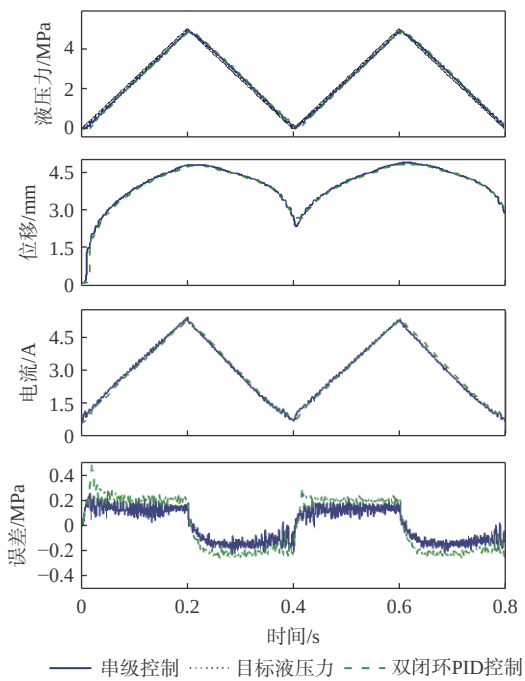


图 8　三角波压力信号试验结果

Fig. 8　Experimental results of triangular wave pressure signal

在三角波目标跟踪过程中, 双闭环 PID 控制的平均误差 e_{μ} 为 0.201 MPa, 串级控制的 e_{μ} 为 0.126 MPa, 相较于双闭环 PID 控制降低了 37.3%。双闭环 PID 控制的 e_{σ} 为 0.056, 串级控制的标准偏差 e_{σ} 为 0.045, 相比于双闭环 PID 控制减小了 19.6%。说明相较于双闭环 PID 控制, 串级控制在维持一定速率增加或降低液压力时有较好的准确性, 也证明了在三角波目标液压力下, 本文的串级控制器在跟踪性能上优于双闭环 PID 控制器。

5　结　论

1) 本文提出一种响应迅速、不恶化轮边空间与簧下质量的直驱泵阀协同式线控制动单元, 通过电磁直线执行器直接驱动的液压泵与主动阀协调实现轮缸压力调节, 为分布式制动系统提供了一种新的方案。

2) 本文设计了基于外环轮缸压力和内环执行器位移的串级控制器, 通过外环滑模控制得到理想执行器位移, 利用内环自适应积分鲁棒控制减小执行器参数不确定性和时变干扰的影响, 提高了系统响应速度与控制精度。

3) 基于新的制动单元搭建了测试平台, 通过阶跃响应、正弦波跟踪、三角波跟踪等工况, 测试了制动单元性能, 试验结果表明, 在响应速度与控制精度上, 本文的串级控制器优于双闭环 PID 控制器, 为线控制动系统控制提供了一种方案。

参考文献（References）

[1] 高峰, 雍加望, 丁能根, 等. 集成电液制动系统助力算法及其功能验证[J]. 北京航空航天大学学报, 2017, 43(3): 424-431.
GAO F, YONG J W, DING N G, et al. Booster algorithm and functionality validation of an integrated electro-hydraulic brake system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(3): 424-431(in Chinese).

[2] ZHAO J, CHEN Z C, ZHU B, et al. Precise active brake-pressure control for a novel electro-booster brake system[J]. IEEE Transactions on Industrial Electronics, 2020, 67(6): 4774-4784.

[3] 何仁, 刘学军, 刘存香. 汽车电磁-液压复合制动技术研究进展[J]. 中国公路学报, 2014, 27(11): 109-119.
HE R, LIU X J, LIU C X. Research progress in electromagnetic-hydraulic hybrid brake technology[J]. China Journal of Highway and Transport, 2014, 27(11): 109-119(in Chinese).

[4] XIONG L, HAN W, YU Z P. Adaptive sliding mode pressure control for an electro-hydraulic brake system via desired-state and integral-antiwindup compensation[J]. Mechatronics, 2020, 68: 102359.

[5] PAN N, YU L Y, WANG Z Z, et al. Design, modeling and simulation of a new compact electro-hydraulic brake system[C]//Proceedings of the SAE Technical Paper Series. Warrendale: SAE International, 2014.

[6] 余卓平, 史彪飞, 卓桂荣, 等. 集成式电子液压制动系统位移压力特性理论研究[J]. 机械工程学报, 2022, 58(22): 294-303.
YU Z P, SHI B F, ZHUO G R, et al. Theoretical research on pressure-position relationship of the integrated electronic-hydraulic brake system[J]. Journal of Mechanical Engineering, 2022, 58(22): 294-303(in Chinese).

[7] 胡东海, 何仁. 电子液压制动系统的安全设计与匹配分析[J]. 农业工程学报, 2015, 9: 77-84.
HU D H, HE R. Safety design and matching analysis of electronic

hydraulic brake system[J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 9: 77-84(in Chinese).

[8] 刘兆勇, 龙一鸣, 厉逸航, 等. 集成式电液制动系统设计及参数匹配的研究[J]. 汽车工程, 2022, 44(9): 1416-1424.
LIU Z Y, LONG Y M, LI Y H, et al. Study on the design and parameter matching of integrated electro-hydraulic braking system[J]. Automotive Engineering, 2022, 44(9): 1416-1424(in Chinese).

[9] 陈晋市, 刘思远, 王同建, 等. 8×8 全电驱动越野车电机液压联合全液压制动系统设计及性能[J]. 兵工学报, 2021, 42(02): 422-429.
CHEN J S, LIU S Y, WANG T J, et al. Design and performance of electro-hydraulic full hydraulic brake system for 8×8 all-electric drive off-road vehicle[J]. Acta Armamentarii, 2021, 42(02): 422-429 (in Chinese).

[10] WANG Z Z, YU L Y, YOU C X, et al. Fail-safe control allocation for a distributed brake-by-wire system considering the driver's behaviour[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2014, 228(13): 1547-1567.

[11] GONG X X, QIAN L X, GE W G, et al. Research on electronic brake force distribution and anti-lock brake of vehicle based on direct drive electro hydraulic actuator[J]. International Journal of Automotive Engineering, 2020, 11(2): 22-29.

[12] 初亮, 赵迪, 李文惠. 电-液制动系统高速开关电磁阀的建模与动态特性仿真[J]. 汽车工程, 2017, 39(1): 61-65.
CHU L, ZHAO D, LI W H. Modeling and dynamic characteristics simulation for fast-switching solenoid valves in electro-hydraulic braking systems[J]. Automotive Engineering, 2017, 39(1): 61-65(in Chinese).

[13] 熊璐, 韩伟, 余卓平, 等. 考虑关键非线性特征的集成式电子液压制动系统主缸液压力精确控制[J]. 机械工程学报, 2019, 55(24): 117-126.
XIONG L, HAN W, YU Z P, et al. Pressure precisely control of master cylinder on integrated-electro-hydraulic brake system considering the critical nonlinear characteristics[J]. Journal of Mechanical Engineering, 2019, 55(24): 117-126(in Chinese).

[14] 石琴, 刘鑫, 应贺烈, 等. 电液线控制动系统压力反步控制算法研究[J]. 汽车工程, 2022, 44(5): 747-755.
SHI Q, LIU X, YING H L, et al. Study on the backstepping control algorithm for the hydraulic pressure in electro-hydraulic brake-by-wire system[J]. Automotive Engineering, 2022, 44(5): 747-755(in Chinese).

[15] JI Y, ZHANG J Z, HE C K, et al. Constraint performance pressure tracking control with asymmetric continuous friction compensation for booster based brake-by-wire system[J]. Mechanical Systems and Signal Processing, 2022, 174: 109083.

[16] TAN C, REN H X, LI B, et al. Design and analysis of a novel cascade control algorithm for braking-by-wire system based on electromagnetic direct-drive valves[J]. Journal of the Franklin Institute, 2022, 359(16): 8497-8521.

[17] HAN W, XIONG L, YU Z P. Braking pressure control in electro-hydraulic brake system based on pressure estimation with nonlinearities and uncertainties[J]. Mechanical Systems and Signal Processing, 2019, 131: 703-727.

[18] 俞军涛, 占昊, 王丽, 等. 压电式高速开关阀控液压缸位置系统[J]. 北京航空航天大学学报, 2021, 47(4): 706-714.
YU J T, ZHAN H, WANG L, et al. Hydraulic cylinder position system controlled by piezoelectric high-speed on-off valve[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(4): 706-714(in Chinese).

[19] JIANG L G, SHI Q, WEI Y J, et al. Electro-hydraulic braking dynamics for pressure demand control of brake-by-wire system[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2024, 238(4): 735-748.

[20] 谭草, 鲁应涛, 葛文庆, 等. 直驱式永磁直线电机深度模糊滑模-自抗扰控制[J]. 西安交通大学学报, 2023, 57(1): 185-194.
TAN C, LU Y T, GE W Q, et al. Depth fuzzy sliding-mode active disturbance rejection control method of permanent magnet linear motor for direct drive system[J]. Journal of Xi'an Jiaotong University, 2023, 57(1): 185-194(in Chinese).

[21] 李波, 黎德祥, 葛文庆, 等. 基于直驱阀的快速响应线控制动系统液压力精确控制[J]. 中国公路学报, 2021, 34(9): 121-132.
LI B, LI D X, GE W Q, et al. Precision control of hydraulic pressure in fast-response brake-by-wire system based on direct-drive valve[J]. China Journal of Highway and Transport, 2021, 34(9): 121-132(in Chinese).

[22] 潘菲, 朱宏玉. 航天器非奇异自适应终端滑模姿轨联合控制[J]. 北京航空航天大学学报, 2020, 46(7): 1354-1362.
PAN F, ZHU H Y. Spacecraft non-singular adaptive terminal sliding mode attitude-orbit coupling control[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(7): 1354-1362(in Chinese).

Pressure cascade control of brake-by-wire unit based on direct drive pump-valve cooperative

TAN Cao[1]，YU Peng[1]，LI Bo[1,*]，LU Jiayu[1]，REN Yunyun[2]

(1. School of Transportation and Vehicle Engineering，Shandong University of Technology，Zibo 255000，China；
2. Shuntai Automobile Co.，Ltd.，Zibo 255000，China)

Abstract：Aiming at the requirements of distributed braking and high-level automatic driving, a brake-by-wire unit based on a direct drive pump-valve cooperative is designed, to realize the wheel cylinder pressure regulation, the hydraulic pump directly driven by the electromagnetic linear actuator coordinates with the active valve. A control-oriented system dynamics model is established to realize the precise control of pressure, and a cascade control method is proposed, which includes an outer loop to control the cylinder pressure and an inner loop to control the plunger position in the pump. An adaptive integral robust controller is created for the inner loop to lessen the impact of time-varying interference and parameter uncertainty, while a sliding mode controller is designed for the outer loop to increase reaction speed. It is proved that the algorithm is Lyapunov stable. The results suggest that the developed brake unit is a possible novel scheme for distributed braking, and the proposed method mechanism can increase braking pressure speed and control accuracy even further.

Keywords：brake-by-wire; hydraulic pressure control; cascade control; sliding mode control; adaptive integral robust control

Received：2023-04-27；**Accepted**：2023-06-06；**Published Online**：2023-06-15 11：44

URL：link.cnki.net/urlid/11.2625.V.20230614.1457.001

Foundation items：National Natural Science Foundation of China (51975341); Outstanding Young Talents Foundation of Shandong Province (ZR2022YQ51); Major Scientific and Technological Innovation Project of Shandong Province (2021CXGC010703); Innovation Team Project of “Qing-Chuang Science and Technology Plan” of Colleges and Universities in Shandong Province (2019KJB027,2022KJ232); Young Technology Talent Supporting Project of Shandong Province (SDAST2021qt20)

*** Corresponding author**. E-mail：libo@sdut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0224

基于 MDAM-GhostCNN 的滚动轴承故障诊断方法

郭俊锋，谭宝宏*，王智明
(兰州理工大学 机电工程学院，兰州 730050)

摘　　要：针对传统故障诊断方法特征提取不充分、计算复杂及在变工况下识别准确率低的问题，提出一种基于混合域注意力机制 (MDAM)-GhostCNN 的滚动轴承故障诊断方法。采用马尔可夫转移场 (MTF) 将轴承振动信号转化为具有时间相关性的二维特征图；利用 Ghost 卷积计算精简的优点，构造出 GhostCNN；设计一种 MDAM，使网络从通道和空间 2 个维度充分捕获特征信息，实现特征通道间相互依赖的同时让网络有效关注特征空间信息。由此，构建出 MDAM-GhostCNN 模型。将 MTF 二维特征图输入到 MDAM-GhostCNN 模型中进行训练并输出诊断结果。采用凯斯西储大学和江南大学 (JNU) 轴承数据集进行实验验证，并对其数据集进行加噪处理。结果表明：在变工况下，所建模型有着更高的识别准确率、抗噪性能和泛化性能。

关 键 词：滚动轴承；故障诊断；马尔可夫转移场；Ghost 卷积；注意力机制
中图分类号：TH133.33
文献标志码：A　　**文章编号**：1001-5965(2025)04-1172-13

现代工业中，机械设备正朝着智能化、精密化、高效化的方向快速发展[1]。滚动轴承作为旋转机械的关键部件，广泛应用于航空航天、车辆、数控机床等重要领域[2]。滚动轴承的意外失效可能导致机器或整个机械系统故障，造成巨大的财产损失，甚至人员伤亡。因此，滚动轴承的健康状态和故障诊断对于工业系统的安全运行至关重要[3-4]。

传统的智能故障诊断方法过度依赖人工主观经验[5]，手动提取特征耗时耗力，且模型泛化能力较差，特征提取不充分，导致模型的诊断准确率较低。而深度学习因其有多层非线性处理单元的深度结构，可以从原始数据中直接提取特征，已成为轴承故障诊断的一种很有前途的工具[6]。常见的深度学习模型包括深度置信网络 (deep belief network, DBN)[7]、自动编码器 (autoencoders, AE)[8]、卷积神经网络 (convolutional neural networks, CNN)[9-10]等。其中，CNN 作为深度学习应用的主要方法，因其具有局部感知和参数共享等优点，所以在故障诊断领域被广泛应用。利用深度学习可以直接将原始信号输入到模型中提取特征并实现故障诊断。Chen 等[11]利用循环谱相干分析的振动信号的二维 CSCoh 映射，将其馈送到 CNN 中用于电机轴承故障诊断。Xu 等[12]将信号用变模态分解 (variational mode decomposition, VMD) 与深度 CNN 相结合，开发了一种用于轴承故障诊断的 VMD-DCNN 模型。Ye 等[13]用变量模型提取 (variational mode extraction, VME)来处理采集到的轴承振动信号，以获得所需的模态分量并消除噪声干扰信息，然后，将信号输入到 I-1DCNN 模型中进行训练以实现故障诊断。

由于原始振动信号存在可变性因素，如环境、工况等，会导致特征提取困难，使得模型的泛化性变差，从而影响诊断结果。因此，一些学者将原始

收稿日期：2023-05-04；**录用日期**：2023-06-20；**网络出版时间**：2023-07-07 11:34
网络出版地址：link.cnki.net/urlid/11.2625.V.20230706.1616.001
基金项目：国家自然科学基金 (51465034)
*通信作者. E-mail：1426779121@qq.com

引用格式：郭俊锋，谭宝宏，王智明. 基于 MDAM-GhostCNN 的滚动轴承故障诊断方法 [J]. 北京航空航天大学学报，2025，51（4）：1172-1184. GUO J F，TAN B H，WANG Z M. Fault diagnosis method of rolling bearing based on MDAM-GhostCNN [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1172-1184（in Chinese）.

振动信号从时域、频域或时频域方面进行处理后与 CNN 相结合进行故障诊断。刘俊锋等[14] 提出一种基于改进的傅里叶模态分解和 Transformer-CNN 故障诊断方法, 利用改进傅里叶模态分解和位置编码对数据样本进行处理, 提升了网络序列的建模能力。Fu 等[15] 通过连续小波变换将原始振动信号转为二维时频图, 然后, 用数据增强技术对标注故障数据进行扩展, 并将生成的样本输入到 CNN 中有效地进行故障诊断。金江涛等[16] 将 CNN 与双向长短期记忆 (bi-directional long-short term memory, BiLSTM) 网络结合, 采用相空间重构法将原始信号转为二维混沌序列, 学习并提取混沌序列中的有效非线性信息, 并输入 CCNN (Chaotic CNN) -BiLSTM 网络中进行故障诊断, 提高了模型诊断准确率。Mao 等[17] 将原始振动信号转为二维格式的频率和平方包络频谱, 并与红外热图像融合后输入到网络中进行齿轮箱跨域故障诊断。上述方法将原始振动信号与 CNN 结合, 在未考虑模型参数量和计算复杂度的情况下, 均具有较好诊断精度。然而, 在实际工业中, 轴承运行工况多变, 且常常混杂噪声, 复杂的模型结构影响故障诊断的效率的同时导致网络泛化性能退化。

针对以上问题, 本文从模型结构和特征提取方面出发, 提出一种基于混合域注意力机制 (mixed domain attention mechanism, MDAM)-GhostCNN 的滚动轴承故障诊断方法。利用轴承振动信号的时间特性, 用马尔可夫转移场 (Markov transfer field, MTF)将其转化为二维特征图, 以减少时间信息丢失; 由 Ghost 卷积构造出 GhostCNN, 减少网络参数量, 提升计算效率; 同时设计出一种 MDAM, 从通道和空间 2 个维度出发充分提取特征信息, 提高模型的适应性, 增强模型特征提取能力。最后, 结合 GhostCNN和 MDAM 构建出 MDAM-GhostCNN 模型, 并将 MTF 二维特征图输入到此模型中进行训练, 优化模型参数后进行诊断并输出诊断分类结果。运用凯斯西储大学和江南大学(Jiang Nan University, JNU)轴承数据集对本文模型进行验证, 实验结果表明, 本文模型在噪声、变工况下有着良好的泛化性能, 鲁棒性能和抗噪性能。

1　相关理论

1.1　马尔可夫转移场

MTF 以马尔可夫链为理论基础考虑时间位置信息, 运用马尔可夫状态转换概率来编码, 将一维振动信号转换为二维图像, 保留了时间序列信号的时间依赖性和频率结构[18]。

先给出一个时间序列 $\hat{X}=\{x_1,x_2,\cdots,x_n\}$, 其中, $x_i(i=1,2,\cdots,n)$表示相应时刻振动信号的幅值, 然后, 将X离散化为Q个区间, 用分位数$q_j(j\in[1,Q])$量化X的每个值, 通过识别分位数, 将每个数值x_i映射到q_i上。利用一阶马尔可夫链计算各时间步长的区间转移, 得到大小为$Q\times Q$的马尔可夫转移矩阵$\boldsymbol{W}$:

$$\boldsymbol{W}=\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1Q} \\ w_{21} & w_{22} & \cdots & w_{2Q} \\ \vdots & \vdots & & \vdots \\ w_{Q1} & w_{Q2} & \cdots & w_{QQ} \end{bmatrix} \tag{1}$$

式中: $q_j(j\in[1,Q])$为分位数; w_{ij}为分位数q_i和分位数q_j的靠近概率, $w_{ij}=p\{x_{t+1}\in q_j|x_t\in q_i\}$。

由于马尔可夫链对时间序列$\hat{X}$及时间步长t_i两者相关性的忽略, 会导致关键信息缺乏, 所以构造马尔可夫转移矩阵使其关注时间信息和位置因素, 并将其升华为 MTF $\boldsymbol{M}$:

$$\boldsymbol{M}=\begin{bmatrix} m_{ij}|p(x_1\in q_i|x_1\in q_j) & \cdots & m_{ij}|p(x_1\in q_i|x_n\in q_j) \\ m_{ij}|p(x_2\in q_i|x_1\in q_j) & \cdots & m_{ij}|p(x_2\in q_i|x_n\in q_j) \\ \vdots & & \vdots \\ m_{ij}|p(x_n\in q_i|x_1\in q_j) & \cdots & m_{ij}|p(x_n\in q_i|x_n\in q_j) \end{bmatrix} \tag{2}$$

式中: m_{ij}为分位数q_i转移到分位数q_j的概率, 即$m_{ij}=p(q_i\to q_j)$。

MTF 通过式(1)和式(2)编码方式将时间序列可视化, 转换为具有时间依赖性的二维特征图。由于分位数转移概率不同, 所以生成 MTF 图像像素点的颜色有深有浅, 将其作为输入可以充分利用神经网络在图像分类中的优势。且一维信号与 MTF 二维特征图为映射关系, 避免一维信息丢失。

1.2　GhostCNN

传统卷积计算量大, 参数多, 网络模型复杂, 提取特征耗时。为提升诊断效率, 引入 Ghost[19] 卷积, 其参数量少, 网络结构轻盈, 计算复杂度低, 可以用更少参数生成更多的特征图, 本文基于 Ghost 卷积构建 GhostCNN。首先, GhostCNN 采用普通的 1×1 卷积对输入的图片进行通道数的压缩, 浓缩特征信息; 然后, 使用 3×3 逐层卷积经过线性变换得到更多的特征图; 最后, 将 2 次卷积所得到的特征图在通道维度上进行堆叠, 组合成新的输出。图 1(a) 为普通卷积生成不含冗余的特征图的过程, 图 1 (b) 为生成的特征图经过恒等运算和廉价的线性运算生成完整特征图的过程, 其中, Identity 表示恒等运算, Φ_1, Φ_2, $\cdots$, Φ_k表示线性运算, Conv 表示卷积。

假设输入尺寸为$H\times W\times C$, 则输出为$H'\times W'\times$

N，其中，H和W分别为特征图的高度和宽度，C为输入通道数，N为输出通道数，卷积核的大小为$K\times K$。则在线性运算中，若特征图的通道数为M'，每个通道数对应的恒等式映射为S'，则得到$N=M'\times S'$特征映射，由于在运算过程中存在一个恒等运算，所以实际的线形运算为$M'\times(S'-1)=N/S'\times(S'-1)$，则GhostCNN与普通卷积的计算量之比如下：

$$R=\frac{\frac{N}{S'}\times H'\times W'\times C\times K\times K+(S'-1)\times\frac{N}{S'}\times H'\times W'\times K\times K}{N\times H'\times W'\times C\times K\times K}=\frac{S'+C-1}{S'\times C} \tag{3}$$

式中：因为S'和C的值大于1，所以R的值小于1，即GhostCNN的计算量相比于普通卷积要小。

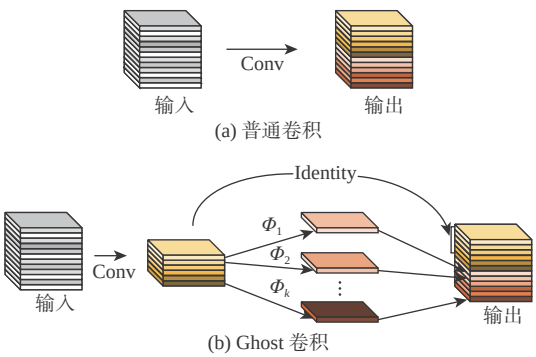


图1 特征生成过程

Fig. 1 Feature generation process

GhostCNN与传统卷积相比主要有以下优点：①降低了目前深度神经网络的计算量，可以从廉价的运算中生成更多的特征映射；②GhostCNN中恒等运算与线性运算并行，很好地保持了与普通卷积输出一致的内在特征图。

1.3 混合域注意力机制的提出

为充分提取特征信息，提出混合域注意力机制，由通道注意力机制[20]和空间注意力机制[21]组合而成，其中，通道注意力机制重点关注特征通道间的信息，增强关键通道特征，抑制无用通道特征，防止过拟合。空间注意力机制通过变换特征空间，捕获特征图任意2个位置之间的空间依赖性，提升网络适应性。

1.3.1 通道注意力机制

通道注意力机制通过获取每个特征通道的重要程度来增强关键特征的权重，减弱无用信息的权重。由于SENet[20]仅用全局平均池化提取到的特征信息单一，故本文在全局平均池的基础上加入全局标准差池，增加特征通道多样性的同时提高特征信息的全面性，充分聚合空间信息。用全连接层捕获通道间的依赖关系，由此提出改进的通道注意力机制（简称ISENet）。ISENet从通道维度利用输入特征的均值和标准差来提高网络捕获全局信息的能力，增加通道特征提取多样化。其结构如图2所示，通道数为C的输入特征$\boldsymbol{X}$在经过ISENet之后，输出特征为$\boldsymbol{Y}$，$\boldsymbol{Y}$的每个通道的重要程度均发生变化，从而使神经网络重点关注权重值大的通道。

实现过程如下：

1）进行特征压缩将原始特征压缩到输出特征空间$Z\in\mathbf{R}^{C\times H\times W}$，其中，$x_C\in\mathbf{R}^{H\times W}$表示输入特征$\boldsymbol{X}$在通道中的特征映射，$F_{\mathrm{sq}}(\cdot)$和$F'_{\mathrm{sq}}(\cdot)$均表示压缩特征。ISENet将每个特征通道经全局平均池化和全局标准差池化压缩为$C\times1\times1$，其表达式分别为

$$z_{\mathrm{c}}=\left(F_{\mathrm{sq}}x_C\right)=\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}x_C(i,j) \tag{4}$$

$$\mu_{\mathrm{c}}=F'_{\mathrm{sq}}(x_C)=\sqrt{\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}(x_C(i,j)-z_{\mathrm{c}})^2} \tag{5}$$

2）进行激励以降低计算量，并通过2个全连接层构建通道间的相关性，为所有通道上的特征映射生成相应的通道关注权值。其表达式为

$$S=F_{\mathrm{ex}}(Z,W)=\sigma(W_1\delta(W_0Z)) \tag{6}$$

式中：$F_{\mathrm{ex}}(\cdot,W)$表示激励映射；W_0和W_1分别为第1全连接层和第2全连接层的参数；δ和σ分别为Relu和Sigmoid函数。

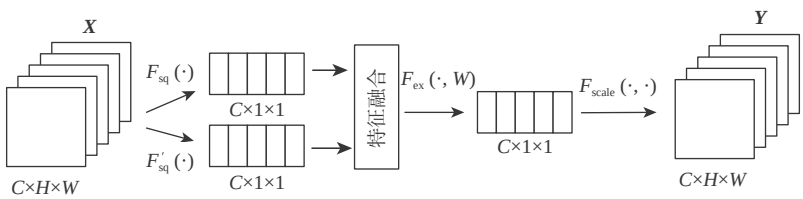


图2 改进通道注意力机制

Fig. 2 Improved channel attention mechanism

3) 对S和$\boldsymbol{X}$进行尺度加权, 得到输出特征张量$\boldsymbol{Y}=\left[y_1,y_2,\cdots,y_C\right]\in\mathbf{R}^{C\times H\times W}$。其元素表达式为

$$y_C=F_{\text{scale}}(x_C,s_C)=x_C s_C \tag{7}$$

式中: $F_{\text{scale}}(\cdot,\cdot)$为通道注意力加权映射函数。

1.3.2　空间注意力机制

仅从通道维度提取特征会忽略特征空间信息, 为提取特征空间维度信息, 本文提出一种空间注意力机制, 重点关注特征空间区域信息。其结构如图 3 所示。

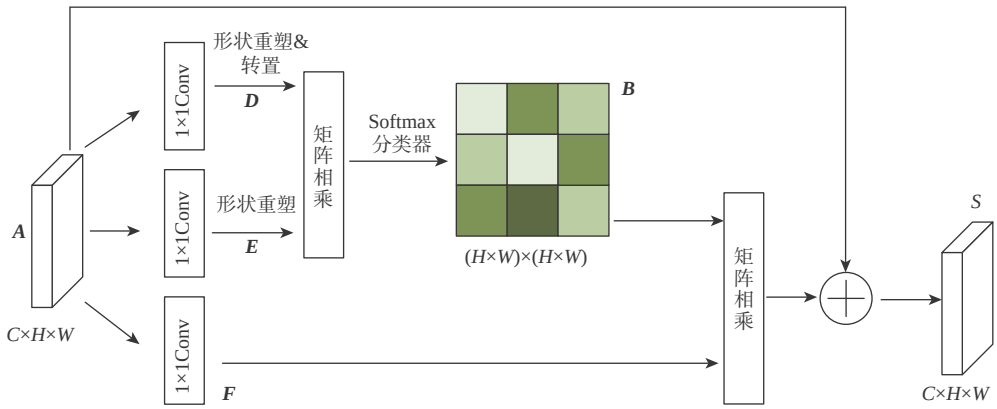


图 3　空间注意力机制

Fig. 3　Spatial attention mechanism

其实现过程如下:

1) 将特征图$\boldsymbol{A}\in\mathbf{R}^{C\times H\times W}$进行线性计算, 得到$\boldsymbol{D}$、$\boldsymbol{E}$、$\boldsymbol{F}$这 3 个特征, $\{\boldsymbol{D},\boldsymbol{E},\boldsymbol{F}\}\in\mathbf{R}^{C\times H\times W}$;

2) 将$\boldsymbol{D}$重塑为$HW\times C$, 将$\boldsymbol{E}$、$\boldsymbol{F}$重塑为$C\times HW$;

3) 将转置后的$\boldsymbol{D}$和$\boldsymbol{E}$进行矩阵相乘, 并通过 Softmax 归一化权重, 得到空间注意力图$\boldsymbol{B}\in\mathbf{R}^{(H\times W)\times(H\times W)}$。矩阵相乘能宏观体现出 2 个位置之间的相似度, 位置相似度越高, 空间注意力图B_{ji}的值越大, 计算式为

$$B_{ji}=\exp\left(D_i,E_j\right)\Big/\sum_{i=1}^{N'}\exp\left(D_i,E_j\right) \tag{8}$$

式中: $N'=H\times W$;

4) 将$\boldsymbol{F}^{\mathrm{T}}$与位置注意力图$\boldsymbol{B}^{\mathrm{T}}$相乘, 并将结果重塑为$\mathbf{R}^{C\times H\times W}$;

5) 将$\mathbf{R}^{C\times H\times W}$与特征$\boldsymbol{A}$进行特征融合得到输出$S\in\mathbf{R}^{C\times H\times W}$,计算式为

$$S_j=\sum_{i=1}^{N'}\left(B_{ji}F_i\right)+A_j \tag{9}$$

空间注意力机制有以下优点: ①帮助模型忽略无关紧要的区域信息, 从而聚焦关键区域, 提高模型的识别和理解准确度; ②空间注意力机制强调重点区域信息, 增强模型的可解释性和通用性。

1.3.3　混合域注意力机制结构

结合改进通道注意力机制和空间注意力机制提出混合域注意力机制, 其结构如图 4 所示。

输入特征图经改进通道注意力机制进行通道间信息的交互, 获取有用特征的通道, 增强网络性能。经空间注意力机制, 将输入特征图进行空间变换, 增强网络对特征图中关键区域的空间感受能力, 提高网络的泛化性和适应性。将改进通道注意力机制和空间注意力机制及输入特征图进行融合, 得到输出特征图。

在模型中输入特征图经改进通道注意力和空间注意力作用后输出特征如式(7)和式(9)所示, 则混合注意力机制输出特征如下:

$$M_{\text{h}}(F)=\text{GAvgpool}\left(y_C+S_j\right)\sigma \tag{10}$$

式中: GAvgpool为全局平均池化函数。

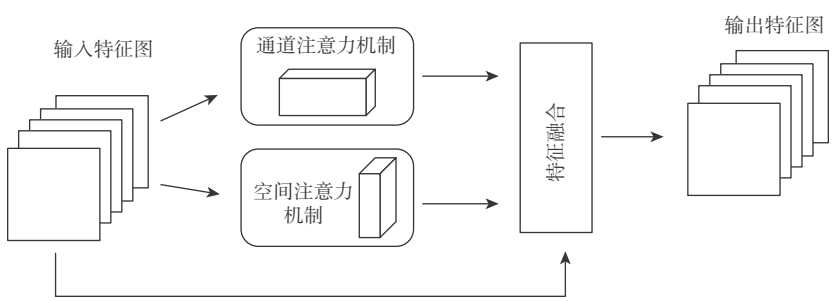


图 4　混合域注意力机制结构

Fig. 4　Mixed domain attention mechanism structure

将通道注意力和空间注意力结合，提出混合域注意力机制，可以使网络既能关注输入数据中的通道维度，又能够更好地将注意力聚集在特定的区域上，提高诊断精度。将 MDAM 嵌入 GhostCNN 中，逐层强化网络训练性能，提高网络对特征的提取能力，更加适用于轴承故障诊断中的二维特征提取。

2 基于 MDAM-GhostCNN 的滚动轴承故障诊断方法

2.1 MDAM-GhostCNN 模型搭建

本文从基于模型的策略出发[22]，对神经网络结构进行简化，基于轻量化网络结构 Ghost 模块提出 GhostCNN，其通过廉价线性计算生成大量特征图，减少了模型参数，提升模型诊断效率的同时增强模型的特征提取能力，为整个故障诊断工作夯实基础。并且，本文基于特征学习策略，结合改进通道注意力机制和空间注意力机制提出一种混合域注意力机制，MDAM 能够从通道和空间 2 个维度更准确地理解输入数据，充分提取关键特征信息，削弱无关紧要信息，提高模型的鲁棒性和适应性，有效解决了普通卷积在噪声和变工况下故障诊断准确率较低、计算复杂、效率低的问题。基于此，本文提出一种基于 MDAM-GhostCNN 的滚动轴承故障诊断方法，MDAM-GhostCNN 的故障诊断模型结构如图 5 所示。

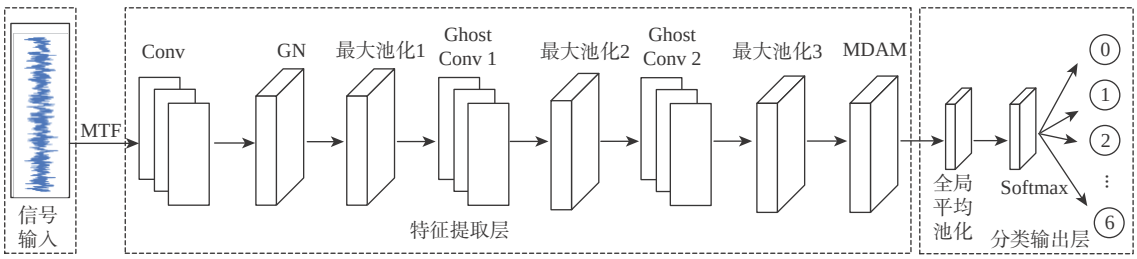


图 5 MDAM-GhostCNN 故障诊断模型

Fig. 5 MDAM-GhostCNN fault diagnosis model

MDAM-GhostCNN 模型主要诊断过程可分为 3 部分：信号输入、特征提取、分类输出。①该模型轴承原始的振动信号经 MTF 转化为具有时间相关性的二维特征图；其次，将 MTF 特征图输入卷积核大小为5×5的卷积层提取特征，并通过群组归一化层 (group normalization, GN) 进行群组归一化。②将群组归一化后的特征通过最大池化层，扩大感知野；最大池化层后接 Ghost Conv1，从特征图冗余的角度出发，可以通过少量计算就可以生成大量特征图；随后进行相同的归一化、池化处理。③将处理后的特征输入 MDAM 模块，由改进通道注意力和空间注意力构成，其中，改进通道注意力机制重点关注特征通道信息，获取有用特征的通道；空间注意力机制重点关注特征空间区域信息，改进通道注意力和空间注意力同时对输入特征进行信息处理，提高网络的泛化性和适应性；将改进通道注意力和空间注意力并联，经全局平均池化整合特征全局空间信息，并防止过拟合。④用 Softmax 分类函数进行故障类型的分类，并输出诊断分类结果。

2.2 MDAM-GhostCNN 模型结构

MDAM-GhostCNN 的模型结构参数如表 1 所示。选择 Relu 激活函数增强模型非线性能力，提高模型稀疏性。使用 GN 进行归一化处理，加快网络收敛速度。用全局平均池化代替全连接层使网络结构正则化，防止过拟合。

表 1 MDAM-GhostCNN 结构

Table 1 MDAM-GhostCNN structure

特征层	卷积核数量	卷积核大小	输出大小
输入			(128, 128, 1)
卷积层	32	5×5	(128, 128, 32)
GN			(128, 128, 32)
最大池化1	32	2×2	(64, 64, 32)
Ghost Conv 1	32	3×3	(64, 64, 32)
GN			(64, 64, 32)
最大池化2	32	2×2	(32, 32, 32)
Ghost Conv 2	32	3×3	(32, 32, 32)
GN			(32, 32, 32)
最大池化3		2×2	(16, 16, 32)
MDAM			(16, 16, 32)
全局平均池化			(1, 1, 32)
分类器	7		(7)

2.3 MDAM-GhostCNN 模型诊断流程

采用 MDAM-GhostCNN 模型进行滚动轴承的故障诊断主要有以下 4 个步骤：数据预处理、模型搭建及参数设置、模型训练、故障分类识别，其流程如图 6 所示。

具体步骤如下：

步骤 1 数据预处理。对采集的原始振动信号

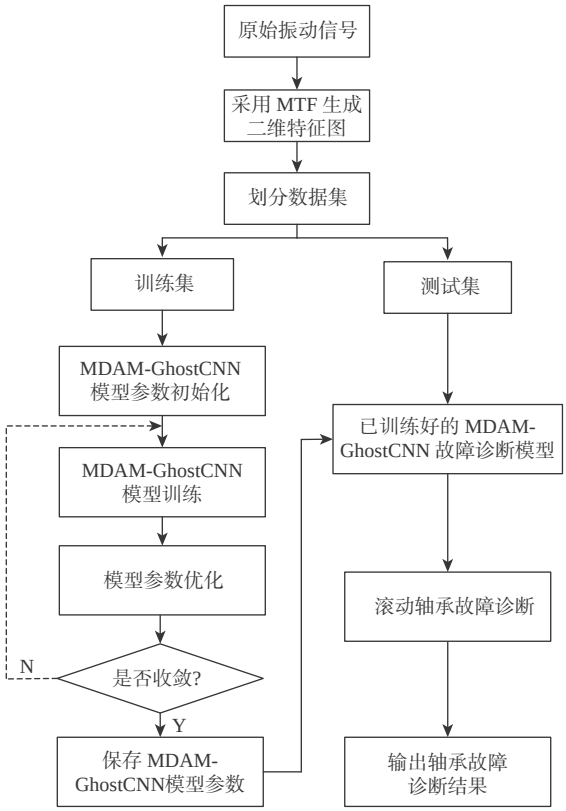


图 6　MDAM-GhostCNN 模型故障诊断流程图

Fig. 6　MDAM-GhostCNN model fault diagnosis flow chart

进行数据预处理, 用 MTF 转换为具有时间序列的二维特征图, 并将样本按一定比例划分为训练集和测试集。

步骤 2　模型搭建及参数初始化。依据 2.1 节搭建的 MDAM-GhostCNN 模型, 设置学习率、批量大小、训练轮数、卷积核大小等相关参数。

步骤 3　模型训练。将训练集输入到网络中进行训练并判断模型是否收敛, 若不收敛, 则优化模型参数, 继续训练。若收敛, 则完成训练, 保存模型参数。

步骤 4　故障分类识别。将测试集输入到优化后的 MDAM-GhostCNN 模型中进行滚动轴承的故障诊断, 并输出分类结果及识别准确率。

3　实验验证

3.1　实验平台介绍

本文实验环境为 Python3.7 解释器, Keras2.4.3 和 Tensorflow2.3.0 深度学习框架搭建模型。在 GTX1660GPU, 英特尔 I7-9700F 处理器上进行实验验证。

3.2　数据集说明和预处理

本文选用 CWRU[23] 公开轴承数据集来进行实验验证。实验装置示意图如图 7 所示。

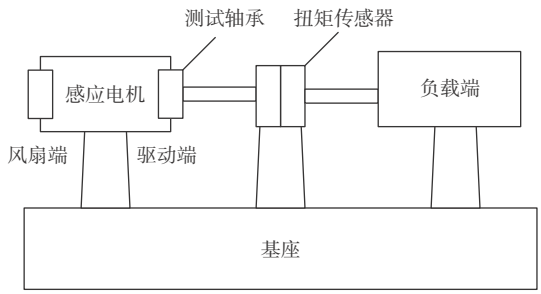


图 7　轴承实验装置示意图

Fig. 7　Schematic diagram of bearing experimental device

实验选用驱动端轴承 SKF6205 的深沟球轴承, 轴承故障为电火花加工单点故障。故障轴承的直径分别为 0.18 mm、0.36 mm, 实验转速为 11 285.16 rad/min, 在 12 kHz 采样频率下分别采集负载为 0.745 7, 1.491 4, 2.237 1 kW 驱动端轴承的内圈、外圈、滚动体故障, 外加一种正常状态, 总共有 7 种类别。

为验证本文模型在变工况下能够达到更高的分类精度, 样本长度取为 2 048 个数据点, 训练集用 240 个样本, 测试集用 100 个样本。CWRU[23] 滚动轴承数据训练集和测试集如表 2 所示, 数据集 A、数据集 B 和数据集 C 的负载分别为 0.745 7, 1.491 4, 2.237 1 kW。

3.3　模型最优参数选取

通过对表 2 中训练集 C 的多次验证, 选择了几组参数进行训练, 并对结果进行比较。对 MDAM-

表 2　CWRU[23] 滚动轴承数据集

Table 2　CWRU [23]rolling bearing data set

损伤位置	标签	损伤直径/mm	样本数量					
			数据集A		数据集B		数据集C	
			训练集	测试集	训练集	测试集	训练集	测试集
正常	0	0	240	100	240	100	240	100
内圈	1	0.18	240	100	240	100	240	100
	2	0.36	240	100	240	100	240	100
外圈	3	0.18	240	100	240	100	240	100
	4	0.36	240	100	240	100	240	100
滚动体	5	0.18	240	100	240	100	240	100
	6	0.36	240	100	240	100	240	100

GhostCNN的学习率、批量大小、卷积核大小进行调整和选择。为快速测试,本节只在一种工况下测试,同时,在每次测试中,除被测参数外其余参数均保持不变。

3.3.1 学习率的选择

学习率作为深度学习中重要的超参数,决定着模型能否收敛到最小值及何时收敛到最小。学习率过大会导致网络无法收敛,模型精度降低;学习率过小会延长网络收敛时间,降低模型训练速度。本文选择常见的学习率进行测试,由图8(a)可知,当学习率为0.000 1时模型精度最高。

3.3.2 批量大小选择

批量大小影响模型的优化程度和训练速度。批量过小会使训练时间变长,训练效率低;过大则会导致模型泛化能力下降。在本节中分别选择批量大小为2、4、8、10、12、16、32来测试MDAM-GhostCNN模型。其结果如图8(b)所示。当其值为12时,模型精度达到最大值,在达到最高精度后,随着数量的增多,准确率开始下降,故批量大小选择12。

3.3.3 卷积核大小选择

卷积核大小对测试精度有很大的影响。在本节中选择卷积核大小为3、5、7、9、11、13进行测试。测试结果如图8(c)所示,由图可知,当卷积核大小为5时准确率最高为99.03%,准确率均高于其他尺寸,故选择卷积核大小为5。

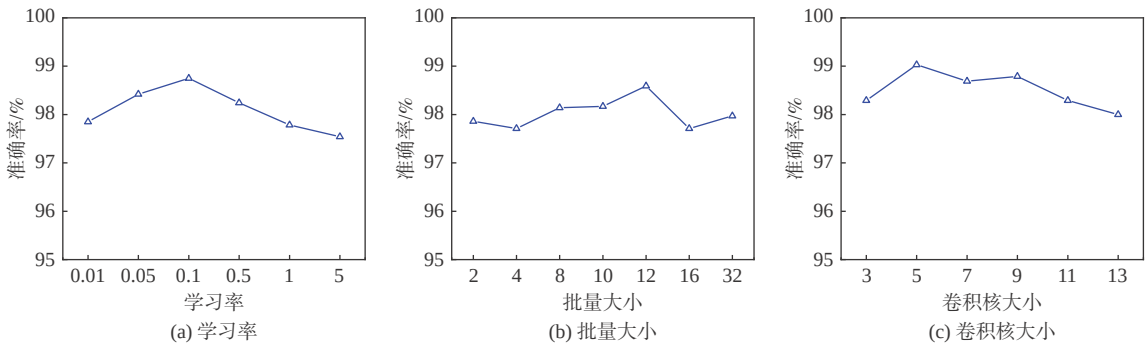


图8 不同参数对诊断精度的影响

Fig. 8 Effects of different parameters on diagnostic accuracy

3.4 变工况下模型的泛化性分析

为验证本文模型在滚动轴承故障诊断领域的有效性,对GhostCNN、宽卷积核卷积神经网络(convolutional neural network with wide convolution kernels, WKCNN)[24]、ResNet[25]、多分支深度可分离卷积神经网络(multi-branch depth separable CNN, MBDS-CNN)[26]、ILeNet-5[27]及本文模型在变工况下进行准确率的测试。为减小误差确保模型诊断的准确率,分别进行10次实验计算平均值。其中,GhostCNN是在MDAM-GhostCNN的基础上去掉混合域注意力机制,选取GhostCNN是为证明MDAM模块的有效性,故消融试验选用GhostCNN。MBDS-CNN[26]的输入也是二维图像,与本文MTF二维特征图输入形成对比。WKCNN[24]是采用加宽卷积核的策略获取更大的感受野的方式搭建模型,宽卷积核与本文模型的常规卷积核形成对比。ResNet[25]是目前应用最广泛的卷积神经网络,已成为故障诊断的研究热点,可与本文模型进行对比。ILeNet-5[27]是一种改进的LeNet-5轴承故障诊断模型。

在实际工业环境中,轴承变负载及噪声是无法避免的,变工况下模型具有良好的诊断性能是确保滚动轴承安全工作的前提,为验证本文模型在变工况条件下的优越性和有效性,在CWRU[23]轴承数据集上进行故障分类识别验证。

通过CWRU[23]轴承数据验证模型在不同负载下的泛化性能,选取0.745 7, 1.491 4, 2.237 1 kW这3种负载下的轴承数据进行验证。数据集信息如表2所示。

图9中A-B表示数据集A为训练集,数据集B为测试集,A、B为不同负载工况的数据集。由图9可知,MDAM-GhostCNN的故障分类平均准确率为99.46%,均高于表中其他5种轴承故障诊断模型,与未添加MDAM的GhostCNN相比准确率提高2.21%,比WKCNN[24]宽卷积核模型高0.69%,较MBDS-CNN[26]高0.75%,较ILeNet-5[27]提高1.14%,比ResNet[25]提高1.99%。故由图9可知,在不同负载下,MDAM-GhostCNN具有优异的分类准确度和泛化性能。

为更直观体现本文模型在训练样本时的诊断状况,其在CWRU[23]轴承测试集上的分类结果的混淆矩阵如图10所示,由图10(a)可知,700个测试集样本中只有3个被分错,具体为3个外圈中度故障样本(标签3)被错分成滚动体中度故障(标签5)。图10(b)中未出现错分样本。以上结果表明,MDAM-

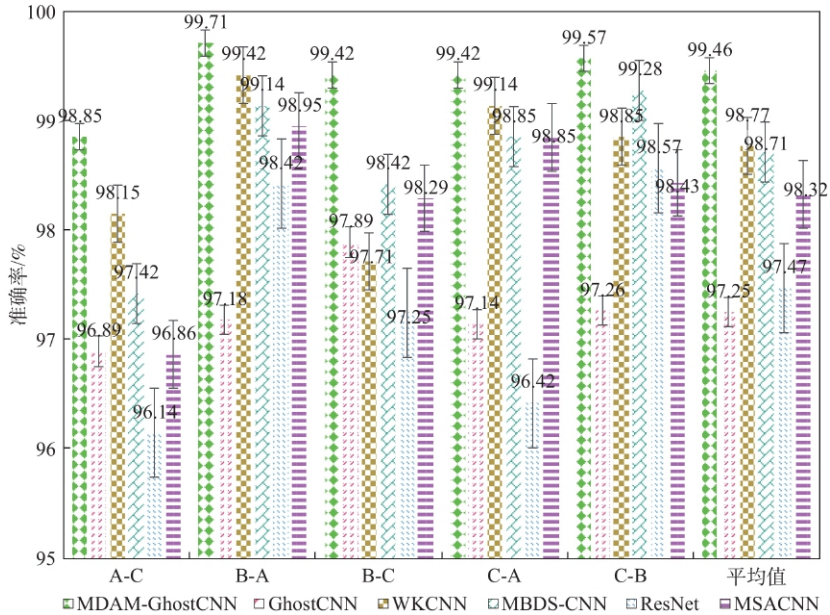


图 9　CWRU[23] 轴承数据集在变负载下不同模型的分类精度

Fig. 9　Classification accuracy of CWRU[23] bearings data set for different models under variable loads

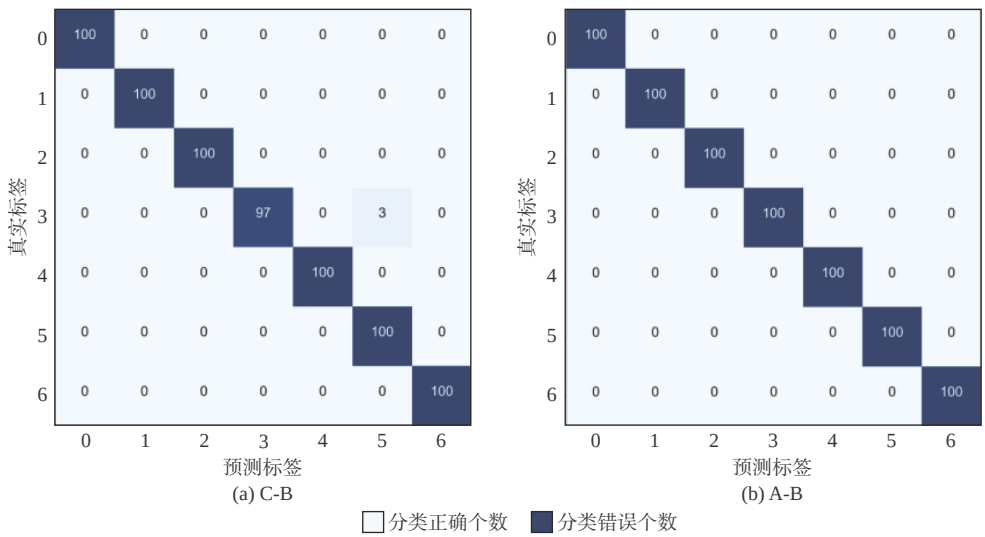


图 10　故障分类结果的混淆矩阵

Fig. 10　Confusion matrix of fault classification results

GhostCNN 模型在不同负载条件下具有良好的识别精度。

3.5　不同模型参数量和训练时间分析

本文模型和其他 5 种模型在数据集 A 上的参数量和训练时间如表 3 所示。训练轮数为 30, 训练集和测试集的样本数量分别为 240 和 100。由表 3 可以看出, 本文模型与参数量最少的 GhostCNN 相比在参数量多 110 800 的情况下, 训练时间仅仅多了 0.11 s, 且参考图 9 中的分类精度, 本文模型的诊断准确率与 GhostCNN 相比提高了 2.21%, 与耗时最长的 ILeNet-5 相比, 训练时间短 0.41 s, 与 MBDS-CNN 相比时间减少 0.97 s, 与 WKCNN 相比参数量少 11 500, 训练时间短 0.25 s, 与 ResNet相比参数量少 73 400, 训练时间短 0.61 s。综上可以说明本文模型参数量较少、训练时间短、模型复杂度低、计算效率高。

表 3　不同模型参数量和训练时间

Table 3　Number of Different model parameters and training time

模型	参数量	训练时间/s
本文模型	11 500	2.84
MBDS-CNN[26]	149 300	3.81
WKCNN[24]	26 900	3.09
ResNet[25]	84 900	3.45
GhostCNN	700	2.73
ILeNet-5[27]	40 900	3.25

3.6 模型抗噪性能分析

在实际的工业环境中，噪声是无法避免的，为更精准地分析模型的抗噪性能，对A-B工况加入高斯白噪声，其中，信噪比为

$$V_{\mathrm{SNR}}=10\lg\frac{P_{\mathrm{s}}}{P_{\mathrm{n}}} \tag{11}$$

式中：P_{s}为信号的功率；P_{n}为噪声的功率。

本节选用0, 2, 4, 6 dB的噪声信号对本文模型及对比模型进行测试，各模型分别测试10次并计算平均值，试验结果如图11所示。由图可知，本文模型在不同信噪比下的平均诊断准确率均高于其他5种模型，具有较好的抗噪性。可以看到信噪比为0 dB时，本文模型平均精度为88.37%，相比于较差的WKCNN[24]高4.57%，随着信噪比的增大，各模型诊断精度均在提高，当信噪比为6 dB时，本文模型平均准确率为99.57%，比效果最好的MBDS-CNN[26]高0.14%。由此可知，相较于其他对比模型，本文模型具有更好的抗噪性能和泛化性能。

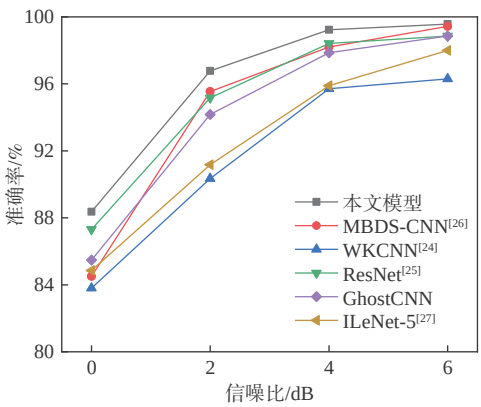


图11 不同模型的信噪比

Fig. 11 Signal-to-noise ratio of different models

3.7 模型训练过程分析

为进一步说明变负载工况下模型的分类效果，引入t-SNE非线性降维算法对模型每一个卷积层及添加MDAM提取到的测试样本的特征进行可视化。以A-B工况为例，结果如图12所示。由图可知，随着GhostCNN和MDAM的加入，模型对故障的分类精度越来越高。在图12(a)中当经过最原始的信号时，特征是不可分的，且特征样本严重重叠，经过图12(b)卷积层1后仅有故障直径为0.36 mm内圈故障被识别；当经过图12(c)Ghost Conv1时，虽有4类被识别分类，分别是轴承故障直径为0.36 mm的内圈故障、轴承故障直径为0.36 mm的滚动体故障、正常状态、轴承故障直径为0.18 mm的外圈故障，但轴承故障直径为0.36 mm的内圈故障与轴承故障直径为0.18 mm的滚动体故障间隔较小，可分性不高，且其余3类样本仍呈现混叠状；当经过图12(d)Ghost Conv2层时，只有少许特征样本出现交叉重叠，且各类故障间隔也在变大；最后，当经过加入混合域注意力机制图12 (e)进行特征权重分配后，各类特征样本呈现聚集状态，且不存在混叠交叉样本，所有类别均被正确识别分类，样本可分性

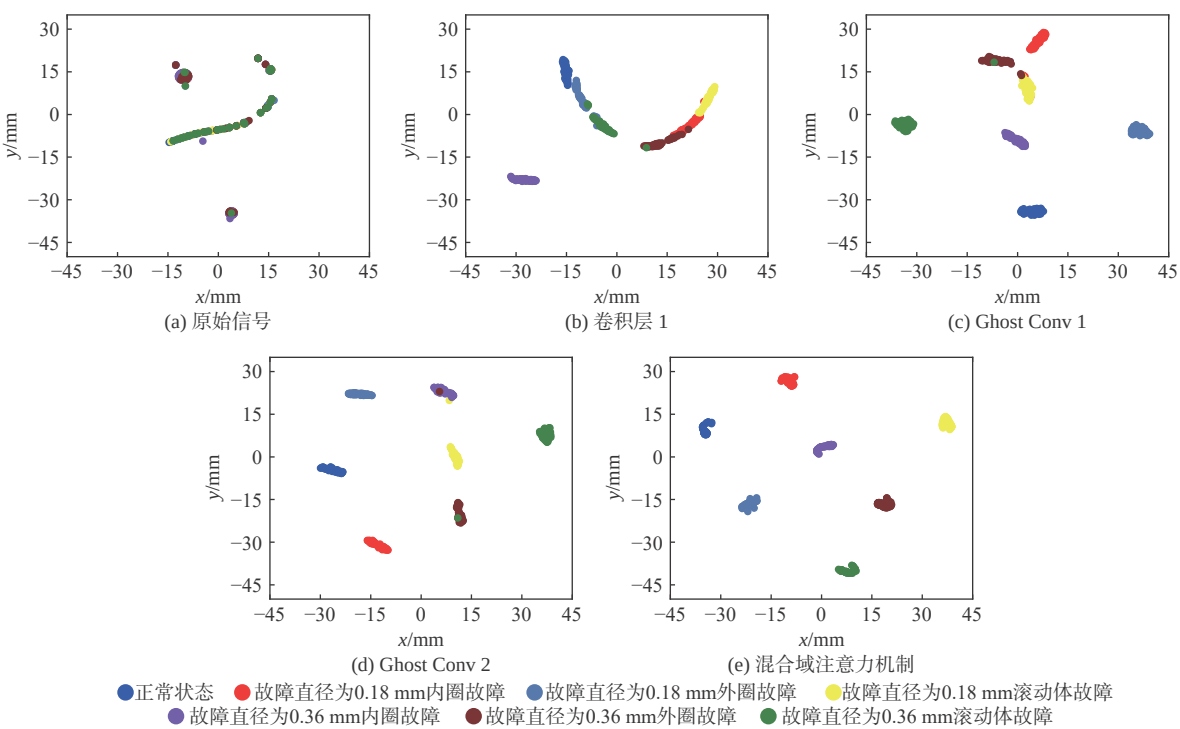


图12 模型训练过程可视化

Fig. 12 Visualization of model training process

提高, 故可实现轴承故障的精确分类。

4　JNU 轴承数据集验证

为进一步验证本文方法的有效性, 对 JUN 轴承数据(详见 http://mad-net.org:8765/explore.html?t=0.%205853151655584721 2)进行了故障分类识别测试。

JNU 轴承数据集采样频率为 50 kHz, 包含有内圈、外圈和滚动体 3 种故障状况外加正常状态共 4 种状况。实验过程中, 在 3 768, 5 024, 6 280 rad/min 这 3 种不同转速下采集振动信号。不同转速被认为不同任务, 分别将转速为 3 768, 5 024, 6 280 rad/min 的振动数据定义为工况 F1、F2 和 F3。根据轴承故障类型不同, 每种转速故障数据分为 4 类。JNU 滚动轴承数据训练集和测试集表 4 所示。

表 4　JNU 轴承数据集说明

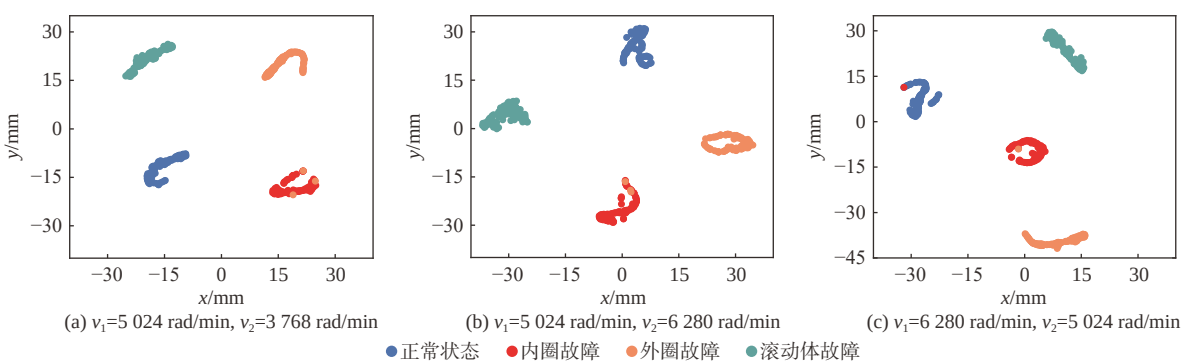
Table 4　JNU rolling bearing data set description

损伤位置	标签	样本数量					
		工况F1		工况F2		工况F3	
		训练集	测试集	训练集	测试集	训练集	测试集
正常	0	240	100	240	100	240	100
内圈	1	240	100	240	100	240	100
外圈	2	240	100	240	100	240	100
滚动体	3	240	100	240	100	240	100

为验证本文模型在变转速下能够达到更高的分类精度, 样本长度取为 2 048 个数据点, 训练集用 240 个样本, 测试集用 100 个样本。在使用不同的模型学习特征之后, 还使用 t-SNE 来获得如图 13 所示的特征效果的二维可视化图。显然, 本文模型在 JNU 轴承数据集上也有良好的分类效果。其中, v_1 为训练集转速, v_2 为测试集转速。

通过 JNU 轴承数据集验证 MDAM-GhostCNN 模型在不同转速下的泛化性, 数据采集如表 4 所示。由图 14 可知, MDAM-GhostCNN 模型在 JNU 轴承数据集上的故障平均分类准确率为 98.99%, 均高于其他 5 种轴承故障诊断模型, 与未添加 MDAM 的 GhostCNN 相比准确率提高 2.34%, 比 WKCNN 宽卷积核模型高 0.63%, 较 MBDS-CNN 高 0.8%, 比 ResNet 提高 1.52%, 与 ILeNet-5 模型相比提高 1.88%, 且从图 14 中的误差线可以看出, MDAM-GhostCNN 的稳定性均优于其他诊断模型。综上, MDAM-GhostCNN 在不同转速下具有更好的泛化性能。

为更精准地分析模型的抗噪性能, 对 F1 和 F2 工况下分别加入 0, 2, 4, 6 dB 的高斯白噪声, 对本文模型及其余 5 种对比模型进行噪声测试。各模型分别测试 10 次并计算平均值, 实验结果如图 15 所示。由图 15 可知, 本文模型在不同信噪比下的平均诊断准确率均高于其他 5 种模型, 具有较好的抗噪性。可以看到信噪比为 0 dB 时, 本文模型平均精度为 84.75%, 比 GhostCNN 高 4.16%, 比 ResNet 高 3.3%, 比 MBDS-CNN 高 1.87%, 比 WKCNN 高 0.9%, 比 ILeNet-5 高 1.81%。而当信噪比为 6 dB 时, 本文模型的信噪比达到了 98.89%, 比效果最好的 ResNet 高 0.61%, 比效果最差的 ILeNet-5 高 2.72%。由此可知, 相较于其他对比模型, 本文模型具有更



(a) v_1=5 024 rad/min, v_2=3 768 rad/min　(b) v_1=5 024 rad/min, v_2=6 280 rad/min　(c) v_1=6 280 rad/min, v_2=5 024 rad/min

图 13　JNU 轴承数据模型训练可视化

Fig. 13　Visualization of JNU bearing data model training

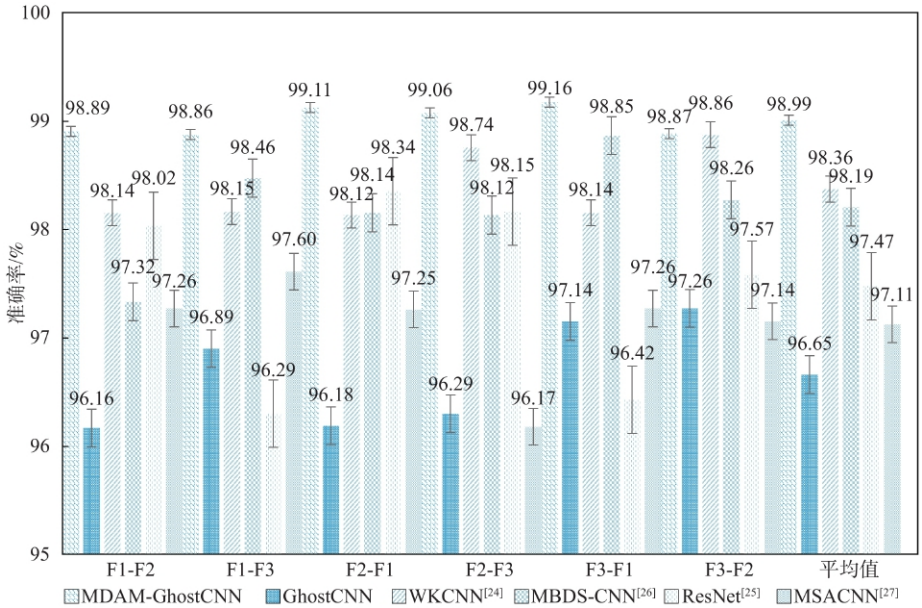


图 14 JNU 数据集轴承变转速下不同模型的分类精度

Fig. 14 Classification accuracy of different models in bearing variable working conditions of JNU data set

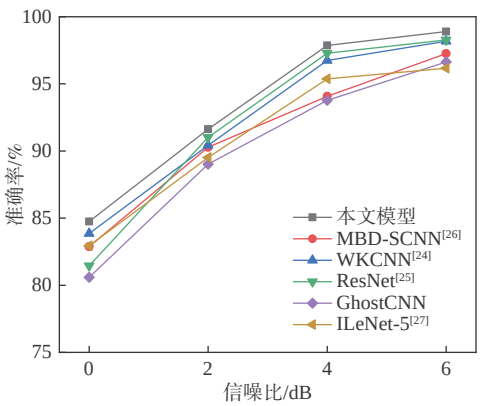


图 15 JNU 数据集上不同模型的信噪比

Fig. 15 Signal-to-noise ratio of different models on the JNU dataset

好的抗噪性能和泛化性能。

5 结 论

本文提出一种名为 MDAM-GhostCNN 的新模型，以解决轴承在变工况、噪声环境下故障诊断识别准确率低的问题。

1）本文模型采用 MTF 获取具有时间依赖性的二维特征图，最大限度地保留了原始信号的时间信息，使神经网络提取到更多的特征。

2）本文模型运用 Ghost 卷积来解决传统神经网络参数量多、运行速度慢及计算复杂的问题。

3）本文模型嵌入了混合域注意力机制以增强 GhostCNN 的特征提取能力，充分提取特征信息。

4）在噪声环境、变负载、变转速下，本文 MDAM-GhostCNN 模型在 CWRU 及 JNU 轴承数据集上均具有较好的泛化性、鲁棒性及抗噪性。

参考文献（References）

[1] HAN T, LIU C, YANG W G, et al. Deep transfer network with joint distribution adaptation: a new intelligent fault diagnosis framework for industry application[J]. ISA Transactions, 2020, 97: 269-281.

[2] HE S Y, HU D Y, YU G, et al. Trackside acoustic detection of axle bearing fault using wavelet domain moving beamforming method[J]. Applied Acoustics, 2022, 195: 108851.

[3] HOU M X, SHI H T. Stator-winding incipient shorted-turn fault detection for motor system in motorized spindle using modified interval observers[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 3505716.

[4] LIU R N, YANG B Y, ZIO E, et al. Artificial intelligence for fault diagnosis of rotating machinery: a review[J]. Mechanical Systems and Signal Processing, 2018, 108: 33-47.

[5] 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).

[6] HOANG D T, KANG H J. A survey on deep lelarning based bearing fault diagnosis[J]. Neurocomputing, 2019, 335: 327-335.

[7] SCHOLARPEDIA G E H. Deep belief networks[J]. Scholarpedia the Peer-Reviewed Open-Access Encyclopedia, 2009, 4(5): 5947.

[8] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323: 533-536.

[9] XU Y, LI Z X, WANG S Q, et al. A hybrid deep-learning model for fault diagnosis of rolling bearings[J]. Measurement, 2021, 169: 108502.

[10] ZHANG Z Q, YANG Q Y, ZI Y Y. Multi-scale and multi-pooling sparse filtering: a simple and effective representation learning method for intelligent fault diagnosis[J]. Neurocomputing, 2021, 451: 138-151.

[11] CHEN Z Y, MAURICIO A, LI W H, et al. A deep learning method

for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks[J]. Mechanical Systems and Signal Processing, 2020, 140: 106683.

[12] XU Z F, LI C, YANG Y. Fault diagnosis of rolling bearing of wind turbines based on the variational mode decomposition and deep convolutional neural networks[J]. Applied Soft Computing, 2020, 95: 106515.

[13] YE M Y, YAN X A, CHEN N, et al. Intelligent fault diagnosis of rolling bearing using variational mode extraction and improved one-dimensional convolutional neural network[J]. Applied Acoustics, 2023, 202: 109143.

[14] 刘俊锋, 俞翔, 万海波, 等. 基于 MFMD 和 Transformer-CNN 的滚动轴承故障诊断方法[J]. 航空动力学报, 2023, 38(6): 1446-1456.
LIU J F, YU X, WAN H B, et al. Fault diagnosis method of rolling bearing using MFMD and Transformer-CNN[J]. Journal of Aerospace Power, 2023, 38(6): 1446-1456(in Chinese) .

[15] FU W L, JIANG X H, LI B L, et al. Rolling bearing fault diagnosis based on 2D time-frequency images and data augmentation technique[J]. Measurement Science and Technology, 2023, 34(4): 045005.

[16] 金江涛, 许子非, 李春, 等. 基于卷积双向长短期记忆网络与混沌理论的滚动轴承故障诊断[J]. 振动与冲击, 2022, 41(17): 160-169.
JIN J T, XU Z F, LI C, et al. Fault diagnosis of rolling bearing based on CCNN-BiLSTMN method[J]. Journal of Vibration and Shock, 2022, 41(17): 160-169(in Chinese).

[17] MAO G, ZHANG Z Z, QIAO B, et al. Fusion domain-adaptation CNN driven by images and vibration signals for fault diagnosis of gearbox cross-working conditions[J]. Entropy, 2022, 24(1): 119.

[18] WANG M J, WANG W J, ZHANG X N, et al. A new fault diagnosis of rolling bearing based on Markov transition field and CNN[J]. Entropy, 2022, 24(6): 751.

[19] HAN K, WANG Y H, TIAN Q, et al. GhostNet: more features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 1577-1586.

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

[21] HUANG Z L, WANG X G, HUANG L C, et al. CCNet: criss-cross attention for semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 603-612.

[22] XU G W, LIU M, JIANG Z F, et al. Online fault diagnosis method based on transfer convolutional neural networks[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(2): 509-520.

[23] Case Western Reserve University. Case Western Reserve University bearing data[EB/OL]. (2018-05-22) [2023-04-28]. https://gitcode.com/open-source-toolkit/1cd2d/?utm_source=tools_gitcode&index=bottom&type=card&webUrl.

[24] 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.

[25] HAO X Y, ZHENG Y, LU L, et al. Research on intelligent fault diagnosis of rolling bearing based on improved deep residual network[J]. Applied Sciences, 2021, 11(22): 10889.

[26] 刘恒畅, 姚德臣, 杨建伟, 等. 基于多分支深度可分离卷积神经网络的滚动轴承故障诊断研究[J]. 振动与冲击, 2021, 40(10): 95-102.
LIU H C, YAO D C, YANG J W, et al. Fault diagnosis of rolling bearings based on multi-branch deep separable convolutional neural networks[J]. Vibration and Shock, 2021, 40(10): 95-102(in Chinese) .

[27] 吴晨芳, 杨世锡, 黄海舟, 等. 一种基于改进的 LeNet-5 模型滚动轴承故障诊断方法研究[J]. 振动与冲击, 2021, 40(12): 55-61.
WU C F, YANG S X, HUANG H Z, et al. An improved fault diagnosis method of rolling bearings based on LeNet-5[J]. Journal of Vibration and Shock, 2021, 40(12): 55-61(in Chinese).

Fault diagnosis method of rolling bearing based on MDAM-GhostCNN

GUO Junfeng，TAN Baohong*，WANG Zhiming

(School of Mechanical and Electronical Engineering，Lanzhou University of Technology，Lanzhou 730050，China)

Abstract：A rolling bearing fault diagnostic approach based on mixture domain attention mechanism (MDAM)-GhostCNN is developed to address the issues of inadequate feature extraction, complicated computation, and low recognition accuracy under varied working conditions in conventional fault detection methods. First of all, the Markov transfer field (MTF) is used to transform the bearing vibration signal into a two-dimensional feature graph with time correlation. Secondly, taking advantage of the simplification of Ghost convolution calculation, GhostCNN is constructed. Then, a MDAM is designed, which makes the network fully capture the feature information from the two dimensions of channel and space, and makes the network pay attention to the feature space information effectively while realizing the interdependence between the feature channels, and construct the MDAM-GhostCNN model. Finally, the MTF two-dimensional feature map is input into the MDAM-GhostCNN model for training and output diagnosis results. Experimental verification and noise processing were performed on the bearing data sets from Jiang Nan University (JNU) and Case Western Reserve University. The results show that under variable working conditions, the constructed model has higher recognition accuracy, noise immunity and generalization performance.

Keywords：rolling bearing；fault diagnosis；Markov transfer field；Ghost convolution；attention mechanism

Received：2023-05-04；**Accepted**：2023-06-20；**Published Online**：2023-07-07 11：34

URL：link.cnki.net/urlid/11.2625.V.20230706.1616.001

Foundation item：National Natural Science Foundation of China (51465034)

* **Corresponding author**. E-mail：1426779121@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0166

基于多层域适应的无标签数据故障诊断方法

王进花[1,*]，刘瑞[1]，曹洁[1,2]

(1. 兰州理工大学 电气工程与信息工程学院，兰州 730050；2. 甘肃省制造信息工程研究中心，兰州 730050)

摘 要：在工业生产中，由于源域数据和目标域数据分布有差异且有标签的故障数据量较少，以至于现有的域适应轴承故障诊断方法大多精度不高。基于此，提出多层域适应神经网络(MDANN)故障诊断方法，用于无标签数据的滚动轴承故障诊断。使用小波包分解与重构(WPT)对原始振动信号进行处理，以降低信号冗余并避免关键信号特征遗失；利用多核最大均值差异(MK-MMD)算法对输入特征值进行差异计算，并通过反向传播更新多层域适应神经网络的参数，使其能够提取域不变特征；为保证无标签目标域数据可以正常参与网络训练，使用最大概率标签作为真实标签的伪标签策略，解决目标域无标签数据无法训练问题，增强模型可靠诊断知识的获取。采用2个公开数据集CWRU和PU进行验证。实验结果表明：所提方法与常见的域适应方法对比具有更高的诊断精度，说明该方法能够有效地学习可迁移特征，拟合2个数据集之间的数据分布差异。

关 键 词：滚动轴承；故障诊断；迁移学习；多层域适应；伪标签策略

中图分类号：TP277; TD453

文献标志码：A **文章编号**：1001-5965(2025)04-1185-10

近年来，随着机器学习的蓬勃发展，基于深度学习的故障诊断逐渐用于工业领域并取得了良好的效果[1-4]。但是，深度学习具有一定的局限性，如要求训练集和测试集必须具有相同的特征分布、标注数据必须充足等。在实际的工业生产中，由于滚动轴承的转速、负荷、扭矩等工作状态多变，这使得原始的振动信号具有不同的特征分布并且部分工况数据采集困难，无法满足以上2个要求。

迁移学习方法被认为是解决上述问题的有效方法之一，其放宽了以上2个前提条件，近年来被用于解决数据缺失状况下的故障诊断问题。Ma等[5]构建了基于AlexNet卷积网络的故障诊断迁移学习模型，并引用了一种新的数据处理方式频率切片小波变换(frequency slicing wavelet transform, FSWT)，用于一维振动信息转换成二维图像，解决模型搭建时间久、诊断精度低的问题。Wu等[6]通过在训练过程中加入额外的噪声干扰来增强故障诊断模型对于工作条件改变的适应性问题。刘飞等[7]将残差网络和迁移学习相结合进行迁移学习故障诊断模型搭建，进行滚动轴承快速故障诊断，极大地缩短了诊断时间。Zou等[8]基于小波包分解与重构(wavelet packet transformation, WPT)的思想进行数据转换，并利用t-sne研究不同工况下迁移学习的可行性，总结了不同工况下轴承故障诊断的技术细节。上述文献利用参数迁移的方法来完成轴承的故障诊断，在保证高精度的同时也降低了模型的训练时间。然而对于一些机器来讲，很难采集到带标记的故障数据。主要原因在于机器从正常运行到故障需要漫长地退化过程，人为很难捕捉到故障时机，即使某个机器发生故障后，也会立马进行替换或修理，避免机器损坏带来经济损失。

为解决无标记数据的训练问题，Zhao等[9]通过

收稿日期：2023-04-06；**录用日期**：2023-06-09；**网络出版时间**：2023-06-29 11:39

网络出版地址：link.cnki.net/urlid/11.2625.V.20230628.1442.003

基金项目：国家重点研发计划(2020YFB1713600)；国家自然科学基金(62063020)；甘肃省自然科学基金(20JR5RA463)

*通信作者. E-mail：wjh0615@lut.edu.cn

引用格式：王进花，刘瑞，曹洁. 基于多层域适应的无标签数据故障诊断方法[J]. 北京航空航天大学学报，2025，51(4)：1185-1194.
WANG J H，LIU R，CAO J. Unlabeled data fault diagnosis method based on multi-domain adaptation[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1185-1194 (in Chinese).

添加软阈值构建深度收缩残差网络提取噪声冗余下的轴承振动数据的特征信息，对输入特征进行边缘分布约束完成故障诊断。Shao 等[10]通过多核最大均值差异 (multi-kernel maximum mean discrepancy, MK-MMD) 距离度量方式减小源域和目标域之间的特征分布差异，提高模型在目标域中的分类精度。Xu 等[4]提出联合均值差异匹配算法建立公共子空间，从而完成对工业过程的故障诊断。上述文献基于一个基本假设即目标域数据充足，但实际工业生产中，数据实用性不强，导致目标域数据不足以训练出高精度模型。因此，本文提出一种基于多层域适应结合伪标签策略的轴承故障诊断方法。该方法通过模型迭代收集源域和目标域数据的可迁移特征，达到对无标签目标域数据的健康状况识别目的。

本文的创新点为：①将小波包分解与重构算法与轴承故障诊断相结合，在数据处理层面对原始振动信号进行特征优化，既避免信号冗余又保证减少特征遗失；②构建多层域适应神经网络 (multi-dimain adaptation neural network, MDANN) 模型，利用源域的诊断知识来识别无标签的目标域样本，跨域边缘分布差异小、类间距离小，能够学习到可迁移特征；③提出多层域适应和伪标签相结合的特征迁移策略，前者逐层减小了源域和目标域学习到的数据可转移特征之间的分布差异，后者利用概率统计思想，取最大概率作为样本标签，解决目标域无标签数据的训练问题。

1 问题描述

本文旨在基于特征迁移思想建立一个轴承故障诊断模型，但源域和目标域的轴承数据来自于不同的数据集，因此，这些数据的特征分布存在严重差异[4,11]。如果直接使用卷积神经网络 (convoluyional neural networks, CNN) 模型学习到的源域特征用于目标域数据的诊断，诊断精度无法保证[12]，如图 1(a) 所示。为解决该问题，在模型搭建的过程中添加一个新的模块，即域自适应模块。通过该模块使得模型可以学习源域数据和目标域数据的可迁移特征，从而实现源域数据提供的诊断知识能够正确识别目标域的样本，如图 1(b) 所示。

使用D^{s}表示源域，使用D^{t}表示目标域。若源域的样本空间$\chi^{\mathrm{s}}\in D^{\mathrm{s}}$，目标域的样本空间$\chi^{\mathrm{t}}\in D^{\mathrm{t}}$，那么分别从源域和目标域获取的样本数据则可以表示为$x^{\mathrm{s}}\in\chi^{\mathrm{s}}$和$x^{\mathrm{t}}\in\chi^{\mathrm{t}}$。设定标签空间$y=\{1,2,\cdots,k\}$，共$k$种健康状况，且源域的健康状况包含目标域的健康状况，即$y^{\mathrm{t}}\subseteq y^{\mathrm{s}}\subseteq y$。同时假设源域和目标域的边缘概率分布分别为$P(x^{\mathrm{s}})$和$Q(x^{\mathrm{t}})$。本文主要目的是利用从源域数据中获取诊断知识，对无标签的目标域进行准确诊断，即使用源域$D^{\mathrm{s}}=\{\chi^{\mathrm{s}},P(x^{\mathrm{s}})\}$的$m$个带标签样本数据$\chi^{\mathrm{s}}=\{x_i^{\mathrm{s}},y_i^{\mathrm{s}}\}_1^m$提供充足的各类健康状况信息，对目标域$D^{\mathrm{t}}=\{\chi^{\mathrm{t}},P(x^{\mathrm{t}})\}$的$n$个不带标签的样本数据$\chi^{\mathrm{t}}=\{x_j^{\mathrm{t}}\}_{j=1}^n$进行故障状态识别。本文所搭建的模型由数据处理、特征提取、域自适应和故障分类 4 部分组成。在数据处理过程中，利用包分解与重构方法对原始振动信息进行再加工。其中，源域数据包含所有的故障类型且标记清晰，提供诊断知识。目标域样本缺少标记，利用源域的诊断知识进行分类。在特征提取方面，将源域数据和目标域数据同时放入搭建的网络中，通过网络迭代从而提取域不变特征。域自适应模块是指利用 MK-MMD 计算 3 层全连接层后输出特征的差异值，并通过反向传播缩小差异，从而获取可迁移特征。最后，对传入分类器的样本特征进行分类。

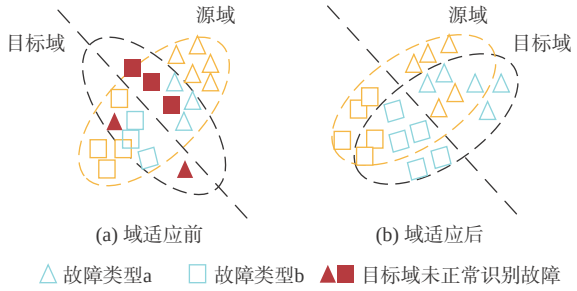


图 1 域适应

Fig. 1 Domain adaptation

2 方法介绍

2.1 小波包分解与重构

深度学习模型通常使用反向传播算法来学习特征，利用多层神经网络逐层提取数据的高层次特征。该方法可以在大数据集上取得很好的效果，但如果数据集较小或存在噪声影响时，深度学习模型容易过拟合，导致泛化能力不足。因此，引入 WPT 算法去除冗余信息和噪声干扰，生成更加稳健的特征表示。

小波包分解与重构提供一种更加精细的信号分析方法，从而将时频平面划分得更细致，同时引入了最优基选择的概念。

在频带经过多层次的划分后，根据被分析信号的特征，自适应地选取最佳基函数与信号相匹配。

从函数理论的角度来看，小波包分解是将信号投影到小波包基函数生成的空间中。从信号处理的角度来看，其是让信号通过系列中心频率不同但带宽相同的滤波器，细化信号的高频部分、抑制噪声并提高信号的分析能力。

设 $x(n)$ 为一段振动信号, $p_i^j(n)$ 表示第 j 层上的第 i 个小波包, 其包系数为 C_i^j。在进行分解和重构的过程中, 通过计算出小波包节点的 j 相对能量 R_{EWPN} 和熵 E_{WPN} 评估最优子结构, 对相关小波包进行筛选。WPT 的过程如下:

$$\begin{cases} p_j^{2i-1} = \sum_{k=1}^{k} h(k-2)p_{j-1}^i(n) \\ p_j^{2i} = \sum_{k=1}^{k} g(k-2l)p_{j-1}^i(n) \end{cases} \tag{1}$$

$$p_j^i(t) = \sum_{k=1}^{k} h(l-2k)p_{j+1}^{2i-1} + \sum_{k=1}^{k} g(l-2k)p_{j+1}^{2i} \tag{2}$$

式中: $h(\cdot)$ 为低通滤波器; $g(\cdot)$ 为高通滤波器; k、l 为分解层数。小波包筛选依据如下:

$$E_{\mathrm{WPN}}(i) = -\sum_{j=1}^{K} P_i^j \log_2(P_i^j) \tag{3}$$

$$R_{\mathrm{EWPN}}(i) = \frac{\sum_{j=1}^{K} \left(C_i^j\right)^2}{\sum_{m=1}^{N}\sum_{j=1}^{K} \left(C_m^j\right)^2} \tag{4}$$

式中: N 为小波包节点总数; K 为每个小波包节点的小波包系数总数。

2.2　多核最大均值差异

MK-MMD 是基于单核最大均值差异的一种改进方法。其将数据分布 x、y 通过核函数 $\varphi(\cdot)$ 映射到希尔伯特空间 H 上, 筛选多个核进行线性组合构成最优的核:

$$d_k^2(x,y) \triangleq \left\| E_x[\phi(x^{\mathrm{s}})] - E_y[\phi(x^{\mathrm{t}})] \right\|_{H_k}^2 \tag{5}$$

m 个核一起定义的特有核可以被表示为

$$K \triangleq \{k = \sum_{u=1}^{m} \beta_u k_u : \sum_{u=1}^{m} \beta_u = 1, \beta_u \geqslant 0, \forall u\} \tag{6}$$

MK-MMD 的损失值可以表示为: $d = \sum_{l=l_1}^{l_3} d_k^2(D_{\mathrm{s}}^l, D_{\mathrm{t}}^l)$。最终通过最小化损失函数来减少网络学习到的可迁移特征的分布差异, 可以表示为

$$\min_{\theta^l} \sum_{l=l_1}^{l_3} d_k^2(D_{\mathrm{s}}^l, D_{\mathrm{t}}^l) \tag{7}$$

式中: $l \subseteq \{l_1, l_2, l_3\}$; θ^l 表示全连接层 F1、F2、F3 的参数集。

2.3　伪标签策略

基于多层域适应策略, 可以减小学习到的可迁移特征的分布差异。而由于目标域数据没有标签导致最后一层全连接层参数无法进行训练。为解决模型无法训练问题, 引入伪标签学习策略。

伪标签策略的核心思想就是将预测概率最大的标签作为真实标签, 并通过网络迭代训练使得伪标签无限接近真实标签。其生成包括样本标签概率预测和伪标签的转换 2 个步骤。其中, 最后一层的 softmax 函数会计算出样本标签的概率分布, 根据该概率分布, 将其转换为该样本最终的标签。具体计算式为

$$y^{\mathrm{s}} = \{1, 2, \cdots, k\} \tag{8}$$

$$y_i^{\mathrm{t}} = \begin{Bmatrix} \max\{y^{\mathrm{s}}, y_i^{\mathrm{t}}\} \\ 0 \end{Bmatrix} \tag{9}$$

式中: y_i^{t} 为目标域第 i 个样本的伪标签。

由于源域和目标域样本都用于训练共享 CNN 参数, 那么伪标签的学习目标优化可表示为

$$\min_{\theta} = \frac{1}{m}\sum_{i=1}^{m} J(y_i^{\mathrm{s}}, F) + \lambda \frac{1}{n}\sum_{j=1}^{n} J(y_j^{\mathrm{t}}, F) \tag{10}$$

式中: m 和 n 分别为源域和目标域样本的数量; $J(\cdot)$ 为交叉熵损失函数; y_i^{s} 为源域中第 i 个样本的真实标签; F 为域共享参数; λ 为权衡参数; θ 为域共享 CNN 的参数集。

3　本文方法

3.1　基本网络架构

模型的基本网络架构由输入层、卷积层、池化层及全连接层 4 部分组成, 网络结构如图 2 所示, 网络中具体参数信息如表 1 所示。

输入层是由 2 个长度为 N 的振动信号组成, 每 2 048 个样本点构成一个数据样本。其中, 源域数据集的样本有标签, 目标域数据集样本无标签。

卷积层进行一维卷积运算[13]。假设卷积核的大小为 $w_{\mathrm{c}} \in \mathbf{R}^{H\times D}$, 第 i 个被切割的样本点表示为 $x_{j-m+1:j}^i \in \mathbf{R}^{H\times D}$, b_{c} 为相应的偏置, 那么一次卷积操作可以描述为

$$c_j = R_{\mathrm{elu}}\left(\sum_{i=1}^{n} x_{j-m+1:j}^i w_{\mathrm{c}} + b_{\mathrm{c}}\right) \tag{11}$$

池化层位于卷积层之后, 用于降低卷积池化的维数, 减少网络训练参数从而有效克服过拟合问题[14]。本文采用最大池化策略, 若池化的最大范围为 m, 池化的向量记为 $\boldsymbol{x}_j$, 池化后的结果记为 $\boldsymbol{l}_j$。一次池化操作可描述为

$$\boldsymbol{l}_j = \max\{\boldsymbol{x}_{jm:j(m+1)}\} \tag{12}$$

原始信号经过 4 次卷积和池化操作后, 利用全连接层将输入的域不变特征映射为一维向量[15]。假设输入的域不变特征在全连接层 F1 前, 则记为

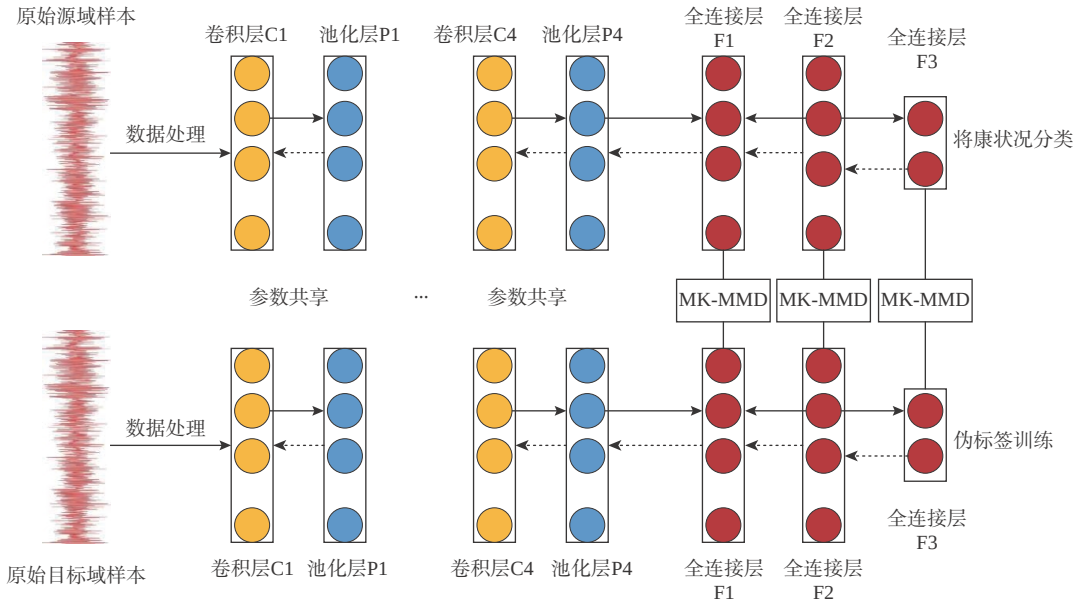


图2 本文方法网络结构

Fig. 2 Network structure of the proposed method

表1 网络参数详解

Table 1 Detailed explanation of network parameters

网络层	网络参数	激活函数	输出尺寸
输入层			2 048×1
卷积层C1	3×1×20	非线性映射Relu	2 046×20
卷积层C2	3×20×20	非线性映射Relu	1 020×20
卷积层C3	3×20×20	非线性映射Relu	508×20
卷积层C4	3×20×20	非线性映射Relu	252×20
池化层P1～P6	2×1		
全连接层F1			5 040×1
全连接层F2	5 040×256		256×1
全连接层F3	256×5	分类函数SoftMax	5×1

$\boldsymbol{x}_i^{\mathrm{F0,D}}$，D表示一维CNN。那么输入到F2层的域不变特征则可由式(13)推导：

$$\boldsymbol{x}_i^{\mathrm{F1,D}}=\sigma_{\mathrm{r}}(\boldsymbol{w}^{\mathrm{F1}}\boldsymbol{x}_i^{\mathrm{F1,D}}+\boldsymbol{b}^{\mathrm{F1}}) \tag{13}$$

经过3次全连接操作后，进行最终的特征分类[16-17]。由分类函数softmax来预测样本x_i^{s}和x_i^{t}的标签，F3层输出的样本标签概率分布可以用式(14)和式(15)推导：

$$\boldsymbol{l}_i^{\mathrm{F3,D}}=[P(y_i^{\mathrm{D}}=1|\boldsymbol{x}_i^{\mathrm{F2,D}})P(y_i^{\mathrm{D}}=2|\boldsymbol{x}_i^{\mathrm{F2,D}})\cdots P(y_i^{\mathrm{D}}=k|\boldsymbol{x}_i^{\mathrm{F2,D}})] \tag{14}$$

$$P(y_i^{\mathrm{D}}=i|\boldsymbol{x}_i^{\mathrm{F2,D}})=\frac{\exp((\boldsymbol{w}_j^{\mathrm{F3}})^{\mathrm{T}}\boldsymbol{x}_i^{\mathrm{F2,D}}+\boldsymbol{b}^{\mathrm{F3}})}{\sum_{j=1}^{n}\exp((\boldsymbol{w}_j^{\mathrm{F3}})^{\mathrm{T}}\boldsymbol{x}_i^{\mathrm{F2,D}}+\boldsymbol{b}^{\mathrm{F3}})} \tag{15}$$

式中：$\boldsymbol{w}_j^{\mathrm{F3}}$和$\boldsymbol{b}^{\mathrm{F3}}$为F3层的权重和偏置；$\boldsymbol{l}_i^{\mathrm{F3,D}}=[l_i^{\mathrm{F3,s}},l_i^{\mathrm{F3,t}}]$为数据样本的标签概率分布。

3.2 网络的训练过程

本文提出的MDANN模型通过联合最小化3个正则化项来进行训练：①源域样本的预测标签与真实标签之间的误差；②目标域数据训练使用伪标签所带来的损失；③跨域数据学习到的域不变特征的多层MK-MMD值。结合3个优化目标，该网络的损失函数定义为

$$L=L_{\mathrm{c}}+\lambda L_{\mathrm{p}}+\mu d_k^2 \tag{16}$$

式中：λ、μ为权衡参数，根据训练结果动态调整。

在训练过程中，使用随机梯度下降(stochastic gradient descent, SGD)[18-19]算法根据损失函数值进行参数优化。设θ_{c}为健康状况分类器的参数，那么式(16)中的损失函数可以改写为

$$L(\theta_{\mathrm{f}}^*,\theta_{\mathrm{c}}^*,\theta_{\mathrm{p}}^*)=\min_{\theta_{\mathrm{f}},\theta_{\mathrm{c}},\theta_{\mathrm{p}}}L_{\mathrm{c}}(\theta_{\mathrm{f}},\theta_{\mathrm{c}})+\lambda L_{\mathrm{p}}(\theta_{\mathrm{p}})+\mu d_k^2(\theta_{\mathrm{f}}) \tag{17}$$

由随机梯度算法和式(15)，参数θ_{f}、θ_{c}、θ_{p}将会按照如下规则更新：

$$\theta_{\mathrm{f}}\leftarrow\theta_{\mathrm{f}}-\varepsilon\left(\frac{\partial L_{\mathrm{c}}}{\partial\theta_{\mathrm{f}}}-\lambda\frac{\partial L_{\mathrm{d}}}{\partial\theta_{\mathrm{f}}}+\mu\frac{\partial d_k^2}{\partial\theta_{\mathrm{f}}}\right) \tag{18}$$

$$\theta_{\mathrm{c}}\leftarrow\theta_{\mathrm{c}}-\varepsilon\frac{\partial L_{\mathrm{c}}}{\partial\theta_{\mathrm{c}}} \tag{19}$$

$$\theta_{\mathrm{p}}\leftarrow\theta_{\mathrm{p}}-\varepsilon\frac{\partial L_{\mathrm{p}}}{\partial\theta_{\mathrm{p}}} \tag{20}$$

式中：ε为学习率，在本文中为0.000 1。

4 实验验证

为证明本文方法的有效性，使用公开数据集设计实验进行相关验证。当网络的损失值不再下降或训练次数达到3 000时，保存网络参数，使用无标签的目标域数据进行测试。由于该网络模型从目

标域中学习到的特征分布与源域样本相似, 因此, 分类器可以正确地对目标域中的未标记样本进行分类。具体训练过程如图 3 所示。

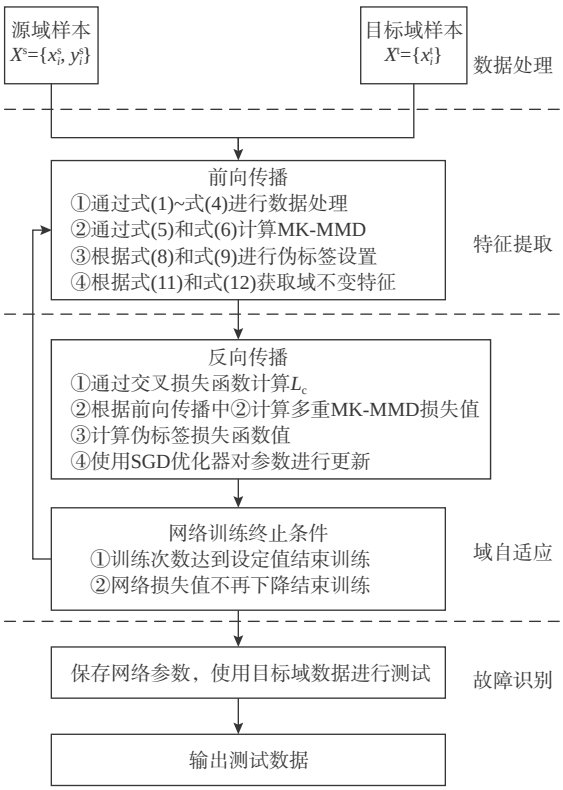


图 3　网络训练流程

Fig. 3　Network training flow

4.1　滚动轴承数据集

CWRU 轴承数据集来自于 SKF 公司型号 6205-2RS 的滚动轴承实验平台的数据采集, 如图 4 所示。本文选择了内圈直径为 0.18 mm 和 0.36 mm, 外圈直径为 0.18 mm和 0.36 mm 及正常状态 5 种健康类型滚动轴承来做故障诊断检测[19-20]。

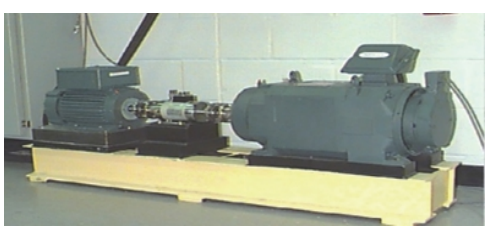
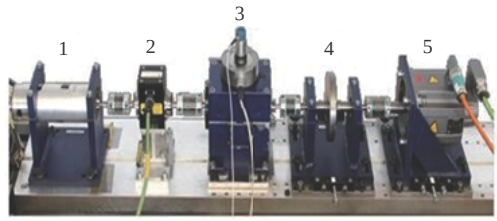


图 4　CWRU 滚动轴承实验平台

Fig. 4　CWRU rolling bearing test platform

PU 轴承数据集是由型号为 6203 的滚动轴承实验平台进行数据采集而成。如图 5 所示, 本文使用其中的电加工放点实验产生的人工故障数据。选取内圈故障 1、故障 2, 外圈故障 1、故障 2, 健康状态 5 种滚动轴承进行健康状况识别。

参与训练和测试的数据集的基本情况如表 2 所示。



1.电机; 2.扭矩测量单元; 3.轴承测量单元; 4.飞轮; 5.负载电机。

图 5　PU 滚动轴承实验平台

Fig. 5　PU rolling bearing test platform

表 2　数据集的基本情况

Table 2　Basic information of datasets

数据集	转速/(r·min^{-1})	负载/N	训练集/测试集（故障样本数）
数据集A(CWRU)	1 700	1 470	1 500/500
数据集B (PU)	1 500	1 000	1 500/500

4.2　权衡参数的确定

为有效提高 MDANN 的特征提取效果, 需要对模型的损失函数做充分的研究。而参数λ、μ作为损失函数的一部分, 对模型的特征提取能力有着重要的影响。因此, 对参数选择的研究非常必要。

分类精度与学习到的 F2 层全连接层的可迁移特征密切相关。通过计算 F2 层输出特征的 MMD 损失值来评估 MDANN 模型选择不同参数的特征迁移效果, 进而衡量参数选择的优劣。动态参数λ和μ分别从数据集合{0,0.05,0.1,0.3,0.5,1}及{0.01,0.05,0.1,0.15,0.2,0.25,0.3,0.35}中取得,最终实验取 10 次训练的平均值。实验结果如图 6 所示, 其中, x 轴表示μ值, y 轴表示对应的 MMD 损失值, 不同颜色的折线表示参数λ的取值。

由图 6 可知, 当权衡参数μ小于 0.15 时, MMD 损失值明显较大, 说明学习到的可迁移特征适应能力弱。一旦将权衡参数λ设置为较大的值, 如大于 0.1, 则跨域差异也难以校正。因此, 选定最终的结果: λ为 0.05, μ为 0.15。

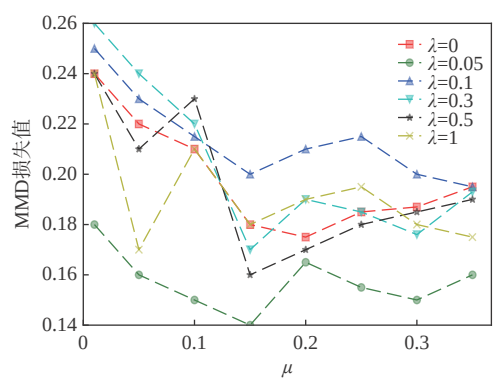


图 6　不同实验参数的诊断结果

Fig. 6　Diagnosis results of different experimental parameters

4.3 不同数据处理对诊断精度的影响

为验证不同数据处理方法对模型诊断精度的影响，采用数据集A、B进行实验。实验中将WPT、短时傅里叶变换(short time Fourier transform, FFT)、连续小波变换(continuous wavelet transform, CWT)及不做任何处理4种数据处理方法进行对比。迁移任务A->B数据集和B->A。在实验过程中，所有方法均调整至最优的参数，并取10次结果的平均值作为最终的结果，具体诊断效果如表3所示。

表3 不同数据处理方法的诊断精度

Table 3 Diagnosis accuracy of different data processing methods

数据处理方法	输入特征	迁移任务	诊断精度/%
不做处理	2 048×1	A->B	62.23
FFT	2 048×1	A->B	71.48
CWT	2 014×1	A->B	73.65
WPT	2 048×1	A->B	82.14
不做处理	2 048×1	B->A	54.21
FFT	2 048×1	B->A	75.47
CWT	2 014×1	B->A	71.56
WPT	2 048×1	B->A	80.37

由表3可知：①在A->B和B->A迁移任务中，对原始振动数据进行数据处理比不进行数据处理所得的诊断精度要高，说明数据处理对于故障诊断的必要性；②使用CWT和FFT可以完成故障诊断，精度在73%左右，效果并不理想；③使用WPT进行数据的拆分并组合的方式可以达到最优的诊断效果，在A->B和B->A的迁移任务中，均达到最优诊断效果，诊断精度分别为82.14%和80.37%，说明本文方法的优越性。

为进一步了解不同参数的分类细节，在进行A->B训练任务时绘制了轴承分类的混淆矩阵，如图7所示。图中，坐标轴上的5个参数代表轴承的5种健康状态，对角线上的所有数据识别正确的样本数目占总样本数据的百分比。由图7可知：①正常状态的轴承基本都能被识别出来；②内圈故障1的诊断效果较差，但易于内圈故障2的诊断混淆，这可能与2个数据集之间的数据特征有关；③4种方法均取到了60%以上的诊断精度，使用WPT方法时达到了最优的分类效果，诊断精度为82.14%。

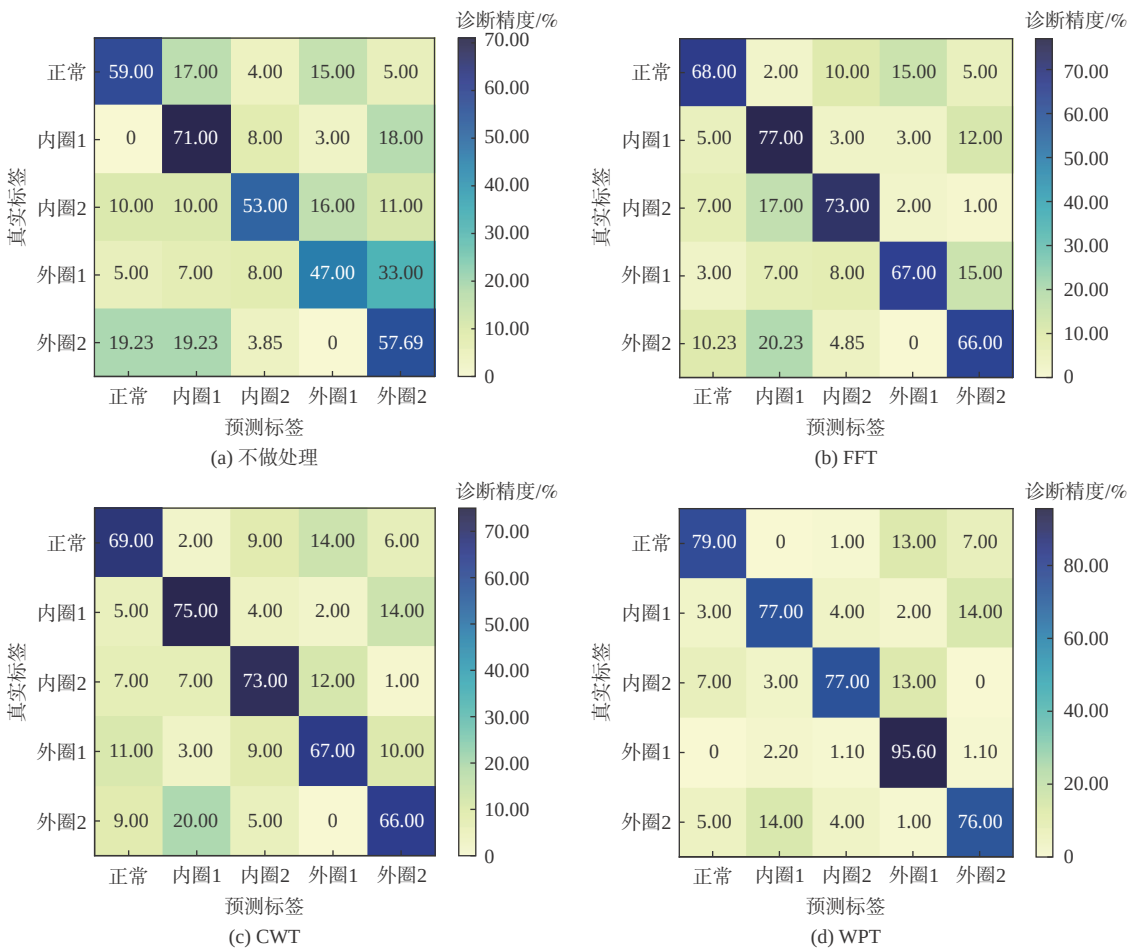


图7 轴承分类的混淆矩阵

Fig. 7 Confusion matrix for bearing classification

4.4　多层域适应对诊断精度的影响

为验证不同层数的域适应算法对诊断精度的影响, 将 MK-MMD 算法分别加在第 1 层全连接层后, 第 1 层、第 2 层全连接层后, 第 1 层、第 2 层、第 3 层全连接层后及不添加 4 种情况做对比。具体故障诊断效果如表 4 所示。

表 4　多层域适应对诊断精度的影响

Table 4　Effect of multi-domain adaptation on diagnosis accuracy

处理策略	训练时间/s	训练任务	诊断精度/%
不添加MK-MMD	93	A->B	37.21
F1处添加MK-MMD	108	A->B	71.32
F1,F2处添加MK-MMD	124	A->B	77.56
F1,F2,F3处添加MK-MMD	147	A->B	82.94
不添加MK-MMD	89	B->A	41.83
F1处添加MK-MMD	97	B->A	75.89
F1,F2处添加MK-MMD	123	B->A	79.87
F1,F2,F3处添加MK-MMD	148	B->A	83.21

由表 4 可知: ①在相同的训练任务下, 随着 MK-MMD 算法逐层加入, 训练时间也随之增多。每多一层加入 MK-MMD 算法来计算输出特征的差异值, 训练时间就会增多 15 s 左右; ②MK-MMD 算法的不断加入使得模型域适应能力明显增强, 分类任务完成的更加出色; ③当所有全连接层都加入 MK-MMD 算法后, 分类效果达到最优, 在 A->B 和 B->A 的迁移任务中, 诊断精度分别为 82.94% 和 83.21%。

通过 t-分布领域嵌入(t-distributed stochastic neighbor embedding, t-SNE)算法, 可以将上述算法全连接层处提取出来的故障特征降维至二维平面, 并以散点图的形式呈现,从而可以对不同算法分类结果进行更为直观的展示。在 A->B 迁移任务下, 4 种不同处理策略提取的特征可视化结果如图 8 所示。

由图 8 可知: ①未加入域适应算法分类效果极差, 各个故障类别之间的界限不够明显, 不同类别之间的特征相互堆叠, 诊断精度只有 37.21%; ②在全连接层处加入 MK-MMD 算法时, 模型的域适应能力得到增强, 可以有效提取 2 个数据集之间的公共特征, 诊断精度明显提高, 为 71.32%; ③随着逐层在全连接处添加域适应算法 MMD 时, 不同类别之间几乎没有特征堆叠, 诊断精度逐步提高。

4.5　不同方法诊断精度对比

为验本文方法的优越性, 将该方法与常见的域适应算法进行对比实验。本文选取迁移成分分析(transfer component analysis, TCA)[21]、深度域混淆(deep domain confusion, DDC)[22] 网络、深度自适应网络 (deep adaptation network, DAN)[23]、与域适应神经网络 (domain adaptive neural network, DANN)[24]4 种迁移学习网络及不添加域适应模块的 MDANN。其中, TCA 主张最小化源域 s 的概率分布 $P(\phi(X^s))$ 和目标域 t 的概率分布 $Q(\phi(X^t))$之间的距离; DDC 的核心理念在于使用 MMD 度量全连接层的第 2 层处的特征差异; DAN 的主要思想是使用 MK-MMD 度量多个全连接层的差异, 并以相加的方式处理计算结果; DANN 是 MDANN 的简化版本, 只进行域混淆, 没有做差异度量。DAN 的网络输入是一维数据, DDC、DANN 的输入是经过小波变换处理后的二维数据。本次的迁移任务是 A->B 和 B->A。在实验过程中, 所有模型均调整至最优的参数, 并取 10 次实验结果的平均值作为最后的结果, 具体诊断效果如表 5 所示。

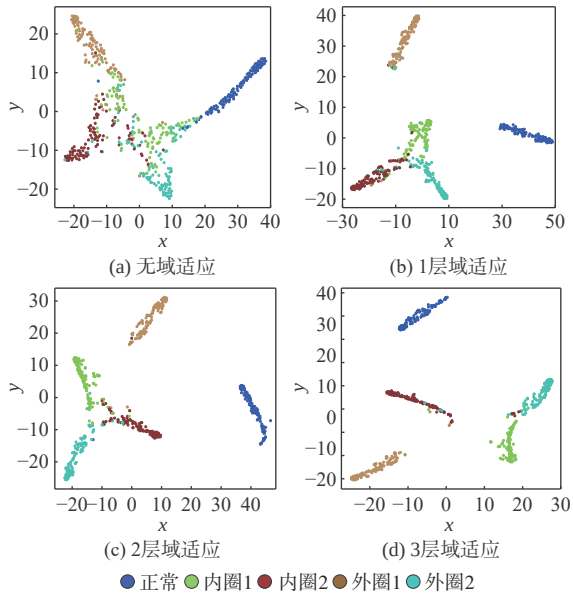


图 8　不同处理策略对于诊断精度影响的 t-SNE 可视化

Fig. 8　t-SNE visualization of effects of different processing strategies on diagnosis accuracy

表 5　不同模型的精度

Table 5　Accuracy of different models

网络模型	是否域适应	迁移任务	输入特征	诊断精度/%
TCA	否	A->B	2 048×1	47.12
DDC	是	A->B	224×224×3	74.32
DAN	是	A->B	2 048×1	76.86
DANN	是	A->B	224×224×3	77.58
MANN	否	A->B	2 048×1	32.15
MDANN	是	A->B	2 048×1	84.17
TCA	否	B->A	2 048×1	28.21
DDC	是	B->A	224×224×3	65.36
DAN	是	B->A	2 048×1	74.21
DANN	是	B->A	224×224×3	75.36
MANN	否	B->A	2 048×1	36.58
MDANN	是	B->A	2 048×1	82.15

由表5可知：①具备域适应的模型能够有效学习到可迁移特征，使得完成迁移任务的效果远远优于未具备域适应功能的网络模型；②仅通过度量输出特征的差异值或仅通过加入域鉴别器进行域混淆虽然可以完成迁移任务，但效果并不理想，诊断精度均在60%～75%之间；③MDANN在迁移任务A->B和B->A中均表现优异，诊断精度分别为84.17%和82.15%，在6个网络模型中排名第1。

对A->B迁移任务进行t-SNE可视化，如图9所示。由图9可知：①TCA、MDANN(未添加域适应模型)各个故障类别之间的界限不够明显，不同类别之间的特征相互堆叠，导致诊断精度不断下降；②DCC、DAN、DANN部分故障类型堆叠较多，导致模型识别困难，诊断性能不理想；③MDANN模型的散点分布图在6个迁移模型中的故障重叠最少，使得故障分辨简单，可以达到最好的分类结果，也进一步说明了该模型在故障诊断领域的优越性。

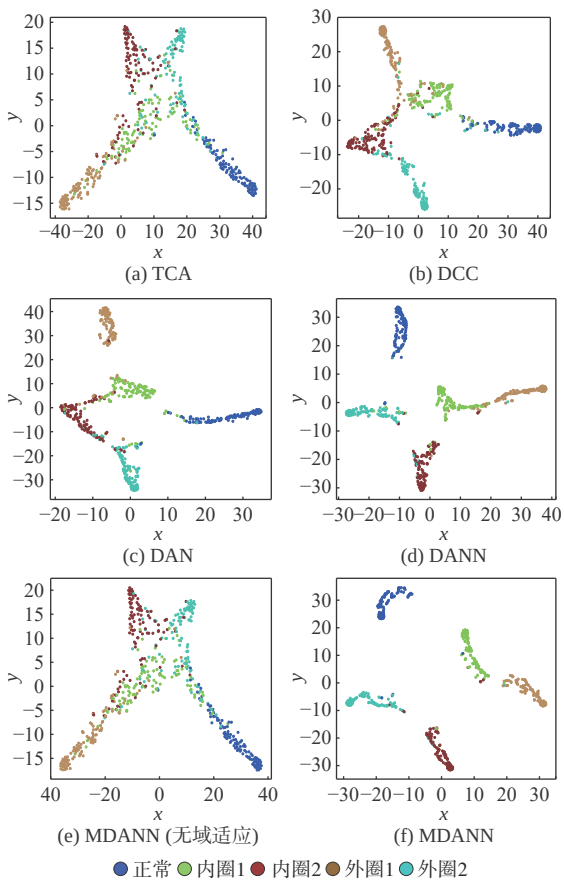


图9 不同模型的t-SNE图

Fig. 9 t-SNE diagrams of different models

5 结 论

本文提出一种基于多层域适应的滚动轴承故障诊断方法，用于解决无标签的故障数据不足的问题，结论如下：

1）鉴于FFT缺乏时间对频率的定位，CWT数据量大，数据具有冗余性的问题，利用WPT对原始轴承数据进行重构，优化原始数据的特征表示，使得网络可以提取深度特征；

2）针对源域数据和目标域数据分布有差异的问题，对全连接层进行多层域适应，使用MK-MMD来度量差异，并使用叠加或相乘方式验证2种策略的优异性；

3）针对目标域没有标签导致最后一个全连接层无法进行训练，使用伪标签策略来解决模型无法训练的问题，从而提高了模型的域适应能力。

未来将对模型内部特征提取层做出优化，从而更好的提取可迁移特征。同时，将本文方法嵌入到单片机内，设计出一套具有实时诊断能力的嵌入式平台，应用于工业故障诊断领域。

参考文献（References）

[1] YU X, LIANG Z T, WANG Y J, et al. A wavelet packet transform-based deep feature transfer learning method for bearing fault diagnosis under different working conditions[J]. Measurement, 2022, 201: 111597.

[2] HU Q, SI X S, QIN A S, et al. Balanced adaptation regularization based transfer learning for unsupervised cross-domain fault diagnosis[J]. IEEE Sensors Journal, 2022, 22(12): 12139-12151.

[3] 赵小强, 张青青, 陈鹏, 等. 基于PSO-BFA和改进Alexnet的滚动轴承故障诊断方法[J]. 振动与冲击, 2020, 39(7): 21-28.
ZHAO X Q, ZHANG Q Q, CHEN P, et al. Rolling bearing fault diagnosis method based on improved Alexnet and PSO-BFA[J]. Journal of Vibration and Shock, 2020, 39(7): 21-28(in Chinese).

[4] XU W, WAN Y, ZUO T Y, et al. Transfer learning based data feature transfer for fault diagnosis[J]. IEEE Access, 2020, 8: 76120-76129.

[5] MA P, ZHANG H L, FAN W H, et al. A novel bearing fault diagnosis method based on 2D image representation and transfer learning-convolutional neural network[J]. Measurement Science and Technology, 2019, 30(5): 055402.

[6] WU Z H, JIANG H K, ZHAO K, et al. An adaptive deep transfer learning method for bearing fault diagnosis[J]. Measurement, 2020, 151: 107227.

[7] 刘飞, 陈仁文, 邢凯玲, 等. 基于迁移学习与深度残差网络的滚动轴承快速故障诊断算法[J]. 振动与冲击, 2022, 41(3): 154-164.
LIU F, CHEN R W, XING K L, et al. Fast fault diagnosis algorithm for rolling bearing based on transfer learning and deep residual network[J]. Journal of Vibration and Shock, 2022, 41(3): 154-164(in Chinese).

[8] ZOU Y S, LIU Y Z, DENG J L, et al. A novel transfer learning method for bearing fault diagnosis under different working conditions[J]. Measurement, 2021, 171: 108767.

[9] ZHAO M H, ZHONG S S, FU X Y, et al. Deep residual shrinkage

networks for fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2020, 16(7): 4681-4690.

[10] SHAO J J, HUANG Z W, ZHU J M. Transfer learning method based on adversarial domain adaption for bearing fault diagnosis[J]. IEEE Access, 2020, 8: 119421-119430.

[11] LI B, TANG B P, DENG L, et al. Joint attention feature transfer network for gearbox fault diagnosis with imbalanced data[J]. Mechanical Systems and Signal Processing, 2022, 176: 109146.

[12] 张根保, 李浩, 冉琰, 等. 一种用于轴承故障诊断的迁移学习模型[J]. 吉林大学学报 (工学版), 2020, 50(5): 1617-1626.
ZHANG G B, LI H, RAN Y, et al. A transfer learning model for bearing fault diagnosis[J]. Journal of Jilin University (Engineering and Technology Edition), 2020, 50(5): 1617-1626(in Chinese).

[13] LIU M K, TRAN M Q, WENG P Y. Fusion of vibration and current signatures for the fault diagnosis of induction machines[J]. Shock and Vibration, 2019, 2019(1): 7176482.

[14] JIE H J, WANDA P. RunPool: a dynamic pooling layer for convolution neural network[J]. International Journal of Computational Intelligence Systems, 2020, 13(1): 66-76.

[15] YANG Z N, WANG X Y, YANG R. Transfer learning based rolling bearing fault diagnosis[C]//Proceedings of the IEEE 10th Data Driven Control and Learning Systems Conference. Piscataway: IEEE Press, 2021: 354-359.

[16] 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述[J]. 计算机学报, 2020, 43(5): 755-780.
XU B B, CEN K T, HUANG J J, et al. A survey on graph convolutional neural network[J]. Chinese Journal of Computers, 2020, 43(5): 755-780(in Chinese).

[17] SHABBEER BASHA S H, DUBEY S R, PULABAIGARI V, et al. Impact of fully connected layers on performance of convolutional neural networks for image classification[J]. Neurocomputing, 2020, 378: 112-119.

[18] 史加荣, 王丹, 尚凡华, 等. 随机梯度下降算法研究进展[J]. 自动化学报, 2021, 47(9): 2103-2119.
SHI J R, WANG D, SHANG F H, et al. Research advances on stochastic gradient descent algorithms[J]. Acta Automatica Sinica, 2021, 47(9): 2103-2119(in Chinese).

[19] CAO X Y, LAI L F. Distributed gradient descent algorithm robust to an arbitrary number of Byzantine attackers[J]. IEEE Transactions on Signal Processing, 2019, 67(22): 5850-5864.

[20] ACHARYA U R, YANTI R, ZHENG J W, et al. Automated diagnosis of epilepsy using CWT, HOS and texture parameters[J]. International Journal of Neural Systems, 2013, 23(3): 1350009.

[21] SHEN J X, SONG D, MA T C, et al. Blade crack detection based on domain adaptation and autoencoder of multidimensional vibro-acoustic feature fusion[J]. Structural Health Monitoring, 2023, 22(5): 3498-3513.

[22] ZHANG W P, ZHANG P, HE X H, et al. Convolutional neural network based two-layer transfer learning for bearing fault diagnosis[J]. IEEE Access, 2022, 10: 109779-109794.

[23] LONG M S, ZHU H, WANG J M, et al. Deep transfer learning with joint adaptation networks[C]//Proceedings of the 34th International Conference on Machine Learning. New York: ACM, 2017: 2208-2217.

[24] WANG Q, MICHAU G, FINK O. Domain adaptive transfer learning for fault diagnosis[C]//Proceedings of the Prognostics and System Health Management Conference. Piscataway: IEEE Press, 2019: 279-285.

Unlabeled data fault diagnosis method based on multi-domain adaptation

WANG Jinhua[1,*], LIU Rui[1], CAO Jie[1,2]

(1. College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China;
2. Gansu Manufacturing Information Engineering Research Center, Lanzhou 730050, China)

Abstract: in industrial production, due to the difference in the distribution of source domain data and target domain data and the small amount of labeled fault data, the accuracy of domain adaptation-based bearing fault diagnosis algorithms proposed in the past is generally not high. In view of this, the multi-domain adaptation neural network (MDANN) fault diagnosis method was proposed in this paper, which was used for rolling bearing fault diagnosis without labeled data. Firstly, the original vibration signal was processed by using wavelet packet transformation (WPT) to reduce signal redundancy and avoid the loss of key signal features. Secondly, the multi-kernel maximum mean discrepancy (MK-MMD) algorithm was used to calculate the difference of input eigenvalues, and the network parameters of MDANN were updated by backpropagation so that the network can extract domain invariant features. Finally, in order to ensure that unlabeled target domain data can participate in network training normally, the maximum probability label was used as a pseudo-label strategy of the real label to solve the problem that unlabeled target domain data cannot be trained and enhance the acquisition of reliable diagnosis knowledge of the model. Two publicly available datasets, CWRU and PU, were used for validation. The experimental results show that the proposed method has higher diagnosis accuracy compared with common domain adaptation methods, which further shows that the method can effectively learn the transferable features and fit the discrepancy in data distribution between the two datasets.

Keywords: rolling bearing; fault diagnosis; transfer learning; multi-domain adaptation; pseudo-label strategy

Received: 2023-04-06; **Accepted**: 2023-06-09; **Published Online**: 2023-06-29 11:39

URL: link.cnki.net/urlid/11.2625.V.20230628.1442.003

Foundation items: National Key Research and Development Program of China (2020YFB1713600); National Natural Science Foundation of China (62063020); Natural Science Foundation of Gansu Province (20JR5RA463)

* **Corresponding author**. E-mail: wjh0615@lut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0245

基于结构参数分析的 Gyroid 换热器性能

杨晓军，张雪丽*，李沛燃

(中国民航大学 航空工程学院，天津 300300)

摘 要：为研究三重周期极小曲面 (TPMS) 在空气-燃油换热器中的流动换热特性，分析结构参数对换热器性能的影响。采用田口方法，针对壁厚、晶格大小和偏移尺寸 3 个结构参数，对努塞尔数和摩擦系数进行综合评估，并与参考的螺旋套管 (TTHC) 换热器性能进行对比。研究得出各个结构参数影响 Gyroid 换热器努塞尔数和摩擦系数方面的主次顺序，并利用矩阵分析法，确定了最优方案。结果表明：对于热侧出口温度，TTHC 换热器比参考的 Gyroid 换热器最高降低 13.13 K；对于压降，TTHC 换热器比参考的 Gyroid 换热器降低 3.02 kPa；对于性能评估系数，热侧通道最高为 14.72，冷侧通道最高为 0.78。

关 键 词：Gyroid 换热器；螺旋套管换热器；田口方法；结构参数；矩阵分析法
中图分类号：TK124
文献标志码：A **文章编号**：1001-5965(2025)04-1195-10

随着燃气涡轮进口温度的不断提高，换热器需要更高效的设计方案。三重周期极小曲面 (triply periodic minimal surface, TPMS) 具有较高的表面积和体积比，能带来更高的换热效率[1]，因此，在换热器领域深入挖掘 TPMS 的应用，是提高其性能和效益的有效方案。

关于 TPMS 换热器的压降和传热性能，国内外学者进行了一些研究。Reynolds[2] 通过数值仿真的方法，研究了 Diamond、Gyroid 和 Primitive 换热器在低雷诺数下压降和传热性能的变化，结果表明：对比努塞尔数，Gyroid>Diamond>平板换热器>Primitive，其中，在压力损失固定时，Primitive 传热性能最好。Wadsoe 和 Holmqvist[3] 通过试验和仿真的方法，研究了 Diamond 和 Gyroid 换热器，评估了换热器的性能，证明了换热器的可制造性。结果表明：与参考换热器相比，TPMS 换热器提供了更高的传热性能，对比压降，Gyroid >参考换热器>Diamond。Kim 和 Yoo[4] 采用试验的方法，研究了介质为水时，Diamond、Gyroid 和 Primitive 换热器的传热性能数据。结果表明：Diamond 换热器提供了最高的传热速率，其次是 Gyroid 和 Primitive 换热器，TPMS 结构的压力损失几乎与传统板式换热器相同。Iyer 等[5] 采用数值仿真的方法，共计算了 7 种 TPMS 和 PNS 结构的换热和流动特性，结果表明：Diamond 的传热性能最好，Primitive 的传热性能最差；TPMS 换热器与管式换热器具有相似的热性能时，体积减少了 3～10 倍。Dixit 等[6] 对 Gyroid 结构制造的紧凑型换热器进行试验研究，结果表明：与热力学等效逆流换热器相比，Gyroid 换热器效率提高了 55%。Dharmalingam 等[7] 采用数值仿真的方法，分析了 Diamond、Fischer Koch S 换热器的热工水力性能，结果表明：Fischer Koch S 换热器在与 Diamond 换热器在具有相似的压力损失下，传热增加了 23%～356%。Peng 等[8] 采用数值仿真的方法，通过新开发的数值模型，分析了 Gyroid 换热器的流体流动和换热特性，证明了换热器结构的可制造性。结果表

收稿日期：2023-05-12；**录用日期**：2023-10-06；**网络出版时间**：2024-10-09 13:15
网络出版地址：link.cnki.net/urlid/11.2625.V.20241009.1033.001
基金项目：中央高校基本科研业务费专项资金 (3122019187)
* **通信作者**. E-mail：zxl982511@163.com

引用格式：杨晓军，张雪丽，李沛燃. 基于结构参数分析的 Gyroid 换热器性能 [J]. 北京航空航天大学学报，2025，51（4）：1195-1204.
YANG X J，ZHANG X L，LI P R. Performance of Gyroid heat exchanger based on structural parameters analysis [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1195-1204（in Chinese）.

明：螺旋结构对增强换热，防止颗粒粘附有非常好的作用，且与传统换热器相比，换热效率提高了7.5倍。Li等[9]采用数值仿真的方法，研究介质为CO_2时，印刷电路板换热器与Gyroid和Diamond换热器的流动和换热特性对比。结果表明：TPMS由于复杂的结构，加剧了湍流动能，强化了换热效果，Gyroid和Diamond换热器的整体的热性能提高了15%～100%。Li等[10]在文献[9]的基础上，将冷却介质更换为水，通过数值仿真的方法，对Gyroid和Diamond换热器与印刷电路板换热器进行了对比研究。结果表明：印刷电路板换热器传热增强的同时，压力损失会明显增加。Gyroid和Diamond换热器的平均传热系数高于印刷电路板换热器，且在雷诺数大于16 000时，Gyroid结构表现出更好的热性能。Femmer等[11]采用试验加数值仿真的方法，在介质为水时，对IWP、Diamond、Gyroid和Primitive换热器进行了结构性能的评估。结果表明：Diamond换热器在压降固定的情况下，具有优异的传热性能，而与平板和中空管相比，TPMS结构热性能高于其一个数量级。Alteneiji等[12]采用数值仿真的方法，分别研究Gyroid和Primitive换热器结构，冷端和热端均为水时的压降和传热性能。结果表明：Gyroid换热器效率比横流式换热器提高35%，最大流动传热系数达到1 400 W/(m^2·K)，Primitive换热器最大流动传热系数达到1 413 W/(m^2·K)，Liang等[13]通过试验和数值仿真的方法，对流动气体为空气时，换热器的性能进行了评估。结果表明：Schwarz-D换热器提供的传热效率比Primitive换热器高41%～52%，比Gyroid换热器高4%～12%。

国内外学者对空-油换热器进行了很多研究。Wen等[14]采用对数平均温差法设计了紧凑型的翅片管换热器，质量为1.207 kg，表面积密度达到697 m^2/m^3，并根据试验结果与关联式之间的差异，建立了管外的传热经验关系式。Liu等[15]采用试验的方法，对空气-燃油螺旋套管（tube-in-tube helical coil，TTHC）换热器进行性能测试，结果表明：在流速为0.005 6 kg/s时，可将热空气温度从823 K降低到519 K，且总结出环空侧的经验关系式。Herring和Heister[16]通过在空-油换热器中使用金属丝盘管插件来提高传热性能，结果表明：当在燃料侧使用电线线圈插件时，可将空气温度提高44%，相对于光管，管内换热系数提高300%。

综上所述，采用TPMS可以较大程度上提高换热器性能，但关于TPMS在空-油换热器的研究较为匮乏，且鲜见研究TPMS的结构参数对性能的影响。杜默等[17]对印刷电路板换热器的结构参数进行多目标性能优化，得到宽为0.29 mm、深为0.39 mm、深宽比为1.34的最优参数组合。故本文采用田口方法，研究Gyroid换热器的结构参数（壁厚、晶格大小和偏移尺寸）对换热器性能的影响，并与参考的TTHC换热器进行对比分析。

1 物理模型与边界条件

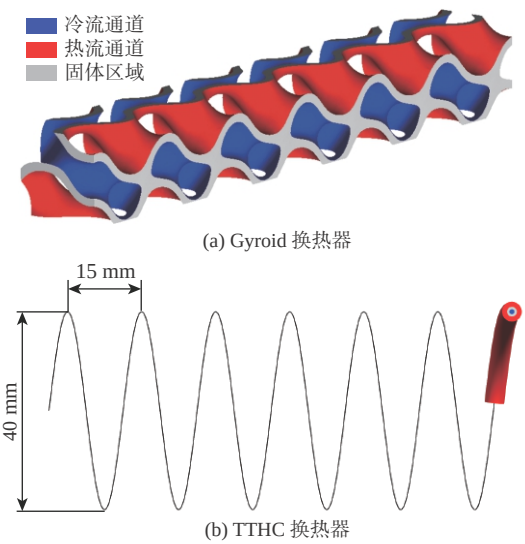
TPMS是生物适应自然环境，进化产生的结构设计方案，是一种平均曲率为0的周期性光滑隐式曲面，由隐式函数控制，参数设计简单，具有很高的结构完整性等特点。本文数值仿真采用在蝴蝶翅膀鳞片中被发现的Gyroid结构，由下列方程控制：

$$K=\cos\left(\frac{2\pi}{L}x\right)\sin\left(\frac{2\pi}{L}y\right)+\cos\left(\frac{2\pi}{L}y\right)\sin\left(\frac{2\pi}{L}z\right)+\cos\left(\frac{2\pi}{L}z\right)\sin\left(\frac{2\pi}{L}x\right) \tag{1}$$

式中：K为偏移参数；L为周期长度。

图1为本文计算采用的2种不同换热器结构，计算区域由空气流经的高温流体域RP-3燃油流经的低温流体域和固体域组成。图1(a)中，Gyroid换热器单元晶格大小为5 mm×5 mm×5 mm，壁厚为0.5 mm，结构偏移量为0.4 mm，横向长度为30 mm。图1(b)为仿真采用的参考TTHC换热器，长度为30 mm。其中，图1(a)Gyroid换热器横向长度与图1(b)TTHC换热器周向长度相同，固体区域具有相同的厚度。

计算域由热流体区域、冷流体区域和不锈钢固体域共同组成。其中，流体域均设置有进出口边界条件，进口设置为质量流量和温度边界，出口设置为压力出口条件，具体参数如表1所示，后续田口



(a) Gyroid换热器

(b) TTHC换热器

图1 2种不同的换热器结构

Fig. 1 Two different heat exchanger structures

方法计算中采用冷侧质量流量为 0.004 kg/s, 热侧质量流量为 0.005 1 kg/s。流体域与固体交界面设置为耦合的边界条件, 固体其他区域设置为绝热的边界条件, 设置如图 2 所示。

表 1　边界条件

Table 1　Boundary conditions

边界	质量流量$\dot{m}$/(kg·s^{-1})	进口温度T_{in}/K	出口压力P_{out}/MPa
冷侧	0.001, 0.002, 0.003, 0.004	288	4.0
热侧	0.005 1, 0.005 3, 0.005 6, 0.005 9, 0.006 1	623	1.0

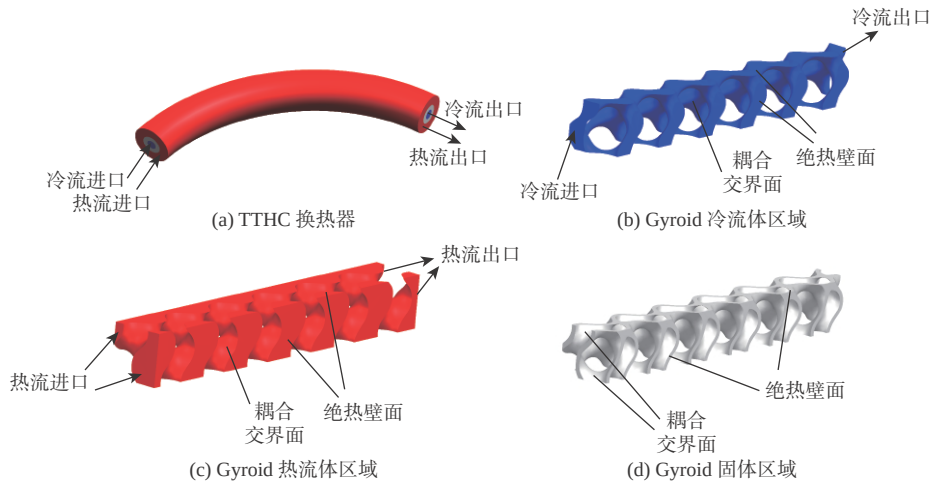


图 2　边界设置

Fig. 2　Boundary setting

计算采用商业软件 ANSYS CFX, 其中, 热空气采用理想空气设置, 冷 RP-3 燃油成分非常复杂, 为计算方便, 很多学者都提出了不同的碳氢化合物组分的燃料替代模型, 本文采用 RP-3 煤油的单组份热物性替代模型(正癸烷), 采用 REFPROP 获取其在 4 MPa 时密度、比热、黏度、导热系数等热物性随温度变化如图 3 所示, 可以看出航空煤油的临界温度在 710 K 附近, 换热后的航空煤油低于该温度, 故不需要考虑相变。

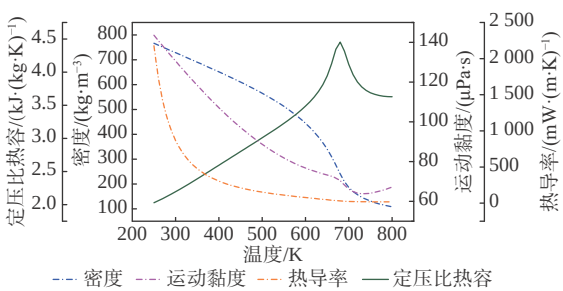


图 3　P=4 MPa 正癸烷热物理特性

Fig. 3　Thermophysical properties of n-decane with P=4 MPa

2　参数的定义

本文主要采用田口方法对 Gyroid 结构的 3 个因素进行分析: 壁厚、晶格大小和偏移尺寸。壁厚是任何换热器设计的主要因素, Tandel 和 Rammohan[18] 指出 LPBF 允许打印厚度大于等于 0.2 mm 的 TPMS, 故为保证其制造可行性, 选择 0.3 mm 作为壁厚设置的下限。晶格大小的变化, 会改变结构表面积和体积之比, 对压降有较大影响。偏移参数的改变, 可以控制 2 个流体通道体积的划分, 影响流体的热量交换和压降变化, 具体结构参数, 如表 2 所示。

表 2　结构参数

Table 2　Structural parameters

单元晶格大小/mm	结构偏移量/mm	壁厚/mm
(5, 5, 5) (7, 7, 7) (8, 8, 8)	0.2, 0.3, 0.4	0.3, 0.4, 0.5

利用正交表设计不同结构参数的组合方案, 通过努塞尔数与摩擦系数来评价换热器性能, 通过信噪比进行分析。信噪比作为衡量产品性能的指标, 值越大, 产品性能越优异。根据 2 个优化指标对换热器性能影响的特性不同, 选择不同的计算式。对于努塞尔数, 越大越好, 所以采用越大越好的计算式, 对于摩擦系数, 越小越好, 所以采用越小越好的计算式。本文所用计算式如下:

1) 水力直径[5]:

$$D_h = \frac{4V}{S} \tag{2}$$

式中: D_h 为水力直径, m; V 为冷流体域或热流体域

体积，m^3；S为流固交界面的表面积，m^2。

2）达西-魏斯巴赫摩擦系数[10]：

$$f=\frac{\Delta P}{\rho_{\mathrm{b}}\left(\frac{L_z}{D_{\mathrm{h}}}\right)\left(\frac{1}{2}U_{\mathrm{b}}^2\right)} \tag{3}$$

式中：f为摩擦系数；ΔP为压降，Pa；ρ_{b}为平均体积密度，kg/m^3；L_z为沿z方向的长度，m；U_{b}为热侧空气平均体积速度，m/s。

3）努塞尔数[10]：

$$Nu=\frac{UD_{\mathrm{h}}}{\lambda_{\mathrm{b}}} \tag{4}$$

其中，平均传热系数为

$$U=\frac{Q}{T_{\mathrm{b}}-T_{\mathrm{w}}} \tag{5}$$

式中：U为平均传热系数，W/K；λ_{b}为热侧空气平均导热系数，W/(m·K)；Q为传热量，W；T_{b}为热侧空气平均体积温度，K；T_{w}为平均壁面温度，K。

4）信噪比F[19]：

越大越好：

$$F=-10\lg\left(\frac{1}{i}\sum\frac{1}{y^2}\right) \tag{6}$$

越小越好：

$$F=-10\lg\left(\frac{1}{i}\sum y^2\right) \tag{7}$$

式中：i为仿真迭代次数；y为每个方案的模拟结果。

5）性能评估系数：

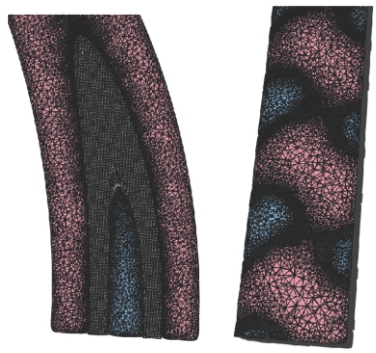
$$P_{\mathrm{EC}}=\frac{Nu/Nu_0}{\sqrt[3]{f/f_0}} \tag{8}$$

式中：P_{EC}为换热器的性能评估系数；Nu和Nu_0分别为Gyroid换热器和TTHC换热器的努塞尔数；f和f_0分别为Gyroid换热器和TTHC换热器的摩擦系数。

3 结果有效性验证

图4为分别对TTHC换热器和Gyroid换热器绘制的网格剖面图，对冷热流体域采用非结构化网格的绘制方法，添加0.4 mm的膨胀层，以保证冷热流体交界面计算准确，湍流模型采用K-Epsilon。为节省计算量，消除网格对计算结果的影响，分别对网格数210万、310万、410万、510万进行比较，数据如表3所示，针对本文的评估目标，可以看出随着网格数的改变，努塞尔数和摩擦系数的结果变化稳定在小数点后两位浮动，故研究选取310万网格进行分析。

图5为正癸烷沿流动方向定压比容仿真数据与REFPROP内数据的对比。本文采用RP-3的单组份替代材料正癸烷进行数值计算，将REFPROP



(a) TTHC 换热器 (b) Gyroid 换热器

图4 计算域网格剖面图

Fig. 4 Mesh profile of calculation domain

表3 网格无关性

Table 3 Grid independence

网格数	努塞尔数	摩擦系数
2.1×10^6	141.838	0.024 6
3.1×10^6	141.842	0.025 1
4.1×10^6	141.844	0.025 3
5.1×10^6	141.841	0.025 0

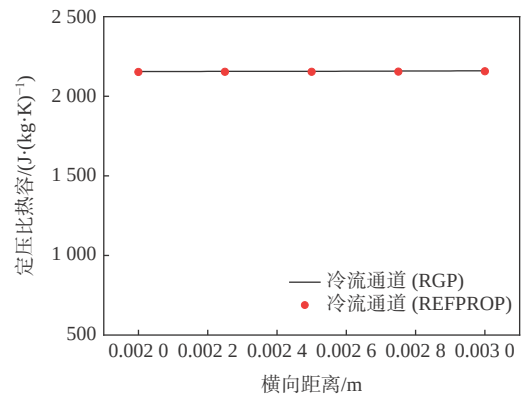


图5 RGP文件准确性验证

Fig. 5 Verification of RGP file accuracy

中正癸烷的热物理特性通过编写RGP文件导入CFX计算。其中，RGP文件中温度变化编写范围为288～623 K，压力变化为3.9～4.1 MPa。为验证RGP文件编写的正确性，选取冷流体区域内，进口y范围变化从0.002～0.003 m的定压比容与REFPROP软件内的热物理特性数据进行对比，可以看出仿真数据和REFPROP软件内数据存在很好的一致性。

由于试验条件的限制，为准确的分析换热器性能，与文献[15]试验获取的结果对比。图6为热侧流体不同质量流量的情况下，努塞尔数与文献[15]试验结果的对比。图7为冷侧流体在不同质量流量下的摩擦系数与文献[15]试验结果的对比。随着质量流量的变化，热侧努塞尔数和冷侧摩擦系数与试验结果对比，均呈相似的变化趋势。努塞尔数仿真结果始终高于试验数据，摩擦系数的仿真结果

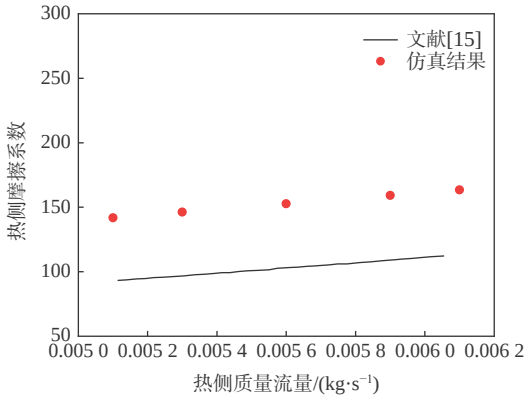


图 6　热侧流体努塞尔数变化

Fig. 6　Change of Nusselt number in hot-side fluid

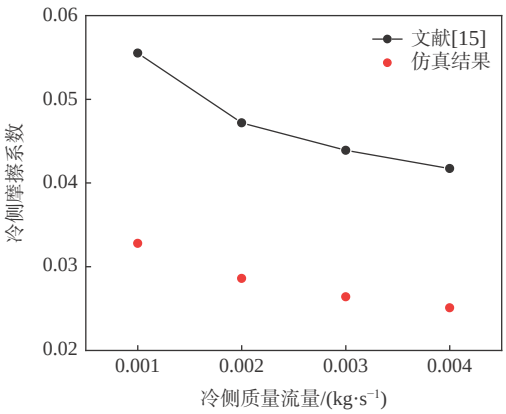


图 7　冷侧流体摩擦系数变化

Fig. 7　Change of friction coefficient of cold-side fluid

始终低于试验数据。这是由于试验采用的换热器结构在热流体外侧还有直径为 5 mm 的固体域包裹, 仿真则设置为绝热壁面, 与试验相比, 减少了对外界的热量损失。且试验采用的 TTHC 换热器为 6.25 圈螺旋, 仿真采用环向长度 30 mm 的 TTHC 换热器, 与试验相比, 减弱了螺旋流动离心效应的损失。综上, 随着质量流量的变化, 误差均保持在稳定的数值, 可以通过修正方式去除, 数值仿真结果可以有效表现换热器的换热特性。

4　结果分析

本文采用田口方法, 研究了 Gyroid 结构在不同结构参数下的换热性能, 设计 3 因素 3 水平的正交表。3 因素为壁厚、晶格大小和偏移尺寸, 三水平对应于壁厚为 0.3 mm、0.4 mm、0.5 mm, 晶格大小为(5 mm、5 mm、5 mm)、(7 mm、7 mm、7 mm)、(8 mm、8 mm、8 mm), 偏移尺寸为 0.2 mm、0.3 mm、0.4 mm。以努塞尔数和摩擦系数为评价指标, 获取其对应的信噪比 F, 如表 4 所示。

4.1　信噪比分析

表 5 为 3 种因素在 3 个水平下的信噪比及相关参数, 表 6 为各个因素在各个水平下信噪比平均值的计算结果, 变化如图 8 和图 9 所示。

对于努塞尔数, 3 种因素的影响大小为晶格大小>偏移尺寸>壁厚, 结构组合性能最优为: 晶格大小为 7 mm(F=82.45), 壁厚为 0.5 mm(F=81.84), 偏移尺寸为 0.4 mm(F=82.38)。

对于摩擦系数, 3 种因素的影响大小为偏移尺寸>晶格大小>壁厚, 结构组合最优为晶格大小为 8 mm(F=−2.22), 壁厚为 0.3 mm(F=−2.82), 偏移尺寸为 0.2 mm(F=−2.24)。

努塞尔数和摩擦阻力分别是评估换热器传热能力和压降损失的 2 个参数。图 8 为努塞尔数的信噪比平均值变化图。努塞尔数的信噪比平均值随着壁厚的增大而逐渐增大, 随着偏移尺寸的增大而增大, 但随着晶格大小的增加, 先增大后减小。当壁厚水平 1 时, 努塞尔数为 12 110.81(F=81.63), 壁厚水平 3 时, 努塞尔数为 12 394.95(F=81.84)。当偏移尺寸为水平 1 时, 努塞尔数为 11 611.02(F=81.28), 当偏移尺寸为水平 3 时, 努塞尔数为 13 273.97(F=82.38)。所以, 对于努塞尔数, 最优的组合结构为晶格大小水平 2、壁厚水平 3、偏移尺寸水平 3。

图 9 为摩擦系数的信噪比平均值变化。摩擦

表 4　Gyroid 换热器的努塞尔数和摩擦系数及对应的信噪比

Table 4　Nusselt number and friction coefficient of Gyroid heat exchanger and corresponding signal to noise

方案	晶格大小/mm	壁厚/mm	偏移尺寸/mm	热侧Nu	冷侧f	信噪比	
						Nu	f
1	5	0.3	0.2	10 894.77	1.10	80.74	−0.83
2	5	0.4	0.3	10 495.93	1.32	80.42	−2.41
3	5	0.5	0.4	11 117.05	2.25	80.92	−7.04
4	7	0.3	0.3	12 095.37	1.81	81.65	−5.15
5	7	0.4	0.4	15 362.57	1.99	83.73	−5.98
6	7	0.5	0.2	12 527.27	1.64	81.96	−4.30
7	8	0.3	0.4	13 342.29	1.33	82.50	−2.48
8	8	0.4	0.2	11 411.02	1.20	81.15	−1.58
9	8	0.5	0.3	13 540.53	1.35	82.63	−2.61

表5 各个水平下努塞尔数计算结果和相应的信噪比

Table 5 Nusselt number calculation results and corresponding signal to noise at each level

水平	晶格大小/mm	壁厚/mm	偏移尺寸/mm	信噪比		
				晶格大小	壁厚	偏移尺寸
1	10 835.92	12 110.81	11 611.02	80.69	81.63	81.28
2	13 328.40	12 423.17	12 043.94	82.45	81.77	81.57
3	12 764.61	12 394.95	13 273.97	82.09	81.84	82.38
极差	2 492.48	312.36	1 662.95	1.76	0.21	1.1

表6 各个水平下摩擦系数计算结果和相应的信噪比

Table 6 Friction coefficient calculation results and corresponding signal to noise at each level

水平	晶格大小/mm	壁厚/mm	偏移尺寸/mm	信噪比		
				晶格大小	壁厚	偏移尺寸
1	1.56	1.41	1.31	−3.43	−2.82	−2.24
2	1.81	1.50	1.49	−5.14	−3.32	−3.39
3	1.29	1.75	1.86	−2.22	−4.65	−5.17
极差	0.52	0.34	0.55	2.92	1.83	2.93

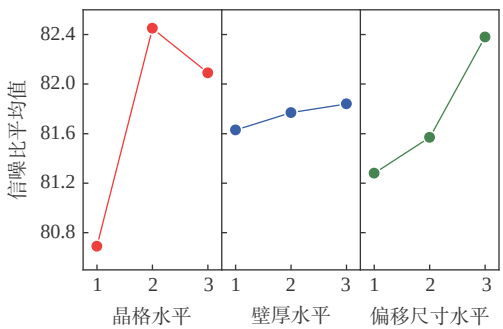


图8 努塞尔数的信噪比平均值

Fig. 8 Average signal to noise ratio of Nusselt numbers

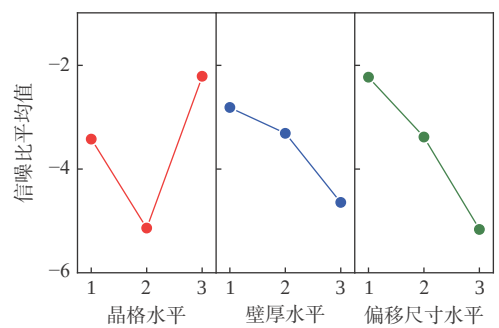


图9 结构参数对摩擦系数的信噪比平均值

Fig. 9 Average signal to noise ratio of structural parameters to friction coefficients

系数的信噪比平均值随着晶格大小的增加先降低后增加，随着壁厚和偏移尺寸的变化，逐渐减小。在壁厚为水平1时，摩擦阻力为1.41(F=−2.82)，壁厚为水平3时，摩擦阻力为1.75(F=−4.65)。在偏移尺寸为水平1时，摩擦阻力为1.31(F=−2.24)，在偏移尺寸为水平3时，摩擦阻力为1.86(F=−5.17)。所以对于摩擦系数，最优的组合为晶格大小水平3、壁厚水平1、偏移尺寸水平1。

4.2 矩阵法分析

对于本文设计的多指标正交方案，采用矩阵分析法[20]建立3层结构模型和层结构矩阵，并得出各因素各水平的权重，确定最优的组合方案。具体计算方程如下：

$$\boldsymbol{\omega} = \boldsymbol{MTS} \tag{9}$$

式中：$\boldsymbol{M}$为指标层矩阵；$\boldsymbol{T}$为因素层矩阵；$\boldsymbol{S}$为水平层矩阵；$\boldsymbol{\omega}$为权重矩阵。

总权重矩阵为2个指标的权矩阵的平均值，总权重矩阵的计算结果为

$$\boldsymbol{\omega} = \frac{1}{2}(\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2) = \frac{1}{2}\left(\begin{bmatrix}0.163\,3\\0.200\,8\\0.192\,3\\0.022\,9\\0.023\,5\\0.023\,4\\0.102\,5\\0.106\,3\\0.117\,2\end{bmatrix} + \begin{bmatrix}0.1208\\0.103\,8\\0.147\,2\\0.086\,9\\0.082\,0\\0.069\,8\\0.151\,2\\0.133\,3\\0.107\,4\end{bmatrix}\right) =$$

$$\begin{bmatrix}0.142\,1\\0.152\,3\\0.169\,8\\0.054\,9\\0.052\,8\\0.046\,6\\0.126\,9\\0.119\,8\\0.112\,3\end{bmatrix} = \begin{bmatrix}A_1\\A_2\\A_3\\B_1\\B_2\\B_3\\C_1\\C_2\\C_3\end{bmatrix}$$

具体计算过程见附录A。由以上计算得出各

个因素对指标影响的最优组合为晶格大小水平 3、壁厚水平 1、偏移尺寸水平 1。

4.3　与 TTHC 换热器对比分析

对于正交设计的结构方案, 表 4 可以看出方案 5 努塞尔数最高, 为 15 362.57(F=83.73), 方案 1 摩擦系数最低, 为 1.10(F=−0.83)。图 10 为热侧质量流量在 0.005 1～0.006 1 kg/s 范围变化时, 参考的 TTHC 换热器和方案 5 的热侧空气出口温度的对比。随着热侧质量流量的增加, 热侧出口温度在逐渐增加。在热侧质量流量 0.005 1 kg/s 时, 方案 5 的热侧出口温度比参考 TTHC 换热器降低 4.5%, 出口平均温度降低 13.13 K。这主要是由于 Gyroid 换热器与 TTHC 换热器内部流动的不同。图 11 为热侧质量流量在 0.005 1～0.006 1 kg/s 范围变化时, 方案 5 热侧温度分布云图及速度流线图。图 12 为在热侧质量流量变化从 0.005 1～0.006 1 kg/s 热侧 TTHC 换热器的速度流线图。Gyroid 换热器流动通道复杂, 流线在通道内呈螺旋流动, 且速度的最大值发生在曲率较大的地方, 随着热侧质量流量的增加, 高流速区域逐渐增加, 低温区域出现位置逐渐后移。对于 TTHC 换热器, 在通道内流线仅有轻微偏转, 没有形成 Gyroid 换热器的螺旋流动

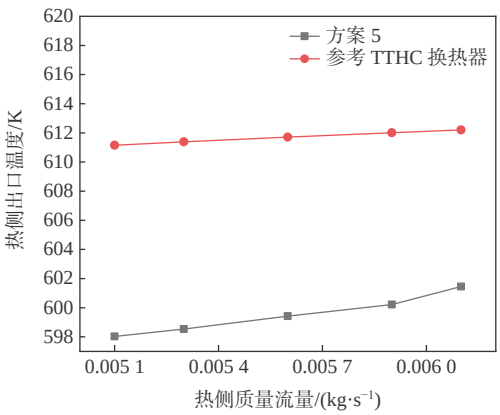


图 10　不同质量流量下空气出口温度对比

Fig. 10　Comparison of air outlet temperatures at different mass flow rates

图 13 为冷侧质量流量在 0.001～0.004 kg/s 变化时, 参考的 TTHC 换热器和方案 1 冷侧压降的变化。随着冷侧质量流量的增加, 两者之间压降差距逐渐拉大, 并在质量流量 0.004 kg/s 时, 压降差距最大为 3.02 kPa, 较 TTHC 换热器压降降低 22.09%。湍流强度的变化对此有很大的影响。图 14 是在冷侧质量流量 0.001～0.004 kg/s 变化时, 冷侧湍流动能沿流向的分布。由于 TTHC 换热器结构呈螺旋状, 无法捕捉合适截面的湍流动能分布, 故截取出口一小部分的湍流动能分布与 Gyroid 换热器对比分析。随着质量流量的增加, 换热器的湍流动能逐渐增加。与 TTHC 换热器相比, Gyroid 换热器具有较强的湍流动能, 这主要是由于 Gyroid 换热器结构

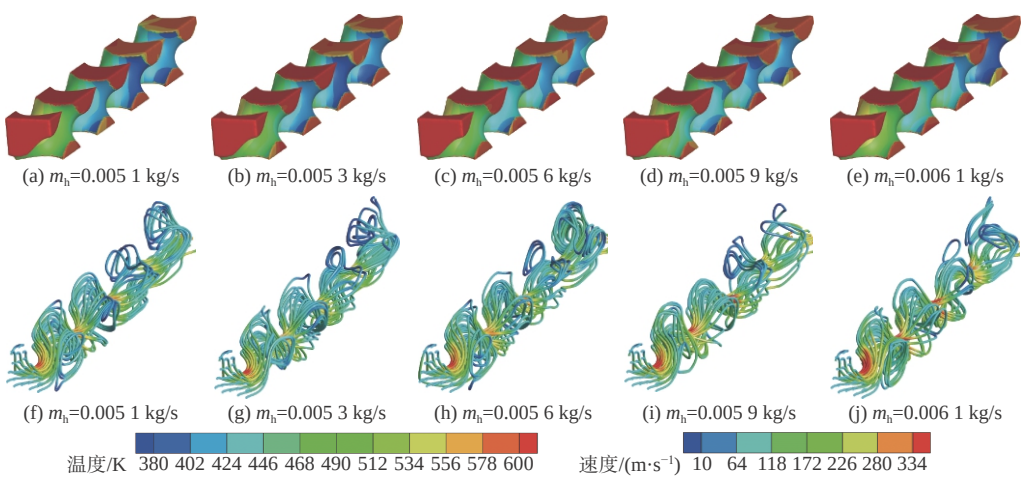


图 11　不同质量流量下 Gyroid 换热器温度云图和速度流线图

Fig. 11　Temperature contour and velocity streamline plot of Gyroid heat exchanger at different mass flow rates

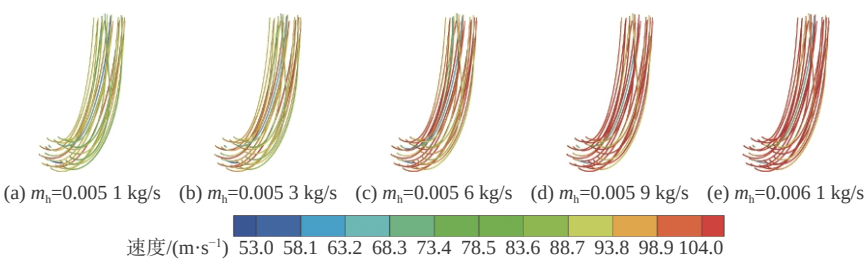


图 12　不同质量流量下 TTHC 换热器的速度流线图

Fig. 12　Velocity streamline of TTHC heat exchanger at different mass flow rates

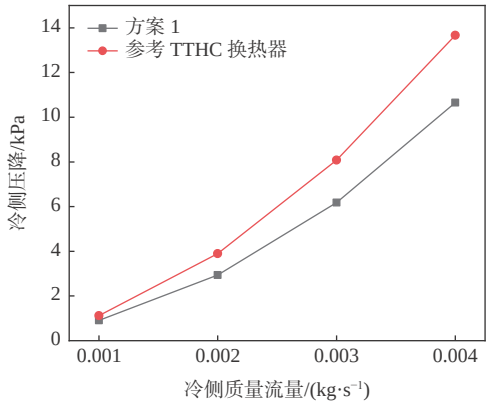


图 13 不同质量流量下燃油压降的变化

Fig. 13 Changes in fuel pressure drop at different mass flow rates

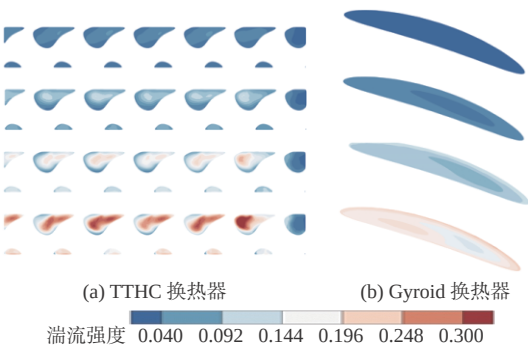


图 14 不同质量流量下湍流动能分布

Fig. 14 Distribution of turbulence kinetic energy at different mass flow rates

更为复杂，掺混强烈。湍流动能增强的同时，传热性能也会增强，但同时也会带来更多的摩擦阻力的损失。

表 7 为各个方案对应的冷热通道的性能评估系数，冷热通道呈一定的差异，主要是由于结构参数设计方案的不同，导致冷热通道相对于 TTHC 换热器的性能评估系数呈现出不同的变化。对于冷侧通道，方案 3 性能评估系数最高，为 0.78，方案 5 次之，为 0.73。对于热侧通道，方案 5 的性能评估系数最高，为 14.72，方案 3 次之，为 12.10。综上所述，结构参数的变化对换热器整体性能会产生比较大的影响，合理化的设计对取得传热和压降的平衡至关重要。

表 7 冷热通道的性能评估系数

Table 7 Performance evoluation coefficient for hot and cold channels

方案	冷侧性能评估系数	热侧性能评估系数
1	0.56	11.59
2	0.55	11.42
3	0.78	12.10
4	0.67	11.70
5	0.73	14.72
6	0.70	11.93
7	0.60	10.90
8	0.74	9.13
9	0.50	10.65

5 结 论

1）对于努塞尔数，各项参数对 Gyroid 换热器影响顺序为晶格大小>偏移尺寸>壁厚，最优方案是晶格大小为 7 mm，壁厚为 0.5 mm，偏移尺寸为 0.4 mm；对于摩擦系数，各项参数对 Gyroid 换热器影响顺序为偏移尺寸>晶格大小>壁厚，最优方案是晶格大小为 8 mm，壁厚为 0.4 mm，偏移尺寸为 0.2 mm。

2）整合数值仿真结果，方案 5 努塞尔数最大，为 15 362.57（F=83.73），方案 1 摩擦系数最低，为 1.10（F=−0.83）。综合各个因素对努塞尔数和摩擦系数的影响，根据权重的大小，确定最优方案为晶格大小水平 3，壁厚水平 1，偏移尺寸水平 1。

3）与 TTHC 换热器相比，方案 5 换热器空气出口温度最高降低 13.13 K，速度流线分布呈螺旋状，换热能力更强；方案 1 压降低于 TTHC 换热器最高 3.02 kPa，湍流动能更强，方案 3 的冷侧通道性能评估系数最高，为 0.78。方案 5 的热侧通道性能评估系数最高，为 14.72。

本文关于 Gyroid 换热器的研究成果希望可以指导换热器的设计，为 TPMS 在空气-燃油换热器上的应用提供参考。

参考文献（References）

[1] VLAHINOS M, O'HARA R. Unlocking advanced heat exchanger design and simulation with nTop platform and ANSYS CFX[EB/OL]. https://www.revolutioninsimulation.org/wp-content/uploads/2020/06/WP_HeatExchangerDesignSimulation.pdf.

[2] REYNOLDS B W. Simulation of flow and heat transfer in 3D printable triply periodic minimal surface heat exchangers[D]. Christchurch: University of Canterbury, 2020.

[3] WADSOE I, HOLMQVIST S. Additively manufactured heat exchangers—development and testing[D]. Lund: Lund University, 2020.

[4] KIM J, YOO D J. 3D printed compact heat exchangers with mathematically defined core structures[J]. Journal of Computational Design and Engineering, 2020, 7(4): 527-550.

[5] IYER J, MOORE T, NGUYEN D, et al. Heat transfer and pressure drop characteristics of heat exchangers based on triply periodic minimal and periodic nodal surfaces[J]. Applied Thermal Engineering, 2022, 209: 118192.

[6] DIXIT T, AL-HAJRI E, PAUL M C, et al. High performance, microarchitected, compact heat exchanger enabled by 3D printing[J]. Applied Thermal Engineering, 2022, 210: 118339.

[7] DHARMALINGAM L K, AUTE V, LING J Z. Review of triply

periodic minimal surface (TPMS) based heat exchanger designs[C]//Proceedings of the 19th International Refrigeration and Air Conditioning Conference at Purdue. West Lafayette : Purdue University, 2022.

[8] PENG H, GAO F, HU W J. Design, modeling and characterization on triply periodic minimal surface heat exchangers with additive manufacturing[C]//Proceedings of the 30th Annual International Solid Freeform Fabrication Symposium. Austin: University of Texas, 2019, 2325–2337.

[9] LI W H, YU G P, YU Z B. Bioinspired heat exchangers based on triply periodic minimal surfaces for supercritical CO_2 cycles[J]. Applied Thermal Engineering, 2020, 179: 115686.

[10] LI W G, LI W H, YU Z B. Heat transfer enhancement of water-cooled triply periodic minimal surface heat exchangers[J]. Applied Thermal Engineering, 2022, 217: 119198.

[11] FEMMER T, KUEHNE A J C, WESSLING M. Estimation of the structure dependent performance of 3-D rapid prototyped membranes[J]. Chemical Engineering Journal, 2015, 273: 438-445.

[12] ALTENEIJI M, ALI M I H, KHAN K A, et al. Heat transfer effectiveness characteristics maps for additively manufactured TPMS compact heat exchangers[J]. Energy Storage and Saving, 2022, 1(3): 153-161.

[13] LIANG D, SHI C W, LI W H, et al. Design, flow characteristics and performance evaluation of bioinspired heat exchangers based on triply periodic minimal surfaces[J]. International Journal of Heat and Mass Transfer, 2023, 201: 123620.

[14] WEN J, HUANG H R, LI H W, et al. Thermal and hydraulic performance of a compact plate finned tube air-fuel heat exchanger for aero-engine[J]. Applied Thermal Engineering, 2017, 126: 920-928.

[15] LIU P H, WANG R T, LIU S B, et al. Experimental study on the thermal-hydraulic performance of a tube-in-tube helical coil air–fuel heat exchanger for an aero-engine[J]. Energy, 2023, 267: 126626.

[16] HERRING N R, HEISTER S D. On the use of wire-coil inserts to augment tube heat transfer[J]. Journal of Enhanced Heat Transfer, 2009, 16(1): 19-34.

[17] 杜默, 孟宝, 潘丰, 等. 考虑制造约束的印刷电路板式换热器优化设计[J]. 北京航空航天大学学报, 2022, 48(10): 1994-2005.
DU M, MENG B, PAN F, et al. Optimal design of printed circuit heat exchanger considering manufacturing constraints[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(10): 1994-2005(in Chinese).

[18] TANDEL S, RAMMOHAN S. Design for additive manufacturing of high performance heat exchangers[D]. Hamilton: Mcmaster University, 2022: 77-78.

[19] JIANG J H, WANG F H, YANG X, et al. Evaluation of the long-term performance of the deep U-type borehole heat exchanger on different geological parameters using the Taguchi method[J]. Journal of Building Engineering, 2022, 59: 105122.

[20] 魏效玲, 薛冰军, 赵强. 基于正交试验设计的多指标优化方法研究[J]. 河北工程大学学报 (自然科学版), 2010, 27(3): 95-99.
WEI X L, XUE B J, ZHAO Q. Optimization design of the stability for the plunger assembly of oil pumps based on multi-target orthogonal test design[J]. Journal of Hebei University of Engineering (Natural Science Edition), 2010, 27(3): 95-99(in Chinese).

附录 A

$$\boldsymbol{M}_1=\begin{bmatrix}10\,835.92 & 0 & 0\\ 13\,328.40 & 0 & 0\\ 12\,764.61 & 0 & 0\\ 0 & 12\,110.81 & 0\\ 0 & 12\,423.17 & 0\\ 0 & 12\,394.95 & 0\\ 0 & 0 & 11\,611.02\\ 0 & 0 & 12\,043.94\\ 0 & 0 & 13\,273.97\end{bmatrix}$$

$$\boldsymbol{T}_1=\begin{bmatrix}0.000\,027 & 0 & 0\\ 0 & 0.000\,027 & 0\\ 0 & 0 & 0.000\,027\end{bmatrix}$$

$$\boldsymbol{S}_1^{\mathrm{T}}=\begin{bmatrix}0.558 & 0.070 & 0.372\end{bmatrix}$$

$$\boldsymbol{M}_2=\begin{bmatrix}0.64 & 0 & 0\\ 0.55 & 0 & 0\\ 0.78 & 0 & 0\\ 0 & 0.71 & 0\\ 0 & 0.67 & 0\\ 0 & 0.57 & 0\\ 0 & 0 & 0.76\\ 0 & 0 & 0.67\\ 0 & 0 & 0.54\end{bmatrix}$$

$$\boldsymbol{T}_2=\begin{bmatrix}0.51 & 0 & 0\\ 0 & 0.51 & 0\\ 0 & 0 & 0.51\end{bmatrix}$$

$$\boldsymbol{S}_2^{\mathrm{T}}=\begin{bmatrix}0.37 & 0.24 & 0.39\end{bmatrix}$$

$$\boldsymbol{\omega}=\frac{1}{2}(\boldsymbol{\omega}_1+\boldsymbol{\omega}_2)=\frac{1}{2}\left(\begin{bmatrix}0.163\,3\\ 0.200\,8\\ 0.192\,3\\ 0.022\,9\\ 0.023\,5\\ 0.023\,4\\ 0.102\,5\\ 0.106\,3\\ 0.117\,2\end{bmatrix}+\begin{bmatrix}0.120\,8\\ 0.103\,8\\ 0.147\,2\\ 0.086\,9\\ 0.082\,0\\ 0.069\,8\\ 0.151\,2\\ 0.133\,3\\ 0.107\,4\end{bmatrix}\right)=$$

$$\begin{bmatrix}0.142\,1\\ 0.152\,3\\ 0.169\,8\\ 0.054\,9\\ 0.052\,8\\ 0.046\,6\\ 0.126\,9\\ 0.119\,8\\ 0.112\,3\end{bmatrix}=\begin{bmatrix}A_1\\ A_2\\ A_3\\ B_1\\ B_2\\ B_3\\ C_1\\ C_2\\ C_3\end{bmatrix}$$

Performance of Gyroid heat exchanger based on structural parameters analysis

YANG Xiaojun，ZHANG Xueli*，LI Peiran

(College of Aeronautical Engineering，Civil Aviation University of China，Tianjin 300300，China)

Abstract：In order to study the flow and heat transfer characteristics of triple periodic minimal surfaces (TPMS) in an air-fuel heat exchanger, the influence of structural parameters on the performance of the heat exchanger was analyzed. By using the Taguchi method, the Nusselt number and friction coefficient were comprehensively evaluated for the three structural parameters of wall thickness, unit cell size, and offset size and compared with the performance of a tube-in-tube helical coil (TTHC) heat exchanger. The primary and secondary order of each structural parameter affecting the Nusselt number and friction coefficient of the Gyroid heat exchanger was obtained, and the optimal scheme was determined by using the matrix analysis method. The results show that compared with that of the TTHC heat exchanger, the hot-side outlet temperature of the Gyroid heat exchanger is reduced by up to 13.13 K, and the pressure drop of the Gyroid heat exchanger is reduced by 3.02 kPa. The highest performance evaluation coefficient of the hot-side channel is 14.72, and that of the cold-side channel is 0.78.

Keywords：Gyroid heat exchanger；tube-in-tube helical coil heat exchanger；Taguchi method；structural parameters；matrix analysis method

Received：2023-05-12；**Accepted**：2023-10-06；**Published Online**：2024-10-09 13：15

URL：link.cnki.net/urlid/11.2625.V.20241009.1033.001

Foundation item：The Fundamental Research Funds for the Central Universities (3122019187)

* **Corresponding author**. E-mail：zxl982511@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0249

基于相邻特征融合与特征解耦的一阶段目标检测

郑剑，贺朝辉，于祥春*
(江西理工大学 信息工程学院，赣州 341000)

摘 要：针对目标检测中特征金字塔网络(FPN)造成的大尺度目标检测精度下降及目标检测2个子任务所需语义特征不一致问题，提出一种基于相邻特征融合(AFF)与特征解耦的网络(AFFDN)模型。该模型中的AFF模块通过多对一连接引入更短的梯度回传路径，缓解了大尺度目标梯度消失的问题；AFF模块通过共享参数和偏移量，有效减小了模型参数量，并增强了多尺度特征语义一致性；相比于基于神经架构搜索的FPN(NAS-FPN)，AFF参数量更小、性能增益更显著。AFFDN模型中的特征解耦模块(FDM)通过动态感受野和全局注意力，在感受野-通道-空间3个维度上进行细粒度特征解耦，为不同分支生成特有的任务相关特征，进而提高目标检测精度。将AFFDN模型应用到不同的一阶段目标检测模型时，在PASCAL VOC和MS COCO2017数据集上与基线模型相比，检测精度分别提升了至少0.9%和2.3%。

关 键 词：目标检测；多尺度特征融合；特征解耦；注意力；可变形卷积
中图分类号：TP391.4
文献标志码：A **文章编号**：1001-5965(2025)04-1205-10

目标检测作为计算机视觉的一项基本且具有挑战性的任务，其本质是对图像中存在的物体进行分类和定位。目前，目标检测算法可大致分为二阶段目标检测算法与一阶段目标检测算法。①二阶段目标检测算法，如Faster R-CNN[1]、Grid R-CNN[2]和Cascade R-CNN[3]等，通过候选区域网络对骨干网络提取的特征信息进行前后景判断，进一步提取候选区域特征，并输入给后续的检测头进行分类。②一阶段目标检测算法，如RetinaNet[4]、全卷积一阶段目标检测(fully convolutional one-stage object detection, FCOS[5])和YOLOX[6]等，是将定位任务视作回归任务，并与分类任务同时进行求解。由于一阶段目标检测算法没有使用区域候选网络提取感兴趣区域特征信息，节省了大量计算资源，在实时性要求高的应用场景下，相较于二阶段目标检测算法，一阶段目标检测算法有着更为广泛的应用[7-9]。

特征金字塔网络[10](feature pyramid network, FPN)作为一种基于Neck的多尺度特征融合方式，有效提升了目标检测算法精度。文献[11-14]在FPN的基础上引入了一条或多条额外路径，通过拓宽FPN获取更多的节点特征信息进行多尺度特征融合，以增强多尺度特征表达能力。但这些基于FPN的改进工作忽视了一个问题，即FPN的总体性能增益是建立在小目标平均精度的提高和大目标平均精度的降低之上。Jin等[15]从优化的角度出发，证明了该现象的出现是因为FPN增加了大尺度目标梯度回传路径的长度，导致了大尺度目标梯度信号更容易消失在模型浅层。双向特征金字塔网络(bidirectional feature pyramid network, Bi-FPN[12])、广义特征金字塔网络(generalized feature pyramid network, G-FPN[13])通过跳跃连接试图减小路径长度，但其大尺度目标梯度回传路径的最短长度仍没

收稿日期：2023-05-15；**录用日期**：2023-12-11；**网络出版时间**：2023-12-27 18:57
网络出版地址：link.cnki.net/urlid/11.2625.V.20231227.1507.004
基金项目：江西省自然科学基金(20224BAB212013)
*通信作者. E-mail：yuxc@jxust.edu.cn

引用格式：郑剑，贺朝辉，于祥春. 基于相邻特征融合与特征解耦的一阶段目标检测[J]. 北京航空航天大学学报，2025，51(4)：1205-1214. ZHENG J，HE Z H，YU X C. One-stage object detection based on adjacent feature fusion and feature decoupling[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1205-1214 (in Chinese).

有减小,这无疑限制了多尺度特征融合策略的表现能力。注意力聚合特征金字塔网络 (attention aggregation feature pyramid network, A^2-FPN[16]) 利用注意力机制引导多尺度特征融合过程,在分割任务上表现出不错的效果。刘树东等[9]通过将注意力机制嵌入多尺度特征融合模块,提高了无人机航拍图像小目标检测精度。Luo 等[14]提出通道注意力引导模块,在特征融合过程中对各层级特征权重进行优化。空间自适应特征融合 (adaptively spatial feature fusion, ASFF[17]) 引入自适应特征融合模块,通过恒等放缩和自适应融合将所有尺度特征信息进行多尺度特征融合,但当 FPN 输出层级增加时,其线性插值倍数的增大会引入过多噪声干扰训练,且参数量大幅上升。Liu 等[18]通过高分辨率特征图,引导相邻的低分辨率特征图进行多尺度特征融合,以补充局部特征细节,但高分辨率特征图在特征融合过程中缺少来自于低分辨率特征图的引导。双重细化特征金字塔网络 (dual refinement feature pyramid network, DRFPN[19]) 提出空间细化模块和通道细化模块,用相邻层级的特征信息来提升空间采样和通道融合的准确率,但参数量和计算量相对较高。

分类任务和定位任务所需的最优特征不一致是一阶段目标检测中另一关键问题。具体来说,分类任务依赖于目标导向的判别特征,以提取丰富的语义信息,定位任务更倾向于检测物体边界,以提取目标边缘信息,这导致在模型优化过程中,2 个任务所需特征空间分布存在着显著差异,导致难以同步优化。Wu[20]和 Song[21]等提出使用解耦头分别执行分类和定位任务,有效缓解分类和定位任务特征空间不匹配的影响,提高模型检测性能。任务对齐一阶段目标检测 (task-aligned one-stage object detection, TOOD[22]) 在解耦头的基础上,将各级输出特征图在通道维度上进行拼接,使用通道注意力对特征在通道维度进行解耦。解耦密集目标检测器 (disentangle your dense object detector, DDOD[23]) 将解耦头 2 个任务分支上的标准卷积全部替换为可变形卷积,使分类和定位分支具有不同的感受野区域,从感受野维度挖掘不同任务所需的最优特征信息,实现基于感受野的特征解耦。但 TOOD 和 DDOD 都存在解耦维度不全面的问题。dynamic head[24]是由全注意力机制组成的检测头,但其参数量较大。路由解耦网络 (disentangled part-object routing network, DPORTNet[25]) 提出解耦胶囊路由,将特征图在水平和竖直方向上分开计算,目的是加快模型推理速度。

针对上述 FPN 造成的大尺度目标检测精度下降和一阶段目标检测 2 个子任务所需语义特征不一致的问题,本文提出一种基于相邻特征融合与特征解耦的网络 (adjacent feature fusion and feature decoupling network, AFFDN) 模型。AFFDN 模型包括相邻特征融合 (adjacent feature fusion, AFF) 模块和特征解耦模块 (feature decoupling module, FDM)。本文提出的 AFFDN 模型的创新点如下:

1) 设计了一种基于 Head 的 AFF 模块,其采取一种新的多对一的连接方式,使各尺度 Head 能够获取到多个相邻 FPN 层级的输出。同时,为大尺度目标梯度信号提供了更短的传播路径,减小 FPN 所造成的大尺度目标检测精度降低。此外,通过参数共享和自适应加权,减小模型参数量,增强多尺度表达能力。

2) 构建 FDM,利用可变形卷积和坐标注意力,将 AFF 模块输出的多尺度特征信息在感受野-通道-空间维度上进行细粒度的特征解耦,更准确地为分类任务和定位任务提供任务相关的最优特征,加快模型训练收敛。

3) 即插即用,大量实验表明,在 PASCAL VOC 数据集和 MS COCO2017 数据集上,AFFDN 模型与主流一阶段目标检测模型整合,如自适应训练样本选择 (adaptive training sample selection, ATSS)、广义焦点损失 (generalized focal loss, GFL) 和 TOOD,能将上述基线模型的检测精度分别提升至少 0.9 和 2.3。

1 本文模型

本文模型遵循一阶段目标检测模型框架设计范式,其由 Backbone、Neck 和 Head 组成,如图 1 所示。

AFFDN 模型在 ATSS 的基础上新增了 AFF 模块和 DFM。其中,Backbone 用于从输入图像提取特征信息,Neck 用于生成多尺度特征,Head 负责预测目标的分类和定位。本文 AFFDN 模型和 ATSS[26]等一阶段目标检测模型的区别主要在于 Head 部分。首先,AFFDN 模型通过多对一的方式连接 FPN 和 Head,获取多个 FPN 层级的输出特征,用以减小梯度回传路径长度。然后,利用参数共享减小模型参数量。在特征融合过程中,采用自适应加权和快速归一化,进一步增强特征的多尺度表达能力。最后,通过 FDM,在感受野、通道和空间 3 个维度对输入特征进行解耦,生成任务相关的最优特征。AFFDN 模型即插即用,可以很容易整合到 ATSS[26]和 GFL[27]等主流一阶段目标检测模型中,

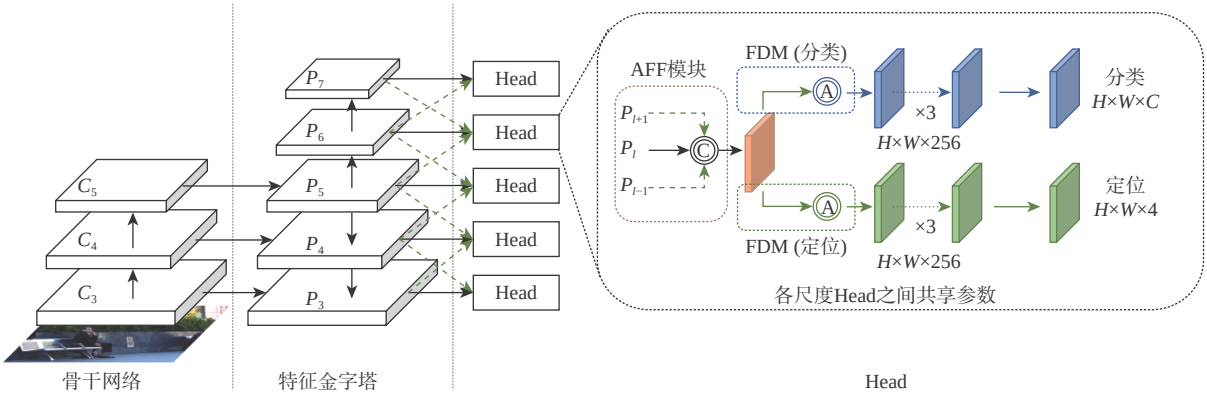


图 1　AFFDN 结构

Fig. 1　Structure of AFFDN

并提升模型检测性能。

1.1　相邻特征融合模块

多尺度特征融合是提升目标检测模型性能的常用策略。在过去几年, FPN 一直被用于各类先进目标检测模型, 后续许多工作都致力于探究如何更好地融合各尺度特征, 但这些工作忽略了引入 FPN 后, 大尺度目标梯度消失和检测精度下降的问题。如图 2 所示, FPN 的总体性能增益建立在小目标平均精度的提高和大目标平均精度的降低之上。

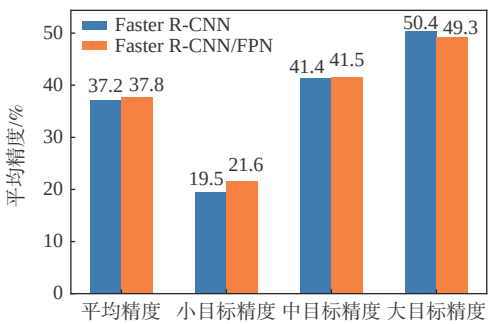


图 2　添加 FPN 后 Faster R-CNN 的性能变化

Fig. 2　Performance change of Faster R-CNN after adding FPN

FPN 与 Head 之间通常是一对一连接, FPN 各层级输出作为对应尺度 Head 的输入, 如图 3(a) 所示。引入 FPN 后, 加大了来自 H_6 和 H_7 的大尺度目标梯度信号传播到模型浅层(如 C_3)的难度。后续一些基于 FPN 的改进工作, 如路径聚合网络 (path aggregation network, PAN[11]), 如图 3(b) 所示。PAN 在 FPN 的基础上增加了一条额外的自底向上的路径, 通过融合更多节点的特征信息, 增强多尺度特征表达能力, 但加大了 FPN 的宽度, 增长了大尺度目标梯度信号传播路径的长度, 进一步加剧了 FPN 对大尺度目标检测精度的影响。

基于上述工作存在的问题, 本文提出 AFF 模块, 通过多对一连接, 使各尺度 Head 能获取到多个相邻尺度 FPN 的输出, 减小大尺度目标梯度回传路径长度, 缓解 FPN 造成的大尺度目标检测精度降低的影响。在特征融合过程中, AFF 模块通过参数共享减小参数量, 采用自适应加权和快速归一化, 进一步增强特征的多尺度表达能力, 提升模型检测精度。AFF 模块梯度回传如图 3(c) 所示, 其结构如图 4 所示。

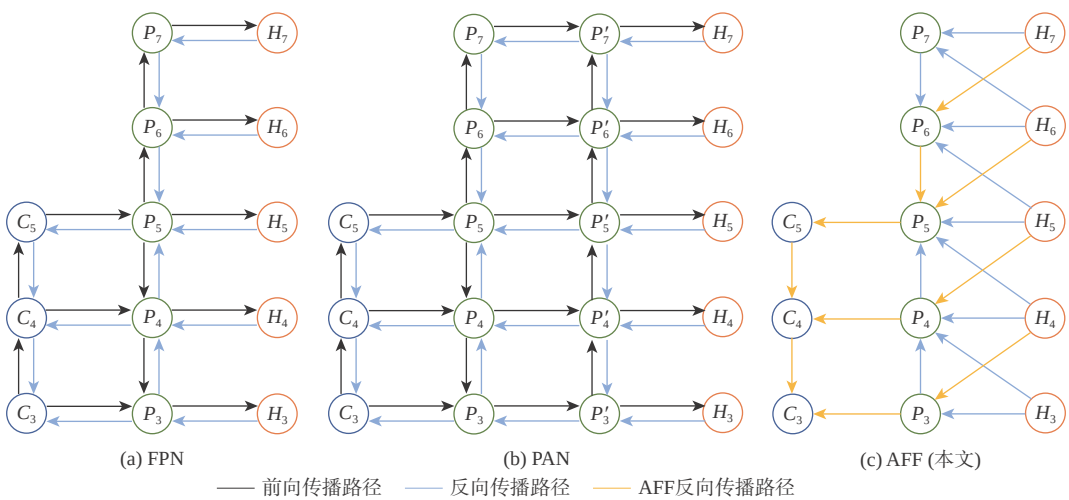


图 3　不同多尺度特征融合方法下梯度信号的传播路径

Fig. 3　Propagation paths of gradient signals under different multi-scale feature fusion methods

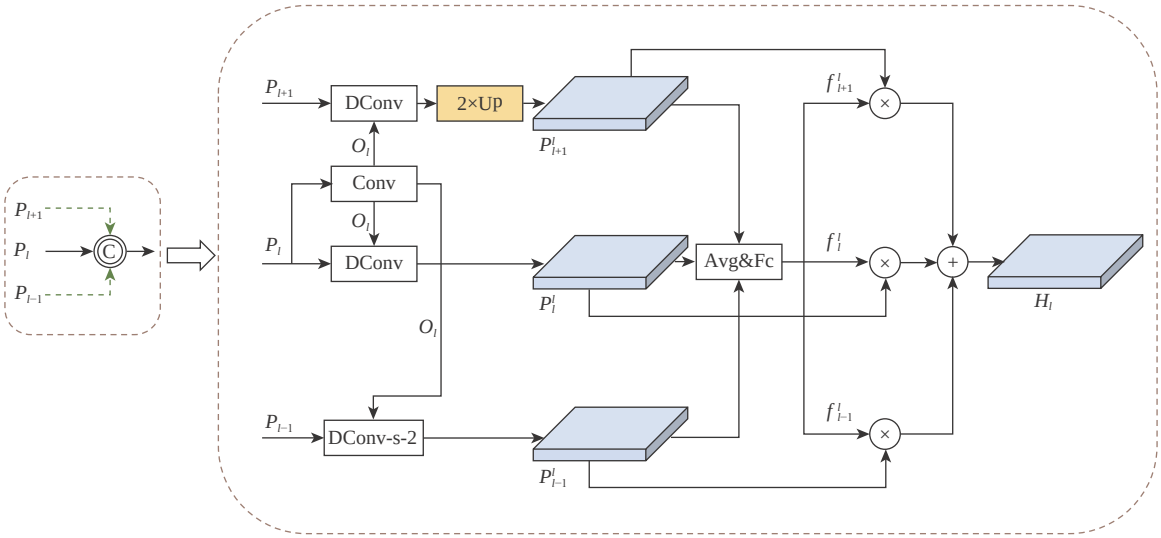


图4 AFF模块

Fig. 4 AFF module

1.1.1 多对一连接

多对一连接指每级Head采用相邻多级FPN的输出作为输入。由于FPN最顶层和最底层输出缺少相邻层级输出，最顶层和最底层Head为二对一连接，其余Head为三对一连接，具体可参考图1中绿线部分。与FPN和Head之间原有的一对一连接方式相比，多对一连接为大尺度和中尺度目标额外提供一条更短的梯度回传路径，如图3(c)所示。为保持整洁，图3(c)中省略了前向传播路径。具体来说，H_l的输入是P_{l-1}、P_l和P_{l+1}，而不仅仅是P_l，P_{l-1}的引入能够减小l层Head梯度信息回传的最短路径长度，减小模型浅层出现梯度消失的可能性。假设图3中的每条线段代表该路径长度为1，则H_7到C_3的最短路径长度在FPN中为6，在PAN中为7，而在本文AFF模块中仅为5。除此之外，每级Head的监督信号可以同时作用于多个FPN层级。由于各尺度目标的监督信号增多，一定程度上减小了梯度信号传播到模型浅层时出现梯度消失的可能性。

1.1.2 基于参数共享的特征采样

AFFDN模型中的Head借鉴了解耦头中参数共享的思想，将各尺度Head中的AFF模块和FDM的参数进行共享，减小了模型参数量。在特征融合过程中，由于FPN各层级输出特征尺寸不一致，需要在融合前对特征进行采样。AFF模块通过卷积操作和双线性插值统一输入尺寸大小，实现上下采样。针对各尺度特征之间存在的语义差距问题，AFF模块在采样过程中，各尺度Head都通过对应FPN层级的输出生成偏移量，并将偏移量共享给该尺度Head中的其余FPN层级输出，通过共享偏移量减小语义差距：

$$O_l = \mathrm{Conv}(P^l_l) \tag{1}$$

$$P^l_{l+1} = \mathrm{Up}(\mathrm{DConv}(P_{l+1}, O_l)) \tag{2}$$

$$P^l_l = \mathrm{DConv}(P_l, O_l) \tag{3}$$

$$P^l_{l-1} = \mathrm{DConv\text{-}s\text{-}2}(P_{l-1}, O_l) \tag{4}$$

式中：l为层级；O_l为第l层Head的共享偏移量；P_l为FPN的第l层输出；P^l_{l+1}、P^l_l和P^l_{l-1}分别表示在第l层Head中经过上采样的P_{l+1}、经过横向连接的P_l和经过下采样后的P_{l-1}；s为卷积的步长；Up表示上采样倍数为2的双线性插值。

1.1.3 自适应加权和快速归一化融合

不同尺度特征在各尺度Head中的重要性不同。AFF模块根据输入特征的不同，通过卷积和池化操作，在各尺度Head上自适应生成动态的权重因子。在生成权重因子的过程中，考虑到Hard Sigmoid避免了指数运算，且计算速度优于Sigmoid，AFF模块采用其作为激活函数。针对多尺度特征相加，实验数据表明了对各尺度特征的权重因子进行归一化的必要性。同理，为避免Softmax中的指数运算，AFF模块采用了Fast normalized fusion[12]实现归一化：

$$f(x) = \mathrm{HardSigmoid}(\mathrm{Conv}(\mathrm{Avg}(x))) \tag{5}$$

$$\mathrm{Add}(x_1, x_2, \cdots, x_n) = \frac{\sum_{i=1}^{n} x_i f(x_i)}{\varepsilon + \sum_{i=1}^{n} f(x_i)} \tag{6}$$

$$H_l = \begin{cases} \mathrm{Add}(P^l_l, P^l_{l+1}) & l = 3 \\ \mathrm{Add}(P^l_{l-1}, P^l_l, P^l_{l+1}) & 4 \leqslant l \leqslant 6 \\ \mathrm{Add}(P^l_{l-1}, P^l_l) & l = 7 \end{cases} \tag{7}$$

式中：x为输入特征；H_l为Head中经过多尺度特征融合处理后的第l层输入；ε为极小常数，用以避免

数值下溢。

1.2　特征解耦模块

由于分类任务和定位任务所需的最优特征不一致, 需要进行特征解耦, 用以生成任务相关特征。本文 FDM 对输入特征在感受野-通道-空间 3 个维度进行细粒度的特征解耦, 提取更优的任务相关特征, 加速训练收敛。

1.2.1　特征解耦模块的基本结构

本文 FDM 很大程度上受 DDOD[23] 和 TOOD[22] 这 2 项工作的启发, 与上述工作不同的是, FDM 舍弃了其中高计算量的部分, 并实现了更细粒度的特征解耦, 如图 5 所示。在感受野维度的特征解耦方面, DDOD 将 Head 中的所有标准卷积替换为了可变形卷积 (Deformed Conv), 如图 5(a) 所示。为减小计算量, FDM 仅将各个任务分支上的首个卷积替换为可变形卷积。同理, FDM 舍弃了 TOOD 在图 5(b) 中对特征进行堆叠的做法。受 TOOD 的启发, 本文 FDM 通过单个可变形卷积和通道-空间注意力, 实现基于感受野-通道-空间的特征解耦, FDM 的基本结构如图 5(c) 中虚线框所示。

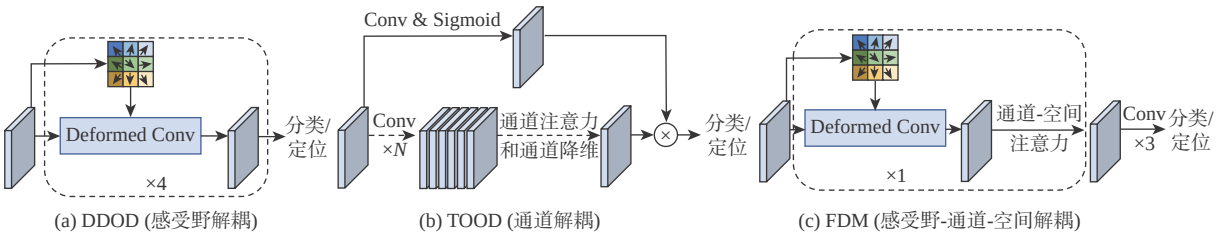


图 5　不同特征解耦方法的对比

Fig. 5　Comparison of different feature decoupling methods

1.2.2　坐标注意力

FDM 中, 通道-空间注意力的具体实现是采用坐标注意力。坐标注意力模块如图 6 所示。

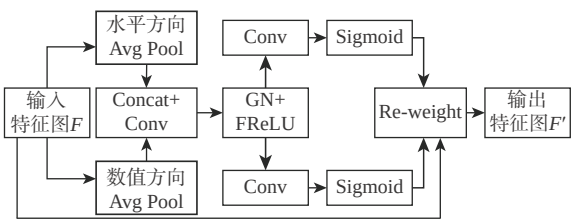


图 6　坐标注意力模块

Fig. 6　Coordinate attention module

坐标注意力的计算过程分为 2 步: ①嵌入坐标信息, ②生成通道-空间注意力。首先, 坐标注意力将全局池化分解为水平和竖直 2 个方向的池化, 实现坐标信息的嵌入; 然后, 为加强这 2 个方向上特征之间的联系, 将池化后的 2 个特征图在空间维度上拼接, 并经过 1×1 卷积、组归一化(GN)和 FReLU 后; 再次, 沿空间维度分为水平和竖直 2 个方向的通道空间注意力图; 最后, 将这 2 个方向上的注意力图通过 Sigmoid 进行数值归一化, 并和原始输入特征相乘, 得到最终输出特征。其中, 使用 FReLU 是为了加强空间感知能力, 而组归一化是为了避免批归一化引起训练和推理不一致。

2　实验结果与分析

2.1　实验数据集与评价指标

实验所采用数据集为目标检测任务中常用的 MS COCO2017 数据集、PASCAL VOC 2007 数据集、PASCAL VOC 2012 数据集。在消融实验中, 添加了用于安检领域的 X-Ray 目标检测数据集, 用以证明本文工作在特殊场景下的适用性。

实验评价指标为类别平均精度(mean average precision, mAP)。mAP 指 IoU 从 0.5 到 0.95 取值, 步长为 0.05, 计算在这些 IoU 下的 AP 的均值。mAP50 指 IoU=0.5 时的 mAP, mAP75 同理。mAPS、mAPM 和 mAPL 分别为小尺度目标、中尺度目标和大尺度目标的 mAP。AP 和 mAP 的计算式如下:

$$P_{\text{recision}} = \frac{T_{\text{P}}}{T_{\text{P}} + F_{\text{P}}} \tag{8}$$

$$R_{\text{ecall}} = \frac{T_{\text{P}}}{T_{\text{P}} + F_{\text{N}}} \tag{9}$$

$$A_{\text{P}} = \int_0^1 P_{\text{recision}} R_{\text{ecall}} \mathrm{d}R_{\text{ecall}} \tag{10}$$

$$m_{\text{AP}} = \frac{\sum_{i=1}^{N} A_{\text{P}i}}{N} \tag{11}$$

式中: N 为类别总数; T_{P} 表示真阳性, 即正样本被正确识别; F_{P} 表示假阳性, 即负样本被错误标识别; F_{N} 表示假阴性, 即正样本被错误识别。

2.2　实验设置

实验环境: 操作系统为 Ubuntu 18.04, GPU 为 4 张 RTX 2080Ti, CUDA 版本为 10.2, Python 版本为 3.8, PyTorch 版本为 1.10。

训练细节: 使用有 ImageNet 预训练权重的 ResNet-50 作为骨干网络并进行初始化。按照标准训练流程进行模型训练, 采用单尺度训练策略, 训

练轮次为12 epochs,表示为1×。初始学习率为0.01,在第8轮和第11轮epoch之后,衰减为原来的1/10。所有实验均在4个GPU上使用SGD优化器进行训练,batchsize为16。训练和推理过程中,图像分辨率为1 333×800。推理时,过滤掉阈值为0.05的背景框,并通过阈值为0.6的非极大值抑制删除冗余框,以获得最终预测结果。

2.3 在MS COCO 2017数据集上的性能比较

本文选取了多个主流的一阶段目标检测模型与AFFDN模型进行比较,并提供了在MS COCO2017 minival数据集上的性能评估报告。报告中,所有模型的测试结果均来自单模单尺度测试,参数量和运算量均采用尺寸为1 280×800的输入图像进行测量得出。如表1所示,在没有添加任何训练技巧和测试增强方法的情况下,AFFDN模型将不同基线模型的检测精度提升了2.3%～2.6%。

2.4 在PASCAL VOC数据集上的性能比较

本文采取多数工作的做法,将PASCAL VOC 2007与PASCAL VOC 2012的训练集拼接在一起,并在PASCAL VOC 2007的验证集上进行测试。实验过程中,图像分辨率为1 000×600,其余训练参数与骨干网络选取均与MS COCO 2017数据集训练流程保持一致,实验数据如表2所示。在表2中,AFFDN模型能将ATSS等基线模型检测精度提高0.9%～1.3%,在飞机、公交车、桌子、火车和电视类别上检测精度提升较大。

2.5 消融实验

本文的消融实验评价结果来源于MS COCO 2017 minival数据集。根据标准训练流程,在MS COCO2017 trainset上进行训练。

表1 不同模型在MS COCO2017 minival数据集上的实验结果对比

Table 1 Comparison of experimental results of different models on MS COCO2017 minival dataset

模型	骨干网络	轮次	mAP/%	mAP50/%	mAP75/%	mAPS/%	mAPM/%	mAPL/%	参数量/MB	运算量
RetinaNet[4]	ResNet-50	12	36.5	55.4	39.1	20.4	40.3	48.1	37.74	239.32×10^9
RetinaNet[4]/CE-FPN[14]	ResNet-50	12	38.0	57.3	40.5	21.2	41.0	47.4	65.02	260.27×10^9
RetinaNet[4]/DRFPN[19]	ResNet-50	12	38.4	58.7	41.6	23.5	42.1	46.7	44.64	410.92×10^9
Double Head[20]	ResNet-50	12	40.1	59.4	43.5	22.9	43.6	52.9	47.12	481.26×10^9
FCOS[5]	ResNet-50	12	38.7	57.4	41.8	22.9	42.5	50.1	32.02	200.50×10^9
ATSS[26]	ResNet-50	12	39.4	57.6	42.8	23.6	42.9	50.3	32.07	205.21×10^9
ATSS[26]/NAS-FPN[28]	ResNet-50	12	40.9	59.2	44.3	24.9	44.3	52.4	38.89	200.54×10^9
ATSS[26]/AF-Head[29]	ResNet-50	12	41.1	59.2	44.5	23.6	44.5	53.7	34.69	210.79×10^9
ATSS[26]/Dy-Head[24]	ResNet-50	12	42.5	60.2	46.2	26.3	46.6	55.3	38.85	112.69×10^9
PAA[30]	ResNet-50	12	40.4	58.4	43.9	22.9	44.3	54.0	32.07	205.30×10^9
GFL[27]	ResNet-50	12	40.2	58.4	43.3	23.3	44.0	52.2	32.22	208.39×10^9
VarifocalNet[31]	ResNet-50	12	41.6	59.5	45.0	24.4	45.1	54.2	32.67	192.86×10^9
DW[32]	ResNet-50	12	42.1	59.9	45.1	24.2	45.3	55.9	32.09	205.73×10^9
TOOD[22]	ResNet-50	12	42.4	59.5	46.1	25.1	45.5	55.5	31.98	184.4×10^9
ATSS(本文)	ResNet-50	12	42.0	60.0	45.7	24.9	45.8	54.5	32.49	186.39×10^9
GFL(本文)	ResNet-50	12	42.5	60.4	46.2	24.6	46.3	55.8	32.64	189.49×10^9
TOOD(本文)	ResNet-50	12	44.7	61.9	48.6	25.6	47.8	58.6	32.40	165.50×10^9

表2 在PASCAL VOC 2007和PASCAL VOC 2012数据集上的实验结果

Table 2 Experiment results on PASCAL VOC 2007 and PASCAL VOC 2012 datasets %

模型	mAP																				平均mAP
	飞机	自行车	鸟	船	瓶子	公交车	汽车	猫	椅子	牛	桌子	狗	马	摩托车	人	盆栽	羊	沙发	火车	电视	
FCOS[5]	79.3	82.1	72.8	62.5	58.0	78.1	83.3	82.7	56.5	71.1	64.0	78.6	78.6	80.1	79.0	43.1	71.4	69.5	77.3	67.7	71.8
ATSS[26]	85.6	82.3	82.4	73.2	70.5	84.5	87.9	88.2	64.9	83.0	70.7	86.7	85.4	83.1	85.1	52.7	81.8	73.6	80.6	79.4	79.1
GFL[27]	84.3	82.7	80.3	69.6	70.6	83.0	87.6	87.5	63.5	85.0	68.9	85.5	84.0	81.9	84.4	52.8	83.4	72.2	84.6	78.0	78.5
TOOD[22]	85.5	84.6	82.7	72.4	71.9	84.7	88.3	88.1	64.4	84.7	71.9	86.6	86.1	84.0	84.9	55.3	81.9	74.5	83.8	80.6	79.8
VarifocalNet[31]	86.2	84.7	82.6	70.6	69.9	86.0	87.6	88.3	63.1	84.3	68.5	85.4	86.7	83.7	84.3	53.8	82.4	71.4	84.7	78.2	79.1
ATSS (本文)	86.6	85.4	82.2	74.8	71.9	87.3	88.1	88.8	65.4	84.9	72.8	86.7	85.6	83.8	85.5	56.3	81.0	72.5	85.2	81.6	80.3
GFL (本文)	87.0	84.1	81.7	74.0	71.3	85.1	87.6	88.6	64.2	85.2	71.2	86.1	85.1	83.5	84.8	54.8	81.4	73.2	85.9	80.8	79.8
TOOD (本文)	87.4	84.1	83.4	73.0	70.6	87.0	88.4	88.3	65.4	88.5	74.5	86.9	87.7	84.2	84.9	55.2	82.8	77.5	86.4	81.0	80.7

2.5.1　AFFDN 模型中各模块的有效性和通用性

实验采用了 2 种主流且差异较大的一阶段目标检测模型作为基线模型, 即 ATSS 和 TOOD, 用以评估本文工作的有效性和通用性, 训练过程严格遵循原始论文中提出的训练流程, 相关实验数据如表 3 所示。

表 3　AFFDN 模型消融实验结果

Table 3　Ablation experiment results of AFFDN model

模型	AFF	FDM	mAP/%	mAP50/%	mAP75/%
ATSS[26]	×	×	39.4	57.6	42.8
	√	×	41.2	59.3	45.0
	×	√	40.8	58.9	44.5
	√	√	42.0	60.0	45.7
TOOD[22]	×	×	42.4	59.5	46.1
	√	×	44.0	61.1	47.6
	×	√	43.5	60.8	47.1
	√	√	44.7	61.9	48.6

FDM 可以分为使用可变形卷积解耦感受野和使用全局注意力解耦通道和空间特征 2 部分。在以 ATSS 为基线模型的消融实验中可以观察到, 感受野解耦能促进通道-空间解耦, 实验数据如表 4 所示。如图 7 所示, 坐标注意力可以捕获长距离信息依赖, 将其引入到 FDM 后, 可更有效地提取到目标的边缘信息, 生成更精确的定位框, 加速定位任务的训练收敛。在以多个不同的基线模型进行实验的过程中, 都观察到类似现象。

表 4　FDM 消融实验结果

Table 4　Ablation experiment results of FDM

通道-空间	感受野	mAP/%	mAP50/%	mAP75/%
×	√	39.4	57.6	42.8
√	×	39.6	57.7	43.0
×	√	40.1	58.5	43.6
√	√	40.8	58.9	44.5

2.5.2　AFF 与 NAS-FPN 的对比

基于 Head 进行多尺度特征融合的 AFF 模块, 能显著提高 ATSS 的 AP, 在中尺度和大尺度目标上更为明显, 如表 5 所示, 这与提出 AFF 模块的目标是一致的, 即减小 FPN 带来的大尺度目标检测精度降低的影响。NAS-FPN[28] 通过神经架构搜索, 在相同等参数量的条件下, 优于 FPN[10]、PAN[11] 等传统人工设计的 Neck 网络架构。AFF 模块使用多对一连接和基于注意力机制的自适应特征融合提高特征多尺度表达能力, 并通过参数共享减小模块参数量。如表 3 所示, AFF 模块在仅增加 0.27 MB 参数量的情况下, 将 ATSS 的 mAP 提高了 1.8%。由于 AFF 模块基于 Head 进行多尺度特征融合, 与 NAS-FPN 等基于 Neck 的多尺度特征融合方法并不冲突, 两者可以搭配使用进一步增强检测性能, 实验数据如表 5 所示。

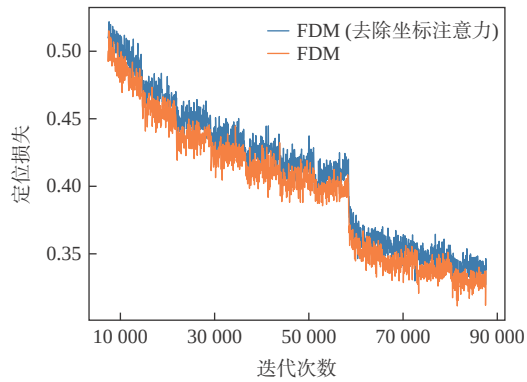


图 7　训练阶段定位损失的变化情况

Fig. 7　Changes in localization loss in training phase

表 5　AFF 与 NAS-FPN 的实验结果对比

Table 5　Comparison of experimental results between AFF and NAS-FPN

模型	mAP/%	参数量/MB	运算量
ATSS[26]	39.4	32.07	205.21×10^9
ATSS[26]/NAS-FPN[28]	40.9	38.89	200.54×10^9
ATSS[26]AFF(本文)	41.2	32.34	208.76×10^9
ATSS[26]/AFF(本文) & NAS-FPN[28]	42.0	39.16	204.00×10^9

2.5.3　权重因子归一化

Bi-FPN[12] 强调了对权重因子归一化的必要性, 本文 AFF 模块同样进行了相应的消融实验。表 6 中, 实验结果表明了权重因子归一化对 AFF 模块的必要性。

表 6　权重因子归一化实验结果对比

Table 6　Comparison of experimental results normalized by weight factor

模型	mAP/%	mAP50/%	mAP75/%
ATSS[26]	39.4	57.6	42.8
ATSS[26]/AFF*(本文)	40.9	58.7	44.5
ATSS[26]/AFF(本文)	41.2	59.3	45.0

注: "*" 表示未采用归一化。

2.5.4　X-Ray 目标检测数据集上的消融实验对比

本节的消融实验数据集来自于面向真实安检场景的违禁品检测挑战赛的水平检测赛道初赛, 该挑战赛于 2023 年由北京航空航天大学、科大讯飞及 OpenI 启智社区联合举办。在赛事过程中, 以 GFL 为基线模型并进行调优, 以达到相对较高的检测精度。AFF 模块和 FDM 提高了优化后基线模型 1% 的检测精度。实验数据如表 7 所示, 表中 mAP50 指标为比赛官网在线评估所得。

2.6　错误修正报告

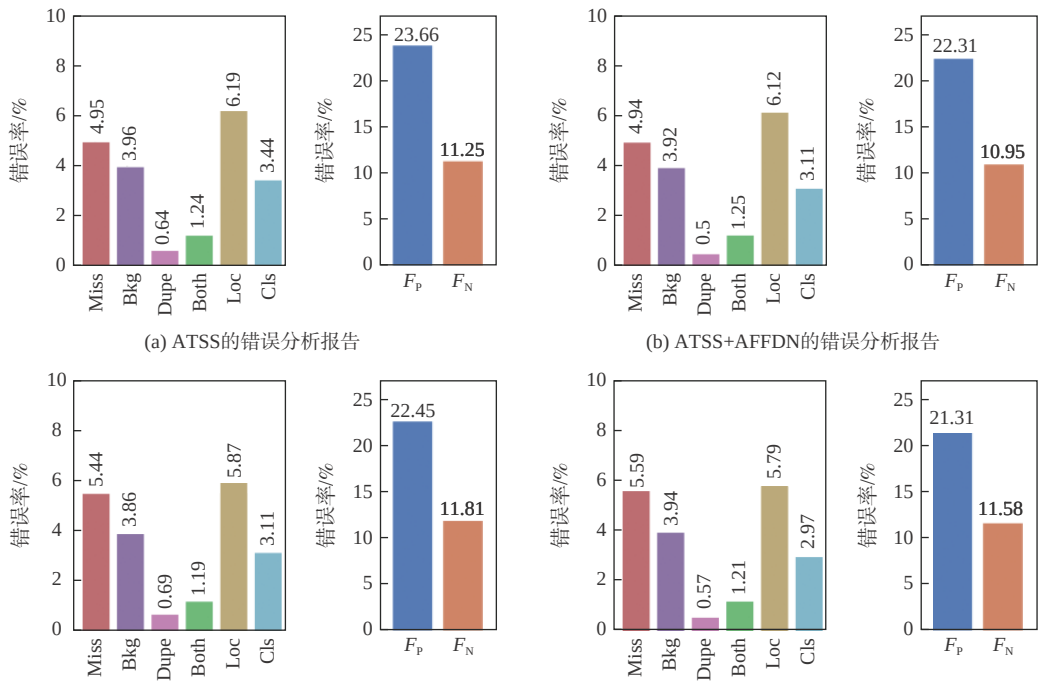
本文借助 TIDE[33] 对 AFFDN 模型进行错误分析。TIDE 将目标检测中的错误归纳为以下 6 类, 分别为分类错误(Cls)、定位错误(Loc)、分类与定

表7 X-Ray目标检测数据集上的消融实验结果

Table 7 Ablation experiment results on X-ray object detection dataset

GFL[27]	SwinL	12 epochs	Mixup, Flip and SoftNMS	AFF(本文)	FDM(本文)	mAP50/%	参数量/MB	运算量
√						68.83	32.22	208.39×10^9
√	√					83.43	203.6	872.59×10^9
√	√	√				87.94	203.6	872.59×10^9
√	√	√	√			89.01	203.6	872.59×10^9
√	√	√	√	√		89.81	203.9	876.14×10^9
√	√	√	√	√	√	90.01	204.0	857.32×10^9

位错误(Both)、重复框错误(Dupe)、背景错误(Bkg)和漏检错误(Miss)。另外，TIDE还提供了F_P和F_N的错误分析，其指标均为越低越好。如图8所示，通过观察AFFDN模型对以上错误的修复情况，可以发现，AFFDN模型有助于减少分类错误、定位错误、重复框错误、假阳性错误和假阴性错误。在分类错误和定位错误上，添加AFFDN模型的ATSS和GFL分类错误和定位错误合计分别降低了0.4%和0.22%，以及在假阳性错误和假阴性错误上，合计分别下降了1.65%和1.37%，这归因于FDM对特征的细粒度解耦，为各分支提供了更适合该分支任务的相关特征，以及AFF模块提供的更丰富的多尺度特征信息，促进了分类任务与定位任务的共同优化。在重复框错误上，分别下降了0.14%和0.12%，这是由于FDM中全局注意力对边界信息的提取，加强了模型对目标边界的感知。



(a) ATSS的错误分析报告 (b) ATSS+AFFDN的错误分析报告

(c) GFL的错误分析报告 (d) GFL+AFFDN的错误分析报告

图8 TIDE生成的错误报告

Fig. 8 Error report generated by TIDE

3 结 论

针对目标检测中FPN造成的大尺度目标检测精度下降和目标检测的2个子任务所需的语义特征不一致问题，本文提出一种新型的基于相邻特征融合与特征解耦的一阶段目标检测模型。

1）相邻特征融合模块基于Head实现多尺度特征融合，通过多对一方式连接FPN和Head，引入更短的梯度传播路径，缓解了引入FPN造成的大尺度目标梯度消失的现象。

2）通过共享参数和偏移量，有效减小相邻特征融合模块的参数量，同时通过注意力机制实现自适应特征融合提升多尺度特征表达能力。

3）特征解耦模块对输入特征从感受野、通道和空间3个维度进行特征解耦，为不同任务提取出了特有的任务相关的细粒度解耦特征，加速了模型训练的收敛，提高了模型检测精度。

4）在PASCAL VOC数据集和MS COCO2017数

据集上, AFFDN 模型可以将主流一阶段目标检测基线模型的检测精度分别至少提高 0.9% 和 2.3%, 并且能减少分类错误、定位错误、假阳性和假阴性等错误。

参考文献（References）

[1] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Piscataway: IEEE Press, 2015: 91-99.

[2] LU X, LI B Y, YUE Y X, et al. Grid R-CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 7355-7364.

[3] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6154-6162.

[4] 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.

[5] TIAN Z, SHEN C H, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 9626-9635.

[6] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-08-06)[2023-04-01]. https://arxiv.org/abs/2107.08430v2.

[7] 李晨瑄, 顾佼佼, 王磊, 等. 多尺度特征融合的 Anchor-Free 轻量化舰船要害部位检测算法[J]. 北京航空航天大学学报, 2022, 48(10): 2006-2019.
LI C X, GU J J, WANG L, et al. Warship's vital parts detection algorithm based on lightweight Anchor-Free network with multi-scale feature fusion[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(10): 2006-2019(in Chinese).

[8] 王榆锋, 李大海. 改进 YOLO 框架的血细胞检测算法[J]. 计算机工程与应用, 2022, 58(12): 191-198.
WANG Y F, LI D H. Improved YOLO framework blood cell detection algorithm[J]. Computer Engineering and Applications, 2022, 58(12): 191-198(in Chinese).

[9] 刘树东, 刘业辉, 孙叶美, 等. 基于倒置残差注意力的无人机航拍图像小目标检测[J]. 北京航空航天大学学报, 2023, 49(3): 514-524.
LIU S D, LIU Y H, SUN Y M, et al. Small object detection in UAV aerial images based on inverted residual attention[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(3): 514-524(in Chinese).

[10] 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.

[11] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8759-8768.

[12] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 10778-10787.

[13] JIANG Y Q, TAN Z Y, WANG J Y, et al. GiraffeDet: a heavy-neck paradigm for object detection[C]//International Conference on Learning Representations. Washington, D. C. : ICLR, 2022: 1-17.

[14] LUO Y H, CAO X, ZHANG J T, et al. CE-FPN: enhancing channel information for object detection[J]. Multimedia Tools and Applications, 2022, 81(21): 30685-30704.

[15] JIN Z C, YU D D, SONG L C, et al. You should look at all objects[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2022: 332-349.

[16] LI R, WANG L B, ZHANG C, et al. A^2-FPN for semantic segmentation of fine-resolution remotely sensed images[J]. International Journal of Remote Sensing, 2022, 43(3): 1131-1155.

[17] LIU S T, HUANG D, WANG Y H. Learning spatial fusion for single-shot object detection[EB/OL]. (2019-11-25)[2023-04-01]. https://arxiv.org/abs/1911.09516v2.

[18] LIU Y, HAN J G, ZHANG Q, et al. Deep salient object detection with contextual information guidance[J]. IEEE Transactions on Image Processing, 2019, 29: 360-374.

[19] MA J L, CHEN B. Dual refinement feature pyramid networks for object detection[EB/OL]. (2020-12-04)[2023-04-01]. https://arxiv.org/abs/2012.01733v2.

[20] WU Y, CHEN Y P, YUAN L, et al. Rethinking classification and localization for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 10183-10192.

[21] SONG G L, LIU Y, WANG X G. Revisiting the sibling head in object detector[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11560-11569.

[22] FENG C J, ZHONG Y J, GAO Y, et al. TOOD: task-aligned one-stage object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 3490-3499.

[23] CHEN Z H, YANG C, LI Q F, et al. Disentangle your dense object detector[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 4939-4948.

[24] DAI X Y, CHEN Y P, XIAO B, et al. Dynamic head: unifying object detection heads with attentions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 7369-7378.

[25] LIU Y, ZHANG D W, LIU N, et al. Disentangled capsule routing for fast part-object relational saliency[J]. IEEE Transactions on Image Processing, 2022, 31: 6719-6732.

[26] ZHANG S F, CHI C, YAO Y Q, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 9756-9765.

[27] LI X, WANG W H, WU L J, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection[J]. Advances in Neural Information Processing Systems,

2020, 33: 21002-21012.

[28] GHIASI G, LIN T Y, LE Q V. NAS-FPN: learning scalable feature pyramid architecture for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 7029-7038.

[29] LIU Z T, SHAO M W, SUN Y T, et al. Multi-task feature-aligned head in one-stage object detection[J]. Signal, Image and Video Processing, 2023, 17(4): 1345-1353.

[30] KIM K, LEE H S. Probabilistic anchor assignment with IoU prediction for object detection[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2020: 355-371.

[31] ZHANG H Y, WANG Y, DAYOUB F, et al. VarifocalNet: an IoU-aware dense object detector[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 8510-8519.

[32] LI S, HE C H, LI R H, et al. A dual weighting label assignment scheme for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 9377-9386.

[33] BOLYA D, FOLEY S, HAYS J, et al. TIDE: a general toolbox for identifying object detection errors[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2020: 558-573.

One-stage object detection based on adjacent feature fusion and feature decoupling

ZHENG Jian，HE Zhaohui，YU Xiangchun*

(School of Information Engineering，Jiangxi University of Science and Technology，Ganzhou 341000，China)

Abstract：In view of reduced large-scale object detection accuracy caused by the feature pyramid network (FPN) in object detection and the inconsistency of the semantic characteristics of the two sub-tasks of object detection, a new model based on adjacent feature fusion (AFF) and feature decoupling network (AFFDN) model was proposed. Firstly, the AFF module in the model introduced a shorter gradient return path by using the many-to-one connection, thereby alleviating the problem of large-scale object gradient disappearance. At the same time, AFF effectively reduced the amount of model parameters and enhanced the semantic consistency of multi-scale features by sharing parameters and offsets. In addition, compared with neural architecture search FPN (NAS-FPN), the parameters of AFF were smaller, and the performance gain was more significant. Secondly, the feature decoupling module (FDM) in the AFFDN used the dynamic receptive field and global attention to decouple fine-grained features in the three dimensions of receptive field, channel, and space, generating unique task-related features for different task branches and thereby improving the accuracy of object detection. Finally, when AFFDN was applied to different one-stage object detection models, the detection accuracy of the baseline model was improved by at least 0.9% and 2.3% on the PASCAL VOC dataset and MS COCO2017 dataset, respectively.

Keywords：object detection; multi-scale feature fusion; feature decoupling; attention; deformable convolution

Received：2023-05-15；**Accepted**：2023-12-11；**Published Online**：2023-12-27 18：57

URL：link.cnki.net/urlid/11.2625.V.20231227.1507.004

Foundation item：Jiangxi Provincial Natural Science Foundation (20224BAB212013)

* **Corresponding author**. E-mail：yuxc@jxust.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0234

基于 Trans-Attention 的飞行区航空器监视数据融合方法

王兴隆[1,*]，尹昊[1]，丁俊峰[2]
(1. 中国民航大学 飞联网重点实验室，天津 300300；2. 南京莱斯信息技术股份有限公司，南京 210001)

摘 要：针对飞行区航空器单一监视源存在监视精度低、位置跳变的问题，提出一种基于 Transformer 和注意力机制的航空器监视数据融合方法。利用 Transformer 的编码器结构分别对各监视源数据进行特征提取，通过注意力机制对不同监视源赋予权重值，经过全连接网络进行回归计算，获得最终的融合结果。选取场面监视雷达 (SMR) 和广播式自动相关监视 (ADS-B) 系统的监视数据作为融合源，多点定位 (MLAT) 数据作为真实标签，实验结果表明：所提方法有效降低了单一监视源的监视误差，且融合效果优于基于注意力机制的长短期记忆网络、循环神经网络和扩展卡尔曼滤波融合方法，平均绝对误差分别提升了 2.81%、16.73% 和 35.80%。

关 键 词：数据融合；Transformer；注意力机制；场面监视雷达；广播式自动相关监视
中图分类号：V351
文献标志码：A **文章编号**：1001-5965(2025)04-1215-09

近年来，民用航空业迅速发展，机场飞行区规模不断扩大，随之也带来一些问题亟须解决。首先，航班量的持续增加影响了机场飞行区的运行效率；其次，在恶劣天气和障碍物遮挡情况下，无法准确、实时地获得航空器的运行状态，从而使得机场飞行区的安全可靠性很难得到有效的保障。

为实现精准监控，保证飞行区的安全高效运行，机场利用场面监视雷达（surface movement radar, SMR）、广播式自动相关监视（automatic dependent surveillance-broadcast, ADS-B）、多点定位（multilateration, MLAT）等监视传感器来获取航空器的信息。20 世纪 60 年代，国外一些大型机场开始将 SMR 用于机场地面监控[1]；国内机场为满足实际需求，也对 SMR 系统开展了一系列研究，并将其应用于各大机场[2]；但其可能因暴雨等天气原因导致目标航迹定位精度变低。ADS-B 通过地面站、机载设备及全球导航卫星系统 (global navigation satellite system, GNSS) 实现对航空器等目标的监视，可作为场面监视的一种补充数据源，为机场管制指挥提供有效的技术支撑手段[3]；但其对 GNSS 的完好性依赖度极高，且无法对目标所处位置加以验证，因此，存在误报、位置可信度较低的问题。MLAT 技术作为当前最先进的场面监视手段及高级场面活动引导与控制系统（advanced surface movement guidance and control system, A-SMGCS）中场面监视系统的核心之一，在国内外多个国家的航空业、军事等领域得到重点研究和应用[4]。戴敏[5]提出一种结合 Tikhonov 正则化的高精度 MLAT 方法，实现高精度的几何精度因子（geometric dilution precision, GDOP）覆盖，对到达时间差（time difference of arrival, TDOA）测量误差具有很好的鲁棒性。

多传感器数据融合技术可以理解为对来自多种监视手段的数据源进行组合优化，经过处理后得到目标的状态信息，融合后得到比单一监视源更高

收稿日期：2023-05-09；**录用日期**：2023-07-14；**网络出版时间**：2023-08-30 10:35
网络出版地址：link.cnki.net/urlid/11.2625.V.20230829.1724.003
基金项目：国家自然科学基金面上项目 (62173332)；国家自然科学基金重点项目 (U2133207)；天津多元基金项目 (21JCYBJCO0700)；中国民航大学校级研究生科研创新项目 (2022YJS086)
* **通信作者**. E-mail：xinglong1979@163.com

引用格式：王兴隆，尹昊，丁俊峰. 基于 Trans-Attention 的飞行区航空器监视数据融合方法 [J]. 北京航空航天大学学报，2025，51（4）：1215-1223.
WANG X L，YIN H，DING J F. Aircraft surveillance data fusion method in flight area based on Trans-Attention[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1215-1223（in Chinese）.

精度的目标状态。传统的数据融合方法有卡尔曼滤波算法、加权法、贝叶斯估计法、最小二乘估计法等。Akram等[6]使用统计估计理论提出一种基于卡尔曼滤波的多传感器数据融合模型,实验结果表明,该模型对目标状态估计时具有较高的精度和鲁棒性。张学军和张其善[7]使用卡尔曼滤波算法将ADS和二次监视雷达(secondary surveillance radar, SSR)数据进行融合,克服了SSR的缺陷,实验结果表明,该算法可提升航空器监视系统的精度和可靠性。Xia等[8]提出一种加权法,用以融合基于速度和卡尔曼状态观测器的误差状态估计,实验证实了该方法不仅可以提升目标状态估计精度,还可以利用来自其他传感器的观测信息,进一步提升可靠性和准确性。刘康等[9]考虑了期望一致性及传感器局部融合精度,提出一种基于多传感器融合的综合权值估计算法,仿真结果表明,该算法提升了融合结果的准确性。李素[10]针对分布式多传感器航迹融合的特点,对局部航迹进行评估,确定各局部航迹的加权因子,采用加权平均方法进行航迹融合,该方法在传感器观测精度相差较大的情况下具有较好的融合性能,提高了系统航迹的精度。Rajasekaran等[11]针对单传感器定位时信号遮挡及杂波干扰的问题,提出一种贝叶斯融合目标定位算法,实验表明,该算法可以有效提高目标定位的鲁棒性和可靠性。郭建刚等[12]针对多表冗余惯导系统的自对准问题,使用最小二乘估计法对惯导仪表数据进行融合,实验证明了该算法在提高系统自对准精度中的有效性。但卡尔曼滤波算法对实际状态转换过程、参数的准确性要求很高且调参过程复杂;加权法只适用于处理线性系统的状态估计,适用范围较为有限;贝叶斯估计法存在难以获取先验知识的问题;最小二乘估计法在拟合非线性数据问题中效果一般。

随着科学技术的不断革新,许多现代数据融合方法开始流行,包括D-S证据理论法、扩展卡尔曼滤波(extended Kalman filter, EKF)法、神经网络法等。周恩帆等[13]针对多传感器网络中采集数据存在的不确定性,提出一种基于D-S证据理论的多传感器数据融合方法。Kaczmarek等[14]提出一种EKF和动态权重矩阵方法,将GNSS、惯性测量单元(inertial measurement unit, IMU)和里程表数据进行融合,结果表明,该方法可以获得目标更精准的位置和方位角信息。宋安宇[15]利用递归神经网络方法处理分类问题和回归问题,判别飞机的机动状态,并进行多雷达监视信息融合,使得融合模型能够及时识别机动状态,达到更精准的融合效果。孟致远[16]提出基于优化AdaBoost的融合方法,将多雷达的监视数据进行融合,融合效果优于单个雷达、BP神经网络融合结果,提升了整体的融合精度。同样,D-S证据理论法不能有效处理矛盾的证据,且具有幂指数增长的计算量的问题;BP神经网络存在过拟合,循环神经网络(recurrent neural network, RNN)模型训练存在梯度消失、长期依赖的问题。

本文提出一种基于Transformer和注意力机制的深度神经网络方法,对多传感器监视数据进行融合。该方法将不同监视源的航迹数据输入Encode层独立训练,利用注意力机制将不同Encode层的输出结果进行融合,充分捕捉不同传感器数据的特征关系,突破了RNN模型不能并行计算、梯度消失及长期依赖的限制。实验结果表明,本文方法有效提升了监视目标位置的准确性和可靠性。

1 飞行区关键监视技术及数据预处理

1.1 飞行区监视技术

本文涉及SMR、ADS-B及MLAT这3种飞行区监视技术,本节分别介绍各自的基本原理和特点。

SMR为一次雷达,通过辐射电磁波接收场面物体的回波反射信号,实现信号覆盖范围内所有关注目标的检测和跟踪[17]。SMR是一种非合作目标监视技术,无需目标装载应答设备,可实时监视飞行区内运动或静止的航空器、车辆、障碍物等目标。SMR报告目标位置的更新周期为1 s,整体具备一致性,可信度较高;但可能因恶劣天气或场面施工造成杂波干扰、回波遮挡、回波多径效应等,继而产生虚假目标或造成正常目标航迹位置发生偏移、中断等现象。此外,SMR安装要求比较高,同时后期维护成本较高。

ADS-B通过飞行区内目标装载并开启的ADS-B广播设备向外发射信号,由ADS-B基站接收信息并生成目标航迹信号,以此来对目标进行定位。ADS-B是一种合作监视技术手段,若未装载或未开启ADS-B机载设备的目标,无法被检测。其数据更新频率高于SMR和MLAT,通常为0.5～1 s;但报告位置信息为ADS-B设备基于GPS定位自主报告的位置,位置可信度低于SMR和MLAT,对于部分设备故障或未校准的目标,报告位置可能会产生较大的偏差。

MLAT通过精确计算目标向3个或3个以上接收基站发出信号的TDOA来定位目标[18],从而实现对飞行区运动或静止目标进行监视。MLAT是一种合作目标监视技术手段,在飞行区监视方面有很大的优势,相较于SMR有更高的定位精度,数据更新率快,相较于ADS-B有更高的可信度。若通过适当的布点,在理论上可覆盖任何区域,且不受天

气影响; 其安装选址要求低于 SMR、安装方便及维护成本低; 此外, 中国北京大兴国际机场安装了全新一代多点监视系统(multilateration surveillance system-airport, NEO-A), 作为机场场面航空器、车辆等目标的主要监视设备[19]。

基于此, 本文将选用 MLAT 的监视数据作为融合模型的真实标签数据。

1.2　数据预处理

监视系统接收到各监视传感器输出的原始报文数据, 这样的数据无法直接用于融合方法, 需要进行一系列预处理。数据预处理通常包括 4 个方面: 数据解析、空间对准、航迹关联及时间对齐。

1) 数据解析。将原始报文的数据部分按字节数组的形式读入, 根据国际标准化组织定义的 ASTERIX 数据信息格式标准进行解析。其中, 类型为 CAT010 的报文报告的数据为场面目标报和服务信息, 包含 SMR 和 MLAT 系统; 类型为 CAT021 的报文用于传输 ADS-B 监视数据。解析流程参考了 ASTERIX官方的协议文档及牛中伟和陈健明对报文的解析方式[20]。

2) 空间对准。解析后得到的 SMR 和 MLAT 数据中的位置信息在笛卡儿坐标系下, 而 ADS-B 数据中的位置信息在 WGS-84 大地坐标系下, 并不在同一坐标系, 这样的数据无法作为融合模型的数据集, 因此, 本文使用高斯克吕格投影方法将这 3 种数据统一到机场平面直角坐标系下。设平面直角坐标系的坐标为(x,y), 则对应投影的计算式为

$$x = X_0 + \frac{1}{2}Ntm_0^2 + \frac{1}{24}(5 - t^2 + 9\eta^2 + 4\eta^4)Ntm_0^4 + \frac{1}{720}(61 - 58t^2 + t^4)Ntm_0^4 \tag{1}$$

$$y = Nm_0 + \frac{1}{6}(1 - t^2 + \eta^2)Ntm_0^3 + \frac{1}{120}(5 - 18t^2 + t^4 + 14\eta^2 - 58\eta^2 t^2)Nm_0^6 \tag{2}$$

$$\eta^2 = \frac{e^2}{1 - e^2}\cos^2 B \tag{3}$$

$$m_0 = l\cos B \tag{4}$$

式中: X_0为监视目标对应纬度的子午线弧长; N为子午圈的曲率半径; t为对应纬度的正切值, 即$t = \tan B$, B为监视目标的纬度; l为监视目标对应经度和中央子午线经度的差; e为椭球的第一偏心率。

3) 航迹关联。经过数据解析和空间对准后, SMR、MLAT 和 ADS-B 报告的目标位置都是离散的坐标点, 航迹关联的目的是: ①区分出属于同一条航迹的位置坐标点, 形成一条完整的航迹; ②将不同监视数据源中同一目标的航迹进行关联。

首先, 处理来自同一数据源的监视数据。对于 SMR 数据, 可以通过解析得到的航迹编号这一字段来筛选出属于同一目标的航迹; 对于 ADS-B 或 MLAT 数据, 可以通过解析得到的航班号这一字段来筛选出属于同一目标的航迹。

然后, 关联来自不同数据源的监视数据。对于 ADS-B 和 MLAT 这 2 种数据源的关联, 其报文都报告航班号信息, 因此, 直接通过航班号这一字段来筛选出属于同一条目标的航迹; 对于 SMR 与 MLAT 或 ADS-B 的相关, SMR 报文无法报告航班号信息, 因此, 不能直接进行关联。本文采用最近邻域(nearest neighbor, NN) 法进行关联, 具体步骤如下:

步骤 1　将同一时间段内的 SMR、MLAT、ADS-B 数据以航迹为单位进行分类。

步骤 2　相同时刻下, 计算每条 SMR 形成的航迹点与 MLAT 或 ADS-B 中每条航迹点距离的平均值。

步骤 3　设定一个阈值, 将距离平均值低于阈值的航迹进行匹配。

步骤 4　按照步骤 3 匹配得到的航迹, 优先选择距离平均值最小的进行关联, 将 2 条航迹进行可视化, 判断其是否为同一条航迹; 若不是, 则选择距离平均值较大的进行关联, 直到找到同一航迹或大于阈值就结束。

4) 时间对齐。由于各个监视源都独立工作, 且监视数据更新频率不同, 输出监视数据的时间戳不同, 在进行数据融合前, 要进行时间对齐。因为 ADS-B 的更新频率大于 SMR 和 MLAT, 所以, 将一条航迹的 SMR 和 MLAT 数据向同一航迹的 ADS-B 数据进行时间上的对齐。以 SMR 向 ADS-B 对齐为例, 采用线性插值法, 时间对齐示意图如图 1 所示。图中, $A(t)$和$S(t)$分别表示 ADS-B 和 SMR 的数据, 可以看到, 在t_1～t_8时间段内, ADS-B 输出了 6 个航迹点, SMR 输出了 4 个航迹点, 音系将 SMR 的数据向 ADS-B 数据对齐, 图中用虚线圆圈表示对齐后的 SMR 数据。

通过该方式, 使得每条航迹中 SMR、MLAT 和 ADS-B 的数据在时间上统一。经过上述 4 个步骤

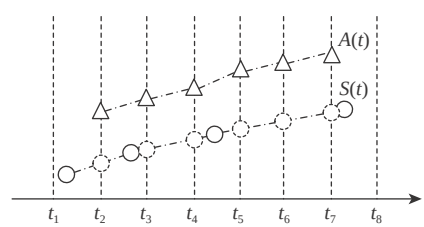


图 1　时间对齐示意图

Fig. 1　Time alignment diagram

处理得到的数据集可以作为后续融合模型的数据集。

2 数据融合模型

机场飞行区航空器各监视传感器的输出信息包括数据源识别码、时间、坐标、速度、航向等。定义$\boldsymbol{X}_t^i$为航空器t时刻航迹特征的六元组数据，表示为

$$\boldsymbol{X}_t^i=(x,y,v_x,v_y,v,\theta) \tag{5}$$

式中：i为监视传感器类型，$\boldsymbol{X}_t^{s1}$为1号SMR数据，$\boldsymbol{X}_t^{s2}$为2号SMR数据，$\boldsymbol{X}_t^{m}$为MLAT数据，$\boldsymbol{X}_t^{a}$为ADS-B数据；x、y、v_x、v_y、v、θ分别为t时刻航空器的横坐标、纵坐标、x方向速度、y方向速度、合成速度、航向。

本文提出基于Trans-Attention的多传感器数据融合模型，其框架结构如图2所示，包含输入层、Encoder层、融合层、Decoder层和输出层。

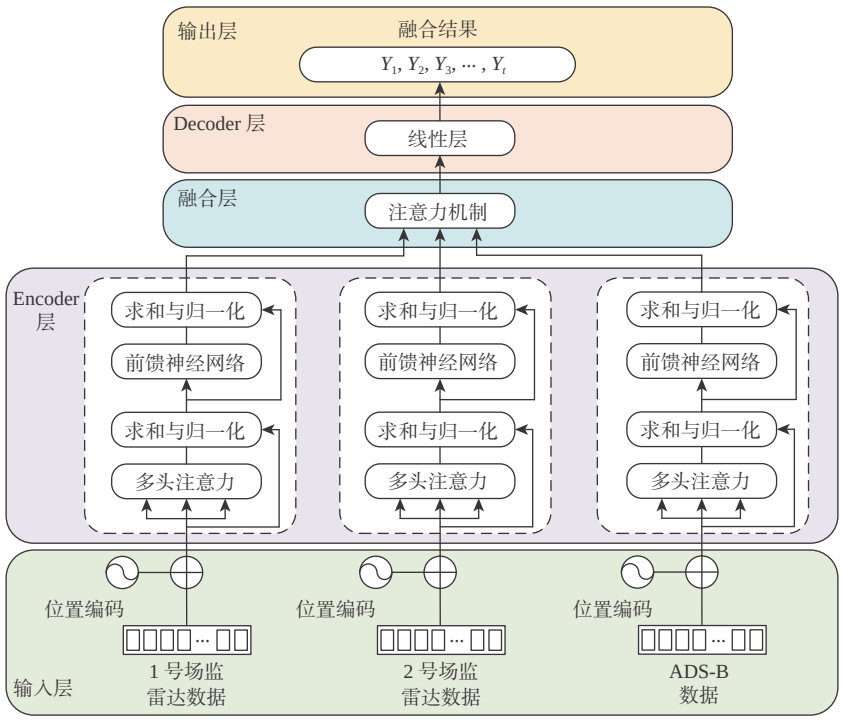


图2 Trans-Attention融合模型结构

Fig. 2 Trans-Attention fusion model structure

2.1 输入层

输入层由数据归一化、传感器数据输入和位置编码组成。

1）数据归一化。传感器数据特征之间差距较大，如果直接采用原始数据输入，会容易导致不平衡，因此，分别对各个传感器数据进行归一化操作。本文选择Max-Min归一化方法对数据做一次线性变换，将其值映射到[0,1]之间，计算式为

$$\boldsymbol{X}_{\text{norm}}=\frac{\boldsymbol{X}-\boldsymbol{X}_{\min}}{\boldsymbol{X}_{\max}-\boldsymbol{X}_{\min}} \tag{6}$$

式中：$\boldsymbol{X}_{\text{norm}}$为归一化数据；$\boldsymbol{X}_{\max}$和$\boldsymbol{X}_{\min}$分别为传感器数据的最大值和最小值。

2）传感器数据输入。将经过归一化处理完成的SMR、MLAT、ADS-B数据构建成数据集，其中，2部SMR和ADS-B数据作为模型融合数据源，MLAT数据作为标签，以连续n个时刻的数据为例，将3个传感器的航迹特征$\{\boldsymbol{X}_t^{s1},\boldsymbol{X}_{t+1}^{s1},\cdots,\boldsymbol{X}_{t+n-1}^{s1}\}$、$\{\boldsymbol{X}_t^{s2},\boldsymbol{X}_{t+1}^{s2},\cdots,\boldsymbol{X}_{t+n-1}^{s2}\}$、$\{\boldsymbol{X}_t^{a},\boldsymbol{X}_{t+1}^{a},\cdots,\boldsymbol{X}_{t+n-1}^{a}\}$作为融合模型的输入。

3）位置编码。在训练前，需要进行位置编码，由于Transformer抛弃了RNN的结构，模型丢失了序列中各个时刻的位置信息，本文采用Vaswani等[21]研究的位置编码方法。其原理是：将生成的不同频率的正弦和余弦数据作为位置编码添加到输入序列中，从而使得模型可以捕捉输入变量之间的位置关系，计算式为

$$P_{(k,2i)}=\sin(k/10\,000^{2i/d}) \tag{7}$$

$$P_{(k,2i+1)}=\cos(k/10\,000^{2i/d}) \tag{8}$$

式中：P为位置编码；k为当前数据在输入序列中的位置；d为位置编码的特征向量的维度；i为数据的第i个的维度。编码向量中，对奇数位置使用余弦函数，偶数位置使用正弦函数。

2.2 Encoder层

Encoder层由注意力机制、求和与归一化和前馈神经网络组成。

1) 注意力机制。注意力机制将查询向量$\boldsymbol{Q}$和一组键值对$\boldsymbol{K}$-$\boldsymbol{V}$映射到输出, 输出的值加权求和, 权重通过$\boldsymbol{Q}$和$\boldsymbol{K}$的相似度计算得到。本文模型中的注意力机制由缩放点积注意力机制和多头注意力机制构成, 通过加入注意力机制, 可以对输入的航迹序列特征进行提取和表达。缩放点积注意力机制的计算过程是: $\boldsymbol{Q}$和$\boldsymbol{K}$向量通过点积方式计算相似度, 通过比例因子$\sqrt{d_k}$处理避免$\boldsymbol{Q}$、$\boldsymbol{K}$内积方差太大导致难以学习的情况, 使用 softmax 函数获取每个$\boldsymbol{K}$的权重, 将权重和其对应的值$\boldsymbol{V}$加权求和输出。计算式为

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\cdot\boldsymbol{V} \tag{9}$$

式中: Attention 为注意力机制计算函数; softmax 为激活函数; d_k为$\boldsymbol{Q}$和$\boldsymbol{K}$的维度。

多头注意力机制通过对$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$分别线性投影到缩放点积注意力机制中, 投影h次后做h次注意力函数运算, 通过并行计算, 将每个输出值拼接后再做一次线性投影得到最终输出。多头注意力机制结构如图 3 所示, 计算式为

$$\boldsymbol{H}_i = \text{Attention}(\boldsymbol{Q}\boldsymbol{W}_i^{\boldsymbol{Q}},\boldsymbol{K}\boldsymbol{W}_i^{\boldsymbol{K}},\boldsymbol{V}\boldsymbol{W}_i^{\boldsymbol{V}}) \qquad i = 1,2,\cdots,h \tag{10}$$

$$\boldsymbol{M}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = C(\boldsymbol{H}_1,\boldsymbol{H}_2,\cdots,\boldsymbol{H}_h)\boldsymbol{W}_{\mathrm{o}} \tag{11}$$

式中: $\boldsymbol{W}_i^{\boldsymbol{Q}}$、$\boldsymbol{W}_i^{\boldsymbol{K}}$、$\boldsymbol{W}_i^{\boldsymbol{V}}$为第$i$个自注意力机制的映射矩阵; $\boldsymbol{H}_i$为第i个自注意力的输出; $\boldsymbol{M}$为多头注意力; C为将多个自注意力头的结果进行拼接的函数; $\boldsymbol{W}_{\mathrm{o}}$为拼接之后的线性映射矩阵。

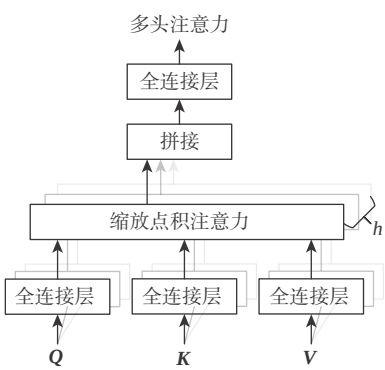


图 3　多头注意力机制结构

Fig. 3　Multi-head attention mechanism structure

2) 求和与归一化。求和就是进行残差连接, 其主要思想是: 通过跨层连接的方式将输入信息加入多层网络后的输出中。一方面, 避免了因注意力层数过多而导致原始数据信息丢失的问题; 另一方面, 当注意力层数增多时, 反向传播经过的层数也会增多, 可能会引起梯度消失的情况, 跨层连接可以保证能迅速达到输入层。

在经过残差连接后, 需要对输出的数据进行层归一化操作, 即将隐藏层的输出归一化为标准正态分布, 这样可以加快模型训练速度, 加速收敛。计算式为

$$\text{LayerNorm}(h) = \alpha\frac{h-\mu_h}{\sqrt{\sigma_h^2+\varepsilon}} + \beta \tag{12}$$

式中: LayerNorm(h)为层归一化的输出; α、β为引入的 2 个参数, 为了弥补归一化操作损失的信息; μ_h和σ_h分别为h的均值和标准差; ε为很小的数, 为了防止出现标准差为 0 的情况。

3) 前馈神经网络。前馈神经网络由 2 层全连接层构成, 第 1 层的激活函数为 ReLU, 第 2 层不使用激活函数。其相当于将每个位置的注意力结果映射到一个更高维的空间, 进一步提取信息。计算式为

$$\text{FFN}(x) = \max(0, x\boldsymbol{W}_1 + b_1)\boldsymbol{W}_2 + b_2 \tag{13}$$

式中: W和b分别为权重矩阵和偏置。

2.3　融 合 层

通过 Encoder 层处理后得到 3 个尺寸大小相等的特征向量$\boldsymbol{X}^{\mathrm{s1}'}$、$\boldsymbol{X}^{\mathrm{s2}'}$、$\boldsymbol{X}^{\mathrm{a}'}$分别表示 3 个传感器的隐藏层向量特征, 即 Encoder 层的输出, 利用注意力机制为其分配权重参数α_{s1}、α_{s2}、α_{a}, 进行自适应加权融合, 最终得到融合后的特征$\boldsymbol{X}'$。计算式为

$$\boldsymbol{X}' = \alpha_{\mathrm{s1}}\boldsymbol{X}^{\mathrm{s1}'} + \alpha_{\mathrm{s2}}\boldsymbol{X}^{\mathrm{s2}'} + \alpha_{\mathrm{a}}\boldsymbol{X}^{\mathrm{a}'} \tag{14}$$

权重参数满足式(15)的约束条件:

$$\alpha_{\mathrm{s1}} + \alpha_{\mathrm{s2}} + \alpha_{\mathrm{a}} = 1 \qquad \alpha_{\mathrm{s1}},\alpha_{\mathrm{s2}},\alpha_{\mathrm{a}} \in [0,1] \tag{15}$$

2.4　Decoder 层

本文 Decoder 层使用全连接网络代替传统 Transformer 的 Decoder 层, 隐藏层的激活函数使用 ReLU 函数。选用全连接网络的优势为: ①传统 Transformer 的 Decoder 层基于自回归逐步生成预测结果, 无法控制输出长度, 使用全连接网络可以指定输出维度, 更利于控制预测的时间步长度; ②传统 Transformer 的 Decoder 层在每个时间步输出预测结果时依赖之前时间步的预测结果, 当之前时间步预测误差较大时, 会严重影响当前时间步的预测精度, 导致若干个时间步后累计误差过大; ③基于自回归的生成方式具有更慢的速度, 影响预测结果的时效性, 而使用全连接网络可以并行生成预测结果, 具有更快的运行效率。

2.5　输 出 层

本文模型最终的输出为航空器的位置坐标, 因此, 输出层有 2 个神经元, 且输出层使用线性激活函数。

3 实验与结果分析

本文提出的Trans-Attention融合模型使用Pytorch框架搭建，版本为1.8.0，Python版本为3.8.13。

3.1 实验设计

1）数据集选取。本文数据集主要包括2部场监雷达、一个MLAT系统及1个ADS-B数据的机场场面监视数据，共计54条不同的航迹数据、118 792个点迹，每条航迹数据都包含4个传感器的监视信息。选用其中50条航迹作为模型的训练数据，剩下的4条航迹作为模型的测试数据。

2）评价指标。本文任务本质上是回归问题，因此，选用均方误差(mean square error, MSE)、均方根误差(root mean squared error, RMSE)、平均绝对误差(mean absolute error, MAE)作为评价指标。3个评价指标值e_{MSE}、e_{RMSE}、e_{MAE}的计算式分别为

$$e_{\mathrm{MSE}}=\frac{1}{m}\sum_{i=1}^{m}(f_i-t_i)^2 \tag{16}$$

$$e_{\mathrm{RMSE}}=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(f_i-t_i)^2} \tag{17}$$

$$e_{\mathrm{MAE}}=\frac{1}{m}\sum_{i=1}^{m}|f_i-t_i| \tag{18}$$

式中：m为样本数量；f_i和t_i分别为第i个轨迹点的真实值和融合值。

3）模型训练。本文方法输入2个SMR、1个ADS-B的航迹特征的六元组数据，最终输出融合后的航空器横纵坐标。具体训练流程如下：

步骤1 设定滑动窗口的大小，并将50条用于训练的航迹数据和4条用于测试的航迹数据进行切分。

步骤2 设定网络的结构参数，包括Encode层深度、多头注意力机制中头的个数、网络训练的轮次等，并根据所选参数对网络中隐藏层权重及偏置矩阵进行初始化，将隐藏层的偏置矩阵初始化为**0**矩阵，随机生成满足方差为0.01且均值为0的数，作为隐藏层权重的初始化值。

步骤3 设定训练时所用的超参数，包括网络学习率、隐藏层丢失率等，并开始对一个epoch的训练过程：①从划分好的训练集中取出一个滑动窗口的数据；②对训练数据进行位置编码和归一化操作；③网络进行前向传播，分别将经过归一化操作后的3个传感器数据输入模型的Encoder层进行学习，同时获取到3个Encoder层的学习结果，将其输入融合层进行融合并输出，将融合层的输出输入Decoder层中进行解码；④网络进行反向传播，预测值与真实值之间的损失通过均方误差损失函数计算，并按照Decoder层、融合层、Encoder层的顺序传播，且内部权重的更新通过Adam优化器实现；⑤重复①～④，直到训练完训练集中的所有数据。

步骤4 重复步骤3的训练过程，训练结束的标志是达到设定训练轮次的最大值，保存训练好的模型。

4）模型参数。模型参数的设定对实验结果的好坏起着不可忽视的作用，具体的参数设置如表1所示。

表1 Trans-Attention融合模型参数设置

Table 1 Trans-Attention fusion model parameter settings

学习率	隐藏层丢失率	训练轮次	多头注意力个数	Encode层深度	批量大小
0.001	0.1	600	6	2 048	8

3.2 结果分析

1）Trans-Attention融合方法结果。由于展示整条航迹的数据过长，本文选取航空器直线滑行和转弯滑行时部分航迹进行融合效果展示，分别如图4和图5所示。可知，基于Trans-Attention的融合方法输出的航迹曲线比单一监视传感器的航迹曲线更接近真实标签数据。

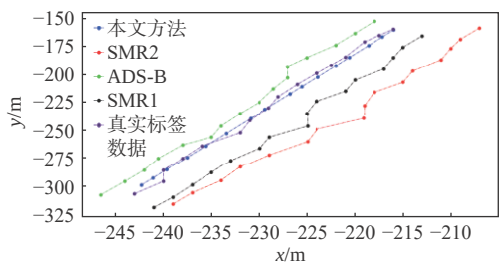


图4 航空器直线滑行时模型融合结果

Fig. 4 Model fusion results during aircraft straight-line taxiing

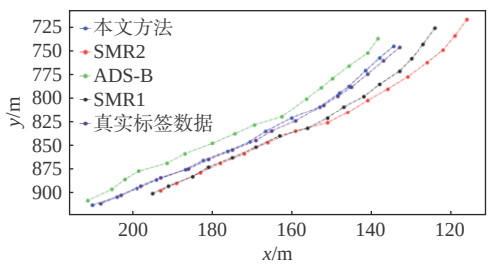


图5 航空器转弯滑行时模型融合结果

Fig. 5 Model fusion results during aircraft turning and taxiing

在3个评价指标上对融合效果进行评估，如表2所示。根据表2，以MAE指标为例，基于Trans-Attention的监视数据融合方法相较于融合前ADS-B、SMR1、SMR2这3种单一监视数据，精度有了明显

表 2　基于 Trans-Attention 融合方法结果数据评估

Table 2　Data evaluation based on Trans-Attention fusion method results

数据来源	MSE/m	RMSE/m	MAE/m
Trans-Attention	4.281 1	2.069 1	2.059 5
ADS-B	14.497 1	3.807 5	4.347 9
SMR1	36.314 5	6.026 2	8.276 9
SMR2	47.447 6	6.888 2	10.657 9

提升, 分别提升了约 52.63%、75.12%、80.68%。

2) 不同融合方法结果对比。为进一步验证本文方法的效果, 与 EKF、RNN 及基于注意力机制的长短期记忆网络 (long short-term memory based on attention mechanism, LSTM-Attention) 融合方法进行对比实验。EKF、RNN 及 LSTM-Attention 融合方法的参数设置如表 3～表 5 所示。

表 3　EKF 融合方法参数设置

Table 3　EKF fusion method parameter settings

参数类型	参数值
初始化状态协方差矩阵 $\boldsymbol{P}_0$	$\begin{bmatrix} 1.0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 \\ 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$
过程噪声协方差 $\boldsymbol{Q}_0$	$\begin{bmatrix} 0.000\,1 & 0 & 0 & 0 & 0 \\ 0 & 0.000\,1 & 0 & 0 & 0 \\ 0 & 0 & 0.000\,09 & 0 & 0 \\ 0 & 0 & 0 & 0.001 & 0 \\ 0 & 0 & 0 & 0 & 0.001 \end{bmatrix}$
SMR1的测量协方差矩阵 $\boldsymbol{R}_1$	$\begin{bmatrix} 6.25 & 0 & 0 \\ 0 & 0.001\,6 & 0 \\ 0 & 0 & 6.25 \end{bmatrix}$
SMR2的测量协方差矩阵 $\boldsymbol{R}_2$	$\begin{bmatrix} 4.622\,5 & 0 & 0 \\ 0 & 0.000\,9 & 0 \\ 0 & 0 & 4.625\,5 \end{bmatrix}$
ADS-B的测量协方差矩阵 $\boldsymbol{R}_3$	$\begin{bmatrix} 2.25 & 0 \\ 0 & 2.25 \end{bmatrix}$

表 4　RNN 融合方法参数设置

Table 4　RNN fusion method parameter settings

学习率	隐藏层丢失率	训练轮次	批量大小	RNN层数
0.001	0.1	300	8	2

表 5　LSTM-Attention 融合方法参数设置

Table 5　LSTM-Attention fusion method parameter settings

学习率	隐藏层丢失率	训练轮次	批量大小	LSTM层数
0.001	0.1	300	8	2

图 6 选取了测试数据中的部分航迹, 将 4 种不同融合方法的融合结果进行比较。可知, 本文方法的结果比 EKF、RNN 及 LSTM-Attention 的结果更准确, 曲线更接近真实标签数据。

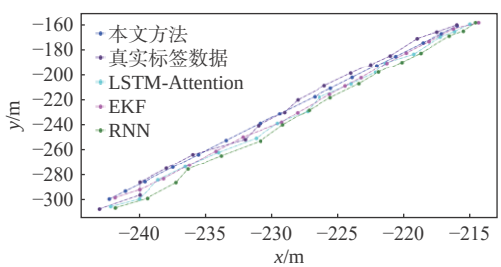


图 6　不同融合方法结果对比

Fig. 6　Comparison of results of different fusion methods

在 3 个评价指标上对融合效果进行评估, 如表 6 所示。

通过对比实验可以发现, EKF 融合方法在 3 个评价指标上表现最差, 主要是由于其主要思想是将非线性函数进行近似线性化处理, 对于本文研究高度非线性系统的线性化效果一般; RNN 融合方法在面对长时间序列数据时会存在梯度消失、长期依赖问题; LSTM-Attention 融合方法能有效解决 RNN 的问题, 且与本文方法都具有强大的长时间序列数据特征提取能力, 但其训练方式迭代, 计算效率低于本文方法, 因此, 本文方法融合的效果最优。

同时, 以 MAE 评价指标为例, 通过计算得出, 本文方法相较于 LSTM-Attention、RNN、EKF 融合方法, 分别提高了 2.81%、16.73%、35.80%。

表 6　不同融合方法结果数据评估

Table 6　Data evaluation form for results of different fusion methods

方法	MSE/m	RMSE/m	MAE/m
本文方法	4.281 1	2.069 1	2.059 5
EKF	9.974 5	3.158 2	3.207 8
LSTM-Attention	4.377 5	2.092 2	2.119 0
RNN	4.969 7	2.229 3	2.473 2

4 结 论

本文基于机场飞行区航空器滑行轨迹数据，采用深度学习方法构建了多传感器监视数据融合模型，结论如下：

1）提出一种基于Trans-Attention的航空器监视数据融合方法，综合考虑各监视源数据的重要度和可信度，弥补了单个监视源监视数据跳变、缺失的问题。

2）本文方法利用Transformer对长时间序列数据敏感、强大的特征提取和并行计算能力，加入注意力机制融合不同监视源数据。与EKF、RNN及LSTM-Attention这3种融合方法相比，融合精度更高。

3）融合输出后的航空器位置信息更加精准、稳定、可靠，帮助管制员更好地监控飞行区内航空器状态，从而保障飞行区运行安全、高效。

参考文献（References）

[1] LANZKRON P, BROOKNER E. Solid state X-band airport surface surveillance radar[C]//Proceedings of the IEEE Radar Conference. Piscataway: IEEE Press, 2007: 670-676.

[2] 卿烈华. 场面监视雷达系统在民用机场的应用[J]. 数字技术与应用, 2018, 36(5): 100-102.
QING L H. Application of scene surveillance radar system in civil airports[J]. Digital Technology and Application, 2018, 36(5): 100-102 (in Chinese).

[3] 王思达, 段昌森. 基于ADS-B的场面监视技术探讨[J]. 空运商务, 2018(6): 59-61.
WANG S D, DUAN C M. Discussion on scene monitoring technology based on ADS-B[J]. Air Transport & Business, 2018(6): 59-61 (in Chinese).

[4] VASYLIEV V M, VASYLIEV D V, NAUMENKO K V. Using data of multilateration surveillance system for aircraft tracking[C]// Proceedings of the 4th International Conference on Methods and Systems of Navigation and Motion Control. Piscataway: IEEE Press, 2016: 279-283.

[5] 戴敏. 结合正则化求解的机场场面多点定位方法[J]. 计算机工程与设计, 2021, 42(10): 2995-3001.
DAI M. Multilateration method for airport surface surveillance based on regularization[J]. Computer Engineering and Design, 2021, 42(10): 2995-3001(in Chinese).

[6] AKRAM M A, LIU P L, TAHIR M O, et al. A state optimization model based on Kalman filtering and robust estimation theory for fusion of multi-source information in highly non-linear systems[J]. Sensors, 2019, 19(7): 1687.

[7] 张学军, 张其善. ATM中的ADS-SSR数据融合研究[J]. 北京航空航天大学学报, 2001, 27(1): 24-27.
ZHANG X J, ZHANG Q S. Data fusion of ADS-SSR in air traffic management[J]. Journal of Beijing University of Aeronautics and Astronautics, 2001, 27(1): 24-27(in Chinese).

[8] XIA X, HASHEMI E, XIONG L, et al. Autonomous vehicle kinematics and dynamics synthesis for sideslip angle estimation based on consensus Kalman filter[J]. IEEE Transactions on Control Systems Technology, 2022, 31(1): 179-192.

[9] 刘康, 何明浩, 韩俊, 等. 基于多传感器的雷达对抗侦察数据融合算法[J]. 系统工程与电子技术, 2023, 45(1): 101-107.
LIU K, HE M H, HAN J, et al. Data fusion algorithm for radar countermeasures and reconnaissance based on multi-sensor[J]. Systems Engineering and Electronics, 2023, 45(1): 101-107(in Chinese).

[10] 李素. 基于航迹质量分析的加权平均融合算法[J]. 现代计算机(专业版), 2018, 24(5): 12-15.
LI S. A weighted average fusion algorithm based on track quality analysis[J]. Modern Computer, 2018, 24(5): 12-15(in Chinese).

[11] RAJASEKARAN R K, AHMED N, FREW E. Bayesian fusion of unlabeled vision and RF data for aerial tracking of ground targets[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2020: 1629-1636.

[12] 郭建刚, 陈鹏, 郑伟. 多表冗余惯导数据融合算法及在自对准中的应用[J]. 北京航空航天大学学报, 2020, 46(12): 2211-2216.
GUO J G, CHEN P, ZHENG W. Data fusion algorithm of multi-sensor redundant inertial navigation and its application in self-alignment[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(12): 2211-2216(in Chinese).

[13] 周恩帆, 马俊, 周永杰, 等. 一种D-S证据理论的多传感器数据融合算法[J]. 小型微型计算机系统, 2022, 43(4): 795-800.
ZHOU E F, MA J, ZHOU Y J, et al. Multi-sensor data fusion algorithm based on D-S evidence theory[J]. Journal of Chinese Computer Systems, 2022, 43(4): 795-800(in Chinese).

[14] KACZMAREK A, ROHM W, KLINGBEIL L, et al. Experimental 2D extended Kalman filter sensor fusion for low-cost GNSS/IMU/Odometers precise positioning system[J]. Measurement, 2022, 193: 110963.

[15] 宋安宇. 基于RNN的航空监视信息融合技术的研究与实现[D]. 北京: 北京邮电大学, 2019.
SONG A Y. Research and implementation of aviation surveillance information fusion technology based on RNN[D]. Beijing: Beijing University of Posts and Telecommunications, 2019(in Chinese).

[16] 孟致远. 基于多神经网络的航空监视信息融合系统的研究与实现[D]. 北京: 北京邮电大学, 2021.
MENG Z Y. The research and implementation of aviation surveillance information fusion system based on multi neural network[D]. Beijing: Beijing University of Posts and Telecommunications, 2021 (in Chinese).

[17] 陈建军, 孙俊, 杨予昊, 等. 机场场面监视雷达系统设计[J]. 现代雷达, 2018, 40(12): 11-14.
CHEN J J, SUN J, YANG Y H, et al. Design of airport surface movement radar system[J]. Modern Radar, 2018, 40(12): 11-14(in Chinese).

[18] LU Y, HUANG R S, XU Z L. Multi-sensor data fusion based on ADS-B and MLAT in approach[J]. Applied Mechanics and Materials, 2014, 602-605: 2491-2494.

[19] 闫峰. 浅析北京大兴国际机场多点监视系统[J]. 电脑知识与技术, 2021, 17(21): 136-138.
YAN F. Analysis of Beijing daxing international airport multipoint

surveillance system[J]. Computer Knowledge and Technology, 2021, 17(21): 136-138(in Chinese).

[20] 牛中伟, 陈健明. 空管场监系统雷达数据格式分析[J]. 信息通信, 2019, 32(1): 156-158.
NIU Z W, CHEN J M. Analysis of radar data format for air traffic control field supervision system[J]. Information & Communications, 2019, 32(1): 156-158(in Chinese).

[21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.

Aircraft surveillance data fusion method in flight area based on Trans-Attention

WANG Xinglong[1,*]，YIN Hao[1]，DING Junfeng[2]

(1. Key Laboratory of Internet of Aircrafts，Civil Aviation University of China，Tianjin 300300，China；
2. Nanjing Les Information Technology Co.，Ltd.，Nanjing 210001，China)

Abstract：An aircraft surveillance data fusion method based on a Transformer and attention mechanism is proposed to address the issues of low monitoring accuracy and position jump in a single surveillance source for aircraft in the flight area. Prior to assigning weight values to various surveillance sources via the attention mechanism, features are first extracted from each surveillance source data using the Transformer's encoder structure. Finally, regression calculations are performed through a fully connected network to obtain the final fusion result. The multilateration (MLAT) data are employed as actual tags, while the surveillance data from the autonomous dependent surveillance-broadcast (ADS-B) system and the surface movement radar (SMR) are chosen as fusion sources. The experimental results show that the proposed method effectively reduces the surveillance error of a single surveillance source, and the fusion effect is better than that of the long short-term memory network based on the attention mechanism, recurrent neural network, and extended Kalman filter fusion methods. The mean absolute error is increased by 2.81%, 16.73% and 35.80% respectively.

Keywords：data fusion；Transformer；attention mechanism；surface movement radar；automatic dependent surveillance-broadcast

Received：2023-05-09；**Accepted**：2023-07-14；**Published Online**：2023-08-30 10：35

URL：link.cnki.net/urlid/11.2625.V.20230829.1724.003

Foundation items：General Program of the National Natural Science Foundation of China (62173332); Key Program of the National Natural Science Foundation of China (U2133207); Diversified Fund Project of Tianjin, China (21JCYBJCO0700); School Level Graduate Research Innovation Project of Civil Aviation University of China (2022YJS086)

*** Corresponding author**. E-mail：xinglong1979@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0177

面向卫星星座的多地面站系统链路分配策略

姜逸飞[1]，何婉霞[2]，刘文正[3]，吴树范[1,*]，魏骁[3]，莫乾坤[1]

(1. 上海交通大学 航空航天学院，上海 200240；2. 上海航天计算机技术研究所，上海 201101；
3. 上海航天系统工程研究所，上海 201101)

摘　　要：针对大规模卫星星座开展通信资源的研究，随着多个低轨星座的成功构建，标志着大规模卫星星座的商用化已经开始。这类大规模卫星星座可以同时兼顾低时延、广覆盖和大流量的优点。高效利用地面站资源成为大规模卫星星座研究的技术瓶颈，针对该问题，提出多星及多站拓扑结构，并结合链路时间资源，提出星地链路时间模型。同时，提出时隙插入策略，以多地面站的整体链路效能作为优化目标，得到多星及多站链路时间表。通过与随机策略和贪婪策略的性能对比，验证了所提策略的可行性和优越性。

关 键 词：卫星星座；地面站；资源分配；网络拓扑；通信链路
中图分类号：V221^{+}.3; TB553
文献标志码：A　　**文章编号**：1001-5965(2025)04-1224-10

随着卫星技术的蓬勃发展，大规模卫星星座已经成为主流研究方向之一。特别是随着多个低轨星座的成功构建，卫星星座的发展已进入商用化进程。Starlink 的 V1 阶段包含了 4 408 颗卫星，Oneweb 的第 1 阶段包含了 576 颗卫星节点[1]。从轨道拓扑而言，前者是由 2 组倾斜轨道和 2 组垂直轨道组成的综合式轨道拓扑；后者是由垂直轨道组成的单一式轨道拓扑。从通信模型而言，前者同时包含了星间链路(inter-satellite-link, ISL)和星地转发链路；后者只包含了星地转发链路。对于中国而言，以“星网”为首的大规模卫星星座包含 2 个阶段：第 1 阶段包含了百余颗卫星，以专用数据传输和功能验证为主；第 2 阶段包含了千余颗卫星，以军民融合应用为主。

对于这些卫星星座的研究涉及了多个技术领域，包含轨道拓扑、网络路由、波束控制、通信技术和资源分配等。轨道拓扑主要针对动力学模型和控制科学；网络路由以各种时空模型为载体，利用不同算法优化路由路径[2-4]；波速控制以多波束模型为载体，利用不同算法优化波位图谱[5]；通信技术主要以物理层为研究对象，包含调制解调、编码、复用和软硬件平台[6-7]；资源分配主要以可用资源为边界条件，通过不同算法实现资源分配[8-9]。这些研究都拥有共同的目标，即提高卫星星座的通信效能。

但是，无论何种通信模式的卫星星座，都需要与地面网络进行大量的数据交互[10-11]，包含用于星座运行维护的遥控、遥测和导航数据，用于移动通信管理的核心网数据和由业务流产生的用户面数据[12]。为满足星地数据交互的要求，地面站系统承担了大量的业务压力。结合信息安全和中国的实际情况，无法全球化部署地面站。因此，在有限的地面站资源内，提高多地面站系统的整体效能是关键。

目前，已有大量针对卫星地面站的研究。一些关于硬件设备的研究包含了天线、图形处理器(graphics processing unit, GPU)加速计算机和通信接

收稿日期：2023-04-13；**录用日期**：2023-08-04；**网络出版时间**：2023-08-30 13:28
网络出版地址：link.cnki.net/urlid/11.2625.V.20230829.1840.004
基金项目：大规模卫星星座传输策略研究 (23GFH-HT01-082)
* **通信作者**. E-mail：shufan.wu@sjtu.edu.cn

引用格式：姜逸飞，何婉霞，刘文正，等. 面向卫星星座的多地面站系统链路分配策略 [J]. 北京航空航天大学学报，2025，51（4）：1224-1233. JIANG Y F，HE W X，LIU W Z，et al. Multi-ground station system-based link allocation strategy for satellite constellation [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1224-1233（in Chinese）.

收机[13-15]。但是, 这些研究主要针对单颗卫星和单个地面站的应用。对于非同步轨道卫星而言, 与固定地面站的可连接时间较短。因此, 单地面站部署将会导致较低的传输效能。为解决该问题, 由多个地面站共同服务一颗卫星的多地面站部署引起了学者的关注[16-18]。这些研究主要以地面网络为载体, 将多个地面站进行协同融合。

对于大规模卫星星座而言, 为多星及多站的拓扑结构分配链路资源是一个 NP 问题。此外, 在有限的国土面积内, 多站部署问题也是一个难点。部署更多的地面站, 可以有效增加链路数量, 但也会造成干扰及增加成本。在多星及多站拓扑的基础上, 引入一个合理的星地链路时间模型可以提高链路分配效率。最终需要一个合适的算法完成星地链路分配策略。

1　多星及多站拓扑

为研究多星及多站拓扑链路分配问题, 具有代表性的多星及多站部署是必不可少的。其中, 大规模卫星星座作为多星部署的应用, 已经成为研究热点。合理的星座拓扑设计需要同时满足全球覆盖和服务质量 (quality-of-service, QoS) 要求。相应地, 合理的多站部署需要同时满足基本国情和通信效能。由于通信效能同时受到链路数量和通信干扰的影响, 多站部署的设计需要追求整体利益。

1.1　星座拓扑

为满足覆盖性和 QoS 的要求, 具有对称特性的 Walker 星座是首选。根据拓扑结构的不同, Walker 星座又可以分为 Walker-delta 拓扑和 Walker-polar 拓扑。其中, Walker-delta 拓扑由多条均匀分布于赤道上的倾斜轨道构成, 每条轨道上均匀分布多颗卫星, 该拓扑可以更好地覆盖中低纬度区域, 为业务热点提供高流量服务, 但无法覆盖两极区域; Walker-polar 拓扑由多条均匀分布于赤道上的垂直轨道构成, 每条轨道上均匀分布多颗卫星, 该拓扑实现全球覆盖, 但是无法满足业务热点的高流量要求。已经完成初步部署的两大商用星座 Starlink 和 Oneweb 是由上述 2 种拓扑构成。Starlink 是由 Walker-delta 拓扑和 Walker-polar 拓扑共同组成的混合式星座, 不同星座拓扑之间不建立通信, 同个星座拓扑内部通过星间链路进行通信; Oneweb 是由 Walker-polar 拓扑构成的单一式星座, 星座拓扑内部不建立通信。表 1 为两大星座的轨道拓扑信息。Starlink 由 3 组 Walker-delta 拓扑 (S.1, S.2, S.3) 和 2 组 Walker-polar 拓扑 (S.4, S.5) 组合构成, 共 4 408 颗卫星; Oneweb 由 1 组 Walker-polar 拓扑 (O.1) 构成, 共 576 颗卫星。

表 1　两大星座轨道拓扑

Table 1　Orbit topology of two constellations

星座	组号	轨道高度/km	轨道数量	卫星总数	轨道倾角/(°)
Starlink	S.1	550	22	1 584	53
	S.2	570	20	720	70
	S.3	540	22	1 584	53.2
	S.4	560	58	348	97.6
	S.5	560	42	172	97.6
Oneweb	O.1	1 200	12	576	87.8

为聚焦本文的研究重点, 具有代表性的卫星星座拓扑是研究基础。经过分析可知, 所有的混合式星座拓扑都可以分解为多个相互独立的单一式星座拓扑。Walker-polar 拓扑的轨道倾角接近 90°, 上行轨道和下行轨道分布于独立的半个赤道, 该拓扑结构使得所有轨道没有重叠。反之, Walker-delta 拓扑的上行轨道和下行轨道相互交织。针对多星及多站部署的研究, 无重叠轨道的 Walker-polar 拓扑可以作为 Walker-delta 拓扑的一种简化情况。综上所述, 选择规模合适的 Walker-delta 拓扑作为研究对象具有典型代表性。

($N/P/F,h,i$) 作为 Walker-delta 拓扑的一种常用表现形式, 在本文被使用。其中, N 为星座拓扑的卫星总数; P 为星座拓扑的轨道数量; F 为相位因子, 是介于 0 和 $P-1$ 之间的整数; h 为轨道高度; i 为轨道倾角。星座拓扑可根据实际需求任意变化, 后续资源分配策略可适用于任意星座拓扑。多星多站资源分配策略可以根据星座部署和地面网络部署自主运行。一个基于 (1 152/24/0,580,58°) 的 Walker-delta 拓扑作为本文的统一模型, 基本拓扑如图 1 所示。每个轨道面为一种颜色: 48 个轨道面; 24 颗卫星每轨; 58°倾角; 580 km 轨道高度。该拓扑可以满足全球 96% 区域的覆盖, 每个地面用户至少被 4 个卫星节点同时覆盖, 其具有典型性, 可以作为统一的星座模型开展研究工作。

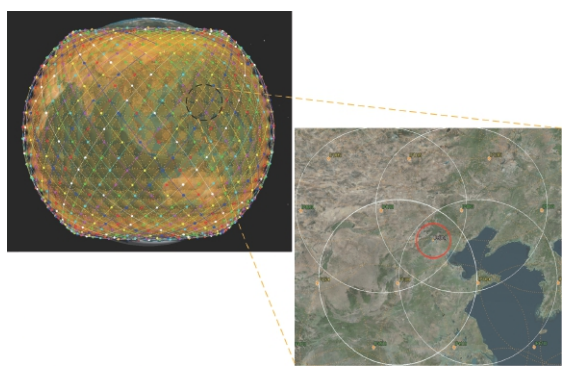


图 1　Walker-delta 星座拓扑

Fig. 1　Topology of Walker-delta constellation

1.2 星地通信模型

对于多站部署问题，需要兼顾链路数量和通信干扰。针对中国的实际情况，地面站部署的边界范围划定为国境区域。多地面站部署的密度越高，可使用的地面站总数就越多，星地链路的可使用资源就越多。但是，随着多地面站部署的密度升高，多星及多站之间的相互干扰也会随之升高。这是由于卫星天线拥有一定的波束宽度，会对周围空间造成干扰。一般地，使用半功率波束宽度 (half power beamwidth, HPBW) 来定义波束的方向性能力。在小于 HPBW 的辐射角度内，天线增益由最大辐射增益决定；在大于 HPBW 的辐射范围时，天线增益会急速衰减。本文天线辐射特性参照标准执行，可以表示为

$$G_{\mathrm{a}}(\varphi)=\begin{cases}G_{\max} & \varphi\leqslant\varphi_{3\mathrm{dB}}\\ G_{\max}-3\left(\dfrac{\varphi}{\varphi_{3\mathrm{dB}}}\right)^{2} & \varphi_{3\mathrm{dB}}<\varphi\leqslant 1.5\varphi_{3\mathrm{dB}}\\ G_{\max}+L_{\mathrm{s}}-25\lg\left(\dfrac{3\varphi}{2\varphi_{3\mathrm{dB}}}\right) & \varphi>1.5\varphi_{3\mathrm{dB}}\end{cases} \tag{1}$$

式中：$G_{\max}$ 为最大辐射增益；φ 为辐射偏角；$\varphi_{3\mathrm{dB}}$ 为 HPBW；L_{s} 为衰减因子。

图 2 为式 (1) 表示的天线辐射方向。

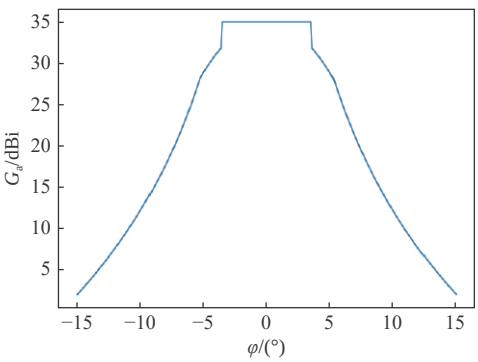


图 2 天线辐射方向

Fig. 2 Antenna radiation direction

多星及多站部署的通信干扰出现于星地链路之间。从理论上，对于下行链路，地面站会收到来自其可视范围内所有卫星发射的电磁能量。介于通信双方之间的电磁能量被定义为接收信号，其余能量称为干扰。实际上，当通信距离或辐射偏角较大时，电磁能量会急剧下降。为聚焦关键问题，本文只研究相邻卫星和相邻地面站之间的干扰问题。对于上行链路，存在相似的干扰情况。

图 3 为多星及多站部署的干扰模型。其中，Sat 1 和 Sat 2 表示 2 颗相邻卫星，Grd 1 和 Grd 2 表示 2 个相邻地面站。Sat 1 和 Grd 1 的波束相互对准，Sat 2 和 Grd 2 的波束相互对准。由 Sat 2 向 Grd 2 发送的

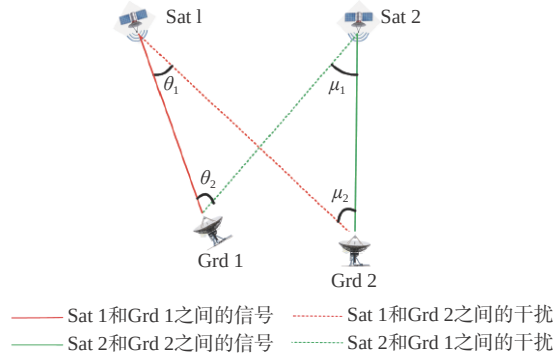


图 3 Sat2 和 Grd1 之间的干扰模型

Fig. 3 Multi-satellite and multi-station interference models

信号会有部分能量作为干扰泄露到 Grd 1。类似地，Grd 2 会收到来自 Sat 1 的干扰；Sat 1 会收到来自 Grd 2 的干扰；Sat 2 会收到来自 Grd 1 的干扰。θ_1 和 θ_2 分别表示 Sat 1 对 Grd 2 和 Grd 1 对 Sat 2 的干扰辐射偏角。类似地，μ_1 和 μ_2 也是干扰辐射偏角。

在假设通信双方天线对准的情况下，信号强度可表示为

$$P_{\mathrm{rec}}=P_{\mathrm{signal}}+G_{\mathrm{asignal}} \tag{2}$$

式中：P_{signal} 为发射信号的功率；G_{asignal} 为收发天线总增益。在理想条件下，收发天线总增益为发射天线和接收天线的最大辐射增益之和。

干扰强度可表示为

$$P_{\mathrm{intf}}=\sum P_{\mathrm{intf}}G_{\mathrm{nintf}} \tag{3}$$

式中：P_{intf} 为干扰信号功率；G_{nintf} 为干扰天线总增益，大小由 θ_1、θ_2、μ_1、μ_2 及式 (1) 共同决定。

根据信号强度和干扰强度，可以得到信号干扰噪声比 (signal and interference noise ratio, SINR)，表示为

$$S_{\mathrm{INR}}=\frac{P_{\mathrm{rec}}}{N_{\mathrm{O}}+P_{\mathrm{intf}}\alpha} \tag{4}$$

式中：N_{O} 为接收机噪声功率；α 为带宽修正因子，介于 0～1 之间。

基于式 (4)，可以得到星地间的信道容量为

$$C_{\mathrm{GSL}}=B\lg(1+S_{\mathrm{INR}}) \tag{5}$$

式中：B 为带宽。

1.3 多站部署

由式 (1)～式 (5) 可知，对于确定的轨道拓扑和多站部署，就可以得知每条星地链路的通信效能。为了一般化，本文假设所有地面站只会和相邻的卫星建立星地链路。在已知卫星拓扑的情况下，以所有星地链路的整体通信效能作为优化目标，可以得到合适的多站部署。

为提高星地链路的整体效能，多地面站将均匀

地分布于允许边界范围内。定义Δl表示相邻地面站之间的经度差和纬度差, 在边界范围内, 以Δl为间隔均匀分布多地面站。将所有地面站与相邻卫星之间的信道价值函数作为对象, 优化Δl的取值。信道价值函数定义为

$$F_{\text{value}}=\frac{\sum_{N_u} C_{\text{GSL}}}{N_u}-\beta N_u \tag{6}$$

式中: N_u为在边界范围内, 以Δl为间隔分布的地面站总数; β为地面站的成本因子, 介于 0～1 之间。定义 $(\Delta l, u)$ 表示多站部署。

本文以中国国境为边界范围, 基于上述理论推导和地理环境限制, 最终选取Δl为 5°。在全国范围内, 共部署了 40 个地面站, 具体分布如表 2 所示。

表 2　多地面站部署分布

Table 2　Deployment of multiple ground stations

编号	纬度/(°)	经度/(°)	编号	纬度/(°)	经度/(°)
1	20	110	21	35	120
2	25	100	22	40	70
3	25	105	23	40	75
4	25	115	24	40	80
5	30	85	25	40	85
6	30	90	26	40	90
7	30	95	27	40	95
8	30	100	28	40	100
9	30	105	29	40	105
10	30	110	30	40	110
11	30	115	31	40	115
12	30	120	32	40	120
13	35	80	33	45	85
14	35	85	34	45	90
15	35	90	35	45	115
16	35	95	36	45	120
17	35	100	37	45	125
18	35	105	38	45	130
19	35	110	39	50	120
20	35	115	40	50	125

本文基于上述 (1 152/24/0, 580, 58°) 的 Walker-delta 拓扑和 (5°, 40) 的多站部署, 研究星地链路资源分配策略。先根据星地可视状态和链路状态建立统一的星地链路时间模型, 再使用时隙插入 (time-sloe-insert, TSI) 策略进行解算, 最终得到星地链路分配时间表。

2　星地链路时间模型

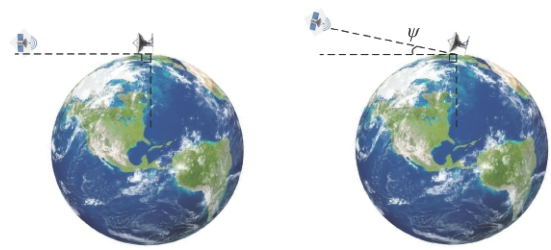
在进行多星及多站链路资源分配之前, 需要进行星地链路计算, 得到星地链路时间信息。但是, 由于 Walker-delta 拓扑的对称性, 同一个地面站可能同时对多颗卫星可见。在可视链路的基础上, 通过时间排序和链路条件筛选, 生成每个地面站的按序无重叠链路时间模型, 该模型可用于后续的资源分配策略。

根据不同的链路状态, 可以将星地链路分为可视链路和状态限制链路。顾名思义, 前者只要满足地球遮挡要求, 即仰角不小于 0°。当卫星和地面站之间的连线高于卫星与地球的切线时, 星地之间处于可视状态。但是, 此时的通信能力很弱, 加上地面遮挡的干扰, 几乎无法建立链路。在此基础上, 根据不同的状态限制, 引入仰角 ψ。从工程而言, 在低干扰和无遮挡的空旷区域, 当 ψ>10°时, 可以建链。对于城市环境而言, 当 ψ>40°时, 链路的通信效能才能稳定, 如图 4 所示。

基于上述分析, 星地链路的状态限制处理流程如图 5 所示, 图中, D 为满足仰角条件的持续时间。

经过上述处理后, 可以得到 4 类状态限制链路, 分别为优选状态链路、短期链路、应急链路和低仰角链路。将优选状态链路作为链路时间模型的输入, 其他链路则按需调用。

虽然优选状态链路拥有较强的通信能力, 但其中存在一个地面站与不同卫星之间的时间重叠链路。为保证在同一时间内, 一个地面站只和一颗卫星存在链路关系, 优选状态链路需要进一步被



(a) 可视链路　　(b) 状态限制链路

图 4　星地链路

Fig. 4　Satellite-ground link

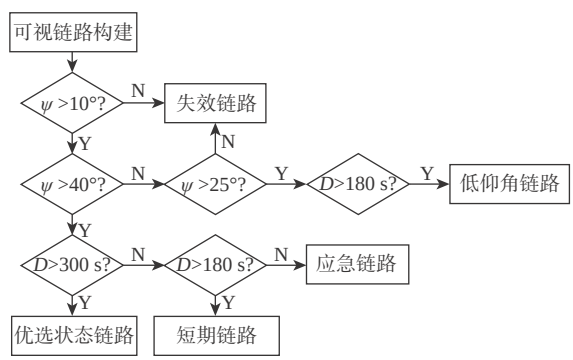


图 5　状态限制链路判断流程

Fig. 5　State limit link judgment process

处理。

构建链路时间模型分为2个步骤：按序排列和重叠解耦。

定义优选状态链路为集合A，将其按地面站分组，形成u个子集$\{a_1,a_2,\cdots,a_u\}$，并且满足：

$$A=a_1\cup a_2\cup\cdots\cup a_u \tag{7}$$

任意子集可表示为

$$a_i=\left\{\left[s_1^i,e_1^i\right],\left[s_2^i,e_2^i\right],\cdots,\left[s_n^i,e_n^i\right]\right\} \tag{8}$$

式中：$i=1,2,\cdots,u$；n为第i个子集的链路数量；s_i^j为第i个子集的第j个链路的起始时间；e_i^j为第i个子集的第j个链路的结束时间。所有时间标准为协调世界时间UTC。按序排列是将每个子集中的所有链路按照起始时间进行排序，得到按序集合$B=\{b_1,b_2,\cdots,b_u\}$。

每个子集可表示为

$$b_i=\left\{\left[{}_bs_1^i,{}_be_1^i\right],\left[{}_bs_2^i,{}_be_2^i\right],\cdots,\left[{}_bs_n^i,{}_be_n^i\right]\right\} \tag{9}$$

式中：${}_bs_i^j$为第i个按序子集的第j个链路的起始时间；${}_be_i^j$为第i个按序子集的第j个链路的结束时间。满足条件：

$${}_bs_k^i>{}_bs_l^i\qquad k>l \tag{10}$$

根据UTC的定义，式(10)表示链路k的开始时间晚于链路l的开始时间。

重叠解耦是将按序子集中所有链路的重叠部分去除。链路重叠的定义如下：

$${}_bs_k^i<{}_be_1^i\qquad k>l \tag{11}$$

式(11)表示，第i个按序子集的第k个链路和第l个链路发生重叠。重复对比所有子集中的所有链路，删除所有重叠链路。

为兼顾所有卫星的连接机会，在重叠删除的过程中，引入禁忌表集合$T=\{t_1,t_2,\cdots,t_d\}$，其中，d为禁忌表的阶数，每个t_i，$i\in[1,d]$表示一个禁忌表。从重叠解耦开始，将无需删除卫星节点装入禁忌表中。每当t_i中的元素个数大于卫星总数的60%，自动开启下一阶禁忌表t_{i+1}。在进行重叠解耦的过程中，保留未进入禁忌表或禁忌阶数低的卫星节点。将包含高阶禁忌卫星节点的链路删除。

经过重叠解耦，得到按序无重叠集合$C=\{c_1,c_2,\cdots,c_u\}$，每个子集可表示为

$$c_i=\left\{\left[{}_cs_1^i,{}_ce_1^i\right],\left[{}_cs_2^i,{}_ce_2^i\right],\cdots,\left[{}_cs_n^i,{}_ce_n^i\right]\right\} \tag{12}$$

式中：n为第i个按序无重叠子集的链路数量；${}_cs_i^j$为第i个按序无重叠子集的第j个链路的起始时间；${}_ce_i^j$为第i个按序无重叠子集的第j个链路的结束时间。所有按序无重叠子集的链路满足以下条件：

$${}_cs_l^i<{}_ce_l^i<{}_cs_k^i<{}_ce_k^i\qquad\forall k>l \tag{13}$$

按序无重叠集合C已是最原始的多星及多站链路分配时间表。如果根据集合C为地面站系统进行任务调度，可以保证不发生重叠冲突。但是，每个子集中的相邻链路之间存在时间间隙问题，定义时间间隙集合如下：

$$G_m=\left\{g_1^m,g_2^m,\cdots,g_{n-1}^m\right\}\qquad m=1,2,\cdots,u \tag{14}$$

每个时间间隙可以定义为

$$g_p^m={}_cs_{p+1}^m-{}_ce_p^m \tag{15}$$

式(15)表示，第m个按序无重叠子集的第p个时间间隔为：该子集的第p个链路的结束时间与第p+1个链路的起始时间之差。在时间间隙内，地面站处于空闲，定义为占空状态。这严重影响了多星及多站链路的通信效能。

根据上述星地链路时间模型，可以定义本文的目标函数为时隙占空比(duty ratio of gap, DRG)：

$$D_{\mathrm{RG}}=\sum_{m\in[1,u]}\left(\frac{\sum\limits_{i\in[1,n-1]}g_i^m}{t_m}\right) \tag{16}$$

式中：u为地面站数量；n表示第m个链路；t_m为第m个链路总工作时间。对应的限制条件为

$$\begin{cases}\sum\limits_{i\in[1,n-1]}g_i^m<t_m\\ t_m>0\end{cases} \tag{17}$$

利用式(16)的目标函数和式(17)的限制条件，对多卫星和多地面站之间的链接资源构建完整模型，并用时隙插入策略进行解算。

3 时隙插入策略

为解决由时间间隙造成的性能下降，本节提出一种时隙插入策略。根据时间间隙的大小分为3种间隙状态：

$$Q=\begin{cases}l_{\mathrm{qgap}} & g<60\\ m_{\mathrm{zgap}} & 60\leqslant g<180\\ l_{\mathrm{dgap}} & g\geqslant 180\end{cases} \tag{18}$$

针对3种间隙状态，分别提出3种子策略。

1) 遮盖策略：对于l_{qgap}而言，延后前链路的结束时间或提早后链路的起始时间，可以有效遮盖时间间隙。

2) 拼接策略：对于m_{zgap}而言，同时延后前链路的结束时间和提早后链路的起始时间，利用前后拼接的方法完成覆盖。

3) 插入策略：对于l_{dgap}而言，将按序集合B中删

除的重叠链路作为备选链路。同时, 调制前链路的结束时间和后链路的起始时间, 并且调整重叠链路的起止时间。将重叠链路与前后链路不发生重叠, 并插入前后链路之间。

遮盖策略和拼接策略可以正常实施的先决条件是原链路仰角 ψ 有一定的通信余量。延迟链路的结束时间或提早链路的起始时间后, 调整后的链路仍然可以保证较高的通信效能。

对于遮盖策略, 对比前链路结束时刻的仰角和后链路起始时刻的仰角, 选择仰角高的链路进行拓展。遮盖链路选择定义如下:

$$l_{\text{select}}=\begin{cases} l_{\text{front}} & \psi_{\text{end}}\geqslant\psi_{\text{begin}} \\ l_{\text{next}} & \psi_{\text{end}}<\psi_{\text{begin}} \end{cases} \tag{19}$$

式中: ψ_{begin} 为后链路起始仰角; ψ_{end} 为前链路结束仰角; l_{front} 为延迟前链路的结束时间; l_{next} 为提早后链路的开始时间。遮盖策略会使用通信资源较多的链路去弥补时间间隙, 提高星地链路通信效能。

对于拼接策略, 前后链路拓展的时间长度可分别表示为

$$E_{\text{next}}=g\frac{\psi_{\text{begin}}}{\psi_{\text{begin}}+\psi_{\text{end}}} \tag{20}$$

$$E_{\text{front}}=g\frac{\psi_{\text{end}}}{\psi_{\text{begin}}+\psi_{\text{end}}} \tag{21}$$

式中: g 为被拼接间隙的时长; E_{next} 为后链路的提前时长; E_{front} 为前链路的延迟时长。拼接策略会同时使用前后链路的通信资源, 但是拓展时长与通信资源成正比, 该策略可以平衡通信负载, 提高整体效能。

对于插入策略, 会从重叠链路中选择一个剩余资源较多的链路。同时调整前链路的结束时间、后链路的起始时间和插入链路的起止时间。3 个链路的持续时间与通信资源相关, 具体表示如下:

$$D_{\text{next}}=\frac{\psi_{\text{begin}}^{\max}}{\psi_{\text{begin}}^{\max}+\psi_{\text{insret}}^{\max}+\psi_{\text{end}}^{\max}}D_{\text{total}} \tag{22}$$

$$D_{\text{front}}=\frac{\psi_{\text{end}}^{\max}}{\psi_{\text{begin}}^{\max}+\psi_{\text{insert}}^{\max}+\psi_{\text{end}}^{\max}}D_{\text{total}} \tag{23}$$

$$D_{\text{insert}}=\frac{\psi_{\text{insert}}^{\max}}{\psi_{\text{begin}}^{\max}+\psi_{\text{insert}}^{\max}+\psi_{\text{end}}^{\max}}D_{\text{total}} \tag{24}$$

$$D_{\text{total}}=D_{\text{next}}+D_{\text{front}}+D_{\text{insert}} \tag{25}$$

式中: D_{next} 为后链路的延续时间; D_{front} 为前链路的延续时间; D_{insert} 为新插入链路的延续时间; $\psi_{\text{begin}}^{\max}$ 为后链路的仰角最高值; $\psi_{\text{end}}^{\max}$ 为前链路的仰角最高值; $\psi_{\text{insert}}^{\max}$ 为新插入链路的仰角最高值。该策略通过在时隙中插入新的链路以弥补时间间隙, 同时调整前后链路的工作时间, 追求整体效能最大化。

本节时隙插入策略包含 3 个子策略, 分别解决 3 类不同的链路间隙问题, 最终以多星及多站链路的整体效能为优化目标, 实现高效星地链路。

时隙插入策略伪代码如下:

输入: A, B, C。

输出: u, T, ψ。

1. For: all link in C
2. if $l==l_{\text{qgap}}$:
3. if $\psi_{\text{end}}\geqslant\psi_{\text{begin}}$:
4. $l_{\text{select}}=l_{\text{front}}$
5. else:
6. $l_{\text{select}}=l_{\text{next}}$
7. extend the l_{select}
8. elif $l==m_{\text{zgap}}$:
9. calculate E_{begin} and E_{end}
10. extend l_{front} and l_{next}
11. elif $l==l_{\text{dgap}}$:
12. calculate D_{begin}, D_{end}, D_{insert}
13. adjust l_{front} and l_{next}
14. insert l_{new}

利用本文策略及算法, 可以得到多星及多站链路资源时间表。在实际星地链路构建阶段, 根据预先计算的时间表, 可以高效完成接入任务。本文策略的性能将通过仿真对比展现。

4　仿真对比

为验证时隙插入策略的有效性, 本节将开展一系列仿真及分析。根据本文的 (1 152/24/0, 580, 58°) Walker-delta 拓扑, 轨道周期大于 90 min, 以 24 h 为回归周期, 每 24 h 卫星运行 16 个轨道周期, 总仿真时间设置为 16 个轨道周期。

为更明确地验证本文策略的性能, 定义如下仿真参数: 卫星激活率、最大地面站时间间隙、星地链路时间、地面站连接时间抖动、地面站总建链时间、地面站平均时间间隙和卫星连通率。

为鲜明地体现本文策略的优越性, 使用性能对比方式, 但由于大规模卫星星座是新兴技术, 目前还处于起步阶段。大部分学者都专注于星座建设和组网通信, 忽略了地面运维技术。本文针对的多星及多站问题结合中国实际国情的需求, 因此, 目前没有完全一致的研究结果可以进行对比分析[19-20]。结合类似问题的研究现状, 本文提出 2 个对比策略, 即随机策略和贪婪策略。

1) 随机策略: 基于按序集合 B, 对每个子集 b_i, 随机选择链路。如果被选的链路与已选的链路存在重叠, 则放弃该备选链路。

2）贪婪策略：基于按序无重叠集合 B,对每个子集 b_i, 首先选择延续时间最长和仰角最高的链路，并且以总链路数量最多为目标，迭代调整链路。

将本文的时隙插入策略、随机策略和贪婪策略共同进行仿真及对比，仿真平台基于 Python3.8 和 Network X 组件。其中，Network X 组件主要用于网络参数表征和仿真参数表征，具体仿真过程的动态数据维护通过时间触发和事件触发的方式实现，具体代码基于 Python 的仿真脚本实现。所有卫星星座的网络特性、链路状态和物理限制都在仿真脚本的传输速率、传输延迟和等待时间等参数中体现。

对于仿真中的星座设计、轨道参数和链路状态，利用自定义脚本设计。其中，星座设计利用标准 Walker 星座构型实现。轨道参数利用 J2 摄动（一阶）外推算法，考虑了由地球非球形引起的轨道根数的长周期变化。链路状态由图 5 所示的链路限制条件获得。

卫星激活率以运行时间为自变量，定义如下：

$$r_{\text{satellite enable}}=\frac{u_{\text{enable}}}{u_{\text{total}}}\times 100\% \tag{26}$$

式中：u_{total}为整个拓扑中卫星的总数量；u_{enable}为已经与地面站建链的卫星数量。该参数反映了一个回归周期内所有卫星被地面站跟踪、捕获和测控的概率。仿真结果如图 6 所示。可知，时隙插入策略以更快的速率完成所有卫星的地面建链，在 10 个周期内，可以保证所有卫星与地面站完成测控和关键数据下传。

最大地面站时间间隙以运行时间为自变量，表示每个地面站的最大无业务时间，反映了多星及多站链路时间表的占空情况。仿真结果如图 7 所示。时隙插入策略的时间间隙低于 0.005 h，优于随机策略和贪婪策略。

星地链路时间以运行时间为自变量，表示每颗卫星与地面建链的平均时间，反映了星地链路的吞吐量。仿真结果如图 8 所示。时隙插入策略的

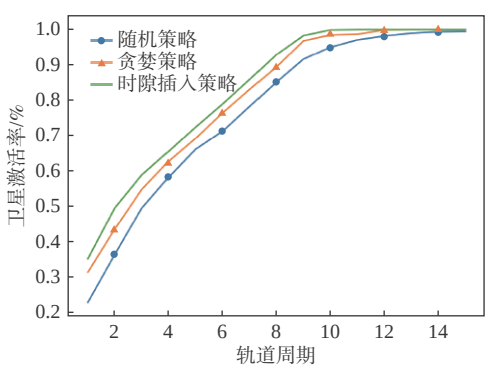


图 6 卫星激活率

Fig. 6 Satellite enabl rate

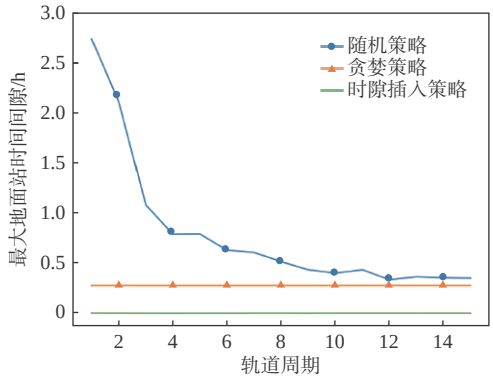


图 7 最大地面站时间间隙

Fig. 7 Maximum gap of ground station

建链时间，无论是绝对值还是增长速率，都优于其他策略。

地面站连接时间抖动表示每个地面站的建链时长抖动，反映地面资源使用的平衡度，是地面负载均衡的体现。仿真结果如图 9 所示。时隙插入策略的地面站连接时间抖动最大值小于 8 min，优于其他策略。本文策略实现了多站资源负载均衡。

地面站总建链时间表示每个地面在仿真周期内总的通信时间，反映了地面站的接入吞吐量。仿

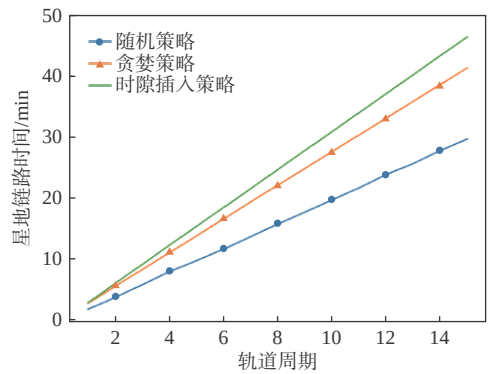


图 8 星地链路时间

Fig. 8 Satellite link to ground time

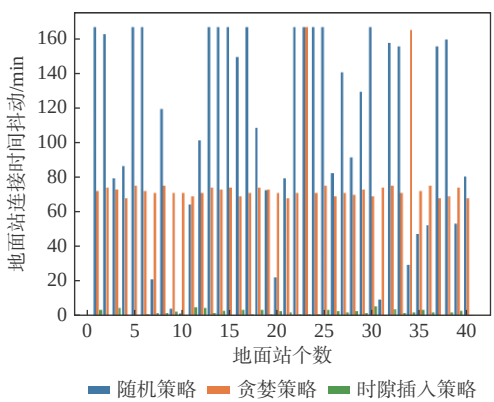


图 9 地面站连接时间抖动

Fig. 9 Ground station link duration jitter

真结果如图 10 所示。一个回归周期内, 时隙插入策略的总建链时间接近 24 h, 优于其他策略。

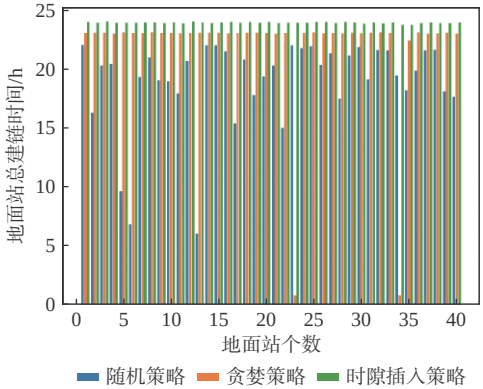


图 10　地面站总建链时间

Fig. 10　Total link duration of ground station

地面站平均时间间隙表示每个地面站在仿真周期内的平均时间间隙, 反映了不同站之间的运行状态和占空情况。仿真结果如图 11 所示。时隙插入策略的平均时间间隙最大值小于 3 min, 优于其他策略。

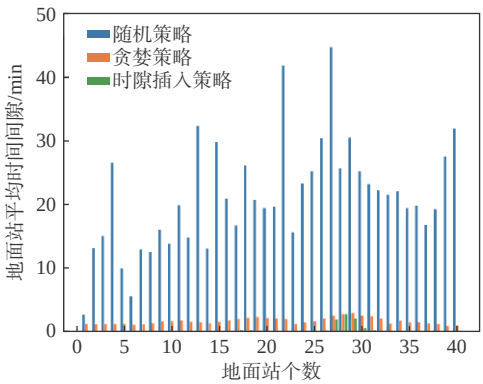


图 11　地面站平均时间间隙

Fig. 11　Ground station average gap

卫星连通率表示每个卫星在运行周期内的星地通信率, 具体表示如下:

$$r_{\text{satellite link}} = \frac{T_{\text{link}}}{T_{\text{simulation}}} \times 100\% \tag{27}$$

式中: $T_{\text{simulation}}$ 为总仿真时长; T_{link} 为卫星与地面站建立通信的时间。该参数反映了所有卫星拥有的星地数据交互机会。仿真结果如图 12 所示。时隙插入策略的卫星连通率以 0.06% 为均值分布, 该均值高于其他策略。

图 6～图 12 的仿真从运行时间、多站部署和卫星星座多个维度入手, 多方面体现了本文策略的性能, 并且通过性能对比的方式, 体现了本文策略的优越性。

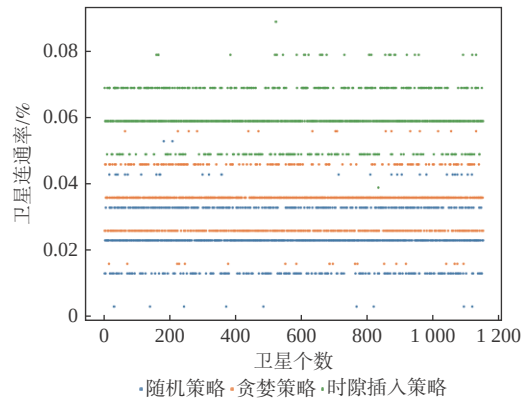


图 12　卫星连通率

Fig. 12　Satellite link rate

5　结　论

1) 本文根据大规模卫星星座的发展现状, 提出多星及多站链路分配研究的需求。

2) 根据主流卫星星座的构建现状, 提出一种基于 (1 152/24/0, 580, 58°)Walker-delta 的星座拓扑。提出星地干扰通信模型, 用于衡量多站部署性能。结合实际国情, 提出 (5°, 40) 的多站部署。

3) 提出一种时隙插入策略用于优化多星及多站链路时间表, 该策略根据不同的时间间隙, 提出 3 个子策略。

4) 利用仿真及对比的方法验证了时隙插入策略的可行性和优越性, 从仿真时间、多站部署和卫星星座 3 个维度进行仿真及对比。

参考文献 (References)

[1] 吴树范, 王伟, 温济帆, 等. 低轨互联网星座发展研究[J]. 北京航空航天大学学报, 2024, 50(1): 1-11.
WU S F, WANG W, WEN J F, et al. Review on development of LEO Internet constellation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(1): 1-11(in Chinese).

[2] JIANG Y F, WU S F, MO Q K. A compass time-space model-based virtual IP routing scheme for NTSN satellite constellations[J]. Chinese Journal of Aeronautics, 2023, 36(9): 280-288.

[3] JIANG Y F, WU S F, MO Q K, et al. An energy sensitive and congestion balance routing scheme for non-terrestrial-satellite-network (NTSN)[J]. Remote Sensing, 2023, 15(3): 585.

[4] HUANG Y X, WU S F, KANG Z Y, et al. Reinforcement learning based dynamic distributed routing scheme for mega LEO satellite networks[J]. Chinese Journal of Aeronautics, 2023, 36(2): 284-291.

[5] TAKAHASHI M, KAWAMOTO Y, KATO N, et al. DBF-based fusion control of transmit power and beam directivity for flexible resource allocation in HTS communication system toward B5G[J]. IEEE Transactions on Wireless Communications, 2021, 21(1): 95-105.

[6] JIANG Y F, HE W X, LIU W Z, et al. A B5G non-terrestrial-network (NTN) and hybird constellation based data collection system (DCS)[J]. Aerospace, 2023, 10(4): 366.

[7] 徐可笛, 徐兆斌, 郭晓旭, 等. 大规模卫星星座组网的码分多址干扰分析[J]. 北京航空航天大学学报, 2024, 50(9): 2885-2892.
XU K D, XU Z B, GUO X X, et al. Analysis of CDMA interference in large-scale satellite constellation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(9): 2885-2892(in Chinese).

[8] XU J, ZHAO Z T, WANG L, et al. A novel deep reinforcement learning architecture for dynamic power and bandwidth allocation in multibeam satellites[J]. Acta Astronautica, 2023, 204: 73-82.

[9] 丁晓进, 严岚亭, 张更新. 一种基于 SDN 的低轨卫星物联网信道资源动态分配方法: CN112272412A[P]. 2021-01-26.
DING X J, YAN L T, ZHANG G X. A dynamic allocation method for low earth orbit satellite IoT channel resources based on SDN: CN112272412A[P]. 2021-01-26(in Chinese).

[10] MAATTANEN H L, HOFSTROM B, EULER S, et al. 5G NR communication over GEO or LEO satellite systems: 3GPP RAN higher layer standardization aspects[C]//Proceedings of the IEEE Global Communications Conference. Piscataway: IEEE Press, 2019: 1-6.

[11] 徐珉, 胡南. 5G 非地面网络组网技术研究[C]//5G 网络创新研讨会会议. 北京:《移动通信》杂志社, 2018: 105-110.
XU M, HU N. Research on 5G non ground network networking technology [C]//Proceedings of the 5G Network Innovation Seminar. Beijing: Mobile Communications Magazine, 2018: 105-110(in Chinese).

[12] 姜逸飞, 吴树范, 莫乾坤. 基于大规模卫星星座的 6G 组网部署及核心网节点动态规划技术[C]//5G 网络创新研讨会会议. 北京: 移动通信, 2022: 403-406.
JIANG Y F, WU S F, MO Q K. 6G networking deployment and dynamic programming technology for core network nodes based on large-scale satellite constellations[C]//Proceedings of the 5G Network Innovation Seminar. Beijing: Mobile Communications Magazine, 2022: 403-406(in Chinese).

[13] HIDAYAT A, GUNAWAN H, SYAHPUTRA NASUTION A, et al. Automatic statistics measurement antenna performance towards virtual ground station[J]. IOP Conference Series: Materials Science and Engineering, 2020, 852(1): 012159.

[14] YOUNG E. GPU-accelerated demodulation for a satellite ground station[D]. Logan: Utah State University, 2019.

[15] TIAN K, WANG Y W, HE X T, et al. Research and application of system equipment fault diagnosis based on satellite ground station[C]//Proceedings of the 2020 International Conference on Computer Engineering and Intelligent Control. Piscataway: IEEE Press, 2020: 66-72.

[16] JIANG Y F, WU S F, MO Q K, et al. A cloud-computing-based portable networked ground station system for microsatellites[J]. Sensors, 2022, 22(9): 3569.

[17] VASISHT D, CHANDRA R. A distributed and hybrid ground station network for low earth orbit satellites[C]//Proceedings of the 19th ACM Workshop on Hot Topics in Networks. New York: ACM, 2020: 190-196.

[18] SCHILLING K, SCHECHNER Y, KOREN I. CloudCT-computed tomography of clouds by a small satellite formation[C]//Proceedings of the 12th IAA Symposium on Small Satellites for Earth Observation. Berlin: Springer, 2019: 7.

[19] HOSSEINIAN M, CHOI J P, CHANG S H, et al. Review of 5G NTN standards development and technical challenges for satellite integration with the 5G network[J]. IEEE Aerospace and Electronic Systems Magazine, 2021, 36(8): 22-31.

[20] DARWISH T, KURT G K, YANIKOMEROGLU H, et al. LEO satellites in 5G and beyond networks: a review from a standardization perspective[J]. IEEE Access, 2021, 10: 35040-35060.

Multi-ground station system-based link allocation strategy for satellite constellation

JIANG Yifei[1]，HE Wanxia[2]，LIU Wenzheng[3]，WU Shufan[1,*]，WEI Xiao[3]，MO Qiankun[1]

(1. School of Aeronautics and Astronautics，Shanghai Jiao Tong University，Shanghai 200240，China；
2. Shanghai Aerospace Computer Technology Institute，Shanghai 201101，China；
3. Shanghai Aerospace Systems Engineering Institute，Shanghai 201101，China)

Abstract：Aiming at the research of large-scale satellite constellation, the research content of communication resources is put forward. With the successful construction of several low-orbit constellations, the commercialization of large-scale satellite constellations has begun. This kind of large-scale satellite constellation can simultaneously take into account the advantages of low delay, wide coverage and large traffic. In order to make efficient use of earth station resources, it has become the technical bottleneck of large-scale satellite constellation research. The multi-star and multi-station topology is proposed to better focus on this problem. Combined with the link time resources, a satellite-earth link time model is proposed. On this basis, the time-slot insertion strategy is proposed, and the multi-star and multi-station link schedules are obtained by taking the overall link efficiency of multi-ground stations as the optimization objective. The feasibility and superiority of the proposed strategy are verified by comparing the performance with random strategy and greedy strategy.

Keywords：satellite constellation; ground station; resource allocation; network topology; communication link

Received：2023-04-13；**Accepted**：2023-08-04；**Published Online**：2023-08-30 13：28

URL：link.cnki.net/urlid/11.2625.V.20230829.1840.004

Foundation item：The Research on Networking Transmission Strategy for Large-Scale Satellite Constellation (23GFH-HT01-082)

* **Corresponding author**. E-mail：shufan.wu@sjtu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0231

TBO模式下基于复杂网络的空中交通复杂性分析

彭娅婷[1,2]，温祥西[1,2]，吴明功[1,2,*]，杨志伟[3]，王楠[3]

(1. 空军工程大学 空管领航学院，西安 710051；2. 国家空管防相撞技术重点实验室，西安 710051；
3. 中国人民解放军93160部队，北京 100800)

摘　　要：由于依据统一间隔标准构建的复杂网络未考虑机型运行的差异性，不能满足基于航迹运行(TBO)下空中交通复杂性分析的精细化要求。为解决该问题，提出一种基于复杂网络区分不同机型的空中交通复杂性分析模型。建立不同机型侧向飞行安全间隔模型，构建航空器精准保护区，优化飞行冲突网络中航空器连边的判定依据。飞行冲突判断在考虑航空器航向和速度等信息的基础上，关注航空器的不同性能与状态，使飞行冲突网络能够更加贴近TBO的运行模式。通过实验仿真TBO运行环境，同时利用厦门高崎国际机场雷达数据进行验证。结果表明：所提模型较改进前的飞行冲突网络，能够精细化航空器间的水平安全间隔标准，降低空域的复杂度，减轻管制员工作负荷，提高空域的运行效率，为航空器自主选择最优化航迹提供更大的空间。

关 键 词：飞行冲突；复杂网络；基于航迹运行；水平间隔标准；保护区模型
中图分类号：V355
文献标志码：A　　**文章编号**：1001-5965(2025)04-1234-11

近年来，民航运输业迅猛发展，同时低空空域改革为通用航空的发展带来了机遇，飞行需求不断增加，空域运行环境日趋复杂，飞行安全压力骤然加剧。为保证飞行安全的同时高效利用有限的空域资源，基于航迹运行(trajectory based operation, TBO)[1]与自由航线空域(free route airspace, FRA)[2]等新一代空管运行概念被提出。在TBO和FRA运行方式中，航空器可根据实际情况自主选取最优化航迹，同时允许根据航空器性能采取不同的间隔标准，大幅增加了飞行流量，提高了空域容量和飞行运行效率[3]。然而，TBO下空域用户数量和机型种类的骤增与用户航线的自由规划，使空情变得更加复杂[4]。因此，迫切需要一种能够反映TBO运行环境下潜在飞行冲突的模型，显示空中交通情况，辅助管制员决策指挥，以实现自由航线TBO。

在空中交通情况评估方面，国内外学者围绕空中交通管理复杂性展开了一系列研究，利用动态密度、内禀性分析和复杂系统建模等方法反映空域运行态势。1976年，Schmidt提出引入管制过程“复杂性或难度”权重系数衡量管制员工作量，由此开启复杂性研究[5]；Delahaye等以航空器对的相对距离与相对速度为基本内禀属性，建立评价体系客观度量典型冲突场景的空中态势及其演化规律[6]；胡明华等[7-8]依据航空器间的迫近效应、汇聚态势和连携关系建立了空域复杂性评估模型，分析空中交通内禀性和几何结构，避免了主观因素的影响，实现了客观描述空中交通复杂性的可能。

复杂系统尤其是复杂网络理论近年来被用于描述空中交通复杂性，为分析空中交通结构和演化进程提供了一条行之有效的研究路线。复杂网络建模的核心是对复杂系统中的节点及节点之间的相互关系进行抽象。王红勇等[9-11]提出以航空器为

收稿日期：2023-05-08；**录用日期**：2023-06-25；**网络出版时间**：2023-07-25 15:39
网络出版地址：link.cnki.net/urlid/11.2625.V.20230725.1328.001
基金项目：国家自然科学基金(71801221)；国家社会科学基金(22XGL001)
*通信作者. E-mail：wuminggong@sohu.com

引用格式：彭娅婷，温祥西，吴明功，等. TBO模式下基于复杂网络的空中交通复杂性分析[J]. 北京航空航天大学学报，2025，51(4)：1234-1244.
PENG Y T，WEN X X，WU M G，et al. Complex network-based air traffic complexity analysis in TBO[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1234-1244 (in Chinese).

节点、航班之间连线为边、节点间的复杂度为权值, 构建二维的加权冲突网络, 用网络的节点度、连通率等衡量空中态势的局部与整体复杂性; 吴明功等[12-14]以飞机为节点, 机载防撞系统 (airborne collision avoidance system, ACAS) 建立通信关系为连边构建飞行状态网络, 选取节点强度、平均聚类系数和介数中心性等网络拓扑指标描述空中动态交通状况, 使用独立主元分析法、层析分析法和熵权法等方法监控空情动态、识别关键节点, 在一定程度上有效降低了主观因素带来的影响, 更好地识别了空中交通复杂性; 毕可心等在该基础上, 引入速度障碍法优化飞行状态网络, 帮助管制员提前对航空器之间的潜在冲突和空域复杂性进行分析[15]。而在这些研究中, 飞行冲突网络的连边是以通信关系或民航现行的统一间隔标准为依据判定, 没有依据航空器性能采取不同的间隔标准进行判定, 不能很好地适用于分析 TBO 运行环境下机型各异的航空器所构成的空中复杂网络。本文在传统基于概率的碰撞风险模型的基础上, 综合考虑 3 种影响因素, 构建了不同机型侧向飞行安全间隔模型, 以此为依据判定飞行冲突关系, 从而对飞行冲突网络结构做出优化, 更适用于分析 TBO 运行环境下的空中交通状况。

因此, 本文提出一种基于 TBO 不同机型的飞行冲突网络模型, 基于概率论和几何法, 综合考虑基于性能的通信、导航与监视 (communication, navigation and surveillance, CNS) 等误差影响, 利用分段查找区间法求解满足目标安全等级(target level of safety, TLS)的水平安全间隔标准, 构建不同机型航空器精准保护区模型。以此为依据利用速度障碍法判定网络连边, 确定网络边权, 引入网络拓扑指标, 分析 TBO 运行环境下的空中交通复杂性。本文模型根据航空器性能采取不同的冲突判定间隔标准, 旨在贴近 TBO 下的空中情况, 同时增大了空域容量, 为实现空域精细化运行打下基础。

1　航空器保护区模型

TBO 模式下的空中交通管理系统由 7 个功能模块组成, 其中, 冲突管理模板是针对航空器等冲突对象, 通过战略冲突管理、配备合理间隔或避免潜在碰撞风险等方式, 消除冲突隐患, 保障飞行安全。飞行冲突探测是研究航空器未来一段时间内是否发生潜在冲突的过程, 由于存在系统误差及无法消除误差的情况, 航空器的探测容易受到外界因素的干扰, 故采用以飞行器为中心建立航空器保护区的方法。目前, 应用较为广泛的保护区模型有圆柱形和椭球形, 本文采用符合实际且便于计算的椭球形保护区模型。依据国际民航组织(International Civil Aviation Organization, ICAO)规定的间隔标准, 椭球体的长焦距 a=10 km, 短焦距 d_v=300 m, 当进入彼此的保护区时则存在冲突。随着 CNS 性能的不断增强、ACAS 的不断完善, 在未来 TBO 模式下, 不能再笼统按照飞行安全间隔标准判断是否存在冲突, 而应为航空器划设更为精细的保护区。

飞机期望航迹存在不确定性误差, 许多不确定因素会导致飞行状态的随机性, 随之影响飞行冲突的判断结果。研究表明, 飞行中期望航迹与实际航迹之间的误差服从正态分布[16]。本文运用飞机质点运动学模型, 基于概率论方法, 引入 CNS、飞行员反应时间和气象风 3 个随机误差项, 推导出航空器间碰撞风险模型, 更加系统、有效地评估最低水平飞行安全间隔标准, 得到更加精准的航空器保护区模型。

1.1　冲突解脱运动过程

飞行过程中, 航空器的飞行轨迹会受到 CNS 性能、飞行员反应时间和气象风 3 个随机因素的影响, 飞机的实际飞行轨迹与期望轨迹有所偏差, 因此, 采用概率建模。同时, 2 个航空器处于同一高度层时, 航空器对头飞行所求得的最小安全间隔满足其他相遇场景的安全要求。因此, 建立不同机型保护区模型时, 应考虑对头飞行的避免碰撞运动过程。

首先, 对航空器冲突解脱的运动过程进行分析。在一段时间内, 同一高度层、不同机型的 2 架航空器 A、B 相向匀速飞行, 如图 1 所示, xOy坐标系中, 航空器机载设备探测到飞行冲突时, 2 架航空器间的距离为 L, 根据飞行基本规则[17]第 71 条, 各自向右避让, 2 架航空器飞行员经过反应时间 $t_{r1(2)}$ 决策和反应之后, 采取做圆周运动的方式进行航向解脱。进而可以把航空器避免碰撞飞行运动过程分为 2 个阶段: 第 1 阶段, 当 t = 0时, 飞行员收到采取解脱策略指令, 因为飞行员存在反应时间 $t_{r1(2)}$, 所

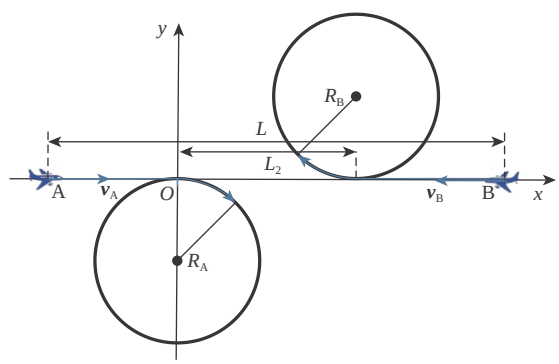


图 1　航空器运动过程示意图

Fig. 1　Schematic diagram of the aircraft movement process

以不改变航空器的航向，沿预定航迹继续飞行；第2阶段，$t=t_{r1(2)}$时，航空器A的位置为(0,0)，航空器B的位置为$(L_2,0)$，飞行员开始执行解脱策略，采取改变航向的措施，左转做匀速圆周运动。假定2架航空器的转弯半径分别为R_A、R_B，转弯的角速度分别为ω_A、ω_B，得到第2阶段改进的飞行运动模型为

$$\begin{cases} Q_A = f(A) + c_A + w_A + r_A \\ Q_B = f(B) + c_B + w_B + r_B \\ D_{AB} = Q_A - Q_B \end{cases} \tag{1}$$

式中：$f(\cdot)$为飞机质点运动学模型，CNS性能满足均值为0、方差为σ_{cns}^2的高斯分布，随机误差函数表示为$c \sim N\left(0,\sigma_{cns}^2\right)$，高空风的随机扰动满足均值为0、方差为$\sigma_w^2$的高斯分布，随机误差函数表示为$w \sim N\left(\mu_w,\sigma_w^2\right)$[18]，随机误差函数$r$服从伽马分布$G(\alpha,\lambda)$[19]，数学期望为$\alpha/\lambda$，方差为$\alpha/\lambda^2$；$D_{AB}$为2架航空器的水平距离，3个随机因素间及2架航空器各自所受随机因素间相互独立；Q为考虑以上3个随机因素下的实际飞机位置坐标。

1.2 碰撞风险模型

本文假定2架航空器处于同一高度层，建立碰撞风险模型，并求解2架航空器水平安全间隔标准。由于高斯随机变量之差仍服从高斯分布，2架航空器在t_1时刻的间隔同样服从高斯分布，均值为2架航空器在t_1时刻的理论间隔的变量，2架航空器之间的水平欧氏距离$d_{AB}=\sqrt{\left(\Delta X_t^2+\Delta Y_t^2\right)}$，$\Delta X_t$与$\Delta Y_t$满足高斯分布：

$$\Delta X_t \sim N(R_A\sin(\omega_A t_1) - L_2 + R_B\sin(\omega_B t_2) + \delta_{Ax} - \delta_{Bx}, \sigma_{Ax}^2 + \sigma_{Bx}^2) = N\left(\mu_x,\sigma_x^2\right) \tag{2}$$

$$\Delta Y_t \sim N(R_A - R_A\cos(\omega_A t_1) - R_B\cos(\omega_B t_2) + R_B + \delta_{Ay} - \delta_{By}, \sigma_{Ay}^2 + \sigma_{By}^2) = N\left(\mu_y,\sigma_y^2\right) \tag{3}$$

式中：t_1和t_2满足$|t_{r1}-t_{r2}|=|t_1-t_2|$；$\delta_{Ax}$为航空器A在$x$方向运行的随机误差。

根据文献[20]，侧向碰撞风险计算式为

$$C_R(t) = 2N_p P(t) \tag{4}$$

式中：N_p为空域或航路上平均每飞行小时内经过的航空器数；$P(t)$为2架航空器的碰撞概率。

当$d_t \leqslant \lambda_A+\lambda_B$时，则视为2架航空器发生碰撞，2架航空器的碰撞概率$P(t)$为其欧氏距离d_t小于$\lambda_A+\lambda_B$的概率，即$P(t)=\{d_t \leqslant \lambda_A+\lambda_B\}$，其中，$\lambda_A$和$\lambda_B$分别为2架航空器碰撞球半径，取其机身长、翼展长和机身高中的最大值。因此，可得第2阶段的任意时刻t_1，航空器间的碰撞概率为

$$P(t) = P_t(d_t \leqslant \lambda_A+\lambda_B) = P_t\left(u \leqslant (\lambda_A+\lambda_B)^2\right) = \int_0^{(\lambda_A+\lambda_B)^2} f'(u)\mathrm{d}u \tag{5}$$

$u=\Delta X_t^2+\Delta Y_t^2$，利用概率论相关知识推导可得$u$的概率密度函数为

$$f'(u) = \begin{cases} \displaystyle\int_0^u \frac{1}{8\pi\sigma_x\sigma_y\sqrt{u-z_2 z_2}} \cdot \left\{\left[\exp\left(-\frac{\left(\sqrt{u-z_2}-\mu_x\right)^2}{2\sigma_x^2}\right)+\exp\left(-\frac{\left(\sqrt{u-z_2}+\mu_x\right)^2}{2\sigma_x^2}\right)\right]\cdot\left[\exp\left(-\frac{\left(\sqrt{z_2}-\mu_y\right)^2}{2\sigma_y^2}\right)+\exp\left(-\frac{\left(\sqrt{z_2}+\mu_y\right)^2}{2\sigma_y^2}\right)\right]\right\}\mathrm{d}z_2 & u>0 \\ 0 & u\leqslant 0 \end{cases} \tag{6}$$

可得，CNS性能、飞行员反应时间和高空风导致的侧向碰撞风险$C_R(t)$为

$$C_R(t) = 2N_p P(t) = 2N_p\int_0^{(\lambda_A+\lambda_B)^2} f'(u)\mathrm{d}u \tag{7}$$

1.3 水平安全间隔的计算

根据ICAO对空中相撞的安全目标等级相关规定[21]，在初始2架航空器水平安全间隔不断变化情况下，解出相应碰撞风险C_R，取C_R为5×10^{-9}时对应的水平安全间隔值作为最低水平安全间隔标准L_{TLS}。

选择最优化方法求解最低水平安全间隔标准L_{TLS}时，粒子群算法、遗传算法和灰狼算法等最优化算法每个步骤产生的解具有随机性，且反复迭代寻找最优解相比二分法和分段查找区间法用时较长，而分段查找区间法相较二分法求解时间更短，能快速地解算不同机型航空器对间的最小安全间隔、预测潜在的飞行冲突和提供冲突解脱指令，因此，选用分段查找区间法对最低水平安全间隔进行求解，计算步骤如下：

步骤1 确定距离查找范围$[0,L_{max}]$及查找区间$\Delta L_1=1\,000$ m。

步骤2 确定第2阶段时间范围$[t_r,t_m]$及步长Δt；计算$C_R(t)$，若在整个时间段满足$C_R(t)$小于等于目标安全等级，则输出此时的双机间距离L_1，否则继续循环。

步骤3 缩小距离查找范围$[L_1-\Delta L_2,L_1]$及查找区间$\Delta L_2=100$ m，跳至步骤2继续循环，进一步缩小查找范围及区间，直至得到符合精度要求$\varepsilon=1$ m的L_{TLS}，即为所求最低水平安全间隔最优解。

定义航空器A、B的保护区为模型所求得的L_{TLS}为长轴、垂直安全间隔为300 m、300 m为短轴的椭球体。保护区长轴的大小与航空器的机型、

所处的 CNS 环境、飞行员反应时间和高空气象风相关。

2　基于 TBO 的飞行冲突网络

构建 TBO 运行模式下的飞行冲突网络, 分析 TBO 模式下空中交通复杂性, 需要解决航空器保护区的确定和网络连边的判断问题。飞行冲突网络 $G=(V,E,W)$是以空域中航空器之间的冲突关系为研究对象的复杂网络, 其中, $V=\{h_1,h_2,\cdots,h_n\}$为网络中由航空器组成的节点集合, $E=\{e_1,e_2,\cdots,e_n\}$为反映航空器之间冲突关系的连边集合, $W=\{w_1,w_2,\cdots,w_n\}$为反映潜在冲突紧迫性的连边权重集合。

2.1　TBO 运行特点

TBO 是以对航空器全生命周期的四维航迹为基础, 在空管、航空公司、机场、航空器等相关方之间实时共享和动态维护航迹动态信息, 进而实现多方协同决策。与现行基于空域扇区的运行方式相比, TBO 具有如下特点:

1) 引入四维航迹管理, 规划和执行飞行计划过程实现一体化, 空管系统对空域资源的使用和管理更为精细。

2) 航空器可根据实际情况自主选择最优航迹, 航空器之间可根据性能采取不同的间隔标准, 实现航迹共享, 提高飞行员情景意识。

3) 利用数据链对空指挥, 实现地面空管系统与机载空管电子系统的协同。

在 TBO 背景下, 航空器间潜在的飞行冲突是无差别的, 因此, 飞行冲突网络为无向网络。

2.2　冲突网络连边的确定

网络中, 航空器节点构成连边需要满足以下 2 个条件: ①位置临近, 当 2 架航空器之间的间隔小于一定范围时, 可能会发生冲突; ②满足速度障碍关系, 即假设一架航空器静止, 并为其设置保护区, 另一架航空器与其相对速度的方向在速度障碍锥内部。

在三维空域中, 为航空器设置椭球形飞行保护区, 当保护区内存在障碍物或其他航空器时, 判定冲突发生。在 TBO 运行背景下, 可根据航空器性能采取不同的间隔标准, 椭球形保护区的长轴焦距为以上模型所求得的最小水平安全间隔 $d_l=L_{TLS}$, 短轴焦距 $d_v=300$ m。设航空器 B 的坐标为 (X_B,Y_B,Z_B), 则飞行保护区椭球面可表示为

$$\frac{(x-X_B)^2}{d_l^2}+\frac{(y-Y_B)^2}{d_l^2}+\frac{(z-Z_B)^2}{d_v^2}=1 \tag{8}$$

航空器节点间建立连边需要通过速度障碍模型进行第 2 次判断。相较于车辆和机器人, 航空器进行机动的空域是三维空间, 因此, 二维的速度障碍法无法适应航空器所有的机动行为, 将其拓展至三维, 如图 2 所示。

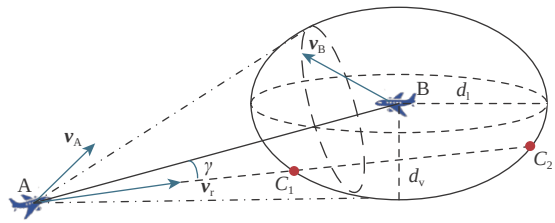


图 2　三维速度障碍冲突探测模型

Fig. 2　Three-dimensional speed barrier conflict detection model

航空器 A 和航空器 B 的速度分别为$\boldsymbol{v}_A$和$\boldsymbol{v}_B$, 利用矢量三角形法得出相对速度$\boldsymbol{v}_r=\boldsymbol{v}_A-\boldsymbol{v}_B$。由航空器 A 向航空器 B 的飞行保护区做切面, 形成的区域称为速度障碍锥。当$\boldsymbol{v}_r$方向在速度障碍锥的内部时, 继续飞行的航空器 A 将会进入航空器 B 的飞行保护区, 如不进行调整, 将会有飞行冲突发生。

$\boldsymbol{v}_r$所在的直线方程为

$$\frac{x-X_A}{\boldsymbol{v}_x}=\frac{y-Y_A}{\boldsymbol{v}_y}=\frac{z-Z_A}{\boldsymbol{v}_z} \tag{9}$$

式中: (X_A,Y_A,Z_A)为航空器 A 的坐标; $(\boldsymbol{v}_x,\boldsymbol{v}_y,\boldsymbol{v}_z)$为$\boldsymbol{v}_r$的方向分量。

联立曲面方程式(8)与直线方程式(9), 如果存在 2 个解, 说明航空器 B 的飞行保护区与$\boldsymbol{v}_r$所在的直线有 2 个交点 C_1 和 C_2; 如果有$\cos\gamma>0$成立, 即$\gamma<90°$, 则$\boldsymbol{v}_r$的方向在速度障碍锥内部, 航空器 A 和航空器 B 之间构成连边:

$$\cos\gamma=\cos\langle\boldsymbol{v}_r,\boldsymbol{l}_{AB}\rangle=\frac{\boldsymbol{v}_r\cdot\boldsymbol{l}_{AB}}{|\boldsymbol{v}_r||\boldsymbol{l}_{AB}|} \tag{10}$$

其中: γ为相对速度$\boldsymbol{v}_r$与航空器 A 和航空器 B 的位置向量$\boldsymbol{l}_{AB}$的夹角; $|\boldsymbol{l}_{AB}|$为航空器 A 和航空器 B 之间的距离。

2.3　冲突网络边权的确定

飞行冲突网络的连边表示空域中航空器之间潜在的飞行冲突, 而边权要能反映冲突的紧迫程度。定义t_c为航空器保持当前的飞行状态, 预计会发生冲突的时间。在设计连边权重时, 需要满足: ①航空器间距离越小, 相对速度越大, 冲突越紧迫, 连边的权重越大, 即边权随t_c的减小而增大; ②随着t_c的减小, 飞行冲突的紧迫性变化越来越剧烈, 权重增加的幅度也越来越大; ③考虑到速度障碍法原理和飞行冲突的现实意义, 连边的权重应为正值。

图 2 中, 相对速度延长线$\boldsymbol{v}_r$与椭球在三维空间有 2 个交点 C_1、C_2, 记航空器距 2 个交点距离为$\left|\overrightarrow{AC_1}\right|$、$\left|\overrightarrow{AC_2}\right|$, 则有

$$t_c=\frac{\min\left\{\left|\overrightarrow{AC_1}\right|,\left|\overrightarrow{AC_2}\right|\right\}}{|\boldsymbol{v}_r|} \tag{11}$$

由于可解脱时间t_c与边权是负相关的关系，结合负指数函数的性质，定义网络边权：

$$h_{ij}=\exp(-t_c) \tag{12}$$

由于t_c为正值，其负指数函数的值域为[0,1]，可同时对飞行冲突网络的边权起到单位化的作用。

3 实验验证与结果分析

3.1 侧向飞行安全间隔模型验证与分析

为验证本文侧向飞行安全间隔模型的有效性，即当2架航空器对头相遇时，所需的安全间隔最大，本文模型求得的最低安全间隔满足其他相遇和运动情况的安全性要求，选取A380和波音B737-800作为研究对象，具体数据如表1所示。速度均选取航空器的巡航速度，转弯坡度选取保证乘客舒适性的较大转弯坡度20°，因为2架航空器所处位置相近，所以，CNS性能均设定为RCP60、RNP10、RSP10[17]，飞行员反应时间t_r服从伽马分布$G(4,5)$[18]，N_p=4架次/h[19]。

利用表1的航空器对性能参数和本文模型计算得出的最低安全间隔d_m=6 227 m，将该距离作为航空器对航向解脱的初始距离，航空器初始位置确定，模拟2架航空器初始飞行方向夹角从0°变化到180°的运动过程，计算每个角度运动过程中的最大侧向碰撞风险变化如图3所示。可以看出，随着2架航空器初始飞行方向夹角变大，2架航空器间侧向碰撞风险也随之变大。当2架航空器相向飞行，即初始飞行夹角为180°时，侧向碰撞风险最大为3.90×10^{-9}次/飞行小时。

表1 最低安全间隔计算模型所需参数值

Table 1 Required parameter values for the minimum safety interval calculation model

航空器	λ_x/m	λ_y/m	λ_z/m	v/(m·s^{-1})	β/(°)	σ_{cns}	σ_w	μ_w	N_p
A380	72.8	79.8	24.1	251	20	64.5	10.8	60.3	16
B737-800	39.5	35.8	12.5	230	20	64.5	11.5	60.3	16

由图3可知，相同初始距离的条件下，双机相向避撞飞行过程中，最大侧向碰撞风险大于其他场景，所需要的安全间隔最大，验证了本文模型适用于其他的飞行场景，所求得的最低安全间隔满足其他情况的飞行安全需求。

为验证模型的安全性和有效性，通过数据模拟的方法，利用蒙特卡罗算法对该航空器对的运动过程进行仿真。将2架航空器的初始距离设置为6 227 m，对其运动过程和碰撞情况进行模拟仿真，

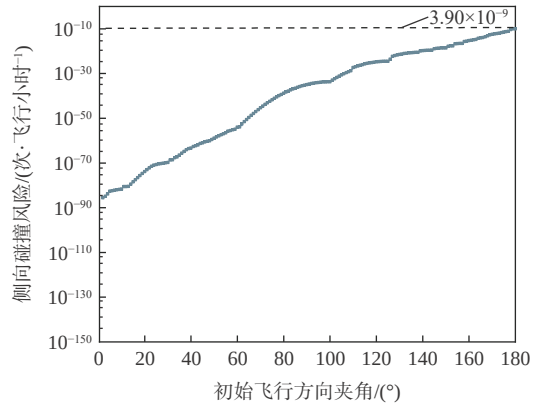


图3 侧向碰撞风险随初始飞行方向夹角变化

Fig. 3 Variation of lateral collision risk with initial flight direction angle

可以得出每1×10^8次仿真中2架航空器出现冲突的次数和平均侧向碰撞风险，如图4所示。在10×10^8次仿真中，出现的冲突次数为1，2架航空器间的侧向碰撞风险概率始终满足小于5×10^9次/飞行小时的飞行安全目标要求，大量的模拟运行有助于提高统计量显著性并提高置信度，验证了最低飞行安全间隔模型的安全性。

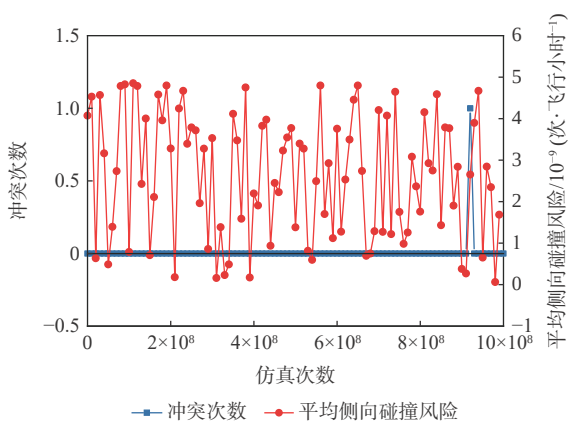


图4 蒙特卡罗仿真

Fig. 4 Monte Carlo simulation

3.2 侧向飞行安全间隔模型应用分析

根据ICAO按照航空器几何尺寸进行分类的规定，中国常见民用航空器被分为等级A～F。基于本文模型，按照民用航空器等级A～F的分类标准，计算得到6个等级的航空器之间的安全间隔标准，如表2所示。计算方法为：选取6个分类等级每个等级中碰撞球半径最大的航空器机型，从等级A到等级F依次为伊尔14、雅克42、空客300、空客330-300、波音747-8和空客380；将这6个机型的性能参数代入本文模型中，其他参数同表1。

国内常见不同性能民用航空器之间的侧向安全间隔标准可参照表2，进一步扩大了本文模型的应用范围，提高了参考价值。

表 2　6 个等级航空器间侧向安全间隔标准

Table 2　Standards for lateral safety interval between six classes of aircraft

航空器等级	侧向安全间隔/m					
	A	B	C	D	E	F
A	2 582	3 779	3 841	3 873	3 930	4 028
B	3 779	4 836	5 229	5 445	5 580	5 643
C	3 841	5 229	5 589	5 790	5 832	6 186
D	3 873	5 445	5 790	5 870	5 995	6 223
E	3 930	5 580	5 832	5 995	6 103	6 252
F	4 028	5 643	6 186	6 223	6 252	6 365

3.3　网络性能对比

假定本文 TBO 模式的空域中有包括民用飞机和通用飞机 5 种不同机型航空器, 查阅资料获取 5 种航空器大小、巡航速度等性能参数, 其他相关参数同表 1, 求算出对应的最小安全间隔, 如表 3 所示。可以看出, 航空器碰撞半径越大, 续航速度越快, 所需的最小安全间隔越大, 通用飞机由于体型小、速度慢, 所需的水平安全间隔较小。同时, 不同的航空器对之间的最小水平安全间隔不同, 同一机型航空器 A380 和 B737-800、A320、Cessna172、DA40 的最小安全间隔不同, 且均小于目前采用的国际水平安全间隔标准 10 km。

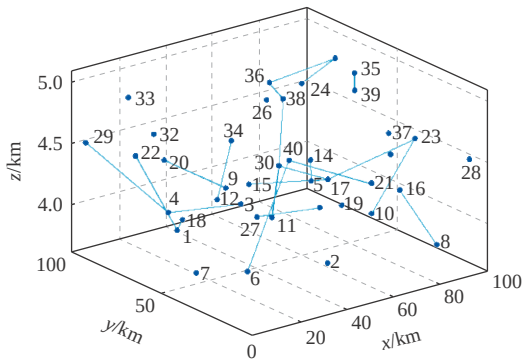
表 3　不同机型组的最小安全间隔

Table 3　Minimum safety distance for different model groups

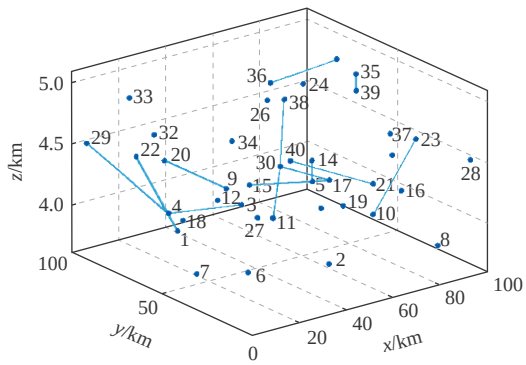
航空器机型	最小水平安全间隔/m				
	A380	B737-800	A320	Cessna172	DA40
A380	6 365	6 227	6 222	4 209	4 361
B737-800	6 227	6 025	6 017	3 884	4 032
A320	6 222	6 017	6 006	6 161	3 875
Cessna172	4 209	3 884	3 875	1 571	1 713
DA40	4 361	4 032	4 024	1 713	1 856

为分析本文模型的性能与特性, 通过 MATLAB 模拟航空器运行的仿真场景, 构建 100 km×100 km×1 km 的三维空域, 为模拟 TBO 运行环境, 在 4 个高度层和高度层之间随机生成表 3 中 5 种不同机型航空器共 40 架。航空器速度设置在 600～800 km/h 范围内, 航向任意, 位置随机均匀分布。

由于潜在的飞行冲突对于 2 架航空器是无差别的, 网络中的连边也没有方向, 飞行冲突网络为无向网络。图 5(a)为基于速度障碍法所构建的冲突网络模型(简称冲突模型), 以 10 km 为航空器椭



(a) 长轴10 km保护区下的飞行冲突网络



(b) 精准保护区下的飞行冲突网络

图 5　网络性能对比

Fig. 5　Network performance comparison

球形保护区长轴所构建的冲突网络。图 5(b)为以表 3 中不同机型组的最小安全间隔为保护区长轴所构建的冲突网络。图中节点为航空器, 实线为冲突连边, 两两连边越粗, 表明网络边权越大, 冲突越紧迫。

对比图 5(a)和图 5(b)的连边数量, 长轴 10 km 保护区下的飞行冲突网络连边为 22, 精准保护区下的飞行冲突网络连边为 13, 相比降低了 40.9% 的网络连边。由于一组场景导致的数据变化有一定的偶然性, 另设置 200 km×200 km 空域内随机生成 10～70 架航空器进行对比重复实验。图 6 为重复

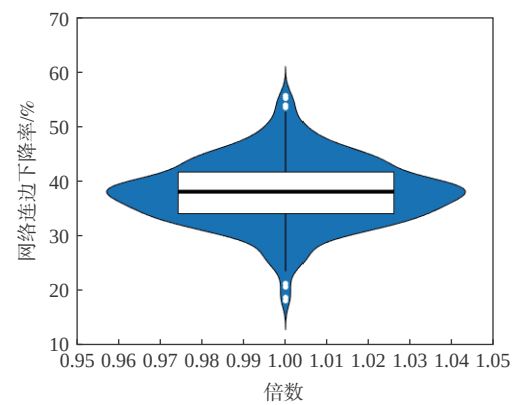


图 6　网络连边下降率分布

Fig. 6　Distribution of network connected edge decline rate

500次实验得出的网络连边下降率的小提琴图。图中，小提琴状区域为核密度曲线，白色矩形为25%、75%分位数，中间横线为中位数，白色点为离群值点，黑色竖线为去掉离群值点后点的上下限。可以看到，数据分布集中于30%～45%之间，总体平均值为38.1%，因此，认为TBO运行模式下，精准保护区下的冲突模型平均降低了38.1%的网络连边。

3.4 网络节点特征分析

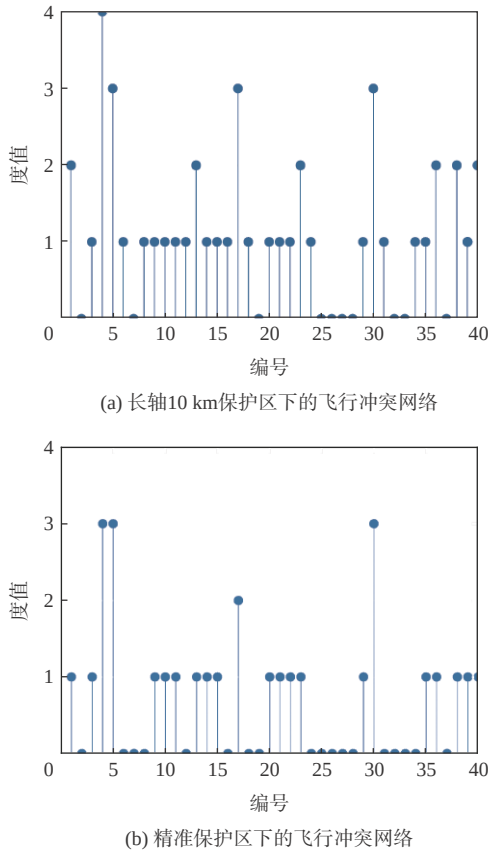
针对图5所构建的2种飞行冲突网络仿真场景，考虑复杂网络指标中反映节点特征的指标：节点的度和点强，对2种冲突网络进行定量分析。

定义1 节点度k_i。节点的度定义为网络中该节点和其他节点之间链接的数目，是反映网络中节点与节点连接情况的拓扑指标。在飞行冲突网络中，各航空器节点的度表示其存在的冲突数量的多少，度值越高，表示节点在网络中存在更多的邻居节点，与更多的行为单元存在关联，重要性相对更高。节点的度可表示为

$$k_i = \sum_{i=1}^{n} l_i \tag{13}$$

式中：k_i为节点度；l_i为节点的连边数目。

对2种保护区构建的飞行冲突网络模型节点的性质进行分析。对比分析2种网络中各个航空器对应的节点度值，如图7所示。



(a) 长轴10 km保护区下的飞行冲突网络

(b) 精准保护区下的飞行冲突网络

图7 网络节点度值分布的对比

Fig. 7 Comparison of the distribution of degree values of network nodes

由图7可知：①长轴10 km保护区下的飞行冲突网络中，孤立节点共10个：2、7、19、25、26、27、28、32、33、37；精准保护区下的飞行冲突网络中，孤立节点共18个：2、6、7、8、16、18、19、24、25、26、27、28、31、32、33、34、37、38。孤立节点数量的增多，一定程度上缓解了空中交通拥堵情况，减轻了管制员的负荷。②长轴10 km保护区下的飞行冲突网络中度值最高的是4节点，度值为4，而精准保护区下的飞行冲突网络中度值最高的为4、5、30节点，其潜在冲突数量为3，节点度的峰值降低了25%；经过保护区变化和速度障碍法判定后，27号节点节点度值降为0，与长轴10 km保护区下的飞行冲突网络得到的结果存在较大的偏差。

定义2 点强s_i。点强是与节点相连的边所对应的权值之和，该值能够反映飞机冲突态势的紧迫程度。节点的点强可表示为

$$s_i = \sum_{j=1}^{n} a_{ij} w_{ij} \tag{14}$$

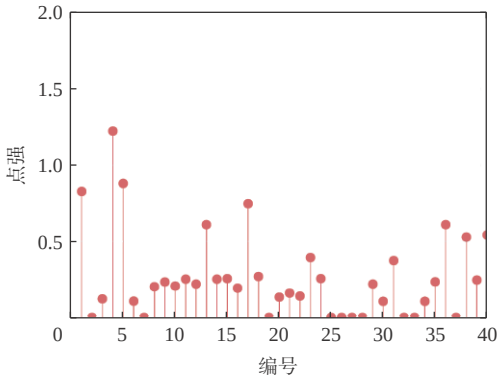
式中：a_{ij}为节点i和节点j的连接关系，若相连$a_{ij}=1$，否则$a_{ij}=0$；w_{ij}为节点v_i和节点v_j连边的权值。在飞行冲突网络中，点强反映了航空器节点潜在冲突整体的严重程度。

节点强度是由航空器节点各个冲突边的权重相加得到，可以较直观地反映航空器当前潜在的冲突情况。在相同场景下，基于不同长轴保护区形成的冲突网络点强分布如图8所示。结合图7可以看出，长轴10 km保护区下的飞行冲突网络中，4号节点的点强最高，其节点度为4，点强为1.24，对应的航空器整体潜在飞行冲突最强；相对比，精准保护区下的飞行冲突网络中，5号节点的点强最高，其节点度为3，点强为0.88。

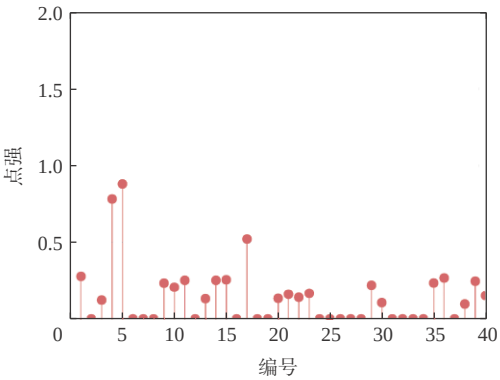
综上所述，无节点度值的节点也同样无点强，节点度值较大的节点的点强也相对较大，这是因为度值较大的节点，其冲突数量较多，冲突数量多使点强增加。相较于长轴10 km保护区下的飞行冲突网络，精准保护区下的飞行冲突网络节点度值和点强都有大幅度减少。因此，精准保护区形成的冲突网络在结构上更加简单，减少单个节点的冲突数量，进而提升TBO运行模式下整个航空网络的运行效率。

3.5 网络整体特性分析

定义3 集聚系数C。集聚系数表示网络中的



(a) 长轴10 km保护区下的飞行冲突网络



(b) 精准保护区下的飞行冲突网络

图 8　网络节点点强分布的对比

Fig. 8　Comparison of point strength distribution of network nodes

节点的邻居节点也互为邻居节点的比例, 反映了网络小集群的完美程度。加权聚集系数是普通聚集系数的加权表达, 定义为

$$c_{\mathrm{w}}(i)=\frac{1}{s_i(k_i-1)}\sum_{j,k}\frac{w_{ij}+w_{ik}}{2}a_{ij}a_{jk}a_{ik} \tag{15}$$

式中: a_{ij}、a_{jk}、a_{ik}为连通三元组的个数; j和k为节点i的邻居节点。

对于整个网络, 平均加权集聚系数可以描述网络整体的冲突汇聚程度, 用 C 表示:

$$C=\frac{1}{N}\sum_i c_i \tag{16}$$

式中: N 为网络中的节点数目。

定义 4　平均路径长度 D。飞行冲突网络中, 从节点v_i和节点v_j之间的路径不止一条, 选取距离最短的一条为最短路径, 该最短路径长度表示为

$$d_{ij}^{w}=1/\sum_{l\in L}\frac{1}{w_l} \tag{17}$$

式中: L 为节点v_i和节点v_j之间最短的一条道路; l 为道路 L 上的一条边; w_l为道路 l 的边权。

整个网络的平均路径长度可表示为

$$D=\frac{1}{N(N-1)}\sum_{i\neq j}d_{ij} \tag{18}$$

式中: d_{ij}为节点i和j之间的最短路径。

在飞行冲突网络中, 平均路径长度反映了网络中各个航空器节点间冲突的严重情况, 与空域的复杂性联系密切。

定义 5　网络效率 N_{E}。网络效率能直观反映整个网络的连通度, 2 个节点间的效率表示为彼此距离的倒数, 而整个网络的效率为每 2 对节点之间效率的平均值。在冲突网络中, 该值能够简易地表示出航空器冲突的扩散广度, 定义为

$$N_{\mathrm{E}}=\frac{1}{N(N+1)}\sum_{i\neq j}1/d_{ij} \tag{19}$$

网络效率 N_{E} 越高, 节点间的联系就越紧密, 网络也相对更复杂。

为计算和分析 2 种网络的整体特性指标, 在已有仿真场景基础上设置网络演化规则: 在 MATLAB 中构建仿真场景, 其中, 空域范围、航速、位置分布、航向均与上述仿真一致。令仿真空域随机生成 50 架 5 种机型的航空器, 设置航空器演化次数为 500, 每次演化时间间隔为 5 s(即每间隔 5 s 记录一次), 航速与航向在每次演化过程中有 30% 的几率发生改变; 在 200 km×200 km 空域范围内的航空器有 50% 的几率离开该空域, 而在所设定空域外的航空器有 50% 的几率进入该空域。分别计算按照规则每步演化后 2 种飞行冲突网络的集聚系数、平均路径长度和网络效率, 得到各指标时间序列中第 1 次到第 500 次演化的数值, 绘制成折线图, 如图 9 所示。

由图 9 可知, 2 种保护区组成的飞行冲突网络中, 各指标数值在演化过程中均在某个水平范围内波动。由长轴 10 km 保护区构成的飞行冲突网络中, 集聚系数以 0.24 均线上下波动, 最大值为 0.43, 平均路径长度以 0.86 均线上下波动, 最大值为 1.43, 最小值为 0.31, 网络效率以 1.71 均线上下波动, 最大值为 2.55, 最小值为 0.24; 精准保护区构成的飞行冲突网络中, 集聚系数以 0.07 均线上下波动, 最大值为 0.16; 平均路径长度以 0.15 均线上下波动, 最大值为 0.28, 最小值为 0.06, 网络效率以 0.25 均线上下波动, 最大值为 0.34, 最小值为 0.05。可以得到结论, 经过有效缩小航空器保护区范围, 为不同机型航空器划定精准的保护区大小, 能减少飞行冲突网络的复杂程度, 降低冲突的紧迫程度, 更加贴合 TBO 运行模式精细化、高效化、可预知的特点。

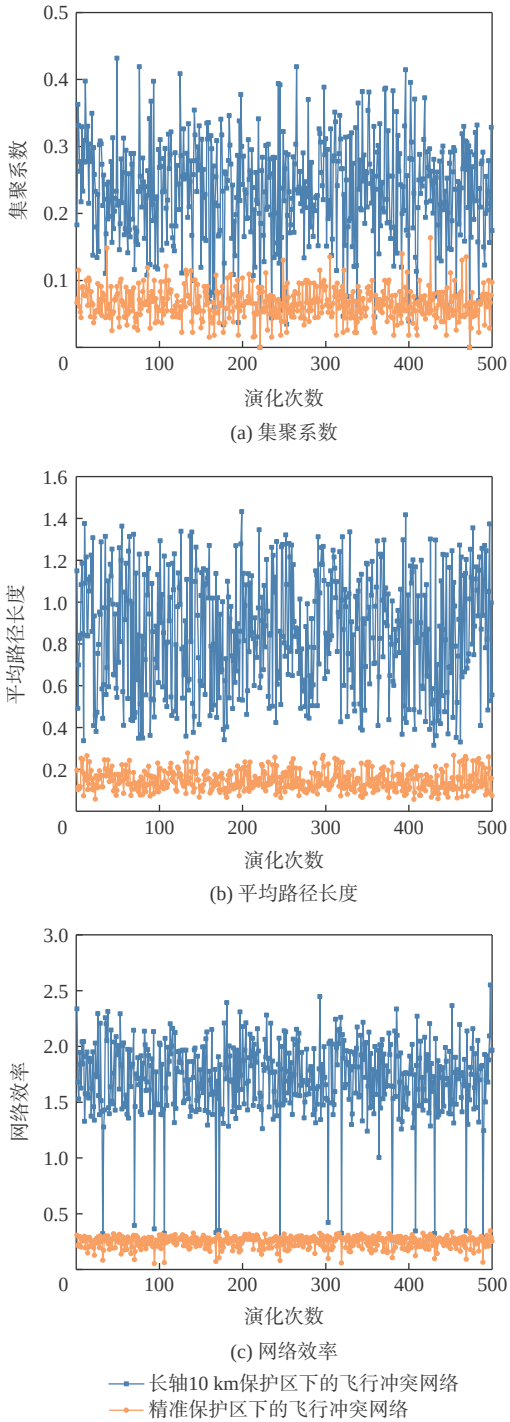


图 9 网络整体特性指标对比分析

Fig. 9 Comparative analysis of overall network characteristics indicators

4 实例仿真与分析

为进一步验证模型的合理性和有效性，选取厦门高崎国际机场周边管制区雷达态势作为实例进行仿真分析。厦门高崎国际机场位于福建省厦门市，是中国东南沿海重要的区域性航空枢纽，同时由于其位于中国的东南沿海，军用飞机也多将其作为任务飞行中的备降机场。以厦门高崎国际机场2021年某日14:10—14:40的飞行雷达数据作为测试样本，将该时段数据每隔10 min记录一次，记录的内容包括飞机的航速、航向和位置等，对每个时刻的空中情况通过MATLAB进行冲突网络建模，如图10所示。图中，实线表示扇区边界，虚线表示进近管制区边界，不同的实心形状表示不同机型的航空器，虚线内方块表示厦门高崎国际机场。

由图10可知，在30 min内，空情呈平稳态势

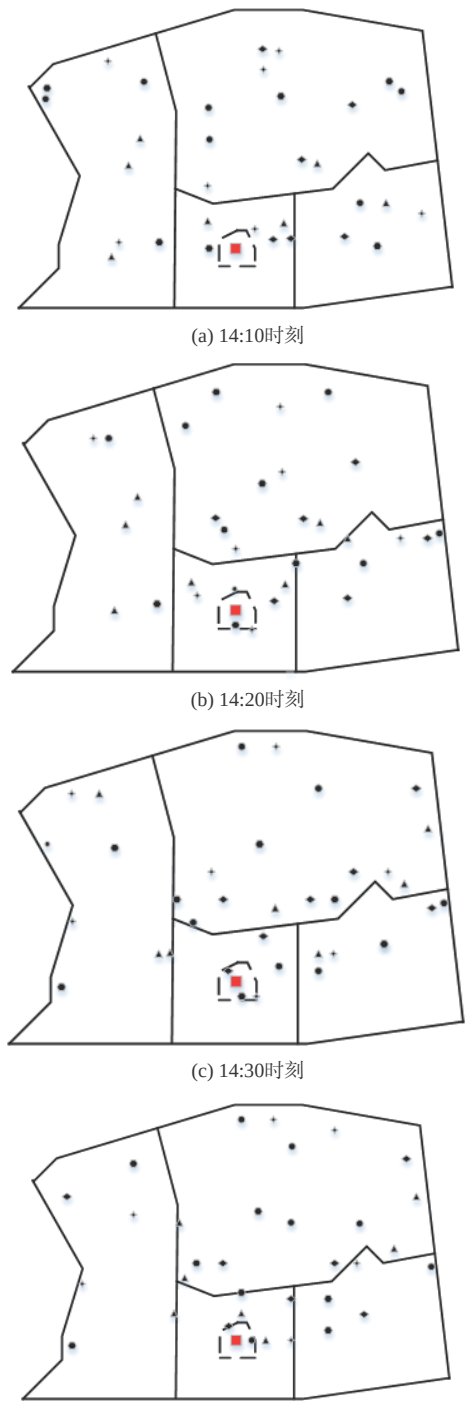


图 10 不同时刻空中运行态势

Fig. 10 Airborne operation situation at different moments

变化, 在该场景下验证本文模型的有效性。针对每个时刻的空中运行态势, 分别利用现行标准间隔形成的保护区和本文提出的不同机型基准保护区判定航空期间的冲突关系, 得到同一时刻 2 个不同的飞行冲突网络, 分别对长轴 10 km 保护区下的飞行冲突网络和精准保护区下的飞行冲突网络的网络指标进行分析, 得到的网络整体性数据如表 4 所示。

表 4　不同时刻的网络指标值

Table 4　Values of network indicators at different moments

时刻	集聚系数		平均路径长度		网络效率	
	长轴10 km保护区下的飞行冲突网络	精准保护区下的飞行冲突网络	长轴10 km保护区下的飞行冲突网络	精准保护区下的飞行冲突网络	长轴10 km保护区下的飞行冲突网络	精准保护区下的飞行冲突网络
12:10	0.437 8	0.268 6	0.781 4	0.257 8	1.644 9	0.195 7
12:20	0.447 4	0.279 3	0.683 2	0.227 9	1.739 0	0.218 9
12:30	0.466 5	0.296 9	0.555 8	0.236 0	1.320 4	0.228 6
12:40	0.408 8	0.232 9	0.651 8	0.178 8	1.438 6	0.231 0

由图 10 和表 4 可知, 4 个连续时刻的空情变化不大, 基于空情构建的飞行冲突网络拓扑结构和网络指标随时间也波动较小。长轴 10 km 保护区得到的飞行冲突网络中, 聚集系数在 0.43 左右, 平均路径长度在 0.67 左右, 网络效率在 1.52 左右。精准保护区得到的飞行冲突网络中, 聚集系数在 0.43 上下, 平均路径长度在 0.67 上下, 网络效率在 0.22 上下。对比厦门高崎国际机场终端区空域仿真得到的 2 种飞行冲突网络的网络指标可以看出, 在现行基于空域扇区运行的模式下, 精准保护区模型能根据航空器性能采取不同的间隔标准, 以此为基础构建的飞行冲突网络冲突连边数量和权值减小, 复杂程度降低, 从而可以减轻管制员工作负荷, 能提高空域的运行效率。

5　结　论

1) 航空器的临近关系是管制员判断潜在飞行冲突和空域复杂情况的重要依据, 本文模型改进了使用通信关系和统一间隔标准判断冲突关系带来的空域资源浪费, 更适用于未来精细化的管理模式。

2) 本文建立不同机型侧向飞行安全间隔模型, 使用蒙特卡罗等方法验证了模型的安全性和有效性, 以此为航空器划设精准保护区, 基于速度障碍法构建飞行冲突网络, 分析网络的拓扑指标衡量空中复杂性。

3) 相较于传统的建模方法, 本文根据不同航空器性能得到精准的保护区模型, 以此判定航空器间的冲突关系, 能够减少 38.1% 的飞行冲突网络连边数量, 降低空域的复杂程度, 同时缩小了水平安全间隔标准, 提高了空域的运行效率, 为以 TBO 为核心的未来空中交通管理提供了理论支撑。

参考文献 (References)

[1] International Civil Aviation Organization. Global air traffic management operational concept: Doc 9854 AN/458[R]. Montreal: International Civil Aviation Organization, 2005.

[2] BENTRUP L. Free routing airspace in Europe implementation concepts and benefits for airspace users[J]. International Conference on Research in Air Transportation, 2016: 1-32.

[3] 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).

[4] 赵嶷飞, 王梦琦. 空中交通工程学理论内涵与关键科学技术[J]. 航空学报, 2022, 43(12): 026537.
ZHAO Y F, WANG M Q. Important theories and critical scientific technology of air traffic engineering[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(12): 026537(in Chinese).

[5] SCHMIDT D K. On modeling ATC work load and sector capacity [J]. Journal of Aircraft, 1976, 13(7): 531-537.

[6] DELAHAYE D, PUECHMOREL S, HANSMAN J, et al. Air traffic complexity map based on non linear dynamical systems[J]. Air Traffic Control Quarterly, 2004, 12(4): 367-388.

[7] 张进, 胡明华, 张晨, 等. 空域复杂性建模[J]. 南京航空航天大学学报, 2010, 42(4): 454-460.
ZHANG J, HU M H, ZHANG C, et al. Airspace complexity modeling[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2010, 42(4): 454-460(in Chinese).

[8] 叶博嘉, 胡明华, 张晨, 等. 基于交通结构的空中交通复杂性建模[J]. 交通运输系统工程与信息, 2012, 12(1): 166-172.
YE B J, HU M H, ZHANG C, et al. Traffic structure-based air traffic complexity modeling[J]. Journal of Transportation Systems Engineering and Information Technology, 2012, 12(1): 166-172(in Chinese).

[9] 王红勇, 赵嶷飞, 温瑞英. 基于复杂网络的空中交通复杂性度量方法[J]. 系统工程, 2014, 32(3): 112-118.
WANG H Y, ZHAO Y F, WEN R Y. Air traffic complexity metrics based on complex networks[J]. Systems Engineering, 2014, 32(3): 112-118(in Chinese).

[10] WANG H Y, WEN R Y, ZHAO Y F. Analysis of topological characteristics in air traffic situation networks[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace En-

gineering, 2015, 229(13): 2497-2505.

[11] WANG H Y, SONG Z Q, WEN R Y, et al. Study on evolution characteristics of air traffic situation complexity based on complex network theory[J]. Aerospace Science and Technology, 2016, 58: 518-528.

[12] JIANG X R, WEN X X, WU M G, et al. A complex network analysis approach for identifying air traffic congestion based on independent component analysis[J]. Physica A: Statistical Mechanics and Its Applications, 2019, 523: 364-381.

[13] 吴明功, 王泽坤, 甘旭升, 等. 基于复杂网络理论的关键飞行冲突点识别[J]. 西北工业大学学报, 2020, 38(2): 279-287.
WU M G, WANG Z K, GAN X S, et al. Identification of key flight conflict nodes based on complex network theory[J]. Journal of Northwestern Polytechnical University, 2020, 38(2): 279-287(in Chinese).

[14] 吴明功, 叶泽龙, 温祥西, 等. 基于复杂网络的空中交通复杂性识别方法[J]. 北京航空航天大学学报, 2020, 46(5): 839-850.
WU M G, YE Z L, WEN X X, et al. Air traffic complexity recognition method based on complex networks[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(5): 839-850(in Chinese).

[15] 毕可心, 吴明功, 张文斌, 等. 基于速度障碍法的飞行冲突网络建模与分析[J]. 系统工程与电子技术, 2021, 43(8): 2163-2173.
BI K X, WU M G, ZHANG W B, et al. Modeling and analysis of flight conflict network based on velocity obstacle method[J]. Systems Engineering and Electronics, 2021, 43(8): 2163-2173(in Chinese).

[16] ERZBERGER H, PAIELLI R, ISAACSON D. Conflict detection and resolution in the presence of prediction error[C]//Proceeding of the 1st USA/Europe Air Traffic Management R&D Seminar. Washington, D.C.: NASA, 1997: 1-15.

[17] 中华人民共和国国务院. 中华人民共和国飞行基本规则[S]. 北京: 中华人民共和国国务院, 2007.
The State Council the People's Republic of China. Flight basic rules of the People's Republic of China[S]. Beijing: The State Council the People's Republic of China, 2007(in Chinese).

[18] MOEK G, LUTZ E, MOSBERG W. Risk assessment of RNP 10 and RVSM in the south Atlantic flight identification regions: 21401-7465[R]. Annapolis: ARINC, 2001.

[19] 王欣, 徐肖豪. 空中飞机侧向间隔标准的初步研究[J]. 中国民航学院学报 (综合版), 2001, 19(1): 1-5.
WANG X, XU X H. Study of lateral separation in air traffic control[J]. Journal of Civil Aviation University of China, 2001, 19(1): 1-5(in Chinese).

[20] 张兆宁, 时瑞军. 自由飞行下基于冲突解脱的碰撞风险模型研究[J]. 安全与环境工程, 2016, 23(2): 157-161.
ZHANG Z N, SHI R J. Study on free flight collision risk model based on conflict resolution[J]. Safety and Environmental Engineering, 2016, 23(2): 157-161(in Chinese).

[21] International Givil Aviation Organization. Convention on international civil aviation: Doc T300[S]. Montreal: International Givil Aviation Organization, 2006.

Complex network-based air traffic complexity analysis in TBO

PENG Yating[1, 2], WEN Xiangxi[1, 2], WU Minggong[1, 2, *], YANG Zhiwei[3], WANG Nan[3]

(1. Air Traffic Control and Navigation College, Air Force Engineering University, Xi'an 710051, China;
2. National Key Laboratory of Air Traffic Collision Prevention, Xi'an 710051, China;
3. Unit 93160 of the People's Liberation Army of China, Beijing 100800, China)

Abstract: Since the complex network constructed based on the unified spacing standard does not take into account the differences in aircraft type operation, it cannot meet the refined requirements of air traffic complexity analysis under trajectory based operation (TBO). A complex network-based air traffic complexity analysis model is suggested as a solution to this issue in order to differentiate between various aircraft types. In order to create an aircraft precision protection zone and optimize the foundation for identifying the aircraft related edges in the flight conflict network, a lateral flight safety interval calculation model is first developed for various aircraft types. Based on the information on aircraft heading and speed, the flight conflict judgment focuses on different performances and status of aircraft, so that the flight conflict network can be closer to the operation mode of TBO. The findings demonstrate that the model can improve the operational efficiency of the airspace, decrease the complexity of the airspace, lessen the workload of controllers, improve the horizontal separation criteria between aircraft, and give aircraft more freedom to select the best course on their own than the previous flight conflict network.

Keywords: flight conflicts; complex networks; trajectory based operations; horizontal separation criteria; protected area model

Received: 2023-05-08; **Accepted**: 2023-06-25; **Published Online**: 2023-07-25 15:39

URL: link.cnki.net/urlid/11.2625.V.20230725.1328.001

Foundation items: National Natural Science Foundation of China (71801221); National Social Science Fund of China (22XGL001)

* **Corresponding author**. E-mail: wuminggong@sohu.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0248

小腔对排气活门快速调压能力的影响

吴豪，刘猛*，王浚

(北京航空航天大学 航空科学与工程学院，北京 100191)

摘 要：飞机座舱进气流量冲击情况下，气动式座舱压力调节系统(PCPCS)可能由于动态调节速度不足而导致座舱压力尖峰，进而出现“压耳”问题。基于此，对PCPCS的工作原理开展分析，指出小腔可能对排气活门的快速运动产生阻碍。建立PCPCS的动力学模型，描述排气活门中控制腔（A腔）和小腔（B腔）的工作状态；开展PCPCS在座舱进气流量冲击下的动态工作仿真，定量展示排气活门的动态工作特征和活门芯的受力变化；讨论小腔定径孔直径、小腔顶部直径、小膜片直径对排气活门应对座舱进气流量冲击的影响。仿真结果表明：适当增大小腔定径孔直径有助于提升排气活门响应速度，小腔顶部直径对排气活门运动速度没有显著影响，减小小膜片直径对排气活门响应速度的提升较为显著。

关 键 词：飞行器环境控制；飞机座舱压力调节；压耳；气动式座舱压力调节系统；排气活门

中图分类号：V245.4

文献标志码：A **文章编号**：1001-5965(2025)04-1245-10

气动式座舱压力调节系统(pneumatic cabin pressure control system, PCPCS)的作用是在飞机飞行过程中自动调节座舱内的空气压力及其变化率，在飞机座舱内部营造一个适宜乘员工作生存的压力环境，保障机上人员的生命安全和舒适度[1]。

PCPCS以空气为信息传递介质，通过感压元件和弹性元件实现力的传递，达到调节座舱排气的目的。该类压力调节系统于20世纪50年代诞生[2]，经历多年迭代[3-5]，于20世纪70年代成型[6]，伴随而生的还有电动式的改进[7-8]和功能性的扩展[9-10]。由于该类压力调节系统结构成熟且不受电磁干扰，受到了广泛的青睐[11-13]。其排气活门采用双模片结构设计，不仅在座舱排气方面具备优良的性能，也使其能够轻松应对负压补气的要求，因此，常用来当作安全排气活门使用[3,9-10]。

中国在座舱压力调节方面起步较晚，于20世纪70年代开始开展PCPCS相关的研究工作[14-15]，所采用的设备为经典的无导杆式的压力调节系统[16-17]，对该气动式压力调节系统的研究也逐渐深入展开。徐杨禾等[16]建立了较为详尽的PCPCS数学模型，并进行了线性化近似，开展了数值仿真研究工作，但受限于计算力和实验条件，数学模型进行了简化处理，主要结果更偏重于定性分析。虽然已过去几十年，但关于PCPCS的分析仍然在持续进行。霍昱旭等[18]使用MATLAB/Simulink平台构建了一套参数修改方便、结构易于调整的动态模型，对系统进行了静态/动态性能的仿真分析。魏晓永[19]采用CFD方法对MA700飞机的PCPCS开展了流动模拟，分析其内部压力分布和速度分布。郑新华[20]建立了较为完整的PCPCS组件模型，采用传递函数的方法研究了系统的动态特性，并使用基于集员描述函数法的稳定性分析方法来分析非线性的压力波动问题。上述分析从不同层面研究了PCPCS的动态特性，但对该系统运行的动态建模仍然以20世纪70年代的简化数学模型为基础，建模分析普遍理想化。为此，Zheng等[21]建立了较为准确的

收稿日期：2023-05-15；**录用日期**：2023-08-14；**网络出版时间**：2023-08-22 10:12
网络出版地址：link.cnki.net/urlid/11.2625.V.20230821.1508.001
*通信作者. E-mail：liumeng@buaa.edu.cn

引用格式：吴豪，刘猛，王浚. 小腔对排气活门快速调压能力的影响[J]. 北京航空航天大学学报，2025，51（4）：1245-1254.
WU H，LIU M，WANG J. Effect of balance chamber on rapid pressure regulation ability of outflow valve[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1245-1254（in Chinese）.

压力调节系统数学模型，考虑了外界环境对系统的干扰，指出流量冲击对座舱压力的波动最为显著。虽然科研人员对压力调节系统的动态特性研究已持续了几十年，但对于核心执行部件，即排气活门的腔体（尤其是排气活门底部的小腔）容积的影响普遍没有过多关注。

由于战斗机的更新迭代，其飞行机动性、发动机引气的快速性也逐渐增强，座舱压力调节系统的动态响应能力随着战斗机的发展而受到考验。飞行员的“压耳”问题始终是困扰战斗机研究PCPCS的关键问题，从20世纪50年代开始至今，国内外不断有文献报道记载座舱“压力尖峰”导致的飞行员耳气压伤的案例[22-25]，尤其是近年爆出的F-18和F-35战机飞行员反映的压耳问题[24,26]，使业内人员又提高了对座舱压力调节性能的重视。压耳问题不仅影响飞行员舒适性，严重时还会造成飞行员骨膜破裂，听力受损[27]。

“压力尖峰”的一大诱因就是短时间内座舱进排气流量不匹配，导致座舱压力快速波动。例如，发动机快速加力过程中，环控系统的发动机引起进量快速增加，座舱的进气量可能快速提升，而排气活门的响应速度较慢，导致座舱的排气量跟不上进气量的变化，此时的座舱压力就会发生快速变化，产生“压力尖峰”。提升排气活门的响应速度是控制“压力尖峰”的一种可行手段。分析排气活门的动态特性，找出影响其动态特性的关键设计参数是提升排气活门响应速度的关键。小腔作为排气活门底部的感压和施力环节，对排气活门动态特性存在影响。

本文旨在开展排气活门的动态建模，分析排气活门运动过程中各个环节对其运动的影响特征，使用仿真的方法研究小腔在排气活门工作时发挥的作用，并着重研究座舱进气流量冲击情况下小腔的物理特性对排气活门运动的影响，为排气活门和小腔的优化设计提供改进建议。

1 PCPCS工作原理及小腔的影响

1.1 PCPCS及排气活门的工作原理

经典的飞机PCPCS如图1所示，主要包括控制机构（余压机构、绝压机构）、排气活门及相关的连通管道。控制机构调节排气活门控制腔（A腔）的压力，改变活门开度，从而调节座舱排气量，达到调节座舱压力的作用。

排气活门作为PCPCS的核心执行机构，其动态性能直接影响座舱压力的动态特性，因此，需要对排气活门开展更深一步的讨论。

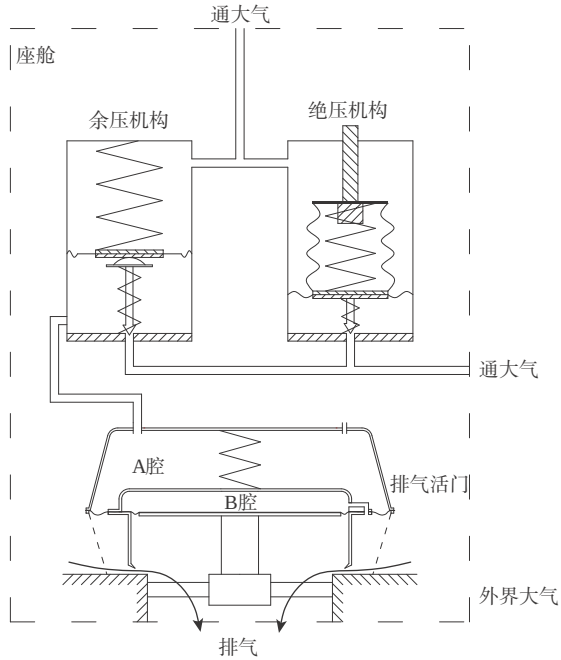


图1 飞机气动式座舱压力调节系统

Fig. 1 PCPCS of aircraft

图2为一种常用的排气活门。A腔是由排气活门的外壳、大膜片及活门芯顶部共同构成的容积腔。A腔有2个开孔，一个连通控制机构，另一个称为A腔定径孔，与座舱连通。在正常工作情况下，座舱内空气有一小部分通过A腔定径孔进入A腔，A腔内的空气也有一小部分通过控制机构流入外界大气，稳定情况下，A腔进排气平衡，形成恒定的A腔气压。活门芯的上部感受A腔压力的同时，还与弹簧相连感受弹簧弹力，活门芯外缘连接大膜片，内部覆盖了小膜片，小膜片与活门芯的顶部构成了一个封闭腔，通常称为小腔（本文称为B腔）。B腔有一个开孔，称为B腔定径孔，与座舱相连通。在内外压力相同时，小膜片收缩，靠近B腔底部；当飞机处于高空时，通常外部气压低于舱内气压，B腔膨胀，将小膜片压在基座挡板上。大小膜片通过感受上下的压差而给活门芯施加作用力。活门芯在A/B腔压差力、弹簧弹力、膜片压差力的

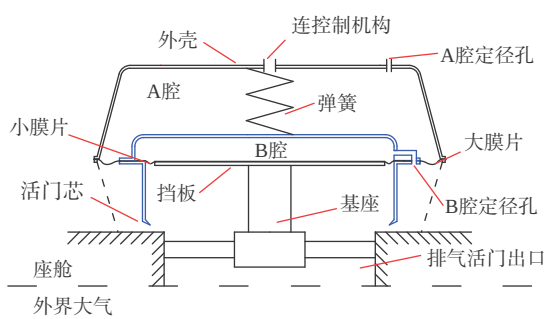


图2 排气活门

Fig. 2 Outflow valve

共同作用下运动, 实现排气活门开启/关闭的操作, 从而调整座舱的排气速度。排气活门的拆解图如图 3 所示。

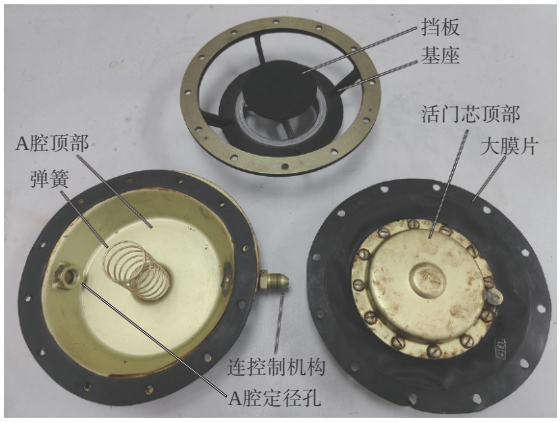


图 3　排气活门拆解图

Fig. 3　Disassembly of outflow valve

1.2　小腔的工作特征

活门芯底部如图 4 所示, 金属部分的活门芯与软质膜片相结合, 活门芯内部与底部膜片连接形成封闭的 B 腔, 通过 B 腔定径孔与座舱相连, 稳态时 B 腔气压与座舱气压相同, 小膜片收缩。当 B 腔内部气压高于外界气压时, 会将 B 腔撑起, 增大 B 腔内部容积。图 5 展示了充气后的 B 腔。充气后的 B 腔底部与基座的挡板接触, 露出 B 腔底部内外缘构成的环形小膜片。小膜片设计的作用是感受座舱与外界大气的压力差, 当座舱内气压低于外界大气压时(即座舱负压), 小膜片的压差力推动活门芯开启, 使外界空气流入座舱, 提升座舱内气压。

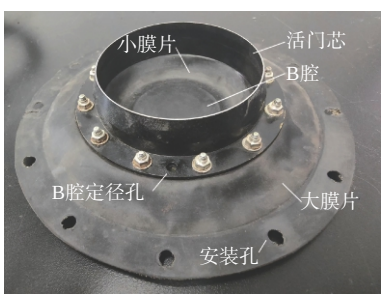


图 4　活门芯底部部分装配零件

Fig. 4　Part of assembly parts at bottom of poppet

B 腔作为防止座舱负压的功能性容积腔, 在排气活门运动过程中, 其内部气压可能影响到活门芯的受力。B 腔对活门芯的受力可能存在如下影响:

1) 稳态时, A 腔一端通过控制机构连接外部大气, 另一端通过 A 腔定径孔连通座舱, 气压会略低于座舱压力; B 腔通过 B 腔定径孔与座舱相连, 气压约等于座舱压力。A 腔气压略低于 B 腔, A 腔与

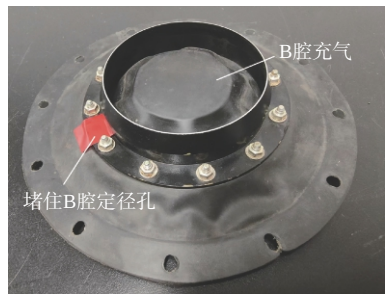


图 5　小腔充气示意图

Fig. 5　Balance chamber inflation

B 腔的压力差作用在活门芯上, 会对活门芯产生额外的向上作用力。

2) 活门芯向上运动时, A 腔压缩快速升压, B 腔会跟随活门芯的运动而向上运动, 但其容积及压力的变化取决于定径孔的净流量及其与 B 腔底部的压力差。A、B 这 2 个容积腔的压力变化可能会改变对活门芯向上的作用力, 但两腔的压力变化率的关系却较为复杂。对于 A 腔而言, 其压力的变化一方面与活门芯顶部表面积有关, 另一方面还受到 A 腔定径孔和连通控制机构的孔径的影响。对于 B 腔而言, 其压力的变化取决于: ①B 腔定径孔的净流量; ②B 腔与底部的压力差; ③B 腔高度与活门开度的关系。这种复杂的影响可能会产生复杂的压差力, 从而在动态过程中对活门开度变化产生复杂的影响。

3) 由于 B 腔压力通常高于外界, 小膜片的压差力通常对活门芯施加的是向下的拉力, 受力面通常为环形。但当活门芯快速运动时, 小膜片可能会脱离基座, 此时的受力面变为整个小膜片的底部圆面, 向下拉力增大, 阻碍活门芯开启。

4) 活门芯向下运动时, A 腔扩大, B 腔压缩, 2 个容积腔的压力变化可能会增大对活门芯向上的作用力, 阻碍活门芯向下关闭。阻碍效果的强度与 A、B 两腔的压差直接相关, 而压差变化的程度也同样由 A 腔、B 腔的定径孔决定。

由此可见, B 腔的存在虽然可以通过 A/B 腔的压力差提供活门芯向上的力, 但也会对排气活门产生向下的拖拽力。B 腔对活门芯运动的影响程度, 以及哪些物理参数会改变 B 腔对活门芯运动的影响, 需要开展定量分析。

2　本文方法

2.1　活门芯动力学分析

为研究 B 腔对活门芯运动的影响, 需要建立活门芯运动的动力学方程。在不考虑排气活门外界加速度变化的环境下, 活门开度的微分方程可以表示为

$$m_{\mathrm{P}}\frac{\mathrm{d}^2y_{\mathrm{P}}}{\mathrm{d}\tau^2}=F_{\mathrm{D1}}+F_{\mathrm{AB}}-F_{\mathrm{S}}-F_{\mathrm{D2}}-F_{\mathrm{V}} \tag{1}$$

式中：m_{P}为活门芯的质量；y_{P}为活门开度；τ为时间；F_{D1}为大膜片由于上下表面压差而施加给活门芯的向上的力；F_{AB}为A腔与B腔的压差作用在活门芯的向上的力；F_{S}为弹簧向下的推力；F_{D2}为小膜片由于上下表面压差而施加给活门芯的向下的力；F_{V}为活门芯在运动过程中受到的黏滞阻力。

大膜片施加的力可以表示为

$$F_{\mathrm{D1}}=(P_{\mathrm{D1}}-P_{\mathrm{A}})A_{\mathrm{D1}} \tag{2}$$

式中：P_{D1}为大膜片下表面的气压；P_{A}为A腔内的气压；A_{D1}为大膜片的有效受力面积。

A腔与B腔的压差力可以表示为

$$F_{\mathrm{AB}}=(P_{\mathrm{B}}-P_{\mathrm{A}})A_{\mathrm{P}} \tag{3}$$

式中：P_{B}为B腔内的气压；A_{P}为活门芯顶部的有效受力面积。

考虑弹簧质量及变形过程中受到的阻尼效果，假设弹簧的质量均匀分布，其推力可表示为

$$F_{\mathrm{S}}=m_{\mathrm{S}}\frac{\mathrm{d}^2y_{\mathrm{S}}}{\mathrm{d}\tau^2}+K_{\mathrm{S}}y_{\mathrm{P}}+F_{\mathrm{S0}}+f_{\mathrm{S}}\frac{\mathrm{d}y_{\mathrm{S}}}{\mathrm{d}\tau} \tag{4}$$

式中：m_{S}为弹簧的质量；y_{S}为弹簧质心的位移；K_{S}为弹簧的弹性系数；F_{S0}为弹簧安装时的预压缩力；f_{S}为弹簧变形时受到的黏滞阻力系数。

小膜片施加的力F_{D2}可以表示为

$$F_{\mathrm{D2}}=(P_{\mathrm{B}}-P_{\mathrm{D2}})A_{\mathrm{D2}} \tag{5}$$

式中：P_{D2}为小膜片下表面的气压；A_{D2}为小膜片的有效受力面积，当B腔高度与活门开度一致时，B腔与基座挡板相贴合，A_{D2}为环形受力面，但当B腔高度小于活门开度时，B腔脱离基座挡板，A_{D2}为圆形受力面。

活门芯运动的黏滞阻力F_{V}可以表示为

$$F_{\mathrm{V}}=f_{\mathrm{P}}\frac{\mathrm{d}y_{\mathrm{P}}}{\mathrm{d}\tau} \tag{6}$$

式中：f_{P}为活门芯运动的黏滞阻力系数。

式(1)～式(6)组成了排气活门的活门芯的动力学方程，可以描述活门芯在指定初始开度、外部压力环境条件下的开度变化情况。

式(2)中，大膜片下表面的气压P_{D1}通常近似为座舱内的气压P_{C}；式(5)中，小膜片下表面的气压P_{D2}通常近似为外界大气压P_{atm}[28]。但文献[21]中指出，在不同活门开度y_{P}和不同座舱余压(P_{C}−P_{atm})的情况下，P_{D1}与P_{C}并不能始终近似为相等，P_{D2}与P_{atm}也存在显著的差异。利用文献[21]的方法，使用CFD工具来分析P_{D1}和P_{D2}在不同活门开度、外界气压、舱内外压差下的变化规律，得到P_{D1}和P_{D2}与上述参数的拟合关系，利用相似准则就可以确定相同形状不同尺寸的排气活门在不同工况下的P_{D1}和P_{D2}。

此外，文献[21]也给出了在不同活门开度y_{P}、不同大气压P_{atm}、不同座舱余压(P_{C}−P_{atm})下流经排气活门的空气流量M_{out}，采用数据拟合的方式可以得到排气活门排气流量的表达式。已知座舱进气流量M_{in}后，就可以确定座舱气压的变化情况：

$$\frac{\mathrm{d}P_{\mathrm{C}}}{\mathrm{d}\tau}=\frac{RT_{\mathrm{C}}}{V_{\mathrm{C}}}(M_{\mathrm{in}}-M_{\mathrm{out}}-M_{\mathrm{leak}}) \tag{7}$$

式中：T_{C}和V_{C}分别为座舱内的空气温度和座舱容积；R为空气的气体常数；M_{leak}为座舱空气的泄漏量。

2.2 控制腔内气压

影响B腔内气压的主要因素包括：①A腔通过控制机构向外界大气的净排气量；②座舱通过A腔定径孔进入A腔的空气量；③A腔容积随活门开度的变化。因此，A腔内气压P_{A}变化的动态方程可以写为

$$\frac{V_{\mathrm{A}}}{RT_{\mathrm{A}}}\cdot\frac{\mathrm{d}P_{\mathrm{A}}}{\mathrm{d}\tau}+\frac{P_{\mathrm{A}}}{RT_{\mathrm{A}}}\cdot\frac{\mathrm{d}V_{\mathrm{A}}}{\mathrm{d}\tau}=M_{\mathrm{d}}-M_{\mathrm{c}} \tag{8}$$

式中：T_{A}和V_{A}分别为A腔内的温度和A腔的容积；$M_{\mathrm{A_d}}$为通过A腔定径孔流入A腔的空气流量；$M_{\mathrm{A_c}}$为A腔通过控制机构向外界大气的空气流量。流量$M_{\mathrm{A_d}}$和$M_{\mathrm{A_c}}$可利用常规定熵绝热喷管的流动方程[29]来计算。

式(8)中，等号左边第2项描述的是A腔容积的变化引起的当量流量。要确定该当量流量，需要知道A腔容积与活门开度y_{P}的关系。文献[21]使用几何方法确定了活门开度y_{P}与A腔容积的关系，根据活门开度的变化计算A腔容积的变化，进而利用式(8)计算A腔气压的变化情况。

2.3 小腔内气压

影响B腔内气压的主要因素包括：①座舱通过B腔定径孔进入B腔的空气量；②B腔容积的变化。

与A腔的计算类似，B腔内气压P_{B}变化的动态方程可以写为

$$\frac{V_{\mathrm{B}}}{RT_{\mathrm{B}}}\cdot\frac{\mathrm{d}P_{\mathrm{B}}}{\mathrm{d}\tau}+\frac{P_{\mathrm{B}}}{RT_{\mathrm{B}}}\cdot\frac{\mathrm{d}V_{\mathrm{B}}}{\mathrm{d}\tau}=M_{\mathrm{d}} \tag{9}$$

式中：T_{B}和V_{B}分别为B腔内的温度和B腔的容积；$M_{\mathrm{B_d}}$为通过B腔定径孔流入B腔的空气流量。

流量M_{d}的计算同样可以采用常规定熵绝热喷管的流动方程，对于B腔容积V_{B}的变化则较为复杂。

如果B腔膨胀，小膜片紧贴基座挡板，则B腔的高度与活门开度一致，此时容积的计算与A腔容积计算的方法相同，根据活门开度y_{P}来求解B腔容

积 V_B。但 B 腔有可能在一些特定情况下没有膨胀或没有紧贴基座挡板, 此时 B 腔的容积计算就需要额外考虑。

影响 B 腔容积的主要因素除了自身压力 P_B 以外, 还有 3 个外界因素: 小膜片下部气压 P_{D2}、座舱压力 P_C 及活门开度 y_P。因此, 计算 B 腔容积变化时, 需要同时判断这 3 个外界因素对 B 腔的影响, 从而确定 B 腔容积的计算遵循哪个特定过程。

1) B 腔容积内气压 P_B 与小膜片下部气压 P_{D2} 的大小关系。

如果 $P_B \geqslant P_{D2}$, 则 B 腔趋于膨胀; 如果 $P_B < P_{D2}$, 则 B 腔趋于收缩。没有其他约束的情况下, B 腔膨胀、收缩的速度接近于声波的传递速度, 此时的 B 腔容积变化满足:

$$\frac{\mathrm{d}V_B}{\mathrm{d}\tau} = \frac{\mathrm{d}V_B}{\mathrm{d}y_B} \cdot \frac{\mathrm{d}y_B}{\mathrm{d}\tau} \tag{10}$$

$$\frac{\mathrm{d}y_B}{\mathrm{d}\tau} = c_{\mathrm{local}} \tag{11}$$

式中: y_B 为 B 腔的高度; c_{local} 为当地声速。

2) B 腔容积内气压 P_B 与座舱压力 P_C 的大小关系。

如果 $P_B \geqslant P_C$, 则 B 腔为放气趋势; 如果 $P_B < P_C$, 则 B 腔为充气趋势。没有外界约束时, 充放气时 B 腔的气压恒定, 此时 B 腔容积变化的计算满足:

$$\frac{\mathrm{d}V_B}{\mathrm{d}\tau} = M_{\mathrm{d}} \frac{RT_B}{P_B} \tag{12}$$

当受到限位约束时, 充放气的容积不变, 即

$$\frac{\mathrm{d}V_B}{\mathrm{d}\tau} = 0 \tag{13}$$

3) B 腔高度 y_B 与活门开度 y_P 的关系。

如果 $y_B = y_P$, 则 B 腔无法再膨胀, 排气活门压缩 B 腔导致的 B 腔容积变化可以写为

$$\frac{\mathrm{d}V_B}{\mathrm{d}\tau} = \frac{\mathrm{d}V_B}{\mathrm{d}y_B} \cdot \frac{\mathrm{d}y_P}{\mathrm{d}\tau} \tag{14}$$

如果 $y_B < y_P$, 则 B 腔可以自由收缩或膨胀, 速度接近于声波的传递速度。

根据上述判断过程来计算 B 腔容积的变化, 再结合 M_{d} 的计算公式, 就可以代入式 (9) 求得 B 腔气压的变化情况。

2.4　计算流程

根据上述活门芯受力、座舱气压、A 腔压力、B 腔压力的动态分析, 可以计算排气活门在不同工况下的动态特性及座舱压力变化的情况, 具体计算流程如图 6 所示。

需要指出, 图 6 所示的计算流程针对的是飞机所处海拔高度固定、控制机构状态不变的情况下, 座舱进气流量发生变化时的排气活门动态变化及座舱气压变化的计算过程。飞机高度发生变化时, 外界气压的变化先经过控制机构传递给 A 腔, 再改变活门开度, 才会对座舱内气压、B 腔气压产生影响。而当座舱进气流量发生变化时, 座舱气压直接发生改变, 随后才是进一步改变活门开度和 A 腔、B 腔的气压, 这种动态冲击最直接, 对座舱压力的影响最显著, 最能直接反映排气活门的动态特性。因此, 分析座舱进气流量冲击下的压力调节系统动态特性对于评价排气活门的动态设计、描述压力调节系统应对压力冲击的能力最为有效。

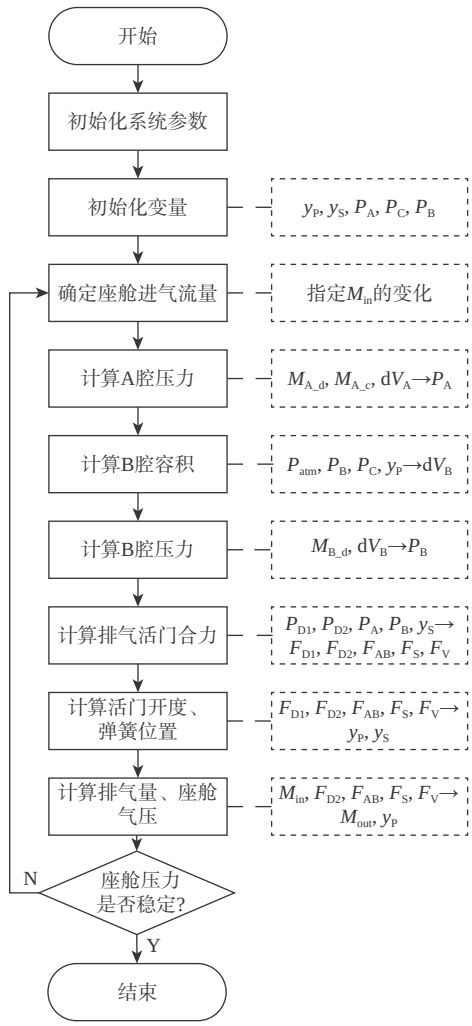


图 6　指定座舱进气流量时的计算流程

Fig. 6　Calculation process when specifying cabin inflow

3　仿真与结果

利用第 2 节的计算方法可以求解座舱进气快速变化情况下的 PCPCS、座舱压力的响应情况。为验证文本方法的准确性, 选取某型号压调器试验数据[30] 作为对比, 开展相同工况的仿真; 开展 B 腔设计参数对座舱压力影响的仿真计算, 使用固定变量法分析不同 B 腔设计参数对座舱压力变化的影响特性, 给出可以提升排气活门应对流量冲击的响应速度的改进建议。

3.1 验证性仿真

文献 [30] 给出了某座舱 PCPCS 的地面流量冲击试验数据。飞机座舱供气流量在约 4 s 内由 100 kg/h 增至 900 kg/h, 座舱压力随即上升, 待 PCPCS 工作稳定后座舱压力变化率恢复平稳。

针对相同工况开展数值仿真。图 7 展示了座舱进气流量冲击的曲线, 图 8 为对应的压力变化率曲线。试验中的传感器采样周期为 1 s, 而仿真数据相对较为连续, 因此, 测量信号与仿真数据存在一定的相位差。

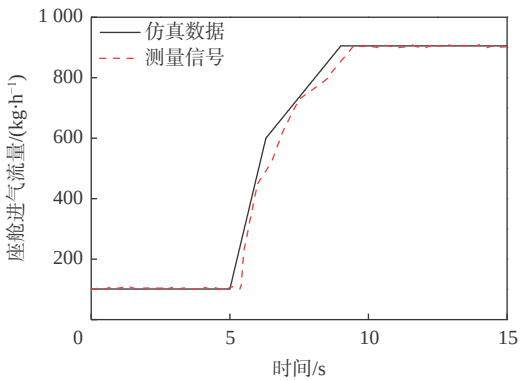


图 7 座舱进气流量冲击

Fig. 7 Cabin inflow impact

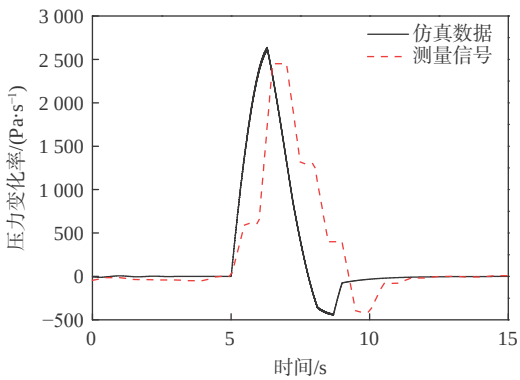


图 8 座舱进气流量冲击下的座舱压力变化率

Fig. 8 Change rate of cabin pressure under cabin inflow impact

从图 7 和图 8 中可以看出, 座舱进气流量的数值仿真与测量信号吻合度较好, 在该情况下, 座舱压力的变化率趋势基本一致。这反映了本文方法的准确性。

由于试验测量信号的采样频率有限, 仿真数据的一些特征(如斜率、尖峰)难以准确体现在测量数据中。这也表明实际试验的短时快速变化过程中, 压力变化率的最大值可能会比测量值更大, 即实际座舱压力变化的恶劣程度要高于测量数据显示的程度。

3.2 小腔设计参数对座舱压力调节的影响

第 2 节分析中提到, B 腔内的气压会影响活门芯的受力, 进而影响座舱压力的动态特性, 因此, 对可能影响 B 腔内气压的设计参数进行分析。选取 B 腔定径孔直径、B 腔顶部直径、小膜片直径 3 个参数作为待研究参数, 开展其对座舱压力调节的动态特性的影响分析。

以某型号排气活门为研究对象, 先计算其在特定流量阶跃情况下的工作特性, 再采用控制变量法改变上述 3 个特定参数, 观察其对座舱压力调节的影响变化。

3.2.1 座舱进气流量冲击模拟

模拟座舱容积为 1.7 m^3, 外界气压为标准大气压, 温度为常温, 初始座舱进气流量为 250 kg/h, 随后给出阶跃信号, 将座舱进气流量信号设置为 1 100 kg/h, 观察压力调节系统及座舱压力的变化情况。

图 9 给出了座舱进气流量阶跃条件下的活门开度变化及 A 腔、B 腔、座舱压力的变化情况。初始状态下, 座舱进气流量为 250 kg/h, 活门开度稳定在略高于 50%, 座舱压力略高于外界大气压, B 腔压力与座舱压力一致, A 腔压力略低于座舱压力。在 t=3 s 时, 座舱进气流量突变至 1 100 kg/h, 此时座舱压力上升, 活门开度随之增加, A 腔被压缩而增压, B 腔由于净流量的增加而增压。由于 A 腔另一端通过控制机构与大气连通, A 腔压力开始逐渐降低。而 B 腔的增压速度逐渐跟不上座舱压力。当排气活门开至最大后, 座舱压力逐渐放缓增长并趋于稳定, A 腔压力也逐渐稳定, B 腔不再膨胀, 通过净流量而快速增压并跟上座舱压力, 整个系统逐渐回归新的平衡状态。

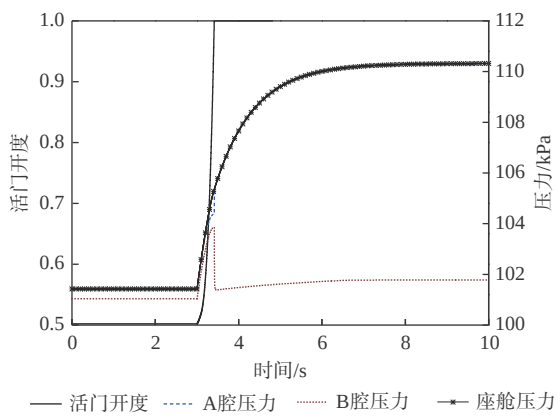


图 9 座舱进气流量阶跃下的排气活门及座舱状态变化

Fig. 9 State changes of outflow valve and cabin under cabin inflow step

图 10 给出了座舱进气流量阶跃过程中活门芯主要受力的变化。t=3 s 前, A/B 腔的压差力和大膜片的推力是推动排气活门开启的力, 弹簧弹力和小膜片的拉力是阻碍排气活门开启的力。t=3 s 后, 座舱遭遇流量冲击, A/B 腔的压差力和大膜片的推力

开始增长, 但由于 B 腔膨胀增压速度放缓, A/B 腔的压差力增长速度放缓; 小膜片的拉力逐渐增大, 但同样由于 B 腔压力增长速度放缓, 导致小膜片内外的压差不再增加, 小膜片的拖拽力也逐渐变小。

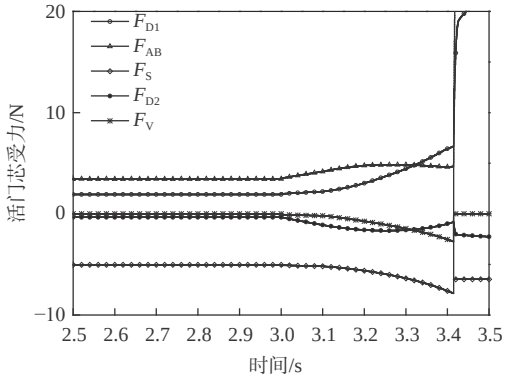


图 10　座舱进气流量阶跃过程中活门芯的受力变化

Fig. 10　Force changes of poppet during cabin inflow step process

整个冲压过程, 排气活门开至最大的用时约为 0.414 65 s, B 腔脱离基座挡板的持续时间约为 0.416 s。从上述座舱进气流量冲击过程中力的变化情况可以看出, 要增加排气活门的开启速度, 需要增大 B 腔向上的推力, 并使小膜片尽量贴在基座挡板上, 削弱小膜片的拉力。

3.2.2　小膜片参数对排气活门动态特性的影响分析

以 B 腔定径孔直径、B 腔顶部直径、小膜片直径这 3 个参数作为自变量, 分别选取初始参数的 50%、75%、125%、150% 作为变化参数, 观察不同变化参数下排气活门调节压力能力的变化情况。

1) B 腔定径孔直径。选取的被试排气活门的初始 B 腔定径孔的直径为 3.5 mm。B 腔定径孔的大小直接决定了 B 腔与座舱的流通能力, 也在一定程度上决定了 B 腔在座舱进气流量冲击下的冲压速度。

在 250 kg/h 的恒定座舱进气流量状态下, 不同 B 腔定径孔对应的活门稳态开度不变, 即稳态时座舱与 B 腔的压力基本一致, 不随定径孔的变化而改变。当发生座舱进气流量冲击时, 排气活门运动的动态特性随 B 腔定径孔直径的变化如图 11 所示。活门从稳定开度到最大开度的耗时随着孔径的增大而逐渐缩短, 对应的平均开启速度也逐渐增加, 在开启过程中, B 腔脱离基座托盘的持续时间也与最大开度用时保持一致的趋势。当 B 腔定径孔增大到一定程度后, 排气活门开启的耗时逐渐趋于平缓。将 B 腔定径孔增大 50%, 排气活门的开启速度从 19.2 mm/s 增至约 20.2 mm/s。

上述分析表明, 适当增加 B 腔定径孔的大小有助于提升排气活门在流量冲击下的开启速度。

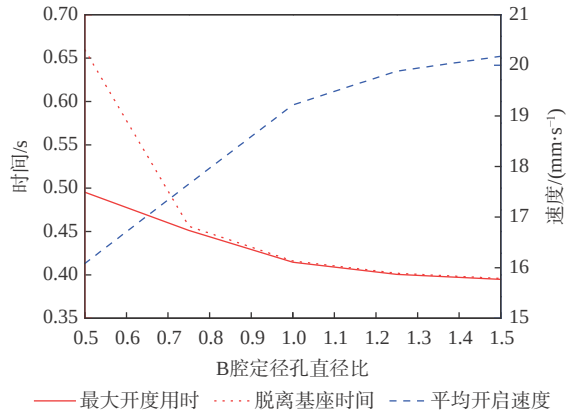


图 11　不同 B 腔定径孔直径对应的排气活门动态特征

Fig. 11　Dynamic characteristics of outflow valve corresponding to different metering hole diameters of balance chamber

2) B 腔顶部直径。选取的被试排气活门的初始 B 腔顶部直径为 0.106 m。B 腔顶部的圆面是 A/B 腔内气压差的作用面, 其大小直接影响了压差力 F_{AB} 的值。

在 250 kg/h 的恒定座舱进气流量状态下, 不同 B 腔顶部直径会产生不同大小的 F_{AB}, 影响活门芯的受力, 因此, 活门稳定时的相对开度也不相同。图 12 展示了 B 腔顶部直径变化时, 相同座舱进气流量下活门稳定状态的相对开度及稳定座舱压力的变化关系。随着 B 腔顶部直径的增加, F_{AB} 增大, 推动排气活门开启的力增加, 排气活门稳定时的开度也相应增加, 这也导致座舱的排气量增大, 稳定时的座舱压力随之减小。

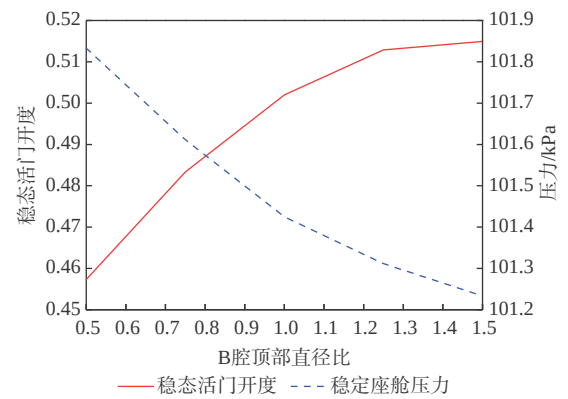


图 12　不同 B 腔顶部直径对应的稳态活门开度和稳定座舱压力

Fig. 12　Steady state opening and cabin pressure corresponding to different top diameters of balance chamber

在座舱进气流量冲击情况下, 不同的 B 腔顶部直径也会使得排气活门的动态特性发生改变。图 13 给出了不同 B 腔顶部直径对应的最大开度用时、平均开启速度及 B 腔脱离基座的持续时间的变化关系。可以看出, B 腔顶部直径的增加有助于缩短排气活门开至最大所消耗的时间, B 腔脱离基座的时

间也随之变短。但由于稳态时活门开度随着 B 腔顶部直径的增加而增大，在座舱进气流量冲击过程中，活门芯的运动行程随着 B 腔顶部直径的增大而减小，活门芯的加速时间随之缩短，反映在图 13 中就是排气活门平均开启速度并没有出现单调变化的规律。

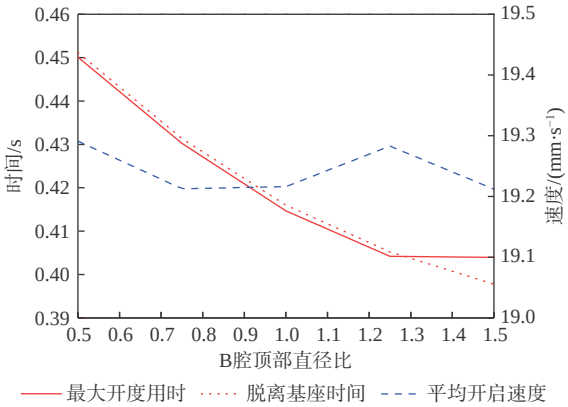


图 13 不同 B 腔顶部直径对应的排气活门动态特征

Fig. 13 Dynamic characteristics of outflow valve corresponding to different top diameters of balance chamber

上述分析表明，虽然增大 B 腔顶部直径可以提升排气活门开启的推动效果，但由于行程较短，对于排气活门开启速度的提升效果并不明显。

3）小膜片直径。选取的被试排气活门的初始小膜片直径为 0.106 m，小膜片内径（基座挡板的外径）为 0.09 m。小膜片直径的大小直接影响了活门芯受到的小膜片力 F_{D2} 的大小。另外，当小膜片贴在基座挡板上时，F_{D2} 的有效作用面为环形面，当小膜片脱离基座挡板时，F_{D2} 的有效作用面变为完整的圆面。为便于分析，在进行直径调整时，环形面的内外径、圆面的直径均按指定比例调整。

在 250 kg/h 的恒定座舱进气流量状态下，改变小膜片的大小引起 F_{D2} 改变，因此，活门稳态开度也发生改变。图 14 为不同小膜片直径情况下活门相对开度及稳定座舱压力的变化关系。随着小膜片直径的增大，F_{D2} 增大，对活门芯向下的拖拽作用增强，活门相对开度变小，对应的座舱压力增大。

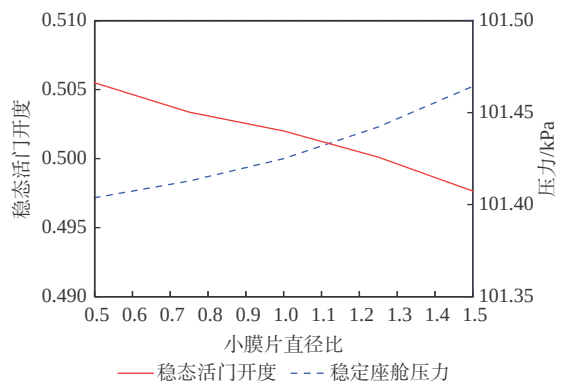


图 14 不同小膜片直径对应的稳态活门开度和稳定座舱压力

Fig. 14 Steady state opening and cabin pressure corresponding to different diameters of balance diaphragm

在座舱进气流量冲击情况下，排气活门的动态特性也会随着小膜片大小的变化发生改变。图 15 为小膜片直径变化时活门最大开度用时、平均开启速度及 B 腔脱离基座的持续时间的变化关系。增大小膜片直径后，小膜片的拖拽效果增强，排气活门开启的耗时及 B 腔脱离基座挡板的耗时都随之增加，对应的排气活门开启速度逐渐降低。将小膜片直径削减 50%，排气活门的开启速度从原来的 19.2 mm/s 提升至约 20.8 mm/s。结合活门芯的受力（见图 10）可以看出，虽然小膜片的拖拽力相对于大膜片力和 A/B 腔的压差力而言并没有很大，但改变小膜片的拖拽力对于排气活门开启速度的影响却较为显著。

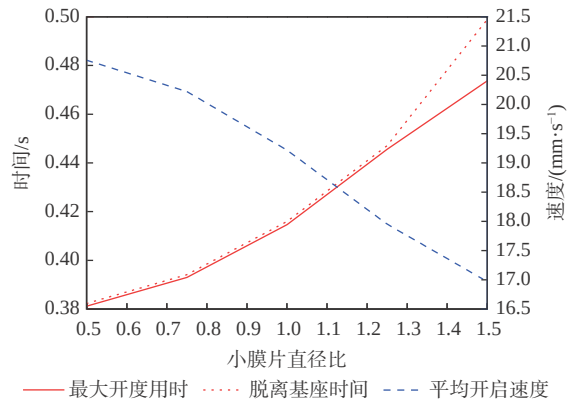


图 15 不同小膜片直径对应的排气活门动态特征

Fig. 15 Dynamic characteristics of outflow valve corresponding to different diameters of balance diaphragm

上述分析表明，减小小膜片的面积对于提升排气活门在座舱进气流量冲击下的开启速度有明显的效果。由于小膜片在排气活门的设计中还起到了负压充气的效果，小膜片的有效作用面积的设计需要慎重考虑：一方面，可以适当减小其面积来提升排气活门开启速度；另一方面，应当设计足够的作用面积来保证座舱负压时，排气活门能被顶开来对座舱反向充气保障结构安全。

4 结 论

本文指出排气活门小腔可能对排气活门的动态响应速度产生阻碍，通过对 PCPCS 的动力学分析，得到以下结论：

1）增大小腔定径孔有助于提升座舱进气流量冲击下的排气活门响应速度，但小腔定径孔增大到一定程度后，对排气活门响应速度的提升效果放缓。

2）增大小腔顶部直径可以增强对排气活门开

启的推动效果, 但同时也缩短了活门芯的运动行程, 因此, 对排气活门开启速度的提升效果不明显。

3）减小小膜片的面积可以显著减弱小膜片对活门芯的拖拽效果, 对排气活门响应速度的提升较为显著。

参考文献（References）

[1] 寿荣中, 何慧姗. 飞行器环境控制[M]. 北京: 北京航空航天大学出版社, 2004.
SHOU R Z, HE H S. Spacecraft optimal control theory and method [M]. Beijing: Beihang University Press, 2004(in Chinese).

[2] FISCHER R A. Pneumatic valve: US2672085A[P]. 1954-03-16.

[3] JENSEN R W. Safety valve: US2672086A[P]. 1954-03-16.

[4] FISCHER R A, KEMPER J M. Dual differential pressure regulating control: US2900890A[P]. 1959-08-25.

[5] FISCHER R A. Pressure regulating mechanism: US2986990A[P]. 1961-06-06.

[6] BURGESS G A, TAYLOR P J. Pressure control system: US3974752A[P]. 1976-08-17.

[7] MAAS B N. Pressure regulating mechanism: US2883920A[P]. 1959-04-28.

[8] JR ANDRESEN J H. Aircraft cabin pressurization system having controlled rate of pressure change: US3450020A[P]. 1969-06-17.

[9] TAYLOR P J. Cabin pressure control systems for general aviation aircraft[J]. SAE Transactions, 1980, 89: 2124-2135.

[10] 孙丽蓉, 郑丽, 李钧. 民机座舱负压差安全释压功能试飞研究[J]. 民用飞机设计与研究, 2019(1): 49-52.
SUN L R, ZHENG L, LI J. Research on cabin negative differential pressure relief test of civil aircraft[J]. Civil Aircraft Design & Research, 2019(1): 49-52(in Chinese).

[11] WHITNEY T J, LUI T L. Aircraft cabin multi-differential pressure control system: US7066808[P]. 2006-06-27.

[12] HORNER D, ARTHURS T R, ARMSTRONG B. Poppet valve for cabin pressure control systems: US8382035[P]. 2013-02-26.

[13] Aeronautics-Guide. Control of cabin pressure-aircraft pressurization systems (Part 3) [EB/OL]. (2017-05-01) [2023-05-11]. https://www.aircraftsystemstech.com/2017/05/control-of-aircraft-cabin-pressure.html.

[14] 北京航空学院五零五教研室. 座舱压力调节器原理[M]. 北京: 北京航空学院五〇五教研室, 1974.
Teaching and Research Office of Beijing Institute of Aeronautics and Astronautics. Principle of cabin pressure regulator[M]. Beijing: 505 Teaching and Research Office of Beijing Institute of Aeronautics and Astronautics, 1974(in Chinese).

[15] 王浚, 徐杨禾. 飞机座舱空气参数控制[M]. 北京: 国防工业出版社, 1980.
WANG J, XU Y H. Aircraft cabin air parameter control[M]. Beijing: National Defense Industry Press, 1980(in Chinese).

[16] 徐扬禾, 孙长祝, 朱东明. 带放大器的气动式座舱压力调节器动态特性计算和分析[J]. 北京航空学院学报, 1983, 9(3): 41-49.
XU Y H, SUN C Z, ZHU D M. Calculatio and analysis of the dynamic characteristic for pneumatic cabin pressure regulator with an amplifier[J]. Journal of Beijing University of Aeronautics and Astronautics, 1983, 9(3): 41-49(in Chinese).

[17] 成杰, 董鹏生. 两种座舱压力控制系统的对比分析[J]. 飞机工程, 2005(4): 42-44.
CHENG J, DONG P S. The comparison and analysis of two cabin pressure control system[J]. Aircraft Engineering, 2005(4): 42-44(in Chinese).

[18] 霍昱旭, 李玉忍, 宋颖慧. 飞机座舱压力气动调节系统的建模与仿真[J]. 计算机仿真, 2014, 31(1): 36-40.
HUO Y X, LI Y R, SONG Y H. Modeling and simulation of aircraft cabin pressure's pneumatic regulator system[J]. Computer Simulation, 2014, 31(1): 36-40(in Chinese).

[19] 魏晓永. 飞机座舱压力调节系统仿真研究[D]. 南京: 南京航空航天大学, 2012.
WEI X Y. Research on key technologies and optimal design of pneumatic cabin pressure regulation system[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012(in Chinese).

[20] 郑新华. 气动式座舱压力调节系统关键技术研究与优化设计[D]: 西安: 西北工业大学, 2016.
ZHENG X H. Key technology and optimization of pneumatic cabin pressure regulating system[D]. Xi'an: Northwestern Polytechnical University, 2016(in Chinese).

[21] ZHENG Y G, LIU M, ZHUANG J J, et al. A more precise modeling method for pneumatic cabin pressure control system considering the ambient environmental impacts[J]. International Journal of Aerospace Engineering, 2022, 2022: 2109994.

[22] BEAR S H. Sinus symptoms originating from modern flight[J]. The Laryngoscope, 1958, 68(6): 1057-1068.

[23] 朱治平, 于庆祥, 郜德成, 等. 模拟座舱压力变化对中耳影响的研究[J]. 航天医学与医学工程, 1989, 2(1): 30-35,75-76.
ZHU Z P, YU Q X, GAO D C, et al. Effects of simulated gabin pressure change on the middle ear in human[J]. Space Medicine & Medical Engineering, 1989, 2(1): 30-35,75-76(in Chinese).

[24] VALERIE I. A fix is coming for a problem that left two F-35 pilots in 'excruciating' pain[EB/OL]. (2019-06-12) [2023-05-11]. https://www.defensenews.com/air/2019/06/12/a-fix-is-coming-for-a-problem-that-left-two-f-35-pilots-in-excruciating-pain/.

[25] 刘剑飞, 王伟. 飞机座舱压力调节系统推油门"压耳"故障原因及解决措施研究[J]. 河南科技, 2019, 38(1): 92-95.
LIU J F, WANG W. Solution and research of push throttle pressure ear fault in cabin pressure regulating system[J]. Henan Science and Technology, 2019, 38(1): 92-95(in Chinese).

[26] MEGAN E. Navy rules out suspected physiological episodes cause while super hornet rates grow in 2019[EB/OL]. (2019-04-04) [2023-05-11]. https://news.usni.org/2019/04/04/navy-rules-out-contamination-as-physiological-episodes-cause-focused-on-air-pressure-as-super-hornet-rates-still-high.

[27] 胡正元, 史秀凤, 范静平. 不同增压速率所引起的耳气压伤特征[J]. 声学技术, 2004, 23(1): 25-28.
HU Z Y, SHI X F, FAN J P. Characteristics of aural barotrauma caused by different compression rate[J]. Technical Acoustics, 2004, 23(1): 25-28(in Chinese).

[28] CHAURASIYA K L, BHATTACHARYA B, VARMA A K, et al. Dynamic modeling of a cabin pressure control system[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2020, 234(2): 401-415.

[29] 廉乐明, 谭羽非, 吴家正, 等. 工程热力学[M]. 5 版. 北京: 中国建筑工业出版社, 2007.
LIAN L M, TAN Y F, WU J Z, et al. Engineering thermodynamics [M]. 5th ed. Beijing: China Architecture & Building Press, 2007(in Chinese).

[30] 崔超群. 针对座舱压力增长率的分析与研究[J]. 中国科技纵横, 2013(20): 183-184.
CUI C Q. Analysis and research on the growth rate of cabin pressure[J]. China Science & Technology Overview, 2013(20): 183-184(in Chinese).

Effect of balance chamber on rapid pressure regulation ability of outflow valve

WU Hao，LIU Meng*，WANG Jun

(School of Aeronautic Science and Engineering，Beihang University，Beijing 100191，China)

Abstract：Under aircraft cabin inflow impact, the inadequate dynamic regulation speed of the pneumatic cabin pressure control system (PCPCS) may lead to a cabin pressure spike, causing “barotrauma”. Therefore, the working principle of the PCPCS was analyzed, and it was pointed out that the balance chamber may hinder the rapid movement of the outflow valve. A dynamic model of the PCPCS was established. The working states of the control chamber (A chamber) and the balance chamber (B chamber) in the outflow valve were described. The dynamic simulation of the PCPCS under cabin inflow impact was carried out. The dynamic working characteristics of the outflow valve and the force changes of the poppet were quantitatively displayed. The effects of the metering hole diameter of the balance chamber, the top diameter of the balance chamber, and the diameter of the balance diaphragm on the outflow valve’s response to cabin inflow impact were discussed. The simulation results show that appropriately increasing the metering hole size of the balance chamber can help to improve the response speed of the valve; the top diameter of the balance chamber has no significant effect on the movement speed of the outflow valve, and reducing the area of the balance diaphragm improves the response speed of the outflow valve significantly.

Keywords：environmental control of aircraft；pressure control of aircraft cabin；barotrauma；pneumatic cabin pressure control system；outflow valve

Received：2023-05-15；**Accepted**：2023-08-14；**Published Online**：2023-08-22 10：12
URL：link.cnki.net/urlid/11.2625.V.20230821.1508.001

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0184

翼伞翼型非定常气动特性数值模拟研究

王震[1,2]，钟伟[1,2,*]，王同光[1,2]，李旭东[1,2]，张红英[1]

(1. 南京航空航天大学 航空学院，南京 210016；2. 南京航空航天大学 江苏省风力机设计高技术研究重点实验室，南京 210016)

摘　　要：翼伞的非定常气动特性关乎飞行稳定性和飞行安全，是一个值得深入研究的问题。采用计算流体力学(CFD)方法对迎角动态变化下的翼伞翼型开展非定常数值模拟，分析平均迎角、迎角振幅和缩减频率对翼伞翼型非定常气动力和流场形态的影响。结果显示：平均迎角、迎角振幅和缩减频率3个参数中的任一个增大，都导致翼伞翼型气动力的非定常性显著增强；平均迎角与迎角振幅共同决定翼型迎角动态变化的范围，显著影响翼伞翼型的动态失速特性；缩减频率的变化对翼伞翼型的升力迟滞效应产生影响，最大升力系数随缩减频率的增加而呈线性增大。流场分析表明：翼伞翼型迎角动态增大阶段的非定常效应，延缓了翼型流动分离的发生，并使分离涡紧贴上翼面呈扁平形状，共同促使翼伞翼型的临界迎角和最大升力系数相对定常情况显著增大。研究成果有助于增进对翼伞翼型非定常气动力和流场规律的理解，为复杂风环境下的翼伞非定常气动力预测和安全性评估提供支撑。

关 键 词：翼伞；翼型；气动性能；动态失速；缩减频率

中图分类号：V244

文献标志码：A　　**文章编号**：1001-5965(2025)04-1255-12

翼伞是一种采用带前缘切口翼型、在前缘及气室内滞止压力的作用下保持气动构型的可控气动力减速装置，具备优良的滑翔性能和可控性，应用于航天器返回、定点空投等多个领域[1-2]。空气动力特性是翼伞滑翔和控制的基础性能，大量学者开展了相关研究，绝大部分是关于稳态飞行的定常气动特性的研究，缺乏非定常气动特性的研究。然而，在实际大气环境下，突风、阵风等瞬态复杂风况普遍存在，翼伞在遭遇该风况时，与稳态风况相比，具有的非定常气动特性关乎翼伞的飞行稳定性和飞行安全，是值得深入研究的问题。例如，当遭遇下击暴流，低空飞行的飞机可能发生坠地危险[3]，而翼伞的飞行速度远低于飞机，遭遇相同的极端风况时，其气动特性受到的影响更大，安全风险也更大。由于翼伞的非定常气动特性很大程度上取决于其所采用翼型的非定常气动特性，深入研究翼伞翼型的非定常气动特性对翼伞的设计和安全性有着重要意义。

已有较多学者应用计算流体力学(computational fluid dynamics, CFD)方法开展了翼伞翼型的二维流场数值模拟研究，为翼伞翼型设计提供了参考。Mittal等[4-5]研究了翼伞前缘切口对翼伞气动性能的影响；聂帅等[6]分析了前缘切口和弧面下反角对翼伞气动性能的影响；李健[7]对后缘下拉状态的翼伞翼型开展了气动特性数值模拟研究。还有一些学者通过风洞试验或CFD数值模拟对三维翼伞模型开展了气动特性研究。Furfaro等[8]和Uddin、Mashud[9]利用风洞试验研究了翼伞前缘切口对翼伞气动特性的影响，并通过改进切口形状提高了翼伞的滑翔比；朱旭和曹义华[10]对不同特征几何参数翼伞模型的流场进行了三维数值模拟，分析了翼伞平面形状对气动性能的影响；王国庆等[11]从大型翼伞伞衣变形过程出发，分析了大型翼伞转弯过程的运动原理，通过理论推导、CFD仿真、空投试验

收稿日期：2023-04-18；**录用日期**：2023-07-03；**网络出版时间**：2023-07-12 09:00
网络出版地址：link.cnki.net/urlid/11.2625.V.20230711.2149.003
***通信作者**. E-mail：zhongwei@nuaa.edu.cn

引用格式：王震，钟伟，王同光，等. 翼伞翼型非定常气动特性数值模拟研究[J]. 北京航空航天大学学报，2025，51(4)：1255-1266.
WANG Z，ZHONG W，WANG T G，et al. Numerical simulation of unsteady aerodynamic characteristics of parafoil airfoil[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1255-1266 (in Chinese).

3种方法得到了大型翼伞的侧力系数与下拉量之间的关系；孙青林等[12]利用CFD技术模拟风雨环境对翼伞飞行性能的影响，模拟结果与NASA风洞试验结果接近，能够较好地描述翼伞在风雨环境下的飞行性能。

关于翼伞翼型的非定常气动特性研究鲜见报道，但对普通航空翼型的非定常研究较多，可以作为参考。Raffel等[13]通过粒子图像测速法，在低速风洞中研究了NACA 0012翼型在俯仰动态失速条件下的非定常流场；宋红军等[14]对二维翼型沉浮、俯仰及沉浮/俯仰联合3种运动方式分别进行了CFD数值模拟研究；闫文辉[15]对NACA 4412翼型低速绕流进行了定常/非定常数值模拟，并与实验数据进行了对比，显示非定常模拟结果与实验数据更加符合；宗涛[16]对NACA 23012翼型的动态失速进行数值模拟研究，分析了翼型在振荡周期内升力系数随迎角的变化，并进一步研究了俯仰运动的关键参数对翼型动态失速的影响，得出了翼型升力系数曲线在不同迎角、振幅、衰减频率下的变化规律；钱炜祺等[17]对典型的振荡翼型轻失速和深失速算例开展了数值分析；Tuncer等[18-19]对翼型在单纯的沉浮运动及复合的俯仰/沉浮运动下的流动分离特性开展了研究。

翼伞翼型与普通航空翼型的差异在于前缘切口，尽管该差异会导致气动特性的不同，但用于普通航空翼型非定常气动特性研究的数值模拟方法仍然适用于翼伞翼型研究。本文借鉴翼伞翼型定常模拟的经验，以及普通航空翼型非定常模拟的经验，对翼伞翼型开展非定常数值模拟，分析平均迎角α_0、迎角振幅$\Delta\alpha$和缩减频率K对翼伞翼型非定常气动特性特别是动态失速的影响。考虑到复杂风况对翼伞翼型气动特性的影响可以等效为迎角和风速瞬态变化的影响，其中又以迎角变化的影响最为显著，本文研究有助于揭示翼伞翼型的非定常气动力特性和流场特征。

1 数值模拟方法

1.1 控制方程

翼伞的滑翔速度远低于声速，因此，可以把空气简化为不可压缩流。本文数值模拟基于的控制方程为不可压缩的雷诺平均Navier-Stokes（Reynolds averaged Navier-Stokes, RANS）方程，表达式为

$$\frac{\partial}{\partial t}\int_{\Omega}\boldsymbol{W}\mathrm{d}\Omega+\oint_{\partial\Omega}(\boldsymbol{F}_{\mathrm{c}}-\boldsymbol{F}_{\mathrm{v}})\mathrm{d}S=0 \tag{1}$$

式中：Ω表示控制体；$\partial\Omega$为包围控制体的封闭表面；S表示控制体的任意边界面元；$\boldsymbol{W}$为守恒变量；$\boldsymbol{F}_{\mathrm{c}}$为对流通量；$\boldsymbol{F}_{\mathrm{v}}$为黏性通量。

$$\boldsymbol{W}=\begin{bmatrix}\rho\\ \rho u\\ \rho v\end{bmatrix} \tag{2}$$

$$\boldsymbol{F}_{\mathrm{c}}=\begin{bmatrix}\rho V_{\mathrm{r}}\\ \rho u V_{\mathrm{r}}+pn_x\\ \rho v V_{\mathrm{r}}+pn_y\end{bmatrix}=\begin{bmatrix}\rho V\\ \rho u V+pn_x\\ \rho v V+pn_y\end{bmatrix}-V_{\mathrm{b}}\begin{bmatrix}\rho\\ \rho u\\ \rho v\end{bmatrix} \tag{3}$$

$$\boldsymbol{F}_{\mathrm{v}}=\begin{bmatrix}0\\ n_x\tau_{xx}+n_y\tau_{xy}\\ n_x\tau_{yx}+n_y\tau_{yy}\end{bmatrix} \tag{4}$$

$$\begin{cases}V_{\mathrm{r}}=V-V_{\mathrm{b}}\\ V=un_x+vn_y\\ V_{\mathrm{b}}=u_{\mathrm{b}}n_x+v_{\mathrm{b}}n_y\end{cases} \tag{5}$$

式中：ρ为密度；p为压强；n_x和n_y分别为控制体表面的单位外法矢$\boldsymbol{n}$的2个分量；u和v分别为速度矢量的2个分量；V_{r}为流体与网格的相对法向速度；u_{b}、v_{b}分别为动网格的运动速度在x、y方向上的分量；τ_{ij}为剪切应力张量，包括层流黏性和湍流黏性的作用，各项具体表达式为

$$\begin{cases}\tau_{xx}=-\dfrac{2}{3}\mu\left(\dfrac{\partial u}{\partial x}+\dfrac{\partial v}{\partial y}\right)+2\mu\dfrac{\partial u}{\partial x}\\ \tau_{yy}=-\dfrac{2}{3}\mu\left(\dfrac{\partial u}{\partial x}+\dfrac{\partial v}{\partial y}\right)+2\mu\dfrac{\partial v}{\partial y}\\ \tau_{xy}=\tau_{yx}=\mu\left(\dfrac{\partial u}{\partial y}+\dfrac{\partial v}{\partial x}\right)\end{cases} \tag{6}$$

其中：μ为黏性系数，是层流黏性系数μ_{L}和湍流黏性系数μ_{T}的和：

$$\mu=\mu_{\mathrm{L}}+\mu_{\mathrm{T}} \tag{7}$$

湍流模拟采用Menter[20]提出的二方程SST k-ω模型，模型的输运方程为

$$\frac{\partial k}{\partial t}+u_j\frac{\partial k}{\partial x_j}=\frac{1}{\rho}\cdot\frac{\partial}{\partial x_j}\left[\left(\mu+\frac{\mu_T}{\sigma k}\right)\frac{\partial k}{\partial x_j}\right]\frac{Ma_\infty}{Re}+\frac{1}{\rho}Pk\frac{Ma_\infty}{Re}-\beta'\omega k\frac{Re}{Ma_\infty} \tag{8}$$

$$\frac{\partial \omega}{\partial t}+u_j\frac{\partial \omega}{\partial x_j}=\frac{1}{\rho}\cdot\frac{\partial}{\partial x_j}\left[\left(\mu+\frac{\mu T}{\sigma\omega}\right)\frac{\partial \omega}{\partial x_j}\right]\frac{Ma_\infty}{Re}+\frac{1}{\rho}P\omega\frac{Ma_\infty}{Re}-\beta'\omega^2\frac{Re}{Ma_\infty}+2(1-F)\frac{1}{\sigma_{\omega2}}\cdot\frac{1}{\omega}\cdot\frac{\partial k}{\partial x_j}\cdot\frac{\partial \omega}{\partial x_j}\cdot\frac{Ma_\infty}{Re} \tag{9}$$

式中：k和ω分别为湍动能和比耗散率；σ_k、σ_ω为扩散项控制系数；β、β^*为湍流模型经验系数；$\sigma_{\omega2}$为耗散项扩散系数；P_k、P_ω为湍动能、耗散率的生成项；F为混合函数；Re为雷诺数；Ma_∞为自由流马赫数[21]。

求解过程采用基于压力的 Coupled 算法处理 RANS 方程中速度和压力的耦合, 即求解动量方程时直接对压力进行隐式离散求解。与 SIMPLE 系列算法相比, Coupled 算法虽然每个时间步的计算量更大, 内存需求也更多, 但有更好的鲁棒性和收敛性, 适用于翼型深失速等 SIMPLE 算法较难收敛的场合, 而且可以采用更大的时间步长进行求解, 从而能够更快地获得流场的收敛解[22]。

1.2　模型与网格

用于本文数值模拟的翼伞翼型是基于 Clark-Y 翼型的前缘开口翼型, 该翼型是翼伞气动特性研究的常用翼型[23], 具体参数为切口长度为 0.05C, 切口与弦线的角度为 135°, 翼型弦长 C=2 m, 如图 1 所示。

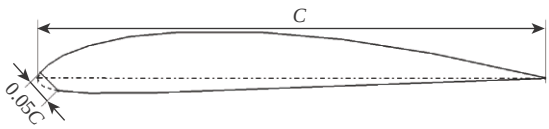


图 1　基于 Clark-Y 翼型的前缘开口翼型构型

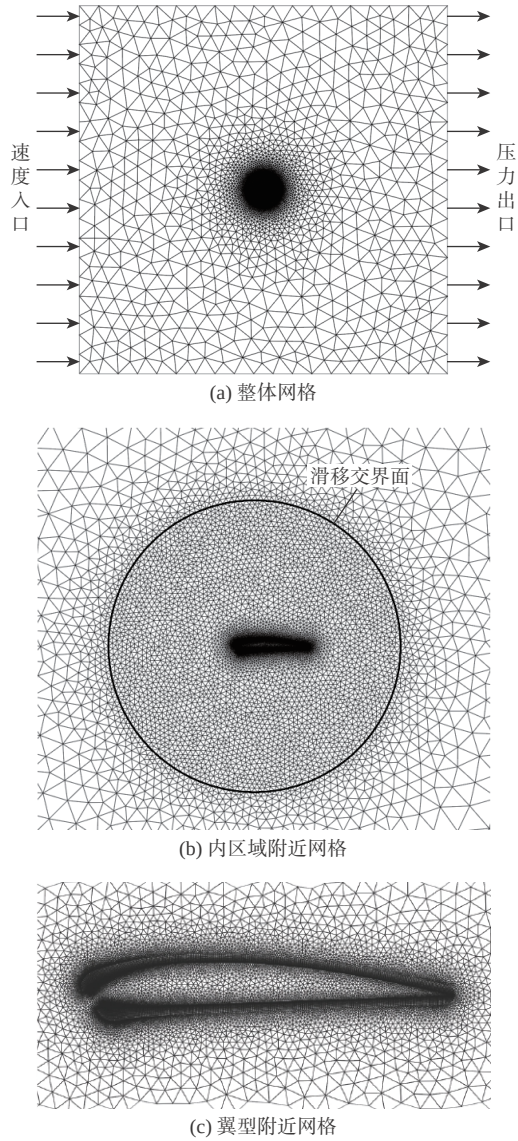
Fig. 1　Clark-Y-based airfoil configuration with an open leading edge

计算域网格如图 2 所示, 分为内外 2 个区域, 网格总数约为 4.9 万。外区域为边长为 40 倍翼型弦长的正方形, 采用三角形网格划分, 左侧边界为入口边界, 指定来流速度、方向和湍流强度, 右侧边界为出口, 指定静压。内区域为半径是 2 倍翼型弦长的圆, 圆心位于翼型的 1/4 弦长位置, 临近翼型外壁面的边界层网格为四边形, 第 1 层网格高度为 10^{-5} m, 相邻网格从内到外的高度增长率为 1.05, 边界层网格共有 35 层。边界层网格以外的区域采用三角形网格划分。

在滑翔过程中, 翼伞迎角的瞬态变化主要由风突变导致。本文数值模拟根据相对运动原理采用翼型俯仰振荡、来流风不变的方式开展, 这是翼型非定常气动特性数值模拟和风洞试验的典型做法。翼型的俯仰振荡通过内区域网格的刚性旋转实现, 内区域网格与外区域网格的交界面采用滑移网格技术实现数据传递。

1.3　计算条件

如果翼伞在飞行过程中遭遇风突变, 即风向或风速在短时间内大幅度变化的现象, 会对翼伞的飞行产生严重影响。风速或风向的突变会造成翼伞飞行迎角的突变。翼型迎角由图 3 所示的速度三角形决定, 其中, $\boldsymbol{V}_0$为来流风速, V_{0y}为$\boldsymbol{V}_0$在 y 方向的分量, V_{0x}为$\boldsymbol{V}_0$在 x 方向的分量, 则迎角 α 在不太大时(如 20°以内)可以表示为



(a) 整体网格

(b) 内区域附近网格

(c) 翼型附近网格

图 2　计算网格

Fig. 2　Computational grid

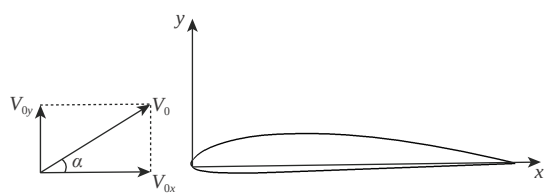


图 3　风速和风向对翼型迎角变化

Fig. 3　Influence of wind speed and wind direction on change of angle of attack of airfoil

$$\alpha = \arctan\left(\frac{V_{0y}}{V_{0x}}\right) \approx \frac{V_{0y}}{V_{0x}} \tag{10}$$

翼伞以小迎角稳定飞行时的V_{0x}比V_{0y}大得多, 当V_{0x}和V_{0y}突变的绝对量相同时, V_{0y}的突变更容易造成迎角的突变, 从而对翼伞气动力和飞行安全产生更严重的影响。非定常空气动力的理论研究中, 常以正弦函数的形式作为非定常变化的输入函数。例如, 可将V_{0y}的非定常变化设为

$$V_{0y}=V_{0y}^{*}+\Delta V_{0y}\sin(\omega t) \tag{11}$$

式中：V_{0y}^{*}为y方向的平均速度；ΔV_{0y}为y方向的速度振幅；ω为速度变化频率；t为时间。

由此得到的迎角α的表达式也是正弦函数的形式：

$$\alpha=\frac{1}{V_{0x}}\left(V_{0y}^{*}+\Delta V_{0y}\sin(\omega t)\right) \tag{12}$$

因此，本文的数值模拟研究中，设定翼型迎角α以正弦规律变化，翼型的转动中心设置在1/4弦长位置，翼型绕其转动中心做正弦振荡，迎角变化规律为

$$\alpha=\alpha_0+\Delta\alpha\sin\left(\frac{2K|\boldsymbol{V}_0|}{C}t\right) \tag{13}$$

$$K=\frac{\omega C}{2|\boldsymbol{V}_0|} \tag{14}$$

翼型在迎角动态变化过程中的气动特性主要由3个因素决定，分别为平均迎角α_0、迎角振幅$\Delta\alpha$和缩减频率K[24]。α_0和$\Delta\alpha$共同决定了最大迎角$\alpha_{\max}$，而$\alpha_{\max}$决定了翼型流动分离的严重程度。K对动态失速的作用会受到边界层分离类型和失速机制的影响，K越高，动态失速和气动力迟滞通常越显著。

在来流风速一定的条件下，平均迎角α_0、迎角振幅$\Delta\alpha$和缩减频率K完全决定了翼型迎角的取值范围。本文设置了5个算例来分析平均迎角、迎角振幅和缩减频率对翼型非定常气动力的影响，如表1所示。平均迎角α_0在翼伞的正常工作迎角范围内取值，涉及α_0=6°和α_0=8°；迎角振幅$\Delta\alpha$的选取使翼伞的迎角变化范围覆盖小迎角和轻失速情况，涉及$\Delta\alpha$=4°和$\Delta\alpha$=6°；缩减频率K的取值范围在0.01～0.051之间，是翼型非定常空气动力研究中普遍关注的缩减频率范围[25]。

非定常模拟的初始流场由平均迎角下的翼型定常流场求解结果提供。非定常求解的物理时间步长Δt设置为翼型每个俯仰周期内有N个时间步，即

$$\Delta t=\frac{2\pi/\omega}{N} \tag{15}$$

表1 算例设置

Table 1 Computational cases

算例	平均迎角α_0/(°)	迎角振幅$\Delta\alpha$/(°)	缩减频率K
算例1	8	6	0.051
算例2	6	6	0.051
算例3	6	4	0.051
算例4	6	6	0.03
算例5	6	6	0.01

式中：$\omega=2K|\boldsymbol{V}_0|/C$。本文数值模拟的弦长$C$=2 m，来流速度$|\boldsymbol{V}_0|$=20 m/s。根据后续时间步长无关性分析，取$N$=600，则算例1～算例3的时间步长为$\Delta t$=0.051 s，算例4的时间步长为$\Delta t$=0.017 5 s，算例5的时间步长为$\Delta t$=0.052 4 s。试算3个俯仰周期后，翼型的气动力和流场呈周期性稳定变化，取第3周期结果进行分析。

1.4 网格和时间步长的无关性分析

1.4.1 网格无关性分析

数量足够且高质量的网格是CFD数值模拟可靠的重要基础。网格数量是否足够，与网格总数有关，还与网格节点的空间分布有关。判断网格数量是否足够，需要通过对单元数递增的网格进行计算测试，直至计算结果不再随网格数量增加而变化时，认为网格数量达到了计算无关性要求，可以用于数值模拟。本文在正式数值模拟前，采用网格单元数依次递增的3套网格进行了网格无关性测试，这3套网格分别被命名为G01、G02和G03，其具体参数如表2所示。

表2 用于网格无关性测试的网格参数

Table 2 Grid parameters for grid-independence testing

网格编号	壁面首层网格高度/m	翼型弦向网格点数	翼型往外的尺度增长率	网格单元总数
G01	10^{-5}	200	1.15	2.3×10^{4}
G02	10^{-5}	250	1.1	3.7×10^{4}
G03	10^{-5}	300	1.05	4.9×10^{4}

使用这3套测试网格进行数值模拟获得的Clark-Y切口翼型在来流速度$|\boldsymbol{V}_0|$=20 m/s下的升力系数C_L曲线如图4所示。可知，G01～G03网格的计算结果基本重合，表明这3套网格都符合网格无关性要求。考虑到对于二维仿真，计算量不是严重的制约因素，从谨慎的角度出发，本文后续数值模拟选取网格密度最高的G03网格开展。

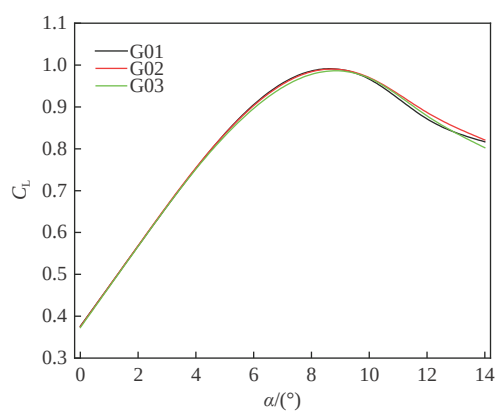


图4 不同网格密度下的升力系数曲线

Fig. 4 Lift coefficient curves for different grid densities

1.4.2　时间步长无关性分析

一般来说, 时间步长过大将导致数值模拟丢失某些非定常流动特征, 相应的, 计算误差经过一定时间的累积, 可能导致求解结果不可靠。因此, 有必要确认数值模拟采用的时间步长足够小, 达到求解结果的时间步长无关性。

选取 α_0=6°, $\Delta\alpha$= 6°, K=0.051 作为测试工况, 以 4 种时间步长设置 (Δt=0.1, 0.05, 0.025, 0.01 s) 开展时间步长无关性的数值模拟分析。非定常求解过程中, 一个时间步内的迭代次数选择较充分的 50 次, 以确保单个时间步内的求解达到收敛。这 4 种时间步长使得每周期内数值模拟的计算步数分别为 63、125、248 和 616。数值模拟测试得到的非定常升力系数曲线如图 5 所示。可知, 时间步长为 0.1 s 的结果与其他时间步长的结果存在较明显的差异; 随着时间步长减小, 结果的差异性也随之减小; 时间步长为 0.025 s 和 0.01 s 的计算结果具有很高的一致性。因此, 对于本文算例, 时间步长小于 0.025 s, 即一个俯仰周期内的计算时间步大于 248 步, 即可认为满足时间步长无关性要求。本文后续数值模拟采用的时间步长设置为保证一个俯仰周期内的计算时间步等于 600 步 (N=600), 根据式 (15) 解出 Δt 的具体值。

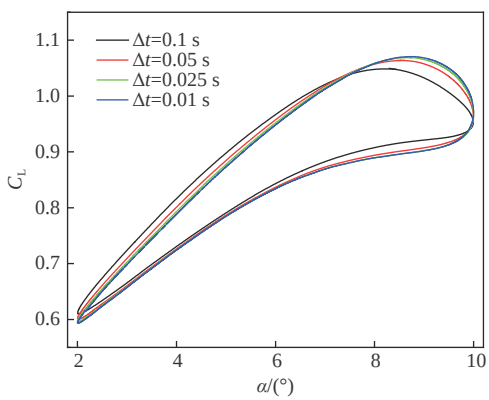


图 5　不同时间步长下数值模拟的升力系数曲线

Fig. 5　Numerically simulated lift coefficient curves at different time steps

1.5　算例验证

本文数值仿真的开口翼型气动实验数据和 Clark-Y 翼型非定常气动力实验鲜有报道, 因此, 算例验证针对 S809 翼型开展。S809 翼型与 Clark-Y 翼型都属于低速翼型, 且两者的临界迎角较为接近。本文采用文献 [26] 给出的 S809 翼型在 C=0.457 m、Re=1×10^6、α_0=8°、$\Delta\alpha$=10°、K=0.052 的实验结果来进行算例验证。

图 6 为采用本文数值模拟方法求解得到的升力系数随迎角变化的迟滞曲线与实验值的比较。

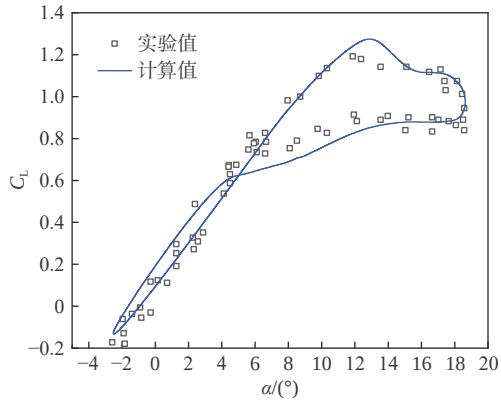


图 6　S809 翼型的升力系数迟滞曲线计算值与实验值对比

Fig. 6　Comparison of calculated and experimental values of lift coefficient hysteresis curves of S809 airfoil

可知, 本文 CFD 计算结果与风洞试验值在气动力变化的规律上呈现出较好的一致性, 各迎角下升力系数的具体量值差异也较小, 表明本文数值模拟方法用于翼型的非定常气动力仿真具有较高的可信度。

2　非定常气动力分析

2.1　平均迎角的影响

算例 1 和算例 2 的迎角振幅和缩减频率等参数相同, 仅平均迎角不同, 下面以算例 1 和算例 2 的数值模拟结果分析平均迎角对翼伞翼型非定常气动力的影响。

数值模拟获得的翼伞翼型升力系数曲线如图 7 所示。可见, 算例 1 和算例 2 所呈现的翼伞翼型非定常俯仰的升力系数变化与定常情况存在显著差异, 主要体现在以下方面: ①非定常升力系数在上仰段普遍大于定常情况, 在下俯段普遍小于定常情况, 在最大迎角附近呈现急剧下降。②非定常升力系数曲线上有 3 个点值得关注, 图 7 中以 $A_1(A_2)$、$B_1(B_2)$ 和 $C_1(C_2)$ 标记。点 $A_1(A_2)$ 是升力系数的峰值点, 点 $B_1(B_2)$ 是升力系数从峰值点急剧下降达到的最低值点, 点 $C_1(C_2)$ 是升力系数从点 $B_1(B_2)$完成快速回升

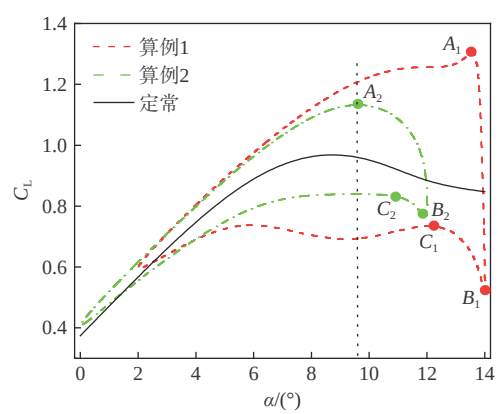


图 7　算例 1 和算例 2 的升力系数曲线

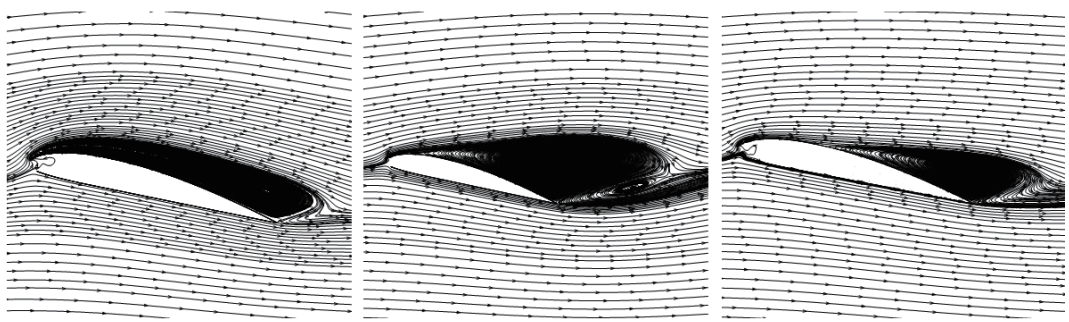
Fig. 7　Lift coefficient curves for cases 1 and cases 2

后的对应点。定常升力系数曲线的变化规律相对简单，不存在 $A_1(A_2)\rightarrow B_1(B_2)\rightarrow C_1(C_2)$这样的变化过程。③在下俯段，小于点 $C_1(C_2)$对应迎角的一段范围内，升力系数变化较为平缓。此外，算例1比算例2的升力系数曲线迟滞环面积大，升力系数峰值及其对应迎角也大。算例1的峰值升力系数达到1.31，对应迎角为13.6°；算例2的峰值升力系数为1.13，对应迎角为9.6°；而定常计算的升力系数峰值为0.98，对应迎角为8°。

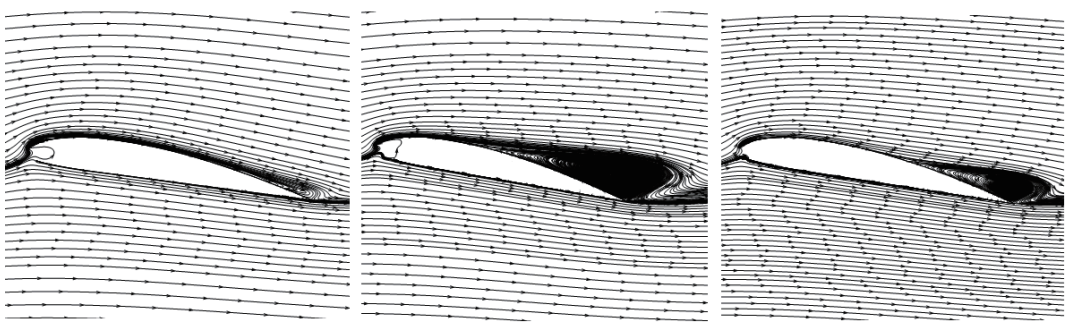
图8显示了算例1和算例2分别在点 A_1、B_1、C_1 和点 A_2、B_2、C_2 对应的流线图。由图8(a)可知，在点 A_1 处，翼型上表面从前缘到后缘均已发生流动分离，但分离涡呈扁平状态，局限于翼型上表面附近区域；由图8(d)可知，在点 A_2 处，翼型上表面尚未发生大面积的流动分离。这表明算例1和算例2的升力系数峰值点所对应的翼型流动状态是不同的，显示出平均迎角 α_0 的变化可导致翼型非定常流动特征的显著差异。由图8(b)和8(e)可知，点 B_1 处翼型上表面呈现大面积的完全分离流动状态，而 B_2 处翼型上表面呈现尾缘分离流动状态，这是算例1在点 B_1 升力系数显著低于算例2在点 B_2 升力系数的原因。由图8(c)和8(f)可知，在点 C_1 和点 C_2 处，算例1仍有较大的尾缘分离涡，而算例2的尾缘分离涡明显小于前者。

由图7可见，算例1在点 A_1 处的升力系数大于算例2在点 A_2 处的升力系数；由图8(a)和图8(d)可知，点 A_1 处的翼型上表面被扁平状分离涡完全覆盖，而点 A_2 处的流动分离并不显著。也就是说，翼型在点 A_1 处发生大面积的流动分离却仍然保持了高升力系数，而在点 A_2 处流动分离不显著却只产生了相比点 A_1 处更低的升力系数。这种现象在定常情况下不会出现，是一种典型的非定常现象。为进一步分析该现象，观察图9所示的点 A_1 和点 A_2 处的翼型压力分布对比。可知，在点 A_2 处，翼型前缘附近的压力系数 C_p 产生一个高的吸力峰值，符合附着流动产生前缘吸力峰值的特征；在点 A_1 处，没有观察到类似的吸力峰值，但在翼型上表面大部分区域有更低的压强，正是大面积的低压区域使翼型产生了更高的升力。与图8(a)所示的流场图进行对照分析，可以认为该大面积低压区域是覆盖在翼型上表面的扁平状分离涡导致的。也就是说，点 A_1 处在完全分离流动状态下仍保持高升力系数，是非定常上仰过程中特有的扁平状分离涡提供“涡升力”的结果。

图7～图9分析了算例1和算例2在不同迎角下的非定常气动力和流场特征，下面分析其在相同迎角下的流场特征。选取迎角 α=9.6°时2个算例的流场开展对比分析，如图10所示。之所以选择 α=9.6°开展分析，是因为在该迎角处，算例2取得升力系数峰值，且2个算例在下俯阶段的升力系数差异最大。由图10(a)和图10(b)可见，在算例1上仰阶段的 α=9.6°处，翼型上表面只在前缘处



(a) 点A_1 (α=13.6°)↑　(b) 点B_1 (α=14.0°)↓　(c) 点C_1 (α=12.0°)↓



(d) 点A_2 (α=9.6°)↑　(e) 点B_2 (α=11.9°)↓　(f) 点C_2 (α=11.0°)↓

图8　算例1和算例2在不同迎角时的流线图

Fig. 8　Streamlines for case 1 and case 2 at various angles of attack

出现较小的分离涡, 在下俯阶段的 α=9.6°处, 翼型上表面后缘处出现明显的分离涡。由图 10(b)和图 10(c)可见, 算例 1 在下俯阶段 α=9.6°处的分离涡明显大于算例 2 在下俯阶段 α=9.6°处的分离涡, 这是算例 1 在下俯阶段的升力系数小于算例 2 对应升力系数的原因。

由以上分析可见, 尽管仿真的 2 个算例迎角只相差 2°, 翼伞翼型的非定常升力系数和流场形态都发生了十分明显的变化, 较大的平均迎角导致显著更大的升力系数迟滞环、临界迎角和最大升力系数。算例 1 的临界迎角(13.6°)比算例 2 的临界迎角(9.6°)增大了 4°, 即临界迎角的增量达到了平均迎角增量的 2 倍, 可见平均迎角的增加对翼伞翼型非定常气动特性的影响并非线性。这种非线性影响的流场原因在于, 非定常条件下, 翼伞翼型表面流动分离在临界迎角附近呈现出与定常条件下不同的形态, 且分离流动状态随迎角的变化相比定常情况更快速和剧烈。

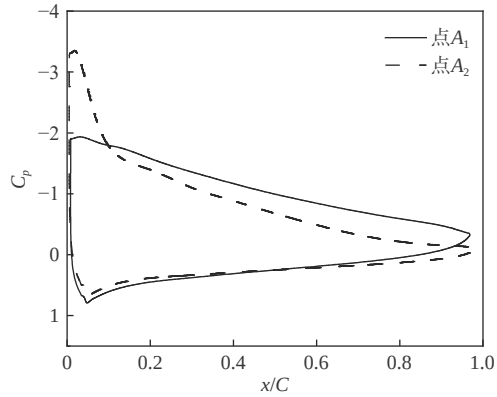
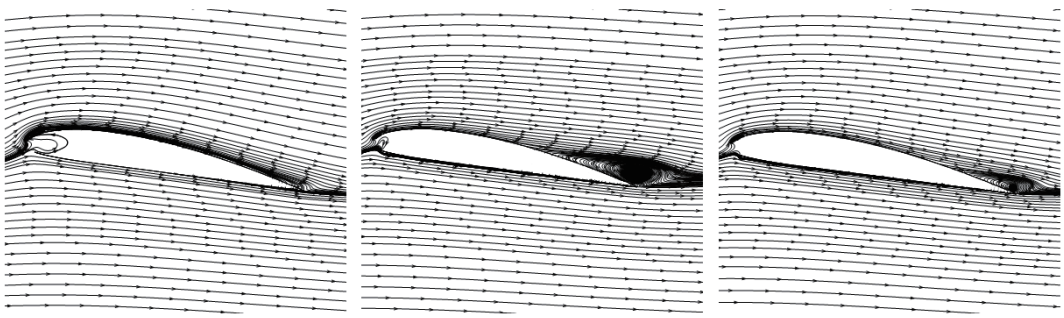


图 9　点 A_1、点 A_2 处的翼型压力分布

Fig. 9　Pressure distribution at point A_1 and point A_2 of airfoil



(a) 算例1↑　　(b) 算例1↓　　(c) 算例2↓

图 10　算例 1 和算例 2 在相同迎角时的流线图(α=9.6°)

Fig. 10　Streamlines for case 1 and case 2 at same angles of attack (α = 9.6°)

2.2　迎角振幅的影响

算例 2 和算例 3 的平均迎角和缩减频率等参数相同, 仅迎角振幅不同, 下面以算例 2 和算例 3 的数值模拟结果分析迎角振幅对翼伞翼型非定常气动特性的影响。

数值模拟获得的算例 2 和算例 3 的升力系数曲线如图 11 所示。其中, 算例 2 的升力系数曲线变化特征已在 2.1 节中分析。算例 3 与算例 2 相比, 平均迎角和缩减频率一致, 迎角振幅小 2°, 即俯仰过程中的最大迎角和最小迎角分别小 2°。两者的升力系数曲线在迎角 8°以下较为接近, 差异主要体现在迎角 8°以上。算例 2 的最大升力系数及临界迎角仅略大于算例 1, 具体可见图 11 中升力系数峰值点 A_2 与点 A_3 之间的位置差异。升力系数的低点 B_2 和点 B_3 则呈现大的位置差异, 表明在这 2 个算例的计算条件下, 相对于升力系数峰值点, 升力系数低点对迎角振幅的变化更敏感。

图 12 显示了算例 2 和算例 3 分别在点 A_2、B_2、C_2 和点 A_3、B_3、C_3 的流线图。由图 12(a) 和图 12(d)可见, 在上仰阶段的升力系数峰值点处, 2 个算例的

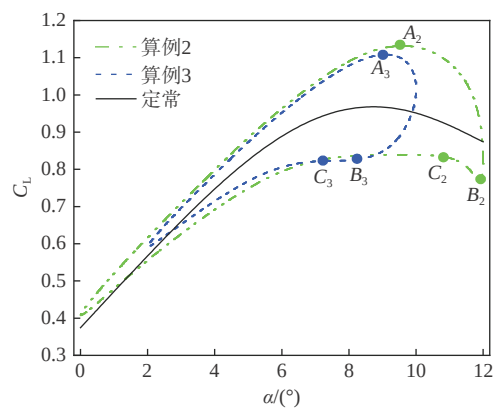
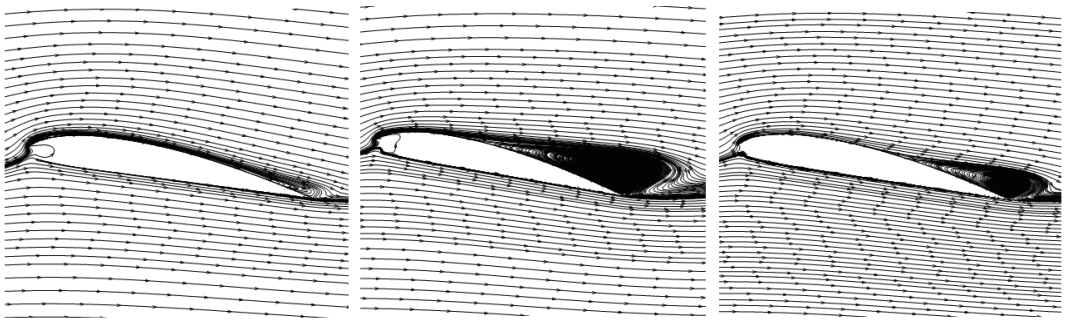


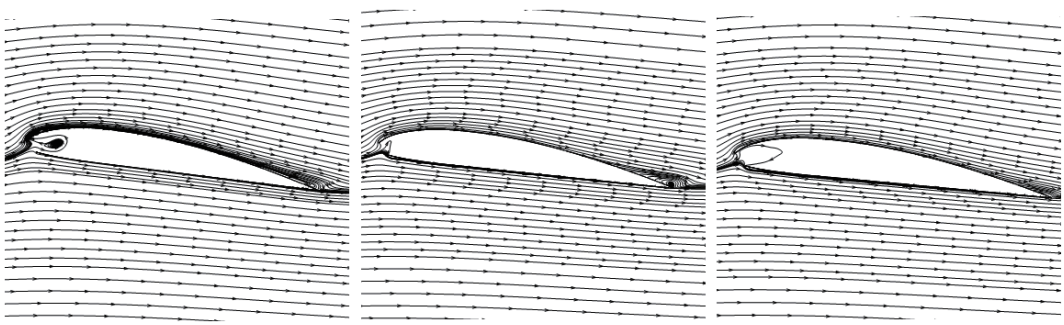
图 11　算例 2 和算例 3 的升力系数曲线

Fig. 11　Lift coefficient curves for case 2 and case 3

流动形态相近, 流动分离不明显; 由图 12(b) 和图 12(e)可见, 算例 2 在点 B_2 处呈现明显的尾缘分离, 而算例 3 在点 B_3 处的流动分离不明显, 这是 2 个算例的升力系数低点产生显著差异的流场原因; 由图 12(c)和图 12(f) 可见, 算例 2 在点 C_2 的尾缘分离涡已经显著减小, 而算例 3 在点 C_3 处没有观察到明显的气流分离。



(a) 点A_2 (α=9.6°)↑ (b) 点B_2 (α=11.9°)↓ (c) 点C_2 (α=11.0°)↓



(d) 点A_3 (α=9.3°)↑ (e) 点B_3 (α=8.2°)↓ (f) 点C_3 (α=7.4°)↓

图 12 算例 2 和算例 3 在不同迎角时的流线图

Fig. 12 Streamlines for case 2 and case 3 at various angles of attack

2.3 缩减频率的影响

算例 2、算例 4 和算例 5 的平均迎角和迎角振幅等参数相同，仅缩减频率不同，下面以算例 2、算例 4 和算例 5 的数值模拟结果分析缩减频率对翼伞翼型非定常气动特性的影响。

数值模拟获得的算例 2、算例 4、算例 5 的升力系数曲线如图 13 所示。可见，缩减频率越大，升力系数曲线迟滞环的面积越大，最大升力系数 ΔC_{Lmax} 越大。为丰富数据以便分析规律，对更多缩减频率（K=0.003、K=0.007、K=0.009、K=0.015、K=0.02）的情况开展了数值模拟。各缩减频率下数值模拟得到的最大升力系数数据如表 3 所示，其中，ΔC_{Lmax} 为与定常模拟相比的最大升力系数的变化比例。图 14 显示了最大升力系数 C_{Lmax} 随缩减频率 K 的

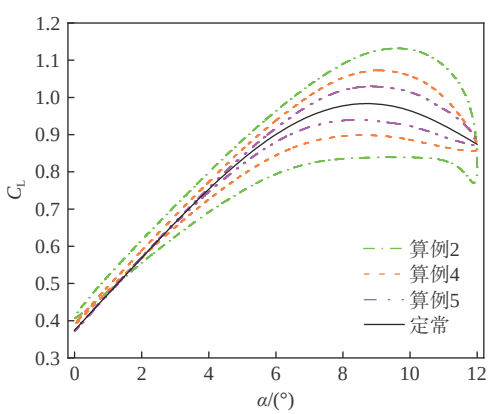


图 13 算例 2、算例 4 和算例 5 的升力系数曲线

Fig. 13 Lift coefficient curves for case 2, case 4, and case 5

表 3 最大升力系数随缩减频率变化的计算数据

Table 3 Variation of maximum lift coefficient with reduced frequency

缩减频率/K	最大升力系数C_{Lmax}	ΔC_{Lmax}/%
0(定常)	0.977	
0.003	0.986	0.9
0.007	0.999	2.2
0.009	1.003	2.7
0.01	1.005	2.9
0.015	1.021	4.5
0.02	1.038	6.3
0.03	1.073	9.8
0.051	1.132	15.9

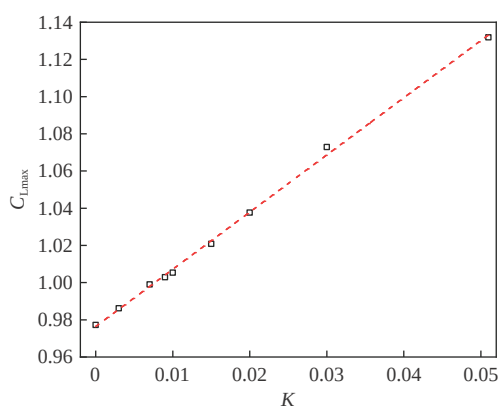


图 14 最大升力系数随缩减频率变化曲线

Fig. 14 Variation of maximum lift coefficient with reduced frequency

变化。可见, C_{Lmax} 随缩减频率的增大而增大, 并且呈线性增长趋势, 图中, K=0 为定常计算结果。这种线性规律的存在有利于通过数学建模描述不同缩减频率下翼伞翼型的非定常气动力。

3　非定常流场分析

从第 2 节的分析可以看出, 算例 1 的非定常效应最为明显, 因此, 对算例 1 在一个俯仰周期中的典型流动特征进行分析。具体选取 8°、12°和 14°这 3 个典型迎角, 分别作出该迎角下定常状态、上仰运动和下附运动时的流线图, 通过对比分析获得非定常流动特征。图 15显示了算例 1 的非定常升力系数曲线及定常状态的升力系数曲线, 并标示出了将要分析的 3 个迎角。之所以选择这 3 个迎角是因为: 迎角为 8°时, 定常升力系数曲线接近峰值, 是尚未失速但临近失速的典型状态, 而此时非定常上仰状态的升力系数仍处于增长趋势, 非定常下附状态的升力系数处于低位, 3 种状态的显著差异性值需从流场的角度进行探究; 迎角为 12°时, 定常状态的升力系数已经处于下降阶段, 是典型的失速状态, 有利于探讨上仰和下附运动对翼伞翼型失速特性的影响; 迎角 14°是算例 1 俯仰运动的最大迎角, 非定常升力系数的数值在该迎角附近从最大值急剧下降, 有必要在该迎角下对比非定常与定常的流场差异。

图 16 显示了迎角 8°时翼伞翼型分别处于定常、上仰和下俯状态的流场图。可见, 定常状态和非定常上仰状态均未出现明显的流动分离, 非定常下俯状态缘处出现较小的分离涡。表明该迎角下流动分离尚不显著, 上仰导致的升力系数增大, 下俯导致的升力系数减小, 主要受翼型附着绕流的非定常效应的影响。

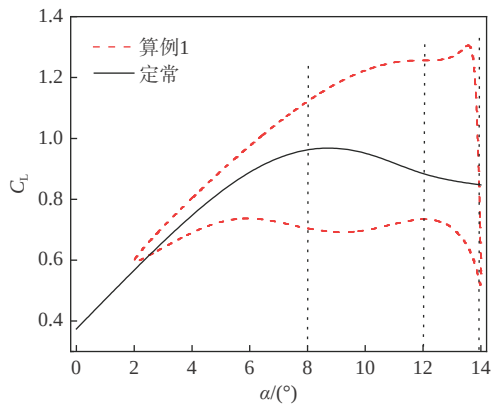
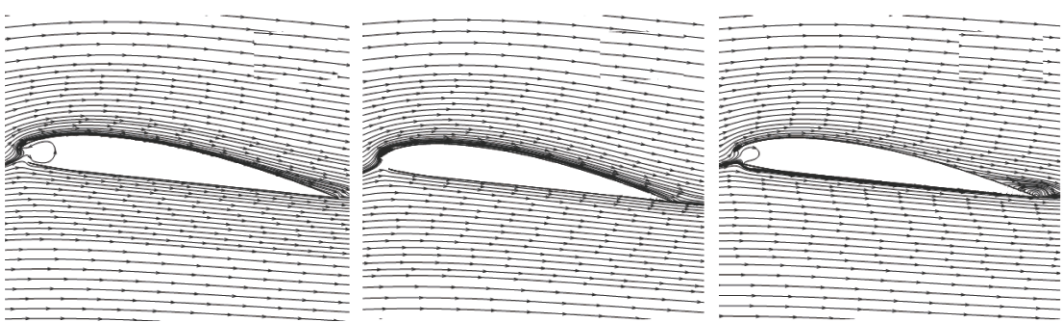


图 15　算例 1 的升力系数曲线

Fig. 15　Lift coefficient curve for case 1



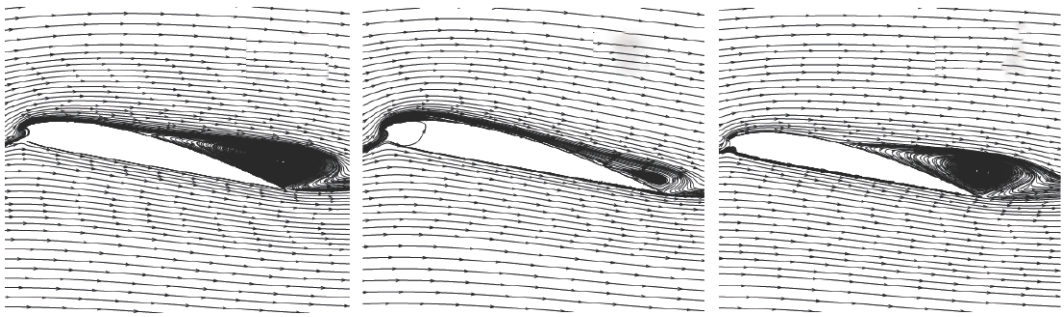
(a) 定常　　(b) 上仰　　(c) 下俯

图 16　α=8°时, 算例 1 与定常计算的流线图

Fig. 16　Streamlines for case 1 and case of steady simulation when α = 8°

图 17 显示了迎角 12°时翼伞翼型分别处于定常、上仰和下俯状态的流场图。可知, 3 个状态都出现了不同程度的流动分离, 其中, 非定常上仰状态在翼型的前缘和尾缘附近出现尺寸相对小的分离涡, 定常状态和非定常下俯状态出现较大尺度的尾缘分离涡。表明该迎角下的上仰运动抑制了翼伞翼型的尾缘分离, 但一定程度促进了前缘分离, 而下俯运动对流动分离的影响不显著。

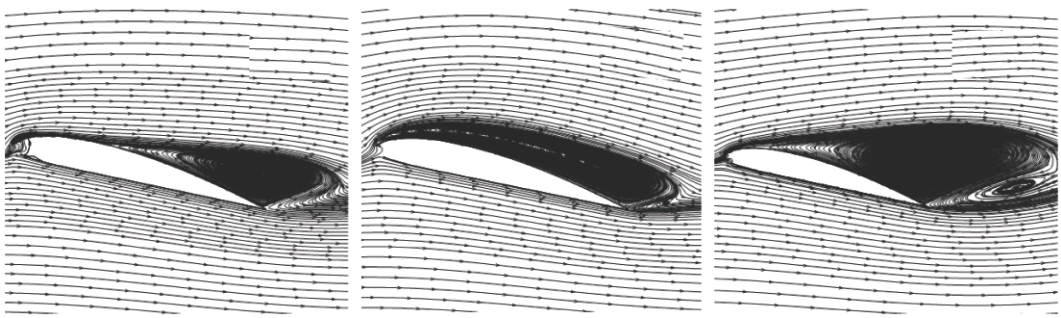
图 18 显示了迎角 14°附近翼伞翼型分别处于定常、上仰和下俯状态的流场图, 其中, 上仰状态取升力系数峰值点处(对应迎角为 13.6°)的对应流场, 定常状态和下附状态的对应迎角为 14°。可知, 3 种状态全部出现大尺度的分离涡, 其中, 非定常上仰和下俯状态已达到从前缘到尾缘的完全流动分离, 分离涡覆盖了翼型的整个上表面。但上仰状态的分离涡形状具有特殊性, 其呈现明显的扁平状, 局限于翼型上表面的临近区域。这表明非定常上仰对流动分离的作用, 不仅体现在影响流动分离的发生, 还体现在显著影响分离涡的形状和尺寸。由上仰阶段转变为下俯阶段时, 翼型上翼面的分离涡体积明显增大, 对应的升力系数急剧减小, 由 1.3 左右降低到 0.5 左右(见图 15)。



(a) 定常 (b) 上仰 (c) 下俯

图 17 α=12°时,算例 1 与定常计算的流线图

Fig. 17 Streamlines for case 1 and case of steady simulation when α = 12°



(a) 定常 (b) 上仰 (c) 下俯

图 18 α=14°附近时,算例 1 与定常计算的流线图

Fig. 18 Streamlines for case 1 and case of steady simulation when α = 14°

4 结 论

为研究翼伞翼型在迎角非定常变化过程中的气动力和流场特征,本文采用 CFD 数值模拟方法对前缘开口的 Clark-Y 翼型进行了非定常俯仰运动的气动特性数值模拟研究,分析了平均迎角、迎角振幅和缩减频率对翼伞翼型非定常气动力的影响,并探讨了对应的流场特征。结论如下:

1) 翼伞翼型的升力在迎角动态变化过程中体现出显著的非定常特性,升力系数在上仰段普遍大于定常情况,在下俯段普遍小于定常情况;平均迎角、迎角振幅和缩减频率 3 个参数中的任一个增大,都导致该非定常升力特性更显著。

2) 平均迎角同迎角振幅一起决定了翼型迎角动态变化的范围,当运动所达的最大迎角(平均迎角+正幅值)增大到导致流动分离发生时,翼伞翼型的动态失速特性显著,主要表现在:临界迎角和最大升力系数明显增大,升力系数在最大迎角附近呈现急剧下降。

3) 缩减频率的变化对翼伞翼型的升力迟滞效应产生显著影响,缩减频率越大,则迟滞环面积越大,且最大升力系数随着缩减频率的增加而呈线性增大,该线性关系有利于对翼伞翼型的非定常气动力进行预测和建模。

4) 翼伞翼型的非定常流场与定常流场的对比分析表明,翼伞翼型迎角动态增大阶段(上仰阶段)的非定常效应:一方面,延缓了翼型流动分离的发生;另一方面,还使分离涡紧贴上翼面呈扁平形状。这 2 个方面的作用导致处于动态失速的翼伞翼型的临界迎角和最大升力系数相对定常情况显著增大。

以上结论有助于增进对翼伞翼型非定常气动力和流场规律的理解,为复杂风环境下的翼伞非定常气动力预测和安全性评估提供支撑,为考虑非定常效应的翼伞设计提供参考。

参考文献(References)

[1] 李春, 吕智慧, 黄伟, 等. 精确定点归航翼伞控制系统的研究[J]. 中南大学学报 (自然科学版), 2012, 43(4): 1331-1335.
LI C, LÜ Z H, HUANG W, et al. Guidance navigation & control system for precision fix-point homing parafoil[J]. Journal of Central South University (Science and Technology), 2012, 43(4): 1331-1335(in Chinese).

[2] TAO J, SUN Q L, ZHU E L, et al. Quantum genetic algorithm based homing trajectory planning of parafoil system[C]//Proceedings of the 2015 34th Chinese Control Conference. Piscataway: IEEE Press, 2015: 2523-2528.

[3] 高振兴. 复杂大气扰动下大型飞机飞行实时仿真建模研究[D]. 南京: 南京航空航天大学, 2009.
GAO Z X. Research on real-time flight simulation of large aircraft in complex atmospheric disturbance[D]. Nanjing: Nanjing Uni-

versity of Aeronautics and Astronautics, 2009(in Chinese).

[4] MITTAL S, SAXENA P, SINGH A. Computation of two-dimensional flows past ram-air parachutes[J]. International Journal for Numerical Methods in Fluids, 2001, 35(6): 643-667.

[5] BALAJI R, MITTAL S, RAI A K. Effect of leading edge cut on the aerodynamics of ram-air parachutes[J]. International Journal for Numerical Methods in Fluids, 2005, 47(1): 1-17.

[6] 聂帅, 曹义华, 田似营. 前缘切口对翼伞气动性能的影响[J]. 航空动力学报, 2016, 31(6): 1477-1485.
NIE S, CAO Y H, TIAN S Y. Effect on parafoil aerodynamic performance of leading edge[J]. Journal of Aerospace Power, 2016, 31(6): 1477-1485(in Chinese).

[7] 李健. 前缘切口对冲压式翼伞的气动力影响[J]. 航天返回与遥感, 2005, 26(1): 36-41.
LI J. The aerodynamic influence of the cutter of the front edge of parafoil[J]. Spacecraft Recovery & Remote Sensing, 2005, 26(1): 36-41(in Chinese).

[8] FURFARO J, HAILEY C, BEHR V, et al. An experimental study of parafoil sections[C]//Proceedings of the14th Serodynamic Decelerator Systems Technology Conference. Reston: AIAA, 1997: 1527.

[9] UDDIN M, MASHUD M. Wind tunnel test of a paraglider (flexible) wing canopy International[J]. Journal of Mechanical & Mechatronics Engineering, 2010, 10(3): 7-18.

[10] 朱旭, 曹义华. 翼伞弧面下反角、翼型和前缘切口对翼伞气动性能的影响[J]. 航空学报, 2012, 33(7): 1189-1200.
ZHU X, CAO Y H. Effects of arc-anhedral angle, airfoil and leading edge cut on parafoil aerodynamic performance[J]. Acta Aeronautica et Astronautica Sinica, 2012, 33(7): 1189-1200(in Chinese).

[11] 王国庆, 滕海山, 王奇. 大型翼伞操纵转弯动力学研究[J]. 航天返回与遥感, 2021, 42(2): 49-58.
WANG G Q, TENG H S, WANG Q. Dynamics research on maneuvering and turning of the large parafoil[J]. Spacecraft Recovery & Remote Sensing, 2021, 42(2): 49-58(in Chinese).

[12] 孙青林, 梁炜, 陶金, 等. 基于 CFD 的风雨环境翼伞动力学建模[J]. 中南大学学报 (自然科学版), 2017, 48(8): 2053-2062.
SUN Q L, LIANG W, TAO J, et al. Dynamic modeling of parafoil in wind and rain environment based on CFD method[J]. Journal of Central South University (Science and Technology), 2017, 48(8): 2053-2062(in Chinese).

[13] RAFFEL M, KOMPENHANS J, WERNERT P. Investigation of the unsteady flow velocity field above an airfoil pitching under deep dynamic stall conditions[J]. Experiments in Fluids, 1995, 19(2): 103-111.

[14] 宋红军, 王肇, 尹协振. 二维翼型非定常运动数值模拟[J]. 水动力学研究与进展 (A 辑), 2004, 19(增刊 1): 896-903.
SONG H J, WANG Z, YIN X Z. Numerical simulation on unsteady motion of 2D airfoil[J]. Chinese Journal of Hydrodynamics, 2004, 19(Sup 1): 896-903(in Chinese).

[15] 闫文辉. NACA4412 翼型低速绕流的定常/非定常计算对比研究[J]. 科学技术与工程, 2017, 17(2): 283-287.
YAN W H. Steady/unsteady numerical study of NACA4412 airfoil around flow[J]. Science Technology and Engineering, 2017, 17(2): 283-287(in Chinese).

[16] 宗涛. 风力机翼型动态失速的数值模拟[J]. 湖北电力, 2017, 41(4): 50-54.
ZONG T. Numerical simulation of airfoil dynamic stall of wind turbine[J]. Hubei Electric Power, 2017, 41(4): 50-54(in Chinese).

[17] 钱炜祺, 符松, 蔡金狮. 翼型动态失速的数值研究[J]. 空气动力学学报, 2001, 19(4): 427-433.
QIAN W Q, FU S, CAI J S. Numerical study of airfoil dynamic stall [J]. Acta Aerodynamica Sinica, 2001, 19(4): 427-433(in Chinese).

[18] TUNCER I, PLATZER M. A computational study of flow separation characteristics and wake profiles behind a flapping airfoil[C]// Proceedings of the 37th Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 1999: 648.

[19] TUNCER I, WALZ R, PLATZER M. A computational study on the dynamic stall of a flapping airfoil[C]//Proceedings of the 16th AIAA Applied Aerodynamics Conference. Reston: AIAA, 1998: 2519.

[20] MENTER F R. Two-equation eddy-viscosity turbulence models for engineering applications[J]. AIAA Journal, 1994, 32(8): 1598-1605.

[21] MENTER F, RUMSEY C. Assessment of two-equation turbulence models for transonic flows[C]//Proceedings of the Fluid Dynamics Conference. Reston: AIAA, 1994: 2343.

[22] ELGAMMI M, SANT T. A new stall delay algorithm for predicting the aerodynamics loads on wind turbine blades for axial and yawed conditions[J]. Wind Energy, 2017, 20(9): 1645-1663.

[23] 续荣华, 王震, 黄及水, 等. 上翼面开缝的翼伞翼型气动特性研究[J]. 航天返回与遥感, 2022, 43(3): 1-11.
XU R H, WANG Z, HUANG J S, et al. Aerodynamic performance study on a parafoil airfoil with an upper surface slit[J]. Spacecraft Recovery & Remote Sensing, 2022, 43(3): 1-11(in Chinese).

[24] EKATERINARIS J A, SØRENSEN N N, RASMUSSEN F. Numerical investigation of airfoil dynamic stall in simultaneous harmonic oscillatory and translatory motion[J]. Journal of Solar Energy Engineering, 1998, 120(1): 75-83.

[25] GUILMINEAU E, PIQUET J, QUEUTEY P. Unsteady two-dimensional turbulent viscous flow past aerofoils[J]. International Journal for Numerical Methods in Fluids, 1997, 25(3): 315-366.

[26] RAMSAY R F, HOFFMAN M J, GREGOREK G M. Effects of grit roughness and pitch oscillation on the S809 airfoil[R]. Golden: National Renewable Energy Laboratory, 1995.

Numerical simulation of unsteady aerodynamic characteristics of parafoil airfoil

WANG Zhen[1, 2], ZHONG Wei[1, 2, *], WANG Tongguang[1, 2], LI Xudong[1, 2], ZHANG Hongying[1]

(1. College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;
2. Jiangsu Key Laboratory of Hi-Tech Research for Wind Turbine Design, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: The unsteady aerodynamic characteristics of parafoils are a matter of flight stability and flight safety and are worthy of in-depth study. In this paper, the computational fluid dynamics (CFD) method was used to carry out an unsteady numerical simulation of the parafoil airfoil under the dynamic change of the angle of attack, so as to analyze the influence of the average angle of attack, the amplitude of the angle of attack, and the reduced frequency on the unsteady aerodynamic force and flow field pattern of the parafoil airfoil. The results show that an increase in any one of the above three parameters leads to a significant increase in the aerodynamic unsteadiness of the parafoil airfoil. The average angle of attack and the amplitude of the angle of attack amplitude jointly determine the dynamic range of the angle of attack of the airfoil, which significantly affects the dynamic stall characteristics of the parafoil airfoil; the change in reduced frequency affects the lift hysteresis effect of the parafoil airfoil, with the maximum lift coefficient increasing linearly with increasing reduced frequency. Further flow field analysis shows that the unsteady effect of the angle of attack of the parafoil airfoil in the dynamic increase stage delays the occurrence of airfoil flow separation on the one hand, and on the other hand, it flattens the separating vortex against the upper airfoil, which together contributes to a significant increase in the critical angle of attack and maximum lift coefficient of the parafoil airfoil relative to the constant case. The results of this essay will help to enhance the understanding of the unsteady aerodynamic force and flow field patterns of parafoil airfoil, support the prediction of unsteady aerodynamic forces and safety assessment of parafoil in complex wind environments.

Keywords: parafoil; airfoil; aerodynamic performance; dynamic stall; reduced frequency

Received: 2023-04-18; **Accepted**: 2023-07-03; **Published Online**: 2023-07-12 09:00
URL: link.cnki.net/urlid/11.2625.V.20230711.2149.003
* **Corresponding author**. E-mail: zhongwei@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0191

基于不确定理论的网络时间可靠性评估方法

马骥[1]，李瑞莹[1, 2, *]，张清源[3]，康锐[1, 2, 3]

(1. 北京航空航天大学 可靠性与系统工程学院，北京 100191；2. 可靠性与环境工程技术重点实验室，北京 100191；
3. 杭州市北京航空航天大学国际创新研究院（北京航空航天大学国际创新学院），杭州 311115)

摘 要：针对现有的网络时间可靠性评估方法往往只考虑固有不确定性的影响，而忽略由于故障信息不足而导致的认知不确定性对可靠性评估结果影响的问题，提出一种基于不确定理论的新方法。根据网络可靠性关注的节点范围，设计了单节点对和多节点对时间可靠度两类度量参数，提出了扩展不确定网络模型可直接实现对节点和链路上的认知不确定性特征建模，并进一步设计了基于最可靠路径和最可靠扩展不确定子网络的单节点和多节点对时间可靠度算法。分别以一个六节点网络、中国教育和科研计算机网 (CERNET) 骨干网络为例，应用所提方法实现了2种时间可靠性指标的评估，验证了所提方法的正确性和有效性。

关 键 词：时间可靠性评估；认知不确定性；扩展不确定网络；不确定理论；确信可靠性
中图分类号：TB114.3
文献标志码：A **文章编号**：1001-5965(2025)04-1267-10

通信、电力、交通等网络设施为人们的生产生活带来了极大的便捷。为分析这类网络化结构系统的可靠性水平，研究者开展了大量研究。最早的网络可靠性研究集中在连通可靠性方面，其关注在规定时间和规定条件下网络保持连通的能力[1]。连通可靠性主要考查网络中节点和链路故障对网络互联互通的影响。考虑到建立网络的目的是要传输信息、物质和能量，除了满足连通要求，网络还需要满足这些负载对网络性能的要求[2-3]。作为网络功能体现，性能(时间、容量、丢包率等)是网络在工程实践中着重考察的内容。因此，研究人员在连通可靠性的基础上加入性能约束条件，逐渐形成了性能可靠性概念[4]。性能可靠性是指在规定条件下和规定时间内，网络提供的性能满足传输要求的能力。目前，常见的性能可靠性参数有时间可靠性，容量可靠性等，其中，时间可靠性作为网络中重要的可靠性参数，关注的是信息、物质和能量在网络中传输时间满足传输时间阈值的能力[4]。

时间可靠性，在通信网络和交通网络里也称为及时可靠性或行程时间可靠性，研究人员提出了很多评估方法。Liu 等[5] 将通信网络中时延表征为离散随机变量并建立随机流网络，基于概率用户均衡模型和时延失效模型，提出通信网络时间可靠性定义和算法。Wu 等[6] 考虑航空电子全双工交换以太网中流量随机性对时延的影响，提出一种基于网络演算的时间可靠性概率算法。Wang 和 Guo[7] 考虑随机攻击和故意攻击对交通网络的影响，给出基于旅行时间系数和延迟系数的时间可靠度定义，基于复杂网络理论对交通网络的时间可靠性进行了评估。Zhang 等[8] 提出线性量化混合模型，分析不同因素对旅行时间可靠性的影响并对高速公路的时间可靠度进行了评估。

上述时间可靠性评估方法均以概率论为基础，计算网络传输时间满足用户要求的概率。在这些

收稿日期：2023-04-21；**录用日期**：2023-07-20；**网络出版时间**：2023-07-28 17:01
网络出版地址：link.cnki.net/urlid/11.2625.V.20230728.1358.001
基金项目：国家自然科学基金 (61773044,62073009)；可靠性与环境工程技术重点实验室基金 (WDZC2019601A301)
* **通信作者**. E-mail：liruiying@buaa.edu.cn

引用格式：马骥，李瑞莹，张清源，等. 基于不确定理论的网络时间可靠性评估方法 [J]. 北京航空航天大学学报，2025，51（4）：1267-1276. MA J，LI R Y，ZHANG Q Y，et al. Network time reliability evaluation method based on uncertainty theory [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1267-1276（in Chinese）.

方法中,故障数据是开展可靠性评估的基础。当故障数据充足时,由于符合概率论中的大数定律条件,可以应用概率论评估网络可靠性,此时系统主要表现为固有不确定性(即随机不确定性)。然而,当故障信息不足时(单个或几个故障数据),此时系统存在认知不确定性[9]。例如,对于一些新研网络,受成本和环境等因素影响,很难获取足够的故障数据[10-13]。这种情况下就无法应用概率论进行网络可靠性评估。

为量化认知不确定性,一些数学理论相继被提出,如证据理论[14]、模糊理论[15]和区间理论[16]等。Mi等[17]开发了一种基于证据网络的网络可靠性分层算法。Qi和Cheng[18]采用区间理论来评估发电网络的不精确可靠性。Zhang等[19]基于模糊理论对网络的旅行时间可靠性进行了评估。然而,模糊理论不满足对偶性,会造成反直觉的可靠性评估结果;基于证据理论或区间理论的方法存在区间扩张等问题,对网络特别是对于大型网络会造成可靠性评估结果不准确[20];为此,Liu[21]在2007年提出不确定理论。不确定理论将只有有限信息的事件表述为不确定变量,用不确定测度(非频率)来描述该事件发生的可能性。由于不确定测度服从对偶性公理和乘积公理,因此,基于不确定理论的可靠性度量可以克服上述理论中存在的对偶性问题及区间扩张的问题。Kang等[22-23]基于不确定理论、概率论和机会理论[24]提出确信可靠性理论,分别可以解决系统存在认知不确定性、固有不确定性和2种不确定性混合场景下的系统可靠性度量问题。在不确定理论描述的认知不确定性分支,Wang等[25]采用不确定变量表达原油海运网络中各个节点的连通性,建立了最可靠路径选择模型,设计了网络连通可靠性评估方法。Hosseini和Pishvaee[26]考虑了灾后交通网络中链路通断和容量的认知不确定性,基于不确定理论对网络的容量可靠性进行了评估。然而由于没有考虑时间约束,且只能评估2个节点之间的端端可靠性,上述方法无法直接应用到时间可靠性评估中。

因此,基于上述分析,本文基于不确定理论对受认知不确定性影响的网络设计时间可靠性度量参数和评估方法。首先,考虑网络特征,给出了单节点对和多节点对时间可靠度定义;然后,建立扩展不确定网络描述受认知不确定性影响的网络,并基于该模型给出了单节点对和多节点时间可靠度算法;最后,以一个简单网络验证了本文方法的准确性并基于某个教育网骨干网络作为案例给出了时间可靠性评估流程。本文方法能够解决认知不确定性条件下网络时间可靠性量化问题,既适用于双向网络,也适用于单向网络。

1 时间可靠性

在确信可靠性理论中,系统确信可靠性是指系统功能的性能裕量大于0的能力[22-23]。其中,性能裕量 $m(t)$ 是系统的性能指标与需求阈值之间的距离,定义如下[27]:

$$m(t)=\begin{cases}p_{\mathrm{th}}-p(t) & p(t)\text{为望小型参数}\\ p(t)-p_{\mathrm{th}} & p(t)\text{为望大型参数}\\ \min(p_{\mathrm{th,U}}-p(t),p(t)-p_{\mathrm{th,L}}) & \\ & p(t)\text{为望目型参数}\end{cases} \tag{1}$$

式中:$p(t)$ 表示系统在 t 时刻的性能参数;p_{th} 表示系统中性能参数阈值,即故障判据;$p_{\mathrm{th,U}}$ 和 $p_{\mathrm{th,L}}$ 分别表示望目型参数阈值的上界和下界。

当 $m(t)>0$ 时,系统的功能满足要求。当系统故障特征存在认知不确定性特征时,其确信可靠度为系统性能裕量大于0的不确定测度,即

$$R_{\mathrm{B}}^{\mathrm{U}}(t)=\mathcal{M}\{m(t)>0\} \tag{2}$$

传输时间是通信网络、交通网络的关键性能参数。基于确信可靠性理论,对于受认知不确定性影响的网络,其时间可靠度为时间裕量大于0的不确定测度。通常,网络可靠性参数还会根据所考察的节点范围进一步分为:端端可靠度、K 端可靠度和全端可靠度。其中,K 端可靠度是指节点集 K 内所有节点之间实现规定功能的概率[28],而端端可靠度和全端可靠度则分别是 K 端可靠度在“节点集 K 只有2个节点”和“节点集 K 为网络中所有节点”情况下的特例[29]。然而,在一些应用场景中,比起关注特定用户群,本文更关心网络能满足多少百分比用户的需求,即关注网络中可靠的节点对数(2个节点为一组节点对)。因此,为了能够表述出不同节点对要求范围下的网络时间可靠度,本文分别定义了单节点对和多节点对时间可靠度。

定义1 单节点对时间可靠度。单节点对时间可靠度是 t 时刻指定节点 v_{s} 与节点 v_{t} 之间的传输时间裕量 $m_{\mathrm{sn}}(t)$ 大于0的不确定测度。

$$R_{\mathrm{sn}}(t)=\mathcal{M}\{m_{\mathrm{sn}}(t)>0\}=\mathcal{M}\{D_{\mathrm{th}}-D_{\mathrm{sn}}(t)>0\} \tag{3}$$

式中:$D_{\mathrm{sn}}(t)$ 表示在 t 时刻节点 v_{s} 与节点 v_{t} 之间的传输时间;D_{th} 表示传输时间阈值。

定义2 多节点对时间可靠度。多节点对时间可靠度是 t 时刻网络中至少 $k(1\leqslant k\leqslant K,K$ 为网络中总节点对数)组节点对的传输时间裕量 $m_{\mathrm{mn}(k)}(t)$ 大于0的不确定测度:

$$R_{\mathrm{mn}(k)}(t)=\mathcal{M}\{m_{\mathrm{mn}(k)}(t)>0\}=\mathcal{M}\{D_{\mathrm{th}}-D_{\mathrm{mn}(k)}(t)>0\} \tag{4}$$

式中: $D_{\mathrm{mn}(k)}(t)$表示 t 时刻 k 组节点对的传输时间, 其值为网络中所有 K 组节点对传输时间的第 k 小值; D_{th}表示传输时间阈值。

需要指出的是, 单节点对时间可靠度不等同于 k=1 时的多节点对时间可靠度, 这是因为前者关注指定节点对, 而后者关注任意一组节点对。此外, 当 k=K 时的多节点对时间可靠度关注的是网络中所有节点之间的传输时间都满足要求的不确定测度。

2　时间可靠性模型与算法

首先, 本节给出网络时间可靠性模型, 然后, 基于该模型分别给出单节点对与多节点对时间可靠度算法。为方便表达, 本文将表达式中的 t 略去, 下文中提及的 R、D 均是指 t 时刻下的可靠度和传输时间。

本文假设:

1) 网络中节点和链路只有故障和正常 2 种状态, 不存在性能降级状态。

2) 网络中各节点、链路上的故障相互独立。

3) 网络中信息、物质和能量传输不存在循环情况。

2.1　模　型

对于受认知不确定性影响的网络, Liu[30] 提出不确定网络的概念, 其描述了网络中链路权重的认知不确定性。然而现实世界中的网络, 节点也有权重且可能存在认知不确定性。虽然可以通过图变换的方式, 将节点扩展成链路, 用链路的权重和认知不确定性表征节点相应特征, 即通过扩展网络规模建立等效的不确定网络, 但是网络规模的扩大无疑会使后续的相关计算变得更为复杂。因此, 本文在不确定网络的基础上定义了扩展不确定网络, 用于描述链路和节点权重存在认知不确定性的网络。下面给出扩展不确定网络的定义。

定义 3　扩展不确定网络。扩展不确定网络是由节点集合 V, 链路集合 E 及节点与链路集合对应的不确定变量集合 W 组成的三元组, 记作 $G(V,E,W)$。

通常来说, $V=\{v_1,v_2,\cdots,v_N\}$, $E=\{e_1,e_2,\cdots,e_M\}$, $W=\{w_1,w_2,\cdots,w_{N+M}\}$。在该网络中, V 和 E 定义了网络的结构, W 表示链路和节点上的权重, 为不确定变量集合。其中, $w_1,w_2,\cdots,w_N$表示节点 1,2,$\cdots$, N上的权重, $w_{N+1},w_{N+2},\cdots,w_{N+M}$表示链路 1,2,$\cdots$, M上的权重。

基于扩展不确定网络定义可以建立时间可靠性模型$G(V,E,D)$, 其中, D 表示链路和节点上的传输时间集合, 具体建模过程如下:

考虑到节点和链路只有正常和故障 2 种状态, 本文将节点和链路状态(正常或故障)表述为布尔不确定变量ξ_i,i=1,2,$\cdots$,N+M。基于专家经验或节点与链路的故障数据, 采用不确定统计方法[21] 可以得到ξ_i分别为 1 或 0 时的不确定测度。对于ξ_i, 有

$$\xi_i=\begin{cases}1 & \text{节点/链路正常, 不确定测度为}\alpha_i\\ 0 & \text{节点/链路故障, 不确定测度为}1-\alpha_i\end{cases} \tag{5}$$

收集无故障情况下各个节点和链路上的传输时间, 记作d_i^{c}。此时节点或链路上传输时间d_i可以表示为d_i^{c}与ξ_i的乘积, 即$d_i=\xi_i\times d_i^{\mathrm{c}}$。对于$d_i$, 有

$$d_i=\begin{cases}d_i^{\mathrm{c}} & \text{节点/链路正常, 不确定测度为}\alpha_i\\ \infty & \text{节点/链路故障, 不确定测度为}1-\alpha_i\end{cases} \tag{6}$$

根据网络连接关系建立扩展不确定网络模型, 并建立传输时间邻接矩阵。传输时间邻接矩阵如下所示:

$$\boldsymbol{A}(\boldsymbol{G})=\begin{bmatrix}d_{1,1} & d_{1,2} & \cdots & d_{1,N}\\ d_{2,1} & d_{2,2} & \cdots & d_{2,N}\\ \vdots & \vdots & & \vdots\\ d_{N,1} & d_{N,2} & \cdots & d_{N,N}\end{bmatrix} \tag{7}$$

为方便邻接矩阵的表示, 根据节点和链路之间的连接关系将d_i转化为$d_{x,y}$, $1\leqslant x,y\leqslant N$。当 x=y 时, $d_{x,y}$表示节点 x 上的传输时间; 当$x\neq y$时, $d_{x,y}$表示节点 x 与节点 y 之间链路上的传输时间。

2.2　单节点对时间可靠度算法

根据定义 1, 计算扩展不确定网络的单节点对时间可靠度。首先, 需要确定扩展不确定网络中节点v_{s}与节点v_{t}之间的传输时间D_{sn}。在扩展不确定网络中, D_{sn}可以表示为

$$D_{\mathrm{sn}}=f_{\mathrm{sn}}(d_1,d_2,\cdots,d_{N+M}) \tag{8}$$

式中: $d_1,d_2,\cdots,d_{N+M}$表示扩展不确定网络 G 中节点和链路上的传输时间; f_{sn}为网络单节点对时间函数。在扩展不确定网络中, 当某个节点或链路上的传输时间增加时, 节点v_{s}和节点v_{t}之间的传输时间只能会出现 2 种可能情况: 保持不变或变大。因此, f_{sn}为一个单增函数。

根据式 (7) 和文献 [31] 中的定理, 可以计算网络中的单节点对时间可靠度。

定理 1[31]　假设 m 为有限的正整数, $\eta_1,\eta_2,\cdots,\eta_m$为独立的取有限值的不确定变量, 即对 i = 1, 2,$\cdots$,m, $\eta_i\in\Omega_i=\{b_1^{(i)},b_2^{(i)},\cdots,b_{h_i}^{(i)}\}$, 其中, $h_i\geqslant 2$, $b_1^{(i)}<$

$b_2^{(i)}<\cdots<b_{h_i}^{(i)}$。如果$f$为一个单调的函数，那么$\eta=f(\eta_1,\eta_2,\cdots,\eta_m)$的分布满足：

$$\mathcal{M}\{\eta\leqslant k\}=\sup_{f(B_1,B_2,\cdots,B_m)\subset(-\infty,k]}\min_{1\leqslant i\leqslant m}\mathcal{M}\{\eta_i\leqslant B_i\} \tag{9}$$

式中：B_i为集合$\{b_1^{(i)},b_2^{(i)},\cdots,b_{h_i}^{(i)}\}$的子集，$i=1,2,\cdots,m$。

根据定理1计算单节点对时间可靠度，具体过程如下：首先，枚举出扩展不确定网络中所有满足时间阈值要求的子网络；然后，计算每个子网络中各个节点和链路都正常情况下的不确定测度；最后，对这些不确定测度取大，得到的值即为单节点对时间可靠度。然而对于较大规模网络，逐一枚举子网络将花费很长的计算时间。因此，上述算法无法直接应用到较大规模网络中。为解决上述问题，本文提出一个更为便捷的单节点对时间可靠度算法：基于最可靠路径的单节点对时间可靠度。

首先，给出最可靠路径的定义。

定义4 最可靠路径。p^*为扩展不确定网络G中从节点v_s到节点v_t的一条满足传输时间阈值要求的路径。如果对于扩展不确定网络G中任意一条从节点v_s到节点v_t且满足传输时间阈值要求的路径p，都存在$R_{\mathrm{sn},p^*}\geqslant R_{\mathrm{sn},p}$，那么称$p^*$为最可靠路径，其中，$R_{\mathrm{sn},p^*}$和$R_{\mathrm{sn},p}$分别是路径$p^*$和任意路径$p$的单节点对时间可靠度。

对于满足传输时间阈值要求的路径p，当该路径的所有节点和链路都连通时，该路径是可靠的。根据定理1，路径p的单节点对时间可靠度为

$$R_{\mathrm{sn},p}=\min_{a_i\in p}\mathcal{M}\{d_i=d_i^{\mathrm{c}}\} \tag{10}$$

式中：a_i表示路径p上的节点或链路；d_i^{c}表示节点或链路正常情况下的传输时间，当$d_i=d_i^{\mathrm{c}}$时表示该节点或链路是连通的。

根据最可靠路径定义，下面给出关于单节点对时间可靠度的定理。

定理2 p^*为扩展不确定网络G中节点v_s到节点v_t的最可靠路径，则有

$$R_{\mathrm{sn},G}=R_{\mathrm{sn},p^*} \tag{11}$$

式中：$R_{\mathrm{sn},G}$为扩展不确定网络G中节点v_s到节点v_t的单节点对时间可靠度；$R_{\mathrm{sn},p*}$为路径p^*的单节点对时间可靠度。

证明 假设在扩展不确定网络G中从节点v_s到节点v_t共有h条满足传输时间阈值要求的路径，分别为$p_1,p_2,\cdots,p_h$，这些路径的可靠度分别为$R_{\mathrm{sn},p_1},R_{\mathrm{sn},p_2},\cdots,R_{\mathrm{sn},p_h}$。在这$h$条路径中，只要有一条路径存在，那么节点$v_s$到节点$v_t$是可靠的。因此下式成立：

$$R_{\mathrm{sn},G}=R_{\mathrm{sn},p_1}\vee R_{\mathrm{sn},p_2}\vee\cdots\vee R_{\mathrm{sn},p_h} \tag{12}$$

根据定义4可知

$$R_{\mathrm{sn},p^*}=R_{\mathrm{sn},p_1}\vee R_{\mathrm{sn},p_2}\vee\cdots\vee R_{\mathrm{sn},p_h} \tag{13}$$

证毕

基于定理2，本文给出扩展不确定网络中节点v_s和节点v_t之间的单节点对时间可靠度算法，具体过程如下。

步骤1 基于扩展不确定网络$G(V,E,D)$，构造不考虑故障情况下的确定网络$G^{\mathrm{c}}(V,E,D^{\mathrm{c}})$，其中$D^{\mathrm{c}}=\{d_1^{\mathrm{c}},d_2^{\mathrm{c}},\cdots,d_{N+M}^{\mathrm{c}}\}$为确定性常量集合，表示无故障情况下各个节点和链路上的时间。

步骤2 在确定网络$G^{\mathrm{c}}(V,E,D^{\mathrm{c}})$中，基于深度优先搜索算法选出节点$v_s$和节点$v_t$之间所有满足传输时间阈值要求的路径$p_1,p_2,\cdots,p_h$。

步骤3 基于扩展不确定网络$G(V,E,D)$，计算出路径$p_1,p_2,\cdots,p_h$对应时间可靠度$R_{\mathrm{sn},p_1},R_{\mathrm{sn},p_2},\cdots,R_{\mathrm{sn},p_h}$。其中，时间可靠度值最大的路径为最可靠路径，最大值为节点v_s和节点v_t的时间可靠度。

单节点对时间可靠度算法的复杂度主要为求解步骤2。由于本文将扩展不确定网络用邻接矩阵表示，基于深度优先搜索算法思想计算指定节点间所有满足要求的路径。此时单节点对时间可靠度算法的复杂度为$\mathrm{O}(N^2)$，其中，N表示网络中的节点数。

2.3 多节点对时间可靠度算法

根据定义2，计算扩展不确定网络G的多节点对时间可靠度。首先，需要得到k组节点对的传输时间$D_{\mathrm{mn}(k)}$。在扩展不确定网络中，存在

$$D_{\mathrm{mn}(k)}=f_{\mathrm{mn}(k)}(d_1,d_2,\cdots,d_{N+M}) \tag{14}$$

式中：$f_{\mathrm{mn}(k)}$为多节点对传输时间函数，是一个单增的函数。

和单节点对时间可靠度算法类似，借助定理1，可以对一些简单网络的多节点对时间可靠度进行评估，但却无法应用到大型网络中。为解决上述问题，本文给出多节点对时间可靠度算法。首先，给出可靠扩展不确定子网络和最可靠扩展不确定子网络的定义。

定义5 可靠扩展不确定子网络。$S_k=(V_k,E_k,D_k)$是扩展不确定网络$G(V,E,D)$的子网络，其中，$V_k\subseteq V$，$E_k\subseteq E$，$D_k\subseteq D$。当该子网络中所有链路和节点都存在时，有k组节点对满足传输时间阈值要求，且去掉任意一条链路或节点都不满足上述要求，那么该子网络称为可靠扩展不确定子网络。

图1为一个具有6个节点和8条链路的扩展不确定网络，各个节点和链路上无故障情况下的传输时间和对应的不确定测度如表1所示。该网络各

个节点之间的传输时间阈值均为 10 ms。那么, 当要求的节点对数为 3 时, 满足该要求的可靠扩展不确定子网络有 16 个, 本文给出 4 个示例, 如图 2 所示。

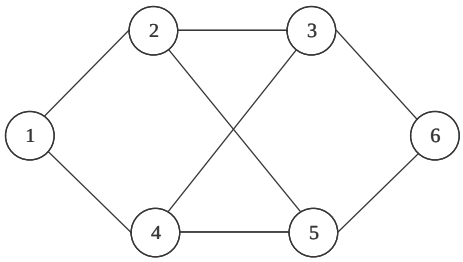


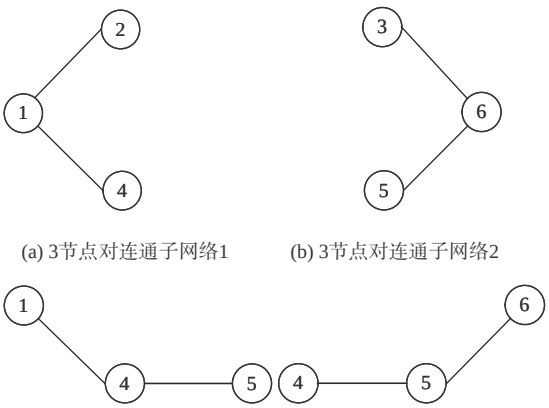
图 1　扩展不确定网络

Fig. 1　Extended uncertain network

表 1　节点和链路信息

Table 1　Node and link information

节点编号	节点时间/ms	节点不确定测度	链路编号	节点时间/ms	节点不确定测度
1	1	0.99	$e_{1,2}$	1.5	0.96
2	1.9	0.97	$e_{1,4}$	1	0.92
3	1.9	0.94	$e_{2,3}$	1	0.93
4	1.9	0.96	$e_{2,5}$	1	0.98
5	1.9	0.95	$e_{3,4}$	1.5	0.91
6	1	0.98	$e_{3,6}$	1.5	0.95
			$e_{4,5}$	2	0.97
			$e_{5,6}$	1	0.94



(a) 3节点对连通子网络1　(b) 3节点对连通子网络2

(c) 3节点对连通子网络3　(d) 3节点对连通子网络4

图 2　可靠扩展不确定网络

Fig. 2　Reliable extended uncertain network

定义 6　最可靠扩展不确定子网络。S_k^*为扩展不确定网络 G 中的一个可靠扩展不确定子网络, 如果对于扩展不确定网络 G 中任意一个可靠扩展不确定子网络, 都存在$R_{\mathrm{mn}(k),S_k^*} \geqslant R_{\mathrm{mn}(k),S_k}$, 那么称$S_k^*$为最可靠扩展不确定子网络。其中, $R_{\mathrm{mn}(k),S_k^*}$和$R_{\mathrm{mn}(k),S_k}$分别是S_k^*和S_k的多节点对时间可靠度.

定理 3　在扩展不确定网络 G 中, S_k^*为最可靠扩展不确定子网络, 则有

$$R_{\mathrm{mn}(k),G} = R_{\mathrm{mn}(k),S_k^*} \tag{15}$$

式中: $R_{\mathrm{mn}(k),G}$和$R_{\mathrm{mn}(k),S_k^*}$分别为扩展不确定网络 G 和最可靠扩展不确定子网络S_k^*的多节点对时间可靠度。

证明　令 $G(V,E,D)$为一个扩展不确定网络。在扩展不确定网络中, 很明显$R_{\mathrm{mn}(k),G} \geqslant R_{\mathrm{mn}(k),S_k^*}$, 因此, 本文只需证明$R_{\mathrm{mn}(k),G} \leqslant R_{\mathrm{mn}(k),S_k^*}$。

根据定理 1, 扩展不确定网络 G 中的多节点对时间可靠度为

$$R_{\mathrm{mn}(k),G} = \sup_{f_{\mathrm{mn}(k)}(B_1,B_2,\cdots,B_{N+m})<D_{\mathrm{th}}} \min_{1\leqslant i\leqslant N+M} \mathcal{M}\{d_i \in B_i\} \tag{16}$$

式中: B_i为集合$\{d_i^{\mathrm{c}},\infty\}$的子集, $i=1,2,\cdots,N+M$。

那么, 肯定存在一组$\{B_i^{\mathrm{c}}\}_{i=1}^{N+M}$, 取值$\{d_i^{\mathrm{c}}\}$或$\{d_i^{\mathrm{c}},\infty\}$, 使得

$$R_{\mathrm{mn}(k),G} = \min_{1\leqslant i\leqslant N+M} \mathcal{M}\{d_i \in B_i^{\mathrm{c}}\} = \min_{1\leqslant j\leqslant J} \mathcal{M}\{d_{i_j} = d_{i_j}^{\mathrm{c}}\} \tag{17}$$

式中: i_j和 J 分别为$\left\{i|B_i^{\mathrm{c}} = \left\{d_i^{\mathrm{c}}\right\}, 1 \leqslant i \leqslant N+M\right\}$的元素和元素规模。

那么由 J 个链路和节点组成的子网络 S 肯定满足 k 组节点对时间可靠的条件。如果子网络 S 中存在一条链路或节点不影响子网络 S 保持 k 组节点对时间可靠, 则将这条链路或节点去除, 得到一个新的子网络S_1, 包含 J-1 个链路和节点。此时S_1肯定满足 k 组节点对时间可靠的条件, 且满足:

$$R_{\mathrm{mn}(k),G} \geqslant R_{\mathrm{mn}(k),S_1} \geqslant \min_{1\leqslant j\leqslant J, j\neq J'} \mathcal{M}\{d_{i_j} = d_{i_j}^{\mathrm{c}}\} \geqslant \min_{1\leqslant j\leqslant J} \mathcal{M}\{d_{i_j} = d_{i_j}^{\mathrm{c}}\} = R_{\mathrm{mn}(k),G} \tag{18}$$

式中: $i_{j'}$为被移除的节点或链路的下标。

此时得到

$$R_{\mathrm{mn}(k),G} = R_{\mathrm{mn}(k),S_1} \tag{19}$$

重复上述删除链路或节点的过程, 直到子网络没有多余的链路和节点可以去除, 且满足有 k 节点对时间可靠的条件。此时的子网络为一个可靠扩展不确定子网络且满足

$$R_{\mathrm{mn}(k),G} = R_{\mathrm{mn}(k),S_k} \leqslant R_{\mathrm{mn}(k),S_k^*} \tag{20}$$

所以, 综上所述, 可知

$$R_{\mathrm{mn}(k),S_k^*} \leqslant R_{\mathrm{mn}(k),G} \leqslant R_{\mathrm{mn}(k),S_k^*} \tag{21}$$

证得$R_{\mathrm{mn}(k),G} = R_{\mathrm{mn}(k),S_k^*}$。　**证毕**

基于定理 3, 对扩展不确定网络 G 中多节点对时间可靠度的计算转化为对S_k^*中多节点对时间可靠度的计算。

定理 4　S_k^*为扩展不确定网络 G 的最可靠扩展不确定子网络。假设网络 G 中共有 K 组节点对, 且这些节点对对应的单节点对时间可靠度分别为$R_{\mathrm{sn},G,1}, R_{\mathrm{sn},G,2}, \cdots, R_{\mathrm{sn},G,K}$, 那么$S_k^*$的多节点对时间可靠

度为

$$R_{\mathrm{mn}(k),S_k^*}=k\text{-}\max\left[R_{\mathrm{sn},G,1},R_{\mathrm{sn},G,2},\cdots,R_{\mathrm{sn},G,K}\right] \tag{22}$$

式中：k-max[·]表示[·]中第k大的值。

证明 令S_k为一个可靠扩展不确定子网络。根据定理1和定义6，S_k中的多节点对时间可靠度为

$$R_{\mathrm{mn}(k),S_k}=\min_{e_i\in S_k}\mathcal{M}\{d_i=d_i^{\mathrm{c}}\} \tag{23}$$

令S_k中的k个最可靠路径分别为$p_1^*,p_2^*,\cdots,p_k^*$。那么存在

$$R_{\mathrm{mn}(k),S_k}=\min_{e_i\in S_k}\mathcal{M}\left\{d_i=d_i^{\mathrm{c}}\right\}=\min_{e_i\in\bigcup_{j=1}^{k}p_j^*}\mathcal{M}\left\{d_i=d_i^{\mathrm{c}}\right\}=\min_{1\leqslant j\leqslant k}\min_{e_h\in p_j^*}\mathcal{M}\left\{d_h=d_h^{\mathrm{c}}\right\}=\min_{1\leqslant j\leqslant k}R_{\mathrm{sn},S_k,j} \tag{24}$$

式中：$R_{\mathrm{sn},S_k,j}$表示子网络S_k中第j组节点对对应的单节点对时间可靠度。

已知在扩展不确定网络G中共有K组节点对，且这些节点对对应的单节点对时间可靠度分别为$R_{\mathrm{sn},G,1},R_{\mathrm{sn},G,2},\cdots,R_{\mathrm{sn},G,K}$。考虑到$S_k$是$G$的子网络，式(24)可以写为

$$R_{\mathrm{mn}(k),S_k}=\min_{1\leqslant j\leqslant k}R_{\mathrm{sn},S_k,j}=\min_{1\leqslant j\leqslant k}R_{\mathrm{sn},G,j} \tag{25}$$

式中：$R_{\mathrm{sn},G,j}$为第j组节点对对应的单节点对时间可靠度。

对于最可靠扩展不确定子网络S_k^*，根据定义6，其多节点对时间可靠度为扩展不确定网络G中第k大单节点对时间可靠度，即

$$R_{\mathrm{mn}(k),S_k^*}=k\text{-}\max\left[R_{\mathrm{sn},G,1},R_{\mathrm{sn},G,2},\cdots,R_{\mathrm{sn},G,K}\right] \tag{26}$$

证毕

基于定理3和定理4，本文给出多节点对时间可靠度算法，具体过程如下。

步骤1 计算出扩展不确定网络G中的节点对总数K。

步骤2 基于单节点对时间可靠度算法，分别计算出扩展不确定网络中K组节点对对应的单节点对时间可靠度。

步骤3 最后，将这些可靠度从大到小依次排序，第k个值为扩展不确定网络G的多节点对(要求节点对数为k)时间可靠度。

多节点对时间可靠度算法的复杂度主要为步骤2中求解扩展不确定网络中K组节点对对应的单节点对时间可靠度。考虑到$K=N(N-1)/2$，因此单节点对时间可靠度算法的复杂度为$\mathrm{O}\left(N^4\right)$，其中，$N$表示网络中的节点数。

3 案例分析

3.1 简单案例

本节基于图1中的网络，分别使用基于定理1的枚举算法和单节点对时间可靠度算法和多节点对时间可靠度算法计算网络的单节点对时间可靠度和多节点对时间可靠度。网络中各个节点和链路上无故障情况下的传输时间和对应的不确定测度如表1所示。假设网络各个节点之间的传输时间阈值均为10 ms。

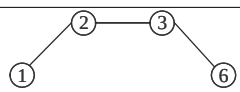
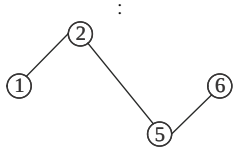
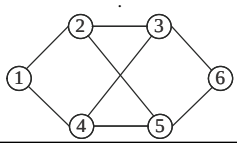
1）单节点对时间可靠度验证。本文以节点1和节点6为目标节点对，计算其对应的单节点对时间可靠度。

根据定理1，首先，需要枚举出图1中所有使得节点1和节点6传输时间满足时间阈值要求的子网络，共有95个这样的子网络。然后，分别计算这95个子网络的可靠度，具体信息如表2所示。最后，对表2中子网络可靠度取大，得到节点1和节点6的单节点对时间可靠度为0.94。

根据单节点对时间可靠度算法，同样也可计算得出节点1和节点6之间的单节点对时间可靠度

表2 满足节点1和节点6传输时间要求的子网络信息

Table 2 Subnetwork information meeting transmission time requirement of node 1 and node 6

编号	子网络	子网络可靠度
1		0.99∧0.96∧0.97∧0.93∧0.94∧0.95∧0.98=0.93
⋮	⋮	⋮
4		0.99∧0.96∧0.97∧0.98∧0.95∧0.94∧0.98=0.94
⋮	⋮	⋮
95		0.99∧0.96∧0.97∧0.93∧0.94∧0.95∧0.98∧0.94∧0.95∧0.91∧0.97∧0.96∧0.98∧0.92=0.91

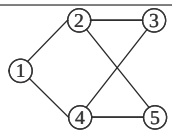
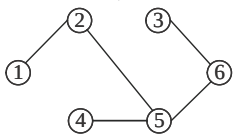
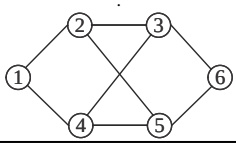
为 0.94, 最可靠路径为 1-2-5-6, 与枚举算法结果相同, 验证了单节点对时间可靠度算法的准确性。

2) 多节点对时间可靠度验证。图 1 的网络共有 15 个节点对, 本文以 k=10 为例, 计算其对应的多节点对时间可靠度。

同样根据定理 1, 首先, 需要找出满足 10 个节点对传输时间要求的子网络, 共有 79 个这样的子图。然后, 分别计算这些子网络的可靠度, 具体信息如表 3 所示。最后, 对这些子网络的可靠度取最大值, 得到 k=10 时多节点对时间可靠度。

表 3　满足 10 个节点对传输时间要求的子网络信息

Table 3　Subnetwork information meeting transmission time requirement of 10 node pairs

编号	子网络	子网络可靠度
1		0.99∧0.96∧0.97∧0.93∧0.94∧0.98∧0.92∧0.96∧0.97∧0.95∧0.91=0.91
⋮	⋮	⋮
4		0.99∧0.96∧0.97∧0.98∧0.95∧0.94∧0.97∧0.96∧0.94∧0.98∧0.95=0.94
⋮	⋮	⋮
79		0.99∧0.96∧0.97∧0.93∧0.94∧0.95∧0.98∧0.94∧0.95∧0.91∧0.97∧0.96∧0.98∧0.92=0.91

由表 3 可知, 编号 4 子网络的可靠度最大。因此, 求得 k=10 时的多节点对时间可靠度为 0.94。同样, 基于算法 2 计算得出 k=10 时的多节点对时间可靠度为 0.94, 与上述计算结果相同, 验证了多节点对时间可靠度算法的准确性。

3.2　教育网案例

对中国教育和科研计算机网(China education and resrarch network, CERNET)骨干网络进行时间可靠度评估。CERNET 是负责实现校园间信息共享和远程教育的计算机互联网络。在 20 个城市间分布有节点, 并通过 22 条链路连接在一起。CERNET 骨干网络的拓扑结构如图 3 所示[32]。

在本文案例中, 假设 CERNET 骨干网络中各个节点和链路都只有故障和正常 2 种状态, 节点和链路上的故障相互独立。基于 CERNET 骨干网络中有限的可靠性数据或专家经验, 根据不确定统计方法可以获取到各个节点和链路是否故障的不确定分布, 进而得到节点和链路正常的不确定测度, 具体方法可以参考文献 [21]。CERNET 骨干网络中各个节点和链路上的信息分别如表 4 和表 5 所示。

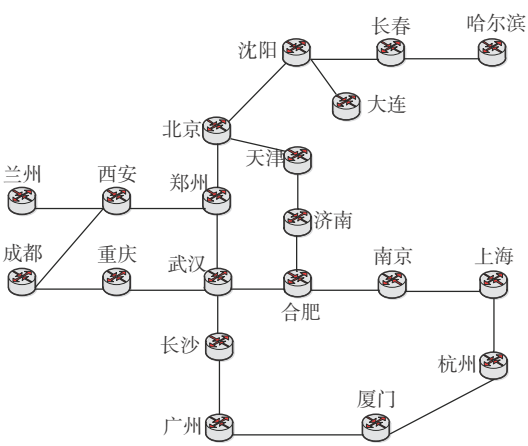


图 3　CERNET 骨干网络拓扑结构[32]

Fig. 3　Topology structure of CERNET backbone network [32]

表 4　节点信息

Table 4　Node information

节点	时延/ms	不确定测度	节点	时延/ms	不确定测度
哈尔滨	2	0.983	杭州	2	0.991
长春	2	0.988	厦门	2	0.989
沈阳	2	0.995	广州	2	0.986
大连	2	0.987	长沙	2	0.989
北京	1	0.991	武汉	2	0.980
天津	2	0.986	郑州	2	0.986
济南	2	0.987	重庆	2	0.987
合肥	2	0.983	成都	2	0.992
南京	2	0.988	西安	2	0.996
上海	2	0.994	兰州	2	0.989

1) 单节点对时间可靠性分析。根据 CERNET 骨干网络中各个节点和链路的信息, 可以建立扩展不确定网络。基于单节点对时间可靠度算法和扩展不确定网络, 通过寻找指定节点对的最可靠路径, 计算相应的单节点时间可靠度。表 6 为时延阈

表5 链路信息

Table 5 Link information

链路	时延/ms	不确定测度	链路	时延/ms	不确定测度
哈尔滨-长春	1	0.986	上海-杭州	0.5	0.981
长春-沈阳	1	0.983	杭州-厦门	3	0.985
沈阳-大连	1	0.978	厦门-广州	2	0.991
沈阳-北京	2	0.986	广州-长沙	2	0.984
北京-天津	0.5	0.987	长沙-武汉	1	0.989
北京-郑州	2	0.987	武汉-郑州	1.5	0.976
天津-济南	1	0.988	武汉-重庆	3	0.982
济南-合肥	1	0.986	郑州-西安	1.5	0.979
合肥-南京	0.5	0.983	重庆-成都	1	0.993
合肥-武汉	1.5	0.979	成都-西安	2.5	0.988

值为35 ms时不同节点对的时间可靠度和最可靠路径。

在时间可靠性评估的过程中，本文寻找的是满足时延阈值要求的最可靠路径。以表6中兰州与合肥之间的通信为例，其共有6条可能的传输路径，其中，满足时延要求的路径包括3条，分别为16.5 ms、22 ms和22 ms，在这些满足时延要求的路径中，最可靠路径为“兰州-西安-郑州-北京-天津-济南-合肥”，因此，这条路径的可靠度决定了兰州与合肥之间的时间可靠度。时延最小路径“兰州-西安-郑州-武汉-合肥”没有被选中的原因是郑州与武汉之间链路的可靠度较低，限制了“兰州-西安-郑州-武汉-合肥”这条路径的可靠度。

时延阈值取值约束了可能的传输路径，以哈尔滨和广州之间的单节点对时间可靠度为例，表7为不同时延阈值下的单节点对时间可靠度和最可靠路径。

由表7可知，不同时延阈值要求下的时间可靠度和最可靠路径不相同。当时延阈值为30 ms时，最可靠路径为“哈尔滨-长春-沈阳-北京-郑州-武汉-长沙-广州”，对应的可靠度为0.976。由于时延阈值较小，此时从哈尔滨到广州满足时延阈值要求的路径只有这一条，此时该路径为最可靠路径且可靠度为0.976。当时延阈值提升到35 ms，满足时延阈值要求的路径变成了2条，由于武汉-郑州这条链路的可靠度较低影响了其所属路径的可靠度，此时最可靠路径为“哈尔滨-长春-沈阳-北京-天津-济南-合肥-武汉-长沙-广州”，其可靠度为0.979。当时延阈值提升到40 ms时，最可靠路径变为了“哈尔滨-长春-沈阳-北京-天津-济南-合肥-南京-上海-杭州-厦门-广州”，此时的可靠度为0.981。

表6 不同节点对之间的单节点对时间可靠度

Table 6 Single-node-pair time reliability between different node pairs

节点对	可靠度	最可靠路径
哈尔滨-广州	0.979	哈尔滨-长春-沈阳-北京-天津-济南-合肥-武汉-长沙-广州
成都-上海	0.983	成都-重庆-武汉-合肥-南京-上海
兰州-合肥	0.981	兰州-西安-郑州-北京-天津-济南-合肥

表7 不同时延阈值下的单节点对时间可靠度

Table 7 Single-node-pair time reliability under different delay thresholds

时延阈值/ms	可靠度	最可靠路径
30	0.976	哈尔滨-长春-沈阳-北京-郑州-武汉-长沙-广州
35	0.979	哈尔滨-长春-沈阳-北京-天津-济南-合肥-武汉-长沙-广州
40	0.981	哈尔滨-长春-沈阳-北京-天津-济南-合肥-南京-上海-杭州-厦门-广州

2）多节点对时间可靠性分析。基于扩展不确定网络和多节点对时间可靠度算法，可以计算多节点对时间可靠度。网络中有20个节点，因此，该网络共有20×(20−1)/2=190组节点对。表8为在时延阈值为35 ms时，要求节点对数k取不同值时，网络的多节点对时间可靠度。

表8 多节点对时间可靠度

Table 8 Multi-node-pair time reliability

要求节点对数	可靠度	要求节点对数	可靠度
19	0.984	114	0.980
38	0.983	133	0.979
57	0.981	152	0.979
76	0.981	171	0.978
95	0.980	190	0.976

注：可靠度指要求达到传输时间要求的节点对数占网络总节点对数的比例。

多节点对时间可靠度随着要求节点对数的增加而减少。根据多节点对时间可靠度算法，多节点对时间可靠度等于网络中第k大的单节点对时间可靠度。因此，当k=10时，网络的多节点对连通可靠度为表中最大值0.984。当k=190时，系统的多节点对时间可靠度为表中最小值0.976，表示网络中所有节点对间的数据传输都及时可靠的信度，即网络的全端时间可靠度。

图4为时延阈值分别为30, 35, 40 ms时，网络的多节点对时间可靠度。由图4可知，当时延阈值的提升时，要求节点对数k为[87,99], [110,126], [156,162]和[185,190]范围内的多节点对时间可靠度有相应的提升，在k取其余值时，时间可靠度则没有变化。这是由于当时延阈值增加时，节点对之间满足时延阈值要求的路径会增加，当新增加路径的可靠度大于原有路径的可靠度时，该节点对的单节点对时间可靠度会增加，进而对应的多节点对可靠度也会提升。而当新增加路径的可靠度不大于

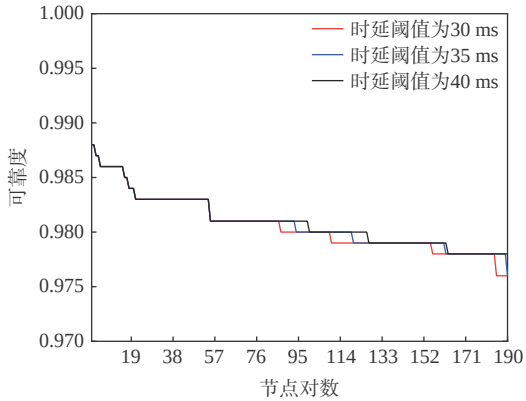


图 4　不同时延阈值下多节点对时间可靠度

Fig. 4　Multi-node-pair time reliability under different delay thresholds

原有路径的可靠度时, 此时对应的多节点对时间可靠度则不会改变。

3) 算法复杂度分析。为评估 2 种算法的计算复杂度, 表 9 以 CERNET 骨干网络为例分别给出单节点对时间可靠度算法和多节点对时间可靠度算法的运行时间。2 种算法均基于 MATLAB 2017b 实现, 且运行环境为 Windows 10 家庭中文版 Intel(R) Core(TM) i5-9300H CPU @ 2.40 GHz, 8 GB RAM。

表 9　2 种算法的运行时间

Table 9　Running time of two algorithms

算法	运行时间/ms
单节点对时间可靠度算法	2.689
多节点对时间可靠度算法	93.61

由表 9 可知, 多节点对时间可靠度算法的运行时间要远高于单节点对时间可靠度算法的运行时间, 与第 2 节中的算法复杂度分析结论一致。2 种算法的运行时间都较低, 能够用于复杂网络的时间可靠度评估。

4　结　论

1) 本文通过建立扩展不确定网络表征节点和链路的认知不确定性,可对存在认知不确定性的网络时间可靠性进行评估。

2) 本文提出单节点对和多节点对 2 种时间可靠性度量及其算法, 可满足不同应用场景需求。

未来将进一步研究节点和链路性能降级情况下的网络时间可靠性评估方法。

参考文献（References）

[1] GAUR V, YADAV O P, SONI G, et al. A literature review on network reliability analysis and its engineering applications[J]. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 2021, 235(2): 167-181.

[2] HONG S, LV C, ZHAO T D, et al. Cascading failure analysis and restoration strategy in an interdependent network[J]. Journal of Physics A: Mathematical and Theoretical, 2016, 49(19): 195101.

[3] HONG S, ZHU J X, BRAUNSTEIN L A, et al. Cascading failure and recovery of spatially interdependent networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2017, 2017(10): 103208.

[4] 黄宁, 伍志韬. 网络可靠性评估模型与算法综述[J]. 系统工程与电子技术, 2013, 35(12): 2651-2660.
HUANG N, WU Z T. Survey of network reliability evaluation models and algorithms[J]. Systems Engineering and Electronics, 2013, 35(12): 2651-2660(in Chinese).

[5] LIU Z, HUANG N, LI D P. An algorithm for delay-reliability in communication networks based on probabilistic user equilibrium model[C]// Proceedings of the International Conference on Information Science and Cloud Computing Companion. Piscataway: IEEE Press, 2013: 135-141.

[6] WU Z T, HUANG N, LI R Y, et al. A delay reliability estimation method for avionics full duplex switched ethernet based on stochastic network calculus[J]. Eksploatacja i Niezawodnosc-Maintenance and Reliability, 2015, 17(2): 288-296.

[7] WANG W X, GUO R J. Travel time reliability of highway network under multiple failure modes[J]. Sustainability, 2022, 14(12): 7256.

[8] ZHANG X X, ZHAO M, APPIAH J, et al. Prediction of travel time reliability on interstates using linear quantile mixed models[J]. Transportation Research Record: Journal of the Transportation Research Board, 2023, 2677(2): 774-791.

[9] JUNG Y, JO H, CHOO J, et al. Statistical model calibration and design optimization under aleatory and epistemic uncertainty[J]. Reliability Engineering & System Safety, 2022, 222: 108428.

[10] BAUDRIT C, DUBOIS D. Practical representations of incomplete probabilistic knowledge[J]. Computational Statistics & Data Analysis, 2006, 51(1): 86-108.

[11] SIMON C, WEBER P. Imprecise reliability by evidential networks[J]. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 2009, 223(2): 119-131.

[12] MUSHARRAF M, BRADBURY-SQUIRES D, KHAN F, et al. A virtual experimental technique for data collection for a Bayesian network approach to human reliability analysis[J]. Reliability Engineering & System Safety, 2014, 132: 1-8.

[13] FILIPPI G, VASILE M, KRPELIK D, et al. Space systems resilience optimisation under epistemic uncertainty[J]. Acta Astronautica, 2019, 165: 195-210.

[14] BEER M, FERSON S, KREINOVICH V. Imprecise probabilities in engineering analyses[J]. Mechanical Systems and Signal Processing, 2013, 37(1-2): 4-29.

[15] ZADEH L A. Fuzzy sets as a basis for a theory of possibility[J]. Fuzzy Sets and Systems, 1999, 100: 9-34.

[16] MOORE R E. Methods and applications of interval analysis[M]. Philadelphia: Society for Industrial and Applied Mathematics, 1979.

[17] MI J H, LU N, LI Y F, et al. An evidential network-based hierarchical method for system reliability analysis with common cause failures and mixed uncertainties[J]. Reliability Engineering & System Safety, 2022, 220: 108295.

[18] QI X J, CHENG Q. Imprecise reliability assessment of generating

systems involving interval probability[J]. IET Generation, Transmission & Distribution, 2017, 11(17): 4332-4337.

[19] ZHANG X F, LI R M, LIU M, et al. Evaluating travel time reliability based on fuzzy logic[J]. Applied Mechanics and Materials, 2011, 97-98: 952-955.

[20] KANG R, ZHANG Q Y, ZENG Z G, et al. Measuring reliability under epistemic uncertainty: review on non-probabilistic reliability metrics[J]. Chinese Journal of Aeronautics, 2016, 29(3): 571-579.

[21] LIU B D. Uncertainty theory[M]. 2nd ed. Berlin: Springer-Verlag, 2007: 205-234.

[22] ZHANG Q Y, KANG R, WEN M L. Belief reliability for uncertain random systems[J]. IEEE Transactions on Fuzzy Systems, 2018, 26(6): 3605-3614.

[23] 康锐. 确信可靠性理论与方法[M]. 北京: 国防工业出版社, 2020: 63-65.
KANG R. Belief reliability theory and methodology[M]. Beijing: National Defense Industry Press, 2020: 63-65 (in Chinese).

[24] LIU Y H. Uncertain random variables: a mixture of uncertainty and randomness[J]. Soft Computing, 2013, 17(4): 625-634.

[25] WANG S, YANG D, LU J. A connectivity reliability-cost approach for path selection in the maritime transportation of China's crude oil imports[J]. Maritime Policy & Management, 2018, 45(5): 567-584.

[26] HOSSEINI A, PISHVAEE M S. Capacity reliability under uncertainty in transportation networks: an optimization framework and stability assessment methodology[J]. Fuzzy Optimization and Decision Making, 2022, 21(3): 479-512.

[27] ZENG Z G, KANG R, WEN M L, et al. A model-based reliability metric considering aleatory and epistemic uncertainty[J]. IEEE Access, 2017, 5: 15505-15515.

[28] KARGER D R. A randomized fully polynomial time approximation scheme for the all-terminal network reliability problem[J]. IAM Review, 2001, 43(3): 499-522.

[29] YEH W C. An improved sum-of-disjoint-products technique for symbolic multi-state flow network reliability[J]. IEEE Transactions on Reliability, 2015, 64(4): 1185-1193.

[30] LIU W. Uncertain programming models for shortest path problem with uncertain arc lengths[C]//Proceedings of the First International Conference on Uncertainty Theory. Urumchi: Journal of the Operations Research Society of China, 2010: 148-153.

[31] 高原. 不确定图与不确定网络[D]. 北京: 清华大学, 2013: 21-26.
GAO Y. Uncertain graph and uncertain network[D]. Beijing: Tsinghua University, 2013: 21-26 (in Chinese).

[32] 李瑞莹, 李枚楠. 基于 Monte Carlo 和启发式算法的网络可靠性分配[J]. 北京理工大学学报, 2014, 34(7): 695-700.
LI R Y, LI M N. Network reliability allocation based on Monte Carlo and heuristic algorithm[J]. Transactions of Beijing Institute of Technology, 2014, 34(7): 695-700 (in Chinese).

Network time reliability evaluation method based on uncertainty theory

MA Ji[1], LI Ruiying[1, 2, *], ZHANG Qingyuan[3], KANG Rui[1, 2, 3]

(1. School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China;
2. Science & Technology on Reliability & Environmental Engineering Laboratory, Beijing 100191, China;
3. Hangzhou International Innovation Institute of Beihang University, Hangzhou 311115, China)

Abstract: The current network time reliability evaluation methods only consider the effect of inherent uncertainty but ignore the impact of epistemic uncertainty due to a lack of failure information on reliability evaluation results. To address this issue, a new method based on uncertainty theory was proposed. Based on the node range for network reliability, two metric parameters including single-node-pair time reliability and multi-node-pair time reliability were designed. An extended uncertain network model was proposed, which could directly model the epistemic uncertainty features on both nodes and links. Two algorithms were proposed to compute single-node-pair and multi-node-pair time reliability based on the most reliable path and the most reliable extended uncertain subnetwork. Finally, the proposed method was proposed to evaluate two time reliability metrics with a six-node network and the China education and research network (CERNET) backbone network as the example, and the results verify the correctness and effectiveness of the method.

Keywords: time reliability evaluation; epistemic uncertainty; extended uncertain network; uncertainty theory; assured reliability

Received: 2023-04-21; **Accepted**: 2023-07-20; **Published Online**: 2023-07-28 17:01

URL: link.cnki.net/urlid/11.2625.V.20230728.1358.001

Foundation items: National Natural Science Foundation of China (61773044,62073009); Fund for Science and Technology on Reliability and Environmental Engineering Laboratory (WDZC2019601A301)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0209

不同气体模型与滑移边界模型下升力体摩阻热流预测

严攀[1]，李沁[1,*]，黄潇[1]，郭晓明[2]，翁谊辉[1]，尤延铖[1]

(1. 厦门大学 航空航天学院，厦门 361000；2. 中国运载火箭技术研究院 空间物理重点实验室，北京 100076)

摘 要：针对高超声速升力体外形的摩阻、热流预测问题，采用所发展的三阶精度本质无波动格式，研究了不同气体模型(完全气体、平衡气体)、滑移边界模型(M-S、Le)和来流条件（高度、马赫数和壁面温度）对摩阻、热流等预测的影响。研究采用不同滑移边界模型和气体模型，对双马赫反射问题及典型高超声速绕流问题绕流进行了数值模拟与分析。结果表明：高保真气体模型、滑移边界模型和高精度计算格式能够使高超声速问题计算具有更好的精度。在此基础上，开展了不同高度、马赫数和壁温的高超声速升力体绕流数值模拟与分析，综合分析了滑移边界模型和气体模型对摩阻、热流等预测的影响。结果表明：不同气体模型之间的结果存在较大差异，平衡气体模型预示的边界层温度更低、边界层厚度更小、壁面热流更大、摩阻系数与总阻力系数稍大，二者差异随高度增加而增大；完全气体模型下，不同滑移边界模型总阻力系数、压心位置和热流分布均存在差异且随高度增加差异有所增大；采用平衡气体模型时，不同滑移模型之间预测结果相近。

关 键 词：高超声速；升力体；滑移边界条件；平衡气体；热流；摩阻
中图分类号：V211.3; V411
文献标志码：A **文章编号**：1001-5965(2025)04-1277-15

高超声速飞行器飞行速度快、突防能力强，是当今世界科技强国重点研究和全力发展的关键技术；高度集成的升力体构型是其基本特征之一。30～70 km空域是高超声速飞行器重要飞行走廊，其典型飞行环境特征是高马赫数、低密度/低雷诺数。强激波所导致的高温高压环境，引起严重的气动阻力、气动加热等，进而影响飞行器性能。因此，准确预测高超声速飞行器气动特性，对型号研制具有极其重要的意义。

朱广生院士在《临近空间高速飞行面临的机遇与挑战》报告(2020年)中，分析了临近空间(20～100 km)飞行器研发面临的五大挑战，其中，挑战4为“气动力、热高精度预示”。朱院士在《CFD在工程中的应用进展及存在问题》报告(2021年)中，进一步指出计算流体力学(computational fluid dynamic, CFD)在航天工程应用面临的五大问题，其中，第3个问题为“高空层流摩阻预示”。表明气动力、热预示存在显著误差是目前亟待解决的、需要集智攻关的重要问题。

相比于风洞试验，开展CFD数值模拟是高超声速流动预测经济且高效的手段。一方面，高超声速飞行环境存在稀薄气体效应、高温真实气体效应等复杂效应，气体物理建模的准确程度直接影响模拟精度；另一方面，数值格式对激波/激波与激波/边界层干扰及流动分离等的准确模拟存在直接影响。因此，发展和应用高保真物理模型并提高格式精度有助于解决高空层流摩阻、热流预示问题。

在气体物理模型方面，完全气体模型忽略了气体分子间相互作用，在高超声速飞行条件下缺乏高温、高压等对物态方程、输运系数等带来影响的考

收稿日期：2023-04-25；**录用日期**：2023-07-22；**网络出版时间**：2023-08-30 16:46
网络出版地址：link.cnki.net/urlid/11.2625.V.20230830.1542.003
***通信作者**. E-mail：qin-li@vip.tom.com

引用格式：严攀，李沁，黄潇，等. 不同气体模型与滑移边界模型下升力体摩阻热流预测[J]. 北京航空航天大学学报，2025，51(4)：1277-1291.
YAN P，LI Q，HUANG X，et al. Friction and heat flux prediction of lift body under different gas models and slip boundary models[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1277-1291 (in Chinese).

虑。作为真实气体模型的一种，平衡气体模型针对气体物理化学反应时间远小于流动特征时间的情况，所描述流动处于热化学平衡状态。Srinivasan等发展的分段函数曲线拟合方法[1-2]是常用的平衡气体模型实现方法之一，其物态方程和输运系数适用温度范围最高分别可达25 000 K和15 000 K，使用十分便捷、高效，在工程领域具有很好的应用前景。除上述气体模型外，在高空稀薄气体绕流中，空气的连续性假设可能部分失效，壁面附近存在速度滑移和温度跳跃等现象。对于滑移流范围流动，工程应用通常采用滑移边界模型对无滑移条件进行改进，其中，常见的有Maxwell速度滑移和Smoluchowski温度跳跃边界条件(后文简称M-S条件)[3-4]。

工程应用中，数值模拟求解高速边界层流动最为普遍的方法之一是二阶TVD有限体积方法。为进一步提高求解精度，基于结构网格的高精度格式如Jiang和Shu提出的WENO格式[5](WENO-JS)成为CFD研究的热点。而WENO-JS格式存在临界点降阶问题，针对该问题，Henrick等基于映射函数法发展了WENO-M类格式[6]，Borges等从改进光滑度量因子角度提出WENO-Z类格式[7]。在此基础上，李沁等提出新的区间临界点精度分析方法，提出具有真正一阶临界点精度恢复能力的、具有尺度无关特性的三阶格式WENO3-PRM^2-1,1[8]和WENO3-Z-ES3[9]，进一步提高了格式的实际计算精度和分辨率，因此，本文计算中采用WENO3-Z-ES3作为无黏通量重构格式。

由上述讨论可知，将工程实用的平衡气体模型与近年来Le等发展的滑移边界条件[10-13]相结合，采用高精度数值模拟方法开展典型高超声速问题摩阻、热流预测研究，具有显著的研究和应用价值。据此，本文针对高超声速升力体简化外形，通过开展不同高度(60 km、65 km、70 km)、不同马赫数(10、12.5、15)和不同壁温(500 K、1 000 K、1 500 K)的数值模拟与分析，综合分析滑移边界模型和平衡气体模型对气动特性预测的影响。

值得指出的是，在高空高超声速流动计算中，虽然采用高精度格式并考虑稀薄、真实气体等效应进行数值模拟具有较大的难度，但由于其对气动力、热等预测所具有的重要影响，开展相应工作是必要的。

1 考虑壁面吸附、物面曲率和黏性热效应的滑移边界模型

1.1 传统M-S滑移边界条件

Maxwell于1879年针对固定平板提出经典Maxwell滑移边界条件[3]：

$$\boldsymbol{u}_{\mathrm{s}}=\boldsymbol{u}_{\mathrm{w}}+\frac{(2-\sigma_{\boldsymbol{u}})}{\sigma_{\boldsymbol{u}}}\lambda\frac{\partial\boldsymbol{u}_{\tau}}{\partial\boldsymbol{n}} \tag{1}$$

式中：$\boldsymbol{u}_{\mathrm{s}}$为滑移速度；$\boldsymbol{u}_{\tau}$为切向速度；$\boldsymbol{u}_{\mathrm{w}}$为壁面速度；$\partial/\partial\boldsymbol{n}$为壁面法向的方向导数；$\sigma_{\boldsymbol{u}}$为动量适应系数；$\lambda$为平均分子自由程，其定义为

$$\lambda=\frac{\mu}{\rho}\sqrt{\frac{\pi}{2RT}} \tag{2}$$

式中：μ为动力黏性系数；R为气体常数；ρ为质量密度；T为温度。

Smoluchowski于1898年提出温度跳跃条件[4]：

$$T_{\mathrm{s}}-\frac{(2-\sigma_T)}{\sigma_T}\frac{2\gamma}{Pr(\gamma+1)}\lambda\frac{\partial T}{\partial n}=T_{\mathrm{w}} \tag{3}$$

式中：T_{s}为跳跃温度；T_{w}为壁面温度；σ_T为温度适应系数；γ为比热容比；Pr为普朗特数。

传统M-S滑移边界条件是最常用的滑移条件，由于其程序实现相对方便，已广泛应用于工程计算。本文计算中，动量适应系数$\sigma_{\boldsymbol{u}}$与温度适应系数σ_T均取1.0。

1.2 Le滑移边界条件

Maxwell滑移边界条件中动量适应系数$\sigma_{\boldsymbol{u}}$反映了壁面反射分子中发生漫反射与镜面反射的比例。在实际问题中，难以确定系数$\sigma_{\boldsymbol{u}}$的取值，因此，带来一定不确定性。为此，Le等将Langmuir壁面吸附理论与M-S滑移边界条件结合，在动量守恒方程中考虑壁面分子吸附的影响，即吸附的分子对总剪切应力不产生贡献，并且考虑曲面效应和热蠕变，发展了Langmuir-Maxwell速度滑移条件及Langmuir-Smoluchowski温度跳跃条件[10]；Le等又在Smoluchowski的温度跳跃条件的基础上，考虑滑移产生的摩擦热(黏性热)的贡献，发展了改进的Smoluchowski温度跳跃条件[11]；在此基础上，Le等分别从Kaniadarkis模型[14]和Patterson切向动量平衡方程推导了Langmuir-Kaniadarkis速度滑移条件[12]和改进的Patterson温度跳跃条件[13]。为方便起见，本文将Langmuir-Kaniadarkis速度滑移条件与改进的第4种Patterson温度跳跃条件合称为Le滑移边界条件。

1.2.1 Langmuir-Kaniadarkis速度滑移条件

Langmuir-Kaniadarkis速度滑移条件为

$$\boldsymbol{u}+\left(\frac{1}{1+\theta}\right)\lambda\nabla_{\boldsymbol{n}}(\boldsymbol{Su})=\boldsymbol{u}_{\mathrm{w}}-\left(\frac{1}{1+\theta}\right)\frac{\lambda}{\mu}\boldsymbol{S}(\boldsymbol{n}\boldsymbol{\Pi}_{\mathrm{mc}})-\frac{3}{4}\frac{\mu}{\rho}\frac{\boldsymbol{S}\nabla T}{T} \tag{4}$$

式中：等号右边第2项和第3项分别为考虑曲面效应和热蠕变的新项；张量$\boldsymbol{S}=\boldsymbol{I}-\boldsymbol{nn}$；$\boldsymbol{\Pi}_{\mathrm{mc}}=$

$\mu\left((\nabla \boldsymbol{u})^{\mathrm{T}}-\frac{2}{3}\mathrm{tr}(\nabla \boldsymbol{u})\boldsymbol{I}\right)$; θ为表面覆盖率, 由 Langmuir 等温吸附模型计算如下。

对于单原子分子:

$$\theta=\frac{\beta p}{1+\beta p} \tag{5}$$

对于双原子分子:

$$\theta=\frac{\sqrt{\beta p}}{1+\sqrt{\beta p}} \tag{6}$$

式中: $\beta=\frac{K_L}{RT_{\mathrm{w}}}$为与表面温度有关的平衡常数, K_L为化学反应的化学平衡常数:

$$K_L=A_{\mathrm{m}}\lambda\exp\left(\frac{D_{\mathrm{e}}}{R_u T_{\mathrm{w}}}\right) \tag{7}$$

式中: A_{m}为通过$N_A\pi d^2/4(\mathrm{m}^2/\mathrm{mol})$测量或计算; d为气体分子的共价直径, 氮气和氩气共价直径分别为 0.142 nm 和 0.21 nm; N_A为阿伏伽德罗常数, 通用气体常数$R=8.314(\mathrm{J/(mol\times K)})$; D_{e}为吸附热的测量值, 对于氩气和氮气$D_{\mathrm{e}}=5\,525\ \mathrm{J/mol}$。

1.2.2　Patterson 温度跳跃条件

由于物体表面存在速度滑移, 因此, 会产生黏性热并对跳跃温度产生影响。Le 等[11] 进一步考虑曲率的影响提出含有黏性热的热流计算式:

$$q=-k\nabla_n T-\boldsymbol{u}_{\mathrm{s}}\boldsymbol{S}(\boldsymbol{n}\boldsymbol{\Pi}) \tag{8}$$

式中: k 为热传导系数。

在此基础上, Le 等[13] 提出的第 4 种 Patterson 温度跳跃条件为

$$T+\frac{1}{2}\cdot\frac{\lambda}{Pr}\left(\frac{2-\sigma_T}{\sigma_T}\right)\frac{T_{\mathrm{w}}}{T}\cdot\frac{\gamma}{(\gamma-1)}\nabla_n T=T_{\mathrm{w}}-\frac{1}{2}\left(\frac{2-\sigma_T}{\sigma_T}\right)\frac{1}{(\gamma-1)C_v}\cdot\frac{T_{\mathrm{w}}}{T}\cdot\frac{\lambda}{\mu}\boldsymbol{S}(\boldsymbol{n}\boldsymbol{\Pi})\boldsymbol{u} \tag{9}$$

式中: C_v为定容比热容, $\boldsymbol{\Pi}=\mu\nabla\boldsymbol{u}+\boldsymbol{\Pi}_{\mathrm{mc}}$。

1.3　算例验证平板绕流

为验证 Le 滑移边界条件的实现并与 M-S 滑移边界条件进行比较, 本文选取 Becker[15] 的 Ma=12.7 平板算例进行验证计算。该算例来流为氩气, 来流参数$T_\infty=64.5$ K, $P_\infty=3.73$ Pa, 平均分子自由程$\lambda_\infty=0.23$ mm, 壁温$T_{\mathrm{w}}=292$ K。计算采用 WENO3-Z-ES3 格式。

图 1 和图 2 分别为采用 M-S 滑移条件和 Le 滑移条件结果和直接模拟蒙特卡罗(direct simulation of Monte Carlo, DSMC)平板结果对比, 其中, DSMC 结果来自文献 [10]。由图可知, Le 滑移边界条件在温度、滑移速度分布上与 DSMC 结果符合较好, 相对于 M-S 滑移边界条件取得显著改进。

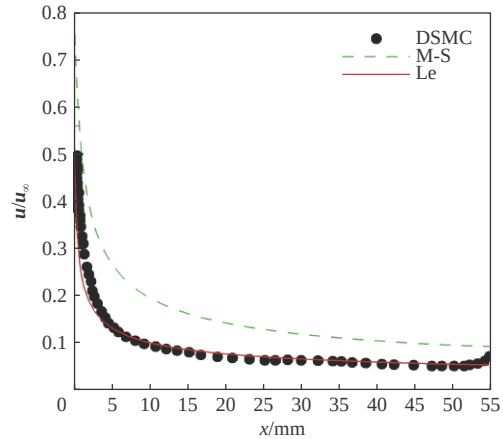


图 1　Ma=12.7 平板绕流壁面滑移速度分布

Fig. 1　Slip velocity distribution of Ma=12.7 flow around a flat plate

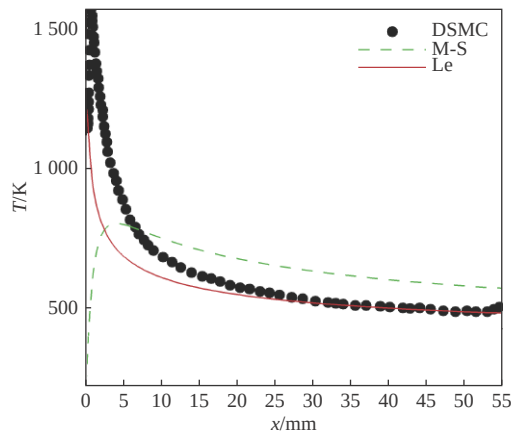


图 2　Ma=12.7 平板绕流壁面近壁面气体温度分布

Fig. 2　Wall gas temperature distribution of Ma=12.7 flow around a flat plate

2　基于曲线拟合的平衡气体模型

平衡气体流动的处理的方法有 2 类: ①基于平衡组分方程和流动控制方程(如平衡常数法、最小自由能法); ②基于实验数据, 采用曲线拟合方法。曲线拟合法公式通常根据高温空气热力学特性表的数据拟合得到, 具有编程相对简单及计算效率高的优点, 广泛用于平衡流动计算。本文采用 Srinivasan 等于 1987 年改进的 Tannehill 曲线拟合法[1-2]。

在 Tannehill 曲线拟合法中, 每个热力学量或运输系数的曲线拟合都是由不同区域中的一系列的分段函数组成。分段函数采用 Grabau 型过渡函数, 其一般形式如下:

$$f(x,y)=f_1(x,y)+\frac{f_2(x,y)-f_1(x,y)}{1\pm\exp(k_0+k_1x+k_2y+k_3xy)} \tag{10}$$

式中: "+" 表示奇函数过渡; "–" 表示偶函数过渡; k_i为常系数; $f_1(x,y)$和$f_2(x,y)$为三次多项式, 表示为

$$f_1(x,y)=p_1+p_2x+p_3y+p_4xy+p_5x^2+p_6y^2+p_7x^2y+p_8xy^2+p_9x^3+p_{10}y^3 \quad (11)$$

$$f_2(x,y)-f_1(x,y)=p_{11}+p_{12}x+p_{13}y+p_{14}xy+p_{15}x^2+p_{16}y^2+p_{17}x^2y+p_{18}xy^2+p_{19}x^3+p_{20}y^3 \quad (12)$$

式中：p_i为常系数。奇、偶函数取法由参数表给出。

2.1 热力学参数的拟合

温度T与$\tilde{\gamma}$的拟合可以采用密度与内能进行拟合（如$T=T(e,\rho)$），也可以采用密度与压力进行拟合（如$T=T(p,\rho)$）。通过拟合计算得到$\tilde{\gamma}$，再采用$a=\sqrt{\tilde{\gamma}RT}$近似计算声速。

热力学参数的拟合公式中，需要将有量纲量用参考量进行无量纲化，以温度关于密度和内能的拟合$T=T(\rho,e)$为例，其拟合式为

$$\lg\frac{T}{T_0}=f(Y,Z) \quad (13)$$

式中：$Y=\lg(\rho/\rho_0)$；$X=\lg(p/p_0)$；$Z=X-Y$；$T_0=151.78$ K；$\rho_0=1.292$ kg/m^3；$p_0=1.0134\times10^5$ N/m^2。

2.2 输运系数的拟合

黏性系数μ可以用密度和温度拟合，也可以采用密度和内能拟合。热传导系数的计算有2种方式：①转换为Pr并结合μ进行计算，其中，Pr采用密度和温度拟合；②采用密度和内能直接拟合k。

输运系数的拟合以黏性系数μ为例，其采用基于密度和温度的拟合关系：$\mu=\mu(\rho,T)$，具体拟合式为

$$\frac{\mu}{\mu_0}=f(X,Y) \quad (14)$$

式中：$X=T/1\,000$；$Y=\lg(\rho/\rho_0)$；$\rho_0=1.243$ kg/m^3；$\mu_0=1.749\times10^{-5}$ (kg·s)/m。

2.3 算例验证圆锥绕流

本文采用Ma=25.3高超声速零迎角圆锥绕流计算验证平衡气体模型实现并与完全气体结果进行对比。圆锥半锥角为10°，其他计算条件如表1所示，计算网格数为200×100。

图3为$x=1$ m处的流场速度和温度分布，图4为壁面摩阻系数和热流系数沿流向分布，对比数据为Tannehill等的三维迎风PNS代码结果[16]（3D_UPS）。可以看到本文平衡气体模型计算得到的参数分布与Tannehill等给出的结果相符，而完全气体结果则显示出与前两者的明显差异。从速度和温度分布上可以看到，2种模型结果具有不同的边界层厚度和激波位置，完全气体模型得到的边界层内

表1 高超声速圆锥绕流计算条件

Table 1 Calculation conditions of hypersonic flow around a cone

气体	Ma	T_∞/K	P_∞/Pa	Re_∞
空气	25.3	252.6	20.35	1.29×10^5

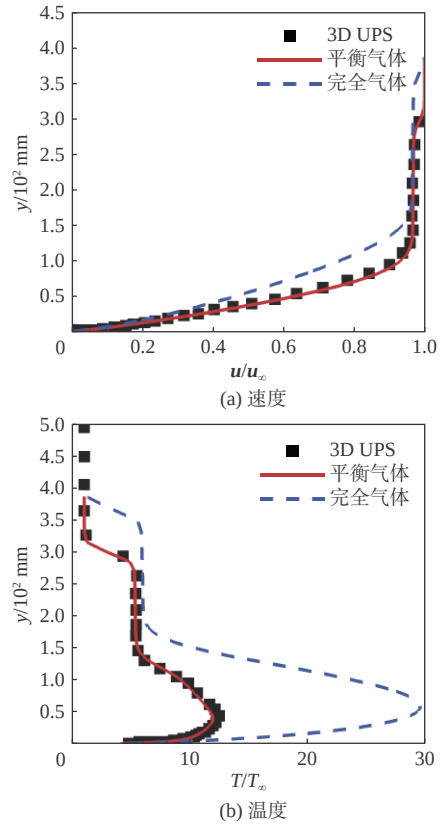


图3 Ma=25.3圆锥绕流x=1 m处截面参数分布

Fig. 3 Cross-section parameter distribution at x = 1 m of Ma=25.3 hypersonic flow around a cone

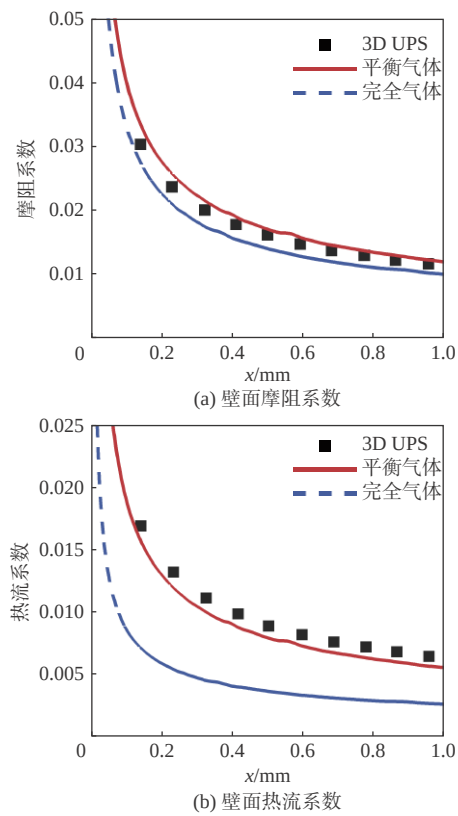


图4 Ma=25.3圆锥绕流壁面参数分布

Fig. 4 Wall parameter distribution of Ma=25.3 hypersonic flow around a cone

温度峰值接近平衡气体的 3 倍。在壁面摩阻系数和热流系数分布上, 完全气体壁面摩阻系数略小于平衡气体, 而壁面热流系数低于平衡气体达 50%左右。

3　三阶精度尺度无关本质无波动格式

李沁等在 WENO3-JS 格式的基础上, 首先基于映射函数法发展了 WENO3-PRM^2-1,1格式[8], 其具有良好的数学形式和性质; 进一步基于 WENO-Z 方法发展了 WENO3-Z-ES3 格式[9], 其具有较好的计算效率。在第 5 节升力体计算中, 2 种格式都给出类似结果, 限于篇幅, 本文仅给出 WENO3-Z-ES3 格式计算结果。

3.1　WENO3-PRM^2-1,1 格式

基于映射函数方法, 李沁等[8] 发展了如下 $P_{n,m}^{n+1}$ 系列映射函数:

$$\begin{cases} P_{n,m;m_1;c_1,c_2}^{L,n+1} = d_k + \dfrac{(\omega-d_k)^{n+1}}{(\omega-d_k)^n + (-1)^{1+n}c_2^L(\omega-d_k)\omega^{m_1} + (-1)^n c_1^L \omega^{m+1}} \\ P_{n,m;m_1;c_1,c_2}^{R,n+1} = d_k + \dfrac{(\omega-d_k)^{n+1}}{(\omega-d_k)^n + c_2^R(\omega-d_k)(1-\omega)^{m_1} + c_1^R(1-\omega)^{m+1}} \end{cases} \tag{15}$$

式中: ω为非线性权; d_k为线性权, 其余各参数取值见文献 [8]。得益于其非线性形式和自由参数, PRM 映射函数解决了其他同类方法无法克服的困难——如何协调函数的平坦性和端点与恒等映射趋近程度的问题, 使得两者可以同时得到兼顾。在文献 [8] 中给出了 PRM 函数满足良好可调节性质($C_{n,m}$特性)的定理及证明。通过选取合适的自由参数, PRM 可以同时具备较高的计算分辨率和较好的稳定性。将 WENO3-JS 格式与三阶 PRM^2-1,1映射函数结合, 得到 WENO3-PRM^2-1,1格式。

3.2　WENO3-Z$_{ES3}$ 格式

基于 WENO-Z 方法, 针对三阶 WENO 格式, 李沁等[8] 进行了格式精度分析并提出新的区间临界点精度分析方法, 其假设临界点出现在格式区间内(现有文献通常假设临界点位于x_j)。根据区间临界点分析方法, 李沁等[8] 证明了格式在传统三点模板$\{x_{j-1}, x_j, x_{j+1}\}$下无法满足一阶临界点恢复要求, 并结合格式恢复精度的充分条件提出了新的全局光滑度量因子τ_4。为进一步提高格式鲁棒性, 李沁等[8] 经过理论分析与数值实验将光滑度量因子β_k也进行了模板拓展, 最终得到 WENO3-Z-ES3 格式如下:

$$\alpha_k = d_k\left(1 + c_\alpha\left(\frac{\tau_4}{\beta_k^{(r)*} + \varepsilon}\right)^2\right) \tag{16}$$

式中: $c_\alpha = 0.4$, $\varepsilon = 1\times10^{-40}$, 全局光滑度量因子 $\tau_4 = \left|\left(2f_{j+1} - 3f_j + f_{j-1}\right)\left(f_{j+2} - 3f_{j+1} + 3f_j - f_{j-1}\right)\right|$, 光滑度量因子$\beta_0^{(2)*} = \beta_0^{(2)} + c_{\beta_0}\left(f_j - 2f_{j-1} + f_{j-2}\right)^2$, $\beta_1^{(2)*} = \beta_1^{(2)} + c_{\beta_1}\left(f_{j+2} - 2f_{j+1} + f_j\right)^2$。文献 [11] 通过若干典型算例表明, WENO3-Z-ES3 格式既具有较高精度又具有较好鲁棒性。

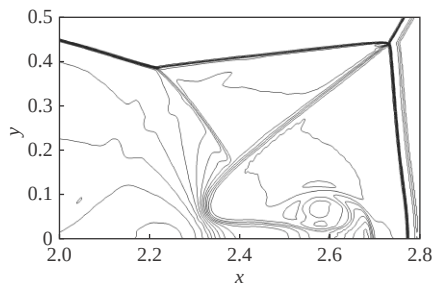
3.3　算例验证

上述 2 种三阶格式在理论上具有较好的计算分辨率和鲁棒性, 为验证其计算表现, 下面给出相关算例计算结果。

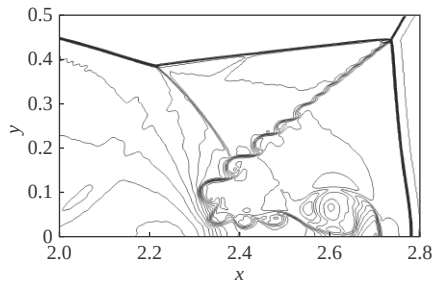
3.3.1　双马赫反射问题

该问题描述了在二维无黏流动中, Ma=10 的平面激波以 60°的入射角撞击壁面并继续传播, 计算域为 [0,3]×[0,1], 网格数为 1 920×480。

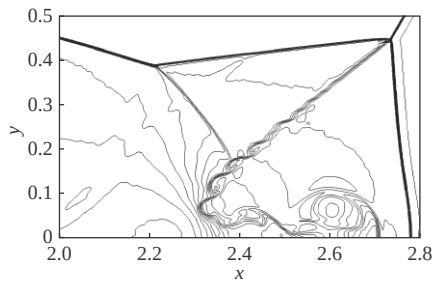
图 5 为双马赫反射问题 WENO3-JS、-PRM^2-1,1 和-Z-ES3 格式密度分布对比, 可以看到 WENO3-PRM^2-1,1、-Z-ES3 格式能分辨出沿滑移线的小尺



(a) WENO3-JS



(b) WENO3-PRM²-1,1



(c)WENO3-Z-ES3

图 5　双马赫反射问题 WENO3-JS、-PRM^2-1,1 和-Z-ES3 格式密度等值线结果比较

Fig. 5　Comparison of density contour results in WENO3-JS, -PRM^2-1,1 and -Z-ES3 schemes for double Mach reflection problems

度卷起涡，分辨率相比 WENO3-JS 格式有明显提高，表明格式具有良好的分辨率、激波捕捉能力和鲁棒性。

3.3.2 *Ma*=11.35 绕中空圆柱-长裙流动

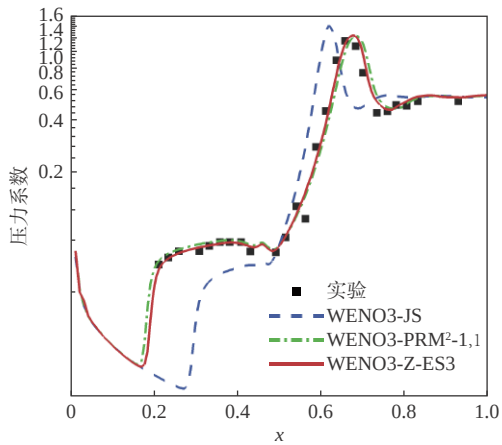
绕中空圆柱-长裙（hollow cylinder extended flare, HCEF）流动是高超声速空气动力学中的一个典型算例，计算条件如表 2 所示，实验数据来自文献 [17]，计算网格数为 256×128。

表 2 *Ma*=11.35 绕中空圆柱-长裙流动计算条件[17]

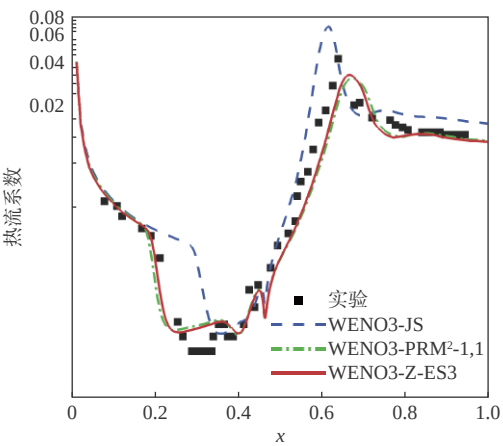
Table 2 Calculation condition of *Ma*=11.35 flow around a hollow cylinder-flare[17]

气体	*Ma*	*Re*/m	T_∞/K	T_w/K
N2	11.35	3.596×10^5	79	294

WENO3-JS、-PRM^2-1,1 和-Z-ES3 格式计算给出的壁面压力系数和热流系数分布如图 6 所示，图 6(a) 显示 WENO3-PRM^2-1,1 格式和 WENO3-Z-ES3 格式压力系数结果和实验符合良好，给出准确的压力平台和峰值预示，而 WENO3-JS 格式除分离区位置与实验偏离外，分离区压力平台预示也出现较大误差。图 6(b) 显示，WENO3-PRM^2-1,1 格式和 WENO3-Z-ES3 格式热流系数结果相似，均与实验符合良好，明显优于 WENO3-JS 格式；WENO3-JS 格式得到的分离区过小，分离区后峰值超出实验值。



(a) 压力系数



(b) 热流系数

图 6 绕 HCEF 流动 WENO3-JS、-PRM^2-1,1 和-Z-ES3 格式壁面参数分布对比

Fig. 6 Wall parameter distribution of flow around HCEF in WENO3-JS, -PRM^2-1,1, and -Z-ES3 schemes

此外，图 6(b) 还显示由于分离区流动结构复杂，所有格式在对热流峰值的预测上与实验结果存在一定偏差。考虑到此处压力峰值近似对应流动再附，从而也应与热流峰值近似对应，而图中实验结果则表明其峰值压力与峰值热流位置并不一致。因此，该处峰值热流的计算、实验的进一步比较分析有待更精细实验的开展。

4 不同气体与滑移模型下升力体摩阻/热流预测

升力体外形是高超声速飞行器常见气动外形之一，本文针对如图 7 所示的升力体外形，网格数为 175×61×181（流向×法向×周向），法向第 1 层网格大小 $\Delta x/L$=0.000 01（L=3.2 m 为升力体轴向总长），端头驻点曲率半径 r/L=0.002 17，升力体迎风面后段存在相对子午面迎风母线的、大小为 4.25°的平面切角，零迎角下该平面与水平面平行；后缘升降舵相对切角平面舵偏为 10°。图中，侧缘线用粗实线条标出。值得指出的是，后缘舵偏角及平面切角大小对流动结构分析、流动参数分布及气动力系数结果有直接影响，由于本文主要关注对比不同滑移边界模型和气体模型结果，限于篇幅暂时仅对 10°舵偏角和 4.25°平面切角情况进行数值计算与分析，考虑不同后缘舵偏角的模拟分析拟在后续研究中进一步开展。计算工况均为 0°迎角；在使用 WENO3-Z-ES3 格式的基础上，分别采用 M-S 和 Le 滑移模型、完全气体和平衡气体模型，进行了不同高度、马赫数、壁温下的数值模拟分析，综合研究不同气体与滑移模型在不同条件对升力体摩阻/热流预示的影响。

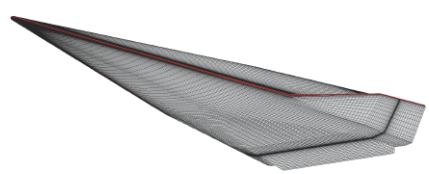


图 7 升力体外形示意图

Fig. 7 Outline of lift body

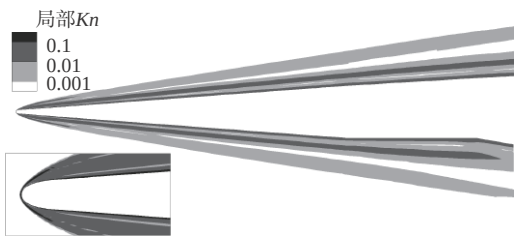
4.1 *Ma*=12.5 不同高度的数值模拟与分析

为研究不同飞行高度条件下，真实气体效应及稀薄气体效应对高超声速升力体绕流摩阻/热流预示的影响，分别进行 60 km、65 km 和 70 km 高度升力体流动的数值模拟，其他计算条件分别为：飞行马赫数为 12.5、迎角为 0°、升力体壁温为 1 000 K。

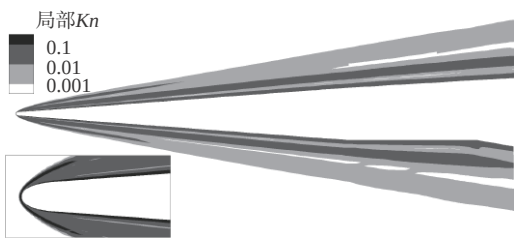
图 8 为升力体流场对称面局部 Kn 数 Kn_{GLL} ($(\lambda/\rho)|\nabla\rho|$)在不同高度下的分布以显示来流稀薄程度, 计算采用平衡气体模型和 Le 滑移边界模型。由图可见, 60 km 时 Kn_{GLL} 在激波附近和边界层内处于滑移流区(0.01～0.1)的范围, 65 km 和 70 km 时头部和边界层附近局部 Kn 几乎都处于滑移流范围内, 并且存在部分区域 Kn_{GLL} 大于 0.1, 这意味着流场局部区域稀薄效应显著, 因此, 考虑稀薄气体效应进行数值模拟是必要的。



(a) 高度为60 km



(b) 高度为65 km



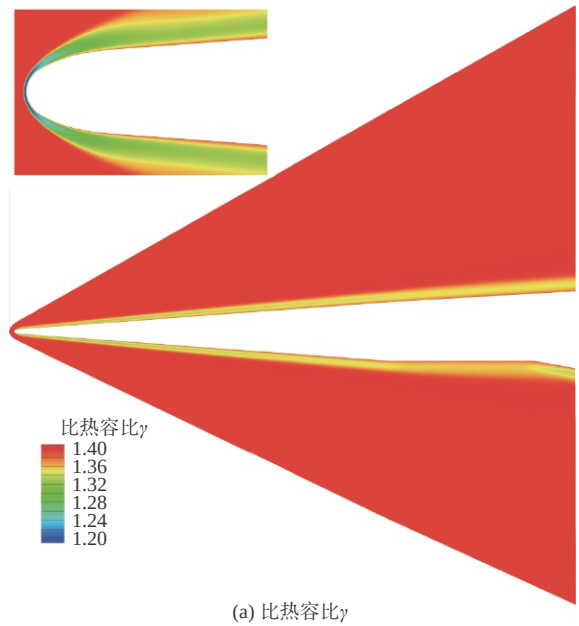
(c) 高度为70 km

图 8　Ma=12.5、壁温 1 000 K 不同高度升力体对称面局部 Kn 分布(平衡气体和 Le 滑移模型)

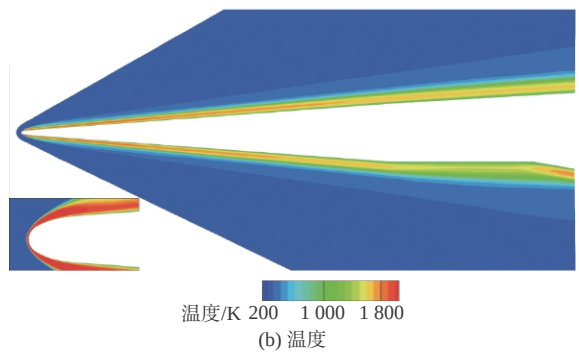
Fig. 8　Local Kn distribution on symmetry plane of lift body at different heights under Ma=12.5 and wall temperature 1 000 K (equilibrium gas and Le slip model)

图 9 为 60 km 高度下采用平衡气体模型和 Le 滑移边界模型的升力体流场对称面比热比 γ 和温度分布云图(65 km 与 70 km 下的结果与此类似故未给出)。相较于完全气体的比热容比 γ 固定为 1.4, 采用平衡气体模型进行数值模拟时, 气体的比热比 γ 将会发生变化, 主要集中在升力体壁面附近, 尤其是在头部激波处, 流场最小比热容比约为 1.2。同时对比同高度下温度分布云图可见, 上述区域温度达到 1 800 K 以上, 此时若采用完全气体假设将造成较大的误差, 需要考虑真实气体效应对流动带来的影响。

图 10 为 60 km 高度下采用平衡气体模型和 Le 滑移边界模型的升力体表面压力等值线分布, 结果显示从升力体头部区域沿着两侧侧缘为高压区。图 11 为 60 km 高度下采用 Le 滑移边界模型的



(a) 比热容比γ



(b) 温度

图 9　Ma=12.5、壁温 1 000 K、60 km 高度下升力体对称面比热容比 γ/温度分布云图(平衡气体和 Le 滑移模型)

Fig. 9　Specific heat ratio γ/temperature distribution on symmetry plane of lift body at 60 km height, Ma=12.5, and wall temperature 1 000 K (equilibrium gas and Le slip model)

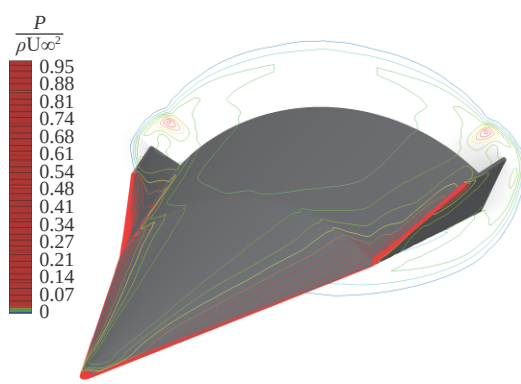


图 10　Ma=12.5、壁温 1 000 K、60 km 高度下升力体表面及底部截面压力等值线分布(平衡气体和 Le 滑移模型)

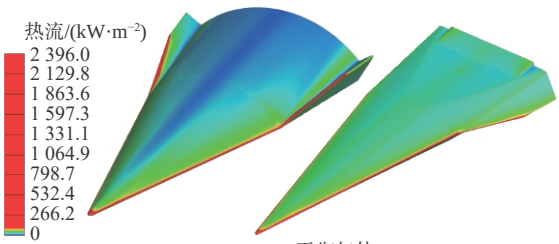
Fig. 10　Pressure contour distribution on surface and bottom sections of lift body at 60 km height, Ma=12.5, and wall temperature 1 000 K (equilibrium gas and Le slip model)

升力体表面热流分布，可以看到高热流集中区域与高压区分布较为一致。对比平衡气体与完全气体模型得到的表面热流分布，可以发现前者表面热流显著高于后者，两者的高热流区域分布则比较一致。

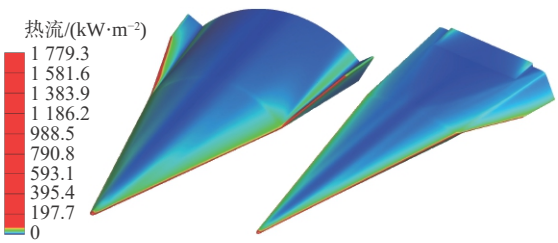
图 12 为采用平衡气体模型和 Le 滑移模型计算得到的升力体表面摩阻系数分布云图，高摩阻系数区域主要分布在升力体头部、侧缘及机翼前缘，此外升力体上下表面的摩阻系数较小；摩阻分布与热流分布特性定性一致。

图 11 和图 12 均显示出升力体上表面机翼内侧有热流、摩阻较大的区域，这与边界层在这些区域中存在较大速度/温度梯度有关，其由边界层变薄和/或外沿参数变大引起。由于升力体上表面激波会与机翼产生激波发生碰撞，进而产生激波/激波干扰现象。

图 13 为不同高度下驻点线上的参数（无量纲压力和无量纲温度，采用来流参数进行无量纲化，参考量分别为：$\rho_\infty U_\infty^2$、T_∞）分布曲线对比。采用平



(a) 平衡气体



(b) 完全气体

图 11 *Ma*=12.5、壁温 1 000 K、60 km 高度下升力体表面热流分布（Le 滑移模型）

Fig. 11 Heat flux distribution on surface of lift body at 60 km height, *Ma*=12.5, and wall temperature 1 000 K (Le slip model)

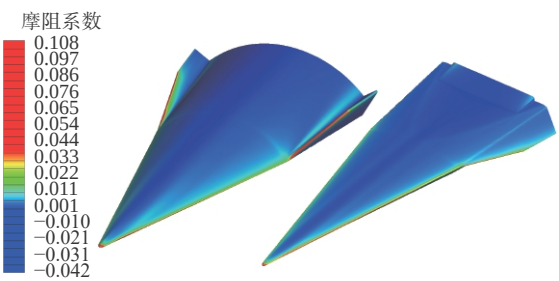
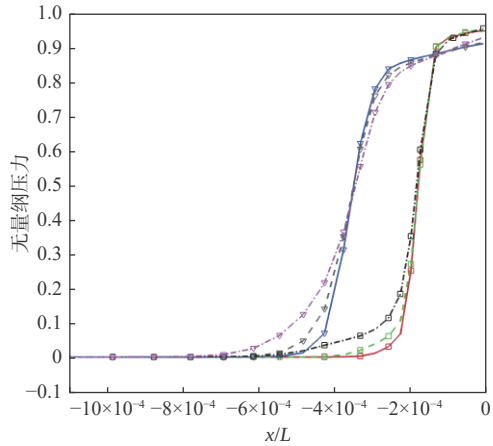
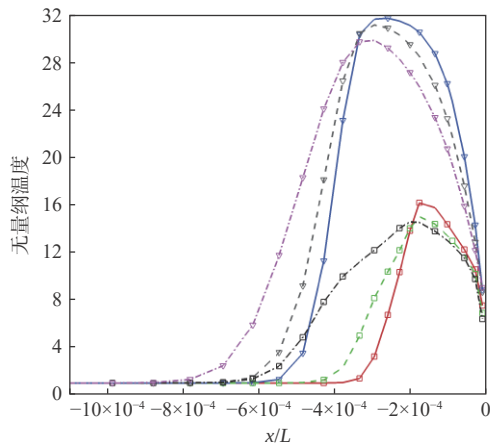


图 12 *Ma*=12.5、壁温 1 000 K、60 km 高度下升力体表面摩阻分布（平衡气体和 Le 滑移模型）

Fig. 12 Friction distribution on surface of lift body at 60 km height, *Ma*=12.5, and wall temperature 1 000 K (equilibrium gas and Le slip model)



(a) 无量纲压力



(b) 无量纲温度

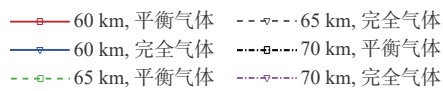


图 13 *Ma*=12.5、壁温 1 000 K 不同高度驻点线参数分布（Le 滑移模型）

Fig. 13 Parameter distribution of stagnation point line at different heights under *Ma*=12.5 and wall temperature 1 000 K (Le slip model)

衡气体模型计算得到的激波脱体距离显著小于完全气体。同一高度下驻点线上平衡气体流动压力与完全气体差别不大，而驻点线上的温度分布二者存在差异，完全气体温度高于平衡气体，前者温度峰值超过后者 1 倍左右。

图 14 为不同高度下平衡气体模型和完全气体模型升力体腹部表面极限流线和壁面压力云图（采用 Le 滑移模型）。完全气体模型结果在不同高度下均出现了明显的分离区，且随着高度增加，分离区逐渐减小；平衡气体模型在 60 km 高度下也出现了流动分离现象，而 65 km 和 70 km 下则几乎不存在分离结构。因此，平衡气体计算给出的分离尺度小于完全气体计算得到结果。

由于图 14(a) 中完全气体和平衡气体的表面分离尺度及结构不同，为此在图 15 中给出 60 km 高度

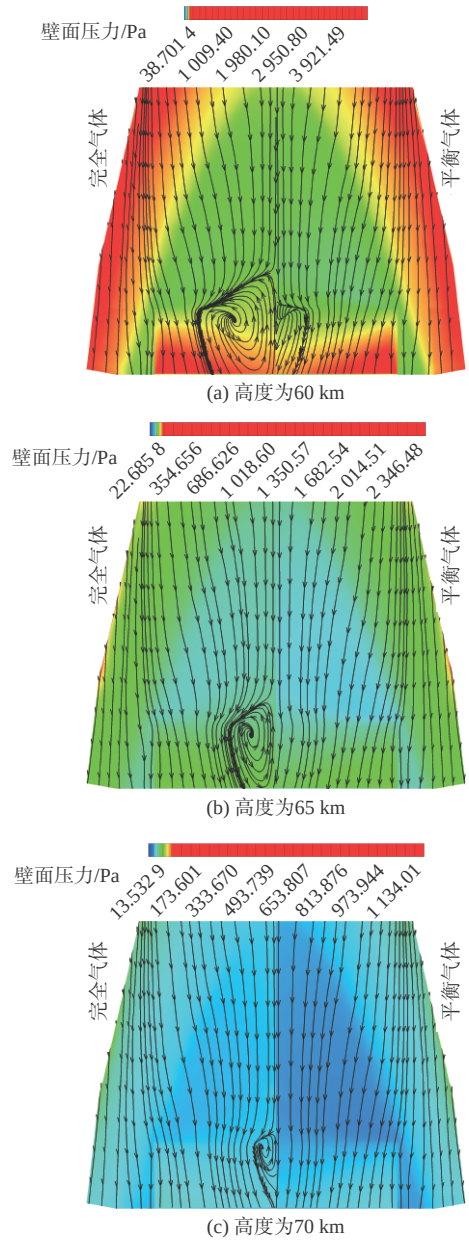


图 14　*Ma*=12.5、壁温 1 000 K 不同高度下升力体腹部壁面压力云图和极限流线(Le 滑移模型)

Fig. 14　Wall pressure distribution and limit streamline on windward surface of lift body at different heights under *Ma*=12.5 and wall temperature 1 000 K (Le slip model)

下、分离区不同壁面法向截面上的压力云图与流线图, 其中, 图 15(a) 和图 15(b) 截面为对称子午面, 图 15(c) 截面经过完全气体计算结果中的表面分离螺旋点。图 15(a) 中的平衡气体截面流线显示分离较小且流线局部呈螺旋型, 图 15(b) 则显示完全气体在该截面存在较大分离但没有出现螺旋型结构, 而图 15(c) 中的截面则显示完全气体截面流动出现了螺旋型结构, 因此, 截面分离流动形态与截面位置相关。同时可以看到完全气体分离区后的高压区范围较大, 而平衡气体的高压区较小, 因此, 平衡气体下游高压区对上游流动的影响更小, 边界层内压

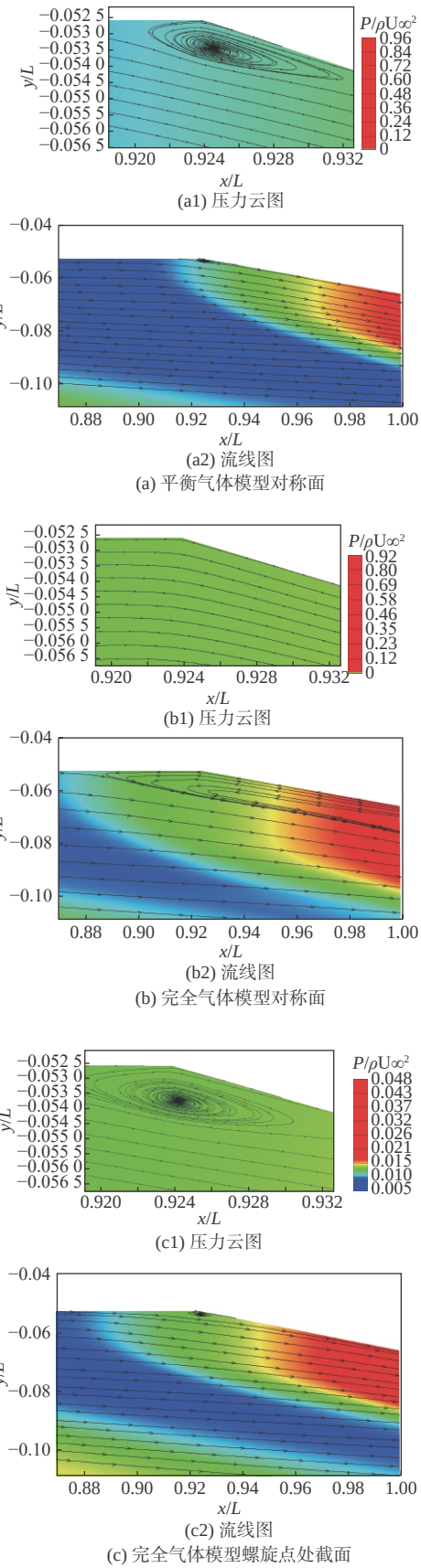


图 15　*Ma*=12.5、壁温 1 000 K、60 km 高度下升力体分离区压力云图和流线(Le 滑移模型)

Fig. 15　Pressure distribution and streamline of lift body at separation zone at 60 km height, *Ma*=12.5, and wall temperature 1 000 K (Le slip model)

力更小、速度更大。因此，导致平衡气体流动的分离区更小。关于不同气体模型对分离与旋涡的影响，后续还将对其中的机制开展深入研究。

表 3 为不同高度下、不同模型的驻点热流。平衡气体的驻点热流相较于完全气体的驻点热流平均高出约 37.5%。不同滑移模型间驻点热流的差距则较小，平均差距仅为 1.4% 左右。

表 3　不同高度下不同气体、滑移模型给出的升力体驻点热流

Table 3　Stagnation point heat flux of lift body by different gas and slip models at different heights　kW/m^2

气体	滑移	高度为60 km	高度为65 km	高度为70 km
平衡	M-S	2 397.25	1 647.80	1 042.67
	Le	2 383.09	1 639.62	1 092.05
完全	M-S	1 800.28	1 178.63	761.17
	Le	1 775.83	1 192.44	759.53

图 16 为不同高度下升力体总阻力系数的变化趋势，平衡气体的阻力系数整体上相较于完全气体更大。平衡气体下 Le 滑移模型与 M-S 滑移模型给出的阻力系数几乎一致，而完全气体下 Le 滑移模型得到的阻力系数则略大于 M-S 滑移模型。

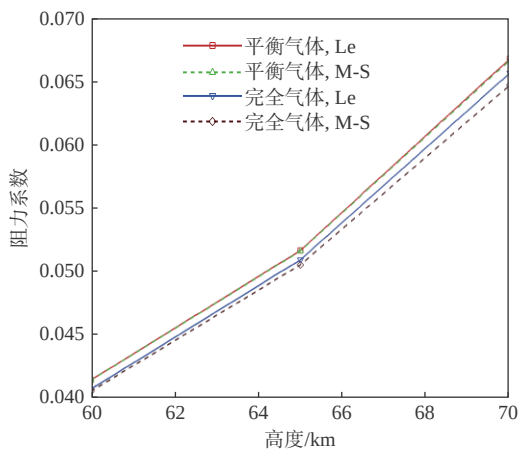


图 16　*Ma*=12.5、壁温 1 000 K 下不同滑移、气体模型升力体总阻力系数随高度变化

Fig. 16　Variation of total drag coefficient of lift body by different slip and gas models with height under *Ma*=12.5 and wall temperature 1 000 K

图 17 为不同高度下升力体压心位置的变化情况，结果显示随着高度的增加，升力体压心向后移动。压心位置大小由俯仰力矩系数除以法向力系数得到，由图 17(b) 和图 17(c) 可知，完全气体模型下，俯仰力矩系数和法向力系数绝对值均随高度增大而减小；平衡气体模型下，俯仰力矩系数和法向力系数绝对值随高度增大先增大后减小；此外，俯仰力矩系数和法向力系数的变化速率不同。上述

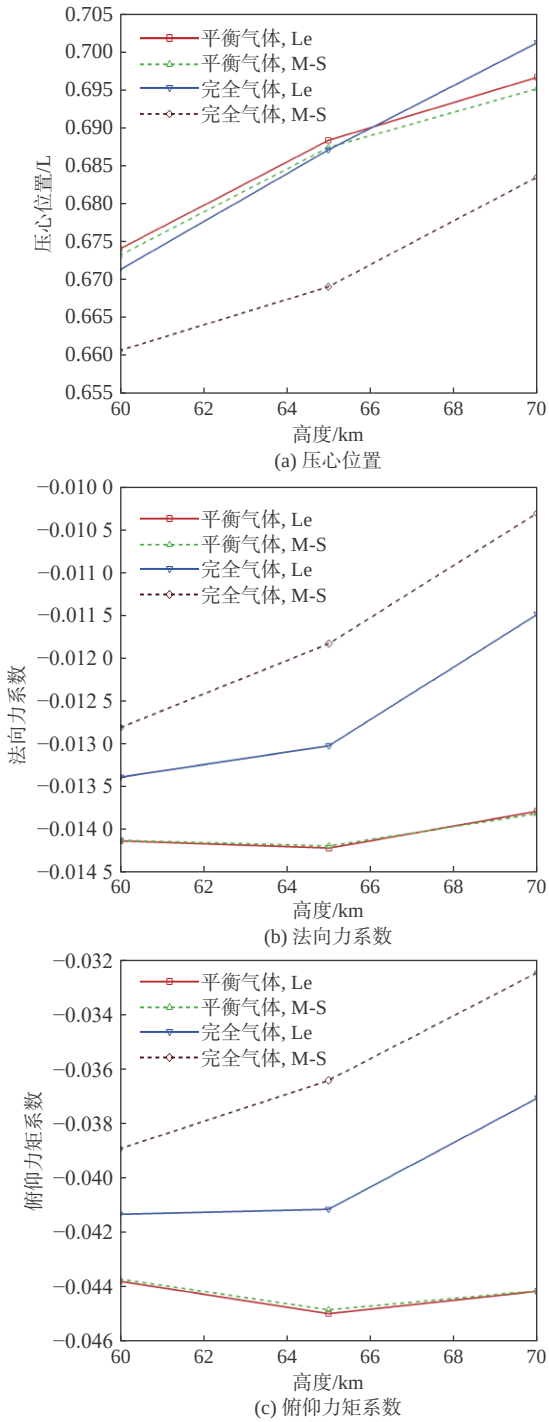


图 17　*Ma*=12.5、壁温 1 000 K 下不同滑移、气体模型升力体压心位置、法向力系数和俯仰力矩系数随高度变化

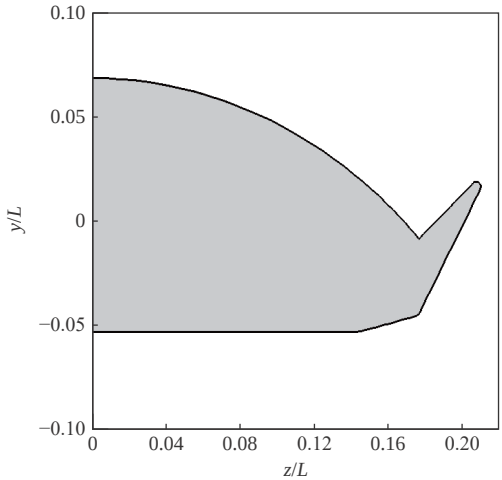
Fig. 17　Variation of position of the center of pressure, normal force coefficient, and pitching moment coefficient of lift body by different slip and gas models with height under *Ma*=12.5 and wall temperature 1 000 K

变化的综合效果体现为压心位置随高度增大后移。完全气体下 2 种滑移模型的压心位置差距较大，Le 滑移模型的压心更靠后；而平衡气体下不同滑移模型的差异则较小。

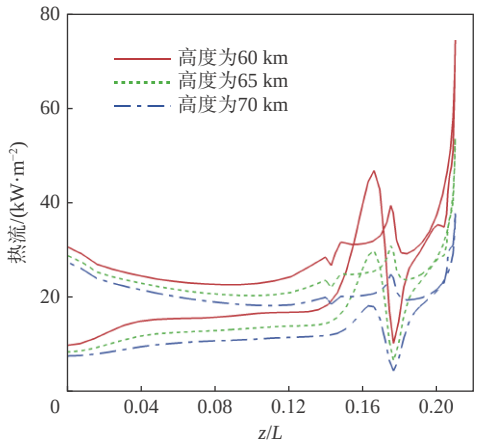
为定量显示不同高度下不同模型计算结果的差异，本文分别选取了 3 个流场特征位置，对沿展

向、流向和外形特征位置的表面参数分布进行分析比较。这 3 个具体位置分别是升力体底面升降舵前缘(即 x/L≈0.92 处)、升力体底部中心线位置和升力体侧缘线(见图 7)。升力体 x/L≈0.92 处轴向剖面壁面热流对比如图 18 所示。图 18(a) 显示了升力体剖面的外形示意图并给出了不同高度下热流的对比,结果显示随着高度的增加,热流逐渐减小,其中,峰值热流下降最为明显,此外上表面热流在中心线附近低于下表面,但在靠近翼的位置出现高于下表面的峰值,这与图 11 给出的热流分布云图一致。图 18(b) 对比了 60 km 高度不同模型的热流计算结果,完全气体模型热流较低且升力体上下表面中心线处热流相对较为接近,而平衡气体的上表面热流明显低于下表面。Le 滑移边界条件和 M-S 滑移边界条件在平衡气体模型下两者热流分布几乎一致,而在完全气体下,Le 滑移边界条件得到的升力体上表面热流大于 M-S 滑移条件得到结果。

升力体底面中心线上热流分布如图 19 所示。图 19(a) 对比了不同高度和气体模型,均采用了

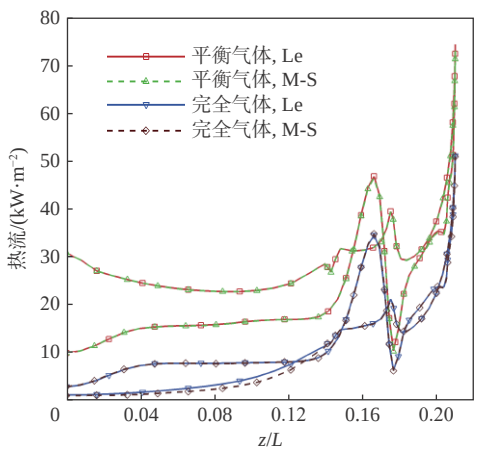


(a1) 外形示意图



(a1) 热流

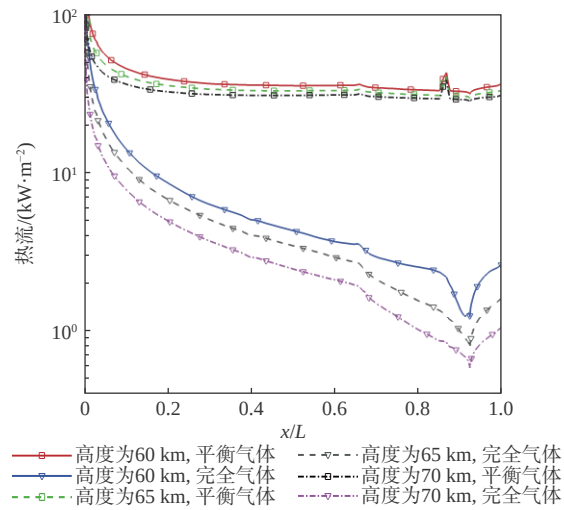
(a) 平衡气体、Le模型下不同高度对比及外形示意图



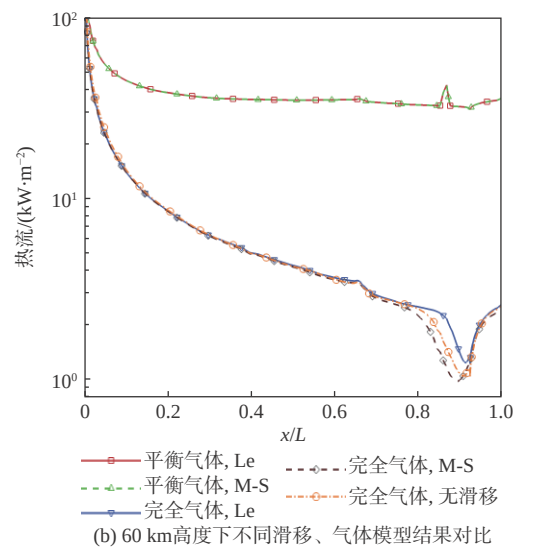
(b) 60 km高度下不同滑移、气体模型热流分布对比

图 18　Ma=12.5、壁温 1 000 K 不同计算条件下升力体 x/L≈0.92 处热流分布对比

Fig. 18　Heat flux distribution at x/L≈0.92 of lift body with different calculation conditions under Ma=12.5 and wall temperature 1 000 K



(a) 不同高度下不同气体模型给出的结果对比 (Le滑移模型)



(b) 60 km高度下不同滑移、气体模型结果对比

图 19　Ma=12.5、壁温 1 000 K 不同高度下升力体底面中心线不同气体、滑移模型给出的热流分布对比

Fig. 19　Heat flux distribution comparison of different gas and slip models on centerline of bottom section of lift body at different heights under Ma=12.5 and wall temperature 1 000 K

Le 滑移边界条件，结果显示随着高度的增加，热流逐渐减小，此外在中心线上上表面热流高于下表面。图 19(b) 给出了 60 km 高度下不同气体模型和滑移模型的热流分布对比，结果显示平衡气体模型下中心线上的热流比较平稳，且 2 种滑移模型的结果几乎一致；完全气体下热流从头部到尾部逐渐减小，且明显小于平衡气体的热流，尾部舵面分离区处热流明显较低，并且不同滑移模型的结果也在此处显现出一定差别。

图 20 为 Le 滑移模型下不同气体模型迎风面中心线摩阻系数的对比，结果显示随着高度的增加，摩阻系数呈现增大的趋势。在 60 km 高度时，平衡气体与完全气体的摩阻系数较为相近，随着高度增加，2 种气体模型的差异增大；在 70 km 高度下，完全气体的摩阻系数小于平衡气体。

图 21 为不同高度升力体侧缘线上的热流分布，图 21(a) 为平衡气体模型下不同滑移边界条件的对比，21(b) 为 Le 滑移模型下不同气体模型的对比。结果显示，热流总体上先逐渐下降，在机翼前缘处跃起，然后在头部脱体激波与翼脱体激波形成激波/激波干扰处（$x/L\approx0.76\sim0.78$）到达峰值，随后逐渐下降；在翼尖前缘（$x/L\approx0.88$）热流快速下降，之后在翼尖上平稳减小。不同滑移模型的侧缘线热流分布几乎无差别，但不同气体模型的热流分布存在显著差异，其中平衡气体模型的热流相较于完全气体更大。

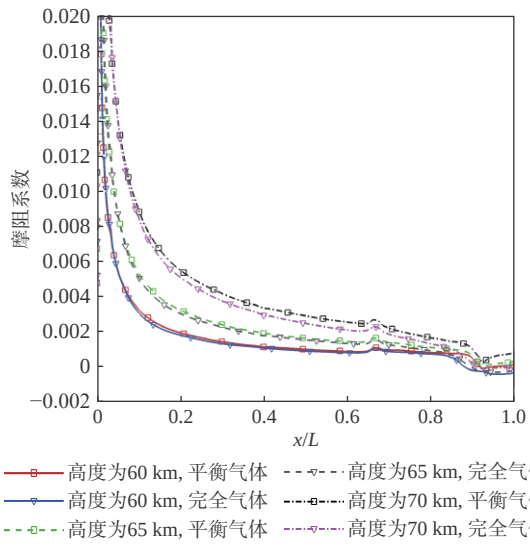
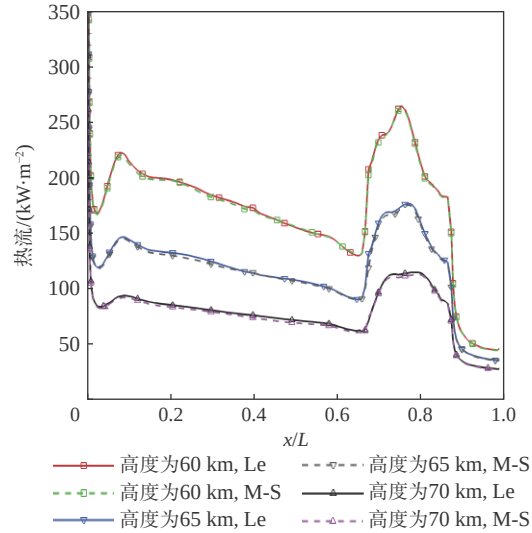
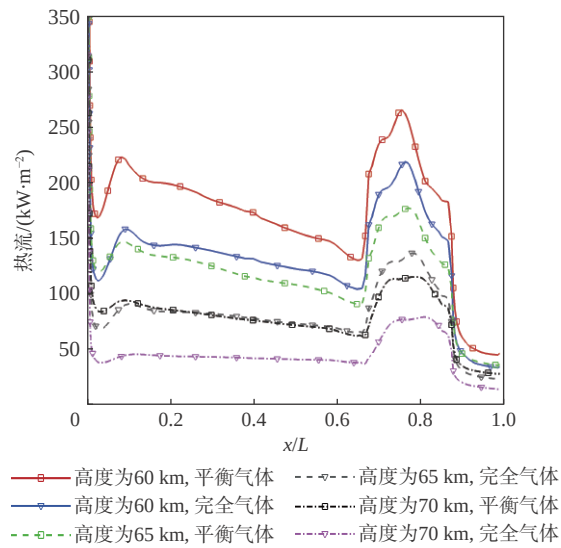


图 20 *Ma*=12.5、壁温 1 000 K 不同高度下升力体底面中心线不同气体模型给出的摩阻系数分布(Le 滑移模型)

Fig. 20 Friction coefficient distribution of different gas models on centerline of bottom section of lift body at different heights under *Ma*=12.5 and wall temperature 1 000 K (Le slip model)



(a) 不同滑移模型结果对比 (平衡气体)



(b) 不同气体模型结果对比 (Le滑移模型)

图 21 *Ma*=12.5、壁温 1 000 K 不同高度下升力体侧缘线不同气体、滑移模型给出的热流分布对比

Fig. 21 Heat flux distribution comparison of different gas and slip models of side edge line of lift body at different heights under *Ma*=12.5 and wall temperature 1 000 K

4.2 高度为 65 km 不同马赫数下的数值模拟与分析

本节分别对 *Ma* 为 10、12.5、15 的升力体流动进行了数值模拟，其他计算条件均为高度 65 km，壁温 1 000 K，采用平衡气体模型。

图 22 为 *Ma*=10 和 *Ma*=15 的下升力体腹部平台壁面压力云图与流线分布，*Ma*=12.5 的结果如图 14(b) 所示。结果显示，平衡气体计算结果在 *Ma*=10 出现了较大的分离涡，而 *Ma*=12.5 与 *Ma*=15 则没有分离区，并且分离区的结构与完全的气体计算的分离区类似。

图 23 为采用不同滑移模型下，升力体总阻力系数随马赫数的分布。结果表明，随着马赫数的增

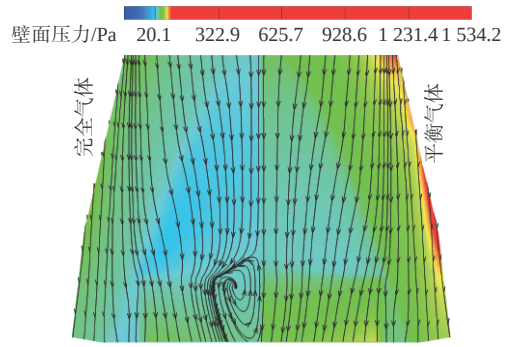


图 22　高度 65 km、壁温 1 000 K 不同马赫数下升力体腹部壁面压力云图和极限流线对比(Le 滑移模型)

Fig. 22　Wall pressure distribution and limit streamline on windward surface of lift body under different Mach numbers at 65 km height and wall temperature 1 000 K (Le slip model)

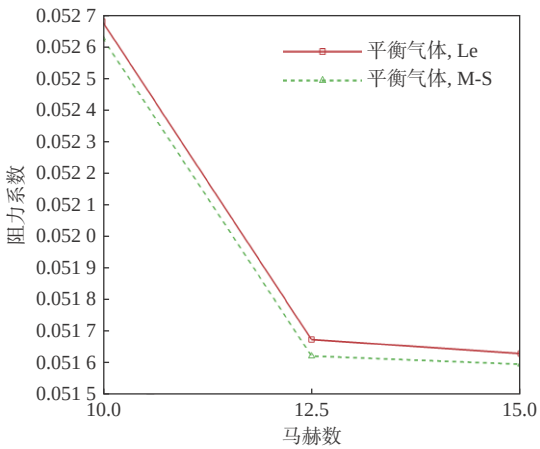


图 23　高度 65 km、壁温 1 000 K、平衡气体模型下不同滑移模型给出的升力体总阻力系数随马赫数变化

Fig. 23　Variation of total drag coefficient of lift body by different slip models under equilibrium gas model with Mach number under 65 km height and wall temperature 1 000 K

大, 阻力系数总的来说逐渐减小, 其中, 从 *Ma*=10 到 *Ma*=12.5 阻力系数减小显著, 而从 *Ma*=12.5 到 *Ma*=15 则变化较小; 相同马赫数下, Le 滑移模型得到的阻力系数略大于 M-S 滑移模型。

图 24 为升力体底部中心线摩阻系数随马赫数的分布, 计算均采用 Le 滑移模型。结果显示, 在升力体底部平台处即 $x/L\approx0.65$附近摩阻系数有一个小的上升; 在底部分离区位置 *Ma*=10 与 *Ma*=12.5 和 *Ma*=15 的摩阻系数分布具有较大差异, 这与其流动结构有关, 具体地: *Ma*=10 流动存在清晰的分离结构, 相应摩阻系数分布出现负值, 而 *Ma*=12.5 和 *Ma*=15 流场则没有明显的分离。

图 25 为迎风面中心线上滑移速度分布。结果显示, 滑移速度随着马赫数的增大而增加, 各马赫数下 M-S 模型的滑移速度均大于 Le 模型。

图 26 为不同马赫数下采用 Le 滑移模型给出的升力体侧缘线上摩阻系数分布。*Ma*=12.5 与 *Ma*=15 的摩阻系数分布接近, 均明显小于 *Ma*=10 的摩阻系数; 此外, 除头部高摩阻区外, 机翼迎风斜面的摩阻也较大。

此外, 本文进行了不同壁温对升力体流场影响的测试。结果显示在考虑温度跳跃边界条件后, 壁温对流动的影响很小。在 65 km 高度、*Ma*=12.5、平衡气体模型下, M-S 滑移边界条件在壁温为 1 000 K 和 1 500 K 下壁面处气体温度差距约为 5 K, 而 Le 滑移边界条件差异仅为 1 K; 阻力系数的变化则小于 0.1%。可见温度跳跃现象对流动的影响显著。

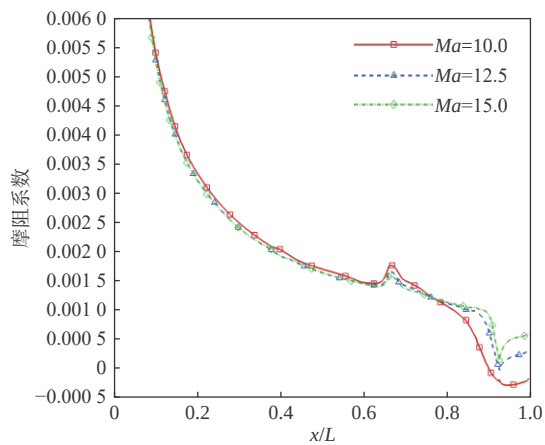


图 24　高度 65 km、壁温 1 000 K 不同马赫数下迎风面中心线摩阻系数分布(Le 滑移模型)

Fig. 24　Frictional coefficient distribution on centerline of windward surface under different Mach numbers at 65 km height and wall temperature 1 000 K (Le slip model)

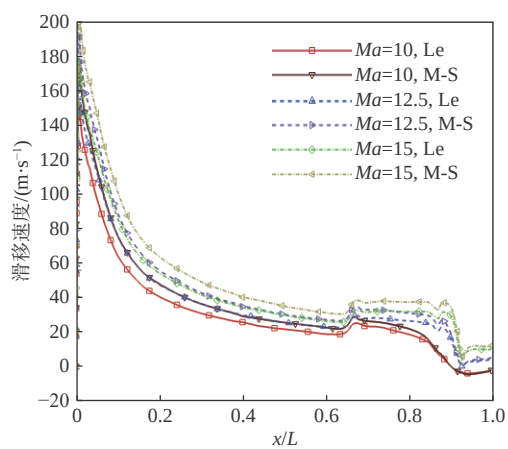


图 25　高度 65 km、壁温 1 000 K 不同马赫数下迎风面中心线不同滑移模型给出的滑移速度分布

Fig. 25　Slip velocity distribution of different slip models on centerline of windward surface under different Mach numbers at 65 km height and wall temperature 1 000 K

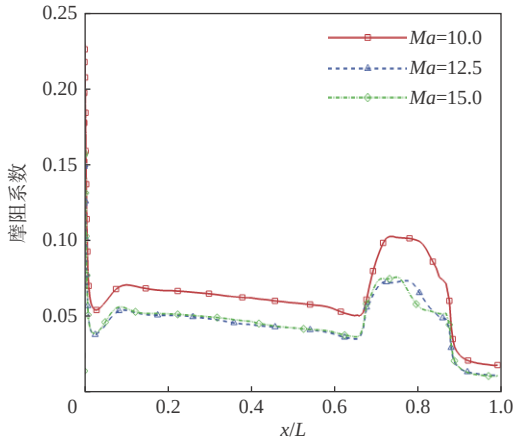


图 26 高度 65 km、壁温 1 000 K 不同马赫数下侧缘线上摩阻系数分布(Le 滑移模型)

Fig. 26 Friction coefficient distribution on side edge line under different Mach numbers at 65 km height and wall temperature 1 000 K (Le slip model)

5 结 论

本文结合高精度格式、平衡气体模型和高保真滑移边界条件,对高超声速升力体外形进行了数值模拟,考虑不同高度、马赫数和壁温下不同滑移边界模型与气体物理模型对升力体外形摩阻、热流等预示的影响,取得如下结论:

1)不同气体模型之间的结果存在较大差异,平衡气体模型预示的边界层温度更低、边界层厚度更小、壁面热流更大、摩阻系数与总阻力系数稍大,二者差异随高度增加而增大。

2)完全气体模型下,不同滑移边界模型总阻力系数、压心位置和热流分布均存在差异且随高度增加差异有所增大。

3)采用平衡气体模型时,不同滑移模型之间预测结果差异较小。

参考文献(References)

[1] SRINIVASAN S, TANNEHILL J C, WEILMUENSTER K. Simplified curve fits for the thermodynamic properties of equilibrium air. NASA RP-1181[R]. Washington, D. C. : NASA, 1987.

[2] SRINIVASAN S, TANNEHILL J C, WEILMUENSTER K J. Simplified curve fits for the transport properties of equilibrium Air: NASA CR-178411[R]. Washington, D. C. : NASA, 1987.

[3] MAXWELL J C. On stresses in rarified gases arising from inequalities of temperature[J]. Philosophical Transactions of the Royal Society of London, 1879, 170: 231-256.

[4] VON SMOLAN M S. Ueber Wärmeleitung in verdünnten gasen[J]. Annalen der Physik, 1898, 300(1): 101-130.

[5] JIANG G S, SHU C W. Efficient implementation of weighted ENO schemes[J]. Journal of Computational Physics, 1996, 126(1): 202-228.

[6] HENRICK A K, ASLAM T D, POWERS J M. Mapped weighted essentially non-oscillatory schemes: achieving optimal order near critical points[J]. Journal of Computational Physics, 2005, 207(2): 542-567.

[7] BORGES R, CARMONA M, COSTA B, et al. An improved weighted essentially non-oscillatory scheme for hyperbolic conservation laws[J]. Journal of Computational Physics, 2008, 227(6): 3191-3211.

[8] LI Q, YAN P, HUANG X, et al. On developing piecewise rational mapping with fine regulation capability for WENO schemes[J]. Journal of Scientific Computing, 2021, 88(3): 75.

[9] LI Q, YAN P, HUANG X, et al. Improvements to enhance robustness of third-order scale-independent WENO-Z schemes[J]. Advances in Applied Mathematics and Mechanics, 2024, 17(2): 373-406.

[10] LE N T P, WHITE C, REESE J M, et al. Langmuir–Maxwell and Langmuir–Smoluchowski boundary conditions for thermal gas flow simulations in hypersonic aerodynamics[J]. International Journal of Heat and Mass Transfer, 2012, 55(19-20): 5032-5043.

[11] LE N T P, VU N A, LOC L T. New type of Smoluchowski temperature jump condition considering the viscous heat generation[J]. AIAA Journal, 2017, 55(2): 474-483.

[12] LE N T, TRAN N H, TRAN T N, et al. New slip boundary condition in high-speed rarefied gas flow simulations[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2020, 234(3): 840-856.

[13] LE N T P, TRAN N H, TRAN T N. Modified Patterson temperature jump condition considering viscous heat generation[J]. International Journal of Heat and Mass Transfer, 2018, 126: 1267-1274.

[14] KARNIADAKIS G, BESKOK A, ALURU N, et al. Microflows and nanoflows: fundamentals and simulation[M]. New York: Springer New York, 2001: 59-66.

[15] BECKER, M. Flat plate flow files and surface measurements from merged layer into transition regime[C]// Proceedings of the seventh International Symposium on Rarefied Gas Dynamics. Pisa: University of Pisa, 1969 : 515–528.

[16] TANNEHILL J C, BUELOW P E, LEVALTS J O, et al. Three-dimensional upwind parabolized Navier-Stokes code for real gas-flows[J]. Journal of Spacecraft and Rockets, 1990, 27(2): 150-159.

[17] JOHN K H, HOLDEN M S, WADHAMS T P. Code validation study of laminarshock/boundary layer and shock/shock interactions in hypersonic flow[C]//Proceedings of the 39th Aerospace Science Meeting and Exhibit. Reston: AIAA, 2001-1031.

Friction and heat flux prediction of lift body under different gas models and slip boundary models

YAN Pan[1], LI Qin[1,*], HUANG Xiao[1], GUO Xiaoming[2], WENG Yihui[1], YOU Yancheng[1]

(1. School of Aerospace Engineering, Xiamen University, Xiamen 361000, China;
2. Science and Technology on Space Physics Laboratory, China Academy of Launch Vehicle Technology, Beijing 100076, China)

Abstract: In view of the friction and heat flux prediction problems in the hypersonic flows around a lift body, the influence of different gas models (perfect gas and equilibrium gas), slip boundary models (M-S and Le), and incoming flow conditions (height, Mach number, and wall temperature) on the prediction of friction and heat flux was studied with the developed third-order weighted essentially non-oscillatory scheme. Firstly, numerical simulation and analysis of double Mach reflection problem and typical hypersonic examples were carried out by different slip boundary models and gas models. The results show that high-fidelity gas models, slip boundary models, and high-precision schemes show better accuracy in the calculation of hypersonic problems. On this basis, numerical simulation and analysis of hypersonic flows around a lift body at different heights, Mach numbers, and wall temperatures are carried out. The effects of slip boundary models and gas models on the prediction of friction and heat flux are comprehensively analyzed. The results show that different gas models are quite different, and the equilibrium gas model gets lower temperature in the boundary layer, smaller thickness of the boundary layer, greater wall heat flux, and slightly larger friction coefficient and total drag coefficient. The difference between the two gas models increases with the increase in height. Under the perfect gas model, the total drag coefficient, position of the center of pressure, and heat flux distribution of different slip boundary models show greater difference, which increases with the increase in height. Under the equilibrium gas model, the results of different slip boundary models are similar.

Keywords: hypersonic flow; lift body; slip boundary condition; equilibrium gas; heat flux; friction

Received: 2023-04-25; **Accepted**: 2023-07-22; **Published Online**: 2023-08-30 16:46
URL: link.cnki.net/urlid/11.2625.V.20230830.1542.003

* **Corresponding author**. E-mail: qin-li@vip.tom.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0235

3种六自由度动力下降凸优化制导方法

王驰[1,2]，刘伟[1,2,*]，高扬[1,2]

(1. 中国科学院大学 航空宇航学院，北京 100049；2. 中国科学院空间应用工程与技术中心 太空应用重点实验室，北京 100094)

摘 要：飞行器动力下降软着陆的一个关键技术在于实时求解六自由度(6-DoF)动力下降制导问题，该问题可以描述为多约束条件下的燃料最省轨迹优化问题。选取飞行时间、时间替代变量、轨迹高度3种自变量建立3种优化模型，将原始轨迹优化问题转化为序列凸优化可解形式进行迭代求解，形成了3种在线制导方法。比较3种制导方法在收敛性、实时性、最优性及求解精度上的差异，结果表明：3种制导方法均能求解六自由度动力下降问题；自变量为飞行时间的制导方法计算时间最短且燃料消耗最少，但需要预先确定动力下降飞行时间；基于其余2类自变量的制导方法能够优化动力下降飞行时间，但均为次优解，且计算时间显著增加；相同离散点数量下3种方法的求解精度相近。若采用序列凸优化作为动力下降在线制导方案，如何确定最优飞行时间、逼近燃料最优解及进一步缩短计算时间等仍有待深入研究。

关 键 词：六自由度；动力下降；自主制导；凸优化；最优飞行时间
中图分类号：V448.131; V448.233
文献标志码：A **文章编号**：1001-5965(2025)04-1292-12

飞行器动力下降的在线制导是回收复用运载火箭的一项关键技术，在线制导算法需要同时具备实时性、收敛性和最优性，才能满足实际工程应用需求。飞行器动力下降问题具有时变、非线性、非凸约束的特点，在线制导算法成为一个难题。Song等[1]总结了动力下降软着陆自主制导方法，划分为解析制导方法、数值优化制导方法和基于学习的制导方法这3类。其中，解析制导方法虽然实时性好、无收敛性问题，但往往不是最优的；数值优化制导方法寻求最优解，但计算量偏大；基于学习的制导方法可以通过离线训练的方式避免数值优化方法在收敛性和计算效率方面的不足，但需要大量的样本数据，且其可靠性仍需进一步的实验验证。近年来，凸优化方法由于具有全局收敛性、局部最优即为全局最优、多项式时间收敛等优点，成为一类受到广泛关注的在线制导算法[2-4]，其潜在应用被推广至月球着陆[5-6]、火星着陆[7-9]及小行星着陆[10-11]。

本文重点关注凸优化方法在动力下降在线制导问题中的应用。针对飞行器动力下降在线制导问题，最初考虑解决简化的三自由度轨迹规划。该问题存在状态方程非线性、约束非凸的特点，因此，不能够直接利用凸优化方法求解，可以利用变量代换、等效转换等方法进行无损凸化[9,12]。Acikmese和Ploen[9]利用变量代换、引入新变量的方法将固定飞行时间三自由度动力下降问题纳入凸优化方法可解框架内，并基于该方法实现了10 km高度动力下降定点软着陆，制导计算时间仅需约16 ms[13]。但制导计算中忽略了气动力，且需要事先确定飞行时间。为了将飞行时间纳入优化框架，一些学者选取其他的自变量建立动力学方程。Liu等[12]引入无量纲能量作为自变量建立动力学方程，通过将非线性状态方程线性化，多次迭代求解再入段三自由度

收稿日期：2023-05-10；**录用日期**：2023-07-21；**网络出版时间**：2023-08-14 12:52
网络出版地址：link.cnki.net/urlid/11.2625.V.20230814.0912.001
基金项目：中国科学院青年创新促进会项目 (292020000040)
*通信作者. E-mail：liuwei@csu.ac.cn

引用格式：王驰，刘伟，高扬. 3种六自由度动力下降凸优化制导方法[J]. 北京航空航天大学学报，2025，51(4)：1292-1303.
WANG C，LIU W，GAO Y. Three convexification-based methods for six-degree-of-freedom powered descent guidance[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1292-1303 (in Chinese).

下降问题，即序列凸优化方法。大量仿真表明：序列凸优化方法能够快速收敛[14-15]，文献 [16-17] 也对序列凸优化方法的收敛性进行了理论研究，分析表明序列凸优化方法可以收敛到满足一阶必要条件的解。Yang 和 Liu[18] 选取高度作为动力学方程的自变量，无需确定飞行时间，解决了考虑气动力的三自由度动力下降问题。然而，三自由度轨迹规划忽略了飞行器姿轨耦合问题，对于基于矢量推进的六自由度动力下降并不适用。

为解决六自由度动力下降在线制导问题，Szmuk 等[19] 利用序列凸优化方法处理非线性状态方程和非凸约束，解决了固定时间六自由度动力下降问题。为确定动力下降的最优飞行时间，Szmuk 和 Acikmese [20] 引入了自由时间变量，通过序列凸优化方法求解，无需事先确定飞行时间。Reynolds 等[21] 进一步考虑了包含状态触发约束的六自由度动力下降问题。为进一步提高序列凸优化方法的实时性，近年来一些学者针对序列凸优化求解算法进行改进[22-23]。虽然序列凸优化方法可以对六自由度动力下降问题进行求解，但目前还没有研究对采用不同自变量构建六自由度问题的序列凸优化求解效果进行比较分析。

针对六自由度动力下降制导问题，本文采用序列凸优化方法进行求解，为探究以不同自变量建立动力学模型的求解效果及其在线制导的适应性，为提出可行的在线制导方案提供参考，本文通过算例的仿真分析比较了以飞行时间、时间替代变量和高度为自变量的 3 种参数化方法，其中，以飞行时间为自变量的参数化方法为固定时间优化，其余 2 种参数化方法为自由时间优化，并针对 3 种方法的收敛性、实时性、最优性和求解精度进行了分析。

1　3 种自变量的飞行器动力学方程

飞行器动力下降可以描述为六自由度动力学与控制问题。选取着陆场坐标系为参考坐标系，动力下降过程飞行时间短，可以将着陆场坐标系视为惯性坐标系。分别取着陆目标点和飞行器质心作为着陆场坐标系 $O_{\mathrm{I}}x_{\mathrm{I}}y_{\mathrm{I}}z_{\mathrm{I}}$ 和飞行器本体坐标系 $O_{\mathrm{B}}x_{\mathrm{B}}y_{\mathrm{B}}z_{\mathrm{B}}$ 的原点，两坐标系示意图如图 1 所示。其中，$\boldsymbol{g}$ 为重力加速度，$\boldsymbol{T}$ 为推力，$\boldsymbol{D}$ 为空气阻力，$\boldsymbol{r}$、$\boldsymbol{v}$ 分别为飞行器位置矢量和速度矢量。

1）飞行时间为自变量的动力学方程。

设飞行器姿态四元数 $\boldsymbol{q}=[q_0 \quad q_1 \quad q_2 \quad q_3]$，姿态角速度为 $\boldsymbol{\omega}$。以飞行时间 $t\in[t_0 \quad t_{\mathrm{f}}]$ 为自变量时，选取状态变量 $\boldsymbol{x}=[m \quad \boldsymbol{r} \quad \boldsymbol{v} \quad \boldsymbol{q} \quad \boldsymbol{\omega}]^{\mathrm{T}}$，控制量 $\boldsymbol{u}=\boldsymbol{T}$，飞行器的动力学方程为

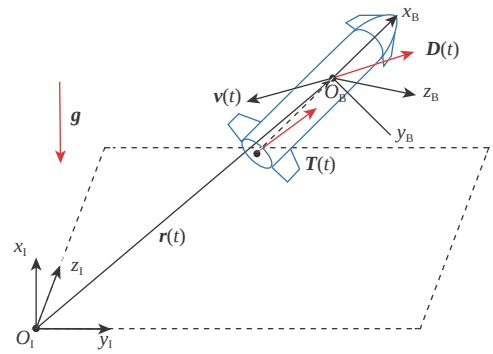


图 1　飞行器本体坐标系与着陆场坐标系示意图

Fig. 1　Schematic diagram of the vehicle body coordinate system and the landing site coordinate system

$$\dot{\boldsymbol{x}}=\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})=\begin{bmatrix}-\alpha\|\boldsymbol{T}\|\\ \boldsymbol{v}\\ \dfrac{\boldsymbol{C}_{\mathrm{IB}}\boldsymbol{T}}{m}+\boldsymbol{g}-\left(\dfrac{1}{2}\rho\|\boldsymbol{v}\|C_{\mathrm{d}}S\right)\boldsymbol{v}\\ \dfrac{1}{2}\boldsymbol{\Omega}(\boldsymbol{\omega})\boldsymbol{q}\\ \boldsymbol{J}^{-1}(\boldsymbol{r}_{\mathrm{T}}\times\boldsymbol{T}-\boldsymbol{\omega}\times\boldsymbol{J}\boldsymbol{\omega})\end{bmatrix} \tag{1}$$

式中：$\alpha=1/(I_{\mathrm{sp}}g_{\mathrm{e}})$；$I_{\mathrm{sp}}$ 为发动机比冲；g_{e} 为地球水平面加速度；ρ 为大气密度；C_{d} 为阻力系数；S 为参考面积；$\boldsymbol{J}$ 为转动惯量；$\boldsymbol{r}_{\mathrm{T}}$ 为质心至推力作用点的矢量；$\boldsymbol{C}_{\mathrm{IB}}$ 为飞行器本体坐标系至着陆场坐标系的方向余弦矩阵；$\boldsymbol{C}_{\mathrm{IB}}$、$\boldsymbol{\Omega}(\boldsymbol{\omega})$ 展开式分别为

$$\boldsymbol{C}_{\mathrm{IB}}=\begin{bmatrix}q_0^2+q_1^2-q_2^2-q_3^2 & 2(q_1q_2-q_3q_0) & 2(q_1q_3+q_2q_0)\\ 2(q_1q_2+q_3q_0) & q_0^2-q_1^2+q_2^2-q_3^2 & 2(q_2q_3-q_1q_0)\\ 2(q_1q_3-q_2q_0) & 2(q_2q_3+q_1q_0) & q_0^2-q_1^2-q_2^2+q_3^2\end{bmatrix}$$

$$\boldsymbol{\Omega}(\boldsymbol{\omega})=\begin{bmatrix}0 & -\omega_x & -\omega_y & -\omega_z\\ \omega_x & 0 & \omega_z & -\omega_y\\ \omega_y & -\omega_z & 0 & \omega_x\\ \omega_z & \omega_y & -\omega_x & 0\end{bmatrix} \tag{2}$$

2）时间替代变量为自变量的动力学方程。

引入时间替代变量 τ：

$$\frac{\mathrm{d}\tau}{\mathrm{d}t}=\frac{1}{\lambda} \tag{3}$$

则有

$$\tau=\frac{1}{\lambda}t+\tau_0 \tag{4}$$

令 $\tau\in[\tau_0 \quad \tau_{\mathrm{f}}]$，则 $t\in[0 \quad \lambda(\tau_{\mathrm{f}}-\tau_0)]$，对式 (1) 进行如下处理：

$$\begin{cases}\dot{\boldsymbol{x}}=\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\tau}\cdot\dfrac{\mathrm{d}\tau}{\mathrm{d}t}=\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\\ \dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\tau}=\dfrac{1}{\dfrac{\mathrm{d}\tau}{\mathrm{d}t}}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})=\lambda\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\end{cases} \tag{5}$$

飞行器动力学方程转化为

$$\dot{\boldsymbol{x}}=\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\tau}=\lambda\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u}) \tag{6}$$

式(6)采用$\tau\in[\tau_0\quad\tau_\mathrm{f}]$作为自变量，飞行时间由$\lambda(\tau_\mathrm{f}-\tau_0)$决定，状态量为$\boldsymbol{x}=[m\quad\boldsymbol{r}\quad\boldsymbol{v}\quad\boldsymbol{q}\quad\boldsymbol{\omega}]^\mathrm{T}$和$\lambda$。

3）高度为自变量的动力学方程。

选取$\boldsymbol{r}=[r_x\quad r_y\quad r_z]$中的高度$r_x$为自变量，飞行器动力学方程转化为

$$\begin{cases}\dot{\boldsymbol{x}}=\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}r_x}\cdot\dfrac{\mathrm{d}r_x}{\mathrm{d}t}=\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\\ \dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}r_x}=\dfrac{\mathrm{d}t}{\mathrm{d}r_x}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})=\dfrac{1}{v_x}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\end{cases} \tag{7}$$

即

$$\dot{\boldsymbol{x}}=\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}r_x}=\frac{1}{v_x}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u}) \tag{8}$$

式中：$r_x\in[r_{x0}\quad r_{xf}]$。由于$\mathrm{d}r_x/\mathrm{d}r_x\equiv1$，因此，状态量$\boldsymbol{x}=[m\quad r_y\quad r_z\quad\boldsymbol{v}\quad\boldsymbol{q}\quad\boldsymbol{\omega}]^\mathrm{T}$。比较式(1)、式(6)和式(8)可以发现，以时间替代变量和高度为自变量的动力学方程分别在以飞行时间为自变量的动力学方程的基础上乘以常变量λ和时变变量$1/v_x$。

2 3种参数优化问题

针对六自由度动力下降制导问题，上述3种动力学方程(即式(1)、式(6)、式(8))，可构建3种参数优化问题。其中，基于以飞行时间为自变量的动力学方程建立的参数优化问题需要给定飞行时间，为固定时间优化问题；其余2种优化问题为自由时间优化问题。3种优化问题的状态约束和控制约束相同，状态约束如式(9)所示，依次为路径约束、最大姿态角约束和最大姿态角速度约束；控制约束如式(10)所示，依次为推力幅值约束和推力摆角约束。

状态约束：

$$\begin{cases}\tan\gamma\left\|\boldsymbol{H}^\mathrm{T}\boldsymbol{r}(t)\right\|_2\leqslant\boldsymbol{e}_\mathrm{u}^\mathrm{T}\boldsymbol{r}(t)\\ \left\|\begin{bmatrix}0&0&1&0\\0&0&0&1\end{bmatrix}\boldsymbol{q}(t)\right\|_2\leqslant\sqrt{\dfrac{1-\cos\theta_{\max}}{2}}\\ \|\boldsymbol{\omega}(t)\|_2\leqslant\omega_{\max}\end{cases} \tag{9}$$

控制约束：

$$\begin{cases}T_{\min}\leqslant\|\boldsymbol{T}(t)\|_2\leqslant T_{\max}\\ \cos(\delta_{\max})\|\boldsymbol{T}(t)\|_2\leqslant\boldsymbol{e}_\mathrm{u}^\mathrm{T}\boldsymbol{T}(t)\end{cases} \tag{10}$$

式中：γ为轨迹锥半锥角的余角；$\theta_{\max}$为最大姿态角；$\omega_{\max}$为最大姿态角速度；$T_{\min}$为最小推力；$T_{\max}$为最大推力；$\delta_{\max}$为最大推力摆角；$\boldsymbol{H}=\begin{bmatrix}0&1&0\\0&0&1\end{bmatrix}^\mathrm{T}$；$\boldsymbol{e}_\mathrm{u}=\begin{bmatrix}1&0&0\end{bmatrix}^\mathrm{T}$。

由于控制量与状态量的耦合会导致求解结果的振荡[12,15,19]，因此，为了对控制量进行解耦，参考文献[19]中的处理方法，在原有动力学方程的基础上进行扩充，引入：

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{T}}{\mathrm{d}\mu}=\dot{\boldsymbol{T}}\\ \dfrac{\mathrm{d}\dot{\boldsymbol{T}}}{\mathrm{d}\mu}=\boldsymbol{u}\end{cases} \tag{11}$$

式中：μ为对应的自变量。

2.1 固定时间优化问题（以飞行时间为自变量）

基于以飞行时间为自变量的动力学方程建立六自由度动力下降参数优化问题。给定飞行时间t_f，自变量范围$t\in[t_0\quad t_\mathrm{f}]$，优化变量为$[m\quad\boldsymbol{r}\quad\boldsymbol{v}\quad\boldsymbol{q}\quad\boldsymbol{\omega}\quad\boldsymbol{T}\quad\dot{\boldsymbol{T}}\quad\ddot{\boldsymbol{T}}]^\mathrm{T}$，形成如下优化问题P1：

$$\begin{cases}\min\{-m(t_\mathrm{f})\}\\ \dot{\boldsymbol{x}}=\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\qquad t\in[t_0\quad t_\mathrm{f}]\end{cases} \tag{12a}$$

边界约束：

$$\begin{cases}m(t_0)=m_0\\ \boldsymbol{r}(t_0)=\boldsymbol{r}_0\qquad\boldsymbol{v}(t_0)=\boldsymbol{v}_0\\ \boldsymbol{q}(t_0)=\boldsymbol{q}_0\qquad\boldsymbol{\omega}(t_0)=\boldsymbol{\omega}_0\\ \boldsymbol{r}(t_\mathrm{f})=\boldsymbol{r}_\mathrm{f}\qquad\boldsymbol{v}(t_\mathrm{f})=\boldsymbol{v}_\mathrm{f}\\ \boldsymbol{q}(t_\mathrm{f})=\boldsymbol{q}_\mathrm{f}\qquad\boldsymbol{\omega}(t_\mathrm{f})=\boldsymbol{\omega}_\mathrm{f}\end{cases} \tag{12b}$$

状态与控制约束见式(9)、式(10)。

固定时间优化预先设定了飞行时间，因此，要寻找最优的飞行时间需要对时间进行遍历，多次求解。

2.2 自由时间优化问题(以时间替代变量为自变量)

基于以时间替代变量为自变量的动力学方程建立六自由度动力下降参数优化问题。自变量范围$\tau\in[\tau_0\quad\tau_\mathrm{f}]$，其中$\tau_0=0$，$\tau_\mathrm{f}=1$，优化变量为$[m\quad\boldsymbol{r}\quad\boldsymbol{v}\quad\boldsymbol{q}\quad\boldsymbol{\omega}\quad\boldsymbol{T}\quad\dot{\boldsymbol{T}}\quad\ddot{\boldsymbol{T}}]^\mathrm{T}$、$\lambda$，其中，$\lambda$可以对飞行时间进行优化。优化问题P2：

$$\begin{cases}\min\{-m(\tau_\mathrm{f})\}\\ \dot{\boldsymbol{x}}=\lambda\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\qquad\tau\in[\tau_0\quad\tau_\mathrm{f}]\end{cases} \tag{13a}$$

边界约束：

$$\begin{cases}m(\tau_0)=m_0\\ \boldsymbol{r}(\tau_0)=\boldsymbol{r}_0\qquad\boldsymbol{v}(\tau_0)=\boldsymbol{v}_0\\ \boldsymbol{q}(\tau_0)=\boldsymbol{q}_0\qquad\boldsymbol{\omega}(\tau_0)=\boldsymbol{\omega}_0\\ \boldsymbol{r}(\tau_\mathrm{f})=\boldsymbol{r}_\mathrm{f}\qquad\boldsymbol{v}(\tau_\mathrm{f})=\boldsymbol{v}_\mathrm{f}\\ \boldsymbol{q}(\tau_\mathrm{f})=\boldsymbol{q}_\mathrm{f}\qquad\boldsymbol{\omega}(\tau_\mathrm{f})=\boldsymbol{\omega}_\mathrm{f}\end{cases} \tag{13b}$$

状态与控制约束见式(9)、式(10)。

2.3 自由时间优化问题（以高度为自变量）

基于以高度为自变量的动力学方程建立六自由度动力下降参数优化问题。自变量范围$r_x\in[r_{x0}\quad r_{xf}]$，其中，$r_{x0}$为初始高度，$r_{xf}=0$。优化变量$\boldsymbol{x}=[m\quad r_y\quad r_z\quad\boldsymbol{v}\quad\boldsymbol{q}\quad\boldsymbol{\omega}\quad\boldsymbol{T}\quad\dot{\boldsymbol{T}}\quad\ddot{\boldsymbol{T}}]^\mathrm{T}$，以高度为自变量的优化问题减少了一个优化变量$r_x$。优化问题P3：

$$\begin{cases}\min\{-m(r_{xf})\}\\ \dot{\boldsymbol{x}}=\dfrac{1}{v_x}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u})\qquad r_x\in[r_{x0}\quad r_{xf}]\end{cases} \tag{14a}$$

边界约束:

$$\begin{cases} m(r_{x0})=m_0 & r_y(r_{x0})=r_{y0} \\ r_z(r_{x0})=r_{z0} & \boldsymbol{v}(r_{x0})=\boldsymbol{v}_0 \\ \boldsymbol{q}(r_{x0})=\boldsymbol{q}_0 & \boldsymbol{\omega}(r_{x0})=\boldsymbol{\omega}_0 \\ r_y(r_{xf})=r_z(r_{xf})=0 & \boldsymbol{v}(r_{xf})=\boldsymbol{v}_f \\ \boldsymbol{q}(r_{xf})=\boldsymbol{q}_f & \boldsymbol{\omega}(r_{xf})=\boldsymbol{\omega}_f \end{cases} \tag{14b}$$

状态与控制约束见式(9), 式(10)。

3　序列凸优化

序列凸优化将非线性优化问题中的非凸约束和目标函数在初始猜测附近线性化, 线性化后的问题为凸优化问题, 通过多次迭代求解凸优化问题可以逼近原问题的最优解。其收敛性在多篇论文中已经得到证明[17,24]。

对于第 2 节中的 3 种参数优化问题, 仅状态方程和推力幅值约束为非凸约束, 为了采用序列凸优化方法求解, 需要将非凸约束线性化, 再将问题离散化求解。下面以自由时间优化问题(以时间替代变量为自变量)为例介绍序列凸优化求解框架。对于以高度为自变量的自由时间优化问题和固定时间优化问题的凸优化求解框架, 可以参考文献 [18-19]。

3.1　非凸约束线性化

以时间替代变量为自变量的动力学方程可表示为

$$\dot{\boldsymbol{x}}(\tau)=\boldsymbol{g}(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\lambda) \tag{15}$$

设初始猜测为 $\hat{\boldsymbol{x}}(\tau)$、$\hat{\boldsymbol{u}}(\tau)$、$\hat{\lambda}$, 线性化后的状态方程可表示为

$$\begin{cases} \dot{\boldsymbol{x}}(\tau)=\boldsymbol{A}(\tau)\boldsymbol{x}(\tau)+\boldsymbol{B}(\tau)\boldsymbol{u}(\tau)+\boldsymbol{\Sigma}(\tau)\lambda+\boldsymbol{z}(\tau) \\ \boldsymbol{A}(\tau)=\left.\dfrac{\partial \boldsymbol{g}(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\lambda)}{\partial \boldsymbol{x}}\right|_{\hat{\boldsymbol{x}}(\tau),\hat{\boldsymbol{u}}(\tau),\hat{\lambda}} \\ \boldsymbol{B}(\tau)=\left.\dfrac{\partial \boldsymbol{g}(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\lambda)}{\partial \boldsymbol{u}}\right|_{\hat{\boldsymbol{x}}(\tau),\hat{\boldsymbol{u}}(\tau),\hat{\lambda}} \\ \boldsymbol{\Sigma}(\tau)=\left.\dfrac{\partial \boldsymbol{g}(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\lambda)}{\partial \lambda}\right|_{\hat{\boldsymbol{x}}(\tau),\hat{\boldsymbol{u}}(\tau),\hat{\lambda}} \\ \boldsymbol{z}(\tau)=\boldsymbol{g}(\hat{\boldsymbol{x}}(\tau),\hat{\boldsymbol{u}}(\tau),\hat{\lambda})-\boldsymbol{A}(\tau)\hat{\boldsymbol{x}}(\tau)-\boldsymbol{B}(\tau)\hat{\boldsymbol{u}}(\tau)- \\ \qquad \boldsymbol{\Sigma}(\tau)\hat{\lambda} \end{cases} \tag{16}$$

将推力幅值约束在初始猜测附近线性化并化简后, 可得

$$T_{\min}\leqslant\frac{\hat{\boldsymbol{T}}^{\mathrm{T}}(\tau)}{\left\|\hat{\boldsymbol{T}}(\tau)\right\|}\boldsymbol{T}(\tau)\leqslant T_{\max} \tag{17}$$

3.2　动力学方程离散化

设自变量 $\tau\in[\tau_0\quad\tau_f]$, 将其均匀离散化为 K 段, 共有 $K+1$ 个节点, 每个节点对应的自变量可以表示为

$$\begin{cases} \tau_k=\tau_0+k\Delta\tau \\ \Delta\tau=\dfrac{\tau_f-\tau_0}{K} \end{cases} \qquad k=0,1,\cdots,K \tag{18}$$

对于以高度为自变量优化, 由于下降过程末端飞行器速度逐渐减小, 将高度均匀离散会导致时间间隔增大, 从而引起末端误差增大。对此采用文献 [18] 中提出的非均匀离散点选择策略进行优化。

控制量 $\boldsymbol{u}$ 采用一阶保持模型离散化, 可以用 $\boldsymbol{u}_k=\boldsymbol{u}(\tau_k)$ 和 $\boldsymbol{u}_{k+1}=\boldsymbol{u}(\tau_{k+1})$ 表示为

$$\begin{cases} \boldsymbol{u}(\tau)=\lambda_k^-(\tau)\boldsymbol{u}_k+\lambda_k^+(\tau)\boldsymbol{u}_{k+1} \\ \qquad \tau\in\left[\tau_k\quad\tau_{k+1}\right], \qquad k=0,1,\cdots,K-1 \end{cases} \tag{19}$$

式中: $\lambda_k^-(\tau)=\left(\dfrac{\tau_{k+1}-\tau}{\tau_{k+1}-\tau_k}\right)$, $\lambda_k^+(\tau)=\left(\dfrac{\tau-\tau_k}{\tau_{k+1}-\tau_k}\right)$。

常用的数值积分方法包括矩形积分、四阶龙格库塔积分、伪谱法等, 仿真结果表明, 不同方法间的差异较小[18,25]。本文中离散系统状态递推关系采用四阶龙格库塔积分方法将连续系统离散化:

$$\begin{cases} \boldsymbol{x}_k=\boldsymbol{x}_{k-1}+\dfrac{\Delta\tau}{6}(\boldsymbol{K}_1+2\boldsymbol{K}_2+2\boldsymbol{K}_3+\boldsymbol{K}_4) \\ \boldsymbol{K}_1=\boldsymbol{g}\left[\tau_{k-1}\quad\boldsymbol{x}_{k-1}\right] \\ \boldsymbol{K}_2=\boldsymbol{g}\left[\tau_{k-1}+\dfrac{\Delta\tau}{2}\quad\boldsymbol{x}_{k-1}+\dfrac{\Delta\tau}{2}\boldsymbol{K}_1\right] \\ \boldsymbol{K}_3=\boldsymbol{g}\left[\tau_{k-1}+\dfrac{\Delta\tau}{2}\quad\boldsymbol{x}_{k-1}+\dfrac{\Delta\tau}{2}\boldsymbol{K}_2\right] \\ \boldsymbol{K}_4=\boldsymbol{g}\left[\tau_{k-1}+\Delta\tau\quad\boldsymbol{x}_{k-1}+\Delta\tau\boldsymbol{K}_3\right] \\ k=1,2,\cdots,K \end{cases} \tag{20}$$

将式 (6) 代入式(20)中, 可得

$$\begin{cases} \boldsymbol{x}_k=\boldsymbol{x}_{k-1}+\dfrac{\Delta\tau}{6}(\boldsymbol{K}_1+2\boldsymbol{K}_2+2\boldsymbol{K}_3+\boldsymbol{K}_4) \\ \boldsymbol{K}_1=\boldsymbol{A}_{k-1}\boldsymbol{x}_{k-1}+\boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1}+\boldsymbol{\Sigma}_{k-1}\lambda+\boldsymbol{z}_{k-1} \\ \boldsymbol{K}_2=\boldsymbol{A}_{\bar{k}}\left(\boldsymbol{x}_{k-1}+\dfrac{\Delta\tau}{2}\boldsymbol{K}_1\right)+\boldsymbol{B}_{\bar{k}}\dfrac{\boldsymbol{u}_{k-1}+\boldsymbol{u}_k}{2}+\boldsymbol{\Sigma}_{\bar{k}}\lambda+\boldsymbol{z}_{\bar{k}} \\ \boldsymbol{K}_3=\boldsymbol{A}_{\bar{k}}\left(\boldsymbol{x}_{k-1}+\dfrac{\Delta\tau}{2}\boldsymbol{K}_2\right)+\boldsymbol{B}_{\bar{k}}\dfrac{\boldsymbol{u}_{k-1}+\boldsymbol{u}_k}{2}+\boldsymbol{\Sigma}_{\bar{k}}\lambda+\boldsymbol{z}_{\bar{k}} \\ \boldsymbol{K}_4=\boldsymbol{A}_k(\boldsymbol{x}_{k-1}+\Delta\tau\boldsymbol{K}_3)+\boldsymbol{B}_k\boldsymbol{u}_k+\boldsymbol{\Sigma}_k\lambda+\boldsymbol{z}_k \\ k=1,2,\cdots,K \end{cases} \tag{21}$$

式中: $\boldsymbol{A}_j=\boldsymbol{A}(\tau_j)$, $\boldsymbol{B}_j=\boldsymbol{B}(\tau_j)$, $\boldsymbol{\Sigma}_j=\boldsymbol{\Sigma}(\tau_j)$, $\boldsymbol{z}_j=\boldsymbol{z}(\tau_j)$, $j=k-1$、$\bar{k}$、k。$\tau_{\bar{k}}=(\tau_{k-1}+\tau_k)/2$。对于 $\boldsymbol{A}(\tau_{\bar{k}})$ 采用式 (22) 进行近似计算, $\boldsymbol{B}_{\bar{k}}$、$\boldsymbol{\Sigma}_{\bar{k}}$、$\boldsymbol{z}_{\bar{k}}$ 同理:

$$\boldsymbol{A}(\tau_{\bar{k}})=\left.\frac{\partial \boldsymbol{g}(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\lambda)}{\partial \boldsymbol{x}}\right|_{\frac{\hat{\boldsymbol{x}}(\tau_{k-1})+\hat{\boldsymbol{x}}(\tau_k)}{2},\frac{\hat{\boldsymbol{u}}(\tau_{k-1})+\hat{\boldsymbol{u}}(\tau_k)}{2},\hat{\lambda}} \tag{22}$$

3.3　序列凸优化子问题

经过线性化、离散化后, 非线性优化问题被转化为凸优化子问题。为保证序列凸优化迭代的稳定, 还需要进行以下处理:

1) 由于将非凸约束在初始猜测 $\hat{\boldsymbol{x}}(\tau)$、$\hat{\boldsymbol{u}}(\tau)$、$\hat{\lambda}$ 处线性化, 因此, 当状态偏离初始猜测较远时, 会

引起较大的线性化误差，甚至导致优化问题无解。为避免这种情况，需要引入如下约束：

$$\begin{cases}\Delta\boldsymbol{x}_k^{\mathrm{T}}\Delta\boldsymbol{x}_k+\Delta\boldsymbol{u}_k^{\mathrm{T}}\Delta\boldsymbol{u}_k\leqslant\eta_k & k=0,1,\cdots,K\\ \left(\lambda-\hat{\lambda}\right)^2\leqslant\delta_\lambda\end{cases}\tag{23}$$

式中：$\Delta\boldsymbol{x}_k=\boldsymbol{x}_k-\hat{\boldsymbol{x}}_k$，$\Delta\boldsymbol{u}_k=\boldsymbol{u}_k-\hat{\boldsymbol{u}}_k$。式(23)形成了一个“信任域”，为了限制信任域的大小，在原有目标函数$-m(\tau_{\mathrm{f}})$的基础上增加惩罚项$w_\eta\|\boldsymbol{\eta}\|_2+w_\lambda\delta_\lambda$。

2）由于线性化后的状态方程与原方程不等价，为避免线性化导致优化问题无解，引入松弛变量$\boldsymbol{\kappa}=[\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2,\cdots,\boldsymbol{\kappa}_K]$，新的状态方程表示为

$$\begin{cases}\boldsymbol{x}_k=\boldsymbol{x}_{k-1}+\dfrac{\Delta\tau}{6}(\boldsymbol{K}_1+2\boldsymbol{K}_2+2\boldsymbol{K}_3+\boldsymbol{K}_4)+\kappa_k\\ k=1,2,\cdots,K\end{cases}\tag{24}$$

同时在原有目标函数$-m(\tau_{\mathrm{f}})$的基础上增加对松弛变量的惩罚项$w_{\boldsymbol{\kappa}}\|\boldsymbol{\kappa}\|_2$。

至此，原非线性优化问题被转化为如下凸优化子问题PP：

$$\begin{cases}\min\{-m(\tau_{\mathrm{f}})+w_{\eta}\|\boldsymbol{\eta}\|_2+w_{\kappa}\|\boldsymbol{\kappa}\|_2+w_\lambda\delta_\lambda\}\\ \quad\boldsymbol{x}_k=\boldsymbol{x}_{k-1}+\dfrac{\Delta\tau}{6}(\boldsymbol{K}_1+2\boldsymbol{K}_2+2\boldsymbol{K}_3+\boldsymbol{K}_4)+\boldsymbol{\kappa}_k\\ \quad k=1,2,\cdots,K\end{cases}\tag{25a}$$

边界约束：

$$\begin{cases}m(\tau_0)=m_0\\ \boldsymbol{r}(\tau_0)=\boldsymbol{r}_0 \qquad \boldsymbol{v}(\tau_0)=\boldsymbol{v}_0\\ \boldsymbol{q}(\tau_0)=\boldsymbol{q}_0 \qquad \boldsymbol{\omega}(\tau_0)=\boldsymbol{\omega}_0\\ \boldsymbol{r}(\tau_K)=\boldsymbol{r}_{\mathrm{f}} \qquad \boldsymbol{v}(\tau_K)=\boldsymbol{v}_{\mathrm{f}}\\ \boldsymbol{q}(\tau_K)=\boldsymbol{q}_{\mathrm{f}} \qquad \boldsymbol{\omega}(\tau_K)=\boldsymbol{\omega}_{\mathrm{f}}\end{cases}\tag{25b}$$

状态约束：

$$\begin{cases}\tan\gamma\|\boldsymbol{H}^{\mathrm{T}}\boldsymbol{r}(\tau_k)\|_2\leqslant\boldsymbol{e}_{\mathrm{u}}^{\mathrm{T}}\boldsymbol{r}(\tau_k)\\ \left\|\begin{bmatrix}0&0&1&0\\0&0&0&1\end{bmatrix}\boldsymbol{q}(\tau_k)\right\|_2\leqslant\sqrt{\dfrac{1-\cos\theta_{\max}}{2}}\\ \|\boldsymbol{\omega}(\tau_k)\|_2\leqslant\omega_{\max}\end{cases}\tag{25c}$$

控制约束：

$$\begin{cases}\boldsymbol{T}_{\min}\leqslant\dfrac{\hat{\boldsymbol{T}}^{\mathrm{T}}(\tau_k)}{\|\hat{\boldsymbol{T}}(\tau_k)\|}\boldsymbol{T}(\tau_k)\leqslant\boldsymbol{T}_{\max}\\ \cos(\delta_{\max})\|\boldsymbol{T}(\tau_k)\|_2\leqslant\boldsymbol{e}_{\mathrm{u}}^{\mathrm{T}}\boldsymbol{T}(\tau_k) \qquad k=0,1,\cdots,K\end{cases}\tag{25d}$$

信任域约束：

$$\begin{cases}\Delta\boldsymbol{x}_k^{\mathrm{T}}\Delta\boldsymbol{x}_k+\Delta\boldsymbol{u}_k^{\mathrm{T}}\Delta\boldsymbol{u}_k\leqslant\eta_k & k=0,1,\cdots,K\\ \left(\lambda-\hat{\lambda}\right)^2\leqslant\delta_\lambda\end{cases}\tag{25e}$$

3.4 初值与迭代

优化变量初值：

$$\begin{cases}\hat{m}_k=\left(\dfrac{K-k}{K}\right)m_{\mathrm{wet}}+\left(\dfrac{k}{K}\right)m_{\mathrm{dry}}\\ \hat{\boldsymbol{r}}_k=\left(\dfrac{K-k}{K}\right)\boldsymbol{r}_0+\left(\dfrac{k}{K}\right)\boldsymbol{r}_{\mathrm{f}}\\ \hat{\boldsymbol{v}}_k=\left(\dfrac{K-k}{K}\right)\boldsymbol{v}_0+\left(\dfrac{k}{K}\right)\boldsymbol{v}_{\mathrm{f}}\\ \hat{\boldsymbol{q}}_k=\begin{bmatrix}1&0&0&0\end{bmatrix}^{\mathrm{T}}\\ \hat{\boldsymbol{\omega}}_k=\begin{bmatrix}0&0&0\end{bmatrix}^{\mathrm{T}}\\ \hat{\boldsymbol{T}}_k=\begin{bmatrix}-\hat{m}_kg&0&0\end{bmatrix}^{\mathrm{T}}\\ \hat{\dot{\boldsymbol{T}}}_k=\begin{bmatrix}0&0&0\end{bmatrix}^{\mathrm{T}}\\ \hat{\ddot{\boldsymbol{T}}}_k=\begin{bmatrix}0&0&0\end{bmatrix}^{\mathrm{T}}\end{cases}\qquad k=0,1,\cdots,K\tag{26}$$

$$\begin{cases}\hat{\boldsymbol{x}}_k^0=\begin{bmatrix}\hat{m}_k^{\mathrm{T}}&\hat{\boldsymbol{r}}_k^{\mathrm{T}}&\hat{\boldsymbol{v}}_k^{\mathrm{T}}&\hat{\boldsymbol{q}}_k^{\mathrm{T}}&\hat{\omega}_k^{\mathrm{T}}&\hat{\boldsymbol{T}}_k^{\mathrm{T}}&\hat{\dot{\boldsymbol{T}}}_k^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}\\ \hat{\boldsymbol{u}}_k^0=\hat{\ddot{\boldsymbol{T}}}_k^{\mathrm{T}} \qquad k=0,1,\cdots,K\\ \hat{\lambda}^0=1\end{cases}\tag{27}$$

式中：m_{wet}为初始质量；m_{dry}为结构质量。

序列凸优化迭代算法如下：

For $i=1:i_{\max}$

由$\hat{\boldsymbol{x}}^{i-1}$、$\hat{\boldsymbol{u}}^{i-1}$、$\hat{\lambda}^{i-1}$计算凸优化子问题PP中的参数$\boldsymbol{A}_j$、$\boldsymbol{B}_j$、$\boldsymbol{\Sigma}_j$、$\boldsymbol{z}_j$，其中，$j=k-1$、$\bar{k}$、$k$，$k=1,2,\cdots,K$。

求解凸优化子问题PP，得到一组解$\boldsymbol{x}^i$、$\boldsymbol{u}^i$、λ^i。

如果$\|\boldsymbol{\eta}\|_2\leqslant\varepsilon_{\boldsymbol{\eta}}$、$\|\boldsymbol{\kappa}\|_2\leqslant\varepsilon_{\boldsymbol{\kappa}}$，迭代误差收敛到误差限以内，迭代停止。

令$\hat{\boldsymbol{x}}^i=\boldsymbol{x}^i$、$\hat{\boldsymbol{u}}^i=\boldsymbol{u}^i$、$\hat{\lambda}^i=\lambda^i$。

End

4 仿真结果及分析

飞行器5 km高度至地面的动力下降过程的参数如表1所示。飞行器初始状态、目标状态参数及约束参数如表2所示。采用cvx软件包[26]中的SDPT3[27]求解器对凸问题进行求解，CPU为3.7 GHz的i7-8700K。

将飞行器视为均质圆柱体，飞行器的转动惯量为

$$\boldsymbol{J}=\begin{bmatrix}\dfrac{1}{2}R^2&&\\&\dfrac{1}{4}R^2+\dfrac{1}{12}L^2&\\&&\dfrac{1}{4}R^2+\dfrac{1}{12}L^2\end{bmatrix}m\tag{28}$$

式中：R为圆柱体底面半径；L为圆柱体长度；m为飞行器质量。

所有仿真中均采用归一化方法进行无量纲处理，以避免不同变量的数量级差距过大可能导致的优化失败。对于任意参数h，其归一化结果$\bar{h}$可以表示为

表 1　仿真参数

Table 1　Simulation parameters

参数	数值
初始质量m_{wet}/kg	65 170
结构质量m_{dry}/kg	25 600
圆柱体底面半径R/m	1.85
圆柱体长底L/m	41.2
最大推力$T_{\max}$/N	845 000
最小推力$T_{\min}$/N	845 000×0.4
比冲I_{sp}/s	310
质心至推力作用点位置矢量$\boldsymbol{r}_{\text{T}}$/m	[−20　0　0]
大气密度ρ/(kg·m^{-3})	1.225
阻力系数C_{d}	1.3
参考面积S/m^2	$\pi r^2, r$ = 1.85 m

表 2　状态参数

Table 2　State parameters

状态参数	数值
初始位置$\boldsymbol{r}_0$/m	[5 000　1 000　0]
初始速度$\boldsymbol{v}_0$/(m·s^{-1})	[−100　−10　0]
初始姿态四元数$\boldsymbol{q}_0$	[1　0　0　0]
初始姿态角速度$\boldsymbol{\omega}_0$/(rad·s^{-1})	[0　0　0]
末端位置$\boldsymbol{r}_{\text{f}}$/m	[0　0　0]
末端速度$\boldsymbol{v}_{\text{f}}$/(m·s^{-1})	[−1　0　0]
末端姿态四元数$\boldsymbol{q}_{\text{f}}$	[1　0　0　0]
末端姿态角速度$\boldsymbol{\omega}_{\text{f}}$/(rad·s^{-1})	[0　0　0]
路径角γ/(°)	4
最大姿态角$\theta_{\max}$/(°)	30
最大姿态角速度$\omega_{\max}$/((°)·s^{-1})	5
最大推力摆角$\delta_{\max}$/(°)	20

$$\bar{h} = h/h_s \tag{29}$$

式中：h_s为参数h的量纲所对应的归一化参数。给出归一化距离参数、归一化时间参数和归一化推力参数，其余量纲的归一化参数可由这 3 个参数导出。

4.1　固定时间优化

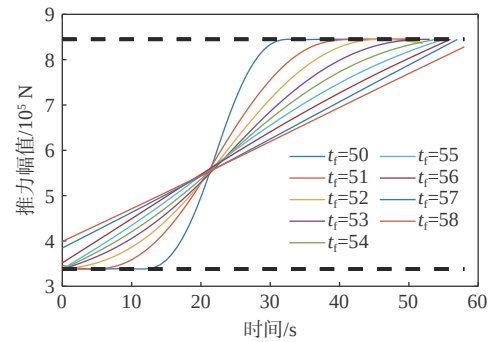
固定时间优化初值参考式 (26) 和式 (27) 给出。仿真参数如表 3 所示。

对于固定时间优化问题，如图 2 所示。随着飞行时间的减少，推力幅值越来越接近开关控制，燃料消耗也越来越少，当t_{f} = 49 s时迭代不收敛；当t_{f} = 50 s时，达到最大迭代次数仍未收敛到误差限以内，但误差逐渐减小。这说明最优飞行时间在 50 s 附近。随着飞行时间逐渐接近最优时间，迭代次数也出现了明显增加。这种情况是由于采用$\boldsymbol{u} = \ddot{\boldsymbol{T}}$为控制量导致的，最优推力表现为开关控制，随着飞行时间逐渐接近最优时间，序列凸优化的解也会逐渐逼近最优解，此时推力的二阶导数在开关点会趋

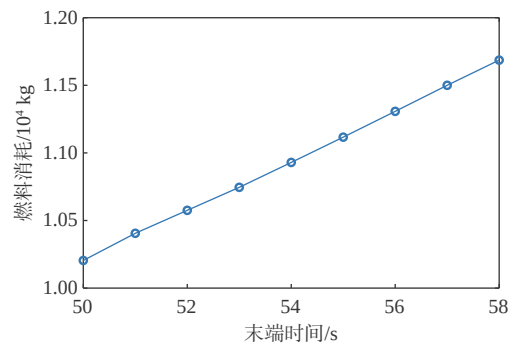
表 3　固定时间序列凸优化参数

Table 3　Fixed time sequential convex optimization parameters

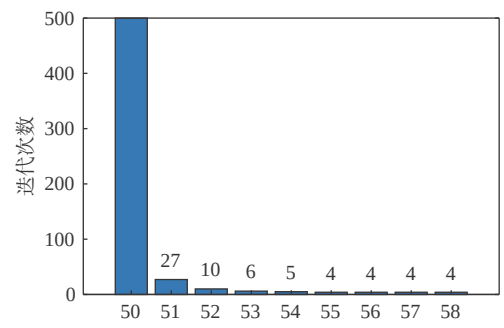
参数	数值
时间t_{f}/s	50,51,⋯,58
归一化距离l_{s}/m	5 000
归一化时间t_{s}/s	50,51,⋯,58
归一化推力T_{s}/N	845 000
离散点个数K	60
信任域罚函数系数w_η	1
线性化误差罚函数系数w_κ	10^4
最大迭代次数$i_{\max}$	500
信任域误差限ε_η	10^{-4}
线性化误差限ε_κ	10^{-8}



(a) 不同末端时间的推力幅值结果



(b) 不同末端时间的燃料



(c) 不同末端时间的迭代次数

图 2　不同末端时间的仿真结果

Fig. 2　Simulation results for different terminal times

近于无穷，迭代次数也会趋近于无穷大。这是控制量与状态量解耦带来的副作用。如果状态量与控制量不解耦，固定时间t_f = 50 s的仿真结果如图3所示，控制力会出现明显的抖振。

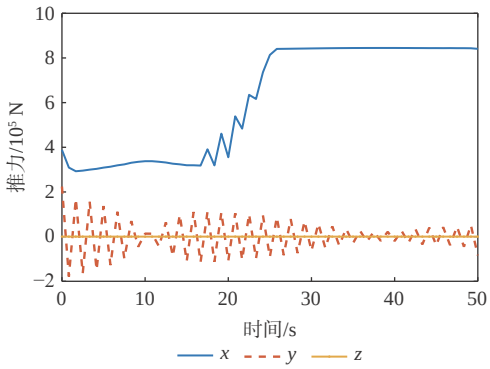


图3 本体系下的推力(控制量与状态量不解耦)

Fig. 3 Thrust under the body coordinate frame(control and state are coupled)

在非线性模型中采用四阶龙格库塔积分得到的着陆误差如表4所示，历次优化均采用归一化方法对数据进行无量纲处理，归一化参数如表3所示。随着时间归一化参数的逐渐减小(由58～50 s)，速度与角速度归一化参数会增大，因此，达到相同的误差限后实际误差会增大，从而导致位置和姿态误差也会略有增大。t_f = 50 s时误差明显增大是由于迭代计算未进入到误差限导致的。

表4 不同末端时间着陆误差

Table 4 Landing error for different terminal times

末端时间/s	位置误差/m	速度误差/(m·s^{-1})	姿态角误差/(°)	角速度误差/((°)·s^{-1})
50	256.94	19.80	4.06	0.046 2
51	95.81	7.40	1.52	0.024
52	55.59	4.29	0.87	0.014 8
53	45.27	3.46	0.68	0.008 5
54	41.46	3.11	0.59	0.005 6
55	38.25	2.83	0.52	0.003 2
56	35.00	2.56	0.46	0.001 4
57	34.03	2.46	0.43	0.000 87
58	34.39	2.43	0.42	0.000 769

4.2 自由时间优化

自由时间优化包括以时间替代变量为自变量和以高度为自变量2种。这2种优化问题均无需考虑对飞行时间的遍历，因此，将其在本节进行仿真对比。以时间替代变量为自变量的序列凸优化参数如表5所示，以高度为自变量优化问题序列凸优化参数如表6所示。优化迭代的初值参考式(26)和式(27)。

由于以高度为自变量序列凸优化仿真中不存

表5 序列凸优化参数（以时间替代变量为自变量）

Table 5 Successive convex optimization parameters (with time substitution variables as independent variables)

参数	数值
归一化距离l_s/m	5 000
归一化时间t_s/s	43
归一化推力T_s	845 000
离散点个数K	60
λ罚函数系数w_λ	10
信任域罚函数系数w_η	0.1
线性化误差罚函数系数w_κ	10^4
最大迭代次数i_{max}	500
信任域误差限ε_η	10^{-4}
线性化误差限ε_κ	10^{-8}

表6 序列凸优化参数（以高度为自变量）

Table 6 Successive convex optimization parameters(with height as independent variables)

参数	数值
归一化距离l_s/m	5 000
归一化时间t_s/s	12
归一化推力T_s	845 000
离散点个数K	60
信任域罚函数系数w_η	1
线性化误差罚函数系数w_κ	10^4
最大迭代次数i_{max}	500
信任域误差限ε_η	10^{-4}
线性化误差限ε_κ	10^{-8}

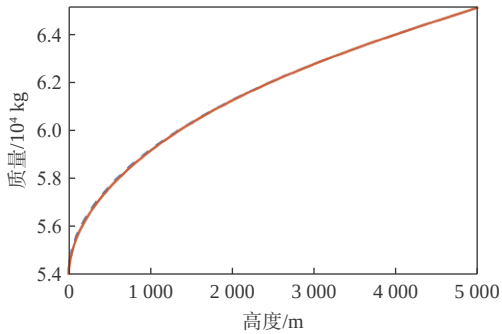
在时间变量，为了便于比较，以高度为横轴自变量对2种方法进行对比。如图4所示，2种方法得到的结果基本一致。但推力幅值均没有表现为开关控制，因此，优化结果均为次优解。出现这种情况的原因有可能与固定时间优化问题中推力幅值曲线未达到开关控制的原因相同，也有可能是由于信任域和松弛变量相关惩罚项的引入使得优化问题与原始问题不等价导致的。

将优化结果代入非线性模型中采用四阶龙格库塔积分进行数值计算，末端状态如表7所示，自由时间优化结果与固定时间t_f = 55 s的优化结果比较接近，在后续分析中选择t_f = 55 s的固定时间优化与自由时间优化方法进行对比。

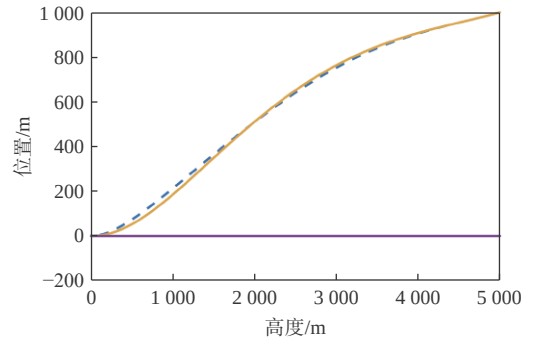
4.3 归一化参数对最优性影响分析

对于固定时间优化问题，时间归一化参数在较大范围内(20～80 s)对优化问题的最优性没有显著影响，但当时间归一化参数大于60 s时，迭代次数会增加。

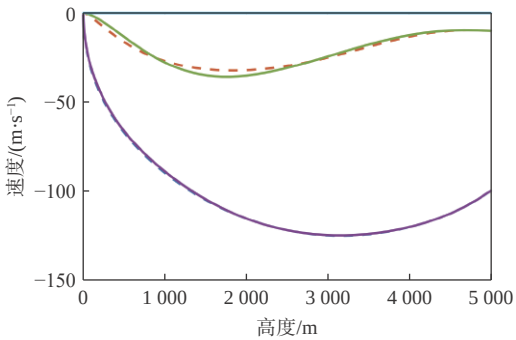
对于自由时间优化问题，2种问题对时间的归



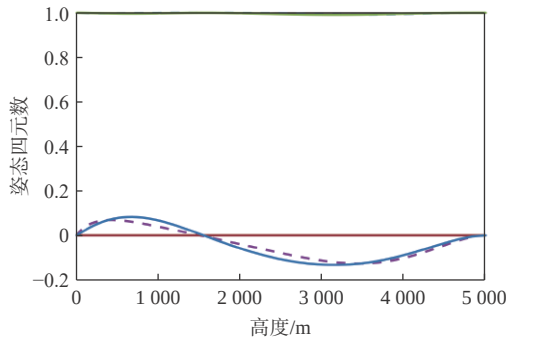
(a) 质量随高度的变化



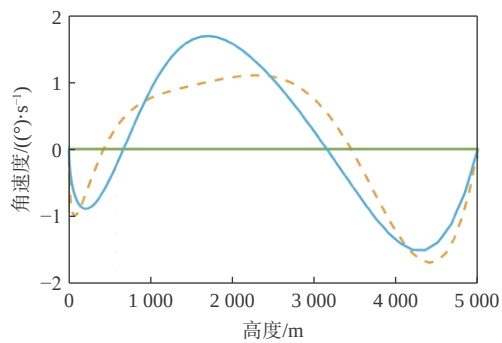
(b) 位置随高度的变化



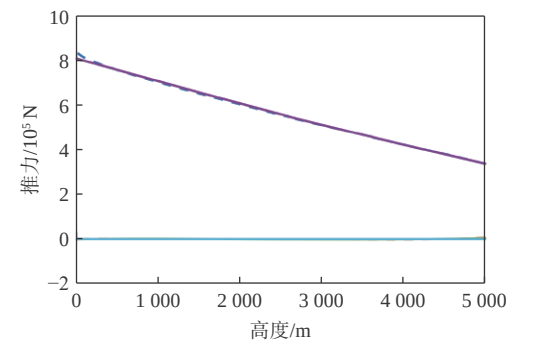
(c) 速度随高度的变化



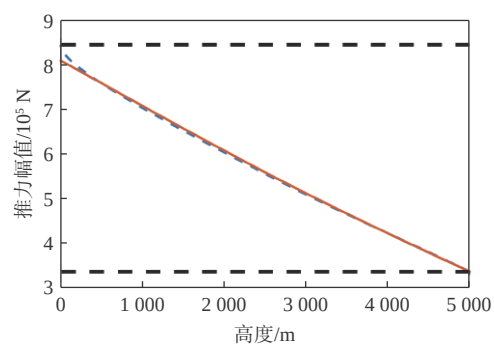
(d) 姿态四元数随高度的变化



(e) 角速度随高度的变化



(f) 本体系下推力随高度的变化



(g) 推力幅值随高度的变化

图 4　2 种自由时间优化对比

Fig. 4　Comparison of two free-final-time optimizations

表 7 自由时间优化末端着陆误差对比

Table 7 Comparison of landing errors for free-final-time optimization

自变量	燃料消耗/kg	位置误差/m	速度误差/(m·s^{-1})	姿态角误差/(°)	角速度误差/((°)·s^{-1})
时间替代变量	11 185.7	35.107 0	2.627 0	0.481 2	0.004 1
高度	11 008.0	39.004 3	5.978 2	0.470 74	0.115 41

一化参数都比较敏感，不同的时间归一化参数会获得不同的局部最优解，需要进行多组仿真选取最优的时间归一化参数。2 种问题改变时间归一化参数的仿真结果如图 5 和图 6 所示，t_s为时间归一化参数。从仿真结果可以看出，当以时间替代变量为自变量优化的时间归一化参数为 43 s 时，以高度为自变量优化的时间归一化参数为 12 s 时（进一步减小无法收敛至误差限内），燃料消耗最小。

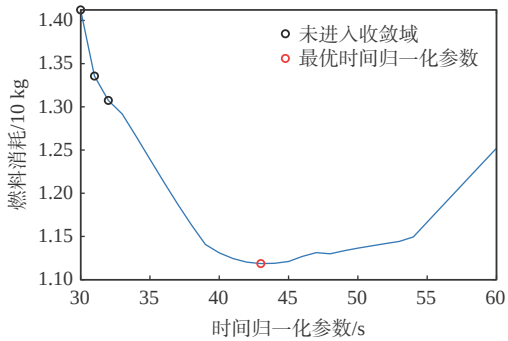


图 5 改变时间归一化参数的仿真结果
（以时间替代变量为自变量优化问题）

Fig. 5 Simulation results of changing the time normalization parameters (Optimisation problem with time substitution variables as independent variables)

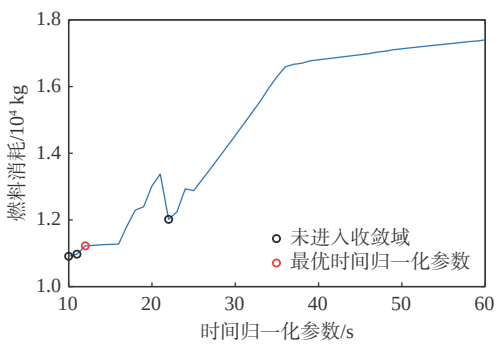


图 6 改变时间归一化参数的仿真结果
（以高度为自变量优化问题）

Fig. 6 The simulation results of changing the time normalization parameters (Optimisation problem with height as independent variables)

4.4 离散点数量对精度、计算时间的影响分析

改变离散点数量$K = 30, 40, \cdots, 100$，针对 3 种优化问题进行仿真，其余仿真参数同 4.1 节和 4.2 节。

从仿真结果可以看出，在末端误差方面，如图 7 和图 8 所示，当离散点数量较小时，3 种优化方法的位置误差和姿态误差都偏大；随着离散点数量的增多，误差也会逐渐减小。当$K < 60$时，随着离散点数量的增加，误差下降较快；当$K > 60$时，随着离散点数量的增加误差下降较缓。$K = 60$时 3 种方法的位置误差和姿态角误差接近。因此，选取$K = 60$作为离散点数量对 3 种方法进行比较是合理的。

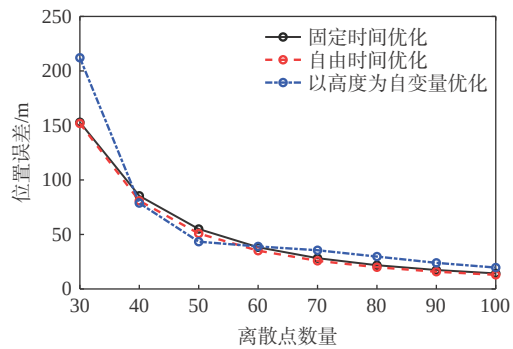


图 7 离散点数量对位置误差的影响

Fig. 7 Effect of the number of discrete points on the position error

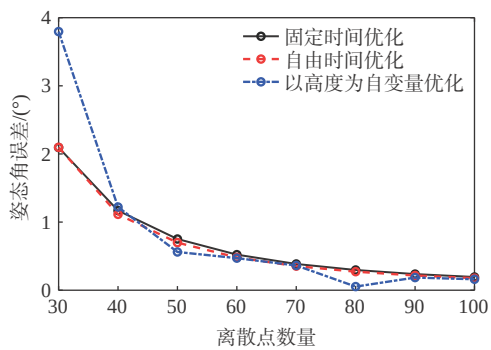


图 8 离散点数量对姿态角误差的影响

Fig. 8 Effect of the number of discrete points on the attitude angle error

在计算时间方面，图 9 和图 10 分别给出了不同离散点数量对计算时间和迭代次数的影响。需要注意的是，本节中的计算时间包含了 cvx 软件调用的时间，相较于实际的求解时间相比偏大，但仍可用于对比分析。对于固定时间优化，迭代次数基本稳定在 4 次，计算时间随着离散点数量的增加而增加；对于以时间替代变量为自变量优化问题，当离散点数量偏小时，迭代次数会增加从而导致计算时间的增加，后续迭代次数稳定后随着离散点数量的增加，计算时间明显增加；对于以高度为自变量优化问题，较少的离散点数量不利于方法的收敛，导

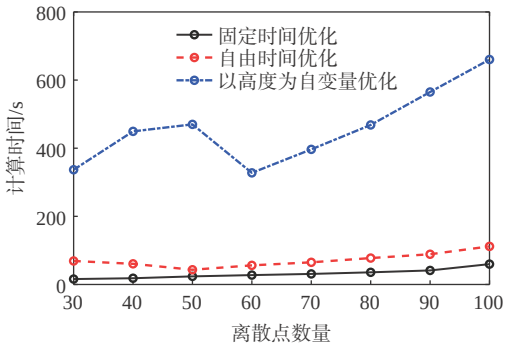


图 9　离散点数量对计算时间的影响

Fig. 9　Effect of the number of discrete points on the computation time

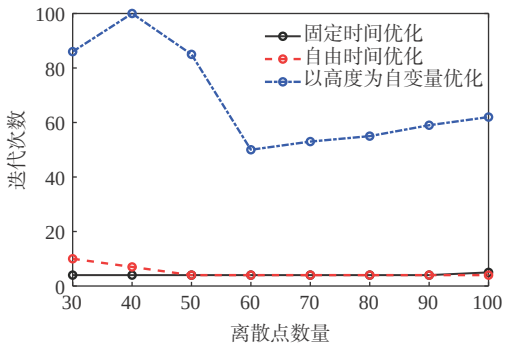


图 10　离散点数量对迭代次数的影响

Fig. 10　Effect of the number of discrete points on the number of iterations

致计算时间较高。从以上分析中可知离散点数量过小或过大均不利于迭代收敛，特别是对于自由时间优化问题。

5　3 种序列凸优化方法的总体对比

本节从收敛性、实时性和最优性 3 个方面对 3 种序列凸优化方法进行对比分析。

1）收敛性。在合适的归一化参数下，3 种方法均可以收敛到指定误差限。优化问题的收敛性受归一化参数影响，归一化参数不仅会影响解的优劣，甚至还会影响问题是否可以求解。因此，无论哪种方法均需要选取合适的归一化参数。首先，归一化参数的选取应保证优化的各状态量与控制量处于同一个数量级，避免数量级差距过大出现病态问题；其次，可以根据仿真结果对归一化参数进行调整，选取较优的归一化参数。本文中对于固定时间优化问题，以最大高度、飞行时间和最大推力归一化来确定长度、时间和质量归一化参数，对于自由时间优化问题，通过多次仿真确定较优的时间归一化参数。

2）实时性。为评估优化问题解算的实际成本，本节中计算时间统计的是 sdpt3 求解器给出的实际解算时间。表 8 给出了 3 种优化问题的计算时间及选取固定时间优化结果作为初值时自由时间优化问题的计算时间。以高度、飞行时间和时间替代变量为自变量优化问题的优化参数数量是递增的，其对应的单次迭代时间也在增加。在飞行时间没有接近最优时间时，固定时间优化收敛速度很快，约 4 次迭代即可收敛，但需要提前给出合适的飞行时间。以时间替代变量为自变量优化问题和以高度为自变量优化问题分别经 4 次迭代和 50 次迭代后收敛至误差限以内。2 种自由时间优化问题相比，两者可行域大小一致，以高度为自变量优化问题状态量更少，动力学模型更复杂，因此，表现的更难以收敛；而以时间替代变量为自变量优化问题相较于固定时间优化问题，动力学模型乘以了一个常优化变量，基本保留了原有的动力学性质，因此，具有和固定时间优化问题相似的迭代收敛次数。虽然以时间替代变量为自变量优化问题的迭代次数远小于以高度为自变量的优化问题，但在求解时间上 2 种方法的平均单次迭代时间分别为 7.64 s 和 0.44 s，总时间分别为 30.57 s 和 21.82 s。以高度为自变量优化问题的总计算时间更短。需要注意的是，序列凸优化算法中罚函数系数的选取也会影响算法的迭代次数和计算时间。

表 8　计算时间对比

Table 8　Comparision of computation time

优化方法	迭代次数	平均单次迭代时间/s	总时间/s
固定时间优化(55 s)	4	0.62	2.46
自由时间优化	4	7.64	30.57
以高度为自变量优化	50	0.44	21.82
自由时间优化(*)	1	7.80	2.46+7.80
以高度为自变量优化(*)	33	0.41	2.46+13.45

注：带*仿真选取55 s固定时间优化结果作为初值。

当选取固定时间优化结果作为初值时，以时间替代变量为自变量优化问题和以高度为自变量优化问题的迭代次数分别为 1 次和 33 次，均大幅减少。总计算时间分别为 10.26 s 和 15.91 s。

自由时间优化问题由于需要对飞行时间进行优化，无论采用以时间替代变量为自变量还是以高度为自变量，总求解时间均远高于固定时间优化。固定时间优化方法在计算时间上更有优势，更适合在线实时计算。

3）最优性。固定时间优化方法可以通过遍历飞行时间逐渐逼近到最优解，推力幅值逐渐逼近开关控制，最优飞行时间不大于 50 s；以时间替代变量为自变量的优化问题解得的飞行时间为 55.35 s；

以高度为自变量的优化问题获得了略优于以时间替代变量为自变量的优化问题的次优解，飞行时间为54.72 s。由于采用了控制量与状态量解耦的求解策略，自由时间优化得到的解均为次优解，相比最优解多消耗约10%燃料。

4）求解精度。求解精度主要受离散点数量影响。在相同的离散点数量下，3种方法的精度相近。离散点数量越多，位置和姿态误差越小，精度越高。但随着离散点个数的增加，精度的提升趋于缓慢，相应的计算时间也会增加，因此需要在精度与计算时间上进行权衡，根据任务需要选取合适的离散点个数。

6 结 论

本文比较了利用序列凸优化求解六自由度动力下降制导问题的3种参数化方法：分别以飞行时间、时间替代变量和高度为自变量建立参数优化问题。与已有研究相比，本文考虑了简单的气动力作用，采用以高度为自变量求解六自由度动力下降问题，并比较了3种方法在收敛性、实时性、最优性和求解精度上的差异。

1）仿真结果显示在合适的归一化参数下，3种方法均可以收敛。

2）以飞行时间为自变量的参数优化需约2.46 s即可完成，远低于自由时间优化问题的计算时间。

3）由于采用控制量与状态量解耦的策略以避免控制力抖振，以飞行时间为自变量的参数优化问题在飞行时间接近最优时间时，迭代次数会明显增加，而自由时间优化问题获得的是次优解，相较于最优解额外消耗约10%的燃料。

4）3种方法在相同离散点数目下的求解精度相近。

5）以高度为自变量的优化方法虽然减少了优化变量个数，在单次迭代计算时间上有明显优势，但总计算时间与以时间替代变量为自变量的优化方法相当，均高于固定时间优化，难以满足实时性要求。

利用序列凸优化求解动力下降制导问题，达到工程可用需求，仍存在以下问题有待进一步研究：如何更快地计算出最优飞行时间；如何实现较少迭代次数逼近开关控制最优解；如何提升迭代计算实时性，如1 s内完成。

参考文献（References）

[1] SONG Z Y, WANG C, THEIL S, et al. Survey of autonomous guidance methods for powered planetary landing[J]. Frontiers of Information Technology & Electronic Engineering, 2020, 21(5): 652-674.

[2] LIU X F, LU P, PAN B F. Survey of convex optimization for aerospace applications[J]. Astrodynamics, 2017, 1(1): 23-40.

[3] LU P, LIU X F. Autonomous trajectory planning for rendezvous and proximity operations by conic optimization[J]. Journal of Guidance, Control, and Dynamics, 2013, 36(2): 375-389.

[4] TANG G, JIANG F H, LI J F. Fuel-optimal low-thrust trajectory optimization using indirect method and successive convex programming[J]. IEEE Transactions on Aerospace and Electronic Systems, 2018, 54(4): 2053-2066.

[5] 邓雁鹏, 穆荣军, 彭娜, 等. 月面着陆动力下降段最优轨迹序列凸优化方法[J]. 宇航学报, 2022, 43(8): 1029-1039.
DENG Y P, MU R J, PENG N, et al. Sequential convex optimization method for lunar landing during powered descent phase[J]. Journal of Astronautics, 2022, 43(8): 1029-1039(in Chinese).

[6] 徐西宝, 白成超, 陈宇燊, 等. 月/火探测软着陆制导技术发展综述[J]. 宇航学报, 2020, 41(6): 719-729.
XU X B, BAI C C, CHEN Y S, et al. A survey of guidance technology for Moon/Mars soft landing[J]. Journal of Astronautics, 2020, 41(6): 719-729(in Chinese).

[7] HARRIS M W, AÇıKMEŞE B. Maximum divert for planetary landing using convex optimization[J]. Journal of Optimization Theory and Applications, 2014, 162(3): 975-995.

[8] BLACKMORE L, ACIKMESE B, SCHARF D P. Minimum-landing-error powered-descent guidance for Mars landing using convex optimization[J]. Journal of Guidance, Control, and Dynamics, 2010, 33(4): 1161-1171.

[9] ACIKMESE B, PLOEN S R. Convex programming approach to powered descent guidance for Mars landing[J]. Journal of Guidance, Control, and Dynamics, 2007, 30(5): 1353-1366.

[10] PINSON R, LU P. Rapid generation of optimal asteroid powered descent trajectories via convex optimization[C]//Proceedings of the AAS/AIAA Astrodynamics Specialist Conference. Reston: AIAA, 2015.

[11] PINSON R M, LU P. Trajectory design employing convex optimization for landing on irregularly shaped asteroids[J]. Journal of Guidance, Control, and Dynamics, 2018, 41(6): 1243-1256.

[12] LIU X F, SHEN Z J, LU P. Entry trajectory optimization by second-order cone programming[J]. Journal of Guidance, Control, and Dynamics, 2016, 39(2): 227-241.

[13] SCHARF D P, AÇIKMEŞE B, DUERI B, et al. Implementation and experimental demonstration of onboard powered-descent guidance[J]. Journal of Guidance, Control, and Dynamics, 2017, 40(2): 213-229.

[14] SZMUK M, ACIKMESE B, BERNING A W. Successive convexification for fuel-optimal powered landing with aerodynamic drag and non-convex constraints[C]//Proceedings of AIAA Guidance, Navigation, and Control Conference. Reston: AIAA, 2016.

[15] WANG Z B, GRANT M J. Constrained trajectory optimization for planetary entry via sequential convex programming[J]. Journal of Guidance, Control, and Dynamics, 2017, 40(10): 2603-2615.

[16] MAO Y Q, SZMUK M, AÇıKMEŞE B. Successive convexification of non-convex optimal control problems and its convergence properties[C]// Proceedings of the IEEE 55th Conference on Decision and Control. Piscataway: IEEE Press, 2016: 3636-3641.

[17] LIU X F, LU P. Solving nonconvex optimal control problems by

convex optimization[J]. Journal of Guidance, Control, and Dynamics, 2014, 37(3): 750-765.

[18] YANG R Q, LIU X F. Fuel-optimal powered descent guidance with free final-time and path constraints[J]. Acta Astronautica, 2020, 172: 70-81.

[19] SZMUK M, EREN U, ACIKMESE B. Successive convexification for Mars 6-DoF powered descent landing guidance[C]//Proceedings of the AIAA Guidance, Navigation, and Control Conference. Reston: AIAA, 2017.

[20] SZMUK M, ACIKMESE B. Successive convexification for 6-DoF Mars rocket powered landing with free-final-time [C]//Proceedings of the AIAA Guidance, Navigation, and Control Conference. Reston: AIAA, 2018.

[21] REYNOLDS T P, SZMUK M, MALYUTA D, et al. Dual quaternion-based powered descent guidance with state-triggered constraints[J]. Journal of Guidance, Control, and Dynamics, 2020, 43(9): 1584-1599.

[22] KAMATH A G, ELANGO P, MCEOWEN S, et al. Customized real-time first-order methods for onboard dual quaternion-based 6-DoF powered-descent guidance[C]//Proceedings of the AIAA SCITECH 2023 Forum. Reston: AIAA, 2023.

[23] CHEN Y S, YANG G W, WANG L, et al. A fast algorithm for onboard atmospheric powered descent guidance[J]. IEEE Transactions on Aerospace and Electronic Systems, 2023, 59(5): 6112-6123.

[24] MAO Y Q, DUERI D, SZMUK M, et al. Successive convexification of non-convex optimal control problems with state constraints[J]. IFAC-PapersOnLine, 2017, 50(1): 4063-4069.

[25] MALYUTA D, REYNOLDS T, SZMUK M, et al. Discretization performance and accuracy analysis for the rocket powered descent guidance problem[C]//Proceedings of the AIAA Scitech 2019 Forum. Reston: AIAA, 2019.

[26] MATTINGLEY J, BOYD S. CVXGEN: A code generator for embedded convex optimization[J]. Optimization and Engineering, 2012, 13(1): 1-27.

[27] TOH K C, TODD M J, TÜTÜNCÜ R H. On the implementation and usage of SDPT3–A Matlab software package for semidefinite-quadratic-linear programming, version 4.0[M]// Handbook on Semidefinite, Conic and Polynomial Optimization. Berlin: Springer, 2011: 715-754.

Three convexification-based methods for six-degree-of-freedom powered descent guidance

WANG Chi[1, 2], LIU Wei[1, 2, *], GAO Yang[1, 2]

(1. School of Aeronautics and Astronautics, University of Chinese Academy of Sciences, Beijing 100049, China;
2. Key Laboratory of Space Utilization, Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, China)

Abstract: A key technology for the safe landing of spacecraft in powered descent is to solve the six degrees of freedom(6-DoF) powered descent guidance problem in real time. This problem is a fuel-optimal problem with multiple constraints. In this paper, three optimization methods are established by three independent variables, namely flight time, time substitution variable, and height. Three onboard navigation techniques are created by transforming the trajectory optimization issue into a form that can be solved by successive convexification algorithms. The comparative analysis shows that all three methods can solve the 6-DoF powered descent problem. Although the flight time must be known beforehand, the optimization with flight time as an independent variable has the maximum computing efficiency and the lowest fuel use. The other two optimizations can optimize the flight time, but both are sub-optimal solutions, and the computation time is greatly increased. The accuracy of the three methods is similar under the same number of discrete points. If successive convexification algorithms are used as the online guidance algorithms for power descent, problems such as how to determine the optimal flight time, approach the optimal fuel solution and shorten the calculation time still need to be further studied.

Keywords: six degrees of freedom; powered descent; autonomous guidance; convex optimization; optimal flight time

Received: 2023-05-10; **Accepted**: 2023-07-21; **Published Online**: 2023-08-14 12:52

URL: link.cnki.net/urlid/11.2625.V.20230814.0912.001

Foundation item: Youth Innovation Promotion Association CAS (292020000040)

* **Corresponding author**. E-mail: liuwei@csu.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0237

运载火箭发射过程中的最大气动载荷预测方法

程胡华[1,*]，武帅[2]，姜祝辉[3]，张入财[4]

(1. 中国人民解放军63729部队，太原 030027；2. 中国人民解放军61741部队，北京 100094；
3. 北京应用气象研究所，北京 100029；4. 中国人民解放军31308部队，成都 610031)

摘 要：火箭发射前3 h的最大气动载荷是决定能否按计划发射的重要指标之一，该值与发射时刻的最大气动载荷差异特征研究极少。通过对2014年12月—2019年12月的时间间隔3.5 h内最大气动载荷差异特征进行分析，发现6.29%样本数的最大气动载荷绝对差超过了500 Pa·rad，主要是由于3.5 h内高空风出现异常增大、异常减小导致的，若未提前发现该变化特征，则可能会影响火箭的飞行安全。针对绝对差超过500 Pa·rad的样本，通过建模订正可得到更接近发射后0.5 h最大气动载荷值，绝对差由713.08 Pa·rad减小到398.22 Pa·rad，相对误差由29.02%减小到16.32%，相关系数由0.27增大到0.71，表明该建模方法有一定的改进作用，有利于提高火箭安全飞行的保障能力。

关 键 词：运载火箭发射；最大气动载荷偏差；最大气动载荷预测；建模精度；安全飞行
中图分类号：V411.8
文献标志码：A **文章编号**：1001-5965(2025)04-1304-09

运载火箭能否按计划发射，除考虑火箭本身是否安全可靠外，还需考虑大气环境对运载火箭飞行的安全影响，特别是高空风引起的最大气动载荷$q\alpha_{\max}$(q为动压头，α为总攻角)是否较小，若$q\alpha_{\max}$过大，则会导致箭体弯曲折断，造成飞行失败。为保障运载火箭飞行成功，一方面，针对运载火箭的飞行控制、动力等方面开展研究[1-4]，及进行数值实验分析[5-7]，李东等[1]提出一种基于多体动力学虚拟样机建模与仿真的方法，有效解决运载火箭姿态动力学模型地面难以验证的困难；另一方面，对大气环境产生的气动载荷特征开展了一系列研究[8-15]，程镇煌[8]指出风载设计中增强火箭原型结构的靶场风载试验和高空风载试验的重要性。程胡华等[14]发现若预报风速偏差较大，火箭最大气动载荷偏差值甚至超过500 Pa·rad，将对火箭安全飞行带来潜在风险；相关研究成果为持续提高火箭飞行成功率提高了有力支撑。

决定火箭是否按计划发射，需提前几天给出$q\alpha_{\max}$预测值，若超过火箭设计阈值，则考虑推迟发射。利用数值天气预报模式提供的高空预报风计算火箭$q\alpha_{\max}$预测值，是目前常用的比较可靠方法[14,16]，该方法可以提供未来第1～11天内$q\alpha_{\max}$预测值，且预测值误差随预报日数延长呈现较明显增大趋势；但当火箭进入临近(0～3 h)发射阶段时，此时基于探空气球高空实况风计算得到的$q\alpha_{\max}$实况值，被认可为最可靠的参考数据。在火箭发射保障过程中，较少考虑$q\alpha_{\max}$在3 h内会出现急剧变化，但实践表明，虽然$q\alpha_{\max}$在短时间内的变化通常较小，但偶尔会出现急剧变化。例如，若高空急流位于发射场上空或附近时，高空急流的南北摆动、下沉或上升等运动，以及对其它天气系统的动力作用[17-18]，将会导致高空风在几小时内出现急剧变化[19]，程胡华等[19]利用3.5 h内高空风数据集进行研究，发现3.5 h内高空风最大风速偏差范围为−24.00～26.00 m/s，最大风速绝对差超过10 m/s的样本数达到31.70%，此时将导致3.5 h内的$q\alpha_{\max}$出

收稿日期：2023-05-12；**录用日期**：2023-06-30；**网络出版时间**：2023-08-30 16:48
网络出版地址：link.cnki.net/urlid/11.2625.V.20230830.1443.002
*通信作者. E-mail：chenghongxi2012@qq.com

引用格式：程胡华，武帅，姜祝辉，等. 运载火箭发射过程中的最大气动载荷预测方法[J]. 北京航空航天大学学报，2025，51(4)：1304-1312. CHENG H H，WU S，JIANG Z H，et al. Maximum aerodynamic load prediction method during launching of vehicle launch[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1304-1312 (in Chinese).

现相应变化。

因此, 考虑到$q\alpha_{\max}$在短时间内可能出现急剧变化特征, 若以火箭发射前 3 h 的$q\alpha_{\max}$是否超过阈值, 判断发射是否按计划进行, 存在一定的安全风险。但截止目前, 针对火箭发射前 3 h 的$q\alpha_{\max}$与发射时刻$q\alpha_{\max}$之间的差异特征, 鲜见相关研究。由于只有间隔 3.5 h 的$q\alpha_{\max}$资料集, 本文对 3.5 h 内$q\alpha_{\max}$差异特征进行分析, 并选取$q\alpha_{\max}$绝对差值超过 500 Pa·rad 的样本作为建模订正对象, 分析原因并提出订正方法, 最后, 对订正效果进行检验。

1　资料及最大气动载荷计算

1.1　资　料

间隔 3.5 h 的$q\alpha_{\max}$数据集(共 302 组), 是基于 2014 年 12 月—2019 年 12 月间隔 3.5 h 高空风计算得到, 该高空风资料在高度 0～10 km、10～20 km 范围内的间隔分别为 0.20 km、0.50 km。建模得到火箭发射后 0.5 h 的$q\alpha_{\max}$数据集, 是基于高空订正风计算得到, 该高空订正风是通过利用美国国家环境预报中心(National Centers for Environmental Prediction, NCEP)/全球预报系统(global forecast system, GFS)预报场资料, 作为天气研究与预报(weather research and forecast, WRF)模式的初始场和侧边界条件, 通过时间积分得到未来 12 h 内的高空风资料, 并结合火箭发射前 3 h 高空风资料, 建模得到高空风订正值。

1.2　最大气动载荷计算

计算最大气动载荷时, 除需要火箭本身参数外, 还需要高空风、大气密度等气象要素, 其中, 高空风通过探空气球、WRF 模式可以直接获取, 大气密度需通过温度、相对湿度和气压进行计算, 其具体计算步骤见文献 [13] 。气动载荷的$q\alpha$值表达式为

$$q\alpha = q\sqrt{(\alpha')^2+(\beta')^2} \tag{1}$$

式中: q为动压头; α为总攻角; α'和β'分别为攻角和侧滑角。$q\alpha$值反映了大气环境对运载火箭飞行作用下产生的气动载荷, 每次火箭飞行过程中产生的最大$q\alpha$值是否超过火箭设计阈值, 是衡量火箭能否安全飞行的重要参数, 关于$q\alpha$的推导过程, 请参见文献 [11]。

利用偏差和相对偏差反映 3.5 h 内高空风、大气密度、最大气动载荷差异特征, 通过偏差、绝对差、相对偏差和相对误差对建模效果进行检验, 其计算公式如下:

$$D = \left[\sum_{i=1}^{n}(x_i - y_i)\right] \Big/ n \tag{2}$$

$$A = \left[\sum_{i=1}^{n}|x_i - y_i|\right] \Big/ n \tag{3}$$

$$B = 100\sum_{i=1}^{n}\left(\frac{x_i - y_i}{y_i}\right) \tag{4}$$

$$R = 100\sum_{i=1}^{n}\left(\frac{|x_i - y_i|}{y_i}\right) \tag{5}$$

式中: D为偏差; A为绝对差; B为相对偏差; R为相对误差; n为样本数, $i = 1,2,\cdots,n$; x为评估量(或比较量); y为基准量。

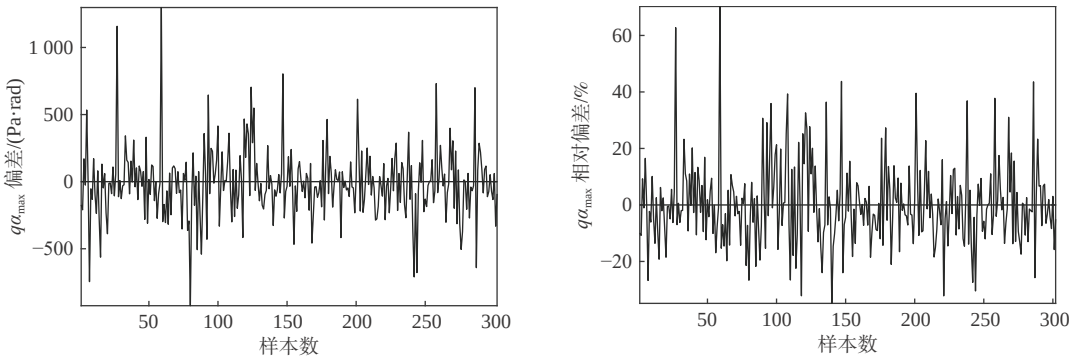
2　3.5 h 内 $q\alpha_{\max}$ 差异特征

火箭发射前 3 h、发射后 0.5 h 的$q\alpha_{\max}$分别记为$q\alpha_{\max}^{b3h}$、$q\alpha_{\max}^{a0.5h}$, 以$q\alpha_{\max}^{b3h}$为基准, 得到 3.5 h 内的$q\alpha_{\max}$偏差、相对偏差分布特征如图 1 所示, 图中黑横线对应的值为 0。$q\alpha_{\max}^{a0.5h}$与$q\alpha_{\max}^{b3h}$之间的偏差存在正值、负值交替变化特征 (见图 1(a)), 其中, 有 135 个样本的$q\alpha_{\max}$偏差为正值 (占总数的 44.70%), 167 个样本的$q\alpha_{\max}$偏差为负值 (占总数的 55.30%), $q\alpha_{\max}$偏差值范围为−924.43～1 297.62 Pa·rad, 平均值为−13.76 Pa·rad, $q\alpha_{\max}$偏差值超过 500 Pa·rad、200 Pa·rad 的样本数分别有 10 个、39 个, 分别占总数的 3.31%、12.91%, $q\alpha_{\max}$偏差值超过−500 Pa·rad、−200 Pa·rad 的样本数分别有 9 个、58 个, 分别占总数的 2.98%、19.21%, 即$q\alpha_{\max}$绝对差值超过 500 Pa·rad、200 Pa·rad 的样本数, 分别占总数的 6.29%、32.12%。$q\alpha_{\max}$相对偏差同样存在正值、负值交替变化特征 (见图 1(b)), 图中的$q\alpha_{\max}$相对偏差值范围为−34.84%～70.26%, 平均值为 0.16%, 其中, $q\alpha_{\max}$相对偏差超过 20%、50% 的样本数分别有 27 个、2 个, 分别占总数的 8.94%、0.66%, $q\alpha_{\max}$相对偏差超过−20%、−50% 的样本数分别有 15 个、0 个, 分别占总数的 4.97%、0.00%。

通过上述分析可知, 本文的 302 组样本中, 大多数样本在 3.5 h 内的$q\alpha_{\max}$差异较小, 偏差值位于−200～200 Pa·rad 范围内占总数的 67.88%, 但分别有 10 个和 9 个样本的$q\alpha_{\max}$偏差值超过了 500 Pa·rad、−500 Pa·rad, 分别占总数的 3.31%、2.98%, 即 6.29% 样本数的$q\alpha_{\max}$绝对差值超过了 500 Pa·rad, 表明这是小概率事件, 但若未能提前发现该变化特征, 将对火箭能否按计划发射的决策、能否安全飞行产生重大隐患。因此, 针对 3.5 h 内$q\alpha_{\max}$偏差值异常偏大的样本开展研究、分析原因并进行建模订正, 具有实际意义。

3　原因分析

在计算$q\alpha_{\max}$的程序中, 大气密度、高空风均为



(a) $q\alpha_{max}$ 偏差变化特征　(b) $q\alpha_{max}$ 相对偏差变化特征

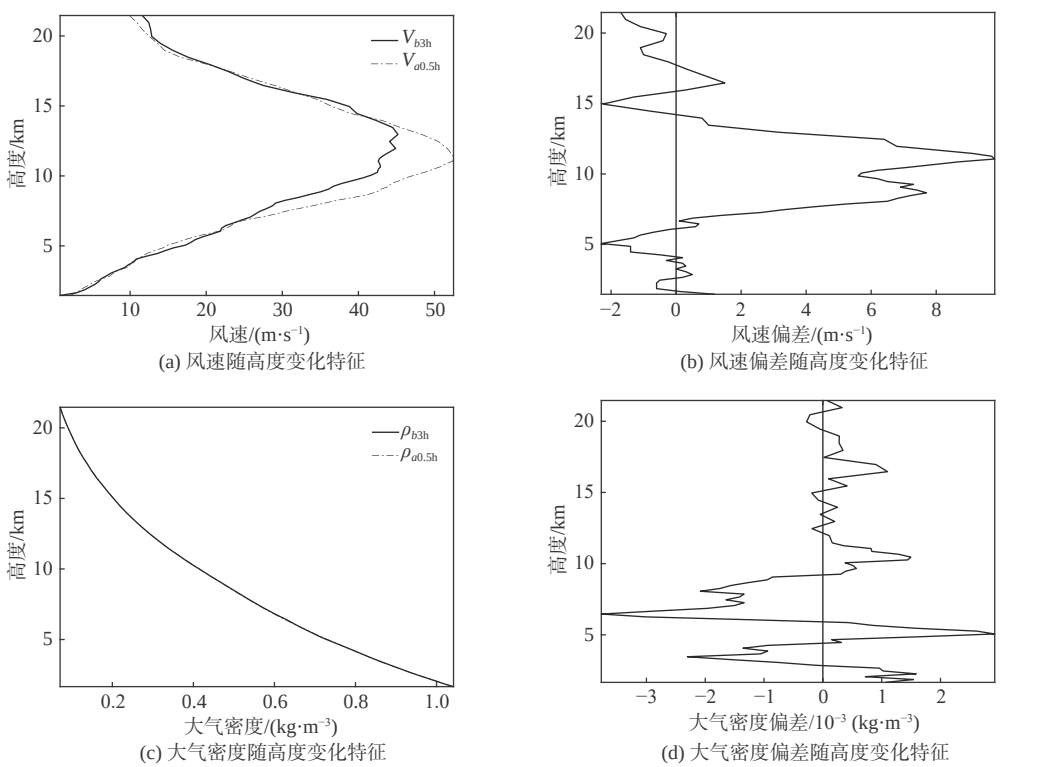
图 1　在 3.5 h 内, $q\alpha_{max}$ 偏差、相对偏差的变化特征

Fig. 1　The variation characteristics of $q\alpha_{max}$ deviation and $q\alpha_{max}$ relative deviation within 3.5 h

所需的大气环境参数。在 3.5 h 内, 火箭本身的状态参数基本不变, 但天气系统会引起大气密度和高空风变化, 导致$q\alpha_{max}$出现相应变化。考虑到研究区域的海拔高度为 1.5 km 左右, 以及火箭$q\alpha_{max}$均出现在飞行高度 20 km 以内, 下面针对上述 19 个样本 ($q\alpha_{max}$ 绝对差超过 500 Pa·rad) 在高度 1.5～21.5 km 范围内, 3.5 h 内的高空风、大气密度差异特征进行分析。

3.1　$q\alpha_{max}$ 偏差值超过 500 Pa·rad

对于$q\alpha_{max}$偏差值超过 500 Pa·rad 的 10 个样本, 火箭发射前 3 h 与发射后 0.5 h 的高空风 (分别记为 V_{b3h}、$V_{a0.5h}$)、大气密度 (分别记为ρ_{b3h}、$\rho_{a0.5h}$) 及偏差随高度变化特征如图 2 所示, 图中黑竖线对应的值为 0.00, 相应统计结果如表 1 所示。V_{b3h}与$V_{a0.5h}$随高度变化趋势基本一致, 均随高度增加表现为增大、减小的变化趋势, 但在某些高度层的数值存在明显差异 (见图 2(a)), 结合表 1 结果可知, 3.5 h 内高空风速最大值明显增大、对应高度降低, 且平均风速增大。$V_{a0.5h}$与V_{b3h}之间的平均偏差随高度变化如图 2(b) 所示, 高度 7.90～12.50 km 范围内的风速偏差较大 (超过 5 m/s), 偏差值范围为 5.20(高度 7.90 km)～9.80 m/s(高度 11.10 km), 平均值为 7.18 m/s。与高空风变化特征不同, ρ_{b3h}与$\rho_{a0.5h}$随高度增加均呈现指数减小变化特征、且基本重合 (见图 2(c)), 火箭发射前 3 h、发射后 0.5 h 的大气密度值范围及平均值基本相同 (见表 1), $\rho_{a0.5h}$与ρ_{b3h}之间的偏差随高度增加表现为正值、负值交替变化特征 (见



(a) 风速随高度变化特征　(b) 风速偏差随高度变化特征

(c) 大气密度随高度变化特征　(d) 大气密度偏差随高度变化特征

图 2　$q\alpha_{max}$偏差值超过 500 Pa·rad 时, 火箭发射前 3 h、发射后 0.5 h 的高空风速及其偏差、大气密度及其偏差随高度变化特征

Fig. 2　When the $q\alpha_{max}$ deviation value exceeds 500 Pa·rad, the upper wind speed and its deviation, the atmospheric density and its deviation change with altitude 3 h before launch and 0.5 h after launch

图 2(d)), 偏差值范围及平均值非常小 (见表 1), 与其原始值相比可忽略不计。根据文献 [13] 和文献 [14] 分析结论可知, 此次 3.5 h 内 $q\alpha_{\max}$ 偏差值超过 500 Pa·rad 的样本, 大气密度差异的贡献很小, 主要是由于高空风速最大值明显增大、对应高度降低, 且平均风速增大, 同时在高度 7.90～11.10 km 内风速快速增大、且高度较低共同导致的。

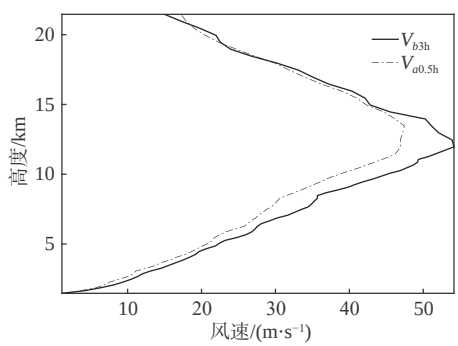
3.2　$q\alpha_{\max}$ 偏差值超过−500 Pa·rad

对于 $q\alpha_{\max}$ 偏差值超过−500 Pa·rad 的 9 个样本, $V_{a0.5h}$ 与 V_{b3h}、$\rho_{a0.5h}$ 与 ρ_{b3h} 及偏差随高度变化特征见图 3 所示, 相应统计结果如表 1 所示。V_{b3h} 与 $V_{a0.5h}$ 随高度变化趋势基本一致, 均随高度增加表现为增大、减小的变化趋势, 但在某些高度层的数值存在较大差异 (见图 3(a)), 从表 1 可看出, 3.5 h 内高空风速最大值存在明显减小、对应高度明显升高, 且平均风速减小; $V_{a0.5h}$ 与 V_{b3h} 之间的平均偏差随高度变化如图 3(b) 所示, 在高度 7.70～12.50 km 范围内的风速偏差较大 (大多数高度层超过−5.00 m/s), 偏差值范围为 −7.21(高度 12.00 km)～−4.22 m/s(高度 8.50 km), 平均值为−5.67 m/s。与高空风变化特征不同, 3.5 h 内的大气密度差异很小, ρ_{b3h} 与 $\rho_{a0.5h}$ 随高度变化基本重合 (见图 3(c)), 大气密度值范围及平均值基本相同 (见表 1), $\rho_{a0.5h}$ 与 ρ_{b3h} 之间的偏差随高度增加同样表现为正值、负值交替变化特征 (见

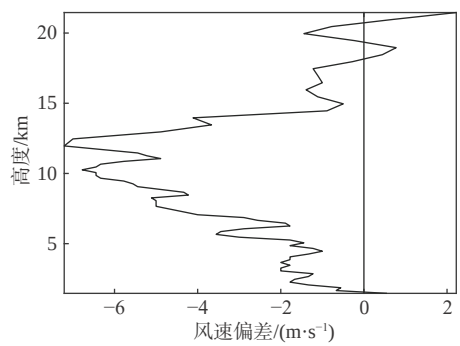
表 1　$q\alpha_{\max}$ 偏差值超过 500 Pa·rad、−500 Pa·rad 时, 火箭发射前 3 h、发射后 0.5 h 的大气环境差异特征

Table 1　When the $q\alpha_{\max}$ deviation value exceeds 500 Pa·rad and −500 Pa·rad, the atmospheric environment difference characteristics of 3 h before rocket launch and 0.5 h after launch

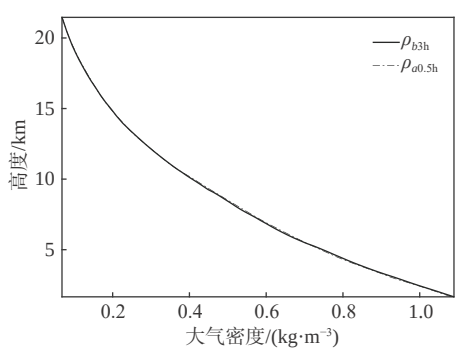
$q\alpha_{\max}$ 偏差值/(Pa·rad)	V_{b3h}			$V_{a0.5h}$		
	数值范围/(m·s^{-1})	对应高度/km	平均值/(m·s^{-1})	数值范围/(m·s^{-1})	对应高度/km	平均值/(m·s^{-1})
>500	0.80～45.19	1.50～13.00	24.59	2.00～52.50	1.50～11.30	26.68
<−500	1.11～54.15	1.50～12.00	30.55	1.67～47.44	1.50～13.50	27.65
$q\alpha_{\max}$ 偏差值/(Pa·rad)	风速偏差			ρ_{b3h}		
	数值范围/(m·s^{-1})	对应高度/km	平均值/(m·s^{-1})	数值范围/(kg·m^{-3})	对应高度/km	平均值/(kg·m^{-3})
>500	−2.30～9.80	5.10～11.10	2.09	0.07～1.04	21.50～1.50	0.51
<−500	−7.21～2.22	12.00～21.50	−2.90	0.07～1.09	21.50～1.50	0.51
$q\alpha_{\max}$ 偏差值/(Pa·rad)	$\rho_{a0.5h}$			大气密度偏差		
	数值范围/(kg·m^{-3})	对应高度/km	平均值/(kg·m^{-3})	数值范围/(kg·m^{-3})	对应高度/km	平均值/(kg·m^{-3})
>500	0.07～1.04	21.50～1.50	0.51	−0.003 8～0.002 9	6.50～5.10	-1.2×10^{-4}
<−500	0.07～1.08	21.50～1.50	0.51	−0.009 2～0.007 1	3.10～7.90	2.9×10^{-4}



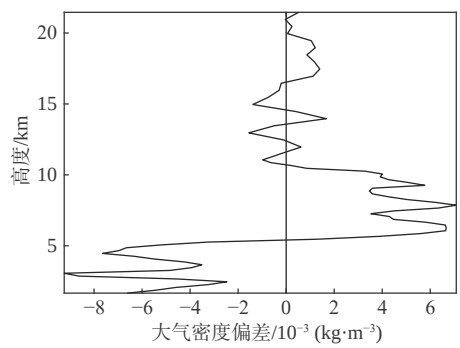
(a) 风速随高度变化特征



(b) 风速偏差随高度变化特征



(c) 大气密度随高度变化特征



(d) 大气密度偏差随高度变化特征

图 3　$q\alpha_{\max}$ 偏差值超过−500 Pa·rad 时, 火箭发射前 3 h 与发射后 0.5 h 的高空风速及其偏差、大气密度及其偏差随高度变化特征

Fig. 3　When the $q\alpha_{\max}$ deviation value exceeds −500 Pa·rad, the upper wind speed and its deviation, the atmospheric density and its deviation change with altitude 3 h before launch and 0.5 h after launch

图 3(d)), 偏差值范围数值及平均值非常小 (见表 1), 与其原始值相比可忽略不计。根据文献 [13] 和文献 [14] 分析结论可知, 此次 3.5 h 内 $q\alpha_{\max}$ 偏差值超过−500 Pa·rad 的样本, 同样表现为大气密度差异的贡献很小, 主要是由于高空风速最大值存在明显减小、对应高度明显升高、且平均风速减小, 同时在高度 7.70～12.50 km内风速出现快速减小、且高度较低共同导致的。

4 建模及分析

通过上述分析可知, 3.5 h 内 $q\alpha_{\max}$ 偏差较大的原因是高空风异常增大或减小导致的, 而大气密度差异可忽略不计。利用 WRF 模式可提前判断高空风变化差异特征, 该模式是目前世界上最先进的中尺度数值天气模式之一, 已被广泛地应用于强对流、高空急流、地面风等天气预报业务及机理研究, 并取得重要成果[18-21]。本文利用 WRF 模式获取 3.5 h 内的高空风变化差异, 经过前期验证, 在 WRF 模式中设计适合本地区的参数化方案 (见表 2), 并设置模式输出产品的时间分辨率为 15 min、水平分辨率为 9 km, 高度 1 000～50 hPa 分成 48 层, 利用 WRF 模式的时间积分模块可得到未来 12 h 内 (时间分辨率为 15 min) 的高空风数据集, 选取所关心的 3.5 h 内高空风差异, 以符号 $\Delta V_{f3.5\mathrm{h}}$ 表示。利用 $V_{b3\mathrm{h}}$ 和 $\Delta V_{f3.5\mathrm{h}}$, 即得到的高空订正风 $V_{f0.5\mathrm{h}}$[19], 计算表达式为

表 2 WRF 模式中的参数化方案设计

Table 2 Parametric scheme design in WRF model

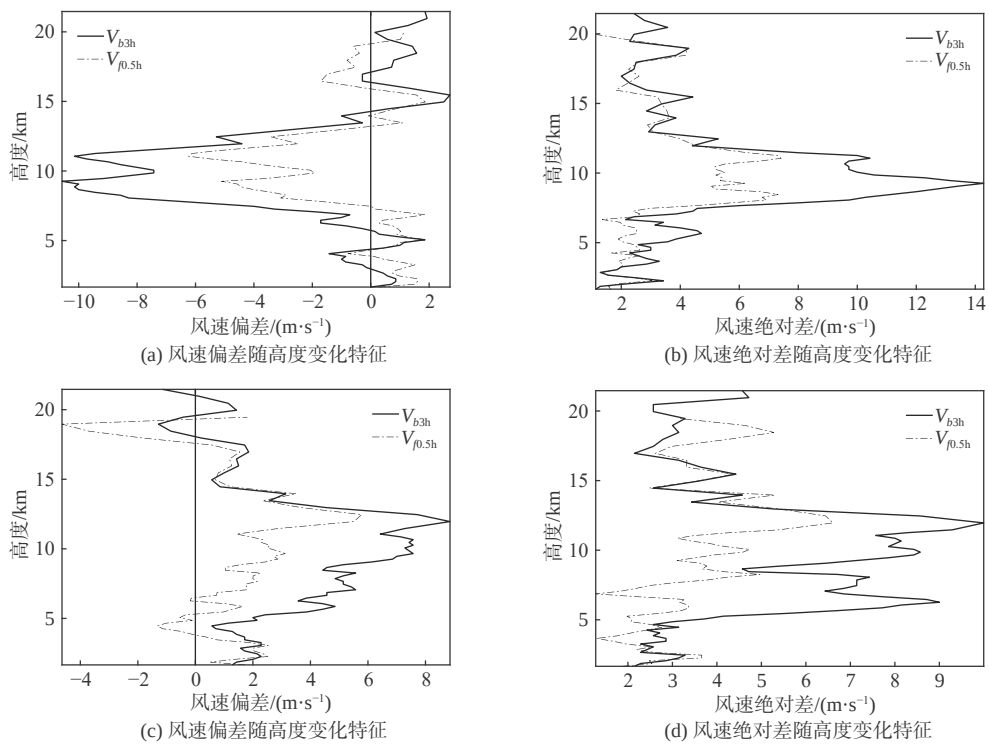
参数化方案	方案设计
微物理过程方案	WDM 5-class scheme
长波辐射方案	rrtm scheme
短波辐射方案	Goddard short wave
近地面层方案	Monin-Obukhov(Janjic) scheme
陆面过程方案	unified Noah land-surface model
边界层方案	Mellor-Yamada-Janjic TKE scheme
积云参数化方案	Betts-Miller-Janjic scheme

$$V_{f0.5\mathrm{h}} = V_{b3\mathrm{h}} + \Delta V_{f3.5\mathrm{h}} \tag{6}$$

式中: $V_{b3\mathrm{h}}$ 为火箭发射前 3 h 高空风; $\Delta V_{f3.5\mathrm{h}}$ 为 WRF 模式得到的 3.5 h 内高空风差异; $V_{f0.5\mathrm{h}}$ 为建模得到的火箭发射后 0.5 h 高空订正风, 该值的精度取决于 $\Delta V_{f3.5\mathrm{h}}$。若 WRF 模式的参数化方案设计合理、且同化更多各类实况资料, 则有利于得到更精准的 $\Delta V_{f3.5\mathrm{h}}$, 是本文建模的重点和难点, 需经过长期的调试验证。针对 $q\alpha_{\max}$ 绝对差值超过 500 Pa·rad 的 19 个样本, 以 $V_{a0.5\mathrm{h}}$、$q\alpha_{\max}^{a0.5\mathrm{h}}$ 为基准, 对建模得到的火箭发射后 0.5 h 高空订正风 $V_{f0.5\mathrm{h}}$ 及其对应的火箭发射后 0.5 h 最大气动载荷 $q\alpha_{\max}^{f0.5\mathrm{h}}$ 精度特征进行分析。

4.1 $V_{a0.5\mathrm{h}}$ 精度分析

$q\alpha_{\max}$ 偏差值超过 500 Pa·rad、−500 Pa·rad 时, $V_{b3\mathrm{h}}$、$V_{f0.5\mathrm{h}}$ 与 $V_{a0.5\mathrm{h}}$ 之间偏差和绝对差随高度变化特征如图 4 所示, 相应的统计结果如表 3 所示, 与 $V_{b3\mathrm{h}}$ 相比, 总体上表现为 $V_{f0.5\mathrm{h}}$ 更接近 $V_{a0.5\mathrm{h}}$。当 $q\alpha_{\max}$ 偏差值超过 500 Pa·rad 时, $V_{b3\mathrm{h}}$、$V_{f0.5\mathrm{h}}$ 的偏差及绝对差随



(a) 风速偏差随高度变化特征 (b) 风速绝对差随高度变化特征

(c) 风速偏差随高度变化特征 (d) 风速绝对差随高度变化特征

图 4 $q\alpha_{\max}$ 偏差值超过 500 Pa·rad、超过−500 Pa·rad 时, $V_{b3\mathrm{h}}$、$V_{f0.5\mathrm{h}}$, 与 $V_{a0.5\mathrm{h}}$ 之间的偏差和绝对差随高度变化特征

Fig. 4 When the $q\alpha_{\max}$ deviation exceeds 500 Pa·rad and exceeds −500 Pa·rad, the characteristics of deviation and absolute difference change with height between $V_{b3\mathrm{h}}$, $V_{f0.5\mathrm{h}}$ and $V_{a0.5\mathrm{h}}$

表 3　$q\alpha_{\max}$ 偏差值超过 500 Pa·rad、−500 Pa·rad 时，V_{b3h}、$V_{f0.5h}$，与$V_{a0.5h}$之间的偏差和绝对差特征

Table 3　When the $q\alpha_{\max}$ deviation value exceeds 500 Pa·rad, −500 Pa·rad, the deviation between V_{b3h}, $V_{f0.5h}$, and $V_{a0.5h}$ and the absolute difference characteristics

$q\alpha_{\max}$偏差值/(Pa·rad)	V_{b3h}偏差/(m·s^{-1})		$V_{f0.5h}$偏差/(m·s^{-1})		V_{b3h}偏差/(m·s^{-1})		$V_{f0.5h}$偏差/(m·s^{-1})	
	数值范围	平均值	数值范围	平均值	数值范围	平均值	数值范围	平均值
>500	−10.57～2.71	−2.61	−6.24～1.87	−0.83	1.14～14.29	5.19	1.13～7.42	3.52
<−500	−1.28～8.85	3.50	−4.64～5.78	1.34	2.14～10.00	5.17	1.28～6.61	3.33

注: 所有偏差均为与$V_{a\,0.5h}$相比较。

高度变化特征如图 4(a) 和图 4(b) 所示, 其偏差、绝对差范围及相应平均值均出现较明显减小 (见表 3), 且在高度 7.70～10.90 km(见图 4(a)), 偏差值范围由−5.43(高度 7.70 km)～−10.57 m/s(高度 9.30 km) 减小到−0.97(高度 7.70 km)～−5.46 m/s(高度 10.90 km), 平均偏差值由−8.64 m/s 减小到−3.37 m/s; 在高度 8.70～10.70 km(见图 4(b)), 绝对差值范围由 9.57(高度 10.70 km)～14.29 m/s(高度 9.30 km) 减小到 5.06 (高度 9.10 km)～6.72 m/s(高度 8.70 km), 平均值由 11.58 m/s 减小到 5.51 m/s。当$q\alpha_{\max}$偏差值超过 −500 Pa·rad 时, V_{b3h}、$V_{f0.5h}$的偏差及绝对差随高度变化特征如图 4(c) 和图 4(d) 所示, 其偏差值、绝对差值范围及相应平均值均同样出现较明显减小 (见表 3), 且在高度 6.50～12.00 km(见图 4(c)), 偏差值范围由 3.86(高度 6.50 km)～8.85 m/s(高度 12.00 km) 减小到−0.14(高度 6.50 km)～5.51 m/s(高度 12.00 km), 平均偏差值由 6.13 m/s 减小到 2.09 m/s; 在高度 5.90～12.00 km(见图 4(d)), 绝对差值范围由 4.57(高度 8.70 km)～10.00 m/s(高度 12.00 km) 减小到 1.28(高度 6.90 km)～6.61 m/s(高度 12.00 km), 平均值由 7.52 m/s 减小到 3.62 m/s。研究结果表明, 当$q\alpha_{\max}$绝对差超过 500 Pa·rad 时, 与V_{b3h}相比, 利用模型得到的$V_{f0.5h}$更接近$V_{a0.5h}$。

4.2　$q\alpha_{\max}^{f0.5h}$精度分析

利用$V_{f0.5h}$得到的$q\alpha_{\max}$, 记为$q\alpha_{\max}^{f0.5h}$。以$q\alpha_{\max}^{a0.5h}$为基准, 分别对$q\alpha_{\max}^{b3h}$、$q\alpha_{\max}^{f0.5h}$的精度特征进行对比分析, 若相比$q\alpha_{\max}^{b3h}$, $q\alpha_{\max}^{f0.5h}$更接近$q\alpha_{\max}^{a0.5h}$, 则表明建模订正具有较好的改进效果。$q\alpha_{\max}^{b3h}$、$q\alpha_{\max}^{f0.5h}$与$q\alpha_{\max}^{a0.5h}$之间的偏差、绝对差、相对偏差及相对误差变化特征如图 5 所示, 其中, 横坐标为样本数 (共 19 个), 前 10 个为$q\alpha_{\max}$偏差值超过 500 Pa·rad 的样本、后 9 个为超过−500 Pa·rad 的样本, 表 4 给出了$q\alpha_{\max}^{b3h}$、

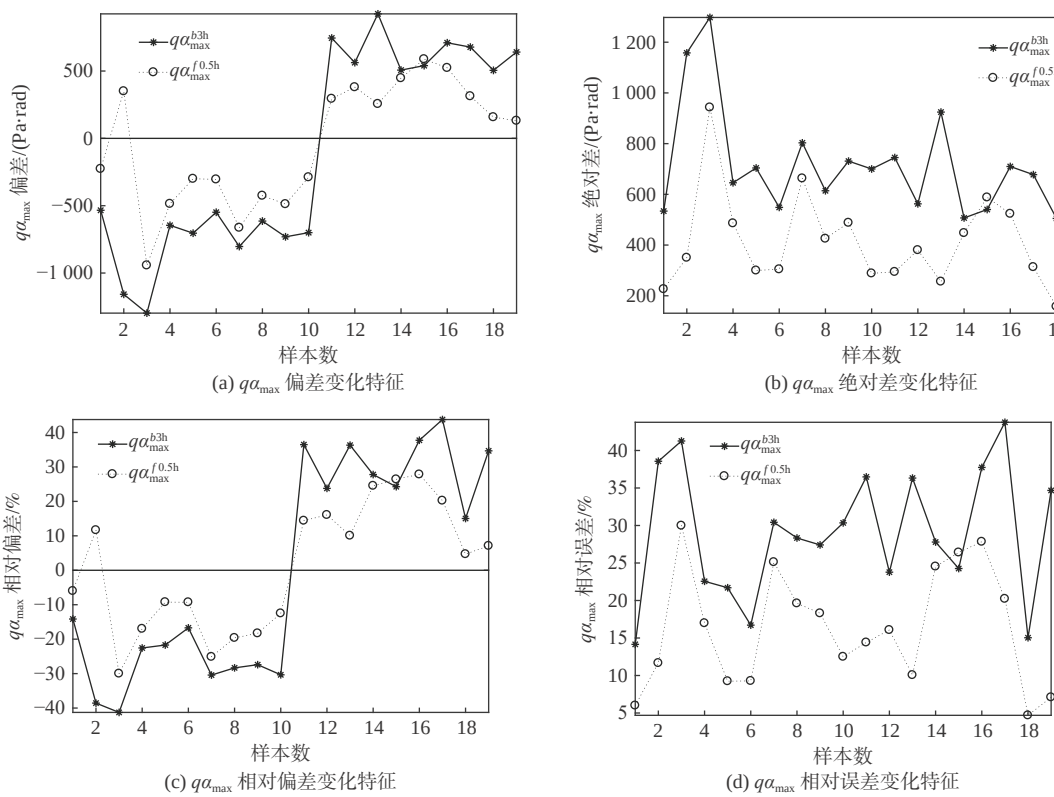


图 5　$q\alpha_{\max}$绝对差值超过 500 Pa·rad 时, $q\alpha_{\max}^{b3h}$、$q\alpha_{\max}^{f0.5h}$与$q\alpha_{\max}^{a0.5h}$之间的偏差、绝对差、相对偏差和相对误差的变化特征

Fig. 5　When the $q\alpha_{\max}$ absolute difference exceeds 500 (Pa·rad), the variation characteristics of deviation, absolute difference, relative deviation and relative error between $q\alpha_{\max}^{b3h}$, $q\alpha_{\max}^{f0.5h}$ and $q\alpha_{\max}^{a0.5h}$

$q\alpha_{\max}^{f0.5h}$与$q\alpha_{\max}^{a0.5h}$之间的偏差、绝对差、相对偏差及相对误差数值范围及平均值，从该表可看出，相比$q\alpha_{\max}^{b3h}$，$q\alpha_{\max}^{f0.5h}$偏差、绝对差、相对偏差及相对误差的数值范围均变小，相应的平均值(相对偏差除外)也减小，表明$q\alpha_{\max}^{f0.5h}$更接近$q\alpha_{\max}^{a0.5h}$。在图5(a)中，相比$q\alpha_{\max}^{b3h}$偏差，有18个样本的$q\alpha_{\max}^{f0.5h}$偏差值偏小，占总数的94.74%，$q\alpha_{\max}^{f0.5h}$绝对差总体上小于$q\alpha_{\max}^{b3h}$(见图5(b))，其中，$q\alpha_{\max}^{b3h}$、$q\alpha_{\max}^{f0.5h}$的绝对差值超过500 Pa·rad的样本数分别为19个、4个，分别占总数的100%、21.05%，减小了78.95%；类似偏差，$q\alpha_{\max}^{f0.5h}$相对偏差和相对误差总体上小于$q\alpha_{\max}^{b3h}$(见图5(c)和图5(d))，其中，相对误差超过20%的样本数分别为16个、6个，分别占总数的84.21%、31.58%，减小了52.63%；相对误差超过30%的样本数分别为9个、0个，分别占总数的47.37%、0，减小了47.37%；$q\alpha_{\max}^{b3h}$、$q\alpha_{\max}^{f0.5h}$与$q\alpha_{\max}^{a0.5h}$之间的相关系数值分别为0.27、0.71，增大了0.44。

通过上述分析，并结合表4统计结果，可知相对$q\alpha_{\max}^{b3h}$，由模型得到的$q\alpha_{\max}^{f0.5h}$更接近$q\alpha_{\max}^{a0.5h}$，即精度有较大提高。

表4 $q\alpha_{\max}$绝对差值超过500 Pa·rad时，$q\alpha_{\max}^{b3h}$、$q\alpha_{\max}^{f0.5h}$与$q\alpha_{\max}^{a0.5h}$之间的数理统计

Table 4 When the $q\alpha_{\max}$ absolute difference exceeds 500 Pa·rad, Mathematical statistics between $q\alpha_{\max}^{b3h}$、$q\alpha_{\max}^{f0.5h}$ and $q\alpha_{\max}^{a0.5h}$

参数	数值范围	平均值
$q\alpha_{\max}^{b3h}$偏差/(Pa·rad)	−1 297.62～924.43	−101.24
$q\alpha_{\max}^{b3h}$绝对差/(Pa·rad)	504.98～1 297.62	713.08
$q\alpha_{\max}^{b3h}$相对偏差/%	−41.27～43.76	0.44
$q\alpha_{\max}^{b3h}$相对误差/%	14.15～43.76	29.02
$q\alpha_{\max}^{f0.5h}$偏差/(Pa·rad)	−942.95～587.99	−35.95
$q\alpha_{\max}^{f0.5h}$绝对差/(Pa·rad)	131.21～942.95	398.22
$q\alpha_{\max}^{f0.5h}$相对偏差/%	−29.99～27.84	0.84
$q\alpha_{\max}^{f0.5h}$相对误差/%	4.69～29.99	16.32

注：所有偏差、绝对差、相对偏差、相对误差均为与$q\alpha_{\max}^{a0.5h}$相比较。

5 结 论

以2014年12月—2019年12月间隔3.5 h的$q\alpha_{\max}$资料集为例，对3.5 h内的$q\alpha_{\max}$差异特征进行统计，针对$q\alpha_{\max}$绝对差值超过500 Pa·rad的样本作为研究对象，分析其产生原因，并提出了建模订正方法，得到主要结论如下：

1）大多数样本在3.5 h内的$q\alpha_{\max}$偏差较小，偏差值位于−200～200 Pa·rad范围内占总数的67.88%；但偶尔会出现异常偏大的现象，其中，6.29%样本数的$q\alpha_{\max}$绝对差值超过了500 Pa·rad。

2）3.5 h内$q\alpha_{\max}$偏差异常偏大的原因，主要是高空风及大风区对应高度均出现异常变化导致。$q\alpha_{\max}$正偏差超过500 Pa·rad时，高空风速最大值明显增大、对应高度降低、平均风速增大，且在较低高度范围内的风速快速增大；$q\alpha_{\max}$负偏差值超过−500 Pa·rad时，高空风速最大值明显减小、对应高度明显升高、平均风速减小，且在较低高度范围内的风速快速减小。

3）相比$q\alpha_{\max}^{b3h}$，利用建模方法得到的$q\alpha_{\max}^{f0.5h}$，更接近$q\alpha_{\max}^{a0.5h}$。在19个样本中，有18个样本的$q\alpha_{\max}^{f0.5h}$绝对差值小于相应的$q\alpha_{\max}^{b3h}$，若以$q\alpha_{\max}^{a0.5h}$为基准，绝对差值超过500 Pa·rad的样本数由19个减少到4个，相对误差超过30%的样本数分别由9个减少到0个，相关系数由0.27增大到0.71，表明该建模订正方法有较大的改进作用。

上述研究表明，提前掌握$q\alpha_{\max}$在3.5 h内可能存在的变化方向(明显增大、明显减小或变化不明显)具有重要实际意义。由于没有发射时刻的$q\alpha_{\max}$，但相关成果对3 h内的$q\alpha_{\max}$差异及建模订正同样具有重要参考价值。

参考文献（References）

[1] 李东, 杨云飞, 胡鹏翔, 等. 运载火箭多体动力学建模与仿真技术研究[J]. 宇航学报, 2021, 42(2): 141-149.
LI D, YANG Y F, HU P X, et al. Research on multibody dynamic modeling and simulation technology for launch vehicles[J]. Journal of Astronautics, 2021, 42(2): 141-149(in Chinese).

[2] PIAO S Y, ZHANG Y H. Frequency-domain response prediction of a launch vehicle coupled with new payloads[J]. Journal of Spacecraft and Rockets, 2021, 58(5): 1557-1562.

[3] 李新田, 陈新民, 陈世立, 等. 固体火箭冲压发动机设计点性能优化分析[J]. 北京航空航天大学学报, 2021, 47(10): 1989-1995.
LI X T, CHEN X M, CHEN S L, et al. Optimal analysis of design point performance of ducted rocket[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(10): 1989-1995 (in Chinese).

[4] 吴杰, 张成, 李淼, 等. 基于凸优化和LQR的火箭返回轨迹跟踪制导[J]. 北京航空航天大学学报, 2022, 48(11): 2270-2280.
WU J, ZHANG C, LI M, et al. Rocket return trajectory tracking guidance based on convex optimization and LQR[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2270-2280(in Chinese).

[5] 王伟龙, 张会强. 预混C2H4/N2O推力室喷注面板热反侵着火现象数值模拟[J]. 清华大学学报(自然科学版), 2020, 60(3): 206-211.
WANG W L, ZHANG H Q. Numerical simulations of ignition by soak-back heat through the injection panel in a premixed C2H4/N2O thruster[J]. Journal of Tsinghua University(Science and Technology), 2020, 60(3): 206-211(in Chinese).

[6] WANG Y B, XU M, AN X M, et al. Numerical study on the trajectory of a long-range flexible rocket with large slenderness ratio[J]. Aerospace Science and Technology, 2021, 117(3): 106959.

[7] 冯瑞, 刘宇, 张章, 等. 火箭整流罩半罩再入过程连续流区气动特性数值研究[J]. 清华大学学报 (自然科学版), 2023, 63(3): 414-422.
FENG R, LIU Y, ZHANG Z, et al. Numerical study on the aerodynamics of a rocket fairing half in the continuum regime of the reentry process[J]. Journal of Tsinghua University(Science and Technology), 2023, 63(3): 414-422(in Chinese).

[8] 程镇煌. 大型火箭风载试验[J]. 上海航天, 1996, 13(4): 3-9.
CHENG Z H. Wind load test for large rocket[J]. Aerospace Shanghai (Chinese & English), 1996, 13(4): 3-9(in Chinese).

[9] 赵人濂, 陈振官, 付维贤. 风切变与运载火箭设计[J]. 宇航学报, 1998, 19(2): 105-108.
ZHAO R L, CHEN Z G, FU W X. Wind shear and rocket design[J]. Journal of Astronautics, 1998, 19(2): 105-108(in Chinese).

[10] 余梦伦. CZ-2E 火箭高空风弹道修正[J]. 导弹与航天运载技术, 2001, 249(1): 9-15.
YU M L. CZ-2E ballistic correction for high altitude wind[J]. Missiles and Space Vehicles, 2001, 249(1): 9-15(in Chinese).

[11] 李效明, 许北辰, 陈存芸. 一种运载火箭减载控制工程方法[J]. 上海航天, 2004, 21(6): 7-9.
LI X M, XU B C, CHEN C Y. An engineering method on the control of decreasing load for a launch vehicle[J]. Aerospace Shanghai, 2004, 21(6): 7-9(in Chinese).

[12] 耿光有, 李东. 由火箭一级飞行弹道分析底部力等动力参数[J]. 导弹与航天运载技术, 2014, 335(5): 10-13.
GENG G Y, LI D. Analysis of dynamic parameters such as base-force for 1st stage of a launch vehicle via the trajectory[J]. Missiles and Space Vehicles, 2014, 335(5): 10-13(in Chinese).

[13] 程胡华, 王益柏, 蔡其发, 等. 大气密度对运载火箭飞行的 $q\alpha_{max}$ 精度影响及建模分析[J]. 宇航学报, 2021, 42(3): 378-389.
CHENG H H, WANG Y B, CAI Q F, et al. Influence of atmospheric density on $q\alpha_{max}$ accuracy of launch vehicle flight and modeling analysis[J]. Journal of Astronautics, 2021, 42(3): 378-389(in Chinese).

[14] 程胡华, 李娟, 肖云清, 等. 风偏差对火箭最大气动载荷精度的影响与分析[J]. 北京航空航天大学学报, 2021, 47(10): 2034-2042.
CHENG H H, LI J, XIAO Y Q, et al. Influence and analysis of wind deviation on rocket maximum aerodynamic load accuracy[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(10): 2034-2042(in Chinese).

[15] 董欣心, 刘莉, 葛佳昊, 等. 捆绑火箭气动载荷分布不确定性分析[J]. 北京航空航天大学学报, 2022, 48(3): 464-472.
DONG X X, LIU L, GE J H, et al. Uncertainty analysis of aerodynamic load distribution on strap-on launch vehicle[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(3): 464-472(in Chinese).

[16] 程胡华, 张军军, 王益柏, 等. 火箭最大气动载荷预报精度及建模订正分析[J]. 航天控制, 2023, 41(4): 77-83.
CHENG H H, ZHANG J J, WANG Y B, et al. Analysis of the precision of the maximum aerodynamic load forecast for rockets and modeling revisions[J]. Aerospace Control, 2023, 41(4): 77-83 (in Chinese).

[17] 全美兰, 刘海文, 朱玉祥, 等. 高空急流在北京“7.21”暴雨中的动力作用[J]. 气象学报, 2013, 71(6): 1012-1019.
QUAN M L, LIU H W, ZHU Y X, et al. Study of the dynamic effects of the upper-level jet stream on the Beijing rainstorm of 21 July 2012[J]. Acta Meteorologica Sinica, 2013, 71(6): 1012-1019(in Chinese).

[18] 孙颖姝, 周玉淑, 王咏青. 一次双高空急流背景下南疆强降水事件的动力过程和水汽源分析[J]. 大气科学, 2019, 43(5): 1041-1054.
SUN Y S, ZHOU Y S, WANG Y Q. Analysis of dynamic process and moisture source on a heavy precipitation event in southern Xinjiang associated with the double upper-level jet[J]. Chinese Journal of Atmospheric Sciences, 2019, 43(5): 1041-1054(in Chinese).

[19] 程胡华, 成巍, 沈洪标, 等. 火箭发射前后 3.5 h 内高空风差异特征及预报[J]. 应用气象学报, 2022, 33(4): 400-413.
CHENG H H, CHENG W, SHEN H B, et al. Upper wind difference characteristics and forecast within 3.5 hours before and after rocket launch[J]. Journal of Applied Meteorological Science, 2022, 33(4): 400-413(in Chinese).

[20] 朱莉, 王曼, 李华宏, 等. 基于 WRF 模式的云南短时强降水物理量特征[J]. 大气科学学报, 2019, 42(5): 755-768.
ZHU L, WANG M, LI H H, et al. Analysis of physical quantity features of Yunnan short-time severe rainfall based on WRF model[J]. Transactions of Atmospheric Sciences, 2019, 42(5): 755-768(in Chinese).

[21] 刘郁珏, 苗世光, 刘磊, 等. 修正 WRF 次网格地形方案及其对风速模拟的影响[J]. 应用气象学报, 2019, 30(1): 70-81.
LIU Y J, MIAO S G, LIU L, et al. Effects of a modified sub-grid-scale terrain parameterization scheme on the simulation of low-layer wind over complex terrain[J]. Journal of Applied Meteorological Science, 2019, 30(1): 70-81(in Chinese).

Maximum aerodynamic load prediction method during launching of vehicle launch

CHENG Huhua[1,*], WU Shuai[2], JIANG Zhuhui[3], ZHANG Rucai[4]

(1. Unit 63729 of PLA, Taiyuan 030027, China;
2. Unit 61741 of PLA, Beijing 100094, China;
3. Beijing Institute of Applied Meteorology, Beijing 100029, China;
4. Unit 31308 of PLA, Chengdu 610031, China)

Abstract: The maximum aerodynamic load 3 h before the launch of the rocket is one of the important data to determine whether the rocket launch can be carried out as planned, and the different characteristics between 3 h before rocket launch and launch time have rarely been studied. The maximum aerodynamic load absolute difference of the 6.29% sample size surpasses 500 Pa·rad, according to an analysis of the features of the maximum aerodynamic load difference during 3.5 hours between December 2014 and December 2019. This is mainly due to the anomalous increase and decrease of the upper wind. If the changing characteristics are not detected in advance, it may affect the safe flight of the rocket; for samples with an absolute difference of more than 500 Pa·rad. The correlation coefficient rose from 0.27 to 0.71, the relative error dropped from 29.02% to 16.32%, and the absolute difference dropped from 713.08 Pa·rad to 398.22 Pa·rad. It shows that the modeling method has a certain improvement effect, which is conducive to improving the guaranteed ability of rocket safety flight.

Keywords: launch vehicle launch; the maximum aerodynamic load deviation; maximum aerodynamic load prediction; modeling accuracy; fly safely

Received: 2023-05-12; **Accepted**: 2023-06-30; **Published Online**: 2023-08-30 16:48
URL: link.cnki.net/urlid/11.2625.V.20230830.1443.002

* **Corresponding author**. E-mail: chenghongxi2012@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0171

基于卷积神经网络的星载粒子鉴别平台及应用

白超平[1,2]，张珅毅[1,2,*]，张鑫[1]，孙越强[1]，张帅[1]，王子婷[1,2]

(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 100149)

摘 要：空间辐射粒子的准确鉴别对科研和工程应用至关重要。现有的粒子鉴别方法，包括探测器望远镜法、静电分析飞行时间法、飞行时间能量法及波形分析能量法，已经在实际应用中取得了良好的效果。结合卷积神经网络(CNN)强大的特征提取与分类能力，有望进一步提升粒子能量测量和种类鉴别的精度。基于空间环境探测载荷常用环境，提出一种用于构建在轨CNN粒子鉴别平台的方法，实现粒子种类鉴别。构建了多维度的输入数据集，借助软件平台完成模型的训练与权值的导出，并通过硬件平台完成波形的推断与数据集的扩充。利用建立好的鉴别平台对实际测试得到的中子和伽马波形数据进行训练和测试，并分析软硬件平台鉴别的准确度，完成了平台的验证工作。鉴别平台的建立和应用为未来空间环境探测中粒子测量和鉴别提供了一种新的思路和方法，具有较强的工程实践意义。

关 键 词：粒子探测；卷积神经网络；星载FPGA平台；脉冲波形鉴别

中图分类号：V11

文献标志码：A **文章编号**：1001-5965(2025)04-1313-11

浩瀚的宇宙空间充斥着各种粒子，来源包括太阳风、宇宙射线、地球辐射带等，因此，各种卫星平台都暴露在空间辐射之中。通过对宇宙中带电粒子的成分、能谱、方向及粒子随空间和时间变化的研究，人们可以得到关于粒子起源、加速和传播过程、寿命和动力学过程等重要信息。此外，空间辐射对各种空间飞行器和航天员的影响不容小觑，粒子可以引起电子设备故障、干扰通信、损坏电子元件、危害人类健康等。因此，研究空间环境中粒子的各种物理特性至关重要，而粒子鉴别则是其中的关键一环。通过粒子鉴别，可以确定粒子的种类、能量和通量等信息，从而评估粒子对航天器和航天员的影响，并可以用于制定相应的防护措施，保障航天器和航天员的安全和健康。因此，粒子鉴别在科学和工程领域中具有极其重要的意义[1]。

基于以上情况，粒子鉴别是各国学者聚焦的一个研究领域。现阶段星载粒子鉴别主要包含4种常用方法：①探测器望远镜法[2](ΔE-E)，主要原理是采用一片或多片固体探测器(solid state detector, SSD)组成叠层，如果第1片传感器足够薄，则可以得到粒子穿透以后沉积能量ΔE，第2片(或多片)传感器足够厚，将粒子剩余能量完全沉积，即可测量出粒子总能量E，通过ΔE vs E二维谱，即可区分出粒子种类。探测器望远镜法简单、可靠，在实际的工程应用中广泛使用，但是需要粒子在透射探测器中损失很少的能量，不适用于较重离子的探测。②静电分析飞行时间法[3](electrostatic anylyzer-time of flight, ESA-TOF)，通过ESA扫描电压V对入射离子的能量电荷比E/q进行筛选，再结合TOF，最终得到离子的质量电荷比。ESA常常用来测试能量较低的带电粒子或中性原子，但由于星载设备空间尺寸的限制，对于中高能粒子需要通过大幅度增

收稿日期：2023-04-10；**录用日期**：2023-07-21；**网络出版时间**：2023-08-18 11:03
网络出版地址：link.cnki.net/urlid/11.2625.V.20230817.1606.001
基金项目：国家自然科学基金(42204180)
*通信作者. E-mail：zsy@nssc.ac.cn

引用格式：白超平，张珅毅，张鑫，等. 基于卷积神经网络的星载粒子鉴别平台及应用[J]. 北京航空航天大学学报，2025，51(4)：1313-1323. BAI C P，ZHANG S Y，ZHANG X，et al. Spaceborne particle identification platform and application based on convolutional neural network[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1313-1323 (in Chinese).

加物理结构来实现探测，这一点无法在轨实现。③飞行时间能量法[4](TOF×E)，其基本原理依据粒子能量公式，即只需测出粒子能量 E 和速度 v(由距离 l 和时间 t 计算可得)便可得到粒子质量从而确定种类。TOF×E 对轻粒子，特别是低能轻粒子，鉴别效果比较好。随着粒子质量的增加，质量的差别越来越小，分辨起来就较为困难。此外，对同质异位素，即具有相同质量数、不同原子序数的核素，采用飞行时间法无法鉴别，必须与其他方法结合起来，才能进行可靠的鉴别。④半导体传感器上升沿能量法[5](PSA×E)或闪烁体探测器波形分析法[6-7]，半导体传感器上升沿能量法是通过带电粒子在半导体传感器中产生的脉冲波形，利用高速数字采集技术作出能量 E 与上升沿时间的二维谱图进行粒子鉴别，而闪烁体探测器波形分析法是通过中子和伽马在闪烁体中产生的发光波形，利用功率谱密度估计(power spectral density estimation, PSD)的分析方法鉴别中子和伽马。虽然该方法尝试使用波形信息，但只使用了少部分数据，并未利用波形的全部信息，可能会由于偶然的测量误差造成结果的偏差。除此之外，以上 4 种方法均需要将在轨的科学数据下传到地面再进行粒子鉴别的分析工作，这对于资源紧张的卫星资源，尤其是深空探测领域并不适用。

目前地面上在许多领域已经开始使用卷积神经网络(convolutional neural networks, CNN)算法来实现一维波形的识别与分类，包括高能物理实验[8]、心电图病理诊断[9]等领域。基于以上背景，通过 CNN 算法，利用空间中带电粒子或中性粒子与传感器作用后产生的全波形组合信息进行粒子鉴别是一种切实可行的手段。

本文拟利用机器学习中 CNN 强大的分类能力建立具有粒子波形数据软件预处理模块、CNN 软件训练模块、粒子波形数据现场可编程门阵列(field-programmable gate array, FPGA)前向推断模块、CNN 训练集扩展模块的粒子鉴别平台。基于 CNN 的粒子鉴别平台与传统的粒子鉴别方法相比，其主要优势有：CNN 擅长于将数据中不同特征进行分离并最终实现分类；CNN 是通过波形整体的数据进行分析，其数据信息更丰富，有望提升精度[10]；CNN 具备良好的扩展性、可移植性及泛化性，因此，可以与传统方法完美融合并改进和提高实际应用的准确性；FPGA 本身具备高并行、低功耗、可重配置的优势[11]，可以实现在轨实时处理与模型更新，降低在轨资源使用率并提升探测准确度。

通过建立 CNN 粒子鉴别平台，可以在传统的粒子鉴别技术之上进一步扩展，实现“CNN+传统粒子鉴别技术”的技术路线，同时弥补传统粒子粒子鉴别和干扰剔除的物理方式，实现软件粒子鉴别和干扰剔除的新方法。此外，通过构建通用的星载粒子波形信号识别 FPGA 平台，可以有效的降低在轨数据率，并兼具可重构和可扩展特性。因此，建立 CNN 的粒子鉴别平台在未来空间科学探索和工程实践中具有较大的意义。

1 CNN 粒子鉴别平台概述

CNN 粒子鉴别平台的构建不仅仅需要考虑粒子鉴别的准确度，还需要考虑在轨运行的特点，如 FPGA 资源、空间粒子通量等信息，最终得到一个最优化且可行的模型架构。

为满足以上要求，CNN 粒子鉴别平台主要由 4 部分构成，分别是粒子波形数据软件预处理模块、CNN 软件训练模块、粒子波形数据 FPGA 前向推断模块、CNN 训练集扩展模块，其平台架构如图 1 所示。

1.1 粒子波形数据软件预处理模块

粒子波形数据软件预处理模块的主要功能是将在轨或地面加速器获得的各类粒子波形数据进行处理，包括数据的归一化、随机采样、随机裁剪及噪声添加，提升模型的鲁棒性和泛化性，并将处理后的数据分成训练数据、验证数据和测试数据。训练数据主要用来完成整个网络结构的权值和偏置训练，验证数据用来确定模型的超参数是否合理，测试数据主要用来完成最终的粒子鉴别和推断，同时度量模型推断的准确率性能。

1.2 CNN 软件训练模块

CNN 软件训练模块主要包括：CNN 网络模块、训练模块、测试验证模块及权值偏置导出和定点数转换模块，分别用来实现模型架构搭建、模型权值和偏置训练、模型推断结果验证及 FPGA 平台权值和偏置导出。

1.3 粒子波形数据 FPGA 前向推断模块

粒子波形数据 FPGA 前向推断模块主要用于实现 FPGA 平台内各类资源的计算以及前向推断工作，包括卷积层、relu 操作、池化层、全连接层的定点数计算和内部 rom 和 RAM 的数据存取操作，并最终输出推断结果。

1.4 CNN 训练集扩展模块

CNN 训练集扩展模块主要用于实现训练样本集的扩充，以此提升鉴别的精度和准确性。在轨获取的波形数据，可以定期下传到地面作为样本集扩充的手段，对模型进行再训练并更新参数，最终通过数据注入的方式更新 FPGA 平台的各层参数，实

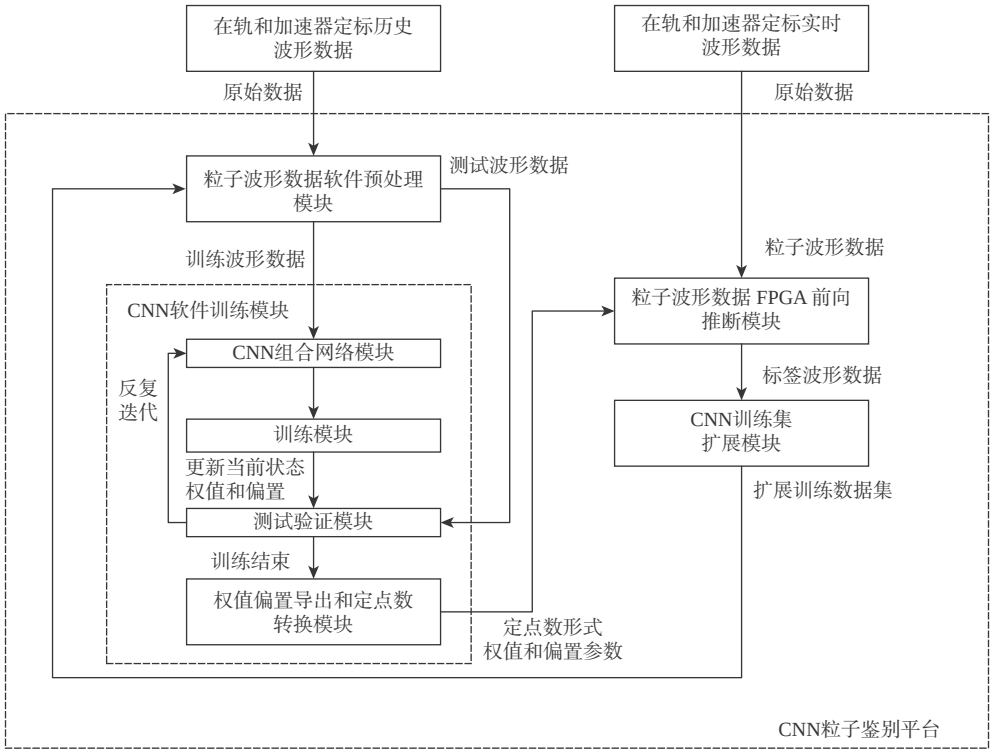


图 1　CNN 粒子鉴别平台结构

Fig. 1　Structure of CNN-based particle identification platform

现天地模型更新的闭环。

2　CNN 粒子鉴别平台建立

CNN 擅于将数据中不同特征进行分离并最终实现分类, 如果当数据集之间存在特征差异, 将会达到较好的分类结果。本节首先理解带电粒子在半导体传感器中沉积波形特征和中性粒子在闪烁体探测器中发光的波形特征, 并利用 CNN 软硬件模型将特征信息进一步提取和分类, 最终实现鉴别和数据集扩充的目的。

2.1　粒子波形数据软件预处理模块的构建

为构建粒子波形数据软件预处理模块, 需要深入了解粒子波形数据的特点, 并依据该特点构建对应的训练数据集, 除此之外通过常规的数据增强方法如随机采样、随机裁剪、噪声添加等手段进一步扩充训练数据集, 以此达到提升模型泛化性和鲁棒性的目的。

2.1.1　粒子波形特征的理论分析

带电粒子探测中半导体探测器是最为常用的手段。由于内部电场的作用, 探测器内电场强度从后端到前端逐渐增加, 当带电粒子由后端入射时, 较重的离子由于其短射程、较高能量损失率及低电场强度, 其径迹上的等离子体侵蚀时间较长。在电荷载流子传输中, 电子迁移率较快, 空穴迁移率较慢。由于是从探测器后端入射, 空穴迁移的平均路程会增加, 从而导致电荷载流子的传输时间增加。当入射能量一定时, 粒子核电荷数和质量数的变化会导致电荷收集时间的差异。较重的粒子产生的电流脉冲持续时间较长、幅值较低、电荷上升时间较长且过零时间较大, 因此, 在传感器中沉积的不同粒子会产生一定的波形差异[12-16], 具体波形特征如图 2 所示, 其中, $i(t)$为电流, $q(t)$为电荷, $U(t)$为电压。

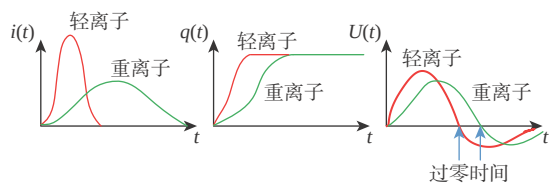


图 2　相同能量不同粒子波形信号示意图

Fig. 2　Schematic diagram of waveform signals for different particles with same energy

针对中性成分而言, 以中子和伽马为例, 有机闪烁体被中子或 γ 射线照射后, 分别产生反冲质子和次级电子而发光, 光的强度很快增强直至最大, 相反的是衰减相对较慢, 衰减应相近似于指数衰减。其发光衰减时间中包含快、慢 2 种成分。当中子或 γ 射线照射进入同一闪烁体后, 在闪烁体中形成的电离密度有所差异从而导致发光衰减时间快慢成分的强度比不同。中子与闪烁体作用所产生的质子激发的荧光中快成分的份额比 γ 射线与闪烁体作用所产生的电子激发的荧光低, 而慢成分的份额比 γ 射线高。正是这一因素, 使得中子与 γ 射

线与闪烁体相互作用所产生荧光的衰减时间有差异，进而导致生成的脉冲形状有所区别[17-18]，具体波形特征如图3所示。

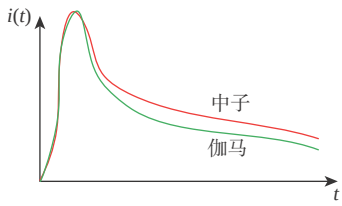


图3 中子和伽马波形示意图

Fig. 3 Schematic diagram of neutron and gamma waveforms

除此之外，由于发光机制的区别，重离子、质子和电子与闪烁体作用产生的脉冲波形同样具备差异性[19]。

综上所述，粒子与传感器作用后产生的波形数据存在一定程度的差异，这为本文使用CNN鉴别提供了可能性。

2.1.2 数据预处理模块的搭建

由2.1.1节的理论分析可知，不同成分的粒子脉冲波形在上升沿或下降沿都具有一定的差异性，为增强模型的普适性和鲁棒性，数据集采用随机裁剪、降采样和噪声添加等数据增强的方式，其基本构成如表1所示。

通过高速采集得到的波形数据需要通过降采样满足模型数据输入长度的要求，此外，为进一步提升模型的鲁棒性，对各个训练集进行随机噪声添加的操作。

表1 训练数据集数据增强信息

Table 1 Data augmentation information for training dataset

序号	训练集内容	数据增强方式
1	波形上升沿信息	降采样+噪声添加
2	波形下降沿信息	降采样+噪声添加
3	全波形信息	降采样+噪声添加

2.2 CNN软件训练模块

软件训练模块主要包括几个部分，分别为CNN网络模块、训练模块、测试验证模块及权值偏置导出和定点数转换模块。

2.2.1 CNN网络模块

构建CNN需要重点确定几个关键参数，分别是输入数据个数、网络层个数、各个卷积层的卷积核通道数、卷积核大小、计算步长、填充数据个数及池化数据大小，为提高软件的灵活性，将除网络层个数外的其他参量以变量形式作为输入，方便进行修改。本文平台以3层卷积层作为基本雏形进行构建，即网络架构为：输入层-卷积层1-池化层1-卷积层2-池化层2-卷积层3-全连接层-输出层（可依据实际类别数量进行调整），如表2所示。

表2 CNN架构信息

Table 2 CNN architecture information

网络层	输入	详细	输出
第1层卷积	1通道M点粒子波形数据	A个卷积核，卷积核元素为$A\times5$	A个通道，每通道M点数据
第1层池化	A个通道，每通道M点数据	X数据最大池化	A个通道，每通道M/X点数据
第2层卷积	A个通道每通道M/X点数据	B个卷积核，卷积核元素为$B\times5$	B个通道，每通道M/X点数据
第2层池化	B个通道每通道M/X点数据	Y数据最大池化	B通道，每通道$M/(XY)$点数据
第3层卷积	B个通道每通道$M/(XY)$点数据	C个卷积核，卷积核元素为$C\times5$	C个通道，每通道$M/(XY)$点数据
全连接	C个通道每通道$M/(XY)$点数据	数据拼接	1通道$CM/(XY)$点数据
输出层			D个种类

注：$M/(XY)$应为正整数。

一维CNN输出的长度计算公式为

$$O=\left[\frac{I-K+2P}{S}\right]+1 \tag{1}$$

式中：I为输入数据长度；K为卷积核尺寸；P为边缘填充数据个数；S为滑动步长；O为单通道输出数据长度。

现阶段采用的模型架构卷积核尺寸K为5，滑动步长S为1，为保证输入和输出数据的维度一致，即I与O相等，边缘填充数据个数为设定为2。

2.2.2 训练模块

训练模块包括粒子波形前向传播计算、损失函数计算与反向梯度计算。依托CNN的架构，对训练数据分别进行前向传播、损失函数计算和反向梯度计算。前向传播是训练参数的前提条件，可以计算分类结果和实际推断的偏差；损失函数计算是为了了解标签数据和计算数据的偏差并以此偏差通过反向梯度计算不断迭代和更新权值参数，达到最终训练的目的。实际运行过程中，首先，确定学习率lr、批尺寸batchsize及数据集训练次数Epoch，以上参数均可通过手动修改变量进行更改，随后通过CrossEntropyLoss()函数计算损失函数、通过backward()进行反向梯度计算并不断迭代训练参数。

2.2.3　测试验证模块

测试验证模块用来分析当前训练状态下粒子鉴别模型实现分类和鉴别的准确度和能力, 通过输入预先处理好的混合测试数据集, 可以得到验证和推断结果的准确率, 从而了解模型训练的有效性, 为确定模型超参数以及确定模型架构提供数据支持。实际操作过程中, 关键变量包括训练集准确度、损失函数值、验证集准确度及测试集准确度。训练集准确度的计算方法为

$$A = B/C \tag{2}$$

式中: A 为训练准确度; B 为训练数据前向推断结果与标签匹配的个数, C 为训练数据总数。

损失函数值的计算方法为

$$D = E/C \tag{3}$$

式中: D 为训练的总体平均损失函数值; E 为总损失函数值; C 为训练数据总数。

验证集准确度的计算方法为

$$F = G/H \tag{4}$$

式中: F 为验证集准确率; G 为验证集前向推断结果与标签匹配的个数; H 验证集数据总数。

测试集准确率计算方法为

$$I = J/K \tag{5}$$

式中: I 为测试集准确率; J 为测试集前向推断结果与标签匹配的个数; K 为测试集数据总数。

除计算中间变量外, 为更为直观的了解整个训练过程的参数变化趋势, 以训练准确度、损失函数值及验证准确度为纵轴, 迭代运行次数 Epoch 为横轴绘制曲线, 更为直观的观察模型迭代训练的状态。

2.2.4　权值偏置导出和定点数转换模块

权值偏置导出和定点数转换模块在训练结束后输出定点数形式的权值和偏置, 并以 COE 文件的方式写入 FPGA 中的 ROM 中。按照模型的基本架构, 定点数输出的参数文件构成如表 3 所示。

表 3　定点数参数信息

Table 3　Fixed-point number parameter information

序号	参数名称	参数排布
1	卷积层1参数Conv1.coe	卷积层1各通道卷积核+各通道偏置
2	卷积层2参数Conv2.coe	卷积层2各通道卷积核+各通道偏置
3	卷积层3参数Conv3.coe	卷积层3各通道卷积核+各通道偏置
4	全连接层参数Full.coe	输出类别对应的全连接层展平后参数+输出类别节点偏置

2.3　粒子波形数据 FPGA 前向推断模块

FPGA 构建 CNN 的一个关键环节就是通过并行运算结构实现卷积层的多通道运算, 卷积计算伪代码如下所示。

For(no = 0;no<N_{of};no++)——Loop4

　For(x=0;x<N_{ox};x++)——Loop3

　　For(ni=0;ni<N_{if};ni++)——Loop2

　　　For(kx=0;kx<N_{kx};kx++)——Loop1

　　　　WaveL(no;x)+=WaveL-1(ni;x+kx)×weightL-1(ni,no;kx);

　　　WaveL(no;x)= WaveL(no;x)+bia(no)

其中, no 表示当前输出通道标签值; N_{of} 表示输出通道总数; x 表示当前滑动窗口标签值; N_{ox} 表示滑动窗口总数; ni 表示当前输入通道标签值; N_{if} 表示输入通道总数; kx 表示当前卷积核标签值; N_{kx} 表示卷积核尺寸; waveL表示输出通道特征值; waveL-1 表示输入通道特征值; weightL-1表示输入通道对应卷积核权重值; bia表示输出通道对应偏置值。

4 个循环的具体含义如下, Loop1 指卷积核内的各元素计算, Loop2 指单个滤波器不同输入通道对应的卷积核计算, Loop3 指单个滤波器在一维数据上的滑动计算, Loop4 指多个滤波器的计算。现阶段采用的并行计算策略为卷积核间全并行, 输入通道全并行, 单个卷积核并行运算, 滑动窗口串行运算。

依据以上并行策略, 脉动阵列是最佳的计算手段, 以 5 元素卷积核和 3 输入数据的单通道输入为例构建脉动阵列计算过程如图 4 所示, 下面的 5 个矩形为计算单元 PE, 包括了 5 个卷积核元素 W_0～W_4, 3 个输入数据按照时钟节拍从左到右依次输入。

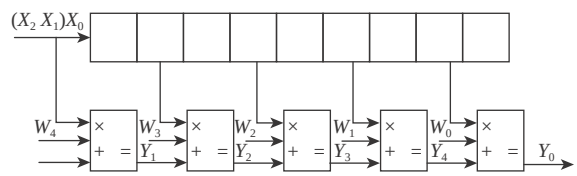


图 4　脉动阵列计算示意图

Fig. 4　Schematic Diagram of pulse array calculation

卷积计算是芯片资源和时间资源占用的主要部分, 定量分析运算资源、了解时间延时是构建平台的关键步骤之一。由于卷积运算主要为乘加运算, 且并行要求较高, 因此在构建运算单元(PE 单

元)时采用FPGA内专用的数字信号处理(digital signal processing, DSP)模块进行计算,其实际构建过程DSP模块占用情况如表4所示,时间资源如表5所示。

表4 DSP需求计算

Table 4 DSP requirement calculation

网络层	输入	输出	需要DSP单元个数
第1层卷积	1通道M点粒子波形数据	A个通道,每通道M点数据	$1\times A\times 5$
第1层池化	A个通道,每通道M点数据	A个通道,每通道M/X点数据	
第2层卷积	A个通道每通道M/X点数据	B个通道,每通道M/X点数据	$A\times B\times 5$
第2层池化	B个通道每通道M/X点数据	B个通道,每通道$M/(XY)$点数据	
第3层卷积	B个通道每通道$M/(XY)$点数据	C个通道,每通道$M/(XY)$点数据	$B\times C\times 5$
全连接	C个通道每通道$M/(XY)$点数据	1通道$CM/(XY)$点数据	$C\times D$
输出层		D个种类	

表5 推断时间计算

Table 5 Inference time calculation

网络层	输入	输出	所需时间(CLK)
第1层卷积	1通道M点粒子波形数据	A个通道,每通道M点数据	M+8
第1层池化	A个通道,每通道M点数据	A个通道,每通道M/X点数据	
第2层卷积	A个通道每通道M/X点数据	B个通道,每通道M/X点数据	M/X+8
第2层池化	B个通道每通道M/X点数据	B个通道,每通道$M/(XY)$点数据	
第3层卷积	B个通道每通道$M/(XY)$点数据	C个通道,每通道$M/(XY)$点数据	$M/(XY)$+8
全连接	C个通道每通道$M/(XY)$点数据	1通道$CM/(XY)$点数据	$M/(XY)$
输出层		D个种类	

通过以上各个步骤,可以首先确定模型架构在FPGA中占用的DSP模块及推断所需时间,判断是否满足实际使用要求。在符合实际使用条件下,通过FPGA的前向推断模块,可以得到网络不同模型不同输出节点的计算值,并通过简单的大小比较就可以最终确定粒子种类。

2.4 CNN训练集扩展模块

通过FPGA推断可以完成波形数据的打标,并实现训练样本集的输出和存储,以此达到样本集数据增强的目的。基于该数据集定期进行模型的再训练并更新参数可以进一步提升粒子鉴别模型的精度以及泛化特性。

3 CNN粒子鉴别平台应用

本节利用中子和伽马波形的实测数据,通过CNN粒子鉴别平台实现中子和伽马鉴别模型的软件训练和FPGA平台推断以及训练集扩展的应用。

3.1 中子和伽马预处理数据构建

中子和伽马数据采用基于CLYC闪烁体传感器构建的实验平台产生的波形数据。为保证CNN输入数据数量的一致,数据的输入长度均设定为100,并均通过归一化规范输入数据。随后采用随机裁剪、噪声添加以及降采样等数据增强方式,产生18 000条训练波形数据(中子和伽马各9 000条)、3 000条验证波形数据(中子和伽马各1 500条),如图5所示,3 000条测试波形数据(中子和伽马各1 500条)如图6所示,覆盖波形上升沿、下降沿及全波形信息。

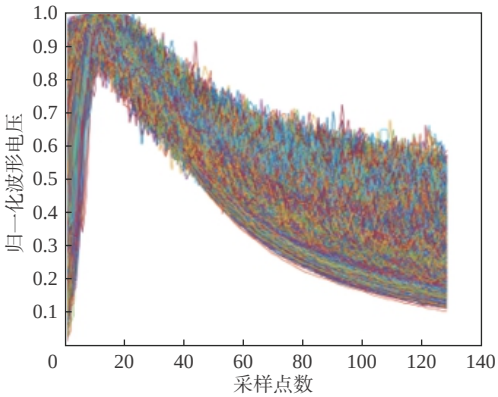
3.2 中子和伽马CNN模型构建

为均衡考虑软件推断准确率与硬件资源,中子和伽马CNN模型基本架构如表6所示。

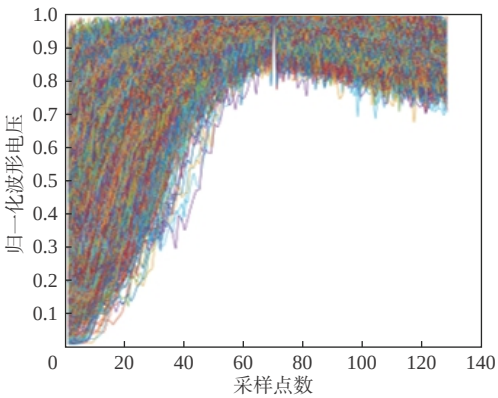
首先确定迭代次数为1 500,该数值在以往的训练过程中基本可以得到较为准确的训练结果;在固定批尺寸的状态下,分别对学习率为0.000 5、0.001和0.005进行测试,验证集准确率曲线如图7所示,可以看到学习率在0.005状态下训练稳定性较差,学习率为0.000 5和0.001状态下稳定性相当但0.001收敛速度更快,因此,选用0.001作为学习率超参数;在迭代次数为1 500,学习率为0.001的参数下,对320、420和520这3种批尺寸进行了测试,如图8所示,批尺寸为420时,验证集准确率曲线较为平稳且收敛,故将批尺寸设为420。

基于以上参数,对整个模型进行训练后,训练精度、损失误差及验证集准确率曲线如图9所示。

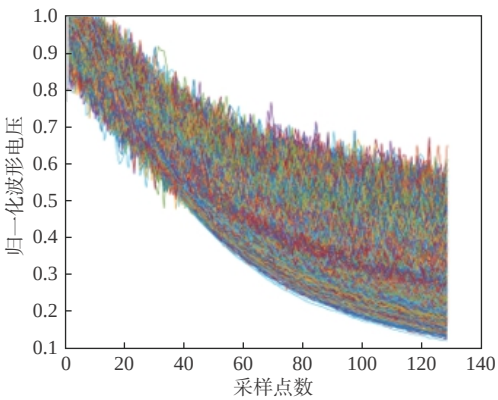
依据训练完成的CNN,对测试波形数据进行前向推断,推断的数据包括中子和伽马,且均包括全



(a) 全波形训练和验证集



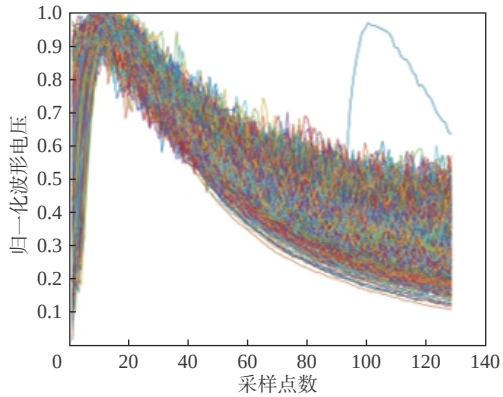
(b) 上升沿训练和验证集



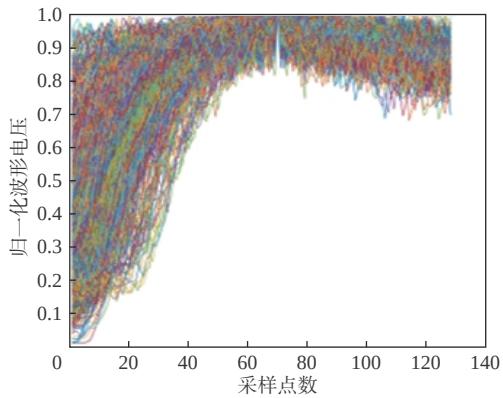
(c) 下降沿训练和验证集

图 5　3 种类型训练和验证数据集

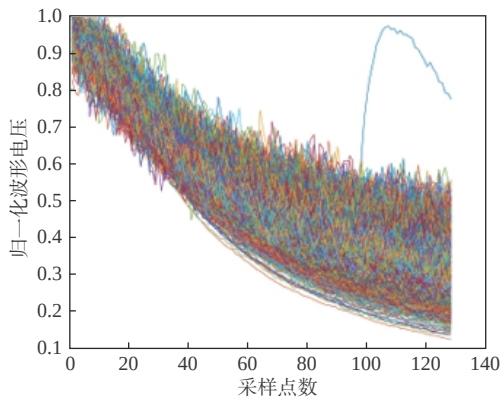
Fig. 5　Three types of training and validation datasets



(a) 全波形测试集



(b) 上升沿测试集



(c) 下降沿测试集

图 6　3 种类型测试数据集

Fig. 6　Three types of test datasets

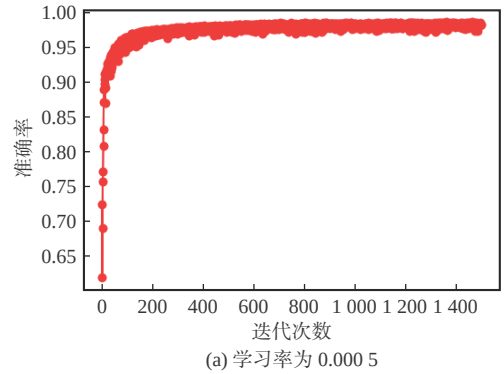
表 6　中子和伽马 CNN 模型架构信息

Table 6　CNN architecture information of neutron and gamma

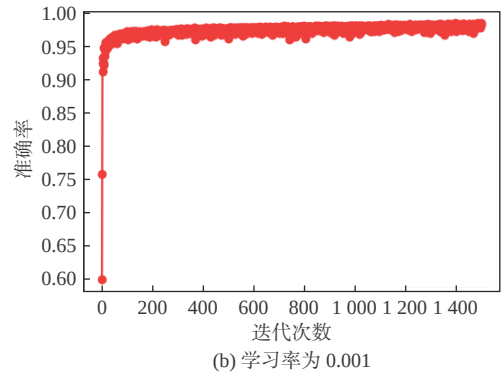
网络层	输入	详细	输出
第1层卷积	1通道100点粒子波形数据	4个卷积核, 卷积核元素为20	4个通道, 每通道100点数据
第1层池化	4个通道, 每通道100点数据	2数据最大池化	4个通道, 每通道50点数据
第2层卷积	4个通道每通道50点数据	8个卷积核, 卷积核元素为40	8个通道, 每通道50点数据
第2层池化	8个通道每通道50点数据	2数据最大池化	8通道, 每通道25点数据
第3层卷积	8个通道每通道50点数据	8个卷积核, 卷积核元素为40	8个通道, 每通道25点数据
全连接	8个通道每通道25点数据	数据拼接	1通道200点数据
输出层			2个种类

波形数据、上升沿数据、下降沿数据和整体测试集进行准确率评估, 评估结果如表 7 所示。

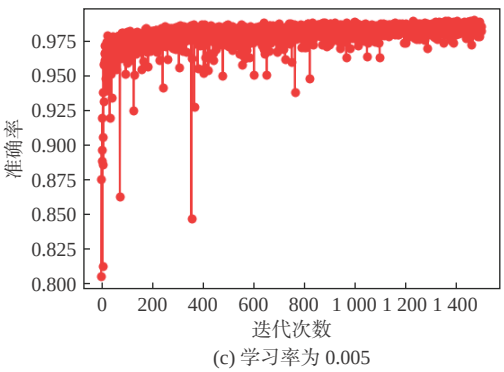
综合表 7 测试准确率的结果, 训练完成的 CNN 对波形的上升沿、下降沿和全波形数据训练精度均



(a) 学习率为 0.000 5



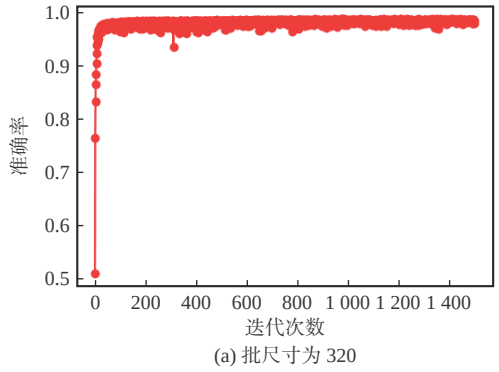
(b) 学习率为 0.001



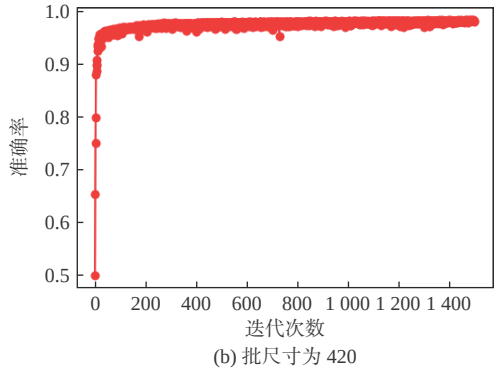
(c) 学习率为 0.005

图 7 不同学习率的验证集准确率曲线

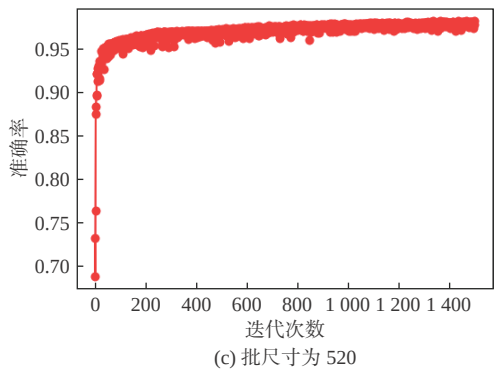
Fig. 7 Validation dataset accuracy curves for different learning rates



(a) 批尺寸为 320



(b) 批尺寸为 420



(c) 批尺寸为 520

图 8 不同批尺寸的验证集准确率曲线

Fig. 8 Validation dataset accuracy curves for different batch sizes

可以达到 97% 以上，因此，当数据质量某一部分较差时，可以选择其他部分数据进行鉴别。

3.3 中子和伽马 FPGA 平台前向推断的实现与训练数据集的扩充

根据第 2 节 DSP 需求计算表格的内容，构建当前架构的 CNN 需要的 DSP 资源如表 8 所示。

本文项目基于XC7Z035 平台的ZYNQ 系列FPGA，其包含 900 DSP 模块，实际使用 DSP 为 516 个，可以满足实际使用要求。

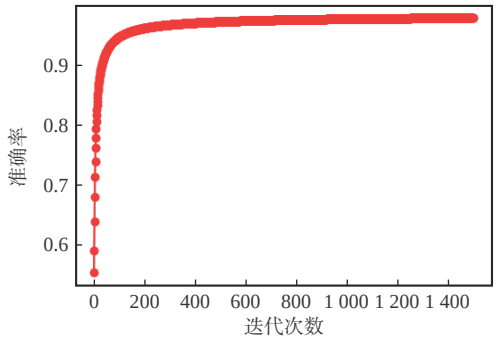
卷积层和全连接计算的理论时间如表 9 所示，运行时钟为 200 M，则总共需要 1.12 us。

考虑到实际运行过程中，还包括了数据读取和写入等时间，为进一步考量 FPGA 平台的实际推断速度，通过仿真测试衡量完成一次推断所需的实际时间，评估结果如图 10 所示。

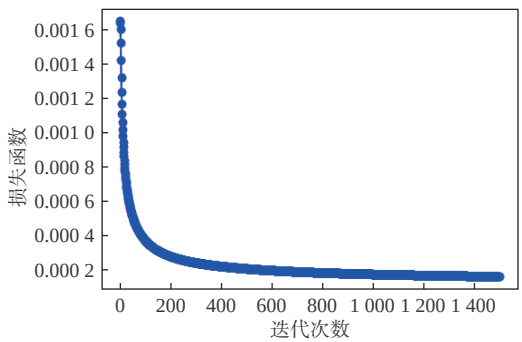
测试结果表明推断一次所需要的时间为 2.95 us，后期通过优化可以进一步降低数据存取所消耗的时间。实际在轨运行中，通过设计几何因子的大小，控制进入探测器中的粒子通量在 10^5/s 以下，即 10 μs 的粒子事件间隔，该平台可以满足实际的使用要求。

为进一步验证 FPGA 平台 CNN 模型推断的正确性，分别输入 3 种类型的波形（上升沿、下降沿和全波形）到同一模型中，并各包含一个中子和伽马的测试样本，观察软件和 FPGA 输出的结果及差异性，结果如表 10 所示。

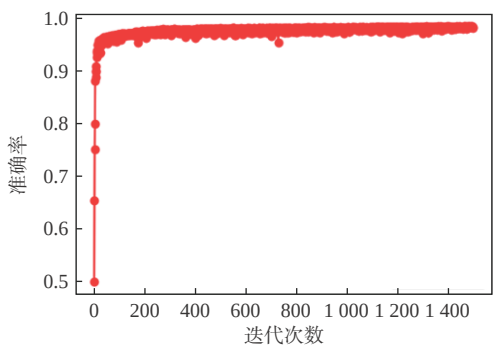
依据结果可知，针对输入数据，软件和 FPGA 均能够正常的判断粒子种类，但由于定点数与浮点数之间的误差，导致两者的计算结果并不一致，后期还需要通过提高精度或其他手段降低误差。



(a) 训练集准确率曲线



(b) 损失误差曲线



(c) 验证集准确率曲线

图 9　训练集和验证集计算结果

Fig. 9　Training and validation set computation results

当 FPGA 完成前向推断和标签标注后, 将 softmax 概率值大于 95% 的数据集作为本文训练数据集的扩展, 代入软件训练模块进行迭代训练, 完

表 7　测试准确率评估

Table 7　Test accuracy evaluation

测试集种类	测试准确率/%
全波形测试集	98
上升沿测试集	99
下降沿测试集	97
整体数据集	98

表 8　DSP 所需资源

Table 8　DSP resource requirements

网络层	输入	输出	需要DSP单元个数
第1层卷积	1通道100点数据	4个通道, 每通道100点数据	20
第2层卷积	4个通道每通道50点数据	8个通道, 每通道50点数据	160
第3层卷积	8个通道每通道25点数据	8个通道, 每通道25点数据	320
全连接	8个通道每通道25点数据	1通道200点数据	16

表 9　推断时间理论分析

Table 9　Theoretical analysis of inference time

网络层	输入	输出	所需时间
第1层卷积	1通道100点数据	4个通道, 每通道100点数据	108
第2层卷积	4个通道每通道50点数据	8个通道, 每通道50点数据	58
第3层卷积	8个通道每通道25点数据	8个通道, 每通道25点数据	33
全连接	8个通道每通道25点数据	1通道200点数据	25



图 10　仿真测试推断时间

Fig. 10　Inference time of simulation test

表 10　软件和 FPGA 推断结果对比

Table 10　Comparison of software and FPGA inference results

输出数据	软件推断输出		FPGA推断输出		Softmax结果		软件判断结果	FPGA判断结果
	模型中子推断节点	模型伽马推断节点	模型中子推断节点	模型伽马推断节点	软件准确率/%	FPGA推断准确率/%		
全波形中子	1.604 7	−2.287 6	0.499 5	−1.191 6	98	84	中子	中子
全波形伽马	−4.595 5	3.560 4	−5.809 5	4.708 2	99	99	伽马	伽马
上升沿中子	3.968 5	−3.671 7	2.280 5	−1.655 7	99	98	中子	中子
上升沿伽马	−3.399 8	4.189 7	−4.468 7	5.475	99	99	伽马	伽马
下降沿中子	2.6	−3.280 6	1.764 8	−2.526 8	99	98	中子	中子
下降沿伽马	−2.354 3	1.367	−3.057 3	1.987 7	97	99	伽马	伽马

成训练数据集的扩充和权值的更新。

4 结 论

1）本文以不同粒子在传感器中沉积后产生的脉冲波形具有差异性为出发点，构建了一套完整的软硬件CNN粒子鉴别平台，包括数据预处理、软件CNN架构、FPGA前向推断平台及数据集扩展模块。该平台可以广泛的适用于粒子鉴别，并可以在传统的方法基础上进行扩展。

2）利用构建好的CNN平台对中子和伽马的粒子波形信息进行鉴别测试，训练集包括了多种波形信息特征的输入，以此达到数据增强的目的，提升模型的鲁棒性和泛化性；利用训练完成的模型对不同种类的测试数据分别进行了准确率测试，测试结果表明，各类输入数据鉴别准确率均可以达到97%以上。

3）在软件模型训练的基础之上，利用导出的权值和偏置数据以定点数形式固化到FPGA平台上，并对比软件推断和FPGA推断，结果表明，软件模型和FPGA模型均可以准确的输出鉴别结果，但存在误差。

在随后的工作中，需要逐步完善CNN粒子鉴别平台，包括进一步扩展数据集种类和数量，譬如从时域和频域2个层面作为输入，或通过仿真分析的方式扩充理论波形数据集；进一步提升模型架构对波形数据的特征提取能力，诸如采用CNN+RNN、组合网络表决模型或多任务输出模型，进一步提升模型复杂度等；进一步提升构建FPGA加速推断平台的效率和准确度，诸如采用vivado HLS软件提升软件开发迭代速度并思考降低由于定点数带来的计算误差等；进一步完成整体测试数据分类可视化及分类错误数据集的分析工作；进一步完成硬件平台全数据的分类测试工作，并与软件分类结果进行对比分析。有理由相信，随着集成电路算力和资源的不断提升，深度学习在粒子鉴别领域的应用也会越来越成熟，其在空间环境探测领域的应用也会逐渐走向前台。

参考文献（References）

[1] 朱金涛. 基于脉冲波形分析的带电粒子鉴别方法研究[D]. 长沙: 国防科学技术大学, 2014.
ZHU J T. Research on charged particle identification method based on pulse waveform analysis[D]. Changsha: National University of Defense Technology, 2014 (in Chinese).

[2] 朱金涛, 刘国福, 杨云, 等. 带电粒子鉴别方法的发展与现状[J]. 核电子学与探测技术, 2014, 34(2): 194-199.
ZHU J T, LIU G F, YANG Y, et al. Development of charged particle identification methods[J]. Nuclear Electronics & Detection Technology, 2014, 34(2): 194-199(in Chinese).

[3] 孔令高, 张爱兵, 田峥, 等. 自主火星探测高集成离子与中性粒子分析仪[J]. 深空探测学报, 2019, 6(2): 142-149.
KONG L G, ZHANG A B, TIAN Z, et al. Integrated ion and neutral particle analyzer for Chinese Mars mission[J]. Journal of Deep Space Exploration, 2019, 6(2): 142-149(in Chinese).

[4] 张惠鸽, 王哲斌, 张文海. 粒子鉴别技术简介[J]. 核电子学与探测技术, 2010, 30(11): 1473-1479.
ZHANG H G, WANG Z B, ZHANG W H. The brief introduction of particle identification[J]. Nuclear Electronics & Detection Technology, 2010, 30(11): 1473-1479(in Chinese).

[5] MAHATA K, SHRIVASTAVA A, GORE J A, et al. Particle identification using digital pulse shape discrimination in a nTD silicon detector with a 1 GHz sampling digitizer[J]. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2018, 894: 20-24.

[6] 侯东辉, 张珅毅, 杨祎罡, 等. 基于CLYC闪烁体的中子能谱测量及反演方法[J]. 北京航空航天大学学报, 2021, 47(1): 106-114.
HOU D H, ZHANG S Y, YANG Y G, et al. Neutron measurement and inversion based on CLYC scintillator[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(1): 106-114 (in Chinese).

[7] WANG Q B, TUO X G, DENG C, et al. Characterization of a Cs 2 LiYCl 6: Ce 3 + scintillator coupled with two silicon photomultiplier arrays of different sizes[J]. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2019, 942: 162339.

[8] 艾鹏程. 基于卷积神经网络的高能物理事例特征信息在线提取算法研究[D]. 武汉: 华中师范大学, 2020.
AI P C. On-line extraction algorithm of high-energy physics case feature information based on convolutional neural network[D]. Wuhan: Central China Normal University, 2020(in Chinese).

[9] 原永朋, 游大涛, 武相军, 等. 类内类间距离在CNN识别心拍类中的应用研究[J]. 计算机工程与应用, 2019, 55(14): 242-248.
YUAN Y P, YOU D T, WU X J, et al. Application of intra-class and inter-class distance in CNN recognition of heart beat[J]. Computer Engineering and Applications, 2019, 55(14): 242-248(in Chinese).

[10] FOBAR D, PHILLIPS L, WILHELM A, et al. Considerations for training an artificial neural network for particle type identification[J]. IEEE Transactions on Nuclear Science, 2021, 68(9): 2350-2357.

[11] MA Y F, CAO Y, VRUDHULA S, et al. Optimizing loop operation and dataflow in FPGA acceleration of deep convolutional neural networks[C]//Proceedings of the ACM/SIGDA International Symposium on Field-Programmable Gate Arrays. New York: ACM, 2017: 45-54.

[12] PAUSCH G, MOSZYŃSKI M, WOLSKI D, et al. Application of the pulse-shape technique to proton-alpha discrimination in Si-detector arrays[J]. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 1995, 365(1): 176-184.

[13] PAUSCH G, ORTLEPP H G, BOHNE W, et al. Identification of light charged particles and heavy ions in silicon detectors by means

of pulse-shape discrimination[J]. IEEE Transactions on Nuclear Science, 1996, 43(3): 1097-1101.

[14] PAUSCH G, BOHNE W, FUCHS H, et al. Particle identification in solid-state detectors by exploiting pulse shape information[J]. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 1992, 322(1): 43-52.

[15] QUARANTA A A, MARTINI M, OTTAVIANI G. The pulse shape and the timing problem in solid state detectors——a review paper [J]. IEEE Transactions on Nuclear Science, 1969, 16(2): 35-61.

[16] 朱金涛, 刘国福, 杨俊, 等. 带电粒子鉴别的脉冲波形分析法比较研究[C]//第十七届全国核电子学与核探测技术学术年会论文集. 兰州: 中国电子学会, 2014: 253-258.
ZHU J T, LIU G F, YANG J, et al. Pulse shape analysis comparison research of charged particle Identification[C]//Proceedings of the 17th National Academic Annual Conference on Nuclear Electronics and Nuclear Detection Technology. Lanzhou: Chinese Institute of Electronice, 2014: 253-258(in Chinese).

[17] DROZ D, TYKHONOV A, WU X, et al. A neural network classifier for electron identification on the DAMPE experiment[J]. Journal of Instrumentation, 2021, 16(7): P07036.

[18] ASTRAIN M, RUIZ M, STEPHEN A V, et al. Real-time implementation of the neutron/gamma discrimination in an FPGA-based DAQ MTCA platform using a convolutional neural network[J]. IEEE Transactions on Nuclear Science, 2021, 68(8): 2173-2178.

[19] LU J, TUO X G, YANG H C, et al. Pulse-shape discrimination of SiPM array-coupled CLYC detector using convolutional neural network[J]. Applied Sciences, 2022, 12(5): 2400.

Spaceborne particle identification platform and application based on convolutional neural network

BAI Chaoping[1, 2]，ZHANG Shenyi[1, 2, *]，ZHANG Xin[1]，SUN Yueqiang[1]，ZHANG Shuai[1]，WANG Ziting[1, 2]

(1. National Space Science Center，the Chinese Academy of Sciences，Beijing 100190，China；
2. University of Chinese Academy of Sciences，Beijing 100049，China)

Abstract: Accurate identification of space radiation particles is crucial for both scientific research and engineering applications. Existing particle identification methods, including detector telescope methods, electrostatic analysis time-of-flight methods, time-of-flight energy methods, and waveform analysis energy methods, have achieved good results in practical applications. However, by leveraging the powerful feature extraction and classification capabilities of convolutional neural networks (CNN), it is expected to further enhance the precision of particle energy measurement and species identification. Based on common space environment detection payloads, this paper proposes a method to build an on-orbit CNN-based particle identification platform for particle species identification. The platform first constructs a multidimensional input dataset, with model training and weight extraction completed through software platforms, and waveform inference and dataset expansion carried out on the hardware platform. The established identification platform is used to train and test neutron and gamma waveform data obtained from actual tests, and the identification accuracy of both the software and hardware platforms is analyzed, completing the platform verification. The establishment and application of this identification platform provide a new approach and method for future particle measurement and identification in space environment detection, with significant engineering practical implications.

Keywords: particle detection; convolutional neural network; spaceborne FPGA platform; pulse waveform identification

Received：2023-04-10；**Accepted**：2023-07-21；**Published Online**：2023-08-18 11：03

URL：link.cnki.net/urlid/11.2625.V.20230817.1606.001

Foundation item：National Natural Science Foundation of China (42204180)

*** Corresponding author**. E-mail：zsy@nssc.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0218

基于时频重排算法的LPI雷达信号识别

姬丽彬[1,2]，朱岩[1,*]，崔天舒[3]，王栋[1,2]，黄永辉[1]

(1. 中国科学院国家空间科学中心 复杂航天系统电子信息技术重点实验室，北京 100190；
2. 中国科学院大学 计算机科学与技术学院，北京 100049；3. 清华大学 北京信息科学与技术国家研究中心，北京 100084)

摘 要：针对低截获概率(LPI)雷达信号在低信噪比(SNR)情况下识别效果不理想、网络模型复杂的问题，提出一种基于时频重排和多尺度残差网络的LPI雷达信号识别方法。该方法以魏格纳-维尔分布(WVD)为基础，通过时频重排算法提高信号的聚集性，得到信号的时频分布图像，输入到多尺度残差网络中完成信号的分类。通过构建多径莱斯衰落信道完成复杂电磁环境仿真，实验结果表明：所提方法在SNR为−8 dB时，对Costas、Frank、P1～P4等13类LPI雷达典型调制样式能达到94%的识别准确率，相比其他方法在低信噪比下具有更好的识别性能。

关 键 词：低截获概率雷达；信号识别；时频重排；多尺度残差网络；调制样式
中图分类号：TN974
文献标志码：A **文章编号**：1001-5965(2025)04-1324-08

作为新体制雷达的典型代表，低截获概率(low probability of intercept, LPI)雷达信号[1]得益于其功率小、隐蔽性高、拦截率低、抗干扰能力强[2]等特性，近年来被广泛用于舰载防空雷达、机载火控雷达。在电子对抗侦察中，电子侦察系统被视为军事作战的“千里眼”，检测和截获信号的能力是评估电子侦察系统发展进程的重要特征之一。在越来越恶化的电磁环境下，自动高效的识别低截获概率信号是现代电子侦察方向一个重要研究课题，对于战场成败具有现实意义。

雷达辐射源脉内调制方式识别的方法，基本可分为特征提取和分类器设计两大步。非平稳信号所使用的特征主要包括时频变换和高阶统计分析，其中，时频变换大多建立在魏格纳-维尔分布(Wigner-Ville distribution, WVD)的基础上。文献[3]利用短时傅里叶变换(short-time Fourier transform, STFT)和*k*-means算法进行信号特征分析的频谱计算，将STFT作为特征用于信号的识别。文献[4]使用崔-威廉姆斯分布(Choi-William distribution, CWD)变换将雷达信号波形转换为时频图像，并从时频图像中提取定向梯度(histogram of oriented gradients, HOG)特征和深度特征进行特征融合，进而完成对雷达信号的分类。文献[5]基于小波变换提取信号时域特征、模糊函数切片特征及循环频谱特征，并通过特征融合处理完成发射端的雷达信号分类。然而上述方法大多在信噪比(signal-to-noise ratio, SNR)高于0 dB时，识别精度较高，随着信噪比降低分类准确率会出现骤降的情况，鲁棒性较差。由于深度学习强大的特征学习能力，越来越多的研究者将其应用于雷达信号调制方式识别方面。文献[6]由残差学习卷积去噪自编码器(residual learning convolutional denoising autoencoder, RLCDAE)和基于噪声级估计(noise level estimation, NLE)的辅助分类网络构成X-net，用于强噪声背景下辐射信号的检测。文献[7]基于一维CNN网络模型，采用伪标签和自步调数据增强，提高了脉内分类的准确性。文献[8]引入一种具有2个不同网络——进化神经网络(evolutionary neural networks, ENN)和卷积神经网络

收稿日期：2023-04-27；**录用日期**：2023-07-21；**网络出版时间**：2023-08-01 09:27
网络出版地址：link.cnki.net/urlid/11.2625.V.20230731.1532.001
基金项目：中国科学院复杂航天系统电子信息技术重点实验室自主部署基金(Y42613A32S)
*通信作者. E-mail：zhuyan@nssc.ac.cn

引用格式：姬丽彬，朱岩，崔天舒，等. 基于时频重排算法的LPI雷达信号识别[J]. 北京航空航天大学学报，2025，51(4)：1324-1331.
JI L B，ZHU Y，CUI T S，et al. LPI radar signal recognition based on time-frequency reassignment algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1324-1331 (in Chinese).

(convolutional neural networks, CNN) 的混合分类器完成自动雷达波形识别系统。尽管深度学习能够对多种雷达信号实现有效的分类, 但现阶段的研究工作大多为了提取信号的细微特征而设计较深的网络模型, 使得算法难以在实际战场环境部署。

本文从信号的时频特性出发, 提出一种基于时频重排算法和平滑伪魏格纳-威尔分布 (smoothed pseudo Wigner-Ville distribution, SPWVD) 的 LPI 雷达信号识别方法, 充分利用重排算法提高时频聚集性的特点与 SPWVD 抑制交叉项的特性, 将重排后的 SPWVD 时频图像作为模型的输入, 利用多尺度残差网络作为分类模型, 对 13 类常见 LPI 雷达信号进行有效分类识别, 在信噪比大于−8 dB 的情况下, 整体平均识别率也可达到 94% 以上。

1　雷达信号特征提取

1.1　时频特征分析

时频变换是分析诸如雷达信号等非平稳信号的有力工具, 时频变换可以用来估计信号在时域和频域的联合分布情况, 表征时间和频率之间的关系。常用的时频分析包括 STFT、WVD、CWD 等。WVD 是较早提出的用于时频分析的方法, 其摆脱了窗函数的束缚, 同时还克服了短时傅里叶变换的缺点, 信号 $s(t)$ 的 WVD 分布定义为[9]

$$\text{WVD-s}(t,f)=\int_{-\infty}^{+\infty}s(t+\tau/2)s^{*}(t-\tau/2)e^{(-\mathrm{j}2\pi f\tau)}\mathrm{d}\tau \tag{1}$$

式中: t 为时间; f 为频率; $s(\cdot)$ 为时间信号; "*" 表示复共轭; τ 为时延。

为陈述时频重排的原理, 使用二次型时频分布[10] 的表达式:

$$\text{QTFD-s}(t,f)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\phi(t',f')\text{WVD-s}(t-t',f-f')\mathrm{d}t'\mathrm{d}f' \tag{2}$$

式中: 核函数 $\phi(t,f)$ 为时频平面的二维低通滤波器。

由式 (2) 可以看出, 时频平面任意点 (t,f) 的值是该点的邻近点 $(t-t',f-f')$ 上的原信号的 WVD-s(t,f) 经过核函数加权后的平均值。时频重排算法是通过将代表信号局部能量分布的几何中心重新定位到时频分布的能量的中心, 从而提高时频图像的聚集性和可读性[11]。时频的重排引起对时频分布的重构, 移动后重心对应的坐标[12] 为

$$\hat{t}(t,f)=t-\frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}t'\phi(t',f')\text{WVD-s}(t-t',f-f')\mathrm{d}t'\mathrm{d}f'}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\phi(t',f')\text{WVD-s}(t-t',f-f')\mathrm{d}t'\mathrm{d}f'} \tag{3}$$

$$\hat{f}(t,f)=f-\frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f'\phi(t',f')\text{WVD-s}(t-t',f-f')\mathrm{d}t'\mathrm{d}f'}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\phi(t',f')\text{WVD-s}(t-t',f-f')\mathrm{d}t'\mathrm{d}f'} \tag{4}$$

信号的 WVD 在时频平面的聚集性最佳, 但是会导致大量交叉项的产生, 对于相位调制信号而言, 跳频幅度较小, 而交叉项会导致细微跳频特征不清晰, 进而影响相位调制信号的识别。SPWVD 算法是对 WVD 算法进行时频域平滑的改进算法[13], 可以独立地控制在时间和频率方向上的平滑, 因为其具有可分离的核函数 $\phi(t,f)=g(t)H(f)$, 则可得 SPWVD 算法的表达式为

$$\text{SPWVD-s}(t,f)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(u)h(\tau)\cdot s(t-u+\tau/2)s^{*}(t-u-\tau/2)e^{(-\mathrm{j}2\pi f\tau)}\mathrm{d}u\mathrm{d}\tau \tag{5}$$

式中: $g(u)$ 为抑制时域方向交叉项的窗函数; $h(\tau)$ 为抑制频域方向交叉项的窗函数。该算法使用窗函数对时域和频域进行滤波, 很好地抑制了交叉项的产生, 但是随着平滑次数的增加, 时频分辨率显著降低。

对于 SPWVD 而言, 其重排的坐标为

$$\hat{t}(t,f)=t-\frac{\text{SPWVD-Tg-h}(t,f)}{\text{SPWVD-g-h}(t,f)} \tag{6}$$

$$\hat{f}(t,f)=f+\mathrm{j}\frac{\text{SPWVD-g-Dh}(t,f)}{\text{SPWVD-g-h}(t,f)} \tag{7}$$

式中: T_g 为乘积算子; D_h 为微分算子。分别定义为 $T_g(t)=tg(t)$ 和 $D_h(t)=\dfrac{\mathrm{d}}{\mathrm{d}t}h(t)$。重排后的时频分布在任意点 (t',f') 处的值是所有重排到这一点的值的和, 时频重排后的 SPWVD 定义如下所示:

$$\text{RSPWVD-s}(t,f)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\text{SPWVD}(t,f)\cdot\delta(t'-\hat{t}(t,f))\delta(f'-\hat{f}(t,f))\mathrm{d}t'\mathrm{d}f' \tag{8}$$

式中: $\delta(t)$ 为狄拉克冲击函数。通过使用快速傅氏变换算法, 每个重排的时频分布就可以被高效计算, 而不会产生很大的计算负担。任何重排分的脊可以利用脊和时频算子之间的关系得到, 提取时频重排后 SPWVD 的脊, 通过在重排后的脊上选择合适的时频点可以有效地筛选出所需信号, 重排后信号的时频聚集性更强, 信号所在时频域的脊和轮廓都更为明显, 同时还大大降低了噪声信号的干扰, 从而保证了不同调制类型信号的分辨能力, 进而提高了算法的抗噪声性能。

图 1 为信噪比为 10 dB 下 13 类 LPI 信号重排平滑伪维格纳维尔分布 (reassigned smoothed pseudo Wigner-Ville distribution, RSPWVD) 时频图像, 可以

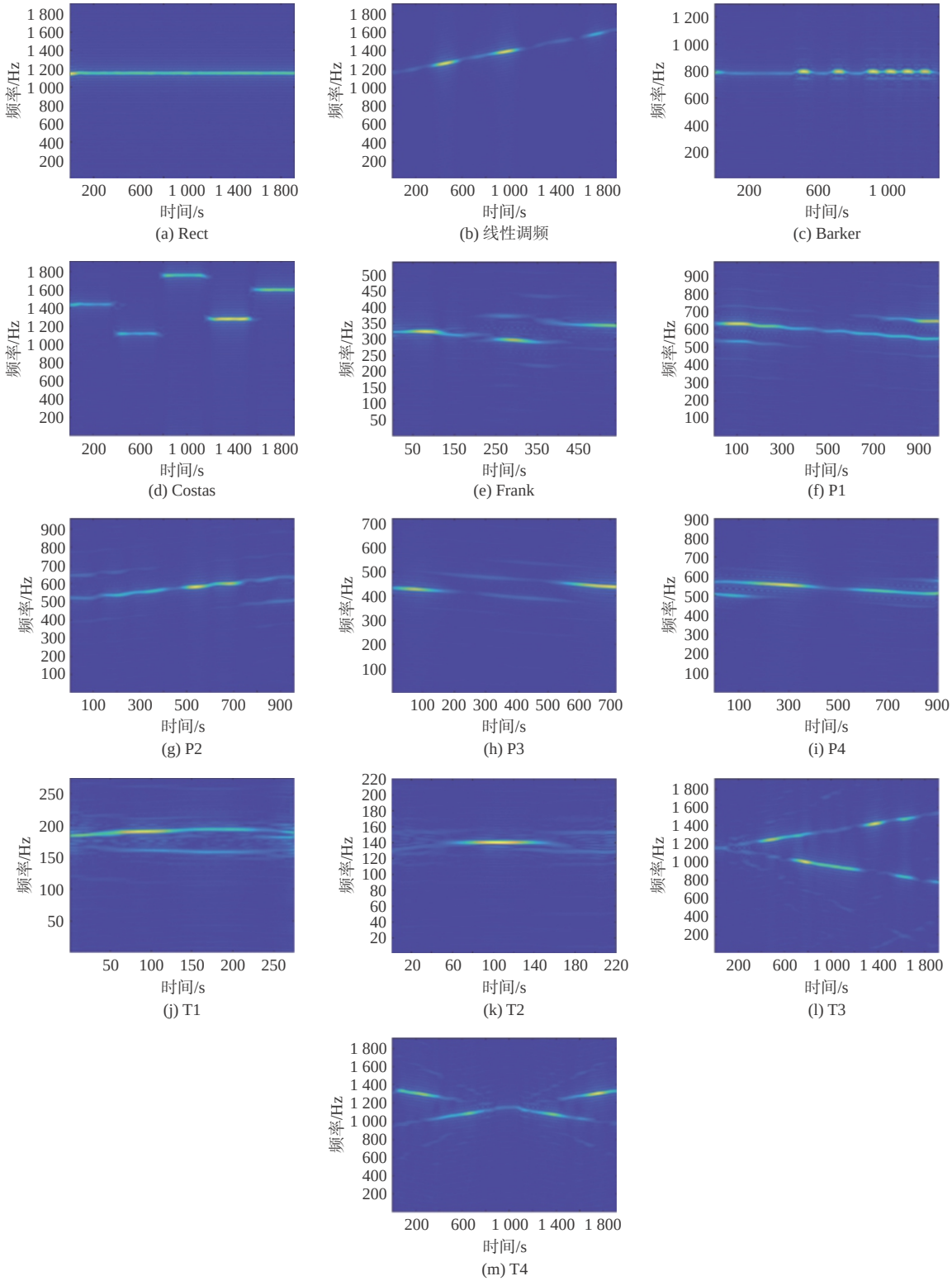


图1 SNR为10 dB时13种雷达信号的RSPWVD时频图像

Fig. 1 RSPWVD time-frequency images of 13 kinds of radar signals with SNR of 10 dB

看出调制方式不同的LPI雷达信号的特征差距较为显著。

1.2 时频图像预处理

雷达信号的原始时频图为RGB图像，而时频图最明显的特征是形状和数值。为区分不同信号，灰度图像足以显示这些特征。除此之外，与由R(红)、G(绿)、B(蓝)3个分量组成的RGB图像相比，灰度图像只有一个分量，彩色向灰度的转换可以大大缩小输入数据，大大降低计算成本。因此，信号经过时频处理后还需进行灰度化、二值化，以及大小的调整，再输入到分类网络进行识别。以0 dB下线性调频信号的RSPWVD时频图像为例，

处理过程如图 2 所示, 具体步骤如下。

步骤 1　通过计算 R、G 和 B 分量的加权和, 将原始 RGB 时频图像转换为灰度图像, 减少通道数。

步骤 2　使用全阈值二值法[14]将灰度图像转换为二值图, 进一步抑制噪声和降低计算量。

步骤 3　利用双三次插值运算将时频图像大小缩减为 128 像素×128 像素, 进一步减少数据量。

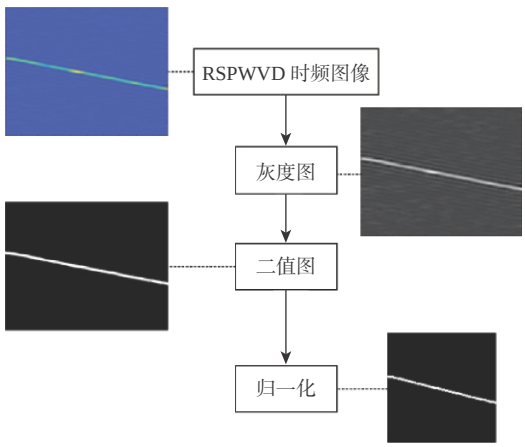


图 2　图像预处理流程

Fig. 2　Flow chart of image preprocessing

2　基于多尺度残差网络的信号识别方法

2.1　多尺度残差网络构建

卷积神经网络因其出色的识别能力而被广泛用于图像和模式识别。CNN 能够跳过复杂的数据预处理, 自动提取信号特征, 拥有更强大的非线性表征能力[15]。

与传统的神经网络相比, CNN 善于处理图像等高维数据。自 2012 年 AlexNet 被提出之后, CNN 层数不断加深, 网络拟合能力理应越来越强, 但越深的网络反而训练误差越高, 网络能力出现退化, 识别率反而降低。残差网络由于其独特的跳跃连接设计, 提供了一定程度的信息完整性, 避免了在信息传递过程中, 传统卷积出现的信息丢失问题, 避免了网络过深导致的性能下降问题[16], 进一步提高了目标分类和识别性能。因此, 本文选择以残差结构为主体, 设计出适合雷达信号识别的深度学习网络模型。

图 3 给出了用于 LPI 雷达波形识别的多尺度残差网络模型, 主要基于非对称卷积核提取多尺度特征信息的思想。在捕获粗特征阶段, 网络由 1 个 3×3 内核大小的卷积 (conv) 层和 2 个 3×1 和 1×3 非对称卷积核组成, 与直接使用大核卷积相比, 非对称卷积核可以减少训练参数的数量, 但不会明显降低提取的特征质量。3×1 和 1×3 这 2 个卷积核并行, 用于多个尺度显式地提取特征。网络的核心组成部分为图 3 右侧所示的相同结构的 3 个处理模块, 这些处理模块级联排列, 以获得多尺度分辨率下的高分辨特征。每个模块都从 max-pooling 层开始, 以降低输入特征图的空间维数, 之后由 2 个具有非对称内核的 conv 层 (即 1×3 和 3×1) 并行排列, 几乎保持了与对称卷积相当的学习效率, 两层特征映射的输出构成深度叠加层通过 Depth concat 连接

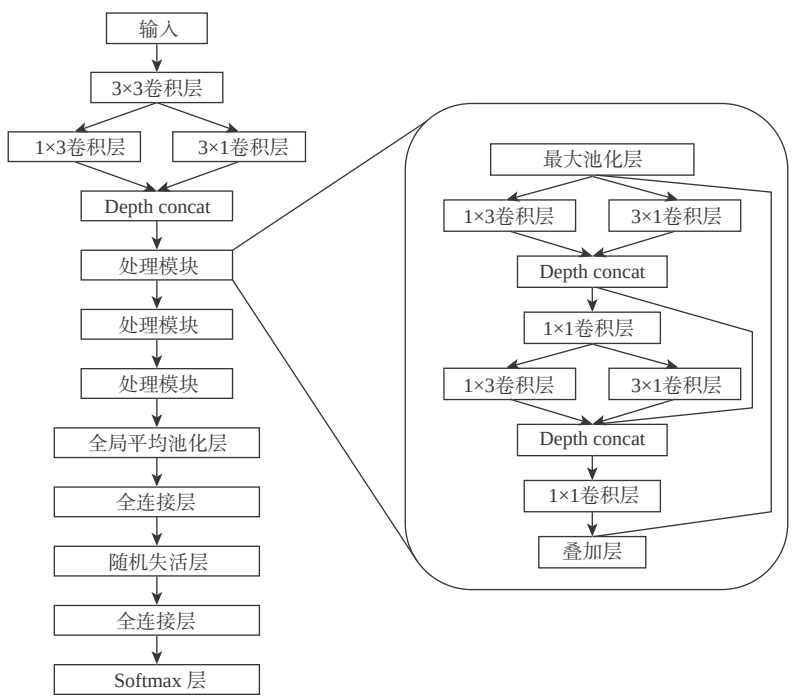


图 3　多尺度残差网络模型

Fig. 3　Multi-scale residual network model

产生输出，之后所连接的1×1单元conv层被用来控制特征的深度大小，这种跨通道参数池结构允许跨通道信息的交互。第2个Depth concat层前由类似的结构组成。Depth concat层之间采用短连接，以避免多个层堆叠造成梯度消失，同时使模块能够学习剩余信息，而不是学习规则结构中的真实输出。特征提取层中每一个卷积层后面都连接一个批量归一化 (batch normalization, BN) 和指数线性单元 (exponential linear unit, eLU)。最后一个模块的输出被传递到全局平均池化 (global average pooling, GAP) 层，通过计算输入映射的每个通道的空间维度上的平均值执行下采样，将多个特征映射到一个低维空间中。在分类网络中，GAP层通常位于全连接 (fully connected, FC) 层之前，以减少参数数量供下层学习[17]。最后的Softmax层用于分类，获得13类LPI雷达信号。

本文所设计的多尺度残差模型通过利用非对称卷积和跳跃连接的思想，使其拥有更为强大的特征提取性能，同时，结合时频分析图像的特点，优化模型深度及输出特征图维度，一定程度上缓解了“过拟合”问题及维度灾难问题。

2.2 信号识别流程

本文构建的基于时频重排和多尺度残差网络的LPI雷达信号识别方法流程如图4所示。具体步骤如下。

1）对LPI雷达信号作RSPWVD时频变换处理，获取信号相应的时频图像。

2）对时频图像完成预处理流程，抑制噪声的同时减少数据的大小。

3）构建多尺度残差网络模型。将预处理后的图像划分为训练集、验证集和测试集。将训练数据送入构建好的卷积神经网络完成训练，并从验证集中抽取图像进行验证。

4）信号识别。将测试数据送入训练好的分类网络中，得到信号的调制类型。

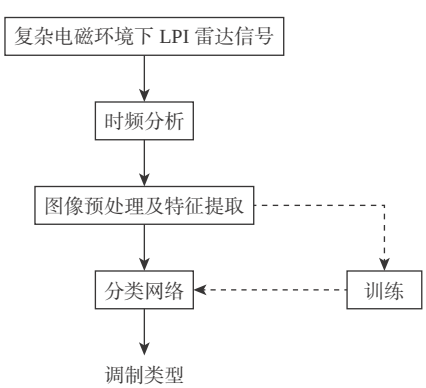


图4 LPI雷达信号调制识别流程

Fig. 4 LPI radar signal modulation recognition process

3 实验结果与分析

3.1 实验条件

硬件配置。Intel(R) Core(TM) i7-7700K CPU@4.20 GHz处理器，16GB内存，NVIDIA GeForce GTX 1080。

数据集。主要面向LPI雷达脉内调制信号。这类雷达大多采用频率编码、相位编码等脉冲压缩技术去完成低截获特性的要求，因此，考虑了13种常见LPI雷达波形：线性调频、矩形、Costas码、Barker码、5种多相码 (Frank、P1、P2、P3和P4) 和4种多时码 (T1、T2、T3和T4)。各信号具体的参数设置如表1所示。网络训练时选用交叉熵损失函数，SGDM优化算法，初始学习率为0.01，epoch取值为12，batchsize设置为128。

表1 波形参数设置

Table 1 Waveform parameters setting

LPI信号调制样式	调制参数	取值范围
All	载频f_c/Hz	$U(f_s/6, f_s/5)$
线性调频	带宽B/Hz	$U(f_s/20, f_s/15)$
	采样个数N	[512, 1 024]
Rect	采样个数N	[512, 1 024]
Costas	跳频序列长度	{3, 4, 5, 6}
	基础频率f_{min}/Hz	$U(f_s/32, f_s/25)$
	采样个数N	[512, 1 024]
Barker	码长L	{7, 11, 13}
	相位子波数c_{pp}	[2, 5]
Frank	相位子波数c_{pp}	[3, 5]
	频率步长M	{6, 7, 8}
P1,P2	相位子波数c_{pp}	[3, 5]
	频率步长M	{6, 8}
P3,P4	相位子波数c_{pp}	[3, 5]
	子码数n_s	{36, 64}
T1,T2	相位状态数N_p	2
	相位波形段数N_g	{4, 5, 6}
	采样个数N	[512, 1 024]
T3,T4	相位状态数N_p	2
	相位波形段数N_g	{4, 5, 6}
	采样个数N	[512, 1 024]
	调制带宽B/Hz	$U(f_s/20, f_s/15)$

该领域的大多数研究[15,18-19]只考虑加性高斯白噪声 (additive white Gaussian noise, AWGN) 对信号产生的影响，与实际信道环境相差甚远。本文在生成数据集时，叠加范围为 $[f_s/6, f_s/5]$ 的载频偏移于信号，并在AWGN的基础上叠加莱斯衰落信道去模拟真实信号的发送接收，具体信道参数如表2所示。数据集考虑了26种不同的信噪比环境，从−15 dB到10 dB，步长为1 dB。每1 dB产生1 000个样本，其中700个 (70%) 作为训练集，150个 (15%) 作为验证集，150个 (15%) 作为测试集。

3.2 实验结果分析

为了验证本文方法在LPI雷达信号识别的有

表 2　信道参数设置

Table 2　Channel parameters setting

信道参数	取值范围
采样频率/Hz	f_s
多径衰减/ns	U(1,1 000)
平均路径增益/dB	U(−20,0)
K-factor	4
最大多普勒频移/Hz	U(10, 1 000)

效性, 设置 2 个综合实验评估对 LPI 雷达波形识别的效果。

1) 实验 1。该实验探究所提出的深度网络获得的 13 种 LPI 波形的识别精度。从图 5 的结果可见, 识别率随着信噪比的增加而增加。部分调制波形, 包括 LFM, Costas 和 P2, 在低信噪比下实现非常高的识别精度, 在−10 dB 信噪比下准确率约为 97.0%。值得注意的是, Costas 和 LFM 在−8 dB 信噪比下达到了 100% 的正确识别率, 因为其时频图像比其他波形更具鉴别性。该识别模型在−4 dB 信噪比下的整体准确率达到 98.8% 左右。为了更细致地分析, 图 6 提供了−8 dB 信噪比下 13 类波形识

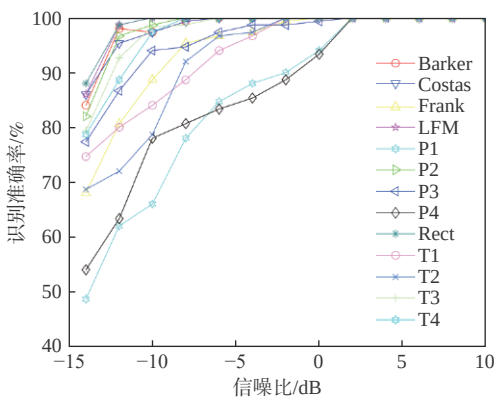


图 5　13 种雷达调制方式的识别准确率

Fig. 5　Recognition accuracy of 13 radar modulation modes

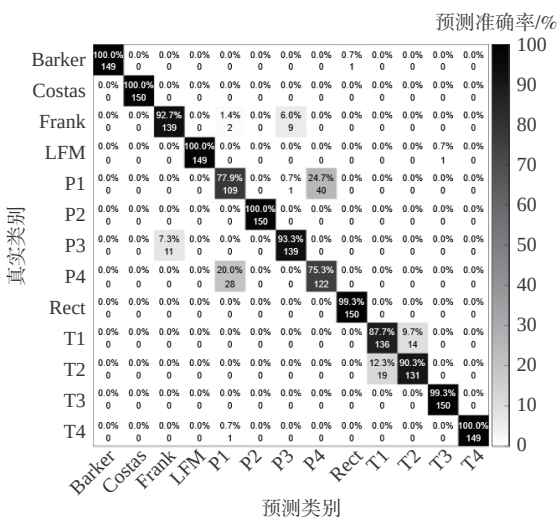


图 6　信号的混淆矩阵(−8 dB)

Fig. 6　Signal confusion matrix (−8 dB)

别的混淆矩阵, Barker、Costas、LFM、P2、T4 码 5 种信号的识别率达到 100%。但由于信道恶化, 多径衰落和加性噪声导致多相码 P1 和 P4 的严重误分类 (约 20.0%～24.7%)。

2) 实验 2。时频变换方面, 在同等条件下比较了本文方法和 CWD、SPWVD、WVD 与 MSST[20] 的识别效果。运行时间对比如表 3 所示, 本文方法基于 SPWVD 改进, 相较 SPWVD 而言运行时间略有上升。对于应用最为广泛的 CWD 方法, 单个信号执行一次耗时需要 0.307 2 s 左右, 而执行本文方法仅需 0.269 s 左右, 运算速度更快, 在保持较好实时性的情况下提高了识别准确率。6 种时频分析方法均采用设计的多尺度残差网络对数据集进行识别, 如图 7 所示, 本文方法具有较强的抑制干扰的高分辨率谱图平滑能力, 比其他算法更适合分析雷达波形。重排后的时频图像要比 SPWVD 有更好的识别效果, 当信噪比低于 0 dB 时, WVD 的识别概率高于 STFT 和 CWD。WVD 表现出良好的性能, 但随着信噪比从−2 dB 开始下降后, 精度迅速下降。MSST 在处理 T 码时效果较差进而大大影响了总体识别率。总体而言, 本文方法的识别准确率在信噪比较低的情况下也表现出较好的识别性能。

表 3　运行时间对比

Table 3　Comparison of running time

时频方法	时间/s
本文方法	0.269 3
CWD	0.307 2
WVD	0.039 2
STFT	0.064 6
SPWVD	0.264 1
MSST	0.199 3

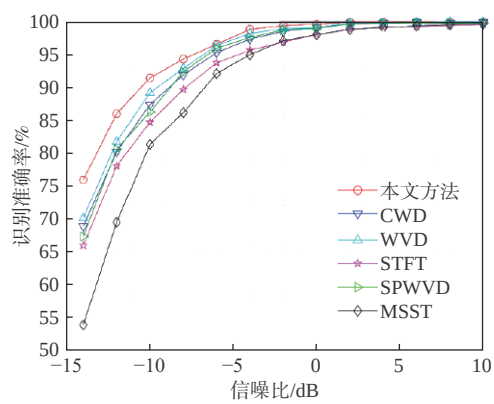


图 7　6 种方法识别准确率对比

Fig. 7　Comparison of recognition accuracy of six methods

4　结　论

针对多径条件下雷达信号时频聚集性骤降, 识

别准确率不高的问题，本文提出一种基于时频重排的LPI雷达信号识别方法，主要结论如下。

1）在SPWVD时频分析的基础上引入时频重排算法，在降低交叉项的干扰的同时高效提高时频聚集性，解决了高噪声环境下时频聚集性较低的问题。

2）在分类网络中，将多尺度残差网络模型应用于LPI雷达信号识别，借鉴残差网络ResNet的思想设计了卷积核大小、并行连接及模块间的跳连方式。

3）实验结果表明，与传统方法相比，本文方法优于其他时频分析方法，同时还具有更高的准确率和抗噪能力，在−8 dB下信号分类准确率也可达到94%。

实际战场环境中，信号复杂多样，后续还需考虑脉冲丢失、虚假脉冲等干扰现象，以及混合调制信号的情况。

参考文献（References）

[1] PACE P E. Detecting and classifying low probability of intercept radar[M]. 2nd ed. Boston: Artech House, 2009.

[2] LIU Y J, XIAO P, WU H C, et al. LPI radar signal detection based on radial integration of Choi-Williams time-frequency image[J]. Journal of Systems Engineering and Electronics, 2015, 26(5): 973-981.

[3] XIAO Z L, YAN Z Y. Radar emitter identification based on novel time-frequency spectrum and convolutional neural network[J]. IEEE Communications Letters, 2021, 25(8): 2634-2638.

[4] QUAN D Y, TANG Z Y, WANG X F, et al. LPI radar signal recognition based on dual-channel CNN and feature fusion[J]. Symmetry, 2022, 14(3): 570.

[5] WANG X Z. Electronic radar signal recognition based on wavelet transform and convolution neural network[J]. Alexandria Engineering Journal, 2022, 61(5): 3559-3569.

[6] CHEN K Y, ZHANG J Y, CHEN S, et al. Automatic modulation classification of radar signals utilizing X-net[J]. Digital Signal Processing, 2022, 123: 103396.

[7] YUAN S B, LI P, WU B, et al. Semi-supervised classification for intra-pulse modulation of radar emitter signals using convolutional neural network[J]. Remote Sensing, 2022, 14(9): 2059.

[8] ZHANG M, DIAO M, GAO L P, et al. Neural networks for radar waveform recognition[J]. Symmetry, 2017, 9(5): 75.

[9] WANG G, CHEN S, HUANG J, et al. Radar signal sorting and recognition based on transferred deep learning[J]. Computer Science and Application, 2019, 9(9): 1761-1778.

[10] BOASHASH B. Time frequency signal analysis and processing: a comprehensive reference[M]. Amsterdam: Elsevier, 2003.

[11] 杨瑾, 郝新红, 陈齐乐. 基于多域特征的多雷达信号自动识别方法[J]. 北京航空航天大学学报, 2024, 50(3): 931-939.
YANG J, HAO X H, CHEN Q L. Automatic recognition method of multi-radar signals based on multi-domain features[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(3): 931-939(in Chinese).

[12] AUGER F, FLANDRIN P. Improving the readability of time-frequency and time-scale representations by the reassignment method[J]. IEEE Transactions on Signal Processing, 1995, 43(5): 1068-1089.

[13] WANG Y, WU X, LI W Z, et al. Analysis of micro-Doppler signatures of vibration targets using EMD and SPWVD[J]. Neurocomputing, 2016, 171: 48-56.

[14] ATAIE R, EMRANI ZARANDI A A, SAFAEI MEHRABANI Y. An efficient inexact full adder cell design in CNFET technology with high-PSNR for image processing[J]. International Journal of Electronics, 2019, 106(6): 928-944.

[15] 崔天舒, 崔凯, 黄永辉, 等. 卷积神经网络卫星信号自动调制识别算法[J]. 北京航空航天大学学报, 2022, 48(6): 986-994.
CUI T S, CUI K, HUANG Y H, et al. Convolutional neural network based algorithm for automatic modulation recognition of satellite signals[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(6): 986-994(in Chinese).

[16] 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, 2023, 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 Astronautics, 2023, 49(2): 254-264(in Chinese).

[17] GAO J P, LU Y, QI J W, et al. A radar signal recognition system based on non-negative matrix factorization network and improved artificial bee colony algorithm[J]. IEEE Access, 2019, 7: 117612-117626.

[18] KONG S H, KIM M, HOANG L M, et al. Automatic LPI radar waveform recognition using CNN[J]. IEEE Access, 2018, 6: 4207-4219.

[19] HOANG L M, KIM M, KONG S H. Automatic recognition of general LPI radar waveform using SSD and supplementary classifier[J]. IEEE Transactions on Signal Processing, 2019, 67(13): 3516-3530.

[20] 王功明, 陈世文, 黄洁, 等. 基于多重同步压缩变换的雷达辐射源分选识别[J]. 现代雷达, 2020, 42(3): 49-56.
WANG G M, CHEN S W, HUANG J, et al. Radar emitter sorting and recognition based on multi-synchrosqueezing transform[J]. Modern Radar, 2020, 42(3): 49-56(in Chinese).

LPI radar signal recognition based on time-frequency reassignment algorithm

JI Libin[1, 2], ZHU Yan[1, *], CUI Tianshu[3], WANG Dong[1, 2], HUANG Yonghui[1]

(1. Key Laboratory of Electronic and Information Technology for Complex Space Systems, National Space Science Center, Chinese Academy of Science, Beijing100190, China;
2. School of Computer Science and Technology, University of Chinese Academy of Science, Beijing 100049, China;
3. Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: In view of the problems of low probability of acquisition (LPI) radar signal recognition in low signal-to-noise ratio (SNR) and complex network model, an LPI radar signal recognition method based on time-frequency reassignment and multi-scale residual network was proposed. The time-frequency reassignment approach is used to enhance the signal's aggregation based on the Wigner-Ville distribution (WVD). The resulting time-frequency distribution image is then fed into the multi-scale residual network to finish the signal's categorization. In addition, the complex electromagnetic environment simulation was completed by constructing a multi-path Rice-fading channel. According to the experimental results, when the SNR is −8 dB, the suggested approach can achieve 94% recognition accuracy for a total of 13 different types of typical LPI radar modulation patterns, including Costas, Frank, P1～P4, etc. Compared with other methods, it has better recognition performance at a low signal-to-noise ratio.

Keywords: low intercept probability radar; signal recognition; time-frequency reassignment; multi-scale residual network; modulation style

Received: 2023-04-27; **Accepted**: 2023-07-21; **Published Online**: 2023-08-01 09:27

URL: link.cnki.net/urlid/11.2625.V.20230731.1532.001

Foundation item: Independent Deployment Foundation of Key Laboratory of Electronic Information Technology for Complex Space Systems, Chinese Academy of Sciences (Y42613A32S)

* **Corresponding author**. E-mail: zhuyan@nssc.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0227

基于奇异值方法的推力矢量/气动舵飞机复合控制律参数优化方法

阮仕龙[1]，董哲[1,*]，孙尧[2]，曲晓雷[3]，霍少泽[1]

(1. 大连理工大学 航空航天学院，大连 116024；2. 沈阳飞机工业（集团）有限公司，沈阳 110034；
3. 沈阳飞机设计研究所，沈阳 110035)

摘 要：推力矢量/气动舵布局的先进飞机存在显著的控制耦合问题，传统的单回路控制参数设计方法在该场景下无法兼顾多控制回路的性能。因此提出一种基于回差矩阵奇异值的控制律参数优化方法，利用时域控制性能指标确定控制参数优化区间，在此基础上用奇异值方法衡量多输入多输出 (MIMO) 系统的稳定裕度，并建立相应的最优目标函数，从而对控制器参数寻优。采用数值仿真验证了该方法的可行性，结果显示所设计的控制参数优化算法相较于传统单回路控制参数设计方法具有更好的时域控制性能与更大的系统稳定裕度。

关 键 词：回差矩阵奇异值；参数寻优；稳定裕度；MIMO 系统；推力矢量
中图分类号：V249.1
文献标志码：A **文章编号**：1001-5965(2025)04-1332-10

推力矢量/气动舵布局的复合控制飞机具有操纵耦合、系统摄动、外界干扰强等特点[1-7]，飞行控制系统设计需要实现精准、强鲁棒性和强稳定性的功能，以完成相应的机动任务。先进飞机推力矢量/气动舵的冗余配置构成了相互耦合的多回路系统，使得传统单入单出系统(single-in single-out, SISO)的控制设计方法无法有效地满足多输入多输出系统(MIMO)的时域性能要求与频域稳定裕度要求。如何在保证闭环系统期望操稳性能的前提下，实现多回路耦合控制系统的参数优化是需要解决的关键问题。

对于推力矢量/气动舵复合控制飞机，通常有多个执行机构控制同一状态的情况，因此如何进行控制分配是本文需要关注的问题之一。气动舵的舵效受大气动压影响很大，影响动压的两个主要因素包括飞机飞行速度及大气密度，速度越高舵效越高；而推力矢量/燃气舵舵效几乎不受速度影响，但推力矢量/燃气舵的偏转会增加发动机的推力损失，所以需要引入合理的分配策略对复合机构进行控制。为此，国内外的研究人员提出了很多不同的解决控制分配问题的方法，如比例分配方法[8-10]、链式融合分配策略[11]、基于能量的最优分配策略[12]等。复合控制分配策略的设计可以更注重工程上的可实现性，如采用气动舵与推力矢量按常数比例获得虚拟指令。将力矩指令按照一定的常值比例分配到各执行机构上，该方式使得力矩分配具有更强的灵活性，可进行人为的增益调度。对于不同的机动任务和飞行需求，可以人为调节分配的比例，如大迎角大机动任务下可以提高推力矢量的比例，从而提高机动性；常规平飞巡航时可以降低推力矢量喷管变化的幅度，从而提高隐身性能；同时对于故障和损失的情况，可以针对故障位置和损失程度

收稿日期：2023-05-08；**录用日期**：2023-07-16；**网络出版时间**：2023-09-08 11:28
网络出版地址：link.cnki.net/urlid/11.2625.V.20230907.1549.006
基金项目：国家自然科学基金重点项目 (U2141229)；基础加强项目 (2019-JCJQ-DA-001-131)
* **通信作者**. E-mail：zhedong@mail.dlut.edu.cn

引用格式：*阮仕龙，董哲，孙尧，等. 基于奇异值方法的推力矢量/气动舵飞机复合控制律参数优化方法 [J]. 北京航空航天大学学报，2025，51（4）：1332-1341. RUAN S L，DONG Z，SUN Y，et al. Parameter optimization method of thrust vector/pneumatic rudder composite control law for aircraft based on singular value method[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1332-1341（in Chinese）.*

重新进行比例的调配。

针对多变量系统的稳定裕度分析，传统的SISO 系统幅值裕度和相角裕度求解方式无法推广到 MIMO 系统中[13]，而奇异值摄动方法则可以有效度量多变量系统的稳定裕度。MIMO 系统常采用系统回差矩阵最小奇异值的理论来确定系统的稳定裕度[14-18]。结合奇异值理论，通过分析系统回差阵即可得到系统的稳定裕度，且基于回差阵奇异值计算稳定裕度的方法的优点在于计算量小，可实现性强。李信栋等[19]针对两输入两输出系统利用奇异值方法计算 MIMO 系统稳定裕度并分析了稳定裕度计算的保守性。吴斌等[20]采用了改进方法对某型飞机的侧向运动系统进行了稳定裕度的分析和比较。段子帆等[21]针对超燃冲压发动机为动力的飞行器设计了基于极点配置的控制器并利用闭环系统回差奇异值方法评估了控制系统的稳定裕度。李帆[22]利用回差矩阵奇异值方法对提出的几种解耦控制方法进行了分析和比较。回差矩阵奇异值摄动方法实现简单，适用于 MIMO 系统的稳定裕度计算，具有较好的工程实用性和高效性。

采用奇异值摄动方法对多变量系统的稳定裕度进行分析已经具备较为成熟的理论基础，但是利用稳定裕度分析结果对多回路系统的耦合控制参数进行优化设计未得到充分研究。

本文基于系统奇异值摄动方法，面向推力矢量/气动舵布局的复合控制飞机 MIMO 系统，首先利用时域控制性能仿真进行控制器参数范围的确定和筛选，然后利用奇异值摄动方法判断系统的频域稳定裕度，将稳定裕度最大的参数作为最优控制器参数，使得系统的鲁棒性、稳定性和控制性能同时得到保证。

1　推矢作动建模与动力学建模

推力矢量喷管为可变动角度尾喷管，通过改变发动机喷口偏转角度从而实现矢量推力。根据上述物理模型建立推力矢量偏转的数学模型，推力矢量下偏为正。俯仰平面推力矢量偏转的坐标示意图如图 1 所示。

图中，$O_c x_E y_E z_E$为机体坐标系，$O_t x_T y_T z_T$为与机体坐标系方向定义相同的推力坐标系，用以注明推力矢量喷管的偏转情况，其中O_t为矢量推力作用点，而O_c为飞机重心所在位置。设推力矢量喷管下偏，产生矢量推力$\boldsymbol{T}_t$；其在各个坐标轴的分量为T_{tvx}和T_{tvy}。规定二元矢量喷管等效俯仰舵为δ_{tz}。矢量推力作用点O_t相对于重心位置O_c在x、y、z方向的距离分别为l_x、l_y、l_z。

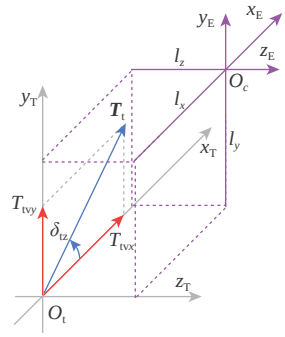


图 1　推力矢量偏转坐标示意图

Fig. 1　Schematic diagram of thrust vector deflection coordinates

由上述示意图可以得到推力矢量偏转时的力模型：

$$\begin{cases} T_{tvx} = T\cos\delta_{tz} \\ T_{tvy} = T\sin\delta_{tz} \end{cases} \tag{1}$$

由图 1 的定义得知，矢量推力作用点O_t在机体坐标系$O_c x_E y_E z_E$下到质心O_c的距离分别为l_x、l_y、l_z，T为发动机产生的推力。那么推力分量引起的三轴偏转力矩可以表示为推力矢量力矩模型：

$$\begin{cases} M_{xt} = -T_{tvy} l_z \\ M_{yt} = T_{tvx} l_z \\ M_{zt} = -T_{tvx} l_y + T_{tvy} l_x \end{cases} \tag{2}$$

如果推力作用点在飞机纵向对称面内，即$l_z = 0$，则可以对推力矢量力矩模型进行简化：

$$\begin{cases} M_{xt} = 0 \\ M_{yt} = 0 \\ M_{zt} = -T_{tvx} l_y + T_{tvy} l_x \end{cases} \tag{3}$$

由此可以得到推力矢量飞机的纵向动力学模型如下：

$$\begin{cases} \dot{V}_x = -V_y\omega_z + (T_{tvx} - D - mg\sin\vartheta)/m \\ \dot{V}_y = V_x\omega_z + (L - mg\cos\vartheta + T_{tvy})/m \\ \dot{\omega}_z = (M_z + M_{zt})/I_{zz} \\ \dot{\vartheta} = \omega_z \\ \dot{h} = V_x\sin\vartheta + V_y\cos\vartheta \\ \dot{S}_g = V_x\cos\vartheta - V_y\sin\vartheta \end{cases} \tag{4}$$

式中：ϑ为俯仰角；ω_z为俯仰角速率；D和L分别为阻力和升力；M_z为俯仰通道的气动力矩，其与气动力矩系数之间可建立关系：$M_z = \bar{q}Slm_z$，S、S和l分别为动压、特征面积和特征长度；m_z即气动力矩系数；M_{zt}为该型飞机的推力矢量机构在俯仰通道中产生的力矩；V_x和V_y分别为飞行器沿着机体坐标系各轴的速度分量。S_g和h分别为飞行器沿着地面系各轴的前向位置和高度位置。m和g分别为飞行器质量和重力加速度。I_{zz}为推力矢量飞机对机体坐标系俯仰轴的转动惯量。

对飞机纵向运动方程组进行小扰动线性化求解，选取状态输出为速度、攻角、俯仰角、俯仰角速度、高度，即$\left[\Delta V \quad \Delta\alpha \quad \Delta\vartheta \quad \Delta\omega_z \quad \Delta h\right]$，控制输入为油门、升降舵和矢量喷管偏转，即油门开度、升降舵、矢量喷管偏转角度$\left[\delta_{\mathrm{th}} \quad \delta_{\mathrm{e}} \quad \delta_{\mathrm{tz}}\right]$，则可以将小扰动线性化方程整理为线性状态空间的形式。

$$\begin{bmatrix}\Delta\dot{V}\\ \Delta\dot{\alpha}\\ \Delta\dot{\vartheta}\\ \Delta\dot{\omega}_z\\ \Delta\dot{h}\end{bmatrix}=\begin{bmatrix}\dfrac{T_{\mathrm{tv}x}^{V}\cos\alpha-T_{\mathrm{tv}y}^{V}\sin\alpha-D^{V}}{m} & g\cos\theta+\dfrac{-T_{\mathrm{tv}x}\sin\alpha-T_{\mathrm{tv}y}\cos\alpha-D^{\alpha}}{m} & -g\cos\theta & 0 & 0\\ \dfrac{-T_{\mathrm{tv}x}^{V}\sin\alpha-T_{\mathrm{tv}y}^{V}\cos\alpha-L^{V}}{mV} & \dfrac{g}{V}\sin\theta+\dfrac{-T_{\mathrm{tv}x}\cos\alpha+T_{\mathrm{tv}y}\sin\alpha-L^{\alpha}}{mV} & -\dfrac{g}{V}\sin\theta & 1 & 0\\ 0 & 0 & 0 & 1 & 0\\ \left(M_z^{V}+M_{\mathrm{zt}}^{V}\right)/J_z & M_z^{\alpha}/J_z & 0 & M_z^{\omega_z}/J_z & M_z^{h}/J_z\\ -\sin\theta & V\cos\theta & -V\cos\theta & 0 & 0\end{bmatrix}\cdot\begin{bmatrix}\Delta V\\ \Delta\alpha\\ \Delta\vartheta\\ \Delta\omega_z\\ \Delta h\end{bmatrix}+\begin{bmatrix}\dfrac{T_{\mathrm{tv}x}^{\delta_{\mathrm{th}}}\cos\alpha-T_{\mathrm{tv}y}^{\delta_{\mathrm{th}}}\sin\alpha}{m} & -\dfrac{D^{\delta_{\mathrm{e}}}}{m} & \dfrac{T_{\mathrm{tv}x}^{\delta_{\mathrm{tz}}}\cos\alpha-T_{\mathrm{tv}y}^{\delta_{\mathrm{tz}}}\sin\alpha}{m}\\ \dfrac{T_{\mathrm{tv}x}^{\delta_{\mathrm{th}}}\sin\alpha+T_{\mathrm{tv}y}^{\delta_{\mathrm{th}}}\cos\alpha}{mV} & -\dfrac{L^{\delta_{\mathrm{e}}}}{mV} & \dfrac{-T_{\mathrm{tv}x}^{\delta_{\mathrm{tz}}}\sin\alpha-T_{\mathrm{tv}y}^{\delta_{\mathrm{tz}}}\cos\alpha}{mV}\\ 0 & 0 & 0\\ M_{\mathrm{zt}}^{\delta_{\mathrm{th}}}/J_z & M_z^{\delta_{\mathrm{e}}}/J_z & M_{\mathrm{zt}}^{\delta_{\mathrm{tz}}}/J_z\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\Delta\delta_{\mathrm{th}}\\ \Delta\delta_{\mathrm{e}}\\ \Delta\delta_{\mathrm{tz}}\end{bmatrix} \tag{5}$$

式中：θ为弹道倾角。

得到线性状态空间模型后，可以通过$\boldsymbol{W}(s)=(s\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{B}$求解其对应的传递函数矩阵形式：

$$\begin{bmatrix}V\\ \alpha\\ \vartheta\\ \omega_z\\ h\end{bmatrix}=\begin{bmatrix}W_{\delta_{\mathrm{th}}}^{V} & W_{\delta_{\mathrm{e}}}^{V} & W_{\delta_{\mathrm{tz}}}^{V}\\ W_{\delta_{\mathrm{th}}}^{\alpha} & W_{\delta_{\mathrm{e}}}^{\alpha} & W_{\delta_{\mathrm{tz}}}^{\alpha}\\ W_{\delta_{\mathrm{th}}}^{\vartheta} & W_{\delta_{\mathrm{e}}}^{\vartheta} & W_{\delta_{\mathrm{tz}}}^{\vartheta}\\ W_{\delta_{\mathrm{th}}}^{\omega_z} & W_{\delta_{\mathrm{e}}}^{\omega_z} & W_{\delta_{\mathrm{tz}}}^{\omega_z}\\ W_{\delta_{\mathrm{th}}}^{h} & W_{\delta_{\mathrm{e}}}^{h} & W_{\delta_{\mathrm{tz}}}^{h}\end{bmatrix}\begin{bmatrix}\delta_{\mathrm{th}}\\ \delta_{\mathrm{e}}\\ \delta_{\mathrm{tz}}\end{bmatrix} \tag{6}$$

式中：传递函数的分母多项式为$|s\boldsymbol{I}-\boldsymbol{A}|$，该项的求解包含全部状态矩阵$\boldsymbol{A}$的信息，建立了气动舵/推力矢量/油门杆控制向量到状态向量的传递函数后，便可以从频域角度定量分析气动舵/推力矢量/油门杆复合操纵机构对飞机状态的耦合影响。$W_{\delta_{\mathrm{e}}}^{\omega_z}$、$W_{\delta_{\mathrm{tz}}}^{\omega_z}$是俯仰通道气动升降舵和推力矢量共同对俯仰角速率进行控制的传递函数，而$W_{\delta_{\mathrm{th}}}^{\omega_z}$说明了油门杆控制对俯仰角速率有影响。$W_{\delta_{\mathrm{th}}}^{V}$是速度控制通道油门杆对速度进行控制的传递函数，而$W_{\delta_{\mathrm{e}}}^{V}$、$W_{\delta_{\mathrm{tz}}}^{V}$说明了气动升降舵和推力矢量机构对速度也有一定的影响。综上所述从频域说明了飞机的速度控制回路和俯仰角控制回路存在耦合特性。

本文以纵向控制通道下速度控制和姿态控制多回路系统作为研究设计和数值仿真对象，采用常值系数比例分配，同时从时域和频域奇异值角度进行控制系统参数优化设计工作。

2 基于奇异值方法的MIMO系统稳定裕度分析方法研究

稳定裕度是评价一个控制系统的重要性能指标，表明系统稳定的可靠程度。系统的稳定裕度包括两个方面：增益稳定裕度和相位稳定裕度。SISO系统的稳定裕度研究较为成熟，很容易从经典的Bode图或奈奎斯特曲线上获得。而对于MIMO系统，多回路交叉耦合，单回路系统稳定裕度的求解方式不再适用。因此设计了基于闭环回差矩阵最小奇异值方法来判断系统的稳定裕度。

闭环系统的稳定性取决于回差矩阵与逆回差矩阵的零点位置。由方程

$$\boldsymbol{H}(s)=[\boldsymbol{I}_m+\boldsymbol{G}(s)\boldsymbol{K}(s)]^{-1}\boldsymbol{G}(s)\boldsymbol{K}(s) \tag{7}$$

式中：$\boldsymbol{G}(s)$为开环传递函数矩阵；$\boldsymbol{K}(s)$为反馈矩阵。式(7)描述的多变量反馈系统为渐近稳定的充分必要条件为

$$n_0+\sum_{i=1}^{m}n_i=0 \tag{8}$$

式中：n_0为系统开环特征多项式位于s右半平面上的零点数；$\sum n_i$表示标量传递函数$g_i(s)k_i(s)(1\leqslant i\leqslant m)$的奈奎斯特图包围临界点$(-1,\mathrm{j}0)$的次数$n_i(1\leqslant i\leqslant m)$之和。

对于多变量系统其回差阵满足关系式：

$$\det(\boldsymbol{I}+\boldsymbol{GK})=\frac{\varphi_{\mathrm{c}}(s)}{\varphi_0(s)} \tag{9}$$

式中：$\varphi_{\mathrm{c}}(s)$为闭环系统特征多项式；$\varphi_0(s)$为开环系统特征多项式。若闭环标称系统是渐近稳定的，说明$\det(\boldsymbol{I}+\boldsymbol{GK})$的分子式$\varphi_{\mathrm{c}}(s)$的根全部具有负实部，因此令$s=\mathrm{j}\omega$，必有$\det(\boldsymbol{I}+\boldsymbol{GK})\neq 0$。

则根据奇异值的性质，意味着

$$\underline{\sigma}(\boldsymbol{I}+\boldsymbol{G})>0 \tag{10}$$

对于标称系统，要求受扰动后仍能保持全部零点在s复平面的左半平面内，且距虚轴有一段距离，描述方法就是矩阵奇异值概念。反馈系统的摄动通常有多种方式，本文采用的是乘法摄动，对于标称系统摄动矩阵$\boldsymbol{L}=\boldsymbol{I}$。

则标称系统回差矩阵定义为

$$\boldsymbol{W}=\boldsymbol{I}+\boldsymbol{K}(s)\boldsymbol{G}(s) \tag{11}$$

闭环系统回差矩阵的最小奇异值反映了该系统抗摄动的能力，因此，可以用来衡量该系统的稳定裕度。

稳定性判据描述如下：

若标称系统稳定，则受扰动系统稳定。如果在全频域范围有下列不等式成立：

$$\bar{\sigma}(\boldsymbol{L}^{-1}-\boldsymbol{I})\leqslant\sigma_{\mathrm{M}}\leqslant\underline{\sigma}(\boldsymbol{I}+\boldsymbol{KG}) \tag{12}$$

式中：σ_{M}是在所有频率范围内闭环系统的$\underline{\sigma}(\boldsymbol{I}+\boldsymbol{KG})$的最小值，上式的左边取决于增益和相位裕度，对于乘法摄动，有式(13)成立：

$$\bar{\sigma}\left(\boldsymbol{L}^{-1}-\boldsymbol{I}\right)=\max_{n}\sqrt{\left(1-\frac{1}{K_n}\right)^2+\frac{2}{K_n}(1-\cos\phi_n)} \tag{13}$$

因此经过上述推导，可求得系统的受控对象传递函数矩阵和控制回路传递函数矩阵，最终得到系统回差矩阵的最小奇异值，可以得到该系统的稳定裕度，计算公式如下：

$$\begin{cases}G_{\mathrm{M}}=-20\lg(1-\sigma_{\mathrm{M}})\\ P_{\mathrm{M}}=2\arcsin\dfrac{\sigma_{\mathrm{M}}}{2}\end{cases} \tag{14}$$

3　耦合多回路系统控制参数双域优化方法设计

对于 MIMO 系统的单回路进行控制系统参数设计方法简单，目前有多种方法可以寻优，但同时考虑多回路系统的控制性能时，由于控制回路耦合，控制器构型不同，控制参数的设计和寻优都较为困难。因此本文提出一种基于双域寻优的控制系统参数优化设计方法，根据期望的控制响应时域品质确定参数优化区间，然后利用频域的奇异值方法从稳定裕度的角度为控制系统参数设计进行寻优，具体内容如图 2 所示。

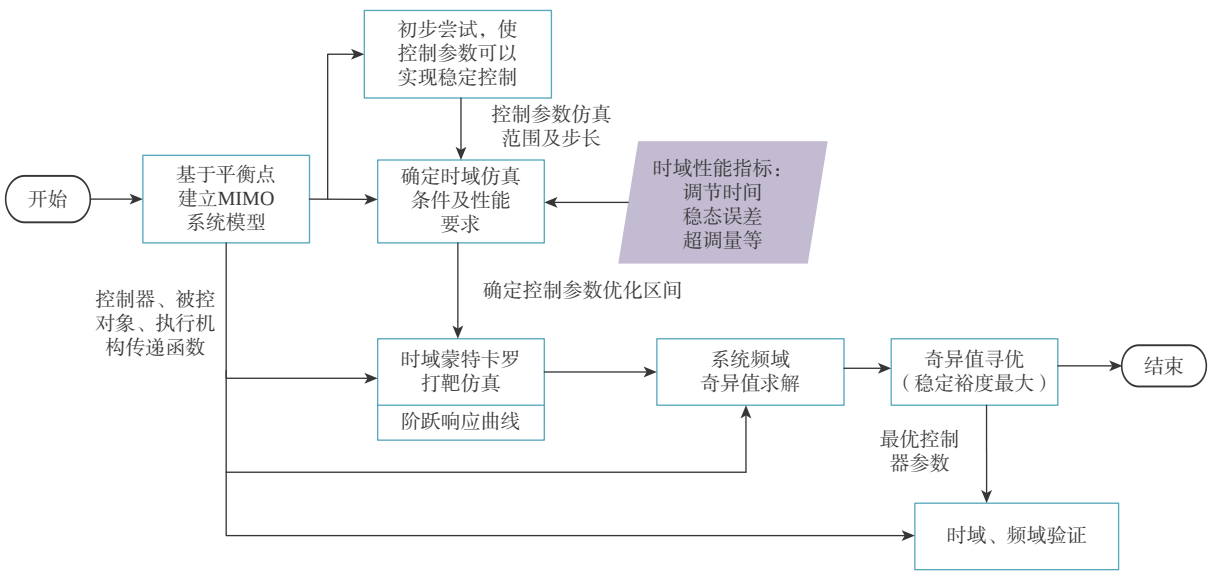


图 2　MIMO 系统控制参数寻优方法

Fig. 2　Optimization method for MIMO system control parameters

3.1　基于时域蒙特卡罗打靶的控制参数初步筛选

如图 2 所示，建立了 MIMO 系统传递函数模型后，需要根据时域性能指标进行控制器参数的初步筛选。

该类飞行器的纵向通道由矢量喷管和气动升降舵共同控制姿态，油门控制速度。因此本文研究如何使这两种执行机构和油门杆协同工作来完成控制系统的参数设计工作，该系统的 PID 控制律如式(15)所示：

$$\begin{cases}M_{zc}=K_{\mathrm{p}}(\vartheta_{\mathrm{cmd}}-\vartheta)+K_{\mathrm{i}}\int(\vartheta_{\mathrm{cmd}}-\vartheta)\\ \delta_{\mathrm{th}}=K(V_{\mathrm{cmd}}-V)+I\int(V_{\mathrm{cmd}}-V)\end{cases} \tag{15}$$

式中：ϑ_{cmd}和V_{cmd}分别为俯仰角速度指令和速度指令，δ_{th}为油门开度大小，K_{p}、K_{i}、K、I分别为俯仰角和速度控制器的参数。这里的控制分配模块为比例分配，即按照一定的比例将控制力矩指令分配到矢量喷管和气动升降舵上。由此可以看到该系统存在两个回路，分别为速度控制回路和姿态控制回路，因此本文对以上耦合回路进行研究。

首先采用试凑法，确定出控制器参数的数量级和大致范围。接着按照标准规定的性能品质级别选择合适的时域性能指标，如调节时间、超调量、稳态误差等，对不同组合的控制器参数进行大量的阶跃信号仿真，筛选出符合时域性能指标的控制参数集合。

3.2 基于多回路系统稳定裕度的控制参数寻优

基于时域控制性能指标获得初步的控制参数优化区间后，将其作为控制参数优化问题的约束，无穷维控制参数寻优问题被转化为有限维非线性参数优化问题，带约束的非线性优化问题一般形式可表示为

$$\begin{cases}\min\ f(x)\\ \quad \text{s.t.}\ c_i(x)=0 \qquad i=1,2,\cdots,m_{\mathrm{e}}\\ \qquad c_i(x)\leqslant 0 \qquad i=m_{\mathrm{e}}+1,m_{\mathrm{e}}+2,\cdots,m\\ \qquad \nabla f(x)+\sum_{i=1}^{m}\lambda_i\nabla c_i(x)=0\end{cases} \tag{16}$$

式中：目标函数f和c_i均为定义于R^n中的二阶连续可微函数。设x^*为上述问题的局部最优解。若在x^*处起作用约束条件，即有$c_i(x^*)=0$成立的约束条件的梯度向量$\nabla c_i(x^*)$线性无关，则存在满足下列Kuhn-Tucker条件的向量组$(x,\boldsymbol{\lambda})=(x^*,\boldsymbol{\lambda}^*)$

$$\begin{cases}\nabla f(x)+\sum_{i=1}^{m}\lambda_i\nabla c_i(x)=0\\ c_i(x)=0 \qquad i\in E\\ c_i(x)\leqslant 0,\lambda_i\geqslant 0,\ \ \lambda_i c_i(x)=0 \quad i\in K\end{cases} \tag{17}$$

式中：$\boldsymbol{\lambda}=\{\lambda_1,\lambda_2,\cdots,\lambda_m\}^{\mathrm{T}}$是拉格朗日乘子向量。上述最优性条件可由序列二次规划（sequential quadratic programming, SQP）算法[23]求解。

求解流程图如图3所示。

图中：控制参数优化过程包括两部分，首先基于控制性能指标和时域拉偏仿真，获得参数优化问题的约束；其次求解参数优化问题的目标函数，即使得系统的最小奇异值最大或使得系统稳定裕度最大。通过SQP算法求解最优控制参数，用于多回路控制器的参数更新。

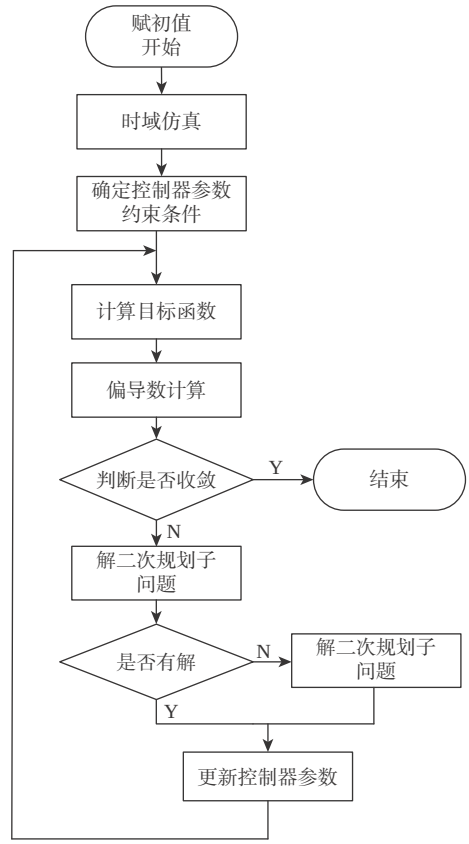


图3 SQP算法流程图

Fig. 3 SQP algorithm flowchart

4 数值仿真校验

4.1 推力矢量MIMO系统仿真模型建立

本文建立了推力矢量飞机的纵向传递函数模型，分别得到油门到速度、油门到俯仰角速度、升降舵到速度、升降舵到俯仰角速度、矢量喷管到速度、矢量喷管到俯仰角速度的传递函数如下：

$$G_i^j=\frac{K_4s^4+K_3s^3+K_2s^2+K_1s+K_0}{s^5+1.49s^4+4.78s^3+0.029s^2+0.024s+6.73\times10^{-5}} \tag{18}$$

式中：i,j分别为系统状态变量及执行机构变量；K_4,K_3,K_2,K_1,K_0分别为传递函数的分子系数，具体数值见表1。

在本文建立的MIMO系统中，每个回路均采

表1 不同传递函数的分子系数

Table 1 Molecular coefficients of different transfer functions

传递函数	K_4	K_3	K_2	K_1	K_0
$G_{\delta_{\mathrm{th}}}^{V}$	6.045	9.012	28.49	−0.033	0.019
$G_{\delta_{\mathrm{th}}}^{\omega_z}$	−0.016	−0.017	0.013	-3.4×10^{-5}	-4.95×10^{-20}
$G_{\delta_{\mathrm{e}}}^{V}$	9.942	13.37	−52.61	22.85	0.024
$G_{\delta_{\mathrm{e}}}^{\omega_z}$	−5.280	−2.329	−0.077	-5.196×10^{-5}	9.712×10^{-18}
$G_{\delta_{\mathrm{tz}}}^{V}$	0.083	−0.115	34.26	−9.072	-7.493×10^{-4}
$G_{\delta_{\mathrm{tz}}}^{\omega_z}$	1.827	0.902	0.035	1.959×10^{-7}	-1.257×10^{-20}

用 PI 控制器, 因此共有 4 个控制器参数。每个控制器的传递函数构型如式(19)所示:

$$G_c = \frac{P_c s + I_c}{s} \tag{19}$$

式中: P_c和I_c分别为 PI 控制器的比例增益和积分增益。

4.2　基于时域控制性能指标的参数优化区间确定

根据工程经验, 对两个回路的阶跃信号确定了如表 2 所示的控制性能指标。

表 2　回路控制性能指标内容及要求

Table 2　Content and requirements of loop control performance indicators

指标要求	阶跃信号	调节时间/s	超调量/%	状态误差/%
速度回路	100 m/s	5	10	1
姿态回路	0.1 rad/s	5	10	5

根据建立的系统模型给出了时域性能指标要求后, 可以根据蒙特卡罗时域仿真确定控制系统参数的优化区间。多回路控制系统参数的仿真范围与仿真步长如表 3 所示。

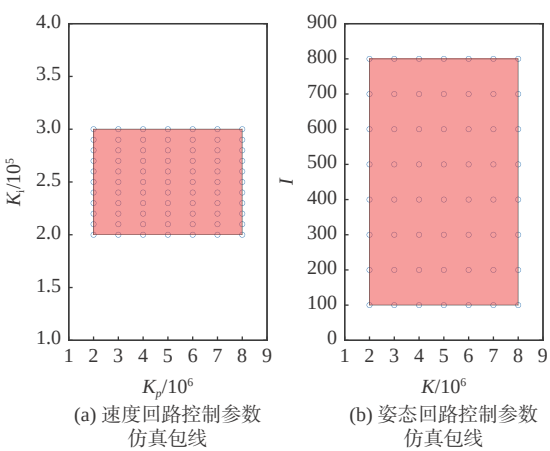
表 3　控制参数优化区间与步长

Table 3　Optimization interval and step size of control parameters

优化区间及步长	速度回路		姿态回路	
	K_p	K_i	K	I
下限	2×10^6	2×10^5	2×10^6	1×10^2
上限	8×10^6	3×10^5	8×10^6	8×10^2
步长	10^6	10^4	10^6	10^2

通过控制参数的蒙特卡罗打靶仿真, 与时域控制性能指标进行对比, 得到控制参数的初步优化区间如图 4 所示, 共有 4 313 组控制器参数组合。

按照上述参数进行大量时域仿真后得出, 共有 3 696 组控制参数满足本节中所要求的时域性能指标。



(a) 速度回路控制参数仿真包线　　(b) 姿态回路控制参数仿真包线

图 4　不同回路控制器参数仿真包线

Fig. 4　Simulation envelope of different loop controller parameters

4.3　基于奇异值稳定裕度分析法的控制参数寻优

在 4.2 节的 3 696 组控制参数集合中, 进行 MIMO 系统奇异值的寻优, 求出奇异值最大时的控制参数即为稳定裕度最大的控制参数。

图 5 为 3 696 组控制参数集合的奇异值求解结果。

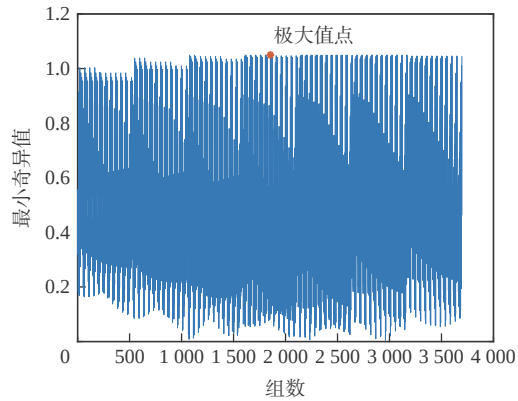


图 5　不同组控制参数时的最小奇异值

Fig. 5　Minimum singular value for different sets of control parameters

由图 5 可以求出奇异值最大的点, 此时奇异值为 1.049 8, 根据式 (14) 求解得到的 MIMO 系统稳定裕度如表 4 所示。

表 4　最优控制器参数下的多输入多输出系统裕度

Table 4　MIMO system margin under optimal controller parameters

幅值裕度/dB	相角裕度/(°)
6.234 2	63.322 2

对于基于比例分配的升降舵/推力矢量分配方式, 进行了不同比例下的奇异值优化结果、幅值裕度和相角裕度的求解, 如表 5 所示, 表中比例情况为升降舵分配的力矩指令与推力矢量分配的力矩指令之比。

表 5　不同分配比例情况下的指标结果

Table 5　Indicator results under different allocation ratios

比例情况	奇异值优化结果	幅值裕度/dB	相角裕度/(°)
5∶5	1.049 8	6.234 2	63.322 2
4∶6	1.054 0	6.252 0	63.606 4
3∶7	1.058 5	6.270 8	63.906 3
2∶8	1.062 4	6.287 5	64.174 4
1∶9	1.066 8	6.306 0	64.471 5

可以看到, 改变比例分配的情况, 对于奇异值优化结果、幅值裕度和相角裕度的影响不大, 但是随着矢量喷管分配到的力矩变多, 频域稳定裕度在变好。

4.4 双域寻优的最优控制参数时域仿真校验

根据4.3节所求的最优参数，可以得到此时速度环控制器参数为：$P=5\times10^6$，$I=2.5\times10^5$，姿态环控制器参数为：$P=6\times10^6$，$I=600$。此时两个回路的时域仿真结果如图6和图7所示。

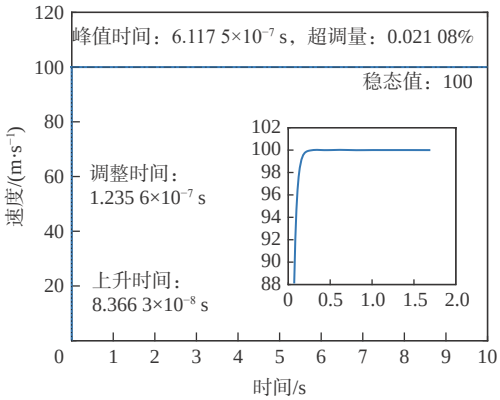


图6 速度环阶跃响应曲线及控制性能

Fig. 6 Speed loop step response curve and control performance

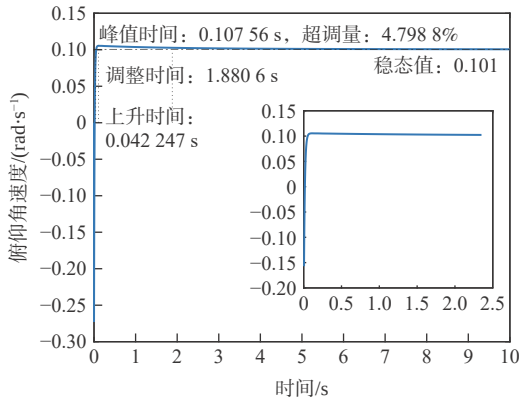


图7 俯仰角速度环阶跃响应曲线及控制性能

Fig. 7 Step response curve and control performance of pitch angular velocity loop

将本文设计的参数优化设计方法得到的结果与常规经验试凑得到的控制参数放入控制器中进行对比仿真分析，并进行了力矩系数±10%、±20%的拉偏仿真，得到如图8和图9所示的控制对比。

可以看到，利用本文的方法，最优参数的时域控制效果得到了保证，在速度环快速性更好，调节时间更快，在俯仰角速度环快速性更好，超调量更小，调节时间更快。同时最优参数在力矩系数拉偏情况下仍具有较好的控制效果，可以说明控制器的鲁棒性。

将常规单输入控制设计结果和多回路控制设计进行对比，结果如图10和图11所示。

由图10和图11的结果可以看到，单输入设计完成后放到多回路系统中仍然具有较好的控制效果。速度环中多回路设计方法的效果明显优于单

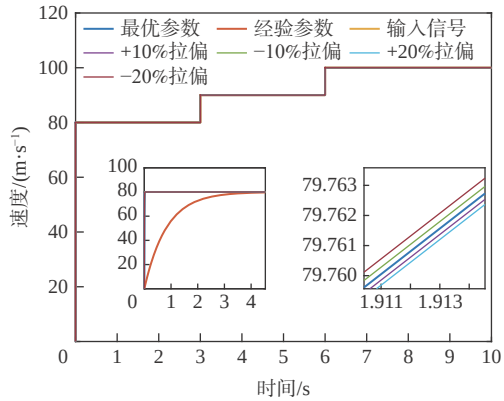


图8 速度环对比曲线

Fig. 8 Speed loop comparison curve

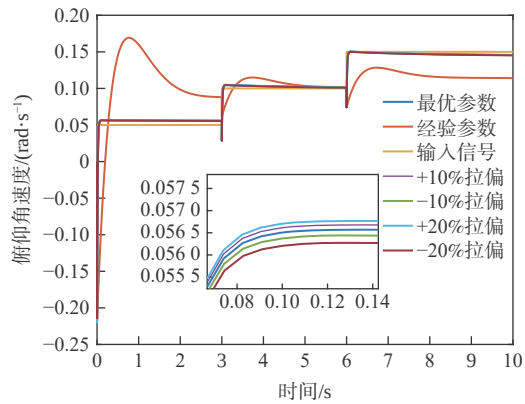


图9 俯仰角速度环对比曲线

Fig. 9 Comparison curve of pitch angular velocity loop

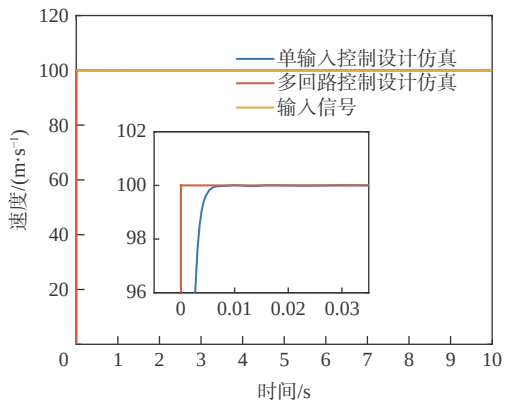


图10 单输入控制设计仿真和多回路控制设计仿真速度对比结果

Fig. 10 Comparison of speed results between single input control design simulation and multi loop control design simulation

输入设计方法，调节时间有大幅减少；俯仰角速度环中多回路设计方法相较于单输入设计方法的超调有减少，但效果不明显。本文基于奇异值提出的控制参数优化方法，以频域稳定裕度的提升为目标进行控制参数优化，而时域性能仅作为参数优化过程中的约束条件，不作为优化的准则，对于频域稳定裕度的提升结果如表6所示。

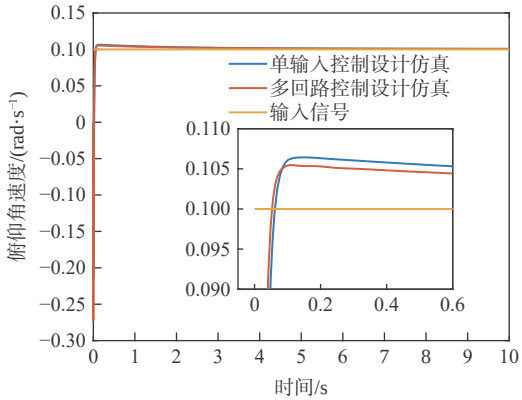


图 11　单输入控制设计仿真和多回路控制设计仿真俯仰角速度对比结果

Fig. 11　Comparison results of pitch angular velocity between single input control design simulation and multi loop control design simulation

表 6　两种控制参数设计方法稳定裕度对比结果

Table 6　Comparison results of stability margin between two control parameter design methods

控制参数设计方法	奇异值	幅值裕度/dB	相角裕度/(°)
单输入	0.827 2	5.235 5	48.859 3
基于奇异值的多回路	1.049 8	6.234 2	63.322 2

注：奇异值、幅值裕度、相角裕度的提升幅度分别为26.9%，19.1%，29.6%。

为更好地比较系统的鲁棒能力，在速度输出端引入随机信号干扰，如图 12 所示，对比最优参数和经验参数的控制效果如图 13 和图 14 所示。

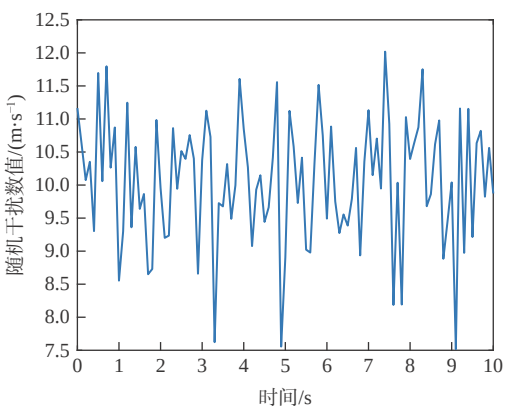


图 12　随机干扰信号曲线

Fig. 12　Random interference signal curve

如图 13 和图 14 所示，加入随机信号干扰后，速度环受到的影响较小，但俯仰角速度环较为敏感，可以看到最优参数下的控制效果虽有较小幅度的震荡，但控制效果仍较好，但经验参数下系统开始震荡，控制效果很差，因此该方法设计得到的控制参数具有一定的优越性，可以保证系统具有时域性能的同时，进一步确保了系统的鲁棒性、稳定性。

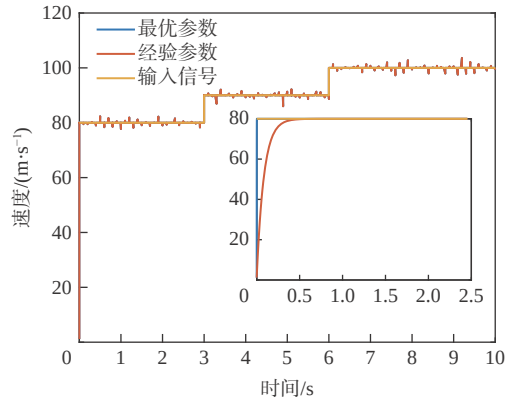


图 13　加入干扰后速度对比曲线

Fig. 13　Speed comparison curve after adding interference

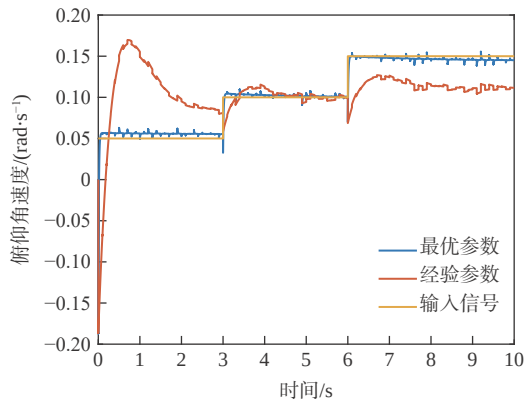


图 14　加入干扰后俯仰角速度环对比曲线

Fig. 14　Comparison curve of pitch angular velocity loop after adding interference

5　结　论

1）针对推力矢量/气动舵飞机纵向 MIMO 控制系统提出了一种控制参数双域优化设计方法，该方法可推广至 MIMO 系统的控制系统参数设计中。

2）控制参数双域优化设计方法基于时域性能要求确定优化参数区间，然后将系统稳定裕度作为 MIMO 耦合控制系统参数优化的目标，采用频域奇异值方法对控制参数进行带约束寻优，从而保证了对多回路系统的稳定控制。

3）通过对比仿真分析，验证了基于奇异值方法得到的 MIMO 系统控制参数具有更好的鲁棒性、稳定性。

参考文献（References）

[1] LIU J J, SUN M W, CHEN Z Q, et al. Super-twisting sliding mode control for aircraft at high angle of attack based on finite-time extended state observer[J]. Nonlinear Dynamics, 2020, 99(4): 2785-2799.

[2] 占正勇，崔玉伟. 多用途作战飞机复合控制技术研究[J]. 飞行力学, 2015, 33(6): 537-541,550.
ZHAN Z Y, CUI Y W. Research on complex control for multi-pur-

pose combat aircraft[J]. Flight Dynamics, 2015, 33(6): 537-541,550 (in Chinese).

[3] 周大鹏, 杨大鹏, 刘然, 等. 一种面向空天飞机再入的智能自适应复合控制方法[J]. 宇航学报, 2022, 43(8): 1109-1119.
ZHOU D P, YANG D P, LIU R, et al. An intelligent adaptive compound control method for aerospace plane reentry[J]. Journal of Astronautics, 2022, 43(8): 1109-1119(in Chinese).

[4] 董哲, 刘凯, 李旦伟, 等. 考虑动态控制分配的空天飞行器再入姿态复合控制设计[J]. 宇航学报, 2021, 42(6): 749-756.
DONG Z, LIU K, LI D W, et al. A dynamic control allocation approach for reentry compound attitude control design of aerospace vehicle[J]. Journal of Astronautics, 2021, 42(6): 749-756(in Chinese).

[5] 董金鲁, 马悦萌, 周荻, 等. 临近空间高超声速飞行器的直接力与襟翼复合滑模控制[J]. 兵工学报, 2023, 44(2): 496-506.
DONG J L, MA Y M, ZHOU D, et al. A composite sliding mode control scheme based on reaction jets and flaps for near-Space hypersonic vehicles[J]. Acta Armamentarii, 2023, 44(2): 496-506(in Chinese).

[6] 郭建国, 彭谦, 周敏. 直接力/气动力复合控制技术发展综述[J]. 航空兵器, 2022, 29(1): 1-13.
GUO J G, PENG Q, ZHOU M. Review on development of direct force/aerodynamic force compound control technology[J]. Aero Weaponry, 2022, 29(1): 1-13(in Chinese).

[7] AZARSKOV V, TUNIK A, SUSHCHENKO O. Design of composite feedback and feedforward control law for aircraft inertially stabilized platforms[J]. International Journal of Aerospace Engineering, 2020, 2020: 8853928.

[8] 丁伟涛, 肖翀, 黄玉平. 直接力与推力矢量复合控制技术研究[J]. 导航定位与授时, 2017, 4(5): 27-31.
DING W T, XIAO C, HUANG Y P. Researsh on the compound control of direct force and thrust vector[J]. Navigation Positioning and Timing, 2017, 4(5): 27-31(in Chinese).

[9] 崔彦凯, 王志刚. 越肩发射空空导弹气动力/推力矢量复合控制器设计[J]. 宇航学报, 2010, 31(4): 1100-1104.
CUI Y K, WANG Z G. Aerodynamic and thrust vector complex controller design of over the shoulder air-to-air missile[J]. Journal of Astronautics, 2010, 31(4): 1100-1104 (in Chinese).

[10] GAO F, TANG S J, SHI J, et al. Compound control allocation strategy of dual aero/jet vane control missile[J]. Journal of Beijing Institute of Technology, 2012, 21(4): 434-441.

[11] 苏润, 赵明. 飞艇矢量推力与舵面复合控制分配研究[J]. 电脑知识与技术, 2017, 13(24): 221-223.
SU R, ZHAO M. Study on the distribution of vector thrust and rudder composite control of airship[J]. Computer Knowledge and Technology, 2017, 13(24): 221-223(in Chinese).

[12] 艾尔, 王衍洋, 屈香菊. 气动/推力矢量控制面融合方式研究[J]. 飞行力学, 2005, 23(4): 20-24.
AI E, WANG Y Y, QU X J. An analysis on the combination of aero/thrust vector control surfaces[J]. Flight Dynamics, 2005, 23(4): 20-24(in Chinese).

[13] 李子木. 基于奇异摄动法的多变量系统稳定性测量[D]. 大连: 大连海事大学, 2017: 9-12.
LI Z M. Stability measurement of multivariable systems based on singular perturbation method[D]. Dalian: Dalian Maritime University, 2017: 9-12 (in Chinese).

[14] 曹玉腾, 倪少波. 再入飞行器多变量稳定裕度研究[J]. 航天控制, 2015, 33(5): 28-32.
CAO Y T, NI S B. Research of reentry aerocraft multivariable stability margin[J]. Aerospace Control, 2015, 33(5): 28-32(in Chinese).

[15] 瞿福存, 史忠科, 戴冠中. MIMO 系统稳定裕度的几个定义[J]. 飞行力学, 2002, 20(2): 6-9.
QU F C, SHI Z K, DAI G Z. Stability margin definitions for MIMO systems[J]. Flight Dynamics, 2002, 20(2): 6-9(in Chinese).

[16] LEHTOMAKI N, SANDELL N, ATHANS M. Robustness results in linear-quadratic Gaussian based multivariable control designs[J]. IEEE Transactions on Automatic Control, 2003, 26(1): 75-93.

[17] 李信栋, 苟兴宇. 由回差阵奇异值求稳定裕度的退化算法[J]. 控制理论与应用, 2016, 33(4): 460-465.
LI X D, GOU X Y. Degraded algorithm for determining stability margin by using singular value of the return difference matrix[J]. Control Theory & Applications, 2016, 33(4): 460-465(in Chinese).

[18] SON J E, MANI A S, LATCHMAN H A. Robustness analysis for MIMO systems with unstructured uncertainties[C]// IEEE ICCA 2010. Piscataway: IEEE Press, 2010: 1333-1337.

[19] 李信栋, 苟兴宇. 多体卫星 MIMO 控制及稳定裕度研究[J]. 航天控制, 2013, 31(5): 25-30.
LI X D, GOU X Y. On MIMO control and stability margin of multibody satellite, China[J]. Aerospace Control, 2013, 31(5): 25-30(in Chinese).

[20] 吴斌, 程鹏. 多变量飞控系统的稳定裕度分析[J]. 航空学报, 1998, 19(6): 657-661.
WU B, CHENG P. Stability margin analysis of multivariable flight control system[J]. Acta Aeronautica et Astronautica Sinica, 1998, 19(6): 657-661 (in Chinese).

[21] 段子帆, 宁国栋, 张红文, 等. 高超声速飞行器横航向 MIMO 系统控制及稳定裕度评估[J]. 战术导弹技术, 2018(2): 70-79.
DUAN Z F, NING G D, ZHANG H W, et al. Lateral MIMO systems control and stability margin evaluation of hypersonic vehicle[J]. Tactical Missile Technology, 2018(2): 70-79(in Chinese).

[22] 李帆. 不确定性系统的解耦控制与稳定裕度分析[D]. 西安: 西北工业大学, 2001: 94-114.
LI F. Decoupling control and stability margin analysis of uncertain systems[D]. Xi'an: Northwestern Polytechnical University, 2001: 94-114 (in Chinese).

[23] PESCETELLI F, MINISCI E, MADDOCK C, et al. Ascent trajectory optimisation for a single-stage-to-orbit vehicle with hybrid propulsion[C]// 18th AIAA/3AF International Space Planes and Hypersonic Systems and Technologies Conference. Reston: AIAA, 2012: AIAA2012-5828.

Parameter optimization method of thrust vector/pneumatic rudder composite control law for aircraft based on singular value method

RUAN Shilong[1]，DONG Zhe[1,*]，SUN Yao[2]，QU Xiaolei[3]，HUO Shaoze[1]

(1. School of Aeronautics and Astronautics，Dalian University of Technology，Dalian 116024，China；
2. Shenyang Aircraft Corporation，Shenyang 110034，China；
3. Shenyang Aircraft Design and Research Institute，Shenyang 110035，China)

Abstract：Control coupling is a major problem for advanced aircraft with thrust vector/aerodynamic rudder layouts, as the performance of numerous control loops cannot be balanced using conventional single loop control parameter design techniques. Therefore, a control law parameter optimization method based on the singular value of the feedback matrix is proposed. The optimization interval of the control parameters is first established using the time domain control performance index. The stability margin of the multiple-in multiple-out (MIMO) system is then measured using the singular value method, and the corresponding optimal objective function is established to optimize the controller parameters. The feasibility of this method was verified through numerical simulation, and the results showed that the designed control parameter optimization algorithm has better time-domain control performance and a larger system stability margin compared to traditional single loop control parameter design methods.

Keywords：return-difference-matrix singular values；parameter optimization；control system stability；MIMO systems；thrust vectors

Received：2023-05-08；**Accepted**：2023-07-16；**Published Online**：2023-09-08 11：28

URL：link.cnki.net/urlid/11.2625.V.20230907.1549.006

Foundation items：National Natural Science Foundation of China (U2141229); National Basic Research Priorities Program of China (2019-JCJQ-DA-001-131)

* **Corresponding author**. E-mail: zhedong@mail.dlut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0238

基于鲁棒最优滑模的非充气轮胎驱动车辆稳定性控制

李田，赵又群*，徐涛，沈亚伟，林棻
(南京航空航天大学 能源与动力学院，南京 210016)

摘　　要：机械弹性电动轮是一种新式非充气轮胎，具有防爆、防胎刺等优点，文章中基于匹配该电动轮的分布式汽车，提出了一种鲁棒最优滑模(ROSM)控制策略，以便提高车辆转向时的横摆稳定性。针对线性二自由度模型，为实现最优控制采用线性二自由度调节器(LQR)，输出初始的附加横摆力矩；考虑到实际行驶中车辆状态是复杂的非线性系统，建立包含不确定参数的车辆动力学模型，并在初始最优控制量的基础上设计一种鲁棒积分滑模控制器，该控制器对不确定参数与外部干扰具有良好的鲁棒性，且仍能实现最优控制；通过MATLAB/Simulink和Carsim联合仿真，对控制方法进行仿真验证，结果表明：双移线工况中，ROSM控制下横摆角速度的平均绝对误差(MAE)、均方根误差(RMSE)与LQR控制相比分别下降了63.83%、65.33%；蛇形工况中，其分别降低了58.38%、60.02%。

关　键　词：非充气轮胎；直接横摆力矩；线性二自由度调节器；鲁棒最优控制；滑模控制
中图分类号：U461.6
文献标志码：A　　**文章编号**：1001-5965(2025)04-1342-10

随着世界各国对环境的日渐重视，不少欧洲国家相继颁布未来燃油车禁售令，纯电动汽车的应用与普及已成为必然趋势。分布式轮毂电机驱动作为新兴的驱动结构，具有四轮转矩独立可控、结构紧凑、智能网联适配性强等优点，极大地提高了车辆稳定性控制的灵活性。

车轮是保证分布式电动车稳定行驶的基础部件，传统充气轮胎在恶劣路面有穿刺、爆胎的风险，南航赵又群团队[1]研制的机械弹性电动轮(mechanical elastic electric wheel, MEEW)具有防爆、防胎刺等优点，并在重型卡车上完成了实车验证。Deng等[2]对机械弹性车轮的结构、力学特性进行了深入研究，张陈曦等[3]充分考虑了该车轮衍生的免充气轮胎的径向刚度特性，运用半逆解法优化尺寸设计，提高了车辆行驶的平顺性；张晨等[4]运用疲劳试验方法对该车轮的耐久性进行了研究。

直接横摆力矩控制(direct yaw moment control, DYC)是辅助驾驶中控制车辆稳定性的常见方法，其关键是准确计算出附加横摆力矩值，常用算法包括模糊控制[5]、PID控制[6]、滑模控制[7]和模型预测控制[8]。Liu等[9]构建了基于无味粒子滤波观测器的前馈系统，通过实时观测车辆横摆角速度与质心侧偏角，计算出精确的附加横摆力矩，提高了车辆稳定行驶的跟踪精度。林棻等[10]采用分层控制器控制车辆的横摆稳定性，上层基于积分联合终端的滑模控制器，获得期望的横摆力矩，下层基于纵向力轴载比例合理分配力矩。Zhu等[11]设计了一种基于粒子群优化的非线性模型预测控制方法，控制

收稿日期：2023-05-11；**录用日期**：2023-09-25；**网络出版时间**：2023-10-11 08:33
网络出版地址：link.cnki.net/urlid/11.2625.v.20231009.1357.004
基金项目：国家自然科学基金项目(52272397,11672127)；中央高校基本科研业务费专项资金资助项目(NP2022408)；高机动防暴车辆技术国家工程实验室开放基金项目(B20210017)；南京航空航天大学科研与实践创新计划项目(20220211)
* **通信作者**. E-mail：yqzhao@nuaa.edu.cn

引用格式：李田，赵又群，徐涛，等. 基于鲁棒最优滑模的非充气轮胎驱动车辆稳定性控制[J]. 北京航空航天大学学报，2025，51(4)：1342-1351. LI T，ZHAO Y Q，XU T，et al. Stability control of vehicles powered by non-pneumatic wheels based on robust optimal sliding mode[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1342-1351 (in Chinese).

精度较高, 但计算复杂耗时, 难以满足车辆辅助驾驶的实时性要求。

滑模控制由于其实现简单、响应迅速、鲁棒性强等优点, 广泛应用于系统工程中的非线性扰动控制。Yan 等[12]、Rangel 等[13] 为了使水下车辆在复杂多变的环境中跟踪三维轨迹, 设计了基于积分滑模的鲁棒控制器。Norsahperi 等[14] 提出了一种改进的最优积分滑模控制策略, 使得机器人在耦合效应、参数变化和外界干扰下仍能保持较高的跟踪性能。Xu 等[15] 针对分布式电动车动力学参数不确定与扰动问题, 提出了最优控制与高阶滑模相结合的鲁棒控制算法。丛森森等[16] 针对底盘横纵向稳定性问题, 以滑模面构建了车辆动态稳定域, 滑模动态控制器始终控制车辆维持在稳定状态, 保证行驶的安全性。梁宝钰等[17] 为了保持车辆在侧风干扰下行驶的稳定性, 基于滑模理论计算出汽车高速行进时所需的附加横摆力矩, 有效提高了车辆抗侧风干扰能力。陈特等[18] 基于非奇异终端滑模控制, 在状态观测基础上实现了主动前轮转向功能, 大大提高了轨迹跟踪的精度。

在上述研究基础上, 本文基于匹配机械弹性车轮的分布式汽车, 提出了一种鲁棒最优滑模[12-19] (ROSM) 的控制策略, 以便提高车辆转向时的横摆稳定性。首先, 基于线性二自由度车辆模型, 设计线性二自由度调节器 (LQR) 控制器, 输出初始最优附加横摆力矩值; 其次, 由于机械弹性车轮实际行驶中侧偏刚度存在不确定性摄动, 且路面状态复杂多变, 从而引入参数不确定的非线性动力学模型, 在此基础上设计了一种基于初始最优控制量的鲁棒最优积分滑模控制策略, 该策略不仅可以实现最优控制目标, 且对不确定参数与外部干扰具有良好的鲁棒性; 最后基于二次规划, 将附加横摆力矩合理分配至各轮胎。

1　动力学模型搭建

1.1　车辆动力学模型建立

本文选取理想的车辆二自由度模型作为参考模型, 如图 1 所示。

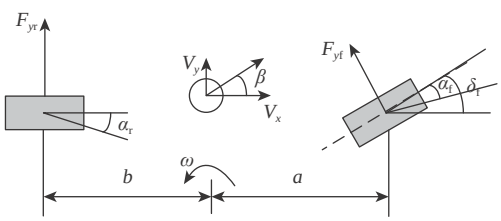


图 1　二自由度车辆动力学模型

Fig. 1　2 DOF vehicle dynamics model

其动力学方程为

$$\begin{bmatrix}\dot{\beta}\\ \dot{\omega}\end{bmatrix}=\begin{bmatrix}\dfrac{k_f+k_r}{mV_x} & \dfrac{ak_f-bk_r}{mV_x^2}-1\\ \dfrac{ak_f-bk_r}{I_z} & \dfrac{a^2k_f+b^2k_r}{I_zV_x}\end{bmatrix}\begin{bmatrix}\beta\\ \omega\end{bmatrix}+\begin{bmatrix}\dfrac{-k_f}{mV_x}\\ -\dfrac{ak_f}{I_z}\end{bmatrix}\delta_f \tag{1}$$

式中: β为质心侧偏角; ω为横摆角速度; V_x为车辆纵向车速; a、b为车辆质心至前、后轴距离; m为整车质量; I_z为横摆转动惯量; δ_f为前轮转角; k_f、k_r为车辆在理想状态下的前、后轴侧偏刚度。

一般而言, 理想质心侧偏角可以取为 0, 而期望的横摆角速度为

$$\omega_d=\frac{V_x/L}{1+K_sV_x^2}\delta_f \tag{2}$$

式中: L为车辆前后轴距; K_s为车辆不足转向参数, $K_s=\dfrac{m}{L^2}\left(\dfrac{a}{k_r}-\dfrac{b}{k_f}\right)$。

由于轮胎附着极限的存在, 理想横摆角速度有一个上限值, 可以表示为

$$\omega_m=\frac{\mu g}{V_x} \tag{3}$$

式中: μ为路面附着系数。

综上, 理想横摆角速度和质心侧偏角可表示为

$$\begin{cases}\omega_d=\begin{cases}\omega & |\omega|<\omega_m\\ \omega_m\mathrm{sign}(\omega) & |\omega|>\omega_m\end{cases}\\ \beta_d=0\end{cases} \tag{4}$$

1.2　非充气轮胎动力学模型

南航赵又群团队研制的机械弹性电动轮 (MEEW)[20] 具有防爆、防胎刺等优点, 能够承受较大的径向载荷, 选取该非充气轮胎作为本文的研究对象。MEEW 结构与试验台如图 2、图 3 所示, MEEW 由轮毂、铰链组、輻轮组成, 橡胶层经硫化处理覆在卡环、弹性环上形成輻轮。

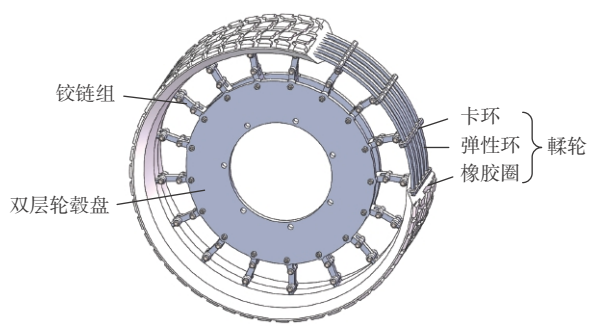


图 2　MEEW 结构建模

Fig. 2　MEEW structural model

为具象表达 MEEW 的轮胎特性, 引入“魔术公式”[21], 分别建立 MEEW 纵向力F_x与侧向力F_y的力学模型。



图 3 台架试验

Fig. 3 Experimental equipment

$$\begin{cases} F_x = D_x \sin\left[C_x \arctan(B_x\phi_x)\right] + S_{vx} \\ \phi_x = (1-E_x)\lambda + (E_x/B_x)\arctan(B_x\lambda) \end{cases} \tag{5}$$

$$\begin{cases} F_y = D_y \sin\left[C_y \arctan(B_y\phi_y)\right] + S_{vy} \\ \phi_y = (1-E_y)\alpha + (E_y/B_y)\arctan(B_y\alpha) \end{cases} \tag{6}$$

式中：B_x、B_y分别为纵向力、侧向力的刚度因子；C_x、C_y分别为纵向力、侧向力曲线的形状因子；D_x、D_y分别为纵向力、侧向力的峰值因子；E_x、E_y分别为纵向力、侧向力曲线的曲率因子；λ为纵向滑移率，α为轮胎的侧偏角；S_{vx}为纵向力曲线的水平方向漂移，S_{vy}为侧向力曲线的垂直方向漂移。具体为

$$\begin{cases} B_x = r_x/(C_xD_x) \quad C_x = b_0 \\ D_x = b_1F_z^2 + b_2F_z \\ r_x = (b_3F_z^2 + b_4F_z)\mathrm{e}^{-b_5F_z} \\ E_x = b_6F_z^2 + b_7F_z + b_8 \end{cases} \tag{7}$$

$$\begin{cases} B_y = r_y/(C_yD_y) \quad C_y = a_0 \\ D_y = a_1F_z^2 + a_2F_z \\ r_y = a_3\sin(2\arctan(F_z/a_4)) \\ E_y = a_6F_z + a_7 \end{cases} \tag{8}$$

式中：F_z为 MEEW 承受的径向载荷；r_x、r_y分别为纵向力、侧向力曲线过原点的斜率。

在样机实验中对 MEEW 分别施加 10 kN、15 kN、20 kN 的径向载荷，并拟合实验结果数据，得到纵向力F_x与侧向力F_y的拟合曲线如图 4、图 5 所示，魔术轮胎公式参数辨识结果如表 1 所示。

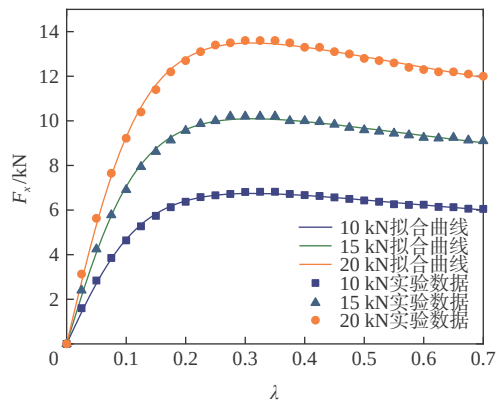


图 4 MEEW 纵向力拟合曲线

Fig. 4 Longitudinal force fitting curve of MEEW

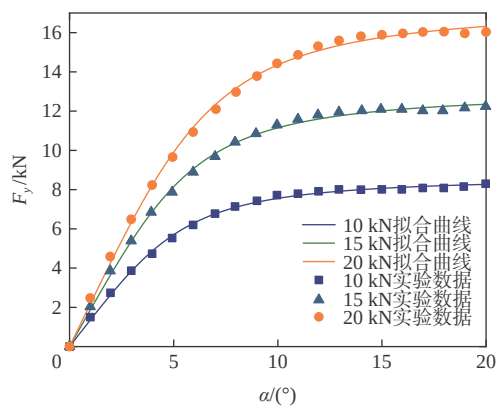


图 5 MEEW 侧向力拟合曲线

Fig. 5 Lateral force fitting curve of MEEW

表 1 MEEW 参数拟合结果

Table 1 Parameter fitting results

F_z/kN	B_x	C_x	D_x	E_x	B_y	C_y	D_y	E_y
10	4.947 7	1.721 6	6 795.14	0.454 1	0.136 4	1.268 2	8 150	−0.098 8
15	3.932 3	2.125 2	10 103.66	0.874 2	0.127 1	1.268 2	12 200	−0.095 2
20	4.807 0	1.739 0	13 495.56	0.441 1	0.111 6	1.268 2	16 310	−0.091 4

2 基于鲁棒最优滑模的上层控制器设计

2.1 基于线性二自由度模型的 LQR 控制方法

轮胎侧偏特性在非线性区域时，轮胎侧偏刚度也随着侧偏角的增大而不断变化，通常为简化问题，假使其依旧保持原有的等效侧偏刚度，本节基于线性二次型 (LQR) 设计直接横摆力矩控制器[22]。基于线性二自由度模型，理想的质心侧偏角与横摆角速度为

$$\begin{bmatrix} \dot{\beta}_d \\ \dot{\omega}_d \end{bmatrix} = \begin{bmatrix} \dfrac{k_f + k_r}{mV_x} & \dfrac{ak_f - bk_r}{mV_x^2} - 1 \\ \dfrac{ak_f - bk_r}{I_z} & \dfrac{a^2k_f + b^2k_r}{I_zV_x} \end{bmatrix} \begin{bmatrix} \beta_d \\ \omega_d \end{bmatrix} + \begin{bmatrix} \dfrac{-k_f}{mV_x} \\ -\dfrac{ak_f}{I_z} \end{bmatrix} \delta_f \tag{9}$$

为保持车身稳定, 轮毂电机驱动对四轮施加直接横摆力矩, 总附加横摆力矩ΔM作用后, 动力学方程更新为

$$\begin{bmatrix}\dot\beta\\ \dot\omega\end{bmatrix}=\begin{bmatrix}\dfrac{k_\mathrm{f}+k_\mathrm{r}}{mV_x} & \dfrac{ak_\mathrm{f}-bk_\mathrm{r}}{mV_x^{\,2}}-1\\ \dfrac{ak_\mathrm{f}-bk_\mathrm{r}}{I_z} & \dfrac{a^2k_\mathrm{f}+b^2k_\mathrm{r}}{I_zV_x}\end{bmatrix}\begin{bmatrix}\beta\\ \omega\end{bmatrix}+\begin{bmatrix}-\dfrac{k_\mathrm{f}}{mV_x}\\ -\dfrac{ak_\mathrm{f}}{I_z}\end{bmatrix}\delta_\mathrm{f}+\begin{bmatrix}0\\ \dfrac{1}{I_z}\end{bmatrix}\Delta M \tag{10}$$

定义理想质心侧偏角与实际响应值的差为$\Delta\beta$, 理想横摆角速度与实际响应值的差为$\Delta\omega$, 具体为

$$\begin{cases}\Delta\beta=\beta_\mathrm{d}-\beta\\ \Delta\omega=\omega_\mathrm{d}-\omega\end{cases} \tag{11}$$

式 (9) 与式 (10) 相减得

$$\begin{bmatrix}\Delta\dot\beta\\ \Delta\dot\omega\end{bmatrix}=\begin{bmatrix}\dfrac{k_\mathrm{f}+k_\mathrm{r}}{mV_x} & \dfrac{ak_\mathrm{f}-bk_\mathrm{r}}{mV_x^{\,2}}-1\\ \dfrac{ak_\mathrm{f}-bk_\mathrm{r}}{I_z} & \dfrac{a^2k_\mathrm{f}+b^2k_\mathrm{r}}{I_zV_x}\end{bmatrix}\begin{bmatrix}\Delta\beta\\ \Delta\omega\end{bmatrix}+\begin{bmatrix}0\\ -\dfrac{1}{I_z}\end{bmatrix}\Delta M \tag{12}$$

令 $\boldsymbol{A}=\begin{bmatrix}\dfrac{k_\mathrm{f}+k_\mathrm{r}}{mV_x} & \dfrac{ak_\mathrm{f}-bk_\mathrm{r}}{mV_x^{\,2}}-1\\ \dfrac{ak_\mathrm{f}-bk_\mathrm{r}}{I_z} & \dfrac{a^2k_\mathrm{f}+b^2k_\mathrm{r}}{I_zV_x}\end{bmatrix}$, $\boldsymbol{B}=\begin{bmatrix}0\\ -\dfrac{1}{I_z}\end{bmatrix}$, $\boldsymbol{X}=\begin{bmatrix}\Delta\beta & \Delta\omega\end{bmatrix}^\mathrm{T}$, $U=\Delta M$, 则式 (12) 可化成标准形式:

$$\dot{\boldsymbol{X}}=\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}U \tag{13}$$

为消除误差, 求出附加横摆力矩最优解U^*, 基于 LQR 理论设计控制方程:

$$\begin{cases}\min J=\dfrac{1}{2}\displaystyle\int_0^\infty(\boldsymbol{X}^\mathrm{T}\boldsymbol{Q}_\mathrm{k}\boldsymbol{X}+U^\mathrm{T}\boldsymbol{R}_\mathrm{k}U)\mathrm{d}t\\ U^*=-\boldsymbol{K}\boldsymbol{X}\\ \boldsymbol{K}=\boldsymbol{R}_\mathrm{k}^{-1}\boldsymbol{B}^\mathrm{T}\boldsymbol{P}_\mathrm{k}\end{cases} \tag{14}$$

式中: $\boldsymbol{K}$为反馈系数矩阵, $\boldsymbol{Q}_\mathrm{k}$矩阵能使系统误差快速逼近 0, $\boldsymbol{R}_\mathrm{k}$矩阵限制系统最大量, 避免输出超出轮毂电机负载的横摆力矩。$\boldsymbol{P}_\mathrm{k}$为 Riccati 方程的解, Riccati 方程为

$$\begin{aligned}&\boldsymbol{A}^\mathrm{T}\boldsymbol{P}_\mathrm{k}+\boldsymbol{P}_\mathrm{k}\boldsymbol{A}+\boldsymbol{Q}_\mathrm{k}-\boldsymbol{P}_\mathrm{k}\boldsymbol{B}\boldsymbol{R}_\mathrm{k}^{-1}\boldsymbol{B}^\mathrm{T}\boldsymbol{P}_\mathrm{k}=\\ &\quad\boldsymbol{A}^\mathrm{T}\boldsymbol{P}_\mathrm{k}+\boldsymbol{P}_\mathrm{k}\boldsymbol{A}+\boldsymbol{Q}_\mathrm{k}-\boldsymbol{P}_\mathrm{k}\boldsymbol{B}\boldsymbol{K}=\boldsymbol{0}\end{aligned} \tag{15}$$

为证明 LQR 控制的稳定性, 定义李雅普诺夫函数为

$$V(\boldsymbol{X})=\boldsymbol{X}^\mathrm{T}\boldsymbol{P}_\mathrm{k}\boldsymbol{X}>0 \tag{16}$$

对$V(X)$求导得

$$\dot V(\boldsymbol{X})=\dot{\boldsymbol{X}}^\mathrm{T}\boldsymbol{P}_\mathrm{k}\boldsymbol{X}+\boldsymbol{X}^\mathrm{T}\boldsymbol{P}_\mathrm{k}\dot{\boldsymbol{X}} \tag{17}$$

当$U=U^*$时,

$$\dot{\boldsymbol{X}}=\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}(-\boldsymbol{K}\boldsymbol{X})=(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{K})\boldsymbol{X} \tag{18}$$

将上式代入式 (17), $\dot V(\boldsymbol{X})$可更新为

$$\begin{aligned}\dot V(\boldsymbol{X})&=(\boldsymbol{A}\boldsymbol{X}-\boldsymbol{B}\boldsymbol{K}\boldsymbol{X})^\mathrm{T}\boldsymbol{P}_\mathrm{k}\boldsymbol{X}+\boldsymbol{X}^\mathrm{T}\boldsymbol{P}_\mathrm{k}(\boldsymbol{A}\boldsymbol{X}-\boldsymbol{B}\boldsymbol{K}\boldsymbol{X})=\\ &\boldsymbol{X}^\mathrm{T}\left(\boldsymbol{P}_\mathrm{k}\boldsymbol{A}+\boldsymbol{A}^\mathrm{T}\boldsymbol{P}_\mathrm{k}-\boldsymbol{P}_\mathrm{k}\boldsymbol{B}\boldsymbol{K}-\boldsymbol{K}^\mathrm{T}\boldsymbol{B}^\mathrm{T}\boldsymbol{P}_\mathrm{k}\right)\boldsymbol{X}=\\ &\boldsymbol{X}^\mathrm{T}\left(-\boldsymbol{Q}_\mathrm{k}-\boldsymbol{K}^\mathrm{T}\boldsymbol{B}^\mathrm{T}\boldsymbol{P}_\mathrm{k}\right)\boldsymbol{X}=\\ &\boldsymbol{X}^\mathrm{T}\left(-\boldsymbol{Q}_\mathrm{k}-\left(\boldsymbol{B}^\mathrm{T}\boldsymbol{P}_\mathrm{k}\right)^\mathrm{T}\boldsymbol{R}_\mathrm{k}^{-1}\left(\boldsymbol{B}^\mathrm{T}\boldsymbol{P}_\mathrm{k}\right)\right)\boldsymbol{X}<0\end{aligned} \tag{19}$$

根据李雅普诺夫稳定性理论, $\dot V(\boldsymbol{X})$负定, 基于 LQR 的闭环动态控制系统趋于稳定。

2.2　基于非线性二自由度模型的鲁棒最优滑模控制

实际行驶中, 车辆轮胎是一个复杂的非线性系统。与普通充气轮胎通过胎压向地面传递力不同, 机械弹性电动轮的接地状态复杂多变, 是由铰链组挤压弹性环来实现触地区域力的传递。如图 6 所示, 假设轮胎的扁平程度不变, 触地区域中线与铰链最小夹角变化范围为 0～(180/n)°(n为铰链组数), 整个过程轮胎侧偏刚度一直变化[23]且摄动程度高于普通轮胎, 因而基于线性模型的稳定性控制缺乏准确性与鲁棒性。

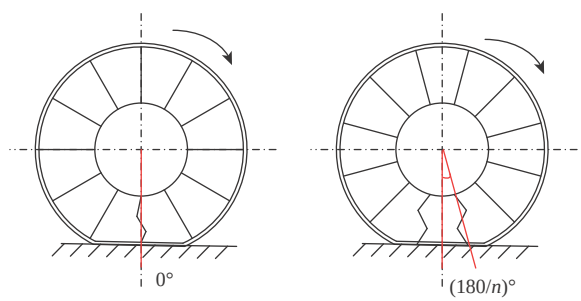


图 6　机械弹性电动轮接地状态

Fig. 6　Grounded state of MEEW

考虑轮胎侧偏刚度、路面状态等因素的变化, 将式 (13) 更新为非线性二自由度模型如下:

$$\dot{\boldsymbol{X}}'=(\boldsymbol{A}+\Delta\boldsymbol{A})\boldsymbol{X}+(\boldsymbol{B}+\Delta\boldsymbol{B})U+\boldsymbol{N} \tag{20}$$

为表征参数的不确定性, 引入时变函数$\Delta\boldsymbol{A}$、$\Delta\boldsymbol{B}$; 为描述随机的路面干扰, 引入时变函数$\boldsymbol{N}$。针对该不确定车辆模型, 若仍沿用前文线性模型下的 LQR 控制器, 则无法保证车辆极限工况下稳定性的控制效果。对此本文设计了一种鲁棒最优滑模控制方法, 与传统的滑模控制不同, 积分滑模控制中避免了趋近阶段。因此, 积分滑模的优点是控制系统总是从初始状态开始就在滑模表面上运行, 可以实现整个动态响应过程的鲁棒性[24]。为该不确定模型设计积分滑模面, 如下所示:

$$S(t)=\boldsymbol{W}(\boldsymbol{X}'(t)-\boldsymbol{X}(t)|_{U=U^*}) \tag{21}$$

式中, $\boldsymbol{W}$为常数矩阵, 且满足: $|\boldsymbol{W}\boldsymbol{B}|\neq0$

对式 (21) 求导得

$$\dot{S}(t)=\boldsymbol{W}((\boldsymbol{A}+\Delta\boldsymbol{A})\boldsymbol{X}+(\boldsymbol{B}+\Delta\boldsymbol{B})U+\boldsymbol{N}-(\boldsymbol{A}-\boldsymbol{BK})\boldsymbol{X})=\boldsymbol{W}((\boldsymbol{A}+\boldsymbol{BK})\boldsymbol{X}+(\boldsymbol{B}+\Delta\boldsymbol{B})U+\boldsymbol{N}) \tag{22}$$

为使切换函数滑动模态稳定，令$\dot{S}(t)=0$，可得等效控制率：

$$U_{\mathrm{eq}}=-(\boldsymbol{W}(\boldsymbol{B}+\Delta\boldsymbol{B}))^{-1}\cdot(\boldsymbol{W}(\Delta\boldsymbol{A}+\boldsymbol{BK})\boldsymbol{X}+\boldsymbol{WN})=-(\boldsymbol{B}+\Delta\boldsymbol{B})^{-1}\cdot((\Delta\boldsymbol{A}+\boldsymbol{BK})\boldsymbol{X}+\boldsymbol{N}) \tag{23}$$

将其代入不确定系统(20)得到：

$$\begin{aligned}\dot{\boldsymbol{X}}'&=(\boldsymbol{A}+\Delta\boldsymbol{A})\boldsymbol{X}+(\boldsymbol{B}+\Delta\boldsymbol{B})U_{\mathrm{eq}}+\boldsymbol{N}=\\&(\boldsymbol{A}+\Delta\boldsymbol{A})\boldsymbol{X}-(\boldsymbol{B}+\Delta\boldsymbol{B})(\boldsymbol{B}+\Delta\boldsymbol{B})^{-1}\cdot\\&((\Delta\boldsymbol{A}+\boldsymbol{BK})\boldsymbol{X}+\boldsymbol{N})+\boldsymbol{N}=\\&(\boldsymbol{A}-\boldsymbol{BK})\boldsymbol{X}\end{aligned} \tag{24}$$

可以看出，式(24)与式(18)一致，表明当不确定系统在所设计的积分滑模面(21)运行时，两者系统状态空间方程相同，未受不确定参数$\Delta\boldsymbol{A}$、$\Delta\boldsymbol{B}$、$\boldsymbol{N}$的影响，即在不确定系统下仍能实现式(14)中目标函数的优化，仍能实现最优控制。在此基础上构造滑模控制率：

$$\begin{cases}U=U^{*}-(\boldsymbol{WB})^{-1}f(x)\mathrm{sgn}(s)=\\\qquad-\boldsymbol{KX}-(\boldsymbol{WB})^{-1}f(x)\mathrm{sgn}(s)\\f(x)=\eta_1\|\boldsymbol{WB}\|\|\boldsymbol{X}\|+\eta_2\|\boldsymbol{WB}\|+\eta_3\end{cases} \tag{25}$$

式中：η_1、η_2、η_3为给定大于0的常数。

为削弱抖振，用饱和函数$\mathrm{sat}(s)$代替式(25)中的符号函数$\mathrm{sgn}(s)$，饱和函数表达式为

$$\mathrm{sat}(s)=\begin{cases}1 & s_2>\Delta\\ s/\Delta & |s_2|\leqslant\Delta\\ -1 & s_2<-\Delta\end{cases} \tag{26}$$

式中：Δ为边界层厚度。

2.3 鲁棒最优滑模控制器稳定性证明

为证明鲁棒最优滑模控制的稳定性，定义李雅普诺夫函数为

$$V=\frac{1}{2}S^{2} \tag{27}$$

对其求导得

$$\dot{V}=S\dot{S}=S\boldsymbol{W}((\boldsymbol{A}+\boldsymbol{BK})\boldsymbol{X}+(\boldsymbol{B}+\Delta\boldsymbol{B})U+\boldsymbol{N})=S\boldsymbol{W}(\Delta\boldsymbol{AX}+\Delta\boldsymbol{B}U+\boldsymbol{N})+S\boldsymbol{WB}(\boldsymbol{KX}+U) \tag{28}$$

在实际行车环境中，不确定项$\Delta\boldsymbol{A}$、$\Delta\boldsymbol{B}$、$\boldsymbol{N}$都是有界的，那么可以定义

$$\begin{cases}\Delta\boldsymbol{A}(t)=\boldsymbol{B}\Delta\boldsymbol{A}'(t)\\\Delta\boldsymbol{B}(t)=\boldsymbol{B}\Delta\boldsymbol{B}'(t)\\\boldsymbol{N}(t)=\boldsymbol{B}\Delta\boldsymbol{N}'(t)\end{cases} \tag{29}$$

式中：$\Delta\boldsymbol{A}'$、$\Delta\boldsymbol{B}'$、$\boldsymbol{N}'$为未知矩阵函数。由于$\Delta\boldsymbol{A}$、$\Delta\boldsymbol{B}$、$\boldsymbol{N}$是有界的，$\boldsymbol{B}$为已知实矩阵，显然$\Delta\boldsymbol{A}'$、$\Delta\boldsymbol{B}'$、$\boldsymbol{N}'$也是有界的，不妨定义总的不确定干扰项$\boldsymbol{\xi}(t)$，具体形式为

$$\boldsymbol{\xi}(t)=\Delta\boldsymbol{A}'(t)\boldsymbol{X}(t)+\Delta\boldsymbol{B}'(t)U(t)+\boldsymbol{N}'(t) \tag{30}$$

由于$\boldsymbol{\xi}(t)$表达式中各项都是有界的，因此可以假定

$$\|\boldsymbol{\xi}(x,t)\|\leqslant\eta_1\|\boldsymbol{X}(t)\|+\eta_2 \tag{31}$$

将式(29)代入式(28)可得

$$\dot{V}=S\boldsymbol{W}(\boldsymbol{B}\Delta\boldsymbol{A}'\boldsymbol{X}+\boldsymbol{B}\Delta\boldsymbol{B}'U+\boldsymbol{BN}')+S\boldsymbol{WB}(\boldsymbol{KX}+U) \tag{32}$$

在满足式(31)的条件下，将控制率式(25)代入式(32)得

$$\begin{aligned}\dot{V}&=S\boldsymbol{W}(\boldsymbol{B}\Delta\boldsymbol{A}'\boldsymbol{X}+\boldsymbol{B}\Delta\boldsymbol{B}'U+\boldsymbol{BN}')+\\&S\boldsymbol{WB}(\boldsymbol{KX}-\boldsymbol{KX}-(\boldsymbol{WB})^{-1}f(x)\mathrm{sgn}(s))=\\&S\boldsymbol{WB}(\Delta\boldsymbol{A}'\boldsymbol{X}+\Delta\boldsymbol{B}'U+\boldsymbol{N}')-S\cdot f(x)\mathrm{sgn}(s)\leqslant\\&S\boldsymbol{WB}(\eta_1\|\boldsymbol{X}\|+\eta_2)-\\&S(\eta_1\|\boldsymbol{WB}\|\|\boldsymbol{X}\|+\eta_2\|\boldsymbol{WB}\|+\eta_3)\mathrm{sgn}(s)\leqslant\\&-\|S\|\cdot\eta_3\leqslant0\end{aligned} \tag{33}$$

综上，根据李雅普诺夫稳定性理论，基于鲁棒最优滑模的闭环动态控制系统趋于稳定。

3 基于二次规划的下层控制器设计

利用分布式电动车各轮毂电机独立可控的特点，采用二次规划算法对上层控制器输出的总附加横摆力矩U进行分配。轮胎附着利用率η是能表征车辆稳定性的参数之一，η越小，车辆越不易抱死，稳定性越强，其表达式为

$$\eta=\frac{F_{xi}{}^{2}+F_{yi}^{2}}{(\mu F_{zi})^{2}}\qquad i=\mathrm{fl,fr,rl,rr} \tag{34}$$

式中：F_{xi}、F_{yi}、F_{zi}分别代表各轮的纵向力、侧向力、垂向力；μ为路面附着系数。

由于直接横摆力矩控制各轮纵向力，故以优化轮胎附着利用率η为目标，设计二次规划目标函数如下：

$$\min J=\min\sum_{i}\eta=\min\sum_{i}\frac{F_{xi}^{2}+F_{yi}^{2}}{(\mu F_{zi})^{2}} \tag{35}$$

依托分布式电动车四轮转矩独立可控的特点，本文中设计的控制器主要通过控制四轮的纵向力分配来达到稳定性控制效果。由于轮胎侧向力与前轮转角、侧偏角等因素相关，DYC无法直接控制其大小，无法进行优化，因此可以将轮胎利用率改写为[25]

$$\min J=\min\sum_{i}\frac{F_{xi}^{2}}{(\mu F_{zi})^{2}}=\min\sum_{i}\frac{T_{xi}^{2}}{(\mu F_{zi}^{r})^{2}} \tag{36}$$

式中：T_{xi}为各轮纵向驱动力；r为车轮滚动半径。车

轮总纵向力矩T_x与总附加横摆力矩U存在如下关系:

$$\begin{cases} T_x = T_{x\mathrm{fl}} + T_{x\mathrm{fr}} + T_{x\mathrm{rl}} + T_{x\mathrm{rr}} \\ U = \dfrac{d}{2r}(-T_{x\mathrm{fl}} + T_{x\mathrm{fr}} - T_{x\mathrm{rl}} + T_{x\mathrm{rr}}) \end{cases} \tag{37}$$

式中: d为轮距。令$\boldsymbol{x} = [\ T_{x\mathrm{fl}} \quad T_{x\mathrm{fr}} \quad T_{x\mathrm{rl}} \quad T_{x\mathrm{rr}}\ ]^{\mathrm{T}}$, $\boldsymbol{b}_{\mathrm{eq}} = [\ T_x \quad U\ ]^{\mathrm{T}}$, $\boldsymbol{A}_{\mathrm{eq}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -\dfrac{d}{2r} & \dfrac{d}{2r} & -\dfrac{d}{2r} & \dfrac{d}{2r} \end{bmatrix}$, 则

$$\boldsymbol{A}_{\mathrm{eq}}\boldsymbol{x} = \boldsymbol{b}_{\mathrm{eq}} \tag{38}$$

考虑到轮胎力不得超过路面提供的附着力, 设置约束条件:

$$-\mu F_{zi} r \leqslant T_{xi} \leqslant \mu F_{zi} r \qquad i = \mathrm{fl,fr,rl,rr} \tag{39}$$

考虑驱动轮的理想输出转矩不得超过轮毂电机的最大输出转矩, 设置约束条件:

$$-T_{\max} \leqslant T_{xi} \leqslant T_{\max} \tag{40}$$

式中, $T_{\max}$由电机外特性曲线得到。

综上所述, 可将式 (36) 的二次规划问题描述成标准形式:

$$\begin{cases} \min\ J = \min \boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{x} \\ \quad \text{s.t.} \quad \boldsymbol{A}_{\mathrm{eq}}\boldsymbol{x} = \boldsymbol{b}_{\mathrm{eq}} \\ \qquad \max(-\mu F_{zi} r, -T_{\max}) \leqslant T_{xi} \leqslant \min(\mu F_{zi} r, T_{\max}) \end{cases} \tag{41}$$

式中:

$$\boldsymbol{G} = \mathrm{diag}\left[\frac{1}{(\mu F_{z\mathrm{fl}} r)^2} \quad \frac{1}{(\mu F_{z\mathrm{fr}} r)^2} \quad \frac{1}{(\mu F_{z\mathrm{rl}} r)^2} \quad \frac{1}{(\mu F_{z\mathrm{rr}} r)^2} \right]$$

该二次规划问题可在 MATLAB 中调用 quadprog 函数求解。

4　仿真验证

为了验证本文设计的鲁棒最优滑模控制器的有效性, 利用 MATLAB/Simulink 与车辆动力学仿真软件 Carsim 搭建联合仿真平台进行试验验证。在 Carsim 中搭建 MEEW 车辆模型, 具体参数如表 2 所示。本文中 ROSM 控制器参数为η_1=100, η_2 = 3 000, η_3=100, $\boldsymbol{R}_k = 3\times10^{-5}$, $\boldsymbol{Q}_k = \mathrm{diag}(20\,000, 20\,000)$, 边界层厚度$\Delta = 10$。

表 2　车辆参数

Table 2　Vehicle parameters

参数	数值
车辆质量 m/kg	1 610
转动惯量 I_z/(kg·m^2)	2 059.2
质心到前轴距离 a/m	1.05
质心到后轴距离 b/m	1.61
轮距 d/m	1.565
轮胎滚动半径 R/m	0.35
前轮等效刚度 k_f/(N·rad^{-1})	−87 002
后轮等效刚度 k_r/(N·rad^{-1})	−79 240

4.1　双移线工况

设置车速为 72 km/h, 路面附着系数为 0.3 的双移线工况。图 7 为横摆角速度对比图, 当无控制时, 曲线发散, 车辆失稳。单一 LQR 控制与鲁棒最优滑模控制都能极大改善车辆行驶的稳定性, 但可明显看出鲁棒最优滑模控制效果曲线更贴近理想值, 横摆角速度与理想值的误差峰值仅为 0.94(°)/s, 而单一 LQR 控制误差峰值为 3.75(°)/s, 降低了 74.93%。

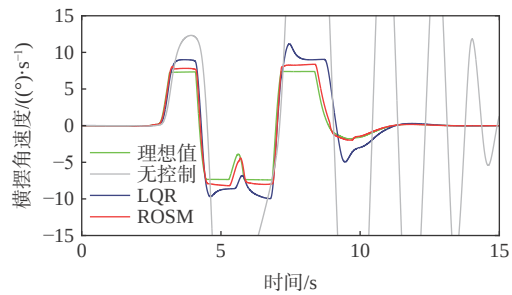


图 7　横摆角速度变化曲线

Fig. 7　Yaw rate variation curve

图 8 为质心侧偏角对比图, 可明显看出鲁棒最优滑模控制效果更好, 质心侧偏角与理想值的误差峰值仅为 3.03°, 而单一 LQR 控制误差峰值为 5.04°, 峰值降低了 39.88%。

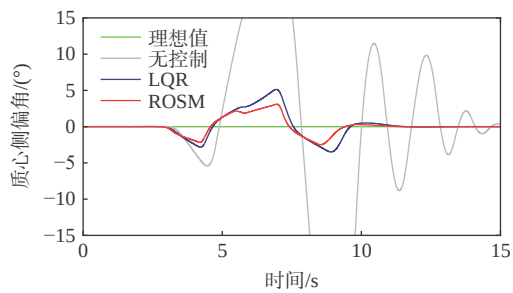


图 8　质心侧偏角变化曲线

Fig. 8　Slide slip angle variation curve

图 9 为鲁棒最优滑模控制下的车辆四轮转矩图。车辆在低附着路面转向时易侧滑、倾翻, 由图可看出在 4 s 与 7 s 左右转向时, 电机执行器动态补偿了轮胎转矩, 极大提升了车辆的稳定性。

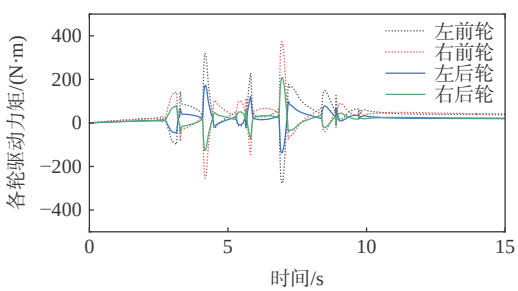


图 9　四轮转矩变化(鲁棒最优滑模控制)

Fig. 9　Four -wheel torque variation (robust optimal sliding mode control)

图10为车速变化对比图，期望车速为72 km/h，LQR控制器与鲁棒最优滑模控制器都能很好地稳定车速，保证车辆的平顺性与驾驶人的乘坐体验。

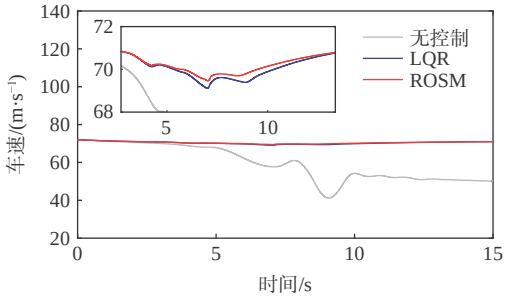


图10 车速变化曲线

Fig. 10 Speed variation curve

图11为不同控制下的“质心侧偏角-质心侧偏角速度”相图。无控制时，参数发散，车辆失稳；LQR控制与鲁棒最优滑模控制都能使曲线收敛，但鲁棒最优滑模控制收敛速度更快、半径更小、效果更佳。

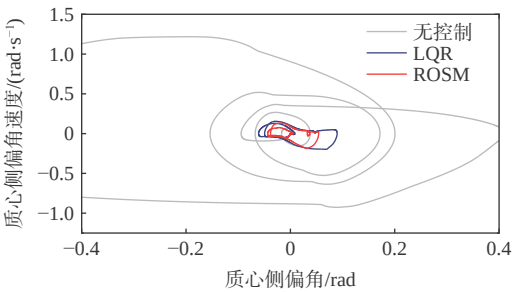


图11 “质心侧偏角-质心侧偏角速度”相图

Fig. 11 ‘Centre-of-mass lateral deflection angle - centre-of-mass lateral deflection angular velocity’ phase Diagrams

4.2 蛇形工况

设置车速为72km/h，路面附着系数为0.3的蛇形工况。从图12、图13可知，在蛇形工况下，ROSM控制下的质心侧偏角、横摆角速度曲线仍更贴合理想值，说明ROSM的稳定性控制效果优于LQR控制。从图14可知，控制器在蛇形转弯处补偿转矩，保证了车辆的操稳性。图15、图16与双移线工况

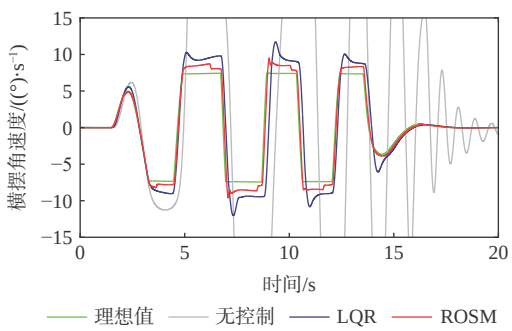


图12 横摆角速度变化曲线

Fig. 12 Yaw rate variation curve

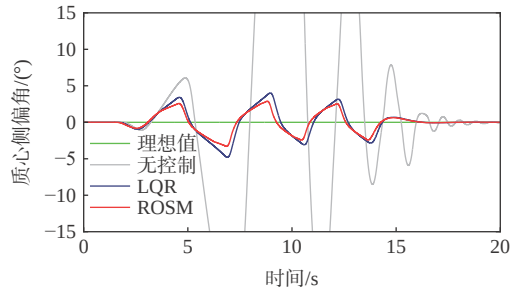


图13 质心侧偏角变化曲线

Fig. 13 Slide slip angle variation curve

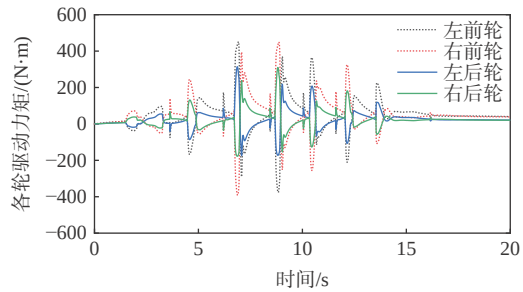


图14 四轮转矩变化(鲁棒最优滑模控制)

Fig. 14 Four -wheel torque variation (robust optimal sliding mode control)

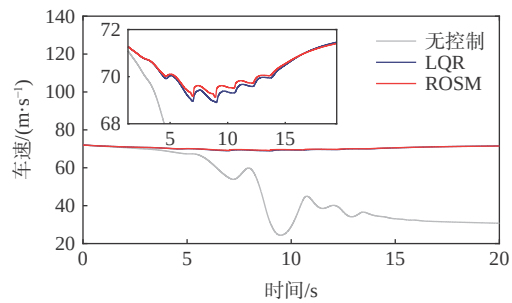


图15 车速变化曲线

Fig. 15 Speed variation curve

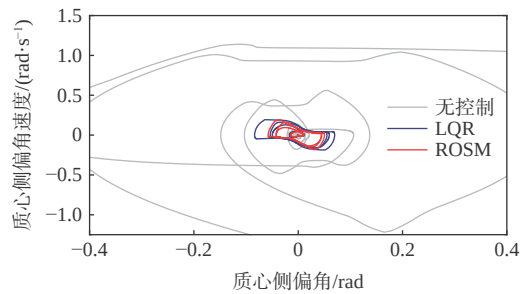


图16 “质心侧偏角-质心侧偏角速度”相图

Fig. 16 ‘Centre-of-mass lateral deflection angle - centre-of-mass lateral deflection angular velocity’ phase Diagrams

类似，ROSM控制下，车速保持效果略好于LQR控制，相图中ROSM控制效果最佳。

横摆角速度与质心侧偏角是评估车辆稳定性的两个重要指标，为了评估二者跟踪理想值的效果，引入平均绝对误差(MAE)δ_{MAE}和均方根误差(RMSE)δ_{RMSE}的评估指标，具体为

$$\delta_{\mathrm{MAE}}=\frac{1}{k}\sum_{i=1}^{k}|x-x_{\mathrm{d}}| \tag{42}$$

$$\delta_{\mathrm{RMSE}}=\sqrt{\frac{1}{k}\sum_{i=1}^{k}(x-x_{\mathrm{d}})^2} \tag{43}$$

式中: x指代ω或β; x_{d}为理想值; k为采样样本数, 计算结果如表 3 所示。

表 3　评估指标对比

Table 3　Comparison of evaluation indicators

工况	$\omega\delta_{\mathrm{MAE}}$	$\omega\delta_{\mathrm{RSME}}$	$\beta\delta_{\mathrm{MAE}}$	$\beta\delta_{\mathrm{RSME}}$
双移线LQR	1.016 1	1.828 5	1.027 0	1.713 9
双移线ROSM	0.367 5	0.633 9	0.707 0	1.166 3
蛇形LQR	1.494 4	2.279 1	1.244 6	1.786 3
蛇形ROSM	0.622 0	0.911 1	0.973 2	1.367 9

从表中可明显看出, 无论双移线工况还是蛇形工况, 鲁棒最优滑模控制下ω与β的 δ_{MAE} 与 δ_{RSME} 都小于相同工况下的单一 LQR 控制, 更能保证车辆在极限工况下行驶的稳定性。双移线工况中, ROSM 控制下横摆角速度的 δ_{MAE}、δ_{RSME} 与 LQR 控制相比分别下降了 63.83%、65.33%; 蛇形工况中, 其分别降低了 58.38%、60.02%, 说明文中设计的 ROSM 控制器控制效果较之 LQR 控制有明显提升。质心侧偏角方面, 双移线下其 δ_{MAE}、δ_{RSME} 与 LQR 控制相比分别下降了 31.16%、31.95%, 而蛇形工况下, 其分别下降了 21.81%、23.42%, 说明鲁棒最优控制器能有效改善车辆的操稳性。

5　结　论

本文针对匹配机械弹性电动轮 (MEEW) 车辆转向时的横摆稳定性控制问题, 提出了一种鲁棒最优滑模的控制策略。首先, 基于线性二自由度车辆模型, 设计线性二自由度调节器 (LQR) 控制器, 输出初始最优控制量; 其次, 考虑到轮胎侧偏刚度、路面状态等因素的变化, 引入参数不确定的非线性动力学模型, 为了处理参数的不确定性与外部干扰, 设计了一种基于初始最优控制量的鲁棒最优积分滑模控制策略 (ROSM), 该策略不仅可以实现最优控制目标, 而且对不确定参数与外部干扰具有良好的鲁棒性; 最后通过 MATLAB/Simulink 和 Carsim 联合仿真, 对控制方法进行仿真验证, 结果表明: 在双移线与蛇形工况下, 鲁棒最优滑模控制车辆稳定性, 其横摆角速度、质心侧偏角与理想值的平均绝对误差、均方根误差皆小于相同工况下的 LQR 控制, 跟踪理想值精度高于 LQR 控制, 控制效果更佳。

参考文献 (References)

[1] 赵又群. 非充气机械弹性安全车轮理论与方法[M]. 北京: 科学出版社, 2020: 12-28.
ZHAO Y Q. Theory and method for non-pneumatic mechanical elastic safety wheel[M]. Beijing: Science Press, 2020: 12-28 (in Chinese).

[2] DENG Y J, ZHAO Y Q, XU H, et al. Finite element modeling of interaction between non-pneumatic mechanical elastic wheel and soil[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2019, 233(13): 3293-3304.

[3] 张陈曦, 赵又群, 冯世林, 等. 伪刚体-柔体耦合的新式免充气轮胎刚度分析[J]. 中国机械工程, 2021, 32(9): 1051-1060,1072.
ZHANG C X, ZHAO Y Q, FENG S L, et al. Stiffness analysis of new type non-pneumatic tires based on pseudo-rigid-flexible body coupling model[J]. China Mechanical Engineering, 2021, 32(9): 1051-1060,1072(in Chinese).

[4] 张晨, 赵又群, 郑鑫, 等. 随机载荷下机械弹性车轮的热力耦合耐久性研究[J]. 中国机械工程, 2021, 32(14): 1669-1676.
ZHANG C, ZHAO Y Q, ZHENG X, et al. Study on thermal coupling durability of mechanical elastic wheels under random loads[J]. China Mechanical Engineering, 2021, 32(14): 1669-1676(in Chinese).

[5] WANG Q G, ZHUANG Y, WEI J N, et al. A driver model–based direct yaw moment controller for in-wheel motor electric vehicles[J]. Advances in Mechanical Engineering, 2019, 11(9): 1687814019877319.

[6] GAN L. Study on yaw moment control for electric vehicle with four-wheel in-wheel motor based on fuzzy PI control[J]. Machinery Design & Manufacture, 2015, 7: 103-107.

[7] ASIABAR A N, KAZEMI R. A direct yaw moment controller for a four in-wheel motor drive electric vehicle using adaptive sliding mode control[J]. Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics, 2019, 233(3): 549-567.

[8] 段敏, 孙小松, 张博涵. 基于模型预测控制与离散线性二次型调节器的智能车横纵解耦跟踪控制[J]. 汽车技术, 2022(8): 38-46.
DUAN M, SUN X S, ZHANG B H. Horizontal and vertical decoupling tracking control based on MPC and DLQR[J]. Automobile Technology, 2022(8): 38-46(in Chinese).

[9] LIU D Y, HUANG S, WU S, et al. Direct yaw-moment control of electric vehicle with in-wheel motor drive system[J]. International Journal of Automotive Technology, 2020, 21(4): 1013-1028.

[10] 林棻, 蔡亦璋, 赵又群, 等. 匹配机械弹性车轮的分布式驱动电动汽车稳定性控制[J]. 机械工程学报, 2022, 58(8): 236-243.
LIN F, CAI Y Z, ZHAO Y Q, et al. Lateral stability control of distributed drive electric vehicle with mechanical elastic wheel[J]. Journal of Mechanical Engineering, 2022, 58(8): 236-243(in Chinese).

[11] ZHU J J, WANG Z P, ZHANG L, et al. Braking/steering coordination control for in-wheel motor drive electric vehicles based on nonlinear model predictive control[J]. Mechanism and Machine Theory, 2019, 142: 103586.

[12] YAN Z P, WANG M, XU J. Robust adaptive sliding mode control

of underactuated autonomous underwater vehicles with uncertain dynamics[J]. Ocean Engineering, 2019, 173: 802-809.

[13] RANGEL M A G, MANZANILLA A, SUAREZ A E Z, et al. Adaptive non-singular terminal sliding mode control for an unmanned underwater vehicle: real-time experiments[J]. International Journal of Control, Automation and Systems, 2020, 18(3): 615-628.

[14] NORSAHPERI N M H, DANAPALASINGAM K A. An improved optimal integral sliding mode control for uncertain robotic manipulators with reduced tracking error, chattering, and energy consumption[J]. Mechanical Systems and Signal Processing, 2020, 142: 106747.

[15] XU T, ZHAO Y Q, DENG H F, et al. Integrated optimal control of distributed in-wheel motor drive electric vehicle in consideration of the stability and economy[J]. Energy, 2023, 282: 128990.

[16] 丛森森, 高峰, 许述财. 基于动态稳定域的车辆横纵向稳定性协同控制[J]. 汽车工程, 2022, 44(6): 900-908.
CONG S S, GAO F, XU S C. Cooperative control of vehicle lateral and longitudinal stability based on dynamic stability region[J]. Automotive Engineering, 2022, 44(6): 900-908(in Chinese).

[17] 梁宝钰, 汪怡平, 刘珣, 等. 基于滑模理论的高速车辆侧风稳定性控制研究[J]. 汽车工程, 2022, 44(1): 123-130.
LIANG B Y, WANG Y P, LIU X, et al. Study on crosswind stability control of high-speed vehicle based on sliding mode theory[J]. Automotive Engineering, 2022, 44(1): 123-130(in Chinese).

[18] 陈特, 徐兴, 蔡英凤, 等. 基于状态估计的无人车前轮转角与横摆稳定协调控制[J]. 北京理工大学学报, 2021, 41(10): 1050-1057.
CHEN T, XU X, CAI Y F, et al. Coordinated control of front-wheel steering angle and yaw stability for unmanned ground vehicle based on state estimation[J]. Transactions of Beijing Institute of Technology, 2021, 41(10): 1050-1057(in Chinese).

[19] BAI R, WANG H B. Robust optimal control for the vehicle suspension system with uncertainties[J]. IEEE Transactions on Cybernetics, 2022, 52(9): 9263-9273.

[20] DENG H F, ZHAO Y Q, FENG S L, et al. Torque vectoring algorithm based on mechanical elastic electric wheels with consideration of the stability and economy[J]. Energy, 2021, 219: 119643.

[21] PACEJKA H. Tire and vehicle dynamics[M]. Amsterdam: Elsevier, 2005: 27-29.

[22] 龙文, 刘豪. 车辆稳定性控制系统 LQR 算法设计[J]. 汽车实用技术, 2021, 46(23): 76-79,110.
LONG W, LIU H. Design of vehicle system dynamics control algorithm based on LQR method[J]. Automobile Applied Technology, 2021, 46(23): 76-79,110(in Chinese).

[23] 郑鑫, 赵又群, 王秋伟, 等. 匹配机械弹性车轮的电子稳定控制器参数分析[J]. 中国机械工程, 2020, 31(23): 2883-2890.
ZHENG X, ZHAO Y Q, WANG Q W, et al. Parameter analysis of electronic stability controller matching mechanical elastic wheels[J]. China Mechanical Engineering, 2020, 31(23): 2883-2890(in Chinese).

[24] ALI MASOOD CHEEMA M, FLETCHER J E, FARSHADNIA M, et al. Sliding mode based combined speed and direct thrust force control of linear permanent magnet synchronous motors with first-order plus integral sliding condition[J]. IEEE Transactions on Power Electronics, 2018, 34(3): 2526-2538.

[25] 王文伟, 赵一凡, 张伟, 等. 多轴轮边驱动铰接客车的横摆稳定性控制策略[J]. 机械工程学报, 2020, 56(14): 161-172.
WANG W W, ZHAO Y F, ZHANG W, et al. Yaw stability control strategy of multi-wheel independent electric articulated bus[J]. Journal of Mechanical Engineering, 2020, 56(14): 161-172(in Chinese).

Stability control of vehicles powered by non-pneumatic wheels based on robust optimal sliding mode

LI Tian，ZHAO Youqun*，XU Tao，SHEN Yawei，LIN Fen

(College of Energy and Power Engineering，Nanjing University of Aeronautics and Astronautics，Nanjing 210016，China)

Abstract：Mechanical elastic electric wheel is a new type of non-pneumatic tire, which has the advantages of explosion-proof, anti-puncture, etc. In this paper, based on the distributed vehicle matching with the electric wheel, a robust optimal sliding mode (ROSM) control strategy is proposed to improve the yaw stability of the vehicle. Firstly, the initial additional yaw moment is calculated using a linear 2-DOF regulator (LQR) control rule for the linear 2-DOF model. Secondly, considering the complex nonlinearity of the actual vehicle system, a vehicle dynamics model with uncertain parameters is established. And a robust integral sliding mode controller is designed on the basis of the initial optimal control law. The improved controller has good robustness to uncertain parameters and external disturbances, which could still achieve optimal control effectiveness. Finally, the control scheme is verified by co-simulations of MATLAB/Simulink and Carsim. The findings demonstrate that, in the simulation of double lane change, the mean absolute error (MAE) and root mean square error (RMSE) of the yaw rate under ROSM control are decreased by 63.83% and 65.33%, respectively, in comparison to LQR control. As for the serpentine condition, they respectively decreased by 58.38% and 60.02%.

Keywords：non-pneumatic wheels；direct yaw moment；LQR；robust optimal control；sliding mode control

Received：2023-05-11；**Accepted**：2023-09-25；**Published Online**：2023-10-11 08：33

URL：link.cnki.net/urlid/11.2625.v.20231009.1357.004

Foundation items：the National Natural Science Foundation of China (52272397,11672127); the Fundamental Research Funds for the Central Universities (NP2022408); the National Engineering Laboratory of High Mobility anti-riot vehicle technology (B20210017); Postgraduate Research & Practice Innovation Program of NUAA (20220211)

* **Corresponding author**. E-mail：yqzhao@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0160

月面应急直接上升交会轨迹快速规划方法

班焕恒[1,2]，周聪[1,2]，闫晓东[1,2,*]

(1. 西北工业大学 航天学院，西安 710072；2. 陕西省空天飞行器设计重点实验室，西安 710068)

摘 要：当月面探测器出现某些突发状况而需要返回环月轨道器或地球时，上升器需具备自主规划出合适的应急上升交会轨迹的能力。以上升交会时间最短为目标函数，基于序列二阶锥规划方法，建立直接上升交会轨迹优化模型，并构建求解直接上升交会轨迹优化问题的凸优化算法。为提高计算效率，对内点法进行定制化改进，主要改进有：线性方程组求解过程定制化改进；内点法热启动；二阶锥规划子问题求解精度动态调整。仿真结果表明：所提月面应急上升交会轨迹优化方法可实现上升器的快速上升交会。与采用通用内点法求解器的传统序列二阶锥规划方法相比，在求解精度不变的情况下，所提的序列二阶锥规划问题求解加速方法可达到9.5倍左右的加速比。综合运用所提方法，有望实现月面上升器的自主在线轨迹规划。

关 键 词：月面应急上升交会；序列二阶锥规划方法；内点法定制化；内点法热启动；在线轨迹规划

中图分类号：V448.2

文献标志码：A **文章编号**：1001-5965(2025)04-1352-15

随着科学技术的发展，人类对于太空的探索愈加频繁，实施月球探测工程具有经济、科技和国家发展战略等方面的重大意义[1]。由于月面工作环境较为恶劣，探测器在进行采样或探测活动时可能会出现某些突发状况而需要应急返回的情况，此时，上升器需要自主规划出合适的应急上升交会轨迹。

对于上升器自主上升交会的研究可分为上升段的研究及交会段的研究。对于上升段轨迹优化的研究，通常采用的是传统优化方法，即间接法和直接法。为减少月面上升的燃料消耗量，文献[2-4]基于极大/极小值推导了月面上升段燃料最优或次优的制导律，但基于极大/极小值原理的间接法对于初始猜想较为敏感，通过间接法求解轨迹优化问题的局限性较大。相较于间接法，直接法在优化问题的转化及求解上更具通用性。在基于直接法的上升段轨迹优化中，研究人员分别基于序列二次规划方法[5-6]和序列凸优化方法[7-9]展开了大量的研究。相比于序列二次规划方法，基于凸优化的方法由于其计算的快速性而受到广泛的关注，但受限于箭载或嵌入式计算机的性能，序列凸优化问题的求解效率仍制约着该方法实时在线应用。

交会段的轨迹优化通常采用智能优化方法及传统优化方法。智能优化方法主要应用于脉冲交会问题的优化[10-12]，对于应急返回情况下的连续推力上升交会适用性较差。传统优化方法主要包括间接法和直接法，基于间接法的交会段轨迹优化[13-14]对于初值的依赖性较大，文献[15-16]对于不同的初值计算方法进行了深入的研究，但相应的方法使得轨迹优化问题的复杂度与计算耗时均有所增加。在基于直接法的交会段轨迹优化中，研究人员应用了序列二次规划方法[17]、非线性规划法[18]及序列凸优化方法[19-23]对交会段的轨迹优化展开了深入的研究，但相关的研究主要聚焦于优化模型的建立，未对优化问题的求解过程进行研究，因而相应的优化问题求解效率较低。

对于月面应急上升交会的轨迹优化问题，如果

收稿日期：2023-04-04；**录用日期**：2023-07-11；**网络出版时间**：2023-07-31 09:25
网络出版地址：link.cnki.net/urlid/11.2625.V.20230728.1648.002
***通信作者**. E-mail：yan804@nwpu.edu.cn

引用格式：班焕恒，周聪，闫晓东. 月面应急直接上升交会轨迹快速规划方法[J]. 北京航空航天大学学报，2025，51(4)：1352-1366.
BAN H H，ZHOU C，YAN X D. Rapid planning method for lunar direct ascent and rendezvous trajectory in emergency cases[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1352-1366 (in Chinese).

仍将上升段与交会段分开研究, 则由于未能考虑上升段和交会段的相互影响, 上升交会耗时长, 在应急返回情况下不适用。目前对于月面上升交会联合轨迹优化的研究较少, 文献 [24] 提出了基于高斯伪谱法与序列二次规划的两阶段联合轨迹优化方法, 但相应的方法求解耗时极长, 难以实现在线应用。

在基于序列凸优化的上升与交会轨迹优化方法中, 目前的研究主要聚焦于优化模型的建立与离散方法的选择上, 而在求解相应的优化问题时, 采用的多为通用求解器, 如 SeDuMi[25]、ECOS[26] 等。尽管成熟的通用求解器极大方便了凸优化方法研究, 但在应用于工程实践时存在较多问题: 一方面通用求解器针对通用问题开发, 其算法复杂且未对具体问题进行特定优化, 导致其计算速度通常较慢, 难以实现在线应用; 另一方面, 复杂的代码结构使其难以转化成为嵌入式或箭载计算机使用的代码。

为提高凸优化算法的计算速度, 研究人员对内点法的代码自动生成定制化方法[27-28] 展开了深入的研究, 但该方法仅对小规模的问题加速效果较好。而且, 受限于嵌入式计算机的储存空间限制, 基于代码自动生成的定制化方法在嵌入式平台上的应用会受到较多限制。随着 GPU 计算加速方法的发展, 利用 CPU-GPU 异构平台对内点法计算过程进行并行加速也受到人们的关注[29], 但相应的方法仅适用于大规模的优化问题。对于内点法计算过程中的方程组求解问题, 文献 [30] 通过降低方程组求解精度的方法提高计算速度, 但此方法可能会影响内点法最终收敛精度, 因此, 仅对于精度要求较低的问题适用。

除了传统的内点法求解器外, 近年来快速发展的一阶算法为凸优化问题求解提供了更多的选择, 如基于交替方向乘子法 (alternating direction method of multipliers, ADMM) 算法[31] 的分裂锥求解器 (splitting conic solver, SCS)[32] 及比例积分投影梯度 (proportional-integral projected gradient, PIPG) 算法[33] 等。相比于内点法, 一阶算法的优点在于容易实现热启动, 但总迭代次数较多, 当要求的精度较高时, 内点法在计算速度上仍具有明显的优势。而且, 内点法的热启动技术近年来也得到较大突破, 文献 [34] 提出了一种内点法的热启动方法, 该方法将前一个优化问题的最优解与冷启动值的组合作为热启动初值, 以加快内点法计算速度。为减少序列二阶锥规划方法的内点法总迭代次数, 文献 [35] 进一步将热启动初值的计算扩展为前一个优化问题的非最优解与冷启动值的组合。文献 [36] 基于二阶锥规划热启动方法提出了一种火星垂直着陆的模型预测控制方法, 但该方法在求解初始优化问题时仍需要根据传统内点法计算高精度解。针对大气层内的轨迹优化问题, 文献 [37] 对内点法中 KKT 矩阵进行稀疏化处理, 同时结合热启动方法, 显著加快了序列二阶锥规划方法的求解速度, 但文中引入了额外的轨迹积分方法作为序列二阶锥规划方法终止条件的判断, 而这又增加了相应的计算耗时。

综上所述, 本文基于序列二阶锥规划方法, 提出了一种月面应急上升交会联合轨迹优化方法, 并对内点法进行改进以提高优化问题的求解效率。主要的改进策略有: ①综合应用定制化、并行及动态精度调整的方法提高内点法 KKT 方程组的求解速度; ②采用热启动策略, 加快内点法的收敛速度; ③设计子问题求解精度动态调整策略降低内点法求解的总迭代次数。仿真结果表明, 本文月面应急上升交会轨迹优化方法可实现上升器与轨道器的短时间上升交会。同时, 本文提出的序列二阶锥规划问题求解加速方法相比通用求解器可实现 9.5 倍左右的计算速度提升, 综合运用本文方法, 有望实现上升器的自主在线轨迹规划。

1　月面应急上升交会轨迹规划问题

本文在月心球面坐标系下建立轨迹优化模型, 月心球面坐标系定义如图 1 所示。图中: 月心球面坐标系的原点 O 在月心, R 为月心到飞行器连线 OS 的长度; Ox 轴在交会轨道面内, 指向上升器位置在交会轨道面内投影的方向; Oy 轴在轨道面内, 指向轨道器经过 Ox 轴时的方向; Oz 轴垂直于交会轨道面, 与 Ox 轴和 Oy 轴成右手坐标系, $Oxyz$ 在上升器起飞后不动; θ 为 x 轴按逆时针方向旋转到飞行器在轨道面内的投影所转过的角度; ϕ 为 OS 与轨道面的夹角。定义单位矢量方向 $[e_r, e_t, e_p]^{\mathrm{T}}$ 如下: e_r 沿 OS 的方向; e_p 在平面 zOS 内, 垂直于 e_r, 指向 ϕ 增大的方向; e_t 垂直于 OS, 指向 θ 增大的方向, 且与 e_r 和 e_p 构成直角坐标系。

本文的月面应急上升交会方案如图 2 所示。图中: 在发生应急状况时, 上升器在采样区等待, 当

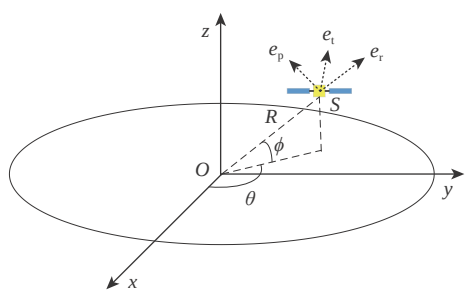


图 1　月心球面坐标系

Fig. 1　Lunar center spherical coordinate system

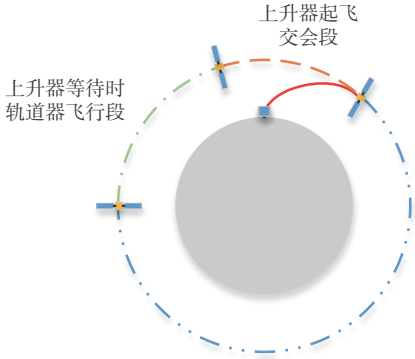


图2 月面应急上升交会示意图

Fig. 2 Lunar ascent and rendezvous process in emergency cases

轨道器飞行到合适的位置时，上升器起飞直接与轨道器交会。

在月心球面坐标系下建立上升器的动力学方程组如下：

$$\begin{cases}\dot{R}(t)=V_r(t)\\ \dot{\theta}(t)=\dfrac{V_t(t)}{R(t)\cos\phi(t)}\\ \dot{\phi}(t)=\dfrac{V_p(t)}{R(t)}\\ \dot{V}_r(t)=\dfrac{V_t^2(t)}{R(t)}+\dfrac{V_p^2(t)}{R(t)}-\dfrac{\mu_{\mathrm{m}}}{R^2(t)}+P\dfrac{u_r(t)}{m(t)}\\ \dot{V}_t(t)=-\dfrac{V_r(t)V_t(t)}{R(t)}+\dfrac{V_t(t)V_p(t)\tan\phi(t)}{R(t)}+P\dfrac{u_t(t)}{m(t)}\\ \dot{V}_p(t)=-\dfrac{V_r(t)V_p(t)}{R(t)}-\dfrac{V_t^2(t)\tan\phi(t)}{R(t)}+P\dfrac{u_p(t)}{m(t)}\\ \dot{m}=-P/(I_{\mathrm{sp}}g_0)\end{cases}\tag{1}$$

式中：R为径向距离；θ为相角；ϕ为极角；V_r、V_t和V_p为上升器的速度在月心球坐标系下的各分量；μ_{m}为月球引力常量；P为上升器的推力；u_r、u_t和u_p为上升器推力方向在月心球坐标系下的各分量；m为上升器的质量；I_{sp}为上升器发动机比冲；g_0为地球引力加速度。

本文假设轨道器在上升器到达交会轨道前，一直处于其停泊轨道上，且轨道器的停泊轨道为圆轨道。在月心球面坐标系下建立轨道器的动力学方程组，在上升器的上升交会阶段，轨道器仅有相角发生改变，即

$$\dot{\theta}_{\mathrm{o}}=\omega_{\mathrm{o}}\tag{2}$$

式中：θ_{o}为轨道器的相角；ω_{o}为轨道器在其停泊轨道上的角速度。

上升器的初始状态约束如下：

$$\begin{cases}R(t_0)=R_0,\theta(t_0)=\theta_0,\phi(t_0)=\phi_0\\ V_r(t_0)=V_{r0}\ ,V_t(t_0)=V_{t0}\ ,V_p(t_0)=V_{p0}\\ m(t_0)=m_0\end{cases}\tag{3}$$

式中：R_0为初始径向距离；θ_0为初始相角；ϕ_0为初始极角；V_{r0}、V_{t0}和V_{p0}为上升器的初始速度在月心球坐标系下的各分量；m_0为上升器初始质量。上升器与轨道器的交会约束可表示如下：

$$\begin{cases}R(t_{\mathrm{f}})=R_{\mathrm{f}},\theta(t_{\mathrm{f}})=\theta_{o0}+\omega_o t_{\mathrm{f}},\phi(t_{\mathrm{f}})=0\\ V_r(t_{\mathrm{f}})=0\ ,V_t(t_{\mathrm{f}})=V_{\mathrm{f}}\ ,V_p(t_{\mathrm{f}})=0\\ m(t_{\mathrm{f}})\geqslant m_{\mathrm{dry}}\end{cases}\tag{4}$$

式中：t_{f}为终端交会时间；R_{f}为交会轨道的月心距；θ_{o0}为轨道器初始相角；$V_{\mathrm{f}}=\sqrt{\mu_{\mathrm{m}}/R_{\mathrm{f}}}$表示交会速度；$m_{\mathrm{dry}}$为上升器结构和载荷质量。

为避免上升器撞到月球表面，对月心距R作如下约束：

$$R(t)\geqslant R_{\mathrm{m}}\tag{5}$$

式中：R_{m}为月球半径。上升器推力大小约束如下：

$$\sqrt{u_r^2(t)+u_t^2(t)+u_p^2(t)}=1\tag{6}$$

为方便求解轨迹优化问题，对式(6)的控制约束作如下松弛：

$$\sqrt{u_r^2(t)+u_t^2(t)+u_p^2(t)}\leqslant 1\tag{7}$$

松弛的等价性可参考文献[9]。

由此可得到月面应急上升交会轨迹优化模型的连续轨迹优化问题(optimal continuous problem, OCP)如下：

$$\begin{cases}\min\ J=t_{\mathrm{f}}\\ \text{s.t. 式(1)、式(3)~式(5)、式(7)}\end{cases}\tag{8}$$

2 月面应急上升交会离散优化问题

由于上升器所携带的燃料较少，而主发动机无法二次启动，上升器仅由主发动机无法完成直接上升交会。因此，本文将直接上升交会分为2段，第1段为主发动机工作段，第2段为推力器工作段。

在各飞行阶段，上升器的推力大小不变，因此，质量仅与时间相关，本文以各段飞行时间为变量，即可略去状态量中的质量变量。为表达的简洁，取$\boldsymbol{X}(t)=[R(t),\theta(t),\phi(t),V_r(t),V_t(t),V_p(t)]^{\mathrm{T}}$代表状态量，$\boldsymbol{U}(t)=[u_r(t),u_t(t),u_p(t)]^{\mathrm{T}}$代表控制量，$\boldsymbol{f}(\boldsymbol{X}(t),\boldsymbol{U}(t))$代表式(1)的动力学方程组的右端项。假设$\boldsymbol{X}_1$和$\boldsymbol{U}_1$为第1段的状态量和控制量，$\boldsymbol{X}_2$和$\boldsymbol{U}_2$为第2段的状态量和控制量。

由于OCP为无限维的非线性优化问题，为对OCP快速求解，需对其进行离散化和凸化处理。

2.1 主发动机工作段

假设主发动机工作段的时间为$t_{1\mathrm{f}}$，离散点数为N_1，则离散时间步长Δt_1表示如下：

$$\Delta t_1=t_{1\mathrm{f}}/(N_1-1)\tag{9}$$

为保证离散精度和优化问题的稀疏性, 本文采用梯形离散法对动力学方程组进行离散化处理。则主发动机工作段的动力学方程组在经过梯形离散后可表示为

$$\begin{aligned}&\boldsymbol{X}_1(k+1)=\boldsymbol{X}_1(k)+\\&\quad\frac{1}{2}\Delta t_1\left(\boldsymbol{f}(\boldsymbol{X}_1(k),\boldsymbol{U}_1(k))+\boldsymbol{f}(\boldsymbol{X}_1(k+1),\boldsymbol{U}_1(k+1))\right)\end{aligned}\tag{10}$$

式中: $\boldsymbol{X}_1(k)=\boldsymbol{X}_1(t_{1(k)})$; $\boldsymbol{U}_1(k)=\boldsymbol{U}_1(t_{1(k)})$。为表达的简洁, 将$\boldsymbol{f}(\boldsymbol{X}_1(k),\boldsymbol{U}_1(k))$简写为$\boldsymbol{f}_1(k)$。

由于$\boldsymbol{f}_1(k)\Delta t_1$为非线性方程组, 需要对其进行泰勒展开线性化处理, 即

$$\begin{aligned}\boldsymbol{F}_1(k)=&\boldsymbol{f}_1^0(k)\Delta t_1^0+\left.\frac{\partial\boldsymbol{f}_1(k)}{\partial\boldsymbol{X}_1(k)}\right|_{\boldsymbol{X}_1^0(k)}(\boldsymbol{X}_1(k)-\boldsymbol{X}_1^0(k))+\\&\left.\frac{\partial\boldsymbol{f}_1(k)}{\partial\boldsymbol{U}_1(k)}\right|_{\boldsymbol{U}_1^0(k)}(\boldsymbol{U}_1(k)-\boldsymbol{U}_1^0(k))+\boldsymbol{f}_1^0(k)(\Delta t_1-\Delta t_1^0)\end{aligned}\tag{11}$$

式中: $\left(\boldsymbol{X}_1^0,\boldsymbol{U}_1^0,\Delta t_1^0\right)$为参考轨迹; $\boldsymbol{f}_1^0(k)$为第k个参考点处的动力学方程组矢量。在求$\boldsymbol{f}_1(k)$对$\boldsymbol{X}_1(k)$的偏导数时, 由于R和ϕ的变化量较小, 因此, 忽略对该两项的求导, 直接取参考轨迹结果计算即可。

为避免由于初始猜想的不精确性及线性化处理带来的伪不可行现象, 对式 (11) 加入虚拟控制量, 即

$$\begin{cases}\begin{aligned}\boldsymbol{F}_1(k)=&\boldsymbol{f}_1^0(k)\Delta t_1^0+\left.\frac{\partial\boldsymbol{f}_1(k)}{\partial\boldsymbol{X}_1(k)}\right|_{\boldsymbol{X}_1^0(k)}(\boldsymbol{X}_1(k)-\boldsymbol{X}_1^0(k))+\\&\left.\frac{\partial\boldsymbol{f}_1(k)}{\partial\boldsymbol{U}_1(k)}\right|_{\boldsymbol{U}_1^0(k)}(\boldsymbol{U}_1(k)-\boldsymbol{U}_1^0(k))+\\&\boldsymbol{f}_1^0(k)(\Delta t_1-\Delta t_1^0)+\boldsymbol{a}_{R1}(k)\end{aligned}\end{cases}\tag{12}$$

在加入虚拟控制量$\boldsymbol{a}_{R1}(k)$后, 应在目标函数中加入对该项的罚函数, 以避免$\boldsymbol{a}_{R1}(k)$对结果的影响, 即

$$\begin{cases}\min\ (N_1-1)\Delta t_1+\gamma_{\mathrm{ar1}}\sum\limits_{k=1}^{N_1}\kappa_{\mathrm{ar1}k}\\\quad\text{s.t. }\boldsymbol{X}_1(k+1)-\boldsymbol{X}_1(k)=\frac{1}{2}(\boldsymbol{F}_1(k)+\boldsymbol{F}_1(k+1))\\\quad\|\boldsymbol{a}_{R1}(k)\|_2\leqslant\kappa_{\mathrm{ar1}k}\end{cases}\tag{13}$$

式中: $\kappa_{\mathrm{ar1}k}$为虚拟控制量的约束变量; γ_{ar1}为虚拟控制量的罚函数系数。为避免虚拟控制量的加入引起的优化问题无界的现象, 应对性能指标相关的变量加入信赖域约束, 即

$$\varpi_{\mathrm{l}\Delta t_1}\Delta t_1^0\leqslant\Delta t_1\leqslant\varpi_{\mathrm{u}\Delta t_1}\Delta t_1^0\tag{14}$$

式中: $\varpi_{\mathrm{l}\Delta t_1}$和$\varpi_{\mathrm{u}\Delta t_1}$为时间信赖域约束的系数。

对于姿态角变化率的约束, 本文假设推力方向沿着上升器的纵轴方向, 则姿态角变化率的约束可转换为对推力方向的约束[38], 如图 3 所示。

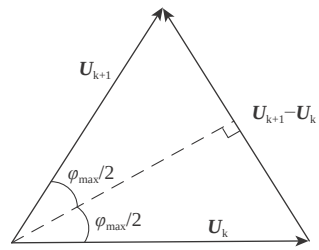


图 3　推力方向几何示意图

Fig. 3　Geometry of thrust direction

图 3 中: $\varphi_{\max}$为离散时间步长内的最大姿态变化角。由图 3 的几何关系可知:

$$\begin{cases}\dfrac{1}{2}\|\boldsymbol{U}(k+1)-\boldsymbol{U}(k)\|_2=\sin\left(\dfrac{\varphi_{\max}}{2}\right)\\\varphi_{\max}=\dfrac{1}{2}\displaystyle\int_{t_k}^{t_{k+1}}\omega_{\max}\mathrm{d}t\end{cases}\tag{15}$$

式中: $\omega_{\max}$为上升器的最大姿态角变化率。则姿态角变化率可表示如下:

$$\|\boldsymbol{U}(k+1)-\boldsymbol{U}(k)\|_2\leqslant2\sin\left(\frac{1}{2}\omega_{\max}\Delta t^0\right)\tag{16}$$

由于本文是在球坐标系下建立的优化模型, 球坐标系随着上升器位置的变化而变化, 因此, 以式 (16) 约束姿态角会由于坐标系的变化而存在偏差。但由于前后 2 个离散点之间的时间很短, 即前后 2 个离散点之间的坐标系偏差很小, 因此, 坐标系变化引起的姿态角约束偏差可以忽略。

为后续表达的简洁, 取$\boldsymbol{G}_1(k)$表示不等式约束, 即

$$(\boldsymbol{G}_1(k)\leqslant\boldsymbol{0}):\begin{cases}R_1(k)\geqslant R_{\mathrm{m}}\\\sqrt{u_{r1}^2(k)+u_{t1}^2(k)+u_{p1}^2(k)}\leqslant1\\\varpi_{\mathrm{l}\Delta t_1}\Delta t_1^0\leqslant\Delta t_1\leqslant\varpi_{\mathrm{u}\Delta t_1}\Delta t_1^0\\\|\boldsymbol{a}_{R1}(k)\|_2\leqslant\kappa_{\mathrm{ar1}k}\\\|\boldsymbol{U}_1(k+1)-\boldsymbol{U}_1(k)\|_2\leqslant2\sin\left(\dfrac{1}{2}\omega_{1\max}\Delta t_1^0\right)\end{cases}\tag{17}$$

式中: $\omega_{1\max}$为主发动机工作段的最大姿态角变化率。

2.2　推力器工作段

假设推力器工作段的时间为$t_{2\mathrm{f}}$, 离散点数为N_2, 则离散时间步长Δt_2表示如下:

$$\Delta t_2=t_{2\mathrm{f}}/(N_2-1)\tag{18}$$

与主发动机工作段类似, 动力学方程组右端项$\boldsymbol{F}_2$如下:

$$\begin{aligned}\boldsymbol{F}_2(k)=&\boldsymbol{f}_2^0(k)\Delta t_2^0+\left.\frac{\partial\boldsymbol{f}_2(k)}{\partial\boldsymbol{X}_2(k)}\right|_{\boldsymbol{X}_2^0(k)}(\boldsymbol{X}_2(k)-\boldsymbol{X}_2^0(k))+\\&\left.\frac{\partial\boldsymbol{f}_2(k)}{\partial\boldsymbol{U}_2(k)}\right|_{\boldsymbol{U}_2^0(k)}(\boldsymbol{U}_2(k)-\boldsymbol{U}_2^0(k))+\\&\boldsymbol{f}_2^0(k)(\Delta t_2-\Delta t_2^0)+\boldsymbol{a}_{R2}(k)\end{aligned}\tag{19}$$

式中：$\boldsymbol{a}_{R2}(k)$表示推力器工作段的虚拟控制量。不等式约束$\boldsymbol{G}_2$如下：

$$(\boldsymbol{G}_2(k)\leqslant\boldsymbol{0}):\begin{cases}R_2(k)\geqslant R_{\mathrm{m}}\\ \sqrt{u_{r2}^2(k)+u_{t2}^2(k)+u_{p2}^2(k)}\leqslant 1\\ \varpi_{\mathrm{l}\Delta t_2}\Delta t_2^0\leqslant\Delta t_2\leqslant\varpi_{\mathrm{u}\Delta t_2}\Delta t_2^0\\ \|\boldsymbol{a}_{R2}(k)\|_2\leqslant\kappa_{\mathrm{ar}2k}\\ \|\boldsymbol{U}_2(k+1)-\boldsymbol{U}_2(k)\|_2\leqslant 2\sin\left(\dfrac{1}{2}\omega_{2\max}\Delta t_2^0\right)\end{cases}\tag{20}$$

式中：$\varpi_{\mathrm{l}\Delta t_2}$和$\varpi_{\mathrm{u}\Delta t_2}$为时间信赖域的系数，$\kappa_{\mathrm{ar}2k}$为推力器工作段的虚拟控制量约束变量，$\omega_{2\max}$为推力器工作段的最大姿态角变化率。

由于直接上升交会方案并非立即起飞，因而本文在终端相角约束中加入了上升器在月面等待的时间t_{w}，以及松弛项s_θ，即：

$$\theta_{\mathrm{o}0}+\omega_{\mathrm{o}}((N_1-1)\Delta t_1+(N_2-1)\Delta t_2)+\omega_{\mathrm{o}}t_{\mathrm{w}}-\theta_2(N_2)-2\pi=s_\theta\tag{21}$$

式中：相角约束减去2π是因为轨道器与上升器的最终相角偏差应为2π。

2.3 离散优化模型

由2.1和2.2节可得月面应急上升交会的离散轨迹优化问题 (optimal discretized problem, ODP) 如下：

$$\begin{cases}\min t_{\mathrm{w}}+(N_1-1)\Delta t_1+(N_2-1)\Delta t_2+\gamma_\theta s_\theta+\gamma_{\mathrm{ar}1}\displaystyle\sum_{k=1}^{N_1}\kappa_{\mathrm{ar}1k}+\gamma_{\mathrm{ar}2}\sum_{k=1}^{N_2}\kappa_{\mathrm{ar}2k}\\ \text{s.t.}\quad \boldsymbol{X}_1(1)=\boldsymbol{X}_0\\ \boldsymbol{U}_1(1)=[1,0,0]^{\mathrm{T}}\\ \boldsymbol{X}_1(k+1)-\boldsymbol{X}_1(k)=\dfrac{1}{2}(\boldsymbol{F}_1(k)+\boldsymbol{F}_1(k+1))\\ \boldsymbol{G}_1(k)\leqslant\boldsymbol{0}\\ \boldsymbol{X}_2(1)=\boldsymbol{X}_1(N_1)\\ \boldsymbol{U}_2(1)=\boldsymbol{U}_1(N_1)\\ \boldsymbol{X}_2(k+1)-\boldsymbol{X}_2(k)=\dfrac{1}{2}(\boldsymbol{F}_2(k)+\boldsymbol{F}_2(k+1))\\ \boldsymbol{G}_2(k)\leqslant\boldsymbol{0}\\ R_2(N_2)=R_{\mathrm{f}},\phi_2(N_2)=0,\\ V_{r2}(N_2)=0,V_{t2}(N_2)=V_{\mathrm{f}},V_{p2}(N_2)=0\\ \theta_{\mathrm{o}0}+\omega_{\mathrm{o}}t_{\mathrm{w}}+\omega_{\mathrm{o}}((N_1-1)\Delta t_1+(N_2-1)\Delta t_2)-\theta_2(N_2)-2\pi=s_\theta\\ s_\theta\geqslant 0,t_{\mathrm{w}}\geqslant 0\\ m_{\mathrm{e}1}(N_1-1)\Delta t_1+m_{\mathrm{e}2}(N_2-1)\Delta t_2\leqslant m_0-m_{\mathrm{dry}}\end{cases}\tag{22}$$

式中：上升器需垂直起飞，因此，$\boldsymbol{U}_1(1)=[1,0,0]^{\mathrm{T}}$；令$s_\theta\geqslant 0$是为了避免性能指标取最小值时出现$s_\theta$取得负无穷的情况；$\gamma_\theta$为$s_\theta$的罚函数系数；$m_{\mathrm{e}1}=P_1/I_{\mathrm{sp}1}g_0$，$m_{\mathrm{e}2}=P_2/I_{\mathrm{sp}2}g_0$，$I_{\mathrm{sp}1}$和$I_{\mathrm{sp}2}$分别为主发动机的比冲和推力器的比冲。

为求解的稳定性，需对上升器的动力学方程组中的变量作归一化处理。径向距离的归一化参数R_{n}为目标轨道高度，相角θ和极角ϕ的数值在10以内，无需归一化；V_r的归一化参数$V_{r\mathrm{n}}=400$；V_t的归一化参数$V_{t\mathrm{n}}$取为V_{f}；V_p的归一化参数$V_{p\mathrm{n}}=30$。为表达的简洁，后文仍取原变量的符号作为归一化后的变量。

本文在优化结果达到以下条件时终止序列二阶锥规划的迭代：

$$\begin{cases}\left|\boldsymbol{X}_1^j-\boldsymbol{X}_1^{j+1}\right|_{\max}\leqslant\delta_{\mathrm{SOCP}},\left|\boldsymbol{U}_1^j-\boldsymbol{U}_1^{j+1}\right|_{\max}\leqslant\delta_{\mathrm{SOCP}}\\ \left|\boldsymbol{X}_2^j-\boldsymbol{X}_2^{j+1}\right|_{\max}\leqslant\delta_{\mathrm{SOCP}},\left|\boldsymbol{U}_2^j-\boldsymbol{U}_2^{j+1}\right|_{\max}\leqslant\delta_{\mathrm{SOCP}}\end{cases}\tag{23}$$

式中：δ_{SOCP}为设定的值；$\boldsymbol{X}_1^j$和$\boldsymbol{X}_1^{j+1}$及$\boldsymbol{U}_1^j$和$\boldsymbol{U}_1^{j+1}$分别为第j次和第j+1次序列优化得到的主发动机工作段的状态量和控制量；$\boldsymbol{X}_2^j$和$\boldsymbol{X}_2^{j+1}$及$\boldsymbol{U}_2^j$和$\boldsymbol{U}_2^{j+1}$表示第j次和第j+1次序列优化得到的推力器工作段的状态量和控制量。

由于动力学方程组的线性化需要初始参考轨迹，本文初始猜想取值方法如下：

$$\begin{cases}R_1^0=R_{\mathrm{m}}/R_n\\ V_{t1}^0=1\\ R_2^0=1\\ V_{t2}^0=1\end{cases}\tag{24}$$

$[\boldsymbol{X}_1,\boldsymbol{U}_1]$和$[\boldsymbol{X}_2,\boldsymbol{U}_2]$中其余变量初始猜想均为0，$t_{1\mathrm{f}}$的初始猜想为150，$t_{2\mathrm{f}}$的初始猜想值根据式(25)计算：

$$t_{2\mathrm{f}}=(m_0-m_{\mathrm{dry}}-m_{\mathrm{e}1}t_{1\mathrm{f}})/m_{\mathrm{e}2}\tag{25}$$

3 内点法定制化改进求解设计

首先对ECOS中的内点法作简要说明，之后再对内点法的计算过程进行加速，以加快二阶锥规划问题的求解速度。

3.1 二阶锥规划问题的内点法求解方法

3.1.1 标准二阶锥规划问题

标准的二阶锥规划问题如下：

$$\begin{cases}\min\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}\\ \quad\text{s.t.}\quad \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\\ \qquad\boldsymbol{G}\boldsymbol{x}+\boldsymbol{s}=\boldsymbol{h}\\ \qquad\boldsymbol{s}\in\mathcal{K}\end{cases}\tag{26}$$

式中：$\boldsymbol{x}$为原变量；$\boldsymbol{s}$为松弛变量；$\boldsymbol{c}\in\mathbf{R}^n$，$\boldsymbol{G}\in\mathbf{R}^{m\times n}$，$\boldsymbol{h}\in\mathbf{R}^m$，$\boldsymbol{A}\in\mathbf{R}^{p\times n}$，$\boldsymbol{b}\in\mathbf{R}^p$为待解问题的矩阵或向量；$\mathcal{K}$为约束锥。

式(26)的对偶问题为

$$\begin{cases}\max\ -\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}-\boldsymbol{h}^{\mathrm{T}}\boldsymbol{z}\\ \text{s.t.}\ \ \boldsymbol{G}^{\mathrm{T}}\boldsymbol{z}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}+\boldsymbol{c}=\boldsymbol{0}\\ \qquad \boldsymbol{z}\in\mathcal{K}\end{cases}\tag{27}$$

式中: $\boldsymbol{y}$和$\boldsymbol{z}$为对偶问题的变量。

3.1.2　内点法流程

求解式 (26) 所示的二阶锥规划问题通常可采用原对偶内点法, 原对偶内点法即根据 KKT 条件的修改形式迭代计算最优解的方法。对于式 (26) 所描述的问题, 其 KKT 条件为

$$\begin{cases}\boldsymbol{c}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}+\boldsymbol{G}^{\mathrm{T}}\boldsymbol{z}=\boldsymbol{0}\\ \boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}=\boldsymbol{0}\\ \boldsymbol{G}\boldsymbol{x}+\boldsymbol{s}-\boldsymbol{h}=\boldsymbol{0}\\ \boldsymbol{s}^{\mathrm{T}}\boldsymbol{z}=0\\ \boldsymbol{s}\in\mathcal{K}\qquad \boldsymbol{z}\in\mathcal{K}\end{cases}\tag{28}$$

内点法的详细计算流程可参考文献 [39] 和文献 [40], 本文仅对关键步骤作简要说明。

1) 中心路径

为便于对 KKT 条件进行求解, 需要对其进行松弛, 在对 KKT 条件进行松弛后可得到中心路径 (29), 对 KKT 条件的求解等价于对中心路径的求解, 中心路径即为满足式 (29) 的$(\boldsymbol{x},\boldsymbol{y},\boldsymbol{s},\boldsymbol{z},\tau,\kappa)$的集合。

$$\begin{cases}\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\\ \boldsymbol{s}\\ \kappa\end{bmatrix}=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{A}^{\mathrm{T}} & \boldsymbol{G}^{\mathrm{T}} & \boldsymbol{c}\\ -\boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{b}\\ -\boldsymbol{G} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{h}\\ -\boldsymbol{c}^{\mathrm{T}} & -\boldsymbol{b}^{\mathrm{T}} & -\boldsymbol{h}^{\mathrm{T}} & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\\ \boldsymbol{z}\\ \tau\end{bmatrix}+\mu\begin{bmatrix}\boldsymbol{q}_x\\ \boldsymbol{q}_y\\ \boldsymbol{q}_z\\ \boldsymbol{q}_\tau\end{bmatrix}\\ (\boldsymbol{W}^{-\mathrm{T}}\boldsymbol{s})\circ(\boldsymbol{W}\boldsymbol{z})=\mu\boldsymbol{e}\qquad \kappa\tau=\mu\\ \boldsymbol{s}\in\mathcal{K}\quad \boldsymbol{z}\in\mathcal{K}\qquad \kappa\geqslant 0\quad \tau\geqslant 0\end{cases}\tag{29}$$

式中: $\mu=(\boldsymbol{s}^{\mathrm{T}}\boldsymbol{z}+\kappa\tau)/(D+1)$, 当$\mu=0$时, 式 (29) 等价于 KKT 条件; D为锥的阶数, 即锥$\mathcal{K}$由D个标准锥组成; τ和κ为额外引入的变量; $\boldsymbol{q}_*$的定义如下:

$$(\boldsymbol{q}_x,\boldsymbol{q}_y,\boldsymbol{q}_z,q_\tau)=\frac{D+1}{\boldsymbol{s}^{\mathrm{T}}\boldsymbol{z}+\tau\kappa}(\boldsymbol{r}_x,\boldsymbol{r}_y,\boldsymbol{r}_z,r_\tau)\tag{30}$$

$$\begin{cases}\boldsymbol{r}_x=-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}-\boldsymbol{G}^{\mathrm{T}}\boldsymbol{z}-\boldsymbol{c}\tau\\ \boldsymbol{r}_y=\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\tau\\ r_\tau=\kappa+\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}+\boldsymbol{h}^{\mathrm{T}}\boldsymbol{z}\\ \boldsymbol{r}_z=\boldsymbol{s}+\boldsymbol{G}\boldsymbol{x}-\boldsymbol{h}\tau\end{cases}\tag{31}$$

式中: 运算符号$\circ$为欧式约当算子[39]。

$\boldsymbol{e}=[\boldsymbol{e}_1^{\mathrm{T}},\boldsymbol{e}_2^{\mathrm{T}},\cdots,\boldsymbol{e}_D^{\mathrm{T}}]^{\mathrm{T}}$, 对于线性锥, $\boldsymbol{e}_i=1$; 对于m维二阶锥, $\boldsymbol{e}_i=(1,\boldsymbol{0}_{m-1})$。$\boldsymbol{W}$为内斯特罗夫-托德 (Nesterov-Todd, NT) 缩放矩阵, 其定义见文献 [39]。

2) 更新方向与更新步长

内点法在求解二阶锥规划问题时, 首先计算变量初值$(\boldsymbol{x}_0,\boldsymbol{y}_0,\boldsymbol{s}_0,\boldsymbol{z}_0,\tau_0,\kappa_0)$, 再根据变量初值计算更新方向$(\Delta\boldsymbol{x},\Delta\boldsymbol{y},\Delta\boldsymbol{s},\Delta\boldsymbol{z},\Delta\tau,\Delta\kappa)$与更新步长$\alpha$。在得到更新方向与更新步长后, 再对变量进行更新:

$$(\boldsymbol{x},\boldsymbol{y},\boldsymbol{s},\boldsymbol{z},\tau,\kappa)=(\boldsymbol{x}_0,\boldsymbol{y}_0,\boldsymbol{s}_0,\boldsymbol{z}_0,\tau_0,\kappa_0)+\alpha(\Delta\boldsymbol{x},\Delta\boldsymbol{y},\Delta\boldsymbol{s},\Delta\boldsymbol{z},\Delta\tau,\Delta\kappa)\tag{32}$$

在完成对变量的更新后, 根据新的变量计算更新方向和更新步长, 不断循环此过程, 直至变量收敛至满足 KKT 条件的解。

在 ECOS 中, 更新方向的计算包括仿射方向、中心化方向、矫正方向 3 个部分。在计算仿射方向时, 令$\mu=0$, 其目的在于让变量尽可能地满足 KKT 条件。计算中心化方向的目的在于让变量更接近中心路径, 以便在下一次迭代中可以取得更大的更新步长。计算矫正方向是为了补偿在计算仿射方向时, 由于对式 (29) 泰勒展开线性化所引入的误差。中心化方向和矫正方向的计算可进行合并, 该合并方向称为组合方向。

仿射方向和组合方向的计算可转化为式 (33) 和 (34) 的线性方程组的计算:

$$\boldsymbol{K}[\Delta\boldsymbol{x}_1,\Delta\boldsymbol{y}_1,\Delta\boldsymbol{z}_1]^{\mathrm{T}}=[-\boldsymbol{c},\boldsymbol{b},\boldsymbol{h}]^{\mathrm{T}}\tag{33}$$

$$\boldsymbol{K}[\Delta\boldsymbol{x}_2,\Delta\boldsymbol{y}_2,\Delta\boldsymbol{z}_2]^{\mathrm{T}}=[\boldsymbol{d}_x,\boldsymbol{d}_y,\boldsymbol{d}_z]^{\mathrm{T}}\tag{34}$$

式中:

$$\boldsymbol{K}=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{A}^{\mathrm{T}} & \boldsymbol{G}^{\mathrm{T}}\\ \boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{G} & \boldsymbol{0} & -\boldsymbol{W}^2\end{bmatrix}\tag{35}$$

式 (33) 与 (34) 称为 KKT 方程组, 其系数矩阵称为 KKT 矩阵。在计算仿射方向和组合方向时, 式 (34) 的右端项会有差别, 详情可参考文献 [26]。

3) 内点法迭代停止条件

当变量迭代至最优解时, 内点法即停止迭代, 一般的, 最优条件的设定如下:

$$\zeta=\max\left\{\frac{\|\boldsymbol{r}_x\|_2}{r_{x\mathrm{s}}\tau},\frac{\|\boldsymbol{r}_y\|_2}{r_{y\mathrm{s}}\tau},\frac{\|\boldsymbol{r}_z\|_2}{r_{z\mathrm{s}}\tau}\right\}\tag{36}$$

式中:

$$r_{x\mathrm{s}}=\max\left\{1,\|\boldsymbol{c}\|_2+\|\boldsymbol{y}\|_2+\|\boldsymbol{z}\|_2\right\}\tag{37}$$

$$r_{y\mathrm{s}}=\max\{1,\|\boldsymbol{b}\|_2+\|\boldsymbol{x}\|_2\}\tag{38}$$

$$r_{z\mathrm{s}}=\max\{1,\|\boldsymbol{h}\|_2+\|\boldsymbol{x}\|_2+\|\boldsymbol{s}\|_2\}\tag{39}$$

其次, 定义对偶间隙与相对对偶间隙分别如式 (40) 和 (41) 所示:

$$\delta_{\mathrm{gap}}=\boldsymbol{s}^{\mathrm{T}}\boldsymbol{z}\tag{40}$$

$$\delta_{\mathrm{relgap}}=\boldsymbol{s}^{\mathrm{T}}\boldsymbol{z}/\upsilon\tag{41}$$

式中: $\upsilon=\max\left\{-\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x},-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{y}-\boldsymbol{h}^{\mathrm{T}}\boldsymbol{z}\right\}$。

当$\upsilon>0$时, 如果$\zeta<o_{\mathrm{feastol}}$, 而且$\delta_{\mathrm{abs}}<o_{\mathrm{abstol}}$或$\delta_{\mathrm{relgap}}<o_{\mathrm{reltol}}$, 即认为求得最优解。$o_{\mathrm{feastol}}$为等式约束残差精度, o_{abstol}为绝对对偶残差精度, o_{reltol}为相对

对偶残差精度。

4) KKT矩阵的变换

在求解式(33)和式(34)所示的KKT方程组时,一般采用直接法进行求解,即先对KKT矩阵进行分解,再根据分解得到的三角矩阵计算变量的值。然而,由于式(35)的KKT矩阵不是正定的,无法对其直接进行快速矩阵分解,因此,需对KKT矩阵做如下转换:

$$\tilde{\boldsymbol{K}}=\boldsymbol{K}+\Delta\boldsymbol{K}=\begin{bmatrix}\delta\boldsymbol{I}_n & \boldsymbol{A}^{\mathrm{T}} & \boldsymbol{G}^{\mathrm{T}}\\ \boldsymbol{A} & -\delta\boldsymbol{I}_p & \boldsymbol{0}\\ \boldsymbol{G} & \boldsymbol{0} & -\boldsymbol{W}^2\end{bmatrix} \tag{42}$$

式中:

$$\Delta\boldsymbol{K}=\begin{bmatrix}\delta\boldsymbol{I}_n & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & -\delta\boldsymbol{I}_p & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0}\end{bmatrix} \tag{43}$$

式中:$\delta\boldsymbol{I}_n$和$\delta\boldsymbol{I}_p$分别为n阶和p阶对角微元单位矩阵,$\delta=7\times10^{-8}$。

5) 内点法流程

内点法计算流程如下:

算法 内点法算法。

输入:二阶锥规划问题的$\boldsymbol{c},\boldsymbol{G},\boldsymbol{h},\boldsymbol{A},\boldsymbol{b}$。

输出:$(\boldsymbol{x},\boldsymbol{y},\boldsymbol{s},z,\tau,\kappa)$。

1. 输入二阶锥规划问题的$\boldsymbol{c},\boldsymbol{G},\boldsymbol{h},\boldsymbol{A},\boldsymbol{b}$;

2. 计算初值$(\boldsymbol{x}_0,\boldsymbol{y}_0,\boldsymbol{s}_0,z_0,\tau_0,\kappa_0)$;

3. 计算残差,式(30);

4. 判断循环条件,当达到最优或问题无解时,退出循环;

5.更新KKT矩阵中的$\boldsymbol{W}$矩阵;

6. KKT矩阵分解;

7. 仿射方向和仿射步长的计算;

8. 中心化参数计算;

9. 更新方向与更新步长的计算;

10. 更新变量$(\boldsymbol{x},\boldsymbol{y},\boldsymbol{s},z,\tau,\kappa)$,回到步骤3。

3.2 线性方程组定制化求解加速方法

在内点法的单次迭代中,仿射方向和组合方向均需计算式(33)和(34)的KKT方程组,但由于式(33)的KKT方程组右端项不发生改变,因此,在计算组合方向时,无需对式(33)进行重复计算。因此,每一次内点法的迭代需要计算3次KKT方程组。而KKT方程组求解是内点法中最耗时的部分,因此,本文主要对KKT方程组的求解进行加速。

3.2.1 定制化矩阵排序

在式(42)中,$\tilde{\boldsymbol{K}}$矩阵为拟正定矩阵,为了减少对$\tilde{\boldsymbol{K}}$矩阵进行分解时产生的填充元数量,需要对$\tilde{\boldsymbol{K}}$矩阵进行重新排序。即

$$\tilde{\boldsymbol{K}}_{\boldsymbol{P}}=\boldsymbol{P}\tilde{\boldsymbol{K}}\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^{\mathrm{T}} \tag{44}$$

式中:$\boldsymbol{P}$矩阵为排序矩阵;$\boldsymbol{L}$为矩阵分解后得到的下三角矩阵;$\boldsymbol{D}$矩阵为矩阵分解后得到的对角矩阵。

为综合排序耗时与排序效果,ECOS采用近似最小度排序算法[41]对$\tilde{\boldsymbol{K}}$矩阵进行重新排序。但近似最小度排序并不一定是填充最少的排序方法,常用的排序方法还有最小度排序[42],局部填充最小排序[43]等。

对于定制化的问题,$\tilde{\boldsymbol{K}}$矩阵的非零元素位置不发生改变,即排序矩阵可提前计算得到。由文献[44]可知,对于任意的排序矩阵$\boldsymbol{P}$,$\boldsymbol{P}\tilde{\boldsymbol{K}}\boldsymbol{P}^{\mathrm{T}}$矩阵的分解是稳定的。因此,在选择排序方法时,无需考虑排序耗时的影响,仅需根据排序效果选取$\boldsymbol{L}$矩阵非零元素个数最少的排序方法即可。

以本文的轨迹规划问题为例,通过取不同的离散点数得到不同维数的KKT矩阵,对比近似最小度排序算法以及局部填充最小排序算法的排序效果如表1所示,表1中离散点数(N_1+N_2)表示式(22)中的离散点数N_1和N_2之和,后文与此相同。可看到,采用局部填充最小排序算法,$\boldsymbol{L}$矩阵的非零元素个数更少。

表1 不同排序方法的矩阵非零元素个数对比

Table 1 Number of non-zero elements of matrix for different ordering methods

离散点数(N_1+N_2)	非零元素个数	
	近似最小度算法	局部填充最小算法
11+21	8 194	7 209
21+41	17 059	14 416
31+61	26 977	21 709
41+81	34 435	28 916

3.2.2 定制化矩阵分解

为加快矩阵分解的速度,一般采用稀疏矩阵Cholesky分解法[45]对$\tilde{\boldsymbol{K}}_{\boldsymbol{P}}$矩阵进行分解。对于定制化的问题,$\tilde{\boldsymbol{K}}_{\boldsymbol{P}}$矩阵的非零元素个数及位置不随着优化问题的求解而改变,因此,可对矩阵的分解过程进行定制化。

本文将$\tilde{\boldsymbol{K}}_{\boldsymbol{P}}$矩阵分解所需的定制化信息,如$\boldsymbol{L}$矩阵的非零元素位置信息等提前计算得出,内置计算结果。本文方法相较于采用代码自动生成的方法而言,需要额外的读取非零元素位置等数据的时间,但代码量小,所需的储存空间也较小。

3.2.3 数据储存定制化

通用求解器在求解相应的优化问题时,一般采用动态内存储存数据。对于动态申请的储存空间,储存空间的申请与释放均需要消耗一部分时间,且此时数据储存于内存中,读取与写入的速度较慢。

然而,对于定制化的问题,$\tilde{\boldsymbol{K}}_{\boldsymbol{P}}$矩阵的非零元素

位置不改变, 排序矩阵不改变, 数据所需的储存空间是确定的。因此, 本文采用静态内存储存数据, 对于静态内存而言, 储存空间申请与释放的耗时少, 且此时数据储存于缓存区, 读取与写入的速度快。

3.2.4　仿射方向的并行计算

在仿射方向时, 需要计算式 (33) 和 (34) 的 2 个线性方程组。为加快仿射方向的计算速度, 可将式 (33) 和 (34) 并行计算, 如图 4 所示。

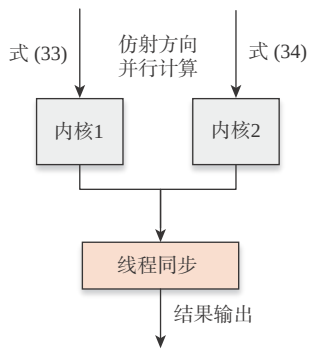


图 4　仿射方向并行计算流程

Fig. 4　Parallel calculation process of affine direction

3.2.5　动态调整 KKT 方程组求解精度

取$\tilde{\boldsymbol{K}}\boldsymbol{\xi}=\boldsymbol{r}$进行分析, 在采用$\tilde{\boldsymbol{K}}$矩阵对 KKT 方程组进行求解时, 由于引入了微元矩阵$\Delta\boldsymbol{K}$, 求解结果会存在偏差。在对求解精度要求较高时, 一般采用迭代法提高 KKT 方程组的求解精度, 具体流程如下:

算法　KKT 方程组求解算法。

输入: $\tilde{\boldsymbol{K}}$、$\boldsymbol{r}$、$\boldsymbol{K}$。

输出: $\boldsymbol{\xi}$。

1. $\tilde{\boldsymbol{K}}\boldsymbol{\xi}=\boldsymbol{r}$;

2. $\tilde{\boldsymbol{K}}\Delta\boldsymbol{\xi}=\boldsymbol{r}-\boldsymbol{K}\boldsymbol{\xi}$;

3. $\boldsymbol{\xi}=\boldsymbol{\xi}+\Delta\boldsymbol{\xi}$;

4. 若$\boldsymbol{\xi}$精度满足要求, 或精度提高缓慢, 则跳出循环; 否则回到步骤 2。

由 KKT 方程组求解算法可知, 求解 KKT 方程组需迭代多次。然而本文研究发现, 在内点法计算初期, 即对偶间隙与相对对偶间隙较大时, 微元δ对组合方向与更新步长的计算结果影响微弱。

因此, 可以对 KKT 方程组的求解精度进行动态调整, 在对偶间隙δ_{gap}与相对对偶间隙δ_{relgap}较大时, 不修正 KKT 方程组的精度; 在δ_{gap}与δ_{relgap}较小时, 再加入对 KKT 方程组求解的迭代修正, 其改进算法如下所示, 其中: ε_{abs}和ε_{rel}为设定的 KKT 方程组修正阈值。

算法　改进 KKT 方程组求解算法。

输入: $\tilde{\boldsymbol{K}}$、$\boldsymbol{r}$、$\boldsymbol{K}$、ε_{abs}、ε_{rel}。

输出: $\boldsymbol{\xi}$。

1. $\tilde{\boldsymbol{K}}\boldsymbol{\xi}=\boldsymbol{r}$;

2. 当$\delta_{\text{gap}}>\varepsilon_{\text{abs}}$且$\delta_{\text{relgap}}>\varepsilon_{\text{rel}}$时, 跳出循环; 否则转到 3;

3. $\tilde{\boldsymbol{K}}\Delta\boldsymbol{\xi}=\boldsymbol{r}-\boldsymbol{K}\boldsymbol{\xi}$;

4. $\boldsymbol{\xi}=\boldsymbol{\xi}+\Delta\boldsymbol{\xi}$;

5. 若$\boldsymbol{\xi}$精度满足要求, 或精度提高缓慢, 则跳出循环; 否则回到步骤 3。

3.3　热启动方法

内点法在求解优化问题时需给优化变量赋初值, 常用的初值计算方法有冷启动法、传统热启动法以及最近提出的序列二阶锥规划改进热启动法 (后称改进热启动法)。

3.3.1　冷启动法

内点法的冷启动一般如下选择冷启动参数:

$$(\boldsymbol{x},\boldsymbol{s},\boldsymbol{y},z,\kappa,\tau)=(\boldsymbol{0},\boldsymbol{e},\boldsymbol{0},\boldsymbol{e},1,1) \tag{45}$$

式中: $\boldsymbol{e}=[\boldsymbol{e}_1^{\mathrm{T}},\boldsymbol{e}_2^{\mathrm{T}},\cdots,\boldsymbol{e}_D^{\mathrm{T}}]^{\mathrm{T}}$为约束锥的单位向量。

3.3.2　传统热启动法

在传统热启动方法中, 内点法初值计算方法如下:

1) 变量$\boldsymbol{x}$和$\boldsymbol{s}$

对变量$\boldsymbol{x}$的初值计算等价于求解以下优化问题:

$$\begin{cases}\min\ \dfrac{1}{2}\|\bar{\boldsymbol{s}}\|^2\\ \text{s.t.}\quad \bar{\boldsymbol{s}}=\boldsymbol{h}-\boldsymbol{G}\boldsymbol{x}\\ \qquad \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\end{cases} \tag{46}$$

变量$\boldsymbol{s}$的初值可由式 (46) 中的$\bar{\boldsymbol{s}}$计算得到, 即

$$\boldsymbol{s}=\begin{cases}\bar{\boldsymbol{s}} & \bar{\boldsymbol{s}}\in\mathcal{K}\\ [\bar{\boldsymbol{s}}+(1+\alpha_p)\boldsymbol{e}]\in\mathcal{K} & \bar{\boldsymbol{s}}\notin\mathcal{K}\end{cases} \tag{47}$$

2) 对偶变量$\boldsymbol{y}$和$\boldsymbol{z}$

对偶变量$\boldsymbol{y}$的初值取以下优化问题的结果:

$$\begin{cases}\min\ \dfrac{1}{2}\|\bar{\boldsymbol{z}}\|^2\\ \text{s.t.}\ \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}+\boldsymbol{G}^{\mathrm{T}}\bar{\boldsymbol{z}}+\boldsymbol{c}=\boldsymbol{0}\end{cases} \tag{48}$$

对偶变量$\boldsymbol{z}$的初值可通过式 (48) 中的$\bar{\boldsymbol{z}}$计算得到, 即

$$\boldsymbol{z}=\begin{cases}\bar{\boldsymbol{z}} & \bar{\boldsymbol{z}}\in\mathcal{K}\\ [\bar{\boldsymbol{z}}+(1+\alpha_p)\boldsymbol{e}]\in\mathcal{K} & \bar{\boldsymbol{z}}\notin\mathcal{K}\end{cases} \tag{49}$$

3) 变量κ和τ

变量κ和τ的初值均取 1。

3.3.3　改进热启动法

在求解序列二阶锥规划问题时, 冷启动内点法及传统热启动内点法未能充分利用前一次序列优化的结果。因此, 文献 [35] 提出了一种序列二阶锥

规划改进热启动方法，该方法取热启动参数如下：

$$\begin{cases} \boldsymbol{x}=\omega\boldsymbol{x}' \\ \boldsymbol{y}=\omega\boldsymbol{y}' \\ \boldsymbol{s}=\omega\boldsymbol{s}'+(1-\omega)\boldsymbol{e} \\ \boldsymbol{z}=\omega\boldsymbol{z}'+(1-\omega)\boldsymbol{e} \\ \kappa=\boldsymbol{s}^{\mathrm{T}}\boldsymbol{z}/D \\ \tau=1 \end{cases} \tag{50}$$

式中：$(\boldsymbol{x}',\boldsymbol{y}',\boldsymbol{s}',\boldsymbol{z}')$为序列二阶锥规划中上一次优化的结果；$\omega$取值接近1。

热启动初值较为理想时，初始$(\boldsymbol{r}_x,\boldsymbol{r}_y,\boldsymbol{r}_z,r_\tau,\mu)$的值较小，此时，内点法收敛到指定精度所需的迭代次数较少，相关的证明可参考文献[35]。

3.4 二阶锥规划子问题求解精度调整策略

在序列二阶锥规划方法中，各二阶锥规划子问题的精度可分为凸化精度与求解精度两部分，凸化精度受前后2次优化量的接近程度、线性化方法、离散方法与离散点数等的影响[46]；求解精度即内点法求解二阶锥规划子问题的精度。

传统的序列二阶锥规划方法在求解各二阶锥规划子问题时，均是让内点法在收敛到同一较高精度时才停止迭代。然而，在序列二阶锥规划初期的计算中，由于初始猜想的不精确性，在经过泰勒展开线性化等凸化方法处理后，二阶锥规划子问题的凸化精度较低，此时将求解精度设置得过高并无意义。

因此，可对各二阶锥规划子问题的求解精度进行动态调整，在二阶锥规划子问题的凸化精度较低时，通过降低内点法的求解精度提高求解速度；在后续求解二阶锥规划子问题时，再逐步提高内点法的求解精度，使得序列二阶锥规划问题最终收敛到高精度的解。

假设内点法初始猜想$(\boldsymbol{x}_0,\boldsymbol{s}_0,\boldsymbol{y}_0,\boldsymbol{z}_0,\kappa_0,\tau_0)$的等式约束残差为$\zeta_0$，对偶间隙为$\delta_{\mathrm{gap0}}$，相对对偶间隙为$\delta_{\mathrm{relgap0}}$，则求解精度可如下设置：

$$\varpi=\min\left\{\frac{\zeta_0}{M},\frac{\delta_{\mathrm{gap0}}}{M},\frac{\delta_{\mathrm{relgap0}}}{M}\right\} \tag{51}$$

$$o_{\mathrm{feastol}}=o_{\mathrm{abstol}}=o_{\mathrm{reltol}}=\max\{\varpi,o_{\max}\} \tag{52}$$

式中：M为设定的精度提高倍数；$o_{\max}$为内点法最高优化精度。

由式(51)和式(52)可知，如果$\varpi>o_{\max}$，内点法求解结果的精度达到初始猜想的精度的$1/M$时退出循环；如果$\varpi<o_{\max}$，内点法求解结果的精度达到最高精度时退出循环。

为避免序列二阶锥规划方法在初始几次的迭代中，因为内点法优化精度ϖ太低而导致前后2个二阶锥规划子问题变化量较小，从而出现序列二阶锥规划过早终止迭代的情况。首先，精度提高倍数M的值不应太小；其次，在序列二阶锥规划方法迭代的终止条件中还应加上内点法优化精度的判断，即

$$\varpi<o_{\min} \tag{53}$$

式中：$o_{\min}$为内点法最低优化终止精度。

3.5 内点法定制化改进的序列二阶锥规划流程

综合3.1节～3.4节的KKT方程组求解加速方法、改进热启动方法及二阶锥规划子问题求解精度调整策略，可得到内点法定制化改进的序列二阶锥规划流程图如图5所示。

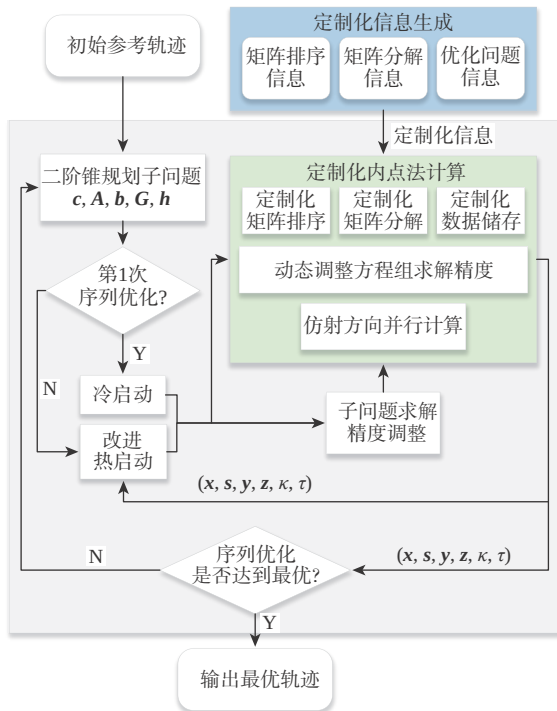


图5 基于定制化改进内点法的序列二阶锥规划流程

Fig. 5 Successive second-order cone programming based on customized and modified interior point method

如图5所示，在实施时，①根据二阶锥规划子问题的结构生成定制化信息（比如矩阵排序的信息，矩阵定制化分解时所需的$\boldsymbol{L}$矩阵非零元素位置信息等）；②根据初始参考轨迹生成二阶锥规子问题的$\boldsymbol{c},\boldsymbol{A},\boldsymbol{b},\boldsymbol{G},\boldsymbol{h}$矩阵及向量；③由于在第1次序列二阶锥规划迭代时无法获得热启动值，因此，第1个二阶锥规划子问题需根据冷启动法计算内点法优化变量的初值，在后续的二阶锥规划子问题的求解中，再根据改进热启动法计算内点法优化变量的初值；④在计算得到内点法优化变量的初值后，根据二阶锥规划子问题求解精度调整策略，计算当前优化问题的求解精度；⑤将内点法变量初值及优化精度赋予定制化内点法求解器中进行求解；⑥在定

制化求解器计算时, 求解器根据得到的定制化信息进行快速矩阵排序、矩阵分解及线性方程组并行计算, 在内点法的计算结果达到设定的二阶锥规划子问题求解精度时, 求解器输出求解结果; ⑦根据求解器的求解结果判断是否达到序列二阶锥规划算法的最优条件, 即式 (23) 及式 (53), 当达到最优条件时, 输出最优轨迹, 否则更新二阶锥规划子问题, 再次求解。

值得注意的是, 定制化信息的计算需要二阶锥规划子问题 $\boldsymbol{A}$ 矩阵和 $\boldsymbol{G}$ 矩阵的非零元素位置信息, 以及 $\boldsymbol{c}$、$\boldsymbol{b}$、$\boldsymbol{h}$ 向量的维数信息, 但定制化信息的计算仅需在整个过程实施一次, 且计算结果可在主程序运行前预置于定制化求解器中。

4　数值仿真与分析

为验证本文的月面应急上升交会轨迹规划方法及基于定制化改进内点法的序列二阶锥规划方法(后文称改进序列二阶锥规划方法), 用改进序列二阶锥规划方法求解 ODP 问题。数值仿真参数如表 2 所示。

表 2　上升器参数

Table 2　Parameters of ascender

上升器总重 m_{wet}/kg	上升器干重 m_{dry}/kg	主发动机推力 P_1/N	主发动机比冲 I_{sp1}/s	推力器总推力 P_2/N	推力器比冲 I_{sp2}/s
800	370	3 000	313.7	960	290

轨道器的轨道参数为: 半长轴 $a_{\text{o}} = 1\,937.4$ km, 偏心率 $\hat{e}_{\text{o}} = 0$, 轨道倾角 $i_{\text{o}} = 43.75°$, 升交点赤径 $\Omega_{\text{o}} = 232°$; 轨道器初始相角 $\theta_{\text{o0}} = 104.7°$。上升器的初始位置为: $R_0 = 1\,737.4$ km, $\theta_0 = 0°$, $\phi_0 = 0.744\,9°$; 由于月球自转, 上升器具有初始速度: $V_{r0} = 0$ m/s, $V_{t0} = 3.382\,5$ m/s, $V_{p0} = 0.007\,2$ m/s。

4.1　求解结果及分析

以表 2 的仿真参数为例, 应用本文的定制化求解器(命名为 MCOS)及改进序列二阶锥规划方法求解 ODP 问题。此外, 为验证本文方法的正确性, 应用通用求解器 ECOS 及传统序列二阶锥规划方法求解 ODP 问题以进行对比分析。为避免降低优化精度对 ECOS 最终优化结果的影响, ECOS 的优化精度仍设置为其默认的 1×10^{-8}; 在 MCOS 中, KKT 方程组修正阈值设置为 $\varepsilon_{\text{abs}} = \varepsilon_{\text{rel}} = 1\times10^{-3}$, 热启动参数 $\omega = 0.992$, 内点法最低优化终止精度 $o_{\min} = 1\times10^{-4}$, 内点法最高优化精度 $o_{\max} = 1\times10^{-8}$, 序列二阶锥规划收敛精度 $\delta_{\text{SOCP}} = 1\times10^{-3}$。

取主发动机工作段离散点数 $N_1 = 11$, 推力器工作段离散点数 $N_2 = 21$。仿真结果如图 6～图 8 所示, 图中改进方法表示以定制化求解器 MCOS 及改进序列二阶锥规划方法求解 ODP 问题的结果, 原方法表示通用求解器 ECOS 及传统序列二阶锥规划方法求解 ODP 问题的结果。

图 6 中, 上升器在月面等待轨道器飞行到相应的位置后再起飞, 月面等待时间为 4 897.4 s, 上升交会飞行时间为 952.4 s, 即上升器在 5 849.8 s 后完成与轨道器的远程交会。在图 7 中, 为实现应急上升交会, 原方法与改进方法均耗尽了燃料。在图 8 中, 由于推力角度变化率的约束, 主发动机工作段

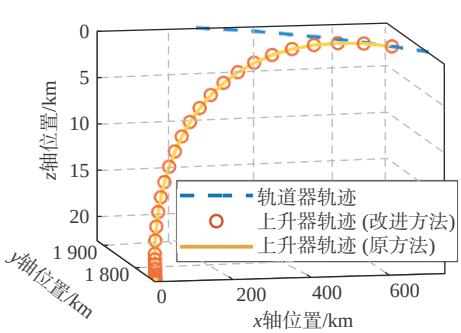


图 6　三维轨迹

Fig. 6　3D trajectories

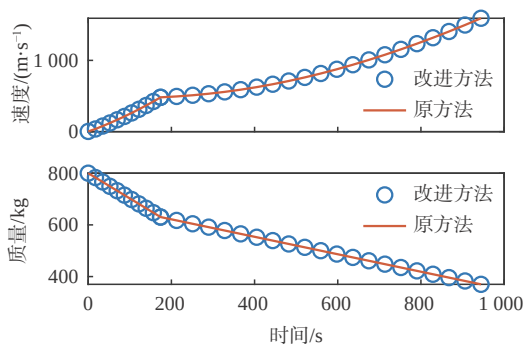


图 7　速度和质量曲线

Fig. 7　Velocity and mass curves

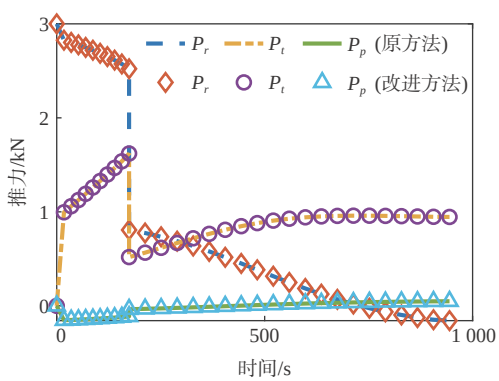


图 8　推力曲线

Fig. 8　Thrust curves

的第1个离散点和第2个离散点的变化是上升器从垂直起飞到转弯完成的过程。由图6～图8可知，原方法和改进方法的求解结果一致。

4.2 计算效率对比分析

4.2.1 热启动效果及分析

以求解ODP为例，对比各种内点法启动方法的迭代次数，取离散节点数$N_1=21$，$N_2=41$，本节的热启动效果分析不涉及二阶锥规划子问题的精度调整，因此，内点法在求解各二阶锥规划子问题时，求解精度均设置为$o_{\text{feastol}}=o_{\text{abstol}}=o_{\text{reltol}}=1\times10^{-8}$，热启动参数$\omega=0.999\,2$，不同启动方法的迭代次数对比如图9所示。

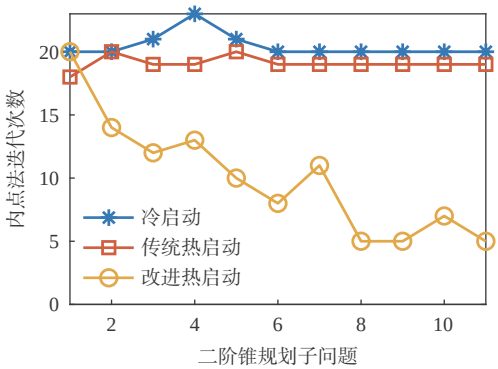


图9 不同启动方法迭代次数

Fig. 9 Iteration times of different starting methods

由图9可知，改进热启动内点法所需的迭代次数相较于冷启动内点法以及传统热启动内点法而言更少，且随着迭代的进行，改进热启动内点法所需的迭代次数呈下降趋势。原因是对于上升交会问题ODP，不等式约束的$\boldsymbol{G}$和$\boldsymbol{h}$及等式约束的$\boldsymbol{A}$和$\boldsymbol{b}$随着序列二阶锥规划迭代的进行变化量越来越小，取$\delta_{\boldsymbol{A}}$，$\delta_{\boldsymbol{b}}$，$\delta_{\boldsymbol{G}}$，$\delta_{\boldsymbol{h}}$如下：

$$\delta_{\boldsymbol{A}}=\left\|\boldsymbol{A}^{i+1}-\boldsymbol{A}^{i}\right\|_F \tag{54}$$

$$\delta_{\boldsymbol{b}}=\left\|\boldsymbol{b}^{i+1}-\boldsymbol{b}^{i}\right\|_2 \tag{55}$$

$$\delta_{\boldsymbol{G}}=\left\|\boldsymbol{G}^{i+1}-\boldsymbol{G}^{i}\right\|_F \tag{56}$$

$$\delta_{\boldsymbol{h}}=\left\|\boldsymbol{h}^{i+1}-\boldsymbol{h}^{i}\right\|_2 \tag{57}$$

式中：$\boldsymbol{A}^i$和$\boldsymbol{A}^{i+1}$为第i个和第i+1个二阶锥规划子问题的等式约束矩阵；$\boldsymbol{b}^i$和$\boldsymbol{b}^{i+1}$为第i个和第i+1个二阶锥规划子问题的等式约束右端项；$\boldsymbol{G}^i$和$\boldsymbol{G}^{i+1}$为第i个和第i+1个二阶锥规划子问题的不等式约束矩阵；$\boldsymbol{h}^i$和$\boldsymbol{h}^{i+1}$为第i个和第i+1个二阶锥规划子问题不等式约束的右端项。优化问题的变化趋势如图10所示。

由图10可知，随着序列二阶锥规划迭代的进行，优化问题的数值变化趋于下降，优化变量趋向于收敛，即本次序列二阶锥规划的结果与上一次序列二阶锥规划的结果越来越接近，因而热启动内点法所需的迭代次数随着序列二阶锥规划迭代的进行越来越少。

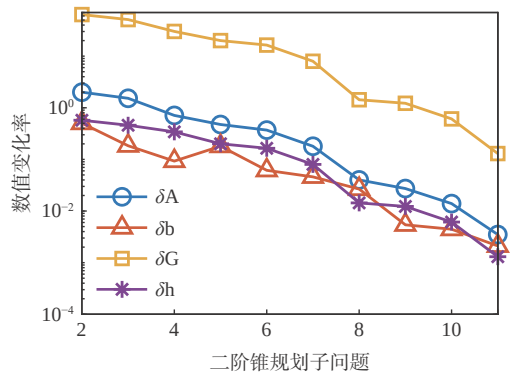


图10 优化问题变化趋势

Fig. 10 Optimization problem evolution trend

4.2.2 二阶锥规划子问题求解精度调整效果及分析

以4.2.1节的仿真参数为例，对比不同的精度提高倍数M，对序列二阶锥规划迭代次数、内点法总迭代次数以及平均每次内点法的迭代次数的影响。对比结果如表3所示。

表3 不同M值下的计算结果对比

Table 3 Comparison of calculation results under different M values

M	序列二阶锥规划迭代次数	内点法总迭代次数
20	15	96
40	15	102
60	15	102
100	10	60
200	10	64
400	10	72
600	12	84
1 000	11	89

由表3可知，M值的大小会对序列二阶锥规划的迭代次数产生影响，在M值较小时，二阶锥规划子问题的求解精度较低，各二阶锥规划子问题的线性化参考轨迹的误差较大，因而序列二阶锥规划的迭代次数较多，内点法迭代次数也较多。在M值增大时，序列二阶锥规划的迭代次数减少，但M值增大到一定程度时，二阶锥规划子问题的求解精度过高，因而内点法的迭代次数也较多。随着M值的增加，总的内点法迭代次数呈先下降后上升的趋势。

4.3 综合加速效果与求解精度对比分析

以4.1节的仿真为例，通过取不同的离散点数得到不同规模的优化问题，对于不同的优化问题，运行100次取平均值计算求解耗时。仿真设置与

4.1 节一致。以台式计算机作为仿真平台, 仿真环境为 Visual Studio 2019, 仿真语言为 C 语言, 台式计算机的 CPU 为 Intel Core i3 10100F(基准频率 3.6 GHz, 最高频率 4.3 GHz, 三级缓存 6 MB), 内存 8 G。不同离散点数的热启动参数ω以及精度提高倍数M如表 4 所示, 统计结果如图 11 及表 5 和表 6 所示, 离散点数 11+21 表示主发动机工作段离散点数为 11 个, 推力器工作段的离散点数为 21 个, 其余离散点数同理。

表 4　不同离散点数下的参数设置

Table 4　Parameter setting under different discrete points

离散点数(N_1+N_2)	热启动参数ω	精度提高倍数M
11+21	0.992 0	200
21+41	0.999 2	100
31+61	0.999 6	100
41+81	0.999 6	100

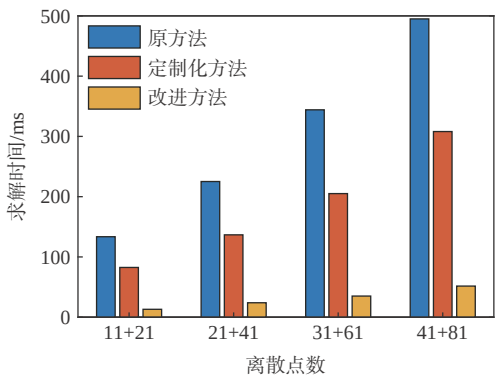


图 11　不同方法的求解耗时统计图

Fig. 11　Statistics of runtime by different methods

表 5　不同方法的求解耗时统计

Table 5　Statistics of runtime by different methods

离散点数 (N_1+N_2)	原方法 耗时/ms	定制化方法 耗时/ms	改进方法 耗时/ms
11+21	133.60	82.60	13.15
21+41	225.20	136.70	24.00
31+61	344.10	205.10	35.10
41+81	495.00	308.10	51.60

表 6　不同方法加速比

Table 6　Speedup ratios of different methods

离散点数 (N_1+N_2)	定制化方法 加速比	改进方法 加速比
11+21	1.62	10.19
21+41	1.65	9.39
31+61	1.68	9.81
41+81	1.61	9.58

为进一步体现本文方法的有效性, 除了与原方法对比外, 本文还增加了与文献 [27] 中定制化方法(后文简称定制化方法)的对比, 但由于文献 [27] 中的 Bsocp 求解器不开源, 因此, 本文仅应用文献 [27] 中的定制化方法的原理(即本文的 3.2.1 节、3.2.2 节和 3.2.3 节中的方法)对 ECOS 进行改进。

定义加速比χ如下:

$$\chi=\frac{\text{加速前的计算耗时}}{\text{加速后的计算耗时}} \tag{58}$$

由图 11、表 5 和表 6 可知, 相比于传统的序列二阶锥规划方法, 在仅应用定制化方法时可达到 1.6 倍的加速比。与文献 [27] 相比, 在应用本文的定制化方法时未出现加速效果随问题规模增大而下降的情况, 这是因为本文在对求解器进行定制化时未通过代码自动生成的方式实现, 无需生成庞大的可执行代码, 因而不会出现文献 [27] 中所述的由于生成的代码量过于庞大而导致的缓存丢失的现象。

对比原方法与改进方法可知, 本文提出的方法相比于传统的序列二阶锥规划方法可达到 9.5 倍左右的加速比。

为进一步对比各种方法的求解精度, 本文通过对优化问题 ODP 取不同的离散点数, 对比优化结果的精度。在计算优化结果的精度时, 本文根据优化得到的控制量进行积分计算, 积分方法为四阶龙格库塔法, 积分步长为 0.01 s。精度对比结果如表 7 所示。

表 7　不同方法的求解精度

Table 7　Solving accuracies of different methods

离散点数 (N_1+N_2)	方法	位置误差/km	速度误差/ ($m\cdot s^{-1}$)
11+21	原方法	16.280 4	33.529 9
	定制化方法[27]	16.280 4	33.529 9
	改进方法	16.155 7	33.268 1
21+41	原方法	7.942 4	16.482 4
	定制化方法[27]	7.942 4	16.482 4
	改进方法	8.118 8	16.841 4
31+61	原方法	5.262 3	10.938 1
	定制化方法[27]	5.262 3	10.938 1
	改进方法	5.312 4	11.029 8
41+81	原方法	4.080 0	8.472 8
	定制化方法[27]	4.080 0	8.472 8
	改进方法	3.933 3	8.172 2

由表 7 可知, 原方法与定制化方法相比, 求解精度完全一样, 这是因为在对求解器进行定制化改进后, 内点法的求解流程没有发生任何改变, 因而求解结果完全相同。但原方法与本文的改进方法求解结果存在偏差, 这是因为改进方法在求解各二阶锥规划子问题时的初值与收敛精度和原方法相比存在偏差, 因而两方法的优化结果也会存在微弱的偏差。由位置误差和速度误差也可看出, 原方法

与改进方法的求解精度差距微弱。

5 结 论

1）本文基于序列二阶锥规划方法，提出了一种月面应急上升交会联合轨迹优化方法，并且应用内点法定制化方法、改进热启动法、二阶锥规划子问题精度调整的方法，对月面上升交会的序列二阶锥规划算法进行加速。

2）仿真表明，在普通台式计算机上，综合应用本文的方法，相较于传统序列二阶锥规划方法，本文方法在不降低求解精度的情况下可达到9.5倍左右的加速比。

参考文献（References）

[1] 欧阳自远. 我国月球探测的总体科学目标与发展战略[J]. 地球科学进展, 2004, 19(3): 351-358.
OUYANG Z Y. Scientific objectives of Chinese lunar exploration project and development strategy[J]. Advances in Earth Science, 2004, 19(3): 351-358(in Chinese).

[2] 李鑫, 刘莹莹, 周军. 载人登月舱上升入轨段的制导律设计[J]. 系统工程与电子技术, 2011, 33(11): 2480-2484.
LI X, LIU Y Y, ZHOU J. Design of guidance law for lunar ascent phase of manned lunar module[J]. Systems Engineering and Electronics, 2011, 33(11): 2480-2484(in Chinese).

[3] 马克茂, 陈海朋. 登月舱上升段最优轨迹设计[J]. 中国空间科学技术, 2013, 33(2): 54-60.
MA K M, CHEN H P. Optimal trajectory design for the lunar module in ascent stage[J]. Chinese Space Science and Technology, 2013, 33(2): 54-60(in Chinese).

[4] HULL D G, HARRIS M W. Optimal solutions for quasiplanar ascent over a spherical moon[J]. Journal of Guidance Control Dynamics, 2012, 35(4): 1218-1224.

[5] MA L, SHAO Z J, CHEN W F, et al. Three-dimensional trajectory optimization for lunar ascent using Gauss pseudospectral method[C]// Proceedings of the AIAA Guidance, Navigation, and Control Conference. Reston: AIAA, 2016.

[6] 张磊. 月面上升的轨迹优化与参数分析[J]. 深空探测学报, 2019, 6(4): 391-397.
ZHANG L. Trajectory optimization and parameter analysis for lunar ascent[J]. Journal of Deep Space Exploration, 2019, 6(4): 391-397(in Chinese).

[7] 宋征宇, 王聪, 巩庆海. 运载火箭上升段推力下降故障的自主轨迹规划方法[J]. 中国科学 (信息科学), 2019, 49(11): 1472-1487.
SONG Z Y, WANG C, GONG Q H. Autonomous trajectory planning for launch vehicle under thrust drop failure[J]. Scientia Sinica (Informationis), 2019, 49(11): 1472-1487(in Chinese).

[8] HAO Z M, ZHANG R. Onboard real-time generation of launch vehicle abort orbits[J]. Journal of Guidance Control Dynamics, 2021, 44(8): 1541-1549.

[9] BENEDIKTER B, ZAVOLI A, COLASURDO G, et al. Convex approach to three-dimensional launch vehicle ascent trajectory optimization[J]. Journal of Guidance Control Dynamics, 2021, 44(6): 1116-1131.

[10] 王华, 唐国金. 用遗传算法求解双冲量最优交会问题[J]. 中国空间科学技术, 2003, 23(1): 26-30.
WANG H, TANG G J. Solving optimal rendezvous using two impulses based on genetic algorithms[J]. Chinese Space Science and Technology, 2003, 23(1): 26-30(in Chinese).

[11] 冉茂鹏, 王青. 一种基于EPSO的航天器交会轨迹优化方法[J]. 宇航学报, 2013, 34(9): 1195-1201.
RAN M P, WANG Q. Spacecraft rendezvous trajectory optimization method based on EPSO[J]. Journal of Astronautics, 2013, 34(9): 1195-1201(in Chinese).

[12] 李君龙, 李松洲, 周荻. 一种多约束条件下的三脉冲交会优化设计方法[J]. 系统工程与电子技术, 2022, 44(8): 2612-2620.
LI J L, LI S Z, ZHOU D. Optimization method for three-impulse rendezvous under multi-constraints[J]. Systems Engineering and Electronics, 2022, 44(8): 2612-2620(in Chinese).

[13] 任远, 崔平远, 栾恩杰. 最优两脉冲行星际轨道转移优化算法[J]. 航空学报, 2007, 28(6): 1307-1311.
REN Y, CUI P Y, LUAN E J. Interplanetary optimum two-impulse transfer trajectories[J]. Acta Aeronautica et Astronautica Sinica, 2007, 28(6): 1307-1311(in Chinese).

[14] YUE X C, YANG Y, GENG Z Y. Indirect optimization for finite-thrust time-optimal orbital maneuver[J]. Journal of Guidance, Control, and Dynamics, 2010, 33(2): 628-634.

[15] 武冠群. 在轨服务航天器交会轨迹优化与近距离安全接近控制研究[D]. 哈尔滨: 哈尔滨工业大学, 2019.
WU G Q. Research on rendezvous trajectory optimization and close-range safe approach control of on-orbit service spacecraft[D]. Harbin: Harbin Institute of Technology, 2019(in Chinese).

[16] 管凯颜. 不同质量航天器在连续推力下快速协同交会探讨[D]. 济南: 山东大学, 2021.
GUAN K Y. Discussion on rapid cooperative rendezvous of spacecraft with different masses under continuous thrust[D]. Jinan: Shandong University, 2021(in Chinese).

[17] 李萌, 龚胜平, 彭坤, 等. 直接优化算法在快速交会组合变轨策略中的应用[J]. 载人航天, 2017, 23(2): 156-162.
LI M, GONG S P, PENG K, et al. Application of direct optimization algorithmin synthetic orbit maneuver strategy of short rendezvous[J]. Manned Spaceflight, 2017, 23(2): 156-162(in Chinese).

[18] 周新耀, 陶勇鹏, 曾亮, 等. 空间轨道交会非线性优化方法研究[J]. 上海航天, 2018, 35(1): 54-58.
ZHOU X Y, TAO Y P, ZENG L, et al. Study on nonlinear optimization of space orbit interception[J]. Aerospace Shanghai (Chinese & English), 2018, 35(1): 54-58(in Chinese).

[19] ORTOLANO N, GELLER D K, AVERY A. Autonomous optimal trajectory planning for orbital rendezvous, satellite inspection, and final approach based on convex optimization[J]. The Journal of the Astronautical Sciences, 2021, 68(2): 444-479.

[20] 靳锴, 罗建军, 郑茂章, 等. 考虑导航误差和摄动影响的椭圆轨道最优交会制导[J]. 控制理论与应用, 2018, 35(10): 1484-1493.
JIN K, LUO J J, ZHENG M Z, et al. Guidance design with navigation errors for relative motion in noncircular perturbed orbits[J]. Control Theory & Applications, 2018, 35(10): 1484-1493(in Chinese).

[21] WANG Z B, GRANT M J. Optimization of minimum-time low-

thrust transfers using convex programming[J]. Journal of Spacecraft and Rockets, 2018, 55(3): 586-598.

[22] WANG Z B, GRANT M J. Minimum-fuel low-thrust transfers for spacecraft: a convex approach[J]. IEEE Transactions on Aerospace and Electronic Systems, 2018, 54(5): 2274-2290.

[23] 王访寒, 周如好, 余薛浩, 等. 基于序列凸优化的上面级远程交会轨迹优化[J]. 上海航天 (中英文), 2022, 39(2): 45-52.
WANG F H, ZHOU R H, YU X H, et al. Trajectory optimization of remote rendezvous for upper stage based on sequential convex optimization[J]. Aerospace Shanghai(Chinese& English), 2022, 39(2): 45-52(in Chinese).

[24] 彭祺擘. 考虑应急返回能力的载人登月轨道优化设计及特性分析[D]. 长沙: 国防科学技术大学, 2012.
PENG Q B. Optimal design and characteristic analysis of manned moon orbit considering emergency return capability[D]. Changsha: National University of Defense Technology, 2012(in Chinese).

[25] STURM J F. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones[J]. Optimization Methods and Software, 1999, 11(1-4): 625-653.

[26] DOMAHIDI A, CHU E, BOYD S. ECOS: an SOCP solver for embedded systems[C]//Proceedings of the 2013 European Control Conference. Piscataway: IEEE Press, 2013: 3071-3076.

[27] DUERI D, ACIKMESE B, SCHARF D P, et al. Customized real-time interior-point methods for onboard powered-descent guidance[J]. Journal of Guidance Control Dynamics, 2017, 40(2): 197-212.

[28] MATTINGLEY J, BOYD S. CVXGEN: a code generator for embedded convex optimization[J]. Optimization and Engineering, 2012, 13(1): 1-27.

[29] JUNG J H, O'LEARY D P. Implementing an interior point method for linear programs on a CPU-GPU system[J]. Electronic Transactions on Numerical Analysis, 2008, 28: 174-189.

[30] WANG J W, ZHANG R, LI H F. Onboard optimization of multi-arc trajectories with constraints on duration of arcs[J]. Acta Astronautica, 2022, 192: 434-442.

[31] BOYD S. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends® in Machine Learning, 2010, 3(1): 1-122.

[32] O'DONOGHUE B, CHU E, PARIKH N, et al. Conic optimization via operator splitting and homogeneous self-dual embedding[J]. Journal of Optimization Theory and Applications, 2016, 169(3): 1042-1068.

[33] ELANGO P, KAMATH A, YU Y, et al. A customized first-order solver for real-time powered-descent guidance[C]//Proceedings of the AIAA Scitech Forum. Reston: AIAA, 2022: 0951.

[34] SKAJAA A, ANDERSEN E D, YE Y Y. Warmstarting the homogeneous and self-dual interior point method for linear and conic quadratic problems[J]. Mathematical Programming Computation, 2013, 5(1): 1-25.

[35] CHEN Y S, YANG G W, WANG L, et al. Iteration complexity of an infeasible interior point methods for seconder-order cone programming and its warmstarting[EB/OL]. (2023-01-24)[2025-02-04]. https://doi.org/10.48550/arXiv.2205.03570.

[36] ZHANG Z Y, ZHAO Q C, DAI F A. A warm-start strategy in interior point methods for shrinking horizon model predictive control with variable discretization step[J]. IEEE Transactions on Automatic Control, 2023, 68(6): 3830-3837.

[37] CHEN Y S, YANG G W, WANG L, et al. A fast algorithm for onboard atmospheric powered descent guidance[J]. IEEE Transactions on Aerospace and Electronic Systems, 2023, 59(5): 6112-6123.

[38] 郝泽明, 张冉, 王嘉炜, 等. 大气层内固体火箭实时轨迹优化方法[J]. 宇航学报, 2021, 42(11): 1416-1426.
HAO Z M, ZHANG R, WANG J W, et al. Real-time atmospheric trajectory optimization for solid rockets[J]. Journal of Astronautics, 2021, 42(11): 1416-1426(in Chinese).

[39] VANDENBERGHE L. The CVXOPT linear and quadratic cone program solvers[EB/OL]. (2010-03-20)[2023-03-28]. http://cvxopt.org/documentation/coneprog.pdf.

[40] DOMAHIDI A. Methods and tools for embedded optimization and control[D]. Zurich: ETH Zurich, 2013.

[41] DAVIS T A, GILBERT J R, LARIMORE S I, et al. A column approximate minimum degree ordering algorithm[J]. ACM Transactions on Mathematical Software, 2004, 30(3): 353-376.

[42] GEORGE A, LIU J W H. The evolution of the minimum degree ordering algorithm[J]. SIAM Review, 1989, 31(1): 1-19.

[43] DUFF I S, ERISMAN A M, REID J K. Direct methods for sparse matrices[M]. New York: Oxford University Press, 2017.

[44] GILL P E, SAUNDERS M A, SHINNERL J R. On the stability of cholesky factorization for symmetric quasidefinite systems[J]. SIAM Journal on Matrix Analysis and Applications, 1996, 17(1): 35-46.

[45] DAVIS T A. Algorithm 849[J]. ACM Transactions on Mathematical Software, 2005, 31(4): 587-591.

[46] GAO Y, SHAO Z J, SONG Z Y. Enhanced successive convexification based on error-feedback index and line search filter[J]. Journal of Guidance Control Dynamics, 2022, 45(12): 2243-2257.

Rapid planning method for lunar direct ascent and rendezvous trajectory in emergency cases

BAN Huanheng[1, 2], ZHOU Cong[1, 2], YAN Xiaodong[1, 2, *]

(1. School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China;
2. Shaanxi Aerospace Flight Vehicle Design Key Laboratory, Xi'an 710068, China)

Abstract: When a lunar detector encounters unexpected conditions and needs a return to the lunar orbiter or Earth, the lunar ascender should be able to plan an appropriate ascent and rendezvous trajectory independently in emergency cases. In this paper, a direct ascent and rendezvous trajectory optimization model was established based on a successive second-order cone programming method, and minimum ascent and rendezvous time was chosen as the objective function. Additionally, a convex optimization algorithm was proposed to solve the ascent and rendezvous trajectory optimization problem. In order to enhance computational efficiency, the interior point method was customized and modified. The main adaptations included: the process of solving linear equations was customized and modified; a warm starting was used in the interior point method; the second-order cone programming subproblem solving accuracy was adjusted dynamically. Simulation results demonstrate that the proposed lunar ascent and rendezvous trajectory optimization method in emergency cases facilitates the rapid ascent and rendezvous process of the lunar ascender. Moreover, when compared to the traditional successive second-order cone programming method employing a general interior point method solver, the proposed acceleration technique achieves a speedup ratio of approximately 9.5 times while maintaining the same level of solving accuracy. Consequently, the method outlined in this paper holds significant potential for enabling autonomous online trajectory planning of lunar ascenders.

Keywords: lunar ascent and rendezvous process in emergency cases; successive second-order cone programming method; customized interior point method; warm starting in interior point method; online trajectory planning

Received: 2023-04-04; **Accepted**: 2023-07-11; **Published Online**: 2023-07-31 09:25
URL: link.cnki.net/urlid/11.2625.V.20230728.1648.002
* **Corresponding author**. E-mail: yan804@nwpu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0180

多智能体仿射编队机动控制的有向交互拓扑优化设计

马苏慧[1]，张栋[1,2,*]，王孟阳[1]，王庭晖[1]，刘送丹[3]

(1. 西北工业大学 航天学院，西安 710072；2. 陕西省空天飞行器设计重点实验室，西安 710072；
3. 中国航天科技创新研究院，北京 100083)

摘 要：研究了多智能体仿射编队机动控制的有向交互拓扑优化设计问题。考虑信息交互成本、信息传播能耗等优化指标，建立了包含拓扑结构构造和权重配置的仿射编队机动有向拓扑优化模型；针对仿射编队机动拓扑结构构造，提出了有向 k 根图的检测方法，实现了有向信息交互拓扑的 d+1 根约束求解，并设计了一种改进 NSGA-II 的拓扑结构构造优化算法，以二维空间中 7 个智能体的编队为例进行仿真验证，结果表明：改进 NSGA-II 的拓扑结构构造优化算法具有更好的优化效果，能够有效地为仿射编队机动控制提供多种可行的有向交互拓扑，并且所生成的交互拓扑能够满足有向 d+1 根图的要求。

关 键 词：多智能体系统；仿射变换；编队控制；交互拓扑设计；有向图；多目标优化
中图分类号：TP273; N945.15
文献标志码：A **文章编号**：1001-5965(2025)04-1367-10

作为集群运动的行为基础，编队控制是多智能体系统协同控制领域的一个热点研究问题，被广泛应用于以无人机、机器人、水下航行器、卫星等为对象的任务场景。为实现几何队形的生成与保持，多数编队控制器均显式地包含期望编队队形信息[1]，其设定的交互拓扑只影响编队误差的收敛过程。不同于对编队信息的直接利用，相似编队控制[2]和仿射编队控制[3]利用特定交互拓扑的性质，将期望的基准编队队形信息隐藏在拓扑权重中，所设计的交互拓扑将直接决定智能体的最终收敛状态并影响编队误差收敛过程，使得智能体的位置状态收敛于标称队形的相似或者仿射映射空间。因此，面对障碍物或者其他突发情况，智能体仍旧能够使用原有的相似或者仿射控制协议及参数设置，去实现经过几何变换的目标编队。为进一步实现目标队形的连续变换，相似编队机动控制[4,5]和仿射编队机动控制[6,7]被提出，其中，前者可以实现二维平面中队形的平移、旋转和缩放变换，后者还可以在三维甚至更高维空间额外地实现几何队形的切变变换。鉴于参数设置对目标队形改变不敏感且队形变换类型丰富，仿射编队控制及其衍生的仿射编队机动控制能够更好地响应动态环境和复杂任务，已成为编队控制理论研究中的一个重要课题。

截至目前，围绕仿射编队控制和仿射编队机动控制，学者已经开展了大量的研究，例如，针对离散[8]、高阶[7,9]、异构[10]、欧拉-拉格朗日[11]的多智能体系统或者考虑时延[12]、外部扰动[13]、通信资源受限[14]等情况设计合适的编队控制方案。然而，上述研究工作主要关心如何设计一个能够实现仿射编队或者仿射编队机动的控制协议，很少研究信息交互拓扑的生成方案及其优化设计。

在 d 维空间中，仿射编队控制和仿射编队机动控制的信息交互拓扑设定相同，均要求拓扑结构是(无向)普遍刚性的或者有向 $d+1$ 根的，并且拓扑权

收稿日期：2023-04-14；**录用日期**：2023-08-10；**网络出版时间**：2023-09-05 11:17
网络出版地址：link.cnki.net/urlid/11.2625.V.20230905.1002.003
基金项目：国家自然科学基金 (61933010)
* **通信作者**. E-mail：zhangdong@nwpu.edu.cn

引用格式：马苏慧，张栋，王孟阳，等. 多智能体仿射编队机动控制的有向交互拓扑优化设计 [J]. 北京航空航天大学学报，2025，51（4）：1367-1376. MA S H，ZHANG D，WANG M Y，et al. Directed interactive topology optimization design for multi-agent affine formation maneuver control [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1367-1376（in Chinese）.

重需要满足一定的应力约束[3,7]。结合刚度图理论和张拉框架的概念，Yang等[15]提出了普遍刚性张拉框架算法，可用于生成仿射编队控制所需的无向拓扑。Xiao等[16]设计了基于混合整数半定规划的无向拓扑优化算法，还为采用领导-跟随结构的仿射编队控制提出了领导选择方案。针对有向拓扑切换的情况，Yang等[17]提出了有向$d+1$根图的重构算法，使得有向边的改变不破坏有向$d+1$根图的本质。而关于仿射编队和仿射编队机动控制所要求的有向$d+1$根图生成设计及优化算法研究很少。

仿射编队机动控制的有向交互拓扑优化设计的最大挑战，是如何保证优化设计生成的拓扑结构是有向$d+1$根图。以有向$d+1$根图为硬性约束，可进行拓扑优化问题模型的构建。该类拓扑优化问题往往采用智能优化算法进行求解。例如，Mondal等[18]设计了二维遗传算法以解决一致性控制中满足有向生成树约束和最小能量消耗目标的拓扑优化问题；崔亚妮等[19]以网络连通度为约束条件，通过综合考虑链路通信质量、网络连接收益和连接成本构建多无人船通信网络拓扑优化控制模型，并提出类电磁机制的离散粒子群算法进行求解；谷晓燕等[20]应用改进的人工蜂群算法，对由编队通信总链长、剩余能量不均衡和最大通信跳数多个目标组成的无人机编队信息交互拓扑优化问题进行求解，生成满足度中心性约束条件的无向交互拓扑。上述工作以0-1矩阵为优化变量进行拓扑优化设计，均认为相应的优化问题是单模态的，未考虑当交互拓扑节点被赋予某些属性时优化问题的多模态特性。

基于此，本文研究了仿射编队机动控制的有向交互拓扑优化问题，主要贡献包括：针对仿射编队机动控制设定的有向拓扑设计提出了拓扑优化问题模型，包括有向$d+1$根图拓扑结构构造和拓扑权重配置两个子问题模型；研究了有向k根图的检测方法，并提供了理论支撑；针对具有多模态特性的拓扑结构构造子问题，设计了改进NSGA-II的拓扑结构构造优化算法，可用于仿射编队控制和仿射编队机动控制的拓扑设计。

1 预备知识

1.1 有向图

有向符号图$\mathcal{G}=(\mathcal{V},\mathcal{E},\boldsymbol{W})$表示由$n$个智能体组成的多智能体系统对应的编队交互拓扑，其中$\mathcal{V}=\{1,2,\cdots,n\}$为顶点集，$\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$为边集，$\boldsymbol{W}=[w_{ij}]_{n\times n}\in\mathbf{R}^{n\times n}$为符号邻接矩阵。如果边$(j,i)\in\mathcal{E}$，则称$j$是$i$的一个(入)邻居。有向图$\mathcal{G}$的符号权重$w_{ij}\neq 0$当且仅当$(j,i)\in\mathcal{E}$，$w_{ij}=0$当且仅当$(j,i)\notin\mathcal{E}$。本文用$\boldsymbol{A}=[a_{ij}]_{n\times n}$表示$\boldsymbol{W}$对应有向图$\mathcal{G}$的交互拓扑结构，即$a_{ij}=1$当且仅当$w_{ij}\neq 0$，$a_{ij}=0$当且仅当$w_{ij}=0$；此时$\boldsymbol{A}$为单位邻接矩阵，仅代表有向图$\mathcal{G}$中顶点之间的连接关系。有向符号图$\mathcal{G}$的符号拉普拉斯矩阵为$\boldsymbol{L}=[l_{ij}]_{n\times n}\in\mathbf{R}^{n\times n}$，其中$l_{ii}=\sum_{j=1}^{n}w_{ij}$，$l_{ij}=-w_{ij}$。

图$\mathcal{G}$的一条有向路径是指边$\{(i_0,i_1),(i_1,i_2),\cdots,(i_{l-1},i_l)\}$的有序组合，其中$i_0,i_1,\cdots,i_l$是不同的顶点。从顶点$s$到顶点$t$所有的有向路径被称为$(s,t)$路。如果两条有向路径中没有相同的顶点，则称它们为点不相交的有向路径；如果两条有向路径中没有相同的边，则称它们为边不相交的有向路径。如果仅含一条边的有向路径的首顶点和尾顶点重合，则称其为自环。如果图$\mathcal{G}$中2个相邻顶点拥有2条或者2条以上方向相同的有向边，那么这些边称为平行边。如果图$\mathcal{G}$具有平行边且无自环，则称图$\mathcal{G}$为有向多重图。给定顶点s和t，对于顶点子集$\mathcal{S}\subset\mathcal{V}-\{s,t\}$，如果图$\mathcal{G}$删除$\mathcal{S}$中所有顶点后得到的图$\mathcal{G}(\mathcal{V}-\mathcal{S})$不含有$(s,t)$路，则称$\mathcal{S}$是$\mathcal{G}$的一个$(s,t)$分离因子。最小$(s,t)$分离因子是指具有最小基数的分离因子。一个$(s,t)$割是一个形如$(\mathcal{S},\bar{\mathcal{S}})$的边集合，其中$\mathcal{S}$和$\bar{\mathcal{S}}$形成$\mathcal{V}$的一个顶点集划分，使得$s\in\mathcal{S}$和$t\in\bar{\mathcal{S}}$。

定义1(k可达)[3]：对于有向图$\mathcal{G}$，如果删除除节点v之外的任意$k-1$个节点后，仍能够找到集合$\mathcal{U}$中的一个节点到节点v的一条有向路径，即从集合$\mathcal{U}$到节点v有k条不连通的有向路径，则称节点v对于非单点集合$\mathcal{U}$是k可达的。

定义2(k根)[3]：对于有向图$\mathcal{G}$，如果存在一个包含k个节点的集合(根集)，除该集合之外的每个节点都是k可达的，则称图$\mathcal{G}$是k根的。

定理1(Menger定理)[21]：设$\mathcal{G}=(\mathcal{V},\mathcal{E})$为有向多重图，$s\in\mathcal{V}$和$t\in\mathcal{V}$是一对不同的顶点，则下列陈述成立：

1) 边不相交(s,t)路的最大数目等于覆盖(s,t)路的所有边的最小数目，且这个最小值可由某(s,t)割$(\mathcal{U},\bar{\mathcal{U}})$达到。

2) 如果边(s,t)不在$\mathcal{E}$中，则点不相交的(s,t)路的最大数目等于最小(s,t)分离因子的顶点数目。

1.2 网络流

一个容量网络$\mathcal{N}=(\mathcal{V},\mathcal{E},c)$是指一个有向图$\mathcal{G}=(\mathcal{V},\mathcal{E})$，其具有定义在边集$\mathcal{E}$上的容量函数$c$：对于每一条边$(i,j)\in\mathcal{E}$有$c_{ij}=c(i,j)\geqslant 0$。网络$\mathcal{N}$的流是定义在$\mathcal{N}$的边集上的函数$x$：对每一条边$(i,j)\in\mathcal{E}$有$x_{ij}=x(i,j)\geqslant 0$。用$b(v)$表示所有从顶点$v$出去的流量与所有进入顶点$v$的流量之差。根据$b(v)$可将网络$\mathcal{N}$中的顶点分为3类：若$b(v)>0$，顶

点v被称作网络$\mathcal{N}$的源点; 若$b(v)<0$, 顶点v被称作网络$\mathcal{N}$的汇点; 若$b(v)=0$, 顶点v被称作网络$\mathcal{N}$的平衡顶点。在网络$\mathcal{N}$中, 如果对于每一条边$(i,j)\in\mathcal{E}$总有$x_{ij}\leqslant c_{ij}$, 并且流x具有平衡向量, 则称x是可行流。最大流问题是指对于只有一个源点s和一个汇点t的网络$\mathcal{N}=(\mathcal{V},\mathcal{E},c)$, 从$s$到$t$的可行$(s,t)$流中寻找流值$|x|=b(s)$最大的流, 该流被称为网络$\mathcal{N}$的最大流。一个$(s,t)$割$(\mathcal{S},\bar{\mathcal{S}})$的容量是一个实数$u(\mathcal{S},\bar{\mathcal{S}})$, 指从$\mathcal{S}$到$\bar{\mathcal{S}}$的所有有向边的容量之和。

定理 2(最大流最小割定理)[21]: 设$\mathcal{N}=(\mathcal{V},\mathcal{E},c)$是一个源点为$s$、汇点为$t$的网络。关于$\mathcal{N}$中的每个可行$(s,t)$流$x$, 有下列相互等价的描述:

1) 流x是一个最大流;

2) $\mathcal{N}(x)$中不存在一条(s,t)路;

3) 存在一个(s,t)割$(\mathcal{S},\bar{\mathcal{S}})$, 使得$|x|=u(\mathcal{S},\bar{\mathcal{S}})$。

1.3　仿射定位

假设由n个智能体组成的多智能体系统包含m个跟随智能体和$n-m$个领导智能体, 对应的顶点集分别表示为$\mathcal{V}_{\rm f}=\{1,2,\cdots,m\}$和$\mathcal{V}_{\rm l}=\{m+1,m+2,\cdots,n\}$。在$d$维空间中, n个智能体的位置向量$\boldsymbol{p}=[\boldsymbol{p}_{\rm f}^{\rm T},\boldsymbol{p}_{\rm l}^{\rm T}]^{\rm T}\in\mathbf{R}^{nd}$表示一个队形, 与$n$个智能体对应的$\mathcal{G}$共同组成编队$\mathcal{F}_{\boldsymbol{p}}=(\mathcal{G},\boldsymbol{p})$。若$\boldsymbol{q}=[\boldsymbol{q}_1^{\rm T},\boldsymbol{q}_2^{\rm T},\cdots,\boldsymbol{q}_n^{\rm T}]^{\rm T}=[\boldsymbol{q}_{\rm f}^{\rm T},\boldsymbol{q}_{\rm l}^{\rm T}]^{\rm T}\in\mathbf{R}^{nd}$表示一个标称队形, 则$\mathcal{F}_{\boldsymbol{q}}=(\mathcal{G},\boldsymbol{q})$被称为标称编队。标称队形$\boldsymbol{q}$的仿射映射定义为

$$\mathcal{A}(\boldsymbol{q})=\left\{\boldsymbol{p}=(\boldsymbol{I}_n\otimes\bar{\boldsymbol{A}})\boldsymbol{q}+[1]_{1\times n}\otimes\bar{\boldsymbol{b}}\,|\,\bar{\boldsymbol{A}}\in\mathbf{R}^{d\times d},\bar{\boldsymbol{b}}\in\mathbf{R}^{d}\right\}\tag{1}$$

式中: $(\bar{\boldsymbol{A}},\bar{\boldsymbol{b}})$为仿射变换。标称队形$\boldsymbol{q}$的仿射张成定义为

$$\tilde{\mathcal{S}}(\boldsymbol{q})=\left\{\sum_{i=1}^{n}a_i\boldsymbol{q}_i:a_i\in\mathbf{R},\sum_{i=1}^{n}a_i=1\right\}\tag{2}$$

定义 3[7]: 标称编队$\mathcal{F}_{\boldsymbol{q}}=(\mathcal{G},\boldsymbol{q})$在$\mathbf{R}^d$空间中若能够实现仿射定位, 如下两个条件需要同时满足:

1) 对于$\mathbf{R}^d$空间中的任意$\boldsymbol{p}=[\boldsymbol{p}_{\rm f}^{\rm T},\boldsymbol{p}_{\rm l}^{\rm T}]^{\rm T}\in\mathcal{A}(\boldsymbol{q})$, 能够被唯一确定;

2) 对于$\mathcal{G}$和$\boldsymbol{p}$, 存在一个对应$\mathcal{G}$的符号拉普拉斯矩阵$\boldsymbol{L}\in\mathbf{R}^{n\times n}$满足$(\boldsymbol{L}\otimes\boldsymbol{I}_d)\boldsymbol{p}=\boldsymbol{0}$。

定理 3[7]: 假设给定n个智能体的标称编队$\mathcal{F}_{\boldsymbol{q}}=(\mathcal{G},\boldsymbol{q})$满足$\mathcal{G}$中领导者集合$\mathcal{V}_{\rm l}$有$d+1$个领导者且$\{\boldsymbol{q}_i\}_{i\in\mathcal{V}_{\rm l}}$可以在$\mathbf{R}^d$中仿射张成, 则该标称编队$\mathcal{F}_{\boldsymbol{q}}=(\mathcal{G},\boldsymbol{q})$在$\mathbf{R}^d$中能够实现仿射定位, 当且仅当$\mathcal{G}$是$d+1$根的。

2　拓扑优化问题模型

2.1　问题分析

多智能体系统应用仿射编队思想实现队形生成和队形机动的基本前提是标称编队$\mathcal{F}_{\boldsymbol{q}}$能够实现仿射定位[3,6,7]。当这一前提被满足时, 分布式控制器的设计还必须找到一个对应的符号拉普拉斯矩阵满足。另外, 如果采用如 Lin 等[3,6,7]提出的仿射编队和编队机动控制方案, 系统还需要进一步配置一个对角矩阵$\boldsymbol{D}=\mathrm{diag}\{d_1,d_2,\cdots,d_n\}$使得$\bar{\boldsymbol{L}}=\boldsymbol{D}\boldsymbol{L}$有$d+1$个零根且其余特征根均有正实部。因此, 根据定理 3 和仿射编队机动控制对$\boldsymbol{L}$的基本要求可知, 当标称队形$\boldsymbol{q}$给定且其中领导者身份确定时, 仿射编队控制的拓扑设计问题可分为拓扑结构构造和权重配置两个子问题。

本文假设标称队形$\boldsymbol{q}$给定且其中领导者身份确定。根据定理 3, 对应的拉普拉斯矩阵$\boldsymbol{L}$可重写为如下形式:

$$\boldsymbol{L}=\begin{bmatrix}\boldsymbol{L}_{\rm ff}\in\mathbf{R}^{m\times m} & \boldsymbol{L}_{\rm fl}\in\mathbf{R}^{(n-m)\times(n-m)}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\tag{3}$$

2.2　拓扑结构构造

由定理 3 可知, d维空间中仿射定位可实现的有向交互拓扑结构要求对应的有向图$\mathcal{G}$是$d+1$根的。然而, 满足这一几何条件的交互拓扑并不是唯一的。交互拓扑中有向边的数量和位置决定着多智能体系统信息交互的成本和速度。基于上述思想, 本文选取以下指标为优化对象。

1) 信息交互成本

信息交互拓扑的每条边对应一条物理链路。而物理链路数量越多, 构建信息交互拓扑的成本就越大。因此信息交互成本f_1定义为

$$f_1=\sum_{i\in\mathcal{V}_{\rm f}}\sum_{j\in\mathcal{V}}Y_{ij}^{A}\tag{4}$$

式中: $Y_{ij}^{A}=a_{ij}(i\in\mathcal{V}_{\rm f},\ j\in\mathcal{V})$为拓扑结构构建问题的决策变量; $\boldsymbol{Y}_A=[Y_{ij}^{A}]$。

2) 信息传播能耗

信息交互拓扑中单条信息交互链路的物理能耗与节点之间的距离相关。故定义信息传播能耗f_2为

$$f_2=\sum_{i\in\mathcal{V}_{\rm f}}\sum_{j\in\mathcal{V}}Y_{ij}^{A}d_{ij}\tag{5}$$

式中: d_{ij}表示智能体i与智能体j之间的欧氏距离。

3) 最大中介中心性

为度量每个节点在交互拓扑中承担信息交互的重要程度, 直接引入复杂网络中中介中心性[22]的定义, 即信息交互拓扑中包含节点i的所有最短路径的条数占所有最短路径条数的百分比。最大中介中心性f_3可描述为

$$f_3=\max\left\{\sum_{j}\sum_{k\neq j}\frac{g_{jik}}{g_{jk}}\right\}\tag{6}$$

式中：g_{jk}为节点j到节点k的有向(最短)路径数量；g_{jik}为经过节点i的节点j到节点k的有向(最短)路径数量。节点i的中介中心性代表信息交互拓扑对节点i的依赖能力。最大中介中心性越小，表示每个节点在信息传播的重要程度差异性越小，拓扑的同步能力增强。

4) 领导平均入度

根据有向$d+1$根图的定义可知，有向$d+1$根图最简单的构造思路是直接使每个跟随者智能体均能够获取所有领导智能体的信息，但这种拓扑结构对领导智能体的依赖程度过高，与其余跟随智能体的协作效率过低。因此，应该控制交互拓扑中领导智能体集合与跟随智能体集合之间的有向路径数量。故定义领导平均入度f_4为

$$f_4=\frac{\sum_{i\in\mathcal{V}_f}\sum_{j\in\mathcal{V}_l}Y_{ij}^A}{(n-d-1)} \tag{7}$$

式中：f_4表示跟随智能体的入邻居为领导智能体的平均水平。

在构建的每个有向拓扑中，还必须满足如下的约束条件。

1) 有向$d+1$根

信息交互拓扑是$d+1$根的是仿射编队机动实现的基本前提。注意到，当领导智能体的数量给定为$d+1$时，当前有向拓扑只可能是1根的、2根的、…、$d+1$根的，不可能是$d+2$根的。因此，有向$d+1$根图可视为一个等式约束。记当前信息交互拓扑$\mathcal{G}$是k根的，则定义有向$d+1$根约束函数H_1表示

$$H_1=\frac{k}{d+1}-1=0 \tag{8}$$

2) 无自环

在协同控制中，往往默认信息交互拓扑不存在自环。令o_1表示当前信息交互拓扑中的自环数量，则定义无自环约束函数H_2为

$$H_2=100o_1=0 \tag{9}$$

3) 可配置非零权重

在拓扑权重配置问题中，拉普拉斯矩阵L需要满足$\boldsymbol{L}[1]_{1\times n}=\boldsymbol{0}$和$(\boldsymbol{L}\otimes\boldsymbol{I}_d)\boldsymbol{q}=\boldsymbol{0}$。这些等式方程组将采用消元法求解，因而图$\mathcal{G}$中每条有向边的非零权重将由等式方程组确定的多个自由变量决定。注意到，标称队形$\boldsymbol{q}$的多样性可能导致所设计的自由变量组中存在零元素，进而使得拓扑结构被破坏。因此，针对当前交互拓扑，记求解$\boldsymbol{L}[1]_{1\times n}=\boldsymbol{0}$和$(\boldsymbol{L}\otimes\boldsymbol{I}_d)\boldsymbol{q}=\boldsymbol{0}$等式方程组约束的自由变量组中零元素的个数为$o_2$，并定义可配置非零权重的约束函数$H_3$为

$$H_3=10o_2=0 \tag{10}$$

注意，o_2的确定无需提前进行拓扑权重设计算法的求解，可直接利用消元法确定。

结合上述指标模型和约束条件，可建立拓扑结构构造的多目标优化模型：

$$\begin{cases}\min\ \boldsymbol{F}=[f_1,f_2,f_3,f_4]^{\mathrm{T}}\\ \text{s.t.}\quad H_1=0,H_2=0,H_3=0\end{cases} \tag{11}$$

式中：$\boldsymbol{F}$为拓扑结构构造问题的优化性能指标向量。

2.3 拓扑权重配置

拓扑结构构造的结果是一个$(n-d-1)\times n$维的0-1矩阵$\boldsymbol{Y}_A$，其对应$n\times n$的单位邻接矩阵$\boldsymbol{A}$。拓扑权重配置的主要目标是给矩阵$\boldsymbol{A}$的非零项赋予非零数值，且赋权结果对应符号邻接矩阵$\boldsymbol{W}$。

1) 满足特征值约束的决策变量

由于$\bar{\boldsymbol{L}}[1]_{1\times n}=\boldsymbol{0}$和$(\bar{\boldsymbol{L}}\otimes\boldsymbol{I}_d)\boldsymbol{q}=\boldsymbol{0}$显然成立，分布式仿射编队协议中$\boldsymbol{L}$和$\boldsymbol{D}$的非零项赋值问题，可直接被处理为寻找满足$\boldsymbol{L}[1]_{1\times n}=\boldsymbol{0}$和$(\boldsymbol{L}\otimes\boldsymbol{I}_d)\boldsymbol{q}=\boldsymbol{0}$的矩阵$\boldsymbol{L}$使其具有$d+1$个零根且其余特征根均有正实部。根据定理1可知，等式方程组$\boldsymbol{L}[1]_{1\times n}=\boldsymbol{0}_n$和$(\boldsymbol{L}\otimes\boldsymbol{I}_d)\boldsymbol{q}=\boldsymbol{0}$存在多组解，并且利用消元法可确定该方程组对应的约束变量和自由变量之间的关系。用$\boldsymbol{Y}_{\mathrm{W}}=[Y_i^{\mathrm{W}}]\in\mathbf{R}^{n_{\mathrm{W}}\times 1}$表示由等式方程组$\boldsymbol{L}[1]_{1\times n}=\boldsymbol{0}$和$(\boldsymbol{L}\otimes\boldsymbol{I}_d)\boldsymbol{q}=\boldsymbol{0}$得到的所有自由变量组成的向量，其中$n_{\mathrm{W}}$为所有自由变量个数。因此，由$\boldsymbol{Y}_{\mathrm{W}}$决定的拉普拉斯矩阵$\boldsymbol{L}$需要满足的特征值约束函数可定义为

$$G_1=\sum_i\max\{0,-\lambda_i(\boldsymbol{L}_{\mathrm{ff}})\}>0 \tag{12}$$

2) 非零权重配置

$H_3=0$约束成立，意味着$\boldsymbol{Y}_{\mathrm{W}}$的所有项可取非零数值。但是，由$\boldsymbol{Y}_{\mathrm{W}}$决定的$\boldsymbol{L}$中处于非决策变量的权值仍有可能被赋0。为完全避免零权重配置的现象，定义相应的约束函数H_4为

$$H_4=100o_3=0 \tag{13}$$

式中：o_3表示由决策变量$\boldsymbol{Y}_{\mathrm{W}}$生成拉普拉斯矩阵$\boldsymbol{L}$中零元素的个数。

最终，针对仿射编队有向交互拓扑的权重配置问题，可建立约束模型：

$$\begin{cases}G_1>0\\ H_4=0\end{cases} \tag{14}$$

3 有向k根图的检测方法

3.1 检测方法

根据定义1和定义2可知，有向k根图的检测问题可被转化为求解领导智能体集合$\mathcal{V}_l$到每个跟

随智能体节点$i\in\mathcal{V}_f$的点不相交有向路径的最小数量。据此可设计有向k根图的检测方法, 具体步骤:

步骤 1　向有向图$\mathcal{G}$增加一个标号为 0 的节点, 且对每一个顶点$i\in\mathcal{V}_l$建立一条有向边$(0,i)$, 记当前有向图为$\tilde{\mathcal{G}}=(\tilde{\mathcal{V}}=\mathcal{V}_l\cup\mathcal{V}_f\cup\{0\},\tilde{\mathcal{E}},\tilde{\boldsymbol{A}})$。

步骤 2　利用顶点撕裂程序由有向图$\tilde{\mathcal{G}}$构造有向图$\tilde{\mathcal{G}}_{\rm ST}=(\tilde{\mathcal{V}}_{\rm ST},\tilde{\mathcal{E}}_{\rm ST},\tilde{\boldsymbol{A}}_{\rm ST})$。顶点撕裂程序为: 对于每个$i\in\tilde{\mathcal{V}}$, 有向图$\tilde{\mathcal{G}}_{\rm ST}$包含顶点$i_s$、$i_t$以及有向边$(i_t,i_s)\in\tilde{\mathcal{E}}_{\rm ST}$, 且对于每条有向边$(i,j)\in\tilde{\mathcal{E}}$, 必存在有向边$(i_s,j_t)\in\tilde{\mathcal{E}}_{\rm ST}$。对于每个$i\in\mathcal{V}_{\rm f}$有$i_s,i_t\in\tilde{\mathcal{V}}_{\rm STf}$, 且对于每个$i\in\mathcal{V}_{\rm l}$有$i_s,i_t\in\tilde{\mathcal{V}}_{\rm STl}$, 则$\tilde{\mathcal{V}}_{\rm ST}=\{0_s,0_t\}\cup\tilde{\mathcal{V}}_{\rm STf}\cup\tilde{\mathcal{V}}_{\rm STl}$。

步骤 3　对于有向图$\tilde{\mathcal{G}}_{\rm ST}$, 令有向边$(i_t,i_s)\in\tilde{\mathcal{E}}_{\rm ST}$的容量为 1, 有向边$(i_s,j_t)\in\tilde{\mathcal{E}}_{\rm ST}$的容量为$\infty$, 可构造网络$\tilde{\mathcal{N}}_{\rm ST}$。

步骤 4　利用任意最大流算法计算节点 0(视为源节点)到每个节点$i_t\in\tilde{\mathcal{V}}_{\rm STf}$(视为汇点)的最大流$l_i$。计算$k=\min_i\{l_i\}$。

当$k=1$或者$k=2$时, 有向图$\mathcal{G}$含有一个有向生成树或者含有 2 根图。此时, 有向k根图的检测方法可应用于一般的一致性控制和基于复数拉普拉斯矩阵的相似编队控制算法的拓扑设计中。

3.2　检测方法的正确性

定理 4: 利用 3.1 节的检测方法对有向交互拓扑$\mathcal{G}$进行检测, 可知$\mathcal{G}$是k根的。

证明: 已知有向拓扑$\mathcal{G}$是k根的充要条件是领导智能体集合$\mathcal{V}_l$到每个跟随智能体节点$i\in\mathcal{V}_{\rm f}$有k条点不相交有向路径。向有向图$\mathcal{G}$加入节点 0 且节点 0 到每个领导智能体节点$i\in\mathcal{V}_{\rm l}$都有一条直达的有向边, 则$\mathcal{V}_{\rm l}$到$i\in\mathcal{V}_{\rm f}$的点不相交路径最大数量k的计算问题等价于求解图$\tilde{\mathcal{G}}$中节点 0 到$i\in\mathcal{V}_{\rm f}$的点不相交路径最大数量计算问题。

邦詹森等[21]在 Menger 定理 (2) 的证明过程中指出, 图$\tilde{\mathcal{G}}$中点不相交的$(0,i)$路($i\in\mathcal{V}_{\rm f}$)的最大数目等于应用顶点撕裂程序构造的有向图$\tilde{\mathcal{G}}$中对应$(0_s,i_t)$路边不相交的最大数目l_i。同时, 根据定理 2 可知, $(0_s,i_t)$路边不相交的最大数目等于以 0 为源点、以$i_t\in\tilde{\mathcal{V}}_{\rm STf}$为汇点的网络最小割集的基数。检测方法中网络$\tilde{\mathcal{N}}_{\rm ST}$的容量设定导致网络$\tilde{\mathcal{N}}_{\rm ST}$中两个不相交顶点之间的有向路所能允许通过的最大流量为 1。因此, $(0_s,i_t)$路边不相交的最大数目等于以 0 为源点和$i_t\in\tilde{\mathcal{V}}_{\rm STf}$为汇点的网络的最大流。

根据上述推导可知, 对所有节点$i_t\in\tilde{\mathcal{V}}_{\rm STf}$均计算对应的$l_i$, 其最小值$k=\min_i\{l_i\}$表明对于顶点子集$\mathcal{V}_{\rm l}$中的每个节点$i\in\mathcal{V}_{\rm f}$都是$k$可达的, 即有向交互拓扑$\mathcal{G}$是$k$根的, 且其根集为$\mathcal{V}_{\rm l}$。

证毕。

3.1 节中检测方法的时间复杂度是由所采取的最大流算法时间复杂度决定的。当最大流算法采用 Ford-Fulkerson 方法的深度优先算法实现时, 有向k根图的检测方法对应实现的时间复杂度为$O(|\mathcal{V}||\mathcal{E}|C_{\max})$, 其中$C_{\max}$为算法找出的最大流。

4　改进 NSGA-II 的拓扑结构构造优化算法

在仿射编队机动控制的拓扑优化设计中, 拓扑结构构造问题是一个多目标约束优化问题, 而权重配置是一个约束满足问题, 其可以被视为多目标约束优化问题的一个特例。由于标称队形的多样性, 这两个优化问题在某些给定条件下是多模态的, 即目标空间的一个最优值可能对应于决策空间中多个数值不同的变量。

为了得到一个真实的帕累托最优解集, 传统多目标优化算法侧重于提高目标空间的多样性、延展性和收敛性; 而多模态多目标优化算法利用解集在决策空间的分布特性, 在传统多目标优化算法的基础上设计特殊的搜索机制和环境选择策略, 进而提高算法寻找多个帕累托最优解集的能力[23]。目前, 多模态多目标优化算法主要以连续决策变量为对象, 通过引入特殊拥挤度距离、决策空间小生境或聚类等机制解决多目标优化的多模态特性。因此, 具有连续决策变量的权重配置问题模型 (14), 可利用一般的多模态多目标优化算法求解; 而拓扑结构构造优化模型 (11) 的决策变量是 0-1 矩阵, 难以直接利用已有的算法进行高效求解。

注意到, 遗传算法是处理二进制编码优化问题的主流算法; NSGA-II 则是由遗传算法衍生的典型多目标优化算法, 具有良好的收敛能力。因此, 本节将针对拓扑结构构造优化模型 (11), 以 NSGA-II[24] 为基本框架, 引入常见的多模态处理机制并进行改进, 提出改进 NSGA-II 的拓扑结构构造优化算法, 其具体流程环节如下。

1) 编码和种群初始化

记当前种群规模为N。算法直接采用二维二进制编码。由于拓扑结构构造问题中有$d+1$根图的约束, 决策矩阵$\boldsymbol{Y}_A$的每一行至少$d+1$个非零元素。算法随机生成个体基因对应二进制编码的每行非零元素数量不少于$d+1$的初始种群。

2) 基于约束支配原则的快速非支配排序

由于 NSGA-II 的快速非支配排序并没有考虑约束条件的存在, 故算法采取 Deb 等[25]提出的基于约束支配原则的快速非支配排序, 其主要思想是首先将种群根据约束违反度进行排序, 然后将同一约

束违反度的个体进行快速非支配排序。其中，个体基因当前的约束违反度计算公式为$V_i^{\mathrm{CV}}=\sum_j|H_j|$。

3）特殊拥挤度计算

Deng等[26]提出的特殊拥挤度距离计算是处理连续变量多模态多目标优化算法的有效机制，可以兼顾决策空间和目标空间的分布情况。对当前种群同一约束支配层级中个体i的特殊拥挤度距离V_i^{SCD}计算公式为

$$V_i^{\mathrm{SCD}}=\begin{cases}\max\{V_i^{\mathrm{CDY}},V_i^{\mathrm{CDF}}\} \\ \quad V_i^{\mathrm{CDY}}>V_i^{\mathrm{CDAY}}\ \text{or}\ V_i^{\mathrm{CDF}}>V_i^{\mathrm{CDAF}} \\ \min\{V_i^{\mathrm{CDY}},V_i^{\mathrm{CDF}}\}\quad \text{otherwise}\end{cases} \tag{15}$$

式中：V_i^{CDY}和V_i^{CDF}分别为当前支配层级中个体i的决策空间和目标空间的拥挤度距离；V_i^{CDAY}和V_i^{CDAF}分别为当前支配层级所有个体决策空间和目标空间的平均拥挤度距离。

为进一步提升二维二进制编码的种群多样性，对Deng等[26]工作中所采用的V_i^{CDY}和V_i^{CDF}进行改进设计。设计当前支配层级中个体i目标空间的拥挤度距离V_i^{CDF}的计算公式为

$$V_i^{\mathrm{CDF}}=\sum_{j=1}^{M}\frac{1}{n_j^{i,f}}\frac{f_j^{i_+}-f_j^{i_-}}{f_j^{\max}-f_j^{\min}} \tag{16}$$

式中：$i_+=\arg\min_k\{f_j^k|f_j^k>f_j^i\}$；$i_-=\arg\max_k\ \{f_j^k|f_j^k<f_j^i\}$；$f_j^{i_+}$为个体$i_+$的第$j$个目标函数取值，$f_j^{i_-}$为个体$i_-$的第$j$个目标函数取值；$f_j^{\max}$和$f_j^{\min}$分别为所有当前支配层个体第$j$个目标函数的最大值和最小值；$n_j^{i,f}$为当前支配层级中与个体$i$的目标$j$分量数值相同的个体数量，其可被称为目标变量的重合因子。特别地，当任意目标j计算i_+或者i_-为空集时，取$f_j^{i_+}-f_j^{i_-}=f_j^{\min}-f_j^{\max}$。不同于Deng等[26]的工作，重合因子$n_j^{i,f}$引入的目的是降低目标函数取值重合的个体被执行操作的优先度，提升当前支配层级目标空间的多样性。同理，可设计当前支配层级中个体i决策空间的拥挤距离V_i^{CDY}计算公式为

$$V_i^{\mathrm{CDY}}=\frac{1}{n^{i,y}}\frac{y^{i_+}-y^{i_-}}{y^{\max}-y^{\min}} \tag{17}$$

式中：$i_+=\arg\min_k\{y^k|y^k>y^i\}$；$i_-=\arg\min_k\{y^k|\ y^k<y^i\}$；$y^{i_+}$和$y^{i_-}$分别为个体$i_+$和个体$i_-$的决策变量距离；$y^{\max}$和$y^{\min}$分别为所有当前支配层个体决策距离的最大值和最小值；$n^{i,y}$为当前支配层级中与个体i的决策变量距离相同的个体数量，其可称为决策变量的重合因子。特别地，当i_+或者i_-为空集时，取$y^{i_+}-y^{i_-}=y^{\max}-y^{\min}$。二进制编码个体$i$的决策变量距离$y^i$定义为

$$y^i=\frac{1}{n_{\min}^{i,y}}\min_{j,i\neq j}\{y_j^i\} \tag{18}$$

式中：y_j^i为个体i和当前支配层中其他个体j的汉明距离；$n_{\min}^{i,y}$为当前支配层级中与i的汉明距离最近的个体数量，称为汉明距离变量的重合因子。y^i是一种改进的汉明距离。

4）二元锦标赛选择

算法直接采用Deng等[26]提出的改进二元锦标赛选择策略，对于随机两个个体的支配层级和特殊拥挤度先后作比较进行选择操作。

5）自适应交叉和变异

交叉操作和变异操作分别选取均匀交叉和均匀变异。均匀交叉中，每对基因座进行相互交换的概率为0.5。为避免算法陷入早熟，提高种群的多样性，设置自适应的交叉概率p_{c}和变异概率p_{m}分别为

$$p_{\mathrm{c}}=p_{\mathrm{c0}}+\frac{(p_{\mathrm{c1}}-p_{\mathrm{c0}})(|R_i-R_j|+1)}{\max(R_i)+\arccos(|V_i^{\mathrm{SCD}}-V_j^{\mathrm{SCD}}|)} \tag{19}$$

$$p_{\mathrm{m}}=p_{\mathrm{m0}}+\frac{(p_{\mathrm{m1}}-p_{\mathrm{m0}})R_i}{\max(R_i)+\arcsin(V_i^{\mathrm{SCD}})} \tag{20}$$

式中：p_{c0}和p_{c1}为交叉概率所允许的最小值和最大值边界；p_{m0}和p_{m1}为变异概率所允许的最小值和最大值边界；R_i为个体i的支配层级。当前种群中，个体i与个体j的支配层级、拥挤度相差越大，这2个个体进行交叉操作的概率越大；个体i的支配层级越靠后、特殊拥挤度距离越小，则个体i发生变异的概率越大。

6）精英保留策略和算法终止条件

算法直接采取NSGA-II的精英保留策略产生下次迭代过程中的父代种群。定义算法最大迭代次数N_{GEN}。如果当前累计迭代次数大于N_{GEN}，则算法终止。

5 仿真与分析

5.1 仿真场景及算法参数设置

在二维空间，考虑由7个智能体组成的多智能体系统，其中跟随智能体编号为1～4，领导智能体编号为5～7。给定编队系统的标称队形为$\boldsymbol{q}=[\boldsymbol{q}_1^{\mathrm{T}},\boldsymbol{q}_2^{\mathrm{T}},\boldsymbol{q}_3^{\mathrm{T}},\boldsymbol{q}_4^{\mathrm{T}},\boldsymbol{q}_5^{\mathrm{T}},\boldsymbol{q}_6^{\mathrm{T}},\boldsymbol{q}_7^{\mathrm{T}}]^{\mathrm{T}}$，其中，$\boldsymbol{q}_1=[0,2]^{\mathrm{T}}$，$\boldsymbol{q}_2=[0,-2]^{\mathrm{T}}$，$\boldsymbol{q}_3=[-2,2]^{\mathrm{T}}$，$\boldsymbol{q}_4=[-2,-2]^{\mathrm{T}}$，$\boldsymbol{q}_5=[4,0]^{\mathrm{T}}$，$\boldsymbol{q}_6=[2,2]^{\mathrm{T}}$，$\boldsymbol{q}_7=[2,-2]^{\mathrm{T}}$。算法参数设置为$N=50$，$N_{\mathrm{GEN}}=100$，$p_{\mathrm{c0}}=0.6$，$p_{\mathrm{c1}}=1$，$p_{\mathrm{m0}}=0.001$和$p_{\mathrm{m1}}=0.1$。

5.2 仿真结果

首先，应用改进NSGA-II的拓扑结构构造优化算法进行求解，拓扑结构构造结果如表1所示，

其对应的拓扑结构如图 1 所示。

从表 1 的结果可知, $\mathcal{G}_1$～$\mathcal{G}_7$的信息交互成本相对最低, $\mathcal{G}_1$的信息传播能耗相对最小, $\mathcal{G}_{19}$的同步能力相对较强, $\mathcal{G}_5$和$\mathcal{G}_{13}$对领导智能体的物理依赖程度相对最小。

另外, $\mathcal{G}_{25}$和$\mathcal{G}_{26}$对应的拓扑结构不相同, 但对应的目标向量数值相同, 这说明应用改进 NSGA-II 的拓扑结构构造优化算法可以对多模态多目标优化问题进行求解。

表 1　拓扑结构构造结果

Table 1　Topology structure construction results

拓扑编号	f_1	f_2	f_3	f_4
$\mathcal{G}_1$	12	43.55	6.00	1.25
$\mathcal{G}_2$	12	43.55	4.83	1.50
$\mathcal{G}_3$	12	44.21	4.00	1.25
$\mathcal{G}_4$	12	44.69	3.50	1.50
$\mathcal{G}_5$	12	44.93	4.00	1.00
$\mathcal{G}_6$	12	45.20	2.50	1.75
$\mathcal{G}_7$	12	46.15	2.00	2.00.
$\mathcal{G}_8$	13	48.73	3.33	1.50
$\mathcal{G}_9$	13	50.15	1.83	2.00
$\mathcal{G}_{10}$	13	50.81	1.67	2.00
$\mathcal{G}_{11}$	13	52.00	1.00	2.50
$\mathcal{G}_{12}$	13	52.00	1.50	2.25
$\mathcal{G}_{13}$	14	47.74	3.50	1.00
$\mathcal{G}_{14}$	14	53.63	2.33	1.75
$\mathcal{G}_{15}$	14	53.87	2.50	1.50
$\mathcal{G}_{16}$	14	54.81	1.33	2.25
$\mathcal{G}_{17}$	14	56.47	1.17	2.25
$\mathcal{G}_{18}$	14	58.32	0.50	2.75
$\mathcal{G}_{19}$	14	59.85	0.00	3.00
$\mathcal{G}_{20}$	15	53.40	3.00	1.25
$\mathcal{G}_{21}$	15	60.00	0.33	2.75
$\mathcal{G}_{22}$	15	60.47	0.83	2.50
$\mathcal{G}_{23}$	15	62.47	0.67	2.50
$\mathcal{G}_{24}$	16	59.05	2.00	1.50
$\mathcal{G}_{25}$	18	68.00	1.83	1.75
$\mathcal{G}_{26}$	18	68.00	1.83	1.75
$\mathcal{G}_{27}$	20	75.48	1.33	2.00
$\mathcal{G}_{28}$	20	77.14	1.00	2.00

针对众多的拓扑结构构造结果, 可根据拓扑设计需求选择合适的拓扑结构。以$\mathcal{G}_1$为例, 对权重配置问题进行求解, 可设计$\mathcal{G}_1$对应的拉普拉斯矩阵子块为

$$\boldsymbol{L}_{\mathrm{ff}}=\begin{bmatrix}1.6 & 0 & 0 & -0.8\\ -0.4 & 1.2 & 0 & 0\\ -0.4 & 0 & 0.4 & -0.2\\ 0.8 & -0.8 & -0.8 & 0.8\end{bmatrix}$$

$$\boldsymbol{L}_{\mathrm{fl}}=\begin{bmatrix}0 & -1.6 & 0.8\\ 0.8 & 0 & -1.6\\ 0 & 0 & 0.2\\ 0 & 0 & 0\end{bmatrix}$$

矩阵块$\boldsymbol{L}_{\mathrm{ff}}$的特征根为 0.071 9, 0.976 7, 1.475 7+0.858 6i 和 1.475 7−0.858 6i, 均在右开半平面。

以上述拓扑优化设计的结果$\boldsymbol{L}$为前提, 应用 Zhao[6] 所提出的二阶多智能体系统时变加速度领导智能体情况下的仿射编队机动控制律进行编队仿真, 运动轨迹和总编队误差范数变化分别如图 2 和图 3 所示。图 2 展示了当前的拓扑方案能够使得智能体的几何队形连续地实现由平移、旋转、伸缩和切变变换组合的复合变换。由图 3 可知, 当前多智能体系统的总编队误差范数会随着时间趋向于 0。因此, 本文所设计的拓扑优化算法能够为一般的仿射编队机动控制方案生成合适的拓扑, 可用作仿射编队机动控制协议的前置拓扑设计算法。

5.3　算法有效性分析

为进一步说明改进 NSGA-II 的拓扑结构构造优化算法的有效性, 将在其基础上不考虑算法改进特点的进化算法称为改进的多模态 NSGA-II (imNSGA-II), 即目标空间的拥挤度计算公式 (16) 中忽略重合因子的引入、决策空间的拥挤度计算公式与目标空间拥挤度计算公式形式一致且交叉概率和变异概率为常数设置。采用决策空间的反向世代距离V_{IGDX}和目标空间的反向世代距离V_{IGDF}[26-27] 分别对算法求解得到的 Pareto 最优解集(Pareto Set, PS)和 Pareto 前沿面(Pareto Front, PF)进行评估。V_{IGDX}和V_{IGDF}计算公式如下:

$$V_{\mathrm{IGDX}}(\mathcal{O},\mathcal{P}^*)=\frac{\sum_{v\in\mathcal{P}^*}\tilde{d}(v,\mathcal{O})}{|\mathcal{P}^*|}\tag{21}$$

$$V_{\mathrm{IGDF}}(\mathcal{O},\mathcal{P}^*)=\frac{\sum_{v\in\mathcal{P}^*}\tilde{d}(v,\mathcal{O})}{|\mathcal{P}^*|}\tag{22}$$

式中: $\mathcal{O}$为算法求解得到的 PS 或者 PF; $\mathcal{P}^*$为沿着真实 PS 均匀分布的点集; $\tilde{d}(v,\mathcal{O})$表示点v到$\mathcal{O}$中点的最小欧氏距离。本文$\mathcal{P}^*$直接选取暴力搜索求得的真实 PS。除此以外, 还用算法求解得到 PS 中个体数量N_{PS}以及具有多模态特征目标向量的个体数量N_{PSMF}衡量算法的优劣。

以 5.1 节的仿真场景为例, 改进 NSGA-II 的拓扑结构构造优化算法采用 5.1 节的参数设置, imNSGA-II 的交叉概率和变异概率分别为 0.9 和

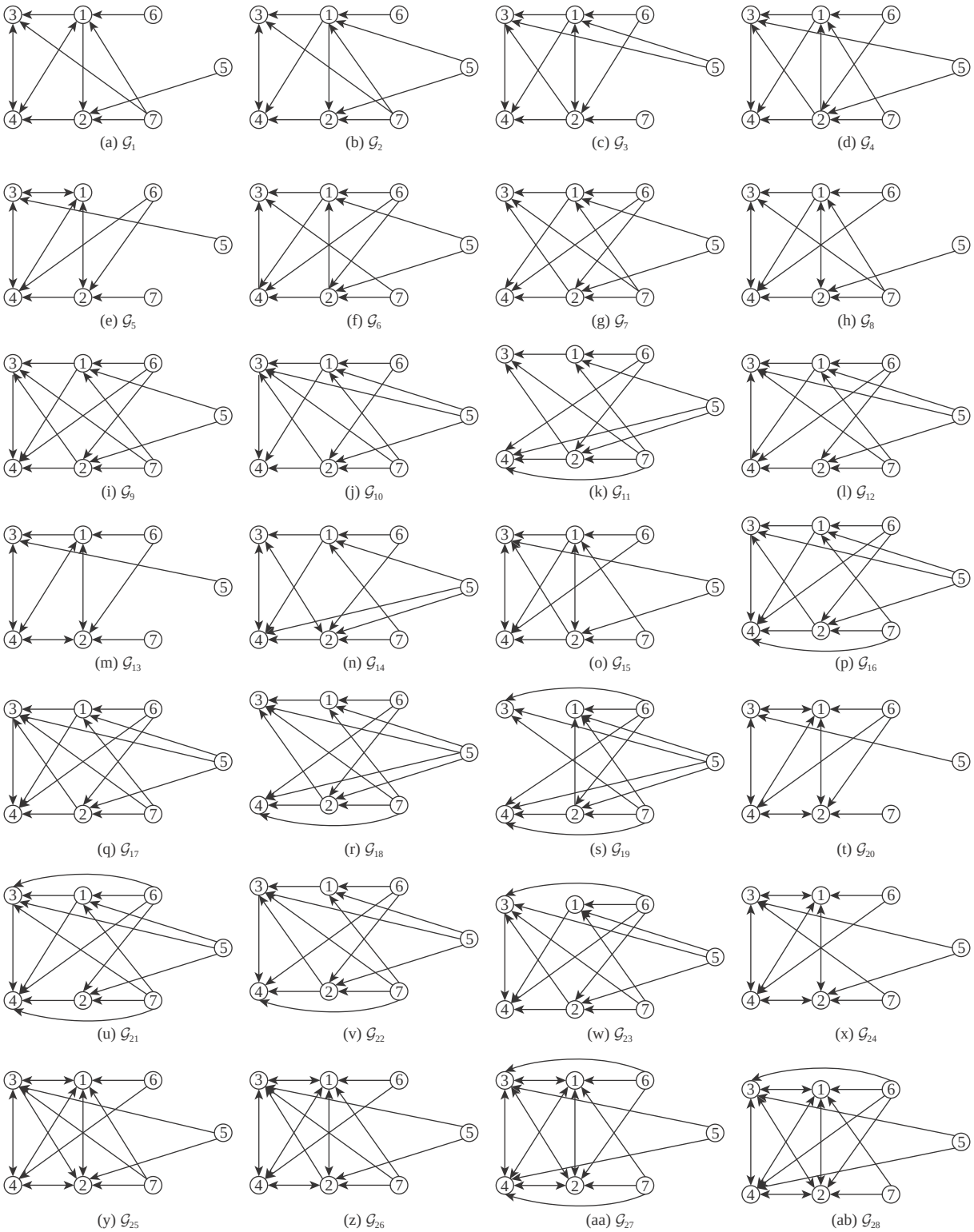


图 1　表 1 对应的拓扑结构

Fig. 1　Topology structure corresponding to Table 1

0.01。针对拓扑结构构造问题，将改进 NSGA-II 的拓扑结构构造优化算法和 imNSGA-II 实现独立运行 100 次，得到的算法评估指标结果如表 2 所示。利用暴力搜索求解拓扑结构构造问题的真实 PS 特征也在表 2 中进行展示。

由表 2 可知，相比于 imNSGA-II，本文算法的 N_{PS} 和 N_{PSMF} 均值结果更接近真实 PS 特征，且 IGDX 和 IGDF 的均值和方差更小。此外，imNSGA-II 单次执行时间为 23.59 s，本文提出的改进 NSGA-II 的拓扑结构构造优化算法单次执行时间为 23.74 s，暴力搜索单次执行时间至少需要 23 h，前 2 种算法执行时间相差不大且远远小于暴力搜索的执行时间。

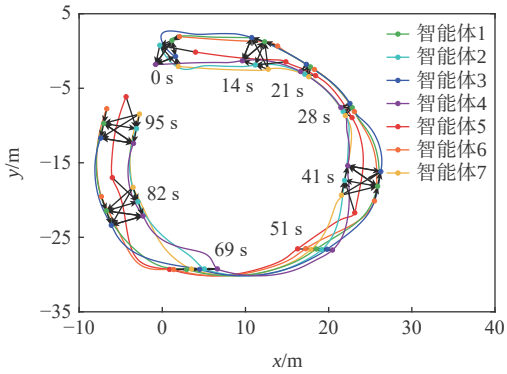


图 2　编队轨迹

Fig. 2　Formation trajectory

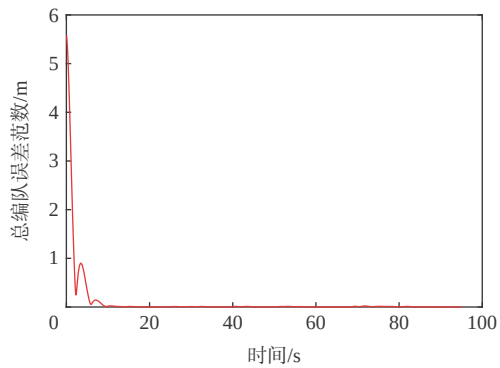


图 3　总编队误差范数

Fig. 3　Total formation error norm

表 2　算法评估指标结果

Table 2　Algorithm evaluation indicator results

算法	N_{PS} 平均值	N_{PSMF} 平均值	IGDX		IGDF		暴力搜索
			平均值	方差	平均值	方差	
imNSGA-II	2.15	0.05	3.185 7	4.974 2	3.780 9	423.723 7	37
本文	23.16	0.58	2.340 1	3.450 8	1.376 0	5.835 9	15

因此，上述算法对比分析说明改进 NSGA-II 的拓扑结构构造优化算法可以有效地解决二进制编码的多模态多目标优化问题，其对拥挤度计算和遗传操作的改进能够进一步改善多模态多目标优化算法的求解效果。

6　结　论

1）提出的有向k根图的约束检测方法，实现了有向信息交互拓扑的$d+1$根约束求解，并且在理论层面验证了该检测方法的可行性，能够直接应用于有向生成树和有向k根图的检测判断。

2）通过多目标优化算法的性能指标，对比评估了改进 NSGA-II 的拓扑结构构造优化算法的有效性和优越性；并且以 7 个智能体的编队为例，给出了拓扑结构构造问题的 28 组解，进行了拓扑权重配置，生成了满足有向$d+1$根图约束的仿射编队机动交互拓扑。

本文仅根据拓扑要求设计了仿射编队机动控制的有向交互拓扑优化方案，未以具体模型为对象考虑控制性能层面的优化目标，这一点将在以后的研究中进一步考虑。

参考文献（References）

[1] DONG X W, HU G Q. Time-varying formation control for general linear multi-agent systems with switching directed topologies[J]. Automatica, 2016, 73: 47-55.

[2] LIN Z Y, DING W, YAN G F, et al. Leader–follower formation *via* complex Laplacian[J]. Automatica, 2013, 49(6): 1900-1906.

[3] LIN Z Y, WANG L L, CHEN Z Y, et al. Necessary and sufficient graphical conditions for affine formation control[J]. IEEE Transactions on Automatic Control, 2015, 61(10): 2877-2891.

[4] HAN Z M, WANG L L, LIN Z Y, et al. Formation control with size scaling *via* a complex Laplacian-based approach[J]. IEEE Transactions on Cybernetics, 2016, 46(10): 2348-2359.

[5] FANG X, LI X L, XIE L H. Distributed formation maneuver control of multiagent systems over directed graphs[J]. IEEE Transactions on Cybernetics, 2022, 52(8): 8201-8212.

[6] ZHAO S Y. Affine formation maneuver control of multiagent systems[J]. IEEE Transactions on Automatic Control, 2018, 63(12): 4140-4155.

[7] XU Y, ZHAO S Y, LUO D L, et al. Affine formation maneuver control of high-order multi-agent systems over directed networks[J]. Automatica, 2020, 118: 109004.

[8] XU Y, LUO D L, YOU Y C, et al. Affine transformation based formation maneuvering for discrete-time directed networked systems[J]. Science China Technological Sciences, 2020, 63(1): 73-85.

[9] CHEN L M, MEI J, LI C J, et al. Distributed leader–follower affine formation maneuver control for high-order multiagent systems[J]. IEEE Transactions on Automatic Control, 2020, 65(11): 4941-4948.

[10] XU Y, LUO D L, LI D Y, et al. Affine formation control for heterogeneous multi-agent systems with directed interaction networks[J]. Neurocomputing, 2019, 330: 104-115.

[11] XU Y, LUO D L, LI D Y, et al. Target-enclosing affine formation control of two-layer networked spacecraft with collision avoidance[J]. Chinese Journal of Aeronautics, 2019, 32(12): 2679-2693.

[12] XU Y, LI D Y, LUO D L, et al. Affine formation maneuver tracking control of multiple second-order agents with time-varying delays[J]. Science China Technological Sciences, 2019, 62(4): 665-676.

[13] LUO Z X, ZHANG P Y, DING X J, et al. Adaptive affine formation maneuver control of second-order multi-agent systems with disturbances[C]// 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV). Piscataway: IEEE Press, 2020: 1071-1076.

[14] CHANG Z Z, WANG J J, LI Z K. Fully distributed event-triggered affine formation maneuver control over directed graphs[J]. IFAC-PapersOnLine, 2022, 55(3): 178-183.

[15] YANG Q K, CAO M, FANG H, et al. Constructing universally rigid tensegrity frameworks with application in multiagent formation control[J]. IEEE Transactions on Automatic Control, 2018, 64(1):

381-388.

[16] XIAO F, YANG Q K, ZHAO X Y, et al. A framework for optimized topology design and leader selection in affine formation control[J]. IEEE Robotics and Automation Letters, 2022, 7(4): 8627-8634.

[17] YANG J Y, XIAO F, CHEN T W. Formation tracking of nonholonomic systems on the special euclidean group under fixed and switching topologies: an affine formation strategy[J]. SIAM Journal on Control and Optimization, 2021, 59(4): 2850-2874.

[18] MONDAL S, TSOURDOS A. Optimal topology for consensus using genetic algorithm[J]. Neurocomputing, 2020, 404: 41-49.

[19] 崔亚妮, 任佳, 杜文才, 等. 多无人船通信网络拓扑优化控制算法[J]. 控制理论与应用, 2016, 33(12): 1639-1649.
CUI Y N, REN J, DU W C, et al. Network topology optimization control algorithm for multiple unmanned surface vehicle[J]. Control Theory & Applications, 2016, 33(12): 1639-1649(in Chinese).

[20] 谷晓燕, 陈亮, 邓香平. 无人机编队信息交互拓扑多目标优化[J]. 电光与控制, 2022, 29(9): 27-31,52.
GU X Y, CHEN L, DENG X P. Multi-objective optimization of UAV formation information interaction topology[J]. Electronics Optics & Control, 2022, 29(9): 27-31,52(in Chinese).

[21] 邦詹森 J, 古廷 G Z. 有向图的理论、算法及其应用[M]. 姚兵, 张忠辅, 译. 北京: 科学出版社, 2009.
BANG-JENSEN J, GUTIN G Z. Digraphs theory, algorithms and applications[M]. YAO B, ZHANG Z F translated. Beijing: Science Press, 2009(in Chinese).

[22] HONG H, KIM B J, CHOI M Y, et al. Factors that predict better synchronizability on complex networks[J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2004, 69(6 Pt 2): 067105.

[23] 岳彩通, 梁静, 瞿博阳, 等. 多模态多目标优化综述[J]. 控制与决策, 2021, 36(11): 2577-2588.
YUE C T, LIANG J, QU B Y, et al. A survey on multimodal multiobjective optimization[J]. Control and Decision, 2021, 36(11): 2577-2588(in Chinese).

[24] DEB K, PRATAP A, AGARWAL S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182-197.

[25] DEB K, JAIN H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints[J]. IEEE Transactions on Evolutionary Computation, 2013, 18(4): 577-601.

[26] DENG W, ZHANG X X, ZHOU Y Q, et al. An enhanced fast nondominated solution sorting genetic algorithm for multi-objective problems[J]. Information Sciences, 2022, 585: 441-453.

[27] ZHOU A M, ZHANG Q F, JIN Y C. Approximating the set of Pareto-optimal solutions in both the decision and objective spaces by an estimation of distribution algorithm[J]. IEEE Transactions on Evolutionary Computation, 2009, 13(5): 1167-1189.

Directed interactive topology optimization design for multi-agent affine formation maneuver control

MA Suhui[1], ZHANG Dong[1, 2, *], WANG Mengyang[1], WANG Tinghui[1], LIU Songdan[3]

(1. School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China;
2. Shaanxi Key Laboratory of Aerospace Flight Vehicle Technology, Xi'an 710072, China;
3. China Academy of Aerospace Science and Innovation, Beijing 100083, China)

Abstract: This paper investigated the directed interactive topology optimization design problem for multi-agent affine formation maneuver control. Firstly, by considering the optimization indexes such as information interaction cost and energy consumption during information spreading, a directed topology optimization model for affine formation maneuver was established, including topology structure construction and weight allocation. Secondly, in view of the topological structure construction for affine formation maneuver, a directed k-rooted graph detection method was proposed, which could realize the solution of $d+1$-rooted constraint for directed information interaction topology. Then, an improved NSGA-II-based topology structure construction optimization algorithm was designed. Finally, a formation of seven agents in two-dimensional space was taken as an example for simulation verification. The results show that the improved NSGA-II-based topology structure construction optimization algorithm has better optimization effects. It can effectively provide a variety of feasible directed interactive topologies for affine formation maneuver control, and the generated interactive topology can meet the requirements of a directed $d+1$-rooted graph.

Keywords: multi-agent system; affine transformation; formation control; interactive topology design; directed graph; multi-objective optimization

Received: 2023-04-14; **Accepted**: 2023-08-10; **Published Online**: 2023-09-05 11:17

URL: link.cnki.net/urlid/11.2625.V.20230905.1002.003

Foundation item: National Natural Science Foundation of China (61933010)

* **Corresponding author**. E-mail: zhangdong@nwpu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0199

基于异常点重构的调频引信抗扫频干扰方法

段乐帆，郝新红*，陈齐乐
(北京理工大学 机电动态控制重点实验室，北京 100081)

摘　　要：针对调频引信对抗扫频式干扰的问题，提出一种基于异常点重构的调频引信抗扫频干扰方法。该方法基于迭代滤波分解(IFD)算法，将引信中频信号分解为本征模态函数(IMF)分量；扫频干扰在IMF分量中表现为异常点，对IMF分量进行异常点检测与消除，并利用长短时记忆(LSTM)网络构建数据重构网络模型，对异常点部分进行数据重构，可以实现对扫频干扰的抑制。通过仿真验证了所提方法的有效性，结果表明该方法提高了调频引信对抗扫频式干扰的能力。

关 键 词：调频引信；扫频式干扰；抗干扰方法；异常点重构；迭代滤波分解；长短时记忆网络
中图分类号：V221+.3; TB553
文献标志码：A　　**文章编号**：1001-5965(2025)04-1377-08

调频引信通过发射频率随时间变化的调频连续波信号，检测无线电波在引信与目标往返传播过程中的信号频率变化来测定引信和目标之间的距离，这种调频定距方法定距精度较高，抗干扰性能好，被广泛应用于常规弹药[1-2]。现代战场电磁环境复杂，在各种电子干扰手段中，扫频式干扰是针对调频体制引信威胁较大的一种干扰样式[3-4]。扫频式干扰信号能够进入引信接收机工作带宽内，干扰信号能量较大，很容易使引信输出信号达到判决门限，进而导致引信误启动。

目前已有一些针对调频引信抗干扰技术的研究成果。文献[5-6]以分数阶傅里叶变换为核心算法来提高调频引信的定距精度。文献[7]提出一种基于弹目交会过程中的谐波时序特征的定距方法，提高调频引信抗干扰性能。文献[8]通过变调制率和双通道相关检测技术以缓解干扰，实现调频引信的精确定距。文献[9]针对扫频式干扰提出了一种基于谐波系数幅值平均的引信抗干扰方法。上述方法从引信系统波形设计或波形特征识别两个方面提高了引信应对扫频式干扰的抗干扰性能，但未能针对干扰信号本身进行抑制处理，高能量的扫频式干扰覆盖目标回波时仍将导致引信“早炸”或“瞎火”。对于受扫频式干扰影响的调频引信中频信号，难以直接进行干扰抑制，通过迭代滤波分解算法将中频信号分解为不同频率成分的分量，有利于检测并抑制扫频式干扰。

本文从引信接收信号处理方面设计抗扫频式干扰方法，针对扫频干扰信号本身进行抑制，提出一种基于异常点重构的调频引信抗扫频干扰方法。该方法采用迭代滤波分解算法将引信中频信号分解为本征模态函数分量，扫频干扰在各个本征模态函数分量中表现为异常点；对分量进行异常点检测与消除，并构建基于长短时记忆网络的数据重构网络模型，对异常点部分进行数据重构，实现对扫频式干扰的抑制。最后通过仿真试验验证了该方法的有效性。

收稿日期：2023-04-21；**录用日期**：2023-06-16；**网络出版时间**：2023-07-11 15:05
网络出版地址：link.cnki.net/urlid/11.2625.V.20230711.0954.002
基金项目：国家自然科学基金(61973037)
*通信作者. E-mail：haoxinhong@bit.edu.cn

引用格式：*段乐帆，郝新红，陈齐乐. 基于异常点重构的调频引信抗扫频干扰方法[J]. 北京航空航天大学学报，2025，51(4)：1377-1384. DUAN L F，HAO X H，CHEN Q L. Anti-sweep jamming method for FM fuze based on outlier reconstruction[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1377-1384 (in Chinese).*

1 调频引信与扫频式干扰

1.1 调频引信谐波定距原理

设调频引信发射信号为 $S_t(t)$，其傅里叶级数形式的表达式为[10]

$$S_t(t)=A\cos(2\pi f_c t)\sum_{m=-\infty}^{\infty}a_m e^{j2\pi m f_M t} \tag{1}$$

式中：A 为发射信号的幅度；f_c 为发射信号的载频；f_M 为调制频率；a_m 为傅里叶系数。

调频引信发射信号的频率受周期波调制，目标回波信号的时频特性图形状与发射信号的一致，但相对于发射信号产生一定的时移 τ。

以三角波线性调频引信为研究对象，分析其时频特性可得一个调制周期内的差频信号由4个部分组成，分别为2个调频信号和2个单频信号。在对差频信号表达式做傅里叶变换后，可得一个调制周期内的差频信号频谱表达式[8]

$$F_b(f,\tau)=A_s\sum_m[(k(m,\tau)+a(m,\tau))\times \delta(2\pi f-2\pi m f_M\pm 2\pi f_d)] \tag{2}$$

式中：A_s 为差频信号的幅度；$k(m,\tau)$ 为两个单频信号的傅里叶系数；$a(m,\tau)$ 为两个调频信号的傅里叶系数；m 为谐波次数；f_d 为多普勒频率。

实际应用场景中，τ 是一个非常小的数值，针对引信差频频谱的研究主要考虑两个单频信号成分即可。由单频信号的傅里叶系数 $k(m,\tau)$，可以推得 m 次谐波达到峰值时，谐波次数 m 与弹目距离 R 的关系式

$$m=\frac{8\Delta F_M R}{c} \tag{3}$$

式中：ΔF_M 为单边调频带宽。由式(3)可得 m 次谐波峰值出现时所对应的定距距离 R。

1.2 扫频式干扰信号

扫频式干扰信号的带宽覆盖引信接收机工作带宽，且在扫频带宽内按照一定的频率步进和驻留时间进行频率扫描。设扫频式干扰信号的起始频率为 f_j，扫频带宽均分为 K 个频段，每个频段带宽为 B、扫频点数为 P，单一频点驻留时间为 T_P，则干扰信号 $S_J(t)$ 可表示为各个频段的扫频干扰的累加和，表达式为

$$S_J(t)=\sum_{k=0}^{K-1}\sum_{p=0}^{P-1}\sum_{u=-\infty}^{\infty}m(t_J)\cos\left(2\pi\left(f_j+\frac{(kP+p)B}{P}\right)t_J\right) \tag{4}$$

式中：

$$t_J=t-kPT_P-pT_P-uPT_P,\ m(t)=\begin{cases}1, 0<t<T_P\\ 0, \text{else}\end{cases}。$$

1.3 扫频式干扰作用下调频引信中频信号

扫频干扰进入引信接收机后，与发射信号进行混频，并通过低通滤波器滤除高频成分后，得到扫频干扰作用下引信中频信号输出[10]

$$S_{IF}(t)=\frac{1}{2}AA_j\cos(2\pi(f_{jn}-f_c)t)\times g_n(t)\sum_M a_m e^{j2\pi m f_M t} \tag{5}$$

式中：A_j 为扫频干扰的载波幅值；f_{jn} 为第 n 个扫频驻留频点；M 为低通滤波器通带内的谐波数；$g_n(t)$ 为门函数：$g_n(t)=\begin{cases}1, nT_p<t\leqslant(n+1)T_P\\ 0, \text{else}\end{cases}$。

由上述表达式可知，如果扫频式干扰的参数设置合理，干扰信号可以同时进入引信的各次谐波通道中，干扰信号能量达到判决门限时，便会干扰引信，导致引信误启动。

2 基于异常点重构的抗干扰方法

2.1 抗干扰总体方案

由前所述，扫频式干扰可同时进入引信的各次谐波通道中，能量相对较大的干扰信号表现为谐波通道中信号的数据异常点。本文设计了基于异常点重构的抗扫频干扰方法，针对谐波通道中受扫频干扰影响产生的数据异常点进行处理。该方法采用迭代滤波分解算法对引信中频信号进行分解，基于长短时记忆网络进行数据异常点的重构。以三角波线性调频引信为研究对象，该方法的总体方案如图1所示。

得到引信中频信号后，采用迭代滤波分解算法将中频信号分解为分属各次谐波通道的本征模态函数分量，再针对本征模态函数分量进行异常点检测和消除，最后基于长短时记忆网络构建数据重构网络模型，对异常点消除部分进行数据重构，最终还原得到抑制扫频干扰的中频信号。

2.2 迭代滤波分解算法原理

迭代滤波分解(IFD)算法是受经验模态分解(empirical mode decomposition, EMD)算法的启发，由Lin等[11-12]提出的一种用于分解和分析非线性非平稳信号的算法技术。IFD算法是EMD算法的补充和替代，两者目的都是将目标信号分解为简单的信号分量，称为本征模态函数(IMF)分量。IFD算法与EMD算法结构相似，但具有EMD算法所没有的严格数学理论基础，能够获取更准确的IMF分量[13]。此外，针对离散的数字信号数据，IFD算法允许通过快速傅里叶变换(fast Fourier transform, FFT)加速信号分解[14]，算法复杂度显著降低至与FFT相同的 $O(N\log N)$。

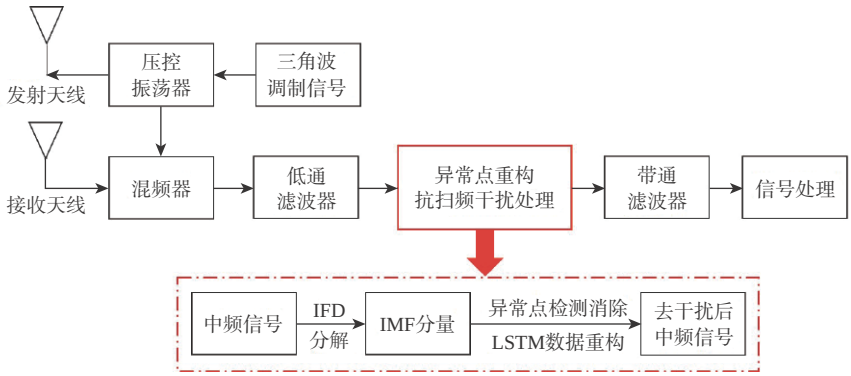


图 1　抗干扰总体方案

Fig. 1　General scheme of anti-jamming

IFD 算法包含两个过程, 分别为内循环和外循环。其中, 内循环是提取出单个 IMF 分量的过程, 外循环用来控制分解得到的 IMF 分量的数量。IFD 算法与 EMD 算法的区别在于, 在筛选 IMF 分量的过程中, IFD 算法使用滑动平均值而非局部平均值, 这通过当前待分解信号与一个滤波函数卷积而得到。

针对引信中频信号进行 IFD 算法处理, 目的就是将中频信号分解为一系列不同频率成分的 IMF 分量, 这些 IMF 分量就是中频信号分解到各次谐波通道内的信号分量。

对于当前待分解信号 $S(x)$, 计算其与滤波函数 $\omega(t)$ 的卷积, 可以得到滑动算子的表达式为

$$\Gamma(S(x))=\int_{-l_n}^{l_n} S(x+t)\omega(t)\mathrm{d}t \tag{6}$$

式中: $\omega(t)$ 为滤波函数; l_n 为滤波区间值, 通过如下表达式得到

$$l_n=2\left\lfloor\lambda\frac{N}{k}\right\rfloor \tag{7}$$

其中: $\lambda\in[1,3]$ 为标量参数; N 为信号采样点数; k 为信号极值点数; $\lfloor\cdot\rfloor$ 表示向下取整。

由当前待分解信号减去滑动算子, 可得到波动算子, 表达式为

$$\kappa(S(x))=S(x)-\Gamma(S(x)) \tag{8}$$

由式 (8) 计算得到波动算子后, 要进行一次内循环停止准则的判断, 若不满足停止准则, 则将该波动算子视为当前待分解信号, 重复式 (6) 和式 (8)。直至满足内循环停止准则, 此时的波动算子即为该次内循环提取出来的 IMF 分量。

理论上, IFD 的内循环可以一直迭代, 但实际应用中为了保证算法效率, 会设置一个内循环迭代的停止准则[14], 计算表达式为

$$\mathrm{SD}:=\frac{\mathrm{IMFs}(x)-\mathrm{IMFc}(x)^2}{\mathrm{IMFc}(x)^2} \tag{9}$$

式中: IMFs(x) 为当前需要进行停止准则判断的分量; IMFc(x) 为前一次提取的 IMF 分量。当计算得此时的 SD 等于或小于预设的停止迭代阈值时, 算法将结束此次内循环, 并将此时的 IMFs(x) 作为该次内循环所提取出来的 IMF 分量。SD 阈值设置得越小, 一次内循环需要进行的迭代次数越多。SD 阈值一般在 0.001～0.2 范围内进行设置[15]。

IFD 的具体算法流程如图 2 所示。

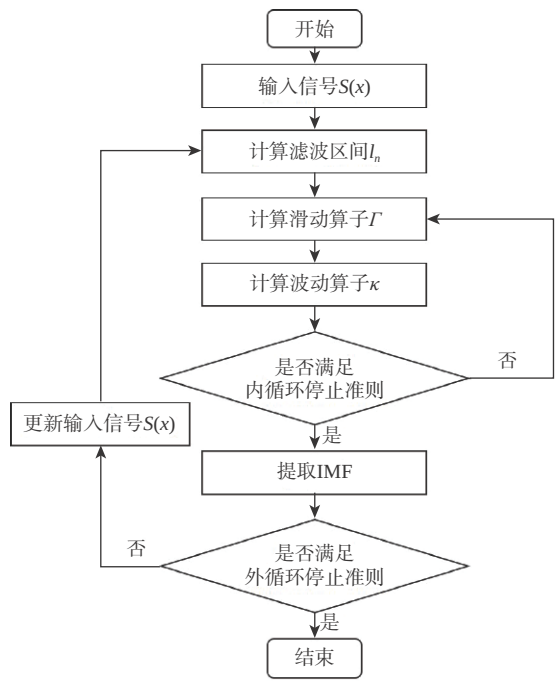


图 2　IFD 算法流程

Fig. 2　Flow chart of IFD algorithm

2.3　数据异常点处理

引信中频信号经过 IFD 算法处理后, 得到分属各次谐波通道的 IMF 分量, 根据引信工作要求, 针对性选择 IMF 分量进行数据异常点检测与消除, 将受扫频干扰影响而产生的异常点进行消除。

扫频式干扰的能量相对较大, 在中频信号经 IFD 算法分解得到的各次谐波通道信号成分的 IMF 分量中, 扫频干扰作用效果表现为 IMF 分量的某段数据值激增, 表明 IMF 分量受扫频干扰影响出

现了数据异常点。

因此，根据数据值激增这一特征，可对IMF分量进行数据异常点检测，对于符合特征的数据判定为异常点并进行消除。本文将同时满足以下两个条件的数据值定义为数据异常点：①与中位数相差超过三倍换算绝对中位差（median absolute deviation, MAD）；②增幅速率超过引信谐波信号正常增幅速率。其中采用滑动窗计算中位数与增幅速率，来进行数据异常点的动态检测。对于判定为异常点的数据将做置零处理。

2.4 长短时记忆网络基本原理

将IMF分量中因扫频干扰而产生的异常点消除后，需要对异常点消除部分进行数据重构，以还原信号数据。对于IMF分量的数据重构，针对其时间序列的属性，本文提出一种基于长短时记忆（LSTM）网络[16]的数据重构方法，通过LSTM网络强大的时间序列预测能力，对IMF分量异常点消除部分进行数据重构。

长短时记忆（LSTM）网络是基于循环神经网络（recurrent neural network, RNN）提出的一种特殊神经网络，LSTM网络具有信息保留和筛选信息的功能，即具有“记忆”和“遗忘”两个功能，这两个功能使得LSTM特别适合处理时间序列问题[17-18]。LSTM单元结构如图3所示。

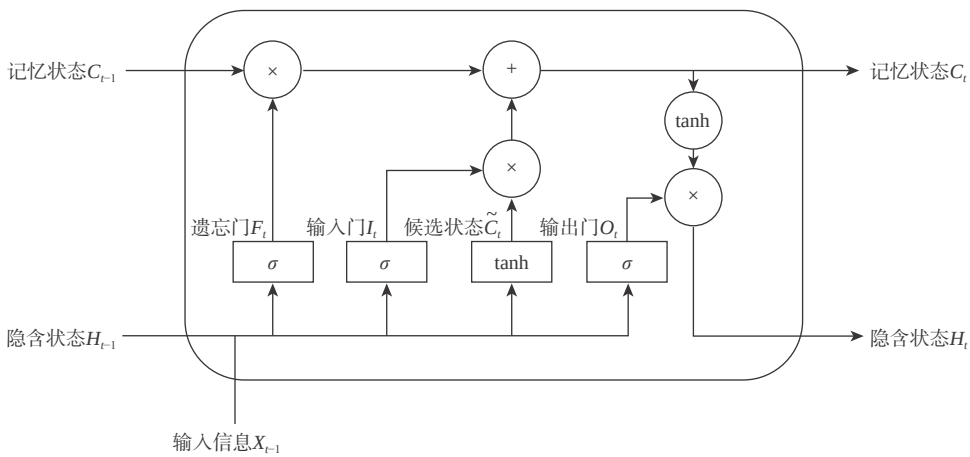


图3 LSTM单元结构

Fig. 3 LSTM unit structure

LSTM的单元结构为链式结构，单元内部有四个网络层，流经LSTM单元的信息将依次通过这四层网络。基于这四层网络的结构，LSTM单元可以进行状态的改变和信息的增添或删减，该功能的实现主要依赖于3个“门”，分别是遗忘门、输入门和输出门。

LSTM在对输入信息进行处理时，第1步要通过一个遗忘门过滤信息。根据输入进遗忘门的H_{t-1}和X_t的信息进行计算，最终输出放弃或保留信息的结果，输出0代表放弃，输出1代表保留。计算表达式如下：

$$F_t=\sigma(W_{\mathrm{F}}[H_{t-1},X_t+b_{\mathrm{F}}]) \tag{10}$$

式中：F_t为遗忘门输出的结果；H_{t-1}为$t-1$时刻的隐含状态；X_t为t时刻的输入信息；W_{F}为遗忘门权重；b_{F}为遗忘门偏置；σ代表Sigmoid激活函数。

第2步是决定添加到单元状态中的新信息，可分为两个步骤：①通过输入门决定更新信息内容；②通过tanh函数层计算得到候选状态$\tilde{C}_t$。两个步骤的计算表达式如下：

$$I_t=\sigma(W_{\mathrm{I}}[H_{t-1},X_t+b_{\mathrm{I}}]) \tag{11}$$

$$\tilde{C}_t=\tanh\left(W_{\tilde{C}_t}\left[H_{t-1},X_t+b_{\tilde{C}_t}\right]\right) \tag{12}$$

式中：I_t为输入门输出的结果；W_{I}为输入门权重；$W_{\tilde{C}_t}$为候选权重；b_{I}、$b_{\tilde{C}_t}$分别为对应结构的偏置。

第3步是对旧单元状态进行更新。旧记忆状态C_{t-1}通过遗忘门进行信息删减，再通过输入门选择更新候选状态$\tilde{C}_t$，得到新记忆状态C_t。计算表达式如下：

$$C_t=F_t\times C_{t-1}+I_t\times\tilde{C}_t \tag{13}$$

最后一步是根据输入单元的H_{t-1}和X_t得到应该输出的状态和特征。该步骤中，H_{t-1}和X_t输入到输出门的Sigmoid激活函数中计算获得判决条件，式(13)得到的新记忆状态C_t输入到tanh函数层计算得到一个向量，该向量与输出门Sigmoid激活函数输出的判决条件相乘，得到更新后的隐含状态H_t。计算表达式如下：

$$O_t=\sigma(W_{\mathrm{O}}[H_{t-1},X_t+b_{\mathrm{O}}]) \tag{14}$$

$$H_t=O_t\times\tanh(C_t) \tag{15}$$

式中: O_t 为输出门输出的结果; W_O 为输出门权重; b_O 为输出门偏置。

2.5　基于 LSTM 网络的数据重构

通过 LSTM 网络对数据进行重构, 其本质是用正常数据样本对 LSTM 网络进行训练, 基于 LSTM 网络的时序预测功能对异常点消除部分进行数据预测, 以完成数据重构。

基于正常 IMF 分量数据样本对 LSTM 网络进行离线训练, 获取最优数据重构网络模型。实际工作过程中, IMF 分量经过异常点检测和消除处理, 分别截取正常数据部分和异常点消除部分, 将正常数据输入数据重构网络模型, 基于正常数据部分对异常点消除部分进行数据预测, 用预测得到数据替换异常点消除部分的数据, 完成数据重构。数据重构流程如图 4 所示。

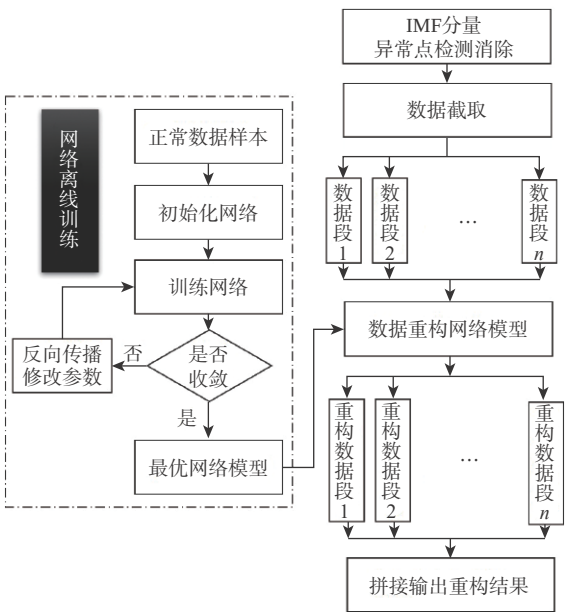


图 4　数据重构流程

Fig. 4　Flow chart of data reconstruction

3　仿真分析

为了验证提出的调频引信抗扫频干扰方法, 本文基于 MATLAB 平台建立三角波调频引信和扫频式干扰的模型, 并就扫频式干扰作用下调频引信未采用和采用异常点重构抗扫频干扰方法的两种情况进行了仿真试验。调频引信及扫频式干扰的仿真参数如表 1 所示。

未施加扫频式干扰时, 仿真弹目距离 50～0 m 的过程, IFD 算法分解阶数设置为 7, 内循环停止准则 SD 值设置为 0.15, 调频引信中频信号经 IFD 算法分解结果如图 5 所示。

仿真试验引信定距距离设置为 3 m, 根据式 (3) 可计算得应选用 2 次谐波定距。2 次谐波对应的谐波频率为 200 kHz, 对分解得到的 IMF 分量做功率谱分析, 由图 6 可得 IMF6 通道内信号分量的能量集中在 200 kHz 附近, 因此在仿真试验中, 选择 IMF6 通道内的 IMF 分量做后续处理。

对调频引信施加干信比为 10 dB 的扫频式干扰, 仿真弹目距离 50～0 m 的过程。未采用异常点重构抗扫频干扰方法时, 引信中频信号直接通过中心频率为 200 kHz 的带通滤波器, 最终得到引信 2 次谐波信号和门限判决输出信号的仿真结果如

表 1　仿真参数

Table 1　Simulation parameters

类别	数值
载波频率/MHz	300
调制频偏/MHz	25
调制频率/kHz	100
采样频率/MHz	5
扫频起始频率/MHz	200
扫频带宽/MHz	200
扫频点数	100
驻留时间/μs	2
干信比/dB	0

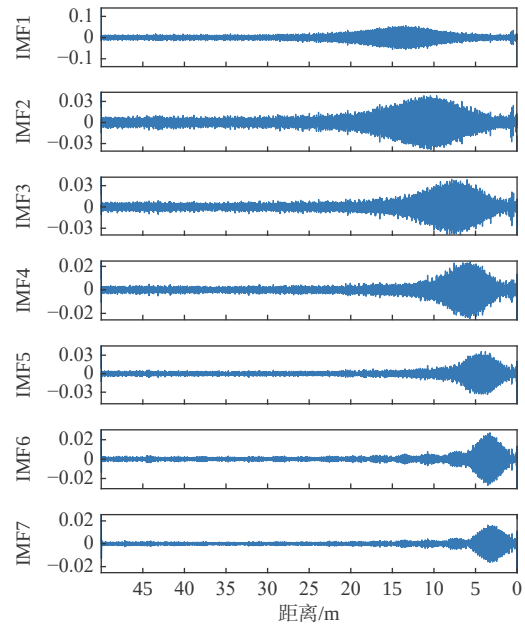


图 5　无扫频干扰时中频信号 IFD 分解结果

Fig. 5　IFD results of intermediate frequency signal without sweep jamming

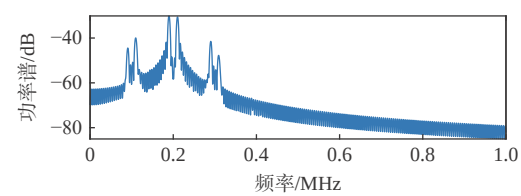


图 6　IMF6 分量功率谱分析

Fig. 6　Power spectrum analysis of IMF6 component

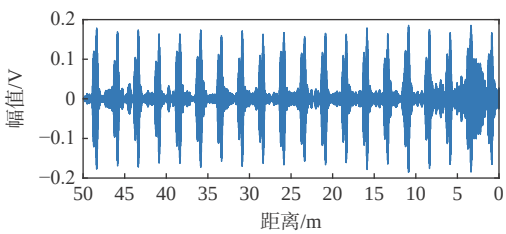
图7所示。

由图7可以看出，在弹目距离50～0 m的过程中，受到扫频式干扰的影响，引信谐波信号在弹目距离并未到达预设的定距距离时，屡次达到门限判决阈值，使门限判决输出信号输出启动信号，导致引信发生误启动。

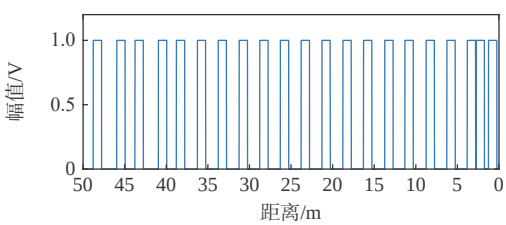
对调频引信施加干信比为10 dB的扫频式干扰，仿真弹目距离50～0 m的过程，并采用异常点重构抗扫频干扰方法。首先对引信中频信号进行IFD算法分解，得到对应谐波频率200 kHz的IMF6通道内信号分量如图8所示。

对IMF6分量进行数据异常点检测与消除处理，结果如图9所示。IMF6分量受扫频干扰影响产生的异常点数据段得到了消除。

基于LSTM网络的基本原理，对数据重构网络



(a) 引信谐波信号



(b) 门限判决输出

图7 扫频式干扰作用下引信输出

Fig. 7 Fuze output under sweep jamming

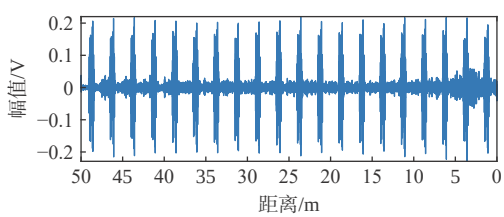


图8 扫频干扰作用下中频信号IMF6分量

Fig. 8 IMF6 component of intermediate frequency signal under sweep jamming

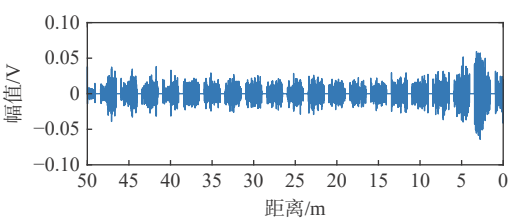


图9 异常点消除后的IMF6分量

Fig. 9 IMF6 component after outlier removal

进行搭建与训练。为了有效重构IMF分量异常点消除部分的数据，针对所输入的IMF分量数据为一维时间序列，采用一层LSTM网络层和一层全连接层的架构对网络进行搭建。网络优化器采用Adam优化算法，回归输出层通过计算半均方误差来测定网络训练过程中的损失。详细的LSTM网络参数如表2所示。

表2 LSTM网络参数

Table 2 LSTM network parameters

类别	数值
输入层特征数	1
LSTM层隐含单元数	50
全连接层输出维数	1
最大训练周期	200
初始学习率	0.01
学习率折损周期	100
学习率折损因子	0.1

初始化网络搭建完成后，基于无干扰时的正常IMF6分量数据，生成训练数据集，对LSTM数据重构网络进行离线训练。训练数据集生成过程中，基于正常IMF6分量进行输入数据和重构数据的截取，共获取1 000组训练数据样本，每组训练数据样本中包含一段输入数据和一段重构数据，两段数据在正常IMF6分量数据中相邻。网络训练过程中，训练数据样本中的输入数据作为网络输入，重构数据作为网络的正确输出，与网络的当前输出对比计算误差，根据误差情况修改网络内部参数，循环迭代训练网络，获取最优数据重构网络模型。

将异常点消除后的IMF分量进行数据截取，输入到训练完成的LSTM数据重构网络中进行数据重构。基于输入到网络的正常数据对异常点消除部分进行预测重构。IMF6分量的数据重构结果如图10所示。

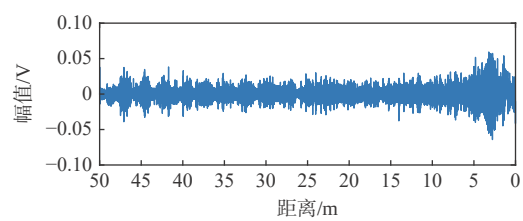


图10 数据重构后的IMF6分量

Fig. 10 IMF6 component after data reconstruction

扫频式干扰作用下，引信中频信号经过异常点重构抗扫频干扰方法处理后，通过中心频率为200 kHz的带通滤波器，最终得到引信2次谐波信号和门限判决输出信号的仿真结果如图11所示。

采用异常点重构抗扫频干扰方法后，在弹目距

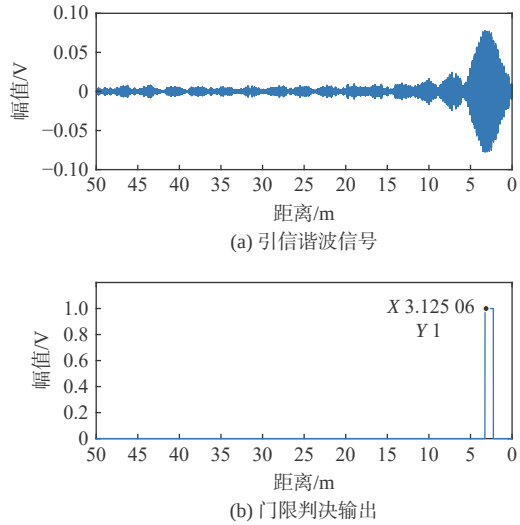


图 11　采用异常点重构抗干扰方法的引信输出

Fig. 11　Fuze output using anti-jamming method by outlier reconstruction

离 50～0 m 的过程中, 引信没有受扫频干扰影响发生误启动, 门限判决输出信号根据 2 次谐波信号的峰值, 在弹目距离 3 m 附近输出启动信号。

针对采用异常点重构抗扫频干扰方法后的引信谐波信号, 与无干扰时引信谐波信号对比进行相似度分析, 表 3 列出了干信比分别为 0 dB、5 dB、10 dB 时引信谐波信号数据重构相似度的情况。

表 3　谐波信号数据重构相似度

Table 3　Similarity of harmonic signal data reconstruction

干信比/dB	数据重构相似度/%
0	93.67
5	92.53
10	90.80

在干信比为 10 dB 的情况下, 将本文方法与基于集成经验模态分解(ensemble empirical mode decomposition, EEMD)[19]算法和奇异谱分析(singular spectrum analysis, SSA)[20]算法的信号分解重构方法进行比较, 由表 4 所示的结果可知, 本文的方法在数据重构和处理时间方面具有显著优势。

表 4　各方法的数据重构相似度与平均处理时间

Table 4　Data reconstruction similarity and average processing time of each method

核心算法	数据重构相似度/%	平均处理时间/s
IFD	90.80	0.002 7
EEMD	84.26	4.940 4
SSA	87.56	1.579 5

综合上述仿真试验结果, 本文提出的基于异常点重构的调频引信抗扫频干扰方法在满足引信信号处理实时性要求的基础上, 可以有效抑制扫频干扰并重构信号数据, 提高了调频引信对抗扫频式干扰的能力。

4　结　论

本文提出并仿真验证了基于异常点重构的调频引信抗扫频干扰方法, 得出如下结论:

1) 该方法基于 IFD 算法将引信中频信号分解为 IMF 分量, 对 IMF 分量进行异常点检测与消除处理, 可将扫频干扰产生的异常点消除, 对扫频干扰进行抑制。

2) 在消除异常点的基础之上, 基于 LSTM 网络对异常点消除部分进行数据重构, 可以有效还原数据信息, 谐波信号数据重构相似度达到 90% 以上, 提高了调频引信在扫频干扰作用下的抗干扰能力。

3) 通过仿真验证, 本文方法算法复杂度和实时性满足引信实际工程需要, 后续将进一步优化算法并进行硬件与实测数据的验证。

参考文献（References）

[1] 崔占忠, 宋世和, 徐立新. 近炸引信原理[M]. 3 版. 北京: 北京理工大学出版社, 2009: 112-119.
CUI Z Z, SONG S H, XU L X. Principle of proximity fuze[M]. 3rd ed. Beijing: Beijing Institute of Technology Press, 2009: 112-119 (in Chinese).

[2] 赵惠昌. 无线电引信设计原理与方法[M]. 北京: 国防工业出版社, 2012: 35.
ZHAO H C. Fundamentals and methodology of radio fuze[M]. Beijing: National Defense Industry Press, 2012: 35 (in Chinese).

[3] GIAQUINTO G, HARTLEY T. ALQ-218(V)(ICAP III)[J]. An electronic Warfare Forecast, 2012 (APR.): 23-28.

[4] RAO G N, SASTRY C S, DIVAKAR N. Trends in electronic warfare[J]. IETE Technical Review, 2003, 20: 139-150.

[5] 岳凯, 郝新红, 栗苹, 等. 基于分数阶傅里叶变换的线性调频引信定距方法[J]. 兵工学报, 2015, 36(5): 801-808.
YUE K, HAO X H, LI P, et al. Research on ranging method for linear frequency modulation radio fuze based on fractional Fourier transform[J]. Acta Armamentarii, 2015, 36(5): 801-808 (in Chinese).

[6] 张天鹏, 刘忙龙, 谢嘉. 基于 FRFT 的调频引信 LFM 干扰抑制改进方法[J]. 探测与控制学报, 2019, 41(2): 22-25.
ZHANG T P, LIU M L, XIE J. Improvement of FM fuze LFM interference suppression based on FRFT[J]. Journal of Detection & Control, 2019, 41(2): 22-25(in Chinese).

[7] 孔志杰, 郝新红, 栗苹, 等. 调频引信谐波时序检测抗干扰方法及实现[J]. 北京航空航天大学学报, 2018, 44(3): 549-555.
KONG Z J, HAO X H, LI P, et al. Harmonic timing sequence detection anti-jamming method and its implementation for FM fuze[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(3): 549-555(in Chinese).

[8] 陈齐乐, 郝新红, 闫晓鹏, 等. 变调制率调频引信双通道相关检测抗数字射频存储干扰方法[J]. 兵工学报, 2019, 40(3): 449-455.
CHEN Q L, HAO X H, YAN X P, et al. Anti-DRFM jamming

method using dual-channel harmonic correlation detection for variable chirp rate FM fuze[J]. Acta Armamentarii, 2019, 40(3): 449-455(in Chinese).

[9] 陈齐乐, 郝新红, 闫晓鹏, 等. 基于谐波系数幅值平均的复合调制引信抗扫频式干扰方法[J]. 北京航空航天大学学报, 2020, 46(7): 1317-1324.
CHEN Q L, HAO X H, YAN X P, et al. Anti sweep jamming method of hybrid modulation fuze based on harmonic coefficient amplitude averaging[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(7): 1317-1324(in Chinese).

[10] 李泽, 闫晓鹏, 栗苹, 等. 扫频式干扰对调频多普勒引信的干扰机理研究[J]. 兵工学报, 2017, 38(9): 1716-1722.
LI Z, YAN X P, LI P, et al. Jamming mechanism of frequency sweep jamming to FM Doppler fuze[J]. Acta Armamentarii, 2017, 38(9): 1716-1722(in Chinese).

[11] LIN L, WANG Y, ZHOU H M. Iterative filtering as an alternative algorithm for empirical mode decomposition[J]. Adv Data Sci Adapt Anal, 2009, 1: 543-560.

[12] HUANG N E, SHEN Z, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proceedings of the Royal Society of London Series A, 1998, 454(1971): 903-998.

[13] CICONE A. Iterative filtering as a direct method for the decomposition of nonstationary signals[J]. Numerical Algorithms, 2020, 85(3): 811-827.

[14] CICONE A, ZHOU H M. Numerical analysis for iterative filtering with new efficient implementations based on FFT[J]. Numerische Mathematik, 2021, 147(1): 1-28.

[15] CICONE A, LIU J F, ZHOU H M. Adaptive local iterative filtering for signal decomposition and instantaneous frequency analysis[J]. Applied and Computational Harmonic Analysis, 2016, 41(2): 384-411.

[16] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9: 1735-1780.

[17] MA X L, TAO Z M, WANG Y H, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data[J]. Transportation Research Part C: Emerging Technologies, 2015, 54: 187-197.

[18] ALI M M, BABAR M I, HAMZA M, et al. Industrial financial forecasting using long short-term memory recurrent neural networks[J]. International Journal of Advanced Computer Science and Applications, 2019, 10(4): 14.

[19] WU Z H, HUANG N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method[J]. Advances in Adaptive Data Analysis, 2009, 1(1): 1-41.

[20] ELSNER J B, TSONIS A A. Singular Spectrum Analysis[M]. Boston: MASpringer US, 1996.

Anti-sweep jamming method for FM fuze based on outlier reconstruction

DUAN Lefan，HAO Xinhong*，CHEN Qile

(Science and Technology on Electromechanical Dynamic Control Laboratory，Beijing Institute of Technology，Beijing 100081，China)

Abstract：An anti-sweep jamming method for frequency modulation (FM) fuze based on outlier reconstruction was proposed for the problem of FM fuze against sweep jamming. Based on the iterative filtering decomposition (IFD) algorithm, the intermediate frequency signal of fuze was decomposed into intrinsic mode function (IMF) components. Sweep jamming was represented as outliers in the IMF components. By detecting and eliminating outliers in the IMF components and using a long short-term memory (LSTM) network to build a data reconstruction network model, the data of outliers was reconstructed, and the sweep jamming could be suppressed. The effectiveness of the proposed method was verified by simulation. The results show that the proposed method improves the ability of FM fuze against sweep jamming.

Keywords：frequency modulation fuze; sweep jamming; anti-jamming method; outlier reconstruction; iterative filtering decomposition; long short-term memory network

Received：2023-04-21；**Accepted**：2023-06-16；**Published Online**：2023-07-11 15:05

URL：link.cnki.net/urlid/11.2625.V.20230711.0954.002

Foundation item：National Natural Science Foundation of China (61973037)

* **Corresponding author**. E-mail：haoxinhong@bit.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0211

基于深度强化学习的离散状态转移算法求解柔性作业车间调度问题

朱家政，王聪，李新凯，董颖超，张宏立*

(新疆大学 电气工程学院，乌鲁木齐 830047)

摘 要：柔性作业车间调度问题(FJSP)作为一种在实际生活中应用广泛的调度问题，对其智能算法具有重要价值。为了解决FJSP，以最小化最大完工时间为优化目标，提出了一种基于近端策略优化的离散状态转移算法(DSTA-PPO)。DSTA-PPO具有3个特点：考虑到FJSP需要同时对工序排序、机器分配同时进行调度安排，结合工序编码和机器编码，设计了一种能够充分表达当前调度问题的状态特征；针对工序排序、机器分配设计了多种基于关键路径的搜索操作；通过强化学习的训练，能够有效地引导智能体选择正确的搜索操作优化当前的调度序列。通过基于不同数据集的仿真实验，验证了算法各环节的有效性，同时在相同算例上以最小化最大完工时间为对比指标与现有算法进行了比较，对比结果表明了所提算法能够在多数算例上以更短的完工时间对算例完成求解，有效地求解了柔性作业车间调度问题。

关 键 词：深度学习；强化学习；离散状态转移算法；近端策略优化算法；柔性作业车间调度

中图分类号：TP181

文献标志码：A **文章编号**：1001-5965(2025)04-1385-10

云制造作为工业4.0中下一代制造系统的新兴范式，在过去的数年中，受到了持续的关注[1]。云制造能够将分布式制造资源虚拟化并整合到云平台中，经过与网络物理系统、物流互联网和人工智能等先进技术的结合[2]，达到按需提供灵活、优质的制造服务[3]。提供一个高效的资源调度更是其中的关键部分，其主要任务是在制造资源上对制造任务进行合理的安排，以达到充分发挥系统性能的目的。本文所研究的FJSP隶属于云制造资源调度任务中的一种[4,5]，虽然其同作业车间调度问题一样都属于NP-hard问题中，但FJSP允许工件有一个或多个加工机器进行选择。因此，FJSP更切合高灵活性、高多样性的云制造模式，有着重要的应用价值[6]。因此，FJSP得到了广泛的关注。

近年来，许多方法已经被应用于解决FJSP。一些研究提出了精确算法，包括Choi等[7]提出的数学编程和混合整数目标编程。然而，由于FJSP的NP-hard特性，精确算法很耗时，而且即使对于小规模的实例也很难达到全局最优。因此，许多研究人员开始开发更有效的解决方法，以获得接近最优的解决方案。Zhang等[8]对粒子群算法和禁忌搜索算法相结合，用以解决FJSP。Nouiri等提出了一种两阶段粒子群算法[9]和基于多智能体的粒子群算法[10]用于解决该问题。Xing等[11]针对FJSP提出了一种基于知识模型的蚁群优化算法，其中，知识模型被用来指导当前的启发式搜索。Jiang等[12]首次将离散的灰狼优化法应用于解决FJSP，其中引入了自适应突变法，目标是最小化完工时间。尽管这些智能

收稿日期：2023-04-26；**录用日期**：2023-07-27；**网络出版时间**：2023-09-12 16:53
网络出版地址：link.cnki.net/urlid/11.2625.V.20230911.1743.001
基金项目：国家自然科学基金(52267010,62263030)；新疆维吾尔自治区自然科学基金资助项目(2022D01C367,2022D01E33)
* **通信作者**. E-mail：zhlxju@163.com

引用格式：朱家政，王聪，李新凯，等. 基于深度强化学习的离散状态转移算法求解柔性作业车间调度问题[J]. 北京航空航天大学学报，2025，51(4)：1385-1394. ZHU J Z，WANG C，LI X K，et al. A deep reinforcement learning based on discrete state transition algorithm for solving fuzzy flexible job shop scheduling problem[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(4)：1385-1394 (in Chinese).

优化算法能够在合理的时间内解决FJSP,但并不能在更复杂的调度问题取得较好的表现。

状态转移算法是Zhou等[13]提出的一种智能型随机性全局优化算法,该算法同其他群智能算法相比具有结构简单、参数少、易理解等优点。董天雪等[14]提出了一种新的离散状态转移算法提高企业员工指派问题的成功率和稳定性。张凤雪等[15]采用状态转移算法对非线性规划问题进行了求解,并验证了算法的有效性。状态转移算法同其他优化算法相比,具有较快的收敛速率和较高的寻优精度,但是,在特定的复杂高维全局优化问题中,状态转移算法表现出较为缓慢的收敛速度和较低的收敛精度,且其在局部收敛方面的能力相对有限。

深度强化学习算法提供了一种可扩展的方法来解决具有共同特征的调度问题。Du等[16]提出了一种基于知识的强化学习同分布式估计算法结合的混合算法求解FJSP。Chen等[17]提出了一种自适应学习的遗传算法求解FJSP,通过强化学习自适应地学习遗传算法中的关键参数以提高求解性能。Lei等[18]提出了一种端到端的深度强化学习算法,利用图神经网络求解FJSP,并通过实验在大规模的FJSP中验证了算法的有效性。张凯等[19]提出一种集成式强化学习算法求解FJSP,通过设计调度规则完成求解。王凌等[20]提出一种协同群智能算法,通过基于信息交互的工序排序搜索和基于Q-learning的机器分配搜索求解分布式绿色柔性作业车间调度问题。可以看出深度强化学习算法已经广泛地应用在FJSP中,并取得较好的效果。

本文结合状态转移算法和深度强化学习算法的优点,提出了一种基于深度强化学习的离散状态转移算法。将离散状态转移算法的三种高效的变换算子添加到深度强化学习算法工件选择的动作中,并添加机器的邻域搜索结构用于确定加工机器。针对缩短柔性作业车间调度的最大完工时间问题,通过深度神经网络训练强化学习智能体在不同的状态下,能够自适应选择合适的离散状态转移动作和机器邻域搜索结构,改进了深度强化学习算法的策略搜索过程,增加了策略空间的多样性,提高了策略搜索的效果。

为了验证DSTA-PPO的有效性,本文在两种测试算例中进行了仿真实验,分别在低维和高维问题中同其他深度强化学习算法完成了对比实验。

1 问题描述

1.1 问题定义

FJSP描述如下:n个工件在m台机器上加工,工件集和机器集分别记为$J_i(i=1,2,\cdots,n)$和$M_k(k=1,2,\cdots,m)$。每个工件都包含有多道不同的工序记为$J_i=\{O_{i1},O_{i2},\cdots,O_{in_i}\}$,每道工序进行加工时须严格按照先后顺序,每个工件中至少包含一道工序具有多个可加工机器。在加工过程中,工件中的工序能够选择任意一台候选机器进行加工。FJSP需要对所有工件工序的加工机器进行选择,并合理地安排加工机器加工工序的顺序,在某些优化目标下合理地安排生产。本文以最小化最大完工时间$C_{\max}$为优化目标,即

$$\min C=\min(\max\{C_i|i=1,2,\cdots,n\})$$

式中:C_i表示工件J_i的完工时间;最大完工时间$C_{\max}=\max\{C_i|i=1,2,\cdots,n\}$。

不同的假设和约束条件会导致不同的柔性作业车间调度问题,本文作出如下假设:

1)起始时刻所有机器为空闲状态,所有工件为待加工状态;

2)同一机器同一时刻中只能加工一个工件;

3)同一工件工序顺序确定不变,不同工件无优先级差别;

4)任意一道工序开始加工后不能中断。

表1为三工件、三机器的FJSP实例。

表1 3×3的FJSP实例

Table 1 3×3 FJSP instance

工件	工序	工序号	工序和加工时间		
			M_1	M_2	M_3
J_1	O_{11}	1	1	2	*
	O_{12}	2	3	2	1
J_2	O_{21}	3	*	2	1
	O_{22}	4	1	2	4
	O_{23}	5	*	*	4
J_3	O_{31}	6	3	2	2
	O_{32}	7	1	*	2

注:“*”表示该工序无法在该机器上进行加工。

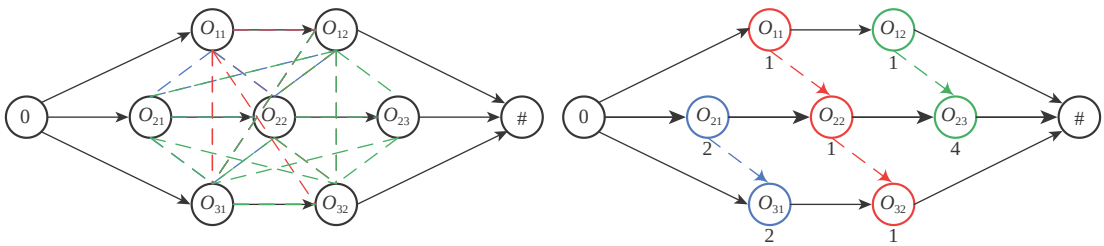
1.2 调度问题的析取图描述

析取图以图的形式来表示FJSP的调度结果,能够更直观地了解工序的加工机器和加工顺序。一个FJSP析取图记为$G=(N,A,E)$,其中,N是包含了调度问题中所有工件工序的顶点、开始顶点和结束顶点(0和#)的集合,即$N=\{O_{ij}|i=1,2,\cdots,n;1\leqslant j\leqslant n_i\}\cup\{0,\#\}$,$A=\{(O_{ij},O_{i,j+1})|i=1,2,\cdots,n;1\leqslant j\leqslant n_i-1\}\cup\{(0,O_{i1})(O_{in_i},\#)|i=1,2,\cdots,n\}$,表示同一工件的两道连续工序之间的合取弧集合,反映了任意工件的工序之间的顺序约束关系;E表示同一台机器上加工的相邻两道工序之间析取弧集合,反映了每台机器上的工序加工顺序和加工序列;对于任意一道工

序ω, 将其在所选机器M_k上的加工时间p_{wk}作为该节点的权重($p_{wk}>0; p_0=p_{\#}=0$), 当工序ω的加工机器是确定的, p_{wk}可简写为p_w。当图G为一个无环图时, 图G能够表示为调度问题的一个可行解。

析取图G路径$(\omega_1,\omega_2,\cdots,\omega_q)$的长度为从最先开始加工的工序$\omega_1$到最后加工的工序$\omega_q$间加工花费时间总和, 即$\sum_{i=2}^{q-1} p_{w_i}$。$L(i,j)$表示工序$i$到工序$j$之间的最长路径的长度。析取图$G$的关键路径为开始至结束顶点间加工时间最长的一段路径, 在关键路径中加工的工序则为关键工序。用$L(0,\#)$表示关键路径长度, 则调度问题的最大完工时间可以表示为$C_{\max}(G)=L(0,\#)$。

以表 1 中的 FJSP 实例为例, 实例的析取图模型如图 1 所示。图 1(a)所示的无向图表示未分配机器和工序时的析取图模型。图 1(b)为实例的一个可行解, 加粗实线表示关键路径, 可以看出其长度为 7, 即最大完工时间$C_{\max}(G)=7$, 且存在如图中加粗显示的一条关键路径, 分别是$0\to O_{21}\to O_{22}\to O_{23}\to\#$, 工序$O_{21}$、$O_{22}$、$O_{23}$是关键工序。



(a) 实例析取图模型　　(b) 实例的调度解

图 1　FJSP 的析取图描述

Fig. 1　Description of FJSP's disjunction graph

2　基础算法

2.1　编码和解码

FJSP 需要确定工序的加工机器、工序的加工顺序, 因此分别采用表示工序排序(operation sequence, OS)和加工设备选择(machine assignment, MA)的两个向量(OSV, MAV)对解进行编码, 如图 2 所示。OSV 中的每个元素为工件序号, 第p次出现表示该工件的第p道工序, MAV 的每个元素对应于 OSV 中相应位置工序的加工机器。为保证解的可行性, MAV 中每个元素的取值都从 OSV 中对应工序的可加工机器集合中选取。

	O_{11}	O_{31}	O_{12}	O_{21}	O_{32}	O_{22}	O_{23}
OSV	1	3	1	2	3	2	2
MAV	2	3	1	2	1	3	2
	M_2	M_3	M_1	M_2	M_1	M_3	M_2

图 2　编码示意图

Fig. 2　Encoding diagram

为求得当前序列对应解的最大完工时间, 需将编码解码为满足约束的调度方案。采用左插法的主动调度理论, 将 OSV 中的工件工序按顺序安排至对应的 MAV 中的机器上, 并遵循最早可加工时间安排, 得到唯一的调度方案。

2.2　搜索算子

为提高算法的寻优能力及寻优策略, 分别针对编码中 OS 的操作排序以及 MA 中的机器分配设计多种有效操作。

2.2.1　针对 OS 的操作设计

状态转移算法是一种创新性的随机性全局优化方法, 其核心思想是将问题的可行解视为一种状态, 并将可行解的更新过程看作从一个状态向另一个状态的变化, 即状态之间发生转移。该算法产生候选解的基本框架如式(1)所示:

$$\begin{cases} x_{k+1}=A_k x_k \oplus B_k u_k \\ y_{k+1}=f(x_{k+1}) \end{cases} \tag{1}$$

为了保持状态转移算法在 FJSP 状态转移过程的可控性、高效性, 使用了 3 种特殊的状态转移算子对 OS 序列中的关键路径上的关键工序序列进行操作。具体如下。

1) 交换变换算子

$$x_{k+1}=A_k^{\text{swap}} x_k \tag{2}$$

式中: $A_k^{\text{swap}}\in\mathbf{R}^{n\times n}$称为交换变换矩阵, 其具有交换变换功能。该算子能够交换x_k中随机两个位置元素。如下式所示:

$$\begin{bmatrix}1\\5\\3\\4\\2\end{bmatrix}=\begin{bmatrix}1&0&0&0&0\\0&0&0&0&1\\0&0&1&0&0\\0&0&0&1&0\\0&1&0&0&0\end{bmatrix}\times\begin{bmatrix}1\\2\\3\\4\\5\end{bmatrix} \tag{3}$$

对 OS 序列中第 2 号和第 5 号位置上的元素进

行了交换。

2）移动变换算子

$$x_{k+1}=A_k^{\text{shift}}x_k \tag{4}$$

式中：$A_k^{\text{shift}}\in\mathbf{R}^{n\times n}$称为移动变换矩阵，其具有移动变换功能。该算子能够移动x_k中某个随机位置到另一个随机位置后面。比如：

$$\begin{bmatrix}1\\2\\4\\5\\3\end{bmatrix}=\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&0&1&0\\0&0&0&0&1\\0&0&1&0&0\end{bmatrix}\times\begin{bmatrix}1\\2\\3\\4\\5\end{bmatrix} \tag{5}$$

将OS序列中第3号位置上的元素移动至第5号位置元素的后面。

3）对称变换算子

$$x_{k+1}=A_k^{\text{sym}}x_k \tag{6}$$

式中：$A_k^{\text{sym}}\in\mathbf{R}^{n\times n}$称为对称变换矩阵，是一个带有对称变换功能的随机0-1矩阵。该算子具有将x_k中随机两个位置之间所有元素对称或倒置的能力。比如：

$$\begin{bmatrix}1\\5\\4\\3\\2\end{bmatrix}=\begin{bmatrix}1&0&0&0&0\\0&0&0&0&1\\0&0&0&1&0\\0&0&1&0&0\\0&1&0&0&0\end{bmatrix}\times\begin{bmatrix}1\\2\\3\\4\\5\end{bmatrix} \tag{7}$$

将OS序列中第2、3、4、5号位置元素进行对称变换，其位置发生倒置。

2.2.2 针对MA的操作设计

在每道工序均有一个或多个加工设备可选择的情况下，MA的搜索空间是巨大的，针对同一OSV的机器分配方案数量为$\prod_{j=1}^{n}\prod_{i=1}^{n_j}|M_{j,i}|$。不同的OSV的最优MAV仍有不同，若依靠随机性的迭代搜索，效率太低，无法寻找到较优的MAV。因此设计了一种关键工序的邻域搜索：跨机器移动工序。跨机器移动关键工序邻域结构示意图如图3所示。在图3所示情况下，对关键工序ω进行跨机器移动，采用直接利用机器空闲时间的方式。由于工件的加工顺序要求，以工序ω的前继工序JP[ω]的最早完工时间EC(JP[ω])和后继工序JS[ω]的最晚开工时间LS(JS[ω])为区间的范围内，在工序ω的候选机器集中查找机器空闲时间。工序ω能否跨机器移动，取决于移至机器位置的空闲时间，只有当式(8)满足时，跨机器移动关键工序ω至i和j中间才不会增大关键路径长度。

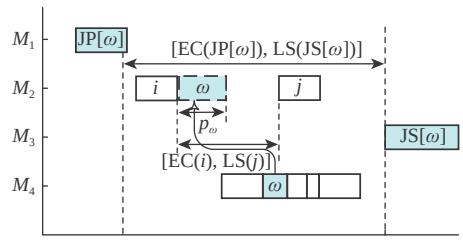


图3 邻域结构示意图

Fig. 3 Schematic diagram of neighborhood structure

$$\{\text{EC(JP}[\omega]),\text{LS(JS}[\omega])\}\cap\{\text{EC}(i),\text{LS}(j)\}>p_\omega \tag{8}$$

2.3 强化学习和PPO

一般而言，强化学习算法能够处理马尔可夫过程。在这个过程中，智能体通过训练和试错与环境进行交互以达到最大化期望的累积长期奖励。一个典型的马尔可夫过程可以表示为一个五元组(S,A,P,γ,R)。在每个决策点t，智能体获取当前时刻状态$s_t\in S$并根据策略$\pi(S\to A)$采取可行的动作。之后，根据状态转移概率$p(s_{t+1}|s_t,a_t)\in P(S\times A\to S)$进入新的状态$s_{t+1}$得到即时奖励$r_t\in R(S\times A\times S\to\mathbf{R})$。$Q_\pi(s,a)$表示在状态$s$下根据给定策略$\pi$采取动作$a$时的期望累积奖励(也称为行动价值函数或$Q$函数)，其定义如下：

$$Q_\pi(s,a)=E\left[r_t+\gamma r_{t+1}+\gamma^2 r_{t+2}+\cdots|s_t=s,\pi\right] \tag{9}$$

强化学习算法中智能体的目标是从所有可能获取的状态中找到一个能使期望累积奖励最大化的最佳策略π^*，其定义如下：

$$\pi^*=\arg\max E_{s|\pi}\left[V_\pi(s)\right] \tag{10}$$

在基于深度强化学习的方法中，策略$\pi(\theta)$总是以具有可微调参数θ的深度神经网络表示。为了保证式(10)在参数更新后旧策略$\pi_{\text{old}}(\theta_{\text{old}})$和新策略$\pi(\theta)$之间呈现单调递增，能够最大化智能体新旧策略间KL散度的信赖域策略优化被提出。而近端策略优化是信赖域策略优化的一种扩展，它保留了信赖域策略优化的稳定性和可靠性，且计算效率更高，实现起来也更简单。PPO使用了一个只需要计算似然的新目标来替代KL散度约束，将其定义为$\text{ratio}_t(\theta)=(\pi(s_t,a_t;\theta))/(\pi_{\text{old}}(s_t,a_t;\theta_{\text{old}}))$，PPO的目标如下所示：

$$\max_\theta E_t\min(\text{ratio}(\theta)A_t,\text{clip}(\text{ratio}_t(\theta),1-\varepsilon,1+\varepsilon)A_t) \tag{11}$$

式中：$A_t=Q_{\pi\text{old}}(s_t,a_t)-V_{\pi\text{old}}(s_t)$为优势函数，$\varepsilon\in(0,1)$为裁剪系数。由于PPO算法可以通过限制策略的更新来降低参数设置的敏感性，从而能够获得良好的性能，本文将采用它作为基本的学习算法。

3　基于深度强化学习的离散状态转移算法

本节分别从状态定义、动作设计、奖励设计等方面对本文所提出的基于深度强化学习的智能优化算法进行介绍。

3.1　状态特征定义

状态特征描述了决策时刻下对当前调度问题解决的状态。因此, 对于状态特征的设计应该能够反映出当前动作的影响, 并有助于后续的动作选择。

在 DSTA 算法中是以当前 OSV 作为状态进行候选解的选取, OSV 与对应的 MAV 解码后, 即可得到当前问题的一个对应调度解, 不同的 OSV 则会产生不同的调度解, 选用 OSV 作为状态是对 FJSP 很好的反应。因此本文所提出的 DSTA-PPO 将在当前 OSV 和 MAV 的基础上, 进行状态特征设计。在决策时刻下状态特征中变化的工件序列, 对 OSV 和 MAV 进行了如图 4 所示的处理, 得到智能体的 State 列表内元素以 ijk 形式进行表示, 其中 i 代表加工工件序号, j 代表加工工件的工序序号, k 代表加工机器的序号。状态特征列表作为深度神经网络的输入, 为更有利于神经网络的训练, 对当前状态列表进行标准化处理。标准化处理的核心在于, 在同环境交互的过程中, 维护一个动态的关于所有经历过的所有状态的 mean 和 std, 并对当前 state 进行标准化处理, 使 state 符合 mean=0, std=1 的正态分布。具体状态标准化处理算法如算法 1 所示。

算法 1　状态标准化处理算法。

输入: s_t, t

输出: Normalized state list s_{t_norm}

Initialization t=0、Mean= 0、S=0、Std=sqrt(S)
if t=0 then
　Mean=s_t
　Std=s_t
end
if $t>0$ then
　Mean_old = Mean
　$\text{Mean} = \text{Mean_old} + (s_t - \text{Mean_old})/t$
　$S = S + (s_t - \text{Mean_old}) * (s_t - \text{Mean})$
　$\text{Std} = \text{sqrt}(S/t)$
end
$s_t = (s_t - \text{Mean})/(\text{Std} + 1e-8)$
t+ = 1

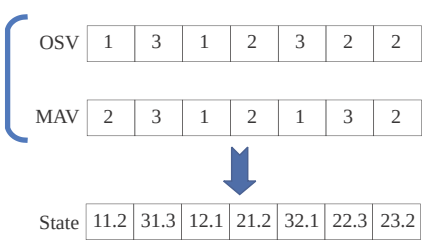


图 4　状态特征列表

Fig. 4　Status feature list

3.2　动作空间设计

在 DSTA-PPO 算法中, 基于 FJSP 问题的特点设计了 4 种动作来提高求解质量, 4 种动作都是针对关键路径上的关键工序进行操作的。

a_1、a_2、a_3 3 种动作都是基于离散状态转移算法的 3 种算子进行设计的。原始的 3 种搜索算子是对关键工序直接进行操作变换, 这种过程为一次状态转移, 进行一次状态转移后, 能够得到 C_n^2 种候选的关键工序。在基于一次状态转移所得候选关键工序集的基础上, 使用相同的搜索算子再次进行状态转移变换, 得到二次状态转移的候选关键工序集, 此时, 若忽略其中重复的候选解, 可得 $C_n^2 \times C_n^2$ 种候选解。通过此种方式能够提升 3 种动作的搜索空间, 提高了智能体对关键工序的探索能力, 其具体执行如算法 2 所示。以上 3 种动作在离散状态转移算法的思想上做了一定改进, 在选定一种动作后执行 SE 次, 得到大小为 SE 的 OSV 集合, 计算所有该集合所对应的完工时间得到解集。此时不再要求解集中一定出现优于当前最优解的结果时才进行状态更新。而是选定解集中表现最好的 OSV 进行状态更新, 即使此时的完工时间不一定优于已知的最优完工时间, 从而避免算法陷入局部最优。

算法 2　a_1, a_2, a_3。

输入: Critic path A, action a

输出: New OSV

for i in SE do
　if $a = a_1$ then
　　Swap(A)
　　Swap(A)
　　Update(OSV)
　end
　if $a = a_2$ then
　　Shift(A)
　　Shift(A)
　　Update(OSV)
　end
　if then
　　Sysmmetry(A)

14. Sysmmetry(A)
15. Update(OSV)
16. end
17. end

a_1、a_2、a_3 3种动作的执行能够保证得到较优的OSV,但MAV在整个过程中是没有变换的,最终得到的调度解也只能保证是在当前MAV下的一种局部最优解。因此添加了跨机器邻域搜索的动作a_4。智能体通过执行a_4能够改变关键工序的加工机器。其具体执行如算法3所示。通过执行动作a_4,提升了智能体的全局寻优能力。

算法3 a_4。

输入: Critic path A,action α

输出: New OSV

1. for critic operation O_{ij} in A do
2. for k in M_k do
3. if [EC(JP[j],LS(JS[f]]]∧[EC(k),LS(k)]> $p_{j,\omega}$
4. MAV[O_{ij}]=k
5. end
6. end
7. end

3.3 奖励设计

在每个决策点,通过对当前OSV、MAV解码后计算可得出最大完工时间,通过式(12)进行奖励计算,当最大完工时间减小时,奖励会相应变大。

$$r_t = 1/C\max_t \tag{12}$$

3.4 DSTA-PPO算法计算过程

DSTA-PPO算法被提出用以训练智能体解决FJSP,提高智能体对当前状态自适应选择动作的能力,提出的方法能够有效地处理工件调度排序问题,避免了处理调度问题时工件间的约束条件。该算法的具体流程如算法4所示。

算法4 DSTA-PPO算法流程。

1. 初始化当前策略网络$\pi(\theta)$和价值网络$v(\omega)$
2. for apoch $=1:L$ do
3. 观察当前状态s_t,提取当前状态特征向量$\phi(s_t)$
4. 选择一个动作$a_1=\pi(\theta_2\phi(s_t))$
5. if $a_1=a_1$ or $a_1=a_2$ or $a_1=a_3$
6. 通过算法2更新OSV
7. end
8. if $a_1=a_4$
9. 通过算法3更新MAV
10. end
11. 计算即时奖励r_t并产生新的状态s_{t+1}
12. if $t\%M=0$
13. 通过算法5更新$\pi(\theta),v(\omega)$
14. end
15. end

算法5 $\pi(\theta),v(\omega)$参数更新。

1.
2. for $i=t-1:-1:t-M$ do
3. $R_i=r_i+\gamma R_{i+1}$
4. $A_i=R_i-v_m(\phi_i;\omega)$
5. $\text{ratio}(\theta)=\dfrac{\pi(\phi_i,a_i;\theta)}{\pi_{\text{old}}(\phi_i,a_i;\theta_{\text{old}})}$
6. end
7. $\text{PPO}(\theta)=-\sum_{i=t-M}^{t-1}\min(\text{ratio}i(\theta)A_i,\text{clip}(\text{ratio}i(\theta),1-\varepsilon,1+\varepsilon)A_i)$
8. $\text{MSE}(\omega)=\sum_{i=t-M}^{t-1}(R_i-v(\phi_i;\omega))^2$
9. 通过梯度下降法更新θ、PPO(θ)
10. 通过梯度下降法更新ω、MSE(ω)
11. 返回θ、ω

4 实例仿真

4.1 实验设置

本文使用Python编程语言在Windows11系统上运行Pycharm实现了所提出的DSTA-PPO算法。硬件条件为AMD Ryzen 5 4600H 3.00 GHz, 16 GB RAM。为验证算法性能,使用了2类不同规模大小的FJSP测试集。第1类测试集为Brdata测试集,包括了10个不同规模的算例。第2类测试集为Hurink测试集,包括了40个不同规模的算例。

4.2 参数设置

本文所提出算法包含4个关键参数:①候选解集大小SE;②动作网络和评价网络学习率α;③batchsize(batch)表示一次完整的训练迭代所收集的经验轨迹;④mibi-batchsize(minibatch)表示从batch中随机选取的小部分样本个数,用于更新网络参数。采用实验设计(design-of-experiment, DOE)方法分析参数对算法性能的影响。为综合参数在不同规模算例上的影响,选取第2类测试集中Vdata_la1、Vdata_la6、Vdata_la11、Vdata_la16、Vdata_la21等5个不同规模的算例进行分析。每个参数设置4个水平,SE$\in\{5,10,20,30\}$,$\alpha\in\{1e-3,1e-4,3e-5,1e-5\}$,batch$\in\{128,256,512,1024\}$,mini_batch$\in\{8,16,32,64\}$。选择正交数组$L_{16}(4^4)$,包含16组参数水平组合。对于每个组合计算同算例最优解之间的误差,算法运行10次并取最优的一次作为该组参数水平的解。运行16个参数水平组合可得解集$E_i(i=1,2,\cdots,16)$。计算所有算例的参数水平组合

的平均误差作为响应值(response value, RV), 如表 2 所示。对表 2 中的数据进行极差分析可得表 3, 通过极差分析, 能够分析各因子间的优势及因子间具体水平的优劣。表 3 中 K 值为某因子某水平时试验数据求和; K_{avg} 为对应的平均值; 最佳水平指某因子的最佳 K_{avg} 对应的水平标号; R 指因子的极差值, 该值越小则该因素对于算法性能影响越大。因子的趋势变化如图 5 所示。

根据表 3 以及图 5 可以看出, SE 对于算法性能影响最大, α、batch、mini-batch 次之。合适的 SE 能够影响算法的多样性和收敛性。过小的α则会使算法收敛过快, 较小的 batch、mini-batch 会使算法性能得到提升。根据以上实验结果和分析, 设置 SE = 20,$\alpha = 3\times10^{-5}$,batch = 128, mini_batch = 8。

4.3　与现有算法比较

为了验证 DSTA-PPO 算法在求解柔性作业车间调度问题的有效性和先进性, 首先在第一类测试集 Brdata 中进行了算法实例仿真, 在第一类测试集中, 同目前已有的使用了相同算例的最新深度强化学习算法 SLGA[17] 和集成强化学习算法[19] 进行了比较, SLGA 算法的相关参数为: 学习率为 0.75, 折扣因子为 0.2, 初始奖励为 1, 贪婪率为 0.85。为了显示算法改进的有效性, 也同 DSTA 算法进行了比较。如表 4 所示, 其中加黑显示的数据代表了 4 种算法在各个算例上求得的最优解, 带星号的表示目前已知的算法所能求得的最优解。在求解最优解方面, DSTA-PPO 算法在 7 个算例中都找到了目前算例所能求得的最优解, 虽然有 3 个算例未能找到所能求得的最优解, 所得解的效果略差于当前的启发式算法, 但深度强化学习算法的优势主要在于当完成训练后, 即可在整个测试集上直接进行应用, 无须像启发式算法一般需要对每个算例分别进行求解。

表 2　正交试验表及各参数组合的响应值

Table 2　Orthogonal test table and response values of various parameter combinations

编号	参数因子组合				RV
	SE	α	batch	mini-batch	
1	1	1	1	1	3.054 784
2	1	2	2	2	3.181 015
3	1	3	3	3	3.080 03
4	1	4	4	4	2.928 553
5	2	1	2	3	2.802 323
6	2	2	1	4	2.575 107
7	2	3	4	1	2.600 353
8	2	4	3	2	2.827 569
9	3	1	3	4	2.297 4
10	3	2	4	3	2.398 384
11	3	3	1	2	2.019 692
12	3	4	2	1	2.095 43
13	4	1	4	2	2.322 646
14	4	2	3	1	2.171 169
15	4	3	2	4	2.145 923
16	4	4	1	3	2.196 415

表 3　参数各水平的极差分析

Table 3　Extreme variance analysis of parameters at different levels

参数因子	K值				K_{avg}值				最佳水平	R	水平数量	每水平重复数
	1	2	3	4	1	2	3	4				
SE	12.24	10.81	8.81	8.84	3.06	2.7	2.2	2.21	3	−0.86	4	4
α	10.48	10.33	9.85	10.05	2.62	2.58	2.46	2.51	3	−0.16	4	4
batch	9.85	10.22	10.38	10.25	2.46	2.56	2.59	2.56	1	−0.13	4	4
mini-batch	9.92	10.35	10.48	9.95	2.48	2.59	2.62	2.49	1	−0.14	4	4

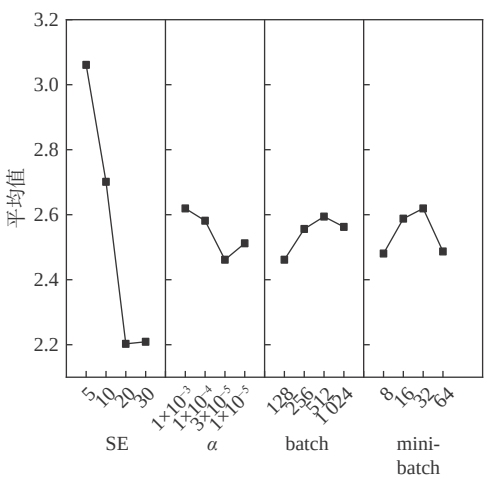


图 5　因素的趋势变化

Fig. 5　Trend changes in factors

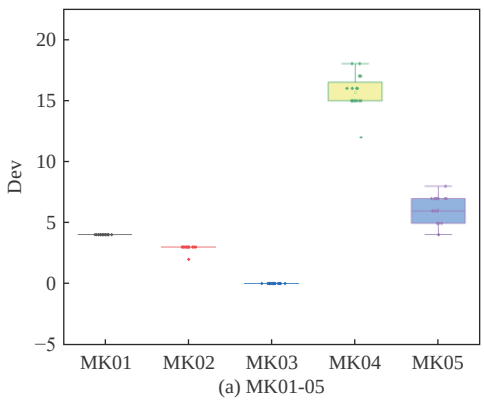
如图 6 所示, 为 DSTA-PPO 算法在 10 个算例上运行 20 次所得结果与算例下界求得的差值所做的箱线图, 可以看出, DSTA-PPO 算法在各个算例中的波动程度都很小, 均值与最优值差距很小。这表明 DSTA-PPO 在求解 FJSP 问题相较其余 3 种算法会有着更好的统计效果。图 7 给出了算法在 MK10 算例上的调度甘特图。

在第 2 类测试集上, 以 Multi-PPO[19] 为基准在 8 组不同规模的算例中进行算法先进性的对比, Multi-PPO 算法参数为: 学习率为 0.001, 裁剪系数为 0.2。对比包括每个算例的调度结果、其同最优解的差距。结果如表 5 所示。表中加黑部分为该算例中表现最好的算法。可以看出, DSTA-PPO 和

表4 第一类测试集对比

Table 4 Comparison of the first type of dataset

数据集	最优解	DSTA	SLGA	集成强化学习	DSTA-PPO
Mk01	36	**40***	**40***	42	**40***
Mk02	24	28	27	31	**26***
Mk03	204	204	**204***	**204***	**204***
Mk04	48	66	**60***	69	**60***
Mk05	168	178	**172***	180	**172***
Mk06	33	73	69	68	**62**
Mk07	133	147	144	153	**142**
Mk08	523	**523***	**523***	**523***	**523***
Mk09	299	327	320	315	**307***
Mk10	165	263	254	230	**225**



(a) MK01-05

(b) MK06-10

图6 箱线图

Fig. 6 Box plot

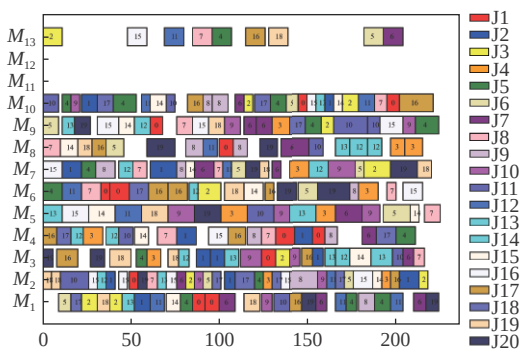


图7 MK10算例调度图

Fig. 7 MK10 example scheduling diagram

Multi-PPO算法表现明显优于DSTA算法。DSTA-PPO算法在28个算例上都取得了比Multi-PPO更优的值。在机器数为10的算例中，随着机器数的增加，DSTA-PPO算法的寻优能力相较Multi-PPO算法变弱，但在最终的平均差距中，DSTA-PPO相较Multi-PPO的值提升了约39%。能够发现，虽然

表5 第二类测试集对比

Table 5 Comparison of the second type of dataset

Dataset		Val				gap/%		
		DSTA	Multi-PPO	DSTA-PPO	UB	DSTA	Multi-PPO	DSTA-PPO
10×5	Vdata_la1	613	610	**573**	570	7.54	7.02	**0.53**
	Vdata_la2	557	555	**536**	529	5.29	4.91	**1.32**
	Vdata_la3	531	532	**487**	477	11.32	11.53	**2.10**
	Vdata_la4	538	530	**510**	502	7.17	5.58	**1.59**
	Vdata_la5	490	507	**467**	457	7.22	10.94	**2.19**
15×5	Vdata_la6	824	820	**802**	799	3.13	2.63	**0.38**
	Vdata_la7	788	757	**755**	749	5.21	1.07	**0.80**
	Vdata_la8	791	782	**765**	765	3.40	2.22	**0.00**
	Vdata_la9	925	879	**859**	853	8.44	3.05	**0.70**
	Vdata_la10	830	862	**808**	804	3.23	7.21	**0.50**
20×5	Vdata_la11	1 102	1 101	**1 074**	1 071	2.89	2.80	**0.28**
	Vdata_la12	960	950	**939**	936	2.56	1.50	**0.32**
	Vdata_la13	1 070	1 053	**1 042**	1 038	3.08	1.45	**0.39**
	Vdata_la14	1 095	1 086	**1 071**	1 070	2.34	1.56	**0.09**
	Vdata_la15	1 117	1 111	**1 092**	1 089	2.57	2.02	**0.28**
10×10	Vdata_la16	**717**	**717**	**717**	717	**0.00**	**0.00**	**0.00**
	Vdata_la17	646	647	**646**	646	0.00	0.15	**0.00**
	Vdata_la18	714	**663**	**663**	663	7.69	**0.00**	**0.00**
	Vdata_la19	705	626	644	617	14.26	1.46	4.38
	Vdata_la20	756	**756**	**756**	756	0.00	**0.00**	**0.00**
15×10	Vdata_la21	903	887	**870**	804	12.31	10.32	**8.21**
	Vdata_la22	849	**793**	798	736	15.35	**7.74**	8.42
	Vdata_la23	970	**858**	888	815	19.02	**5.28**	8.96
	Vdata_la24	891	883	**834**	775	14.97	13.94	**7.61**
	Vdata_la25	896	883	**820**	756	18.52	16.80	**8.47**
20×10	Vdata_la26	1 195	**1 089**	1 100	1 054	13.38	**3.32**	4.36
	Vdata_la27	1 231	**1 123**	1 133	1 084	13.56	**3.60**	4.52
	Vdata_la28	1 249	**1 106**	1 121	1 070	16.73	**3.36**	4.77
	Vdata_la29	1 121	**1 049**	1 048	994	12.78	**5.53**	5.43
	Vdata_la30	1 256	**1 117**	1 122	1 069	17.49	**4.49**	4.96
30×10	Vdata_la31	1 688	1 561	**1 553**	1 520	11.05	2.70	**2.17**
	Vdata_la32	1837	**1 693**	1 697	1 658	10.80	**2.11**	2.35
	Vdata_la33	1 630	1 531	**1 530**	1 497	8.88	2.27	**2.20**
	Vdata_la34	1 731	**1 562**	1 565	1 537	12.62	**1.63**	1.82
	Vdata_la35	1 711	**1 574**	1 582	1 549	10.46	**1.61**	2.13
15×15	Vdata_la36	1 100	985	**967**	948	16.03	3.90	**2.00**
	Vdata_la37	1 171	**1 028**	1 035	986	18.76	**4.26**	4.97
	Vdata_la38	1 061	948	**943**	943	12.51	0.53	**0.00**
	Vdata_la39	1 109	976	**959**	922	20.28	6.18	**4.01**
	Vdata_la40	1 041	968	**955**	955	9.01	1.36	**0.00**

注：Ave.Gap：DSTA、Multi-ppo、DSTA-PPO分别为9.55、4.20、2.58。

在部分算例上 DSTA-PPO 算法测试时的效果略差于 Multi-PPO, 但整体的性能明显优于 Multi-PPO。图 8 给出了算法在 Vdata_la40 上的调度甘特图。

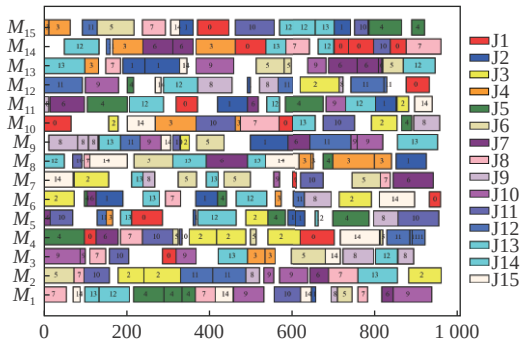


图 8　Vdata_la40 算例调度图

Fig. 8　Vdata_la40 Arithmetic scheduling diagram

5　结　论

本文将深度强化学习算法同状态转移算法结合, 提出了一种基于深度强化学习的状态转移算法, 针对不同规模的 FJSP, 以最小化最大完工时间为调度目标进行了算例测试和算法对比。结果表明: 本文提出的 DSTA-PPO 能够有效地解决 FJSP, 并在大多数测试算例中提升效果显著。经过设计的状态特征和动作空间能够有效地使智能体在经过训练后高效地完成调度操作, 产生合适的调度策略。该算法不足之处在于: 首先 2 种算法的结合增加了整体算法的复杂性, 其次 2 种算法结合后参数增多, 增加了调参和训练的难度。进一步工作将在此基础上继续改进, 优化代码结构, 以提高 DSTA-PPO 的训练效率, 对其动作空间继续优化, 以提高其在更大规模调度问题中的泛化性和鲁棒性。

参考文献（References）

[1] THAMES L, SCHAEFER D. Software-defined cloud manufacturing for industry 4.0[J]. Procedia CIRP, 2016, 52: 12-17.

[2] GOLPÎRA H, KHAN S A R, SAFAEIPOUR S. A review of logistics Internet-of-things: current trends and scope for future research [J]. Journal of Industrial Information Integration, 2021, 22: 100194.

[3] LIU Y K, WANG L H, WANG X V, et al. Scheduling in cloud manufacturing: state-of-the-art and research challenges[J]. International Journal of Production Research, 2019, 57(15-16): 4854-4879.

[4] FANG Y L, PENG C, LOU P, et al. Digital-twin-based job shop scheduling toward smart manufacturing[J]. IEEE Transactions on Industrial Informatics, 2019, 15(12): 6425-6435.

[5] ZHANG Y F, WANG J, LIU S C, et al. Game theory based real - time shop floor scheduling strategy and method for cloud manufacturing[J]. International Journal of Intelligent Systems, 2017, 32(4): 437-463.

[6] GAO K Z, CAO Z G, ZHANG L, et al. A review on swarm intelligence and evolutionary algorithms for solving flexible job shop scheduling problems[J]. IEEE/CAA Journal of Automatica Sinica, 2019, 6(4): 904-916.

[7] CHOI I C, CHOI D S. A local search algorithm for jobshop scheduling problems with alternative operations and sequence-dependent setups[J]. Computers & Industrial Engineering, 2002, 42(1): 43-58.

[8] ZHANG G H, SHAO X Y, LI P G, et al. An effective hybrid particle swarm optimization algorithm for multi-objective flexible job-shop scheduling problem[J]. Computers & Industrial Engineering, 2009, 56(4): 1309-1318.

[9] NOUIRI M, BEKRAR A, JEMAI A, et al. Two stage particle swarm optimization to solve the flexible job shop predictive scheduling problem considering possible machine breakdowns[J]. Computers & Industrial Engineering, 2017, 112: 595-606.

[10] NOUIRI M, BEKRAR A, JEMAI A, et al. An effective and distributed particle swarm optimization algorithm for flexible job-shop scheduling problem[J]. Journal of Intelligent Manufacturing, 2018, 29(3): 603-615.

[11] XING L N, CHEN Y W, WANG P, et al. A knowledge-based ant colony optimization for flexible job shop scheduling problems[J]. Applied Soft Computing, 2010, 10(3): 888-896.

[12] JIANG T H, ZHANG C. Application of grey wolf optimization for solving combinatorial problems: job shop and flexible job shop scheduling cases[J]. IEEE Access, 2018, 6: 26231-26240.

[13] ZHOU X J, YANG C H, GUI W H. State transition algorithm[J]. Journal of Industrial & Management Optimization, 2012, 8(4): 1039-1056.

[14] 董天雪, 阳春华, 周晓君, 等. 一种求解企业员工指派问题的离散状态转移算法[J]. 控制理论与应用, 2016, 33(10): 1378-1388.
DONG T X, YANG C H, ZHOU X J, et al. A novel discrete state transition algorithm for staff assignment problem[J]. Control Theory & Applications, 2016, 33(10): 1378-1388(in Chinese).

[15] 张凤雪, 阳春华, 周晓君, 等. 基于控制周期计算的锌液净化除铜过程优化控制[J]. 控制理论与应用, 2017, 34(10): 1388-1395.
ZHANG F X, YANG C H, ZHOU X J, et al. Optimal control based on control period calculation for copper removal process of zinc solution purification[J]. Control Theory & Applications, 2017, 34(10): 1388-1395(in Chinese).

[16] DU Y, LI J Q, CHEN X L, et al. Knowledge-based reinforcement learning and estimation of distribution algorithm for flexible job shop scheduling problem[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2022, 7(4): 1036-1050.

[17] CHEN R H, YANG B, LI S, et al. A self-learning genetic algorithm based on reinforcement learning for flexible job-shop scheduling problem[J]. Computers & Industrial Engineering, 2020, 149: 106778.

[18] LEI K, GUO P, ZHAO W C, et al. A multi-action deep reinforcement learning framework for flexible Job-shop scheduling problem[J]. Expert Systems with Applications, 2022, 205: 117796.

[19] 张凯, 毕利, 焦小刚. 集成强化学习算法的柔性作业车间调度问题研究[J]. 中国机械工程, 2023, 34(2): 201-207.
ZHANG K, BI L, JIAO X G. Research on flexible job-shop scheduling problems with integrated reinforcement learning algorithm[J]. China Mechanical Engineering, 2023, 34(2): 201-207(in Chinese).

[20] 王凌, 王晶晶. 考虑运输时间的分布式绿色柔性作业车间调度协同群智能优化[J]. 中国科学: 技术科学: 1-15.

A deep reinforcement learning based on discrete state transition algorithm for solving fuzzy flexible job shop scheduling problem

ZHU Jiazheng，WANG Cong，LI Xinkai，DONG Yingchao，ZHANG Hongli*

(School of Electrical Engineering，Xinjiang University，Urumqi 830047，China)

Abstract：The study of the intelligent algorithms for the flexible job shop scheduling issue (FJSP), a scheduling problem with a broad range of application backgrounds, is very relevant both academically and practically. To address FJSP with the objective of minimizing the maximum completion time, this paper proposes a discrete state transfer algorithm based on proximal policy optimization (DSTA-PPO). DSTA-PPO has the following three characteristics. Considering that FJSP requires simultaneous scheduling arrangements for operation sequencing and machine assignment. This state feature can adequately express the current scheduling problem that was designed by combining operation coding and machine coding. Various critical path based search operations have been designed for operation sequencing and machine allocation. Reinforcement learning training is an efficient way to direct intelligence to choose the best search operation to maximize the current scheduling sequence.. The effectiveness of each component of the algorithm is verified through simulation experiments on different datasets. Furthermore, a comparison is conducted with existing algorithms using the objective of minimizing the maximum completion time in the same instances. The comparison results show that the suggested method successfully resolves the flexible job shop scheduling issue by typically achieving shorter completion times.

Keywords：deep learning；reinforcement learning；discrete state transition algorithm；proximal policy optimization；flexible job shop scheduling

Received：2023-04-26；**Accepted**：2023-07-27；**Published Online**：2023-09-12 16：53

URL：link.cnki.net/urlid/11.2625.V.20230911.1743.001

Foundation items：National Natural Science Foundation of China (52267010,62263030); Natural Science Foundation of Xinjiang Uygur Autonomous Region (2022D01C367,2022D01E33)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0213

岸基 GNSS-R 码测高反演模型

宁保蛟[1]，王娜子[1,*]，荆丽丽[1]，高凡[1,2]，孔亚慧[1]，贺匀峤[1]

(1. 山东大学 空间科学研究院，威海 264209；2. 山东省光学天文与日地空间环境重点实验室，威海 264209)

摘 要：全球导航卫星系统反射测量 (GNSS-R) 海面测高作为 GNSS 技术应用领域之一，其凭借全天时、低成本、高时空分辨率等特点在海洋遥感领域占据极大的优势。但受 GNSS 码片宽度的限制，岸基 GNSS-R 码延迟海面测高的最优精度仅为分米级，难以满足大地测量学领域对海面高的精度要求。提出利用傅里叶级数拟合法对码延迟测高结果进行优化，并于山东省威海市开展 GNSS-R 码测高实验进行验证，分别对 QZSS L2C 信号、QZSS L5 信号及 GPS L5 信号的码延迟测高结果进行拟合处理并计算反演精度，对比原始测高结果发现：3 段不同时长的原始测高结果经过傅里叶级数拟合优化后，其均方根误差 (RMSE) 从 70～90 cm 提升至 9～15 cm。在此基础上，利用拟合所得傅里叶级数曲线对未来 24 h 内的海面高进行预测，与验潮站数据相比，预测结果仍保持较高精度，RMSE 为 13～18 cm，相关系数大于 0.97。结果表明所提出的方法不仅能对码延迟测高结果进行优化，能对存在数据缺失的情况进行补充，还能对未来海面高度进行高精度的预测。

关 键 词：岸基；全球导航卫星系统反射测量；码延迟；海平面高度；傅里叶级数拟合

中图分类号：P228.4

文献标志码：A **文章编号**：1001-5965(2025)04-1395-09

由于全球变暖日益加剧，冰川融化导致海平面不断上升，因此对海平面变化的监测和研究越来越重要。海平面高度是衡量海平面变化的重要指标，对人类社会有着重要的研究意义，其观测手段也日益发展。传统监测方法有验潮站和卫星雷达测高等，其优点主要是观测精度较高，但时空分辨率低且成本较高。随着全球卫星导航系统 (Global Navigation Satellite System, GNSS) 的不断发展，GNSS-R (Global Navigation Satellite System-Reflectometry) 技术应运而生，GNSS-R 可以应用在海洋遥感方面，包括海面风场[1-5]、海面测高[6-10]、海面溢油[11-13] 以及海冰监测[14-17] 等不同方向。其中的 GNSS-R 海面测高技术与传统监测手段相比，具有低成本和高时空分辨率等优势。

随着 GNSS-R 技术的发展，国内外学者开展了一系列岸基、空基和星基的 GNSS-R 观测实验，空间分辨率不断提高。其中的岸基 GNSS-R 测高可以为近海海域海面高度监测提供一种新方式。由于岸基 GNSS-R 成本低且实验便捷，因此可以很好地解决传统监测方法成本高等问题，同时开展岸基 GNSS-R 研究还可以辅助星载观测。岸基 GNSS-R 测高是将观测设备架设在观测海域附近的海岸上来实现对海面高度的监测。根据观测量的不同，GNSS-R 测高法主要有两种：一种是基于多路径信噪比的海面测高方法，即 GNSS-IR (GNSS Interferometric Reflectometry) 测高法。GNSS-IR 测高法仅可

收稿日期：2023-04-26；**录用日期**：2023-07-21；**网络出版时间**：2023-08-09 16:33

网络出版地址：link.cnki.net/urlid/11.2625.V.20230809.1442.001

基金项目：山东省重点研发计划 (重大科技创新工程)(2021ZDSYS01)；国家自然科学基金 (41704017,41604003)；山东省自然科学基金 (ZR2022MD046)

* **通信作者**. E-mail：wnz@sdu.edu.cn

引用格式：宁保蛟，王娜子，荆丽丽，等. 岸基 GNSS-R 码测高反演模型 [J]. 北京航空航天大学学报，2025，51（4）：1395-1403.

NING B J，WANG N Z，JING L L，et al. The retrieval model of shore-based GNSS-R code altimetry [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1395-1403 （in Chinese）.

在卫星高度角较低时观测，且整个观测弧段得到一个反演结果，因此造成其有效观测值的数目大幅减少、时间分辨率降低。另一种是基于相位差的海面测高方法，包括载波相位法和码相位法。在基于相位差的测高方法中，当海面波动较大时，接收机无法稳定跟踪载波相位，然而码相位测高方法可以相对不受海面环境的影响。因此本文致力于开展岸基GNSS-R码相位测高方法研究。

1993年，Martin-Neira[18]提出了Paris(Passive Reflectometry and Interferometry System)概念，证实了GNSS-R的可行性。1997年9月，Martin-Neira团队在荷兰基于双天线码相位测高方法首次开展岸基海面测高实验，最终得到5 s内观测结果误差约为3.3 m，测高精度是L1C/A码片长度的1%[19]。1999年，Treuhaft等[20]通过相位测量的方法得到了误差为2 cm的湖面测高结果。2002年9月，王鑫等[21]在中国福建进行了我国首个岸基GNSS-R实验，实验最终得到有效波高参数。2010年，白伟华等[22]通过开环相位测高技术对湖面高度进行反演，其测高精度可达到厘米级。Carreno-Luengo等[8]在2012年实验通过码测高得到了亚米级的海面高结果。2015年，Zhang等[23]在浙江大洋山海域开展基于北斗二代信号的岸基海面测高研究，其结果表明测高精度达到了亚米级。2021年，储倜等[24]开展岸基GNSS-R海面测高实验，实验表明基于QZSS(Quasi-Zenith Satellite System) L1 C/A码和L1C码的海面测高精度最优可达0.63 m和0.4 m。2022年，侯金华等[25]发现基于B1C码的GNSS-R海面测高最优精度达到分米级，据10 s观测数据的平滑结果显示：单颗卫星最优测高结果的RMSE为0.634 m，多卫星观测值加权组合最优结果的RMSE为0.538 m。

尽管GNSS-R码测高可提供高时空分辨率和稳定的测高结果，但受GNSS码片宽度限制和测量误差的影响，经过码延迟计算的岸基GNSS-R测高结果精度在米级，最优可达分米级[26,27]。这样的测高精度难以满足大地测量学研究中对海面高度精度的要求，例如，从海面高度获取海洋重力场数据往往需要海面高度的精度为厘米级。文献[6,23,24]使用滑动平均法对海面高度结果进行后处理，但其结果不能如实准确地反映海平面变化。因此，本文致力于研究GNSS-R码测高反演模型来提高海面高精度。在山东省威海市开展3次GNSS-R码测高实验，通过提出的傅里叶级数拟合法对原始测高结果进行后处理。结果发现，该方法不仅能够大幅度提高GNSS-R码测高反演精度，弥补数据缺失，还能对海平面高度进行预测。

1 GNSS-R测高

1.1 GNSS-R测高原理

GNSS-R码相位海面测高技术的关键是分别利用上视和下视的接收天线获取直、反射中频信号、通过自主研发的软件接收机[6]处理中频信号得到直、反射信号的码相位路径延迟、结合岸基GNSS-R海面测高模型由码相位路径延迟计算得到反射面到直射接收天线的高度。再通过直射天线获取的信号利用传统定位方法如精密单点定位(PPP)等得到直射天线到参考椭球面的高度H，最后计算得到海平面高度值。

GNSS-R码相位海面测高实验的几何结构如图1所示：h是反射天线相位中心到海面的垂直高度、θ为GNSS卫星高度角、d代表直射天线和反射天线的相位中心距离、$\Delta\rho$是直射信号和反射信号之间的路径延迟(上述100个路径延迟的中值)、ε为海面高度测量误差，根据图1所示的几何关系可得

$$H_{\mathrm{SSH}} = H - \frac{\Delta\rho - d\sin\theta}{2\sin\theta} + \varepsilon \tag{1}$$

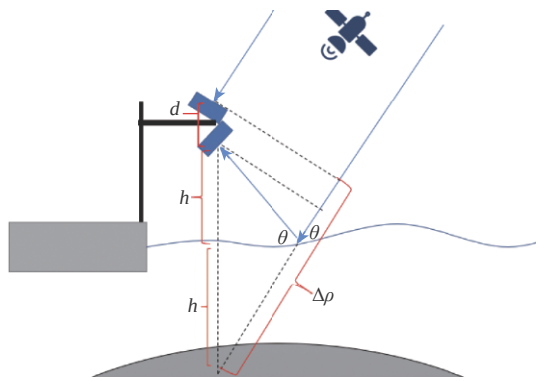


图1 岸基GNSS-R码延迟海面测高模型

Fig. 1 Shore-based GNSS-R code delayed sea level altimeter model

文献[28]研究表明，高度角对GNSS-R测高结果有一定的影响，且在低高度角时误差较大。因此一般采用高高度角数据进行计算，本文数据处理限制高度角设置为30°。为保证反射信号来自海面，设置观测方位角为140°～300°。

其中，GNSS-R信号数据处理部分简述如下：直射和反射天线接收到的右旋直射信号和左旋反射信号，经过射频前端的滤波器和放大器滤波放大后，再下变频为中频信号，后经模数转换为数字中频信号。数字中频信号传至软件接收机的基带信号处理模块实现下变频，剥离载波后成为基带信号。与本地复制码进行相干积分后输出相应的多普勒频移和码相位信息，最后计算得到反射信号相

对于直射信号的码延迟$\Delta\rho$。本文软件接收机采用的相干积分时长为 10 ms, 再将 1 s 内所得的 100 个路径延迟取中值作为该秒的最终路径延迟。详细描述见文献 [28]。

1.2　傅里叶级数拟合法后处理 GNSS-R 反演高度

通过上述 GNSS-R 码测高得到的海面高度最优可达分米级。在海面高度研究领域, 上述所得分米级的测高结果难以满足大地测量领域的需求。同时, 上述测高结果的时间分辨率是由原始中频数据的采样率和接收机对数据处理时的相干积分共同决定的。利用本文的接收机处理, 得到的海面高时间分辨率为 1 s。而每 1 s 的海面高实际上是由 10 ms 的信号经由数据处理得到的。10 ms 的反射信号会受到海面瞬时浪高、瞬时波高等的影响进而对海面高结果产生误差。另外, 受 GNSS 卫星分布以及限制高度角的影响, GNSS-R 测高结果并不是每秒都可得到, 甚至可能会出现较长时间的数据缺失。为了对 GNSS-R 原始码测高结果进一步处理, 弥补数据缺失以及提高 GNSS-R 海面测高精度, 我们提出利用傅里叶级数拟合法对 GNSS-R 原始测高数据进行建模处理。该方法的基本原理简述如下。

傅里叶级数是一种基于函数形态和函数参数拟合的方法, 通过将复杂的函数分解为一系列简单的正弦周函数的级数相加, 可将离散数据进行拟合、插值和滤波等, 可大大提高函数拟合的精度。其表达式如下:

$$f(x)=\sum_{i=0}^{m}[a_i\cos(ixw)+b_i\sin(ixw)] \tag{2}$$

式中: x为待拟合数据; i 为某一阶数; m 为总阶数; w为频率; a_i和b_i为系数。

而在经典潮汐调和分析模型中, 海平面高度可以认为是一系列余弦函数线性叠加的结果, 每一个余弦函数都代表了一个分潮:

$$Z(t)=S_0+\sum_{j=1}^{J}(H_j\cos(\sigma_j t-g_j)) \tag{3}$$

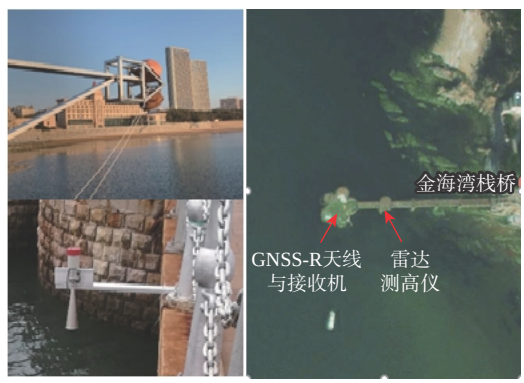
式中: $Z(t)$是时刻 t 观测到的水位; S_0是平均海平面; σ_j、H_j、g_j分别为第 j 分潮对应的角频率、振幅和迟角。

若考虑经由 1.1 节计算得到的 GNSS-R 码测高原始结果, 将其视为离散数据x, 代入式(2), 选择阶数 m, 即可得到拟合后的海面高度$f(x)$。对比式(2)和式(3)可以发现, 若将式(3)展开, 可化为式(2)的形式, 也就是说, 可以用傅里叶级数(式(2))来表示 GNSS-R 海面高度。因此, 本文提出用傅里叶级数拟合 GNSS-R 测高结果, 以提高 GNSS-R 海面高观测精度。

2　GNSS-R 实验概况

2.1　实验场景

为验证本文理论内容, 分别采用了不同时长的 3 段实验数据: 第 1 段于 2020 年 11 月 5 日在山东威海一栈桥处(N:37°32′ E:122°2′)开展双天线岸基 GNSS-R 海面测高实验采集到的 GPS L5 近 8 h 数据。第 2 段于 2021 年 3 月 30 日在同地采集到的 QZSS L2C 频点 6 h 数据, 两次实验场地如图 2 所示。第 3 段于 2023 年 5 月 27 日距离栈桥直线距离约 3 km 外的小石岛采集到的 QZSS L5 24 h 数据, 实验地点如图 3 所示。



(a) 接收和反射天线及高精度雷达测高仪　　(b) 栈桥俯视图

图 2　2020 年 11 月 5 日及 2021 年 3 月 30 日实验环境及实验地点

Fig. 2　Experimental environment and experimental location on 5 November 2020 and 30 March 2021



(a) 实验仪器架设　　(b) 实验地区

图 3　2023 年 5 月 27 日实验环境及实验地点

Fig. 3　Experimental environment and experimental location on 27 May 2023

从图 2 中可以看出, GNSS-R 天线以及接收机架设于深入海岸线约 100 m 的海边栈桥, 距其约 20 m 的位置处安置了雷达测高仪。图 3 中距离该实验地点 5 m 处有一验潮站, 采用验潮站数据作为比对

数据。该实验地点与栈桥直线距离约为3 km。

3次实验的实验场地附近没有遮挡，采集数据时海面较为平静。采集到的中频数据通过笔记本电脑保存到移动硬盘中，并用1.2节所描述的软件接收机进行进一步处理，最终获得海面高度结果。

2.2 实验地点潮汐变化

中国渤海、东海、黄海的多数地点为半日潮型，即一个太阴日内出现两次高潮和两次低潮，其涨潮过程和落潮过程的时间也几乎相等约为6 h，本实验海域即为半日潮型。

表1是GNSS-R码相位海面测高实验地点附近海域在3段实验期间的潮汐变化表。

表1 3次实验期间实验地点潮汐变化表

Table 1 Table of tidal change at experimental site during the three experiments

潮汐	潮时			潮高(峰值)/cm		
	第1段	第2段	第3段	第1段	第2段	第3段
干潮	05:35	04:29	00:40	37	45	84
满潮	11:27	10:17	07:26	222	209	239
干潮	17:53	16:06	14:26	42	55	65
满潮	23:48	22:25	20:15	221	242	196

从表1中可知：实验期间每天出现2次满潮和2次干潮，满潮与干潮时的潮高峰值第1段相差约为180 cm、第2段相差约为150 cm、第3段相差约为170 cm。为了验证所述方法的正确性，采用了包含不同海面高变化过程的3段GNSS-R数字中频信号进行处理：第1段18:00−24:00时共6 h仅包含涨潮阶段的数据，第2段14:00及22日包含部分落潮及涨潮过程的数据，第3段包含完整的涨、落潮阶段的24 h数据。

3 实验结果分析

通过1.1节的原理处理3段数字中频信号得到反射信号与直射信号的路径差，再结合海面测高反演几何模型反演得到海平面高度 h_{SSH}。通过处理QZSS GEO卫星的L2C数据、GPS L5数据以及QZSS L5数据得到海平面高度结果以此验证GNSS-R码相位海面测高实验的准确性与可实用性。

图4中黑色的散点代表GNSS-R码测高中经由1.1节内容处理得到的原始海面测高结果，蓝线代表比对数据(雷达测高仪或者验潮站数据)。从图3可以看出：QZSS GEO卫星L2C码和L5码测高结果与比对数据相符合，残差值主要集中在±1 m之间。GPS L5码测高结果为当天可获取的GPS MEO卫星PRN10、PRN23、PRN26和PRN27卫星数据反演所得，受到实验时高度角和方位角选

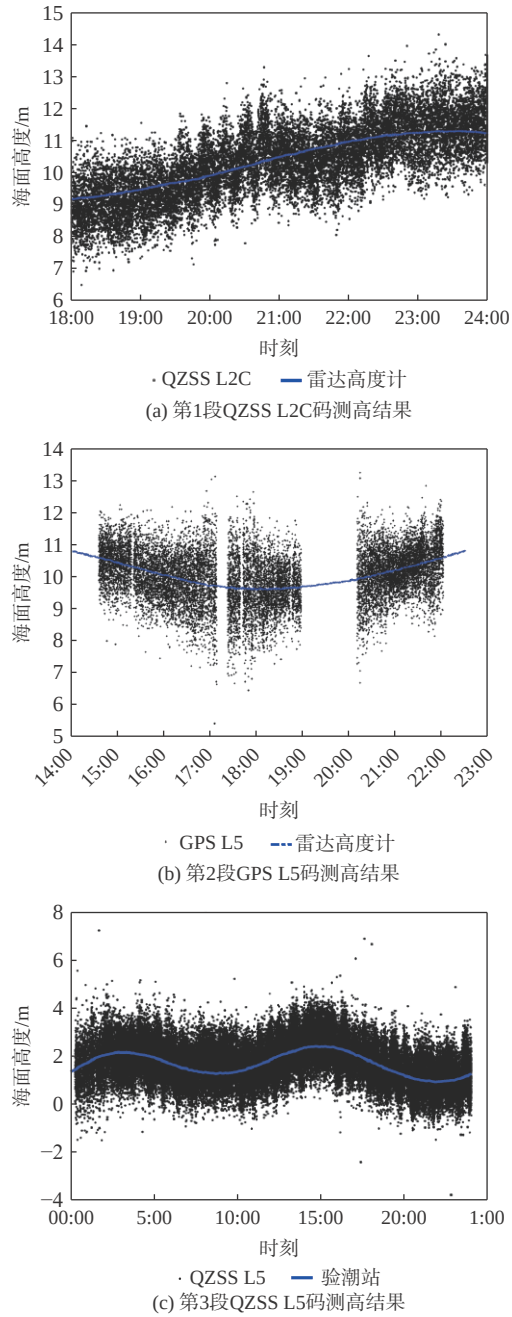


图4 3段数据码测高结果与海面真实高度对比

Fig. 4 Sea levels derived from GNSS-R code-delay altimetry of the three data codes are compared with the real sea surface height

取的限制，当天GPS L5码仅有4颗可见卫星，其测高结果并不能覆盖全天时观测。通过3段反演的海面高度数据可以看出海平面涨潮幅度约2 m，这与当地潮汐变化表的情况基本一致。最后，经过计算得到3段GNSS-R码测高结果的精度如表2所示。由表2可知，整体而言，GNSS-R码测高精度为70～90 cm，GEO卫星测高结果均值偏差较小，而MEO卫星结果均值偏差较大。

3.1 傅里叶级数拟合测高结果

由式(2)可知，傅里叶级数随着阶数 m 的取值不同，拟合结果也不同，因此对于3段数据进行了

表 2　3 段实验数据的 GNSS-R 码测高精度统计

Table 2　Statistical analysis of the three experimental data for GNSS-R code-delay altimetry

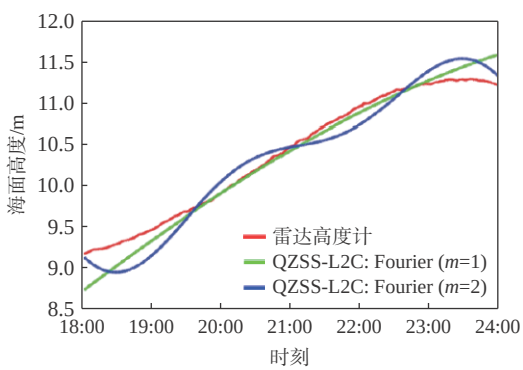
数据类型	均值/m	RMSE/m
QZSS L2C	−0.046 7	0.761 8
GPS L5	0.287 7	0.808 3
QZSS L5	−0.043 8	0.860 6

不同阶数的傅里叶级数拟合。图 5～图 7 为 3 段数据采用不同阶数傅里叶级数拟合 GNSS-R 测高结果与比对数据(雷达高度计或者验潮站数据)的比较, 表 3 总结了 3 段数据的比较精度。

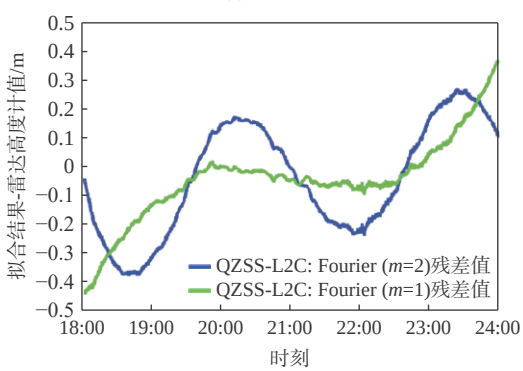
对于第 1 次获取的近 6 h 的 QZSS L2C 码测高结果进行 1 阶和 2 阶傅里叶级数拟合, 拟合曲线与雷达高度计结果如图 5 所示。从图中可以看出, 相对于 2 阶傅里叶级数拟合, 1 阶级数拟合结果与雷达高度计数据更符合。

表 3 中加粗数据表示最优匹配统计结果。从表 3 可知, 1 阶级数拟合结果较 2 阶精度更高, 均值偏差为−0.045 1 m, 均方根误差为 0.148 1 m, 相关系数为 99.33 %。

对于第 1 段获取的近 8 h 的 GPS L5 码测高结果进行 1 阶和 2 阶傅里叶级数拟合, 拟合曲线与雷



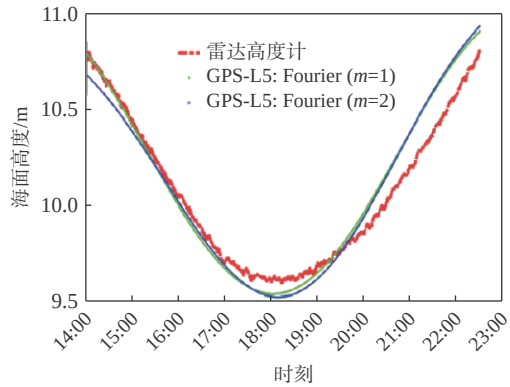
(a) 海面高度值



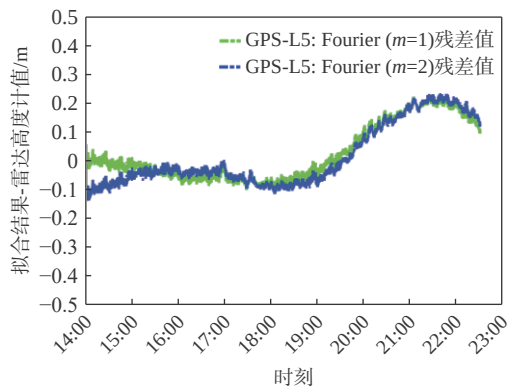
(b) 拟合结果与雷达高度计数据的差值

图 5　QZSS L2C1/2 阶傅里叶级数拟合值与雷达高度计数据比较

Fig. 5　First- and Second-Fourier series fitting data of QZSS GEO L2C



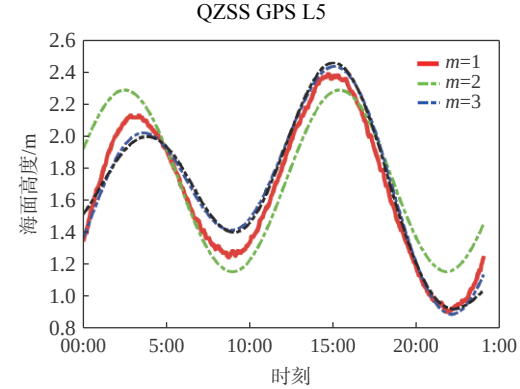
(a) 海面高度值



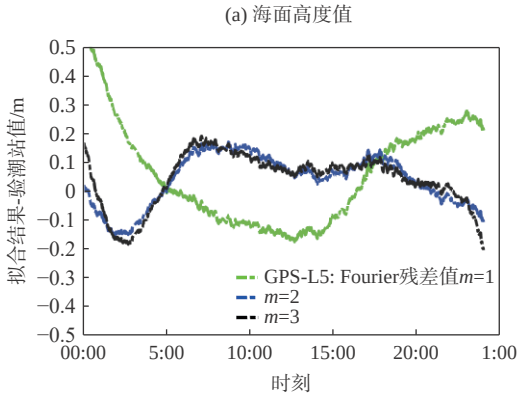
(b) 拟合结果与雷达高度计数据的差值

图 6　GPS L51/2 阶傅里叶级数拟合值与雷达高度计数据比较

Fig. 6　First- and Second-Fourier series fitting data of QZSS GPS L5



(a) 海面高度值



(b) 拟合结果与验潮站数据的差值

图 7　1 阶、2 阶和 3 阶傅里叶级数拟合的 QZSS GEO L5 码测高结果与验潮站数据对比

Fig. 7　First-, Second- and Third- Fourier series fitting data of QZSS GEO L5

表3 3段实验数据的傅里叶级数拟合结果与比对数据对比精度统计

Table 3 Statistical analysis of the Fourier series fitting results comparing with in-situ data

数据类型	均值/m	RMSE/m	R^2/%
QZSS L2C(m=1)	**−0.045 1**	**0.148 1**	**99.33**
QZSS L2C(m=2)	−0.042 9	0.196 1	98.14
GPS L5(m=1)	**0.025 6**	**0.101 1**	**98.07**
GPS L5(m=2)	0.013 8	0.107 7	97.36
QZSS L5(m=1)	0.061 0	0.186 1	91.60
QZSS L5(m=2)	**0.039 2**	**0.096 1**	**97.99**
QZSS L5(m=3)	0.041 7	0.101 6	97.76

达高度计结果如图6所示。从图6可以看出，2阶傅里叶级数拟合和1阶级数拟合结果相差不大，但在14:00-16:00期间，1阶傅里叶级数与雷达高度计数据更符合。从表3可知，1阶级数拟合结果较2阶精度稍高，均值偏差为0.025 6 m，均方根误差为0.101 1 m，相关系数为98.07 %。

第3段QZSS L5数据获得的24 h海面高结果，可以描述一整个潮汐涨落的过程。图7展示了对该GNSS-R反演结果进行1、2和3阶傅里叶级数拟合所得的曲线结果及其与验潮站数据对比的残差值。同时，图7(a)也画出了验潮站数据以进行比较。由图4可知，与验潮站峰值相比，当m=1时，拟合波形峰值存在超前/滞后现象且存在较大的偏差；而m=2和3时峰值存在较小的超前/滞后现象，但相对于m=3，m=2时偏差更小。因此取m=2时拟合波形与验潮站数据可实现最优匹配。由表3可知，m=2时，傅里叶级数拟合结果相对验潮站值的精度最高，均值偏差为0.039 2 m，均方根误差为0.096 1 m，相关系数为97.99 %。

综合3段数据反演结果可知，最优傅里叶级数拟合的阶数并不是固定的，对于较短时间的6～8 h GNSS-R海面高结果来说，1阶傅里叶级数可实现最优匹配，对于24 h的海面高结果，2阶傅里叶级数可实现最优匹配。这可能与数据量的长度有关：较短时间的数据难以完全反映海面高的周期性变化，采用较低阶拟合即可实现最优匹配。因为GNSS-R测高结果存在星间偏差，对测高结果建立模型，模型的拟合系数需要根据不同导航卫星等进行调整。根据文中3段数据的分析，时间长短对拟合时最优阶数的选取有影响，因此，可以推测对同一颗卫星结果不同观测时间段数据进行拟合模型参数可能也有所差别。综合上述分析及比较表2和表3数据可得：GNSS-R码测高结果经傅里叶级数拟合后，与雷达高度计或者验潮站数据相比，精度由70～90 cm提升为9～15 cm，且相关系数大于0.97。因此，傅里叶级数拟合法不仅能大幅度提高码延迟海面测高精度，经拟合所得的海面高结果与比对数据相似且变化平滑，能如实准确地反映出海平面变化。另外，由于高度角和方位角限制，原始GPS L5码测高结果中存在有短期的数据缺失(如图4(b)17:00-17:30以及19:00-20:00区间)，而采用傅里叶级数拟合后，该段数据结果仍可通过拟合曲线所得(如图6所示)。因此，傅里叶级数拟合方法也可补充原始观测结果中的数据缺失问题。

3.2 利用傅里叶级数拟合曲线进行海面高预测

经由3.2节获取的傅里叶级数拟合曲线的自变量为时间，结果为海面高，因此，为进一步验证本文方法的正确性，采用QZSS L52阶傅里叶级数拟合曲线对海面高结果进行预测，并将预测值与验潮站数据进行比较，比较结果如图8和表4所示。图8显示：经由第3段GNSS-R码测高数据建立的2阶傅里叶级数拟合曲线在未来24 h内(2023年5月28日)对海面高的预测结果与验潮站数据变化一致。预测结果在前12 h内存在有较大均值偏差，后12 h偏差较小，但整体海面高变化与验潮站结果保持一致。由表4可知，预测结果精度较高，均方根误差维持在13～18 cm，相关系数大于0.95。所得结果表明：傅里叶级数拟合法不仅可以提升GNSS-R码测高的精度，还可以对未来24 h内海面

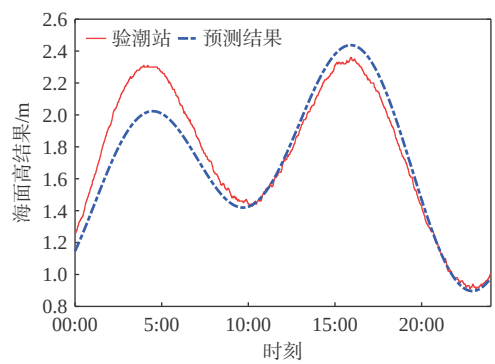


图8 2023年5月28日2阶傅里叶级数拟合曲线计算的海面高度与验潮站数据比较

Fig. 8 Comparison between sea levels calculated from Second-order Fourier series fitting curve and tide gauge data on May 28th, 2023

表4 QZSS L5模型预测海面高与验潮站结果比较

Table 4 Statistical analysis of the sea levels from Fourier series fitting model comparing with in-situ data

时间	均值/m	RMSE/m	R^2/%
未来1 h	−0.130 7	0.132 1	99.64
未来12 h	−0.151 2	0.185 6	97.15
未来24 h	−0.047 6	0.141 9	95.16

高度结果进行预测, 且预测精度较高。

4 结　论

伴随着 GNSS 的高速发展, GNSS-R 的重要性日益明显, 其具有全天时、低成本、高时空分辨率等优势, 但受 GNSS 码片宽度限制和测量误差的影响, 岸基 GNSS-R 码测高原始精度难以满足大地测量学研究中对海面高精度的要求。

本文通过岸基 GNSS-R 码相位海面测高实验, 发现该技术原始测高结果精度为分米级。提出利用傅里叶级数拟合法建立模型对原始测高结果进行后处理, 发现其能大幅度提高 GNSS-R 码测高反演精度并能补充 GNSS-R 测高缺失数据, 且拟合曲线能对未来海面高度进行预测。本文的主要结论如下。

1) 通过对岸基 GNSS-R 码测高 3 段包含不同海面高变化过程的数据进行处理发现: 原始测高结果的 RMSE 为 70～90 cm, 测量精度较低, 且由于高度角和方位角的限制, MEO 卫星结果数据存在缺失。经过傅里叶级数拟合法建立海面高反演模型后 RMSE 提升为 9～15 cm, 且相关系数大于 0.97。海面高精度得到大幅提升, 且可以补充原始测高结果中的少量数据缺失。

2) 文中傅里叶级数拟合的最优阶数会随着拟合数据的长短有变化, 这可能是因为较短的数据难以显示海面高的周期性变化。

3) 利用所建立的傅里叶级数拟合曲线可以实现对未来 24 小时内海面高度值的预测, 且预测精度较高。

本文所提出的傅里叶级数拟合法可以较好地拟合用正、余弦函数表达的数据序列, 因此其不仅可以与 GNSS-R 测高数据相结合进而提高精度, 也可应用到卫星测高等数据的优化中。

致　谢 感谢威海市海洋与渔业监测减灾中心 (http://www.weihai.gov.cn/art/2023/3/17/art_125777_3487711.html) 提供的小石岛验潮站数据。

参考文献 (References)

[1] VALENCIA E, ZAVOROTNY V U, AKOS D M, et al. Using DDM asymmetry metrics for wind direction retrieval from GPS ocean-scattered signals in airborne experiments[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 52(7): 3924-3936.

[2] RUF C S, GLEASON S, MCKAGUE D S. Assessment of CYGNSS wind speed retrieval uncertainty[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 12(1): 87-97.

[3] RUF C S, BALASUBRAMANIAM R. Development of the CYGNSS geophysical model function for wind speed[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 12(1): 66-77.

[4] FOTI G, GOMMENGINGER C, JALES P, et al. Spaceborne GNSS reflectometry for ocean winds: first results from the UK TechDemoSat-1 mission[J]. Geophysical Research Letters, 2015, 42(13): 5435-5441.

[5] CLARIZIA M P, RUF C S, JALES P, et al. Spaceborne GNSS-R minimum variance wind speed estimator[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(11): 6829-6843.

[6] GAO F, XU T H, MENG X Y, et al. A coastal experiment for GNSS-R code-level altimetry using BDS-3 new civil signals[J]. Remote Sensing, 2021, 13(7): 1378.

[7] ZHANG G D, XU Z C, WANG F, et al. Evaluation and correction of elevation angle influence for coastal GNSS-R ocean altimetry[J]. Remote Sensing, 2021, 13(15): 2978.

[8] CARRENO-LUENGO H, PARK H, CAMPS A, et al. Submeter ocean altimetry with GPS L1 C/a signal[C]// 2012 IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2012: 7071-7074.

[9] SOISUVARN S, JELENAK Z, SAID F, et al. The GNSS reflectometry response to the ocean surface winds and waves[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 9(10): 4678-4699.

[10] RUFFINI G, SOULAT F, CAPARRINI M, et al. The Eddy Experiment: accurate GNSS-R ocean altimetry from low altitude aircraft[J]. Geophysical Research Letters, 2004, 31(12): 261-268.

[11] 吴军, 张波, 洪学宝, 等. 基于北斗卫星反射信号的海面溢油探测方法及试验[J]. 北京航空航天大学学报, 2017, 43(1): 193-199.
WU J, ZHANG B, HONG X B, et al. Sea surface oil spill detection method using BeiDou satellite reflected signal and its experiment[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(1): 193-199(in Chinese).

[12] VALENCIA E, CAMPS A, RODRIGUEZ-ALVAREZ N, et al. Using GNSS-R imaging of the ocean surface for oil slick detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2012, 6(1): 217-223.

[13] LI C, HUANG W M, GLEASON S. Dual antenna space-based GNSS-R ocean surface mapping: oil slick and tropical cyclone sensing[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 8(1): 425-435.

[14] ZHU Y C, TAO T Y, LI J Y, et al. Spaceborne GNSS-R for sea ice classification using machine learning classifiers[J]. Remote Sensing, 2021, 13(22): 4577.

[15] ZHANG Y, MENG W T, GU Q M, et al. Detection of Bohai Bay Sea ice using GPS-reflected signals[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 8(1): 39-46.

[16] YAN Q Y, HUANG W M, MOLONEY C. Neural networks based sea ice detection and concentration retrieval from GNSS-R delay-Doppler maps[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(8): 3789-3798.

[17] ALONSO-ARROYO A, ZAVOROTNY V U, CAMPS A. Sea ice detection using U. K. TDS-1 GNSS-R data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(9): 4989-5001.

[18] MARTIN-NEIRA M. A passive reflectometry and interferometry

system (PARIS): Application to ocean altimetry[J]. ESA Journal, 1993, 17(4): 331-355.

[19] MARTIN-NEIRA M, CAPARRINI M, FONT-ROSSELLO J, et al. The PARIS concept: an experimental demonstration of sea surface altimetry using GPS reflected signals[J]. IEEE Transactions on Geoscience and Remote Sensing, 2001, 39(1): 142-150.

[20] TREUHAFT R N, LOWE S T, ZUFFADA C, et al. 2-cm GPS altimetry over crater lake[J]. Geophysical Research Letters, 2001, 28(23): 4343-4346.

[21] 王鑫, 孙强, 张训械, 等. 中国首次岸基 GNSS-R 海洋遥感实验[J]. 科学通报, 2008, 53(5): 589-592.
WANG X, SUN Q, ZHANG X X, et al. First China ocean reflection experiment using coastal GNSS-R[J]. Chinese Science Bulletin, 2008, 53(5): 589-592(in Chinese).

[22] 白伟华, 孙越强, 朱光武, 等. 利用岸基 GNSS-R 信号反演湖面高度[C]//第一届中国卫星导航学术年会, 2010: 496-502.
BAI W H, SUN Y Q, ZHU G W, et al. Lake surface height retrieve using coastal GNSS-R signals[C]//The first China Satellite Navigation Conference, 2010: 1646-1652(in Chinese).

[23] ZHANG Y, TIAN L M, MENG W T, et al. Feasibility of code-level altimetry using coastal BeiDou reflection (BeiDou-R) setups[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(8): 4130-4140.

[24] 储倜, 贺凯飞, 高凡, 等. 基于 QZSS L1 信号的岸基 GNSS-R 码延迟海面测高[J]. 北京航空航天大学学报, 2022, 48(4): 691-697.
CHU T, HE K F, GAO F, et al. Coastal GNSS-R ocean altimetry based on code delay of QZSS L1 signal[J], Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(4): 691-697(in Chinese).

[25] 侯金华, 贺凯飞, 高凡, 等. 岸基 BDS-R 海面测高及其观测值加权方法[J]. 北京航空航天大学学报, 2024, 50(3): 1015-1026.
HOU J H, HE K F, GAO F, et al. Shore-based BDS-R sea surface altimetry and weighting method of its observed values[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(3): 1015-1026(in Chinese).

[26] 孟馨悦. GNSS-R 实时软件接收机设计及其海陆应用[D]. 济南: 山东大学, 2022.
MENG X Y. Design of GNSS-R real-time software receiver and its land and sea application[D]. Jinan: Shandong University, 2022 (in Chinese).

[27] 贺匀峤. 高精度 GNSS-R 海面测高基带信号处理研究[D]. 济南: 山东大学, 2022.
HE Y Q. Research on baseband signal processing of high precision GNSS-R sea surface altimetry[D]. Jinan: Shandong University, 2022 (in Chinese).

[28] WANG N Z, HE K F, GAO F, et al. Analysis of GNSS-R code-level altimetry using QZSS C/A, L1C, and BDS B1C signals and their combinations in a coastal experiment[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16: 4549-4564.

The retrieval model of shore-based GNSS-R code altimetry

NING Baojiao[1]，WANG Nazi[1,*]，JING Lili[1]，GAO Fan[1,2]，KONG Yahui[1]，HE Yunqiao[1]

(1. Institute of Space Sciences，Shandong University，Weihai 264209，China；
2. Shandong Key Laboratory of Optical Astronomy and Solar-Terrestrial Environment，Weihai 264209，China)

Abstract：With the continuous innovation of GNSS technology, its application fields continue to expand. Because of its all-weather capabilities, low cost, and high spatiotemporal resolution, Global Navigation Satellite System-Reflectometry (GNSS-R) sea surface altimetry has many benefits in the field of ocean remote sensing. However, due to the limitation of GNSS chip width, the optimal accuracy of shore-based GNSS-R code-delay sea level measurements is only decimeter level, which makes it difficult to meet the accuracy requirements of Geodesy for sea surface height. In order to maximize the GNSS-R code-delay altimetry results, this research suggested using the Fourier series fitting method. To validate the suggested approach, GNSS-R code altimetry experiments were conducted in Weihai City, Shandong Province. The code-delay altimetry results of QZSS L2C signal, QZSS L5 signal and GPS L5 signal are respectively fitted and processed, and the inversion accuracy is calculated. When compared to the original altimetry findings, it is discovered that the RMSEs of the three original altimetry values with varying lengths have improved from 70-90 cm to 9-15 cm following Fourier series fitting. On this basis, the sea surface heights in the next 24 hours are predicted using the fitted Fourier series curve. Compared with the tide gauge data, the prediction results still maintain high accuracy, with RMSE of 13-18 cm and correlation a coefficient greater than 0.97. The above results indicate that the method proposed in this article can not only optimize the code-delay height measurement results and supplement the missing data, but also make high-precision predictions for future sea levels.

Keywords：shore-based; GNSS-R; code delay; sea surface height; Fourier series fitting

Received：2023-04-26；**Accepted**：2023-07-21；**Published Online**：2023-08-09 16：33

URL：link.cnki.net/urlid/11.2625.V.20230809.1442.001

Foundation items：Key Research and Development Program of Shandong Province (Major Technological Innovation Project) (2021ZDSYS01); National Natural Science Foundation of China (41704017,41604003); Shandong Provincial Natural Science Foundation (ZR2022MD046)

* **Corresponding author**. E-mail：wnz@sdu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0215

融合 LBP 与并行注意力机制的微表情识别方法

李帅超[1]，李明泽[1]，孙嘉傲[1]，卢树华[1, 2, *]

(1. 中国人民公安大学 信息网络安全学院，北京 102600；2. 公安部安全防范技术与风险评估重点实验室，北京 102600)

摘 要：针对面部微表情变化强度弱、背景噪声干扰及特征区分度较小等问题，提出了一种融合 LBP 与并行注意力机制的微表情识别网络。该网络将 RGB 图像输入密集连接改进的 Shuffle Stage 分支提取面部全局特征，增强上下文语义信息关联；将 LBP 图像输入多尺度分层卷积神经网络构成的局部纹理特征分支，提取细节信息；双分支特征提取后，在网络后端引入并行注意力机制提高特征融合能力，抑制背景干扰，专注微表情特征兴趣区域；所提方法在 CASME、CASME II 和 SMIC 等 3 个公开数据集上进行了测试，识别准确率分别达到了 85.18%、74.53% 和 81.19%；实验结果表明，所提方法有效提高了微表情识别准确率，优于当前诸多先进方法。

关 键 词：微表情识别；密集连接；Shuffle Stage 分支；多尺度分层卷积；并行注意力机制
中图分类号：TP391
文献标志码：A **文章编号**：1001-5965(2025)04-1404-11

微表情是一种局部的、短暂的、微弱的面部表情，持续时间只有 1/25～1/5 s[1]，是人们无意识做出的面部动作，可以反映人物内心的真实情感。早在 1966 年 Haggard 等[2]最先发现微表情，后来心理学家 Ekman 等[3]对这种微弱的面部动作做出了定义并开发一套微表情训练工具(micro expression training tool, METT)。与宏表情不同的是，微表情是人们在潜意识下不由自主做出的表情，且难以被抑制住，所以微表情具有较强的客观性和真实性，在公安审讯、测谎、临床诊断等方面应用前景较为广阔。人工识别微表情准确率低且费时，因为微表情难以被一般人类肉眼发现，须经受专业训练的人员才有较大概率识别[4]。为了帮助人们更好地认识微表情，Ekman 等[5]开发出面部动作编码系统(facial action coding system, FACS)，将面部肌肉运动定义为运动单元(action unit, AU)。然而，人工进行微表情识别仍然较为困难，随着计算机技术的不断提高，人们试图运用计算机视觉技术进行微表情自动识别。

传统的微表情识别一般基于图片，通常采用手工设计并提取特征，包括局部二值模式 (linear binary pattern, LBP)[6]、光流直方图(histograms of oriented optical flow, HOOF) 和梯度直方图 (histogram of oriented gradient, HOG)。早期，这些方法在一些数据集上都取得了较好的结果[7-10]。近年来，随着深度学习技术在计算机视觉领域中的深入发展，研究人员开始尝试将其应用于微表情识别，与传统特征提取方法相比，深度学习方法可以得到更高层次的语义信息，能够显著提高人脸微表情识别的准确率。

然而，微表情识别是一项特殊的任务，由于微表情具有强度低、数据集样本类别不平衡和表情特征相似度大等特点，其识别准确率有待进一步提高。受文献 [11-14] 的启发，本文以微表情样本容量小和上下文语义信息相似度高为起点，尝试在有限的数据量下，提高样本利用率，增强网络模型学习特征的能力，考虑到网络架构的特征提取性能和计算复杂度等因素，设计 RGB 全局特征提取分支

收稿日期：2023-04-27；**录用日期**：2023-06-30；**网络出版时间**：2023-07-13 15:10
网络出版地址：link.cnki.net/urlid/11.2625.V.20230713.1227.001
基金项目：中国人民公安大学安全防范工程双一流创新研究专项 (2023SYL08)
* **通信作者**. E-mail：lushuhua@ppsuc.edu.cn

引用格式：李帅超，李明泽，孙嘉傲，等. 融合 LBP 与并行注意力机制的微表情识别方法 [J]. 北京航空航天大学学报，2025，51（4）：1404-1414. LI S C，LI M Z，SUN J A，et al. A micro expression recognition method integrating LBP and parallel attention mechanism [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（4）：1404-1414 （in Chinese）.

和 LBP 局部特征提取分支, 引入并行注意力关注微表情兴趣区域, 提出一种融合 LBP 与并行注意力机制的微表情识别网络, 并在相关数据集上进行了验证与测试, 识别准确率优于当前诸多主流方法, 具有较高的先进性。

1　相关工作

微表情识别方法主要有两大类: 基于传统手工特征的方法和基于深度学习的方法。

1.1　传统手工特征方法

手工提取特征方法通过手动设计视觉描述符来捕捉微表情的独特特征, 再将其输入分类器进行识别。Pfister 等[7] 首次运用 LBP 和机器学习算法实现了微表情识别。为了精确处理面部微表情特征, Liu 等[8] 提出一种主方向平均光流特征提取法, 该方法包括从微表情方向序列上提取光流并取平均值。有研究者借助 Gabor 滤波和支持向量机识别不同类别的微表情[9]。Ma 等[10] 介绍了定向光流区域直方图(region histogram of oriented optical flow, RHOOF), 从相邻帧中提取 HOOF 特征以检测顶点帧, 使用顶点帧与起始帧之间的 RHOOF 特征以提高识别率。

1.2　深度学习方法

深度学习方法一般采用神经网络, 可以实现端到端自动提取深层特征, 对数据进行分类和预测。针对数据集无顶点帧标签的问题, Liu 等[15] 采用轻量级网络 SqueezeNet 对无顶点帧标签的数据集实现可靠的顶点帧定位, 并运用三维卷积网络提取时空特征。针对微表情局部细节特征难以关注的问题, Zhao 等[16] 提出 ME-PLAN 深度学习框架, 通过与表情相关的知识迁移和情景训练来学习精确的微表情特征。Zhou 等[17] 采用双流 inception 网络, 注重获取特定表情的显著特征和判别特征, 并通过融合特定微表情的特征来更好地识别微表情。Li 等[18] 提出了一种深度局部整体网络, 提取局部丰富的时空特征和全局特征, 通过子网络融合用于微表情识别。

注意力机制可有效改进网络模型, 对图像中重要特征分配更高的权重, 在数据的通道和空间维度引入注意力机制以抑制背景干扰, 从而增强网络对关键特征的感知能力。Li 等[19] 将 LBP 纹理特征与注意力机制相结合, 让模型在特征提取时, 对输入的不同区域分配不同的注意力权重, 从而增强对有用特征的关注。Rodríguez 等[20] 将空间注意力与来自不同卷积层的多尺度信息融合在一起, 以表示图像中每个区域的重要性。Su 等[21] 使用组件感知注意力模块, 突显微表情的相关区域, 以有效地捕捉动机信息和非刚性变形。3D Residual Network 将时空和通道信息相结合来更好地识别微表情[22]。

2　基本原理

2.1　网络结构

本文提出的融合 LBP 与并行注意力机制微表情识别网络结构如图 1 所示, 工作流程分为数据预处理、特征提取、特征融合及分类 4 个部分。数据预处理 (dataset preprocessing) 包括人脸裁剪、人脸对齐、数据扩增和 LBP 处理等操作。特征提取网络包括全局特征提取模块 (global feature extraction module, GFEM) 和局部纹理特征提取模块 (local texture feature extraction module, LTFEM); 通过注意力特征融合模块 (convolutional attention feature fusion module, CAFFM) 将两种特征有效融合在一起, 以提高人脸微表情的识别性能。

2.2　全局特征提取模块

2.2.1　密集连接

为了加强特征复用, 充分利用浅层特征, 降低过拟合风险和缓解梯度消失的问题, 本文引入密集连接[11]。如图 2 所示, 在网络中级联第 N 层和第

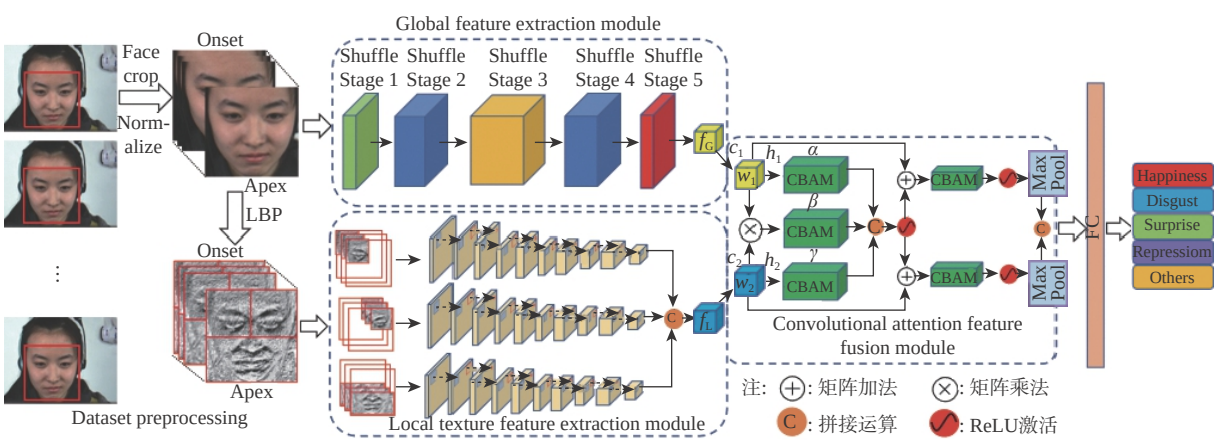


图 1　融合 LBP 与并行注意力机制的微表情识别网络结构

Fig. 1　Structure of micro-expression recognition network integrating LBP and parallel attention mechanism

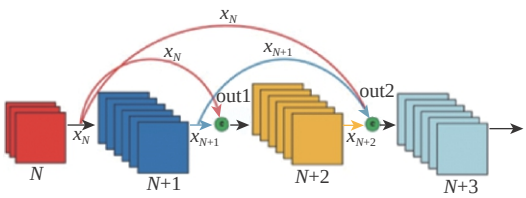


图2 密集连接结构

Fig. 2 Structure of dense connection convolution layers

N+1层的输出，在通道维度上得到特征$\boldsymbol{x}_{N+2}$，然后将$\boldsymbol{x}_{N+2}$、$\boldsymbol{x}_{N+1}$和$\boldsymbol{x}_N$级联，输入到N+3层生成特征$\boldsymbol{x}_{N+3}$。计算公式如下：

$$\boldsymbol{x}_{N+1}=\mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}(\boldsymbol{x}_N))) \tag{1}$$

$$\mathrm{out1}=x_N\copyright x_{N+1} \tag{2}$$

$$\boldsymbol{x}_{N+2}=\mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}(\mathrm{out1}))) \tag{3}$$

$$\mathrm{out2}=x_N\copyright x_{N+1}\copyright x_{N+2} \tag{4}$$

$$\boldsymbol{x}_{N+3}=\mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}(\mathrm{out2}))) \tag{5}$$

式中：Conv为卷积运算；BN为正则化操作；ReLU为非线性运算；©为Channel级连接操作。

2.2.2 全局特征提取模块

全局特征提取模块以ShuffleNet V2网络为基准框架[12]，并将输入特征通道分成两个分支，stride为1时，单元中一个分支保持不变，另一分支由2个1×1普通卷积和一个3×3深度可分离卷积组成；stride为2时，移除通道分离，单元中一个分支由一个1×1普通卷积和一个3×3深度可分离卷积组成，另一分支由2个1×1普通卷积和一个3×3深度可分离卷积组成。其中的模块结构包括了ShuffleNet V1中的基本单元和新的单元，这些单元在网络中通过堆叠的方式逐渐扩大感受野以构建更深的网络。网络通过两个主要的特征提取阶段来捕捉全局特征，其中包含了分组卷积、Channel Shuffle和池化等操作，通过特征通道的混洗操作，使得不同分组的特征可以进行交流和融合，并在第2和4阶段引入密集连接增强了特征的多样性和全局感知能力，结构如图3所示。此外，ShuffleNet V2中的全局平均池化操作将整个特征图转换为固定大小的特征向量进行降维，可以有效地提取全局图像语义信息，进一步加强全局特征的表示能力。

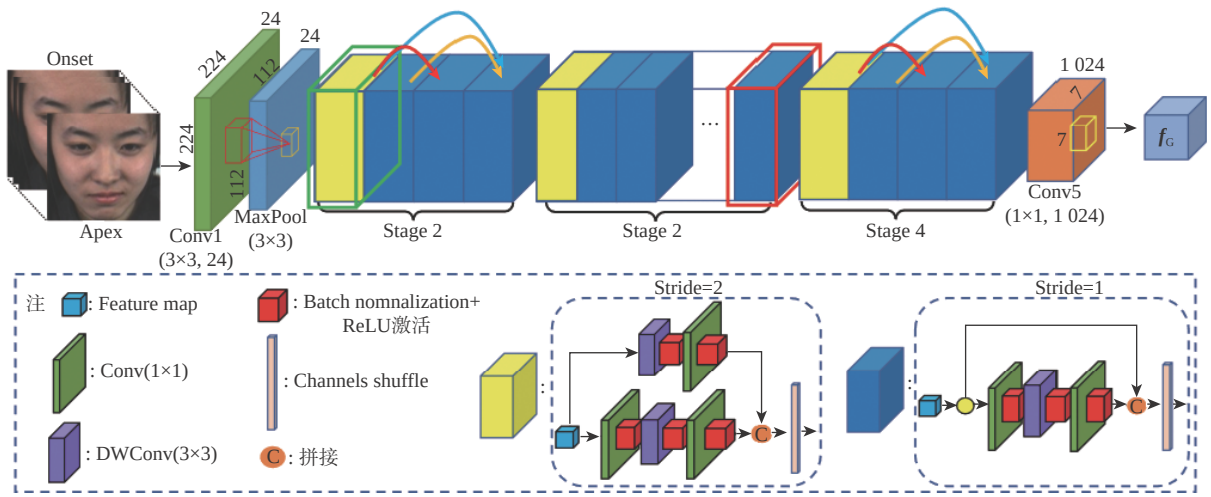


图3 全局特征提取模块

Fig. 3 Global feature extraction module

全局特征具体的提取过程为：将整张RGB图片序列输入到该分支网络中进行全面的特征提取操作，在每个阶段中采用分组卷积和Channel Shuffle操作。分组卷积机制让不同组的特征图进行卷积操作，从而捕捉到不同组之间的全局信息；Channel Shuffle将不同组的特征图进行混合和交互，促进不同组之间的信息流动；全局平均池化层会将整个特征图转换为固定大小的特征向量，从而获取更全局的人脸特征$\boldsymbol{f}_G$，最后输出的全局特征图可视化如图4所示，粉红色区域表示提取到的人脸全局信息。从提取过程上看，该分支网络引入密集连接机制进行优化，加强特征信息之间的交流，保留了样本特征的全局性质，减少过拟合，使模型仍能保持轻量化，提高识别准确性。

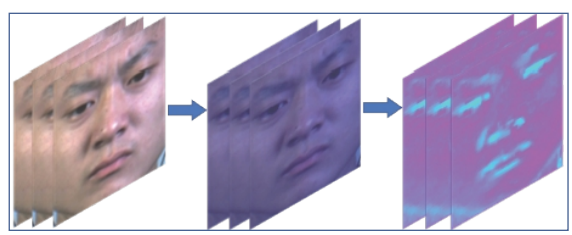


图4 全局特征图可视化

Fig. 4 Visualization of global feature map

2.3　局部纹理特征提取模块

模型提取全局特征时容易丢失图像中的细节信息, 基于多尺度分层卷积神经网络 (multiscale hierarchical convolutional neural network, MHCNN) 的微表情识别算法设计了局部纹理特征提取模块, 单个 MHCNN 网络由 5 层卷积层组成, 结构如图 5 所示。

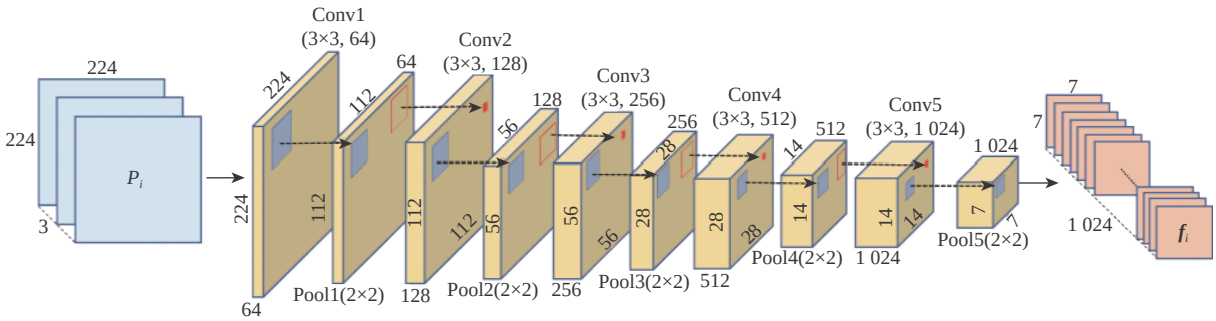


图 5　多尺度分层卷积神经网络

Fig. 5　Multiscale hierarchical convolutional neural network

局部纹理特征具体的提取过程为: ①将整张 LBP 图像分成左眼、右眼和嘴 3 个局部区域, 并分别输入到 MHCNN 网络中, 分别捕捉到不同区域之间的局部信息, 局部特征提取后分别得到 $\boldsymbol{f}_1$、$\boldsymbol{f}_2$ 和 $\boldsymbol{f}_3$ 等 3 个特征值。②为了便于后期的特征融合, 将 $\boldsymbol{f}_1$、$\boldsymbol{f}_2$ 和 $\boldsymbol{f}_3$ 按原位拼接得到输出特征图 $\boldsymbol{f}_{\mathrm{L}}$。具体计算公式如下:

$$\boldsymbol{f}_{\mathrm{L}} = \boldsymbol{f}_1 \copyright \boldsymbol{f}_2 \copyright \boldsymbol{f}_3 \tag{6}$$

式中: ©表示特征图拼接运算。最后输出的局部特征图可视化如图 6 所示, C 中深色区域表示提取到的局部细节信息, 其余部分则是未提取到的信息。从提取过程上看: 并行 MHCNN 有效地提取了 3 个局部位置的特征, 分支网络可以专心学习微表情 LBP 图像序列的局部纹理特征。

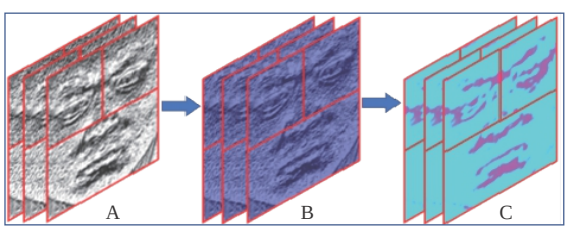


图 6　局部特征图可视化

Fig. 6　Visualization of local texture feature map

2.4　注意力特征融合模块

双分支网络模型采用卷积神经网络提取特征, 缺乏有效的注意力机制使整体网络关注更重要的细节特征。因此, 本文引入 CBAM 模块[11] 对网络结构进行优化。CBAM 包含两个子模块, 分别是通道注意力模块 (channel attention module, CAM) 和空间注意力模块 (spatial attention module, SAM), CAM 主要通过计算每个通道的重要性来自适应地缩放每个通道的特征图, SAM 主要关注不同空间位置之间的相互依赖关系。将这两类模块进行串行连接, 可以集中关注特征图中关键信息。

输入的 RGB 图和 LBP 纹理图经过 GFEM 和 LTFEM 之后, 分别得到全局特征图 $\boldsymbol{f}_{\mathrm{G}}$ 和局部特征图 $\boldsymbol{f}_{\mathrm{L}}$。其中 $\boldsymbol{f}_{\mathrm{G}} \in \mathbf{R}^{h_1 \times w_1 \times c_1}$, $\boldsymbol{f}_{\mathrm{L}} \in \mathbf{R}^{h_2 \times w_2 \times c_2}$, h_1、w_1、c_1 和 h_2、w_2、c_2 分别表示全局特征图的高度、宽度和通道数以及局部特征图的高度、宽度和通道数。特征图中过多的相加和张量积运算会导致计算量过大, 相比之下用特征拼接的融合方法效率高且适用于该算法的双分支网络, 可以充分利用两种特征的互补性。

受文献 [13-14] 的启发, 引入 CBAM 模块处理特征图, 本文设计了一种基于卷积块注意力机制的特征融合模块 (based on convolutional block attention feature fusion module, CAFFM), 结构图如图 7(a) 所示, CAFFM 模块使用 CBAM 的堆叠方式来融合全局和局部特征。具体的融合方法如下。

1) 全局特征 $\boldsymbol{f}_{\mathrm{G}}$ 和局部特征 $\boldsymbol{f}_{\mathrm{L}}$ 进行哈达玛积后, 将三种特征分别经过 α, β, γ 不同系数的 CBAM 模块之后, 再进行拼接和激活得到拼接特征 $\boldsymbol{H}$。计算公式如下:

$$\begin{aligned}\boldsymbol{H} = \mathrm{ReLU}\{\mathrm{Concat}[&\alpha * \mathrm{CBAM}(\boldsymbol{f}_{\mathrm{G}}),\\ &\beta * \mathrm{CBAM}(\boldsymbol{f}_{\mathrm{G}} \otimes \boldsymbol{f}_{\mathrm{L}}), \gamma * \mathrm{CBAM}(\boldsymbol{f}_{\mathrm{L}})]\}\end{aligned} \tag{7}$$

式中: ReLU 为非线性运算; CBAM(·) 表示对特征图进行处理; Concat 为拼接操作; ⊗为哈达玛积运算; *表示数乘运算。

2) 拼接特征 $\boldsymbol{H}$ 分别与全局特征 $\boldsymbol{f}_{\mathrm{G}}$ 和局部特征 $\boldsymbol{f}_{\mathrm{L}}$ 进行加权后再经过 CBAM 模块和激活, 这里采用最大池化将特征图进行聚合, 得到新的全局特征 $\boldsymbol{F}_{\mathrm{G}}$ 和新的局部纹理特征 $\boldsymbol{F}_{\mathrm{L}}$, 以便最终特征的互补融合。计算公式如下:

$$\boldsymbol{F}_{\mathrm{G}} = \mathrm{MaxPool}\{\mathrm{ReLU}[\mathrm{CBAM}(\boldsymbol{H} \oplus \boldsymbol{f}_{\mathrm{G}})]\} \tag{8}$$

$$\boldsymbol{F}_{\mathrm{L}} = \mathrm{MaxPool}\{\mathrm{ReLU}[\mathrm{CBAM}(\boldsymbol{H} \oplus \boldsymbol{f}_{\mathrm{L}})]\} \tag{9}$$

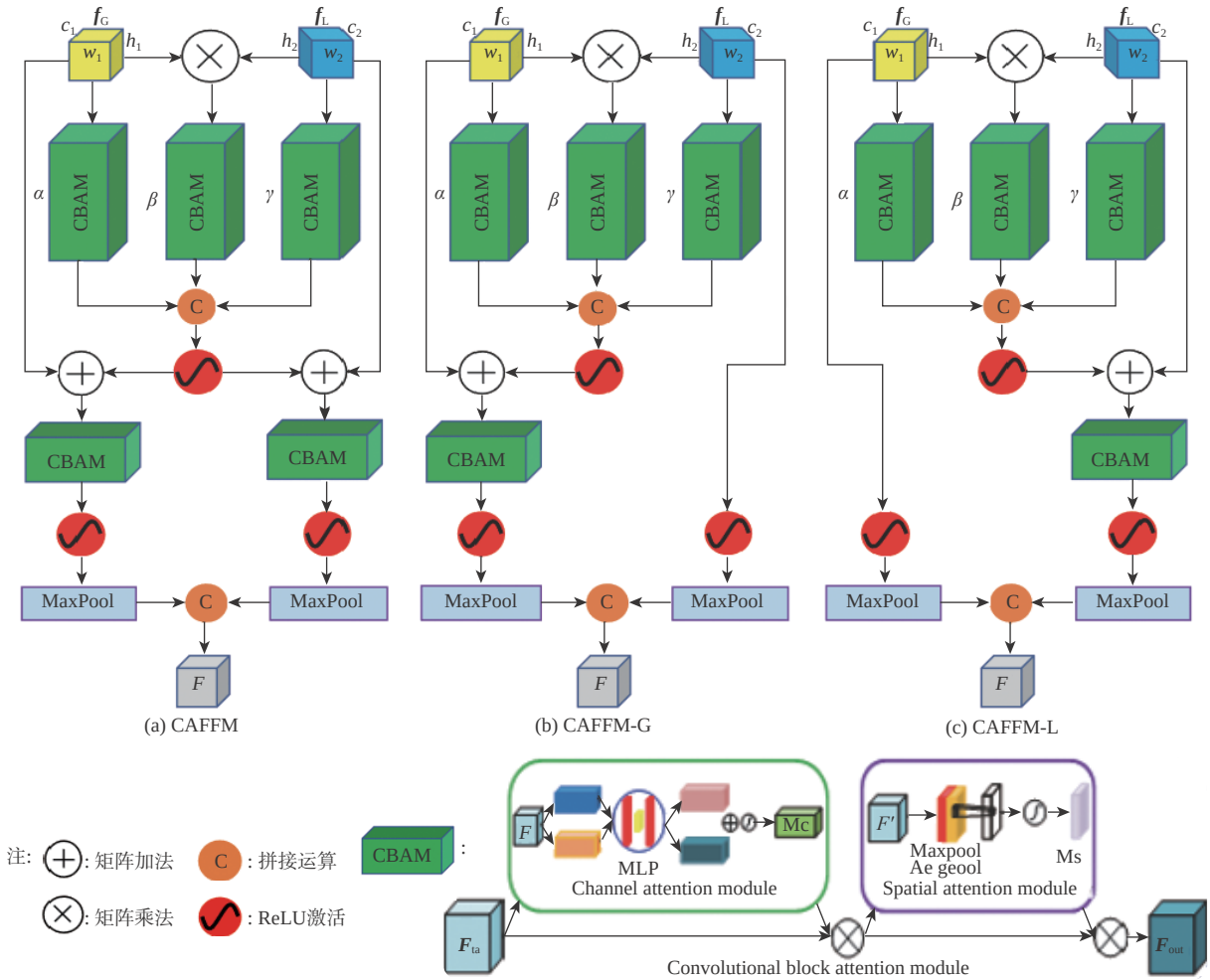


图 7 基于卷积块注意力机制的特征融合模块

Fig. 7 Feature fusion module based on convolutional block attention mechanism

式中: MaxPool(x) 表示最大池化运算; ⊕为矩阵加法运算。

3) 将新的全局特征 $\boldsymbol{F}_{\mathrm{G}}$ 和局部特征 $\boldsymbol{F}_{\mathrm{L}}$ 进行拼接得到最终的微表情特征 $\boldsymbol{F}$。计算公式如下:

$$\boldsymbol{F}=\mathrm{Concat}(\boldsymbol{F}_{\mathrm{G}},\boldsymbol{F}_{\mathrm{L}}) \tag{10}$$

从整体结构上看, CAFFM 模块有效地实现了两种特征的融合, 使模型关注关键特征并抑制无关信息。为了研究全局特征和局部特征对融合效果的影响, 在 CAFFM 模块上进行了改动, 得到另外两种不同的融合策略, 分别如图 7(b)和图 7(c)所示, 称为 CAFFM-G 和 CAFFM-L。与 CAFFM 不同的是, CAFFM-G 模块在加权运算时, 拼接特征 $\boldsymbol{H}$ 只加权全局特征 $\boldsymbol{f}_{\mathrm{G}}$, 局部特征保持不变; 而 CAFFM-L 模块在加权运算时, 拼接特征 $\boldsymbol{H}$ 只加权局部特征 $\boldsymbol{f}_{\mathrm{L}}$, 全局特征保持不变。

3 实验与评估

3.1 数据集

为评估所提出的网络模型的性能, 本文在 3 个典型数据集(CAMSE[23]、CASME II[24] 和 SMIC[25])上进行实验。

3.1.1 SMIC 数据集

SMIC 数据集是由芬兰奥卢大学公开发布的数据集。该数据集包含 HS、VIS 及 NIR 3 个子数据集, 分别用分辨率为 1 280×720 且帧率为 100 帧/s 的高帧率摄像机、分辨率为 640×480 且帧率为 25 帧/s 的普通相机和分辨率为 640×480 且帧率为 25 帧/s 的红外相机进行拍摄。样本数分别为 164 个、71 个、71 个, 由来自 3 个种族的 16 位受试者参与了微表情数据采集。由于 HS 子数据集样本数多帧率高, 所以本文采用 HS 子数据集进行实验, 样本分为 3 类别, 即积极 51 个、消极 66 个和惊讶 40 个, 其中训练集样本数为 130 个, 测试集样本数为 34 个。

3.1.2 CASME 数据集

CASME 数据集由中国科学院心理研究所公开发布的数据集。该数据集总共有 195 个微表情样本, 35 位受试者参与微表情数据采集。样本包含 8 种情感类别的标签, 即高兴、厌恶、悲伤、蔑视、

恐惧、紧张、惊讶和压抑。该数据集的分辨率 A 部分为 1 280×720, B 部分为 640×480, 帧率为 60 帧/s, 样本视频中人脸尺寸为 150×190。由于 B 部分数据样本数量多, 本文将 B 部分数据集采用 3 分类进行实验, 即积极 5 个、消极 82 个和惊讶 14 个, 其中训练集样本数为 80 个, 测试集样本数为 21 个。

3.1.3　CASME II 数据集

CASME II 数据集为 CASME 的扩展, 设备采用高速相机。该数据集包含 249 个微表情样本, 共 26 位受试者参与微表情数据采集。样本包含 5 种情感类别的标签, 即高兴 32 个、压抑 27 个、惊讶 28 个、厌恶 63 个、其他 99 个, 样本情感类别数量相比之前变得更均衡。图片分辨率为 640×480, 帧率为 200 帧/s, 样本人脸尺寸为 280×340。其中训练集样本数为 199 个, 测试集样本数为 50 个。

3.2　数据预处理

微表情开始于起始帧, 并在顶点帧达到最高强度。研究表明, 通过比较这两帧, 能有效提取微表情特征, 同时避免冗余信息[26]。本文使用起始帧、顶点帧和结束帧来捕捉微表情运动变化的全过程。由于 3 个数据集中的数据分辨率、码率和其他指标存在差异, 本实验对不同数据集的数据进行了同等处理。

使用 OpenCV 和 Dlib 工具包提取人脸, 定位人脸标记点, 将人脸裁剪为 224×224 像素的 RGB 图像, 通过仿射变换将图片进行水平对齐, 并对样本进行数据扩增。凭借 LBP 算子像素点进行量化, 将裁剪好的 RGB 图像转化为 LBP 纹理图。

3.3　实验参数

实验环境是在 Ubantu18.04 系统上使用 NVIDIA GeForce GTX 1 080 Ti 下运行, 使用 Python3.8 搭建基础环境并运行程序, 训练的过程使用交叉熵损失函数和 Adam 算法进行优化。样本图像尺寸为 224 像素×224 像素, 采用数据扩增策略以避免过拟合。训练时学习率设置为 1.0×10^{-4}, β_1 设为 0.9, β_2 设为 0.999, Batchsize 设置为 32, 指数衰减率为 1.0×10^{-5}, 总批次设置为 300, L2 正则化设为 0.01, CBAM 压缩率为 8。

3.4　结果和分析

为了研究所提方法的微表情识别准确率, 训练与测试在 SMIC、CASME 和 CASME II 数据集上完成, 并将测试结果与诸多先进方法进行对比, 实验结果如表 1 和表 2 所示。由表 1 能够看出, 所提方法在 CASME 数据集上识别准确率达到 85.18%, 相较于 LCBP-STGCN[27]、MFAPL + MK-SVM[28]、3D Residual Network[22]、LEARNet[29] 等方法提高了 3.92%、5.77%、4.18%、4.56%。所提方法在 CASME 数据集上具有较好的竞争性。

表 1　CASME 数据集识别准确率

Table 1　CASME dataset recognition accuracy

方法	准确率/%
STCLQP[30]	63.49
HIGO[31]	57.81
FHOFO[32]	67.20
MDMO[8]	68.25
Macro2Micro[33]	67.72
STRCN-G[34]	70.90
AKMNet[35]	75.66
LEARNet[29]	80.62
3D Residual Network[22]	81.00
Meta-MMFNet[36]	69.59
MFAPL + MK-SVM[28]	79.41
LCBP-STGCN[27]	81.26
本文	**85.18**

表 2　CASME II 和 SMIC 数据集识别准确率

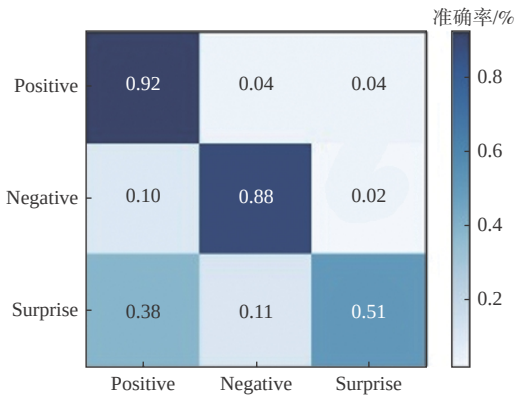
Table 2　CASME II and SMIC dataset recognition accuracy

方法	SMIC 准确率/%	CASME II 准确率/%
MSMMT[37]	78.30	71.31
AMAN[38]	79.87	75.40
KTGSL[39]	72.58	**75.64**
Later[40]	73.17	70.68
SLSTT-Mean[41]	73.17	73.79
STLBP-IIP[42]	60.37	62.75
DISTLBP-IIP[42]	63.41	64.78
3D-CNNs (with transfer learning)[43]	66.30	65.90
DSSN[44]	63.40	70.78
SSSN[44]	63.41	71.19
AKMNet[35]	72.56	67.06
TSCNN-I[26]	72.74	74.05
KFC-MER[21]	65.85	72.76
FR[17]	57.90	62.85
Knowledge Distillation[45]	76.06	72.61
本文	**81.19**	74.53

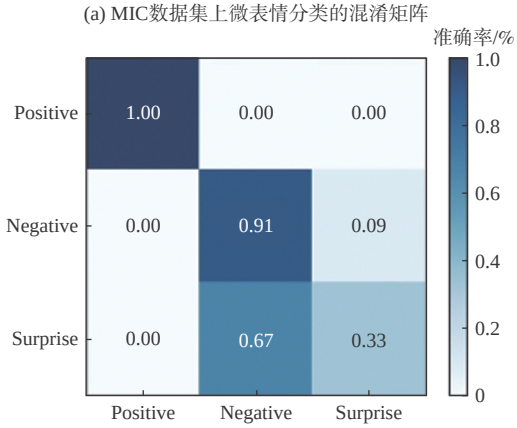
由表 2 能够看出, 所提方法在 SMIC 数据集上识别准确率达到 81.19%, 相较于 MSMMT[39]、AMAN[40]、KTGSL[41]、Later[42] 等方法提高了 2.89%、1.32%、8.61%、8.02%, 在所比较的几种先进算法中, 达到了最高。此外, 所提方法在 CASME II 数据集上取得了 74.53% 的准确率, 相较于 MSMMT[39]、Later[42]、SLSTT-Mean[43] 这些新方法提高了 3.22%、3.85%、0.74%, 与目前 75.64% 最高准确率的 KTGSL 方法也较接近, 综上表明所提方法具有较好的先进

性。本文方法采用了一种学习策略，而KTGSL采用两组核化的稀疏学习，并用两组学习策略从两组特征中学习更多的判别特征。因此所得准确率更高，但两组学习策略所需时间有所增加。所提方法与AMAN方法有差距，AMAN放大了微表情图像，扩大了微表情强度，而所提方法则采用原始数据图像，这可能是所提方法准确率低于AMAN的关键因素。

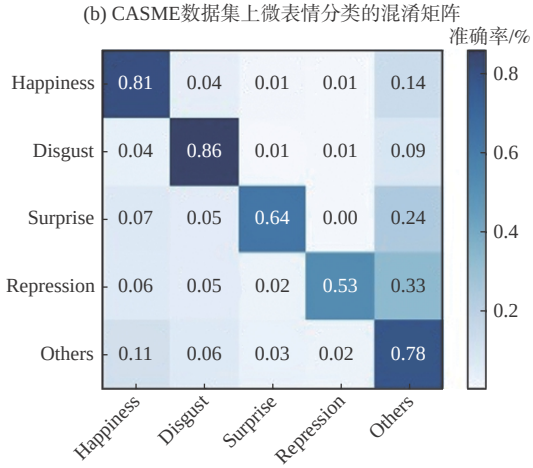
图8展示了所提模型在CASME、SMIC及CASME II数据集上的混淆矩阵。由图8(a)和图8(b)可知，SMIC和CASME数据集在识别过程中，微表情“Surprise”容易同“Positive”和“Negative”发生误判，这可能是因为数据集中“Surprise”的微表情样本较少，使模型欠缺相应表情的训练，导致准确率较低。从图8(c)可以看出，在识别CASME II数据集时，模型的识别率相对较低，这是由于微表情有较多相似的AU，如图9所示，给出一些AU发生的运动部位：AU4(皱眉)、AU9(皱鼻子)、AU10(拉动上嘴唇)、AU14(嘴角收缩)、AU15(嘴角下扬)和AU34(鼓气)，例如，微表情“Surprise”和“Repression”的发生部位都在嘴角以及邻近部位，这与嘴角运动单元AU14、AU15和AU34等有关，所以这些AU运动部位相近容易发生混淆，导致识别率较低。



(a) MIC数据集上微表情分类的混淆矩阵



(b) CASME数据集上微表情分类的混淆矩阵



(c) CASME II数据集上微表情分类的混淆矩阵

图8 在SMIC、CASME、CASME II数据集上的混淆矩阵

Fig. 8 Confusion matrix on SMIC, CASME, CASME II datasets

3.5 消融实验

为了进一步研究所提方法各模块的性能，所提模型在SMIC、CASME和CASME II数据集上进行了消融实验，实验时确保了每个样本数量相同，在全局特征提取模块(GFEM)、局部纹理特征提取模块(LTFEM)和特征融合模块(CAFFM)上，从准确率、模型参数量、浮点运算量和推理速度等方面对不同优化策略进行比较分析，测试结果如表3所示。在CASME数据集上双分支特征网络融合之后，准确率得到显著提升，分别引入密集连接(DEN)和构建CAFFM后参数量增长0.04M和0.8M，同时准确率提高2.08%和1.40%，GFEM+ LTFEM+DEN+CAFFM达到最高准确率85.18%，相较于两个单分支网络提升了7.68%和8.65%。在CASME II数据集上的双分支网络的准确率也得到提升，双分支网络引入DEN后准确率提升了4.85%，融合CAFFM后参数量增长0.8M并且准确率达到74.33%，但相比于双分支网络仅使用CAFFM的准确率相差0.34%，同样具有较好的准确率。引入DEN和CAFFM模块之后运算量增加到5.258G，推理速度变为95.8FPS，其在SMIC数据集上模型准确率亦有较大提升，准确率最高达到了81.19%，相较于单独使用双分支网络提升了4.72%。此外，从表3可以看出，随着优化策略的引入，模型参数量和计算量均有所增加，推理速度有所降低，整体上看，推理速度下降幅度较模型参量增长幅度要小，尚能满足实时性的要求。值得一提的是，微表情识别模型的参数量和准确率之间的平衡性是较为复杂的问题，相关工作仍需要进一步探索。

本文分别设计了3种融合策略对特征进行学习融合，为了进一步分析3个权重系数对模型最后识别的准确率的影响，表4、表5和表6分别列举了CAFFM 3种融合策略方法中α, β, γ这3个权重

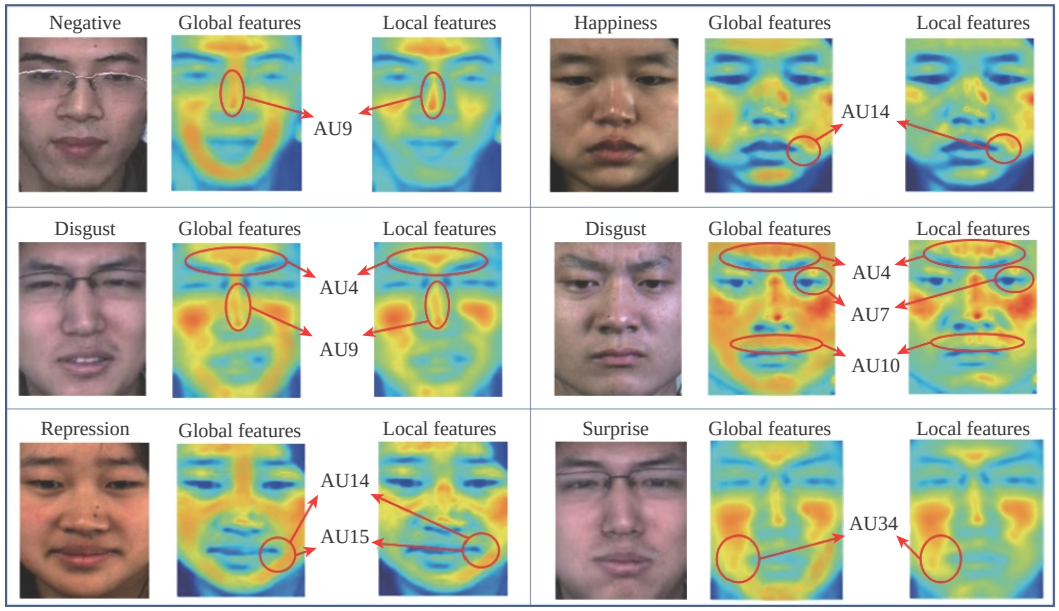


图 9　AU 与微表情可视化关系示例

Fig. 9　Example diagram of the visual relationship between AU and micro-expressions

表 3　CASME、CASME II 和 SMIC 数据集上的消融实验

Table 3　Ablation experimental research on CASME, CASME II and SMIC datasets

方法	CASME 准确率/%	CASME II 准确率/%	SMIC 准确率/%	参数量/M	浮点运算量/G	推理速度/(帧·s^{-1})
GFEM	77.50	66.31	73.38	2.59	1.778	111.5
LTFEM	76.53	62.69	76.90	6.26	3.287	106.3
GFEM+LTFEM	81.41	66.34	76.47	8.13	4.927	104.5
GFEM+LTFEM+DEN	83.49	71.19	77.39	8.17	5.073	102.2
GFEM+LTFEM+CAFFM	82.81	**74.67**	79.76	8.46	5.165	97.3
GFEM+LTFEM+DEN+CAFFM	**85.18**	74.33	**81.19**	8.97	5.258	95.8
GFEM+LTFEM+CAFFM-L	84.52	74.52	79.71	8.49	5.140	97.3
GFEM+LTFEM+DEN+CAFFM-L	84.14	74.55	77.81	8.78	5.232	95.7
GFEM+LTFEM + CAFFM-G	81.63	72.23	77.76	8.50	5.151	97.5
GFEM+LTFEM +DEN+CAFFM-G	83.26	74.05	77.52	8.79	5.243	95.8

表 4　CAFFM 不同权重比在 3 个数据集上的准确率

Table 4　Accuracy of different weight ratios of CAFFM on 3 datasets

α:β:γ	CASME 准确率/%	CASME II 准确率/%	SMIC 准确率/%
1:1:1	**82.82**	**74.67**	80.19
2:1:1	80.78	72.71	**80.76**
1:2:1	78.37	73.67	79.71
1:1:2	79.56	73.28	79.90

表 5　CAFFM-L 不同权重比在 3 个数据集上的准确率

Table 5　Accuracy of different weight ratios of CAFFM-L on 3 datasets

α:β:γ	CASME 准确率/%	CASME II 准确率/%	SMIC 准确率/%
1:1:1	83.53	**74.53**	**79.71**
2:1:1	**84.15**	74.47	78.86
1:2:1	82.96	74.36	78.81
1:1:2	82.44	74.22	77.76

表 6　CAFFM-G 不同权重比在 3 个数据集上的准确率

Table 6　Accuracy of different weight ratios of CAFFM-G on 3 datasets

α:β:γ	CASME 准确率/%	CASME II 准确率/%	SMIC 准确率/%
1:1:1	**82.96**	72.23	77.76
2:1:1	**82.96**	73.86	78.38
1:2:1	81.63	**74.56**	79.57
1:1:2	82.44	73.61	**80.24**

系数的比值在 CASME, CASME II 和 SMIC 这 3 个数据集上的准确率。如表 4 所示, CAFFM 模块 CASME 和 CASME II 数据集的准确率在 α:β:γ=1:1:1 时最高; SMIC 数据集的准确率最高。这可能是因为 SMIC 数据集的分辨率在 α:β:γ=2:1:1 时较高, 提高全局特征提取的权重更有优势, 因此得到的准确率会更高。

如表5所示，CAFFM-L模块CASME II和SMIC的准确率在α:β:γ=1:1:1时最高；CASME的准确率在α:β:γ=2:1:1时最高。可能的原因是CAFFM-Ll模块融合时的全局特征所占比重较小，缺乏足够的全局维度信息。同时，CASME数据集的样本类别分布不均匀，增大全局特征的融合占比可以提升准确率。

如表6所示，CAFFM-G模块，CASME的准确率在α:β:γ=1:1:1和α:β:γ=2:1:1时最高；CASME II的准确率在α:β:γ=1:2:1时最高；SMIC的准确率在α:β:γ=1:1:2时最高。这可能的原因在于CAFFM-G模块融合时局部特征所占比重较小，缺乏局部特征信息，所以在增大局部特征融合比重时，SMIC和CASME II的准确率都得到了提升。但由于CASME数据集原始RGB样本的帧率较小，以及局部特征提取模块算法性能的限制，因此在增大局部特征融合比重时会引入噪声。通过以上实验，可以证明全局特征和局部特征之间存在互补性，二者融合得到的微表情特征更加准确。

4 结 论

1）本文提出了一种融合LBP与并行注意力机制的微表情识别网络，该网络可提取RGB全局特征和LBP局部纹理特征，通过注意力特征融合模块得到更有效的微表情特征，同时，在全局特征提取模块引入密集连接机制增强特征复用。

2）所提方法在数据集CASME、SMIC和CASME II上准确率分别为85.18%、81.19%和74.53%，结果较当前诸多微表情识别先进方法表现优异，表明所提方法可有效提升微表情识别准确率。

3）所提方法随着优化策略引入，模型参数量和计算量均有所增加，推理速度有所降低，介于95.7～111.5帧/s，整体上可满足实时性要求。

在下一步研究工作中，需进一步探索模型识别准确率与推理速度之间的平衡性，在实际应用中提升模型性能。

参考文献（References）

[1] PORTER S, TEN BRINKE L. Reading between the lies: identifying concealed and falsified emotions in universal facial expressions[J]. Psychological Science, 2008, 19(5): 508-514.

[2] HAGGARD E A, ISAACS K S. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy[M]// Methods of Research in Psychotherapy. Boston: Springer US, 1966: 154-165.

[3] EKMAN P, FRIESEN W V. The repertoire of nonverbal behavior: categories, origins, usage, and coding[J]. Semiotica, 1969, 1(1): 49-98.

[4] FRANK M, HERBASZ M, SINUK K, et al. I see how you feel: Training laypeople and professionals to recognize fleeting emotions[C]//The Annual Meeting of the International Communication Association. New York: Academic Press, 2009: 1-35.

[5] EKMAN P, FRIESEN W V. Facial action coding system: a technique for the measurement of facial actions[J]. Rivista Di Psichiatria, 1978, 47(2): 126-138.

[6] OJALA T, PIETIKÄINEN M, HARWOOD D. A comparative study of texture measures with classification based on featured distributions[J]. Pattern Recognition, 1996, 29(1): 51-59.

[7] PFISTER T, LI X B, ZHAO G Y, et al. Recognising spontaneous facial micro-expressions[C]// 2011 International Conference on Computer Vision. Piscataway: IEEE Press, 2011: 1449-1456.

[8] LIU Y J, ZHANG J K, YAN W J, et al. A main directional mean optical flow feature for spontaneous micro-expression recognition [J]. IEEE Transactions on Affective Computing, 2015, 7(4): 299-310.

[9] WU Q, SHEN X B, FU X L. The machine knows what you are hiding: an automatic micro-expression recognition system[M]// D' MELLO S, GRAESSER A, SCHULLER B, et al. eds. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 152-162.

[10] MA H Y, AN G Y, WU S J, et al. A Region Histogram of Oriented Optical Flow (RHOOF) feature for apex frame spotting in micro-expression[C]// 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS). Piscataway: IEEE Press, 2017: 281-286.

[11] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 2261-2269.

[12] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[M]// Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 122-138.

[13] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[M]// Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 3-19.

[14] CHENG R, RAZANI R, TAGHAVI E, et al. $(AF)_2$-S3Net: attentive feature fusion with adaptive feature selection for sparse semantic segmentation network[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 12542-12551.

[15] LIU S H, REN Y S, LI L T, et al. Micro-expression recognition based on SqueezeNet and C3D[J]. Multimedia Systems, 2022, 28(6): 2227-2236.

[16] ZHAO S R, TANG H Y, LIU S F, et al. ME-PLAN: a deep prototypical learning with local attention network for dynamic micro-expression recognition[J]. Neural Networks, 2022, 153: 427-443.

[17] ZHOU L, MAO Q R, HUANG X H, et al. Feature refinement: an expression-specific feature learning and fusion method for micro-expression recognition[J]. Pattern Recognition, 2022, 122: 108275.

[18] LI J T, WANG T, WANG S J. Facial micro-expression recognition based on deep local-holistic network[J]. Applied Sciences, 2022, 12(9): 4643.

[19] LI J, JIN K, ZHOU D L, et al. Attention mechanism-based CNN for

facial expression recognition[J]. Neurocomputing, 2020, 411: 340-350.

[20] RODRÍGUEZ P, VELAZQUEZ D, CUCURULL G, et al. Pay attention to the activations: a modular attention mechanism for fine-grained image recognition[J]. IEEE Transactions on Multimedia, 2019, 22(2): 502-514.

[21] SU Y T, ZHANG J Q, LIU J, et al. Key facial components guided micro-expression recognition based on first & second-order motion[C]// 2021 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2021: 1-6.

[22] GAJJALA V R, REDDY S P T, MUKHERJEE S, et al. MERANet: facial micro-expression recognition using 3D residual attention network[C]// Proceedings of the Twelfth Indian Conference on Computer Vision, Graphics and Image Processing. New York: ACM, 2021: 1-10.

[23] YAN W J, QI W, LIU Y J, et al. CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces[C]// 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). Piscataway: IEEE Press, 2013: 1-7.

[24] YAN W J, LI X B, WANG S J, et al. CASME II: an improved spontaneous micro-expression database and the baseline evaluation[J]. PLoS One, 2014, 9(1): e86041.

[25] LI X B, PFISTER T, HUANG X H, et al. A spontaneous micro-expression database: inducement, collection and baseline[C]// 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). Piscataway: IEEE Press, 2013: 1-6.

[26] SONG B L, LI K, ZONG Y, et al. Recognizing spontaneous micro-expression using a three-stream convolutional neural network[J]. IEEE Access, 2019, 7: 184537-184551.

[27] WANG Y, HAN J F, GUO Z Q. RETRACTED: LCBP-STGCN: a local cube binary pattern spatial temporal graph convolutional network for micro-expression recognition[J]. Journal of Intelligent & Fuzzy Systems, 2023, 44(2): 1601-1611.

[28] CEN S X, YU Y, YAN G, et al. Multi-task Facial Activity Patterns Learning for micro-expression recognition using Joint Temporal Local Cube Binary Pattern[J]. Signal Processing: Image Communication, 2022, 103: 116616.

[29] VERMA M, VIPPARTHI S K, SINGH G, et al. LEARNet: Dynamic imaging network for micro expression recognition[J]. IEEE Transactions on Image Processing, 2019, 29: 1618-1627.

[30] HUANG X H, ZHAO G Y, HONG X P, et al. Spontaneous facial micro-expression analysis using Spatiotemporal Completed Local Quantized Patterns[J]. Neurocomputing, 2016, 175: 564-578.

[31] LI X B, HONG X P, MOILANEN A, et al. Towards reading hidden emotions: a comparative study of spontaneous micro-expression spotting and recognition methods[J]. IEEE Transactions on Affective Computing, 2017, 9(4): 563-577.

[32] HAPPY S L, ROUTRAY A. Fuzzy histogram of optical flow orientations for micro-expression recognition[J]. IEEE Transactions on Affective Computing, 2017, 10(3): 394-406.

[33] PENG M, WU Z, ZHANG Z H, et al. From macro to micro expression recognition: deep learning on small datasets using transfer learning[C]// 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). Piscataway: IEEE Press, 2018: 657-661.

[34] XIA Z Q, HONG X P, GAO X Y, et al. Spatiotemporal recurrent convolutional networks for recognizing spontaneous micro-expressions[J]. IEEE Transactions on Multimedia, 2019, 22(3): 626-640.

[35] PENG M, WANG C, GAO Y, et al. Recognizing micro-expression in video clip with adaptive key-frame mining[EB/OL]. arXiv preprint arXiv: 2009.09179, 2020.

[36] GONG W J, ZHANG Y, WANG W, et al. Meta-MMFNet: meta-learning-based multi-model fusion network for micro-expression recognition[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2023, 20(2): 39.

[37] WANG F, LI J, QI C, et al. Multi-scale multi-modal micro-expression recognition algorithm based on transformer[EB/OL]. arXiv abs/: 2301.02969, 2023.

[38] WEI M T, ZHENG W M, ZONG Y, et al. A novel micro-expression recognition approach using attention-based magnification-adaptive networks[C]// ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2022: 2420-2424.

[39] WEI J S, LU G M, YAN J J, et al. Learning two groups of discriminative features for micro-expression recognition[J]. Neurocomputing, 2022, 479: 22-36.

[40] HONG J, LEE C, JUNG H. Late fusion-based video transformer for facial micro-expression recognition[J]. Applied Sciences, 2022, 12(3): 1169.

[41] ZHANG L F, HONG X P, ARANDJELOVIĆ O, et al. Short and long range relation based spatio-temporal transformer for micro-expression recognition[J]. IEEE Transactions on Affective Computing, 2022, 13(4): 1973-1985.

[42] HUANG X H, WANG S J, ZHAO G Y, et al. Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection[C]// 2015 IEEE International Conference on Computer Vision Workshop (ICCVW). Piscataway: IEEE Press, 2015: 1-9.

[43] ZHI R C, XU H R, WAN M, et al. Combining 3D convolutional neural networks with transfer learning by supervised pre-training for facial micro-expression recognition[J]. IEICE Transactions on Information and Systems, 2019, 102(5): 1054-1064.

[44] KHOR H Q, SEE J, LIONG S T, et al. Dual-stream shallow networks for facial micro-expression recognition[C]// 2019 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2019: 36-40.

[45] SUN B, CAO S M, LI D L, et al. Dynamic micro-expression recognition using knowledge distillation[J]. IEEE Transactions on Affective Computing, 2020, 13(2): 1037-1043.

A micro expression recognition method integrating LBP and parallel attention mechanism

LI Shuaichao[1]，LI Mingze[1]，SUN Jiaao[1]，LU Shuhua[1, 2, *]

(1. College of Information and Cyber Security，People's Public Security University of China，Beijing 102600，China；
2. Key Laboratory of Security Technology and Risk Assessment Ministry of Public Security，Beijing 102600，China)

Abstract：This research proposes a micro expression recognition network that incorporates LBP and parallel attention method to address the issues of small feature discrimination, background noise interference, and weak intensity of facial micro-expression changes. The network inputs the RGB image into the densely connected improved Shuffle Stage branch to extract the global features of the face and enhance the association of contextual semantic information. The LBP image is input into the local texture feature branch composed of a multi-scale layered convolutional neural network to extract detailed information. Following extraction of the dual-branch feature, the network backend implements a parallel attention technique to enhance feature fusion capabilities, reduce background noise, and concentrate on the micro-expression feature's interest region. The proposed method is tested on three public data sets including CASME, CASME II and SMIC, and the recognition is accurate The rates reached 85.18%, 74.53% and 81.19% respectively. The experimental results show that the proposed method effectively improves the accuracy of micro expression recognition, which is better than many current advanced methods.

Keywords：micro expression recognition；dense connection；Shuffle Stage branch；multi-scale layered convolution；parallel attention mechanism

Received：2023-04-27；**Accepted**：2023-06-30；**Published Online**：2023-07-13 15：10

URL：link.cnki.net/urlid/11.2625.V.20230713.1227.001

Foundation item：Double First-Class Innovation Research Project for People's Public Security University of China (2023SYL08)

* **Corresponding author**. E-mail：lushuhua@ppsuc.edu.cn

《北京航空航天大学学报》征稿简则

《北京航空航天大学学报》是北京航空航天大学主办的以航空航天科学技术为特色的综合性自然科学学术期刊(月刊)。本刊以反映航空航天领域研究成果与动态、促进学术交流、培养科技人才和推动科技成果向社会生产力转化为办刊宗旨。本刊为中国自然科学技术核心期刊,并被 Ei Compendex 等国内外权威文献检索系统收录。本刊向国内外公开发行,为进一步提高办刊质量和所刊出文章的学术水平,特制定本简则。

1 论文作者及内容

1.1 本刊面向海内外所有学者。

1.2 主要刊载与航空航天科学技术有关的材料科学与工程、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、计算机科学与应用技术、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 405,407 房间
电　　话:(010)82316698,82316981,82315499,82317448
E-mail: jbuaa@ buaa. edu. cn
http://bhxb. buaa. edu. cn
http://www. buaa. edu. cn

集群成员期刊

《北京航空航天大学学报》
第五届编辑委员会

主　任（主　编）：赵沁平

（以下按姓氏笔画为序）

副主任（副主编）：丁希仑　王少萍　孙志梅　李秋实　李焕喜　杨嘉陵
苗俊刚　相　艳　徐立军　钱德沛　曹晋滨

编　委：马殿富　王　琪　王　聪　邓小燕　王青云　王荣明　刘　宇
刘　红　江　洁　刘　强　闫　鹏　朱天乐　刘铁钢　齐铂金
陈万春　邹正平　苏东林　杨世春　沈成平　邱志平　宋知人
杨树斌　张晓林　杨晓奕　杨继萍　李惠峰　吴新开　张瑞丰
杨照华　宋凝芳　周　锐　林宇震　林贵平　战　强　姚仰平
胡庆雷　赵秋红　段海滨　赵巍胜　席　平　郭　宏　徐　洁
徐世杰　郭洪波　康　锐　翟　锦　熊华钢

编辑部主任：张　嵘

编　　辑：孙　芳　万玉敏　王　茜　卞欢欢　李亚泰

北京航空航天大学学报
Beijing Hangkong Hangtian Daxue Xuebao
（原《北京航空学院学报》）
（月刊　1956年创刊）
第51卷　第4期　2025年4月

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)
(Monthly, Started in 1956)
Vol.51 No.4 April 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

微信公众号　超星电子书　QQ群

国内定价：50.00元／期