

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

北京航空航天大学学报

JOURNAL OF BEIJING UNIVERSITY OF
AERONAUTICS AND ASTRONAUTICS

低空智能感知与安全技术——低空经济专刊

基于自适应改进蜚螂算法的高原森林灭火
多无人机路径规划

朱培 吕孝龙 陈凯森 宋睿 张剑高 邵荃

9

2026
Vol.52

》专刊背景

随着“低空经济”快速发展，低空空域已成为物流配送、旅游观光、城市管理（包括应急救援、城市安防、电力巡检等）、城市空中交通、应急响应等领域的重要战略资源。eVTOL（电动垂直起降飞行器）、无人机、飞行汽车等新型低空运载工具的广泛应用，对低空环境感知和低空安全提出了更高要求。低空智能感知技术涵盖低空环境多源感知与建模、低空自主导航与决策控制、低空空域管理、群体智能与协同感知等关键方向，是保障低空飞行安全、提升运行效率、实现规模化应用的核心。

当前，低空智能感知技术研究仍面临诸多挑战：复杂/动态环境下感知能力、多传感器协同能力、异构数据融合效率、极端天气或电磁干扰下鲁棒性等亟待提升，进而导致低空安全仍然存在短板。与此同时，人工智能、大数据、云计算、物联网、边缘计算、5G/6G 通信等新一代信息技术的突破，为低空智能感知技术发展以提供了新路径，对于提升低空安全至关重要。

鉴于此，为促进该领域的前沿探索与交叉融合，《北京航空航天大学学报》拟推出“低空智能感知与安全技术——低空经济”为主题的专刊，征集高质量原创研究成果，推动理论创新、技术落地和应用推广。热忱欢迎相关研究领域的专家学者提交高质量的研究论文，并关注此专刊的出版。



本专刊聚焦低空智能感知技术的理论突破、技术革新与应用实践，涵盖（但不限于）以下研究方向：

1 低空环境多源感知与建模

- 毫米波雷达、激光雷达、视觉传感器等多源异构数据融合
- 动静态目标（无人机、鸟类、气球等）检测与跟踪

2 低空自主导航与决策控制

- 动态避障路径规划与实时决策算法
- 人机协同下的意图理解与交互策略

3 低空空域管理技术

- 有人-无人航空器协同管理
- 低空交通流量感知与调度

4 标准体系与安全治理

- 智能感知系统的适航认证与标准化
- 复杂场景下的态势感知与风险评估

5 智能感知算法与系统设计

- 模型轻量化边缘部署
- 群体智能与协同感知

6 典型应用场景实践

- 城市空中交通（UAM）的感知-规划-控制
- 无人机物流配送中的末端精准感知
- 灾害救援中的低空快速响应系统

目次

低空智能感知与安全技术——低空经济专刊

基于自适应改进蜣螂算法的高原森林灭火多无人机路径规划	朱培, 吕孝龙, 陈凯森, 宋睿, 张剑高, 邵荃 (2987)
基于 YOLOv11-drone 的复杂背景下反无人机系统目标检测算法	薛珊, 杨文达, 李其贵, 任万飞 (3001)
面向无人机支线物流的起降场容量模型与协同调度	江波, 李蝶, 郑远, 李诚龙, 张照轩 (3011)
基于运行时保证的 DAA 系统运行安全保障方法设计	董磊, 王琦, 陈曦, 刘嘉琛 (3023)
导航辅助雷达量测增强的扩展型低空飞行器跟踪	李坤坤, 黎克波, 曹锐, 刘远贺, 胡爽 (3039)
民用无人系统运行场景架构设计与建模技术	郭泰, 安一平, 舒振杰, 蒋觉义, 江雨航 (3052)
面向无人机平台的 CNN-Transformer 混合模型训练后量化方法	赵呈功, 胡纪远, 杨星宇, 陈磊, 姜宏旭 (3066)
细粒度语义增强的民航图文跨模态检索方法	刘姝妍, 何柳, 曾江辉 (3075)
基于广播代理重加密的低空无人机运行数据安全共享方案	周泽全, 季晓海, 罗喜伶, 毛剑, 王珺珺 (3089)
密集无人机目标雷达非凸 DOA 估计方法	何贵栋, 雷鹏, 王俊 (3100)
基于大模型属性解析的空地行人检索方法	谭荃戈, 王蓉, 廖曼丞, 李欣 (3108)
位置感知与局部预交互的空-地行人重识别方法	毕艺瀚, 李冲, 王蓉 (3117)
基于 MTL-sLSTM 的低空载具锂电池 RUL 预测	张慧珊, 沈吉英, 刘东升, 周志凯, 胡一凡, 徐扬波 (3125)
面向不规则结构的低空感知模型剪枝算法	杨子, 庄连生 (3136)
低空通信环境下基于双层博弈的 LCU 级视频比特分配算法	张雅霏, 刘露露, 于金池, 朱迪锋, 龚轩, 高毅 (3146)
面向低空起降场的 FOD 检测与识别	郑皓文, 胡海苗, 徐壮, 何壮, 胡昊鑫 (3153)
低空安防场景下多目标高精度实时三维检测技术	翁佳楠, 徐壮, 李旻晖, 叶德茂 (3163)
基于伪标签增强与多特征融合的毫米波雷达无人机检测	鲁亚宁, 张生军, 于钱米, 李洪明 (3172)
基于无人机视角的轻量化绝缘子缺陷检测算法	胡道婕, 倪翠, 吴春坡, 刘著, 王朋 (3183)
面向低空场景的光场高斯泼溅表示及空间-角度同时超分辨率方法	孟佳涵, 杨驰, 李传君, 林丽莉, 周文晖 (3189)
基于模型数据混合驱动的干扰磁场补偿方法	张德慧, 张金生, 李婷, 马啸宇, 廖守亿, 倪静, 魏豪 (3202)
基于深度强化学习的高速无人机智能路由策略	寿一凡, 刘东升, 陈亚辉 (3211)
时空网格下多源要素建模与风险量化分析	刘龙浩, 茹乐, 汪洪桥, 王文飞, 张政浩, 李一帆 (3223)

CONTENTS

Low-Altitude Intelligent Sensing and Safety Technology: Special Issue on the Low-Altitude Economy

Multi-UAV path planning for forest fire-fighting in plateau based on adaptive improved dung beetle optimization algorithm
..... ZHU Pei, LYU Xiaolong, CHEN Kaisen, SONG Rui, ZHANG Jiangao, SHAO Quan (2987)

Object detection method for anti-UAV systems in complex backgrounds based on YOLOv11-drone
..... XUE Shan, YANG Wenda, LI Qigui, REN Wanfei (3001)

Capacity model and collaborative scheduling of droneports for drone regional logistics
..... JIANG Bo, LI Die, ZHENG Yuan, LI Chenglong, ZHANG Zhaoxuan (3011)

Design of operational security guarantee method for DAA system based on runtime assurance
..... DONG Lei, WANG Qi, CHEN Xi, LIU Jiachen (3023)

Extended low-altitude aircraft tracking via navigation-assisted radar measurement augmentations
..... LI Kunkun, LI Kebo, CAO Rui, LIU Yuanhe, HU Shuang (3039)

Architecture design and modeling technology for civil unmanned system operation scenarios
..... GUO Tai, AN Yiping, SHU Zhenjie, JIANG Jueyi, JIANG Yuhang (3052)

A post-training quantization method for CNN-Transformer hybrid models on UAV platforms
..... ZHAO Chenggong, HU Jiyuan, YANG Xingyu, CHEN Lei, JIANG Hongxu (3066)

Fine-grained semantic-enhanced cross-modal image-text retrieval method for civil aviation
..... LIU Shuyan, HE Liu, ZENG Jianghui (3075)

Secure sharing scheme for low-altitude UAV operational data based on broadcast proxy re-encryption
..... ZHOU Zequan, JI Xiaohai, LUO Xiling, MAO Jian, WANG Junjun (3089)

Nonconvex radar DOA estimation method for dense UAV targets
..... HE Guidong, LEI Peng, WANG Jun (3100)

Large model attribute parsing-based aerial-ground pedestrian retrieval method
..... TAN Quange, WANG Rong, LIAO Mancheng, LI Xin (3108)

Aerial-ground person re-identification method via position-aware and local pre-interaction
..... BI Yihan, LI Chong, WANG Rong (3117)

Remaining useful life prediction of lithium-ion batteries in low-altitude vehicles based on MTL-sLSTM
..... ZHANG Huishan, SHEN Jiying, LIU Dongsheng, ZHOU Zhikai, HU Yifan, XU Yangbo (3125)

Low-altitude perception models pruning algorithm with irregular structures
..... YANG Zi, ZHUANG Liansheng (3136)

LCU-level video bit allocation algorithm based on two-layer game in low-altitude communication environment
..... ZHANG Yafei, LIU Lulu, YU Jinchu, ZHU Difeng, GONG Xuan, GAO Yi (3146)

FOD detection and recognition for low-altitude takeoff and landing sites
..... ZHENG Haowen, HU Haimiao, XU Zhuang, HE Zhuang, HU Haoxin (3153)

High-accuracy real-time 3D multi-target detection for low-altitude security scenarios
..... WENG Jia'nan, XU Zhuang, LI Yanghui, YE Demao (3163)

Pseudo-label enhanced multi-feature fusion for millimeter-wave radar UAV detection
..... LU Yaning, ZHANG Shengjun, YU Qianmi, LI Hongming (3172)

Lightweight insulator defect detection algorithm based on UAV perspective
..... HU Xiaojie, NI Cui, WU Chunpo, LIU Zhu, WANG Peng (3183)

Light field gaussian splatting representation and joint spatial-angular super-resolution for low-altitude scenarios
..... MENG Jiahua, YANG Chi, LI Chuanjun, LIN Lili, ZHOU Wenhui (3189)

An interference magnetic field compensation method based on a model-data hybrid drive
..... ZHANG Dehui, ZHANG Jinsheng, LI Ting, MA Xiaoyu, LIAO Shouyi, NI Jing, WEI Hao (3202)

Intelligent routing strategy for high-speed UAVs via deep reinforcement learning
..... SHOU Yifan, LIU Dongsheng, CHEN Yahui (3211)

Multi-source element modeling and risk quantitative analysis under spatio-temporal grid
..... LIU Longhao, RU Le, WANG Hongqiao, WANG Wenfei, ZHANG Zhenghao, LI Yifan (3223)

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0846

基于自适应改进蜣螂算法的高原森林灭火 多无人机路径规划

朱培*, 吕孝龙, 陈凯森, 宋睿, 张剑高, 邵荃

(南京航空航天大学 民航学院, 南京 211106)

摘 要: 针对高原森林复杂环境下无人机灭火救援路径规划问题, 提出基于自适应改进蜣螂 (AIDBO) 算法的多无人机路径规划方法。基于数字高程模型 (DEM) 数据建立高原森林山区三维地理空间模型, 考虑复杂地形、环境风、海拔等因素的影响, 以最小飞行时间、风险代价、能耗代价为目标函数, 建立从救援临时基地起飞, 抵达可能存在灾害区域的多约束条件下多无人机路径规划模型; 通过混沌映射与反向学习优化种群初始化, 引入鲸鱼优化算法 (WOA) 的螺旋搜索策略丰富位置更新机制, 采用自适应柯西变异增强种群摆脱局部最优的能力, 提出一种 AIDBO 算法进行求解。结果表明: AIDBO 算法平均适应度较传统蜣螂 (DBO) 算法、改进蜣螂 (IDBO) 算法分别提升 10.59%、9.82%; 与平原海拔 340 m 环境相比, 海拔 2 400 m 环境下无人机最长飞行时间略微增加了 0.997%, 而最大飞行能耗代价增加了 58.96%; 同时, 风向和风速也会对飞行时间和能耗有一定影响。所提 AIDBO 算法能显著优化高原森林多无人机路径, 为高原森林无人机灭火救援调度提供决策支持。

关 键 词: 多无人机; 路径规划; 高原森林; 灭火救援; 改进蜣螂算法; 自适应

中图分类号: U116; V328.4

文献标志码: A

文章编号: 1001-5965(2026)09-2987-14

高原森林复杂环境下, 因地形复杂、地面救援受限等, 灭火救援难度大, 而无人机具有快速机动、不受地形影响的优点, 将逐步成为重要的灭火救援手段。然而, 高海拔低气压、陡峭地形及强风突变等高原特殊条件, 给无人机的飞行性能与安全等带来严峻挑战。因此, 开展面向高原森林复杂环境的灭火无人机路径规划优化研究可为实际灭火救援调度决策提供理论支持。

针对森林火灾环境下的救援路径规划问题, 潘卫军等^[1]以救援航空器升限、航空器最小离地高度、燃油量等性能为约束限制, 提出一种自适应动态调整权重的路径规划求解算法; 郑玉斌等^[2]考虑部

分可观测环境、历史观测信息、高维动作空间和奖励稀疏性等因素, 提出基于门控循环单元-近端策略优化 (gated recurrent unit-proximal policy optimization, GRU-PPO) 的多无人机协同追捕方法; 滕菲等^[3]考虑状态空间、动作空间、动态障碍物和安全约束等因素, 提出基于改进 PPO 算法的无人机动态避障规划方法; 朱元军等^[4]考虑城市建筑障碍、运行风险、路径长度等因素, 提出基于风险约束 A^* (risk-constrained A^* , RC- A^*) 算法的无人机城市空域安全路径规划方法; 朱弥雪^[5]提出了考虑无人机转弯、悬停能耗、地形坡度、避障能力等的无人机飞行路径优化模型; 余顺金等^[6]提出一种基于改进人工势

收稿日期: 2025-12-08; 录用日期: 2026-03-06; 网络出版时间: 2026-04-02 09:01

网络出版地址: link.cnki.net/urlid/11.2625.V.20260401.0915.002

基金项目: 国家自然科学基金 (52202416); 民航应急科学与技术重点实验室开放基金 (NJ2022022)

*通信作者. E-mail: zpei@nuaa.edu.cn

引用格式: 朱培, 吕孝龙, 陈凯森, 等. 基于自适应改进蜣螂算法的高原森林灭火多无人机路径规划 [J]. 北京航空航天大学学报, 2026, 52 (9): 2987-3000. Zhu P, Lyu X L, Chen K S, et al. Multi-UAV path planning for forest fire-fighting in plateau based on adaptive improved dung beetle optimization algorithm [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 2987-3000 (in Chinese).

场法的森林防火无人机路径规划模型;李锐君等^[7]考虑灭火无人机路径长度、障碍物威胁、航程高度及水平转角等因素,提出了基于粒子群(particle swarm optimization, PSO)算法的灭火无人机路径规划方法;Gao 等^[8]建立了考虑误差修正约束和航迹约束的多目标规划模型。除地形、气象环境等对无人机性能的影响,森林火场风险威胁同样影响其飞行航线规划。Demir、Radmanesh 和 Aggarwal^[9-11]等基于火灾风险地图和火场辐射威胁,提出集群无人机路径规划方法;Çoğay 等^[12]提出了一种以无人机能耗最小、火点覆盖率最大为目标,考虑无人机飞行安全高度、飞行时间及能耗约束的路径规划模型;Hasan 和 Xu^[13-14]等分析了风、高温及地形特征对火灾发生的影响,提出了综合考虑无人机、地形、动态障碍物及气象数据的路径规划方法。当在高海拔森林山区环境下执行救援任务时,随着海拔增加,空气密度降低,将对无人机动力学性能产生影响,进而可能影响无人机飞行时间和能耗成本,然而,现有研究较少考虑高海拔低气压等特殊环境的影响。

在无人机路径规划方法方面,元启发式算法因其良好的全局搜索能力和对复杂约束问题的适应性而被广泛应用。经典元启发式算法中,PSO 算法结构简单,但易陷入局部最优;遗传算法(genetic algorithm, GA)全局搜索能力强但收敛慢;鲸鱼优化算法(wale optimization algorithm, WOA)在探索与开发间平衡较好,但在高维复杂环境中仍可能过早收敛。而受蜣螂的滚球、舞蹈、觅食和繁殖等行为启发而被提出的蜣螂(dung beetle optimization, DBO)算法^[15]有着较强的全局探索与局部开发能力,但在复杂环境下,传统 DBO 算法在全局优化方面存在一定局限。众多学者对 DBO 算法进行了改进。Lyu 等^[16]采用混沌映射增强种群多样性,并提出一种全局探索策略取代了 DBO 算法的滚动阶段,改善了信息交换,并最大限度地减少了参数依赖性;Chen 等^[17]将指数递减的惯性权重纳入算法中,提高了收敛速度和探索能力;Zhu 等^[18]提出了产卵蜣螂和觅食蜣螂数量之间的收敛因子和动态平衡,使得算法能够在早期阶段专注于全局搜索;Shen 等^[19]使用 Beta 分布动态生成反射解决方案,以探索更多的搜索空间,并允许个体跳出局部最优值,但增加了算法求解时间;刘微等^[20]引入正余弦搜索算法,减少结果陷入局部最优的可能。然而,现有路径规划研究中往往较多考虑无人机自身约束限制,但面对高原复杂环境下缺乏自适应能力,算法易陷入局部最优,或全局搜索时难以快速找到最优解。

因此,本文针对高原环境下灭火救援无人机路径规划问题,建立多约束下的多目标路径规模模

型,并提出自适应改进蜣螂(adaptive improved dung beetle optimization, AIDBO)算法求解,为实际高原山区灾害救援提供决策支撑。

1 问题描述与模型建立

1.1 问题描述与假设

高原森林地区的复杂地形和强风会对无人机的飞行速度产生影响,且高海拔和低气压会导致大气密度变化,进而影响无人机的升力和其他空气动力学特性,导致飞行过程中需要更多的能量以保持稳定飞行。为确保无人机能够快速、安全地到达救援区域并有效执行任务,必须对无人机的飞行路径进行优化,综合考虑飞行时间、能量消耗和潜在风险,确保救援任务的高效完成。

基于上述问题,本文做如下假设:

- 1) 通过各种技术手段已知救援存在区域范围;
- 2) 路径规划任务的起点和终点已知;
- 3) 无人机飞行空速保持恒定;
- 4) 无人机携带的设备(如传感器、摄像头等)产生的能耗忽略不计,仅计算飞行所消耗的能耗;
- 5) 无人机飞行环境的环境信息已知;
- 6) 所有无人机机型相同;
- 7) 不考虑无人机发生故障等情况。

1.2 三维地理空间建模

为更真实地模拟森林地形环境,采用栅格法基于真实数字高程模型(digital elevation model, DEM)数据,结合函数模拟模型对真实山地环境进行建模。将基准地形函数(式(1))和山峰函数(式(2))作为一种噪声源加入 DEM 数据中,用来模拟地形的微小起伏,增加地形的复杂性。

$$Z_1(x,y)=\sin(y+a)+b\sin(x)+c\cos\left(d\sqrt{x^2+y^2}\right)+e\cos(y)+f\sin\left(g\sqrt{x^2+y^2}\right)\quad(1)$$

式中: $Z_1(x,y)$ 为使用基准地形函数公式生成的高程数据; a 、 b 、 c 、 d 、 e 、 f 、 g 为常数,用于调节函数的形状和特征,不同的取值组合可以得到不同的地形特征,本文依据参考文献[21]分别取值 8.0, 0.1, 0.1, 0.4, 1.0, 0.1, 0.1; x 、 y 为投影水平坐标点。

$$Z_2(x,y)=\sum_{i=1}^kh_{0i}\exp\left[-\left(\frac{x-x_{0i}}{x_{si}}\right)^2-\left(\frac{y-y_{0i}}{y_{si}}\right)^2\right]\quad(2)$$

式中: $Z_2(x,y)$ 为使用山峰函数公式生成的高程数据; h_{0i} 取决于第 i 座山体的高度; (x_{0i}, y_{0i}) 用于描述第 i 座山峰的基本形状,如坡度衰减值; k 为山峰数量; x_{si} 、 y_{si} 表示第 i 座山峰分别沿 x 、 y 轴方向与倾斜坡度相关的参数,决定山峰的陡峭程度,取值越小山峰越陡峭,取值越大山峰越平缓。

将原始 DEM 数据、基准地形数据与山峰地形数据融合, 融合后的地形数据可以更为精确地模拟出复杂的森林地形。融合后的地形数据可表示为

$$z(x,y)=Z_0(x,y)+Z_1(x,y)+Z_2(x,y) \tag{3}$$

式中: $z(x,y)$ 为最终融合后的地形高程数据; $Z_0(x,y)$ 为原始 DEM 高程数据。

1.3 模型建立

1.3.1 目标函数

高原森林救援无人机路径规划需考虑多种影响因素, 本文以飞行时间、风险代价、能耗代价最小为目标函数, 考虑飞行安全距离, 建立从起飞站点到灾害区域的多无人机路径规划模型。

1) 飞行时间代价。无人机飞行时间代价用于衡量无人机在完成任务时所需的最短飞行时间。时间代价模型以降低执行任务的总时间为目标, 通过优化路径规划以尽可能缩短无人机在任务执行过程中花费的时间。第 $i-1$ 个路径点到第 i 个路径点之间的时间 $T_{i-1,i}$ 可以通过式(4)表达:

$$T_{i-1,i}=\frac{l_{i-1,i}}{v_{d,i-1,i}} \tag{4}$$

式中: $l_{i-1,i}$ 为路径上第 $i-1$ 个路径点到第 i 个路径点之间的距离; $v_{d,i-1,i}$ 为无人机地速。在无风情况下, 无人机的飞行速度(空速)不受环境风速的影响, 此时空速与地速相等。然而, 当有风存在时, 无人机的地速则为空速与风速的矢量之和。

无人机地速 $v_{d,i-1,i}$ 表示为

$$\begin{bmatrix} \cos \theta_{i-1,i} \\ \sin \theta_{i-1,i} \end{bmatrix} v_{d,i-1,i} = \begin{bmatrix} \cos \varphi_{i-1,i} & \cos \phi_{i-1,i} \\ \sin \varphi_{i-1,i} & \sin \phi_{i-1,i} \end{bmatrix} \begin{bmatrix} v_a \\ v_{wind,i-1,i} \end{bmatrix} \tag{5}$$

式中: v_a 为无人机空速; $v_{wind,i-1,i}$ 为第 $i-1$ 个路径节点到第 i 个路径节点的风速; $\theta_{i-1,i}$ 为地速航向角; $\varphi_{i-1,i}$ 为空速航向角; $\phi_{i-1,i}$ 为第 $i-1$ 个路径节点到第 i 个路径节点的风向。

无人机总飞行时间表示为

$$T=\sum_{i=1}^n T_{i-1,i} \tag{6}$$

因此, 为获取无量纲飞行时间代价, 定义实际飞行时间 T 与理论最短直线飞行时间 T_0 的比值为飞行时间代价, 如下:

$$f_T=\frac{T}{T_0} \tag{7}$$

式中: $T_0=L_0/v_d$, T_0 为起始点到目标点直线飞行时间, L_0 为起始点到目标点距离。 $f_T=1$ 表示无人机无干扰沿直线飞行, 为理想下限; $f_T>1$ 则反映了规避地形威胁, 大气威胁保持飞行安全等约束所必须付

出的额外时间成本; 最小化飞行时间代价 f_T 即是在满足所有安全与物理约束的前提下, 尽可能逼近最短时间路径, 其取值范围为接近 1 的最优值。

2) 风险代价。山区中由于地形的阻挡, 通信信号可能存在盲区, 导致可能失去与地面控制站通信的风险。为降低无人机飞向盲区的概率, 将整个无人机飞行的林场环境按照通信盲区、地形威胁等因素预先划分为多个风险等级的区域, 并定义风险值范围 (0,1) 区间。为综合评估整个飞行路径的风险, 可以将路径上每一个点的风险值加权求和, 得到路径的总风险值 $R(C_1)$, 表示为

$$R(C_1)=\sum_{i=1}^n (R(C_1^i) l_{i-1,i}) / 1\,000 \tag{8}$$

式中: $R(C_1^i)$ 为路径上第 i 个点的风险值; $l_{i-1,i}$ 为路径上第 $i-1$ 个路径点到第 i 个路径点之间的距离, 表示权重; $R(C_1)$ 取值范围处于 (0,1) 区间。

同时, 森林山区存在的风切变等气象环境将影响飞行安全, 将此类区域定义为大气威胁区。为尽可能避免飞入大气威胁区, 本文将其视为圆柱体, 半径为 R_w , 如图 1 所示。图中: O_w 为第 w 个大气威胁区的中心点位置。同时, 为减少由于过度避开大气威胁区而增加的飞行路径, 增加安全冗余度, 对大气威胁区半径进行裕度扩展, 扩展宽度为 D_{mar} , 无人机主要在扩展区内飞行。因此, 大气威胁区对无人机的威胁成本代价 $R(C_2)$ 可用式 (9) 计算:

$$\begin{cases} R(C_2)=\sum_{i=1}^n \sum_{w=1}^q K_{i-1,i,w} \\ K_{i-1,i,w}=1-\frac{g_{i-1,i}^w}{R_w+D_{mar}} \quad R_w < g_{i-1,i}^w < R_w+D_{mar} \end{cases} \tag{9}$$

式中: $K_{i-1,i,w}$ 为第 $i-1$ 个路径点到第 i 个路径点之间的威胁代价; $g_{i-1,i}^w$ 为第 $i-1$ 个路径点到第 i 个路径点连线到 O_w 点的垂直距离, 即在区间 (R_w, R_w+D_{mar}) 范围内。因此, $K_{i-1,i,w}$ 取值为 (0,1)。大气威胁值 $R(C_2)$ 为有限个大气威胁区的威胁代价值总和。

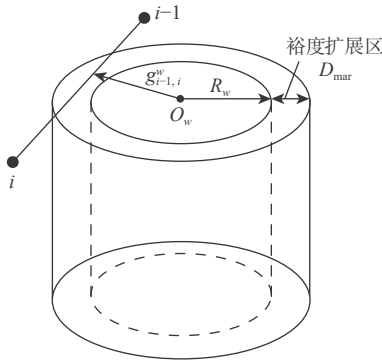


图 1 大气威胁区示意图

Fig. 1 Schematic diagram of atmospheric threat zone

无人机路径风险代价 f_R 可表示为

$$f_R = R(C_1) + R(C_2) \quad (10)$$

3) 能耗代价。飞行能耗影响续航时间和任务执行效率, 决定灭火救援任务能否有效完成。无人机从第 $i-1$ 个路径点到第 i 个路径点所消耗功率 $P_{i-1,i}$ 可表示为

$$P_{i-1,i} = P_{i-1,i}^{\text{blade}} \left(1 + \frac{3v_{d,i-1,i}^2}{U_{\text{tip}}^2} \right) + \frac{P_{i-1,i}^{\text{induced}} v_0}{v_{d,i-1,i}} + \frac{1}{2} d_0 \rho_{i-1,i} s A v_{d,i-1,i}^3 \quad (11)$$

式中: $P_{i-1,i}^{\text{blade}}$ 和 $P_{i-1,i}^{\text{induced}}$ 分别为无人机悬停状态下的叶型功率和悬停状态下的诱导功率; U_{tip} 为旋翼叶片的叶尖速度; v_0 为悬停状态下平均转子诱导速度; d_0 为机身阻力比; s 为转子固体度; A 为转子盘面积; ρ 为大气密度。

$P_{i-1,i}^{\text{blade}}$ 和 $P_{i-1,i}^{\text{induced}}$ 分别表示为

$$P_{i-1,i}^{\text{blade}} = \frac{\delta}{8} \rho_{i-1,i} s A \Omega^3 R_r^3 \quad (12)$$

$$P_{i-1,i}^{\text{induced}} = (1 + \kappa) \frac{W^{\frac{3}{2}}}{\sqrt{2} \rho_{i-1,i} A} \quad (13)$$

式中: δ 为剖面阻力系数; Ω 为叶片角速度; R_r 为转子半径; κ 为对感应功率的增量修正系数; W 为无人机质量。大气密度 ρ 会随着海拔升高而降低, 其与温度及大气压力的关系可表示如下:

$$\rho = \frac{P_a}{0.2869 (T_{\text{temp}} + 273.1)} \quad (14)$$

式中: T_{temp} 为摄氏温度; P_a 为环境压强。

依据大气压力与海拔间的经验公式:

$$P_a = 101.3 \left[1 - 0.0255 \frac{H}{1000} \left(\frac{6357}{6357 + \frac{H}{1000}} \right) \right]^{5.256} \quad (15)$$

式中: H 为海拔。

因此, 考虑海拔影响的无人机总能耗可表示为

$$\sum_{i=1}^n E_{i-1,i} = \sum_{i=1}^n \frac{l_{i-1,i}}{v_{d,i-1,i}} P_{i-1,i} \quad (16)$$

式中: $l_{i-1,i} = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 + (z_i - z_{i-1})^2}$ 为各路径段长度; $P_{i-1,i}$ 为无人机消耗功率。

无人机的能耗代价 f_E 可表示为

$$f_E = \frac{\sum_{i=1}^n E_{i-1,i}}{E_{s \rightarrow p}} \quad (17)$$

式中: $\sum_{i=1}^n E_{i-1,i}$ 为无人机总能耗; $E_{s \rightarrow p}$ 为起始点到目标点直线距离能耗。 $f_E = 1$ 表示无人机沿直线无干

扰飞行的能耗代价; $f_E > 1$ 则反映了规避地形威胁, 大气威胁保持飞行安全等约束需付出的额外能耗成本。因此, 最小化 f_E 即在满足所有安全与物理约束的前提下, 尽可能逼近最小能耗代价, 取值范围为接近 1 的最优值。

4) 总代价目标函数。综上, 单架高原森林灭火无人机路径规划模型目标函数可表示为

$$f = \omega_1 f_T + \omega_2 f_R + \omega_3 f_E \quad (18)$$

式中: $\omega_1 + \omega_2 + \omega_3 = 1$ 。

多灭火无人机路径规划模型目标函数表示为

$$f' = \omega_1 f_{T,\text{sum}} + \omega_2 f_{R,\text{sum}} + \omega_3 f_{E,\text{sum}} \quad (19)$$

式中: $f_{T,\text{sum}}$ 、 $f_{R,\text{sum}}$ 、 $f_{E,\text{sum}}$ 分别为所有无人机总飞行时间代价、总风险代价、总能耗代价, 即所有无人机代价求和值。本文中, 为体现“打早、打小”的火灾扑救原则对快速响应的要求, 时间权重设置为 $\omega_1 = 0.4$; 为确保路径安全, 避免无人机受到大气威胁而失联或坠毁, 风险权重设置为 $\omega_2 = 0.2$; 由于高海拔对无人机续航能力影响较大, 保证无人机能耗极为关键, 能耗权重设置为 $\omega_3 = 0.4$ 。

1.3.2 约束条件

1) 飞行高度约束。设无人机在第 i 个路径点飞行高度为 h_i , 此时对应的地形高度为 $h_{i,m}$, 无人机需要与山体保持一定高度距离 h_{safe} 防止撞击山体, 同时, 飞行高度应小于最大飞行高度值 H_{max} 。

$$h_{i,m} + h_{\text{safe}} \leq h_i \leq H_{\text{max}} \quad (20)$$

2) 俯仰角约束。为确保无人机能够在不同飞行阶段保持适当的姿态并有效应对外部环境的变化, 必须对俯仰角进行合理的约束。

$$0 \leq \mu_i = \frac{|z_i - z_{i-1}|}{\sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}} \leq \tan \mu_{\text{max}} \quad (21)$$

式中: μ_i 为无人机在第 i 个路径点的俯仰角; μ_{max} 为最大俯仰角。

3) 能耗约束。无人机的任务执行能力受到电池容量和能耗效率的制约, 需设定给定任务的能耗约束, 确保其具备足够续航时间完成任务。

$$\sum_{i=1}^n E_{i-1,i} \leq E_{\text{total}} \quad (22)$$

式中: E_{total} 为无人机路径规划任务总能耗。

4) 无人机间安全约束。多无人机执行任务时, 需设置安全距离限制防止碰撞。将无人机体积简化为质点, 并假设 2 个无人机分别为无人机 p 、无人机 q , 每架无人机之间的距离为 d_{pq} , 将无人机之间的最小安全距离设定为 $d_{pq,\text{min}}$, 则需满足:

$$d_{pq} \geq d_{pq,\text{min}} \quad (23)$$

2 算法设计

为解决陷入局部最优的问题,本文提出了一种AIDBO算法,提升其全局搜索能力和适应性。该算法通过基于混沌映射与反向学习的种群初始化,减少初始解的集中性;引入WOA的螺旋搜索策略,丰富位置更新机制;采用自适应柯西变异策略增强蜣螂种群摆脱局部最优解的能力,有效扩展搜索空间范围。

2.1 传统蜣螂算法

DBO算法将种群分为4类:滚球蜣螂、繁殖蜣螂、觅食小蜣螂、偷窃蜣螂。

2.1.1 滚球蜣螂位置更新机制

滚球蜣螂分为有障碍模式和无障碍模式。当无障碍时,执行全局搜索,快速搜索整个三维空间,找到初步的候选路径点,其位置更新如下:

$$X_i(t+1) = X_i(t) + k\alpha X_i(t-1) + b_n \Delta X \quad (24)$$

式中: t 为迭代次数; $X_i(t)$ 和 $X_i(t+1)$ 分别为第 t 次和第 $t+1$ 次迭代时滚球蜣螂的位置信息; $k \in (0, 0.2]$ 为偏转系数; α 为自然数,用于表示是否偏离原来的方向, $\alpha = 1$ 时表示没有偏离, $\alpha = -1$ 时表示偏离原来的滚球方向; b_n 为属于 $(0, 1)$ 的自然系数,取 0.3; $\Delta X = |X_i(t) - X^w|$ 用于模拟光强,即距离全局最差解 X^w 的距离。

当蜣螂遇到障碍物会通过舞蹈重新定位以寻找新的路径,其舞蹈行为可用正切函数模拟,从而获取新的滚动方向。滚球蜣螂的位置更新如下:

$$X_i(t+1) = X_i(t) + \tan \theta_d |X_i(t) - X_i(t-1)| \quad (25)$$

式中: $\theta_d \in [0, \pi]$ 。

2.1.2 繁殖蜣螂位置更新机制

繁殖蜣螂进行局部区域精细搜索,以找到安全通过障碍物的路径,其位置更新如下:

$$B_i(t+1) = X^* + (B_i(t) - L_b^*) \times b_1 + (B_i(t) - U_b^*) \times b_2 \quad (26)$$

$$\begin{cases} L_b^* = \max((1 - R_c)X^*, L_b) \\ U_b^* = \min((1 + R_c)X^*, U_b) \end{cases} \quad (27)$$

式中:“ $\times$ ”表示逐元素相乘; $B_i(t)$ 和 $B_i(t+1)$ 分别为第 t 次和第 $t+1$ 次迭代时繁殖蜣螂的位置信息; L_b^* 和 U_b^* 分别为产卵最优区域的下界和上界; $R_c = 1 - t/t_{\max}$, R_c 为动态变化的参数,用于控制产卵区域随迭代次数动态缩小的速度, $t_{\max}$ 为最大的迭代次数; X^* 为当前种群的局部最优位置; b_1 、 b_2 为 $1 \times D$ 大小的独立随机向量, D 为优化问题的维度。

2.1.3 觅食小蜣螂位置更新机制

觅食小蜣螂负责探索新区域,通过大范围搜索帮助跳出局部最优,增加搜索多样性,位置更新方式如下:

$$X_i(t+1) = X_i(t) + (X_i(t) - L_b^b)C_1 + (X_i(t) - U_b^b) \times C_2 \quad (28)$$

式中:

$$\begin{cases} L_b^b = \max((1 - R_c)X^b, L_b) \\ U_b^b = \min((1 + R_c)X^b, U_b) \end{cases} \quad (29)$$

式中: L_b^b 和 U_b^b 分别为小蜣螂最优觅食区的下界和上界; X^b 为全局最优位置; $C_1 \sim N(0, 1)$ 为随机数并服从正态分布; C_2 为 $(0, 1)$ 的随机向量。

2.1.4 偷窃蜣螂位置更新机制

偷窃蜣螂在路径优化的中后期,偷窃蜣螂通过模仿其他个体的优秀解来改进自身路径,位置更新公式如下:

$$X_i(t+1) = X^b + Sg(|X_i(t) - X^*| + |X_i(t) - X^b|) \quad (30)$$

式中: g 为 $1 \times D$ 大小的正态分布随机向量; S 为常数。

2.2 自适应改进蜣螂算法

2.2.1 种群初始化改进

由于在初始化阶段缺乏参考的全局最优信息,初始种群应尽可能均匀覆盖搜索空间,以减少对全局探索的依赖,从而更快速地找到全局最优解。传统 DBO 算法的初始种群通常采用随机生成方式,可能导致种群分布不均、多样性不足,影响全局搜索效率。首先,采用 Tent 混沌映射生成具有遍历性和均匀性的初始解,避免种群过早聚集,如下:

$$\varphi_{n+1} = \begin{cases} \frac{\varphi_n}{\gamma} & 0 < \varphi_n < \gamma \\ \frac{1 - \varphi_n}{1 - \gamma} & \gamma \leq \varphi_n \leq 1 \end{cases} \quad (31)$$

式中:混沌映射状态值 $\gamma \in (0, 1)$; φ_n 为当前迭代步的混沌状态值; φ_{n+1} 为下一步的混沌状态值。

其次,为增强种群的多样性和探索能力,加快算法的收敛速度,并有效避免陷入局部最优,采用反向学习方法生成与当前解对立的解(反向解),并对比这 2 个解的适应度值,选择适应度较优的解进行下一次迭代。然而,由于 2 个解的距离为固定值,缺少随机性,引入随机因子 J 进一步增强该策略在全局搜索能力:

$$X_{nd} = L_b + y_{nd}(U_b - L_b) \quad (32)$$

$$\hat{X}_{nd} = U_b + L_b - JX_{nd} \quad (33)$$

式中: X_{nd} 为第 n 个种群的第 d 维数; y_{nd} 为由 Tent 混沌变换产生的混沌序列; U_b 和 L_b 分别为搜索的上界

和下界; $\hat{\mathbf{X}}_{nd}$ 为 $\mathbf{X}_{nd}$ 的反向解, 即反向种群个体; J 为 $(0,1)$ 间的随机数。将原始种群与其对应的反向种群合并生成新种群, 计算新种群中每个个体的适应度函数值, 对其进行排序, 最终选择适应度最优的 N 个个体作为新的初始种群。

2.2.2 融合 WOA 螺旋搜索策略的位置更新方法

WOA 凭借其独特的螺旋搜索策略, 在处理复杂优化问题时展现了强大的潜力。DBO 算法中, 繁殖蛱蝶的位置更新主要依赖于局部最优信息, 在复杂高维环境中易导致搜索范围受限。为提升 DBO 算法的性能, 本文将 WOA 的螺旋搜索策略融入繁殖蛱蝶的位置更新机制, 从而增强其全局搜索能力, 在复杂的优化环境中实现更高效的探索与开发。螺旋搜索公式如下:

$$\mathbf{X}_i(t+1) = \mathbf{D}'e^{cr^*} \cos(2\pi r^*) + \mathbf{X}^* \tag{34}$$

$$\mathbf{D}' = |\mathbf{X}^* - \mathbf{X}_i(t)| \tag{35}$$

式中: $\mathbf{D}'$ 为个体和当前最优位置之间的距离; r^* 为 $[0,1]$ 间的随机数; c 为决定对数螺旋形状的常量系数。当 c 取值较大时, 易导致衰减过快从而陷入局部最优, 当 c 取值较小时会导致收敛较慢。因此, 引入动态螺旋搜索形状的参数 cr^* , 表示为

$$cr^* = \exp\left(c \cdot \cos\left(\frac{\pi t}{t_{\max}}\right)\right) \tag{36}$$

因此, 繁殖蛱蝶位置更新公式表示为

$$\mathbf{B}_i(t+1) = \mathbf{X}^* + e^{cr^*} \cos(2\pi r^*) (\mathbf{B}_i(t) - \mathbf{L}_b^*) \times \mathbf{b}_1 + e^{cr^*} \cos(2\pi r^*) (\mathbf{B}_i(t) - \mathbf{U}_b^*) \times \mathbf{b}_2 \tag{37}$$

2.2.3 自适应柯西变异策略

在 DBO 算法的迭代收敛后期, 蛱蝶种群通常会集中在最优个体附近, 可能导致整体多样性的降低, 陷入局部最优。本文引入了一种自适应柯西变异策略, 用于调整蛱蝶种群的位置, 该策略不仅增强了算法摆脱局部最优解的能力, 还扩展了搜索空间范围, 其公式如下:

$$\mathbf{Y}^b(t) = \mathbf{X}^b(t) + \mathbf{X}^b(t) \cdot \text{Cauchy}(0, 1) \tag{38}$$

式中: $\mathbf{Y}^b(t)$ 为第 t 次迭代中当前最优位置在柯西变异后的解; $\text{Cauchy}(0, 1)$ 为柯西变异算子, 表示为

$$\text{Cauchy}(x) = \frac{1}{\pi} \left(\frac{1}{x^2 + 1} \right) \quad -\infty < x < \infty \tag{39}$$

因此, 本文提出的 AIDBO 算法是在传统 DBO 算法基础上, 融合了混沌映射与反向学习的种群初始化策略、WOA 螺旋搜索策略和自适应柯西变异策略。

2.3 算法求解流程

本文基于上述改进策略提出了 AIDBO 算法求

解流程。在算法求解中, 为保证生成的路径满足约束条件, 采用剔除策略对无效路径进行约束处理, 即在种群初始化阶段, 对每个个体进行约束检查, 剔除不满足约束的个体并重新生成, 直至初始种群全部满足约束条件; 在每次迭代中, 滚球蛱蝶、繁殖蛱蝶、觅食小蛱蝶和偷窃蛱蝶的位置更新后, 再次进行约束检查, 若新位置违反约束, 则将该个体标记为无效, 不参与本次迭代的适应度评估和后续的变异操作。通过这种方式, 确保算法搜索过程中的所有候选路径均为物理可行解。路径规划问题求解流程如图 2 所示, 算法求解具体步骤如下:

- 输入:** 最大迭代次数 $T_{\max}$, 种群大小 N 。
输出: 最优位置 $\mathbf{X}^b$ 及对应的适应度 $f_{\min}$ 。
1. 利用混沌映射与反向学习进行种群初始化
 $i \leftarrow 1, 2, \dots, N$ 。
 2. while $t \leq T_{\max}$ do
 3. for i 属于滚球蛱蝶 do
 4. $\alpha = \text{rand}(1)$
 5. if $\alpha \leq 0.9$ then
 6. 使用式 (24) 更新滚球蛱蝶位置。
 7. 检查新位置是否满足约束, 若不满足则重新生成或标记为无效, 无效值不参与后续步骤。
 8. else
 9. 使用式 (25) 模拟存在障碍物情况下的位置更新。

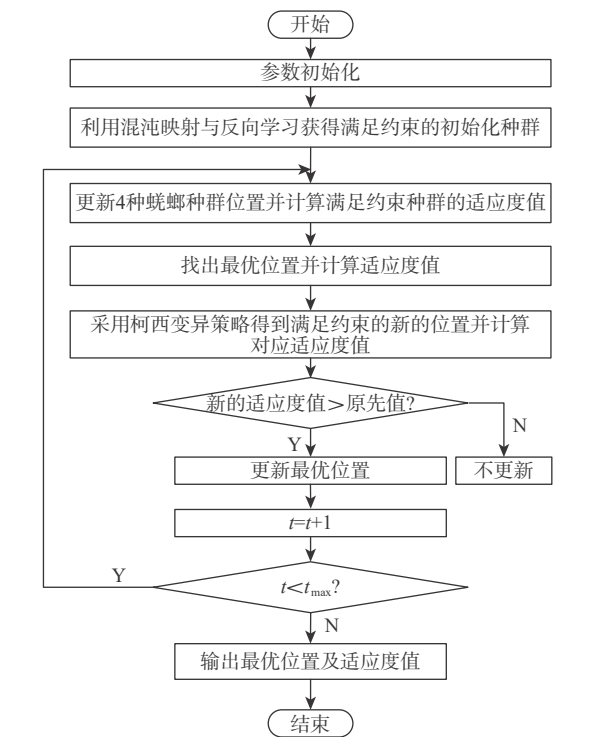


图 2 自适应改进蛱蝶优化算法流程
Fig. 2 Flowchart of adaptive improved dung beetle optimization algorithm

10. 检查新位置是否满足约束, 若不满足则重新生成或标记为无效, 无效值不参与后续步骤。

11. end if

12. end for

13. for i 属于繁殖蜣螂 do

14. 使用式 (27) 确定繁殖蜣螂更新范围, 并使用式 (37) 更新繁殖蜣螂位置。

15. 检查新位置是否满足约束, 若不满足则重新生成或标记为无效, 无效值不参与后续步骤。

16. end for

17. for i 属于觅食小蜣螂 do

18. 使用式 (29) 确定小蜣螂觅食更新范围, 并使用式 (28) 更新小蜣螂位置。

19. 检查新位置是否满足约束, 若不满足标记为无效, 无效值不参与后续步骤。

20. end for

21. for i 属于偷窃蜣螂 do

22. 使用式 (30) 更新偷窃蜣螂位置。

23. 检查新位置是否满足约束, 若不满足标记为无效, 无效值不参与后续步骤。

24. end for

25. 选择当前最优解, 利用式 (38) 实现自适应柯西变异。

26. 检查变异后的解是否满足约束, 若不满足则放弃变异, 保留原解。

27. if 变异后的解比当前解更优 then

28. 更新最优解。

29. end if

30. $t=t+1$

31. end while

32. return 适应度 $f_{\min}$

3 算例分析

3.1 算法有效性

3.1.1 基准函数测试

为验证本文提出的 AIDBO 算法的有效性, 选用 IEEE 发布的 CEC200 测试集作为评估标准。测试中选取 3 个单峰基准函数和 3 个多峰基准函数进行验证。其中, F_1 、 F_2 、 F_3 为单峰函数, 用于评估算法的收敛速度及局部开发能力; F_4 、 F_5 、 F_6 为多峰函数, 用来测试算法的全局探索性能。函数具体表达式如表 1 所示。

实验设置最大迭代次数为 500, 种群规模为 30。滚球蜣螂、繁殖蜣螂、觅食小蜣螂、偷窃蜣螂的数量分别设置为 6、6、7、11。对 WOA、传统 DBO 算法、IDBO 算法^[17]、3 类改进策略组合算法和 AIDBO 算法分别进行 10 次独立测试。其中, 3 类

表 1 测试函数表达式

Table 1 Test Function Expression

函数	表达式	维度	搜索范围
F_1	$F_1(x) = \sum_{i=1}^n x_i^2$	30	$[-100, 100]$
F_2	$F_2(x) = \sum_{i=1}^n x_i + \prod_{i=1}^n x_i $	30	$[-10, 10]$
F_3	$F_3(x) = \sum_{i=1}^n \left(\sum_{j=1}^i x_j^2 \right)^2$	30	$[-100, 100]$
F_4	$F_4(x) = \sum_{i=1}^n i x_i^4 + \text{random}[0, 1]$	30	$[-1.28, 1.28]$
F_5	$F_5(x) = \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i) + 10]$	30	$[-5.12, 5.12]$
F_6	$F_6(x) = -20 \exp \left(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \right) - \exp \left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i) \right)$	30	$[-32, 32]$

不同策略组合的算法是为验证改进策略对算法的影响: ①DBO-1, 在传统 DBO 算法基础上增加混沌映射与反向学习的种群初始化策略; ②DBO-2, 在传统 DBO 算法基础上增加融合 WOA 螺旋搜索策略; ③DBO-3, 在传统 DBO 算法基础上增加混沌映射与反向学习的种群初始化策略和融合 WOA 螺旋搜索策略。WOA 参数 b_h 、 p 分别设置为 1、0.5; DBO 算法和 IDBO 算法中, 根据参考文献 [18] 将参数 k 、 w 、 b 、 λ 分别设置为 0.1、0.9、0.3、0.5; AIDBO 算法中, 根据参考文献 [16, 18], 将参数 k 、 w 、 b 、 λ 、 γ 、 c 分别设置为 0.1、0.9、0.3、0.5、0.5、1。从图 3 的结果可看出, 在 6 个经典测试函数实验中, 相较于其他算法, AIDBO 算法在性能上表现出显著优势。在单峰和多峰函数中, AIDBO 算法的平均值更优, 具有更好的鲁棒性, 能够快速且稳定地达到最优解, 而其他算法的收敛速度较为缓慢, 且通常难以接近最优值。

3.1.2 实例场景

为验证本文路径规划模型与算法的有效性和合理性, 选择云南省丽江市 (平均海拔 2 400 m) 某森林区域为例, 其地形以山地、丘陵为主, 沟壑纵横, 地形起伏显著, 气候昼夜温差大、季风明显, 植被以云南松为主要建群种, 伴有栎类等阔叶树, 是典型高原森林山区。本文通过地理空间数据云公开平台获取 DEM 数据; 利用 ArcGIS 截取某 4 000 m×4 000 m 区域 (地图分辨率为 30 m, 高度 1 590~2 444 m) 用于仿真验证, 将该区域作为无人机执行任务的路径规划环境, 使用 MATLAB 对该环境进行模拟展示。本文采用某型号多旋翼灭火无人机执行任务, 无人机性能参数均来自文献 [22], 如表 2 所示。设置 5 架无人机, 起始点坐标、目标点坐标、大气威胁区参数及风向风速等环境参数设置如

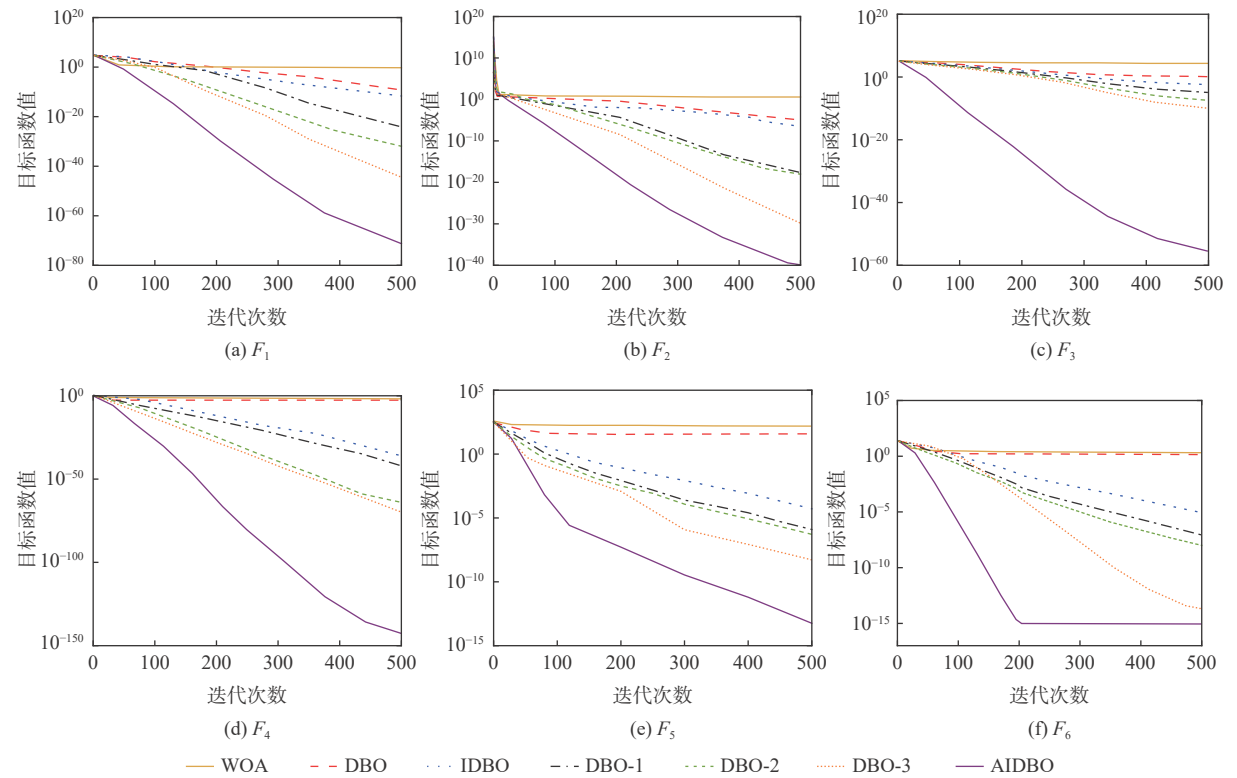


图 3 基准测试收敛曲线

Fig. 3 Benchmark convergence curves

表 2 无人机性能参数设置^[22]

Table 2 UAV performance parameter configuration^[22]

参数	数值
转子半径 R_t/m	0.5
转子盘面积 A/m^2	0.785
剖面阻力系数 δ	0.008
叶片角速度 $\Omega/(\text{rad}\cdot\text{s}^{-1})$	180
机身阻力比 d_0	0.6
转子固体度 s	0.05
感应功率增量修正系数 κ	0.1
无人机总质量 W_q/kg	60
无人机空速 $v_a/(\text{m}\cdot\text{s}^{-1})$	15
最大飞行高度 $H_{\max}/\text{m}$	300
安全距离 h_{safe}/m	80
无人机之间的安全距离 $d_{pq,\min}/\text{m}$	10

表 3 环境参数设置

Table 3 Environment parameter configuration

参数	设置
无人机1起始点坐标/m	(600,3 620,2 000)
无人机2起始点坐标/m	(600,3 600,2 000)
无人机3起始点坐标/m	(600,3 580,2 000)
无人机4起始点坐标/m	(580,3 600,2 000)
无人机5起始点坐标/m	(620,3 600,2 000)
目标点坐标/m	(3 000,300,2 300)
大气威胁区1坐标 i/m	(1 300,1 600)
大气威胁区1半径 r_1/m	200
大气威胁区2坐标 i/m	(2 000,1 500)
大气威胁区2半径 r_2/m	200
风速 $(\text{m}\cdot\text{s}^{-1})$	8
风向	西北(以y轴正向为北,x轴正向为东)
温度/ $^{\circ}\text{C}$	15
大气威胁区裕度扩展宽度 D_{mur}/m	50

表 3 所示。不同区域风险值设置如图 4 所示。

3.1.3 算法对比验证分析

为验证模型和算法在求解实例场景路径规划的有效性,采用不同算法进行路径规划求解,并对比分析算法性能。

首先,为验证 AIDBO 算法中改进策略关键参数的可信性,避免经验依赖性,分别对涉及的混沌映射参数 γ 和螺旋搜索常数 c 等关键参数进行敏感性分析,其他参数依据文献数据保持默认值,计算其最优适应度、平均适应度和标准差,结果如表 4 和表 5

所示。

根据表 4 的结果,当 $\gamma=0.5$ 时,算法获得最优适应度值和最低平均适应度值,表明此参数设置下的初始种群质量最高。 γ 值控制 Tent 混沌映射的序列分布特性, γ 值偏离 0.5 时,序列分布偏向 0 或 1,导致初始种群在解空间一侧过度集中,降低了搜索效率。根据表 5 的结果, $c=1.0$ 时,螺旋搜索能够在探索新区域和开发已知区域之间取得最佳平衡,特

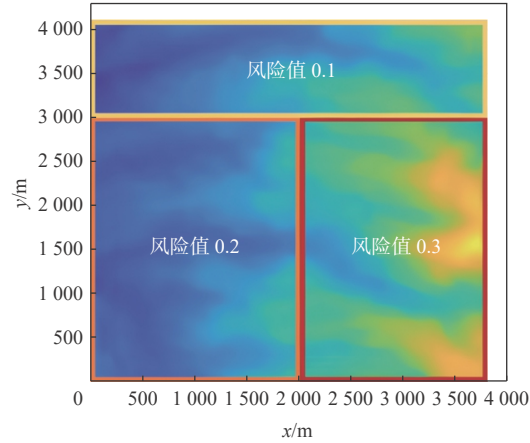


图 4 不同区域风险值设置

Fig. 4 Setting of risk values for different areas

表 4 混沌映射参数 γ 对算法性能的影响

Table 4 Influence of chaotic mapping parameter γ on algorithm performance

γ	最优适应度值	平均适应度值	标准差
0.3	3.305	3.425	0.235
0.5	2.984	3.224	0.251
0.7	3.412	3.589	0.312

表 5 螺旋搜索参数 c 对算法性能的影响

Table 5 Influence of spiral search parameters c on algorithm performance

c	最优适应度值	平均适应度值	标准差
0.5	3.284	3.476	0.267
1.0	2.984	3.224	0.251
2.0	3.396	3.512	0.301

别适合高原复杂地形中多威胁区的路径规划问题。 c 值过小则探索不足, 易陷入局部最优; c 值过大则过度探索, 收敛速度慢。

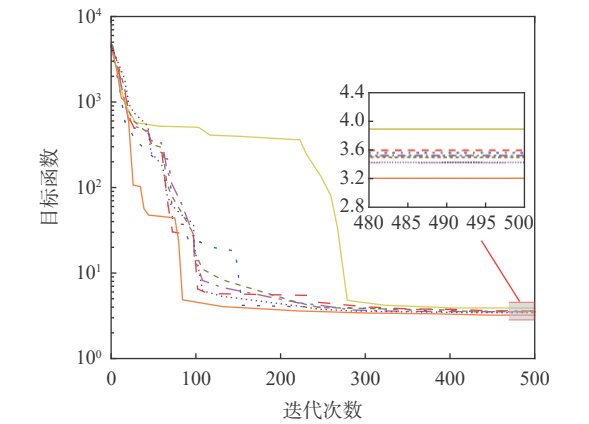
其次, 为验证不同算法求解路径规划问题的优劣, 对比了不同算法在实例场景路径规划模型求解性能。表 6 为 7 种算法在运行 10 次后的最优适应度、平均适应度及适应度标准差; 同时, 图 5 给出了各算法路径规划的收敛曲线。

从表 6 和图 5 对比结果可以看出, AIDBO 算法在最优适应度、平均适应度及收敛性能上均表现最好, 其中, 最优适应度值和平均适应度值分别为 2.984 和 3.224。相较于 WOA、DBO 和 IDBO 算法, AIDBO 算法的平均适应度分别优化了 24.30%、10.59% 和 9.82%, 最优适应度分别优化了 20.34%、11.51%、8.24%。与 WOA 和 DBO 算法相比, IDBO 算法和本文提出的 3 种改进策略算法在适应度上均有所提升, 特别是 DBO-3 算法, 这说明融合改进策略的蜣螂在跳出局部最优陷阱方面有一定优势。在运行时间上, 3 种改进策略算法和 AIDBO 算

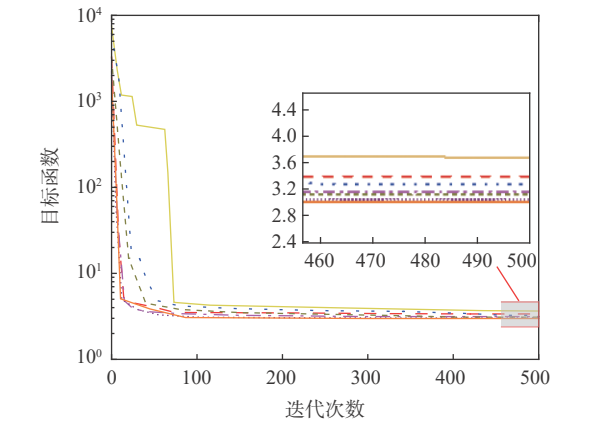
表 6 算法性能对比

Table 6 Algorithm performance comparison

算法	最优适应度值	平均适应度值	标准差	平均运行时间/s
WOA	3.746	4.259	0.415	28.7
DBO	3.372	3.606	0.330	25.2
IDBO	3.252	3.575	0.285	26.9
DBO-1	3.275	3.523	0.367	25.8
DBO-2	3.198	3.512	0.325	26.5
DBO-3	3.102	3.401	0.298	26.3
AIDBO	2.984	3.224	0.251	26.5



(a) 平均值迭代收敛曲线



(b) 最优值迭代收敛曲线

— WOA - - - DBO-2 - - - DBO DBO-3
..... IDBO — AIDBO - - - DBO-1

图 5 7 种算法路径规划收敛曲线

Fig. 5 Convergence curves of seven algorithms for path planning

法求解时间接近, 但均小于 WOA 算法, 这表明尽管 AIDBO 算法引入多种改进策略, 但未导致运行时间的显著增加。同时, AIDBO 算法的适应度标准差远低于其他算法, 表明其优化过程更加稳定, 能够在多次运行中保持较高的一致性和可靠性。因此, AIDBO 算法综合性能均优于其他算法, 表现出较强的全局搜索能力和收敛效率, 不仅提升了跳出局部最优的能力, 还改善了收敛精度和寻优稳定

性,验证了其在解决复杂优化问题中的应用潜力。

图 6 给出了 WOA、DBO、IDBO、AIDBO 这 4 种算法求解的无人机三维路径轨迹结果,表 7 给出了 4 种算法求解的路径规划子目标函数中每架无人机的具体飞行时间、风险代价和能耗结果。从图 6 中看出,AIDBO 算法在无人机路径规划中展现出最优的综合性能,其在 5 架无人机(UAV1~

UAV5)的飞行时间上均最短,总的风险代价值(0.486)和总的能耗代价值(1.044×10^5 kJ)整体低于其他对比算法。结果表明,AIDBO 算法在全球搜索能力和局部优化能力之间实现了良好的平衡,能够有效避免陷入局部最优解。相比其他 3 种算法,AIDBO 算法表现出更高的稳定性和优越性,能够在复杂地形条件下规划出更优的路径。

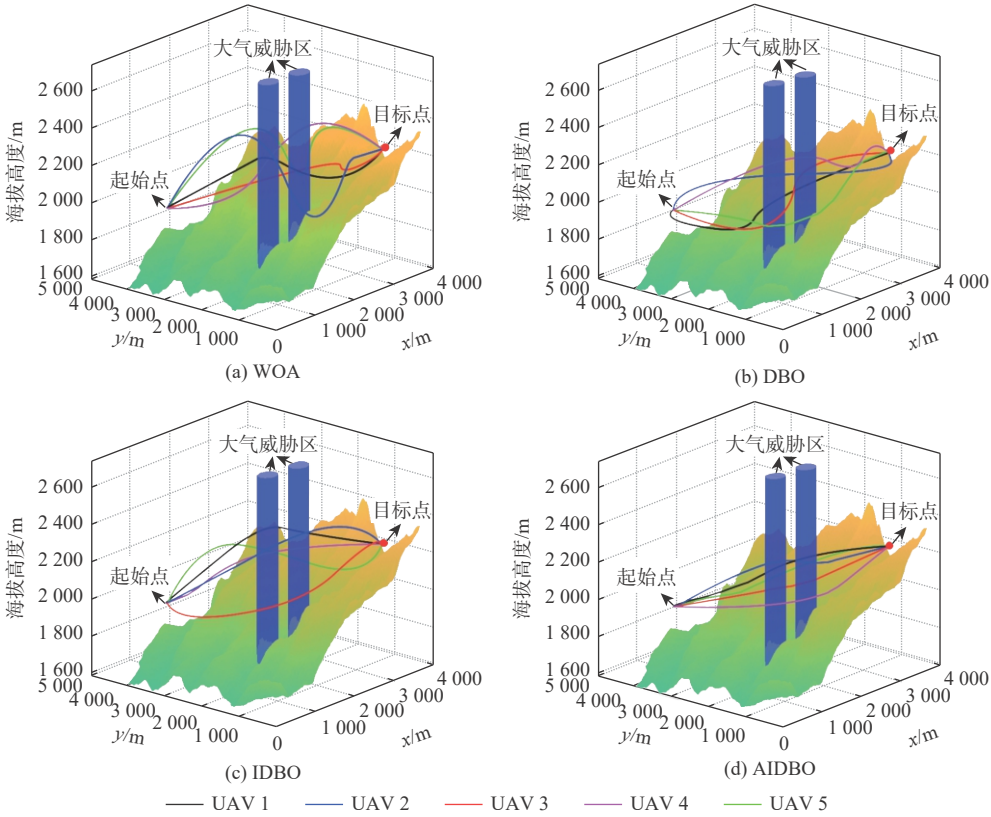


图 6 不同算法求解路径规划结果

Fig. 6 Results of path planning solved by different algorithms

表 7 不同算法路径规划结果对比				
Table 7 Comparison of path planning results of different algorithms				
算法	UAV	飞行时间/s	风险代价	能耗代价/ 10^4 kJ
WOA	UAV1	309.996	0.810	2.29
	UAV2	442.991	0.130	2.67
	UAV3	333.989	0.120	2.54
	UAV4	369.994	0.130	2.31
	UAV5	366.980	0.120	2.32
DBO	UAV1	346.398	0.125	2.40
	UAV2	320.627	0.116	2.39
	UAV3	306.521	0.108	2.42
	UAV4	314.929	0.113	2.51
	UAV5	334.831	0.122	2.39
IDBO	UAV1	306.998	0.102	2.13
	UAV2	305.944	0.103	2.19
	UAV3	312.998	0.099	2.23
	UAV4	333.909	0.107	2.22
	UAV5	317.997	0.108	2.11
AIDBO	UAV1	280.997	0.094	2.08
	UAV2	276.812	0.096	2.10
	UAV3	273.999	0.095	2.04
	UAV4	302.994	0.103	2.13
	UAV5	296.976	0.098	2.09

3.2 结果分析讨论

为探讨不同环境因素对无人机路径选择及能耗的影响及进一步验证本文方法的有效性,通过改变海拔、风速风向等参数,探究各因素影响。

3.2.1 海拔影响

为对比分析不同海拔对无人机飞行时间和能耗的影响,考虑平均海拔 340 m 和 2 400 m 代表平原和高原山区,同时固定西北风 8 m/s 环境风参数,其余风险参数不变的情况下,在较高海拔飞行会大幅增加能耗,而飞行时间在不同海拔条件下差异不显著,不同海拔下无人机路径规划结果如表 8 所示,而无人机三维路径轨迹如图 7 所示。

根据对比结果可以发现,海拔对无人机飞行时间影响较小,但对能耗的影响较为显著。根据表 8 所示,在海拔 340 m 和 2 400 m 环境下,无人机最长飞行时间分别为 301.982 s 和 304.993 s,最大飞行能耗代价分别为 1.34×10^4 kJ 和 2.13×10^4 kJ。因此,相

表 8 不同海拔路径规划结果对比

Table 8 Comparison of path planning results at different altitudes				
平均海拔/m	UAV	飞行时间/s	风险代价	能耗代价/ 10^4 kJ
340	UAV1	281.984	0.097	1.31
	UAV2	301.982	0.103	1.29
	UAV3	283.999	0.096	1.29
	UAV4	279.998	0.095	1.34
	UAV5	299.997	0.098	1.29
2 400	UAV1	280.997	0.094	2.08
	UAV2	276.812	0.096	2.10
	UAV3	273.999	0.095	2.04
	UAV4	304.993	0.103	2.13
	UAV5	296.976	0.098	2.09

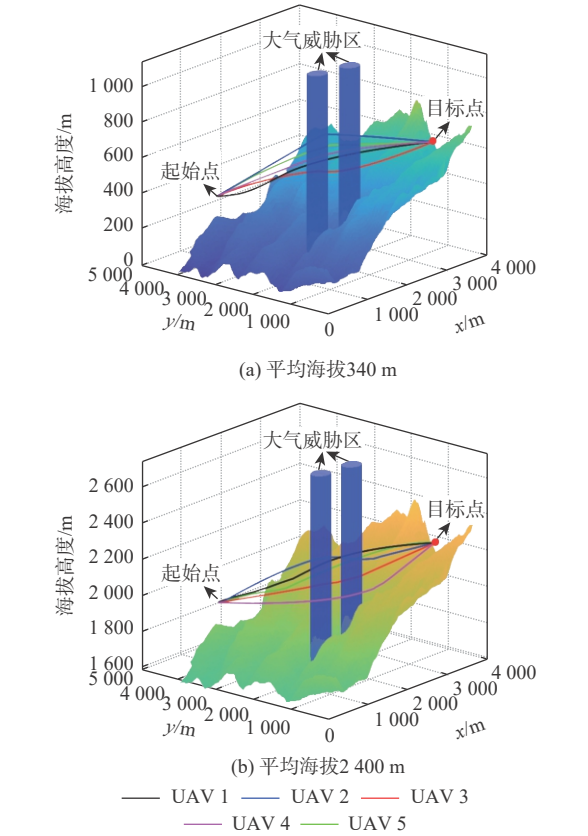


图 7 不同海拔路径规划结果

Fig. 7 Results of path planning by different altitudes

对于平原环境而言,在高海拔环境下,无人机最长飞行时间略微增加了 0.997%,能耗增加了 58.96%。这一结果表明,在高海拔稀薄空气条件下,动力系统需要更高的功率输出以维持飞行。因此,在高原环境下进行灭火救援时,需要重视无人机的能耗保障,对其能源管理进行优化。

3.2.2 风速风向影响

针对风影响下的无人机路径规划问题,考虑表 9 中参数设置,探究不同风速风向对无人机能耗的

表 9 风速风向影响参数设置

Table 9 Parameter settings for investigating impact of wind speed and direction			
工况	平均海拔/m	风速/($\text{m}\cdot\text{s}^{-1}$)	风向
1	2 400	8	西北风(顺风)
2	2 400	8	东南风(逆风)
3	2 400	10	西北风(顺风)
4	2 400	10	东南风(逆风)

影响。

表 10 为不同风速和风向下无人机路径规划的各子目标函数中每架无人机的具体飞行时间、风险代价和能耗结果,图 8 给出了不同风速和风向下无人机三维路径轨迹结果。通过结果对比分析可知,随着工况条件的变化(工况 1~工况 4),无人机集群的飞行风险代价显著上升,尤其在工况 3 和工况 4 中,部分无人机风险代价增至 0.237 以上,最高达 0.91;同时,5 架无人机的总能耗代价也从工况 1 的 1.044×10^5 kJ 逐步上升至工况 4 的 1.265×10^5 kJ。结果表明,风速与风向的恶化会明显增加路径规划的风险与能耗,其中,工况 1 在安全性与能耗方面均表现最优。由图 8 可以看出,风速和风向对无人机任务执行性能的影响具有一定规律性和显著性,顺风条件下飞行时间短于逆风,但两者航线差异较

表 10 不同风速风向路径规划结果对比

Table 10 Comparison of path planning results at different wind speed and direction paths					
工况	UAV	飞行时间/s	风险代价	能耗代价/ 10^4 kJ	总能耗代价/ 10^5 kJ
1	UAV1	280.997	0.094	2.08	1.044
	UAV2	276.812	0.096	2.10	
	UAV3	273.999	0.095	2.04	
	UAV4	302.994	0.103	2.13	
	UAV5	296.976	0.098	2.09	
2	UAV1	284.953	0.094	2.25	1.122
	UAV2	286.330	0.099	2.25	
	UAV3	278.997	0.097	2.20	
	UAV4	304.996	0.103	2.29	
	UAV5	291.997	0.096	2.23	
3	UAV1	294.979	0.104	2.38	1.181
	UAV2	279.958	0.103	2.21	
	UAV3	281.996	0.237	2.49	
	UAV4	305.943	0.446	2.31	
	UAV5	285.998	0.756	2.42	
4	UAV1	289.959	0.334	2.52	1.265
	UAV2	304.989	0.719	2.46	
	UAV3	284.990	0.099	2.48	
	UAV4	322.988	0.839	2.75	
	UAV5	287.916	0.910	2.44	

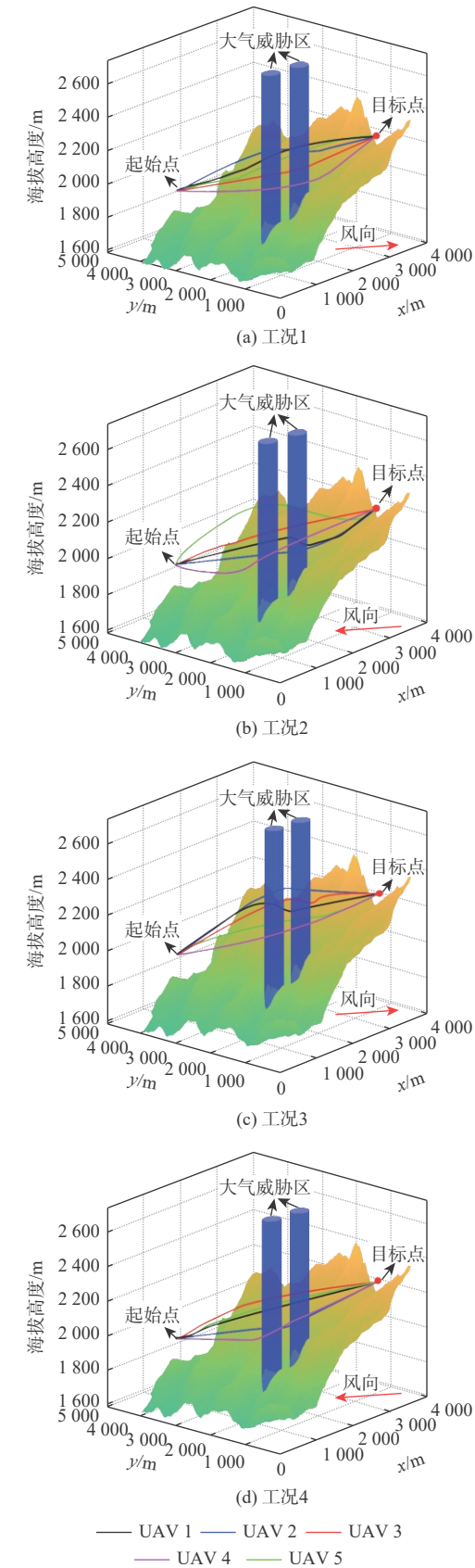


图 8 不同风速风向工况下路径规划结果
Fig. 8 Results of path planning by different wind speed and direction conditions

机总能耗代价增加 7.11%。能耗的提升主要归因于逆风条件下无人机需要额外的功率来克服气流阻力,从而导致更高的能耗。

在相同逆风方向下,不同风速对无人机飞行时间和能耗的影响尤为显著。根据表 10 所示,风速 8 m/s 时无人机最大飞行时间为 304.996 s,飞行总能耗代价为 1.122×10^5 kJ,而风速 10 m/s 时分别为 322.988 s 和 1.265×10^5 kJ。因此,相对于 8 m/s 风速,10 m/s 风速下,最大飞行时间和飞行总能耗代价分别增加了 5.90% 和 12.75%。这种差异表明,高风速环境会加剧无人机任务执行的复杂性,不仅影响任务完成时间,还显著提升了能耗成本,尤其是在长距离任务中,这种累积效应更为明显。此外,风速和风向不仅影响无人机的能耗和时间,还间接提高了风险成本。逆风条件下,飞行时间延长,使无人机在高风险区域的暴露时间增加,从而加大任务风险。

4 结 论

本文针对高原森林复杂环境下无人机路径规划问题,考虑海拔、环境风和威胁区等因素,构建了以最小飞行时间、风险代价及能耗为目标函数的多无人机路径规划模型,并提出 AIDBO 算法进行求解。主要结论如下:

1) 实例算法验证中,相较于 WOA、DBO 和 IDBO 算法,AIDBO 算法的平均适应度分别优化了 24.30%、10.59% 和 9.82%,最优适应度分别优化了 20.34%、11.51%、8.24%;证明 AIDBO 具有更好的鲁棒性,能够快速且稳定地达到最优解。

2) 海拔主要对无人机能耗影响显著,与平均海拔 340 m 环境相比,平均海拔 2 400 m 时无人机最长飞行时间增加了 0.997%,能耗增加了 58.96%。

3) 风速和风向对无人机飞行时间和能耗成本均产生影响。在 8 m/s 风速下,逆风比顺风能耗增加 7.47%;而 10 m/s 风速条件下,最大飞行时间相较于 8 m/s 风速增加了 5.90%,能耗增加幅度达到 12.75%。

由于高原森林环境复杂、影响因素较多,使得本文研究仍存在一些不足。未来研究可进一步考虑动态风场环境影响且无人机动力学性能受环境影响动态变化,采用实时动态调整飞行策略提供鲁棒性。同时,针对无人机失效和碰撞风险,设计更加安全的路径调整策略。

参考文献 (References)

[1] 潘卫军,黄园晶,骆晨,等.森林火灾多机型救援三维航迹规划[J].

小。在 8 m/s 风速下,逆风比顺风无人机总能耗代价增加 7.47%,而在 10 m/s 风速下,逆风比顺风无人

科学技术与工程, 2023, 23(23): 10152-10159.

Pan W J, Huang Y J, Luo C, et al. 3D route planning for forest fire rescue by multi-aircraft[J]. Science Technology and Engineering, 2023, 23(23): 10152-10159(in Chinese).

[2] 郑玉斌, 李新凯, 孟月, 等. 基于 GRU-MAPPO 的多无人机协同追捕方法[J/OL]. 北京航空航天大学学报, 2025(2025-07-02)[2025-11-19]. <https://link.cnki.net/urlid/11.2625.V.20250702.1040.001>.

Zheng Y B, Li X K, Meng Y, et al. Multi-UAV cooperative pursuit method based on GRU-MAPPO[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2025(2025-07-02) [2025-11-19]. <https://link.cnki.net/urlid/11.2625.V.20250702.1040.001>(in Chinese).

[3] 滕菲, 王迎春, 姚永辉, 等. 基于深度强化学习的无人机动态避障规划[J/OL]. 北京航空航天大学学报, 2025(2025-05-23)[2025-11-19]. <https://link.cnki.net/urlid/11.2625.V.20250523.1437.002>.

Teng F, Wang Y C, Yao Y H, et al. UAV dynamic obstacle avoidance planning based on deep reinforcement learning[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2025(2025-05-23)[2025-11-19]. <https://link.cnki.net/urlid/11.2625.V.20250523.1437.002>(in Chinses).

[4] 朱元军, 李妍, 张学军, 等. 基于风险约束的城市空域无人机安全路径规划[J/OL]. 北京航空航天大学学报, 2025(2025-03-27) [2025-11-19]. <https://link.cnki.net/urlid/11.2625.V.20250327.1510.001>.

Zhu Y J, Li Y, Zhang X J, et al. Risk-constrained safe path planning for unmanned aerial vehicles in urban airspace[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2025(2025-03-27)[2025-11-19]. <https://link.cnki.net/urlid/11.2625.V.20250327.1510.001>(in Chinese).

[5] 朱弥雪. 基于无人机的林火检测研究[D]. 呼和浩特: 内蒙古工业大学, 2023.

Zhu M X. Research on UAV-based forest fire detection[D]. Hohhot: Inner Mongolia University of Tehchnology, 2023(in Chinese).

[6] 余顺金, 冯建辉, 秦建明, 等. 基于改进人工势场的无人飞行器三维路径规划[J]. 云南电力技术, 2024, 52(3): 64-68.

Yu S J, Feng J H, Qin J M, et al. Three-dimensional path planning of UAV based on improved artificial potential field[J]. Yunnan Electric Power, 2024, 52(3): 64-68(in Chinese).

[7] 李锐君, 董素鸽. 基于粒子群算法的无人机灭火路径规划仿真[J]. 计算机仿真, 2023, 40(9): 43-48.

Li R J, Dong S G. Simulation of UAV fire extinguishing path planning based on particle swarm optimization algorithm[J]. Computer Simulation, 2023, 40(9): 43-48(in Chinese).

[8] Gao J F, Zheng Y, Ni K, et al. Fast path planning for firefighting UAV based on A-star algorithm[J]. Journal of Physics: Conference Series, 2021, 2029(1): 012103.

[9] Demir K, Tumen V, Kosunalp S, et al. A deep reinforcement learning algorithm for trajectory planning of swarm UAV fulfilling wild-fire reconnaissance[J]. Electronics, 2024, 13(13): 2568.

[10] Radmanesh M, Sharma B, Kumar M, et al. PDE solution to UAV/UGV trajectory planning problem by spatio-temporal estimation during wildfires[J]. Chinese Journal of Aeronautics, 2021, 34(5): 601-616.

[11] Aggarwal R, Soderlund A, Kumar M, et al. Risk-aware path planning for unmanned aerial systems in a spreading wildfire[J]. Journal of Guidance, Control, and Dynamics, 2022, 45(9): 1692-1708.

[12] Çoğay S, Seçinti G. Phoenix: aerial monitoring for fighting wildfires[J]. Drones, 2023, 7(1): 19.

[13] Hasan Z, Kumar S, Patel V, et al. UAV Path planning for wildfires-sustainably fighting wildfires with automated path planning for UAVs: NASA/TM—20220015156[R]. Washington, D. C. : NASA, 2022.

[14] Xu Y Q, Li J M, Zhang F Q. A UAV-based forest fire patrol path planning strategy[J]. Forests, 2022, 13(11): 1952.

[15] Xue J K, Shen B. Dung beetle optimizer: a new meta-heuristic algorithm for global optimization[J]. The Journal of Supercomputing, 2023, 79(7): 7305-7336.

[16] Lyu L X, Jiang H, Yang F. Improved dung beetle optimizer algorithm with multi-strategy for global optimization and UAV 3D path planning[J]. IEEE Access, 2024, 12: 69240-69257.

[17] Chen Q, Wang Y J, Sun Y F. An improved dung beetle optimizer for UAV 3D path planning[J]. The Journal of Supercomputing, 2024, 80(18): 26537-26567.

[18] Zhu F, Li G S, Tang H, et al. Dung beetle optimization algorithm based on quantum computing and multi-strategy fusion for solving engineering problems[J]. Expert Systems with Applications, 2024, 236: 121219.

[19] Shen Q W, Zhang D M, Xie M S, et al. Multi-strategy enhanced dung beetle optimizer and its application in three-dimensional UAV path planning[J]. Symmetry, 2023, 15(7): 1432.

[20] 刘微, 任腾腾, 韩广雨, 等. 多策略改进蜣螂优化算法及其应用[J]. 电子测量技术, 2024, 47(12): 109-121.

Liu W, Ren T T, Han G Y, et al. Multi strategy improvement of dung beetle optimization algorithm and its application[J]. Electronic Measurement Technology, 2024, 47(12): 109-121(in Chinese).

[21] 李晓萌. 高原狭窄区域下多无人机协同飞行控制[D]. 南京: 南京航空航天大学, 2019.

Li X M. Cooperative flight control of multi-UAVs in narrow plateau area[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2019(in Chinese).

[22] Zeng Y, Xu J, Zhang R. Energy minimization for wireless communication with rotary-wing UAV[J]. IEEE Transactions on Wireless Communications, 2019, 18(4): 2329-2345.

Multi-UAV path planning for forest fire-fighting in plateau based on adaptive improved dung beetle optimization algorithm

ZHU Pei*, LYU Xiaolong, CHEN Kaisen, SONG Rui, ZHANG Jiangao, SHAO Quan

(College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: Aiming at the problem of UAV fire-fighting and rescue path planning in the complex environment of plateau forest, a multi-UAV path planning method based on an adaptive improved dung beetle optimization (AIDBO) algorithm was proposed. Based on the digital elevation model (DEM) data, a three-dimensional geospatial model of the plateau forest and mountainous area was established. Considering the influence of complex terrain, environmental wind, altitude and other factors, a multi-UAV path planning model under multiple constraints was established, taking the minimum flight time, risk cost and energy consumption as the objective function. Chaotic mapping and reverse learning optimize the population initialization, the whale optimization algorithm (WOA) introduces the spiral search strategy to enhance the position update mechanism, and adaptive Cauchy mutation improves the population's capacity to eliminate local optima. An AIDBO algorithm is proposed to solve the problem. The results showed that the average fitness of AIDBO was 10.59% and 9.82% higher than that of the original dung beetle optimization (DBO) algorithm and the improved dung beetle optimization (IDBO) algorithm, respectively. Compared with a 340 m plain environment, the maximum flight time of the UAV in a 2 400 m altitude environment is slightly increased by 0.997%, while the energy consumption is increased by 58.96%. At the same time, the wind direction and speed will also have a certain impact on the flight time and energy consumption. In addition to providing decision assistance for UAV firefighting and rescue scheduling in the plateau forests, the adaptive improved dung beetle algorithm presented in this research may greatly optimize the multi-UAV path in the plateau forests.

Keywords: multi-UAV; path planning; plateau forest; fire-fighting and rescue; improved dung beetle optimization algorithm; adaptive

Received: 2025-12-08; **Accepted:** 2026-03-06; **Published Online:** 2026-04-02 09:01
URL: link.cnki.net/urlid/11.2625.V.20260401.0915.002
Foundation items: National Natural Science Foundation of China (52202416); Opening Fund of Key Laboratory of Civil Aviation Emergency Science & Technology (NJ2022022)
*** Corresponding author.** E-mail: zpei@nuaa.edu.cn

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0423

基于 YOLOv11-drone 的复杂背景下反无人机系统目标检测算法

薛珊*, 杨文达, 李其贵, 任万飞

(长春理工大学 机电工程学院, 长春 130022)

摘 要: 低空经济蓬勃发展的同时, 无人机对大型公共场所的安全威胁逐步升级。对无人机准确、快速的目标检测成为反无人机系统的关键和首要环节。针对复杂背景下小目标无人机检测时出现的检测精度低、漏检、错检等问题, 提出一种基于改进 YOLOv11 的反无人机系统目标检测算法 YOLOv11-drone。自主构建复杂背景下小目标无人机数据集; 设计部分空间通道协同注意力模块 C2SCSPA, 通过结合通道注意力及空间注意力提高模型对无人机特征的提取能力; 提出一种针对小目标的多层次特征金字塔网络 (STFPN), 通过浅层语义的充分运用, 极大减少了模型参数量, 并提高了模型对小目标的关注度; 将模型损失函数更换为 EIoU, 提高回归精度和收敛速度。在自建数据集上进行实验, 结果表明: 所提 YOLOv11-drone 算法的检测精度达到 95.5%, 相较原算法提高了 4.1%, 且参数量降低了 62%, 检测速度达到 83 帧/s, 适用于复杂背景下小目标无人机的检测任务。在公共数据集 VisDrone2019 上进行实验验证, 即使参数量只有原算法的 38%, 但改进算法的 mAP@0.5 仍提高了 0.6%, 验证了改进算法具有良好的适用性与泛化性。

关 键 词: 反无人机; 小目标检测; 多尺度特征; YOLOv11; 损失函数

中图分类号: TP391.4; V19

文献标志码: A

文章编号: 1001-5965(2026)09-3001-10

低空经济的蓬勃发展正在重塑全球产业格局, 无人机物流、城市空中交通、应急救援、农业植保等应用场景加速落地。根据中国民用航空局公布的数据, 2023 年, 全国民用无人机注册数量达 126.7 万架, 同比增长 32.2%。然而, 随着空域活动密度指数级增长, 低空安全威胁进一步升级, 对反无人机系统进行升级更新迫在眉睫。

反无人机系统是一种专门用于对抗无人机威胁的综合防御体系, 其工作过程分为 4 个关键阶段: 预防、探测、决策和反制^[1]。无人机的目标检测作为反无人机系统的重要环节, 准确地探测出非合作无人机是进一步实施反制措施的基础, 有着至关重要的作用^[2]。

目前, 无人机探测主要依赖以下 4 种技术: 雷达探测、无线电频谱探测、光电探测和声波探测。相

较于其他反无人机探测手段, 基于光电探测 (可见光与红外) 的无人机探测技术具有直观显示、普适性高、成本低等优点, 在反无人机系统中广泛应用。

然而, 基于光电的无人机探测技术在复杂背景下对小目标无人机进行精准探测仍面临多重挑战。无人机常飞行于城市、森林、建筑群等复杂环境中, 背景纹理多样、动态目标杂多, 导致目标与背景之间区分度低。传统检测模型易受高频纹理、鸟类等干扰, 严重影响目标探测的可靠性。而且无人机在远距离成像中所占像素极少, 往往不足数十像素, 成为了小目标。这类小目标缺乏显著结构和纹理特征, 导致主流检测模型 (如 YOLO 等) 在特征提取阶段难以保留有效信息, 尤其在深层网络中, 特征易被稀释, 造成漏检。同时, 无人机目标的遮挡

收稿日期: 2025-07-01; 录用日期: 2025-11-28; 网络出版时间: 2025-12-09 09:59

网络出版地址: link.cnki.net/urlid/11.2625.V.20251208.1659.002

基金项目: 吉林省自然科学基金面上项目 (20240101070JC)

* 通信作者. E-mail: 1660348815@qq.com

引用格式: 薛珊, 杨文达, 李其贵, 等. 基于 YOLOv11-drone 的复杂背景下反无人机系统目标检测算法 [J]. 北京航空航天大学学报, 2026, 52 (9): 3001-3010. Xue S, Yang W D, Li Q G, et al. Object detection method for anti-UAV systems in complex backgrounds based on YOLOv11-drone [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3001-3010 (in Chinese).

及拍摄角度的多样都影响到反无人机系统对无人机目标的准确探测。因此,针对复杂背景下的小目标无人机检测成为了亟待解决的问题。

复杂背景下的小目标检测任务作为深度学习领域的研究热点,一些学者提出了不同的改进。吕琮堂等^[3-5]使用 k-means 聚类算法重新设计更符合真实无人机目标的预测框,从而提高了算法的识别效率和定位准确度。白宇等^[6]对数据集中图像做了切片操作,使得小目标在数据集中的占比提高,从而使模型更好地学习到小目标特征,以提高小目标的检测精度。刘玉萍等^[7]设计了一种新的全局到局部双向特征金字塔网络(global-local bidirectional feature pyramid network, GLBiFPN),提升多尺度特征融合的效果。李彬等^[8]在 YOLOv11n 基础上进行了改进,通过多尺度膨胀卷积等操作,加强了对无人机小目标细节特征的提取。虽然上述方法一定程度上提高了算法性能,但复杂背景下的小目标无人机检测仍然存在错检、漏检等问题。

因此,针对该问题,本文以性能良好的 YOLOv11 单阶段目标检测算法作为基础算法进行改进,提高复杂背景下小目标无人机检测精度,解决错检、漏检等问题。本文主要贡献如下:

1) 设计了部分空间通道协同注意力模块 C2SC-SPSA,通过引入空间和通道协同注意力(spatial and channel synergistic attention, SCSA)机制,结合通道注意力及空间注意力提高模型对无人机特征的提取能力,同时提高模型对无人机目标的关注度。

2) 提出了一种针对小目标特征金字塔网络(small target FPN, STFPN),将 YOLOv11 原有的 FPN 进行修改,通过对输出特征图尺寸的调整及浅层特征的重复利用,提升模型对小目标的检测效果。

3) 引入 EIoU 损失函数,解耦预测框与真实框的宽高比损失项,使模型更快地朝着减小高度差值的方向更新参数,从而加快模型收敛速度、提高模型回归精度。

1 基于 YOLOv11 的反无人机目标检测算法

目前,基于深度学习的目标检测算法大体分为 2 类。双阶段目标检测算法通过预先生成候选框提高了检测精度,但由于计算量过大导致检测速度过低,其代表算法有 R-CNN^[9]、Faster R-CNN^[10]、Mask R-CNN^[11]等。单阶段目标检测算法直接进行分类与定位,以端到端的形式完成任务,速度较快,但精确度较低,其代表算法有 SSD^[12]及 YOLO 系列算法^[13-14]等。

在众多目标检测算法中, YOLO 具有版本多、相对准确度高、检测速度快、适应性强、网络结构简单等优点,备受关注^[15]。YOLO 系列算法经过历年的迭代更新,目前已发布了许多版本,其性能也在逐渐提升。作为 YOLO 系列最新推出的算法, YOLOv11 正逐步应用于各种视觉场合。本文选用 YOLOv11 算法作为反无人机系统目标检测部分,进行了反无人机系统的目标检测任务,其网络结构及各模块输出结果如图 1 所示。

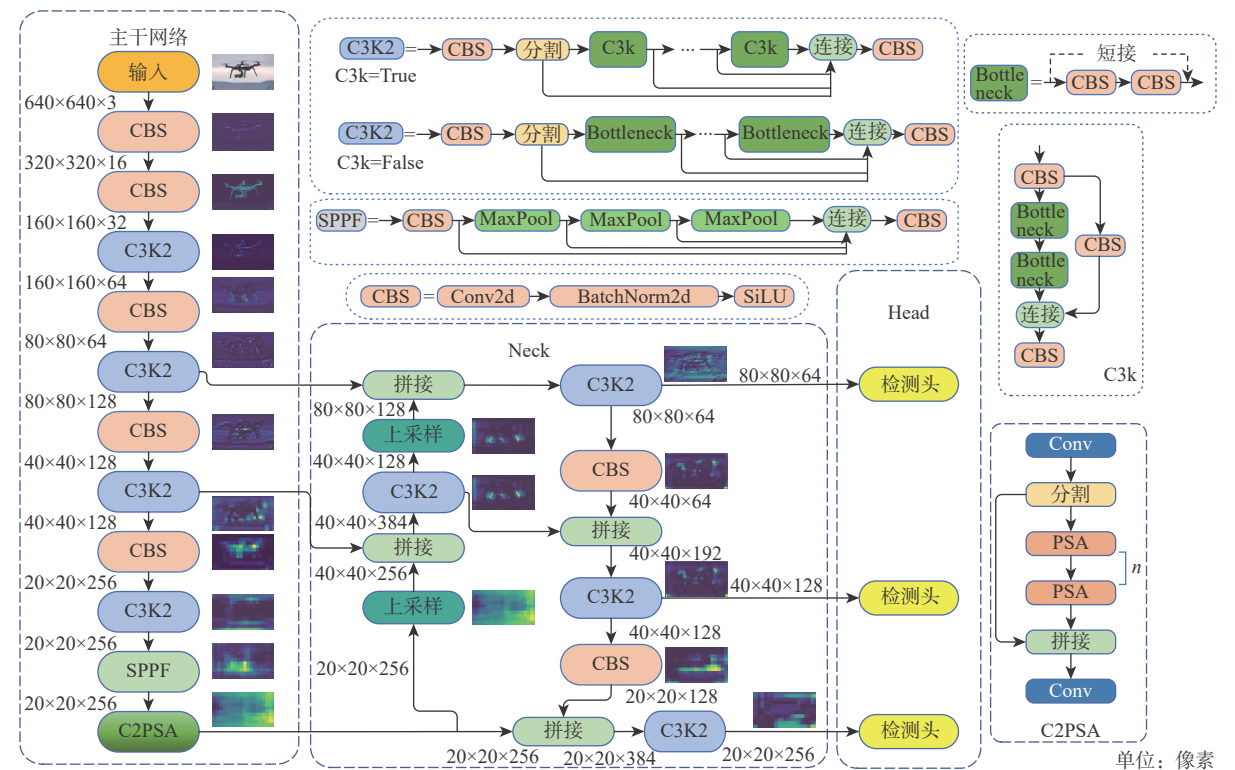


图 1 YOLOv11 算法结构

Fig. 1 Structure of YOLOv11 algorithm

YOLOv11 算法以 CSPDarknet53 为主干网络, 通过 5 次标准卷积模块 CBS 及 C3K2 模块进行下采样和特征提取操作生成不同尺度的特征图。下采样后, 通过引入空间金字塔池化快速 (spatial pyramid pooling-fast, SPPF) 模块, 增加了特征表达的多样性。通过 C2PSA 模块增强模型对输入图像的特征提取能力。YOLOv11 算法的 Neck 部分应用路径聚合网络 (path aggregation network, PANet), 使得低层和高层特征可以直接相互影响。最终, 生成的融合特征图传递给后续的检测头, 用于目标检测。YOLOv11 的检测头采用

解耦结构, 2 条分支分别预测类别信息和位置信息。

2 改进的 YOLOv11-drone 反无人机目标检测算法

复杂背景下, 小目标无人机的检测常常受到图像背景信息的干扰, 且因目标尺度较小导致目标检测精度偏低, 经常出现错检、漏检等现象。针对这一问题, 本文提出一种基于特征融合的小目标反无人机检测算法 YOLOv11-drone, 改进后的算法结构如图 2 所示。图中: S 为步长。

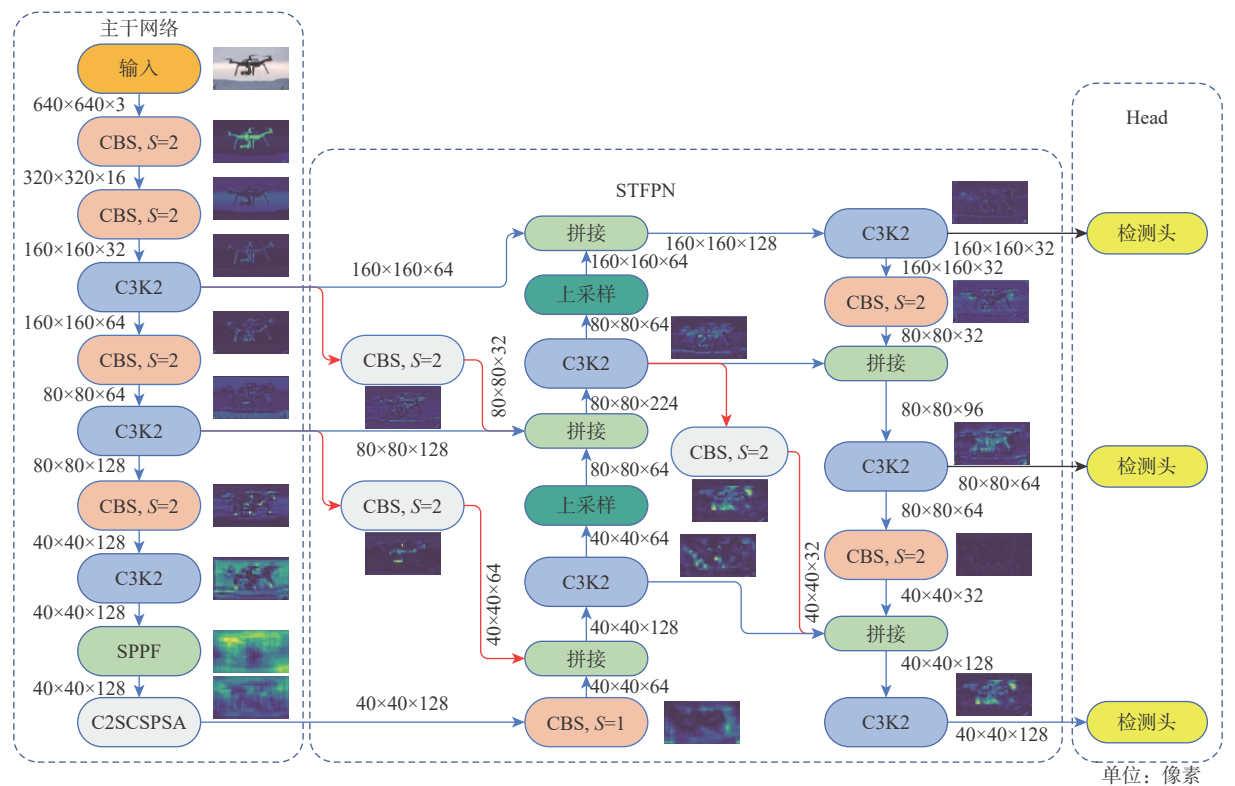


图 2 YOLOv11-drone 算法结构
Fig. 2 Structure of YOLOv11-drone algorithm

首先, 针对图像背景复杂问题设计了 C2SCSPA 模块, 通过协调注意力机制, 同时利用空间与通道信息, 提高模型对小尺度无人机目标的关注度; 其次, 针对无人机目标尺度较小的问题, 通过更换特征图输出尺度及对浅层特征再次融合的方式, 提高小目标的检测效果; 最后, 为进一步提高检测精度与收敛速度, 引入 EIou 作为新的损失函数。

2.1 C2SCSPA 模块设计

C2PSA 模块是 YOLOv11 中用于增强特征提取的高级模块, 结合了跨阶段局部 (cross stage partial, CSP) 结构和部分自注意力 (partial self-attention, PSA) 注意力机制, 其结构可见图 1。虽然该模块提升了特征提取能力, 但并未对空间信息及通道信息做过

多关注, 因此, 本文引入 SCSPA 模块^[16]对 C2PSA 模块进行改进设计。SCSPA 模块结构如图 3 所示。图中: B 为批次大小, C 为通道数, H 、 W 分别为特征图高度、宽度, H' 、 W' 分别为改变后的特征图高度、宽度, $N'=H\times W$ 。SCSPA 模块主要由共享多语义空间注意力 (shared multi-semantic spatial attention, SMSA) 模块及渐进通道自注意力 (progressive channel-wise self-attention, PCSA) 组成。

SMSA 模块通过多尺度深度共享卷积 MS-DWconv1d 先对 4 个子特征进行核尺寸为 3、5、7、9 的深度可分离一维卷积, 再利用共享卷积进行对齐, 捕捉每个子特征内部独特的语义空间结构。不同大小的卷积核允许其在不同感受野下提取空间

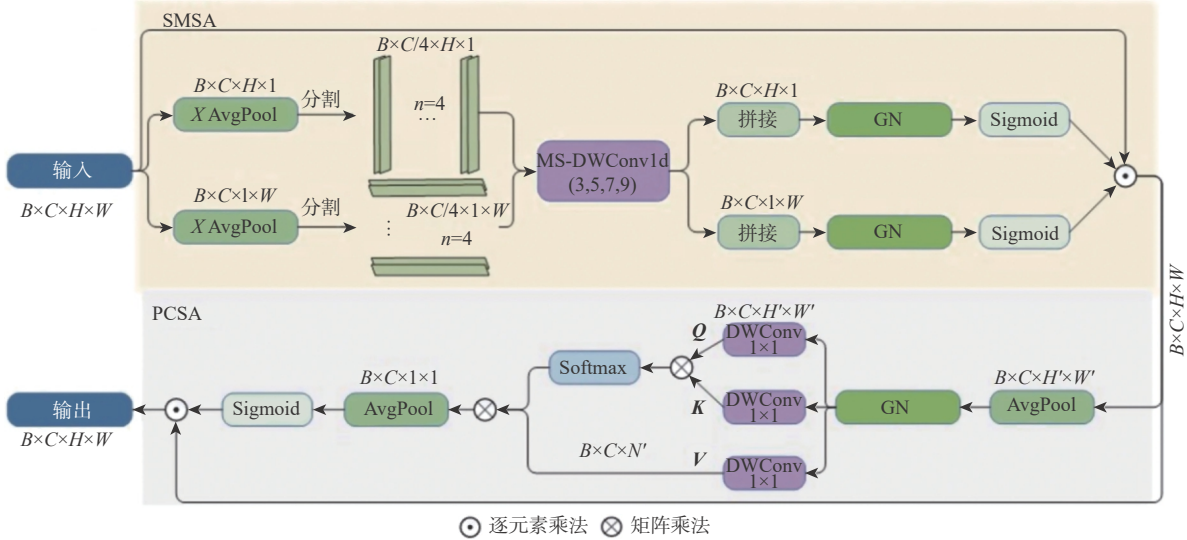


图3 SCSPA 模块结构

Fig. 3 Structure of SCSPA module

结构信息,使其能够同时捕捉局部、全局语义信息及充分的上下文关系,这对于小目标的检测有很大帮助。随后,将不同语义的子特征进行拼接,并采用分组归一化 (group normalization, GN) 进行归一化操作,缓解子特征之间的语义干扰,最终利用 Sigmoid 激活函数生成空间注意力。

PCSA 先采用基于平均池化 AvgPool 的渐进式压缩方式,降低后续计算成本的同时保留并利用 SMSA 提取的多语义空间信息,随后采用输入感知的自注意力机制,结合 SMSA 提供的空间信息,对通道特征进行更精细的自注意力调整,增强了通道注意力在语义一致性和区分性上的表现。

SCSPA 模块运用 SMSA 与 PCSA 的串联使用,实现了空间和通道的协同效应,通过空间注意力引导通道学习,使得通道能更好地关注重要的空间区域,同时通道自注意力还可进一步增强空间结构中的细节表现,在目标检测任务中表现出色。

改进后的 SCSPSA 模块将 PSA 模块中的多头注意力机制更换为 SCSPA 机制,通过空间与通道协同注意力提升模型提取特征能力,提高对无人机目标的关注度。改进后的 SCSPSA 结构如图 4 所示。其数学公式描述如下:该模块输入为 $X \in \mathbf{R}^{B \times C \times H \times W}$,其中, B 为图片数, C 为通道数, H 为高度, W 为宽度,先进行初始投影及通道分割。

$$F = \text{Conv}_{1 \times 1}(X) \in \mathbf{R}^{B \times C \times H \times W} \quad (1)$$

$$F_1, F_2 = \text{Split}(F) \quad (2)$$

随后,分割后的点 F_1 进入 SCSPA 模块与前馈网络 (feed-forward network, FFN) 中进行处理。得到结果 F'_1 与 F_2 进行拼接操作后输出,其过程如下:

$$F'_1 = \text{BatchNorm}(F_1 + \text{SCSPA}(F_1)) \quad (3)$$

$$F'_1 = \text{BatchNorm}(F'_1 + \text{FFN}(F'_1)) \quad (4)$$

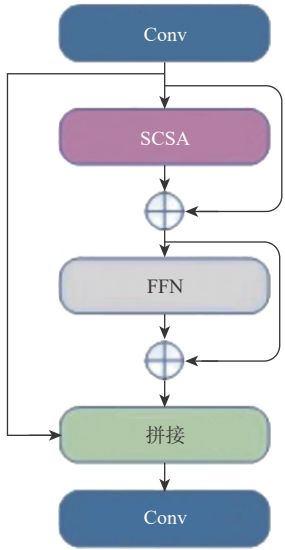


图4 SCSPSA 模块结构

Fig. 4 Structure of SCSPSA module

$$Y = \text{Conv}_{1 \times 1}(\text{Concat}(F'_1, F_2)) \quad (5)$$

式中: Split 表示通道分割操作; BatchNorm 表示批量归一化; $\text{Conv}_{1 \times 1}$ 表示卷积核大小为 1×1 的卷积操作; Concat 表示拼接操作。

最后,结合 CSP 结构及 SCSPSA 模块得到本文改进的 C2SCSPSA 注意力模块,其结构如图 5 所示。首先,进行 CSP 分割,将输入张量 $X \in \mathbf{R}^{B \times C \times H \times W}$ 按照维度分为 2 份,其中, $X_1, X_2 \in \mathbf{R}^{B \times \frac{C}{2} \times H \times W}$ 。主路经过 n 个 SCSPSA 模块串联处理 (本文采用 $n=1$) 后与另一路径进行拼接操作后输出,其表达式如下:

$$X_1, X_2 = \text{Split}(X) \quad (6)$$

$$Y = \text{Conv}_{1 \times 1}(\text{Concat}(\text{SCSPSA}(X_1), X_2)) \quad (7)$$

SCSPSA 模块运用 CSP 结构及 SCSPA 机制的结合,增强了模型对多尺度特征的提取能力,提高了

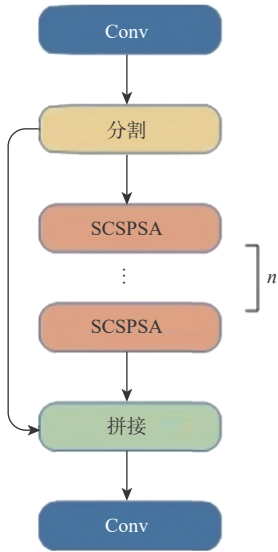


图 5 C2SCSPSA 模块结构
Fig. 5 Structure of C2SCSPSA module

复杂背景下模型对小目标无人机的关注度。

2.2 STFPN 设计

YOLOv11 采用路径聚合特征金字塔网络 (path aggregation feature pyramid network, PAFPN) 结构进行多尺度特征融合, 提高模型对不同尺度目标的检测精度。这种方法虽然很好地聚合了主干网络输出的不同尺度的特征信息, 但在面对小尺度目标时, 由于多层的下采样导致图像分辨率降低, 浅层特征语义往往无法充分传递到检测部分, 降低了模型对于浅层语义需求较高的微小目标的检测精度。因此, 本文设计了针对小目标检测的 STFPN, 保留更多的浅层空间特征, 提高模型对小目标的检测精度。改进前后网络的特征融合变化如图 6 所示。

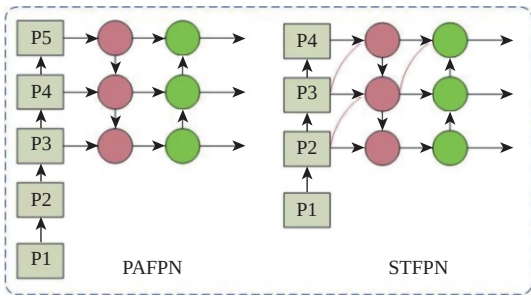


图 6 改进前后特征融合部分对比
Fig. 6 Comparison of feature fusion part before and after improvement

YOLOv11 的主干网络运用标准卷积和 C3K2 模块对输入的分辨率为 640 像素×640 像素的图像进行特征提取。过程中, 对输入图像进行 8 倍、16 倍和 32 倍下采样, 输出的 P3、P4、P5 被用于后续目标融合网络及检测头部分, 其尺寸分别为 80 像素×80 像素、40 像素×40 像素、20 像素×20 像素。当输入的目标尺寸小于 32 像素×32 像素时, 主干网络输出的 P5 特征图上将忽略该目标的信息。

过多的下采样操作不仅不会对小目标的检测效果有所提升, 反而会因为网络过深导致空间信息的部分丢失, 影响小目标的检测效果。

因此, 本文将特征融合部分整体上移, 删去主干网络中最后一组 32 倍的下采样操作 (CBS 和 C3K2), 使主干网络最多进行 16 倍的下采样, 得到 40 像素×40 像素的特征图, 降低网络参数数量的同时, 避免了因网络过深导致的小尺度目标信息丢失问题; 为充分运用对小目标检测更有帮助的浅层特征, 将主干网络中 4 倍下采样得到的 160 像素×160 像素特征图引入 FPN。同时, 分别对 P2、P3 进行额外的下采样操作, 并与原特征融合层中的特征图进行通道层次的融合操作, 使得模型能够精确捕捉到尺度较小目标的细节特征。

2.3 EIou 损失函数

YOLOv11 的预测框回归采用 CIoU 损失函数, 其计算公式如下:

$$L_{CIoU} = 1 - J_{IoU} + \frac{d^2(b, b^{gt})}{c^2} \alpha v \tag{8}$$

$$J_{IoU} = \frac{|b \cap b^{gt}|}{|b \cup b^{gt}|} \tag{9}$$

$$v = \frac{4}{\pi^2} \left(\arctan \left(\frac{w^{gt}}{h^{gt}} \right) - \arctan \left(\frac{w}{h} \right) \right)^2 \tag{10}$$

式中: d 为真实框与预测框中心点的欧氏距离; b 表示预测框; b^{gt} 表示真实框; c 为真实框与预测框最小外接矩形的对角线距离; α 为权重参数; v 为度量宽高比一致性; J_{IoU} 为真实框与预测框的面积交并比; w^{gt} 为真实框的宽度; h^{gt} 为真实框的高度; w 为预测框的宽度; h 为预测框的高度^[17]。CIoU 不仅考虑了交并比, 还考虑了真实框和预测框的中心距离及宽高比的差异。但是, CIoU 难以度量预测框与真实框宽高之间的真实差异, 且在预测框和真实框几乎完全重合 (IoU 接近 1) 时, 各项变化非常微小, 可能会导致梯度消失问题。

因此, 本文引进 EIou 损失函数^[18]进行预测框的回归损失计算。EIou 将 CIoU 中预测框与真实框的宽高比损失项拆分成预测框与真实框的宽度

差值和高度差值, 即用 $\frac{d^2(w, w^{gt})}{C_w^2} + \frac{d^2(h, h^{gt})}{C_h^2}$ 替换 CIoU 中的 αv 项, 加快模型收敛速度。计算公式如下:

$$L_{EIou} = 1 - J_{IoU} + \frac{d^2(b, b^{gt})}{c^2} + \frac{d^2(w, w^{gt})}{C_w^2} + \frac{d^2(h, h^{gt})}{C_h^2} \tag{11}$$

式中: C_w 为预测框与真实框最小外接矩形的宽度; C_h 为预测框与真实框最小外接矩形的高度。

如图 7 所示, 当预测框只需要调整高度时, 由于 EIou 分别明确了宽、高差值的损失, 预测框在回

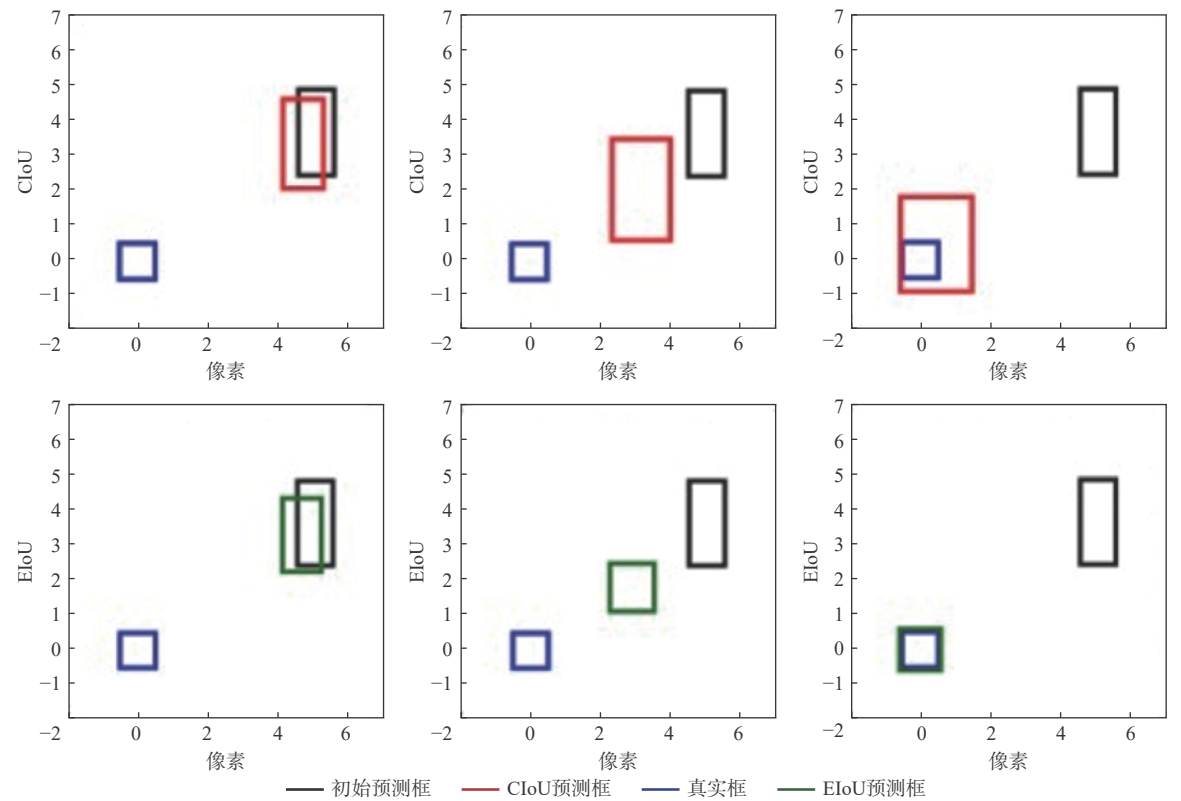


图 7 EIoU 损失函数示意图

Fig. 7 Schematic diagram of EIoU loss function

归过程中可以仅改变高度值,而 ClIoU 则会同时改变预测框的宽度导致回归效果受到影响。因此,EIoU 损失函数可以使模型更快地朝着减小高度差值的方向更新参数,而不会受到宽高比综合因素的干扰,从而加快模型收敛速度,提高模型回归精度。

3 实验参数与结果分析

3.1 实验数据集

本文针对复杂背景下小目标无人机检测出现的问题进行改进,在数据集选择上以自建数据集为主,突出复杂背景及小尺度无人机。其中,图像背景多以复杂的公共安全区域为主,如学校、游乐场、公园,且无人机目标均是远距离拍摄。同时,保证采集无人机图像数据时,无人机飞行背景中均有楼房、树木、游乐设施等复杂背景。无人机型号为大疆 Mavic2 和大疆御 Mavic Air 这 2 款机型。此外,为提高数据集的可靠性,引入部分大连理工大学提出的无人机数据集 DUT-ANTI-UAV^[19],其中包含不同背景、不同型号无人机图像,且多为小尺度无人机图像,与自建数据集合并后,本文数据集共包含 6 761 张图像,按照 7:3 的比例划分为训练集和验证集,其中大部分是复杂背景下小目标无人机图像。图 8 为部分无人机图像数据。

3.2 评价指标

本文对模型性能进行评估主要涉及 4 方面指



图 8 部分数据集示意图

Fig. 8 Schematic diagram of a partial dataset

标:①平均精度均值(mean average precision, mAP)体现检测精度;②参数量体现模型规模;③浮点运算次数衡量运算效率;④模型检测速度。mAP 作为核心指标,其计算过程综合了准确率 P 、召回率 R 及各类别的平均精度 A_p ^[20],计算公式如下:

$$P = \frac{T_p}{T_p + F_p} \times 100\% \tag{12}$$

$$R = \frac{T_p}{T_p + F_N} \times 100\% \tag{13}$$

$$A_p = \int_0^1 P(R) dR \tag{14}$$

$$m_{AP} = \frac{1}{l} \sum_{i=1}^l A_{p,i} \tag{15}$$

式中: T_p 表示模型正确检测出的目标数; F_p 表示模

型错误检测出的目标数; F_N 表示模型未能正确检测出的目标数; l 为数据集中所有样本的类别数^[21]。本文用到的 mAP@0.5 表示 IoU 设为 0.5 时得到的 mAP 值。此外, 为验证模型对小目标的检测效果, 引入 COCO 数据集中对小目标的平均精度评价指标 mAP@0.5:0.95(area=small), 该标准表示在 32 像素×32 像素以下的目标在不同阈值 (从 0.5 到 0.95, 步长为 0.05) 上的平均 mAP。

3.3 实验环境与参数设置

实验操作系统为 Windows 操作系统, CPU 为 13th Gen Intel(R) Core(TM) i5-13600KF; GPU 为显存 12 GB 的 NVIDIA GeForce RTX 4070; 编程语言为 Python3.9.19; 深度学习框架为 Pytorch2.4.0; CUDA 为 12.1; CUDNN 为 9.0.0。

深度学习模型中, 超参数的取值会影响模型的检测性能。为验证本文改进的泛化性, 模型超参数均设为经验值。训练轮数为 100 轮, 训练批次大小为 16, 图像输入尺寸为 640 像素×640 像素, 具体参数如表 1 所示。

3.4 实验结果与分析

3.4.1 消融实验

为评估 YOLOv11-drone 算法改进的有效性, 以 YOLOv11n 为基准算法, 在自主构建的复杂背景下小目标无人机数据集上开展消融实验。通过控制变量法分别验证各优化单元对 mAP@0.5、网络参数量、浮点运算次数及检测速度的影响。实验结果

表 1 实验超参数	
Table 1 Experimental hyperparameters	
超参数	数值
初始学习率	0.01
循环学习率	0.01
动量	0.937
批量大小	16
图片尺寸/(像素×像素)	640×640
训练轮次	100
预热学习轮数	3.0
预热训练动量	0.8
IoU训练阈值	0.7

如表 2 所示。可知, 各模块改进对检测精度均有提升, STFPN 的改进显著降低了模型参数量的同时, 对目标检测精度也有较大提升, 证明了本文改进的有效性。将注意力模块改进及特征融合部分改进后与原算法进行热力图可视化对比实验, 实验结果如图 9、图 10 所示。图中, 颜色越红, 表示模型对该区域的关注度越高。可知, 在复杂背景且目标较小的情况下, 改进后模型对目标的关注度相较于原模型均有提高。同时, 为直观体现本文模型的改进效果, 在复杂背景且目标占比较小情况下对图像进行可视化对比实验, 实验结果如图 11 所示。可以看出, 在复杂背景干扰且目标尺度较小情况下, 原模型会出现漏检、错检等问题, 改进后模型可减少此问题发生。

表 2 消融实验结果对比								
Table 2 Comparison of ablation experiment results								
实验序号	C2SCSPSA	STFPN	EIoU	参数量	浮点运算量/次	mAP@0.5/%	mAP@0.5:0.95-small/%	检测速度/(帧·s ⁻¹)
1	×	×	×	2.59×10 ⁶	6.4×10 ⁹	91.4	24.6	89
2	√	×	×	2.54×10 ⁶	6.4×10 ⁹	92.0	25.0	88
3	√	√	×	0.99×10 ⁶	9.3×10 ⁹	94.6	39.6	79
4	√	√	√	0.99×10 ⁶	9.3×10 ⁹	95.5	39.6	83

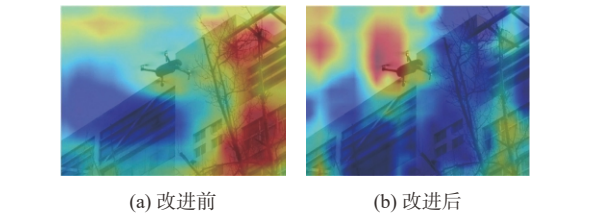


图 9 改进注意力机制前后小目标热力图对比

Fig. 9 Comparison of heatmaps for small targets before and after improving attention mechanism

3.4.2 不同检测算法对比实验

为系统验证 YOLOv11-drone 算法的优越性, 将其与前沿轻量化目标检测算法 (YOLOv7-tiny、YOLOv8n、YOLOv9-tiny、YOLOv10n 及 YOLOv11s) 进行对比分析, 实验结果如表 3 所示。可知, 本文

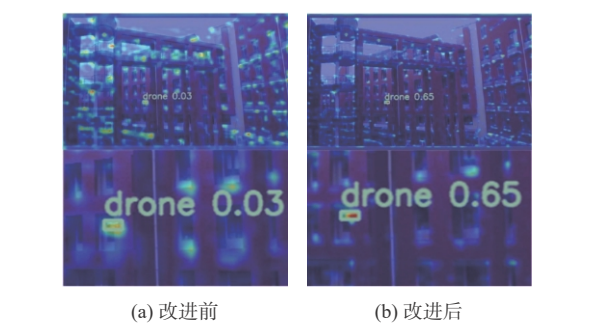


图 10 改进特征融合部分前后小目标热力图对比

Fig. 10 Comparison of heatmaps of small targets before and after improving feature fusion

改进的 YOLOv11-drone 算法相比于原算法参数量降低 62%, 相较于其他算法, 参数量均有大幅度降

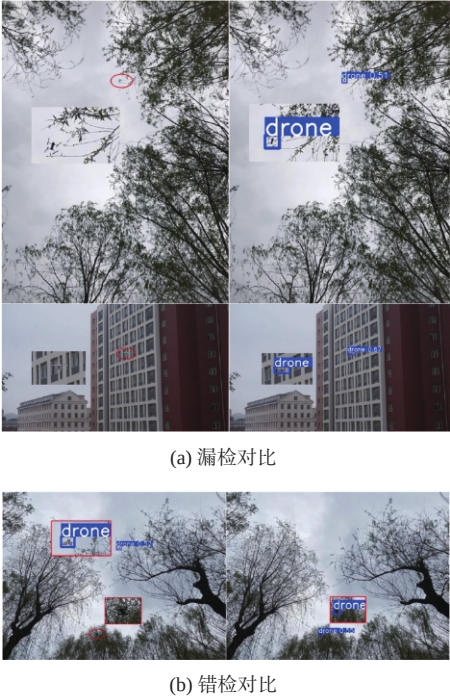


图 11 改进算法检测效果可视化对比

Fig. 11 Visual comparison of detection performance of improved algorithms

表 3 不同算法对比实验结果

Table 3 Comparison experiment results of different algorithms

算法	参数量	浮点运算量/次	mAP@0.5/%	检测速度/(帧·s ⁻¹)
YOLOv7-tiny	6.01×10 ⁶	13.2×10 ⁹	89.0	66
YOLOv8n	3.01×10 ⁶	8.2×10 ⁹	91.2	71
YOLOv9-tiny	2.65×10 ⁶	11.0×10 ⁹	93.0	60
YOLOv10n	2.70×10 ⁶	8.3×10 ⁹	91.7	65
YOLOv11n	2.59×10 ⁶	6.4×10 ⁹	91.4	89
YOLOv11s	9.43×10 ⁶	21.5×10 ⁹	94.7	65
YOLOv11-drone	0.99×10 ⁶	9.3×10 ⁹	95.5	83

低。本文改进算法在检测精度上也表现出色，相较于原算法提升了 4.1%。虽然增加了额外的计算量导致检测速度略有降低，但仍可满足后续反无人机系统跟踪任务的实时性要求。

3.4.3 泛化能力实验

为验证本文设计算法的泛化能力，应用本文提出的 YOLOv11-drone 算法在 VisDrone2019 无人机航拍数据集上进行实验验证。改进前后算法的 mAP@0.5 对比如表 4 所示。实验结果表明，改进算

表 4 VisDrone2019 数据集对比结果

Table 4 Comparison results on VisDrone2019 dataset

算法	mAP@0.5/%	参数量	检测速度/(帧·s ⁻¹)
YOLOv8n	30.1	3.01×10 ⁶	168
YOLOv10n	29.0	2.70×10 ⁶	150
YOLOv11n	30.2	2.59×10 ⁶	146
YOLOv11-drone	30.8	0.99×10 ⁶	112

法在参数量降低一半以上的情况下，检测的平均精度仍较基线算法提高了 0.6%，证明了本文改进算法具有一定的泛化能力。

4 结 论

反无人机系统工作的关键及首要环节是对无人机进行目标检测。本文提出了一种基于 YOLOv11-drone 的反无人机目标检测算法，用以解决传统模型在复杂背景下小目标无人机的检测任务中出现的检测精度低、错检和漏检等问题，为后续反无人机的跟踪与监管任务提供有力支持。

1) 在主干网络中利用 SCSA 机制与 C2PSA 机制的融合，提高模型对复杂背景下无人机目标的关注度，提高模型检测精度。

2) 设计了针对小目标的 STFPN，增强了对浅层特征信息的利用，降低模型参数量 62%，同时大幅提高模型对小目标的检测精度。

3) 采用 EIoU 作为模型损失函数，进一步提高其收敛速度和精度。

以 mAP@0.5、参数量、浮点运算次数、检测速度等指标，对本文改进算法与原算法进行对比实验。实验结果表明，改进算法的 mAP@0.5 提升至 95.5%，证明了本文改进算法在复杂背景下小目标无人机检测中的可行性与先进性，能够为反无人机系统的后续任务提供支持。同时，在公共数据集 VisDrone2019 上进行泛化实验，实验结果证明了本文改进在小目标检测任务中的适用性。

参考文献（References）

[1] 邱小剑, 骆博雅, 付珍, 等. 国内外反无人机技术发展综述[J]. 战术导弹技术, 2024(5): 63-73.
Qiu X J, Luo B Y, Fu Z, et al. An overview on development of domestic and foreign anti-UAV technology[J]. Tactical Missile Technology, 2024(5): 63-73(in Chinese).
[2] 谢学立, 席建祥, 卢瑞涛, 等. 动态区域聚焦的反无人机红外长时跟踪算法[J]. 北京航空航天大学学报, 2025, 51(9): 3039-3051.
Xie X L, Xi J X, Lu R T, et al. Long-term infrared object tracking algorithm based on dynamic region focusing for anti-UAV[J]. Journal of Beijing University of Aeronautics and Astronautics, 2025, 51(9): 3039-3051(in Chinese).
[3] 薛珊, 张亚亮, 吕琼莹, 等. 复杂背景下的反无人机系统目标检测算法[J]. 吉林大学学报(工学版), 2023, 53(3): 891-901.
Xue S, Zhang Y L, Lyu Q Y, et al. Anti-unmanned aerial vehicle system object detection algorithm under complex background[J]. Journal of Jilin University (Engineering and Technology Edition), 2023, 53(3): 891-901(in Chinese).
[4] 薛珊, 卢涛, 吕琼莹, 等. 基于多尺度融合和轻量化网络的无人机目标检测算法[J]. 湖南大学学报(自然科学版), 2023, 50(8): 82-93.

Xue S, Lu T, Lü Q Y, et al. Drone target detection algorithm based on multi-scale fusion and lightweight network[J]. Journal of Hunan University (Natural Sciences), 2023, 50(8): 82-93(in Chinese).

[5] 薛珊, 王亚博, 吕琼莹, 等. 基于 YOLOX-drone 的反无人机系统抗遮挡目标检测算法[J]. 工程科学学报, 2023, 45(9): 1539-1549.

Xue S, Wang Y B, Lü Q Y, et al. Anti-occlusion target detection algorithm for anti-UAV system based on YOLOX-drone[J]. Chinese Journal of Engineering, 2023, 45(9): 1539-1549(in Chinese).

[6] 白宇, 周艳媛, 安胜彪. 改进 YOLOv5 的无人机小目标检测方法研究[J]. 计算机工程与应用, 2024, 60(10): 276-284.

Bai Y, Zhou Y Y, An S B. Research on UAV small object detection method improved by YOLOv5[J]. Computer Engineering and Applications, 2024, 60(10): 276-284(in Chinese).

[7] 刘玉萍, 尚翠娟, 李明明. 改进 YOLOv11 的无人机小目标检测算法[J]. 计算机工程与应用, 2025, 61(15): 124-131.

Liu Y P, Shang C J, Li M M. Improved YOLOv11 algorithm for small target detection in UAVs[J]. Computer Engineering and Applications, 2025, 61(15): 124-131(in Chinese).

[8] 李彬, 李生林. 改进 YOLOv11n 的无人机小目标检测算法[J]. 计算机工程与应用, 2025, 61(7): 96-104.

Li B, Li S L. Improved YOLOv11n small object detection algorithm in UAV view[J]. Computer Engineering and Applications, 2025, 61(7): 96-104(in Chinese).

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

[10] 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.

[11] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2980-2988.

[12] 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.

[13] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the 14th IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 779-788.

[14] Wang C Y, Bochkovskiy A, Liao H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//Proceedings of the 14th IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 7464-7475.

[15] 陈金林, 吴一全, 苑玉彬. 无人机视角下目标检测的 YOLO 系列算法研究进展[J/OL]. 北京航空航天大学学报, 2024(2024-08-26) [2025-06-20]. <https://doi.org/10.13700/j.bh.1001-5965.2024.0420>.

Chen J L, Wu Y Q, Yuan Y B. Research progress of YOLO series algorithms for object detection from UAV perspective[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2024 (2024-08-26) [2025-06-20]. <https://doi.org/10.13700/j.bh.1001-5965.2024.0420>(in Chinese).

[16] Si Y Z, Xu H Y, Zhu X Z, et al. SCSSA: exploring the synergistic effects between spatial and channel attention[EB/OL]. (2024-11-12) [2025-06-20]. <https://arxiv.org/abs/2407.05128>.

[17] 徐慧智, 郝东升, 徐小婷, 等. 基于深度学习的高速公路小目标检测算法[J]. 吉林大学学报 (工学版), 2025, 55(6): 2003-2014.

Xu H Z, Hao D S, Xu X T, et al. Expressway small object detection algorithm based on deep learning[J]. Journal of Jilin University (Engineering and Technology Edition), 2025, 55(6): 2003-2014(in Chinese).

[18] Zhang Y F, Ren W Q, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.

[19] Zhao J, Zhang J S, Li D D, et al. Vision-based anti-UAV detection and tracking[EB/OL]. (2022-05-22) [2025-06-20]. <https://arxiv.org/abs/2205.10851>.

[20] 薛珊, 张振, 吕琼莹, 等. 基于卷积神经网络的反无人机系统图像识别方法[J]. 红外与激光工程, 2020, 49(7): 20200154.

Xue S, Zhang Z, Lv Q Y, et al. Image recognition method of anti UAV system based on convolutional neural network[J]. Infrared and Laser Engineering, 2020, 49(7): 20200154(in Chinese).

[21] 侯公羽, 陈钦煌, 杨振华, 等. 基于改进 YOLOv5 的安全帽检测算法[J]. 工程科学学报, 2024, 46(2): 329-342.

Hou G Y, Chen Q H, Yang Z H, et al. Helmet detection method based on improved YOLOv5[J]. Chinese Journal of Engineering, 2024, 46(2): 329-342(in Chinese).

Object detection method for anti-UAV systems in complex backgrounds
based on YOLOv11-drone

XUE Shan*, YANG Wenda, LI Qigui, REN Wanfei

(School of Mechanical and Electrical Engineering, Changchun University of Science and Technology, Changchun 130022, China)

Abstract: Amid the rapid expansion of the low-altitude economy, UAVs present escalating security threats to large-scale public spaces. Accurate and rapid target detection of drones has become the primary and critical component in anti-drone systems. To address the critical challenges of low detection accuracy, high miss rate, and false positives in small UAV detection within complex environments, this study proposes YOLOv11-drone, an optimized object detection framework based on an enhanced YOLOv11 architecture specifically designed for counter-UAV applications. First, an autonomous dataset of small-target drones in complex backgrounds was constructed. Second, to improve the model’s ability to extract features for drone targets, a partial spatial-channel cooperative attention module C2SCSPSA was created that combines channel attention with spatial attention processes. Furthermore, a small-target-oriented multi-level feature pyramid network (STFPN) was proposed. By fully leveraging shallow semantic information, this architecture significantly reduces the model’s parameter count while improving its focus on small targets. Finally, the model’s loss function was replaced with the EIoU metric, which simultaneously enhances regression accuracy and accelerates convergence speed. The proposed improved algorithm was evaluated on the self-constructed dataset. Experimental results demonstrate that the YOLOv11-drone model achieves a detection accuracy of 95.5%, representing a 4.1% improvement over the baseline algorithm. Additionally, the model reduces parameter count by 62% while attaining an inference speed of 83 frames per second. These advancements confirm its efficacy for small-target drone detection in complex backgrounds. To further evaluate generalization capability, experiments were conducted on the public VisDrone2019 benchmark. The suggested technique shows strong cross-domain applicability and generalization performance, achieving a 0.6% improvement in mAP@0.5 while using just 38% of the baseline model’s parameters.

Keywords: anti-UAV; small object detection; multi-scale features; YOLOv11; loss function

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0422

面向无人机支线物流的起降场容量模型与协同调度

江波^{1,2}, 李蝶³, 郑远^{4,*}, 李诚龙¹, 张照轩³

(1. 中国民用航空飞行学院 飞行技术学院, 广汉 618307; 2. 中国民用航空飞行学院 国产民机飞行与运行支持四川省工程研究中心, 广汉 618307; 3. 中国民用航空飞行学院 空中交通管理学院, 广汉 618307; 4. 中国民用航空飞行学院 计算机与人工智能学院, 广汉 618307)

摘 要: 针对公共垂直起降场终端区空域资源紧张与运行效率不足的问题, 考虑多场站协同调度机制缺失及支线物流区域的负载均衡性需求, 以精准评估容量和优化区域协同调度能力为目标, 提出一种串联耦合排队模型与静态协同调度策略。构建差异化排队模型, 刻画各环节阻塞效应, 突破传统独立分析局限; 在降落排队系统中引入非抢占式优先级策略, 保障高优先级任务; 设计“静态流量预分配+动态容流调配”策略, 通过基于排队论的双路径决策实现多场站协同。仿真结果表明: 所提串联耦合排队模型的整体容量评估偏差控制在 6% 以内, 显著优于传统独立分析模型, 且在等待层空域容量评估中, 同样显著优于基于冲突阈值的模型; 非抢占式优先级策略下, 高优先级任务平均延误时间始终低于 30 s, 但随其在总任务中占比的增加, 系统平均延误显著上升; 静态协同调度策略在突发流量下使区域负载均衡度提升至 95.9%, 无人机总延误时间平均减少 39.49%。研究成果可为低空物流网络规划与实时调度提供理论支撑。

关 键 词: 无人机; 支线物流; 垂直起降场; 排队论; 容量评估; 协同调度

中图分类号: V355

文献标志码: A

文章编号: 1001-5965(2026)09-3011-12

近年来, 倾转四旋翼无人机凭借其垂直起降与高速巡航能力相结合的独特优势, 在支线配送场景中逐渐成为提升区域物流效率和创新能力的工具。2022 年 1 月 25 日, 中国民用航空局正式向顺丰旗下大型无人机公司丰鸟科技颁发支线物流无人机试运行许可和经营许可^[1]。同年, 《“十四五”通用航空发展专项规划》也明确提出要打通航空物流“干支末”网络, 推动无人机有机融入县乡村三级物流网络体系^[2]。其中, “干线运输”依托大型无人机解决跨省际长距离贯通问题; “末端配送”以轻小型无人机实现“最后一公里”短途配送; 而“支线物流”作为连接干线枢纽与末端节点的关键环节, 既能完成城市之间大宗货物的配送或从总

仓库到各分仓库之间的包裹配送, 又能精确整合末端分散的物流需求^[3-5]。在实际运行过程中, 支线物流还呈现出显著的任务目的地区域化特性, 即在特定目的区域内通常会部署多个公共无人机起降场, 以实现区域内物流需求的多起降场动态协同承接。

然而, 随着无人机数量的急剧增加和支线物流波次性运输需求的凸显, 公共无人机起降场终端区空域资源日益紧张, 其运行效率与空域容量之间的结构性失衡问题已愈发凸显。目前, 国内外学者针对终端区容量评估的研究大多聚焦于民航终端区空域和机场跑道, 其评估方法体系主要涵盖四大技术路径^[6]: 基于计算机仿真的动态建模方法、基于

收稿日期: 2025-07-01; 录用日期: 2025-10-20; 网络出版时间: 2025-12-31 09:10

网络出版地址: link.cnki.net/urlid/11.2625.V.20251230.1130.003

基金项目: 国家自然科学基金青年基金项目 (52502410); 中国民用航空飞行学院本科科研创新基金项目 (X202510624277)

* 通信作者. E-mail: ranchozy@cafuc.edu.cn

引用格式: 江波, 李蝶, 郑远, 等. 面向无人机支线物流的起降场容量模型与协同调度 [J]. 北京航空航天大学学报, 2026, 52 (9): 3011-3022.
Jiang B, Li D, Zheng Y, et al. Capacity model and collaborative scheduling of droneports for drone regional logistics [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3011-3022 (in Chinese).

管制员工作负荷的认知分析方法、基于历史数据的统计推演方法及基于数学模型的解析计算方法。针对低空无人机的特殊性,传统评估方法面临显著局限性。一方面,支线物流任务具有静态可规划性和优先级分层特性,但可获取的历史飞行数据样本匮乏;另一方面,无人机普遍采用自主飞行模式,消除了传统民航对管制员工作负荷的强依赖,使得基于管制员工作负荷与历史数据统计的评估方法难以直接迁移使用。在此背景下,基于计算机仿真模型和基于数学模型评估方法逐渐成为低空无人机起降场容量评估的核心技术手段^[7]。Guerreiro 等^[8]提出了一种适用于城市空中交通(urban air mobility, UAM)的先到先服务垂直起降调度算法,通过对地面设施进行时间维度建模,得到了城市空中交通垂直起降场容量和吞吐量;Bulusu 等^[9]从安全性和性能角度出发提出了一种基于阈值的空域容量评估方法,将无人机空域容量界定为不同阈值条件下对应数值中的最小值;与此同时,赵焱飞等^[10]也提出了一种基于冲突阈值的评估模型,将标准冲突时间对应的最大运行架次作为空域容量;Hoekstra 等^[11]则从空间限制角度出发提出“地理矢量引导”概念,并通过限制空域内无人机的三维速度矢量的任何一个分量来降低交通复杂度,从而提高无人机空域容量。然而,这些方法尚处于起步阶段,通常仅考虑部分实际场景约束,较少考虑支线物流场景中任务优先级分层和无人机自身能耗等实际因素,使得在实际应用中存在一定的局限性。

与此同时,国内外有关学者针对传统多机场和垂直起降场终端区调度问题已开展多维度深入研究并取得显著进展。在传统民航多机场调度领域,Barnhart 等^[12]基于先到先服务原则提出了兼顾公平性与系统效能的多机场终端区航班协同管理策略。Jacquillat 等^[13]在航班排序模型中纳入容量利用率指标,通过动态剖析航班延误、航班队列与机场容量间的关联机制,深入探究多机场运行的时空交互影响。Chevalley 等^[14]创新性地整合了管制员管理的离场间隔工具与战术离场调度工具,构建了协同优化模型以提升多机场群离场航班运行效率。万旅等^[15]基于双层编码遗传算法框架,建立了兼顾总延误时间与运行风险值双目标优化的终端区航班调度模型。在垂直起降场调度研究方面,Chen 等^[16]开发了一种能实现起降场终端空域与地面运行综合管理的优化调度算法。常静等^[17]提出了一种基于动态优先级的无人机起飞时刻调度方法,通过熵权法和模糊聚类确定任务优先级,以最小化总延误成本为目标构建模型,并采用遗传算法

求解。张洪海等^[18]针对城市环境下物流无人机飞行计划调配问题,以无人机运输成本和延误成本最小为目标,建立多约束物流无人机飞行计划预先调配模型。然而,以上传统多机场及垂直起降场的调度方法在支线物流无人机起降调度中,均存在一定的局限性。这些方法大多未充分考虑支线物流无人机起降的特殊需求,导致在实际应用中难以直接实现高效、精准的调度,无法充分满足支线物流无人机起降调度复杂且多变的场景要求。

综上,尽管排队论已应用于机场终端区与低空起降场容量评估中,但往往基于一系列简化假设,未考虑环节间的阻塞效应和任务优先级,难以直接迁移到具有优先级分层需求的支线物流任务中。同时,在调度策略方面,缺乏对无人机支线物流网络特性的适配,特别是在处理静态可预知但空间分布不均的波次性需求时,难以实现区域内多起降场协同的流量精准预分配和动态再平衡。因此,本文聚焦于无人机支线物流网络的核心节点——公共垂直起降场,以排队论为核心方法,从理论建模与实际应用双维度开展研究。首先,构建“降落—地面周转—起飞”串联耦合排队模型,并针对这 3 个模块分别建立差异化排队模型,突破传统方法独立分析各环节的局限性,精准刻画环节间阻塞效应对整体容量的影响。其次,在降落排队系统中创新性地引入非抢占式优先级(non-preemptive priority, NP-P)策略,确保高优先级任务安全的同时提升系统整体运行效率。最后,设计“静态流量预分配+动态容流调配”的多起降场协同调度策略,充分利用支线物流任务目的地区域化的特性,在运行前基于任务优先级和实际容量对区域内各起降场的流量进行科学预分配,在运行中基于排队论的双路径决策进行动态调配,高优先级任务选择预期等待时间最短的起降场,低优先级任务采用归一化队长最短策略,有效解决单点过载和区域负载不均问题,提升支线物流网络的整体效率。

1 空域结构

针对支线物流高频次、集散批量和中短途的运输需求,本文特别选择了一种适用于低空货运无人机的“起降分离”终端区空域结构^[19]。该结构采用续航与载重均衡的倾转四旋翼构型无人机,以满足多架无人机同时进出场的需求,并确保无人机运行的安全性与高效性。如图 1 所示,终端进近圈作为终端空域的边界,从上到下依次为等待高度层、降落高度层、起飞高度层、中心的起飞通道及地面停机坪。其中,等待高度层为因区域突发流量过载需

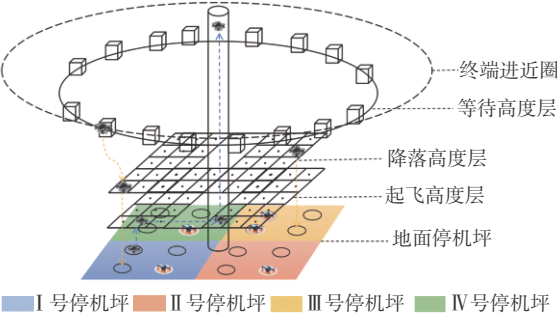


图 1 无人机起降场空域结构

Fig. 1 Airspace structure of droneport

临时悬停的无人机提供缓冲,降低高优先级物流任务的延误风险;降落高度层用于无人机下降时的缓冲,帮助无人机平稳地从巡航高度过渡到接近地面的低空区域;起飞高度层用于无人机在地面完成起飞加速后,攀升至该高度层进入中心的起飞通道,该通道为无人机提供了离场的专用路径,确保无人机能够按照既定流程和安全间隔有序起飞,避免与其他无人机发生冲突;地面停机坪被划分为多个独立的区域,支持多架次无人机并行周转,保障了终端区的运作效率。

在此基础上,为保障终端区无人机运行的安全性与高效性,精准评估空域容量并建立多维运行协同调度机制至关重要。低空空域容量通常定义为:在某特定空域环境与安全等级条件下,单位时间内能够接收的最多航空器架次^[20]。然而,在支线物流场景中,单纯依赖理论容量阈值难以全面反映系统运行状态,因为在适当的可接受延误水平下,起降场的实际服务吞吐量更能体现其运行效能。因此,采用双重评估方法,一方面通过静态计算确定理论容量上限,另一方面,运用差异化耦合排队模型对不同延误阈值下的实际吞吐量进行建模分析。进一步地,针对支线物流网络的空间负载不均特性,在评估单一场站容量的基础上,需进一步拓展至区域协同视角,通过制定有效的静态与动态流量调配策略,平衡各场站负载并提升整体运行效率。

2 终端区容量评估

终端区容量评估是无人机公共垂直起降场高效运行的核心环节,需从理论与实际 2 个维度构建评估体系。理论层面聚焦理想化条件下的极限服务能力,通过构建降落、起飞、地面 3 个模块的独立模型,界定各环节在连续服务、无延误场景下的最大架次输出能力,为系统运行提供理论上的最大容量参考。实际评估则以排队论为核心框架,将终端区视作“降落—地面周转—起飞”的串联耦合系统,充分考虑无人机到达与服务时间的随机性、环

节间阻塞效应及可接受延误水平等关键因素,实现对终端区容量的精准化评估,使评估结果更贴合支线物流运行场景需求。

2.1 理论容量评估

在终端区域的运行管理中,准确评估其理论容量是实现空域资源优化配置、保障无人机安全有序降落的关键所在。本文在理论层面将容量定义为单位时间内垂直起降场所能提供的最大无人机服务架次,并围绕这一定义,重点构建以下核心评估模型。

1) 降落容量模型。根据终端区不断到达的无人机数量,在单位时间内观测无人机的最大降落服务次数,并通过式(1)评估终端区空侧降落总容量:

C_d = c_1 \frac{T_{unit}}{\bar{T}_p^d} \tag{1}

式中: C_d为降落系统总容量; c_1为降落通道的数量; T_{unit}为单位时间; \bar{T}_p^d为平均降落服务时间,每个降落通道能够完成的最大降落服务次数主要取决于每次降落所需的平均服务时间 \bar{T}_p^d。

2) 起飞容量模型。根据终端区执行任务的无人机相继起飞数量,在单位时间内观测无人机的最大起飞服务次数,并通过式(2)评估终端区空侧起飞总容量:

C_t = c_2 \frac{T_{unit}}{\bar{T}_p^t} \tag{2}

式中: C_t为起飞系统总容量; c_2为起飞通道数量; \bar{T}_p^t为平均起飞服务时间。

3) 地面容量模型。根据上述终端区空域结构,本文所构建的地面容量模型未考虑滑行道配置等因素的影响,其容量大小主要取决于地面无人机平均周转时间(包括装卸、充电等环节)、地面服务人员数量及停机坪数量,具体通过式(3)来进行定义:

C_g = \min(n_g, n_s) \frac{T_{unit}}{\bar{T}_g} \tag{3}

式中: C_g为地面系统总容量; n_g为停机坪数量; \bar{T}_g为平均地面服务时间; n_s为地面服务人员数量,且一个地面服务人员同一时刻仅服务一架无人机。

理论容量评估为无人机起降场在理想化条件下的最大服务能力提供了定量依据,是系统资源规划与运行优化的基础参考。在后续的实际容量评估与协同调度实验中,理论容量将作为系统物理资源的上限约束,用于判断实际运行容量是否合理,并确保调度策略在资源边界内可行。

2.2 实际容量评估

排队论^[21]最早是 1909 年丹麦数学家 Erlang 为研究电话网络场景而提出的,经过几十年的发展,

理论体系已逐步趋于完备,其排队过程包含输入过程、排队规则、服务机构 3 部分,如图 2 所示。顾客在到达服务机构前,若无排队则直接接受服务后离去;若有排队则遵循排队规则和服务规则,依次通过排队结构获得服务后离去。

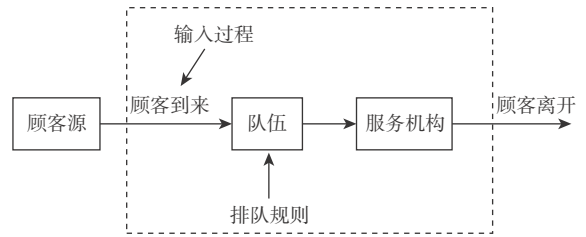


图 2 排队过程

Fig. 2 Queueing process

起降场终端区本质上是一个由降落、地面周转、起飞 3 个环节串联构成的串联耦合系统,该系统的核心特征在于各环节到达流与延误的交互制约机制:地面系统作为枢纽,承担无人机周转任务(装卸、充电等),其输入率由降落系统的实际输出决定,而输出率则直接影响起飞系统的输入流量,同时,地面容量的不足不仅会触发降落系统的服务降级,还可能成为整体系统的关键瓶颈。起飞系统作为出口,其输入流量依赖于地面系统的稳定输出,若起飞通道服务能力不足,队列积累将进一步加剧上游环节的延误传递。鉴于每架无人机的着陆时间和起飞时间是随机和相互独立的,服从负指数分布^[7],本文基于泊松过程的可叠加性,分别建立降落、地面和起飞差异化耦合排队模型。此外,为进一步优化容量评估的流程,对模型进行合理简化,并作出如下假设:

1) 机位布局满足无人机安全起降规范,无需另行计算安全间隔距离。

2) 无人机完成周转作业后即刻离场,不再占用停机坪资源。

2.2.1 降落排队系统

对于降落排队系统,无人机的相继到达时间间隔和服务时间分别服从参数为 λ_1 和 μ_1 的负指数分布,且无人机的到达流为无限源。考虑到空中悬停等待的进场无人机可能因延误增加运行风险,系统采用 NP-P 策略,与抢占式优先级策略不同, NP-P 策略不允许中断正在进行的服务,从而避免了服务切换带来的额外时间开销与系统不确定性,更适合起降场这种对运行连续性和安全性要求较高的场景。基于 NP-P 策略的运行机制及起降场无人机降落过程的实际特征,该系统符合排队论中的多服务台排队模型,可表示为 $M/M/c_1/n_1/\infty/NP-P$, 其中,第 1 个 M 代表无人机到达时间服从负指数分布,

第 2 个 M 表示降落通道的服务时间也服从负指数分布, c_1 为降落通道数量, n_1 为系统排队容量上限, ∞ 表示无限顾客源, NP-P 表示非抢占式优先级策略。系统总到达率 λ_1 由 2 部分组成:高优先级任务平均到达率 λ_h 和低优先级任务平均到达率 λ_l 。

当地面所有停机位均被占用时,降落后的无人机无法释放降落通道,即降落排队系统的有效服务率会受到下一阶段地面排队系统阻塞概率的影响。因此,根据平衡状态,可推导得出降落通道的实际服务率、系统状态概率及各项运行指标,具体计算公式如下。

1) 降落通道的实际服务率 μ_1^a 。

$$\mu_1^a = \mu_1(1 - P_c^g) \tag{4}$$

2) 系统空闲的概率 P_0^d 和系统的阻塞概率 P_c^d 。

$$P_0^d = \left[\sum_{j=0}^{c_1-1} \frac{(\lambda_1/\mu_1^a)^j}{j!} + \sum_{j=c_1}^{n_1} \frac{(\lambda_1/\mu_1^a)^j}{c_1!c_1^{j-c_1}} \right]^{-1} \tag{5}$$

$$P_c^d = \frac{(\lambda_1/\mu_1^a)^{n_1}}{c_1!c_1^{n_1-c_1}} P_0^d \tag{6}$$

3) 高、低优先级任务服务强度 ρ_h 、 ρ_l 及总服务强度 ρ_1 。

$$\rho_h = \frac{\lambda_h(1 - P_c^d)}{c_1\mu_1^a} \tag{7}$$

$$\rho_l = \frac{\lambda_l(1 - P_c^d)}{c_1\mu_1^a} \tag{8}$$

$$\rho_1 = \frac{\lambda_1(1 - P_c^d)}{c_1\mu_1^a} = \frac{(\lambda_h + \lambda_l)(1 - P_c^d)}{c_1\mu_1^a} \tag{9}$$

4) 高、低优先级无人机平均等待时间 $W_{q,h}^d$ 、 $W_{q,l}^d$ 及总平均等待时间 W_q^d 。

$$W_{q,h}^d = \frac{\rho_h\rho_1}{\lambda_h(1 - \rho_h)(1 - P_c^d)} \tag{10}$$

$$W_{q,l}^d = \frac{\rho_l\rho_1}{\lambda_l(1 - \rho_h)(1 - \rho_l)(1 - P_c^d)} \tag{11}$$

$$W_q^d = \frac{\lambda_h W_{q,h}^d + \lambda_l W_{q,l}^d}{\lambda_h + \lambda_l} \tag{12}$$

5) 可接受延误阈值下的降落排队系统容量 C_{ac}^d 。

$$C_{ac}^d = \frac{\rho_1(\rho_h + \rho_l) - \rho_1^2}{T_{ac}^d(1 - \rho_h)(1 - \rho_l)(1 - P_c^d)} \tag{13}$$

式中: μ_1 为单位时间内单个降落通道服务强度; P_c^g 为地面排队系统单位时间内的阻塞概率,在本模型中用于修正降落通道的实际服务率; T_{ac}^d 为可接受延误阈值,是降落排队系统为保障运行效率与服务质量所设定的最大允许平均等待时间,其等待的平均时长不得超过该阈值,否则系统服务能力将被判定为无法满足实际需求。

2.2.2 地面排队系统

对于地面排队系统, 假设地面服务人员为停机坪上无人机提供服务的过程, 依然满足泊松流的特征; 并且由于无人机进入地面排队系统的速率 λ_2 并非无限, 而是由降落系统的输出情况决定。一旦所有停机坪皆被占用, 地面排队系统便不再接纳新的无人机。由此可见, 垂直起降场停机坪的运行特征与排队论中有限源的 $M/M/c_2/n_g$ 排队模型相契合。在该模型中, c_2 表示地面服务人员数量, n_g 表示地面停机坪数量, 即系统排队容量上限为 n_g 。根据平衡状态, 可推导出系统状态概率及各项运行指标, 具体计算公式如下。

1) 稳态时系统大小的概率 P_k^g 及系统空闲概率 P_0^g 。

$$P_k^g = \begin{cases} \frac{(\lambda_2/\mu_2)^k}{k!} P_0^g & 0 \leq k < c_2 \\ \frac{(\lambda_2/\mu_2)^k}{c_2! c_2^{k-c_2}} P_0^g & c_2 \leq k \leq n_g \end{cases} \quad (14)$$

$$P_0^g = \left[\sum_{k=0}^{c_2-1} \frac{(\lambda_2/\mu_2)^k}{k!} + \sum_{k=c_2}^{n_g} \frac{(\lambda_2/\mu_2)^k}{c_2! c_2^{k-c_2}} \right]^{-1} \quad (15)$$

2) 系统的阻塞概率 P_c^g 。

$$P_c^g = \frac{(\lambda_2/\mu_2)^{n_g}}{c_2! c_2^{n_g-c_2}} P_0^g \quad (16)$$

3) 平均正在接受服务的无人机数量 L_s 及平均等待时间 W_q^g 。

$$L_s = \frac{\lambda_2(1-P_c^g)}{\mu_2} \quad (17)$$

$$W_q^g = \frac{\sum_{k=c_2+1}^{n_g} (k-c_2) P_k^g}{\lambda_2(1-P_c^g)} \quad (18)$$

4) 可接受延误阈值下的地面排队系统容量 C_{ac}^g 。

$$C_{ac}^g = \frac{\sum_{k=c_2+1}^{n_g} (k-c_2) P_k^g}{T_{ac}^g(1-P_c^g)} \quad (19)$$

式中: λ_2 为地面排队系统到达率; μ_2 为单位时间内单个地面服务人员的服务强度; k 为地面排队系统中的任务数量; T_{ac}^g 为地面排队系统的可接受延误阈值。

2.2.3 起飞排队系统

对于起飞排队系统, 每架无人机完成地面服务后按照泊松过程请求离场, 当起飞通道空闲时, 无人机垂直上升至起飞高度层, 并在起飞高度层通过平飞进入起飞通道, 从而完成无人机的离场。与此同时, 在同一时刻该起降场可满足 c_3 架无人机的起飞, 未被地面服务占用的停机坪数量为该系统排队

队列上限。因此, 起飞排队系统符合排队论中 $M/M/c_3/n_2$ 有限源排队模型, 其中, c_3 表示起飞通道数量, n_2 为起飞排队系统的排队容量上限。根据平衡状态, 可推导出系统排队容量上限、系统状态概率及各项运行指标, 具体计算公式如下。

1) 系统排队容量上限 n_2 。

$$n_2 = n_g - L_s \quad (20)$$

2) 系统稳态大小概率 P_r^t 及系统空闲概率 P_0^t 。

$$P_r^t = \begin{cases} \frac{(\lambda_3/\mu_3)^r}{r!} P_0^t & 0 \leq r < c_3 \\ \frac{(\lambda_3/\mu_3)^r}{c_3! c_3^{r-c_3}} P_0^t & c_3 \leq r \leq n_2 \end{cases} \quad (21)$$

$$P_0^t = \left[\sum_{r=0}^{c_3-1} \frac{(\lambda_3/\mu_3)^r}{r!} + \sum_{r=c_3}^{n_2} \frac{(\lambda_3/\mu_3)^r}{c_3! c_3^{r-c_3}} \right]^{-1} \quad (22)$$

3) 系统阻塞概率 P_c^t 及平均等待时间 W_q^t 。

$$P_c^t = \frac{(\lambda_3/\mu_3)^{n_2}}{c_3! c_3^{n_2-c_3}} P_0^t \quad (23)$$

$$W_q^t = \frac{\sum_{r=c_3+1}^{n_2} (r-c_3) P_r^t}{\lambda_3(1-P_c^t)} \quad (24)$$

4) 可接受延误阈值下的起飞排队系统容量 C_{ac}^t 。

$$C_{ac}^t = \frac{\sum_{r=c_3+1}^{n_2} (r-c_3) P_r^t}{T_{ac}^t(1-P_c^t)} \quad (25)$$

式中: λ_3 为起飞排队系统到达率; μ_3 为单位时间内单个起飞通道的服务强度; r 为起飞排队系统中的任务数量; T_{ac}^t 为起飞排队系统的可接受延误阈值。

在起降场终端区运行管理体系中, 降落过程、地面周转过程与起飞过程 3 个环节共同构成了一个相互关联的有机整体, 任一环节的容量限制都会成为系统运行的瓶颈。因此, 在评估无人机终端区的实际容量时, 需明确无人机在各环节所能承受的最大可接受延误阈值, 并依据延误曲线确定该时间对应的到达流量, 以此作为各环节的实际运行容量, 而终端区的整体容量, 实际上是由这三者中容量最小的一环所决定, 即取三者中的最小值作为系统运行的实际容量。

$$C_a = \min(C_{ac}^d, C_{ac}^g, C_{ac}^t) \quad (26)$$

式中: C_a 为起降场终端区的运行实际容量; C_{ac}^d 、 C_{ac}^g 、 C_{ac}^t 分别为最大可接受延误阈值下, 通过降落排队系统模型(式(13))、地面排队系统模型(式(19))、起飞排队系统模型(式(25))计算得到的各环节最大可承载流量。

3 基于容流的多起降场协调调度

在无人机支线物流网络日益发展的背景下,单个起降场难以应对波次性高峰和空间负载不均的双重挑战,多起降场协同调度成为提升区域运行效率的关键所在。同时,由于支线物流任务可预知性高但时空分布不均的特性,调度机制遵循“先静态分配、后动态调节”的逻辑,即在运行前通过基于优先级分层的静态流量预分配奠定基础框架,运行过程中引入基于排队论的双路径决策方法,实现物流需求与空域资源的精准匹配。在这一调度体系中,实际容量作为核心依据直接指导流量的分配与动态调整,而理论容量则作为物理资源上限,用于确保调度方案的可行性与系统安全性。二者相互支撑、协同作用,共同保障无人机物流系统的高效与稳定运行。

3.1 静态流量分配

为适配支线物流任务静态可规划性及目的地区域化特性,本文提出基于优先级分层的静态流量预分配方法。该方法的核心在于将无人机的任务级别精细化地划分为高(如医疗物资配送)、低(普通包裹)2个层级,并严格遵循“高优先级任务优先保障、容量比例公平分配”的核心原则,以实现区域内各起降场初始流量的科学配置。

首先,在起降场容量允许的范围内,高优先级任务将获得优先分配的权利,以确保其能够优先分配至起降场。根据式(27)计算单位小时内分配给第*i*个起降场的高优先级任务量。

$$\lambda_{h,i} = T_h \frac{C_{a,i}}{C_{a,m}} \quad (27)$$

式中: $\lambda_{h,i}$ 为第*i*个起降场的高优先级任务量; T_h 为单位小时内到达流中高优先级任务总量; $C_{a,i}$ 为单位小时内第*i*个起降场的实际容量。假设区域内共有*m*个起降场,则 $C_{a,m}$ 为单位小时内*m*个起降场的实际容量总和。

其次,为保障区域内各起降场间资源的公平分配,低优先级任务的分配也将根据各起降场的实际容量进行比例分配。通过式(28)、式(29)计算得到单位小时内分配给第*i*个起降场的低优先级任务量及总任务量。

$$\lambda_{l,i} = T_l \frac{C_{a,i} - \lambda_{h,i}}{C_{a,m} - T_h} \quad (28)$$

$$\lambda_i = \lambda_{h,i} + \lambda_{l,i} \quad (29)$$

式中: $\lambda_{l,i}$ 和 λ_i 分别为第*i*个起降场的低优先级任务量和总任务量; T_l 为单位时间内到达流中低优先级

任务总量。同时,为保证实际运行时的空域资源合理利用与运行安全性,各个起降场单位小时内的静态流量预分配结果要小于对应的起降场单位小时内实际容量,即 $0 \leq \lambda_i < C_{a,i}$ 。

3.2 动态流量调配

针对支线物流目的地区域化场景下的波次性突发需求及区域负载动态不均衡的问题,静态流量分配与实时容量之间极易产生显著失配。为此,本文在静态流量预分配的基础上提出一种基于起降场负载率自适应触发阈值机制的双路径决策方法,如图3所示。当新任务到达时,系统首先判断当前流量总和是否超过起降场容量与自适应触发阈值的乘积。若未超过,则按静态预分配方案运行;若超过,该策略便通过优先级加权的队列最短(priority-weighted join the shortest queue, P-JSQ)优先机制,使高优先级任务选择预期等待时间最短的起降场,以减少关键任务的延迟;低优先级任务则遵循传统JSQ策略,选择归一化队长最短的起降场,以分配起降场的突发流量,进而平衡区域内的系统负载。容流调配流程如图3所示。

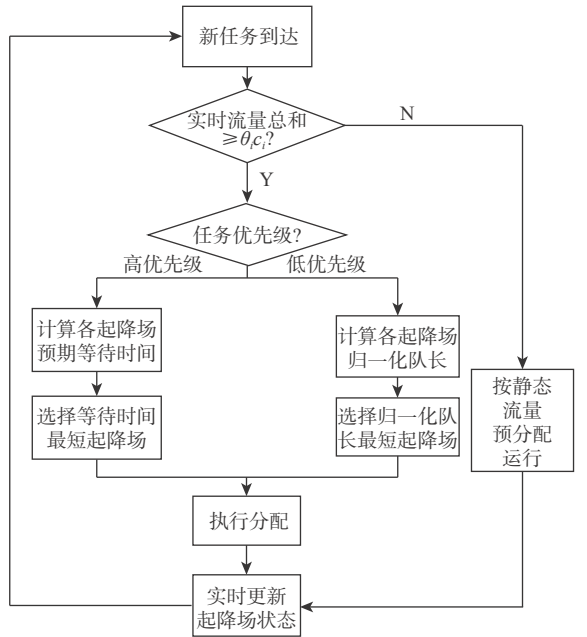


图3 容流调配流程

Fig. 3 Flowchart for capacity-flow regulation

3.2.1 基于起降场负载率的自适应触发机制

为精准刻画起降场整体繁忙程度与调度触发敏感度的动态关联,通过式(30)对各起降场自适应触发阈值进行数学推导。

$$\theta_i = \min \left(\frac{C_{a,i}}{\lambda_i}, 0.85 \right) \quad (30)$$

式中: θ_i 为第*i*个起降场使用的自适应触发阈值。

同时,基于起降场终端区所构建的“降落一地

面周转一起飞”串联耦合系统模型, 结合式(12)、式(18)及式(24)的数学推导, 通过式(31)计算可得第 i 个起降场的预期等待时间。

$$W_i = W_q^d + W_q^g + W_q^t \quad (31)$$

式中: W_i 为第 i 个起降场的预期等待时长。

同时, 根据排队论中串联耦合系统的分层排队特性, 可通过式(32)~式(34)分别计算第 i 个起降场各环节排队的排队队长。

$$L_i^d = \lambda_h W_{q,h}^d + \lambda_l W_{q,l}^d \quad (32)$$

$$L_i^g = \sum_{k=c_2+1}^{n_g} (k - c_2) P_k^g \quad (33)$$

$$L_i^t = \sum_{r=c_3+1}^{n_t} (r - c_3) P_r^t \quad (34)$$

式中: L_i^d 、 L_i^g 和 L_i^t 分别为第 i 个起降场的降落排队队长、地面周转排队队长和起飞排队队长; n_g 为地面停机坪数量, 对应地面排队系统排队容量上限。

然而, 区域内每个起降场的实际容量存在显著差异, 仅依靠绝对队长进行调度决策将导致容量较大的起降场因绝对队长较长而被误判为过载状态, 而容量较小的起降场则可能因绝对队长较短而被低估负载, 故通过式(35)对绝对队长进行归一化处理, 以消除不同起降场容量差异的影响。

$$N_i = \frac{L_i^d + L_i^g + L_i^t}{C_{a,i}} \quad (35)$$

式中: N_i 为第 i 个起降场的归一化队长。

3.2.2 调度效果评估指标

为量化区域内多场站协同调度中的流量分布均匀性, 结合排队论中容量利用率的统计特性, 通过式(36)~(38)分别计算单一起降场的容量利用率、区域平均容量利用率及负载均衡度指标。

$$v_i = \frac{\lambda_i^a}{C_{a,i}} \quad (36)$$

$$\bar{v} = \frac{\sum_{i=1}^m v_i}{m} \quad (37)$$

$$E = 1 - \frac{1}{\bar{v}} \sqrt{\frac{1}{m} \sum_{i=1}^m (v_i - \bar{v})^2} \quad (38)$$

式中: λ_i^a 为第 i 个起降场的实时到达流量; v_i 为第 i 个起降场的容量利用率; $\bar{v}$ 为 m 个起降场的平均容量利用率; $E(0 \leq E \leq 1)$ 为 m 个起降场的负载均衡度, $E = 1$ 时表示完全均衡, $E = 0$ 时表示极度不均衡。

4 仿真实验

为全面验证本文方法在支线物流实际应用场景下的有效性和优势, 开展了一系列实验仿真。本节从容量评估准确性、优先级策略有效性、区域内多起降场协同调度性能 3 个方面进行研究, 通过对比分析, 不仅揭示了本文串联耦合排队模型在容量评估中的高精度特性, 还充分展示了 NP-P 策略对高优先级任务无人机的保障能力及多起降场协同调度策略在应对区域内物流波次性突发需求和降低系统总延误方面的卓越表现。同时, 为确保实验结果的可复现性, 所有测试均在统一软硬件环境的 Win10 系统中完成, 并使用 Python 3.11.4 作为数据处理和算法实现的主要编程语言。

由于无人机在空域内悬停等待的过程中, 容易受到多机间气动耦合干扰效应、低空风场、电磁环境、导航定位误差的影响^[22], 参考相关研究的理论依据与相关类似研究的经验^[22-24], 在本文后续开展的实验中, 默认等待层的半径 $r_1 = 108 \text{ m}$ 。

4.1 容量准确性评估对比实验

为保障无人机支线物流起降场容量评估的精准性, 本文以传统独立分析模型与冲突阈值模型为对照对象开展性能对比实验, 系统验证所提串联耦合排队模型在多评估维度下的核心优势。

4.1.1 与传统独立分析模型的对比实验

为验证串联耦合模型对“降落—地面周转—起飞”环节交互效应的刻画能力, 本文采用蒙特卡罗方法构建终端区仿真运行模型, 全面模拟小型倾转四旋翼无人机在终端区的整个运行周期, 包括降落、地面周转和起飞等各个阶段。同时, 选取文献[7]中基于排队论的独立分析模型作为对照组, 进行终端区容量评估的对比实验。

1) 基础设施及无人机运行参数设置。根据等待层的半径和无人机安全水平间隔^[18], 结合三角函数关系式(39), 可得等待高度层上的等待位数量 $n_1 = 34$ 、停机坪数量 $n_g = 12$ 、降落通道数量 $c_1 = 1$ 、地面服务人员数量 $c_2 = 5$ 、起飞通道数量 $c_3 = 1$ 。其中, 离场所需时间 $t_{\text{dep}} = 30 \text{ s}$ 为无人机从地面停机坪起飞至离开起飞通道的时间; 降落所需时间 $t_{\text{apr}} = 30 \text{ s}$ 为无人机从进入终端区到着陆所需的降落服务时间。同时, 为更真实地模拟实际场景中可能出现的地面中转延迟, 每次实验随机在 $(120, 240) \text{ s}$ 范围内生成地面中转时间。

$$n_1 = \frac{\pi}{\arcsin\left(\frac{D_x}{2r_1}\right)} \quad (39)$$

式中: D_x 为无人机安全水平间隔。

2) 仿真运行环境与初始条件设置。蒙特卡罗实验方法是计算不同规模无人机来流下终端区整体的平均延误时间,通过选取可接受的延误时间来确定对应的到达流,此时的到达流即为终端区的空域容量。每次仿真运行时长为3 600 s,采用蒙特卡罗方法进行100次重复实验取平均值。无人机到达间隔服从指数分布。初始条件设置为终端区内无无人机,所有资源(降落通道、起飞通道、停机位和地面服务人员)均处于空闲状态。

根据本文所构建的串联耦合排队模型和对照组排队模型,可计算得到降落、地面周转和起飞3个子系统的容量评估值,如表1和表2所示。其中,降落、地面周转和起飞排队系统对应的理论容量分别为120架次/h、103.97架次/h、120架次/h,并作为实际容量的物理资源上限。本文的降落排队系统容量增速显著低于地面服务与起飞系统,表明其容量瓶颈集中于降落排队系统,需通过增加降落通道或优化优先级策略来提升效率。而对照组的地面服务系统容量增速滞后于起飞系统,其瓶颈位于地面服务排队系统,需通过扩大停机坪规模或缩短地面服务时间优化。

表1 本文在不同延时条件下的各排队系统容量

Table 1 Capacity of each queuing system under different delays in this research

可接受延误时间/s	容量/(架次·h ⁻¹)		
	降落	地面	起飞
30	60.0	68.8	61.1
60	77.5	84.6	86.0
90	85.6	95.4	105.0
120	89.8	115.2	115.0
150	92.6	129.6	119.4
180	94.7	155.4	120.0

表2 对照组在不同延时条件下的各排队系统容量

Table 2 Capacity of each queuing system under different delays in control group

可接受延误时间/s	容量/(架次·h ⁻¹)		
	降落	地面	起飞
30	60.0	64.6	61.1
60	79.8	73.7	86.0
90	90.0	80.8	105.0
120	96.0	83.6	115.0
150	100.2	85.5	119.4
180	102.8	87.3	120.0

如图4所示,3组数据均随着可接受延误时间的增加而趋于平缓,符合排队论中“服务效率趋近

饱和时边际增益递减”的规律,这对于支线物流运营规划具有重要参考价值。值得注意的是,本文串联耦合排队模型得到的终端区整体容量值,与蒙特卡罗仿真结果(仿真组)结果的相对偏差始终不超过6%,而对照组模型的偏差则达到12%以上,这证实了本文提出的串联耦合排队建模方法能更精准刻画降落—地面周转—起飞环节的交互影响,避免了各模块独立分析导致的容量高估或低估问题。

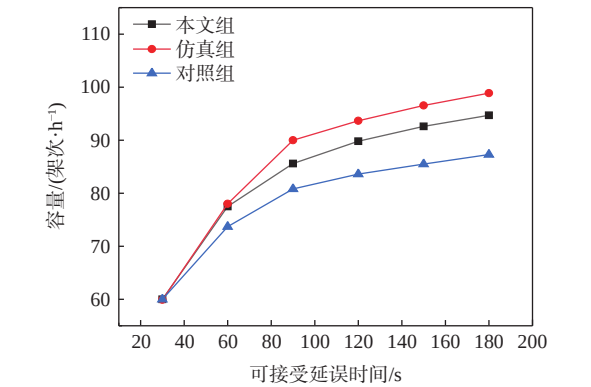


图4 不同延时下的系统容量变化曲线

Fig. 4 System capacity variation curves under different delays

4.1.2 与冲突阈值模型的对比实验

本节将基于冲突阈值模型的容量评估方法作为对照,以验证本文所提容量评估模型在支线物流场景中的通用优势。基于冲突阈值模型容量评估方法的核心逻辑是以可接受的冲突水平作为空域容量的度量标准,并通过仿真手段量化单位时间内空域能安全容纳的最大无人机架次^[10]。由于本文所提空域构型在设计时已经考虑了无人机的安全分离,其起飞通道与降落通道在同一时间仅允许一架无人机使用,不会发生冲突。因此,本节容量评估对比实验仅针对空域中等待高度层部分开展,分别使用本文提出的排队论模型和基于冲突阈值模型的评估方法对等待高度层空域容量进行评估,并对比蒙特卡罗实验得出的容量基准,判断2种方法评估的准确性。具体的实验设置如下:

- 1) 基于排队论的容量评估方法详见2.2.1节。
- 2) 基于冲突阈值模型的容量评估方法是:设置不同规模的无人机到达流,计算每种到达流下等待高度层无人机数量最多时的两机水平间隔,随着无人机到达流的增加,等待高度层无人机间隔逐渐缩小,直到缩小至最小安全间隔前对应的来流即为空域容量。
- 3) 蒙特卡罗实验方法是:计算不同规模无人机来流下的平均等待时间,随着无人机到达流的增加,平均等待时间逐渐增大,直到增大至无人机可接受延误阈值下对应的来流即为空域容量。

上述 3 个实验的参数设置如下: 无人机来流从 40 架次/h 逐渐增加至 140 架次/h; 停机坪数量 $n_g = 12$ 、地面服务人员数量 $c_2 = 5$; 降落通道数量 $c_1 = 1$; 等待层半径 $r_1 = 108\text{ m}$; 地面中转时间在 (120, 240)s 范围内随机生成; 无人机最小安全间隔为 30 m ^[10], 无人机可接受延误阈值时间为 90 s 。3 种不同策略下的容量变化如图 5 所示。

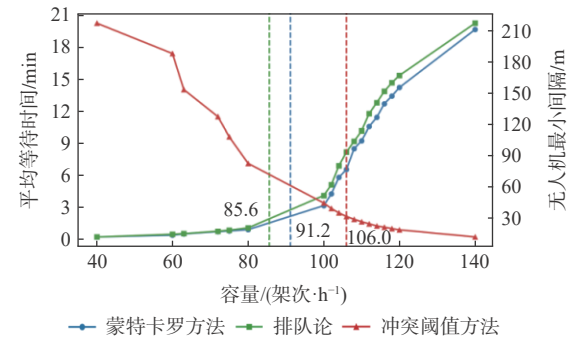


图 5 3 种不同策略下的容量变化曲线

Fig. 5 Capacity variation curves under three different strategies

实验结果表明, 当可接受延误时间为 90 s 时, 基于蒙特卡罗仿真的起降场实际容量为 91.2 架次/h , 本文串联耦合排队论模型的容量评估值为 85.6 架次/h ; 而当无人机最小间隔为 30 m 时, 基于冲突阈值模型的容量评估值为 106 架次/h 。三者对比可见, 冲突阈值模型的评估结果显著高于实际容量与本文模型评估值, 且本文模型与实际容量的偏差仅为 6.1% , 远低于冲突阈值模型 16.2% 的偏差。通过详细分析基于冲突阈值模型容量评估方法的平均等待时间发现, 当到达流来到 106 架次/h 时, 无人机平均等待时间已经远超可接受的延误时间。因此, 造成基于冲突阈值容量评估方法结果偏大的核心原因是未充分考虑无人机在空中的续航能力。

4.2 不同任务优先级比例下的系统性能实验

为验证本文所提 NP-P 策略的有效性, 基于排队论对降落排队模型 $M/M/c_1/n_1/\infty/\text{NP-P}$ 和 $M/M/c_1/n_1/\infty/\text{FCFS}$ 进行比较, 即在相同条件下对比 2 种策略的性能差异, 并分析系统内无人机平均延误时间的变化规律及优先级策略的实际效果。实验将控制无人机到达时剩余电量充足, 根据任务优先级将到达无人机分为 2 组 (任务优先级 1 的无人机优先于任务优先级 2 的无人机)。具体实验参数设置如下:

- 1) 基础设施及无人机运行相关参数沿用 4.1.1 节实验中的设置。
- 2) 无人机飞行任务优先级设置。通过随机生成的方式, 设定 3 组到达流中无人机任务的优先级占比。其中, ω_1 代表高优先级任务占比比例, ω_2 代表低优先级任务占比比例, 则 3 组到达流的优先级

占比分别为: $\omega_1 = 0.3, \omega_2 = 0.7$; $\omega_1 = 0.5, \omega_2 = 0.5$; $\omega_1 = 0.7, \omega_2 = 0.3$ 。

1) 结合图 6 分析可知, 基于优先级的 NP-P 策略相较传统的 FCFS 策略展现出显著优势。FCFS 策略因无任务紧急度区分, 高负载时易出现“优先级倒置”, 导致高优先级任务滞留, 而图 6 中曲线表明, 无论无人机任务的优先级比例如何变化, NP-P 策略的高优先级无人机平均延误时间均显著低于 FCFS 策略, 验证了优先级感知调度在物流场景中的适用性。

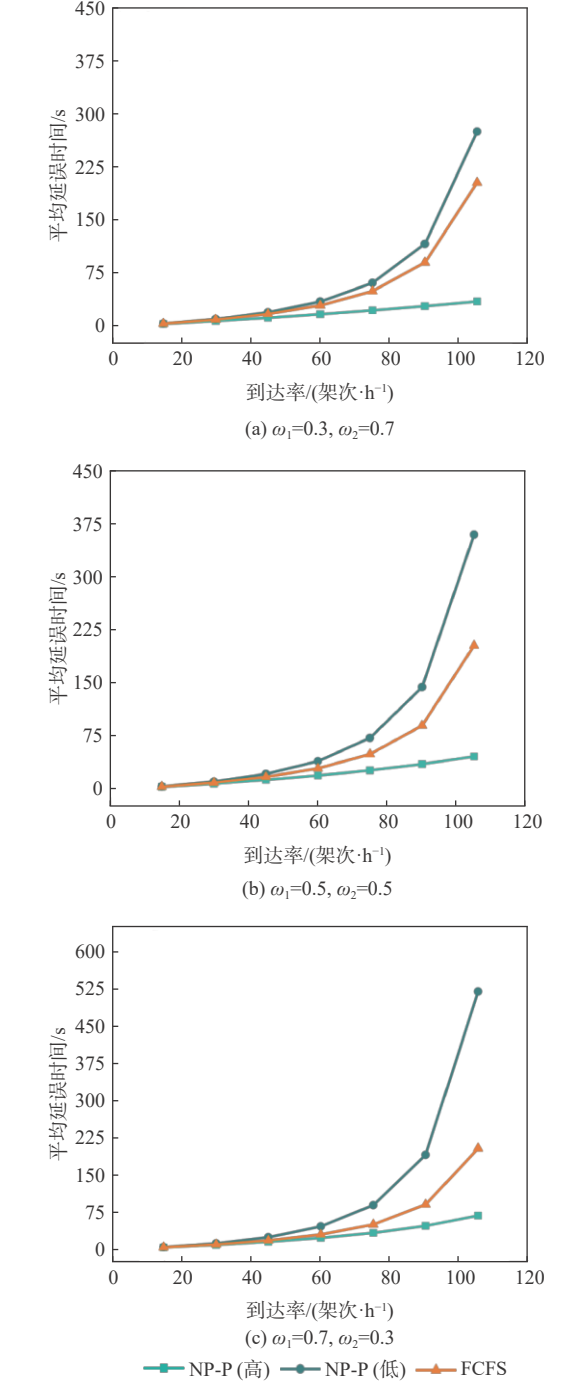


图 6 2 种策略下的无人机平均延误时间曲线

Fig. 6 Curves of average waiting duration of drones under two strategies

2) 随着到达流中高优先级任务占比的不断增加,无人机的平均延误时间均呈现上升趋势。其中,图 6(c) 中高优先级任务和低优先级任务无人机的平均延误时间分别为 28.60 s、126.04 s。相较于图 6(a) 中的对应数据,分别上升了 53.85% 和 68.84%,这体现出高优先级任务占比的提高会加重整体的延误情况。不过,高优先级任务的平均延误时间一直保持在较低水平,且其绝对值始终维持在 30 s 以内,这充分证实了本实验所设定的 NP-P 策略能够有效确保高优先级任务无人机的运行效率和运行安全。但需要注意的是,低优先级任务的无人机会随着高优先级任务占比的增加而面临着较大的运行压力。

4.3 多起降场协同调度实验

为系统评估基于“静态流量预分配+动态容流调配”协同机制的多起降场协同调度策略对提升支线物流网络整体效能的作用,本文在静态流量预分配的基础上,重点验证 P-JSQ 策略在应对突发物流需求、解决区域负载不均及优化系统延误方面的实际效能。具体设置如下:

1) 基础设施及无人机运行相关参数沿用 4.1.1 节实验中的设置。其中,高低优先级任务占比比例为: $\omega_1 = 0.3, \omega_2 = 0.7$ 。同时,当某一起降场的实时流量总和大于该起降场基于负载率的自适应阈值与容量的乘积时,便会启动 P-JSQ 策略。

2) 静态流量预分配。假设区域内有 4 个起降场及可接受延误水平为 90 s 时,基于上述串联耦合排队论模型(式(13)、式(19)和式(25)),可获取单位时间内各起降场的实际容量。同时,假设单位小时内初始到达任务总量为 180 架次/h,则 $T_h = 54$ 架次/h, $T_l = 126$ 架次/h,依据优先级分层静态流量预分配公式(式(27)~式(29)),计算单位时间内各起降场的任务总量,具体情况如表 3 所示。

表 3 起降场终端区静态流量预分配				
Table 3 Droneport terminal area static flow pre-allocation				架次/h
起降场	高优先级任务总量	低优先级任务总量	预分配任务总量	
起降场a	13.0	30.0	43.0	85.6
起降场b	17.0	39.0	56.0	110.0
起降场c	14.0	32.0	46.0	92.0
起降场d	10.0	25.0	35.0	70.0

由图 7 可知,当目的区域内起降场 a 遭遇物流波次性突发需求导致容量过载时,“静态流量预分配+动态容流调配”的双层机制能够迅速响应,并通过基于起降场负载率的自适应触发阈值机制使起降场 a 的实时流量总和降至实际容量与阈值乘

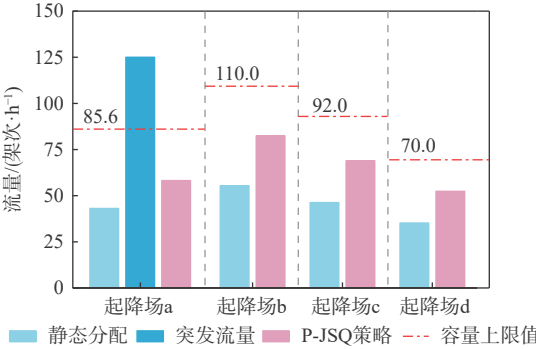


图 7 物流波次性突发需求下的多起降场协同调度
Fig. 7 Multi-droneport coordinated scheduling under logistic wave-like sudden demands

积以下。在整个响应过程中, P-JSQ 策略展现出了卓越的计算效率,平均耗时仅为 1.2 ms,这一特性使其在需要极低延迟决策的实时调度系统中具备显著优势。

与此同时,邻近起降场 b、c、d 的容量利用率分别提升至 68.0%、74.9%、74.9%。而且,得益于动态容流调配对负载的再平衡作用,目的区域内的负载均衡度也提升至 95.9%,实现了区域内流量的动态平衡,并避免了邻近场因分流导致的二次过载。

同时,为定量分析起降场在不同物流负载强度下,采用 P-JSQ 策略对无人机平均延误时间的改善效果,以起降场 a 作为研究对象、到达流作为变量,通过上文构建的降落排队模型和调度方法分别计算存在 P-JSQ 策略和不存在 P-JSQ 策略 2 种情况下的平均延误时间数据进行对比实验,如图 8 所示。当到达流为 80 架次/h 和 100 架次/h,起降场 a 的自适应触发阈值均为 0.85,此时,实时流量总和大于起降场容量与自适应触发阈值的乘积,启动 P-JSQ 策略。

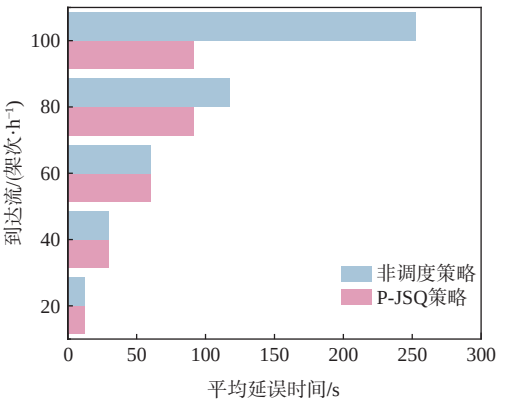


图 8 2 种策略下的无人机平均延误时间
Fig. 8 Average drone delay time under two strategies

实验结果表明,在实时流量总和超过起降场容量与自适应触发阈值的乘积时,采用 P-JSQ 策略可显著改善系统性能,其无人机平均延误时间成本最

大降幅可达 160.7 s。相较于单起降场非调动策略, 无人机总延误时间平均减少了约 39.49%, 有效缓解了支线物流网络在突发流量高峰下的队列拥塞问题。

5 结 论

1) 本文提出的串联耦合排队模型经蒙特卡罗仿真验证, 其整体容量评估偏差控制在 6% 以内, 显著优于对照组模型, 有效解决了传统方法独立分析各环节导致的容量偏差问题。同时在等待高度层空域容量评估实验中, 本文模型偏差远小于基于冲突阈值的模型, 验证了本文的容量评估模型在支线物流场景中的通用趋势。

2) 在降落排队系统中引入的 NP-P 策略, 在保障高优先级任务平均延误始终低于 30 s 的同时, 验证了优先级比例上升对系统平均延误的影响规律, 为物流任务分级管理提供了量化依据。

3) 所设计的“静态流量预分配+动态容流调配”协同调度策略, 在应对支线物流突发需求时, 可使区域负载均衡度提升至 95.9%, 并让无人机的总延误时间平均降低 39.49%, 显著增强了支线物流网络在突发流量冲击下的系统韧性。

本文的多起降场协同调度策略对支线物流场景的静态可规划性与波次性需求适配性强, 但未考虑极端天气等外部扰动因素, 未来需进一步融合动态环境约束优化调度策略。

参考文献 (References)

[1] 中国民用航空局. 民航局颁发支线物流无人机商业试运行牌照 [EB/OL]. (2022-01-25)[2025-04-17]. https://www.caac.gov.cn/XW-ZX/MHYW/202201/t20220125_211269.html.
Civil Aviation Administration of China. CAAC issues commercial trial operation license for regional logistics drones[EB/OL]. (2022-01-25)[2025-04-17]. https://www.caac.gov.cn/XWZ/X/MHYW/202201/t20220125_211269.html(in Chinese).

[2] 中国民用航空局. 民航局关于印发《“十四五”通用航空发展专项规划》的通知[EB/OL]. (2022-06-13)[2025-04-17]. https://www.caac.gov.cn/XXGK/XXGK/FZGH/202206/t20220613_213643.html.
Civil Aviation Administration of China. Notice of the CAAC on issuing the "14th Five-Year Plan" special plan for general aviation development[EB/OL]. (2022-06-13)[2025-04-17]. https://www.caac.gov.cn/XXGK/XXGK/FZGH/202206/t20220613_213643.html(in Chinese).

[3] 郑文娟. 低空物流发展前景与建议[J]. 空运商务, 2025(3): 16-21.
Zheng W J. Development prospect and suggestion of low-altitude logistics[J]. Air Transport & Business, 2025(3): 16-21(in Chinese).

[4] 许建华, 李权. 支线货运无人机的发展现状、运行方式与关键技术分析[J]. 航空工程进展, 2022, 13(4): 1-10.
Xu J H, Li Q. Development, running mode and key technologies analysis of regional cargo drones[J]. Advances in Aeronautical Science and Engineering, 2022, 13(4): 1-10(in Chinese).

[5] 袁于斐. 基于主从模式的支线-末端无人机协同物流调度方法研究[D]. 长沙: 中南大学, 2023: 22-23.
Yuan Y F. Research on the method for logistics scheduling of collaborative feeder-end UAV based on master-slave mode[D]. Changsha: Central South University, 2023: 22-23(in Chinese).

[6] 张洪海, 夷珈, 李姗, 等. 低空空域容量评估研究综述[J]. 交通运输工程学报, 2023, 23(6): 78-93.
Zhang H H, Yi J, Li S, et al. Review on research of low-altitude airspace capacity evaluation[J]. Journal of Traffic and Transportation Engineering, 2023, 23(6): 78-93(in Chinese).

[7] Zhang H H, Fei Y H, Li J Y, et al. Method of vertiport capacity assessment based on queuing theory of unmanned aerial vehicles[J]. Sustainability, 2023, 15(1): 709.

[8] Guerreiro N M, Hagen G E, Maddalon J M, et al. Capacity and throughput of urban air mobility vertiports with a first-come, first-served vertiport scheduling algorithm[C]//Proceedings of the AIAA Aviation 2020 Forum. Reston: AIAA, 2020: 2903.

[9] Bulusu V, Polishchuk V, Sengupta R, et al. Capacity estimation for low altitude airspace[C]//Proceedings of the 17th AIAA Aviation Technology, Integration, and Operations Conference. Reston: AIAA, 2017: 4266.

[10] 赵巍飞, 张志玮. 基于冲突阈值的城市物流无人机空域容量评估[J]. 中国民航大学学报, 2021, 39(5): 1-6.
Zhao Y F, Zhang Z W. Estimation of urban logistics UAV airspace capacity based on conflict threshold[J]. Journal of Civil Aviation University of China, 2021, 39(5): 1-6(in Chinese).

[11] Hoekstra J M, Ellerbroek J, Sunil E, et al. Geovectoring: reducing traffic complexity to increase the capacity of UAV airspace[C]//Proceedings of the International Conference for Research in Air Transportation. Barcelona: ICRAT, 2018: 1-8.

[12] Barnhart C, Bertsimas D, Caramanis C, et al. Equitable and efficient coordination in traffic flow management[J]. Transportation Science, 2012, 46(2): 262-280.

[13] Jacquillat A, Odoni A R. Endogenous control of service rates in stochastic and dynamic queuing models of airport congestion[J]. Transportation Research Part E: Logistics and Transportation Review, 2015, 73: 133-151.

[14] Chevalley E, Parke B, Kraut J, et al. Scheduling and delivering aircraft to departure fixes in the NY metroplex with controller-managed spacing tools: AIAA 2015-2428[R]. Reston: AIAA, 2015.

[15] 万旅, 邵莹. 基于双层遗传编码的多机场航班调度优化研究[J]. 航空计算技术, 2023, 53(5): 39-43.
Wan L, Shao Q. Research on multi-airport flight scheduling optimization based on two-layer coding genetic algorithm[J]. Aeronautical Computing Technique, 2023, 53(5): 39-43(in Chinese).

[16] Chen S, Wei P, Krois P, et al. Arrival management for high-density vertiport and terminal airspace operations[C]//Proceedings of the 2023 Air Traffic Control Association Annual Conference. Washington, D. C.: Air Traffic Control Association, 2023.

[17] 常静, 张启钱, 张洪海, 等. 城市物流无人机起飞时刻排序问题研究[J]. 武汉理工大学学报 (交通科学与工程版), 2024, 48(3): 435-440.
Chang J, Zhang Q Q, Zhang H H, et al. Research on the scheduling problem of urban logistics UAV's take-off time[J]. Journal of Wuhan University of Technology (Transportation Science & Engin-

ering), 2024, 48(3): 435-440(in Chinese).

[18] 张洪海, 任真苹, 冯讴歌, 等. 城市低空物流无人机飞行计划预先调配[J]. 系统工程与电子技术, 2023, 45(9): 2802-2811.

Zhang H H, Ren Z P, Feng O G, et al. Logistics unmanned aerial vehicle flight plan pre-allocation in urban low-altitude airspace [J]. Systems Engineering and Electronics, 2023, 45(9): 2802-2811 (in Chinese).

[19] Zheng Y, Li D, Shen Z, et al. Optimized airspace structures and sequencing method for urban logistics droneport[J]. Aerospace, 2025, 12(3): 257.

[20] Bulusu V, Polishchuk V. A threshold based airspace capacity estimation method for UAS traffic management[C]//Proceedings of the Annual IEEE International Systems Conference. Piscataway: IEEE Press, 2017: 1-7.

[21] 陆传费. 排队论[M]. 北京: 北京邮电学院出版社, 1994.

Lu C L. Queueing theory[M]. Beijing: Beijing Institute of Posts and Telecommunications Press, 1994(in Chinese).

[22] Cui Q Q, Liu M, Huang X Y, et al. Coarse-to-fine visual autonomous unmanned aerial vehicle landing on a moving platform[J]. Biomimetic Intelligence and Robotics, 2023, 3(1): 100088.

[23] Xie H, Ji X H, Yin J N, et al. Standardization of safety separation for multi-category unmanned aerial vehicles in low-altitude airspace operations[J]. Applied Sciences, 2025, 15(13): 7501.

[24] 刘聪, 孟庆昆, 张宗卫. 小型四旋翼无人机悬停状态气动干扰与安全间隔研究[J]. 中国安全生产科学技术, 2022, 18(11): 177-183.

Liu C, Meng Q K, Zhang Z W. Research on aerodynamic interaction and safety separation of small quadcopter under hovering state[J]. Journal of Safety Science and Technology, 2022, 18(11): 177-183(in Chinese).

Capacity model and collaborative scheduling of droneports for drone regional logistics

JIANG Bo^{1, 2}, LI Die³, ZHENG Yuan^{4, *}, LI Chenglong¹, ZHANG Zhaoxuan³

(1. College of Flight Technology, Civil Aviation Flight University of China, Guanghan 618307, China;

2. Sichuan Provincial Engineering Research Center of Domestic Civil Aircraft Flight and Operation Support, Civil Aviation Flight University of China, Guanghan 618307, China;

3. College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan 618307, China;

4. School of Computer Science and Artificial Intelligence, Civil Aviation Flight University of China, Guanghan 618307, China)

Abstract: Droneport terminal airspace faces the critical issues of scarce airspace resources and insufficient operational efficiency. A series-coupled queuing model and a static-dynamic collaborative scheduling strategy are established to achieve accurate capacity assessment and effective regional collaborative scheduling, taking into account the lack of collaborative scheduling mechanisms for multiple droneports and the need for load balancing in regional logistics areas. First, a differentiated queuing model is constructed to characterize blockage effects in each operational link, breaking through the limitations of traditional independent analysis. Second, a non-preemptive priority strategy is introduced into the landing system to safeguard high-priority tasks. Finally, a "static flow pre-allocation + dynamic capacity-flow regulation" mechanism is designed to achieve multi-droneport collaboration through dual-path decision-making based on queuing theory. The static results show that the overall capacity assessment deviation of the proposed series-coupled queuing model is controlled within 6%, which is significantly superior to the traditional independent analysis model. Moreover, in the capacity assessment of the waiting layer airspace, it also significantly outperforms the conflict threshold-based model. Under the non-preemptive priority strategy, the average delay time of high-priority tasks consistently stays below 30 seconds. However, with the increase in their proportion of total tasks, the system's average delay rises significantly. Under surge traffic conditions, the cooperative scheduling approach lowers overall drone delay by 39.49% on average and increases regional load balance to 95.9%. The research results can provide theoretical support for low-altitude logistics network planning and real-time dispatch.

Keywords: drone; regional logistics; droneport; queueing theory; capacity evaluation; collaborative scheduling

* Corresponding author. E-mail: ranchozy@cafuc.edu.cn

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0122

基于运行时保证的 DAA 系统运行安全保障方法设计

董磊^{1,2}, 王琦^{1,3}, 陈曦^{1,2,*}, 刘嘉琛^{1,3}

(1. 中国民航大学 民航航空器适航审定技术重点实验室, 天津 300300; 2. 中国民航大学 科技创新研究院, 天津 300300;

3. 中国民航大学 安全科学与工程学院, 天津 300300)

摘 要: 针对无人驾驶航空器在城市空中交通 (UAM) 场景下的避让问题, 提出了一个基于运行时保证的探测与避让模块 (RTA-DAAM), 通过探测外部环境信息源判断航空器状态, 并提供控制指令和告警信息, 保证航空器在 UAM 场景下执行任务的高效性和安全性。对 RTA-DAAM 中的先进避让模块和备用避让模块的控制方法进行设计, 同时考虑探测、决策和告警功能协同配合的重要性, 定义航空器自身的净空边界、预防边界、障碍物的安全边界, 划分不同的航空器状态, 完成决策与告警逻辑的设计。仿真实验结果表明, RTA-DAAM 可使航空器在仅损失 3.40% 效率的基础上, 确保其始终存在期望分离距离, 以保障 DAA 系统运行安全。小型无人机的实飞结果表明, RTA 技术的引入可提升无人机的安全时间占比至 96.9%, 最大限度地确保了无人机的运行安全。

关 键 词: 运行时保证; 探测与避让系统; 无人驾驶航空器; 安全运行限制; 决策与告警

中图分类号: V243

文献标志码: A

文章编号: 1001-5965(2026)09-3023-16

“低空经济”作为近年来新兴的经济形态, 涵盖了无人机物流、通用航空等诸多前沿领域, 正处于发展的黄金时期, 中国正计划扩大低空经济的规模, 以确保在新兴低空技术赛道占据领先地位^[1]。追溯其源, 为了解决城市道路日益严重的拥堵问题, NASA 于 2018 年提出了城市空中交通 (urban air mobility, UAM) 的概念, 基于先进空中交通工具在城市中的高效运作, 缓解地面交通压力, 提高运输效率^[2]。无人驾驶航空器作为先进交通工具的核心成员之一, 具有高效、经济、环保等优点, 在为城市居民提供更加便携的通勤方式的同时, 也为城市规划提供了新的思路, 在未来有着广阔的发展空间^[3-4]。出于对航空器运行安全的考虑, 防止可控飞行撞地和空中碰撞等事故的发生, 良好的探测与避让 (detect and avoid, DAA) 能力是必不可少的, 其概念定义为“无人驾驶航空器监视跟踪保持良好距离并避免与其他航空器相撞的能力”^[5]。DAA 系统对外部状态进行

探测感知, 面对障碍物时提供机动避让指令, 同时提供预警警告, 作为防止航空器发生碰撞的关键系统之一, 在保证航空器安全方面发挥着巨大的作用^[6]。

在多系统的协同作用下, 无人驾驶航空器能够以一定的安全保障措施解决部分问题, 然而, 未来 UAM 场景下交通愈加复杂, DAA 系统应具备更好的性能以应对突发状况^[7]。伴随着近年来人工智能技术的发展与进步, 打造新一代智能航空成为了发展的必然趋势。航电系统智能化的应用研究正在受到重视与青睐, 工业和信息化部等部门于 2023 年发布的《绿色航空制造业发展纲要 (2023—2035 年)》指出, 未来应建成具有完整性、先进性、安全性的绿色航空制造体系, 以无人化、电动化、智能化为技术特征的新型通用航空设备实现商业化、规模化应用^[8]。因此, DAA 系统与人工智能技术的集成满足中国航空业未来发展前景, 能够为实现高效的空中交通运输任务提供一条可行的道路。

收稿日期: 2025-03-11; 录用日期: 2025-04-25; 网络出版时间: 2025-06-11 09:28

网络出版地址: link.cnki.net/urlid/11.2625.V.20250610.1954.003

基金项目: 中央高校基本科研业务费专项资金 (3122022QD07, 3122022044)

*通信作者. E-mail: chen_x@cauc.edu.cn

引用格式: 董磊, 王琦, 陈曦, 等. 基于运行时保证的 DAA 系统运行安全保障方法设计 [J]. 北京航空航天大学学报, 2026, 52 (9): 3023-3038.

Dong L, Wang Q, Chen X, et al. Design of operational security guarantee method for DAA system based on runtime assurance [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3023-3038 (in Chinese).

无人驾驶航空器 DAA 系统智能化水平的提升和通信导航监视技术的发展将有助于对城市低空空域更精细化的管理和使用,从而最终实现 UAM 的美好蓝图。然而,机器学习、自适应等智能算法的策略及工作过程被称作黑盒模型^[9-10],可解释性的缺失使其集成的系统会出现不安全、不可控、难理解等问题。同样的,航电系统的智能化虽然可以推动航空领域迈向新的层次,但现有民机系统安全性设计与评估方法难以保证应用启发式算法或人工智能方法的功能满足安全需求^[11]。随着未来城市低空空域环境不确定性的增加和 DAA 系统复杂程度的提升,现有方法难以保证 DAA 系统复杂组件在安全关键场景中不会发生失效或故障,无人驾驶航空器的空中碰撞风险可能会持续增加,其运行过程的安全性难以保证在较高的水平上。

因此,有必要引入安全风险减缓技术将 DAA 系统的剩余风险控制 在人们可接受的范围内^[12]。运行时保证(runtime assurance, RTA)作为一种保证系统安全运行的新兴技术,初步解决了一些复杂系统行为难预测、难解释、不安全、非预期结果的问题,在航空航天、智能汽车、机器人等领域的研究正趋于火热^[13-14]。Cofer 等^[15]针对支持学习组件难以采用传统方法认证的问题,引入 RTA 技术架构对飞机滑行的地面运动进行设计,满足了飞机滑行时的安全需求;He 等^[16]提出了结合高级统计分析框架和 R2U2 监视器的运行时监控架构,能够高效地完成飞机自主中心线跟踪任务;Ghori 等^[17]将 RTA 与综合模块化航电系统进行交互,以小型四旋翼无人机进行实际测试,通过故障注入的方式测试了实际情况下 RTA 技术带来的影响;Abate 等^[18]基于航天器非确定性动力学模型的基础上设计了 RTA 过滤器,通过对动力学可达集的实时计算,确保系统始终处于先验安全终端集 的范围内;Avram 等^[19]对基于 RTA 技术架构的四旋翼飞行器非线性自适应控制算法开展了分析研究,提出了四旋翼自适应控制架构,以完整性监测的方法增强了 RTA 技术架构的鲁棒性和故障敏感性;Chen 等^[20]提出了基于 RTA 的自动驾驶汽车控制框架,并在密集交通场景下验证了框架的有效性。RTA 技术在各领域具备一定的研究基础,但在 UAM 场景下,针对无人驾驶航空器的避让问题,相关的案例研究仍然较为匮乏。

综上所述,当前无人驾驶航空器的 DAA 系统因引入先进技术而难以落地使用且存在一定碰撞风险。基于此,本文提出了一种基于 RTA 的 DAA 系统运行安全保障方法,保证了无人驾驶航空器在 UAM 场景下能够兼顾效率与安全,也为复杂功能

难以应用于航空业的问题提供了一条可能的解决思路。具体地,提出了一个基于运行时保证的探测与避让模块(runtime assurance-detect and avoid module, RTA-DAAM),通过探测外部环境信息源确定系统的安全运行限制,控制航空器规避障碍物,并与地面站对接以便及时提供人工干预,为无人驾驶航空器的避让难题提供了新的安全保障思路;在 DAA 系统的决策处理单元融入 RTA 技术,设计了切换与告警逻辑,实现了从探测、决策到告警过程中信息处理与控制执行的协同性,能够为航空器在不牺牲效率的前提下提供运行安全保障。

1 基本概念

1.1 探测与避让系统

DAA 系统是为了协助无人机操作员在国家空域系统中安全操作无人机的职责而开发的,用于帮助操作员保证无人机与其他障碍物之间存在一定的安全距离^[21-22]。区别于传统的空域规划,无人驾驶航空器系统(unmanned aircraft system, UAS)考虑到空域划分及空管分类重叠严重等现象,DAA 系统所属的 UAS 运行环境如图 1 所示。图中:AGL 表示距地面高度,UTM 表示无人机交通管理。DAA 系统运行中,不仅需要处理静态楼宇等障碍物的避让规划,还需要考虑其他机型对本机的影响。因此,在 DAA 系统的决策过程中,应开展静态与动态障碍物结合的避让设计,以满足航空器在 UAM 场景中的安全需求。

DAA 系统架构如图 2 所示。通过多监视设备完成航空器自身及周围环境状态的探测,同时接收空中交通管理系统及地面站获取的障碍物状态信息,由 DAA 模块对目前任务规划的航迹信息开展分析,完成探测阶段。根据安全需求判断障碍物是否对本机产生影响,若不满足,则执行航迹规划,直至能够完成对障碍物的避让,完成决策阶段。告警阶段与决策阶段近似同步进行,出现风险时提供相应的告警方式,警示工作人员准备应急措施以应对突发状况^[23]。

1.2 运行时保证技术框架

2017 年,美国材料与试验学会(American Society of Testing Materials, ASTM)组织发布了 ASTM-F3269 标准文件^[24],提出的 RTA 基础框架如图 3 所示,该框架在航空领域的提出旨在为使用 RTA 技术实现复杂系统的认证提供指导^[25]。典型 RTA 技术框架由复杂功能、备用功能、输入管理器、安全监视器和 RTA 开关组成。输入管理器负责接收并处理信息;复杂功能指复杂度较高,安全难以进行验证,信

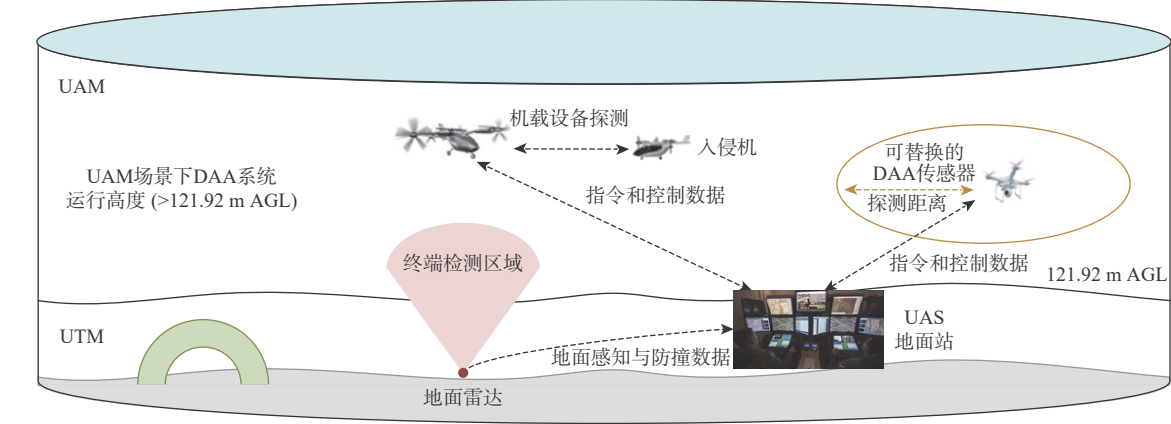


图 1 DAA 系统运行环境

Fig. 1 DAA system operating environment

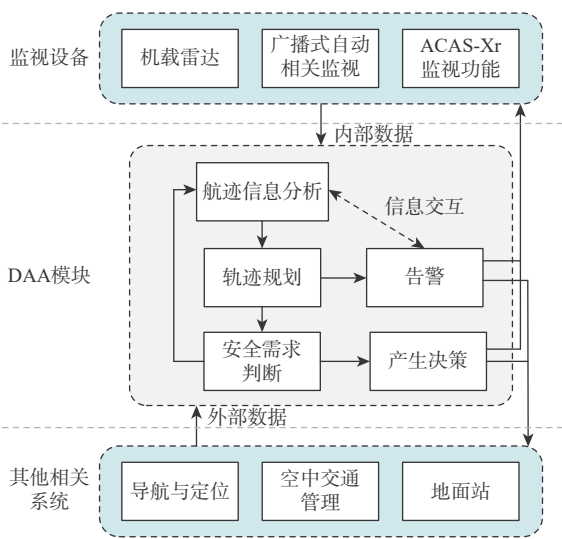


图 2 DAA 系统架构

Fig. 2 DAA system architecture

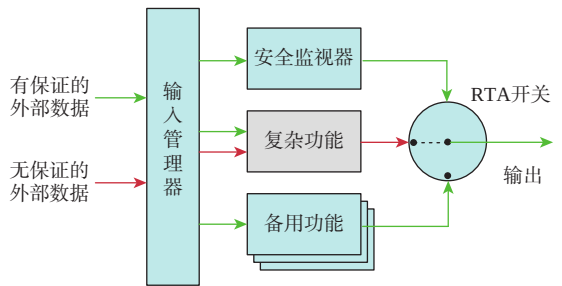


图 3 RTA 基础框架

Fig. 3 RTA foundation framework

程度较低的系统、设备、功能或算法; 备用功能指安全性和可靠性较高, 经验证的系统、设备、功能或算法; 由安全监视器监控系统状态并输出指令控制 RTA 开关, 执行复杂功能和备用功能之间的输出切换。

2 基于运行时保证的 DAA 模块

本文将 RTA 技术与 DAA 系统中的 DAA 主要

模块进行结合, 提出了一个运行安全保障模块 RTA-DAAM, 如图 4 所示。RTA-DAAM 通过对航空器状态和 2 种控制指令进行监控, 以权衡任务执行过程的效率和安全。先进避让模块的设计以性能为优先进行考虑, 迎合新一代智能航空发展, 引入新兴技术, 采用在线学习的方式完善优化自身指令, 但其学习的过程难以解释, 安全性、可靠性难以进行验证。备用避让模块的设计以安全为优先进行考虑, 能够在任务场景下保证其控制指令的稳定性, 并为航空器提供安全保障。在此前提下, 备用避让模块的性能和任务效率可能会低于先进避让模块。

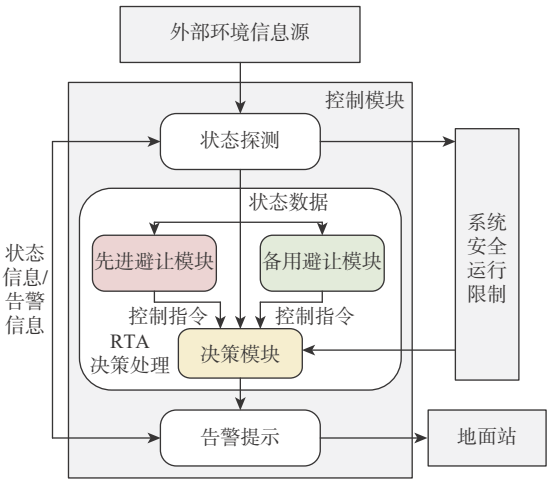


图 4 RTA-DAAM 架构

Fig. 4 RTA-DAAM architecture

- 由图 4 可得, RTA-DAAM 由以下部分组成:
- 1) 外部环境信息源。指无人驾驶航空器在执行任务过程中的环境信息, 包括可视作静态障碍物的楼宇、除本机外的可视为动态障碍物的入侵机、电子地理围栏、本机传感设备能够探测到的距离、角度、动态轨迹等数据。
 - 2) 控制模块。该部分融合了 DAA 模块的主要

功能,用于提供控制指令和告警信息。其中,状态探测单元用于接收本机传感器从外部环境信息源取得的环境数据,结合本机的状态完成探测工作;RTA 决策处理单元用于判断航空器安全水平,并执行先进避让模块或备用避让模块的控制指令输出;告警提示单元根据本机状态和控制指令输出提供相关告警信息。

3) 系统安全运行限制。DAA 系统保障航空器运行安全应满足相应的限制条件,完善安全限制条件并作为基础理念提供给决策模块,使无人驾驶航空器在复杂场景下时刻具备保护本机安全的能力。

4) 地面站。实时接收无人驾驶航空器自身状态和告警信息,若出现突发状况,在必要时进行人工干预,为航空器任务安全提供协助及安全保障。

根据对效率和安全的不同需求确定先进避让模块和备用避让模块的控制方法,航空器在执行任务过程中,由状态探测单元获取外部环境及本机状态信息,基于不同情景分析需满足的系统安全运行限制,处理并提供给决策模块;通常情况下,先进避让模块工作并输出控制指令,以满足航空器对任务效率的需求,在实时监控的过程中,若决策模块根据已有信息判断得出本机存在一定碰撞风险,则切换到备用避让模块输出控制指令,同时将相应告警信息发送至地面站,短暂地降低效率来保证本机自身安全;在决策模块判断本机脱离风险状况后,切换回先进避让模块提供控制指令,继续执行任务。

除此之外,RTA-DAAM 的设计满足以下功能:

1) 备用避让模块应采用易解释、易验证的控制方法,避免备用控制指令安全性、可靠性难以确定的情况存在。

2) 决策模块应在切换 2 种模块的指令输出中加入时间裕度,以防止系统在 2 种控制指令之间短时间内反复切换,导致危险情况的发生。

3) 告警提示单元可基于航空器自身状态对风险程度进行分级处理,以提供更加准确的告警信息。

4) 地面站实时接收无人驾驶航空器的自身状态和场景因素,可自行判断并介入航空器的控制指令切换,在必要时可直接接管航空器的控制权,成为航空器运行安全的最后一层保障。

3 基于 RTA 的航空器自主避让方法

对应 RTA-DAAM 控制模块开展无人驾驶航空器的自主避让方法研究,建立简化的无人驾驶航空器的运动学模型如下:

$$\begin{cases} \dot{x} = V \cos \alpha \cos \beta \\ \dot{y} = V \cos \alpha \sin \beta \\ \dot{z} = V \sin \alpha \\ \dot{V} = a_x - g \sin \alpha \\ \dot{\alpha} = \frac{a_z + g - g \cos \alpha}{V} \\ \dot{\beta} = \frac{a_y}{V \cos \alpha} \end{cases} \quad (1)$$

式中: V 为航空器的飞行速度; α 和 β 分别为航空器的俯仰角和偏航角; g 为重力加速度; a_x 、 a_y 、 a_z 分别为航空器沿 x 、 y 、 z 轴的加速度。

为了更直观地判断航空器的避让效果,搭建障碍物的碰撞模型,使用标准凸面体包络对楼宇、入侵机或其他风险因素进行等效处理,楼宇等静态障碍物可视作平行六面体或圆柱体,入侵机、空中障碍物视作球体。障碍物等效包络方程如下:

$$\Gamma(\eta) = \left(\frac{x-x_0}{A+R_0} \right)^{2D} + \left(\frac{y-y_0}{B+R_0} \right)^{2E} + \left(\frac{z-z_0}{C+R_0} \right)^{2F} \quad (2)$$

式中: 常数 A 、 B 、 C 决定障碍物的大小,常数 D 、 E 、 F 决定障碍物的形状; R_0 为航空器自身半径; $\eta_0 = [x_0, y_0, z_0]$ 为障碍物的中心位置。

因此,航空器碰撞判据可表示如下:

1) $\Gamma(\eta) > 1$ 时,航空器与障碍物之间无接触,无碰撞事件发生。

2) $\Gamma(\eta) \leq 1$ 时,航空器与障碍物接触,发生碰撞事件,对航空器自身安全产生威胁。

3.1 先进避让模块设计

3.1.1 基于 IFDS 的航空器自主避让方法

扰动流体动态系统(interfered fluid dynamical system, IFDS)算法被提出用于解决无人机复杂地形下的三维路径规划问题^[26]。该算法的原理源于自然界的水流避石现象,通过模拟水流自然流向方式,判断障碍物位置、速度等信息,从而规划出一条平滑的避障轨迹^[27]。

在初始坐标系 O 下,航空器位置为 $\eta = (x, y, z)^T$,目标点位置为 $\eta_d = (x_d, y_d, z_d)^T$,飞行速度常数为 V 的情况下,航空器初始速度矢量 $u(\eta)$ 和航空器到目标位置的距离 $d(\eta)$ 表示为

$$u(\eta) = \begin{bmatrix} \frac{V(x-x_d)}{d(\eta)} & \frac{V(y-y_d)}{d(\eta)} & \frac{V(z-z_d)}{d(\eta)} \end{bmatrix}^T \quad (3)$$

$$d(\eta) = \sqrt{(x-x_d)^2 + (y-y_d)^2 + (z-z_d)^2} \quad (4)$$

对于环境中存在的 N 个障碍物,所有障碍物对航空器的初始速度矢量的扰动影响使用总扰动矩阵 $\overline{M}(\eta)$ 表示为

$$\overline{M}(\eta) = \sum_{n=1}^N \omega_n(\eta) M_n(\eta) \quad (5)$$

式中: $\omega_n(\eta)$ 为第 n 个障碍物的扰动权重系数, 航空器与障碍物之间距离越大, 则扰动权重系数越小, 反之亦然。

$\omega_n(\eta)$ 与障碍物 n 的关系如下:

$$\omega_n(\eta) = \begin{cases} 1 & n = 1 \\ \prod_{i=1, i \neq n}^N \frac{\Gamma_i(\eta) - 1}{(\Gamma_i(\eta) - 1) + (\Gamma_n(\eta) - 1)} & n \neq 1 \end{cases} \quad (6)$$

对每个障碍物的扰动矩阵 $\mathbf{M}_n(\eta)$ 进行求解, 定义 $\mathbf{t}_n(\eta)$ 为障碍物的径向法向量, 同时构建垂直于径向法向量 $\mathbf{t}_n(\eta)$ 且互相垂直的 2 个向量 $\mathbf{h}_{n,1}(\eta)$ 和 $\mathbf{h}_{n,2}(\eta)$, 表示为

$$\mathbf{h}_{n,1}(\eta) = \begin{bmatrix} \frac{\partial \Gamma_n(\eta)}{\partial y} & -\frac{\partial \Gamma_n(\eta)}{\partial x} & 0 \end{bmatrix}^T \quad (7)$$

$$\mathbf{h}_{n,2}(\eta) = \begin{bmatrix} \frac{\partial \Gamma_n(\eta)}{\partial x} \cdot \frac{\partial \Gamma_n(\eta)}{\partial z} \\ \frac{\partial \Gamma_n(\eta)}{\partial y} \cdot \frac{\partial \Gamma_n(\eta)}{\partial z} \\ -\left(\frac{\partial \Gamma_n(\eta)}{\partial x}\right)^2 - \left(\frac{\partial \Gamma_n(\eta)}{\partial y}\right)^2 \end{bmatrix}^T \quad (8)$$

由于 $\mathbf{t}_n(\eta)$ 、 $\mathbf{h}_{n,1}(\eta)$ 和 $\mathbf{h}_{n,2}(\eta)$ 两两正交, $\mathbf{h}_{n,1}(\eta)$ 和 $\mathbf{h}_{n,2}(\eta)$ 组成的平面为 H , 可将 3 个向量组成的空间看作一个新的坐标系 O' , H 上的所有单位向量可表示为

$$\mathbf{h}'_n(\eta) = [\cos \theta_n \quad \sin \theta_n \quad 0]^T \quad (9)$$

式中: θ_n 为方向系数, 代表每一切向量与坐标系 O' 中 x' 轴的夹角, 影响切线扰动的方向。

使用坐标变换矩阵 $\mathbf{R}$ 将 $\mathbf{h}'_n(\eta)$ 变换回初始坐标系 O 下:

$$\mathbf{h}_n(\eta) = \mathbf{R} \mathbf{h}'_n(\eta) \quad (10)$$

则单个障碍物的扰动矩阵 $\mathbf{M}_n(\eta)$ 表示为

$$\mathbf{M}_n(\eta) = \mathbf{I} - \frac{\mathbf{t}_n(\eta) \mathbf{t}_n(\eta)^T}{\left| \Gamma_n(\eta) \right|^{\frac{1}{\rho n}} \mathbf{t}_n(\eta)^T \mathbf{t}_n(\eta)} + \frac{\mathbf{h}_n(\eta) \mathbf{t}_n(\eta)^T}{\left| \Gamma_n(\eta) \right|^{\frac{1}{\sigma n}} \left| \mathbf{h}_n(\eta) \right| \left| \mathbf{t}_n(\eta) \right|} \quad (11)$$

式中: ρ 和 σ 分别为障碍物的排斥系数和切向系数, 2 个系数越大, 则航空器做出规避动作的时间越早, 需要合理调节以权衡任务路径长度与飞行安全的重要性。

采用总扰动矩阵 $\bar{\mathbf{M}}(\eta)$ 对 $\mathbf{u}(\eta)$ 进行修正获得扰动速度 $\bar{\mathbf{u}}(\eta)$, 以执行规避动态障碍物的指令:

$$\bar{\mathbf{u}}(\eta) = \bar{\mathbf{M}}(\eta) (\mathbf{u}(\eta) - \mathbf{v}(\eta)) + \mathbf{v}(\eta) \quad (12)$$

式中: $\mathbf{v}(\eta)$ 为动态障碍物对航空器的速度威胁。在

第 n 个动态障碍物的速度影响下, $\mathbf{v}(\eta)$ 可表示为

$$\mathbf{v}(\eta) = \sum_{n=1}^N \omega_n(\eta) \mathbf{e}^{1-\Gamma_n(\eta)} \mathbf{v}_n \quad (13)$$

此时, 对位置进行积分获得从当前航路点 η_i 到下一时刻航路点 η_{i+1} 的位置关系:

$$\eta_{i+1} = \eta_i + \bar{\mathbf{u}}(\eta) \Delta T \quad (14)$$

式中: ΔT 为时间变化量。

3.1.2 基于 TD3 算法的系数优化

由于具有与人类相似的奖励与惩罚机制, 可通过与环境的交互不断学习从而实现最优决策, 深度强化学习在近年来被广泛用于多领域研究, 其学习过程通常采用马尔可夫决策过程 (Markov decision process, MDP) 表示^[28-29]。MDP 为五元组, 表示为 $\{S, A, P, R_{\text{MDP}}, \gamma\}$, 其中, S 为环境状态的集合即状态空间, A 为 Agent 根据环境和策略做出的决策集合, 即动作空间, P 为状态转移模型, $P_{ss'}^a = P[S_{t+1} = s' | S_t = s, A_t = a]$ 表示 Agent 从状态 s 下采取动作 a 后到达状态 s' 的概率, R_{MDP} 为奖励函数, $r = R_{\text{MDP}}[S_t = s, A_t = a]$ 表示 Agent 采取动作后环境对 Agent 的反馈, 即得到的奖励, $\gamma \in (0, 1]$ 为折扣因子, 用以权衡当前奖励与后续奖励之间的权重关系。

综合考虑无人驾驶航空器在城市复杂场景下的交通流量大、航路交互等问题, 应快速且稳定地执行路径规划决策, 为了弥补 IFDS 算法在复杂场景下应对实时碰撞风险的局限性, 将 IFDS 相应系数作为动作选取, 开展系数优化研究。在深度强化学习的发展过程中, SAC、DDPG、TD3 等算法具有各自的特点, 在处理不同问题时有着不同程度的表现。文献[30]采用 DDPG 算法对系数进行优化, 适用于高维连续动作空间, 在处理路径规划、自动驾驶、机器人控制等问题方面有着广泛的应用研究, 但其仍存在估计值函数大于真实值函数的问题。

为了解决过估计问题, TD3 算法基于 AC 算法网络框架^[31], 在 DDPG 的基础上进行了改良, 采用 2 套网络估算不同的 Q 值, Critic 网络 Q_{θ_k} 在状态 s 和动作 a 情况下计算 Q 值 $Q_{\theta_k}(s, a)$, Critic 目标网络 $Q_{\theta'_k}$ 用于计算下一状态 s' 和目标动作 a' 情况下的目标 Q 值 $Q_{\theta'_k}(s', a')$ 。 ϕ 和 θ_k 分别为 Actor 网络和 Critic 网络的参数, ϕ' 和 θ'_k 分别为 Actor 目标网络和 Critic 目标网络的参数。采用较小的目标 Q 值计算目标值函数 f , 从而抑制过高估计的问题发生。

$$f = r + \gamma \min_{k=1,2} Q_{\theta'_k}(s', a') \quad (15)$$

通过最小化损失函数更新 2 个 Critic 网络参数, 可表示为

$$\theta_k \leftarrow \arg \min_{\theta_k} B_1^{-1} \sum (Q_{\theta_k}(s,a) - f)^2 \Big|_{(k=1,2)} \quad (16)$$

在得到准确的值函数估计后,对策略进行延迟更新,避免 Actor 更新时预期取得而实际未取得最优值的情况发生。为使值估计更加准确,将服从截断正态分布的噪声 ε 加入目标策略,则目标动作 a' 可表示为

$$a' = \pi_{\phi'}(s') + \varepsilon \quad (17)$$

式中: $\pi_{\phi'}$ 表示网络参数为 ϕ' 的 Actor 目标网络。

采用的网络结构中, Actor 网络由输入层、全连接层、激活函数层和输出层组成,输入维度由航向、位置、速度等状态向量组成,输出为排斥系数、切向系数等动作向量, Critic 网络中的每个 Q 网络由输入层、全连接层、激活函数层、添加层和输出层组成,输入维度由状态向量和动作向量组成,输出维度则对应 Q 值。

基于 TD3 算法的系数优化方法伪代码如下:

输入: 航空器航向、位置、速度,障碍物形状、速度,航空器与障碍物之间距离。

输出: 排斥系数、切向系数、方向系数。

随机化 ρ 、 σ 和 θ_n 。

初始化 Actor 网络、Critic 网络及网络参数、经验回放池 D 。

初始化 Actor 目标网络、Critic 目标网络及网络参数。

for $l=1, L$ do

更新环境信息

for $u_T=1, U_T$ do

获取外部环境及航空器状态信息

执行动作 a , 获得奖励 r 和状态 s'

从 D 中抽样 s' , 通过 Actor 目标网络获得目标动作 a' , 并通过 Critic 目标网络计算并选用较小的 Q 值, 根据式(15)得到 f

利用 f 更新 Critic 网络参数(式(16))

使用策略梯度法将网络参数按 Q 值增大的方向进行更新:

$$\nabla_{\phi} J(\phi) = B_1^{-1} \sum \nabla_{\phi} Q_{\theta_1}(s,a) \Big|_{a=\pi_{\phi}(s)} \nabla_{\phi} \pi_{\phi}(s)$$

根据延迟更新策略对目标网络进行更新, 表示为

$$\begin{aligned} \theta'_k &\leftarrow \tau \theta_k + (1 - \tau) \theta'_k \Big|_{(k=1,2)} \\ \phi' &= \tau \phi + (1 - \tau) \phi' \\ \tau &\leq 1 \end{aligned}$$

end for

end for

3.1.3 基于 CIFDS 的航空器自主避让方法

航空器在低空域下面临的危险状况趋向复杂多变,为减少基于 IFDS 的航空器路径规划时间,提高避障效率,同时增加规划路径的平滑度,将 TD3 算法与 IFDS 算法进行结合,形成复杂扰动流体动态系统 (complex interfered fluid dynamical system, CIFDS) 算法,以强化学习训练的方式对 IFDS 算法中的障碍物排斥系数、切向系数和方向系数进行调节,在不同环境下做出高效且合理的避障逻辑决策。CIFDS 算法架构如图 5 所示。

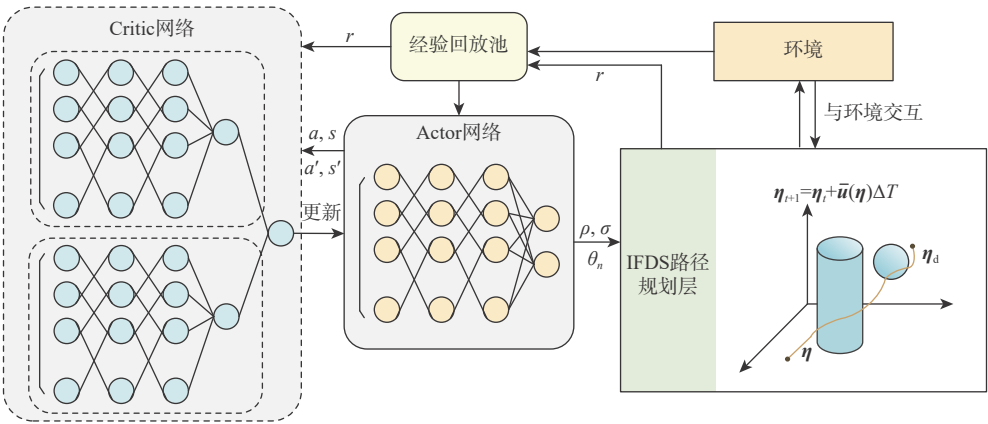


图 5 CIFDS 算法架构
Fig. 5 CIFDS algorithm architecture

航空器被视为智能体,所处的状态空间包括航空器与障碍物的位置关系、航空器与目标点的位置关系和航空器靠近障碍物的速度,为 ρ 、 σ 和 θ_n 赋予合适的常值范围,执行动作后存在以下情况:航空器到下一状态的路径中接触到障碍物,

给予惩罚;航空器到下一状态的路径中不存在障碍物,对路径长度即任务效率进行优化;航空器到下一状态的路径中未接触障碍物,但距离障碍物过近,则给予另一类惩罚。其对应的奖励机制分别表示为

$$r_1 = \|\eta_{t+1} - \eta_o\| - \lambda_1 \quad \|\eta_{t+1} - \eta_o\| \leq d \quad (18)$$

$$r_2 = \|\eta_{t+1} - \eta_o\| + \lambda_2 \quad \|\eta_{t+1} - \eta_o\| > d \quad (19)$$

$$r_3 = \|\eta_{t+1} - \eta_o\| - \lambda_1 + \lambda_3 \quad \|\eta_{t+1} - \eta_o\| < d + \lambda_3 \quad (20)$$

式中: η_{t+1} 为航空器下一状态的路径点; η_o 为距离最近的障碍物中点; d 为障碍物中点到障碍物表面的距离幅值; λ_1 、 λ_2 、 λ_3 为可设定的正数权重。 $\lambda_1 > d$, λ_3 为设定的障碍物外的安全范围半径, 为保证第 3 种情况对航空器做出惩罚, 故使得 $\lambda_3 < \lambda_1$ 。因此, 奖励值函数可表示为

$$r = R_{MDP}[s, a] = r_1 + r_2 + r_3 \quad (21)$$

得到奖励值后, 将其与状态、动作等数据存储在经验回放池中, 将状态、动作输入到 Critic 网络计算并筛选得到 Q 值, 计算目标值函数并更新 Critic 网络参数, 使用策略梯度对 Actor 网络进行更新, 结合 IFDS 的扰动矩阵计算扰动速度, 分析得到下一时刻的航路点, 通过判断是否到达规定的目标点选择继续循环或完成任务。CIFDS 运行流程如图 6 所示。

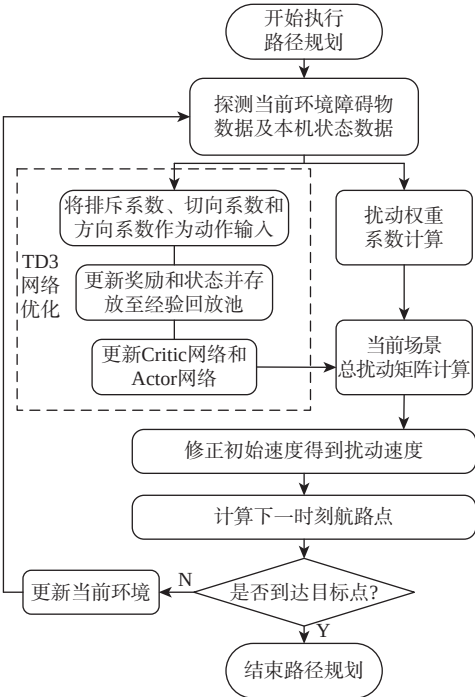


图 6 CIFDS 运行流程

Fig. 6 CIFDS operation flowchart

3.2 备用避让模块设计

备用避让模块的控制方法应具备高水平的安全保障能力, 在 RTA-DAAM 中发挥安全保险的作用, 同时考虑到备用避让方法在可解释、可靠性等方面应满足公众的信任, 因此, 参考无人机主流规划方法 EGO-Planner 作为备用避让方法以输出安全的控制指令^[32]。由于不需要建立栅格距离场, 路径

规划时间缩短, 能够提供更大的反应空间, 同时仅需在每次规划中迭代判断路径是否发生碰撞, 未发生则继续进行优化, 发生则施加反向推力, 能够得到符合航空器运动的平滑且安全的避让路径, 因此, 在安全可靠和解释方面符合备用避让模块的设计理念, 避免风险事件的发生。

EGO-Planner 分析本机与障碍物或入侵机的位置关系, 在此时间步长上, 将本机与障碍物的位置关系作为实时跟踪的路径参考, 对基本的可行路径进行双向的调整以实现与障碍物的分离, 找出最优指令执行并传播至下一时间步长, 规划高安全保障的避让路径。

如图 7 所示, 对于轨迹上接近障碍物的可控点寻求可行路径, 并由路径与障碍物内部接触的可控点 U 向障碍物表面点 K 建立单位向量 w , 则定义航空器与障碍物的距离 D_K 为

$$D_K = (U - K)^T w \quad (22)$$

对于需要远离障碍物的距离 W 施加代价项 $(D_K - W)^2$ 推动 U 沿 w 运动, 进而避开障碍物。

避让路径的轨迹优化过程中, 使用平滑度、碰撞度和动力学可行度来优化轨迹, 目标函数表达式为
$$\min_v J_t = \delta_s J_s + \delta_c J_c + \delta_d J_d \quad (23)$$
 式中: δ 为惩罚项的权重; J_s 、 J_c 和 J_d 分别为平滑度惩罚、碰撞度惩罚和动力学可行度惩罚。

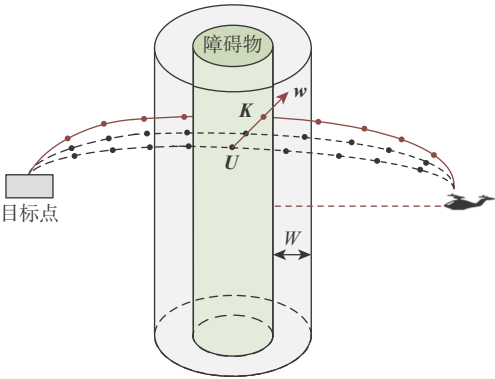


图 7 EGO-Planner 规划示意图

Fig. 7 EGO-Planner planning schematic

3.3 决策与告警逻辑设计

在 RTA-DAAM 中, 决策模块接收航空器状态信息, 并能够控制正向开关和反向开关以分别确保安全性和减少效率损失。在决策模块的作用下, 先进避让方法的设计可以追求性能和效率, 而免于受困解释方面的难题之中, 同时能够使备用避让方法的设计简单化, 减轻其在路径规划性能上的负担, 更加注重航空器的安全问题, 保证航空器满足系统安全运行限制。决策模块架构设计如图 8 所示。

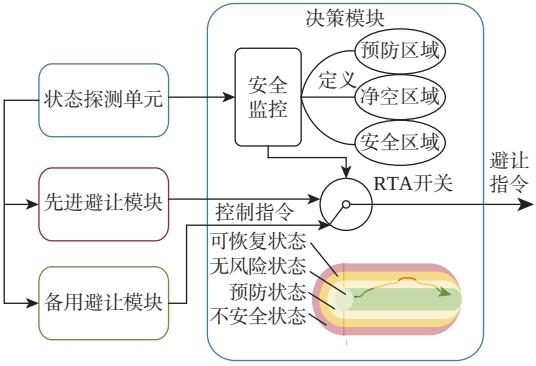


图 8 决策模块运行示意图

Fig. 8 Schematic diagram of operation of decision-making module

为尽可能地维持 2 种避让模块之间的平滑切换,应保证 2 种控制方法的输出指令接口一致,任务状态同步,而使得切换后能够立刻掌握当前环境信息及航空器自身运动状态信息,并以此作为输入,避免突加指令导致航空器稳定性降低,若系统切换过程存在响应时间,航空器可短暂悬停以确保自身安全。

考虑到 DAA 系统中告警功能作为保护无人驾驶航空器安全的重要手段,其提供的信息在 UAS 中需要清楚且能够获得地面站支持与反馈,因此,有必要将基于 RTA 的避让决策逻辑与告警逻辑进

行融合,有助于提高无人驾驶航空器在复杂运行场景下的安全性。

为了保证 RTA-DAAM 提供的告警信息清晰直观,且便于地面站工作人员选择是否介入航空器的控制方法,定义告警等级如表 1 所示。

表 1 告警等级
Table 1 Alarm levels

告警等级	告警名称	采取措施
0	无告警	地面站可选择性干预
I	预防告警	地面站可选择性干预
II	纠正告警	地面站可选择性干预
III	危险告警	地面站干预

如图 9 所示,参考 NASA 提出的无人机探测与避让告警逻辑 (detect and avoid alerting logic for unmanned systems, DAIDALUS)^[5],考虑航空器的运行环境及自身运动特性等因素,定义无人驾驶航空器自身在水平面上的预防边界和净空边界,同时定义障碍物或入侵机的安全边界,该安全边界与障碍物的距离可看作期望分离距离,并作为 DAA 系统运行安全限制用以判断本机与障碍物或入侵机是否存在碰撞风险,将航空器状态进行划分,无风险状态、预防状态和可恢复状态作为安全状态,存在碰撞风险时航空器处于不安全状态。

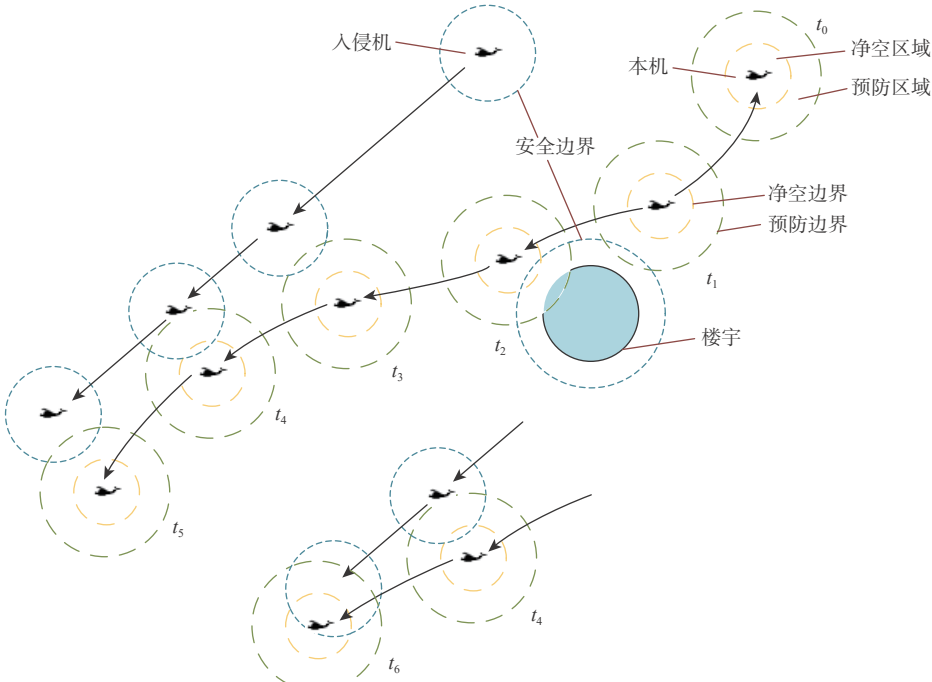


图 9 基于 RTA 的决策与告警逻辑示意图

Fig. 9 Schematic diagram of RTA-based decision-making and alerting logic

同时,各边界的划分与设定依据航空器的机动能力及运行空域密度,运行速度较大的航空器应存在更大的期望分离距离,繁忙的空域需要按照实际

情况对各个边界进行适当调整。决策与告警模块的切换时间裕度的设定应综合考虑系统特性、物理约束和安全需求,其取值应包含最大系统响应时间

和安全余量时间,用以避免系统在短时间内反复切换而引发的不安全行为。

因此,RTA-DAAM 决策与告警逻辑如下:

1) L1: 在 t_0 时刻下,本机的净空边界和预防边界都与障碍物安全边界不存在接触,此时本机不存在碰撞风险,处于无风险状态,由先进避让模块输出控制指令,告警模块不输出告警信息。

2) L2: t_1 和 t_3 时刻下,本机的预防边界与障碍物安全边界存在接触,本机进入预防状态,安全监控对障碍物状态信息进行实时跟踪,可继续由先进避让模块输出控制指令,告警模块输出预防告警,目的在于将信息提供给 UAS,并警示地面站工作人员提供一定的实时状态与路径监视。

3) L3: 在 t_2 和 t_4 时刻下,本机的净空边界与障碍物安全边界接触,代表航空器在当前路径规划上可能存在一定的碰撞风险,本机进入可恢复状态,告警模块提供纠正告警,并由安全监控单元控制 RTA 开关进行正向切换,采用备用避让模块输出避让指令,短暂时间内将无人驾驶航空器的任务执行侧重点由效率变为安全。

4) L4: t_5 时刻下,本机的净空边界脱离障碍物的安全边界,即航空器的碰撞风险降低,本机进入预防状态,告警模块的输出为预防告警,此时由安全监控单元进行判断,若在该时刻前一定的时间裕度内无控制指令的切换,则控制 RTA 开关进行反向切换,由先进避让模块提供控制指令,航空器任务执行的侧重点变回效率优先;若在该时刻前一定时间裕度内已经进行过控制指令的切换,则经过该时间裕度再次进行航空器本机与环境的状态判断。

5) L5: 在空域密集度高或突发状况下,2 种避让模块出现故障或难以有效降低航空器碰撞风险,如 t_6 时刻所示,航空器自身接触障碍物边界,航空器与障碍物之间距离即将小于期望分离距离,存在碰撞风险,则本机进入不安全状态,告警模块输出危险告警并发送至地面站,由地面工作人员干预航空器的控制工作,提供人工引导,使航空器悬停或寻找其他安全路线,直至航空器脱离障碍物安全边界,以此最大程度保证航空器的安全,防止碰撞事件的发生。同时,若发生指令冲突等状况,人工干预作为底层保护逻辑可直接输出系统避让指令。

4 实验验证与分析

本节通过实验的方式对 RTA-DAAM 提供的避让控制方法和决策告警逻辑进行验证,测试避让方法的有效性和 RTA 技术对 DAA 系统的安全保障能力。

4.1 仿真实验平台设置

采用 unity3D 作为仿真平台,基于 ML-Agents 库训练和调试先进避让方法,降低仿真环境与现实场景的差异。模拟 UAM 低空环境下,无人驾驶航空器执行应急消防任务,前往目标点完成灭火协助工作。因此,在该场景下,航空器具有任务执行效率的较高要求,同时要保证自身的安全运行,高效且安全地完成路径规划,以此为例开展仿真实验验证。

如图 10 所示,将仿真场景建立为正方形区域,定义该区域的边界为电子围栏,用于限制航空器的运行范围,其中,障碍物包含圆柱形和平行六面体的静态楼宇,以及可视为动态入侵机的球体,在避免入侵机模型与楼宇模型碰撞的基础上,将入侵机的路径进行随机化处理。同时,在电子围栏内定义一定距离的初始区域,使航空器的初始点与目标点在区域内随机初始化。

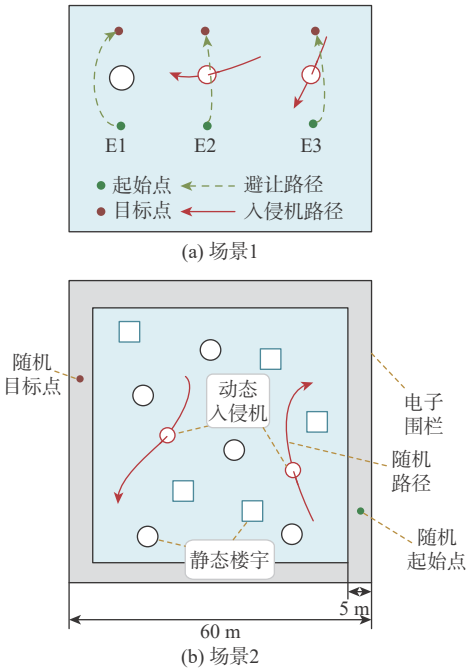


图 10 仿真场景设置
Fig. 10 Simulation scene setting

根据 RTA-DAAM 的切换逻辑规则,规定障碍物表面到安全边界的距离为障碍物安全边界宽度,也作为航空器的期望分离距离,由于距离参数的设定会对结果产生一定影响,结合航空器的运行速度及障碍物的密集程度,合理设定本机的净空边界半径、预防边界半径和障碍物安全边界宽度。强化学习训练参数与环境边界参数定义如表 2 所示。同时,为保证切换时间裕度的有效性,使切换时间裕度扩展至 1.5 s。航空器通过双目相机传感器获取环境信息,探测本机与障碍物的相对距离,以完成 RTA-DAAM 的探测功能。

表 2 参数设置
Table 2 Parameter settings

参数	数值
本机速度/(m·s ⁻¹)	0.8
入侵机速度/(m·s ⁻¹)	0.8
障碍物安全边界宽度/m	2.5
本机净空边界半径/m	1
本机预防边界半径/m	1.5
Actor学习率	2×10 ⁻⁴
Critic学习率	1×10 ⁻³
训练轮数/回合数	5 000
批次大小/个	256
经验池容量/个	1×10 ⁶
折扣因子	0.99
噪声截断值	0.5
策略噪声	0.2

4.2 先进避让方法性能测试

根据先进避让方法设计及参数设置开展 CIFDS 算法训练, 训练过程的奖励函数情况如图 11 所示。CIFDS 算法训练过程早期存在大幅波动, 表明智能体在随机探索环境中频繁触发碰撞惩罚, 在 1 500 回合数左右基本实现了收敛, 策略达到了局部最优。

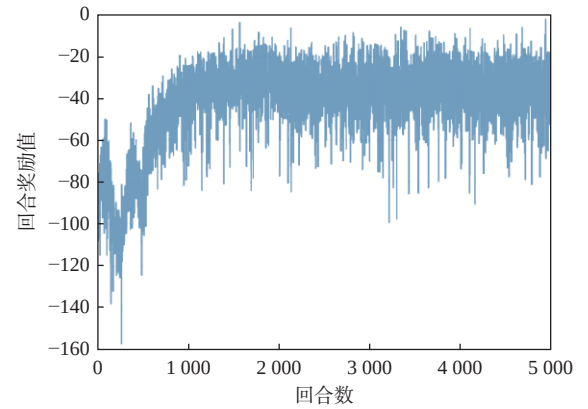


图 11 回合奖励值变化
Fig. 11 Episode reward changes

通过对 CIFDS 算法的路径规划效率进行验证, 以证明该算法作为先进避让方法的合理性。在场景 1 中, 对 CIFDS 与 IFDS 算法规划的路径进行对比, 在面对圆柱形静态障碍物、横向运动的球形障碍物及对向运动的球形障碍物 3 种情况下, CIFDS 与 IFDS 算法的最大飞行角度变化与路径长度对比如表 3 所示。

可以看出, 在规避静态楼宇时, CIFDS 和 IFDS 算法规划的避让路径中, 航空器运行的最大航迹倾角和避让路径长度存在较小差异, 而在规避动态入侵机的情况下, CIFDS 规划出航空器运行路线的最大航迹倾角小于 IFDS, 能够保证航空器执行任务

表 3 CIFDS 与 IFDS 算法路径规划对比
Table 3 Comparison of path planning under CIFDS and IFDS algorithms

障碍环境	最大飞行角度变化/(°)		路径长度/m	
	CIFDS	IFDS	CIFDS	IFDS
E1	26.8	27.2	22.1	23.4
E2	28.5	32.4	23.7	26.6
E3	27.3	30.9	22.9	25.5

时的飞行姿态具有较高的稳定性, 满足安全运行限制, 同时, CIFDS 规划的避让路径长度小于 IFDS 规划的避让路径长度, 使得航空器能够在更短的时间内完成对障碍物的避让, 具有更高的效率, 满足先进避让方法的设计理念与需求, 以此验证了 CIFDS 作为先进避让方法的合理性。

4.3 RTA-DAAM 效果测试与验证

本节基于场景 2 的仿真环境开展先进避让方法、备用避让方法及 RTA-DAAM 避让方法的任务路径对比实验, 由于 IFDS 算法在复杂环境下难以安全地完成任务, 稳定性和效率不及上述 3 种避让方法, 不将其考虑进对比实验中。

航空器在 3 种避让方法执行任务下, 从初始点起航空器与目标点之间的距离变化如图 12 所示。

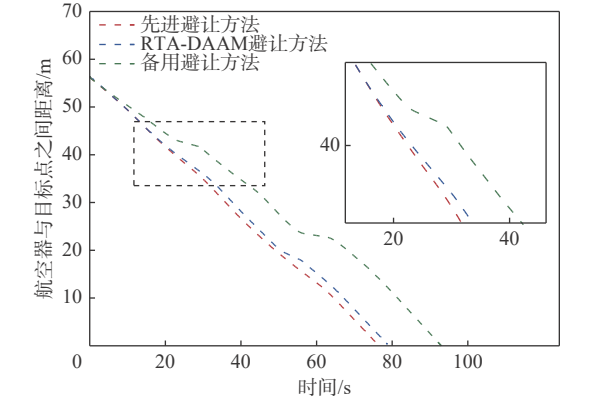


图 12 航空器与目标点之间距离变化
Fig. 12 Distance variation between aircraft and target point

如图 12 所示, 在同一起始点和目标点的任务执行过程中, 先进避让方法展现了更高的效率, 能够更快地到达目标点。相比之下, 备用避让方法在执行效率上明显落后, 任务执行效率低于其他 2 种方法。RTA-DAAM 避让方法在任务过程初期与先进避让方法采取的策略相似, 但随后在任务过程中进行了方法切换, 导致其后续的变化曲线与先进避让方法产生了差异。由表 4 可知, 备用避让方法在任务执行时间和路径长度上相较于先进避让方法损失了 21.57% 的效率, 而 RTA-DAAM 避让方法则仅损失了 3.40% 的效率。

根据前述定义的航空器状态对 3 种避让方法

表 4 避让方法效率对比

Table 4 Comparison of efficiency of avoidance methods

避让方法	任务执行时间/s	路径长度/m
先进避让方法	76.5	61.2
RTA-DAAM避让方法	79.1	63.3
备用避让方法	93.0	74.4

进行安全分析,如图 13 所示。先进避让方法在无风险状态上的占比相对较低,且在任务执行过程中,航空器有进入不安全状态的风险,存在一定的碰撞可能性。相比之下,备用避让方法在执行过程中虽然可恢复状态的占比较少,但无风险状态和预防状态的占比较高,且完全避免了进入不安全状态,确保了任务的安全性。尽管 RTA-DAAM 避让方法完成任务所需的时间相对较长,导致其在无风险状态和预防状态占总运行时间的比例上略低于先进避让方法,但同样能够避免进入不安全状态,从而有效规避了潜在的碰撞风险。这表明在权衡任务执行时间和安全性方面,RTA-DAAM 避让方法展现出了良好的平衡性。

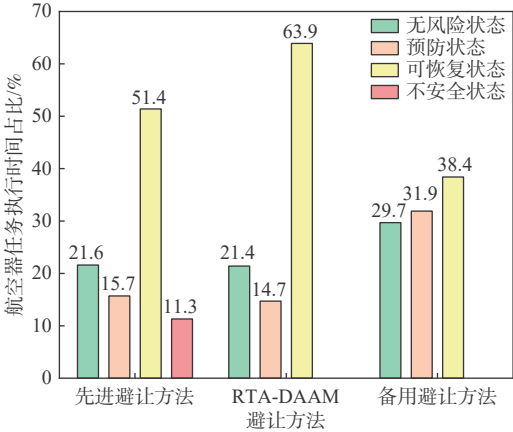


图 13 航空器任务执行时间占比

Fig. 13 Percentage of aircraft operational status time

在执行任务的过程中,备用避让方法虽然以牺牲效率为代价确保了极高的安全性,但对于紧急消防工作这类对时间要求极为严格的任务而言,效率的提升显得尤为重要。为尽可能追求效率的同时保障航空器的安全,对先进避让方法和 RTA-DAAM 避让方法的任务路径规划开展进一步分析,如图 14 和图 15 所示。

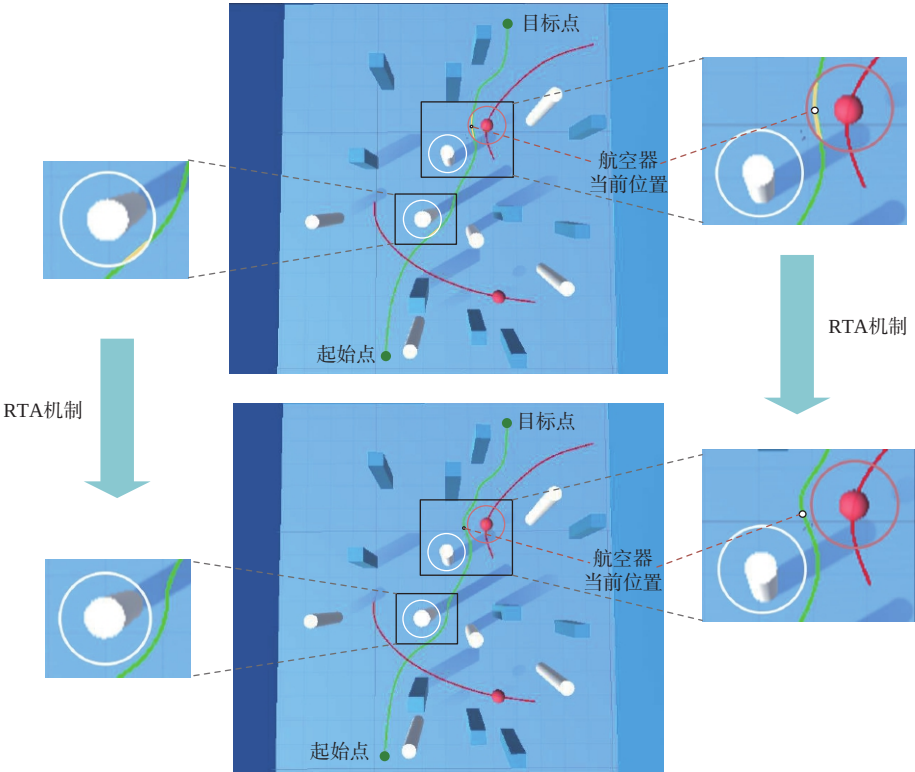


图 14 航空器任务路径规划(俯视图)

Fig. 14 Aircraft mission path planning (top view)

在先进避让方法和 RTA-DAAM 避让方法作用下,航空器与障碍物之间的最小安全距离如图 16 所示。以 RTA-DAAM 避让方法的 3 个关键任务路径点为例开展分析,航空器在起步阶段,默认采用

高安全性的备用避让方法开始执行任务,由于初始位置时航空器净空边界与障碍物边界较为接近,航空器短暂地保持与障碍物的安全距离,随后开始逐渐远离障碍物安全边界,决策模块切换至效率更高

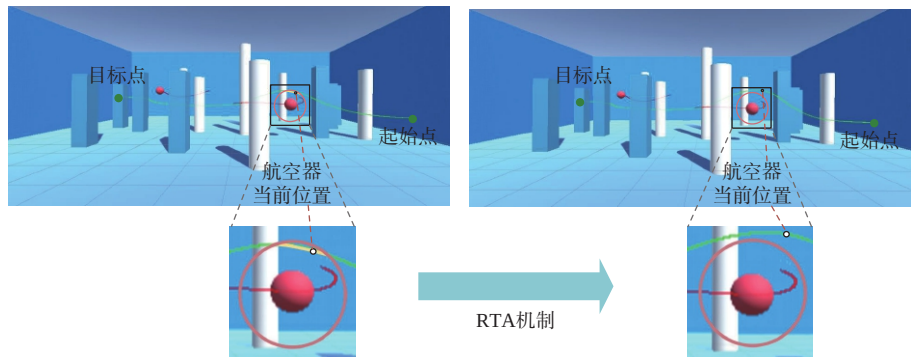


图 15 航空器任务路径规划(左视图)
Fig. 15 Aircraft mission path planning (left view)

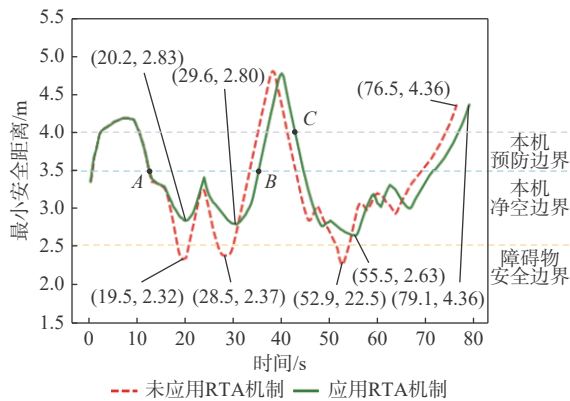


图 16 航空器与障碍物之间的最小安全距离
Fig. 16 Minimum safe distance between aircraft and obstacles

的先进避让方法规划路径,因此,在点 *A* 前 2 种避让方法形成的曲线十分接近;航空器抵达点 *A* 时,面对横向运动的入侵机,根据探测的距离数据判断出航空器净空边界与入侵机的安全边界接触,触发逻辑 L3,由决策模块立即切换至备用避让方法以确保安全,同时发出纠正告警;航空器在点 *A* 与点 *B* 之间的任务过程中,RTA-DAAM 避让方法能够保证航空器自身不进入障碍物安全边界,从而保持期望分离距离的存在;抵达点 *B* 时,航空器净空边界远离障碍物安全边界,触发逻辑 L4,决策模块再次切换回先进避让方法输出控制指令,并调整告警级别为预防告警;航空器运动至点 *C* 时,预防边界与障碍物安全边界接触,触发逻辑 L2,虽然此时并未立即构成直接威胁,但系统仍然从无告警状态转变为提供预防告警,以警示地面站工作人员密切关注航空器的运行状态,确保任务的安全执行。

同时,RTA-DAAM 的避让方法切换过程及提供的告警输出机制如图 17 所示。为确保 DAA 系统控制执行稳定性,在 2 种避让方法切换节点设置了时间裕度,有效避免在场景下短时间内反复切换导致的振荡问题。

根据 RTA-DAAM 的切换逻辑、任务效率及与

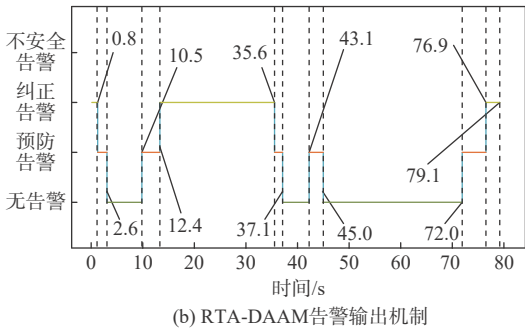
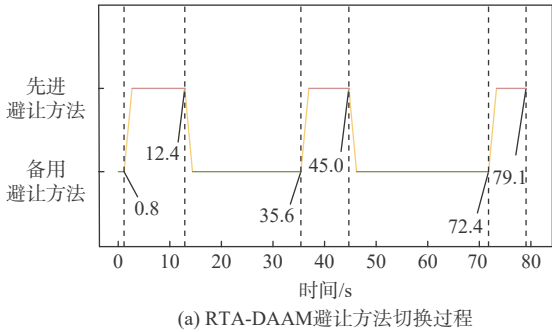


图 17 RTA-DAAM 决策与告警过程
Fig. 17 RTA-DAAM decision-making and alerting processes

障碍物间的最小安全距离效果分析可以得出,RTA-DAAM 可以全程保证航空器不侵入到障碍物安全边界,维持期望分离距离,即满足系统安全运行限制条件,该方法能够保障航空器的运行安全。

为验证 RTA-DAAM 的环境适应能力,采用不同场景类型开展测试。如表 5 所示,在密集度低的场景下,航空器与障碍物之间存在较少的冲突,2 种

表 5 不同密度场景下的航空器路径规划对比
Table 5 Comparison of aircraft path planning in different density scenarios

场景类型	安全时间占比/%		路径长度/m	
	先进避让方法	RTA-DAAM 避让方法	先进避让方法	RTA-DAAM 避让方法
低密集度	100.0	100.0	60.8	61.4
中密集度	91.9	98.5	62.1	65.8
高密度度	77.2	93.8	65.6	74.7

方法的作用下均可以保证航空器的安全; 在密集度适中的场景下, 为航空器的起始点和目标点间设置一定的冲突, RTA-DAAM 避让方法规划的任务过程中, 安全时间占比要高于先进避让方法; 密集度高的场景下, 存在大量的冲突现象, RTA-DAAM 避让方法能够保证大部分时间内航空器处于安全状态, 而位于不安全状态时, 以人工干预的方式进行了路线规划, 因此牺牲了部分效率用以保证航空器的安全。

4.4 小型无人机实飞验证

为进一步验证 RTA 对无人驾驶航空器 DAA 系统的运行安全保障效果, 采用 FS-310 小型无人机在室内开展实验, 考虑到室内场景下无人机运行场地有限, 且需保证实验人员及设备的安全, 用纸板箱堆叠作为静态障碍物, 采用气球作为动态障碍物缓慢地移动, 并在实验场地四周添置防护网。

出于场地限制, 仅考虑无人机的净空边界与障碍物的安全边界, 净空边界为无人机为圆心周围的 0.15 m, 而障碍物安全边界定义为相机采集到的障碍物周围 0.3 m 的距离, 通过通信接口地址调试将飞控与机载电脑连接, 并将 RTA 提供的控制指令切换逻辑部署在硬件平台上, 在图 18 所示的场景下进行实验。



图 18 无人机运行场景

Fig. 18 Scenarios of drone operations

设定同一起始点和目标区域, 在定点模式下通过遥控器使无人机起飞至 1 m 高度, 切至 offboard 模式使无人机进行自主路径规划, 在抵达目标区域时切换回定点模式, 保证无人机安全降落

在预定降落区域内。将飞控记录的飞行数据通过 QGroundControl 导出, 形成的路径如图 19 所示。

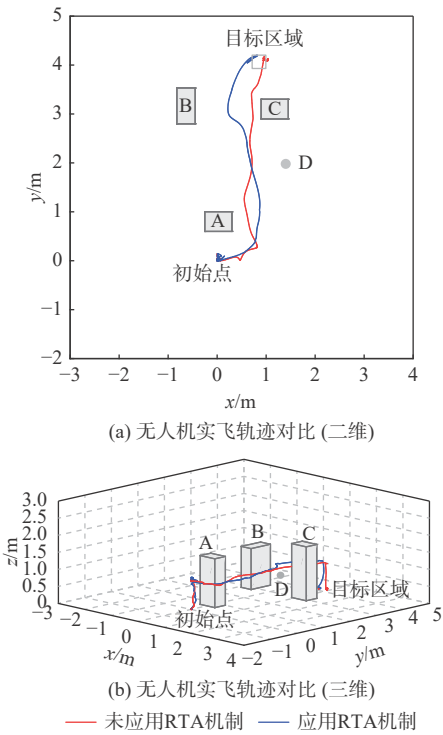


图 19 无人机飞行轨迹对比

Fig. 19 Comparison of drone flight trajectories

在 RTA 的作用下, 无人机与障碍物之间存在更大的安全距离。无人机通过相机捕获场景信息, 面对长方体障碍物时, 需要计算障碍物可能影响无人机安全的最远距离, 因此, 路径规划的部分位置时有较大的偏航角。由于无人机飞行过程受指令的控制执行的影响, 同时在按照规划路径运动中, 与动态障碍物对向接近, 为避开动态障碍物调整了实飞路径, 实飞路径存在一定的波动, 同时无人机到达目标区域后的下降过程也存在些许误差, 该误差处于可接受的范围内。

无人机与障碍物之间的距离如图 20 所示。无人机位于点 D 时触发了逻辑 L4, 随即取消纠正告警并执行了切换动作; 而抵达点 E 后, 无人机根据逻辑 L3 切换至备用避让方法继续执行任务, 即便在短暂时间后无人机净空边界离开了障碍物安全边界, 由于时间裕度的合理设置, 系统避免了频繁的方法切换, 持续采用安全的备用避让方法, 直至时间裕度结束后, 再根据距离判断并进行避让方法的切换; 无人机到达点 F 后, 因其自身进入障碍物安全边界而存在碰撞风险, 逻辑 L5 触发, 无人机发出危险告警, 由实验人员通过控制器介入并辅助无人机离开了障碍物安全边界。对无人机的运行状态时间占比进行分析, RTA 技术的引入显著提升了无人机的安全性, 使其在障碍物安全距离外运行的

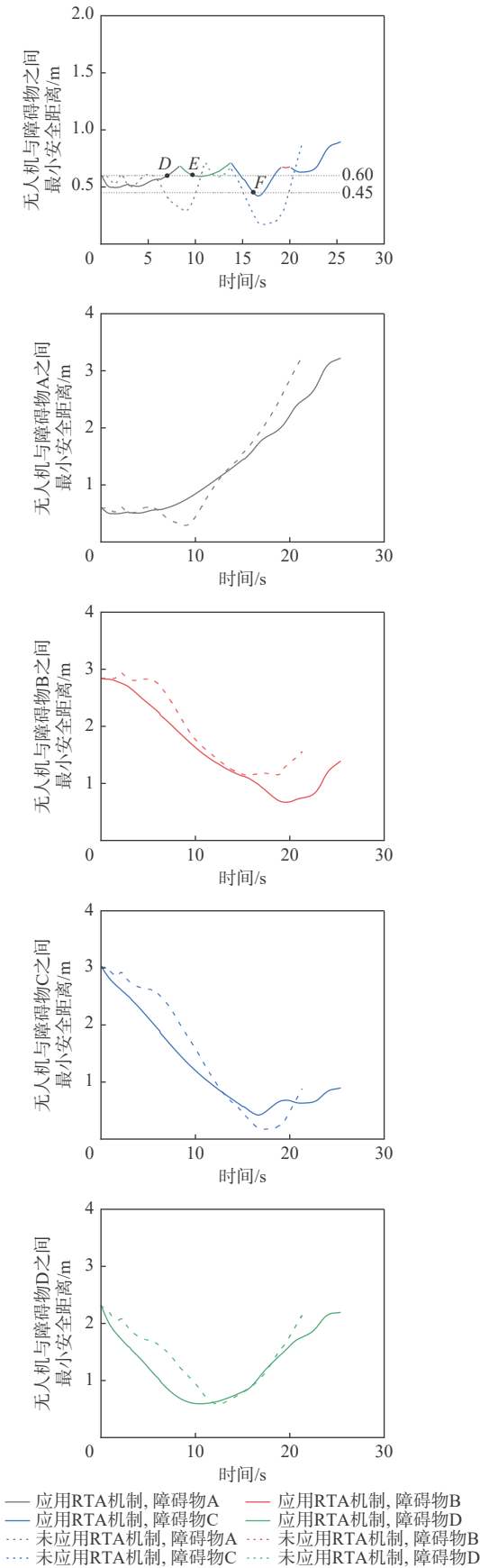


图 20 无人机与障碍物之间的最小安全距离

Fig. 20 Minimum safe distance between a drone and an obstacle

时间即安全时间占比高达 96.9%。而在这一比例之外的时间段,也有实验人员提供必要的辅助控制,

从而最大限度地确保了无人机的运行安全。

通过对比航空器仿真实验和无人机实机实验中的最小安全距离数据,可以看到,在 RTA 技术的作用下,能够使航空器在任务执行过程中兼顾效率与安全,在几乎不牺牲效率的同时为 DAA 系统提供了运行安全保障。

5 结 论

本文针对无人驾驶航空器 DAA 系统的控制决策存在效率与安全的权衡难题,提出了一个运行安全保障模块 RTA-DAAM,旨在确保无人驾驶航空器在飞行环境中能够同时实现高效与安全运行。

1) 为提高 IFDS 算法避让过程的效率和平滑度,引入 TD3 算法构成 CIFDS 算法;同时为 RTA-DAAM 设计决策与告警逻辑,实现 DAA 系统探测、决策到告警的一体化协同配合,有助于增强航空器的安全性和飞行稳定性。

2) 仿真实验结果表明,RTA-DAAM 能够使航空器在仅损失 3.40% 效率的基础上,确保航空器始终存在期望分离距离,以保障 DAA 系统运行安全。

3) 通过在小型无人机上进行实验,证明了 RTA 技术的引入能够提升无人机安全时间占比至 96.9%,引导地面站工作人员的介入也可以进一步防止危险事件的发生,提高无人驾驶航空器的运行安全性。

参考文献 (References)

[1] Huang X. The small-drone revolution is coming: scientists need to ensure it will be safe[J]. Nature, 2025, 637(8044): 29-30.

[2] Thippavong D P, Apaza R, Barmore B, et al. Urban air mobility airspace integration concepts and considerations[C]//Proceedings of the Aviation Technology, Integration, and Operations Conference. Reston: AIAA, 2018: 3676.

[3] Michelmann J, Straubinger A, Al Haddad C, et al. Urban air mobility 2030+: pathways for UAM—a scenario-based analysis[EB/OL]. [2025-03-01].<https://www.semanticscholar.org/paper/URBAN-AIR-MOBILITY-2030%2B%3A-PATHWAYS-FOR-UAM-%E2%80%93-A-Michelmann-Straubinger/dde5cd91a5cd19fd5cec68241fb59bc954e2eed0>.

[4] Cheng A W, Witzberger K E, Phojanamongkolkij N, et al. Urban air mobility (UAM) airspace research roadmap-systems engineering approach to managing airspace evolution towards UML-4[C]//AIAA Aviation 2022 Forum. Reston: AIAA, 2022.

[5] Muñoz C, Narkawicz A, Hagen G, et al. DAIDALUS: detect and avoid alerting logic for unmanned systems[C]//Proceedings of the IEEE/AIAA 34th Digital Avionics Systems Conference. Piscataway: IEEE Press, 2015.

[6] Davies J, Wu M H. Comparative analysis of ACAS-Xu and DAIDALUS detect-and-avoid systems: NASA/TM-2018-219773 [R]. Washington, D. C. : NASA, 2018.

[7] Ryu J Y, Lee H, Lee H T. Detect and avoid AI system model using a deep neural network[C]//Proceedings of the IEEE/AIAA 41st Digital Avionics Systems Conference. Piscataway: IEEE Press, 2022: 1-8.

[8] 工业和信息化部, 科学技术部, 财政部, 中国民用航空局. 绿色航空制造业发展纲要 (2023—2035 年)[R]. 北京: 工业和信息化部, 科学技术部, 财政部, 中国民用航空局, 2023.

Ministry of Industry and Information Technology, Ministry of Science and Technology, Ministry of Finance, Civil Aviation Administration of China. Guideline for green aviation manufacturing development (2023—2035)[R]. Beijing: Ministry of Industry and Information Technology, Ministry of Science and Technology, Ministry of Finance, Civil Aviation Administration of China, 2023(in Chinese).

[9] 曹宏业, 刘潇, 董绍康, 等. 面向强化学习的可解释性研究综述[J]. 计算机学报, 2024, 47(8): 1853-1882.

Cao H Y, Liu X, Dong S K, et al. A survey of interpretability research methods for reinforcement learning[J]. Chinese Journal of Computers, 2024, 47(8): 1853-1882(in Chinese).

[10] Castelvocchi D. Can we open the black box of AI?[J]. Nature, 2016, 538(7623): 20-23.

[11] Perez-Cerrolaza J, Abella J, Borg M, et al. Artificial intelligence for safety-critical systems in industrial and transportation domains: a survey[J]. ACM Computing Surveys, 2024, 56(7): 1-40.

[12] 董磊, 刘嘉琛, 陈曦, 等. 面向适航符合性的智能航电系统认证研究进展[J]. 航空工程进展, 2023, 14(3): 26-40.

Dong L, Liu J C, Chen X, et al. Research progress of AI-based avionics system certification for airworthiness compliance[J]. Advances in Aeronautical Science and Engineering, 2023, 14(3): 26-40 (in Chinese).

[13] Fuller J G. Run-time assurance: a rising technology[C]//Proceedings of the AIAA/IEEE 39th Digital Avionics Systems Conference. Piscataway: IEEE Press, 2020: 1-9.

[14] 董磊, 王琦, 陈曦, 等. 运行时保证技术的研究现状与发展综述[J]. 计算机应用, 2025, 45(3): 1003-1015.

Dong L, Wang Q, Chen X, et al. Survey of research status and development of runtime assurance technology[J]. Journal of Computer Applications, 2025, 45(3): 1003-1015(in Chinese).

[15] Cofer D, Amundson I, Sattigeri R, et al. Run-time assurance for learning-based aircraft taxiing[C]//Proceedings of the AIAA/IEEE 39th Digital Avionics Systems Conference. Piscataway: IEEE Press, 2020: 1-9.

[16] He Y N, Schumann J, Yu H F. Toward runtime assurance of complex systems with AI components[J]. PHM Society European Conference, 2022, 7(1): 166-174.

[17] Ghorl S, Khamvilai T, Feron E, et al. Runtime assurance for distributed avionics architecture[C]//Proceedings of the IEEE/AIAA 41st Digital Avionics Systems Conference. Piscataway: IEEE Press, 2022: 1-6.

[18] Abate M, Mote M, Dor M, et al. Run time assurance for spacecraft attitude control under nondeterministic assumptions[J]. IEEE Transactions on Control Systems Technology, 2024, 32(3): 862-873.

[19] Avram R, Zhang X D, Muse J A, et al. Nonlinear adaptive control of quadrotor UAVs with run-time safety assurance[C]//Proceedings of the AIAA Guidance, Navigation, and Control Conference. Reston: AIAA, 2017: 1896.

[20] Chen S D, Sun Y W, Li D C, et al. Runtime safety assurance for learning-enabled control of autonomous driving vehicles[C]//Proceedings of the International Conference on Robotics and Automation. Piscataway: IEEE Press, 2022: 8978-8984.

[21] 高雅琪. 无人机系统中 DAA 模块的研究和设计实现[D]. 成都: 电子科技大学, 2022.

Gao Y Q. Research design and implementation on DAA module of UAV system[D]. Chengdu: University of Electronic Science and Technology of China, 2022(in Chinese).

[22] Owen M P, Panken A, Moss R, et al. ACAS Xu: integrated collision avoidance and detect and avoid capability for UAS[C]//Proceedings of the IEEE/AIAA 38th Digital Avionics Systems Conference. Piscataway: IEEE Press, 2020: 1-10.

[23] Borshchova I, Ellis K. DAAMSIM: a simulation framework for establishing detect and avoid SYSTEM requirements[J]. Drone Systems and Applications, 2022, 10(1): 266-286.

[24] ASTM International. Standard practice for methods to safely bound behavior of aircraft systems containing complex functions using run-time assurance: ASTM F3269-21[S]. West Conshohocken: ASTM International, 2021.

[25] Torens C, Nagarajan P, Schirmer S, et al. Certification aspects of runtime assurance for urban air mobility: AIAA 2024-1464[R]. Reston: AIAA, 2024.

[26] Wang H L, Lyu W T, Yao P, et al. Three-dimensional path planning for unmanned aerial vehicle based on interfered fluid dynamical system[J]. Chinese Journal of Aeronautics, 2015, 28(1): 229-239.

[27] Yao P, Wang H L, Liu C. 3-D dynamic path planning for UAV based on interfered fluid flow[C]//Proceedings of the IEEE Chinese Guidance, Navigation and Control Conference. Piscataway: IEEE Press, 2015: 997-1002.

[28] Li Y X. Deep reinforcement learning: an overview[EB/OL]. (2018-11-26)[2025-03-01]. <https://arxiv.org/abs/1701.07274>.

[29] Szepesvári C. Algorithms for reinforcement learning[M]. Berlin: Springer, 2022.

[30] 吴健发, 王宏伦, 王延祥, 等. 无人机反应式扰动流体路径规划[J]. 自动化学报, 2023, 49(2): 272-287.

Wu J F, Wang H L, Wang Y X, et al. UAV reactive interfered fluid path planning[J]. Acta Automatica Sinica, 2023, 49(2): 272-287 (in Chinese).

[31] 顾雪平, 刘彤, 李少岩, 等. 基于改进双延迟深度确定性策略梯度算法的电网有功安全校正控制[J]. 电工技术学报, 2023, 38(8): 2162-2177.

Gu X P, Liu T, Li S Y, et al. Active power correction control of power grid based on improved twin delayed deep deterministic policy gradient algorithm[J]. Transactions of China Electrotechnical Society, 2023, 38(8): 2162-2177(in Chinese).

[32] Zhou X, Wang Z P, Ye H K, et al. EGO-Planner: an ESDF-free gradient-based local planner for quadrotors[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 478-485.

Design of operational security guarantee method for DAA system
based on runtime assurance

DONG Lei^{1, 2}, WANG Qi^{1, 3}, CHEN Xi^{1, 2, *}, LIU Jiachen^{1, 3}

- (1. Key Laboratory of Civil Aircraft Airworthiness Technology, Civil Aviation University of China, Tianjin 300300, China;
- 2. Science and Technology Innovation Research Institute, Civil Aviation University of China, Tianjin 300300, China;
- 3. College of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China)

Abstract: A runtime assurance-detect and avoid module (RTA-DAAM) is proposed to address the avoidance issue of unmanned aerial vehicles in the urban air mobility (UAM) scenario. It does this by identifying external environmental information sources and providing control commands and alarm information to guarantee the effectiveness and safety of aircraft in the UAM scenario. The control methods of advanced and standby collision avoidance modules in RTA-DAAM are designed. Considering the importance of coordination of detection, decision-making and warning functions, defining the aircraft’s own clearance boundaries, prevention boundaries, and obstacle safety boundaries, and dividing the different aircraft states so as to complete the design of the decision-making and warning logics. The simulation results show that RTA-DAAM can ensure the desired separation distance of the aircraft at a loss of only 3.40% of efficiency, so as to ensure the safety of the DAA system. The implementation of RTA technology can increase UAV safety time to 96.9%, providing UAV safety to the greatest extent possible, according to the real flight results of tiny UAVs.

Keywords: runtime assurance; detect and avoid system; unmanned aerial vehicles; safe operating limits; decision and alarm

导航辅助雷达量测增强的扩展型低空飞行器跟踪

李坤坤¹, 黎克波^{2,3,*}, 曹锐¹, 刘远贺⁴, 胡爽¹

(1. 中国人民解放军 63660 部队, 洛阳 471099; 2. 国防科技大学 空天科学学院, 长沙 410072;

3. 太空系统运行与控制全国重点实验室, 长沙 410073; 4. 空军航空大学, 长春 130000)

摘 要: 合作式浮低空飞行器飞行测试中, 单站高分辨雷达对飞行器表面部分观测时存在雷达量测稀疏(乃至缺失)、空间完备性差, 致使飞行器跟踪精度欠佳的问题, 对此, 提出一种融合飞行器导航信息辅助雷达量测增强的飞行器跟踪方法。根据飞行器结构扩展特点, 利用随机矩阵模型建模其近似为椭圆体的三维扩展形态; 分别建立飞行器部分观测模型及雷达测量模型, 进一步设计多种利用飞行器导航信息辅助雷达稀疏量测数据增强策略, 以提升量测数据量及空间完备性; 基于贝叶斯滤波框架实现部分观测下的飞行器状态估计。仿真实验结果表明: 所提方法不仅提升了受试飞行器的跟踪精度及飞行态势感知精度, 还兼有多源测控信息融合利用度高及计算成本低等优点。

关 键 词: 低空飞行器; 扩展目标跟踪; 随机矩阵; 部分观测; 稀疏量测; 数据增强

中图分类号: V217⁺.1; TN959.1

文献标志码: A **文章编号:** 1001-5965(2026)09-3039-13

“低空经济”^[1]作为一种新质生产力催生的综合经济形态,其广阔的开发前景引起社会各界的广泛关注,近年来已跻身成为国家重点布局发展的战略性新兴产业,其主要使能技术是高度 3 000 m 以下空域范围内运行的有人/无人机、电动垂直起降飞行器及其他通用航空飞行器。其中,浮低空飞行器作为一种携载运力大、续航时间久的飞行器,被视为物流运输、旅游观光及其他低空服务行业的有力技术选择。因此,运营使用能够稳定飞行的浮低空飞行器成为必然要求。

在浮低空飞行器投入使用前,须根据相关测试规范开展飞行测试,以检验其各项飞行性能指标^[2]。测试过程中,为保障受试飞行器稳定飞行,通常除安装机载组合导航设备实时上传飞行器导航信息(含位置、速度、姿态)至测试中心分析飞行器在空状态外,也配有地面测量雷达提供外部测控支持。地面雷达可自主搜索或利用测试中心转发的合作式飞行器导航信息作为外引导快速跟踪飞行器。

导航信息与雷达测量信息一起“高低”搭配,综合实时监测受试飞行器飞行态势,并预测可能发生的故障坠落、偏航离出等异常飞行事件,为测试人员及时应对提供决策支撑。

一般地,浮空类飞行器结构体积设计较为庞大,以充盈足够轻气体来提供所需的浮升及驻空动力,并平衡机载设备系统重力和减少推进能源消耗,从而使得飞行器长时飞行成为可能^[3]。在此情况下,飞行器的外形尺寸势必扩展增大,因此,随着现代传感技术的快速更新发展,测量雷达的分辨能力不断得到提升,受试飞行器将不再因传统雷达分辨力限制简单地作为“点/小目标”看待,而是具有一定空间分布形态的、占据多个雷达分辨率单元的“扩展型”目标。此时,受试飞行器表面分布有多个散射中心,雷达单帧将获得多量测数据。这些量测信息的处理使得除飞行器运动状态(如位置、速度乃至加速度)可以估计外,其轮廓、大小、方向等扩展形态特征的解析估计也成为可能,从而为飞行

收稿日期: 2025-05-09; 录用日期: 2025-06-20; 网络出版时间: 2025-06-25 18:26

网络出版地址: link.cnki.net/urlid/11.2625.V.20250625.0953.001

基金项目: 国家自然科学基金面上项目(12472359)

* 通信作者. E-mail: likeboreal@nudt.edu.cn

引用格式: 李坤坤, 黎克波, 曹锐, 等. 导航辅助雷达量测增强的扩展型低空飞行器跟踪 [J]. 北京航空航天大学学报, 2026, 52 (9) : 3039-3051.

Li K K, Li K B, Cao R, et al. Extended low-altitude aircraft tracking via navigation-assisted radar measurement augmentations [J].

Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9) : 3039-3051 (in Chinese) .

器飞行态势提供更多感知信息,这属于经典的扩展目标跟踪 (extended target tracking, ETT) 问题^[4-30]。

关于 ETT 问题研究, Mahler^[4] 根据随机有限集理论建立了扩展目标概率假设密度滤波器。基于该理论,此后相继发展出现了其他扩展目标滤波算法^[5-6,13],但这些算法多聚焦于扩展目标的运动状态,少有考虑目标的扩展形态特征。Koch^[7] 最早研究了带有扩展形态估计的 ETT 问题,提出原理简单、鲁棒性强的对称正定随机矩阵模型来描述目标的扩展形态。之后,国内外关于目标扩展形态建模取得了相当丰硕的研究成果^[8-23]。Hu^[10]、Tuncer^[11] 等利用固定/变化数目的多椭圆拟合的随机矩阵模型来逼近不规则形状目标的扩展形态; Baum^[12]、Zea^[13] 等先后提出了随机超表面模型及水平集随机超表面模型描述星凸形目标的扩展形态,这些模型随后得到了其他研究者的拓展应用^[14-15]; Baum^[16]、Yang^[17] 等相继建立和发展了乘性噪声模型,实现了扩展形态参数的显式表达与滤波估计; Kumru^[18]、Xu^[19] 等借助高斯过程模型学习观测数据的空间分布特征,进一步提升了星凸形目标形态表征的灵活性。此外,近年来参数化模型^[20]、B 样条拟合模型^[21]、alpha-shape 模型^[22]、二阶自回归模型^[23] 等多种模型也相继被用于建模不规则形状目标的扩展形态。

上述扩展模型结合各种滤波方法取得了良好的目标跟踪效果,但普遍针对二维扩展目标,鲜少涉及三维扩展目标的跟踪。一方面,三维 ETT 一般需要多传感器联合提供大量的多角度完备观测数据^[18,21],这与二维 ETT 研究通常取完全观测数据的假设一致,但多传感器观测经济成本高,且对多传感器管理及其量测数据融合的架构与方法提出了较高要求^[21,24-25]。相比之下,仅靠单一传感器虽降低了观测成本,但观测角度局限性大,获取信息量有限,通常面临量测数据稀疏、空间完备性差的部分观测难题^[26-27],严重制约了目标跟踪精度。另一方面,现有研究大多假设传感器量测在目标扩展表面是均匀或高斯分布的,且目标检测是稳定连续的。这种假设在三维 ETT 问题中具有极大的局限性,因为三维目标表面部分观测下势必存在可观测面与不可观测面量测点源分布差异大的现象^[18,26-27],且由于观测杂波和干扰的存在,可能出现扩展表面上散射点部分可分辨而产生稀疏量测,以及目标检测失败而造成量测缺失的情况^[28]。这些因素都对三维目标扩展形态的估计及整体认知产生不利影响。因此,对于本文三维扩展型浮低空飞行器目标,受单站雷达对其部分观测影响,目标运动状态及扩展形态的估计精度出现瓶颈,迫切需要研究有效的跟踪方法。

鉴于此,本文从合作式浮低空飞行器测试中地面高分辨测量雷达外测保障角度出发,考虑受试飞行器主体外形通常绕体轴呈旋转对称形式、具有流线椭圆体型的典型扩展先验特点,采用描述椭圆(体)的对称正定随机矩阵模型建模飞行器扩展形态,并以雷达测量数据作为跟踪核心数据,将接收的飞行器机载组合导航数据作为跟踪辅助数据。高分辨雷达的优点是不仅可以借助高精度测量数据获取受试飞行器航迹位置、速度乃至加速度等运动特征,还可以利用高分辨能力获取的单帧多量测数据解析其飞行态势感知所需的轮廓、大小、方向等扩展形态特征;缺点是单站雷达跟踪测量三维扩展型飞行器存在上述部分观测难题,导致单纯依靠雷达量测数据无法较为准确地估计飞行器运动特征及扩展形态特征。而组合导航设备的优点是数据精度高、质量稳定可靠,可连续提供受试飞行器航迹位置、速度及飞行姿态等运动特征信息,缺点是单纯依靠导航数据无法掌握飞行器轮廓、大小等扩展形态特征。

基于上述优缺点考虑,本文一方面利用雷达接收的飞行器导航位置数据引导其快速搜索跟踪目标,实现外测保障的即时响应;另一方面,通过融合利用飞行器导航位置、姿态数据,设计多种借助该导航信息对单站雷达稀疏量测数据进行增强策略,以有效扩充雷达量测数据量并提升其空间完备性。最终,基于增强雷达量测数据和高斯逆威沙特滤波提升飞行器运动状态、扩展形态的估计精度,为受试飞行器跟踪及其飞行态势感知提供重要技术与信息支撑。

1 飞行器跟踪系统状态模型

1.1 运动状态模型

假设第 $k-1$ 时刻受试飞行器在三维 (即 $d=3$) 空间的质心运动状态 $\mathbf{x}_{k-1}$ 为服从高斯分布的随机变量:

$$p(\mathbf{x}_{k-1}|\mathbf{Z}^{k-1}) = N(\mathbf{x}_{k-1}; \mathbf{x}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}) \tag{1}$$

式中: $\mathbf{x} = [x^p, v_x, y^p, v_y, z^p, v_z]^T$, x^p 、 y^p 、 z^p 为各向质心位置坐标, v_i 为对应的速度; $\mathbf{Z}^{k-1} = \{\mathbf{Z}_j\}_{j=1}^{k-1}$ 表示由第 1 至第 $k-1$ 时刻累积的雷达量测集; $\mathbf{x}_{k-1|k-1}$ 和 $\mathbf{P}_{k-1|k-1}$ 分别为该时刻运动状态后验均值和状态协方差矩阵。

考虑到浮低空飞行器结构体积大、运动惯性强、动态响应速度低^[3],故采用近似匀速模型描述其运动特性,建立运动状态演化模型:

$$\begin{cases} \mathbf{x}_k = \Phi_k \mathbf{x}_{k-1} + \mathbf{w}_k \\ \mathbf{w}_k \sim N(\mathbf{w}_k; \mathbf{0}, \mathbf{Q}_k) \end{cases} \tag{2}$$

式中: Φ_k 和 $\mathbf{Q}_k$ 分别为运动状态转移矩阵和高斯过

程白噪声 $\mathbf{w}_k$ 的协方差矩阵,表示为

$$\Phi_k = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \otimes \mathbf{I}_d \quad (3a)$$

$$\mathbf{Q}_k = \begin{bmatrix} T^4/4 & T^3/2 \\ T^3/2 & T^2 \end{bmatrix} \otimes (\sigma^2 \mathbf{I}_d) \quad (3b)$$

其中: T 为雷达采样周期; $\mathbf{I}_d$ 为 $d \times d$ 维单位矩阵; σ 为过程噪声系数; $\otimes$ 表示 Kronecker 乘积。

1.2 扩展形态模型

根据随机矩阵理论^[7-9],记第 $k-1$ 时刻受试飞行器扩展形态为对称正定的随机矩阵变量 $\mathbf{X}_{k-1} \in \mathbf{R}^{d \times d}$,并假设 $\mathbf{X}_{k-1}$ 服从逆威沙特分布:

$$p(\mathbf{X}_{k-1} | \mathbf{Z}^{k-1}) = \text{IW}(\mathbf{X}_{k-1}; \mathbf{v}_{k-1|k-1}, \mathbf{V}_{k-1|k-1}) \quad (4)$$

式中: $\text{IW}(\cdot)$ 表示逆威沙特分布; $\mathbf{v}$ 为逆威沙特分布自由度; $\mathbf{V}$ 为分布参数矩阵,且由逆威沙特分布期望性质有 $\mathbf{X}_{k-1} = \mathbf{V}_{k-1|k-1} / (\mathbf{v}_{k-1|k-1} - 2d - 2)$ 。

由于飞行器发生充盈气体泄漏而导致扩展形态的事件概率较低,此外由前述飞行器运动特性可认为短时间内扩展形态的方向变化较小,可假设飞行器扩展形态短时间内近似时不变^[7-9,29],并采用威沙特分布描述其演化特性:

$$p(\mathbf{X}_k | \mathbf{X}_{k-1}) = W(\mathbf{X}_k; \delta_k, \mathbf{A}_k \mathbf{X}_{k-1} \mathbf{A}_k^T) \quad (5)$$

式中: $W(\cdot)$ 表示威沙特分布; $\delta_k > d-1$ 为威沙特分布的自由度,用于描述扩展形态演化的不确定性程度; $\mathbf{A}_k \in \mathbf{R}^{d \times d}$ 为描述扩展形态演化模式的可逆矩阵,可取为 $\mathbf{A}_k = \mathbf{I}_d / \sqrt{\delta_k}$ 。

2 部分观测下的测量模型

2.1 坐标系及姿态定义

定义相关坐标系及受试飞行器姿态。如图1所示,定义雷达坐标系为:以雷达站址为原点 O_R , $O_R z_R$ 铅垂向上, $O_R x_R$ 和 $O_R y_R$ 位于水平面内,分别指向正东与正北。定义飞行器机体坐标系为:以飞行器质心为原点 O_B , $O_B x_B$ 沿主体轴指向头部, $O_B z_B$ 位于体对称平面内且垂直于 $O_B x_B$ 指向上方, $O_B y_B$ 指向

由右手定则确定。

定义受试飞行器导航姿态为 $\boldsymbol{\theta} = [\varphi, \psi, \gamma]^T$, 俯仰角 φ 为 $O_B x_B$ 与 $O_R x_R y_R$ 平面的夹角,取该平面以上为正,偏航角 ψ 为 $O_B x_B$ 在 $O_R x_R y_R$ 平面上的投影线与 y_R 轴的夹角,取偏东方向为正,滚转角 γ 为 $O_B x_B z_B$ 平面与包含 x_B 轴的铅垂面的夹角,取绕 x_B 轴正向旋转为正。

2.2 飞行器部分观测模型

如图1所示,随着受试飞行器运动航迹及姿态的时间变化,雷达探测视线在机体坐标系中不断变化。受该探测视角限制及飞行器结构自遮挡影响,飞行器可观测表面也随之发生改变(图1中深色区代表飞行器可观测面)。本节从相对几何关系出发建立飞行器部分观测模型。

记飞行器长半轴长为 a , 短半轴长分别为 b 和 c (考虑到飞行器主体外形旋转对称特点,有 $b=c$), 第 k 时刻雷达坐标系下的飞行器质心位置为 $\mathbf{x}_{p,k} = [x_k^p, y_k^p, z_k^p]^T$ 。定义该时刻不考虑受试飞行器姿态的扩展形态表面方程为

$$f(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k) = \frac{(\tilde{x}_k - x_k^p)^2}{a^2} + \frac{(\tilde{y}_k - y_k^p)^2}{b^2} + \frac{(\tilde{z}_k - z_k^p)^2}{c^2} - 1 \quad (6)$$

式中: $\tilde{x}_k$ 、 $\tilde{y}_k$ 、 $\tilde{z}_k$ 分别为雷达坐标系下飞行器扩展表面上任一点的3个方向位置坐标。

记第 k 时刻飞行器表面完全观测、部分观测下的雷达散射点(即雷达量测源)集分别为 $\mathcal{Z}_k = \{\mathbf{z}_k^r\}_{r=1}^{N_k}$ 、 $\mathcal{Z}'_k = \{\mathbf{z}_k^m\}_{m=1}^{N'_k}$, 且 $m \subseteq r \in [1, N_k]$, 其中, $\mathbf{z}_k^r = [\tilde{x}_k^r, \tilde{y}_k^r, \tilde{z}_k^r]^T$ 为扩展表面上第 r 散射点在雷达坐标系下的空间位置, N_k 、 N'_k 分别为完全观测、部分观测下的散射点个数,且满足 $N'_k < N_k$ 。记高分辨雷达的空间分辨力为 τ , 根据以下步骤可近似建立飞行器部分观测模型,并确定第 k 时刻雷达稀疏量测集 $\mathbf{Z}_k$ 。

步骤1 确定完全观测下的散射点集 $\mathcal{Z}_k$ 。借鉴文献[21]划分扩展目标表面控制点的思想,根据雷达空间分辨力 τ 将飞行器表面分别绕/沿体轴横/纵向等间隔划分为 s_1 、 s_2 个子区间:

$$\begin{cases} \frac{2\pi a + 4(a-b)}{s_1} \geq \tau \\ \frac{2\pi b}{s_2} \geq \tau \text{ 或 } \frac{2\pi c}{s_2} \geq \tau \end{cases} \quad (7)$$

则理想条件下每个子区间应包含一个散射点 $\mathbf{z}_k^r$, 且其位置可由式(8)确定:

$$\begin{bmatrix} \tilde{x}_k^r \\ \tilde{y}_k^r \\ \tilde{z}_k^r \end{bmatrix} = \mathbf{I}_B^R(\varphi_k, \psi_k, \gamma_k) \begin{bmatrix} a & & \\ & b & \\ & & c \end{bmatrix} \begin{bmatrix} \cos \alpha_k^r \cos \beta_k^r \\ \cos \alpha_k^r \sin \beta_k^r \\ \sin \alpha_k^r \end{bmatrix} + \mathbf{x}_{p,k} \quad (8)$$

$r = 1, 2, \dots, N_k$

式中: $\alpha_k^r, \beta_k^r \in [0, 2\pi)$ 分别为飞行器机体坐标系下第 r 散射点 $\mathbf{z}_k^r$ 相对质心 O_B 的俯仰角、方位角,且满足

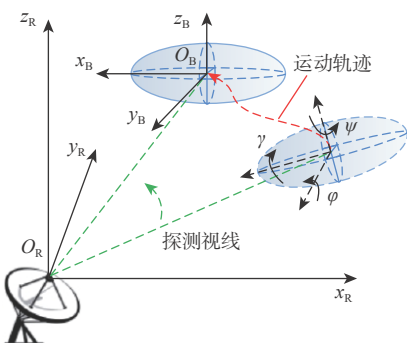


图1 坐标系示意图

Fig. 1 Schematic diagram of coordinate systems

等角间隔 $\Delta\alpha'_k = 2\pi/\varsigma_1$ 及 $\Delta\beta'_k = 2\pi/\varsigma_2$; $\mathbf{L}_B^R$ 为机体坐标系至雷达坐标系转换矩阵, 表示为

$$\mathbf{L}_B^R(\varphi, \psi, \gamma) = (\mathbf{L}_R^B(\varphi, \psi, \gamma))^T = (\mathbf{E}_X(\gamma)\mathbf{E}_Y(-\varphi)\mathbf{E}_Z(\pi/2 - \psi))^T \quad (9)$$

式中: $\mathbf{E}_i(\theta)$ 表示绕 $i = X, Y, Z$ 轴正向旋转角度 θ 得到的旋转矩阵, 表达式可见文献[26]。考虑到飞行器主体外形横/纵向尺寸绕/沿体轴非均匀变化、部分弱散射点会合并为强散射点及散射点部分可分辨等情况, 飞行器表面完全观测下散射点集 $\mathcal{Z}_k$ 应由少于 $N_k = \varsigma_1\varsigma_2$ 个散射点组成。

步骤2 确定部分观测下的散射点集 $\mathcal{Z}'_k$ 。对 $\mathcal{Z}_k$ 中任一散射点 $\mathbf{z}'_k$, 当其位置矢量与该点处的椭圆体法线矢量:

$$\perp(\mathbf{z}'_k) = \begin{bmatrix} \frac{\partial f(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k)}{\partial \tilde{x}_k} \\ \frac{\partial f(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k)}{\partial \tilde{y}_k} \\ \frac{\partial f(\tilde{x}_k, \tilde{y}_k, \tilde{z}_k)}{\partial \tilde{z}_k} \end{bmatrix}_{\tilde{x}_k = \tilde{x}'_k, \tilde{y}_k = \tilde{y}'_k, \tilde{z}_k = \tilde{z}'_k} \quad (10)$$

满足假设H1时, 可确定该散射点 $\mathbf{z}'_k$ 位于飞行器的可观测面, 即雷达探测视线与飞行器朝向雷达一侧表面的交点; 否则, 位于飞行器的不可观测面。

$$\begin{cases} \text{H0: } \text{dot}(\mathbf{z}'_k, \perp(\mathbf{z}'_k)) < 0 \text{ 或 } \theta < \pi/2 \\ \text{H1: } \text{dot}(\mathbf{z}'_k, \perp(\mathbf{z}'_k)) \geq 0 \text{ 或 } \theta \geq \pi/2 \end{cases} \quad (11)$$

式中: H0表示假设 $\mathbf{z}'_k$ 不可观测, H1表示假设该点可观测; $\text{dot}(\cdot, \cdot)$ 表示求点积操作; θ 为2个向量的空间夹角。根据式(10)、式(11), 遍历 $\mathcal{Z}_k$ 中所有散射点可确定飞行器表面部分观测下的散射点集 $\mathcal{Z}'_k$ 。

步骤3 确定部分观测下的雷达稀疏量测集 $\mathbf{Z}_k$ 。

假设雷达量测为独立同高斯分布的散射点位置量测, 且量测数目 n_k 服从均值为 κ 的泊松分布^[8-9]。记雷达目标检测概率为 ρ , 可确定飞行器表面部分观测下的雷达稀疏量测集 $\mathbf{Z}_k$:

$$\begin{cases} \chi_0: \mathbf{Z}_k = \emptyset & \text{w.p. } 1 - \rho \\ \chi_1: \mathbf{Z}_k = \mathcal{Z}'_k \left\{ \mathbf{z}'_k \right\}_{r=\text{rand}\{1:N'_k\}}^{n_k} & \text{w.p. } \rho \end{cases} \quad (12)$$

式中: χ_0 表示假设第 k 时刻雷达检测失败, χ_1 表示假设雷达检测成功; w.p. 表示 with probability (以概率); $r = \text{rand}\{1:N'_k\}$ 表示从部分观测下散射点集 $\mathcal{Z}'_k$ 的序号集 $\{1:N'_k\}$ 中随机选取 $n_k = \text{Poisson}(\kappa)$ 个, $\text{Poisson}(\cdot)$ 表示泊松分布。另外, 量测数目满足 $n_k \ll N'_k < N_k$ 。

2.3 雷达测量模型

对上述雷达稀疏量测集 $\mathbf{Z}_k = \{\mathbf{z}'_k\}_{r=1}^{n_k}$ 采用如下测量模型^[9,25,29]:

$$\mathbf{z}'_k = \mathbf{H}_k \mathbf{x}_k + \boldsymbol{\omega}_k^r \quad \boldsymbol{\omega}_k^r \sim N(\boldsymbol{\omega}_k^r; 0, \mathbf{B}_k \mathbf{X}_k \mathbf{B}_k^T) \quad (13)$$

式中: $\mathbf{H}_k = \tilde{\mathbf{H}}_k \otimes \mathbf{I}_d$ 为测量矩阵, $\tilde{\mathbf{H}}_k = [1, 0]$ 为单维物理空间的测量矩阵; $\boldsymbol{\omega}_k^r$ 为综合考虑飞行器扩展形态与雷达位置测量噪声影响的独立同分布高斯白噪声, 其方差为 $\mathbf{B}_k \mathbf{X}_k \mathbf{B}_k^T$, $\mathbf{B}_k$ 为描述雷达观测飞行器扩展形态失真情况的可逆矩阵, 并近似取 $\mathbf{B}_k = (\varepsilon \mathbf{X}_{k|k-1} + \mathbf{R}_k)^{1/2} \mathbf{X}_{k|k-1}^{-1/2}$, ε 为描述扩展形态影响雷达量测散布的尺度因子; $\mathbf{X}_{k|k-1} = \mathbf{V}_{k|k-1}/(v_{k|k-1} - 2d - 2)$ 为扩展形态预测矩阵; $\mathbf{R}_k$ 为雷达位置测量噪声协方差矩阵。

由式(13), 量测似然概率密度函数可表示为

$$p(\mathbf{Z}_k | \mathbf{x}_k, \mathbf{X}_k) = \prod_{r=1}^{n_k} N(\mathbf{z}'_k; \mathbf{H}_k \mathbf{x}_k, \mathbf{B}_k \mathbf{X}_k \mathbf{B}_k^T) \sim N(\bar{\mathbf{z}}_k; \mathbf{H}_k \mathbf{x}_k, \mathbf{B}_k \mathbf{X}_k \mathbf{B}_k^T / n_k) W(\bar{\mathbf{Z}}_k; n_k - 1, \mathbf{B}_k \mathbf{X}_k \mathbf{B}_k^T) \quad (14)$$

式中: $\bar{\mathbf{z}}_k$ 为量测质心; $\bar{\mathbf{Z}}_k$ 为量测散射矩阵, 分别表示为

$$\bar{\mathbf{z}}_k = \frac{1}{n_k} \sum_{r=1}^{n_k} \mathbf{z}'_k \quad (15a)$$

$$\bar{\mathbf{Z}}_k = \sum_{r=1}^{n_k} (\mathbf{z}'_k - \bar{\mathbf{z}}_k)(\mathbf{z}'_k - \bar{\mathbf{z}}_k)^T \quad (15b)$$

3 融合导航信息的量测数据增强

数据增强方法^[31]是通过人工或算法手段对原始数据进行扩充或变换以生成多样化样本数据的技术, 对于小样本数据及分布不平衡数据的处理具有重要意义。受该方法思想启发, 本节针对受试浮低空飞行器部分观测下的单站雷达量测稀疏且空间完备性差难题, 设计了一种简洁但行之有效的解决方案: 融合利用雷达接收的测试中心转发合作式飞行器的高精度组合导航信息, 对雷达量测数据实施如图2所示的3种增强策略。

3.1 量测数据增强策略

以第 k 时刻雷达稀疏量测集 $\mathbf{Z}_k$ 中任一量测 $\mathbf{z}'_k$ 的数据增强为例, 对各策略思路分别阐述如下:

1) 增强策略1。如图2(a)所示, 利用飞行器质心 O_B 导航位置 $\mathbf{x}_{n,k} = [x_{n,k}^p, y_{n,k}^p, z_{n,k}^p]^T$ (该位置可事先由导航装置安装位置与合作式飞行器中心相对位置关系确定) 及导航姿态 $\boldsymbol{\theta}_k$ 确定的飞行器主轴空间指向, 对位于飞行器表面附近的 $\mathbf{z}'_k$ (记为0号量测) 作3次绕主轴旋转, 分别得到 90° 、 180° 、 270° 旋转角度下的图2(a)中1、2、3号量测:

$$\mathbf{z}_k^{r-j} = \mathbf{x}_{n,k} + \mathbf{L}_B^R(\varphi_k, \psi_k, \gamma_k) \mathbf{E}_X(j\pi/2) \mathbf{L}_R^B(\mathbf{z}'_k - \mathbf{x}_{n,k}) \quad j = 1, 2, 3 \quad (16)$$

式中: 上标 $r-j$ 表示对第 r 量测 $\mathbf{z}'_k$ 作第 j 次旋转。遍历

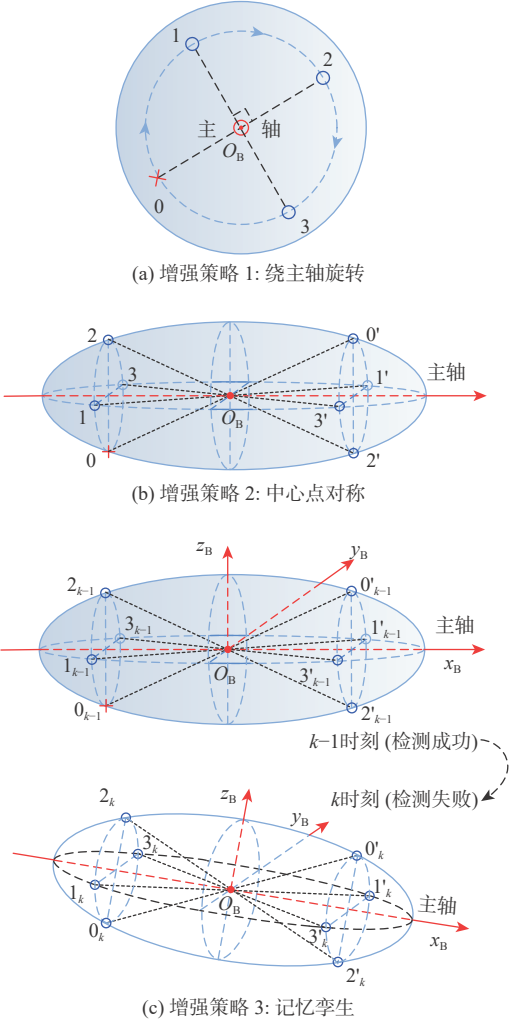


图 2 3 种雷达量测数据增强策略示意图

Fig. 2 Schematic diagrams of three augmentation strategies for radar measurements

$\mathbf{Z}_k$ 中其余量测, 得到增强量测集 $\mathbf{Z}_k^{\text{Au1}} = \{\mathbf{z}_k^{r'}\}_{r'=1}^{4n_k}$, 上标 Au1 表示第 1 种增强策略。此时, 在飞行器绕体轴的横向表面附近增加了 3 倍原始量测数据量, 并提升了空间完备性。

2) 增强策略 2。如图 2(b) 所示, 进一步利用飞行器质心 O_B 导航位置 $\mathbf{x}_{n,k}$, 对上述增强量测集 $\mathbf{Z}_k^{\text{Au1}}$ 中任一量测 $\mathbf{z}_k^{r'}$ (记为 r 号量测) 作中心点对称:

$$\mathbf{z}_k^{r'} = [\tilde{x}_k^{r'}, \tilde{y}_k^{r'}, \tilde{z}_k^{r'}]^T = 2\mathbf{x}_{n,k} - \mathbf{z}_k^{r'} = [2x_{n,k}^p - \tilde{x}_k^{r'}, 2y_{n,k}^p - \tilde{y}_k^{r'}, 2z_{n,k}^p - \tilde{z}_k^{r'}]^T \quad (17)$$

得到图中对应 r' 号量测, 遍历 $\mathbf{Z}_k^{\text{Au1}}$ 中其余量测, 得到增强量测集 $\mathbf{Z}_k^{\text{Au2}} = \{\mathbf{z}_k^{r'}\}_{r'=1}^{8n_k}$, 上标 Au2 表示第 2 种增强策略。此时, 在飞行器沿体轴的纵向表面附近继续增加 4 倍原始量测数据量, 并提升了空间完备性。

3) 增强策略 3。如图 2(c) 所示, 对于第 k 时刻雷达检测飞行器失败而量测集为空的特殊情况, 利用 $k-1$ 、 k 这 2 个时刻的高精度飞行器导航姿态 $\boldsymbol{\theta}_{k-1}$ 、 $\boldsymbol{\theta}_k$, 对第 $k-1$ 时刻得到的增强量测集 $\mathbf{Z}_{k-1}^{\text{Au2}}$ 中任一量测 $\mathbf{z}_{k-1}^{r'}$ (记为 r'_{k-1} 号量测) 作记忆孪生:

$$\mathbf{z}_k^{r'} = \mathbf{x}_{n,k} + \mathbf{L}_B^R(\varphi_k, \psi_k, \gamma_k) \mathbf{L}_{B,k-1}^{\text{Bk}} \mathbf{L}_{R,k-1}^{\text{Bk-1}} (\mathbf{z}_{k-1}^{r'} - \mathbf{x}_{n,k-1}) \quad (18)$$

得到第 k 时刻在机体坐标系下具有相同空间位置关系的孪生量测 $\mathbf{z}_k^{r'}$ (记为 r'_k 号量测), 其中, $\mathbf{L}_{B,k-1}^{\text{Bk}}$ 为机体坐标系由第 $k-1$ 时刻至第 k 时刻的转换矩阵:

$$\mathbf{L}_{B,k-1}^{\text{Bk}} = \mathbf{E}_X(\gamma_{k-1} - \gamma_k) \mathbf{E}_Y(\varphi_{k-1} - \varphi_k) \mathbf{E}_Z(\psi_{k-1} - \psi_k) \quad (19)$$

遍历 $\mathbf{Z}_{k-1}^{\text{Au2}}$ 中其余量测, 得到增强量测集 $\mathbf{Z}_k^{\text{Au3}} = \{\mathbf{z}_k^{r'}\}_{r'=1}^{8n_k}$, 上标 Au3 表示第 3 种增强策略。此时, 第 k 时刻雷达检测失败而量测集为空的信息不足情况得到了很好的补偿。特别地, 考虑到前 2 种增强策略一般可较好满足跟踪量测数据量及空间完备性要求, 故第 3 种增强策略仅在雷达检测失败情况下使用。

3.2 量测数据增强复杂度

采用浮点运算数的复杂度分析方法评估雷达量测增强的复杂度。表 1 为文献 [32] 给出的 4 种矩阵运算的浮点运算数。

表 1 4 种矩阵运算的浮点运算数^[32]

Table 1 Flops of four kinds of matrix operations ^[32]		
矩阵运算	示例矩阵对象	浮点运算数
加减法	$\mathbf{M}_1 \in \mathbf{R}^{n \times m}, \mathbf{M}_2 \in \mathbf{R}^{n \times m}$	$O(nm)$
乘法	$\mathbf{M}_1 \in \mathbf{R}^{n \times m}, \mathbf{M}_2 \in \mathbf{R}^{m \times l}$	$O(2nml - nl)$
求逆	$\mathbf{M}_1 \in \mathbf{R}^{n \times n}$	$O(n^3)$
Cholesky 分解	$\mathbf{M}_1 \in \mathbf{R}^{n \times n}$	$O(n^3/3)$

由式 (16)~式 (19) 可知, 本文设计的 3 种雷达量测数据增强策略主要涉及表 1 中的矩阵加减法、乘法等 2 种运算, 因此, 对稀疏量测集 $\mathbf{Z}_k = \{\mathbf{z}_k^{r'}\}_{r'=1}^{n_k}$, 可计算得到表 2 所示的各雷达量测增强策略的复杂度。由于增强策略 3 对应的检测失败情况受雷达检测概率 ρ 影响, 3 种量测增强策略的总复杂度为 $O(\rho(51n_k + 12n_k) + (1 - \rho)51n_k) = O((12\rho + 51)n_k)$ 。考虑稀疏量测数量 n_k 值较小, 本文设计的雷达量测数据增强策略具有较低的复杂度。

表 2 3 种量测数据增强策略的复杂度

Table 2 Flops of three kinds of measurements data augmentation strategy	
增强策略	浮点运算量
1	$O(51n_k)$
2	$O(12n_k)$
3	$O(51n_k)$

4 基于贝叶斯框架的飞行器跟踪

从贝叶斯滤波角度, 通过计算目标运动状态 $\mathbf{x}_k$ 、扩展形态 $\mathbf{X}_k$ 的联合后验条件概率密度函数, 可

获得扩展目标的状态估计 $\mathbf{x}_{k|k}$ 和 $\mathbf{X}_{k|k}$ 。

$$p(\mathbf{x}_k, \mathbf{X}_k | \mathbf{Z}^k) = \frac{p(\mathbf{Z}_k | \mathbf{x}_k, \mathbf{X}_k) p(\mathbf{x}_k, \mathbf{X}_k | \mathbf{Z}^{k-1})}{\iint p(\mathbf{Z}_k | \mathbf{x}_k, \mathbf{X}_k) p(\mathbf{x}_k, \mathbf{X}_k | \mathbf{Z}^{k-1}) d\mathbf{x}_k d\mathbf{X}_k} \quad (20)$$

因此, 基于贝叶斯滤波框架, 本文受试浮低空飞行器的跟踪估计分为以下时间更新、测量更新 2 个过程。

4.1 时间更新过程

第 $k-1$ 时刻, 根据该时刻累积雷达量测集 $\mathbf{Z}^{k-1}$ 对第 k 时刻飞行器状态进行一步预测:

$$p(\mathbf{x}_k, \mathbf{X}_k | \mathbf{Z}^{k-1}) \approx p(\mathbf{x}_k | \mathbf{Z}^{k-1}) p(\mathbf{X}_k | \mathbf{Z}^{k-1}) \quad (21)$$

由式(21)可见, 飞行器状态一步预测可分为运动状态预测和扩展形态预测 2 部分。

1) 运动状态预测。由式 (1)、式 (2), 飞行器运动状态预测密度为

$$p(\mathbf{x}_k | \mathbf{Z}^{k-1}) = N(\mathbf{x}_k; \mathbf{x}_{k|k-1}, \mathbf{P}_{k|k-1}) \quad (22)$$

式中:

$$\mathbf{x}_{k|k-1} = \Phi_k \mathbf{x}_{k-1|k-1} \quad (23a)$$

$$\mathbf{P}_{k|k-1} = \Phi_k \mathbf{P}_{k-1|k-1} \Phi_k^T + \mathbf{Q}_k \quad (23b)$$

2) 扩展形态预测。由式 (4)、式 (5), 飞行器扩展形态预测密度为

$$p(\mathbf{X}_k | \mathbf{Z}^{k-1}) = \int p(\mathbf{X}_k | \mathbf{X}_{k-1}) p(\mathbf{X}_{k-1} | \mathbf{Z}^{k-1}) d\mathbf{X}_{k-1} \approx \text{IW}(\mathbf{X}_k; \mathbf{v}_{k|k-1}, \mathbf{V}_{k|k-1}) \quad (24)$$

详细推导过程可见文献 [9]。式中:

$$\mathbf{v}_{k|k-1} = \frac{2\delta_k(\lambda_{k-1} + 1)(\lambda_{k-1} - 1)(\lambda_{k-1} - 2)}{\lambda_{k-1}^2(\lambda_{k-1} + \delta_k)} + 2d + 4 \quad (25a)$$

$$\mathbf{V}_{k|k-1} = \frac{\delta_k}{\lambda_{k-1}} (\mathbf{v}_{k|k-1} - 2d - 2) \mathbf{A}_k \mathbf{X}_{k-1} \mathbf{A}_k^T \quad (25b)$$

$$\lambda_{k-1} = \mathbf{v}_{k-1|k-1} - 2d - 2 \quad (25c)$$

从而由前述扩展形态所服从逆威沙特分布的期望性质, 可知飞行器扩展形态预测为

$$\mathbf{X}_{k|k-1} = \mathbf{V}_{k|k-1} / (\mathbf{v}_{k|k-1} - 2d - 2) \quad (26)$$

4.2 测量更新过程

第 k 时刻, 获得雷达量测集 $\mathbf{Z}_k = \{\mathbf{z}_k^i\}_{i=1}^{n_k}$ 形成累积量测集 $\mathbf{Z}^k$ 后, 将式 (14)、式 (21) 代入式 (20) 并通过因式分解可得^[9,29]

$$p(\mathbf{x}_k, \mathbf{X}_k | \mathbf{Z}^k) \approx N(\mathbf{x}_k; \mathbf{x}_{k|k}, \mathbf{P}_{k|k}) \text{IW}(\mathbf{X}_k; \mathbf{v}_{k|k}, \mathbf{V}_{k|k}) = p(\mathbf{x}_k | \mathbf{Z}^k) p(\mathbf{X}_k | \mathbf{Z}^k) \quad (27)$$

由式(27)可见, 飞行器状态测量更新可分为运动状态更新和扩展形态更新 2 部分。

1) 运动状态更新。

$$p(\mathbf{x}_k | \mathbf{Z}^k) = N(\mathbf{x}_k; \mathbf{x}_{k|k}, \mathbf{P}_{k|k}) \quad (28)$$

式中:

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k(\bar{\mathbf{z}}_k - \mathbf{H}_k \mathbf{x}_{k|k-1}) \quad (29a)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{G}_k \mathbf{K}_k^T \quad (29b)$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \mathbf{G}_k^{-1} \quad (29c)$$

$$\mathbf{G}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{I}_d |\mathbf{B}_k|^2 / n_k \quad (29d)$$

2) 扩展形态更新。

$$p(\mathbf{X}_k | \mathbf{Z}^k) = \text{IW}(\mathbf{X}_k; \mathbf{v}_{k|k}, \mathbf{V}_{k|k}) \quad (30)$$

式中:

$$\mathbf{v}_{k|k} = \mathbf{v}_{k|k-1} + n_k \quad (31a)$$

$$\mathbf{V}_{k|k} = \mathbf{V}_{k|k-1} + \mathbf{Y}_k + \mathbf{B}_k^{-1} \bar{\mathbf{Z}}_k \mathbf{B}_k^{-T} \quad (31b)$$

$$\mathbf{Y}_k = \mathbf{G}_k^{-1} (\bar{\mathbf{z}}_k - \mathbf{H}_k \mathbf{x}_{k|k-1}) (\bar{\mathbf{z}}_k - \mathbf{H}_k \mathbf{x}_{k|k-1})^T \quad (31c)$$

从而可知飞行器扩展形态更新为

$$\mathbf{X}_{k|k} = \mathbf{V}_{k|k} / (\mathbf{v}_{k|k} - 2d - 2) \quad (32)$$

需要说明的是, 对于 3.1 节采用量测增强策略得到的增强雷达量测集, 可用其替代上述 $\mathbf{Z}_k$ 开展测量更新, 也就是说, 基于雷达量测增强的跟踪方法的时间更新、测量更新 2 个过程之间增加了量测增强过程。考虑到 $\mathbf{X}_{k|k}$ 无法直观地描述受试飞行器飞行态势感知所需的扩展形态特征信息, 基于 $\mathbf{X}_{k|k}$ 表征椭圆 (体) 形轮廓的事实, 可采用特征值分解方法对 $\mathbf{X}_{k|k}$ 进行特征提取^[30], 将所得特征值的平方根作为飞行器的半轴尺寸, 将最大特征值对应的特征向量作为飞行器主 (体) 轴指向, 从而实现受试飞行器扩展形态特征的在线估计。

5 仿真实验验证

5.1 仿真场景构建

仿真生成受试浮低空飞行器在开阔地带的一段含上升、浮低空驻空巡航、下降的典型低空测试轨迹, 如图 3 所示。其中, 巡航段包含 3 次转弯阶段及多次平直飞行阶段, 高度维持在距地面 800 m, 巡航速度 20 m/s, 且巡航阶段速度方向与飞行器主轴指向假设一致。设飞行器被雷达跟踪起点在地面雷达坐标系下为 $(-1115.6, 0, 759.2)$ m, 跟踪终点为 $(-331.7, 498.8, 674.3)$ m。设飞行器长半轴长 a 为 20 m, 短半轴长 b 为 8 m。

设置 3 种不同质量雷达、导航数据, 以通过两两之间比较的方式检验数据源质量变化条件下本文方法的有效性, 分别为: ①数据质量 1。雷达位置测量噪声方差 $\mathbf{R} = 0.25 \mathbf{I}_d$ (单位为 m^2), 飞行器机载组合导航定位精度 0.25 m、姿态精度 2 mrad。②数据质量 2。雷达位置测量噪声方差 $\mathbf{R} = 0.25 \mathbf{I}_d$ (单位为 m^2),

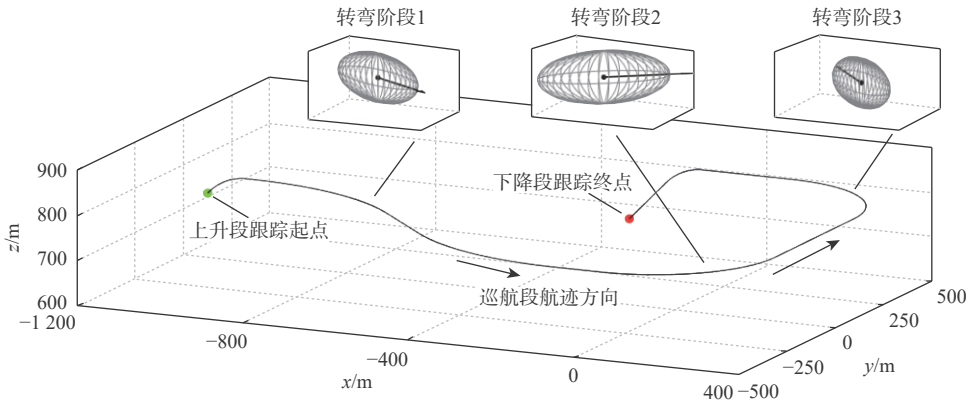


图 3 受试飞行器典型低空测试航迹示意图

Fig. 3 Schematic diagram of typical test trajectory of low-altitude aircraft

飞行器机载组合导航定位精度 1 m、姿态精度 8 mrad。
③数据质量 3。雷达位置测量噪声方差 $\mathbf{R} = 4\mathbf{I}_d$ (单位为 m^2)，飞行器机载组合导航定位精度 1 m、姿态精度 8 mrad。此外，设地面测量雷达采样周期 $T = 0.5 \text{ s}$ ，检测概率 $\rho = 0.8$ ，雷达量测数目服从均值 $\kappa = 8$ 的泊松分布，且量测在飞行器可观测表面附近随机分布，尺度因子 $\varepsilon = 1.4$ 。假设测试过程中机载组合导航设备工作正常且导航数据传输稳定，雷达测量信息与飞行器导航信息在进行融合时已完成时空配准。

对随机矩阵建模及滤波，考虑到飞行器扩展形态时间演化的不确定性很低，取其演化所服从威沙特分布的自由度 $\delta = 100$ 。滤波初始化中，飞行器质心位置初值取作跟踪初始时刻的飞行器质心导航位置，质心速度初值依跟踪前 2 帧的质心导航位置作差分获取，质心运动状态协方差初值为 $\mathbf{P}_0 = \text{diag}([20, 20, 20, 4, 4, 4]^2)$ ；飞行器扩展形态所服从逆威沙特分布的自由度、参数矩阵初值分别取 $\nu_0 = 9$ 、 $\mathbf{V}_0 = \text{diag}([80, 80, 80]^2)$ ，过程噪声系数取 $\sigma = 1$ 。

5.2 估计性能评价

采用均方根误差 (root mean square error, RMSE) 评价飞行器质心运动状态 (位置、速度) 的估计性能：

$$\Xi_k^i = \sqrt{\frac{1}{N_r} \sum_{l=1}^{N_r} \|\hat{\mathbf{S}}_{k,l}^i - \mathbf{S}_{k,l}^i\|^2} \quad i = 1, 2 \quad (33)$$

式中： $i = 1, 2$ 分别表示质心位置、速度等运动状态变量； k 表示跟踪第 k 时刻 (帧)； N_r 为蒙特卡罗仿真次数，本文取 200； $\mathbf{S}$ 为飞行器质心运动状态真值； $\hat{\mathbf{S}}$ 为状态估计值； l 表示第 l 次蒙特卡罗运行。

采用平均估计偏差 (average estimation deviation, AED) 评价飞行器扩展形态特征参数 (半轴长、主轴指向) 的估计性能：

$$\Lambda_k^j = \frac{1}{N_r} \sum_{l=1}^{N_r} (\hat{\varsigma}_{k,l}^j - \varsigma_{k,l}^j) \quad j = 1, 2, 3 \quad (34)$$

式中： $j = 1, 2, 3$ 分别表示飞行器长半轴长、短半轴长、主轴指向等扩展形态特征参数； ς 为飞行器扩展形态特征参数真值； $\hat{\varsigma}$ 为参数估计值。

5.3 实验结果分析

每次蒙特卡罗运行中，假设前 10 帧雷达对飞行器检测成功以实现跟踪稳定起批。此外，人为设置第 46 帧 (位于巡航转弯阶段 1)、第 156 帧 (位于巡航转弯阶段 2)、第 246 帧 (位于巡航转弯阶段 3) 雷达检测失败而各检测失败帧的前、后 2 帧检测成功的稳定现象，其余检测帧按雷达检测概率 $\rho = 0.8$ 随机生成雷达检测成功/失败现象。

5.3.1 雷达量测增强

以数据质量 2 条件下的飞行器巡航转弯阶段 1 第 45 帧雷达检测成功及其下一帧第 46 帧检测失败的雷达稀疏量测数据增强为例，图 4 展示了单次仿真运行下的量测增强结果。由图 4(a) 可以看到，在第 45 帧雷达检测成功时刻，雷达量测受部分观测影响在受试飞行器可观测表面附近分布稀疏，且空间完备性差。在融合飞行器导航位置、姿态数据相继实施增强策略 1 (绕主轴旋转)、增强策略 2 (中心点对称) 后，雷达量测在绕飞行器体 (主) 轴的横向、沿体轴的纵向等表面附近均实现了数据量大扩充，并提升了空间完备性，达到了“扩充样本数据量”及“平衡样本数据空间分布”的增强效果，从而说明了 2 种增强策略的有效性。图 4(b) 展示了第 46 帧雷达检测失败时刻通过实施增强策略 3 (记忆孪生)，实现了与第 45 帧具有机体坐标系下相同空间位置关系的量测孪生目的，从而很好地补偿了由于雷达检测失败致使量测缺失的不足。其他雷达检测失败时刻的稀疏量测增强效果与之类似，限于篇幅不作展示。最终，增强雷达量测可为飞行器状态估计提供有力信息支撑。

由上述雷达量测数据增强过程可知，本文基于雷达量测增强的飞行器跟踪方法是在利用飞行器

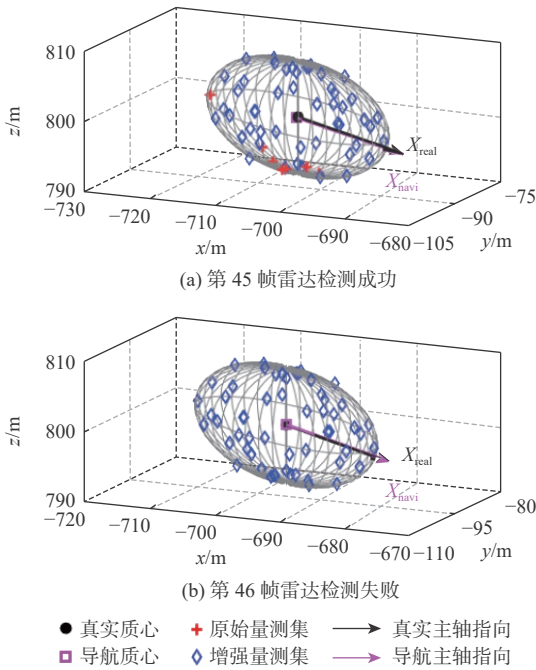


图 4 转弯阶段 1 雷达稀疏量测数据增强结果

Fig. 4 Augmentation results for sparse radar measurements in turning phase 1

导航位置数据作为雷达外引导实现搜索跟踪之外,进一步融合利用导航位置及姿态数据对雷达稀疏量测数据进行增强的过程。因此,可以从导航位置、姿态数据的单帧融合利用次数出发,分析比较量测增强与否时的导航信息融合利用度。由于量测不增强时的跟踪代表雷达仅利用导航位置数据作为外引导跟踪飞行器,雷达单帧测量周期内(以第 k 时刻为例)无论检测成功与否均仅利用 1 次导航位置数据 $\mathbf{x}_{n,k}$ 。而对于量测增强时的飞行器跟踪,由式(16)可知,增强策略 1 中雷达量测 $\mathbf{z}_k^r$ 分别融合利用当前时刻的导航位置数据 $\mathbf{x}_{n,k}$ 及导航姿态数据 $\boldsymbol{\theta}_k = [\varphi_k, \psi_k, \gamma_k]^T$ 各 1 次;由式(17)可知,增强策略 2 中 $\mathbf{z}_k^r$ 再次融合利用 1 次 $\mathbf{x}_{n,k}$;由式(18)、式(19)可知,增强策略 3 中 $\mathbf{z}_k^r$ 分别融合利用当前时刻的 $\mathbf{x}_{n,k}$ 和 $\boldsymbol{\theta}_k$ 各 1 次及第 $k-1$ 历史时刻的 $\mathbf{x}_{n,k-1}$ 和 $\boldsymbol{\theta}_{k-1}$ 各 1 次。考虑增强策略 3 对应的雷达检测失败情况受检测概率 $\rho = 0.8$ 影响,故本文方法单帧的导航位置数据总融合利用度为 $\rho(1+1) + (1-\rho)(1+1) = 0.8 \times 2 + 0.2 \times 2 = 2$ 次,单帧的导航姿态数据总融合利用度为 $\rho \times 1 + (1-\rho) \times (1+1) = 0.8 \times 1 + 0.2 \times 2 = 1.2$ 次。综合比较而言,本文方法具有多源测控信息融合利用度高的优势。

5.3.2 状态联合估计

首先,为清楚展示和比较量测增强与否时受试飞行器的位置、扩展形态滤波估计效果,图 5 以数据质量 2 条件下的飞行器状态联合估计为例,取等

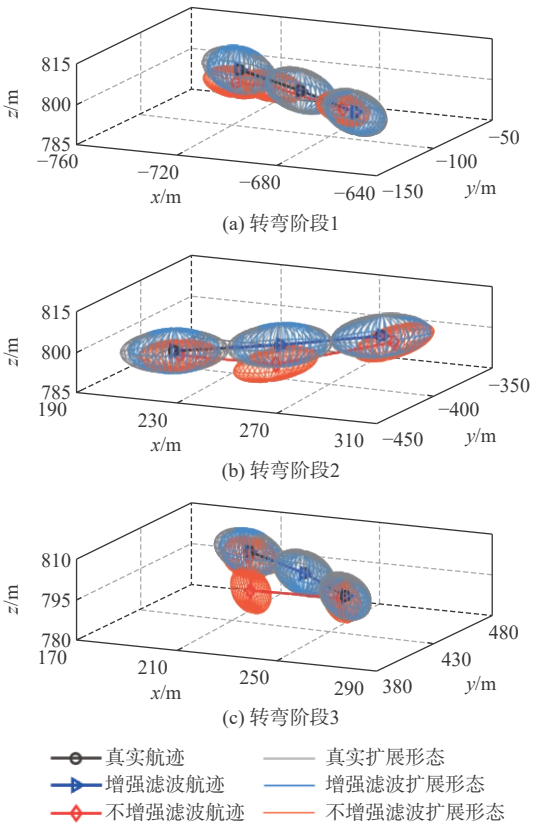


图 5 飞行器位置、扩展形态估计局部结果

Fig. 5 Partial estimation results for position, and extension shape of aircraft

间隔 4 帧分别给出了单次仿真运行中 3 次低空巡航转弯阶段(转弯阶段 1 取第 42 帧、46 帧、50 帧,转弯阶段 2 取第 152 帧、156 帧、160 帧,转弯阶段 3 取第 242 帧、246 帧、250 帧)的滤波估计局部结果。可以看到,受单站雷达对飞行器部分观测的影响,各阶段基于雷达原始稀疏量测的飞行器跟踪方法对其位置状态的估计结果始终是有偏的,对飞行器扩展形态的估计始终是不完整且小于真实扩展形态的,并在各检测失败时刻(第 46 帧、156 帧、246 帧)利用滤波预测值作为估计结果时会出现较大偏差,这是因为雷达稀疏量测分布在飞行器靠近雷达一侧的部分可观测表面附近,使得基于随机矩阵建模及滤波获得的飞行器质心位置估计(即量测集的滤波质心)和扩展形态估计(即量测集的滤波协方差椭圆体)会朝地面雷达方向产生偏移,且后者相比真实飞行器扩展形态会产生一定缩放^[26-27]。在融合飞行器导航信息辅助雷达量测增强后,由于量测数据量及空间完备性均大幅提升,同时雷达目标检测失败导致的量测缺失也得到了很好的补偿,因而飞行器位置、扩展形态等状态估计值分别能与其真实位置、扩展形态连续地产生较好重叠,验证了基于雷达量测增强的飞行器跟踪方法对飞行器位置、扩展形态等状态具有良好的、鲁棒的联合估

计性能。

其次, 对 3 种数据质量条件下空间坐标系各方向的受试飞行器速度分量进行估计, 图 6 取等间隔 4 帧分别展示了单次仿真运行的估计结果。直观地, 可以看到各数据质量条件下量测不增强时, 直接利用原始雷达稀疏量测滤波得到的各速度分量估计均存在较大的振荡, 尤其是在利用速度状态一步预测值作为估计值的雷达检测失败各时刻, 而在实施雷达量测增强后, 由于量测数据质量提升使得速度分量估计更为接近真实值且估计曲线更为平滑, 验证了基于雷达量测增强的飞行器跟踪方法对飞行器速度状态鲁棒估计的有效性。还可以看到, 相比于数据质量 2、3 条件下各速度分量估计的相近结果, 数据质量 1 条件下速度分量估计逼近真实的效果更佳, 不仅说明了良好的数据质量对于提升飞行器运动状态估计的重要性, 也说明飞行器导航

数据质量变化相比于雷达位置量测数据质量变化对飞行器速度估计结果的影响更大。

上述结论也可以由图 7 所示的飞行器运动状态估计 RMSE 曲线验证。因为各数据质量条件下, 量测不增强时滤波得到的飞行器位置、速度 RMSE 均明显高于利用增强雷达量测时的相应 RMSE, 且相比于后者的数值稳定性明显下降, 尤其是在人为设置的第 46 帧、第 156 帧、第 246 帧等雷达检测失败时刻。另外, 表 3 统计给出了蒙特卡罗仿真运行的飞行器运动状态估计 RMSE 结果。其中, 数据质量 1、2、3 条件下基于量测增强的跟踪方法相比量测不增强的飞行器位置估计 RMSE 分别提升了 99.71%、98.92% 及 98.94%, 速度估计 RMSE 分别提升了 96.18%、85.28% 及 85.13%, 而数据质量 1、2 相比于数据质量 3 条件下的量测增强后飞行器位置估计 RMSE 另经计算亦可知分别提升了 74.34% 及 0.66%, 速度估计 RMSE 则分别提升了 75.63% 及 5.65%, 这些进一步验证了基于雷达量测增强的跟踪方法改善飞行器运动状态估计精度的有效性, 并说明了飞行器导航数据质量变化相比于雷达位置量测数据质量变化对飞行器运动状态估计结果的影响更大。

最后, 对 3 种数据质量条件下的受试飞行器各扩展形态特征参数进行估计, 图 8 展示了 200 次蒙特卡罗仿真运行的参数估计平均结果, 表 4 给出了各参数的平均估计偏差 (AED) 结果。可以看到, 各数据质量条件下量测不增强时, 飞行器各扩展形态特征参数估计结果均与其真实值存在较大偏差, 而采用增强雷达量测可以在滤波稳定后得到较为准

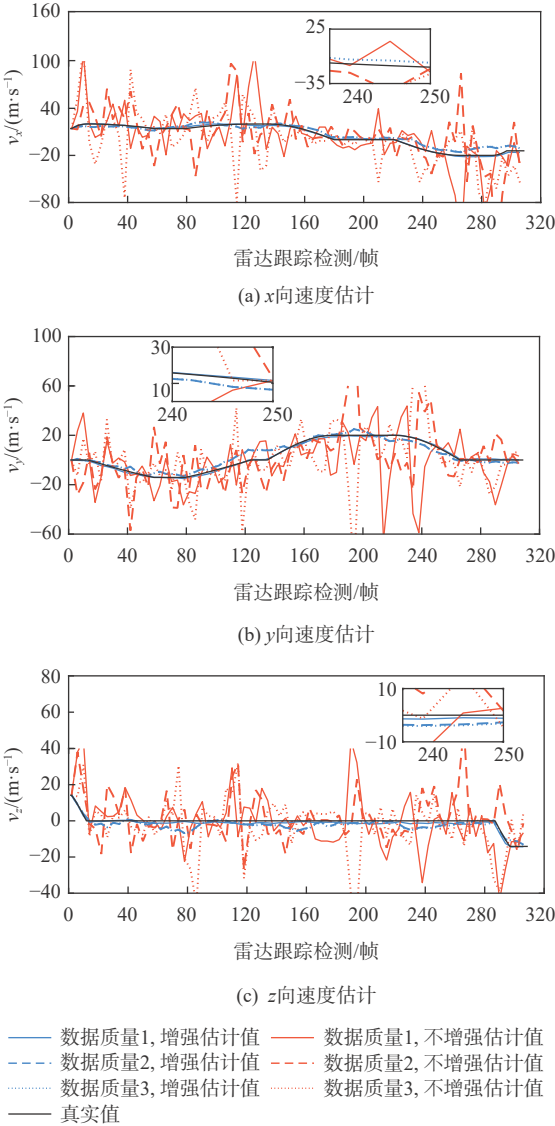


图 6 飞行器速度估计结果

Fig. 6 Estimation results for velocity of aircraft

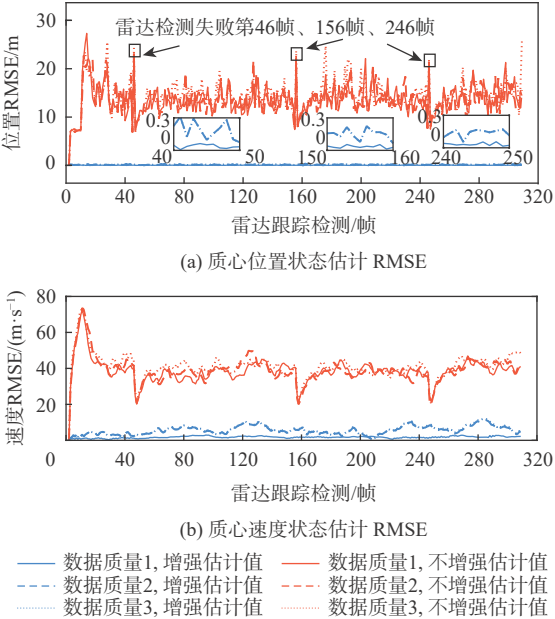


图 7 飞行器运动状态估计 RMSE

Fig. 7 Estimation RMSE for kinematic state of aircraft

表 3 飞行器运动状态估计 RMSE 比较

Table 3 Comparison of estimation RMSE for kinematic state of aircraft

数据质量	位置RMSE/m		位置RMSE提升/%	速度RMSE/(m·s ⁻¹)		速度RMSE提升/%
	不增强	增强		不增强	增强	
1	13.668	0.039	99.71	38.392	1.468	96.18
2	13.926	0.151	98.92	38.607	5.683	85.28
3	14.304	0.152	98.94	40.515	6.023	85.13

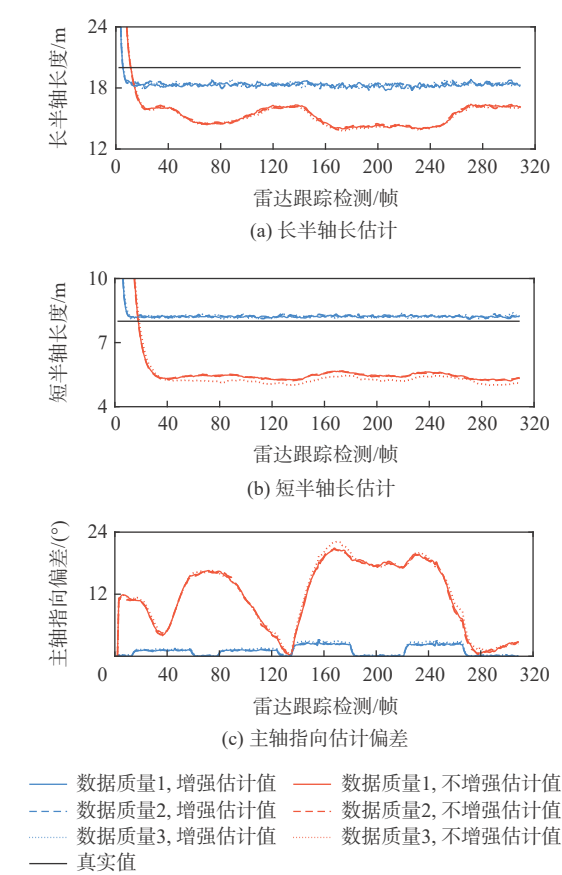


图 8 飞行器扩展形态参数估计平均结果

Fig. 8 Average estimation results for extension shape parameters of aircraft

确的扩展形态参数估计,其中数据质量 1、2、3 条件下对飞行器长半轴长参数估计的 AED 相比量测不增强时分别减少 65.22%、64.98% 及 65.23%,对飞行器短半轴长参数估计的 AED 则分别减少 80.88%、80.62% 及 83.57%,对主轴指向估计偏差的 AED 则分别减少 90.69%、90.42% 及 89.27%。还可以看到,3 种数据质量条件下各扩展形态特征参数

估计的结果特别相近,除短半轴长参数估计精度未严格随数据质量提升而提升外,其余参数估计精度均略有小幅提升,这一情况也可以由表 4 中不同数据质量条件下各扩展形态特征参数估计的 AED 差异很小反映。其中,数据质量 1、2 相比于数据质量 3 条件下的量测增强后飞行器长半轴长参数估计 AED 分别提升了 1.11% 及 0.41%,短半轴长参数估计 AED 则分别下降了 5.34% 及 6.31%,主轴指向角度估计 AED 则分别提升了 15.34% 及 12.60%。这些结果验证了基于雷达量测增强的跟踪方法改善飞行器扩展形态估计精度的有效性,并说明了雷达及导航数据质量变化均对飞行器扩展态估计结果的影响较小。

此外,由图 8(a)、图 8(b) 注意到,相比于飞行器短半轴长估计,3 种数据质量条件下,本文基于雷达量测增强的跟踪方法滤波稳定后对长半轴长的估计偏差略大,且估计值始终小于真实值,这是因为按本文雷达量测在飞行器可观测表面随机分布的假设,原始雷达稀疏量测在飞行器头/尾端点等局部区域分布的比例较小,而主要分布在占飞行器主体很大比例的机身区域,因此,按设计策略完成量测增强后,雷达量测在飞行器绕体轴的横向表面附近分布较多,但在飞行器沿体轴的纵向头/尾端点等表面附近局部区域的量测分布相对而言依然较少(见图 4),造成估计的飞行器扩展形态在体(主)轴方向或纵向尺寸相比真实形态更为偏短一些(见图 5),使得长半轴参数估计相比真实值呈现较大负偏差,这也是本文量测增强策略的固有不足之处。另外,由图 8(c) 看到,3 种数据质量条件下本文基于雷达量测增强的跟踪方法对飞行器主轴指向的估计偏差均远小于不增强时的估计偏差,但在各转弯阶段出现一定程度上升,这是由于受飞行器近似匀速运动模型

表 4 飞行器扩展形态参数估计 AED 比较

Table 4 Comparison of estimation AED for extension shape parameters of aircraft

数据质量	长半轴长/m		减少比/%	短半轴长/m		减少比/%	主轴指向/(°)		减少比/%
	不增强	增强		不增强	增强		不增强	增强	
1	-4.847	-1.686	65.22	-1.135	0.217	80.88	11.623	1.082	90.69
2	-4.848	-1.698	64.98	-1.130	0.219	80.62	11.655	1.117	90.42
3	-4.904	-1.705	65.23	-1.254	0.206	83.57	11.909	1.278	89.27

(式 (2))、飞行器近似时不变扩展形态演化模型 (式 (5)) 及飞行器巡航速度方向与主轴指向假设一致的共同影响, 飞行器主轴空间指向在转弯机动航段的平均估计偏差相比巡航平直飞行航段会出现增加。总体而言, 本文基于雷达量测增强的跟踪方法对飞行器扩展形态特征参数估计性能的可观提升仍具有不可忽视的意义。

5.3.3 计算消耗时间

在相同仿真环境下运行本文基于雷达量测增强与否的飞行器跟踪方法, 其中, 笔记本平台配置为 64 位 win10 操作系统, 内存 32 GB, Intel(R) Core (TM) i7-11800H @2.30 GHz 处理器。在不采用加速技术情况下, 以数据质量 2 条件下的计算消耗时间分析为例, 统计并比较 200 次蒙特卡罗仿真运行的雷达量测增强与否的平均单帧运算时间、平均单帧量测数, 结果如图 9 及表 5 所示。

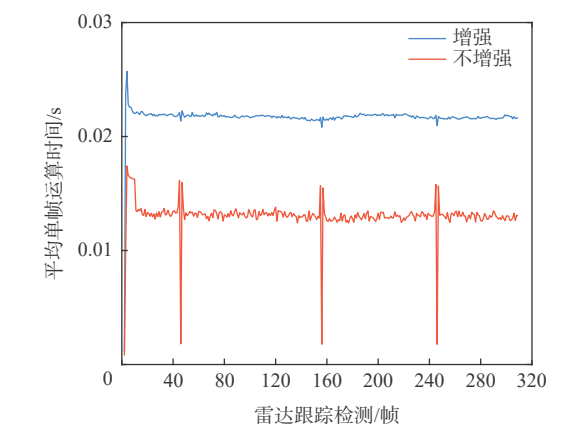


图 9 平均单帧运算时间曲线

Fig. 9 Curves of average computational time per scan

表 5 平均单帧运算时间和平均单帧量测数比较

Table 5 Comparison of average single-frame computational times and average single-frame measurement count

方法	平均单帧运算时间/s	平均单帧量测数/个
不增强	0.013	6.419
增强	0.021	63.853

可以看到, 基于雷达量测增强的跟踪方法的平均单帧运算时间整体高于不增强时方法, 但在前 10 帧时, 二者运算耗时均较高, 另外, 量测不增强时方法在人为设置的第 46 帧、第 156 帧、第 246 帧等雷达检测失败时刻出现耗时低值, 而在其前、后 2 个人为设置雷达稳定检测成功帧均出现耗时峰值。分析这些现象原因, 在前 10 帧时间雷达检测成功以实现目标跟踪稳定起批的实验假设条件下, 2 种方法需执行完整的滤波流程。其中, 基于雷达量测增强的跟踪方法还额外增加了量测增强计算过程; 而在各雷达检测失败帧, 量测不增强时方法

滤波过程中仅需执行状态一步预测, 无需执行量测增强、状态测量更新等计算过程, 在其前、后 2 个设置的稳定检测成功帧则需完整执行这 3 个计算过程, 其余检测帧由于按检测概率 ρ 随机生成雷达检测成功/失败现象, 故执行 1~3 个计算过程不等, 因而 2 种方法的平均单帧运算时间呈现出图 9 所示的耗时变化曲线。

由表 5 统计结果可见, 基于雷达量测增强的跟踪方法的单帧平均耗时相比量测不增强时增加 61.54%, 但二者均远小于雷达采样周期, 从而保证了跟踪方法的实时性。分析其原因, 一方面, 基于随机矩阵建模的 ETT 方法滤波方程简单, 相比于传统点目标跟踪方法仅是增加了目标扩展形态的滤波过程, 且该过程具有的封闭表达式利于计算机执行, 无需额外借助优化技术, 因而计算复杂度增加较少; 另一方面, 量测增强后平均单帧量测数相比量测不增强时虽增加 894.75%, 但由 3.2 节分析可知, 雷达量测增强具有较低的计算复杂度, 且由 4.2 节滤波流程可知, 增强雷达量测最终仅会使得滤波测量更新过程中的量测均值 $\bar{z}_k$ 、量测散射矩阵 $\bar{Z}_k$ 这 2 个批处理公式增加运算耗时, 且该耗时相对滤波总体耗时十分有限。综合而言, 本文基于雷达量测增强的跟踪方法具有较低的计算成本。

6 结 论

1) 扩展型浮低空飞行器飞行测试中运动航迹及姿态的变化会引起飞行器部分可观测表面的改变, 造成地面单站高分辨雷达量测数据稀疏且空间不完备, 制约了受试飞行器跟踪精度。

2) 融合合作式飞行器导航信息辅助雷达稀疏量测增强, 有效扩充了雷达量测数据量及提升了其空间完备性, 降低了雷达稀疏量测或雷达检测失败对受试飞行器跟踪精度的不利影响。

3) 利用增强雷达量测数据并基于随机矩阵模型及贝叶斯滤波框架, 可实现飞行器运动状态及扩展形态特征参数的良好估计, 从而提升了受试飞行器跟踪精度及其飞行态势感知精度。

4) 相比于雷达仅利用飞行器导航位置数据作为外引导跟踪受试飞行器, 融合导航位置、姿态数据以辅助雷达量测增强的飞行器跟踪方法进一步增加了在状态估计精度、鲁棒性及多源测控信息融合利用度等方面的优势, 且计算成本适中, 具有潜在的工程应用价值。

本文假设雷达量测随机分布于飞行器可观测表面, 下一步考虑结合受试飞行器电磁散射特性, 在跟踪方法中纳入飞行器散射中心的实际统计分布特征。

参考文献 (References)

[1] Huang X. The small-drone revolution is coming: scientists need to ensure it will be safe[J]. *Nature*, 2025, 637(8044): 29-30.

[2] 杨雪榕, 杨雅君, 朱俊. 飞行器试验及数据统计处理方法[M]. 北京: 国防工业出版社, 2024: 7-8.

Yang X R, Yang Y J, Zhu J. Aircraft experiment and data statistical processing method[M]. Beijing: National Defense Industry Press, 2024: 7-8(in Chinese).

[3] 翟嘉琪, 杨希祥, 邓小龙, 等. 不确定风场下平流层浮空器全局路径规划[J]. *北京航空航天大学学报*, 2023, 49(5): 1116-1126.

Zhai J Q, Yang X X, Deng X L, et al. Global path planning of stratospheric aerostat in uncertain wind field[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2023, 49(5): 1116-1126(in Chinese).

[4] Mahler R. PHD filters for nonstandard targets, I: extended targets[C]//*Proceedings of the 12th International Conference on Information Fusion*. Piscataway: IEEE Press, 2009: 915-921.

[5] Granström K, Natale A, Braca P, et al. Gamma Gaussian inverse Wishart probability hypothesis density for extended target tracking using X-band marine radar data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(12): 6617-6631.

[6] Du H C, Xie W X, Liu Z X, et al. Track-oriented marginal Poisson multi-Bernoulli mixture filter for extended target tracking[J]. *Chinese Journal of Electronics*, 2023, 32(5): 1106-1119.

[7] Koch J W. Bayesian approach to extended object and cluster tracking using random matrices[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2008, 44(3): 1042-1059.

[8] Feldmann M, Fränken D, Koch W. Tracking of extended objects and group targets using random matrices[J]. *IEEE Transactions on Signal Processing*, 2011, 59(4): 1409-1420.

[9] Lan J, Li X R. Tracking of extended object or target group using random matrix: new model and approach[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2016, 52(6): 2973-2989.

[10] Hu Q, Ji H B, Zhang Y Q. Tracking of maneuvering non-ellipsoidal extended target with varying number of sub-objects[J]. *Mechanical Systems and Signal Processing*, 2018, 99: 262-284.

[11] Tuncer B, Orguner U, Özkan E. Multi-ellipsoidal extended target tracking with variational Bayes inference[J]. *IEEE Transactions on Signal Processing*, 2022, 70: 3921-3934.

[12] Baum M, Hanebeck U D. Random hypersurface models for extended object tracking[C]//*Proceedings of the IEEE International Symposium on Signal Processing and Information Technology*. Piscataway: IEEE Press, 2010: 178-183.

[13] Zea A, Faion F, Baum M, et al. Level-set random hypersurface models for tracking nonconvex extended objects[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2016, 52(6): 2990-3007.

[14] Sun L F, Liu Z Y, Zhang D K, et al. Maneuvering extended object tracking based on generalized conversion nonlinear filtering with random weight cubature rule sampling[J]. *Measurement*, 2025, 251: 117231.

[15] Ma T L, Sun H Y, Wang K X, et al. A novel set-member filter for maneuver non-ellipsoid extended target tracking with unknown but bounded noises[J]. *IEEE Sensors Journal*, 2025, 25(8): 13708-13718.

[16] Baum M, Faion F, Hanebeck U D. Modeling the target extent with multiplicative noise[C]//*Proceedings of the 15th International Conference on Information Fusion*. Piscataway: IEEE Press, 2012: 2406-2412.

[17] Yang S S, Baum M. Tracking the orientation and axes lengths of an elliptical extended object[J]. *IEEE Transactions on Signal Processing*, 2019, 67(18): 4720-4729.

[18] Kumru M, Ozkan E. Three-dimensional extended object tracking and shape learning using Gaussian processes[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2021, 57(5): 2795-2814.

[19] Xu M D, Yang C Q, Cao X M, et al. Irregular extended target tracking with unknown measurement noise covariance[J]. *Signal Processing*, 2024, 225: 109600.

[20] Garcia N, Fascista A, Coluccia A, et al. Cramér-Rao bound analysis of radars for extended vehicular targets with known and unknown shape[J]. *IEEE Transactions on Signal Processing*, 2022, 70: 3280-3295.

[21] 陈振, 李翠芸, 李想. B 样条曲面三维扩展目标跟踪算法[J]. *西安电子科技大学学报(自然科学版)*, 2023, 50(2): 101-111.

Chen Z, Li C Y, Li X. Algorithm for tracking the 3D extended target based on the B-spline surface[J]. *Journal of Xidian University (Natural Science)*, 2023, 50(2): 101-111(in Chinese).

[22] 陈辉, 边斌超, 连峰, 等. 基于 Transformer 复杂运动辨识的机动星凸形扩展目标跟踪方法[J]. *雷达学报*, 2024, 13(3): 629-645.

Chen H, Bian B C, Lian F, et al. A novel method for tracking complex maneuvering star convex extended targets using Transformer network[J]. *Journal of Radars*, 2024, 13(3): 629-645(in Chinese).

[23] Chen H, Bian B C, Lian F, et al. Maneuvering extended target tracking method based on transformer network[J]. *Measurement*, 2025, 240: 115474.

[24] Jiao H, Yan J K, Pu W Q, et al. Wideband sensor resource allocation for extended target tracking and classification[J]. *IEEE Transactions on Signal Processing*, 2025, 73: 55-66.

[25] Zhang X X, Lan J. Measurement combination estimator for multi-sensor extended object tracking using random matrix[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2024, 60(1): 698-715.

[26] Hoher P, Wirtensohn S, Baur T, et al. Extended target tracking with a lidar sensor using random matrices and a virtual measurement model [J]. *IEEE Transactions on Signal Processing*, 2022, 70: 228-239.

[27] Cao X M, Lan J, Liu Y S, et al. Tracking of rectangular object using key points with regionally concentrated measurements[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(6): 5312-5327.

[28] 梁苑, 戚国庆, 陈烨, 等. 不完全量测下事件触发水面扩展目标跟踪[J]. *兵工学报*, 2024, 45(4): 1219-1228.

Liang Y, Qi G Q, Chen Y, et al. Event-triggered surface extended target tracking with intermittent measurements[J]. *Acta Armamentarii*, 2024, 45(4): 1219-1228(in Chinese).

[29] Granstrom K, Bramstang J. Bayesian smoothing for the extended object random matrix model[J]. *IEEE Transactions on Signal Processing*, 2019, 67(14): 3732-3742.

[30] Wang S H, Men C K, Li R X, et al. Extended target-tracking algorithm based on variational Bayes and axis estimation theory[J].

IEEE Transactions on Instrumentation and Measurement, 2025, 74: 8507216.

[31] 张晔, 侯毅, 欧阳克威, 等. 单变量序列数据分类方法综述[J]. 系统工程与电子技术, 2023, 45(2): 313-335.

Zhang Y, Hou Y, Ouyang K W, et al. Survey of univariate sequence data classification methods[J]. Systems Engineering and Electronics, 2023, 45(2): 313-335(in Chinese).

[32] 李春辉, 马健, 杨永建, 等. 低复杂度自适应容积分卡尔曼滤波算法[J]. 北京航空航天大学学报, 2022, 48(4): 716-724.

Li C H, Ma J, Yang Y J, et al. Low-complexity adaptive cubature Kalman filter algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(4): 716-724(in Chinese).

Extended low-altitude aircraft tracking via navigation-assisted radar measurement augmentations

LI Kunkun¹, LI Kebo^{2, 3, *}, CAO Rui¹, LIU Yuanhe⁴, HU Shuang¹

- (1. Unit 63660 of Chinese People’s Liberation Army, Luoyang 471099, China;
2. College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410072, China;
3. State Key Laboratory of Space System Operation and Control, Changsha 410073, China;
4. Aviation University of Air Force, Changchun 130000, China)

Abstract: In the flight test of cooperative aircraft floating at low altitudes, a single-station high-resolution radar can get sparse (even missing) and spatially incomplete measurements, while the radar observes a partial surface of the aircraft. This will lead to poor tracking accuracy for the tested aircraft. To solve this problem, this paper proposes a tracking method using augmented radar measurements by fusing aircraft navigation information. First, a symmetric positive definite random matrix is used to approximate the aircraft’s three-dimensional extension shape, which is roughly ellipsoidal, based on the structurally extended characteristic of the aircraft identified by the high-resolution radar. Subsequently, the partially observed aircraft model and the radar measurement model are both established. Further, multiple strategies are designed to augment sparse radar measurements using assisted aircraft navigation information, such that the radar measurement data is improved in both amount and spatial completeness. Finally, an implementation of aircraft state estimation under partial observation is given within the Bayesian filtering framework. The outcomes of the simulation experiment show that the proposed method improves the tested aircraft’s tracking accuracy and, consequently, awareness accuracy for its flight condition. Moreover, the proposed method has the advantages of high fusion usage of multi-source measurement information and a low computational cost.

Keywords: low-altitude aircraft; extended target tracking; random matrix; partial observation; sparse measurements; data augmentation

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0361

民用无人系统运行场景架构设计与建模技术

郭泰*, 安一平, 舒振杰, 蒋觉义, 江雨航

(中国航空综合技术研究所, 北京 100028)

摘 要: 民用无人系统运行场景是全空间无人体系及低空经济产业价值的汇聚地, 是推动新产业规模化发展的关键, 同时作为典型的复杂体系, 具有边界动态、多级涌现、利益权衡、持续演化等特征, 必须按照体系工程的新思维开展架构设计, 确保体系安全、高效运行。通过对民用无人系统运行场景架构设计与建模技术研究, 提出运行场景“6+6”静动态要素元模型, 从能力架构、运行架构、系统架构、标准架构对运行场景进行多视角架构设计和建模, 形成一套从要素识别、架构设计到架构建模的完整方法, 从而为新型基础设施建设、无人系统研制及选型、政策法规和标准规范的制定提供指导, 为无人系统运行场景的搭建和产业体系化发展提供技术支撑。

关 键 词: 低空经济; 全空间无人体系; 运行场景; 架构建模; 无人系统

中图分类号: V279; F562

文献标志码: A

文章编号: 1001-5965(2026)09-3052-14

“全空间无人体系”和“低空经济”是国家在新业态新领域的战略部署, 是发展战略新兴产业和新质生产力的典型代表。中共中央办公厅、国务院办公厅《关于完善市场准入制度的意见》^[1]中明确提出“率先推动海陆空全空间智能无人体系应用和标准建设”, 低空经济也连续两年被写入中央政府工作报告中, 近期中央还出台了包括完善监管、促进消费、健全体系、发展产业、加强立法、保障空域等在内的多项政策助推低空经济高质量发展。在地方层面, 各省市地方政府加速布局, 截至 2024 年底, 各地共出台全空间无人体系及低空经济相关政策文件 225 部, 从行动指南、规划条例、扶持政策等不同视角加速基础设施和应用场景的建设落地^[2]; 2025 年全国 30 省份将低空经济直接写入政府工作报告, 囊括了产业升级、先导示范、设备更新、基础设施建设等诸多发力点。从产业培育和发展角度, 无人系统场景涵盖无人系统全生命周期, 包括无人系统研制、验证(含检测认证)、运行(正常、非正常、维修保障)、报废处置等环节, 其中, 无人系统

运行场景是指对无人机、无人车、无人船、电动垂直起降航空器 (electric vertical takeoff and landing aircraft, eVTOL) 等海陆空多种无人系统及其配套基础设施一体运行和监管的新型综合应用体系, 运行场景的突破将是全空间无人体系及低空经济取得规模化发展的核心, 是产业价值的主要聚集区, 也是带动产业链的核心增长点。

一方面, 运行场景建设是让技术和产品走向产业化的关键载体^[3], 场景驱动能够引领开创性技术的突破并推动空域资源管理改革^[4]; 另一方面, 针对运行场景当前面临的普及率较低、消费潜力尚未充分激发等体系性问题^[5], 亟需形成统一的架构, 推动基础设施建设、无人系统研制、政策法规和标准规范等协同发展。当前国内外对于全空间无人体系及低空经济的相关研究大多集中在发展路径^[6]、飞行器关键技术^[7]、空域管理^[8]、立体交通^[9]、基础设施^[10]、信息安全^[11]、产业生态^[12]、法律保障^[13]等方面。对于运行场景的总体性、架构性技术研究则略显单薄, 主要以论述的方式对于场景模式^[14]、场景拓展^[15]、

收稿日期: 2025-06-14; 录用日期: 2025-10-04; 网络出版时间: 2025-11-28 17:45

网络出版地址: link.cnki.net/urlid/11.2625.V.20251128.1603.003

* 通信作者. E-mail: guotai101@126.com

引用格式: 郭泰, 安一平, 舒振杰, 等. 民用无人系统运行场景架构设计与建模技术 [J]. 北京航空航天大学学报, 2026, 52 (9): 3052-3065.

Guo T, An Y P, Shu Z J, et al. Architecture design and modeling technology for civil unmanned system operation scenarios [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3052-3065 (in Chinese).

场景赋能^[16]、场景驱动新质生产力发展^[17]等进行阐释。

新兴产业发展早期经常面临运行场景不明确或受限等情况^[18], 无人系统运行场景作为典型的复杂系统之系统(system of systems, SoS), 其子系统之间相互依赖、相互影响、连接复杂^[12], 且各类子系统要素之间的运行规则仍处于不断探索阶段。需要从体系工程角度、结合架构理论, 对运行场景进行多视角分析, 明确运行场景的内涵边界、总体架构, 从而为运行场景中各类新型基础设施、无人设备、法规标准等建设提供顶层指引。因此, 本文从运行场景基本逻辑出发, 对运行场景静态、动态要素进行研究, 并将体系工程与典型架构框架^[19]相结合, 对运行场景总体架构进行多视角设计。结合相关建模工具和方法, 对运行场景架构进行建模, 形成无人系统运行场景架构模型, 为后续基于模型开展各类系统或产品研制的运行概念、需求捕获及功能分析等关键政策法规和标准规范建设提供基础。

1 场景分析

1.1 场景要素分析

“场景”一词在不同的领域和背景下, 具有不同的涵义, 在分布式交互仿真标准 IEEE 1278-1993 中, 场景被定义为对初始条件和重大事件时间线的描述^[20]。美国国防部高层级架构术语中强调场景应识别主要实体及其能力、行为及随着时间推移与所有相关环境条件的相互作用^[21]。北约科学与技术组织建模与仿真 053 小组将场景定义为在与相关事件有关的指定时间框架内对假设或真实区域、环境、手段、目标和事件的描述^[22]。根据国际标准化组织(International Organization for Standardization, ISO)从需求工程角度对运行场景的定义, 运行场景“描述想象中的事件序列, 包括产品或服务与其环境和用户之间的互动, 以及产品或服务组件之间的互动”^[23]。目前场景分析在民机需求捕获和确认中是已被广泛认可的技术方法^[24], 但系统工程或建模仿真领域对场景的定义, 主要针对特定目标系统(system of interest, SOI), 从用户和运行视角描述 SOI 与外部环境或利益相关方的交互关系。全空间无人体系及低空经济无人系统的运行场景是典型的系统之系统, 包括多个系统在内的一组相互作用的系统或系统要素, 可提供任何一个组成系统都无法单独完成的独特能力。因此, 通过借鉴其他领域对场景的定义, 结合无人系统运行场景作为系统之系统的特征, 对运行场景定义如下: “特定区域内各系统、服务及主体与外部环境、目标和用户的互

动, 以及各系统、服务及主体之间的互动”, 该定义并不仅着眼于特定系统或服务层级, 而是站在体系层级给出规范定义。

根据对不同组织“场景”定义的要素分析, 可划分为静态和动态 2 类, 静态类重点描述场景内的对象及环境, 动态类重点描述场景过程, 在此基础上结合典型架构框架元模型和民用飞机运行场景建模经验, 结合无人系统运行场景实际对典型场景要素进行精化, 提出运行场景“6+6”要素元模型, 如图 1 所示, 具体如下。

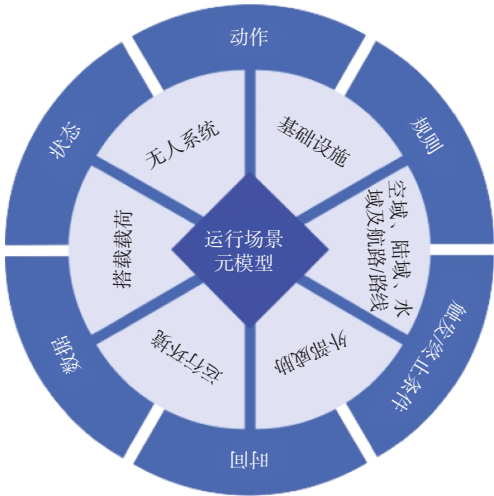


图 1 无人系统运行场景要素元模型

Fig. 1 Element meta-model of unmanned systems operational scenarios

静态要素。主要体现运行场景中各类系统、服务和主体, 主要包括:

- 1) 无人系统: 包括低空航空器如各类无人机、eVTOL 等新构型新能源航空器、无人车、无人船等无人系统产品;
- 2) 基础设施: 包括数据、网络、物理等不同类型的新型基础设施, 作为无人系统运行场景的使能系统或服务;
- 3) 搭载载荷: 包括人员、货物、各类探测感知设备等;
- 4) 空域、陆域、水域及航路/路线: 指运行场景所处的空域、陆域、水域条件及所划设的运行航路或路线;
- 5) 运行环境: 包括地理环境、气象环境、电磁环境、产业环境等;
- 6) 外部威胁: 指可能对运行场景产生安全风险的外部非合作目标如黑飞无人机、热气球、风筝、障碍物、漂浮物等。

动态要素。静态要素构成了运行场景的基础结构, 动态要素驱动场景内静态要素的交互, 主要包括:

- 1) 时间: 指场景的时间线, 包括开始时间、结束时间等;
- 2) 规则: 指静态要素行为应遵循的具体规则, 通常以法规、政策、标准等形式进行约束;
- 3) 状态: 主要用于描述各类静态要素在一定时间范围内维持不变的行为;
- 4) 动作: 指静态要素在场景中的具体行为, 包括自身行为和对其他主体或对象的行为;
- 5) 数据: 指静态要素执行动作或状态转换过程中交互、产生的数据;
- 6) 触发/终止条件: 用于驱动场景动作或要素不同状态的转换。

上述根据对运行场景的要素分析, 提出了场景要素的元模型定义, 为规范化定义无人系统运行场景提供了理论基础。

1.2 场景框架定义

基于本文构建的场景要素元模型, 从区域发展全空间无人体系及低空经济角度, 对运行场景框架

进行定义, 先对特定区域下的无人系统运行场景的特点进行分析。

- 1) 区域管辖: 各地方政府发展全空间无人体系及低空经济主要以管辖区域范围开展, 一般划分为省市级、地市级、县级甚至乡镇级等, 需要对不同区域范围的场景进行统筹考虑。
- 2) 行业融合: 按照不同区域内不同行业及对应的职能管理, 各类运行场景在特定区域共同运行但接受不同部门管理。
- 3) 环境差异: 不同区域带来的基础设施条件不同, 运行环境不同, 空域条件不同, 外部威胁不同, 同行业场景在不同区域需要因地制宜。
- 4) 创新驱动: 全空间无人体系及低空经济运行场景中各类静态要素和动态要素均需要技术创新、产品创新、规则创新的有效支持。

根据以上特征分析, 站在特定区域对全空间无人体系及低空经济运行场景框架进行构建, 如图 2 所示。主要包括 4 个层面。

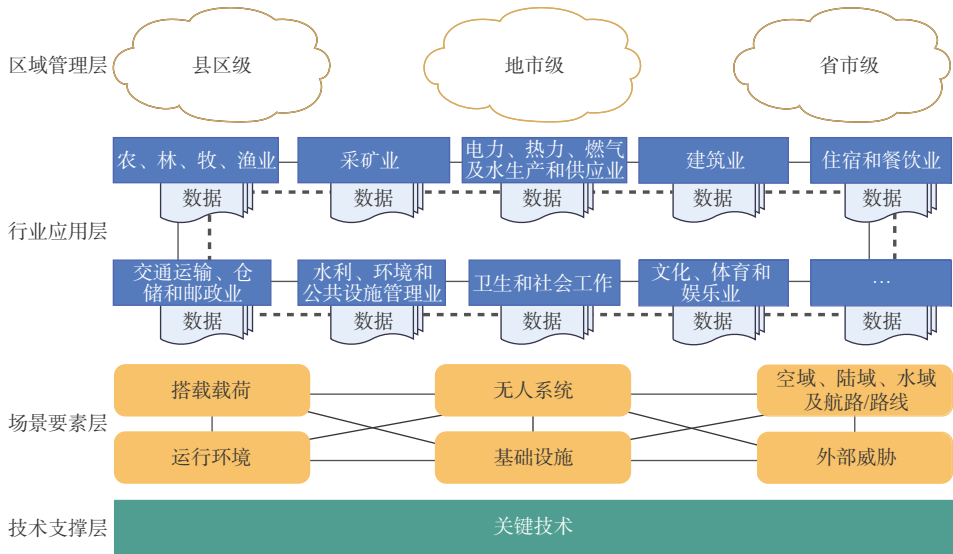


图 2 区域发展视角下运行场景框架

Fig. 2 Framework of operational scenarios from perspective of regional development

- 1) 技术支撑层: 指支撑场景静态要素本身产品创新及动态运行创新的关键技术发展, 依赖于全国乃至全球新技术突破及应用。
- 2) 场景要素层: 指不同地方特定区域范围内依据场景元模型的具体特征化, 如空域航路要素应结合特定区域当前条件及与空域管理机构协商后具体定义。
- 3) 行业应用层: 指国民经济不同行业通过对场景要素的体系化组织形成的经济、社会活动, 这也是场景驱动下全空间无人体系及低空经济的核心。
- 4) 区域管理层: 根据特定区域范围及行政区划

的不同, 对所管辖范围内采取各类关键技术不同场景要素在不同行业的应用进行统筹设计、运营和管理。

2 架构分析

当前有关全空间无人体系及低空经济的架构分析大多停留在产业层面, 描述场景、技术与空域间的共同演化逻辑^[4], 从供给侧、需求侧和监管侧角度对总体架构进行划分, 将产业、技术和生产要素划归为供给侧, 将场景划归为需求侧, 政策、法规、标准等划归为监管侧^[25]。本文结合美国国防部架构

框架 (Department of Defense architecture framework, DoDAF)、统一架构框架 (unified architecture framework, UAF) 等典型架构框架, 提出以“政策法规和标准规范”为核心, “目标-运行-实现”三位一体的架构元模型(见图 3)。其中, “目标”对应能力视角 (capability viewpoint, CV), 表示运行场景作为体系能够在一定区域范围内实现的可持续性目标; “运行”对应运行视角 (operational viewpoint, OV), 表示运行场景中各类元素的动态行为, 核心包括运行活动和运行状态等; “实现”对应系统视角 (system viewpoint, SV) 和服务视角 (service viewpoint, SvcV), 表示为实现目标所必备的系统组成及需要的服务。“政策法规和标准规范”对应标准视角 (standards viewpoint, StdV), 既包括当前政策法规和标准规范, 也包括未来的需求预测, 从而保证运行场景的安全、高效、兼容等目标的实现。此外, 在运行和实现层面, 存在各类数据、信息及资源的交互, 可通过数据与信息视角 (data and information viewpo-



图 3 运行场景架构元模型

Fig. 3 Architecture meta-model of operational scenarios

int, DIV)进行表达。

2.1 能力架构

结合运行场景架构元模型, 围绕全空间无人体系及低空经济“安全底线”、“运行效能”这 2 大目标构建能力架构, 如图 4 所示, 主要包括以下 7 种能力。

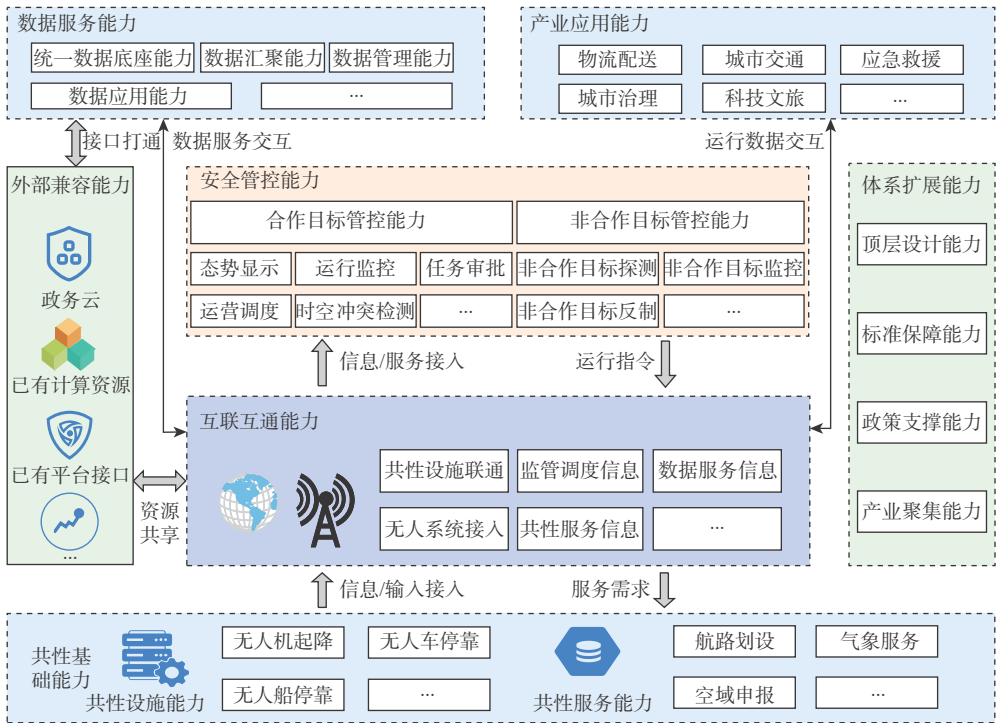


图 4 无人系统运行场景能力架构

Fig. 4 Capability architecture of unmanned systems operational scenarios

- 1) 共性基础能力: 包括共性设施能力和共性服务能力, 为无人设备提供起降停靠等功能支撑, 同时保障无人设备执行任务所需的各类服务。
- 2) 互联互通能力: 在运行场景内保障基础设施与无人设备、基础设施之间的互联互通, 通过数据服务交互支撑数据服务能力, 以运行数据交互支撑产业应用能力, 与外部兼容能力共同作用实现资源

- 共享。
- 3) 安全管控能力: 对运行场景内合作目标实行态势、任务、调度等管控, 对非合作目标进行探测、监控和反制, 保障场景安全稳定运行。
 - 4) 数据服务能力: 实现数据汇聚、数据管理、数据应用等服务能力, 能够实现体系内统一的数据底座和数据服务。

- 5) 产业应用能力: 依托城市治理、物流配送、科技文旅等不同领域的应用场景, 实现产业链结构升级, 打造新领域新业态。
- 6) 外部兼容能力: 实现与已有计算资源、平台接口、政务云等系统的兼容和资源共享, 并与数据服务能力共同实现接口打通。
- 7) 体系拓展能力: 通过顶层设计、标准保障、政策支撑、产业聚集等方式实现体系的更新迭代, 使其适应技术进步和经济社会发展的相关要求。

2.2 运行架构

根据能力架构所构建的体系目标, 运行架构主要从运行视角描述体系的动态过程, 如图 5 所示,

- 主要特点如下。
- 1) 数据和运行管理“双驱动”: 一方面串联无人系统、新型基础设施、航路/路线划设及通信、感知、服务网, 另一方面整合政府、运营商等外部资源。
- 2) 航路、路线划设作为独立模块: 空域、水域、水域及相应的航路、路线作为运行场景的核心静态要素, 在运行架构中作为独立模块, 强化对特定区域资源的统一规划和管理。
- 3) 政策和标准的持续迭代: 在体系运行过程中, 需要遵从现有政策法规和标准规范的约束, 同时通过不断实践形成新政策法规的建议和标准规范, 推动体系持续优化完善。

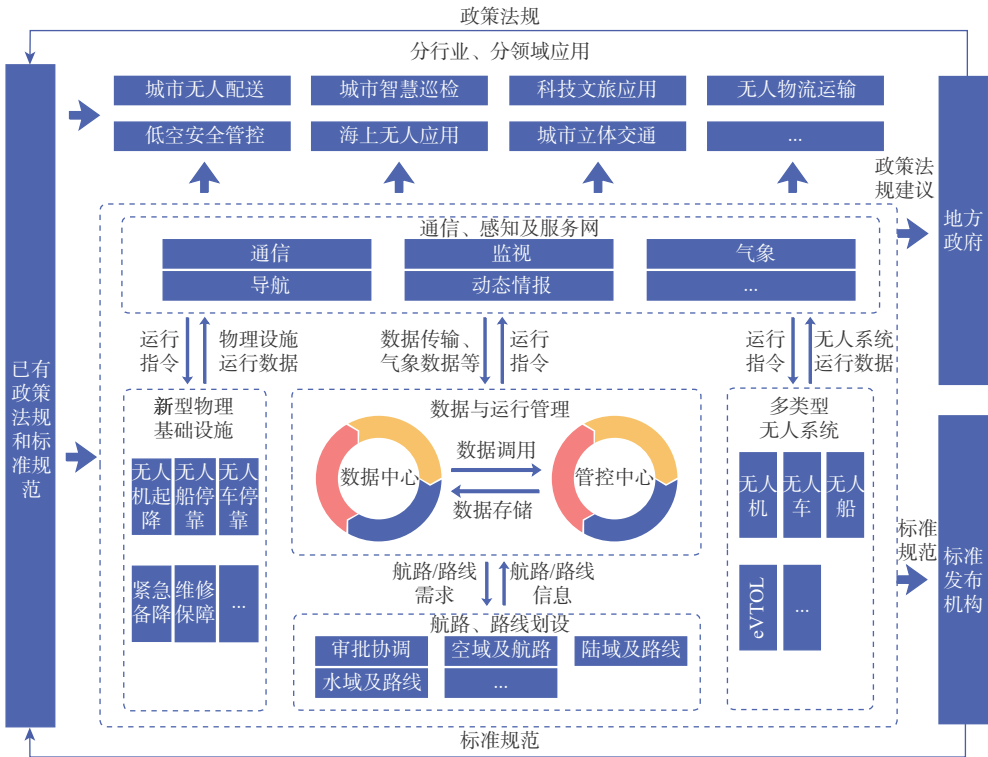


图 5 无人系统运行场景运行架构

Fig. 5 Operational architecture of unmanned systems operational scenarios

2.3 系统/服务架构

综合能力架构和运行架构需要, 从体系实现角

度搭建系统/服务架构, 如图 6 所示, 主要包括 4 个部分。

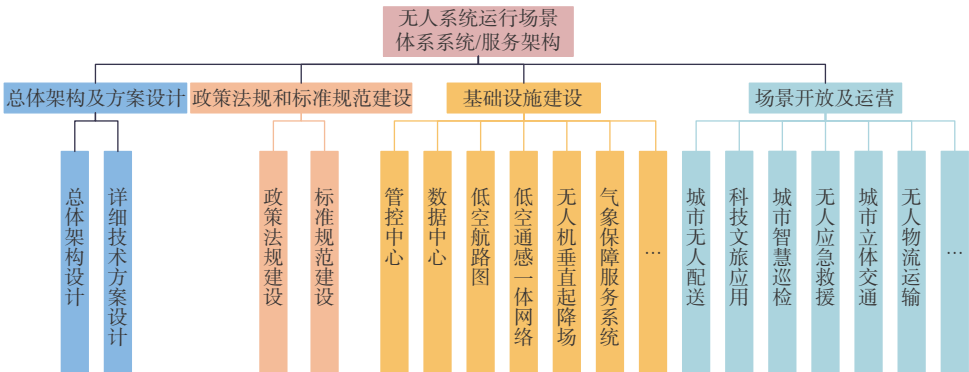


图 6 无人系统运行场景系统架构

Fig. 6 Systems architecture of unmanned systems operational scenarios

- 1) 总体架构及方案设计: 从顶层视角对特定区域内复杂体系开展总体架构设计, 并开展详细方案及场景设计, 用于指导后续建设及体系升级。
- 2) 政策法规和标准规范建设: 围绕基础设施建设、场景开放及运营管理研究政策建议及标准制定, 与架构及方案设计共同支撑体系拓展能力。
- 3) 新型基础设施建设与实施: 保障运行场景运行活动的正常运转, 为其余 6 种能力提供支撑。其中, 基础设施的建设可根据不同区域的特点、性能及经济可承受性等选取不同技术路径的基础设施,

- 同时由于建设主体的不同, 部分基础设施如通信、监视、导航、气象等可通过服务的方式提供。
- 4) 场景开放及运营: 依托体系设计、政策、标准、基础设施, 鼓励不同行业运营主体开展相应场景建设, 前期主要由政府引导、培育, 后期主要由市场驱动。

2.4 标准架构

标准架构是保障全空间无人体系及低空经济场景安全、有序运行的核心, 如图 7 所示, 主要包括 4 个方面的内容。

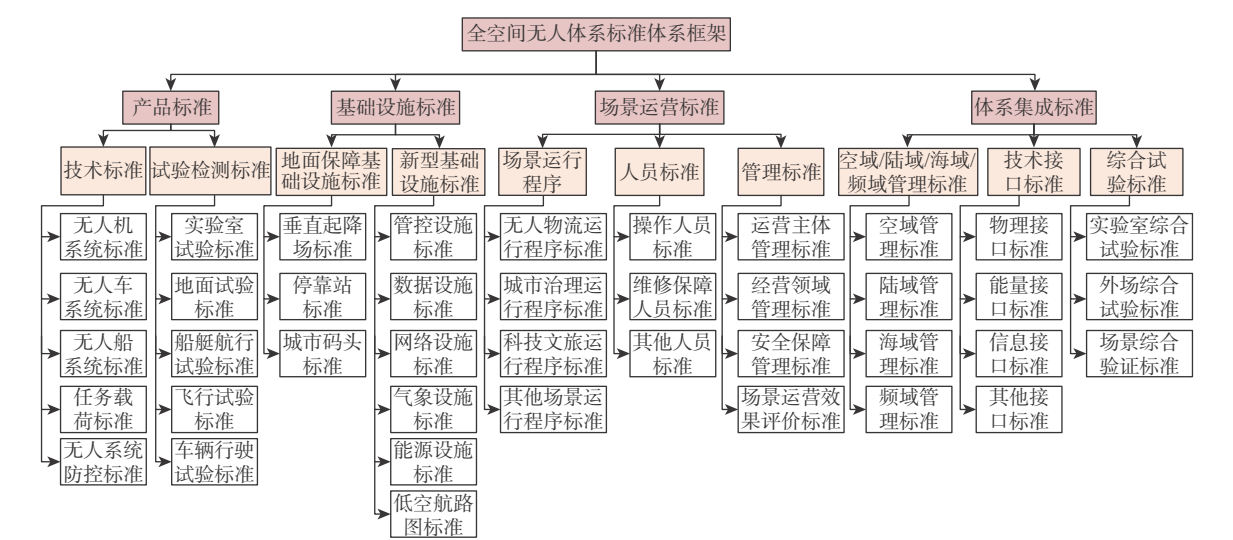


图 7 无人系统运行场景标准架构

Fig. 7 Standards architecture of unmanned systems operational scenarios

- 1) 产品标准: 包括无人系统产品本身的技术标准及相应的试验验证标准。
- 2) 基础设施标准: 主要包括地面保障基础设施和各类新型基础设施标准。
- 3) 场景运营标准: 主要包括运行程序、人员标准、管理标准 3 类, 规范各领域场景的有序运行。
- 4) 体系集成标准: 该类标准为全空间无人体系和低空经济的核心标准, 重点对空域/陆域/海域/频域管理、技术接口、综合试验进行规范。

3 架构建模

针对第 2 节构建的能力架构、运行架构、系统/服务架构、标准架构, 参考相关运行场景建模方法^[26], 综合 DoDAF2.0、UAF 架构框架及系统建模语言(systems modeling language, SysML)提出统一的架构建模方法, 在建模前应通过全景视图(all viewpoint, AV)对全空间无人体系及低空经济所涉及的概念和术语如数据底座等进行规范化定义, 建模总体技术路径如图 8 所示。

3.1 能力视角

通过 CV-2 视图对全空间无人体系及低空经济能力架构进行构建, 形成 7 大能力域对应的能力分解结构, 具体如图 9 所示。

根据 CV-2 能力分解结构模型, 以及运行架构中相关运行场景活动, 可通过 CV-6 建立体系能力与运行活动之间的关联关系, 如图 10 所示。以“互联互通能力”中“数据服务信息”为例, 在正常场景中“传输飞行数据”、“传输航路气象评估数据”、“传输降落指令”、“传输任务指令”等均属于“数据服务信息”的能力范畴, 待运行活动完整构建后, 即可对“数据服务信息”能力进行完整定义。

3.2 运行视角

为使所构建的运行架构所属数据实现结构化, 更好与其他视角数据进行关联, 通过运行活动分解结构(OV-5a)、运行活动逻辑关系(OV-5b)、状态转移描述(OV-6b)、详细行为描述(OV-6c)等视图对运行架构进行构建。

从全空间无人体系及低空经济运行活动分解

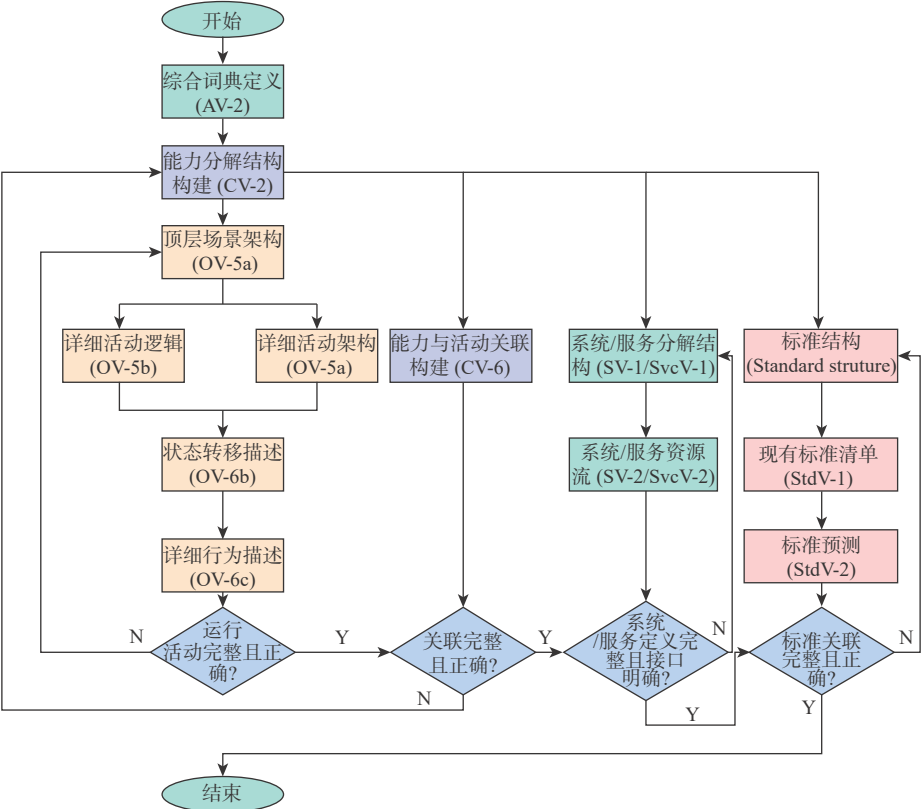


图 8 架构建模技术路径

Fig. 8 Technique path of architecture modelling

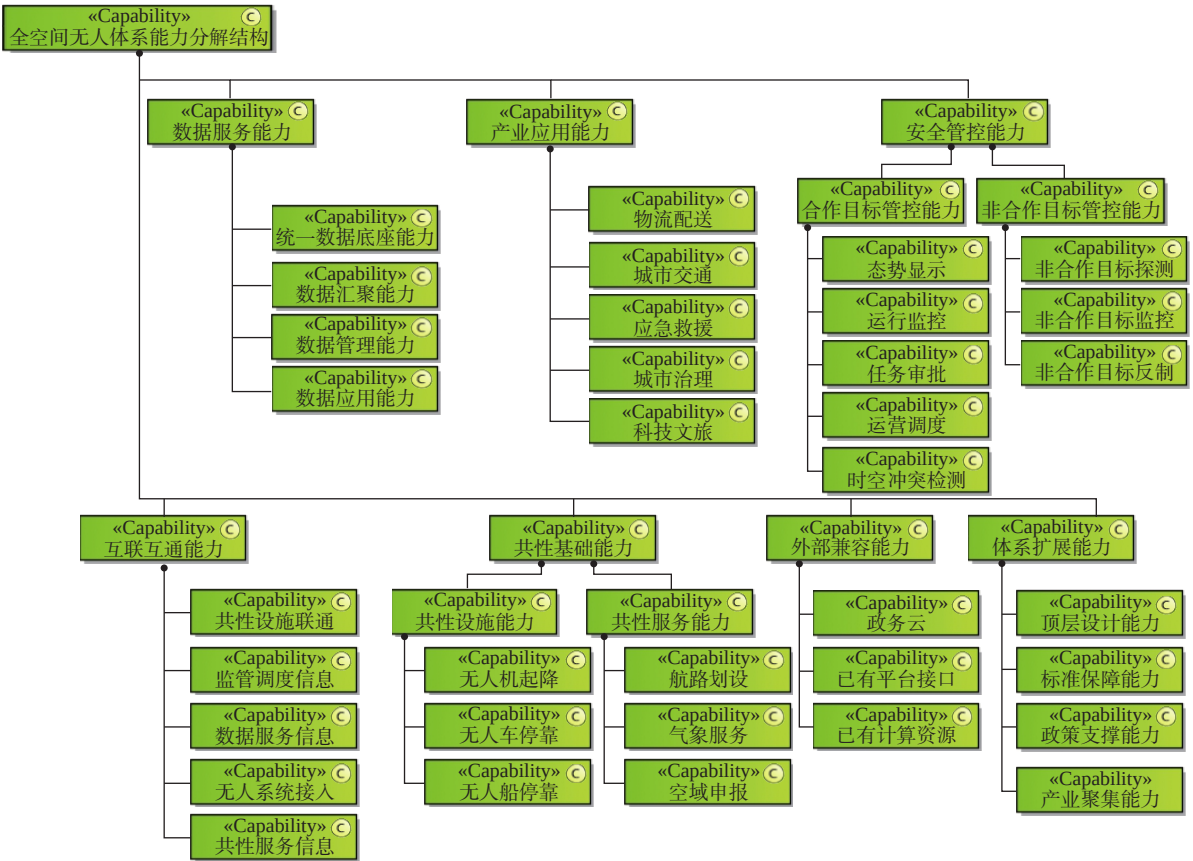


图 9 体系的能力分解结构(CV-2)

Fig. 9 Capability decomposition structure (CV-2)

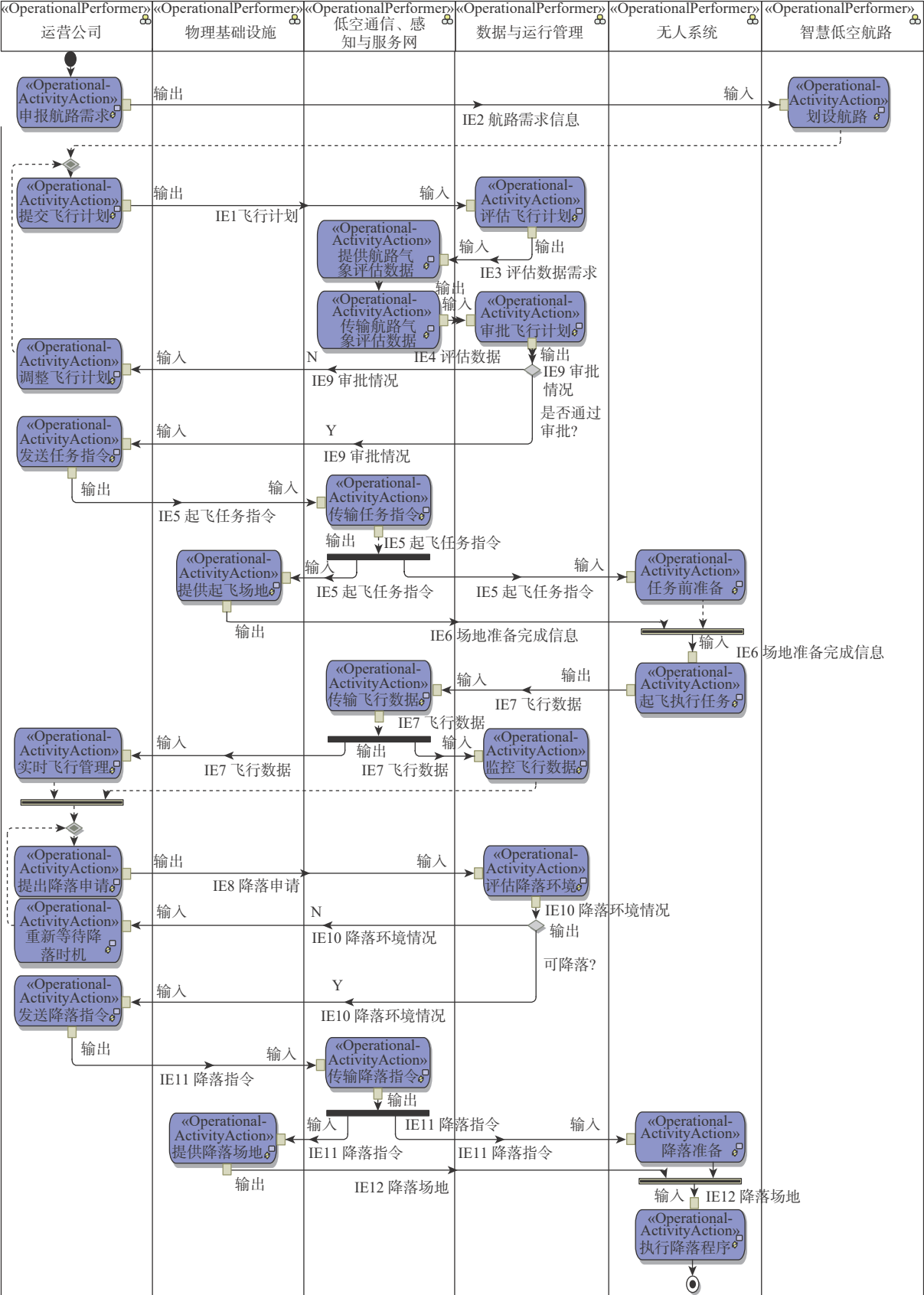


图 12 通用的正常场景活动逻辑(OV-5b)
Fig. 12 Generic normal scenario activity(OV-5b)

3.3 系统/服务视角

系统/服务视角的架构建模与 2.3 节一致,在此

将通感一体网络、低空气象等均作为系统进行处
理,通过 SV-1 方式对系统分解结构进行构建。在

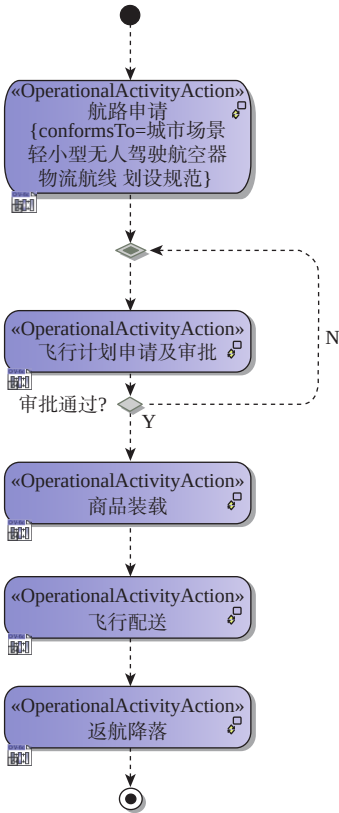


图 13 无人配送场景活动逻辑(OV-5b)

Fig. 13 Unmanned delivery scenario activity(OV-5b)

SV-1 基础上,采用 SV-2 对基础设施之间的关联关系进行构建,如图 16 所示。其中,场景部分主要区别在于无人系统、任务载荷的不同,在此不再赘述。

3.4 标准视角

由于 DoDAF 架构框架没有单独的标准结构视图,在架构建模过程中借鉴 UAF 标准视图中标准结构视图对标准架构进行构建,内容与 2.4 节一致,在此不再重复。

通过标准视图 StdV-1 模型可对当前全空间无人体系已有标准及其对应的系统元素进行映射,形成现有标准清单,建模示例如图 17 所示。基于当前已有标准、政策等,在开展系统各类元素建设时,应符合所采用的标准和政策要求。

标准架构建模的核心价值在于通过实现各类元素与标准之间的显性化关联关系。尤其是全空间无人体系及低空经济作为战略新兴产业,政策法规和标准规范缺乏,需要基于所构建的运行架构、系统架构,对各类元素的标准需求进行识别,具体识别方法如图 18 所示。其中,产品标准、基础设施标准主要来自系统/服务架构,场景运营标准主要来自运行架构,体

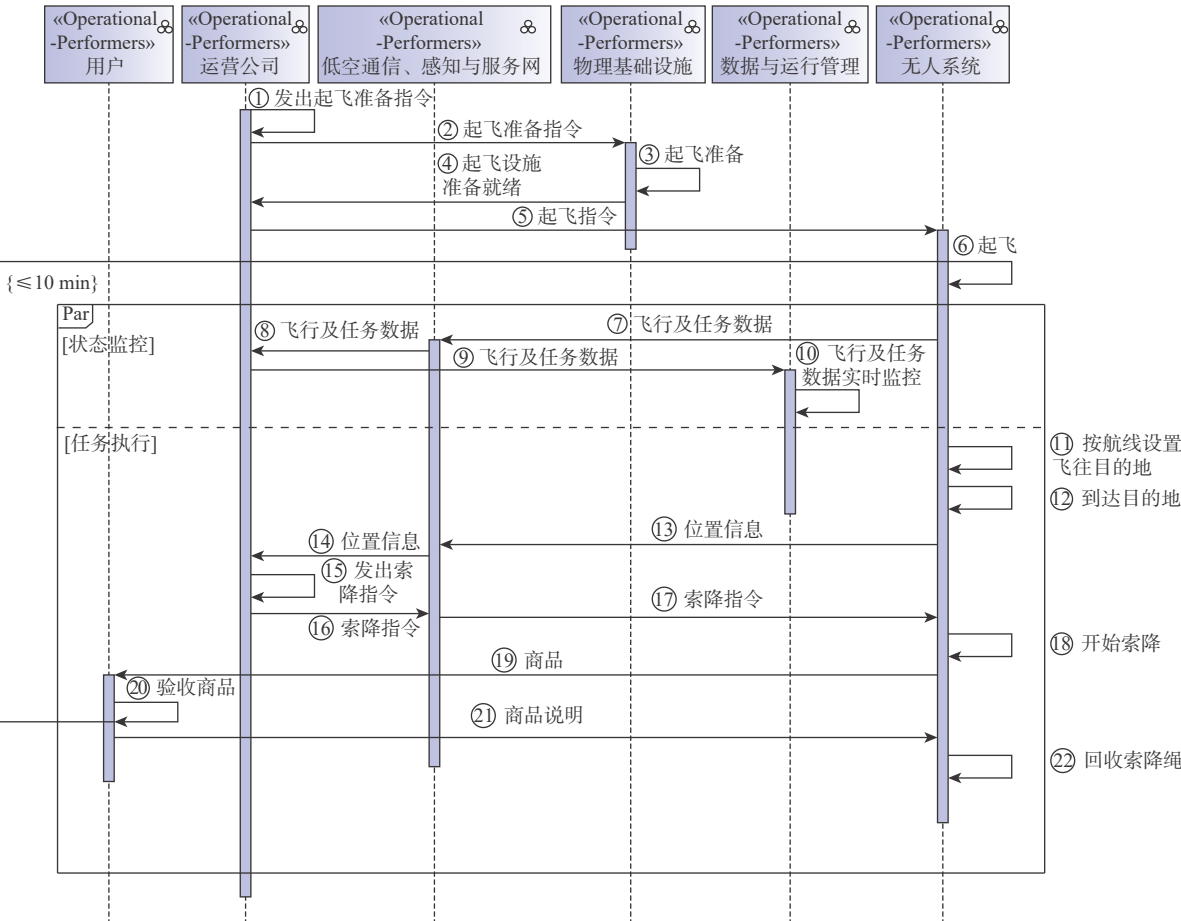


图 14 无人配送场景详细行为描述(OV-6c)

Fig. 14 Detailed behavioural description for unmanned delivery scenario(OV-6c)

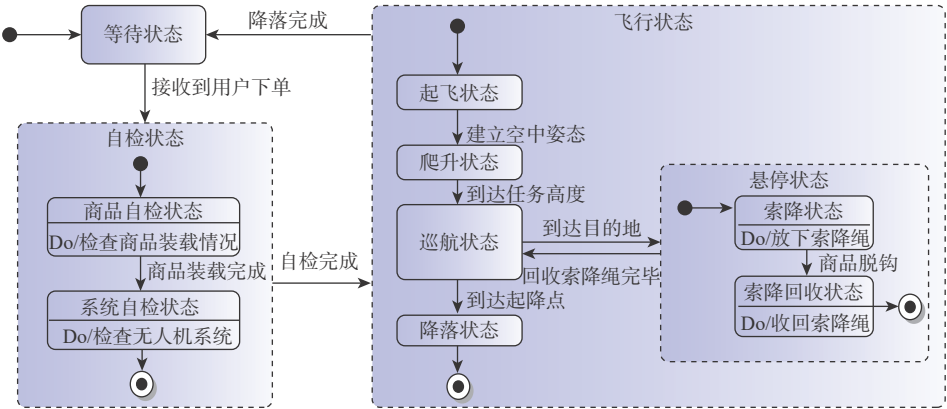


图 15 无人配送场景状态转移描述(OV-6b)

Fig. 15 State transition description for unmanned delivery scenarios(OV-6b)

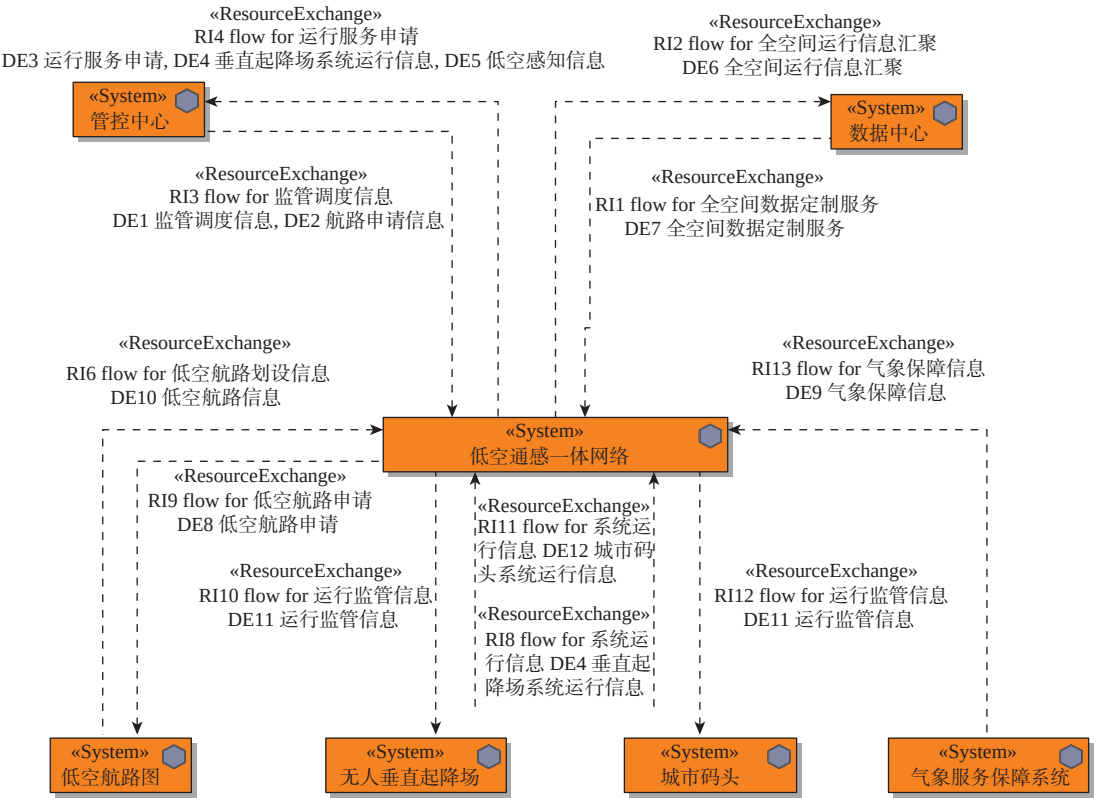


图 16 基础设施接口 SV-2 模型

Fig. 16 SV-2 model of infrastructure interface

#	System element	▽ Standard / Policy
1	低空航路图	SUSE F0005-2024民用无人驾驶航空器城市低空运行航路划设规范
2	无人垂直起降场	SUSE F0004-2024无人驾驶航空器垂直起降场技术要求
3	数据中心	SUSE F0003-2024全空间无人体系数据中心通用技术要求
4	RI3 Resource Exchange:flow for 监管调度信息[管控中心 -> 低空通感一体网络]	SUSE F0002-2024全空间无人体系管控中心通信接口要求
5	RI4 Resource Exchange:flow for 运行服务申请[低空通感一体网络 -> 管控中心]	SUSE F0002-2024全空间无人体系管控中心通信接口要求
6	管控中心	SUSE F0001-2024全空间无人体系管控中心通用技术要求

图 17 现有标准模型(StdV-1)

Fig. 17 Existing standard models(StdV-1)

系集成标准则会同时涉及系统/服务架构和运行架构。

根据对各类系统元素、交互关系、运行活动中标准需求的识别,可按照标准结构不同领域的划分,对拟制定的标准和政策进行规划,具体如图 19 所示。

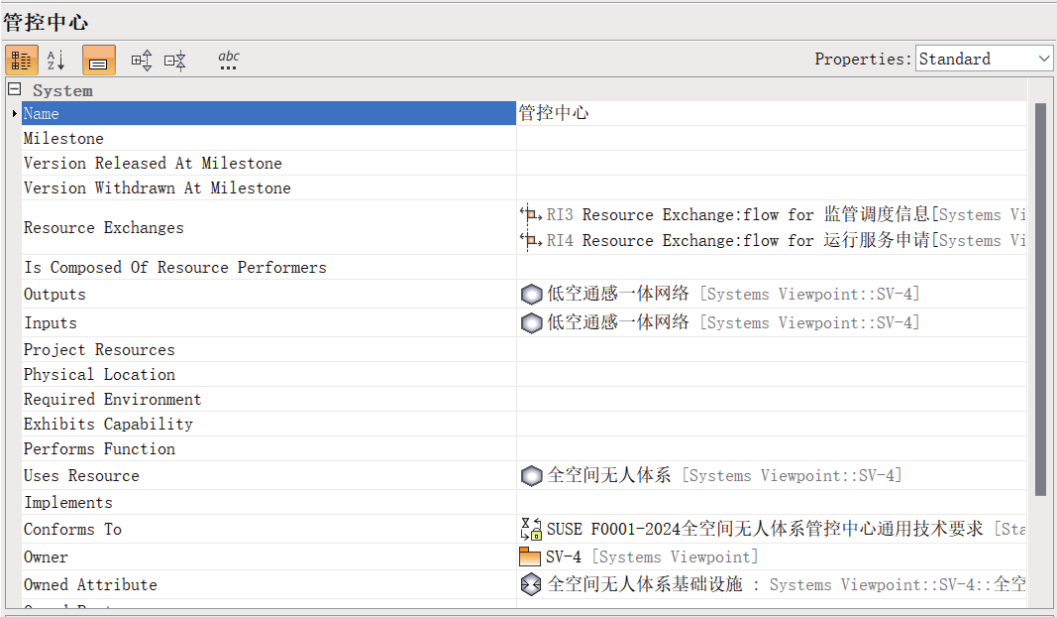


图 18 针对特定对象的标准需求识别

Fig. 18 Standard requirement identification for specific object

#	Technology area	From: 2025-06-30 To: 2026-06-29	From: 2026-06-30 To: 2027-06-29	From: 2027-06-30 To: 2028-06-29
1	数据设施标准	全空间无人体系数 据定义及分类		
2	网络设施标准	全空间无人体系低 空通感一体网络通 用要求		
3	气象设施标准	全空间无人体系低 空气象保障服务系 统通用要求		
4	管控设施标准		全空间无人体系 管控中心信息显 示布局通用要求	
5	技术接口标准		无人系统空地数 据传输内容与格 式	

图 19 标准预测模型(StdV-2)

Fig. 19 Standards forecast models(StdV-2)

4 结 论

围绕全空间无人体系及低空经济背景下的无人系统运行场景复杂体系,通过开展场景分析、架构分析、架构建模 3 方面的研究,主要形成以下结论:

1) 通过对运行场景要素分析,提出了运行场景“6+6”要素元模型,并站在特定区域发展角度提出运行场景框架,对运行场景复杂体系进行结构化解析,为各地区开展运行场景的分析与设计提供有效参考;

2) 在场景框架基础上,从架构分析角度提出运行场景架构元模型,并对全空间无人体系及低空经济运行场景的能力架构、运行架构、系统架构、标准架构进行设计,从不同视角对复杂体系架构进行定义,改变传统单一视角的架构定义,推动对复杂

体系多视角架构的统一理解;

3) 研究形成一种面向全空间无人体系及低空经济运行场景的架构建模方法,实现能力视角、运行视角、系统视角及标准视角不同要素的关联追溯,为后续基于模型开展各类基础设施或无人系统产品研制的运行概念、需求捕获及功能分析提供框架,尤其为支撑体系建设发展的关键政策法规和标准规范建设提供方法支撑,保障无人系统运行场景的安全、健康发展。

参考文献 (References)

[1] 中共中央办公厅, 国务院办公厅. 关于完善市场准入制度的意见 [EB/OL]. (2024-08-01)[2025-06-02]. https://www.qstheory.cn/yao wen/2024-08/21/c_1130194466.htm.
Opinions of the General Office of the Central Committee of the Communist Party of China , the General Office of the State. Coun-

cil on improving the market access system[EB/OL]. (2024-08-01)[2025-06-02]. https://www.qstheory.cn/yaowen/2024-08/21/c_1130194466.htm(in Chinese).

[2] 中国信息协会低空经济分会. 低空经济发展报告 (2024-2025)[R]. 北京: 中国信息协会低空经济分会, 2025: 28-38.

Low-altitude Economy Branch of the China Information Industry Association. Low-altitude economic development report (2024-2025)[R]. Beijing: Low-altitude Economy Branch of the China Information Industry Association, 2025: 28-38(in Chinese).

[3] 欧阳桃花. 低空经济的技术创新与场景创新[J]. 人民论坛·学术前沿, 2024(15): 57-68.

Ouyang T H. Technological innovation and scenario innovation of low-altitude economy[J]. Frontiers, 2024(15): 57-68(in Chinese).

[4] 欧阳桃花, 郑舒文. 基于共同演化的低空经济产业生态策略研究: 以“低空航空器+”为例[J]. 北京航空航天大学学报 (社会科学版), 2024, 37(5): 109-119.

Ouyang T H, Zheng S W. Industrial ecological strategies of low-altitude economy based on co-evolution: taking “low-altitude aircraft+” as an example[J]. Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition), 2024, 37(5): 109-119(in Chinese).

[5] 张晓兰. 我国低空经济发展面临的问题与政策建议[EB/OL]. (2024-12-30)[2025-06-10]. http://www.ndrc.gov.cn/wsdwhfz/202412/t20241230_1395328_ext.html.

Zhang X L. The issues and policy recommendations regarding the development of the low-altitude economy in our country[EB/OL]. (2024-12-30)[2025-06-10]. http://www.ndrc.gov.cn/wsdwhfz/202412/t20241230_1395328_ext.html(in Chinese).

[6] 张夏恒. 新质生产力背景下低空经济高质量发展的机理与路径[J]. 苏州大学学报 (社会科学版), 2025, 46(1): 112-122.

Zhang X H. Research on the mechanism and path of high-quality development of low-altitude economy under the background of new-quality productivity[J]. Journal of Soochow University (Philosophy and Social Science Edition), 2025, 46(1): 112-122(in Chinese).

[7] 褚鹏, 王军, 刘树光, 等. 基于人工智能技术的低空飞行器管控关键技术研究[J]. 航空兵器, 2023, 30(2): 120-124.

Chu P, Wang J, Liu S G, et al. Research on key technologies of low altitude vehicle management and control based on artificial intelligence technology[J]. Aero Weaponry, 2023, 30(2): 120-124(in Chinese).

[8] 刘泉, 陈瑶瑶, 洪晓苇, 等. 面向无人机的城市低空空域规划的国际经验[J]. 城市规划学刊, 2024(5): 64-70.

Liu Q, Chen Y Y, Hong X W, et al. International experience of urban low altitude airspace planning for drones[J]. Urban Planning Forum, 2024(5): 64-70(in Chinese).

[9] Huang C Q, Fang S F, Wu H, et al. Low-altitude intelligent transportation: system architecture, infrastructure, and key technologies [J]. Journal of Industrial Information Integration, 2024, 42: 100694.

[10] 廖小罕, 徐晨晨, 叶虎平. 低空经济发展与低空路网基础设施建设的效益和挑战[J]. 中国科学院院刊, 2024, 39(11): 1966-1981.

Liao X H, Xu C C, Ye H P. Benefits and challenges of constructing low-altitude air route network infrastructure for developing low-altitude economy[J]. Bulletin of Chinese Academy of Sciences, 2024, 39(11): 1966-1981(in Chinese).

[11] Li Z R, Gao Z, Wang K Y, et al. Unauthorized UAV countermeas-

ure for low-altitude economy: joint communications and jamming based on MIMO cellular systems[J]. IEEE Internet of Things Journal, 2025, 12(6): 6659-6672.

[12] 李晓华. 政府引导、产业生态构建与低空经济发展[J]. 改革, 2025(2): 21-35.

Li X H. Government guidance, construction of industrial ecosystem, and the development of low-altitude economy[J]. Reform, 2025(2): 21-35(in Chinese).

[13] 孔得建, 袁泽. 低空经济政策法律体系的现状、经验与展望[J]. 北京航空航天大学学报 (社会科学版), 2024, 37(5): 85-95.

Kong D J, Yuan Z. Current situation, experience and prospect of the policies and legal system of low-altitude economy[J]. Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition), 2024, 37(5): 85-95(in Chinese).

[14] 樊一江, 李卫波. 我国低空经济阶段特征及应用场景研究[J]. 中国物价, 2024(4): 98-103.

Fan Y J, Li W B. The development stage and application scenarios of China’s low-altitude economy[J]. China Price Journal, 2024(4): 98-103(in Chinese).

[15] 庄苗. 人工智能赋能低空经济: 应用场景与未来方向[J]. 人民论坛·学术前沿, 2024(15): 38-44.

Zhuang Z. AI empowers low-altitude economy application scenarios and future directions[J]. Frontiers, 2024(15): 38-44(in Chinese).

[16] Yang H L, Zheng M T, Shao Z L, et al. Intelligent computation offloading and trajectory planning for 3-D target search in low-altitude economy scenarios[J]. IEEE Wireless Communications Letters, 2025, 14(4): 949-953.

[17] 王丹, 罗章松. 低空经济赋能农业新质生产力发展: 角色扮演、现实壁垒与破解之道[J]. 农林经济管理学报, 2025, 24(2): 165-173.

Wang D, Luo Z S. Low-altitude economy empowers development of new quality productivity in agriculture: role playing, realistic barriers and solutions[J]. Journal of Agro-Forestry Economics and Management, 2025, 24(2): 165-173(in Chinese).

[18] 李晓华. 以应用场景创新推动低空经济发展[J]. 中国军转民, 2024(23): 3.

Li X H. Promoting the development of low-altitude economy through application scenario innovation[J]. Defence Industry Conversion in China, 2024(23): 3(in Chinese).

[19] 王灿, 郭齐胜, 穆歌, 等. 军事领域体系结构研究综述[J]. 火力与指挥控制, 2023, 48(2): 1-11.

Wang C, Guo Q S, Mu G, et al. Review of system of systems architecture research in military domain[J]. Fire Control & Command Control, 2023, 48(2): 1-11(in Chinese).

[20] IEEE Computer Society. IEEE standard for information technology-Protocols for distributed interactive simulation applications—Entity information and interaction: IEEE 1278-1993[S]. Piscataway: IEEE Press, 1993.

[21] Durak U, Topçu O, Siegfried R, et al. Scenario development: a model-driven engineering perspective[C]//Proceedings of the 4th International Conference on Simulation and Modeling Methodologies, Technologies and Applications. Piscataway: IEEE Press, 2014: 117-124.

[22] NATO Science and Technology Organization. Rapid scenario generation for simulation applications[EB/OL]. (2023-07-04)[2025-06-09]. <https://www.sto.nato.int/document/rapid-scenario-generation-for->

simulation-applications/.

[23] ISO/IEC/IEEE. Systems and software engineering—Life cycle processes—Requirements engineering: ISO/IEC/IEEE 29148: 2018(E) [S]. Piscataway: IEEE Press, 2018: 1-104.

[24] Society of Automotive Engineers. Guidelines for development of civil aircraft and systems: ARP4754B[S]. Washington, D. C.: Society of Automotive Engineers, 2023: 1-172.

[25] 卢方周, 闫坤, 李文宇, 等. 低空经济总体架构与提升路径[J]. 信息通信技术与政策, 2024, 50(11): 2-10.

Lu F Z, Yan K, Li W Y, et al. Overall architecture and enhancement pathways of low-altitude economy[J]. Information and Communications Technology and Policy, 2024, 50(11): 2-10(in Chinese).

[26] 江雨航, 郭泰, 宫基, 等. 民用飞机应急场景分析与建模方法[J]. 北京航空航天大学学报, 2024, 50(3): 839-849.

Jiang Y H, Guo T, Gong Q, et al. Analysis and modeling method of civil aircraft emergency scenario[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(3): 839-849(in Chinese).

Architecture design and modeling technology for civil unmanned system operation scenarios

GUO Tai*, AN Yiping, SHU Zhenjie, JIANG Jueyi, JIANG Yuhang

(AVIC China Aero-polytechnology Establishment, Beijing 100028, China)

Abstract: Operational scenarios for civil unmanned systems act as hubs of the industrial value of all-space unmanned system and low-altitude economy, playing a crucial role in scaling up new industries. This typical complex system features dynamic boundaries, multi-level emergence, balanced stakeholder interests, and continuous evolution. Modern systems engineering principles must be used in the design of such systems' architecture to guarantee their safe and effective operation. Our research into the architecture design and modeling of civil unmanned system scenarios produced a "6+6" meta-model, capturing static and dynamic elements. In addition to developing a thorough process that covers everything from element identification to architecture design and modeling, we created a multi-perspective architecture design and modeling including capability, operational, system, and standard architectures. This approach not only provides guidance for the construction of new infrastructure, the research, development, and selection of unmanned systems, and the formulation of policy and standard creation, but also concurrently offers technical support for the establishment of unmanned system operation scenarios and the systematic development of the related industry.

Keywords: low-altitude economy; all-space unmanned systems; operational scenarios; framework modeling; unmanned systems

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0411

面向无人机平台的 CNN-Transformer 混合模型训练后量化方法

赵呈功^{1,2}, 胡纪远^{1,3}, 杨星宇⁴, 陈磊⁵, 姜宏旭^{3,*}

(1. 北京航空航天大学杭州创新研究院, 杭州 310056; 2. 北京航空航天大学 软件学院, 北京 100191;
3. 北京航空航天大学 计算机学院, 北京 100191; 4. 内蒙古大学 计算机学院 (软件学院), 呼和浩特 010021;
5. 杭州萧山国际机场飞行区中心, 杭州 311241)

摘 要: 近年, 无人机在各类低空生产、服务场景作用愈发重要。由于边缘计算平台的存储、计算资源受限, 视觉模型的边缘侧运行性能短板成为算法应用于以上场景的技术瓶颈。针对卷积神经网络 (CNN)-Transformer 混合模型激活值长尾分布导致量化精度严重退化问题, 提出高效训练后量化方法。该方法通过激活值噪声补偿 (ANC) 和自适应难度迁移抑制异常值对量化精度的影响, 在提升模型量化精度同时提高其推理速度。实验结果表明: 所提方法在典型模型的 8 bit 量化场景下, 精度损失小于 1%, 量化后推理速度最高提升超过 200%。所提方法显著提升了边缘侧设备的视觉模型推理性能, 为低空无人机平台上的模型部署提供了技术支撑。

关键词: 训练后量化; 模型压缩; 推理加速; 异常值感知; 边缘计算

中图分类号: V221⁺.3; TB553

文献标志码: A

文章编号: 1001-5965(2026)09-3066-09

近年来, 随着低空经济的快速发展, 无人机在低空物流、空中巡检、农林监测、应急通信等场景中得到广泛应用。这些任务往往对视觉感知系统提出了“高实时性、低功耗、轻量化”的严格要求: 无人机等低空平台通常搭载片上系统 (system-on-chip, SoC) 或嵌入式 AI 芯片, 计算资源和电池续航十分有限, 却需在复杂环境下保持对目标的高精度识别和快速响应。视觉 Transformer (vision Transformer, ViT) 等基于自注意力机制的深度学习模型在图像分类、目标检测和语义分割等计算机视觉任务中取得了卓越性能, 但其庞大的参数量和计算开销导致在嵌入式平台上推理时出现内存占用大、功耗高、延迟长等问题, 难以满足无人机对实时推理的苛刻需求。

为兼顾高精度与高效率, 研究者提出了融合卷积神经网络 (convolutional neural network, CNN) 和

视觉 Transformer 优势的混合架构, 通过卷积层负责局部特征提取并与轻量化自注意力模块结合, 实现了更优的性能, 在保持全局信息建模能力的同时, 有效降低了计算复杂度^[1]。这类模型已在多种轻量级设计中得到验证, 显示出在移动设备上兼顾准确率与速度的潜力。即便如此, 这些模型在实际部署时仍面临高计算量和较大内存占用的瓶颈。

量化技术通过将网络参数和激活值从高精度浮点数映射为低位宽整数表示, 将浮点运算转换为整数运算, 有效降低了模型存储与计算开销, 已经成为压缩与加速模型的主流手段之一。传统针对标准 Transformer 架构的量化方法 (如 FQ-ViT、PTQ4ViT 等) 在通用视觉任务上取得了显著成果, 但由于混合架构中深度可分离卷积产生的激活分布往往具备长尾特性, 这些方法难以保证在混合模型上实现低误差量化。此外, 许多硬件加速方案需要对原始

收稿日期: 2025-06-27; 录用日期: 2025-09-25; 网络出版时间: 2025-09-30 13:44

网络出版地址: link.cnki.net/urlid/11.2625.V.20250930.0919.001

* 通信作者. E-mail: jianghx@buaa.edu.cn

引用格式: 赵呈功, 胡纪远, 杨星宇, 等. 面向无人机平台的 CNN-Transformer 混合模型训练后量化方法 [J]. 北京航空航天大学学报, 2026, 52 (9): 3066-3074. Zhao C G, Hu J Y, Yang X Y, et al. A post-training quantization method for CNN-Transformer hybrid models on UAV platforms [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3066-3074 (in Chinese).

网络结构进行插桩或引入特定算子,从而增加了部署复杂性,并对无人机算力与功耗提出挑战。因此,如何在引入额外算子、保持网络结构不变的前提下,实现混合视觉模型的高效量化,并在嵌入式平台上兼顾实时性与精度,成为亟待解决的关键课题。

针对上述挑战,本文提出了一种面向混合视觉模型的异常值感知的量化(outlier-aware quantization, OAQ)方法。该方法基于对激活值分布特性和权重量化难度的深入分析,设计了 2 种互补优化策略:①在推理阶段引入激活值噪声补偿(activation noise compensation, ANC),通过向激活值添加自适应噪声,使深度可分离卷积输出的长尾分布得到平滑,降低极端值对量化区间的拉拽效应;②通过数学等效变换,将激活端高动态范围的量化难度转移至权重量化,并结合分层缩放与不同粒度的量化方案,对权重进行优化,从而简化激活端的量化过程,并避免插入额外算子。这 2 项策略在不改变模型结构的前提下,有效降低了推理时的计算复杂度和内存开销,显著提高了模型在嵌入式平台上的实时推理能力,同时保持了量化后的准确性。本文主要创新如下:

- 1) 从无人机边缘平台资源受限的角度出发,提出了针对混合架构激活分布长尾特性的噪声补偿机制,实现了对量化边界的自适应优化,显著提高了复杂低空环境下的目标检测鲁棒性;
- 2) 通过数学等效变换将激活端的高动态范围量化压力转移至权重端,并采用分层缩放策略,减少了激活端量化误差,无需修改网络结构即可在嵌入式设备上高效部署;
- 3) 本文在主流的 CNN-Transformer 混合架构模型上进行了大量实验。结果表明,在 INT8 量化条件下,本文 OAQ 方法在图像分类任务上实现了接近无损的预测精度,量化后精度下降小于 1%;

1 相关工作

1.1 模型架构发展

无人机等边缘感知设备对视觉模型的实时性和资源消耗提出了更高要求。深度学习模型的架构演变正是为满足这类应用需求,从最早的卷积神经网络到近年来兴起的视觉 Transformer 模型。最初,卷积神经网络凭借局部特征提取的能力在计算机视觉领域取得了显著的成功^[2-4],如 AlexNet^[2]、VGG^[4]、Resnet^[5]等。但 CNN 在长程依赖处理上存在局限^[6],这促使了对新架构的探索。于是, Swin Transformer^[6]、ViT^[7]、DeiT^[8]等模型应运而生。这

些模型通过自注意力机制取代了卷积层,使得模型能够更有效地建模整个图像的全局依赖关系^[9]。与传统的卷积神经网络不同, ViT 将输入图像分割成固定大小的 patch 序列,然后,通过多层 Transformer 编码器来捕捉全局依赖关系。这种设计不仅简化了模型结构,还显著增强了对长距离特征关联的理解能力,并在大规模数据集上训练时展示了超越传统 CNN 的性能,取得了领先成果^[10-11]。这些 Transformer 模型,凭借其强大的全局信息建模能力,已经成为深度学习中的一种主流方法。

Transformer 模型的计算复杂度高及参数规模大。于是研究人员探索了结合 CNN 与 Transformer 架构的方法,旨在利用 CNN 的高效特征提取能力及 Transformer 处理长程依赖的优势,从而提出了多种轻量化设计方案,提升其在移动设备上的部署效率。例如, MobileViT^[12]通过融合局部卷积与全局自注意力机制,实现了高效分类,特别适合移动设备; EdgeNeXT^[13]通过引入分裂深度转置注意力编码器,实现了在较低计算需求下优异性能表现; EfficientFormer^[14]通过重新设计 ViT 架构,采用维度一致的设计范式和延迟驱动的瘦身方法,显著提升了推理速度。CNN-Transformer 混合架构通过结合局部和全局信息的优势,在特定应用场景中提供了良好的平衡。然而,即便在混合架构下,深度可分离卷积和自注意力模块的组合仍会带来较高的计算开销与内存占用,制约了无人机等低空平台的长续航与实时性能。由此可见,针对上述混合视觉模型的进一步压缩与加速,成为低空经济环境中模型部署的关键技术挑战。

1.2 训练后量化

无人机或其他低空平台对视觉模型的部署环境往往资源受限,模型推理必须兼顾实时性与功耗控制。量化技术通过将网络参数和激活值从浮点数表示映射为低位宽整数表示,可显著降低模型的存储需求与计算开销,是边缘设备实现高效推理的重要手段。量化可分为量化感知训练(quantization-aware training, QAT)^[15-16]和后训练量化(post-training quantization, PTQ)^[17-19]。QAT 可在量化后保持较高的准确率,但需要访问完整的训练数据集,并进行额外的训练。PTQ 直接将预训练神经网络的高精度权重和激活值压缩到低精度值来加速推理,而不需要重新训练,比 QAT 更加灵活与高效。

近年来,深度学习模型的量化研究从传统的卷积神经网络逐步扩展到 Transformer 模型,推动了 PTQ 在不同架构中的广泛应用。例如, Jacob 等^[20]通过整数算术推理并优化训练过程,提升了设备延

迟与准确性之间的平衡。Nagel 等^[21]提出的改进 AdaRound 权值取整机制,通过优化权重量化方法,在 ResNet 模型中实现了更高效的 4 位权重量化。对于 Transformer 模型量化,Liu 等^[17]通过引入排名损失保持自注意力机制的顺序,并采用混合精度量化方案,优化视觉 Transformer 的存储和计算成本。随后,PTQ4ViT^[19]提出了一种双重均匀量化方法,通过后训练量化有效减少了视觉 Transformer 中的量化误差,同时提高了推理效率。

然而,面对 CNN-Transformer 混合架构的 ViT 时,现有 PTQ 技术在低空场景下的应用仍存在挑战。鉴于此, Lee 等^[22]提出一种用于混合视觉 Transformer 的后训练量化方法 Q-HyViT,针对混合 ViT 中的高动态范围、零点溢出、不同的归一化类型和参数较少等问题,通过桥接块重构和误差最小化,自动选择量化尺度因子、粒度及量化方案,显著减少了量化误差。Kim 等^[23]提出了一种面向硬件友好的 CNN-Transformer 混合网络训练后量化方法 HyQ,通过引入量化感知的分布缩放来解决卷积层中的通道间方差导致的离群值问题,并在 Transformer 模块中用整数线性近似代替了计算量大的 softmax 函数。虽然 Q-HyViT^[22]可以避免量化 CNN-Transformer 混合模型出现精度严重退化的问题,但是混合量化对推理加速是非常不友好的,需要有专用的内核支持。HyQ^[23]使用整数线性近似代替标准的 softmax 函数,减少了计算量,但是在目前主流的推理框架中,整数线性近似的 softmax 函数支持尚不完善,导致实际应用时需要额外的定制化实现。

2 本文 OAQ 方法

模型量化旨在将模型的权重和激活从浮点数转换为低精度的整数表示,从而减少模型的存储和计算需求。量化器一般使用均匀对称量化,因为大多数硬件平台都很好地支持均匀量化,其将输入激活值 X 映射到 b 位整数值 $Q(X|b)$ 上:

$$Q(X|b) = \text{clip}(\lfloor X/s \rfloor + z, 0, 2^b - 1) \tag{1}$$

式中: clip 为裁剪函数,其作用是将结果限制在区间 $[0, 2^b - 1]$ 内; s 为比例因子, z 为零点,均为量化所需要的参数,通常由 X 中元素的下限 l 和 X 中元素的上限 u 确定,可以表示为

$$\begin{cases} l = \min(X_i) \\ u = \max(X_i) \end{cases} \tag{2}$$

$$s = (u - l) / (2^b - 1) \tag{3}$$

$$z = \text{clip}(\lfloor -l/s \rfloor, 0, 2^b - 1) \tag{4}$$

在边缘计算中,高效的模型设计至关重要,因此,很多轻量化模型在设计时采用了深度可分离卷积来减少计算复杂度。深度可分离卷积在参数量和计算量方面具有显著优势,但是由深度可分离卷积在处理每个输入通道时,独立计算每个通道的特征图。这导致了通道之间的相互依赖性较弱,通道间数据的方差较大,激活分布存在长尾特性。EfficientFormer-L1 中间阶段和最后阶段全连接层的卷积层的输入激活的 95% 的激活都集中在 0 附近,但是数据范围分别为 $[-0.17, 19.57]$ 和 $[-55, 22.34]$,具有非常明显的长尾分布。这种极端的激活经过式 (1) 的计算后,会将大部分的数据映射为 0,影响量化后模型的整体性能。在本节中,介绍一种两阶段的训练后量化方法,无需手工设计参数,可实现在精度退化很少的情况下量化模型。为实现这一目标,在 2.1 节和 2.2 节设计了 2 种针对异常值的量化方法,包括: ANC 使得激活值的分布更加平坦;自适应难度迁移 (adaptive difficulty migration, ADM) 通过将量化难度迁移到权重,使得激活值更易量化。如图 1 所示,其中, Y_{fp32} 为浮点模型的参考输出; $\hat{Y}_{wqaq}$ 为融合 ANC/ADM 参数后的量化模型输出; $L(\cdot)$ 表示损失函数,用于衡量量化输出与浮点输出之间的差异,并据此更新 ANC/ADM 参数。

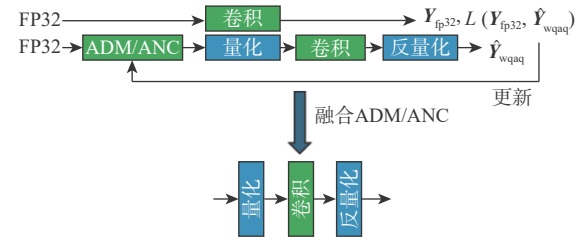


图 1 ANC/ADM 的优化过程

Fig. 1 Optimization process of ANC/ADM

2.1 激活值噪声补偿

激活值噪声补偿的核心思想是在对视觉 Transformer 的激活值进行量化之前,向激活值中添加噪声补偿。这种噪声补偿能够改变原始激活值的分布,使其更加平坦,减少长尾效应。本文将噪声应用于输入激活值 X 和权重 $W \in \mathbf{R}^{C_{out} \times C_{in} \times K_h \times K_w}$, C_{in} 为输入通道数; C_{out} 为输出通道数; K_h 与 K_w 分别为卷积核的高度和宽度。使用噪声补偿的卷积层操作表示如下:

$$Y = WX + B = W(X - \beta) + (B + W\beta) \tag{5}$$

式中: Y 为输出; 偏置 $B \in \mathbf{R}^{1 \times C_{out}}$; 噪声 $\beta \in \mathbf{R}^{1 \times C_{in}}$ 。在量化之前,需要初始化噪声补偿。噪声补偿的初始化需要结合激活值的分布,并且噪声补偿不宜过大,因为过大的噪声补偿会使激活值波动剧烈,导

致量化后模型精度退化严重的问题。因此,设计了如下初始化噪声补偿的方式:

$$\beta_i = \mu_i = \frac{1}{N} \sum_{j=1}^N x_{ij} \quad (6)$$

式中: x_{ij} 为第 i 个通道的第 j 个激活值; N 为激活值的数量; β_i 为第 i 个通道噪声补偿。该噪声补偿 β_i 通过优化过程进行调整,从而使激活值的分布更加适合量化,减少极端值的影响。为优化噪声补偿 β , 本文定义了一个损失函数,该函数包括量化误差和激活值分布平滑 2 个方面。其中,量化误差损失度量了加噪声模型与未加噪声模型之间的量化误差。具体地,令 $Q(\cdot)$ 为量化操作,则量化误差损失函数 $L_{\text{quant}}(\beta)$ 被定义为

$$L_{\text{quant}}(\beta) = \frac{1}{N} \sum_{j=1}^N |Q(Y_{\text{noisy}}^{(j)}) - Q(Y_{\text{orig}}^{(j)})|^2 \quad (7)$$

式中: Y_{noisy} 和 Y_{orig} 分别为卷积层带噪声和未带噪声的输出。此外,为避免激活值分布的极端波动,通过计算激活值的动态范围和方差来平滑其分布,则平滑损失 $L_{\text{smooth}}(\beta)$ 被定义为

$$L_{\text{smooth}}(\beta) = \lambda_1 |X_{i,\max} - X_{i,\min}| + \lambda_2 \sigma_i^2 \quad (8)$$

式中: λ_1 和 λ_2 为超参数,用来控制动态范围和方差的权重; $|X_{i,\max} - X_{i,\min}|$ 为激活值的动态范围, σ_i^2 为激活值的方差。最终的综合损失函数为

$$L_{\text{total}}(\beta) = L_{\text{quant}}(\beta) + L_{\text{smooth}}(\beta) \quad (9)$$

噪声补偿仅在量化前进行优化,在获得最佳噪声补偿后,本文将噪声补偿与激活值结合以此来调整每个通道的激活值分布。在校准阶段,使用损失函数(9)进行优化以获取最佳的 ANC。在校准完成后,将 ANC 合并到量化器中,并对该卷积层的偏置进行更改。此外,在推理阶段,噪声补偿作为固定参数被应用于计算,不再进行动态更新,这使得推理过程的计算效率保持高效。

2.2 自适应难度迁移

激活值添加噪声补偿后,其分布整体变得更加平坦。然而,仍然存在部分通道的激活值相较于其他通道依然偏大,这种不均匀性可能在量化过程中导致精度损失。此外,权重的分布通常更均匀且数值范围较小,具备更高的量化友好性。由此本文提出自适应难度迁移,不需要人为设置难度迁移的因子。其核心思想是在量化之前,使用迁移因子将激活值的量化难度迁移到权重中,使激活值容易量化。本文将迁移因子应用于输入激活 X 和权重 W 。使用迁移因子的卷积层操作表示如下:

$$Y = WX + B = W\delta(X/\delta) + B \quad (10)$$

式中: 难度迁移因子 $\delta \in \mathbf{R}^{1 \times C_m}$ 。确定最佳的难度迁移因子 δ 是整个量化过程中至关重要的一步。为使得难度迁移因子在减少激活值量化难度的同时不过多增加权重的量化难度,本文设计了一个有效的优化策略,以在训练过程中自动调节 δ , 使其适应不同通道的激活值分布特性。首先,本文使用激活值的标准差初始化难度迁移因子。对于每个通道,初始化迁移因子为该通道激活值的标准差 σ_{x_i} :

$$\delta_i = \sigma_{x_i} = \frac{1}{N} \sum_{j=1}^N (X_j - \mu_{x_i})^2 \quad (11)$$

为确定最佳的迁移因子,本文采用基于量化误差的优化方法。在训练过程中,本文不断调整迁移因子的值,使得量化误差最小化。量化误差可以通过均方误差(mean squared error, MSE)来度量:

$$L_{\text{quant}} = \frac{1}{N} \sum_{i=1}^N (\hat{Y}_i - Y_i)^2 \quad (12)$$

式中: N 表示校准集中样本数, $\hat{Y}_i$ 为量化后的输出, Y_i 为真实输出。通过优化损失函数 L_{quant} , 迁移因子 δ 会被调整到最佳值,从而确保量化过程中的误差最小化。

与 2.1 节中操作类似, δ 一旦确定后,在推理过程中是固定的。得益于 2.1 节对于激活值的处理,本文可以更快地获得最佳的 δ 。在校准阶段,使用平均绝对误差(mean absolute error, MAE)进行优化以获取最佳的 ADM。在校准完成后,将 ADM 合并到量化器中,并调整卷积层的权重。此外,与 FQ-ViT^[18] 和 Q-HyViT^[22] 不同的是, OAQ 采用了均匀量化方法,因此,不需要依赖专门的硬件,可以直接在主流的边缘设备上运行,极大地提高了其实际应用的可行性。

3 实验结果与分析

3.1 数据集和评价指标

本文选择的是用于分类任务的 imagenet-1k 数据集,有 1.2×10^6 个训练图像和 1 000 个类的 50 000 个验证图像。此外,为获得最佳的噪声补偿和难度迁移因子,本文从训练集中挑选出 1 000 个未标记的图像作为 PTQ 方法的校准集。该校准集可确保量化参数得到有效的微调,从而提高量化模型的性能。本文对预训练的 CNN-Transformer 混合架构的视觉模型执行统一的 8 位 PTQ。本文量化的模型的来源包括两类,一类是从 timm 库提供的 MobileViT^[12] 和 MobileViTv2^[24],另一类是开源代码仓库中的 EfficientFormer^[14]。本文使用官方提供的浮点模型的 Top-1 准确率作为基线。本文也量化了全连接

层,包括第 1 个投影层和最后 1 个预测层。因为模型中并不是每一层的激活值都有异常激活值,所以本文只为出现异常值的层添加了噪声补偿及量化难度迁移。

3.2 实验平台及参数设置

为确保本文实验结果的准确性与可复现性,所有实验均在统一的软硬件环境下进行。实验所采用的计算平台操作系统为 Ubuntu22.04.4LTS,处理器为 Intel Core i9-13900K,搭载 NVIDIA GeForce RTX 4090 显卡,具备强大的并行计算能力和显存资源。在软件环境方面,实验基于 Python 3.10.15 编写,并使用 PyTorch 2.5.0 深度学习框架,配套 CUDA 12.4 作为 GPU 加速后端。实验中使用的是完整的校准集,学习率为 0.001;对于常规线性模块采用层级量化,对于 LayerNorm 等非线性模块采用通道级量化。上述环境配置为获取自适应噪声与难度迁移因子提供了稳定高效的支持,确保了实验结果的可靠性与一致性。

3.3 性能分析

为全面评估本文 OAQ 方法在模型量化中的性能优势,本文与目前最先进的几种量化方法进行了对比,包括 FQ-ViT^[18]、PTQ4ViT^[19]、Q-HyViT^[22]、HyQ^[23]、RepQ-ViT^[25]、EasyQuant^[26]。通过对比分析,展示了本文 OAQ 方法在多种 CNN-Transformer 混合模型上的量化效果,并进一步证明了其在精度方面的优越性。

本文选择了多个常见的模型架构进行比较测试,涵盖了从轻量级到较大体积的多种网络结构,如 MobileViT 系列模型和 EfficientFormer 模型。如表 1~表 3 所示,在 8 bit 量化过程中,本文 OAQ 方法的表现明显优于当前领先的量化方法,尤其是在 MobileViT-s 模型上,精度仅退化了 0.31%,这一结果相较于其他量化方法具有显著的优势。例如,RepQ-ViT 在该模型上表现较差,精度下降幅度较大,而本文 OAQ 方法能够有效抑制量化误差,显著降低精度退化。

表 1 MobileViT 模型量化后精度

Table 1 Accuracy of MobileViT after quantization

方法	权重/激活值	精度Top-1/%		
		MobileViT-xs	MobileViT-xs	MobileViT-s
全精度	32/32	69.00	74.80	78.40
FQ-ViT ^[18]	8/8	66.46	68.28	77.67
PTQ4ViT ^[19]	8/8	37.75	65.52	68.19
RepQ-ViT ^[24]	8/8	1.85	41.96	50.01
Q-HyViT ^[22]	8/8	67.20	73.89	77.72
HyQ ^[23]	8/8	68.15	73.99	77.93
EasyQuant ^[25]	8/8	36.13	73.16	74.21
本文OAQ方法	8/8	68.65	74.35	78.09

表 2 MobileViTv2 模型量化后精度

Table 2 Accuracy of MobileViTv2 after quantization

方法	权重/激活值	精度Top-1/%		
		MobileViTv2-50	MobileViTv2-75	MobileViTv2-100
全精度	32/32	70.20	75.60	78.10
FQ-ViT ^[18]	8/8	67.66	69.56	77.15
PTQ4ViT ^[19]	8/8	39.39	65.54	51.02
RepQ-ViT ^[24]	8/8	26.60	55.52	40.85
Q-HyViT ^[22]	8/8	69.89	75.29	77.63
HyQ ^[23]	8/8	69.16	74.47	76.63
EasyQuant ^[25]	8/8	66.80	62.91	69.34
本文OAQ方法	8/8	69.97	75.31	77.86

值得注意的是,随着模型规模的增大,本文 OAQ 方法在保持高精度恢复方面表现得更加稳健。在 MobileViT-xs 模型和 MobileViT-s 模型上,本文 OAQ 方法的量化精度不仅与主流动态量化方案不相上下,而且在实际应用中能够将性能损失进一步压缩至极小范围。本文 OAQ 方法未引入额外算子,且

自适应噪声补偿与量化难度迁移在训练阶段参数一旦确定后即保持不变,因此,在低算力芯片上也能稳定部署而不增加额外计算开销。通过这些实验,可以清楚地看到,本文 OAQ 方法在量化大规模模型时具有更强的鲁棒性,并且能够处理更复杂的网络架构。

在 MobileViTv2 和 EfficientFormer 等模型上, 本文 OAQ 方法在量化精度方面表现优异。如表 2 和表 3 所示, 本文 OAQ 方法在各模型上均获得最佳 Top-1 精确, 尤其在结构更复杂的 EfficientFormer 模型中, 有效抑制了异常激活值带来的精度损失。同时, 如图 2 所示, 经 INT8 量化后, 这些模型在 Jetson AGX Xavier 上推理速度显著提升, 帧率整体高于未量化模型, 且模型规模越大加速越明显, 验证了本

文 OAQ 方法在边缘设备上的高效性与适用性。

3.4 消融实验

为验证本文 OAQ 方法中 ANC 和 ADM 的作用及其协同效果, 本文对 MobileViT 系列模型进行了全面的消融实验。通过逐步移除这些模块并观察量化精度的变化, 本文分析了 ANC 和 ADM 各自的贡献及其协同作用对量化性能的提升效果。图 3 展示了各组件的重要性及其在不同模型中的性能

表 3 EfficientFormer 模型量化后精度
Table 3 Accuracy of EfficientFormer after quantization

方法	权重/激活值	Top-1精度/%		
		EfficientFormer-L1	EfficientFormer-L3	EfficientFormer-L7
全精度	32/32	80.20	82.40	83.30
FQ-ViT ^[18]	8/8	66.63	81.85	82.47
PTQ4ViT ^[19]	8/8	79.14	80.31	81.56
RepQ-ViT ^[24]	8/8	79.67	81.98	82.47
HyQ ^[23]	8/8	78.55	82.26	82.66
EasyQuant ^[25]	8/8	78.42	81.03	82.27
本文OAQ方法	8/8	80.09	82.14	83.15

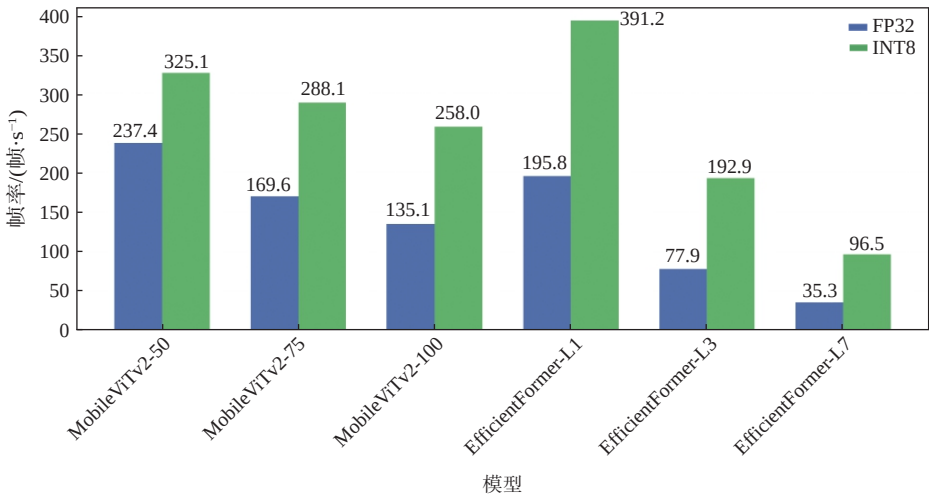


图 2 比较量化前后各模型的推理速度

Fig. 2 Comparison of inference speed before and after quantization across different models

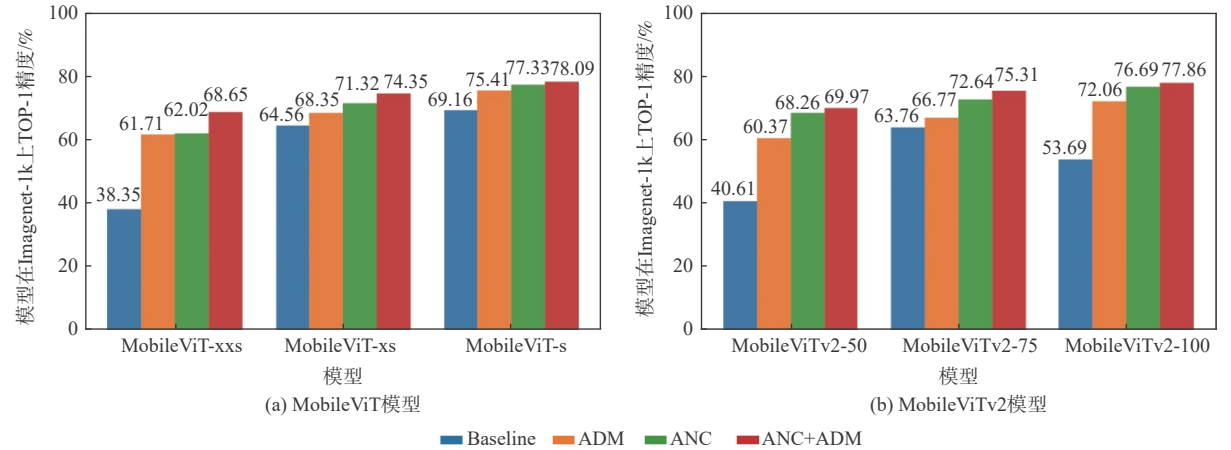


图 3 ANC 和 ADM 在 MobileViT 模型和 MobileViTv2 模型上的消融实验结果

Fig. 3 Ablation results of ANC and ADM on MobileViT and MobileViTv2 model

表现,其中,Baseline表示在未引入ANC与ADM机制时,量化后模型所达到的精度。

移除ANC的实验分析。当移除ANC后,MobileViT系列模型的量化精度出现了一定程度的下降。例如,在MobileViT-s模型上,Top-1精度下降了2.68%,在MobileViT-xs模型和MobileViT-xxs模型上分别下降了6.00%和6.94%。这表明,虽然ADM依然可以对激活值的难度迁移进行优化,但没有ANC提供的激活值平坦化作用,模型在处理异常值时会受到更大的影响,导致量化误差增加。

移除ADM的实验分析。相比于移除ANC,移除ADM对模型量化精度的影响更加显著。例如,在MobileViT-s模型上,Top-1精度下降了0.76%,在MobileViT-xs模型和MobileViT-xxs模型上分别下降了6.63%和3.03%。这表明,ADM在量化中效应更大,尤其是在处理复杂特征分布和高动态范围的激活值时,其优化效果尤为显著。没有ADM时,即使ANC对激活值分布进行了平滑处理,仍然会造成较大的量化误差。尤其是在MobileViT-xs模型和MobileViT-xxs模型这样的较小模型中,激活值分布本身已经相对复杂,移除ADM后精度下降幅度更大,进一步凸显了ADM的重要性。

ANC和ADM的协同作用。实验结果表明,ANC和ADM的协同作用显著增强了量化的整体表现。具体而言,ANC的引入不仅抑制了激活值中的异常值,还通过动态范围的收敛和平坦化,降低了激活值分布对量化的干扰,为ADM学习到最优的难度迁移因子奠定了基础。如图3所示,在包含ANC和ADM这2个组件时,MobileViT模型和MobileViTv2模型的量化精度相比移除ANC或ADM时都有所

提升,充分验证了二者的互补性和协同优化能力。

总之,ANC的异常值平滑作用为ADM提供了更加平稳的激活值分布,从而使ADM能够更高效地学习难度迁移因子。而ADM则进一步优化了难度较高的特征分布,将ANC的改进效果最大化。两者的联合优化显著增强了本文OAQ方法在量化过程中对激活值异常的鲁棒性。

3.5 激活值分布优化

激活分布中的异常值会放大动态范围,从而加剧量化误差,因此,通过抑制异常激活、使分布更加平坦,是降低量化难度的一种有效策略^[27]。为验证激活噪声补偿和自适应难度迁移在优化激活值分布、减小异常值对量化精度影响方面的有效性,本文对EfficientFormer模型中不同阶段的激活值分布进行了分析。如图4所示,图中的激活值分布分别对应于EfficientFormer模型中间阶段全连接层和最后阶段全连接层的卷积输入激活值。

首先,图4中的第1行展示了EfficientFormer模型中间阶段全连接层激活值分布的变化。如图4(a)所示,原始激活值呈现出显著的长尾分布,存在较多的异常值,这些异常值会严重影响量化精度。如图4(b)所示,在引入噪声补偿后,激活值的分布明显趋于平坦,极端值的数量显著减少,激活值的范围更加均匀。这表明,激活噪声补偿通过对激活值添加噪声补偿,成功降低了激活值的方差,并有效减小了异常值的影响。如图4(c)所示,进一步引入自适应难度迁移后,激活值的分布变得更加集中,且没有明显的异常值或长尾效应。图4(d)~图4(f)展示了EfficientFormer最后阶段全连接层的激活值分布,其优化趋势与中间阶段相似,进一步

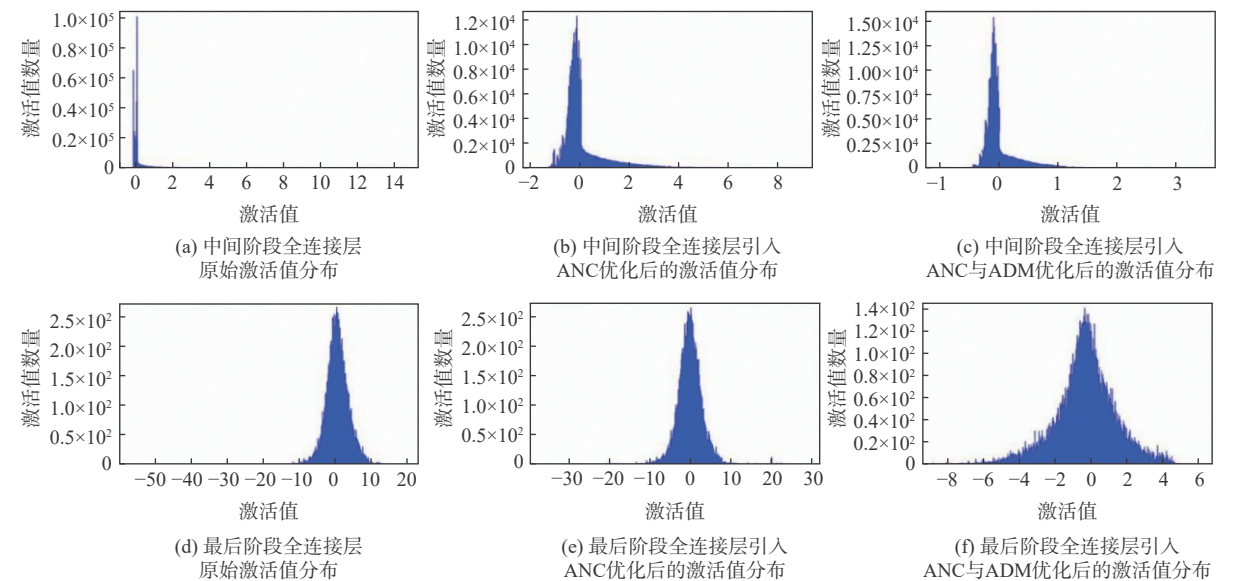


图4 Efficientformer 中间和最后阶段全连接层卷积的激活值分布变化

Fig. 4 Change in activation value distribution of fully connected layer convolutions in middle and final stages of Efficientformer

证明了激活噪声补偿和自适应难度迁移策略在不同网络深度上的鲁棒性和有效性。

如图 4(d)所示,最后阶段原始的激活值分布呈现出宽泛且低矮的形态。虽然分布形态相对对称,但峰值较低,表明激活值分散度高,存在一定数量的离群点造成动态范围的扩大,为量化带来了挑战。如图 4(e)所示,在应用激活噪声补偿后,激活值分布得到了显著的集中。其分布范围大幅收缩,峰值高度显著增加,且形态变得更加陡峭。这一结果有力地证明了激活噪声补偿通过降低激活值的离群值,有效地将分散的分布平坦化并收敛到零值附近,从而为后续的低比特量化提供了更为紧凑的量化区间。最后,如图 4(f)所示,在进一步集成自适应难度迁移策略后,激活值分布达到了最优的集中程度。分布的范围进一步被压缩,峰值高度达到最高,分布形态高度集中且对称。自适应难度迁移通过将激活值量化的难度转移到权重参数上,允许激活值本身向更利于量化的分布变化,从而实现了

对激活值分布的彻底优化,使得异常值基本消除。这种高度集中的分布形态极大地减小了量化误差的潜在风险,为实现高精度低比特量化奠定了基础。

激活值噪声补偿和自适应难度迁移在优化激活值分布、减少异常值方面发挥了关键作用。激活值噪声补偿通过降低激活值的方差和消除异常值,使激活值分布更加平坦;而自适应难度迁移则将量化难度从激活值迁移到权重中,进一步简化了激活值的量化难度,有效保障了模型的量化性能。实验结果表明,这 2 种方法的结合成功解决了输入激活值分布异常的问题,有利于减少量化误差,并为高效部署模型提供了强有力的激活值分布优化工具。

4 结 论

1) 本文 OAQ 方法通过激活值噪声补偿和量化难度迁移抑制异常值以降低量化误差,实现 CNN-Transformer 混合模型在边缘侧计算平台的高精度、快速推理。

2) 多项基于典型模型的实验表明,本文 OAQ 方法显著提升了量化后的推理精度并保持了量化模型的高效推理特性,为在无人机等计算资源受限、对推理效率存在较高要求的设备上部署模型提供了高精度、高效率的量化方案,而无需引入额外算子支持。

参考文献 (References)

[1] Dai Z, Liu H, Le Q V, et al. Coatnet: marrying convolution and attention for all data sizes[J]. *Advances in neural information processing systems*, 2021, 34: 3965-3977.

[2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.

[3] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]//*Proceedings of the Computer Vision-ECCV*. Berlin: Springer, 2014: 818-833.

[4] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2025-06-20]. <https://arxiv.org/abs/1409.1556>.

[5] Shafiq M, Gu Z Q. Deep residual learning for image recognition: a survey[J]. *Applied Sciences*, 2022, 12(18): 8972.

[6] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Piscataway: IEEE Press, 2022: 9992-10002.

[7] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2020-12-22)[2025-06-20]. <https://arxiv.org/abs/2010.11929>.

[8] Touvron H, Cord M, Douze M, et al. Training data-efficient image transformers & distillation through attention[EB/OL]. (2020-12-23)[2025-06-20]. <https://arxiv.org/abs/2012.12877>.

[9] Selva J, Johansen A S, Escalera S, et al. Video transformers: a survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(11): 12922-12943.

[10] Hashiguchi R, Tamaki T. Vision transformer with cross-attention by temporal shift for efficient action recognition[EB/OL]. (2022-04-01)[2025-06-21]. <https://arxiv.org/abs/2204.00452>.

[11] Manzari O N, Ahmabadi H, Kashiani H, et al. MedViT: a robust vision transformer for generalized medical image classification[J]. *Computers in Biology and Medicine*, 2023, 157: 106791.

[12] Mehta S, Rastegari M. MobileViT: light-weight, general-purpose, and mobile-friendly vision transformer[EB/OL]. (2021-10-05)[2025-06-21]. <https://arxiv.org/abs/2110.02178>.

[13] Maaz M, Shaker A, Cholakkal H, et al. EdgeNeXT: efficiently amalgamated CNN-transformer architecture for Mobile vision applications[C]//*Proceedings of the Computer Vision-ECCV Workshops*. Berlin: Springer, 2023: 3-20.

[14] Evangelidis G, Hu J, Li Y Y, et al. EfficientFormer: vision transformers at MobileNet speed[C]//*Proceedings of the Advances in Neural Information Processing Systems 35*. San Diego: Neural Information Processing Systems Foundation, Inc., 2022: 12934-12949.

[15] Zhang D Q, Yang J L, Ye D, et al. LQ-nets: learned quantization for highly accurate and compact deep neural networks[C]//*Proceedings of the Computer Vision-ECCV*. Berlin: Springer, 2018: 373-390.

[16] Dong P Y, Lu L, Lyu C, et al. PackQViT: faster sub-8-bit vision transformers via full and packed quantization on the mobile[C]//*Proceedings of the Advances in Neural Information Processing Systems 36*. San Diego: Neural Information Processing Systems Foundation, Inc., 2023: 9015-9028.

[17] Liu Z H, Wang Y H, Han K, et al. Post-training quantization for vision transformer[EB/OL]. (2021-06-27)[2025-06-21]. <https://arxiv.org/abs/2106.14156>.

[18] Lin Y, Zhang T Y, Sun P Q, et al. FQ-ViT: post-training quantization for fully quantized vision transformer[EB/OL]. (2021-11-27)

[2025-06-21]. <https://arxiv.org/abs/2111.13824>.

[19] Yuan Z H, Xue C H, Chen Y Q, et al. PTQ4ViT: post-training quantization for vision transformers with twin uniform quantization[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2022: 191-207.

[20] Jacob B, Kligys S, Chen B, et al. Quantization and training of neural networks for efficient integer-arithmetic-only inference[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2704-2713.

[21] Nagel M, Ali Amjad R, Van Baalen M, et al. Up or down adaptive rounding for post-training quantization[C]//Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 7197-7206.

[22] Lee J, Kwon Y, Park S, et al. Q-HyViT: post-training quantization of hybrid vision transformers with bridge block reconstruction for IoT systems[J]. IEEE Internet of Things Journal, 2024, 11(22): 36384-36396.

[23] Kim N J, Lee J, Kim H. HyQ: hardware-friendly post-training quantization for CNN-transformer hybrid networks[C]//Proceedings of the 33th International Joint Conference on Artificial Intelligence. New York: ACM, 2024: 4291-4299.

[24] Mehta S, Rastegari M. Separable self-attention for mobile vision transformers[EB/OL]. (2022-06-06)[2025-06-21]. <https://arxiv.org/abs/2206.02680>.

[25] Li Z K, Xiao J R, Yang L W, et al. RepQ-ViT: scale reparameterization for post-training quantization of vision transformers[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2024: 17181-17190.

[26] Wu D, Tang Q, Zhao Y L, et al. EasyQuant: post-training quantization via scale optimization[EB/OL]. (2020-06-30)[2025-06-21]. <https://arxiv.org/abs/2006.16669>.

[27] Chen J M, Chao Y H, Wang Y J, et al. QuantTune: optimizing model quantization with adaptive outlier-driven fine tuning[EB/OL]. (2024-03-11)[2025-06-21]. <https://arxiv.org/abs/2403.06497>.

A post-training quantization method for CNN-Transformer hybrid models on UAV platforms

ZHAO Chenggong^{1, 2}, HU Jiyuan^{1, 3}, YANG Xingyu⁴, CHEN Lei⁵, JIANG Hongxu^{3, *}

- (1. Hangzhou Innovation Institute of Beihang University, Hangzhou 310056, China;
- 2. College of Software, Beihang University, Beijing 100191, China;
- 3. College of Computer Science, Beihang University, Beijing 100191, China;
- 4. College of Computer Science (College of Software), Inner Mongolia University, Hohhot 010021, China;
- 5. Hangzhou Xiaoshan International Airport, Hangzhou 311241, China)

Abstract: In recent years, drones have played increasingly important roles in various low-altitude production and service scenarios. Due to the limited storage and computing resources of edge computing platforms, the deployment performance of vision models becomes a bottleneck for various algorithm applications. The long-tail distribution of activation values in convolutional neural network (CNN)-Transformer hybrid models causes substantial quantization accuracy deterioration, an efficient post-training quantization method is proposed, which this study addresses by suggesting an effective post-training quantization technique. This method uses activation noise compensation (ANC) and adaptive difficulty migration to suppress the impact of outliers on quantization accuracy and improve inference efficiency. According to experimental results, the proposed method achieves less than 1% accuracy loss on typical models under 8-bit quantization, and it improves inference time after quantization by at most 200%. In summary, the proposed method significantly enhances vision model inference performance on edge devices and supports model deployment on low-altitude drone platforms.

Keywords: post-training quantization; model compression; inference acceleration; outlier-aware; edge computing

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0549

细粒度语义增强的民航图文跨模态检索方法

刘姝妍^{1,2}, 何柳^{1,2,3,*}, 曾江辉^{1,2}

(1. 中国航空综合技术研究所 标准数据技术研究部, 北京 100028; 2. 中国航空综合技术研究所 工业互联网应用创新中心, 北京 100028; 3. 复旦大学 计算机科学技术学院, 上海 200438)

摘 要: 低空经济和民用航空领域在安全保障任务中高度依赖图像与文本等跨模态信息的高效关联。然而, 主流跨模态检索模型虽然在通用数据集上表现优异, 但对细粒度语义理解能力要求较高的领域表现不佳。为解决以上问题, 基于可获取的民航领域数据集, 提出一种细粒度语义增强的跨模态检索方法, 构建涵盖数据处理、模型构建与训练的完整流程。基于多模态大模型对文本描述进行优化增强, 构建涵盖丰富语义信息的跨模态检索数据集; 基于主流跨模态检索框架设计模型, 引入细粒度特征提取模块、类监督及关键语义信息掩码等策略, 增强模型的细粒度语义特征表达能力。2个数据集上的实验结果验证了所提方法的有效性, 可为低空经济场景下的跨模态检索需求提供可借鉴的技术路径。

关键词: 低空经济; 民用航空; 跨模态检索; 文本描述优化; 细粒度语义增强

中图分类号: V219; TP391.3

文献标志码: A

文章编号: 1001-5965(2026)09-3075-14

低空经济作为战略性新兴产业的重要引擎, 依托各种有人驾驶、无人驾驶等新型航空器, 正在加速拓展物流配送、应急救援和巡检执法等多样化应用场景^[1]。与传统民用航空领域类似, 低空经济的业务场景同样面临高度复杂的运行环境和严格的任务时效, 其运行与安全保障都依赖海量的多源信息, 包括结构化的飞行数据, 如机型、轨迹、飞行计划等, 以及非结构化的维修记录、技术文档和故障图像等。在面对故障诊断、维修支持或应急处置等安全保障任务时, 尽管这些图文异构的数据蕴含丰富的知识价值, 但由于缺乏高效的语义关联机制, 一线人员往往需要耗费大量时间在不同系统和资料之间切换, 影响任务的响应效率与决策质量。

针对这一问题, 传统的做法主要分为两类: 一类是基于人工经验的方式, 由具备专业背景的人员通过主观判断, 对现场图像内容进行识别并匹配相

关文档。这类方式虽然在经验丰富的人员手中能够保证一定的准确性, 但处理效率低, 难以在大规模、高频次任务环境中推广。此外, 低空经济作为处于成长期的新兴产业, 尚未建立完备的知识体系和技术规范, 部分从业人员缺乏系统化的知识支撑, 在复杂任务中难以形成精准判断。另一类是基于统计学习方法, 通过相关分析或因子分析探索不同模态的共享语义结构。这类方法能够在一定程度上减轻人工负担, 但在处理大规模非结构化数据和捕捉细粒度语义等方面存在明显局限。

在此背景下, 人工智能在赋能低空经济高质量发展的过程中展现出了强大的潜力^[2], 其中, 基于深度学习的图文跨模态检索技术^[3-4]为解决上述问题提供了重要技术路径, 其核心目标是将图像与文本映射到共享语义空间, 实现多模态数据之间的隐式对齐与语义互通。相比传统方法, 跨模态检索技术

收稿日期: 2025-08-06; 录用日期: 2025-09-29; 网络出版时间: 2025-10-09 13:51

网络出版地址: link.cnki.net/urlid/11.2625.V.20251009.1203.001

基金项目: 工业和信息化部民机专项科研项目 (MJ22-3N21)

*通信作者. E-mail: heliu1219@126.com

引用格式: 刘姝妍, 何柳, 曾江辉. 细粒度语义增强的民航图文跨模态检索方法 [J]. 北京航空航天大学学报, 2026, 52 (9): 3075-3088.

Liu S Y, He L, Zeng J H. Fine-grained semantic-enhanced cross-modal image-text retrieval method for civil aviation [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3075-3088 (in Chinese).

可以对多源数据进行整合分析,快速从知识库中定位所需的操作规程、故障案例和应急预案等关键信息,从而提供智能决策,提升低空经济场景下安全故障任务的执行效率。

目前,图像-文本跨模态检索领域常用的 MS-COCO^[5]、Flickr30k^[6] 等数据集为模型训练提供了丰富的图文对资源,主流模型如对比语言-图像预训练(contrastive language-image pre-training, CLIP)模型^[7]、融合前对齐图像与文本表征(align the image and text representations before fusing, ALBEF)模型^[8]和快速语言-图像预训练(fast language-image pre-training, FLIP)模型^[9]等在这些数据集上取得了较高精度,展现了优异的跨模态学习能力。然而,将这些模型直接应用于低空经济和民航等专业领域时,面临 2 方面的问题:①在数据层面,通用数据集缺乏对航空器、飞行状态、运行环境等细粒度语义的刻画,难以覆盖专业任务所需的知识要素,即便是从互联网获取的民航图文对数据,也普遍存在文本描述泛化及语义对齐度低等问题,导致训练样本噪声较高,限制了模型的有效学习;②在模型层面,主流跨模态检索模型主要依赖于对图像与文本的粗粒度语义建模,难以捕捉图文内容中的细粒度差异,例如,模型不仅需要区分“飞机”与“直升机”这样的宏观类别,还需要识别图像中具体的航空器型号和部件状态等,以支持面向特定任务的知识检索与决策。现有模型在这种需要精细语义对齐的任务上表现不佳,导致其在垂直领域场景下的泛化能力和实际应用价值受到限制。

为此,本文提出了一种细粒度语义增强的图像-文本跨模态检索方法,考虑到通用航空是低空经济的发展主体^[10],与低空经济在视觉场景和运行过程等方面存在一定的相似性,本文充分利用现有民航的数据资源,构建带有细粒度语义描述的图像-文本数据集,并在此基础上训练具备领域适用能力的跨模态检索模型。该方法不仅能够有效支撑民航场景下的知识检索与任务决策,也为未来向低空经济场景的迁移应用提供了可借鉴的技术路径。本文的主要工作与创新点如下:

- 1) 提出一种细粒度语义增强的图像-文本跨模态检索方法,构建涵盖数据处理及模型设计和训练的完整流程,实现在跨模态领域下精细的文本与视觉语义匹配;
- 2) 在数据集处理方面,引入多模态大模型对文本描述进行细粒度优化,生成语义信息丰富、结构清晰的图像-文本对数据集,显著提升训练数据的质量与跨模态对齐的一致性;

- 3) 在模型设计方面,基于主流跨模态检索框架进行针对性改进,引入细粒度特征提取模块、类监督与关键语义信息掩码策略,增强模型对细粒度语义信息的表征能力,提升模型的跨模态检索效果。

1 国内外相关工作

1.1 基于传统统计分析的跨模态检索方法

早期的跨模态分析主要依赖于统计学习方法构建不同模态间的公共表示空间。例如,利用典型相关分析(canonical correlation analysis, CCA)学习 2 个模态变量投影后的最大相关方向来进行关联建模。然而,该方法受限于线性假设,且无法对原始变量进行有效重构。随后,Li 等^[11]提出的跨模态因子分析(cross-modal factor analysis, CFA)通过最小化不同模态样本之间的距离来探索其背后的潜在共享因子,增强了模态间语义一致性。

为提升对非线性关系的建模能力,Zhang 等^[12]进一步引入核方法,提出了核典型相关分析(kernel CCA, KCCA),利用核技巧将原始特征映射到高维空间,有效捕捉模态间更复杂的语义相关性。这些传统方法为跨模态检索提供了理论基础,但在大规模非结构化数据处理与语义建模方面仍存在瓶颈。

1.2 基于深度学习的跨模态检索方法

随着神经网络的发展,跨模态检索从浅层特征对齐迈向了语义级别的深度融合检索。Frome 等^[13]提出的 DeViSE 模型引入卷积神经网络(convolutional neural network, CNN)与 Skip-Gram 模型^[14]联合训练图像与文本嵌入表示,通过排序损失学习共享语义空间。随后, Lee 等^[15]提出的堆叠交叉注意力网络(stacked cross attention network, SCAN)模型采用区域级特征与单词序列间的交叉注意力,将每个局部特征与来自另一个模态的所有局部特征进行比对,得到局部特征间的相似度矩阵,最终计算得到不同模态数据之间的相似度。该方法可以获取图像-文本间细粒度的对应关系,但这种细粒度交互方式的计算复杂度非常高,模型的推理速度慢。

为提升推理效率,Wehrmann 等^[16]提出自适应跨模态嵌入(adaptive cross-modal embeddings, ADAPT)模型,基于一种模态的全局信息改进另一种模态的表示,此外还引入了中央凹模块取代堆叠式交叉注意力机制,显著减少计算负担。Qu 等^[17]针对图像-文本匹配中的多视角描述问题,提出上下文感知多视角摘要网络(context-aware multi-view summarization network, CAMERA)模型,设计一种多视角匹配方案,将多视角的图像特征与相应的文本特征进行匹

配, 结合自注意力机制和门控机制, 自适应地提取上下文增强的片段特征。

近年来, 视觉语言预训练(vision-language pre-training, VLP)已成为跨模态任务的主流技术路线, 通过在海量图像-文本对上进行大规模预训练, 模型在多个下游任务上展现出强大的迁移能力。代表性模型之一是通用图像-文本表征(universal image-text representation, UNITER)模型^[18], 其对 4 个图像-文本数据集(COCO、Visual Genome、Conceptual Captions 和 SBU Captions)使用 4 种预训练任务进行大规模的联合学习, 包括图像-文本匹配、词区域对齐、掩码语言建模和掩码区域建模任务, 在当时的多个图像-文本对数据集上取得最先进的检索效果。Li 等^[19]提出了对象语义对齐预训练(object-semantics aligned pre-training, Oscar)模型, 使用图像中检测到的对象标签作为锚点来建立图像和文本模态之间的联系, 这种方法简单高效, 模型在诸多视觉-文本任务数据集上获得了超越 UNITER 的表现。CLIP^[7]模型作为多模态大模型的经典之作, 其基于互联网上获取到的 4 亿组图像-文本对, 使用对比学习任务进行训练, 在 MS-COCO 和 Flickr30k 数据集上取得了当时的最优性能。ALBEF 模型^[8]在 CLIP 模型的基础上通过引入图文匹配和掩码语言建模任务并结合动量蒸馏, 有效缓解了大规模图文数据中的噪声问题。

随后, 研究者们提出了一系列改进模型, 自举语言-图像预训练(bootstrapping language-image pre-training, BLIP)模型^[20]和对比式描述生成器(contrastive captioner, CoCa)模型^[21], 通过结合生成式监督与对比学习, 在跨模态检索和图文生成等任务上取得了效果提升。FLIP 模型^[9]通过低分辨率预训练显著降低了算力开销, 在保持精度的前提下提升了训练效率。细粒度语言引导的图像表征(fine-grained language-informed image representations, FLAIR)模型^[22]通过精细描述结合文本条件的注意力机制以增强模型对复杂语义场景的理解能力。基于 Sigmoid 损失的语言-图像预训练(sigmoid loss for language-image pre-training, SigLIP)系列模型^[23]改进了损失函数和训练策略, 提升了多语种跨模态检索的效果。蒸馏式 CLIP(distill CLIP, DCLIP)模型^[24]借助蒸馏和交叉注意力机制提升了图文匹配精度。增强型语言-图像预训练(enhanced language-image pre-training, ELIP)模型^[25]则利用文本引导的视觉提示作为重排序模块, 有效增强了模型在复杂和分布外场景下的检索表现。这些工作共同推动了跨模态检索向更高效与更具普适性的方向发展。

2 民航领域图文检索数据集构建

2.1 现有数据集的问题

当前主流跨模态检索方法在预训练阶段通常依赖通用图像-文本对数据集及从互联网上爬取的大规模数据集, 如 LAION-5B 数据集^[26]。这类数据虽然规模庞大, 并能有效赋予模型基本的图文对齐能力, 但在民航等专业领域存在显著局限: 首先, 通用数据集缺乏航空器型号、飞行状态、运行环境等专业概念, 难以覆盖民航任务所需的关键知识要素; 其次, 互联网图文数据虽然数量庞大, 但语义噪声较高, 文本描述与图像内容往往弱相关, 难以支撑模型的细粒度语义建模需求。

具体来说, 在互联网可收集的民航领域图像-文本对主要分为 2 类: 第 1 类来源于社交平台账户, 用户将日常拍摄的飞机图像发布在网络上并配文, 如图 1(a)所示。发布该图像的用户配文是“伊丹航空, 大飞机”。该配文包含了飞机所属的航空公司, 但是通常十分简略, 或者包含冗余信息, 难以用于跨模态检索。第 2 类来源于民航领域的报告文件或新闻报道等, 如图 1(b)所示。该图像的文本描述为“等待救援的事故飞机”, 此类数据的文本描述通常与图像内容紧密相关, 但缺少关键信息, 如事故类型及周边环境细节等, 无法满足细粒度语义的检索需求。



(a) 来自于社交平台 (b) 来自于报告文件/新闻报道

图 1 互联网中的民航领域图像-文本对数据示例

Fig. 1 Examples of image-text pairs in civil aviation field on internet

针对上述问题, 为获取能够更精准地描述图像内容的文本描述用于后续跨模态检索模型的训练, 本文提出了一种基于大语言模型(简称大模型)的文本描述优化方案, 利用先进的多模态大模型, 通过其强大的图像理解和自然语言生成能力, 在保留原始文本关键信息的基础上, 为图像生成更为详细且符合领域语义的描述。同时, 将图像附近的上下文段落作为引导信息, 帮助大模型进行更深层次的语义理解, 从而生成粒度更细、领域相关性更强的文本描述。

如图 2 所示的图像为例, 其原始描述仅为“事故飞机”, 该描述过于简单, 虽然从图像中可以感知

事故背景,但文本描述本身并未提供足够的语义细节。在跨模态检索任务中,模型通常只能依赖文本本身进行图像内容理解,缺乏上下文信息辅助,因此,难以实现高精度的检索。相比之下,橙色框所示的文本描述“一架带有马来西亚航空公司标志的大型飞机在地面上,周围有急救车和人员”,则更为完整地刻画了图像信息,不仅明确了飞机所属航空公司,还补充了场景与环境细节,更贴近图像的真实语义内容。这类详细、结构清晰的文本描述对于提高跨模态检索模型的语义理解能力和检索精度至关重要。详细的实现流程见 2.3 节。

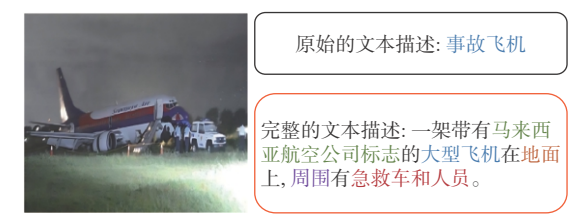


图 2 图像-文本对数据示例

Fig. 2 Example of image and text pair data

2.2 数据集预处理

本文选取公开收集的《世界民航事故调查跟踪系列报告》^[27]及 FGVC Aircraft 民航细粒度图像分类数据集^[28],构建跨模态检索数据集。《世界民航事故调查跟踪系列报告》数据集涵盖民航领域的典型事故案例,其图像具有较强的领域相关性,但原始文本描述不够精细,缺乏细粒度的语义信息。FGVC Aircraft 数据集包含大量已标注的航空器图像,覆盖多种机型及细分类别,能够为模型提供丰富的航空器分类信息,但不具备面向跨模态检索任务所需的文本描述信息。下面分别介绍构建流程。

1) 民航事故数据集

该数据集中的图像-文本对从《世界民航事故调查跟踪系列报告》中提取,首先,将事故报告的 PDF 文档批量转化为易于解析的 WORD 文档;其次,使用自定义的 xpath 规则对文档的章节、段落、包含的图像、图像位置等信息进行解析;随后,遍历文档中的所有图像,并在图像所处段落周围寻找最近的包含“图”字的文本段落作为该图像的图题。最终,图题格式为“图像所在事故报告名、标题名、图像自身的图题”,三部分内容通过“#”进行分隔,如某个提取后图像的图题为《世界民航事故调查跟踪 1002#美国航空公司 MD-82 飞机圣路易斯发动机起火事故调查报告#图 1: 灭火后的发动机》(在后续实验过程中被译制为英文进行模型训练)。

为丰富文本语义信息,本文在提取图题的同时

提取了该图像的上下文段落,用于后续基于大模型的文本描述生成和增强。为确保原始图像、图题和上下文段落的匹配关系准确,本文引入了人工检查与修正环节。该环节由 3 名具备民航专业背景的人员分工完成,采用逐条核验的方式,对全部提取的内容进行人工检查与修正。检查内容包括图题是否正确对应图像、上下文段落是否与图像语义相关等,若发现不相关则进行修正或删除。经统计,约有 20% 的样本需要人工调整,调整内容主要集中在对于上下文段落的修正,整体人工检查与修正环节耗时约 15 h。最终共获取 1 656 组图像样本、相关图题及上下文段落,包括民航事故现场图像及飞机航线图、机场布局图等非实物图像,并按照 8:2 的比例划分训练集和测试集。本部分数据集记为“Accidents”。

2) FGVC Aircraft 数据集

FGVC Aircraft 数据集专注于飞机种类的精细区分,共包含 10 000 张高质量飞机图像,涵盖 100 个不同的型号类别。官方将数据集按照 1:1:1 的比例进行划分,其中,训练集 3 334 张,验证集和测试集分别 3 333 张。此外,该数据集提供 4 种不同粒度的标签体系,包括制造商(如波音)、家族(如波音 737)、变体(如波音 737-700)及模型(如波音 737-76J),由于某些型号在视觉上几乎没有区别,这个粒度不用于评估。在本实验中采用的粒度是变体,即每个类别对应一个特定的飞机型号。每个变体类别均包含 100 张图像,具有良好的类内一致性和类间区分性,适合用于细粒度检索任务中。

针对数据特点,民航事故数据集与 2.1 节介绍的互联网数据集中“文本缺乏细节”的第 2 类问题相似,而 FGVC Aircraft 数据集则与“文本过于简略”的第 1 类问题一致,二者均无法支撑高精度的跨模态检索。本文在此基础上对两类数据进行文本描述生成与语义增强以作为细粒度跨模态检索模型的数据支撑。表 1 总结了这 2 类数据存在的问题。

表 1 民航领域数据集存在的问题			
Table 1 Summary of problems with civil aviation datasets			
数据集	图像	文本	类别
Accidents	存在非实物图像	粗粒度	无
FGVC Aircraft ^[28]	图像内容高度相似	无	有

2.3 数据集构建

2.3.1 文本描述生成与增强

通过 2.2 节方法提取出的图题可以作为图像的初始文本描述,但直接用跨模态检索模型的训练

文本存在严重局限。为此, 本文引入具有视觉-语言理解能力的多模态大模型对图像内容进行精细化文本描述生成。自 2018 年 OpenAI 推出 GPT 模型^[29]以来, 该系列模型通过在大规模无监督文本数据上的预训练, 构建了强大的语言建模能力, 推动了自然语言处理领域在文本生成、问答系统与对话交互等方面的突破性发展。在此基础上, OpenAI 于 2024 年 5 月发布了新一代多模态大模型 GPT-4o, 其采用统一架构处理文本、图像、语音等多模态输入, 具备更强的图像理解和自然语言生成能力, 适用于跨模态语义建模任务。因此, 本文基于 GPT-4o 大模型并精心设计提示词构建文本描述增强机制。首先, 生成覆盖整体语义的基础描述; 其次, 考虑到较长文本描述包含丰富的细节和对图像的全面理解, 能够提供比短区域描述更多的信息^[30]; 再通过细节补充与语言优化提升描述的准确性, 从而在数据构建阶段有效提升图像-文本对的表达质量, 增强下游检索模型的语义感知能力。

本文将数据集中的图像分别输入到 GPT-4o 模型并以少样本提示的方式进行文本描述生成, 具体流程如图 3 所示。对于一幅图像, 首先, 生成基础文本描述, 要求 GPT-4o 准确呈现图像的客观内容。随后, 生成增强文本描述, 提示词更加精细, 且会基于数据集的原始图题提供给模型一些额外信息。对于不同的数据集, 提示词并不相同。具体而言, 在民航事故数据集中, 本文在与 GPT-4o 的对话过程中会告诉 GPT-4o 该事故图像出自哪一场事故, 让 GPT-4o 基于该信息对图像进行增强文本描述生成。在 FGVC Aircraft 数据集中, 本文则提供给 GPT-4o 该图像对应的飞机类别信息, 让 GPT-4o 基于该信息进行生成。以民航事故数据集为例,

基于 GPT-4o 的文本描述生成与增强阶段的基础提示词如图 4 所示, 同时要求其不要描述虚构的内容。



图 4 基于 GPT-4o 的文本描述生成与增强提示词示例
Fig. 4 Example of caption generation and enhanced prompts based on GPT-4o

考虑到更精细的提示词可能带来偏离事实的风险, 本文在生成增强描述后引入自动化的一致性检查环节。已有研究表明, GPT-4 等强大的大模型与人工专家作为评判者具有较高的一致性^[31], 因此, 在采用 GPT-4o 生成增强文本描述后, 将其与原始图像信息、基础文本描述进行对比并输出 2 项内容: 事实一致性得分(1~10 分)和是否存在细节信息增益的判断。仅当一致性得分 ≥ 8.5 分(允许少量噪声)且存在细节信息增益时, 才保留增强文本描述, 从而确保文本描述的丰富性与事实性。

此外, 尽管在提示词中明确要求 GPT-4o 不要生成虚构的内容, 但在实际生成过程中, 模型仍可能使用带有推测性的词语引入幻觉信息。为进一步提升生成文本描述的质量, 本文参考 LLMDeT 方法^[30]中对于生成文本描述的质量保证策略, 引入了后处理环节, 包括删除带有典型猜测性词的描述子句, 如“建议”、“可能”和“似乎”等, 以及过滤重复和无意义的输出, 例如: “图中是一架飞机, 一架飞机”等。

2.3.2 类别标签构建

民航事故数据集仅包含图像与图像标题, 缺乏类似 FGVC Aircraft 数据集的具体的类别信息, 同时该数据集中存在较多的非实物图, 可能会限制图像语义的表达能力。为解决这一问题, 本文为民航事故数据集中的所有图像-文本对构建了类别标签, 不仅能够有效区分实物图与非实物图, 还能对图像内容进行更细粒度的语义结构化标注, 从而为跨模态检索模型提供更明确的先验知识, 支撑后续的模式训练与检索任务。

通过对事故报告中提取出的图像-文本对数据进行人工分析, 发现事故报告中的图像总体上可以分为事故飞机、涉事部件、飞行轨迹、飞行环境、事故现场痕迹等五大类。类别标签具体构建过程

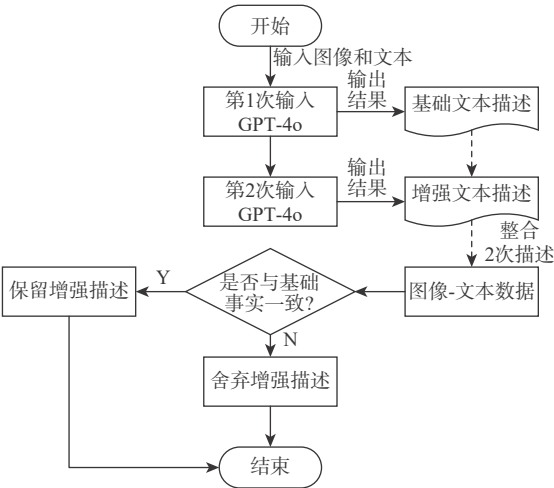


图 3 基于 GPT-4o 的文本描述生成和增强流程
Fig. 3 Text description generation and enhancement process based on GPT-4o

如下:首先,使用词云工具对提取出的图像标题进行分析,按照出现频率对标题中的短语进行排序;其次,手动标注最常出现的 200 条短语所对应的具体图像类别,并对所获得的短语进行模板增强,作为每个类别的语义闭包,如表 2 所示;最后,使用在通用领域预训练的文本相似度 Sentence BERT 模型^[32]计算图像标题和每个类别的语义闭包之间的语义相似度,取相似度最高的类别作为该图像的分类,若相似度均低于一定阈值,则归类为“其他”类别。

至此,基于以上流程,本文构建了一个具备丰富细粒度语义描述信息和对应类别标签的民航领域图像-文本数据集。该数据集有效缓解了《世界民航事故调查跟踪系列报告》^[27]数据集中原始图题描述粒度较粗的问题,同时为原本不包含文本描述的 FGVC Aircraft 数据集生成了高质量的文本描

表 2 语义闭包部分示例	
Table 2 Some examples of semantic closure	
语义闭包	部分示例
事故飞机	民机, 货机, 飞机, 客机, 事故飞机, 涉事飞机
涉事部件	发动机, 高压压气机, 风挡玻璃, 冷却圈, 氧气软管
事故现场	机场图, 事故现场, 坠机现场着陆轨迹, 接地痕迹
飞行轨迹	飞行轨迹, 进近轨迹, 飞行剖面, 进近图, 着陆过程
飞行环境	卫星云图, GOES雷达图, 跑道布局图, 红外卫星云图

述。图 5 展示了本文构建数据集的示例, 并与 Flickr30k 跨模态检索数据集进行了对比, 可以看出, 相较于以通用场景为主且语义粒度较粗的公开数据集, 本文数据集在民航领域专业性、类别可区分性和语义描述粒度等方面具有显著优势, 能够为后续面向细粒度语义提取的跨模态检索模型训练提供坚实的数据基础。



图 5 民航图文数据集和 Flickr30k 图文数据集示例

Fig. 5 Examples of civil aviation image-text dataset and Flickr30k image-text dataset

3 本文模型

3.1 模型架构

本文在对主流图像-文本跨模态检索框架研究的基础上, 针对 2.3 节构建的民航领域图像及其生成的细粒度文本描述, 提出了一种针对性的细粒度语义特征增强的跨模态检索模型, 包括细粒度语义增强的视觉特征提取模块、关键信息掩码的文本特征抽取模块及多模态特征融合模块等, 模型的整体架构如图 6 所示。

在训练数据准备阶段, 模型的输入包括图像、

文本描述及通过 2.3.2 节构建的类别标签。整体流程上, 模型由图像分支、文本分支和多模态融合分支 3 部分构成。在图像分支中, 输入图像一方面经过残差卷积网络 (ResNet)^[33] 提取浅层特征 v^s , 随后, 输入到图像细粒度显著特征提取模块, 得到细粒度局部特征 v^f , 同时, ResNet 特征 v^s 在类监督 (class supervision, CLS) 损失下进行类别判别监督。另一方面, 图像经过基于自注意力机制的全局特征提取模块获取全局语义特征 v^g , v^f 和 v^g 两者融合形成视觉表示 F_v ; 在文本分支中, 输入的文本描述经过关键信息掩码的文本特征提取模块得到文本表

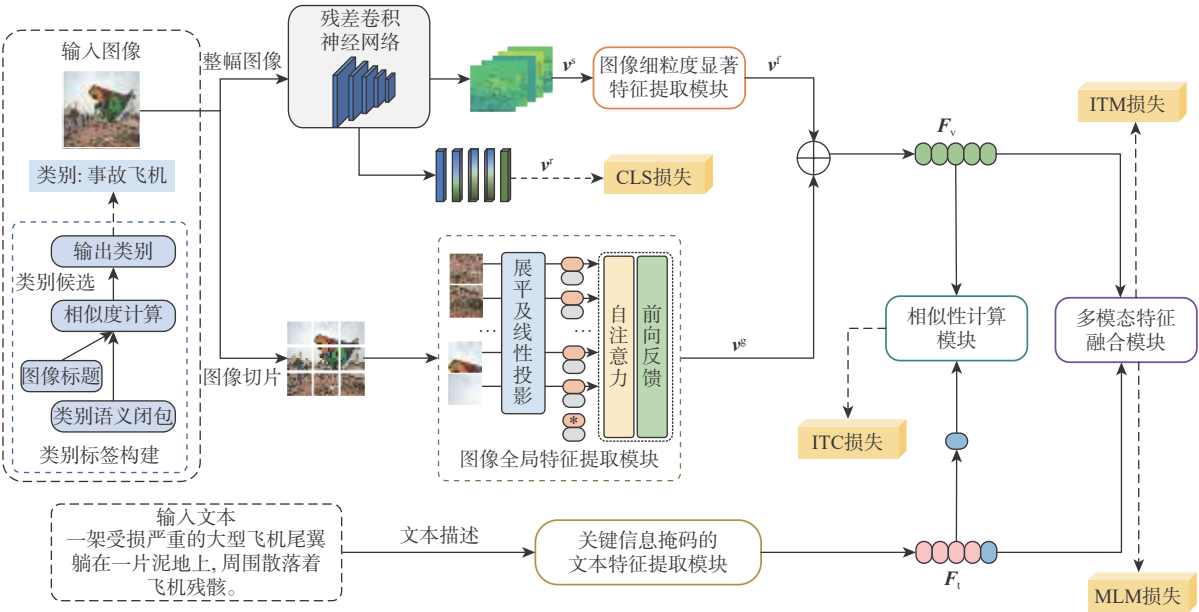


图 6 本文模型架构

Fig. 6 The proposed model architecture

示 F_t ; 在多模态融合分支中, 视觉表示 F_v 和文本表示 F_t 在相似性计算模块中计算图文对比 (image-text contrastive, ITC) 损失, 实现跨模态空间的对齐, 同时二者在多模态特征融合模块中进行交互学习, 并在图文匹配 (image-text matching, ITM) 损失与掩码语言建模 (masked language modeling, MLM) 损失的联合监督下进行优化。最终, 本文模型能够在捕捉图像-文本全局语义关联的同时, 兼顾对于细粒度信息的判别能力。在推理阶段, 模型通过相似性计算模块对输入图像特征 F_v 与文本描述特征 F_t 进行相似度度量, 从而实现跨模态检索。

3.1.1 细粒度语义增强的视觉特征提取模块

民航图像通常包含大量具有专业语义的局部视觉元素, 例如, 机身型号、维修部件及故障迹象等, 这些细节对于区分外观相似但语义不同的图像至关重要。然而, 传统仅依赖全局特征的表示方法往往会忽视这些关键细节, 且难以有效建模图像的全局上下文信息。为此, 本文在视觉特征提取部分设计了细粒度语义增强模块, 包含 2 个子模块: ①基于 ResNet 模型的细粒度显著特征提取模块, 利用卷积的局部感受野与平移不变性, 有助于突出区域细节与类别敏感特征; ②基于自注意力机制的全局特征提取模块, 用于感知图像的全局语义信息与长距离依赖关系。二者结合使得模型能够同时捕捉局部差异与全局关联, 从而获得更具判别力的视觉表示。

细粒度显著特征提取模块的结构如图 7 所示, 输入图像 $I \in \mathbf{R}^{H \times W \times C}$, 首先, 通过 ResNet 模型提取浅层特征 v^s 。随后, 为捕获不同感受野下的区域细节特征, 依次进行多尺度卷积处理, 得到多个尺度的

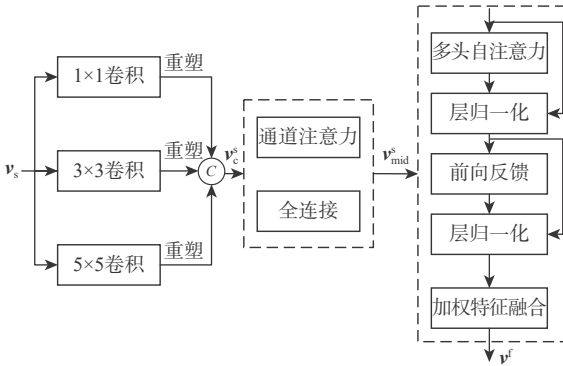


图 7 细粒度显著特征提取模块

Fig. 7 Fine-grained salient feature extraction module

响应特征:

$$v_i^s = \text{Pool}_i(\text{ReLU}(\text{BN}(\mathbf{W}_k * v_s))) \quad i = 1, 2, 3 \quad (1)$$

式中: $\mathbf{W}_k$ 表示卷积核权重; $*$ 表示卷积运算; $\text{Pool}_i(\cdot)$ 表示自适应池化操作, 将不同尺度特征调整至相同空间大小; $\text{BN}(\cdot)$ 为批归一化操作。接着, 将多尺度特征在通道维度拼接并输入通道注意力模块 M_{SE} ^[34], 以自适应调整各通道的重要性:

$$v_{\text{se}}^s = (\mathbf{I} + M_{\text{SE}}(v_c^s)) \otimes v_c^s \quad (2)$$

式中: $\mathbf{I}$ 为全 1 向量; $\otimes$ 表示逐元素相乘; v_c^s 表示多尺度响应特征在通道维度拼接后形成的融合特征; v_{se}^s 表示经通道注意力加权增强后的多尺度显著特征。加权后的特征经全连接层降维得到 v_{mid}^s , 并通过多头自注意力机制进一步融合上下文信息, 最后得到细粒度显著特征表示 v^f :

$$\begin{cases} v_1^f = \text{LN}_1(v_{\text{mid}}^s + \text{Att}(v_{\text{mid}}^s)) \\ v_2^f = \text{LN}_2(v_1^f + \text{FFN}(v_1^f)) \\ v^f = \mathbf{W}_f \cdot v_2^f \end{cases} \quad (3)$$

式中: $\text{LN}(\cdot)$ 表示层归一化; $\text{Att}(\cdot)$ 为多头自注意力机制, 用于建特征间的上下文依赖; $\text{FFN}(\cdot)$ 表示前馈神经网络; $\mathbf{W}_i$ 为可学习的加权矩阵, 用于最终的特征映射。

考虑到民航场景中往往存在密集区域及多语义要素, 单纯依赖局部特征容易导致全局上下文缺失, 限制模型的跨模态语义对齐能力。在视觉特征提取阶段, 同步引入基于自注意力机制的视觉Transformer(vision Transformer, ViT)模型^[35]作为全局特征抽取模块。ViT具备良好的全局建模能力, 能够捕获图像中远距离区域的依赖关系。

具体而言, 如图6所示, 输入图像首先经过区域块层级切分, 被划分为 N 个固定大小的图像块, 每个图像块被展平为嵌入向量, 构成初始图像块序列表示:

$$\mathbf{I}_0 = [\mathbf{x}_{\text{cls}}; \mathbf{x}_1; \mathbf{x}_2; \cdots; \mathbf{x}_N] + \mathbf{E}_{\text{pos}} \quad (4)$$

式中: $\mathbf{x}_{\text{cls}}$ 为可学习的类别标记嵌入向量; $\mathbf{x}_i$ 为第 i 个图像块的嵌入表示; $\mathbf{E}_{\text{pos}}$ 为位置编码。该序列随后输入至由多层Transformer编码器堆叠构成的ViT模块中, 提取输出向量作为全局语义特征表示:

$$\mathbf{v}^g = \text{ViT}(\mathbf{I}) \quad (5)$$

该表示融合了图像全局范围内的空间语义信息, 为跨模态语义对齐提供上下文一致的图像表征。

为融合2种互补的视觉特征, 将细粒度特征 $\mathbf{v}^f$ 和全局特征 $\mathbf{v}^g$ 进行拼接构建最终的视觉表示 $\mathbf{F}_v$:

$$\mathbf{F}_v = \mathbf{v}^f \oplus \mathbf{v}^g \quad (6)$$

式中: $\oplus$ 表示特征拼接。

融合后的视觉特征 $\mathbf{F}_v$ 具备丰富的全局语义信息, 并对关键局部信息进行了增强, 从而提升模型在面对精细化文本描述时的视觉语义表达能力。

3.1.2 关键语义信息掩码的文本特征提取模块

3.1.1节生成的细粒度文本描述包含较多与航空器型号、场景要素和周边环境等相关的语义信息。若采用传统的文本特征提取方式, 容易导致这些关键信息被稀释, 削弱视觉与文本特征的对齐效果。为此, 本文构建了关键信息掩码的文本特征提取模块, 以实现关键语义的增强表征。

在文本编码方面, 本文采用预训练的BERT模型^[36]提取文本特征, 利用其良好的上下文建模能力, 增强模型对长距离文本上下文语义的理解。考虑到文本描述中蕴含的语义信息通常比视觉端要少, 过深堆叠的模型存在过拟合的风险, 因此, 选取BERT模型的前6层作为编码模块, 以获取恰当的特征表达。具体而言, 输入文本 T 经分词切分与嵌入表示后得到序列 $\mathbf{T}'$, 与位置编码 $\mathbf{E}_{\text{pos}}$ 相加后输入

BERT模型, 得到文本表示 $\mathbf{F}_t$:

$$\mathbf{F}_t = \text{BERT}(\mathbf{T}' + \mathbf{E}_{\text{pos}}) \quad (7)$$

基于民航数据集的处理结果, 对民航文本中出现的高频关键词与类别词等, 本文进一步引入关键信息语义掩码策略, 结合随机掩码策略, 共同提升模型在处理精细文本描述时对关键语义单元的建模能力。对于处理后的文本描述序列, 本文首先进行随机候选标记, 以10%的概率随机选择部分位置为候选掩码点, 其次, 进行预设候选标记, 对于民航事故数据集, 基于2.3.2节中获取的高频短语词典, 选择其中的词作为候选掩码点; 对于FGVC Aircraft数据集, 以类别词作为候选掩码点, 随后进行掩码操作, 对已标记的候选位置以一定的概率进行随机替换或保留原词, 如图8所示。构造的掩码输入序列可表示为

$$\hat{\mathbf{T}} = \text{Mask}(\mathbf{T}', M_{\text{rand}} \cup M_{\text{pre}}) \quad (8)$$

式中: M_{rand} 为随机采样的掩码候选点; M_{pre} 为由预设词典确定的掩码候选点, $\text{Mask}(\cdot)$ 表示进行掩码操作。

通过上述预设引导与随机引导相结合的方式, 模型不仅能够聚焦文本中的关键语义信息, 还在面对冗余信息或细粒度异构目标时具备更强的判别能力。

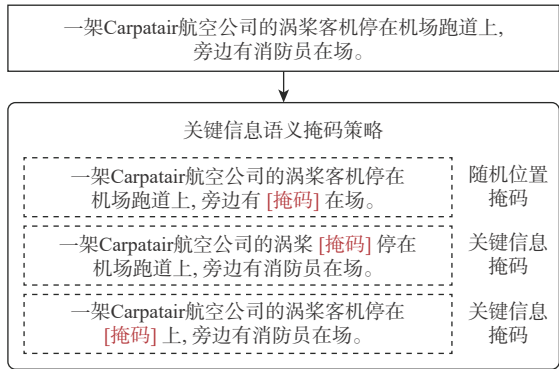


图8 关键信息语义掩码策略

Fig. 8 Key information semantic masking strategy

3.1.3 多模态特征提取模块

尽管视觉和文本特征可以分别通过独立的编码分支获得, 但由于模态间的表示差异显著, 直接用于相似度计算往往难以实现语义空间的对齐, 尤其是在涉及精细文本描述与多尺度图像目标的场景中, 问题更加突出。为增强视觉与文本特征表达的一致性, 本文在视觉与文本编码模块之后引入多模态特征提取模块, 包括相似性计算子模块和多模态特征融合子模块。

相似性计算子模块用于计算视觉特征 $\mathbf{F}_v$ 与文本特征 $\mathbf{F}_t$ 间的相似度, 并在训练阶段构建ITC损失函数, 引导模型学会将匹配的图像-文本对在特征

空间中靠近,不匹配的则拉远,从而保证视觉与文本特征在表示空间中的一致性。多模态特征融合子模块依托 BERT 模型的后六层构建,对视觉特征 F_v 与文本特征 F_t 进行深度交互建模。具体而言,该模块以 F_t 序列作为输入嵌入,随后在 BERT 模型各层中引入视觉表示 F_v ,并将其作为跨模态注意力的上下文,在交互过程中逐层融入到文本序列中,得到融合表示 F_{vt} ,从而实现二者的深度融合。该过程可以表示为

$$F_{vt} = \text{BERT}(F_t, F_v) \quad (9)$$

在训练阶段,该模块结合两类监督信号进行优化,一方面是采用关键语义信息掩码策略,通过 MLM 损失对文本序列中的关键语义词进行掩码,并结合视觉特征完成上下文填充预测,从而引导模型学习图像与文本语义间的依赖关系。此外,为强化模态间的判别能力,另一方面是引入 ITM 任务,对融合表示进行二分类判别,以细化模态间的对齐边界并提升判别能力,从而实现跨模态的深度融合。

在推理阶段,多模态特征融合子模块不再启用。给定查询图像或文本描述,首先获取其对应的视觉特征 F_v 或文本特征 F_t ,随后将其输入至相似性计算子模块,与检索库中预先提取的所有文本特征或图像特征计算相似度并排序,从而得到跨模态检索结果。

3.2 多目标损失函数优化

为全面提升模型在民航场景下的跨模态检索效果,本文设计了一套联合优化的多目标训练策略,综合引入 ITC 损失、ITM 损失、MLM 损失及 CLS 损失 4 类任务,以实现细粒度视觉和文本语义特征表达和跨模态语义特征对齐能力的同步增强。

ITC 损失以视觉特征和文本特征为输入,通过最小化匹配样本之间的距离,最大化不匹配样本之间的距离来训练模型,使得匹配样本在嵌入空间中更加接近,而不匹配的样本则更加远离。具体表示如下:

$$L_{\text{ITC}} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n y_{i,j} d_{i,j}^2 + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (1 - y_{i,j}) \max(0, \varepsilon - d_{i,j}^2) \quad (10)$$

式中: $y_{i,j}$ 表示图像样本 i 和文本样本 j 是否为匹配图文对,若为匹配样本对,则 $y_{i,j} = 1$, 否则为 0; $d_{i,j}$ 为图像和文本特征之间的欧几里得距离; ε 为用于控制样本对之间距离的阈值。式(10)中,等式右侧第 1 项约束匹配样本对的特征距离,使其尽可能接近;等式右侧第 2 项约束不匹配样本对的特征距离,

$\max(0, \varepsilon - d_{i,j}^2)$ 表示取 0 和 $\varepsilon - d_{i,j}^2$ 的较大者,即当不匹配样本间距离 $d_{i,j}^2 < \varepsilon$ 时,损失为 $\varepsilon - d_{i,j}^2$, 迫使其在特征空间中拉远。

基于对比学习过程中得到的相似度概率矩阵,引入 ITM 任务,用于判别给定的图文对是否匹配。具体而言,对于每个图文对样本,将其与 1 个正样本(真实匹配的图文对)和 2 个负样本(随机采样的不匹配图文对)一起输入多模态特征提取模块,并在输出端通过一个二分类器进行判别,得到预测的匹配概率分布 H^{pred} , 监督信号由真实标签向量 H^{true} 提供,其中,正样本对应的分量为 1,其余分量为 0。ITM 损失函数采用交叉熵形式进行优化,表示如下:

$$L_{\text{ITM}} = - \sum_{i=1}^3 H_i^{\text{true}} \times \ln(H_i^{\text{pred}}) \quad (11)$$

式中: $H^{\text{true}} \in \{0, 1\}^3$, 即仅有一个分量为 1,其余为 0,表示正样本匹配的位置。通过最小化该损失,模型能够学习到在表征空间中区分匹配与不匹配样本,从而增强模态间的判别能力。

考虑到民航场景中大量文本描述存在精细的语义信息,为显式强化文本端的上下文建模能力与图文间的依赖关系,引入 MLM 任务。具体而言,在输入文本序列中随机遮蔽部分关键词,模型需依托视觉信息完成被遮蔽词的上下文重建。令被掩码词位置集合为 M ,模型接收融合视觉特征 F_v 与掩码文本序列 $\hat{T}$ 的联合输入,需预测每个被遮蔽位置 $i \in M$ 的原始词 t_i 。掩码语言建模损失可表示为

$$L_{\text{MLM}} = - \sum_{i \in M} \ln p(t_i | \hat{T}, F_v; \theta) \quad (12)$$

式中, $p(t_i | \hat{T}, F_v; \theta)$ 表示在模型参数 θ 下,基于视觉信息 F_v 与掩码文本输入 $\hat{T}$ 对位置 i 的原始词 t_i 的预测概率。该任务显式建模图像信息与文本语义间的对应关系,从而提升文本表示对视觉上下文的感知能力,尤其适用于民航场景中精细语义词与视觉局部细节的对齐需求。

此外,为有效地进行跨模态检索并充分利用数据集类别信息,本文在视觉编码阶段引入类监督任务,引导模型学习具有类别敏感性的视觉表示。具体而言,输入图像首先经过 ResNet 模型进行特征提取,并输入至分类器进行类别预测,采用交叉熵损失进行优化,表示如下:

$$L_{\text{CLS}} = - \sum_{i=1}^{n_c} y_i \ln(y'_i) \quad (13)$$

式中: y_i 为真实类别标签; y'_i 为模型预测的类别概率; n_c 为对应类别数。最终,总体训练目标由上述

四项损失加权组合而成,表示为

$$L = \lambda_1 L_{\text{ITC}} + \lambda_2 L_{\text{ITM}} + \lambda_3 L_{\text{MLM}} + \lambda_4 L_{\text{CLS}}$$

式中: λ_1 、 λ_2 、 λ_3 和 λ_4 为各项损失的权重超参数。

4 实验结果与分析

4.1 数据集与实验细节

本文针对低空经济和民航领域的跨模态检索需求,构建了一个细粒度的图文跨模态检索数据集。数据来源包括公开收集的《世界民航事故调查跟踪系列报告》^[17]和FGVC Aircraft数据集。原始数据经过清洗与预处理等步骤,并统一图像分辨率以适配模型输入。考虑到抽取出的图像-文本对数量有限,且真实场景下的文本描述存在高度多样性,本文在数据构建阶段进行多轮文本增强机制,针对每张图像重复5轮文本描述优化过程,因此,2个数据集的每张图像均对应5个文本描述,以提升模型在复杂文本输入下的泛化与鲁棒能力。

本文实验均基于Python编程语言及PyTorch深度学习框架,在单张Tesla V100 GPU的硬件环境上展开。所有输入图像被统一调整为256像素×256像素的大小,并在训练过程中引入随机翻转、旋转等数据增强策略,以增强模型对图像分布变化的适应性。生成的视觉和文本特征均被编码为512维的向量,作为后续模块的统一输入。在训练阶段,本文设置批处理大小为32,优化器采用Adam进行优化,初始学习率为 10^{-4} ,并采用逐步衰减策略,每训练10轮衰减至原来的0.1倍。同时,为保证特征提取主干网络的稳定性,训练初期采用冻结

训练策略,即固定BERT和ViT编码器的预训练参数,仅优化新加入的模块,随后解冻全部参数进入联合训练。4种损失函数共同指导优化,权重超参数设置为 $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1$ 。最终在验证集上选取性能最优的参数用于实际的算法评估与应用。

在性能评估方面,为衡量模型在跨模态检索任务中的效果,采用广泛使用的召回率进行性能评测^[37],包括 $R@k(k=1、5及10)$ 和平均召回率(mean Recall, mR)的2类指标。其中, $R@k$ 衡量在召回的前 k 个检索结果中包含真实匹配项的比例。mR则对各个 $R@k$ 值在不同样本上的平均表现进行统计,作为衡量整体模型性能的核心指标,能更准确地反映在多样化检索场景中的泛化能力。

4.2 实验结果

本文模型采用了基于区域块层级的特征构建方式^[37],在视觉端,全局特征提取模块将输入图像切分为若干固定大小的图像块;在文本端,文本特征提取模块对输入文本进行分词切分,得到文本序列表示。在本部分实验中,将本文模型与主流的同样基于区域块层级的跨模态检索模型CLIP、ALBEF、FLIP及BLIP在构建的民航领域的2个数据集(文本描述均使用增强文本描述)上进行实验,对比图像-文本跨模态检索指标以验证本文模型的有效性。

实验结果如表3所示,最优结果加粗显示。可以看出,本文模型相比主流跨模态检索模型均取得了更优表现。在Accidents数据集上,本文模型的mR取得了38.72%;在FGVC Aircraft数据集上,mR取得了24.44%,充分证明了本文模型的有效性。

表 3 不同模型在不同数据集上的实验结果

Table 3 Experimental results of different model on different datasets

%

数据集	模型	R@1		R@5		R@10		mR
		以图像检索文本	以文本检索图像	以图像检索文本	以文本检索图像	以图像检索文本	以文本检索图像	
Accidents	CLIP ^[7]	12.99	15.59	25.68	28.64	34.44	38.49	25.97
	ALBEF ^[8]	16.73	18.37	36.56	40.24	47.13	55.65	35.78
	FLIP ^[9]	17.52	18.67	37.46	41.51	48.34	56.74	36.71
	BLIP ^[20]	18.43	19.21	38.37	42.36	47.73	57.82	37.32
	本文模型	20.54	21.39	39.88	43.81	48.34	58.33	38.72
FGVC Aircraft ^[28]	CLIP ^[7]	4.92	4.22	10.74	14.55	19.56	25.75	13.29
	ALBEF ^[8]	6.15	6.11	18.33	24.84	32.16	37.80	20.90
	FLIP ^[9]	7.41	6.83	19.62	25.62	33.39	38.97	21.97
	BLIP ^[20]	8.56	7.61	20.79	27.30	33.75	39.42	22.91
	本文模型	10.62	9.56	22.92	28.59	34.62	40.35	24.44

4.3 消融实验

为全面验证本文提出的从数据集处理到模型构建的跨模态检索方法的有效性,设计了2组消融

实验,分别从文本描述增强策略和模型构建1个方面验证对整体效果的贡献。

第1组消融实验面向基于GPT-4o的文本描述

增强策略对整体检索的提升效果验证。对于 Accidents 数据集, 使用本文模型结构, 在不同的文本描述配置下进行对比实验, 包括原始标题(虽然仅提供较有限的图像语义信息, 但其作为弱基线可用于验证文本描述质量对检索性能的影响趋势)、基础文本描述和增强文本描述。实验结果如表 4 所示, 最优结果加粗显示, 可以看出, GPT-4o 生成的 2 类文本描述在检索效果方面均显著优于原始标题, 证明优化后的文本描述能够为模型提供更丰富的语义信息, 从而有效提升检索性能。此外, 由于在生

成过程中为 GPT-4o 提供了更多的前置信息, 增强文本描述在确保不偏离事实的情况下, 相比基础文本描述在检索效果上进一步提高, 表明更丰富的语义信息能够有效增强模型对图像内容的理解和匹配能力。

对于 FGVC Aircraft 数据集, 由于其没有图像对应的原始标题, 本文只列出 GPT-4o 生成的基础文本描述和增强文本描述的检索效果对比, 实验结果如表 5 所示, 最优结果加粗显示。结论与 Accidents 数据集一致, 基于增强文本描述可以获得更高的检索精度。

表 4 基于 GPT-4o 的文本描述优化策略的消融实验结果 (Accidents 数据集)

Table 4 Ablation experiment results of text description optimization strategy based on GPT-4o (Accidents dataset) %

文本描述	R@1		R@5		R@10		mR
	以图像检索文本	以文本检索图像	以图像检索文本	以文本检索图像	以图像检索文本	以文本检索图像	
原始标题	11.48	13.53	31.11	33.72	38.97	44.253	28.84
基础文本描述	17.22	18.21	37.16	41.63	47.43	55.650	36.22
增强文本描述	20.54	21.39	39.88	43.81	48.34	58.330	38.72

表 5 基于 GPT-4o 的文本描述优化策略的消融实验结果 (FGVC Aircraft 数据集)

Table 5 Ablation experiment results of text description optimization strategy based on GPT-4o (FGVC Aircraft dataset) %

文本描述	R@1		R@5		R@10		mR
	以图像检索文本	以文本检索图像	以图像检索文本	以文本检索图像	以图像检索文本	以文本检索图像	
基础文本描述	7.32	7.01	19.23	25.83	32.55	37.11	21.51
增强文本描述	10.62	9.56	22.92	28.59	34.62	40.35	24.44

第 2 组消融实验在 2 个数据集上基于增强文本描述分别验证细粒度特征提取模块、类监督策略及关键语义信息掩码策略的有效性。为逐步分析各模块对模型性能的影响, 设置如下 4 种实验配置。

1) 基础跨模态检索。基于双路单模态编码器的跨模态检索模型, 也即视觉端采用预训练的 ViT 模型, 文本端采用 BERT 模型。

2) 细粒度特征提取模块。在基础跨模态检索模型的基础上, 引入细粒度特征提取模块, 以提升对局部目标的特征表达能力。

3) 类监督策略。在细粒度特征提取模块的基础上, 引入类监督指导优化策略, 通过对 ResNet 输出的特征进行类别判别监督, 引导视觉端捕捉更具区分性的语义特征。

4) 关键信息语义掩码策略。结合多模态特征融合模块, 在文本端引入随机掩码和针对关键信息的预设掩码策略, 强化文本端对关键语义特征的表达。

消融实验结果如表 6 所示, 最优结果加粗显示, 验证了本文模型中各个模块的有效性。在基础跨模态检索模型上, 依次引入细粒度特征提取模

表 6 本文模型各模块的消融实验结果

Table 6 Ablation experiment results of the proposed model modules %

数据集	模块	R@1		R@5		R@10		mR
		以图像检索文本	以文本检索图像	以图像检索文本	以文本检索图像	以图像检索文本	以文本检索图像	
Accidents	基础跨模态检索	15.71	16.61	34.74	38.37	44.11	53.78	33.89
	细粒度特征提取模块	18.13	18.43	37.46	41.33	46.22	56.62	36.37
	类监督策略	19.64	20.12	38.07	42.72	48.64	57.22	37.74
	关键语义信息掩码策略	20.54	21.39	39.88	43.81	48.34	58.33	38.72
FGVC Aircraft ^[28]	基础跨模态检索	5.40	5.16	16.56	22.53	30.42	36.27	19.39
	细粒度特征提取模块	7.35	6.92	19.35	25.27	32.79	38.59	21.71
	类监督策略	9.10	7.85	21.51	27.34	33.67	39.57	23.17
	关键语义信息掩码策略	10.62	9.56	22.92	28.59	34.62	40.35	24.44

块、类监督策略和关键信息语义掩码策略, 均能带来检索效果的提升。最终结果表明, 针对民航领域数据集, 本文模型通过对视觉与文本语义特征进行细粒度增强, 显著提升了跨模态检索任务的效果。

5 结 论

1) 本文针对民航领域跨模态检索数据集稀缺与描述不足的问题, 基于大模型的文本描述增强策略, 构建了细粒度的数据集, 为后续跨模态检索模型的训练提供了高质量数据支撑, 也为其他垂直领域的数据集构建提供了可借鉴的方法。

2) 本文提出一种针对性的细粒度语义增强的跨模态检索模型, 设计细粒度特征提取模块、关键语义信息掩码策略及类监督策略, 实现了视觉与文本特征的精细对齐, 提升了本文模型在复杂检索场景下的检索性能。

3) 在 2 个数据集上与多种主流跨模态检索方法开展了全面的对比实验和消融实验。实验结果证明了本文模型的有效性。

低空经济中无人机、电动垂直起降飞行器等新型航空设备带来的图像数据量庞大, 运行环境复杂, 将对检索系统的实时性与精细化语义理解能力提出更高要求。本文的技术路径同样适用于低空经济场景, 可以为安全保障任务的跨模态检索需求提供技术参考。未来的研究工作将重点探索轻量化模型适配和跨场景知识迁移等, 以进一步提升模型在低空经济等新兴场景下的实用性与检索效率。

参考文献 (References)

[1] 郭辰阳, 敖万忠, 吕宜宏. 充分把握发展机遇, 加快推进低空经济高质量发展[J]. 财经界, 2022(25): 36-38.

Guo C Y, Ao W Z, Lyu Y H. Fully grasp the development opportunities and accelerate the high-quality development of low-altitude economy[J]. Money China, 2022(25): 36-38(in Chinese).

[2] 王淑鹤. 人工智能时代低空经济高质量发展的挑战 and 应对策略[J]. 产业创新研究, 2025(14): 39-41.

Wang S H. Challenges and countermeasures of high-quality development of low-altitude economy in the era of artificial intelligence[J]. Industrial Innovation, 2025(14): 39-41(in Chinese).

[3] 张若愚, 聂婕, 宋宁, 等. 基于布局化-语义联合表征遥感图文检索方法[J]. 北京航空航天大学学报, 2024, 50(2): 671-683.

Zhang R Y, Nie J, Song N, et al. Remote sensing image-text retrieval based on layout semantic joint representation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(2): 671-683(in Chinese).

[4] 王丹, 张峰, 张辉, 等. 基于信息互补与交叉注意力的跨模态检索方法[J]. 计算机应用研究, 2025, 42(7): 2032-2038.

Wang D, Zhang F, Zhang H, et al. Information complementarity and cross-attention for cross-modal retrieval[J]. Application Research of

Computers, 2025, 42(7): 2032-2038(in Chinese).

[5] Vinyals O, Toshev A, Bengio S, et al. Show and tell: lessons learned from the 2015 MSCOCO image captioning challenge[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 652-663.

[6] Plummer B A, Wang L W, Cervantes C M, et al. Flickr30k entities: collecting region-to-phrase correspondences for richer image-to-sentence models[J]. International Journal of Computer Vision, 2017, 123(1): 74-93.

[7] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//Proceedings of the International Conference on Machine Learning. Cambridge: JMLR, 2021: 8748-8763.

[8] Li J, Selvaraju R, Gotmare A, et al. Align before fuse: vi-sion and language representation learning with momentum distillation[J]. Advances in Neural Information Processing Systems, 2021, 34: 9694-9705.

[9] Li Y H, Fan H Q, Hu R H, et al. Scaling language-image pre-training via masking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 23390-23400.

[10] 刘光才, 黄利萍, 李章萍, 等. 通用航空视角下我国低空经济的发展研判及欧美发展启示[J]. 中国物价, 2025(6): 42-48.

Liu G C, Huang L P, Li Z P, et al. Analysis and prospects of China's low-altitude economy development from the general aviation perspective and enlightenment from European and American development[J]. China Price Journal, 2025(6): 42-48(in Chinese).

[11] Li D G, Dimitrova N, Li M K, et al. Multimedia content processing through cross-modal association[C]//Proceedings of the Eleventh ACM International Conference on Multimedia. New York: ACM, 2003: 604-611.

[12] Zhang H, Liu Y, Ma Z G. Fusing inherent and external knowledge with nonlinear learning for cross-media retrieval[J]. Neurocomputing, 2013, 119: 10-16.

[13] Frome A, Corrado G S, Shlens J, et al. Devise: a deep visual-semantic embedding model[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. New York: ACM, 2013: 2121-2129.

[14] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[EB/OL]. (2023-01-16)[2025-07-20]. <https://arxiv.org/abs/1301.3781>.

[15] Lee K H, Chen X, Hua G, et al. Stacked cross attention for image-text matching[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2018: 212-228.

[16] Wehrmann J, Kolling C, Barros R. Adaptive cross-modal embeddings for image-text alignment[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12313-12320.

[17] Qu L G, Liu M, Cao D, et al. Context-aware multi-view summarization network for image-text matching[C]//Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM, 2020: 1047-1055.

[18] Chen Y C, Li L J, Yu L C, et al. UNITER: universal image-text representation learning[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2020: 104-120.

[19] Li X J, Yin X, Li C Y, et al. Oscar: object-semantics aligned pre-

training for vision-language tasks[C]//Proceedings of the Computer Vision-ECCV. Berlin: Springer, 2020: 121-137.

[20] Li J N, Li D X, Xiong C M, et al. BLIP: bootstrapping language-image pre-training for unified vision-language understanding and generation[C]//Proceedings of the International Conference on Machine Learning. Cambridge: JMLR, 2022: 12888-12900.

[21] Yu J H, Wang Z R, Vasudevan V, et al. CoCa: contrastive captioners are image-text foundation models[EB/OL]. (2022-06-14)[2025-07-30]. <https://arxiv.org/abs/2205.01917>.

[22] Xiao R, Kim S, Georgescu M I, et al. FLAIR: VLM with fine-grained language-informed image representations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2025: 24884-24894.

[23] Tschannen M, Gritsenko A, Wang X, et al. SigLIP 2: multilingual vision-language encoders with improved semantic understanding, localization, and dense features[EB/OL]. (2025-02-20) [2025-08-01]. <https://arxiv.org/abs/2502.14786>.

[24] Csizmadia D, Codreanu A, Sim V, et al. Distill CLIP (DCLIP): enhancing image-text retrieval via cross-modal transformer distillation[EB/OL]. (2025-06-16) [2025-08-01]. <https://arxiv.org/abs/2505.21549>.

[25] Zhan G Q, Liu Y P, Han K, et al. ELIP: enhanced visual-language foundation models for image retrieval[EB/OL]. (2025-05-27) [2025-08-01]. <https://arxiv.org/abs/2502.15682>.

[26] Beaumont R, Cherti M, Coombes T, et al. LAION-5B: an open large-scale dataset for training next generation image-text models[C]//Proceedings of the Advances in Neural Information Processing Systems 35. New York: ACM, 2022: 25278-25294.

[27] 中国民航大学. 世界民航事故调查跟踪系列报告: 2010-2021[R]. 天津: 中国民航大学, 2010-2021.
Civil Aviation University of China. World civil aviation accident investigation tracking report series: 2010-2021[R]. Tianjin: Civil Aviation University of China, 2010-2021(in Chinese).

[28] Maji S, Rahtu E, Kannala J, et al. Fine-grained visual classification of aircraft[EB/OL]. (2013-06-21) [2025-08-01]. <https://arxiv.org/abs/1306.5151>.

[29] Openai, Achiam J, Adler S, et al. GPT-4 technical report[EB/OL]. (2024-03-04)[2025-08-01]. <https://arxiv.org/abs/2303.08774>.

[30] Fu S H, Yang Q Z, Mo Q J, et al. LLMDet: learning strong open-vocabulary object detectors under the supervision of large language models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2025: 14987-14997.

[31] Zheng L M, Chiang W L, Sheng Y, et al. Judging LLM-as-a-judge with MT-bench and chatbot arena[C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. New York: ACM, 2023: 46595-46623.

[32] Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-networks[EB/OL]. (2019-08-27) [2025-08-01]. <https://arxiv.org/abs/1908.10084>.

[33] 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.

[34] 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.

[35] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2021-06-03)[2025-08-01]. <https://arxiv.org/abs/2010.11929>.

[36] Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Kerrville: Association for Computational Linguistics 2019: 4171-4186.

[37] 张振兴, 王亚雄. 图文跨模态检索研究综述[J]. 北京交通大学学报, 2024, 48(2): 23-36.
Zhang Z X, Wang Y X. A survey on image-text cross-modal retrieval[J]. Journal of Beijing Jiaotong University, 2024, 48(2): 23-36 (in Chinese).

Fine-grained semantic-enhanced cross-modal image-text
retrieval method for civil aviation

LIU Shuyan^{1, 2}, HE Liu^{1, 2, 3, *}, ZENG Jianghui^{1, 2}

- (1. Department of Standard and Data Technology Research, China Aero-Polytechnology Establishment, Beijing 100028, China;
- 2. Industrial Internet Application Innovation Center, China Aero-Polytechnology Establishment, Beijing 100028, China;
- 3. School of Computer Science, Fudan University, Shanghai 200438, China)

Abstract: In the fields of low-altitude economy and civil aviation, security assurance tasks heavily rely on the efficient correlation of cross-modal information such as images and texts. However, while mainstream cross-modal retrieval models perform well on general datasets, they underperform in these areas, which require high levels of fine-grained semantic understanding. Based on existing civil aviation datasets, a cross-modal retrieval approach with fine-grained semantic augmentation is suggested as a solution to this problem, creating a whole pipeline that includes data processing, model development, and training. First, text descriptions are optimized and enhanced based on a large multimodal model to construct a cross-modal retrieval dataset containing rich semantic information. Second, a model is created using a popular cross-modal retrieval framework. To improve the model's ability to express fine-grained semantic features, techniques such class supervision, key semantic information masking, and a fine-grained feature extraction module are introduced. Experimental results on two datasets verify the effectiveness of the proposed method, providing a reference technical path for cross-modal retrieval in low-altitude economy scenarios.

Keywords: low-altitude economy; civil aviation; cross-modal retrieval; text description optimization; fine-grained semantic enhancement

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0283

基于广播代理重加密的低空无人机运行 数据安全共享方案

周泽全^{1, 2, 3}, 季晓海^{1, 2, 3}, 罗喜伶^{1, 2, 3, *}, 毛剑^{2, 4}, 王珺珺¹

(1. 北京航空航天大学 电子信息工程学院, 北京 100191; 2. 北京航空航天大学 杭州创新研究院, 杭州 310000; 3. 空地一体新航行系统技术全国重点实验室, 北京 100191; 4. 北京航空航天大学 网络空间安全学院, 北京 100191)

摘 要: 在开放动态的低空物联网中, 无人机 (UAV) 运行数据共享面临着数据窃听、重放攻击和未授权访问等安全威胁, 可能严重危及低空运行安全, 保障该数据安全共享至关重要。现有方案利用身份基广播加密实现点对多点数据安全共享。然而, 该类方案往往需预设共享组容量, 限制了共享灵活性; 此外, 在解密过程中, 共享者需为组内其他成员执行额外的数学运算, 导致解密计算负担较重。因此, 提出一种基于广播代理重加密的低空无人机运行数据安全共享方案。所提方案利用身份基广播加密与对称加密技术构建一种混合加密机制, 通过密文重加密将点对点共享转化成点对多点共享。在点对多点共享过程中, 所提方案支持用户仅持有固定长度的私钥即可实现无状态的数据群组共享且无需预设组容量, 兼顾共享安全性、灵活性和高效性。实验结果表明: 在点对点共享中, 所提方案加解密计算开销固定不变; 在点对多点共享中, 相比次优方案, 所提方案重加密密钥生成时间开销降低 40% 以上, 重加密密文解密运行时间约为固定的 3.6 ms。

关 键 词: 无人机; 数据共享; 数据安全; 加密算法; 密文

中图分类号: V355; TP391

文献标志码: A

文章编号: 1001-5965(2026)09-3089-11

低空经济是国家战略性新兴产业, 是国家大力发展的新质生产力之一。低空物联网作为低空经济发展的数字化底座, 是城市空中交通 (urban air mobility, UAM) 等应用的基础技术保障, 其中, 无人机作为低空物联网服务的核心平台, 已初步应用到包括商业、工业和消防等多个领域, 如物流配送、无人巡检、应急救援等^[1]。截至 2024 年底, 中国无人机运营企业已超过 1.9 万家, 累计注册无人机数量已超过 215.8 万架; 2024 年, 全年累计无人机飞行时长达 2 666 万小时, 同比增长 15%^[2]。加快建设完善低空物联网体系, 提升无人机运行安全性, 是打造低空新质生产力的必要前提。

低空物联网是由数字化、网联化、智能化的飞

行器及设备、信息物理基础设施、数据信息网络及应用服务系统构成的综合性体系, 在“端网云用”总体架构下实现对低空空域的泛在感知、广域互联和智能管服应用, 支撑典型应用场景的安全高效运行^[3]。当前, 中国低空物联网运行仍处于起步阶段, 无人机主要由各自的运营人独立控制, 数据流转方式相对单一, 即数据从无人机回传至地面控制站。然而, 这种隔离运行模式难以满足未来低空多主体协同运行的需求。国家相关主管部门和地方政府正在加快推进 UAM 的试点工作, 并逐步建设低空交通管理平台, 迫切需要通过实现无人机飞行计划制定、态势监视、低空气象与情报服务、间隔管理、应急处置、任务申报审批等关键信息的安全高效共享。

收稿日期: 2025-05-12; 录用日期: 2025-10-31; 网络出版时间: 2025-11-28 14:57

网络出版地址: link.cnki.net/urlid/11.2625.V.20251128.1055.001

基金项目: 浙江省“尖兵”研发攻关计划 (2022C01055); 国家自然科学基金 (62172027); 浙江省自然科学基金 (LZ23F020013)

* 通信作者. E-mail: luoxiling@buaa.edu.cn

引用格式: 周泽全, 季晓海, 罗喜伶, 等. 基于广播代理重加密的低空无人机运行数据安全共享方案 [J]. 北京航空航天大学学报, 2026, 52 (9): 3089-3099. Zhou Z Q, Ji X H, Luo X L, et al. Secure sharing scheme for low-altitude UAV operational data based on broadcast proxy re-encryption [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3089-3099 (in Chinese).

从无人机运行数据空地协同共享的角度来看,无人机节点通过空地通信和地地通信,将运行数据(如飞行计划、位置信息等)共享给多个地面网络节点^[4]。然而,由于网络的动态开放性,无人机运行数据在多主体共享过程中容易遭受多种安全威胁^[5]。例如,攻击者通过窃听获取无人机回传的飞行数据,如航迹、任务目标等,可能导致机密信息泄露;攻击者重放历史数据,可能导致地面系统接收到错误的运行状态或环境信息,影响飞行控制甚至引发飞行事故;非法用户可能非法访问敏感数据,发起更深层次的攻击。2024 年 1 月,中国交通运输部发布了《民用无人驾驶航空器运行安全管理规则》,明确要求在无人机运行数据的共享过程中,应有效防范数据窃听泄露等安全风险^[6]。因此,亟需围绕无人机运行数据多主体安全共享技术展开深入研究,以构建更加安全高效的低空智能网联体系。

Nikooghadam 等提出了一种无人机点对点安全共享方案,支持用户与无人机协商会话密钥以保障通信安全性^[7]。Khan 等引入超椭圆曲线加密机制,实现了点对点高强度安全共享^[8]。Liu 等则采用哈希验证码技术,增强了信息交换过程中的可信性和完整性^[9]。然而,这类点对点共享方案在面向点对点数据共享场景下,往往会导致无人机计算与通信开销显著增加,限制了系统的扩展能力。为实现更高效的多用户访问控制,属性基加密(attribute-based encryption, ABE)被广泛应用于数据共享场景中。该类方案允许数据拥有者依据属性集合定义访问策略并加密数据,仅当接收方用户的属性满足该策略时,方可使用其属性私钥完成解密,从而实现灵活的点对多点数据共享^[10-14]。然而,ABE 方案在访问策略变化或用户动态加入/退出的情形下,通常伴随高昂的密钥更新与管理开销,难以适用于资源受限且动态变化的无人机运行环境。

身份基广播加密(identity-based broadcast encryption, IBBE)通过免除公钥证书的使用,显著降低了密钥管理成本,因而在多主体数据共享场景中具有较高的实用价值^[15]。Liu 等将 IBBE 技术引入会话密钥分发系统,实现了通过一次加密操作即可完成点对多点的数据共享^[16]。Maiti 等结合基于身份的加密(identity-based encryption, IBE)与拉格朗日插值多项式,提出了一种适用于多主体环境的数据共享方案,但该方案引入了大量的乘法群运算,导致计算开销较高^[17]。Zhao 等设计了一种基于身份的广播签名机制,在动态网络环境中同时实现数据的机密性与身份认证功能^[18]。Kim 等构建了一种密文大小恒定的 IBBE 方案,但其共享者在解密过程

中的计算开销与共享用户数量呈线性关系,难以满足多主体共享场景的效率需求^[19]。Deng 等提出了一种重加密方案:首先,计算仅个体授权用户访问的 IBE 密文,然后,可通过重加密转换的 IBBE 密文,从而兼顾个体访问控制与多用户共享^[20]。Chen 等利用拉格朗日插值多项式构建了匿名可撤销的广播代理重加密方案,但该方案依赖大量昂贵的幂运算,导致计算开销较大^[21]。Zhong 等提出一种通过代理重加密机制将 IBE 密文转换为 IBBE 密文的方案,实现了接收方恒定计算成本,但要求提前固定共享组容量,降低了系统的灵活性^[22]。Zhou 等设计了一种具备匿名性的代理广播重加密方案,支持动态用户匿名变更,但其系统公共参数规模与共享者数量线性相关,降低了系统扩展性^[23]。Ge 等构建了一种支持用户撤销的代理广播重加密方案,通过数据发送方生成的重加密密钥实现灵活的访问控制,但在解密阶段仍需共享者对组内其他用户执行额外的数学运算,如代数运算和哈希运算,导致解密计算负担较重^[24]。

综上所述,点对多点的共享环境下,点对点共享方案需要为每个共享者分别协商密钥,会引发较大的计算与通信开销,难以满足多主体用户高效共享的需求。ABE 方案虽然提供了灵活的多主体共享访问控制能力,但访问策略的变更将显著增加密钥管理与更新的复杂度,不适用于资源受限且动态的无人机运行环境。相较而言,IBBE 方案显著降低了密钥管理成本。然而,现有 IBBE 共享方案通常存在以下不足:①系统公共参数的规模随共享者数量线性增长导致多数方案需预设共享组容量,限制了系统可扩展性;②在解密过程中,每个共享者需为组内其他成员执行额外的数学运算,导致接收端计算负担较重。

针对上述问题与挑战,本文提出一种低空无人机运行数据安全共享(unmanned aerial vehicle operational data secure sharing, UODSS)方案。该方案融合身份基广播加密机制与高效的对称加密技术,设计了一种高性能的广播代理重加密算法,支持用户仅持有固定长度的私钥即可完成去状态化的共享组共享,且无需预设组容量,从而提升系统的可扩展性与灵活性。通过引入密文重加密,UODSS 方案实现了点对点、点对多点 2 种数据共享模式,能够有效抵御窃听、重放攻击及未授权访问等安全威胁,高效地保障运行数据在动态开放网络环境下共享的安全性。实验结果表明,UODSS 方案的密钥管理开销始终保持恒定且可忽略不计;在点对点共享中,加解密算法计算开销达到 $O(1)$,

运行时间分别约为 4.02 ms 和 0.74 ms; 在点对多点共享中, 相比较次优方案, 重加密密钥生成时间开销降低 40% 以上, 重加密密文解密算法约为固定的 3.6 ms, 优于现有方案。

1 问题描述

1.1 系统模型

低空无人机运行数据安全共享系统主要涉及 5 个实体, 包括: 低空交通管理平台 (low-altitude air traffic management, LATM)、无人机、云平台、代理人及多主体共享者, 不同实体的具体描述如图 1 所示。

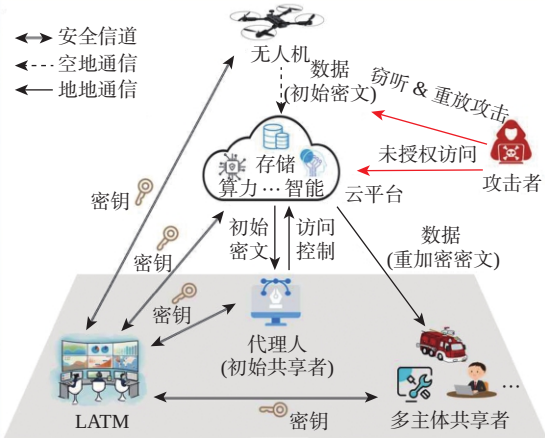


图 1 低空无人机运行数据安全共享系统模型

Fig. 1 System model for secure sharing of low-altitude UAV operational data

- 1) 低空交通管理平台: LATM 由政府部门管控, 主要负责无人机运行数据共享安全机制的实施。例如, 设定共享系统的安全等级、初始化系统的公开参数、验证新加入系统用户的身份并通过安全信道分发相应的身份私钥等。
- 2) 无人机: 作为数据的产生者, 负责初始加密运行过程中生成的各类数据, 如态势监视、任务日志等, 并将初始密文上传至云。无人机往往具有有限的资源, 如算力、存储和续航等。
- 3) 云平台: 作为数据存储与计算中心, 负责存储无人机上传的数据, 并提供数据管理、分析和智能服务等。同时, 云为系统用户提供安全高效的数据访问服务, 即用户可通过网络设备随时随地访问云数据。
- 4) 代理人: 是由无人机授权的数据初始共享者, 一般情况下由无人机地面控制站或运营人担任, 可根据不同应用场景进行灵活设定。代理人不仅能访问解密初始密文数据(点对点), 还能根据场景需求通过重加密设定访问控制, 以实现多主体数据共享(点对多点)。

5) 多主体共享者: 泛指需要访问共享无人机运行数据的多个系统用户, 如无人机运行服务提供方、行业监管方、研究机构等多类主体用户。

低空无人机运行数据安全共享系统模型如图 1 所示, 由 LATM 初始化整个系统, 并通过安全信道(如 Diffie-Hellman 协议)为每个系统用户分发对应的身份私钥。无人机设定数据代理人, 利用其身份加密数据后将初始密文上传到云平台。代理人根据不同的需求设定访问策略, 请求云重加密初始密文。任何满足访问策略的系统用户均可以从云平台上下载重加密密文, 并利用自己的身份私钥解密。

1.2 通信模型

本文的系统模型中不同的网络节点间的通信主要涉及到 2 种类型, 即空地通信与地地通信, 如图 2 所示。

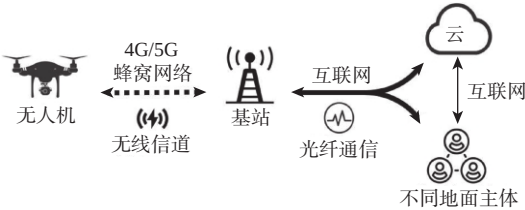


图 2 通信模型

Fig. 2 Communication model

- 1) 空地通信: 无人机主要依托 4G/5G 蜂窝网络通过地面基站与地面节点(如云)建立通信链路, 以传输运行数据。对于远程或特殊环境下的无人机, 还可结合卫星通信或专用无线链路以增强通信覆盖范围和稳定性。
 - 2) 地地通信: 地面控制中心、LATM 和云等地面节点之间的通信主要通过互联网(如 TCP/IP 协议族)进行。相较于无线信道, 光纤通信具备更高的带宽和更低的延迟。
- 1.3 威胁模型
- 假设共享系统中每个实体都配备了可信平台模块(trusted platform module, TPM)用于安全存储秘密值, 如身份私钥。由于网络动态开放, 系统易遭受多样安全威胁。本文主要考虑以下威胁:
- 1) 窃听攻击: 攻击者可能监听空地通信或地地通信, 试图获取传输中的无人机运行数据信息。
 - 2) 云数据泄露: 由于云平台存储无人机上传的数据, 攻击者可能利用云存储漏洞或内部获取泄露数据。
 - 3) 访问策略绕过: 恶意系统用户可能尝试访问获取非授权敏感数据, 非法扩大共享权限。
 - 4) 重放攻击: 攻击者可能捕获并重复发送无人

机的运行数据, 试图欺骗系统并影响正常运行。

1.4 设计目标

本文旨在为低空物联网构建一个高效的 UODSS 方案, 设计目标主要包括:

- 1) 多主体数据共享: 允许构建共享群组, 实现点对多点的数据共享; 同时, 应无需事先确定共享组的容量。
- 2) 无状态接收者: 每个系统用户只保留一个无状态解密(身份)密钥。当共享组变化时, 组中的用户不需要更新他们的解密密钥, 同时解密不需要为组中的其他成员执行数学运算。
- 3) 防重放机制: 防止攻击者重新发送已传输的数据信息。
- 4) 机密性: 确保无人机运行数据在传输、存储和共享过程中仅授权方可访问, 防止未授权的敌手获取、窃取或泄露敏感数据。

2 无人机运行数据安全共享方案

本节提出一种基于身份的广播代理重加密算法, 并进一步利用该算法构建 UODSS 方案, 主要包含 3 个阶段: 初始化、初始共享及多主体共享。初始化阶段主要由 LATM 负责计算生成整个 UODSS 方案的参数, 包括方案公开参数、主密钥对及每个用户的身份私钥。在初始共享阶段中, 无人机首先选择一个初始共享者(也称代理人), 然后, 利用该共享者的身份加密运行数据得到初始密文, 最后, 将该密文上传到云。在多主体共享阶段, 由代理人设定建立共享群组, 并为初始密文生成重加密密钥发送给云, 由云对初始密文进行重加密, 只有共享组用户可利用自己的身份私钥解密重加密密文, 从而实现无人机运行数据的多主体安全共享。

2.1 初始化阶段

初始化阶段提供 3 个算法部署整个 UODSS 方案, 包括: PmsGen、MasterGen 和 Register。

- 1) PmsGen(κ) $\rightarrow$ P : 给定安全参数 κ , 算法随机选择一个素数 p , 并构建一个双线性映射群系统 $(p, \mathbf{Z}_p, G, G_T, \hat{e})$, 其中, $\mathbf{Z}_p$ 表示一个整数群, G 和 G_T 表示乘法循环群, $\hat{e}$ 表示一个双线性映射, 即 $\hat{e}: G \times G \rightarrow G_T$ 。然后, 算法构建 2 个哈希函数 $\mathcal{H}_1: \{0,1\}^* \rightarrow \mathbf{Z}_p^*$ 和 $\mathcal{H}_2: G_T \rightarrow G$, 并选择对称加、解密算法函数 $\mathcal{E}$ 和 $\mathcal{D}$, 如 AES 算法。算法最后从群 G 中随机选取 2 个生成元 $h, u \leftarrow G$, 输出公开参数为 $P = (\mathbf{Z}_p, G, G_T, \hat{e}, \mathcal{H}_1, \mathcal{H}_2, h, u, \mathcal{E}, \mathcal{D})$ 。
- 2) MasterGen(P) $\rightarrow$ (MPK, MSK): 给定公开参数 P , 算法首先随机选择一个整数 $\alpha \leftarrow \mathbf{Z}_p^*$ 和一个乘法循环群生成元 $g \leftarrow G$ 。然后, 算法计算 $h^\alpha \in G$ 、 $u^\alpha \in G$

和 $v \leftarrow \hat{e}(g, h) \in G_T$ 。最后, 输出主公钥 $\text{MPK} = (h^\alpha, u^\alpha, v)$ 和主私钥 $\text{MSK} = (g, \alpha)$ 。

3) Register(P, MSK, d) $\rightarrow s_d$: 给定公开参数 P 、主私钥 $\text{MSK} = (g, \alpha)$ 和一个身份 d , 算法计算并输出身份私钥 s_d 如下:

$$s_d \leftarrow \frac{1}{g^{\alpha + \mathcal{H}_1(d)}} \tag{1}$$

在本阶段, LATM 首先运行 PmsGen 算法生成公开参数 P , 初始化整个共享方案。然后, LATM 运行 MasterGen 算法生成主公钥 MPK 和主私钥 MSK, 公开主公钥并私密保存主私钥。在 UODSS 方案中, 所有系统用户均需向 LATM 发送自己的身份 d , 请求认证注册。身份验证成功后, LATM 运行 Register 算法为身份 d 生成对应的身份私钥 s_d , 并通过安全信道发送给对应的用户。具体过程如图 3 所示。

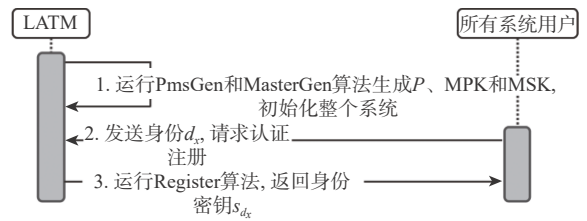


图 3 初始化阶段
Fig. 3 Initialization phase

2.2 初始共享阶段

本阶段提供 2 个算法: Enc 和 Dec, 实现无人机运行数据的初始共享。

1) Enc($P, \text{MPK}, m, d_u, d_a, d_c$) $\rightarrow C$: 给定公开参数 P 、主公钥 $\text{MPK} = (h^\alpha, u^\alpha, v)$ 、明文 m 、数据拥有者身份 d_u 、初始共享者身份 d_a 及云平台身份 d_c , 算法随机选择一个随机元素 $K \leftarrow G_T$ 作为对称密钥, 并加密明文 m 生成密文主体如下:

$$C_m = \mathcal{E}_K(m) \tag{2}$$

然后, 算法随机选择一个随机数 $r_1 \leftarrow \mathbf{Z}_p^*$, 并根据当前时间 t , 计算密文头如下:

$$\begin{cases} C_0 = t \parallel d_u \parallel d_a \parallel d_c \\ C_1 = (h^\alpha \cdot h^{\mathcal{H}_1(d_a)})^{r_1} \\ C_2 = v^{r_1 \cdot \mathcal{H}_1(C_0)} \cdot K \\ C_3 = (u^\alpha \cdot u^{\mathcal{H}_1(d_a)})^{r_1} \end{cases} \tag{3}$$

最后, 输出初始密文 $C = (C_0, C_1, C_2, C_3, C_m)$ 。

2) Dec(P, C, s_{d_a}) $\rightarrow m/\perp$: 给定公开参数 P 、初始密文 $C = (C_0, C_1, C_2, C_3, C_m)$ 和初始共享者的身份私钥 s_{d_a} , 算法恢复出对称密钥 K 如下:

$$\hat{e}(s_{d_a}, C_1) = v^{r_1} \tag{4}$$

$$C_2/(v^{r_1})^{\mathcal{H}_1(C_0)} = K \quad (5)$$

然后, 使用密钥 K 解密成功, 则输出 $m = \mathcal{D}_K(C_m)$; 否则, 解密失败, 则输出无效明文 $\perp$ 。

在本阶段, 无人机设定初始共享者 (也称代理人) 为 d_a (如设定为无人机地面控制站), 该代理人的职责是进一步实现和管理无人机运行数据的多主体共享策略。给定一条运行数据 m , 无人机利用代理人身份 d_a 运行 Enc 算法对数据加密, 得到初始密文 C 。然后, 无人机将该初始密文通过空地通信上传到云。作为初始共享者, 代理人可利用自己的私钥 s_{d_a} 运行 Dec 算法解密密文, 实现点对点数据共享。具体过程如图 4 所示。

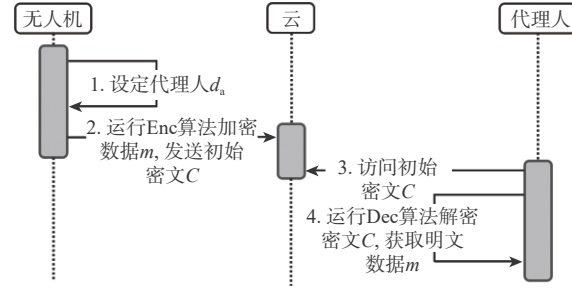


图4 初始共享阶段

Fig. 4 Initial sharing phase

2.3 多主体共享阶段

本阶段提供 3 个算法: ReKeyGen、ReEnc 及 Dec2, 实现无人机运行数据的多主体共享。

1) $\text{ReKeyGen}(P, \text{MPK}, S, s_{d_a}) \rightarrow k_{d_a \rightarrow S}$: 给定公开参数 P 、主公钥 $\text{MPK} = (h^\alpha, u^\alpha, v)$ 、多主体共享组 $S = \{d_i\}_{1 \leq i \leq n}$ 和代理人身份私钥 s_{d_a} , 算法首先随机选择一个随机数 $r_2 \leftarrow \mathbf{Z}_p^*$, 并计算 $A = \{a_i\}_{1 \leq i \leq n}$, 其中:

$$a_i = \mathcal{H}_1(d_i) \parallel (h^\alpha)^{r_2} (h)^{r_2 \mathcal{H}_1(d_i)} \quad (6)$$

最后, 算法选择一个随机数 $r_3 \leftarrow \mathbf{Z}_p^*$, 计算并输出重加密密钥 $k_{d_a \rightarrow S} = (k_1, k_2, k_3)$, 其中:

$$\begin{cases} k_1 = \mathcal{H}_2(v^{r_2}) \cdot h^{r_3} \\ k_2 = s_{d_a} u^{r_3} \\ k_3 = A \end{cases} \quad (7)$$

2) $\text{ReEnc}(P, C, k_{d_a \rightarrow S}) \rightarrow R$: 给定公开参数 P 、初始密文 $C = (C_0, C_1, C_2, C_3, C_m)$ 和重加密密钥 $k_{d_a \rightarrow S} = (k_1, k_2, k_3)$, 重新计算初始密文头得到 $R = (R_0, R_1, R_2, R_3, R_4, C_m)$, 其中:

$$\begin{cases} R_0 = C_0 \\ R_1 = k_1 \\ R_2 = C_2^{1/\mathcal{H}_1(C_0)} \cdot \hat{e}(k_2, C_1^{-1}) \\ R_3 = C_3 \\ R_4 = k_3 \end{cases} \quad (8)$$

最后, 算法输出重加密密文为 R 。

3) $\text{Dec2}(P, R, d_x, k_{d_a \rightarrow S}) \rightarrow m/\perp$: 给定公开参数 P 、重加密密文 R 、一个共享者身份 d_x 和对应的身份私钥 s_{d_x} , 算法首先计算 $\mathcal{H}_1(d_x)$ 并从 k_3 中找到 $\mathcal{H}_1(d_x) \parallel h^{r_2(\alpha + \mathcal{H}_1(d_x))}$ 。如果 $d_x \in S$, 那么令 $y_{d_x} = h^{r_2(\alpha + \mathcal{H}_1(d_x))}$, 算法进一步恢复出对称密钥 K 如下:

$$\hat{e}(s_{d_x}, y_{d_x}) = v^{r_2} \quad (9)$$

$$R_1/\mathcal{H}_2(v^{r_2}) = h^{r_3} \quad (10)$$

$$(R_2 \hat{e}(R_3, h^{r_3}))^{\mathcal{H}_1(R_0)} = K \quad (11)$$

最后, 使用密钥 K 解密成功, 则输出 $m = \mathcal{D}_K(C_m)$; 否则, 解密失败, 则输出无效明文 $\perp$ 。

在本阶段, 假设代理人设定了 n 个主体用户可访问无人机运行数据, 构建了一个多主体共享序列 $S = \{d_i\}_{1 \leq i \leq n}$ 。然后, 代理人 d_a 运行 ReKeyGen 算法生成重加密密钥 $k_{d_a \rightarrow S}$, 并将该密钥发送给云。收到 $k_{d_a \rightarrow S}$ 之后, 云运行 ReEnc 算法对初始密文 C 进行重加密, 生成重加密密文 R 。此时, 任何共享组中的合法身份, 即 $d_x \in S$, 均可以利用其身份私钥运行 Dec2 算法解密重加密密文。

具体过程如图 5 所示。当有新成员加入或旧成员退出共享组时, 采用的更新策略如下:

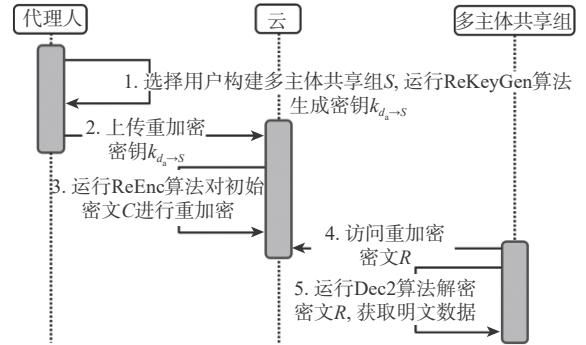


图5 多主体共享阶段

Fig. 5 Multiagent sharing phase

1) 新成员 d^* 加入: 代理人只需为新成员 d^* 计算参数 a_{d^*} , 并将该参数加入 $k_3 = A$ 中, 随后由云对 R_4 进行更新即可添加新成员的共享权限。

2) 旧成员退出: 代理人需要为新的共享组重新生成一个重加密密钥 $k_{d_a \rightarrow S}$ 发送给云用以更新密文, 从而更新共享组权限控制。

本文方案中任何用户的身份私钥都是无状态的; 因此, 当共享组变化时, 组中的用户不需要更新他们的身份密钥。当某用户身份私钥泄漏时, 用户需重新选取公布新的身份, 并为该身份申请新的身份私钥, 同时通知代理人更新密文。

2.4 方案讨论

在 UODSS 方案中, 密文主体采用对称加密, 而

密文头则利用所提基于身份的广播代理重加密技术加密对称密钥 K 。这种混合加密机制仅需管理少量非对称密钥、动态生成对称密钥,从而显著降低密钥管理负担。同时,它利用高效的对称加密处理数据主体,加快了密文主体的加密与解密过程,可有效降低计算开销(尤其是在数据量较大的情况下)。根据实际运行场景,该机制既支持无人机在每条密文中使用不同的对称密钥,以增强数据安全性,又允许在固定时间间隔内更新对称密钥,在保证安全性的前提条件下,进一步减少密文头的计算开销。

为应对未来复杂的应用场景, UODSS 方案提供了 2 种共享模式: 点对点共享(初始共享阶段)和点对多点共享(多主体共享阶段)。这 2 种共享模式可并行执行,以减少多主体数据共享的延迟。例如,代理人可以根据访问策略提前生成重加密密钥并存储在云端,云在接收到无人机上传的密文后,即可直接执行重加密,从而低延迟地实现多主体数据共享。此外, UODSS 方案可根据无人机的不同场景(如通信条件、共享需求)灵活调整不同的共享服务,以进一步降低数据共享延迟: ①无人机直接将初始密文发送给初始共享者,由初始共享者执行重加密并共享给多主体用户; ②无人机自身作为初始共享者,直接接收其他无人机的数据或执行重加密,实现多无人机协同或多主体共享,无需依赖中间代理。

3 安全性分析

3.1 正确性

本文方案的正确性可通过以下定理证明。

定理 1 给定公开参数 P 、主公钥 MPK 和主私钥 MSK, 若初始密文 $C = (C_0, C_1, C_2, C_3, C_m)$ 由算法 $\text{Enc}(P, \text{MPK}, m, d_u, d_a, d_c)$ 生成且身份私钥 s_{d_a} 由算法 $\text{Register}(P, \text{MSK}, d_a)$ 生成, 则初始密文的解密算法 $\text{Dec}(P, C, s_{d_a})$ 总能输出正确的明文数据 m 。

证明 $\text{Dec}(P, C, s_{d_a})$ 算法的第 1 步是计算 $\hat{e}(s_{d_a}, C_1) = v^{r_1}$, 该式的正确性推导如下:

$$\hat{e}(s_{d_a}, C_1) = \hat{e}\left(g^{\frac{1}{\alpha + \mathcal{H}_1(d_a)}}, h^{r_1(\alpha + \mathcal{H}_1(d_a))}\right) = v^{r_1} \tag{12}$$

第 2 步是计算 $\frac{C_2}{(v^{r_1})^{\mathcal{H}_1(C_0)}} = K$, 该式的正确性推导如下:

$$\frac{C_2}{(v^{r_1})^{\mathcal{H}_1(C_0)}} = \frac{v^{r_1 \mathcal{H}_1(C_0)} \cdot K}{(v^{r_1})^{\mathcal{H}_1(C_0)}} = K \tag{13}$$

因此, 使用对称密钥 K 能成功解密 $m = \mathcal{D}_K(C_m)$ 。

证毕

定理 2 给定公开参数 P 、主公钥 MPK、共享组 $S = \{d_i\}_{1 \leq i \leq n}$ 及初始密文 C , 若重加密密钥 $s_{d_a \rightarrow S}$ 由 $\text{ReKeyGen}(P, \text{MPK}, S, d_a)$ 生成、重加密密文 R 由 $\text{ReEnc}(P, C, k_{d_a \rightarrow S})$ 生成、身份私钥 s_{d_x} 由 $\text{Register}(P, \text{MSK}, d_x)$ 生成且 $d_x \in S$, 则解密算法 $\text{Dec2}(P, R, d_x, s_{d_x})$ 总能输出正确的明文数据 m 。

证明 $\text{Dec2}(P, R, d_x, s_{d_x})$ 算法首先获取 $y_{d_x} = h^{r_2(\alpha + \mathcal{H}_1(d_x))}$ 计算 $\hat{e}(s_{d_x}, y_{d_x}) = v^{r_2}$ (详见定理 1)。然后, 算法计算下式:

$$\frac{R_1}{\mathcal{H}_2(v^{r_2})} = \frac{\mathcal{H}_2(v^{r_2}) \cdot h^{r_3}}{\mathcal{H}_2(v^{r_2})} = h^{r_3} \tag{14}$$

紧接着, 算法计算式 (15) 如下:

$$\begin{aligned} (R_2 \hat{e}(R_3, h^{r_3}))^{\mathcal{H}_1(R_0)} = & \left(C_2^{\frac{1}{\mathcal{H}_1(C_0)}} \cdot \hat{e}(k_2, C_1^{-1}) \cdot \hat{e}(R_3, h^{r_3}) \right)^{\mathcal{H}_1(R_0)} = \\ & \left(v^{r_1} \cdot K^{\frac{1}{\mathcal{H}_1(C_0)}} \cdot \hat{e}(k_2, C_1^{-1}) \cdot \hat{e}(R_3, h^{r_3}) \right)^{\mathcal{H}_1(R_0)} \end{aligned} \tag{15}$$

在式 (15) 中, $v^{r_1} \cdot \hat{e}(k_2, C_1^{-1})$ 部分的正确性推导如下:

$$\begin{aligned} v^{r_1} \hat{e}(k_2, C_1^{-1}) &= v^{r_1} \hat{e}\left(s_{d_a} u^{r_3}, (h^{\alpha} h^{\mathcal{H}_1(d_a)})^{-r_1}\right) = \\ & v^{r_1} v^{-r_1} \hat{e}(u^{r_3}, h^{-r_1(\alpha + \mathcal{H}_1(d_a))}) = \hat{e}(u^{r_3}, h^{-r_1(\alpha + \mathcal{H}_1(d_a))}) = \\ & \hat{e}(u, h)^{-r_1 r_3 (\alpha + \mathcal{H}_1(d_a))} \end{aligned} \tag{16}$$

在式 (15) 中, $\hat{e}(R_3, h^{r_3})$ 部分的正确性推导如下:

$$\begin{aligned} \hat{e}(R_3, h^{r_3}) &= \hat{e}\left((u^{\alpha} u^{\mathcal{H}_1(d_a)})^{r_1}, h^{r_3}\right) = \hat{e}(u^{r_1(\alpha + \mathcal{H}_1(d_a))}, h^{r_3}) = \\ & \hat{e}(u, h)^{r_1 r_3 (\alpha + \mathcal{H}_1(d_a))} \end{aligned} \tag{17}$$

因此, 可继续推导式 (15) 如下:

$$\begin{aligned} \left(v^{r_1} \cdot K^{\frac{1}{\mathcal{H}_1(C_0)}} \cdot \hat{e}(k_2, C_1^{-1}) \cdot \hat{e}(R_3, h^{r_3}) \right)^{\mathcal{H}_1(R_0)} = & \left(K^{\frac{1}{\mathcal{H}_1(C_0)}} \right)^{\mathcal{H}_1(R_0)} = K \end{aligned} \tag{18}$$

因此, 使用对称密钥 K 能成功解密 $m = \mathcal{D}_K(C_m)$ 。

证毕

定理 3 当每个系统实体均按照公开参数 P 执行 UODSS 方案, 该方案能够正确实现多主体数据共享、无状态接收者及防重放机制。

证明 根据定理 2, 代理人可有效构建共享组 $S = \{d_i\}_{1 \leq i \leq n}$, 当 $d_x \in S$ 时方可解密成功。同时, 公开参数 $P = (\mathcal{Z}_p, G, G_T, \hat{e}, \mathcal{H}_1, \mathcal{H}_1, h, u, \varepsilon, D)$ 没有限制共享组的容量。因此, UODSS 方案正确实现了多主体数据共享的属性。

然后, 根据 Register 算法, 身份私钥 $s_d = g^{1/(\alpha + \mathcal{H}_1(d))}$ 独立于其他用户私钥。因此, 共享组变化时, 共享组成员无需更新自己的身份密钥。在解密算法中, 共享者只需根据自己的身份定位 $\mathcal{H}_1(d_x) \parallel h^{r_2(\alpha + \mathcal{H}_1(d_x))}$,

不需要为组中的其他成员执行任何数学运算。因此, UODSS 方案正确实现了无状态接收者的属性。

最后, 密文头的 $C_0 = t \parallel d_u \parallel d_a \parallel d_c$, 并通过哈希函数将其嵌入到了 C_2 字段中。UODSS 方案通过引入时间戳 t , 可以有效防止重放攻击。因此, UODSS 方案正确实现了防重放机制。

证毕

3.2 机密性

Delerablée 提出一种基于身份的广播加密算法 (用 DIBBE 表示), 证明了该算法能在随机谰言机模型下安全抵御选择明文攻击 (chosen ciphertext attack, CPA)^[25]。本节利用该算法论证本文 UODSS 方案的机密性。

定理 4 如果 DIBBE 算法在随机谰言机模型下能够安全抵御 CPA, 那么本文 UODSS 方案的初始加密在随机谰言机模型下也能安全抵御 CPA。

证明 给定 DIBBE 算法的公开参数 P_K^D 、主私钥 M_{SK}^D 及身份私钥 s_d^D 如下:

$$P_K^D = (h, h^\alpha, \dots, h^{\alpha^j}, g^\alpha, v) \quad (19)$$

$$M_{SK}^D = (g, \alpha) \quad (20)$$

$$s_d^D = g^{1/\alpha + \mathcal{H}_1(d)} \quad (21)$$

可通过随机选择 $z \leftarrow \mathbf{Z}_p$ 计算 $u = h^z$, 构建 UODSS 方案的主公钥 MPK, 主私钥 MSK 和身份私钥如下:

$$M_{PK} = (h^\alpha, u^\alpha = (h^\alpha)^z, v) \quad (22)$$

$$M_{SK} = (g, \alpha) \quad (23)$$

$$s_d = g^{1/\alpha + \mathcal{H}_1(d)} \quad (24)$$

然后, 在给定一个身份 d^* 的情况下, DIBBE 计算 $E_{nc}^D(P_K^D, d^*, m)$ 得密文 $C^D = (C_1^D, C_2^D, C_m)$:

$$C_1^D = (g^\alpha)^{-r_1} C_2^D = h^{r_1(\alpha + \mathcal{H}_1(d^*))} C_m = \varepsilon_K(m) \quad (25)$$

由于只有一个身份, 此时可移除 C_1^D ; DIBBE 算法的安全之处则在于使用身份私钥 s_{d^*} 才能恢复出对称密钥 $K = \hat{e}(s_{d^*}, C_2^D) = v^{r_1}$ 。UODSS 方案的 $E_{nc}(P, M_{PK}, m, d_u, d^*, d_c)$ 算法计算初始密文 $C = (C_0, C_1, C_2, C_3, C_m)$ 如下:

$$\begin{cases} C_0 = t \parallel d_u \parallel d^* \parallel d_c \\ C_1 = h^{r_1(\alpha + \mathcal{H}_1(d^*))} \\ C_2 = v^{r_1 \cdot \mathcal{H}_1(C_0)} \cdot K \\ C_3 = u^{r_1(\alpha + \mathcal{H}_1(d^*))} \\ C_m = \varepsilon_K(m) \end{cases} \quad (26)$$

式中: C_0 只包含时间戳和公开的身份信息; C_1 则等价于 C_2^D , 只有使用 d^* 的身份私钥 s_{d^*} 才能恢复出 v^{r_1} , 同时才能进一步从 C_2 中恢复出对称密钥 K ; C_3 用于实现重加密, 独立于初始密文安全性。因此, 在随

机谰言机模型下, 本文 UODSS 方案初始加密的安全性等价于 DIBBE 算法, 能够安全抵御选择明文攻击。

证毕

定理 5 如果 DIBBE 算法在随机谰言机模型下能够安全抵御选择明文攻击, 那么本文 UODSS 方案的重加密密钥无法解密初始密文。

证明 给定一个代理身份 d^* , 重加密密钥 $k = (k_1, k_2, k_3)$ 如下:

$$\begin{cases} k_1 = \mathcal{H}_2(v^{r_2}) \cdot h^{r_3} \\ k_2 = s_{d^*} u^{r_3} \\ k_3 = \{a_i\}_{1 \leq i \leq n} \end{cases} \quad (27)$$

根据定理 4, 只有身份私钥 s_{d^*} 才能恢复明文。上式中只有 k_2 包含了身份私钥 s_{d^*} 的信息。但是由于 r_3 是一个秘密随机数, 敌手无法直接从 k_2 中提取出 s_{d^*} 。若直接采用 k_2 解密, 则得到

$$\hat{e}(k_2, C_1) = v^{r_1} \cdot \hat{e}(u, h)^{r_1 r_3 (\alpha + \mathcal{H}_1(d^*))} \quad (28)$$

由于 r_1 也是秘密随机值, 将 $\hat{e}(u, h)^{r_1 r_3 (\alpha + \mathcal{H}_1(d^*))}$ 可看作一个随机元素 $U \leftarrow G_T$ 。因此, 式 (28) 等价于 C_2 , 依旧无法解密初始密文。

证毕

定理 6 如果 DIBBE 算法在随机谰言机模型下能够安全抵御选择明文攻击, 那么本文 UODSS 方案的初始密文重加密能安全抵御 CPA。

证明 给定初始密文 C 、重加密密钥 $k_{d^* \rightarrow S} = (k_1, k_2, k_3)$, 其中 $S = \{d_i\}_{1 \leq i \leq n}$, 重加密文得到 $R = (R_0, R_1, R_2, R_3, R_4, C_m)$ 如下:

$$\begin{cases} R_0 = C_0 = t \parallel d_u \parallel d^* \parallel d_c \\ R_1 = k_1 = \mathcal{H}_2(v^{r_2}) \cdot h^{r_3} \\ R_2 = C_2^{1/\mathcal{H}_1(C_0)} \cdot \hat{e}(k_2, C_1^{-1}) \\ R_3 = C_3 = u^{r_1(\alpha + \mathcal{H}_1(d^*))} \\ R_4 = k_3 = \{a_i\}_{1 \leq i \leq n} \\ C_m = \varepsilon_K(m) \end{cases} \quad (29)$$

根据定理 4 和定理 5, 为恢复出对称密钥 K , 不仅要消除 v^{r_1} 还要消除 $U = \hat{e}(u, h)^{r_1 r_3 (\alpha + \mathcal{H}_1(d^*))}$ 。利用重加密密钥中 k_2 嵌入的 s_{d^*} 可消除 v^{r_1} , 因此, 难点在于如何安全的消除 U 。对于共享组 S 中每个身份 d_i , 重加密密钥生成算法利用 y_i 如下:

$$y_i = h^{r_2(\alpha + \mathcal{H}_1(d_i))} \quad (30)$$

其中, y_i 等价于 C_1 , 根据定理 4, 只有对应的身份私钥 s_{d_i} 才能恢复出 v^{r_2} 、从 R_1 中恢复出 h^{r_3} 、并进一步结合 R_3 消除 U 。因此, 在随机谰言机模型下, 初始密文重加密能安全抵御 CPA。

证毕

4 性能分析

4.1 理论分析

本文方案与 CZL^[21]、ZCD^[23] 和 GLX^[24] 等现有方案的计算对比如表 1 和表 2 所示, 其中, n 表示共享组中所有的身份数量, t_e 表示 G 群上的幂运算时间, t_b 表示双线性映射运算时间。理论分析忽略哈希运算、 G 群上的乘法运算、 G 群上的求逆运算和整数群上的数学运算等, 由于它们的开销远小于 G 群上的幂运算和双线性映射运算。

表 1 初始化阶段与初始共享阶段各算法计算对比

Table 1 computation comparison of algorithms in initialization phase and initial sharing phase

方案	不同算法的计算开销			
	MasterGen	Register	Enc	Dec
CZL ^[21]	$(2n+1)t_e+t_b$	t_e	$7t_e$	t_b
ZCD ^[23]	$(2n+1)t_e+3t_b$	t_e	$5t_e$	t_b
GLX ^[24]	$2nt_e+t_b$	t_e	$4t_e$	t_b
本文方案	$2t_e+t_b$	t_e	$5t_e$	t_e+t_b

表 2 多主体共享阶段各算法计算对比

Table 2 Computation comparison of algorithms in multi-subject sharing phase

方案	不同算法的计算开销		
	ReKeyGen	ReEnc	Dec2
CZL ^[21]	$(n^2+n+5)t_e$	t_b	$(n-1)t_e+2t_b$
ZCD ^[23]	$(2n+5)t_e$	$2t_b$	nt_e+4t_b
GLX ^[24]	$(2n+6)t_e$	$2t_b$	nt_e+3t_b
本文方案	$(n+4)t_e$	t_e+t_b	t_e+2t_b

在初始化阶段中, 由于现有方案往往提前设定了共享组的容量大小为 n , 因此, 它们的 MasterGen 算法均达到了 $\mathcal{O}(n)$ 的计算复杂度。由于 UODSS 方案并没有提前固定共享组容量, 因此实现了 $\mathcal{O}(1)$ 的计算复杂度。对于 Register 算法, 本文方案与对比方案均计算身份私钥 $s_d \leftarrow g^{1/\alpha+\mathcal{H}_1(d)}$, 因此都需要计算 1 次幂运算, 整体计算复杂度为 $\mathcal{O}(1)$ 。在初始共享阶段中, 所有方案的 Enc 算法和 Dec 算法都达到了 $\mathcal{O}(1)$ 的计算复杂度。对于 Enc 算法, 本文方案比 CZL 方案少了 2 次幂运算、但比 GLX 方案多了 1 次幂运算; 对于 Enc 算法, 本文方案多了 1 幂运算。实际运行中, 1 次幂运算的时间开销在 1 ms 左右, 因此, 本文方案依旧保持良好的性能。

在多主体共享阶段中, 由于所有方案都需要在 ReKeyGen 算法中嵌入共享组所有身份的信息, 本文方案、ZCD 方案和 GLX 方案的计算复杂度达到了 $\mathcal{O}(n)$, CZL 方案的计算复杂度达到了 $\mathcal{O}(n^2)$ 。具

体而言, 本文方案依旧比 ZCD 方案和 GLX 方案少了约 n 次幂运算。对于 ReEnc 算法, 本文方案的计算开销略大于对比方案, 但是依旧保持了 $\mathcal{O}(1)$ 的计算复杂度。对于 Dec2 算法, 对比方案由于需要利用所有共享组的身份信息, 因此, 它们的计算复杂度均是 $\mathcal{O}(n)$; 本文方案只需要定位到对应的身份便可完成解密, 因此, 本文方案的计算开销为 $\mathcal{O}(1)$ 。

4.2 实验分析

本节开展实验以评估本文方案的实际运行性能。所有实验均采用 GoLang 程序设计语言, 并运行在 MacOS 系统 (Ventura 13.5.2, 32 GB RAM, 8 Intel Core i9 @2.30 GHz)。实验采用 PBC 库构建双线性映射系统 $(p, \mathbf{Z}_p, G, G_T, \hat{e})$, 所有的测试结果取 10 次试验的平均值。

4.2.1 初始化阶段实验分析

不同方案 MasterGen 算法在不同 $|\mathcal{S}|$ 值 (共享主体数量 n) 下的运行时间对比如图 6 所示。3 种对比方案 CZL、ZCD 和 GLX 的运行时间随着 $|\mathcal{S}|$ 值的增加而逐步增大; 随着共享主体数量增加, 加密算法的运行时间呈现明显的线性增长。本文方案的 MasterGen 算法运行时间显著低于其他 3 个方案; 在共享主体数量增加的情况下, 运行时间保持恒定且保持较低水平, 仅约 4.05 ms。不同方案 Register 算法均保持恒定, 仅约 1.06 ms。

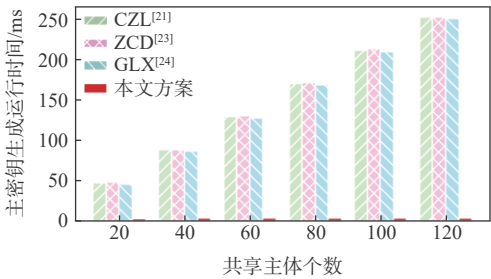


图 6 不同方案 MasterGen 算法的运行时间对比

Fig. 6 Runtime comparison of MasterGen algorithm for different schemes

4.2.2 初始共享阶段实验分析

不同方案的初始加密 Enc 算法计算复杂度均为 $\mathcal{O}(1)$, 实际运行时间对比如图 7 所示。不同方案的运行时间的波动均保持较小的水平, 且具有较优的性能表现。本文方案在 Enc 算法的执行时间上具有明显的优势, 表现出较短运行时间, 仅约 4.98 ms, 适合资源受限无人机节点加密。GLX 方案的 Enc 算法运算时间约为 4.02 ms, 虽然略优于本文方案, 但是该方案其他算法性能相较于本文方案则表现不足。不同方案的初始密文解密 Dec 算法计算复杂度也均是 $\mathcal{O}(1)$, 实际运行时间相差无几, 在 0.6~0.75 ms 之间。

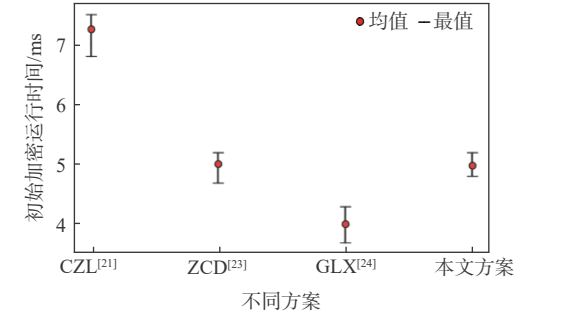


图 7 不同方案Enc算法的运行时间对比

Fig. 7 Runtime comparison of Enc algorithm for different schemes

4.2.3 多主体共享阶段实验分析

在不同的 $|S|$ 值条件下, 不同方案ReKeyGen算法的运行时间对比如表 3 所示。随着共享主体数量增加, ZCD、GLX 和本文方案的ReKeyGen算法的运行时间均呈现明显的线性增长, 且本文方案拥有最低的时间开销; CZL 方案则随着共享主体数量增加, 时间呈现指数级增长, 开销快速增加到了秒级, 远高于另外 3 种方案。相比次优方案, 本文方案ReKeyGen算法计算开销降低了 40% 以上。

表 3 不同方案 ReKeyGen 算法的运行时间对比

Table 3 Runtime comparison of ReKeyGen algorithm for different schemes

S 值	运行时间/ms			
	CZL ^[21]	ZCD ^[23]	GLX ^[24]	本文方案
20	513.61	54.74	55.92	28.74
40	1.95×10^3	102.42	104.23	49.35
60	4.33×10^3	151.80	153.01	70.44
80	7.66×10^3	197.95	199.77	91.32
100	12.72×10^3	246.70	247.83	112.14
120	17.12×10^3	295.76	297.41	132.68

不同方案的重加密ReEnc算法计算复杂度均是 $\mathcal{O}(1)$, 时间对比如图 8 所示, 均具有较优的性能表现。CZL 与本文方案的平均时间开销恒定在微妙级, 分别为 601.57 us 和 747.56 us, 相差可忽略不计;

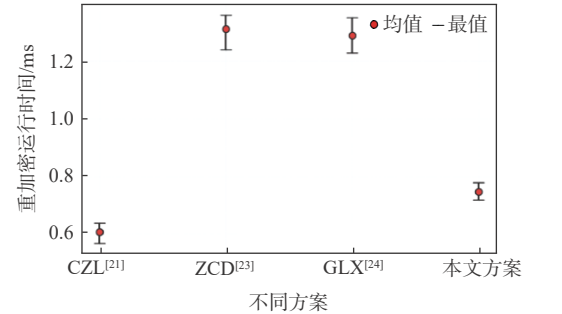


图 8 不同方案ReEnc算法的运行时间对比

Fig. 8 Runtime comparison of ReEnc algorithm for different schemes

另外 2 种方案的平均时间开销则分别为 1.31 ms 和 1.28 ms。

随着共享主体数量增加, 不同方案的时间开销如图 9 所示。3 种对比方案 CZL、ZCD 和 GLX 的Dec2算法的运行时间呈现明显线性增长, 且相差可忽略; 本文方案的时间开销不随共享主体数量变化而变化, 保持恒定的 3.6 ms。

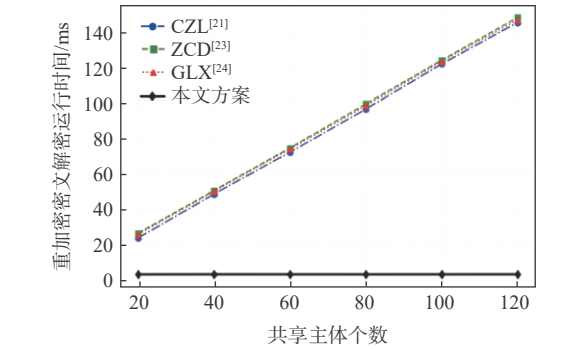


图 9 不同方案Dec2算法的运行时间对比

Fig. 9 Runtime comparison of Dec2 algorithm for different schemes

5 结 论

- 1) 本文方案初始化阶段无需提前设定共享组容量, 且主密钥生成算法的计算开销恒定, 可有效提升无人机运行数据共享的可扩展性。
- 2) 初始共享阶段加解密算法的时间开销分别约为固定的 4.02 ms 和 0.74 ms, 可有效保证点对点共享过程中的机密性和效率。
- 3) 在多主体共享阶段中, 重加密密钥生成算法相比次优方案, 时间开销降低 40% 以上, 重加密密文解密算法约为固定的 3.6 ms, 远优于具有线性开销的对比方案, 可有效保证点对多点共享过程中的安全性和高效性。

参考文献 (References)

[1] 中国信息协会低空经济分会. 低空经济发展报告 (2024—2025) [EB/OL]. (2025-02-28)[2025-04-17]. <http://www.ciia.org.cn/news/26549.cshtml>.
China Information Association Low Altitude Economy Branch. Low altitude economy development report (2024—2025)[EB/OL]. (2025-02-28)[2025-04-17]. <http://www.ciia.org.cn/news/26549.cshtml>(in Chinese).

[2] 宋志勇. 高质量发展通用航空和低空经济[EB/OL]. (2025-01-03) [2025-04-17]. https://www.gov.cn/zhengce/202501/content_6996109.htm.
Song Z Y. High-quality development of general aviation and the low altitude economy [EB/OL]. (2025-01-03) [2025-04-17]. https://www.gov.cn/zhengce/202501/content_6996109.htm(in Chinese).

[3] 低空产业联盟. 低空智能网联体系参考架构[EB/OL]. (2024-11-

11)[2025-04-17].https://www.miit-edic.org.cn/art/2024/11/11/art_1644_11070.html.
Low Altitude Industry Alliance. Reference architecture for low altitude intelligent network system [EB/OL]. (2024-11-11) [2025-04-17]. https://www.miit-edic.org.cn/art/2024/11/11/art_1644_11070.html(in Chinese).

[4] 叶海军, 王国峰, 冯志勇. 有人无人网络化信息系统动态协同技术研究[J]. 通信学报, 2023, 44(7): 185-196.
Ye H J, Wang G F, Feng Z Y. Research on dynamic cooperative technology of manned and unmanned networked information system[J]. Journal on Communications, 2023, 44(7): 185-196 (in Chinese).

[5] Zhou Z Q, Luo X L, Mao J, et al. Blockchain-assisted efficient data sharing scheme with accountability and privacy-preserving for Internet of drones networks[J]. IEEE Transactions on Vehicular Technology, 2025, 74(2): 3253-3268.

[6] 中国交通运输部. 民用无人驾驶航空器运行安全管理规则 [EB/OL]. (2024-01-01)[2025-04-17]. <https://xxgk.mot.gov.cn/2020/gz/202401/W020240103570150577134.pdf>.
Ministry of Transport of the People's Republic of China. Operational safety management rules for civil unmanned aircraft [EB/OL]. (2024-01-01) [2025-04-17]. <https://xxgk.mot.gov.cn/2020/gz/202401/W020240103570150577134.pdf>(in Chinese).

[7] Nikooghdam M, Amintoosi H, Islam S H, et al. A provably secure and lightweight authentication scheme for Internet of Drones for smart city surveillance[J]. Journal of Systems Architecture, 2021, 115: 101955.

[8] Khan M A, Ullah I, Alkhalifah A, et al. A provable and privacy-preserving authentication scheme for UAV-enabled intelligent transportation systems[J]. IEEE Transactions on Industrial Informatics, 2022, 18(5): 3416-3425.

[9] Liu Z Q, Guo J J, Huang F R, et al. Lightweight trustworthy message exchange in unmanned aerial vehicle networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(2): 2144-2157.

[10] George D R, Sciancalepore S, Zannone N. Privacy-preserving multi-party access control for third-party UAV services[C]//Proceedings of the 28th ACM Symposium on Access Control Models and Technologies. New York: ACM, 2023: 19-30.

[11] Han P B, Sui A N, Wu J. A secure and efficient access-control scheme based on blockchain and CP-ABE for UAV swarm[J]. Drones, 2025, 9(2): 148.

[12] Wang Z Z, Li X C, Cheng H T, et al. An access control scheme for UAV swarm based on CP-ABE[C]//Proceedings of the 4th International Conference on Autonomous Unmanned Systems. Berlin: Springer, 2025: 172-185.

[13] Ruan C H, Hu C Q, Li X W, et al. A revocable and fair outsourcing attribute-based access control scheme in metaverse[J]. IEEE Transactions on Consumer Electronics, 2024, 70(1): 3781-3791.

[14] Yang H J, Xu Q C. Fine-grained access control for lightweight data sharing of UAVs[C]// Proceedings of the IEEE Cyber Science and Technology Congress. Piscataway: IEEE Press, 2024: 476-479.

[15] Gayathri Y M, Rekha K S. Comparative analysis of identity-based-broadcast encryption with attribute-based encryption for reduced storage cost of multi users in a public cloud[J]. Proceedings of the 16th International Engineering and Computing Research Conference, 2024, 2729: 030001.

[16] Liu L, Zhang Y Q, Li X J. KeyD: secure key-deduplication with identity-based broadcast encryption[J]. IEEE Transactions on Cloud Computing, 2021, 9(2): 670-681.

[17] Maiti S, Misra S. P2B: privacy preserving identity-based broadcast proxy re-encryption[J]. IEEE Transactions on Vehicular Technology, 2020, 69(5): 5610-5617.

[18] Zhao Y N, Wang Y P, Liang Y H, et al. Identity-based broadcast signcryption scheme for vehicular platoon communication[J]. IEEE Transactions on Industrial Informatics, 2023, 19(6): 7814-7824.

[19] Kim J, Susilo W, Au M H, et al. Adaptively secure identity-based broadcast encryption with a constant-sized ciphertext[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(3): 679-693.

[20] Deng H, Qin Z, Wu Q H, et al. Identity-based encryption transformation for flexible sharing of encrypted data in public cloud[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3168-3180.

[21] Chen L Q, Zhang M, Li J G. Conditional identity-based broadcast proxy re-encryption with anonymity and revocation[J]. IEEE Transactions on Reliability, 2025, 74(3): 3573-3584.

[22] Zhong H, Zhang S, Cui J, et al. Broadcast encryption scheme for V2I communication in VANETs[J]. IEEE Transactions on Vehicular Technology, 2022, 71(3): 2749-2760.

[23] Zhou Y J, Cao Z F, Dong X L. Dynamic identity-based broadcast proxy re-encryption for data sharing in autonomous vehicles[J]. Transactions on Emerging Telecommunications Technologies, 2023, 34(11): e4801.

[24] Ge C P, Liu Z, Xia J Y, et al. Revocable identity-based broadcast proxy re-encryption for data sharing in clouds[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 18(3): 1214-1226.

[25] Delerablée C. Identity-based broadcast encryption with constant size ciphertexts and private keys[C]// Proceedings of the Advances in Cryptology-ASIACRYPT 2007. Berlin: Springer, 2007: 200-215.

Secure sharing scheme for low-altitude UAV operational data based on broadcast proxy re-encryption

ZHOU Zequan^{1, 2, 3}, JI Xiaohai^{1, 2, 3}, LUO Xiling^{1, 2, 3, *}, MAO Jian^{2, 4}, WANG Junjun¹

(1. School of Electronic and Information Engineering, Beihang University, Beijing 100191, China;

2. Hangzhou Innovation Institute, Beihang University, Hangzhou 310000, China;

3. State Key Laboratory of CNS/ATM, Beijing 100191, China;

4. School of Cyber Science and Technology, Beihang University, Beijing 100191, China)

Abstract: In the open and dynamic low-altitude intelligent network, unmanned aerial vehicle (UAV) operational data sharing faces security threats such as data eavesdropping, replay attacks, and unauthorized access, which may seriously jeopardize low-altitude operation security. It is crucial to guarantee the secure sharing of UAV operational data. Existing schemes utilize identity-based broadcast encryption to achieve point-to-multipoint data security sharing. Nevertheless, the sharing flexibility is sometimes restricted by these schemes' requirement to preset the sharing group capacity. Moreover, in the decryption process, the sharer needs to perform additional mathematical operations for other group members, which leads to a heavy burden of decryption computation. In this paper, we propose a secure sharing scheme for low-altitude UAV operational data based on broadcast proxy re-encryption. Through ciphertext re-encryption, the suggested technique creates a hybrid encryption mechanism that converts point-to-point sharing into point-to-multipoint sharing by utilizing identity-based broadcast encryption and symmetric encryption. In point-to-multipoint sharing, the scheme supports stateless data group sharing with fixed-length private keys and without preset group capacity, taking into account the security, flexibility, and efficiency of sharing. Experimental results show that in point-to-point sharing, the encryption and decryption computation overhead is fixed; in point-to-multipoint sharing, compared with the suboptimal scheme, the re-encryption key generation overhead is reduced by more than 40%, and the re-encryption ciphertext decryption algorithm costs about 3.6 milliseconds.

Keywords: unmanned aerial vehicle; data sharing; data security; encryption algorithm; ciphertext

Received: 2025-05-12; **Accepted:** 2025-10-31; **Published Online:** 2025-11-28 14:57

URL: link.cnki.net/urlid/11.2625.V.20251128.1055.001

Foundation items: Zhejiang Provincial " JIANBING " Research and Development Project (2022C01055); National Natural Science Foundation of China (62172027); Zhejiang Provincial Natural Science Foundation of China (LZ23F020013)

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

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0387

密集无人机目标雷达非凸 DOA 估计方法

何贵栋¹, 雷鹏^{1,*}, 王俊²

(1. 北京航空航天大学 电子信息工程学院, 北京 100191; 2. 北京航空航天大学杭州创新研究院, 杭州 310000)

摘 要: 随着低空经济的蓬勃发展, 无人机 (UAV) 在物流、巡检、监测等领域得到广泛应用, 其规模化部署带来了对目标检测与识别能力的更高要求。然而, 在实际场景中, 密集无人机常导致多目标角度接近、回波能量差异显著, 给传统到达角估计方法带来较大挑战。为此, 提出一种基于非凸优化的到达角 (DOA) 估计方法, 利用协方差矩阵特征值分析自适应估计近角目标数目, 避免对先验信息的依赖; 首先构建引入拉普拉斯先验的非凸稀疏约束模型, 在保留角度离散化的同时弱化栅格失配影响, 并通过迭代优化实现超分辨率估计; 所设计的非凸正则项在强弱目标共存条件下具备抑制主瓣干扰与提升弱目标检测能力的特性, 从而增强了对回波能量不一致的鲁棒性。仿真实验表明: 所提方法在复杂无人机探测中的有效性与鲁棒性具有更高的估计成功率与估计精度。

关 键 词: 无人机目标; 到达角估计; 角度超分辨; 非凸优化; 雷达探测

中图分类号: TN953.5; V279

文献标志码: A

文章编号: 1001-5965(2026)09-3100-08

随着低空经济的迅速发展, 无人机在物流配送、智能巡检、环境监测、应急救援等多个领域展现出广泛应用前景^[1]。无人机系统灵活高效, 但其大规模、集群式部署也带来了监管与安全的挑战, 对其空域管理、实时监测和目标识别提出了更高要求。作为关键感知手段, 雷达因其全天候、远距离、高精度的探测能力, 成为实现无人机目标探测与跟踪的重要手段^[2]。其中, 基于阵列信号处理的到达角 (direction of arrival, DOA) 估计技术在无人机目标定位与识别中扮演着至关重要的角色^[3]。

在实际场景中, 由于无人机集群运行密集、姿态多样, 常出现多个目标角度接近, 且因尺寸、材质、姿态差异, 其回波能量相差较大^[4]。这种近角分布+能量差异显著的复杂信号环境, 对 DOA 估计方法提出了双重挑战: 一方面需具备足够的角度分

辨率以区分紧邻目标, 另一方面需具备弱目标的检测能力, 以避免其被强目标谱峰掩盖。

经典的 DOA 估计方法主要包括基于特征空间分解的子空间方法和波束形成方法。其中, 多信号分类法 (multiple signal classification, MUSIC)^[5] 和基于旋转不变性的信号参数估计方法 (estimation of signal parameters via rotational invariance techniques, ESPRIT)^[6] 是子空间方法的典型代表, 它们通过对阵列接收数据的协方差矩阵进行特征分解, 将信号子空间与噪声子空间加以区分, 并利用噪声子空间的正交性在角度域上实现超分辨率估计。该类方法在中高信噪比、目标角度间隔较大的条件下具有优良的分辨性能, 计算复杂度适中, 因而在传统阵列信号处理领域得到了广泛应用。然而, 其性能高度依赖于快拍数和信噪比, 在弱信号、近角目标或能量动态范围较大的场景中, 谱峰容易发生融合或失

收稿日期: 2025-06-20; 录用日期: 2025-08-06; 网络出版时间: 2025-08-22 15:47

网络出版地址: link.cnki.net/urlid/11.2625.V.20250822.1454.001

基金项目: 实时毫米波成像芯片与设备研制 (2023C01148); 毫米波视频智能感知创新团队 (TD2022006)

*通信作者. E-mail: buaaray@gmail.com

引用格式: 何贵栋, 雷鹏, 王俊. 密集无人机目标雷达非凸 DOA 估计方法 [J]. 北京航空航天大学学报, 2026, 52 (9): 3100-3107.

He G D, Lei P, Wang J. Nonconvex radar DOA estimation method for dense UAV targets [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3100-3107 (in Chinese).

真, 导致分辨失败。此外, 当阵列存在幅相误差或通道不一致时, 子空间方法的稳定性也会显著下降。数字波束形成(digital beamforming, DBF)^[7]方法则直接利用阵列加权合成波束形成角度谱, 具有实现简单、硬件友好等优点, 但其角度分辨率受限于阵列孔径, 易出现主瓣展宽、旁瓣抬升等现象, 难以满足密集目标的高分辨需求。

为进一步提升稀疏目标条件下的分辨能力, 压缩感知(compressed sensing, CS)^[8-9]方法被引入 DOA 估计领域。其基本思想是将连续角度空间离散化为有限网格, 并利用 L_1 范数正则化实现稀疏谱的精确重构。典型的 CS 方法, 如基于 L_1 范数最小化的正交匹配追踪(orthogonal matching pursuit, OMP)^[10]、迭代阈值(iterative shrinkage-thresholding, IST)方法^[11]等, 能够在快拍数较少、信噪比较低的情况下仍保持较高的分辨率, 适用于部分稀疏场景。然而, 这类方法存在一个固有问题——栅格失配误差^[12-14], 即当真实目标位置不在预设角度网格点上时, 会导致估计结果产生偏差, 严重影响分辨精度和稳定性。为缓解该问题, 原子范数最小化(atomic norm minimization, ANM)方法^[15-19]被提出, 通过在连续角度域建立稀疏模型, 理论上实现了网格无关的最优稀疏估计, 能够在角度间隔较小的情况下保持较强的分辨能力。然而, ANM 方法默认各原子具有相同权重, 未能有效建模实际信号中强弱目标能量差异, 因此, 在强弱目标共存的场景下, 弱目标的检测性能往往不足。同时, ANM 方法的凸优化过程计算量较大, 尤其是在大规模阵列或多目标场景下, 求解效率和鲁棒性仍存在一定限制^[20-21]。

针对上述挑战, 本文提出一种基于非凸优化的 DOA 估计方法, 旨在同时提升角度分辨率与弱目标检测能力。该方法首先基于接收数据的协方差矩阵进行特征值分析, 实现对无人机目标数目的自适应估计, 从而避免对先验信息的依赖。在此基础上, 基于拉普拉斯型^[22]非凸稀疏先验构建 DOA 估计模型, 通过迭代优化实现超分辨角度估计, 并具备对无人机目标能量不一致的鲁棒性。与 ANM 等连续建模方法不同, 本文保留了角度离散化方式, 但结合非凸正则策略减弱了栅格失配影响, 并显著提升了在能量不一致条件下的谱分辨能力。仿真实验设计了多个典型的多无人机目标场景, 包括近角分布与强弱目标共存情况, 并在不同信噪比下与多种方法进行对比分析。结果表明, 本文方法复杂条件下, 仍能稳定分辨多个近角无人机目标, 并保持较高的估计成功率与精度, 验证了本文方法在复

杂无人机探测任务中的有效性与鲁棒性。

1 无人机集群的雷达阵列信号模型

考虑一个基于调频连续波(frequency modulated continuous wave, FMCW)雷达的单基地系统, 用于对密集编队飞行的无人机群体目标进行探测与角度估计。雷达发射的 Chirp 信号可表示为

$$s_{tx}(t) = \exp \{j(2\pi f_c t + \pi \mu t^2)\} \quad 0 \leq t \leq T_c \quad (1)$$

式中: f_c 为载波频率; $\mu = B/T_c$ 为调频斜率, B 为带宽, T_c 为每个脉冲的持续时间。设接收阵列由 M 元均匀线阵, 阵元间距 $d_r = \lambda/2$, 共发射 N 个脉冲用于提取多普勒信息。假设目标场景中不存在 K 架无人机, 每架无人机可视作为一个点目标。第 k 个无人机目标的距离为 R_k 、径向速度为 v_k 、空间方向为 θ_k 。

FMCW 雷达接收到混频后的中频信号经过低通滤波后, 快时间信号中包含无人机距离信息, 而慢时间维度上的多脉冲观测则包含多普勒频率信息。在第 m 个阵元、第 n 个脉冲下, 对应的快时间采样点 t_q 的接收信号为

$$s_{n,m}(t_q) = \sum_{k=1}^K \alpha_k \exp \left\{ -j2\pi \left(f_b^k t_q + f_d^k n T_c - \frac{m d_r \sin \theta_k}{\lambda} \right) \right\} + w_{n,m}(t_q) \quad (2)$$

式中: $f_b^k = 2\mu R_k/c$ 为目标的拍频, 反映其径向距离; $f_d^k = 2v_k/\lambda$ 为多普勒频率; α_k 为复数回波幅度, 含雷达散射截面与传播衰减; θ_k 为第 k 个无人机目标的角度; $w_{n,m}(t_q)$ 为加性白噪声。

对每个脉冲的快时间数据 $s_{n,m}(t_q)$ 进行距离向快速傅里叶变换(fast Fourier transform, FFT), 提取出第 b 个距离单元的频谱:

$$S_{n,m}[b] = \sum_{q=0}^{Q-1} s_{n,m}(t_q) e^{-j2\pi b q/Q} \quad (3)$$

式中: Q 为快时间采样点数; b 为距离单元索引。然后对 N 个脉冲在慢时间维进行 FFT 提取多普勒频谱, 得到二维距离-多普勒谱:

$$\mathbf{Z}_m[b, d] = \sum_{n=0}^{N-1} S_{n,m}[b] e^{-j2\pi d n/N} \quad (4)$$

式中: d 表示多普勒频率单元的索引。最终形成的二维数据张量 $\mathbf{Z}_m[b, d]$ 中, 每个距离-多普勒单元对应一个 $M \times 1$ 的阵列观测向量, 表示该单元在各阵元接收到的复数幅度。

在常规检测流程中, 采用二维恒虚警检测方法(constant false alarm rate, CFAR)^[23] 在 (b, d) 平面上进行门限判决。若某个单元的功率满足式(5), 则可

判断该距离-多普勒单元是否存在目标回波。

$$\begin{cases} H1: |Z_m[b, d]|^2 \geq T_{\text{CFAR}} \frac{1}{N_{\text{ref}}} \sum_{(i, j) \in W} |Z_m[i, j]|^2 \\ H0: |Z_m[b, d]|^2 < T_{\text{CFAR}} \frac{1}{N_{\text{ref}}} \sum_{(i, j) \in W} |Z_m[i, j]|^2 \end{cases} \quad (5)$$

式中: W 表示 CFAR 参考单元集合; N_{ref} 为参考单元数量; T_{CFAR} 恒虚警门限比例因子; $H1$ 和 $H0$ 分别表示检测问题中的 2 类统计假设: $H0$ 目标不存在(该单元仅包含噪声), $H1$ 为目标存在(该单元中含有目标回波信号)。

然而在无人机群体目标场景中, 多个目标可能落入同一个 (b, d) 单元中, 即同一距离和速度范围内存在多个角度差异较小的目标。此时, 二维 FFT 无法区分这些目标, 仅在该单元中形成一个强能量峰值。为实现进一步分辨, 本文提取该单元在空间阵列维度上的观测向量:

$$\mathbf{x} = [Z_0[b, d], Z_1[b, d], \dots, Z_{M-1}[b, d]]^T = \sum_{k=1}^{K_b} \beta_k \cdot \mathbf{a}(\theta_k) + \mathbf{n} \quad (6)$$

式中: K_b 表示该单元中实际存在的目标数; $\mathbf{a}(\theta_k)$ 为阵列导向矢量:

$$\mathbf{a}(\theta_k) = \left[1, e^{-j \frac{2\pi d}{\lambda} \sin \theta_k}, \dots, e^{-j \frac{2\pi d}{\lambda} (M-1) \sin \theta_k} \right]^T \quad (7)$$

其中: β_k 为第 k 个目标在该单元中的复数反射幅度; $\mathbf{n}$ 为噪声向量。

在式 (6) 中, 目标角度信息完全嵌入在导向矢量 $\mathbf{a}(\theta_k)$ 中。若目标数 K_b 已知, 理论上可采用 MUSIC 或 ESPRIT 等子空间方法进行估计; 若 K_b 未知或角度过近、信噪比偏低, 则这些方法往往失效, 难以准确分辨多个无人机目标。

2 基于非凸优化的 DOA 估计方法

本节首先基于接收数据协方差矩阵的特征值分析, 实现无人机目标数目的自适应估计; 随后, 构建了基于拉普拉斯范数诱导的非凸稀疏 DOA 估计模型, 并给出了相应的迭代求解框架。

2.1 无人机数目估计

在实际应用中, 无人机目标数量 K_b 往往未知。为开展高分辨方向估计, 本文首先需要估计当前距离-多普勒单元中实际存在的无人机数量 K_b 。该问题可形式化为对阵列观测向量中的稀疏度进行估计。传统的稀疏重构方法通常依赖人为设定稀疏度参数, 然而在实际场景中, 目标数量往往未知或动态变化, 若估计不足则可能遗漏目标, 估计过多则可能引入噪声伪峰。因此, 稀疏度的准确估计对

于后续稀疏建模质量至关重要。

为此, 本文采用基于 Marčenko-Pastur 边界^[24]理论的特征值统计法来估计信号子空间维度, 即估计目标数 K_b 。设当前距离-多普勒单元对应的阵列观测矩阵 $\mathbf{Y} \in \mathbf{C}^{M \times L}$, 其中, M 为阵列元数, L 为快拍数(或独立采样数), 则其协方差矩阵估计为

$$\mathbf{R}_Y = \mathbf{Y} \mathbf{Y}^H / L \quad (8)$$

对协方差矩阵 $\mathbf{R}_Y$ 进行特征值分解, 记其特征值按照降序排列为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ 。

在纯噪声假设下, Marčenko-Pastur 定理指出协方差矩阵的特征值将集中分布在一个有限区间内, 其上限可由下式给出:

$$\lambda_{\text{MP}}^{\max} = \sigma_n^2 (1 + \sqrt{\beta})^2 \quad (9)$$

式中: σ_n^2 为噪声功率, $\beta = M/L$ 为维数比。由于实际噪声功率不可知, 本文使用最小一部分特征值的平均作为 σ_n^2 的估计, 即

$$\hat{\sigma}_n^2 = \frac{1}{M-k} \sum_{i=k+1}^M \lambda_i \quad (10)$$

式中: k 为一个较小的初始信号源数猜测(如 $k=5$ 或取为 $0.5M$ 以内)。然后, 据此计算理论上限 $\lambda_{\text{MP}}^{\max}$ 。最后, 所有大于该上限的特征值数量即为估计的目标数:

$$\hat{K}_b = \{i | \lambda_i > \lambda_{\text{MP}}^{\max}\} \quad (11)$$

2.2 非凸 DOA 估计模型

在估计出该单元内可能存在的目标数 K_b 后, 不妨设其真实角度为 $\{\theta_1, \theta_2, \dots, \theta_{K_b}\}$, 对应的阵列流形向量分别为 $\{\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_{K_b})\}$ 。为了进行方向估计, 通常预先设定一个包含真实目标方向的完备角度搜索网格 $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$, 构造角度域上的阵列流形矩阵:

$$\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_N)] \in \mathbf{C}^{M \times N} \quad (12)$$

式中: $\mathbf{a}(\theta_n)$ 表示方向 θ_n 下的阵列响应向量。由于目标数远小于角度网格数量, 目标方向对应的系数在 Θ 上呈现稀疏性, 因此, 可以建立如下角度域稀疏观测模型:

$$\mathbf{Y} = \mathbf{A} \mathbf{X} + \mathbf{N} \quad (13)$$

式中: $\mathbf{X} \in \mathbf{C}^{N \times L}$ 为待估的角度系数矩阵, 仅在对目标角度的行上非零; $\mathbf{N} \in \mathbf{C}^{M \times L}$ 表示加性噪声项。该模型刻画了快拍观测数据在角度域的稀疏表达, 方向估计问题即转化为从 $\mathbf{Y}$ 中恢复出具有行稀疏结构的 $\mathbf{X}$, 进而确定非零行所对应的角度位置。

若采用最直接的稀疏建模方式, 即基于 L_0 范数的稀疏约束模型, 则可写为

$$\begin{cases} \min_x \|Y - AX\|_F^2 \\ \text{s.t. } \|X\|_{2,0} \leq r \end{cases} \quad (14)$$

式中: $\|X\|_{2,0}$ 表示非零行数, 等价于目标个数; r 表示稀疏度, 一般情况下可认为 $r = \hat{K}_b$ 。然而 L_0 范数非凸、非连续, 求解难度极高, 属于 NP-hard 问题。为此, 文献中常通过引入连续可导的稀疏正则函数对其进行近似^[10-14]。常见的凸替代如 L_1 范数虽然具备计算简洁的优势, 但在近角、高相干、信号幅度差异较大时往往恢复性能不足, 容易出现虚假目标、弱目标漏检等问题。

为克服上述限制, 本文引入一种基于拉普拉斯先验构造的非凸稀疏正则项, 更精确地逼近 L_0 行稀疏结构, 同时具备较强的稀疏激励能力与抗主瓣干扰能力。具体而言, 对角度索引第 i 行的快拍向量 $\mathbf{x}_i \in \mathbb{C}^{1 \times L}$, 定义稀疏约束函数为

$$\rho(\mathbf{x}_i) = 1 - \exp(-\|\mathbf{x}_i\|_1/\gamma) \quad (15)$$

式中: $\gamma > 0$ 为调节参数。该函数在小幅度区域诱导稀疏性较强, 有助于抑制虚假目标; 在大幅度区域呈现饱和性, 避免过度惩罚真实目标。

基于该正则项, 本文构建如下非凸稀疏优化模型:

$$\min_x \frac{1}{2} \|Y - AX\|_F^2 + \lambda \sum_{i=1}^N \rho(\mathbf{x}_i) \quad (16)$$

式中: $\lambda > 0$ 为正则化系数, 用于平衡数据拟合与稀疏结构的权重。由于正则项 $\rho(\cdot)$ 非凸且不可分离, 该优化问题无法直接采用解析形式求解。为此, 本文设计了一种近似求解策略, 结合残差反馈、非凸扰动构造与稀疏结构投影, 构成一种迭代更新的硬阈值方法。

该方法首先将目标函数分解为数据保真项与稀疏正则项两部分。对于数据项 $f(\mathbf{X}) = \frac{1}{2} \|Y - AX\|_F^2$, 其梯度可近似表示为

$$\nabla f(\mathbf{X}) = -A^H(Y - AX) \quad (17)$$

结合当前估计值构造残差反馈方向:

$$\mathbf{B}^k = \mathbf{X}^k + A^H(Y - A\mathbf{X}^k) \quad (18)$$

对于稀疏正则项 $\lambda \sum_{i=1}^N \rho(\mathbf{x}_i)$, 其元素级导数为

$$\frac{\partial \rho(\mathbf{x})}{\partial x} = \frac{\lambda}{\gamma} \cdot \exp(-|x|/\gamma) \cdot \text{sign}(x) \quad (19)$$

从而构造扰动向量:

$$\mathbf{v}^k = \alpha \lambda \cdot \text{sign}(\mathbf{X}^k) - \frac{\lambda}{\gamma} \cdot \exp(-|\mathbf{X}^k|/\gamma) \cdot \text{sign}(\mathbf{X}^k) \quad (20)$$

将两者合成, 构造更新方向:

$$\mathbf{D}^k = \mathbf{B}^k + \mathbf{v}^k/2 \quad (21)$$

为提高更新稳定性, 引入自适应步长

$$\mu_k = \frac{\|\mathbf{B}^k\|_F^2}{\|\mathbf{A}\mathbf{B}^k\|_F^2 + \varepsilon} \quad (22)$$

式中: $\varepsilon > 0$ 为防止除 0 的微小常数。最后, 对更新结果施加硬阈值稀疏投影:

$$\mathbf{X}^{k+1} = H_r(\mathbf{X}^k + \mu_k \cdot \mathbf{D}^k) \quad (23)$$

式中: $H_r(\cdot)$ 表示行硬阈值操作, 即保留行范数最大的 r 行, 其余清零, 保持稀疏结构。

通过对式 (18)~式 (23) 的不断迭代, 直到满足收敛条件或达到最大迭代次数, 即可得到最终的 DOA 估计结果。

此外, 本文提出的非凸稀疏建模方法本身即具备一定的能量不一致鲁棒性。具体而言, 非凸稀疏正则在小幅值区域具有较大的梯度响应, 能在优化过程中优先“激活”弱目标, 从而有效避免其被压缩为 0; 而对大幅值强目标的正则惩罚则相对饱和, 有助于减弱其主导地位。这种“强目标抑制、弱目标提升”的非对称响应机制, 使得非凸方法能实现对不同能量级别目标的稳定恢复。

3 仿真实验

本节通过仿真实验验证本文方法的有效性。雷达发射 FMCW 信号, 具体雷达参数如表 1 所示。

表 1 雷达仿真参数	
Table 1 Radar simulation parameters	
雷达参数	数值
雷达载频 f_c /GHz	77
调频斜率 μ /(MHz·μs ⁻¹)	58
距离采样数	256
Chirp 数	40
脉冲重复周期/μs	100
帧数	300
快拍数	4
带宽 B /GHz	1.5
距离分辨率/m	0.1
速度分辨率/(m·s ⁻¹)	0.5
虚拟阵元数	86
角度分辨率/(°)	1.33

获得雷达回波后, 用本文方法进行 DOA 估计, 并且与 MUSIC、DBF、OMP、ANM 方法进行对比。

实验 1 为验证方法的超分辨能力, 本文设定一个近角测试场景: 假设在 CFAR 检测后的某一距离单元内, 存在 3 个角度接近的无人机, 信噪比为 5 dB (高斯白噪声背景下), 3 个目标入射角分别为 -6°、-1°和 0°, 且其回波能量相等。图 1 展示了 5 种

DOA 估计方法在该场景下的估计结果。

图 1 中红色虚线表示 3 个无人机目标的真实方向,可见,3 个无人机回波能量相同,体现了等强度目标的场景设定;蓝色线则分别表示各 DOA 估计方法输出的归一化谱强度。

图 1(a)~图 1(d) 分别展示了 MUSIC、DBF、OMP 及 ANM 这 4 种方法在本实验场景下的估计结果。可以看出, MUSIC 和 DBF 能获得 2 个谱峰,但 -1° 和 0° 处的目标响应存在谱峰融合现象,角度谱分辨能力有限,存在较宽主瓣,无法有效区分 3 个目标; OMP 和 ANM 虽然具有较高的谱峰尖锐性与光滑性,但同样未能正确识别 -1° 和 0° 目标,说明在该信噪比下它们仍面临分辨性能不足的问题。图 1(e) 为本文方法的估计结果,可以看出,其能够准确分辨出 3 个目标的位置,谱峰位置与真实角度高度一致,且无明显主瓣展宽或假峰出现,展现出优越的角度分辨能力。

实验 2 为验证方法在回波能量不一致时的性能,本文设定一个测试场景:假设在 CFAR 检测后的某一距离单元内,存在 3 个角度接近的无人机,信噪比为 5 dB(高斯白噪声背景下),3 个目标的入射角分别为 -6° 、 -1° 和 2° ,且其回波能量不一致。图 2 展示了 5 种 DOA 估计方法在该场景下的估计结果。

由图 2 中红色虚线可见,3 个无人机目标的回波能量存在明显不一致, -1° 方向目标的能量最强, -6° 次之, 2° 最弱,构成了一个典型的强弱目标共存场景。

图 2(a)~图 2(c) 分别展示了 MUSIC、DBF 和 OMP 的估计结果。3 种方法均在 -1° 处形成明显的主峰,对该最强目标的检测较为准确;在 -6° 附近也能形成次级谱峰,幅度略低;然而,对于最弱的 2° 目标,3 个方法均未能有效分辨,其响应被噪声或旁瓣淹没,无法体现出谱峰,表现出对能量动态范围变化的不敏感或鲁棒性不足。

图 2(d) 显示了 ANM 方法的估计结果。可以看出, ANM 成功识别出能量较强的 -6° 和 -1° 这 2 个目标,峰值清晰且尖锐,但 2° 弱目标被完全淹没,未在谱图中体现出来,说明 ANM 在弱目标旁存在较强目标时分辨能力受限,存在“谱峰遮蔽”问题。

图 2(e) 为本文方法的估计结果。可以看到,3 个目标的角度均被准确定位,且谱峰强度与目标能量基本对应。即使是 2° 的弱目标,也能在谱图中形成清晰而准确的响应,未被强目标遮蔽,体现出较强的动态能量分辨能力和稳定性。

实验 3 为评价本文方法在不同信噪比条件下

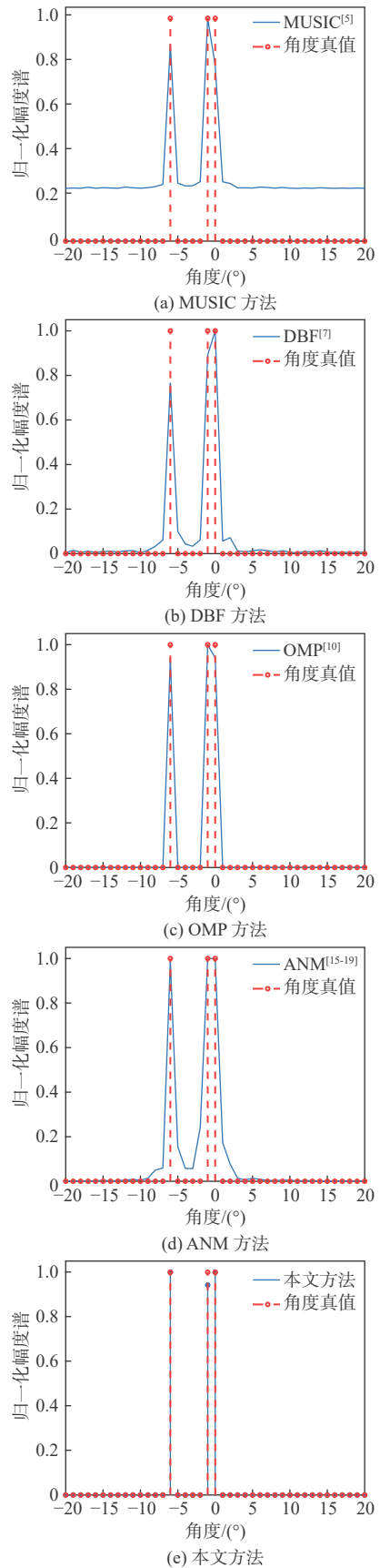


图 1 MUSIC、DBF、OMP、ANM 和本文方法在三目标近角场景下的 DOA 谱估计结果

Fig. 1 DOA spectrum estimation results of MUSIC, DBF, OMP, ANM and the proposed method in a three closely spaced target scenario

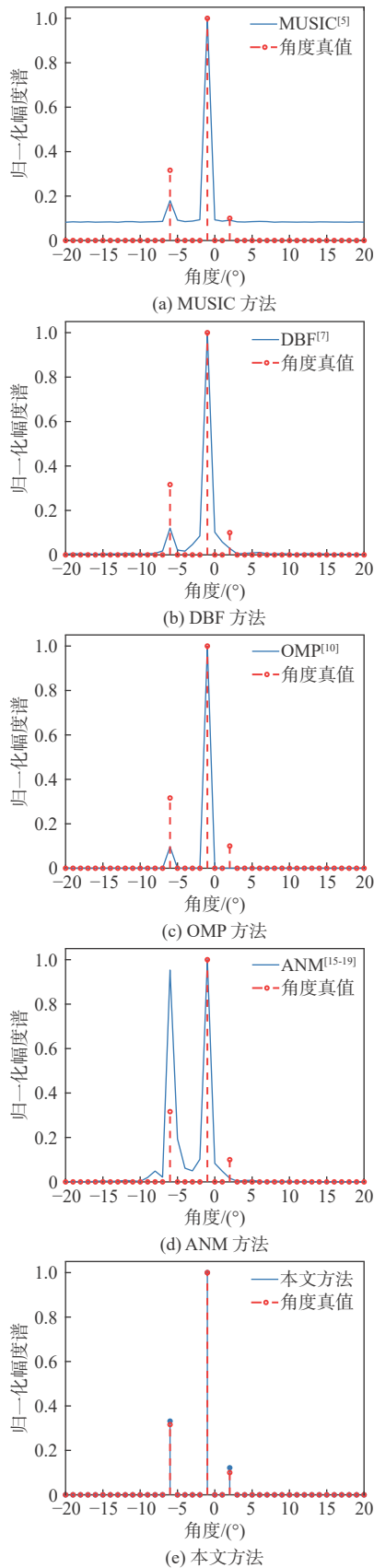


图 2 MUSIC、DBF、OMP、ANM 和本文方法在三目标能量不一致场景下的 DOA 谱估计结果

Fig. 2 DOA spectrum estimation results of MUSIC, DBF, OMP, ANM and the proposed method in a three-target scenario with unequal target powers

的稳健性与平均性能, 本文设计了 Monte Carlo 模拟实验: 在每次仿真中, 假设场景中存在 3 个无人机目标, 其角度分别位于 $[-11^{\circ}, -6^{\circ}]$ 、 $[-1^{\circ}, 0^{\circ}]$ 与 $[1^{\circ}, 2^{\circ}]$ 的区间内, 表示 3 个近角目标簇; 同时, 为模拟不同目标的能量差异, 其回波能量在 $[0.1, 10]$ 范围内均匀随机分布。噪声为加性高斯白噪声, 信噪比在 $0 \sim 25$ dB 范围内变化, 步长为 5 dB。

对于每个信噪比设置, 进行 500 次 Monte Carlo 实验, 采用估计成功率和均方根误差 (root mean square error, RMSE) 作为性能评价指标。成功率定义为: 若估计结果中成功检测出所有 3 个目标, 且每个估计角度与对应真实角度之间的误差小于某一容差阈值 (如 1°), 则视为一次成功估计。

在计算 RMSE 时, 只使用估计成功的估计结果。图 3 给出了不同 DOA 方法在各信噪比下的估计成功率曲线。可以看出, 本文方法在各信噪比条件下均表现出较高的成功率, 平均估计成功率超过其他方法约 9.3%; 在低信噪比区间 ($0 \sim 5$ dB) 下依然保持较强的鲁棒性, 相比其他方法具有更优的稳健性。

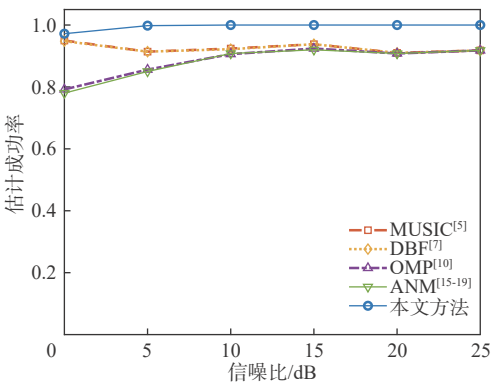


图 3 MUSIC、DBF、OMP、ANM 及本文方法在不同信噪比下的 DOA 估计成功率对比结果

Fig. 3 Comparison of DOA estimation success rates for MUSIC, DBF, OMP, ANM and the proposed method under different SNRs

图 4 给出了不同 DOA 方法在各信噪比下的 RMSE 曲线。可以看出, 本文方法在各信噪比条件下均保持较低的角度估计误差, 整体精度优于其他对比方法, 平均提升约 0.12° 。相比之下, 其他方法在目标角度接近或能量不一致的情况下, 谱峰易出现畸变或偏移, 导致估计误差显著增大。

尽管本文方法在近角强弱目标共存的复杂场景中表现出较高的分辨能力与鲁棒性, 但仍存在一定局限性。首先, 在极低信噪比条件下, 目标数估计易受噪声特征值扰动影响而产生偏差, 从而影响后续角度估计精度。其次, 当阵列存在较大幅相误差或目标呈现非点源散射特性时, 所采用的理想阵列与点目标假设可能不再完全成立, 导致模型拟合

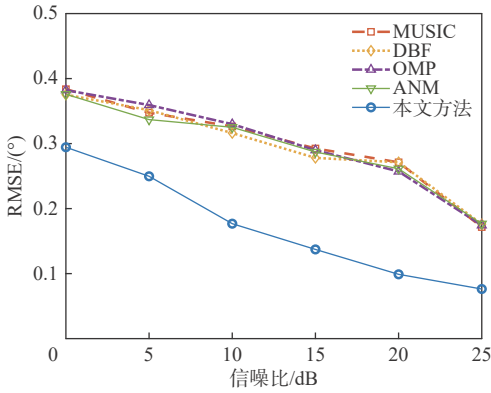


图4 MUSIC、DBF、OMP、ANM 及本文方法在不同信噪比下的 DOA 估计 RMSE 对比结果

Fig. 4 Comparison of DOA estimation RMSE for MUSIC, DBF, OMP, ANM and the proposed method under different SNRs

精度下降,可能造成角度估计错误。最后,相较于部分凸优化方法,该方法在超大规模阵列或长时间观测数据处理时的计算开销仍有提升空间。因此,后续工作将围绕阵列误差建模、弱信号条件下的鲁棒目标数估计及高效求解策略优化等方向展开进一步研究。

4 结 论

本文针对近角强弱目标共存条件下无人机目标的角度估计问题,提出一种基于非凸优化的 DOA 估计方法,并设计仿真实验对本文方法进行了验证。结论如下:

- 1) 在目标角度间隔较小(相差 1°)、能量差异显著时,本文方法能够准确分辨,明显优于 MUSIC、DBF、OMP 和 ANM 方法;
- 2) 在 0~25 dB 条件下表现出较高的估计成功率,平均估计成功率超过其他方法约 9.3%;
- 3) 在 0~25 dB 条件下均保持较低的角度估计误差,整体精度优于其他对比方法,平均提升约 0.12°。

参考文献 (References)

[1] 裴德馨. 低空经济中无人机产业发展概述[J]. 商业全球化, 2024(3): 117-122.
Qiu D X. Overview of the development of the UAV industry in the low-altitude economy[J]. Business and Globalization, 2024(3): 117-122(in Chinese).
[2] Ahmad B I, Rogers C, Harman S, et al. A review of automatic classification of drones using radar: key considerations, performance evaluation, and prospects[J]. IEEE Aerospace and Electronic Systems Magazine, 2024, 39(2): 18-33.
[3] Almasri S A, Johannsen N L, Hoeher P A. Direction-of-arrival estimation for unmanned aerial vehicles and aircraft transponders using a multi-mode multi-port antenna[J]. Sensors, 2024, 24(11): 3452.

[4] Zhang C H, Wang W J, Hong X, et al. A multi-sources DOA localization method based on UAV cluster systems[J]. IEEE Transactions on Vehicular Technology, 2024, 73(6): 8681-8692.
[5] Schmidt R. Multiple emitter location and signal parameter estimation[J]. IEEE Transactions on Antennas and Propagation, 1986, 34(3): 276-280.
[6] Roy R, Kailath T. ESPRIT-estimation of signal parameters via rotational invariance techniques[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1989, 37(7): 984-995.
[7] Capon J. High-resolution frequency-wavenumber spectrum analysis[J]. Proceedings of the IEEE, 1969, 57(8): 1408-1418.
[8] Malioutov D, Cetin M, Willsky A S. A sparse signal reconstruction perspective for source localization with sensor arrays[J]. IEEE Transactions on Signal Processing, 2005, 53(8): 3010-3022.
[9] Stoica P, Babu P. Sparse estimation of spectral lines: grid selection problems and their solutions[J]. IEEE Transactions on Signal Processing, 2012, 60(2): 962-967.
[10] Yang Z, Xie L H. Enhancing sparsity and resolution via reweighted atomic norm minimization[J]. IEEE Transactions on Signal Processing, 2016, 64(4): 995-1006.
[11] Hu Y Q, Sun S Q. IHT-inspired neural network for single-snapshot DOA estimation with sparse linear arrays[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2024: 13081-13085.
[12] Jagannath R, Leus G, Pribić R. Grid matching for sparse signal recovery in compressive sensing[C]//Proceedings of the 9th European Radar Conference. Piscataway: IEEE Press, 2013: 111-114.
[13] Yang Z, Xie L H, Zhang C S. Off-grid direction of arrival estimation using sparse Bayesian inference[J]. IEEE Transactions on Signal Processing, 2013, 61(1): 38-43.
[14] Xenaki A, Gerstoft P. Grid-free compressive beamforming[J]. The Journal of the Acoustical Society of America, 2015, 137(4): 1923-1935.
[15] Fu J, Cao Y H, Yeo T S, et al. A super-resolution method based on iterative weighted atomic norm minimization for UAV swarms[J]. IEEE Transactions on Aerospace and Electronic Systems, 2025, 61(2): 4669-4684.
[16] Li R, Yang J C, Dai Z, et al. High-resolution and robust one-bit direct-of-arrival estimation via reweighted atomic norm estimation[J]. Sensors, 2024, 24(18): 5936.
[17] Zhu H G, Feng W K, Feng C Q, et al. Deep unfolded gridless DOA estimation networks based on atomic norm minimization[J]. Remote Sensing, 2023, 15(1): 13.
[18] Gao S Z, Ma H, Liu H W, et al. A gridless DOA estimation method for sparse sensor array[J]. Remote Sensing, 2023, 15(22): 5281.
[19] Gao S L, Wang M H, Zhang Z, et al. Efficient gridless 2D DOA estimation based on generalized matrix-form atomic norm minimization[J]. Electronics Letters, 2024, 60(10): e13212.
[20] Tian Z, Zhang Z, Wang Y. Low-complexity optimization for two-dimensional direction-of-arrival estimation via decoupled atomic norm minimization[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2017: 3071-3075.
[21] Wei Z Y, Wang W, Dong F W, et al. Gridless one-bit direction-of-

- arrival estimation via atomic norm denoising[J]. IEEE Communications Letters, 2020, 24(10): 2177-2181.
- [22] Cui A G, He H Z, Xie Z Q, et al. Iterative difference hard-thresholding algorithm for sparse signal recovery[J]. IEEE Transactions on Signal Processing, 2023, 71: 1093-1102.
- [23] Richards M A. Fundamentals of radar signal processing[M]. New York: McGraw-Hill, 2005.
- [24] Couillet R, Pascal F, Silverstein J W. The random matrix regime of Maronna's M-estimator with elliptically distributed samples[J]. Journal of Multivariate Analysis, 2015, 139(C): 56-78.

Nonconvex radar DOA estimation method for dense UAV targets

HE Guidong¹, LEI Peng^{1,*}, WANG Jun²

(1. School of Electronic and Information Engineering, Beihang University, Beijing 100191, China;

2. The Hangzhou Innovation Institute of Beihang University, Hangzhou 310000, China)

Abstract: With the rapid development of the low-altitude economy, unmanned aerial vehicles (UAV) have been widely deployed in fields such as logistics, inspection, and surveillance. Their large-scale deployment imposes increasing demands on target detection and identification capabilities. However, in real-world situations, UAV swarms frequently result in widely separated arrival angles and notable variations in echo energy, which provide major difficulties for conventional direction of arrival (DOA) estimate techniques. To address this, this paper proposes a DOA estimation algorithm based on non-convex optimization. First, the number of closely spaced targets is adaptively estimated via eigenvalue analysis of the covariance matrix, thereby avoiding dependence on prior information. Then, a non-convex sparse constraint model incorporating a Laplacian prior is constructed, which preserves angle discretization while mitigating the off-grid effect, and achieves super-resolution angle estimation through iterative optimization. In the presence of coexisting strong and weak targets, the suggested non-convex regularization term can improve robustness to echo power imbalance by suppressing main-lobe interference and increasing weak target detection. Simulation results validate the effectiveness and robustness of the proposed method in complex UAV detection scenarios, demonstrating higher estimation accuracy and success rate.

Keywords: unmanned aerial vehicles targets; direction of arrival estimation; angle super-resolution; non-convex optimization; radar sensing

Received: 2025-06-20; **Accepted:** 2025-08-06; **Published Online:** 2025-08-22 15:47

URL: link.cnki.net/urlid/11.2625.V.20250822.1454.001

Foundation items: Development of Real-Time Millimeter-Wave Imaging Chips and Devices (2023C01148); Millimeter-Wave Video Intelligent Perception Innovation Team (TD2022006)

* **Corresponding author.** E-mail: buaaray@gmail.com

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0834

基于大模型属性解析的空地行人检索方法

谭荃戈¹, 王蓉¹, 廖曼丞², 李欣^{1,*}

(1. 中国人民公安大学 信息网络安全学院, 北京 100038; 2. 中国人民公安大学 侦查学院, 北京 100038)

摘 要: 可在监控探头分布稀疏区域协同分析无人机拍摄图像与固定监控视频图像, 基于行人图像的跨视角身份认定实现对目标持续追踪的巡检任务。无人机俯视角与地面监控的平视角存在显著差异, 导致基于传统监控场景设计的图像检索算法在地空跨视角场景中迁移性能受限。现有空地行人检索方法主要聚焦于缓解跨视角引起的表现差异, 而对行人属性特征的挖掘与分析尚未充分深入。为解决上述问题, 提出基于大模型属性解析的空地行人检索方法。基于多模态大模型构建行人解析模块, 生成细粒度语义属性并设计属性三元组损失实现跨视角语义对齐; 引入视角解耦架构, 通过分层减法分离剥离视角相关特征, 并施加正交损失约束特征独立性; 引入多尺度扩增 Transformer, 结合多尺度扩张注意力与全局自注意力优化计算复杂度与感受野平衡, 降低模型参数量。在 AG-ReID.v1、AG-ReID.v2、CARGO 数据集上的实验表明: 所提方法能够有效提升空地行人检索任务在 Rank-1、mAP 与 mINP 指标上的性能, 证明了该方法的有效性。

关 键 词: UAV; 行人检索; 空地图像; 属性解析; 多模态大语言模型

中图分类号: TP391.4; V279

文献标志码: A

文章编号: 1001-5965(2026)09-3108-09

无人机巡检技术的发展为现代安防技术体系提供了新方法。无人机俯视角相机可快速覆盖传统地面监控无法企及的区域, 实时捕捉人员动态。空地行人检索技术能够将航拍行人目标的视觉特征与地面监控行人特征进行跨视角关联匹配, 在监控探头稀疏区域实现无人机与固定监控探头协同的行人持续搜索与轨迹追踪。

经典的行人检索方法多基于卷积神经网络构建, 基于部件的卷积基线(part-based convolutional baseline, PCB)网络^[1]、多粒度网络(multiple granularity network, MGN)^[2]通过多分支结构和细粒度特征学习显著提升性能, 文献[1,3]通过融入注意力机制增强行人特征表征能力, 在 Market1501^[4]、MSMT17 多场景多时段行人重识别数据集^[5]等地面固定视角数据集上取得突破。然而, 卷积神经网络

的有限感受野难以应对复杂场景的时空关联建模, 这推动 Transformer^[6] 架构在行人检索领域的应用: 基于 Transformer 的目标重识别模型(Transformer-based object re-identification, TransReID)^[7] 及其衍生模型^[8-11] 通过全局自注意力机制捕获特征长距离依赖, 在遮挡和姿态变化场景表现出色。尽管这些行人检索方法能够对不同摄像头视角下的同一行人进行有效检索, 但其固定部署模式仍无法应对稀疏摄像头覆盖区域的目标搜索需求, 因此, 无人机的普及催生空地协同行人检索^[12] 研究。

空地行人检索作为新兴方向, 致力于解决无人机俯视角与地面监控视角间的跨平台身份关联难题。该领域面临双重挑战: 首先, 视角差异引发的人体几何畸变, 使经典行人检索模型的跨视角匹配精度下降; 其次, 无人机影像分辨率较低, 导致细粒

收稿日期: 2025-12-02; 录用日期: 2026-01-25; 网络出版时间: 2026-04-03 10:45

网络出版地址: link.cnki.net/urlid/11.2625.V.20260402.1729.004

基金项目: 国家自然科学基金(62076246)

* 通信作者. E-mail: lixin@ppsuc.edu.cn

引用格式: 谭荃戈, 王蓉, 廖曼丞, 等. 基于大模型属性解析的空地行人检索方法[J]. 北京航空航天大学学报, 2026, 52(9): 3108-3116.

Tan Q G, Wang R, Liao M C, et al. Large model attribute parsing-based aerial-ground pedestrian retrieval method[J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52(9): 3108-3116 (in Chinese).

度特征提取困难。Schumann 等^[12]构建首个空地行人检索数据集并提出一种基于深度学习的行人检索方法。Nguyen 等^[13]提出一种基于软属性解释的双流模型,并在后续工作中扩展为可解释三流模型^[14]。Zhang 等^[15]设计视角解耦 Transformer 网络从全局特征中分离视角特征,并通过 3D 模拟创建了 CARGO 数据集。Wang 等^[16]通过引入扩散模型生成不同视角之间的视图特定特征来检索目标人员。上述方法主要针对地空视角的行人视觉特征差异进行分析,其中,基于属性的方法是通过模型获取的“软属性”,需要人工对图像中的每个行人进行属性标注,并且训练模型预测行人属性。属性行人检索通过引入语义先验突破细粒度识别瓶颈。渐进式注意力增强网络 (progressive attention-enhanced network, PAENet)^[17]引入了 8 个属性作为注释来学习细粒度的语义属性信息。Tan 等^[18]设计了多个模板指导多模态大模型生成图像标签用于模型训练。Yang 等^[19]利用扩散模型构建了包含用于行人检索的多属性与语言搜索 (multi-attribute and language search, MALS) 数据集,并同时进行了文本匹配学习和属性学习。这些方法已经证明引入属性信息对于地面监控视角下的行人目标身份认定具有优势。

上述方法已经证明已有工作有效提升了模型在空对地场景中的特征判别能力、泛化能力、对行人外观显著变化的适应能力。然而,低空无人机视角与地面监控视角之间存在显著差异,使得模型容易忽略对行人检索至关重要的细节特征,制约了检索精度的提升。为挖掘行人细节特征,已有方法通过分析行人属性提升模型细节特征判别力,但应用于地空行人检索方法中的属性需要人工进行标注,且不同视角图像的造成属性预测准确度较低。因此,本文通过引入多模态大模型进行空地行人属性解析与属性三元组损失约束模型学习一致性的身份特征,通过属性描述语义缓解视角差异带来的影响,减少对大规模人工属性标注的依赖。

本文的主要创新点如下:

- 1) 提出多模态零样本属性解析机制,利用多模态大模型解析空地行人图像,构建属性三元组损失。
- 2) 引入视角解耦架构,有效分离视角相关/无关特征。
- 3) 引入多尺度扩增 Transformer^[20],通过多尺度扩增注意力 (multi-scale dilated attention, MSDA) 与多头自注意力 (multi-head self-attention, MHSA) 模块级联优化计算,降低计算参数量,提升模型运行速度。

1 行人属性解析空地行人检索方法

1.1 本文方法总体架构

本文提出一种基于大模型属性解析的空地行人检索方法,该方法的整体框架如图 1 所示。模型采用双分支结构,分别处理地面监控视角和无人航拍视角的输入图像。首先,基于多模态大模型的属性解析模块对两类视角图像进行细粒度语义属性(如衣着颜色、背包类型等)的联合提取与对齐,并构建属性三元组损失以增强跨视角属性一致性。在此基础上,网络通过分层减法分离单元正交损失函数组成的视角解耦架构,显式分离视角相关特征与视角无关特征,从而抑制视角变异对身份判别的影响。特征提取主干采用多尺度扩增 Transformer 网络,其低层网络通过 MSDA 块捕获局部细节特征,而高层网络利用全局自注意力机制建模长程依赖关系,在计算效率与特征表达能力之间取得平衡。最终,通过融合属性监督、视角解耦约束及多尺度特征表示,模型能够在跨视角、低分辨率的无人机巡检场景下实现鲁棒的行人检索。

1.2 多模态大模型属性解析

针对地空跨视角行人检索中存在的视角差异大、类内变化显著等问题,本文提出基于多模态大模型的属性解析方法,通过编码结构化的属性特征引入语义信息,引导相似语义的行人图像缩小特征间距离。利用 GLM-4V-Flash^[21] 的强大语义理解能力提取结构化属性特征,并结合属性三元组损失增强模型对判别性属性的学习。

首先,给定输入图像 $I \in \mathbf{R}^{H \times W \times 3}$, 采用 GLM-4V-Flash 模型进行细粒度属性解析。该模型通过视觉-语言预训练具备优异的跨模态理解能力,可输出 9 类结构化属性描述,包括性别及上衣、下衣、鞋、配饰的类别与颜色:

$$A = \{\mathbf{a}_{\text{gender}}, \mathbf{a}_{\text{upper_type}}, \dots, \mathbf{a}_{\text{accessory_color}}\} \quad (1)$$

为将离散属性转化为连续的嵌入特征,本文设计了分层属性编码架构。首先,为每个属性类别建立独立的嵌入矩阵 $\mathbf{E}_i \in \mathbf{R}^{d_i \times k_i}$ (k_i 为第 i 类属性的取值数量),将各属性 $\mathbf{a}_i$ 映射为嵌入向量 $\mathbf{e}_i = \mathbf{E}_i \cdot \mathbf{a}_i$ 。随后将全部 9 类属性嵌入拼接为复合向量:

$$\mathbf{e}_{\text{concat}} = \text{Concat}(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_9) \quad (2)$$

并通过全连接层 $\mathbf{W}_f \in \mathbf{R}^{d_a \times \sum d_i}$ 与 ReLU 激活函数生成最终属性特征:

$$\mathbf{f}_{\text{attr}} = \text{ReLU}(\mathbf{W}_f \cdot \mathbf{e}_{\text{concat}}) \quad (3)$$

为进一步增强属性特征的判别能力,设计属性

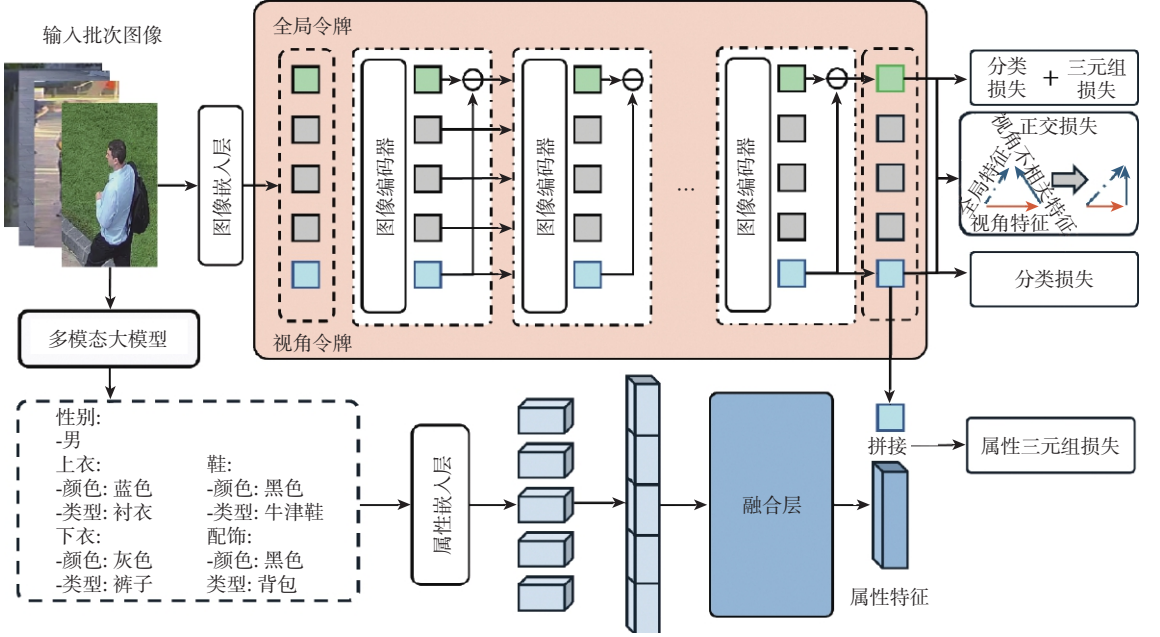


图1 本文方法整体框架

Fig. 1 Overall framework of the proposed method

指导的三元组损失。该方法将全局特征 f_{global} 与属性特征 f_{attr} 拼接为融合特征:

$$\tilde{f} = [f_{\text{global}} || f_{\text{attr}}] \quad (4)$$

在此基础上计算改进的三元组损失:

$$L_{\text{attr}} = \max(0, D(\tilde{f}_a, \tilde{f}_p) - D(\tilde{f}_a, \tilde{f}_n) + \alpha) \quad (5)$$

式中: $\tilde{f}_a$ 表示锚点样本特征; $\tilde{f}_p$ 表示正样本特征; $\tilde{f}_n$ 表示负样本特征; $D(\cdot)$ 采用余弦距离度量; 边界值 α 根据正负样本的属性相似度动态调整:

$$\alpha = \alpha_{\text{base}} + \lambda_{\alpha}(1 - \text{sim}(\mathbf{A}_p, \mathbf{A}_n)) \quad (6)$$

其中: α_{base} 表示基准边界值; $\mathbf{A}_p$ 表示正样本属性; $\mathbf{A}_n$ 表示负样本属性; λ_{α} 表示平衡权重; $\text{sim}(\cdot)$ 计算两样本属性的 Jaccard 相似度; λ 用于控制属性相似度对边界的影响程度。该机制使模型对属性差异小的困难样本对施加更严格的约束, 从而提升特征判别性。

1.3 视角解耦架构

为应对无人机巡检中地面平视视角与高空俯视视角间的剧烈差异对行人特征表达的干扰, 本文引入视角解耦架构, 其核心是通过结构化的特征分离机制显式解耦视角相关特征与视角无关的身份特征。该架构以 Transformer 为基础框架, 融入了 2 个可学习令牌——全局令牌 t_m 用于聚合全局身份语义, 视角令牌 t_v 专用于捕捉拍摄视角特性(如俯视造成的形变、地面视角的姿态变化)。在特征处理过程中, 核心在于分层减法分离机制的设计: 输入令牌序列经多层 Transformer 块逐级处理, 每层首先通过自注意力操作更新包括 t_m 和 t_v 在内的所有

令牌表示; 接着, 执行特征解耦操作, 即从更新后的元令牌表示中显式减去对应的视角令牌表示:

$$t_m^{(j+1)} = t_m^{(j+1)} - t_v^{(j+1)} \quad (7)$$

这种层级递进的减法策略使网络在特征学习过程中持续剥离身份表示中的视角噪声, 使深层 t_m 愈发聚焦于与视角无关的鲁棒行人特征。

为确保模型学习进行正确的解耦计算, 引入正交损失 L_o 作为约束:

$$L_o = \frac{1}{|B|} \sum_{i=1}^{|B|} \frac{|\langle t_m^{(N+1)}(x_i), t_v^{(N+1)}(x_i) \rangle|}{\|t_m^{(N+1)}(x_i)\|_2 \cdot \|t_v^{(N+1)}(x_i)\|_2} \quad (8)$$

式中: B 为一个批次的样本集合; $|B|$ 为获取批次数; x_i 为批次中的第 i 个样本。该损失函数强制最终输出的元令牌(身份特征)与视角令牌(视角特征)在向量空间尽可能正交, 从而消除特征间的残留相关性。在监督信号上, 全局令牌 $t_m^{(N+1)}$ 接受行人身份标签的监督(结合交叉熵损失 L_c^i 与三元组损失 L_t^i 优化判别力), 而视角令牌 $t_v^{(N+1)}$ 则通过交叉熵损失 L_v^i 学习区分视角类别(如无人机/地面)。整体优化目标整合身份识别、视角分类与解耦约束, 并与 1.2 节构建的属性三元组损失共同组成完整的目标函数:

$$L_{\text{total}} = L_{\text{attr}} + L_c^i + L_t^i + \lambda_L (L_v^i + L_o) \quad (9)$$

其中: λ_L 为超参数平衡任务权重; L_{attr} 为基于多模态属性解析的监督项。推理阶段仅使用精炼的元令牌 $t_m^{(N+1)}$ 作为行人检索特征向量。

1.4 多尺度扩增 Transformer

多尺度扩增注意力机制模块由残差结构的深

度可分离卷积、归一化层、堆叠的扩增 Transformer、特征拼接层及最后的归一化与多层感知机(multi layer perceptron, MLP)层构成, 如图 2 所示。针对俯视角下行人特征分散的特性, 扩增 Transformer 在滑动窗口内执行差异化空洞采样: 对于位置 (i, j) 的查询, 仅对以空洞率 r 的稀疏邻域进行键值提取(如 $(r = [1, 2, 3])$ 对应 3×3 至 7×7 感受野), 其坐标集合由 $(i' = i + pr, j' = j + qr)$ 定义 $(p, q \in [-1, 1])$ 。模型利用了浅层注意力矩阵的局部性与稀疏性先验, 在避免全局计算冗余的同时, 通过通道分组的多头并行机制同步捕获行人局部纹理(小感受野)与肢体构型(大感受野)的跨尺度关联——不同头部分配独立空洞率后, 各组输出经拼接与线性投影融合为统一特征表达, 显著提升对高空俯视尺度压缩的鲁棒性。

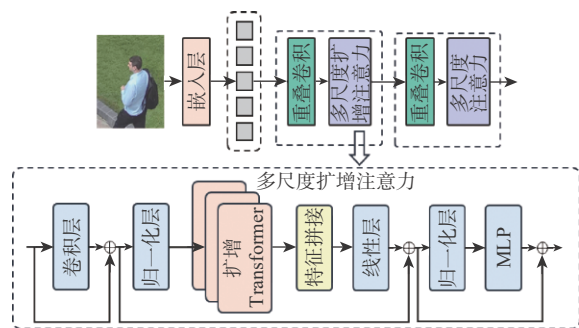


图 2 多尺度扩增注意力
Fig. 2 Multi-scale dilated attention

2 实验与结果分析

2.1 数据集与评估标准

本文在空地行人检索任务中采用两大主流基准进行性能验证:

AG-ReID.v1^[12]数据集由一架在 15~45 m 高度飞行的无人机和大学校园内的地面摄像机拍摄, 包含 388 个行人身份的 21 983 张图像, 每个行人身份带有 15 个软属性。作为领域内首个小规模真实数据集, 重点评估跨视角泛化能力, 其核心协议包括空对地检索(A-G; 无人机→地面监控)和地对空检索(G-A: 地面监控→无人机)2 种定向模式变化场景。

AG-ReID.v2^[14]包含了 1 615 个行人身份的 100 502 张图像, 由无人机、监控探头和可穿戴设备拍摄。具有地面检索(G-G)、空对空检索(A-A)、空对地检索(A-G)、空对穿戴设备检索(A-W)、地对空检索(G-A)、穿戴设备对空检索(W-A)6 种测试协议。本文采用其中的空对地检索(A-G)、地对空检索(G-A)协议进行实验。

CARGO^[15]作为大规模合成数据集, 包括 5 个

地面摄像头、8 个空中摄像头、5 000 个身份及 108 563 张图像。CARGO 涵盖 4 类协议: 全视角混合检索(ALL)、纯地面域内匹配(G-G)、纯无人机域内匹配(A-A)及空-地跨视角泛化性测试(A-G)。

实验遵循 AG-ReID 通用评估框架, 联合采用三重指标全面量化模型性能: Rank-1 反映首候选命中率, 直接衡量跨视角检索精准性; 平均精度均值(mean average precision, mAP)综合评估召回率与查准率的平衡性; 平均逆负惩罚(mean inverse negative penalty, mINP)^[22]重点检验模型对相似干扰项的判别鲁棒性, 三者共同构建覆盖精度、鲁棒性与泛化能力的评测体系。

2.2 实验设置

本文方法在 NVIDIA GeForce RTX 4090 GPU 硬件平台基于 PyTorch 框架实现, 所有输入图像统一预处理为 256 像素×128 像素分辨率。本文方法构建了一行人结构化属性文件, 其中, 以 json 格式记录了多模态大模型解析的行人属性信息。训练模型采用 AdamW 优化器进行参数优化。针对 2 个数据集的训练超参数如下:

对于 AG-ReID.v1 数据集, 设置最大训练轮数为 80, 基准学习率为 0.01, 权重衰减系数为 0.000 5, 批次大小为 64, 预热策略采用初始比例因子为 0.1 配合 1 000 步渐进式学习率调整。

对于 AG-ReID.v2 数据集, 设置最大训练轮数为 200, 基准学习率为 0.001, 权重衰减系数为 0.000 5, 批次大小为 64, 并采用预热学习策略。

对于 CARGO 数据集, 设置最大训练轮数为 120, 基准学习率为 0.008, 权重衰减系数设置为 0.000 1, 批次大小为 128, 预热阶段采用比例因子为 0.01 并扩展至 2 000 步渐进调整, 以适应更大数据规模。

2.3 对比实验

本文方法在 AG-ReID.v1、AG-ReID.v2 和 CARGO 数据集上进行了测试。表 1 展示了本文方法在 AG-ReID.v1 数据集上的测试结果与现有方法的比较。实验结果表明, 本文方法在空对地检索和地对空检索中各指标上都达到了最优效果(见表中加黑数值)。相比 SD-ReID, 本文方法在空对地检索的 Rank-1、mAP、mINP 这 3 种指标上分别提升了 1.19%、0.42%、0.09%; 在地对空检索中分别提升了 0.35%、0.09%、0.31%。实验结果证明了本文方法在空地行人检索中身份有效判别了行人身份特征。

表 2 展示了本文方法在 AG-ReID.v2 数据集上与现有方法的性能比较, 加黑数值为最优结果。本文方法的 Rank-1 和 mAP 这 2 种指标在空对地(A-G)检索中较 V2E 方法提升了 0.15%、0.53%; 在地对空

表 1 AG-ReID.v1 数据集实验结果

Table 1 Experimental results on AG-ReID.v1 dataset

方法	Rank-1/%		mAP/%		mINP/%	
	A-G	G-A	A-G	G-A	A-G	G-A
OSNet ^[23]	72.59	74.22	58.32	60.99		
Bo ^[24]	70.01	71.20	55.47	58.83		
SBS ^[25]	73.54	73.70	59.77	62.27		
VV ^[26-27]	77.22	79.73	67.23	69.83	41.43	42.37
ViT ^[28]	81.28	82.64	72.38	73.35		
TransReID ^[7]	81.80	83.40	73.10	74.60		
FusionReID ^[29]	80.40	82.40	71.40	74.20		
CLIP-ReID ^[30]	79.44	84.20	70.55	73.05		
PCL-CLIP ^[31]	82.16	86.90	73.11	76.28		
Explain ^[13]	81.47	82.85	72.61	73.39		
VDT ^[15]	83.00	84.62	74.06	76.28	50.31	49.51
DTST ^[32]	83.48	84.72	74.51	76.05	49.86	50.04
SD-ReID ^[16]	85.16	85.97	75.40	77.02	51.35	50.42
本文方法	86.35	86.32	75.82	77.11	51.44	50.73

(G-A)检索中较 V2E 方法提升了 0.17%、0.35%。实验结果表明,本文方法在身份较多、场景更多样的 AG-ReID.v2 数据集上也能取得良好效果,具有实际应用价值。

表 3 CARGO 数据集实验结果

Table 3 Experimental results on CARGO dataset

方法	Rank-1/%				mAP/%				mINP/%			
	A-G	ALL	G-G	A-A	A-G	ALL	G-G	A-A	A-G	ALL	G-G	A-A
SBS ^[25]	31.25	50.32	72.31	67.50	29.00	43.09	62.99	49.73	18.71	29.76	48.24	29.32
PCB ^[33]	34.40	51.00	74.10	55.00	30.40	44.50	67.60	44.60	20.10	32.20	55.10	27.00
BoT ^[24]	36.25	54.81	77.68	65.00	32.56	46.49	66.47	49.79	21.46	32.40	51.34	29.82
MGN ^[2]	31.87	54.81	83.93	65.00	33.47	49.08	71.05	52.96	24.64	36.52	55.20	36.78
VV ^[26-27]	31.25	45.83	72.31	67.50	29.00	38.84	62.99	49.73	18.71	39.57	48.24	29.32
AGW ^[22]	43.57	60.26	81.25	67.50	40.90	53.44	71.66	56.48	29.39	40.22	58.09	40.40
ViT ^[28]	43.13	61.54	82.14	80.00	40.11	53.54	71.34	64.47	28.20	39.62	57.55	47.07
VDT ^[15]	45.00	60.58	76.79	82.50	42.08	54.61	71.97	64.67	29.66	41.37	61.42	45.09
DTST ^[32]	50.63	64.42	78.57	80.00	43.39	55.73	72.40	63.31	29.46	41.92	62.10	44.67
SD-ReID ^[16]	53.12	65.06	81.25	82.50	46.44	57.47	74.08	67.70	33.03	44.21	63.10	51.90
本文方法	54.57	66.25	83.46	82.73	47.01	58.83	75.04	68.21	33.52	45.51	63.60	52.24

2.4 消融实验

为验证引入的 3 项改进措施的有效性,在 AG-ReID.v1 数据集上进行了消融实验,实验结果如表 4 所示,最优结果加粗显示。将 TransReID 作为基线模型,逐步加入属性解析模块、视角解耦架构,将 Transformer 层替换为多尺度扩增 Transformer 进行空地行人检索实验。实验结果表明,引入属性解析模块及属性三元组损失后,A-G 检索任务 Rank1、mAP、mINP 分别提升 7.48%、8.37%、5.19%,验证了属性监督对不同视角身份认定的强

表 2 AG-ReID.v2 数据集实验结果

Table 2 Experimental results on AG-ReID.v2 dataset

方法	Rank-1/%		mAP/%	
	A-G	G-A	A-G	G-A
MGN ^[2]	82.09	84.21	70.17	72.41
BoT ^[24]	80.73	79.46	71.49	69.67
SBS ^[25]	81.96	84.10	72.04	73.89
V2E ^[14]	88.77	87.86	80.72	78.51
本文方法	88.92	88.03	81.25	78.86

表 3 展示了本文方法在 CARGO 数据集上的测试结果与现有方法的比较,加黑数值为最优结果。实验结果表明,本文方法在 4 种检索方式中各指标上都达到了最优效果。相比目前的最优方法 SD-ReID,本文方法在空对地检索的 Rank-1、mAP、mINP 这 3 种指标上分别提升了 1.45%、0.57%、0.49%;在混合检索场景中分别提升了 1.19%、1.36%、1.30%;在地对地检索场景中分别提升了 2.21%、0.96%、0.50%;在空对空场景中分别提升了 0.23%、0.51%、0.34%。实验结果表明,本文方法通过解析行人图像属性有效地通过语义信息提升了模型身份判别能力,在不同检索场景中都提升了检索精度。

化作用;视角解耦架构叠加分层减法分离与正交损失模块,A-G 检索 3 种指标较前阶段提升 6.09%、5.23%、4.34%,改善视角差异导致的特征混淆;最后,整合多尺度扩增 Transformer,模型在 A-G 检索中 3 种指标分别达到了 86.35%、75.82%、51.44%,检索效果与前一阶段达到了相当水平,但参数量降低了 11.7%,提升了计算速度降低了算力需求。消融实验表明,本文引入的 3 项改进对空地行人检索的精度及计算速度起到了优化作用。

表 4 消融实验

Table 4 Ablation experiment

方法	Rank-1/%		mAP/%		mINP/%		参数量
	A-G	G-A	A-G	G-A	A-G	G-A	
TransReID	72.63	78.36	62.61	65.37	41.59	42.30	8.050×10^{-7}
+属性	80.11	82.64	70.98	72.22	46.78	47.59	1.904×10^{-8}
+视角解耦	86.20	86.83	76.21	77.38	51.12	50.10	1.957×10^{-8}
+扩增Transformer	86.35	86.32	75.82	77.11	51.44	50.73	1.972×10^{-8}

为验证本文方法的效率优势与改进的实际部署价值, 在 AG-ReID.v1 数据集的测试实验中进行效率参数消融分析, 实验结果如表 5 所示。实验结果表明, 属性三元组损失由于加入了文本编码器, 因此, 模型参数量有比较大的提升, 推理速度有一定下降, 但仍然能保证模型推理的实时性。加入视角解耦架构改进微量增加了模型参数, 几乎未影响模型推理效率。将 Transformer 网络改进为扩增 Transformer 参数量下降至 1.729×10^8 , 实现了参数的精简, 提升了模型的计算效率。

表 5 模型推理效率参数

Table 5 Model inference efficiency parameters

方法	帧率/(帧·s ⁻¹)	推理时间/s	参数量	浮点运算速度/ 10 ⁹ s ⁻¹
TransReID	1 363	7.33×10^{-4}	8.050×10^7	11.13
+属性	758	1.32×10^{-3}	1.904×10^8	14.44
+视角解耦	749	1.33×10^{-3}	1.957×10^8	14.71
+扩增 Transformer	916	1.09×10^{-3}	1.729×10^8	12.04

2.5 可视化结果

为直观展示本文方法的行人检索效果, 在 AG-ReID.v1 数据集上进行了可视化实验, 展示了本文方法与 VDT 方法的检索效果对比及多模态大模型解析行人属性信息, 如图 3 所示。

如图 3(a)所示, 本文方法和 VDT 方法都正确检索出了相同身份的行人图像。VDT 方法检索出的匹配度前 5 的图像都是颜色风格相似的图像, 本文方法在成功检索颜色风格相似图像的基础上还检索出了阴影下的同一行人图像。结果表明, 本文方法的属性解析约束成功将颜色风格变化较大但行人属性相似的行人身份进行了正确关联。

如图 3(b)所示, 本文方法和 VDT 方法在本组检索结果中都出现了检索错误的行人图像。由于地面探头拍摄背景及行人形态与无人机拍摄的差异较大, VDT 方法匹配度前 4 的图像没有检索出身份正确的行人。本文方法由于进行了属性解析能够利用语义信息进行更准确的匹配并且引入多尺度扩增 Transformer, 能够更好地提取小像素图像(如匹配度第 3 的图像)的行人特征。然而, 由于无



图 3 空地行人检索可视化结果

Fig. 3 Visualization results of air-ground pedestrian retrieval

人机拍摄行人图像存在大量小像素行人图像, 多模态大模型对这类图像的属性解析效果较差, 对于具有相似属性的行人存在检索错误的情况。

为检验视角解耦结构架构对特征分类的优化效果, 进行了 t-SNE 特征降维可视化与特征空间热力图可视化分析。

如图 4(a)~图 4(c) 所示, 红圈用于同一区域内 2 种方法的对比, 红圈内视角解耦架构的空地、地空行人检索方法的无人机与地面监控图像特征点相较基线方法距离更近, 表明本文方法特征空间中跨视角差异较小, 模型对于空间变化敏感度低。而基线架构的散点图红蓝散点倾向于分为两类, 如图 4(b) 和图 4(d) 所示, 表明特征受不同视角影响较大。

图 5 展示了 AG-ReID.v1 测试的特征空间热力图结果。图 5(a) 为视角解耦方法测试的特征向量分布, 图 5(b) 为基线架构测试的特征向量分布。如

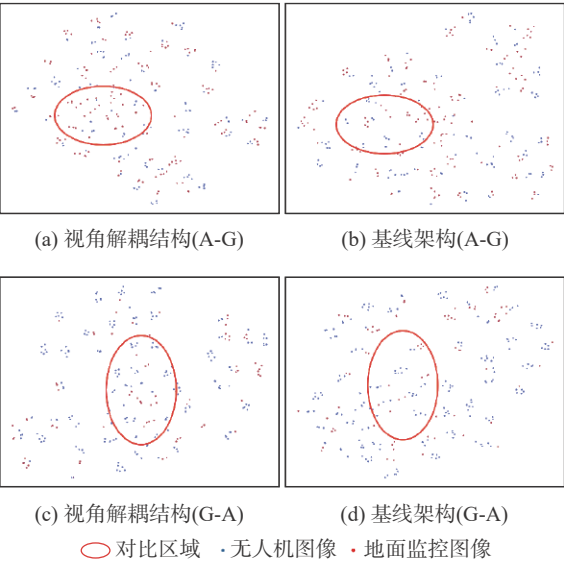


图 4 AG-ReID.v1 测试 t-SNE 可视化结果

Fig. 4 t-SNE visualization results of AG-ReID.v1 testing

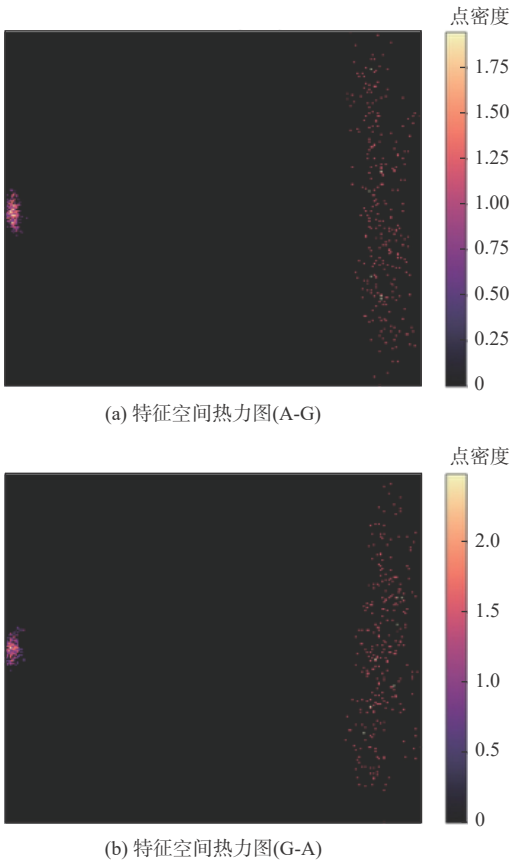


图 5 AG-ReID.v1 测试特征空间热力图

Fig. 5 Feature space heatmap of AG-ReID.v1 testing

可视化结果所示, 视角解耦方法的特征密度中心更重合, 表明跨视角对齐更好; 基线方法的特征分布较稀疏, 表明跨视角表征稳定性较差。

3 局限性

尽管本文方法在构建基于属性解析的地空行

人检索框架时, 创新性地引入了多模态大模型的属性预测能力、视角解耦结构及扩增 Transformer 编码器, 但仍面临一个普遍存在于跨视角行人检索特别是无人机场景的技术挑战: 对低分辨率无人机图像进行可靠属性解析的固有局限性。无人机在高空远距离捕获的图像, 其中包含行人的图像区域分辨率显著低于近处的地面监控视图, 导致行人目标的像素信息稀疏、细小纹理特征高度模糊化。这种信息缺失会干扰属性解析模块的性能, 特别是在预测精细或视觉区分度低的属性类别(如细粒度的衣着、鞋款、颜色)时, 可能引入不可避免的预测偏差或不可置信度。尽管视角解耦机制旨在分离与视角无关的强鉴别特征, 扩增 Transformer 在有限参数下捕捉更优的跨尺度特征表示, 但这些架构上的改进仍然依赖于解析模块所提供的、源自目标像素的初步语义理解。因此, 由低分辨率图像导致的属性解析置信度下降与潜在错误, 可能在模型学习阶段成为偏差源, 影响视角无关特征的表征力, 并在最终的相似性度量中传导为检索错误。这突显了在低空成像条件下提升模型对低质信息鲁棒性的必要性, 也是未来研究需重点攻克的难题之一。

4 结 论

本文针对无人机巡检场景下空地视角差异导致的行人检索性能瓶颈, 提出一种属性语义引导的跨视角行人检索框架。主要结论如下:

- 1) 多模态大模型解析空地行人属性, 通过属性三元组损失有效约束网络匹配属性相似行人, 显著提升俯视场景检索的精度。
- 2) 引入分层减法分离与正交损失约束, 进行解耦式特征建模, 有效剥离视角相关干扰。
- 3) 引入多尺度扩增 Transformer 优化 Transformer 网络, 利用多尺度注意力机制提升小像素图像的表征能力, 在保证模型准确度的同时降低了参数量。

实验验证表明, 本文方法在 AG-ReID.v1 与 CARGO 数据集在空地、地空行人检索的 Rank1、mAP、mINP 指标上均优于当前主流模型, 验证了该方法在无人机巡检场景下进行行人检索的有效性。

参考文献 (References)

[1] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[C]//Proceedings of the Computer Vision-ECCV 2018. Berlin: Springer, 2018: 501-518.

[2] Wang G S, Yuan Y F, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification[C]//

Proceedings of the 26th ACM International Conference on Multimedia. New York: ACM, 2018: 274-282.

[3] Rao Y M, Chen G Y, Lu J W, et al. Counterfactual attention learning for fine-grained visual categorization and re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2022: 1005-1014.

[4] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2016: 1116-1124.

[5] Wei L H, Zhang S L, Gao W, et al. Person transfer GAN to bridge domain gap for person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 79-88.

[6] 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. Berlin: Springer, 2017: 6000-6010.

[7] He S T, Luo H, Wang P C, et al. TransReID: Transformer-based object re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2022: 14993-15002.

[8] Zhang G W, Zhang P P, Qi J Q, et al. HAT: hierarchical aggregation Transformers for person re-identification[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 516-525.

[9] Chen Y, Xia S X, Zhao J Q, et al. ResT-ReID: Transformer block-based residual learning for person re-identification[J]. Pattern Recognition Letters, 2022, 157: 90-96.

[10] Zhu H W, Ke W J, Li D, et al. Dual cross-attention learning for fine-grained visual categorization and object re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 4682-4692.

[11] Zhu K, Guo H Y, Zhang S L, et al. AAformer: auto-aligned transformer for person re-identification[J].IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(12): 17307-17317.

[12] Schumann A, Metzler J. Person re-identification across aerial and ground-based cameras by deep feature fusion[J]. Automatic Target Recognition XXVII, 2017, 10202: 102020A.

[13] Nguyen H, Nguyen K, Sridharan S, et al. Aerial-ground person re-ID[C]//Proceedings of the IEEE International Conference on Multimedia and Expo. Piscataway: IEEE Press, 2023: 2585-2590.

[14] Nguyen H, Nguyen K, Sridharan S, et al. AG-ReID. v2: bridging aerial and ground views for person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 2896-2908.

[15] Zhang Q, Wang L, Patel V M, et al. View-decoupled transformer for person re-identification under aerial-ground camera network [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 22000-22009.

[16] Wang Y H, Hu X, Wang L X, et al. SD-ReID: view-aware stable diffusion for aerial-ground person re-identification[EB/OL]. (2025-05-14)[2025-10-10]. <https://arxiv.org/abs/2504.09549>.

[17] Zheng A H, Pan P, Li H C, et al. Progressive attribute embedding for accurate cross-modality person re-ID[C]//Proceedings of the 30th ACM International Conference on Multimedia. New York: ACM, 2022: 4309-4317.

[18] Tan W T, Ding C X, Jiang J Y, et al. Harnessing the power of MLLMs for transferable text-to-image person ReID[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 17127-17137.

[19] Yang S Y, Zhou Y N, Zheng Z D, et al. Towards unified text-based person retrieval: a large-scale multi-attribute and language search benchmark[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM, 2023: 4492-4501.

[20] Jiao J Y, Tang Y M, Lin K Y, et al. DilateFormer: multi-scale dilated Transformer for visual recognition[J]. IEEE Transactions on Multimedia, 2023, 25: 8906-8919.

[21] Zeng A H, Xu B, Wang B W, et al. ChatGLM: a family of large language models from GLM-130B to GLM-4 all tools[EB/OL]. (2024-07-30)[2025-10-10]. <https://arxiv.org/abs/2406.12793>.

[22] Ye M, Shen J B, Lin G J, et al. Deep learning for person re-identification: a survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 2872-2893.

[23] Zhou K Y, Yang Y X, Cavallaro A, et al. Learning generalisable omni-scale representations for person re-identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 2021: 3069237.

[24] Luo H, Gu Y Z, Liao X Y, et al. Bag of tricks and a strong baseline for deep person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2019: 1487-1495.

[25] He L X, Liao X Y, Liu W, et al. FastReID: a pytorch toolbox for general instance re-identification[C]// Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM, 2023: 9664-9667.

[26] Kuma R, Weill E, Aghdasi F, et al. Vehicle re-identification: an efficient baseline using triplet embedding[C]// Proceedings of the International Joint Conference on Neural Networks. Piscataway: IEEE Press, 2019: 1-9.

[27] Kumar R, Weill E, Aghdasi F, et al. A strong and efficient baseline for vehicle re-identification using deep triplet embedding[J]. Journal of Artificial Intelligence and Soft Computing Research, 2020, 10(1): 27-45.

[28] Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. (2021-06-03)[2025-10-10]. <https://arxiv.org/abs/2010.11929>.

[29] Wang Y H, Zhang P P, Liu X H, et al. Unity is strength: unifying convolutional and transformer features for better person re-identification[J]. IEEE Transactions on Intelligent Transportation Systems, 2025, 26(3): 3713-3723.

[30] Li S Y, Sun L, Li Q L. CLIP-ReID: exploiting vision-language model for image re-identification without concrete text labels[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(1): 1405-1413.

[31] Li J C, Gong X J. Prototypical contrastive learning-based CLIP fine-tuning for object re-identification[EB/OL]. (2025-01-14)[2025-10-10]. <https://arxiv.org/abs/2310.17218>.

[32] Wang Y H, Pishgar M. Dynamic token selection for aerial-ground person re-identification[EB/OL]. (2024-12-25)[2025-10-10]. <https://arxiv.org/abs/2412.15411>.

arxiv.org/abs/2412.00433.

[33] Sun Y F, Zheng L, Li Y L, et al. Learning part-based convolutional

features for person re-identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(3): 902-917.

Large model attribute parsing-based aerial-ground pedestrian retrieval method

TAN Quange¹, WANG Rong¹, LIAO Mancheng², LI Xin^{1,*}

- (1. College of Information Network Security, People’s Public Security University of China, Beijing 100038, China;
- 2. College of Criminal Investigation, People’s Public Security University of China, Beijing 100038, China)

Abstract: Synergistic analysis of drone-captured imagery and fixed surveillance video enables continuous tracking of targets in inspection tasks, achieving cross-view person re-identification in areas where surveillance probes are sparsely distributed. The transferability of image retrieval algorithms created for conventional surveillance situations to aerial-ground cross-view environments is limited by the notable discrepancy between the horizontal view of ground surveillance and the bird’s-eye view of drones. Existing aerial-ground pedestrian retrieval methods primarily focus on mitigating the appearance discrepancies caused by cross-view variations, while the mining and analysis of person attribute characteristics remain insufficiently explored. To address these issues, this paper proposes a aerial-ground pedestrian retrieval method based on large model attribute parsing. A person parsing module is constructed based on a multi-modal large model to generate fine-grained semantic attributes, and an attribute triplet loss is designed to achieve cross-view semantic alignment. A view decoupling architecture is introduced to separate view-specific features through hierarchical subtraction, with an orthogonal loss applied to constrain feature independence. A multi-scale dilated Transformer is incorporated, combining multi-scale dilated attention with global self-attention to optimize the balance between computational complexity and receptive field, thereby reducing model parameters. The effectiveness of the methodology is confirmed by experiments on the AG-ReID.v1, AG-ReID.v2, and CARGO datasets, which show that the suggested strategy successfully increases performance on Rank-1, mAP, and mINP metrics in aerial-ground pedestrian retrieval tasks.

Keywords: UAV; pedestrian retrieval; aerial-ground imagery; attribute parsing; multimodal large language models

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0833

位置感知与局部预交互的空-地行人重识别方法

毕艺瀚, 李冲, 王蓉*

(中国人民公安大学 信息网络安全学院, 北京 100038)

摘 要: 针对现有空-地行人重识别方法中存在的行人特征判别性较差、模型泛化能力不足等问题, 提出一种位置感知与局部预交互的空-地行人重识别方法。提出位置偏置自注意力机制, 通过将位置信息融入注意力分数, 引导模型聚焦输入序列中关键位置的令牌, 增强模型对图块之间空间位置关系的感知能力, 提升行人特征的判别性与鲁棒性; 设计提示-局部预交互模块, 通过一次较低成本的预交互, 建立提示语义与局部特征之间的细粒度关联, 增强提示向量对当前视角图像局部具体细节的感知能力, 提升模型的局部细节重建能力; 引入标签平滑正则化训练策略, 通过将原始 One-Hot 编码的硬标签转换为软标签, 使模型学习更平滑、泛化性更强的特征表示, 缓解模型过拟合问题, 提升模型的性能与泛化能力。在空-地行人重识别公开数据集 LAGPeR 上进行大量实验, 结果表明, 所提方法能够提升重识别性能, 具有一定有效性。

关 键 词: 空-地行人重识别; Transformer; 自注意力机制; 局部特征预交互; 标签平滑正则化
中图分类号: TP391.4; V279

文献标志码: A **文章编号:** 1001-5965(2026)09-3117-08

行人重识别(person re-identification, ReID)旨在利用计算机视觉技术, 实现跨视域特定行人的准确匹配与有效追踪^[1], 在智能监控、安全防范等领域得到了广泛应用^[2]。随着深度学习的发展, 现有研究在传统行人重识别领域取得了显著进展, 一定程度上解决了光照变化^[3]、目标遮挡^[4-5]与低图像分辨率差异^[6-7]等问题。然而, 这些研究主要基于单一平台摄像头网络采集的数据集, 通常由固定的地面摄像头或闭路电视系统组成。这一特点使得这些方法在实际场景中的性能高度依赖于地面摄像头的部署密度与覆盖范围。在城市等人流密集的区域, 密集的摄像头覆盖有助于实现有效的检索; 而在市郊、农田、森林等人流稀疏的区域, 稀疏的摄像头部署会显著降低追踪效果, 可能导致目标踪迹丢失。

随着无人机(unmanned aerial vehicle, UAV)的快速发展, 在无人机上部署摄像头为提升摄像头的覆盖密度提供了一种灵活的解决方案。同时, 这种方法还能增强智能监控的态势感知能力^[8]。因此,

研究在监控网络中同时结合地面摄像头与无人机摄像头的空-地行人重识别(aerial-ground person re-identification, AGPreID), 更符合实际监控场景的需求。

区别于传统行人重识别, 空-地行人重识别的核心挑战在于跨摄像头视角变化导致的显著视觉差异, 这些视觉差异会阻碍鲁棒性行人特征的提取, 降低行人图像匹配的准确性。目前, 一些研究尝试解决该问题。Nguyen等^[9]第1次从深度学习的视角对空-地行人重识别任务进行研究, 提出了一种利用软属性引导模型训练的空-地行人重识别方法, 通过利用行人性别、上衣颜色、背包类型等语义属性信息, 使模型有效地学习更具身份判别性的行人特征, 增强模型对空-地场景中行人外观显著变化的适应能力; 同时, 构建了空-地行人重识别公开数据集 AG-ReID, 为空-地行人重识别研究提供了重要的基准。贝俊仁等^[10]针对空、地视角差异导致的行人外观显著变化与不对齐问题, 提出一种基于隐式解码对齐的空-地行人重识别方法, 通过

收稿日期: 2025-12-02; 录用日期: 2025-12-26; 网络出版时间: 2026-03-03 11:32

网络出版地址: link.cnki.net/urlid/11.2625.V.20260303.0942.001

基金项目: 中央高校基本科研业务费专项资金(2024JKF12)

*通信作者. E-mail: dbdxwangrong@163.com

引用格式: 毕艺瀚, 李冲, 王蓉. 位置感知与局部预交互的空-地行人重识别方法[J]. 北京航空航天大学学报, 2026, 52(9): 3117-3124.

Bi Y H, Li C, Wang R. Aerial-ground person re-identification method via position-aware and local pre-interaction[J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52(9): 3117-3124 (in Chinese).

深度特征学习自动解码并对齐地面摄像头视角与无人机摄像头视角下行人图像的关键语义信息,无需额外的姿态或部件标注即可实现跨视角的鲁棒性行人特征匹配,有效提升了模型在空-地场景中的特征判别能力。Wang 等^[11]利用扩散模型的生成能力增强模型对视角变化的鲁棒性,提出一种基于视角感知稳定扩散的空-地行人重识别方法,通过引入视角感知的条件控制机制,并利用文本或视角标签引导稳定扩散模型生成更具身份一致性的跨视角行人图像,弥补真实数据中视角样本的不足,并学习视角不变的特征表示,有效提升了模型在跨视角场景中的判别能力与泛化能力。Mei 等^[12]针对无人机监控场景中标注数据稀缺的问题,提出了一种基于无监督学习的空-地行人重识别框架,通过引入邻域感知协作学习与渐进图匹配策略,以无监督学习的方式挖掘跨视角行人图像之间的隐含关联;同时,通过设计协作跨模态关联学习模块缓解特征差异、实现模态之间的软对齐,并利用最小行人距离转换策略量化地面摄像头视角与无人机摄像头视角之间的最优行人相似度,在空-地行人重识别公开数据集上取得了较为先进的性能。

目前,空-地行人重识别的研究仍处于起步阶段。上述方法虽然在一定程度上解决了地面摄像头视角与无人机摄像头视角下行人图像的匹配难题,但是仍存在行人特征判别性较差、模型泛化能力不足等问题。

针对上述问题,本文提出一种位置感知与局部预交互的空-地行人重识别方法。该方法使用视觉 Transformer^[13](vision Transformer, ViT)作为主干网络,首先,提出位置偏置自注意力机制,提升行人特征的判别性与鲁棒性;其次,设计提示-局部预交互模块,提升模型的局部细节重建能力;最后,引入标签平滑正则化训练策略,缓解模型过拟合问题,提升模型的泛化能力。本文的主要创新点如下:

- 1) 提出位置偏置自注意力机制(positionally biased self-attention mechanism, PBSAM),通过在标准自注意力机制的查询矩阵上添加一个可学习的偏置向量,将位置信息融入注意力分数,引导模型聚焦输入序列中关键位置的令牌,增强模型对图块之间空间位置关系的感知能力,提升行人特征的判别性与鲁棒性。
- 2) 设计提示-局部预交互模块(prompt-local pre-interaction module, PLPIM),通过将提示向量与局部特征进行快速对齐,建立提示语义与局部特征之间的细粒度关联,增强提示向量对当前视角图像局部具体细节的感知能力,提升模型的局部细节重建能力。
- 3) 将标签平滑正则化^[14](label smoothing regularization, LSR)训练策略引入交叉熵损失,通过引入

一个平滑因子,将原始 One-Hot 编码的硬标签转换为软标签,使模型学习更平滑、泛化性更强的特征表示,减少模型对训练数据的过度自信、缓解模型过拟合问题,提升模型的性能与泛化能力。

1 本文方法

如图 1 所示,本文提出的位置感知与局部预交互的空-地行人重识别方法采用编码器-解码器的 Transformer 框架。编码器采用视觉解耦的 Transformer^[15](view-decoupled Transformer, VDT)网络。相较于传统的 ViT 网络,本文方法通过引入视角令牌,并在每层对类别令牌进行分层解耦,有效分离类别令牌中的视角相关特征与视角不变特征,同时从输入中提取局部特征。解码器由提示重校准模块^[16](prompt re-calibration module, PRM)、提示-局部预交互模块与局部特征优化模块^[16](local feature refinement module, LFRM)组成。其中,PRM 基于当前视角信息自适应生成并重新校准不同视角的提示向量;PLPIM 将重新校准的提示向量与当前视角的局部特征进行预交互;LFRM 利用 PLPIM 预交互的提示向量对局部特征进行解码。

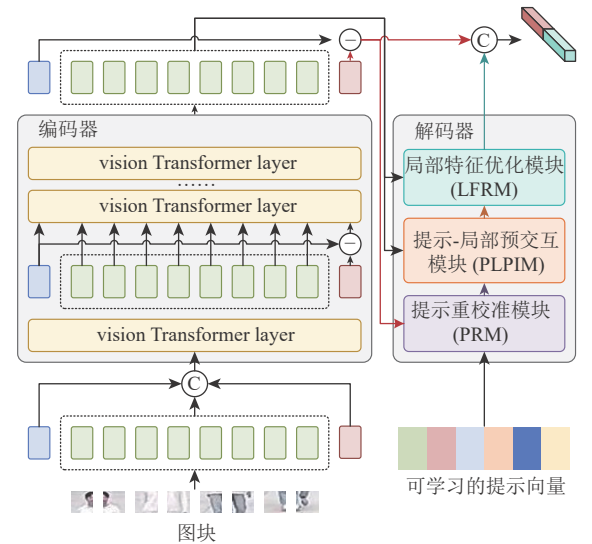


图1 本文方法框架

Fig. 1 Architecture of the proposed method

1.1 位置偏置自注意力机制

为提取更具判别性与鲁棒性的行人特征,本文设计了位置偏置自注意力机制,如图 2 所示,对 Transformer 编码器中的自注意力机制进行了关键改进。在标准自注意力机制中,首先,输入经过 QKV 线性层生成查询矩阵、键矩阵与值矩阵;其次,通过查询矩阵与键矩阵的转置的点积运算,计算注意力分数;然后,利用 softmax 函数将注意力分数归一化为注意力权重;最后,将注意力权重与值

矩阵相乘, 再经过 Projection 线性层投影得到最后的输出:

$$\text{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{softmax}\left(\boldsymbol{Q}\boldsymbol{K}^{\text{T}} / \sqrt{d_{\text{k}}}\right) \boldsymbol{V} \quad (1)$$

式中: $\boldsymbol{Q}$ 为查询矩阵; $\boldsymbol{K}$ 为键矩阵; $\boldsymbol{V}$ 为值矩阵; d_{k} 为键的维度。

如图 2 所示, 位置偏置自注意力机制在标准自注意力机制中引入了位置偏置。如式 (2) 所示, 作为一种高效的位置感知机制, 位置偏置通过在查询矩阵上添加一个可学习的偏置向量, 将位置信息融入注意力分数, 引导模型聚焦输入序列中关键位置的令牌, 增强模型对图块之间空间位置关系的感知能力, 提升行人特征的判别性与鲁棒性。

$$\tilde{\boldsymbol{Q}} = \boldsymbol{Q} + \boldsymbol{P} \quad (2)$$

式中: $\tilde{\boldsymbol{Q}}$ 为添加位置偏置向量后的查询矩阵; $\boldsymbol{P}$ 为位置偏置。

由式 (2) 可得引入位置偏置后的自注意力机制:

$$\text{Attention}(\tilde{\boldsymbol{Q}}, \boldsymbol{K}, \boldsymbol{V}) = \text{softmax}\left((\boldsymbol{Q} + \boldsymbol{P})\boldsymbol{K}^{\text{T}} / \sqrt{d_{\text{k}}}\right) \boldsymbol{V} \quad (3)$$

将引入位置偏置后自注意力机制的注意力分数拆分为两项:

$$(\boldsymbol{Q} + \boldsymbol{P})\boldsymbol{K}^{\text{T}} / \sqrt{d_{\text{k}}} = \boldsymbol{Q}\boldsymbol{K}^{\text{T}} / \sqrt{d_{\text{k}}} + \boldsymbol{P}\boldsymbol{K}^{\text{T}} / \sqrt{d_{\text{k}}} \quad (4)$$

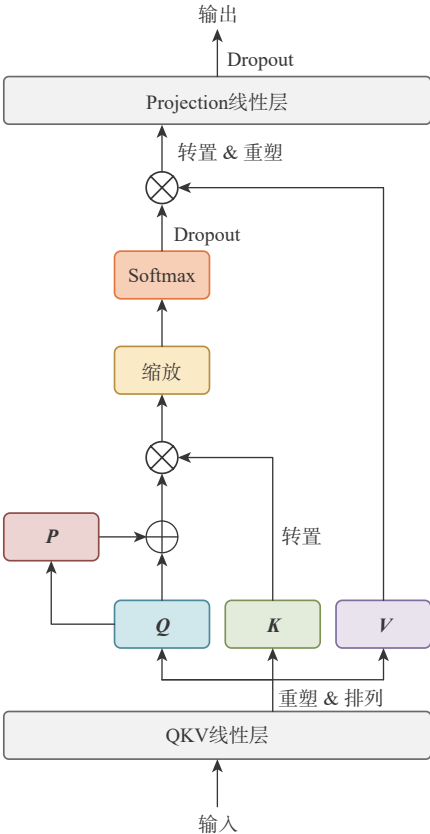


图 2 位置偏置自注意力机制结构

Fig. 2 Architecture of the positionally biased self-attention mechanism

式中: 第 1 项为标准自注意力机制的注意力分数, 即基于内容的寻址; 第 2 项为位置偏置引入的位置相关性分数, 即基于位置的寻址。在第 2 项中, 不考虑缩放因子 $1/\sqrt{d_{\text{k}}}$ 的约束, 通过位置偏置与键矩阵的转置的点积运算, 得到位置相关性分数矩阵:

$$\boldsymbol{S}_{\text{pos}} = \boldsymbol{P}\boldsymbol{K}^{\text{T}} \quad (5)$$

使用 $\boldsymbol{S}_{\text{pos},ij}$ 衡量第 i 个查询向量与第 j 个键向量的位置相关性后, 在模型的反向传播阶段, 损失函数 L 对位置偏置的梯度计算式为

$$\frac{\partial L}{\partial \boldsymbol{P}} = \sum_{i,j} \frac{\partial L}{\partial A_{ij}} \cdot \frac{\partial A_{ij}}{\partial \boldsymbol{S}_{\text{pos},ij}} \cdot \frac{\partial \boldsymbol{S}_{\text{pos},ij}}{\partial \boldsymbol{P}} \quad (6)$$

式中: A_{ij} 为相应的注意力权重。可以看出, 随着位置偏置逐批次更新, 模型动态优化位置相关性分数, 引导注意力权重聚焦行人判别性区域, 从而最小化损失。

1.2 提示-局部预交互模块

为提升模型的局部细节重建能力, 本文设计了提示-局部预交互模块, 如图 3 所示。PLPIM 通过将重新校准的提示向量与当前视角的局部特征进行快速对齐, 使提示向量在输入 LFRM 之前与当前视角的局部特征建立细粒度关联, 使 LFRM 能够专注于更精细的特征融合与解码, 优化解码器的训练动态与计算效率, 提升模型重建局部细节的质量与对视角变化的鲁棒性。

PRM 利用全局视角信息与跨视角差分信息对可学习的提示向量进行自适应重新校准, 导致提示向量缺乏对当前视角图像局部具体细节的直接感知能力。在 PLPIM 中, 首先, 将提示向量投影为查询矩阵、位置增强的局部特征投影为键矩阵、原始局

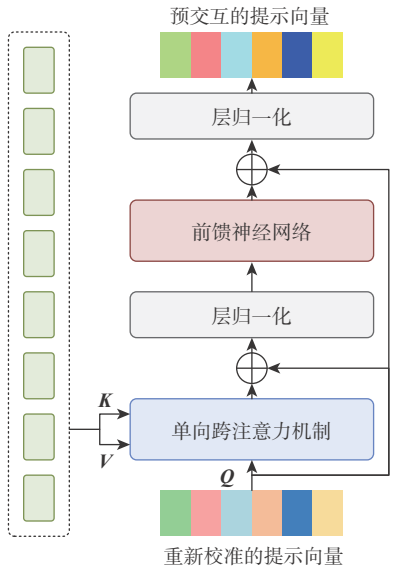


图 3 提示-局部预交互模块结构

Fig. 3 Architecture of the prompt-local pre-interaction module

部特征投影为值矩阵,使用单向跨注意力机制建立提示向量与当前视角局部特征之间的细粒度关联:

Attn($\mathbf{P}, \mathbf{X}$) =

$$\text{softmax} \left\{ \left(\mathbf{P} \mathbf{W}_Q \right) \left[\left(\mathbf{X} + \mathbf{P}_E \right) \mathbf{W}_K \right]^T / \sqrt{d_k} \right\} \left(\mathbf{X} \mathbf{W}_V \right) \mathbf{W}_O \quad (7)$$

式中: $\mathbf{P}$ 为提示向量; $\mathbf{X}$ 为局部特征; $\mathbf{P}_E$ 为位置编码; $\mathbf{W}_Q$ 为查询投影矩阵; $\mathbf{W}_K$ 为键投影矩阵; $\mathbf{W}_V$ 为值投影矩阵; $\mathbf{W}_O$ 为输出投影矩阵。

其次,将跨注意力机制的输出与输入 PLPIM 的提示向量进行残差连接,并使用层归一化进行处理;然后,利用前馈神经网络(feed-forward neural network, FFN)对归一化后的提示向量进行非线性增强:

$$\text{FFN}(\mathbf{x}) = \mathbf{W}_2 \cdot \text{GELU}(\mathbf{W}_1 \cdot \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2 \quad (8)$$

式中: $\mathbf{x}$ 为输入 FFN 的提示向量; $\mathbf{W}_1$ 为 FFN 中第 1 个线性层的权重矩阵; $\mathbf{b}_1$ 为 FFN 中第 1 个线性层的偏置向量; $\mathbf{W}_2$ 为 FFN 中第 2 个线性层的权重矩阵; $\mathbf{b}_2$ 为 FFN 中第 2 个线性层的偏置向量。

最后,将 FFN 的输出与输入 PLPIM 的提示向量进行残差连接,再经过层归一化处理得到既保留原始语义,又融合当前视角图像局部具体细节的提示向量,如图 3 所示。

PLPIM 通过一次较低成本的预交互,建立提示语义与局部特征之间的关联,增强提示向量对当前视角图像局部具体细节的感知能力,提升模型的局部细节重建能力。

1.3 标签平滑正则化

为缓解模型过拟合问题,提升模型的泛化能力,本文将标签平滑正则化引入交叉熵损失。标签平滑正则化是分类任务中一种广泛使用的训练策略,通过将原始 One-Hot 编码的硬标签转换为软标签,使模型学习更平滑、泛化性更强的特征表示,减少模型对训练数据的过度自信、缓解模型过拟合问题,提升模型的性能与泛化能力。

1.3.1 交叉熵损失

作为图像分类任务中一种广泛使用的损失函数,交叉熵损失通过度量模型的预测结果与真实数据之间的相似程度评估模型的预测性能。交叉熵损失的计算式为

$$L_{\text{cross-entropy}} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^M y_{i,k} \ln p_{i,k} \quad (9)$$

式中: M 和 N 分别为类别数量与样本数量; $p_{i,k}$ 为第 i 个样本预测为类别 k 的概率; $y_{i,k}$ 为第 i 个样本在第 k 个类别上的真实标签。

1.3.2 引入标签平滑正则化的交叉熵损失

标准交叉熵损失约束模型对真实类别预测出

趋近于 1 的概率,这种训练目标容易使模型对预测结果过度自信,并使模型学习的特征表示过于尖锐,导致模型的泛化能力不足。

标签平滑正则化通过引入一个平滑因子,对真实类别的概率进行重新分配,即在真实类别上分配略小于 1 的概率,同时在其余类别上均匀分配少量概率,引导模型在训练过程中关注样本的整体分布,而非仅仅依赖严格的硬标签,使模型在类别之间学习更具鲁棒性的特征表示,缓解模型过拟合问题,提升模型的泛化能力。引入标签平滑正则化的交叉熵损失计算式为

$$L_{\text{cross-entropy}}^{\text{LSR}} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^M \left[(1-\varepsilon) y_{i,k} + \frac{\varepsilon}{M} \right] \ln p_{i,k} \quad (10)$$

式中: ε 为平滑因子。

2 实验与分析

2.1 数据集与评价指标

2.1.1 数据集

在空-地行人重识别公开数据集 LAGPeR^[16]上对本文方法进行实验验证。LAGPeR 数据集由 14 个地面摄像头与 7 个无人机摄像头采集得到,包含 4 231 个行人的 63 841 张图像。其中,训练集包含 2 708 个行人的 40 770 张图像;测试集包含 1 523 个行人,选择 3 046 张图像作为待查询图像,其余图像作为查询库图像。同时,构建了一个全面的空-地行人重识别任务设置,包括空对地(aerial to ground, A→G)、地对空(ground to aerial, G→A)与地对全(ground to aerial+ground, G→A+G)3 种实验设置。LAGPeR 数据集的基本情况如表 1 所示。

表 1 LAGPeR 数据集情况介绍
Table 1 Introduction to LAGPeR dataset

实验设置	子集	视角	摄像头个数	行人个数	图像张数
A→G	Train	A+G	12	2 708	40 770
	Query	A	3	1 523	3 046
	Gallery	G	6	1 523	15 533
G→A	Query	G	6	1 523	3 046
	Gallery	A	3	1 523	7 717
G→A+G	Query	G	6	1 523	3 046
	Gallery	A+G	9	1 523	20 204

2.1.2 评价指标

使用 CMC 曲线^[17]、mAP^[18]与 mINP^[19]作为模型性能的评价指标,对本文方法进行实验验证。其中,CMC 曲线通过计算在不同排序的匹配结果中,对待查询的目标行人进行匹配的准确率,较为直观地反映模型的性能;mAP 通过计算在所有查询的匹配结果中,每次对待查询的目标行人进行匹配的平均准确率,较为全面地反映模型的性能;mINP 通过

计算在所有查询的匹配结果中, 对每个待查询的目标行人进行所有正确匹配需克服的平均负样本难度, 反映模型匹配最困难样本的效率。

2.2 实验环境与参数设置

在 GEFORCE RTX 4090 GPU 上使用 PyTorch 1.4.0 深度学习框架, 对本文方法进行实验验证。模型使用 ViT 作为主干网络, 并使用在 ImageNet^[20] 上预训练的参数。模型训练阶段, 输入图像的尺寸设置为 256 像素×128 像素, 总周期数设置为 120, 批量大小设置为 64, 学习率设置为 8×10^{-3} , 平滑因子设置为 0.1。使用 SGD 优化器^[21] 对模型参数进行更新, 权重衰减参数设置为 1×10^{-4} 。

2.3 实验结果与分析

2.3.1 对比实验

为验证本文方法的先进性, 在 LAGPeR 数据集

上进行对比实验, 结果如表 2 所示。表中加粗数据为相应实验设置下的最优结果。可以看出, 在 A→G 实验设置下, 模型的 Rank-1 准确率(在所有测试查询中, 模型返回的相似度最高(排名第 1)的那张恰好是正确目标的概率)达到了 41.99%、mAP 达到了 30.52%、mINP 达到了 12.32%, 均取得了最优结果; 在 G→A 实验设置下, 模型的 Rank-1 准确率达到了 34.93%、mAP 达到了 33.23%、mINP 达到了 24.02%, 均取得了最优结果; 在 G→A+G 实验设置下, 模型的 Rank-1 准确率达到了 22.23%、mAP 达到了 18.17%, 取得了与最优结果接近的次优结果, mINP 达到了 7.24%, 取得了最优结果。综合来看, 本文方法在 LAGPeR 数据集的 3 种实验设置下表现最优, 表明本文方法能够提升空-地行人重识别性能, 具有一定先进性。

表 2 在 LAGPeR 数据集上的对比实验结果
Table 2 Comparative experimental results on LAGPeR dataset %

方法	主干网络	Rank-1			mAP			mINP		
		A→G	G→A	G→A+G	A→G	G→A	G→A+G	A→G	G→A	G→A+G
ViT ^[13]	ViT	38.67	32.04	18.88	27.25	30.69	15.31			
TransReID ^[22]	ViT	38.80	33.00	22.90	28.80	32.10	18.80			
CLIP-ReID ^[23]	CLIP	24.40	21.30	12.30	17.60	20.80	10.20			
MIP ^[24]	ViT	39.30	33.90	21.00	29.30	32.60	17.30			
AG-ReID ^[9]	ViT	40.48	32.96	22.03	28.89	31.91	17.89			
VDI ^[15]	ViT	40.15	33.55	19.50	28.97	31.98	16.45			
SeCap ^[16]	ViT	39.89	33.68	21.04	28.67	32.00	16.80	11.16	23.04	6.59
本文方法	ViT	41.99	34.93	22.23	30.52	33.23	18.17	12.32	24.02	7.24

2.3.2 消融实验

为验证 3 个模块的有效性, 在 LAGPeR 数据集上进行消融实验, 结果如表 3 所示。可以看出, 使用位置偏置自注意力机制后, 在 A→G 实验设置下, 模型的 Rank-1 准确率提升了 0.39%、mAP 提升了 0.53%、mINP 提升了 0.28%; 在 G→A 实验设置下, 模型的 Rank-1 准确率降低了 0.39%、mAP 提升了 0.12%、mINP 提升了 0.14%; 在 G→A+G 实验设置下, 模型的 Rank-1 准确率提升了 0.53%、mAP 提升了 0.62%、mINP 提升了 0.20%, 表明位置偏置自注意力机制能够提升行人特征的判别性与鲁棒性。加入提示-局部预交互模块后, 在 A→G 实验设置下, 模型的 Rank-1 准确率提升了 1.38%、mAP 提升了 1.05%、mINP 提升了 0.72%; 在 G→A 实验设置下, 模型的 Rank-1 准确率提升了 1.38%、mAP 提升了 0.69%、mINP 提升了 0.31%; 在 G→A+G 实验设置下, 模型的 Rank-1 准确率提升了 0.92%、mAP 提升了 0.49%、mINP 提升了 0.03%, 表明提示-局部预交互模块能够提升模型的局部细节重建能力。引入标签平滑正则化训练策略后, 在 A→G 实验设

置下, 模型的 Rank-1 准确率降低了 0.10%、mAP 提升了 0.29%、mINP 降低了 0.08%; 在 G→A 实验设置下, 模型的 Rank-1 准确率提升了 0.56%、mAP 提升了 0.34%、mINP 提升了 0.35%; 在 G→A+G 实验设置下, 模型的 Rank-1 准确率提升了 0.27%、mAP 提升了 0.42%、mINP 提升了 0.43%, 表明标签平滑正则化训练策略能够缓解模型过拟合问题, 提升模型的泛化能力。同时使用位置偏置自注意力机制、加入提示-局部预交互模块、引入标签平滑正则化训练策略后, 在 A→G 实验设置下, 模型的 Rank-1 准确率提升了 2.10%、mAP 提升了 1.85%、mINP 提升了 1.16%, 均取得了最优结果; 在 G→A 实验设置下, 模型的 Rank-1 准确率提升了 1.25%, 取得了与最优结果接近的次优结果, mAP 提升了 1.23%、mINP 提升了 0.98%, 取得了最优结果; 在 G→A+G 实验设置下, 模型的 Rank-1 准确率提升了 1.19%、mAP 提升了 1.37%、mINP 提升了 0.65%, 均取得了最优结果。综合来看, 本文方法在 LAGPeR 数据集的 3 种实验设置下表现最优(见表中加粗数据), 表明 3 个模块协同工作能够提升空-地行人重识别性能,

表 3 在 LAGPeR 数据集上的消融实验结果

Table 3 Results of ablation study on LAGPeR dataset

PBSAM	PLPIM	LSR	Rank-1			mAP			mINP			%
			A→G	G→A	G→A+G	A→G	G→A	G→A+G	A→G	G→A	G→A+G	
√	√	√	39.89	33.68	21.04	28.67	32.00	16.80	11.16	23.04	6.59	
			40.28	33.29	21.57	29.20	32.12	17.42	11.44	23.18	6.79	
			41.27	35.06	21.96	29.72	32.69	17.29	11.88	23.35	6.62	
			39.79	34.24	21.31	28.96	32.34	17.22	11.08	23.39	7.02	
√	√		41.92	34.50	21.50	29.74	33.03	17.40	11.48	23.93	6.80	
√	√	√	41.99	34.93	22.23	30.52	33.23	18.17	12.32	24.02	7.24	

具有一定有效性。

2.3.3 可视化分析

本文方法在 LAGPeR 数据集上的 CMC 曲线如

图 4 所示。可以看出,同时使用 PBSAM、PLPIM、LSR 训练策略后,本文方法在 LAGPeR 数据集的 3 种实验设置下表现最优。

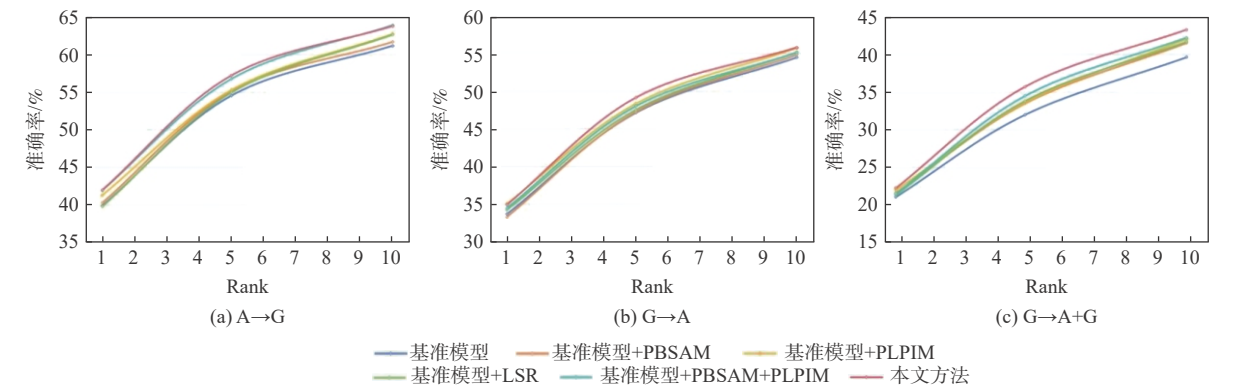


图 4 在 LAGPeR 数据集 3 种实验设置下的 CMC 曲线

Fig. 4 CMC curves under 3 experimental settings of LAGPeR dataset

本文方法在 LAGPeR 数据集上的可视化结果如图 5 所示。其中,Query 表示待查询的目标行人

图像,1~5 表示使用本文方法得到的前 5 位匹配结果。绿色边框表示该匹配结果与待查询的目标行人是同一个行人,红色边框表示该匹配结果与待查询的目标行人不是同一个行人。

可以看出,对于大部分空-地行人图像(ID 1 至 ID 4),本文方法基本能够匹配所有正确的样本。然而,对于更具挑战性的小部分空-地行人图像(ID 5 与 ID 6),如质量较差或存在大面积物体遮挡的空-地行人图像,本文方法会混淆一些难以判别的样本。这主要是由于图像噪声、物体遮挡等因素导致模型难以从空-地行人图像中提取行人关键特征。

3 结 论

本文提出一种位置感知与局部预交互的空-地行人重识别方法,旨在解决现有空-地行人重识别方法中存在的行人特征判别性较差、模型泛化能力不足等问题,主要结论如下:

1) 位置偏置自注意力机制能够通过注意力分数中融入位置信息,引导模型聚焦输入序列中关键位置的令牌,增强模型对图块之间空间位置关系的感知能力。

2) 提示-局部预交互模块能够通过快速对齐提示向量与局部特征,建立提示语义与局部特征之间



图 5 在 LAGPeR 数据集上的可视化结果

Fig. 5 Visualization results on LAGPeR dataset

的细粒度关联, 增强提示向量对当前视角图像局部具体细节的感知能力, 提升模型的局部细节重建能力。

3) 标签平滑正则化训练策略能够通过将原始 One-Hot 编码的硬标签转换为软标签, 使模型学习更平滑、泛化性更强的特征表示, 缓解模型过拟合问题, 提升模型的性能与泛化能力。

4) 本文方法能够实现较为优异的空-地行人重识别性能: 在 $A \rightarrow G$ 与 $G \rightarrow A$ 实验设置下, 模型的 Rank-1 准确率、mAP 与 mINP 均取得了最优结果; 在 $G \rightarrow A+G$ 实验设置下, 模型的 Rank-1 准确率与 mAP 取得了接近最优结果的次优结果、mINP 取得了最优结果。

参考文献 (References)

[1] 孙义博, 张文靖, 王蓉, 等. 基于通道注意力机制的行人重识别方法[J]. 北京航空航天大学学报, 2022, 48(5): 881-889.
Sun Y B, Zhang W J, Wang R, et al. Pedestrian re-identification method based on channel attention mechanism[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(5): 881-889(in Chinese).

[2] 宋晓勇, 孙学宏, 刘丽萍, 等. 基于多细粒度双流网络的行人重识别[J]. 电子测量与仪器学报, 2025, 39(8): 250-257.
Song X Y, Sun X H, Liu L P, et al. Pedestrian re-identification based on a multi-granularity dual-stream network[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(8): 250-257 (in Chinese).

[3] Huang Y K, Zha Z J, Fu X Y, et al. Illumination-invariant person re-identification[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM, 2019: 365-373.

[4] Hou R B, Ma B P, Chang H, et al. VRSTC: occlusion-free video person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 7176-7185.

[5] Huang H J, Li D W, Zhang Z, et al. Adversarially occluded samples for person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 5098-5107.

[6] Li X, Zheng W S, Wang X J, et al. Multi-scale learning for low-resolution person re-identification[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2016: 3765-3773.

[7] Wang Y, Wang L Q, You Y R, et al. Resource aware person re-identification across multiple resolutions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8042-8051.

[8] Schumann A, Metzler J. Person re-identification across aerial and ground-based cameras by deep feature fusion[J]. Automatic Target Recognition XXVII, 2017, 10202: 102020A.

[9] Nguyen H, Nguyen K, Sridharan S, et al. Aerial-ground person re-ID[C]//Proceedings of the IEEE International Conference on Multimedia and Expo. Piscataway: IEEE Press, 2023: 2585-2590.

[10] 贝俊仁, 张权, 赖剑煌. 基于隐式解码对齐的空地行人重识别方法[J]. 自动化学报, 2025, 51(9): 1988-2000.
Bei J R, Zhang Q, Lai J H. Implicit decoder alignment for aerial-ground person re-identification[J]. Acta Automatica Sinica, 2025, 51(9): 1988-2000(in Chinese).

[11] Wang Y H, Hu X, Wang L X, et al. SD-ReID: view-aware stable diffusion for aerial-ground person re-identification[EB/OL]. (2026-05-14)[2026-05-29]. <https://arxiv.org/abs/2504.09549>.

[12] Mei L, Cheng Y W, Chen H X, et al. Unsupervised aerial-ground re-identification from pedestrian to group for UAV-based surveillance[J]. Drones, 2025, 9(4): 244.

[13] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. (2021-06-03)[2025-12-02]. <https://arxiv.org/abs/2010.11929>.

[14] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 2818-2826.

[15] Zhang Q, Wang L, Patel V M, et al. View-decoupled transformer for person re-identification under aerial-ground camera network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 22000-22009.

[16] Wang S N, Wang Y L, Wu R Q, et al. SeCap: self-calibrating and adaptive prompts for cross-view person re-identification in aerial-ground networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2025: 22119-22128.

[17] Wang X G, Doretto G, Sebastian T, et al. Shape and appearance context modeling[C]//Proceedings of the IEEE 11th International Conference on Computer Vision. Piscataway: IEEE Press, 2007: 1-8.

[18] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2016: 1116-1124.

[19] Ye M, Shen J B, Lin G J, et al. Deep learning for person re-identification: a survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 2872-2893.

[20] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009: 248-255.

[21] Kingma D P, BA J. Adam: a method for stochastic optimization [EB/OL]. (2017-01-30)[2025-12-02]. <https://arxiv.org/abs/1412.6980>.

[22] He S T, Luo H, Wang P C, et al. TransReID: Transformer-based object re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2022: 14993-15002.

[23] Li S Y, Sun L, Li Q L. CLIP-ReID: exploiting vision-language model for image re-identification without concrete text labels[EB/OL]. (2023-01-01)[2025-12-02]. <https://arxiv.org/abs/2211.13977>.

[24] Wu R Q, Jiao B L, Wang W X, et al. Enhancing visible-infrared person re-identification with modality- and instance-aware visual prompt learning[EB/OL]. (2024-06-18)[2025-12-02]. <https://arxiv.org/abs/2406.12316>.

Aerial-ground person re-identification method via position-aware and local pre-interaction

BI Yihan, LI Chong, WANG Rong*

(School of Information and Cyber Security, People’s Public Security University of China, Beijing 100038, China)

Abstract: In response to the issues of poor discriminative power in pedestrian features and insufficient model generalization in existing aerial-ground person re-identification methods, this paper proposes an aerial-ground person re-identification method via position-aware and local pre-interaction. A positionally biased self-attention mechanism is proposed, which incorporates positional information into attention scores to guide the model in focusing on tokens at key positions within the input sequence. This enhances the model's perception of spatial relationships between image patches and improves the discriminability and robustness of pedestrian features. A prompt-local pre-interaction module is created that uses a single, inexpensive pre-interaction to create fine-grained linkages between prompt semantics and local attributes. This strengthens the prompt vector's ability to perceive specific local details in the current view and enhances the model's capability for fine-grained detail reconstruction. The label smoothing regularization training strategy is introduced, converting original one-hot encoded hard labels into soft labels. This encourages the model to learn smoother and more generalizable feature representations, mitigates overfitting, and improves overall model performance and generalization. The validity of the proposed strategy is confirmed by extensive trials on the public aerial-ground person re-identification dataset LAGPeR, which show that it effectively enhances Re-Identification performance.

Keywords: aerial-ground person re-identification; Transformer; self-attention mechanism; local feature pre-interaction; label smoothing regularization

基于 MTL-sLSTM 的低空载具锂电池 RUL 预测

张慧珊¹, 沈吉英², 刘东升^{1,*}, 周志凯³, 胡一凡¹, 徐扬波¹

(1. 浙江工商大学 计算机科学与技术学院, 杭州 310020; 2. 杭州市科技信息研究院, 杭州 310053;

3. 杭州数政科技有限公司, 杭州 310020)

摘 要: 针对低空经济载具锂离子电池在多工况耦合下的剩余寿命 (RUL) 预测问题, 提出一种基于多任务学习的标量长短期记忆网络 (MTL-sLSTM) 预测方法。该方法通过多源异构输入层融合跨飞行工况的多维度时序数据, 采用层级堆叠的 sLSTM 结构作为共享特征提取器, 实现跨时间尺度的深度特征融合与非线性退化共性特征提取; 结合任务 ID 硬编码索引动态调整网络权重, 通过多任务学习机制自适应捕捉不同工况的退化模式, 同步实现跨工况知识共享与独立 RUL 预测。实验结果表明: MTL-sLSTM 的预测性能显著优于注意力混合专家模型 (AttMoE)、双通道 LSTM (Dual-LSTM) 等 6 种时序方法, 在 eVTOL 数据集上均方根误差 (RMSE) 较对比方法降低 60.7%~92.3%, 验证了多任务学习机制在增强复杂工况下退化特征泛化能力与提升预测精度方面的有效性。

关键词: 剩余寿命预测; 多任务学习; 长短期记忆网络; 锂离子电池; 电动垂直起降飞机; 低空经济载具

中图分类号: TM912; V242.2; V241.6; V271.3

文献标志码: A

文章编号: 1001-5965(2026)09-3125-11

低空经济作为战略性新兴产业, 依托无人机、电动垂直起降飞行器 (electric vertical takeoff and landing, eVTOL) 等载具在物流、救援等领域快速发展。2024 年中国市场规模已超 6 000 亿元, 预计 2026 年突破万亿, 全球市场同步呈现指数增长。锂离子电池作为核心动力组件, 其剩余使用寿命 (remaining useful life, RUL) 的精准预测直接影响飞行安全。美国联邦航空管理局 (Federal Aviation Administration, FAA) 报告指出, 2018—2023 年 23% 的电动航空器事故与电池衰退相关, 容量衰减超 20% 的电池电压崩溃风险骤增 3.7 倍。

低空载具的“充电-飞行”双模态运行场景对 RUL 预测提出特殊需求: 充电工况可通过恒流/恒温模式下的容量增量曲线等参数建模^[1-3]; 飞行工况

则需应对动态三阶段放电 (高倍率起飞/着陆与中倍率巡航), 尤其是着陆阶段的高负载冲击加剧容量与内阻的耦合衰减。研究指出结合飞行中电压动态特性和温度场数据可提升预测精度^[4], 但现有方法多聚焦单一工况建模, 尚未建立多工况耦合预测体系。

近年来, 航空动力电池 RUL 预测研究取得显著进展。纳米磷酸铁锂电池的航空适用性验证为电池选型奠定基础^[5], 长短期记忆网络 (long short-term memory, LSTM)、灰色 Verhulst-神经网络等模型在小样本场景展现潜力, 而卷积神经网络-长短期记忆网络-深度神经网络 (convolutional neural network-long short-term memory-deep neural network multimodal network, CNN-LSTM-DNN) 多模态网络虽将误差降

收稿日期: 2025-04-03; 录用日期: 2025-04-29; 网络出版时间: 2025-06-04 11:09

网络出版地址: link.cnki.net/urlid/11.2625.V.20250604.0851.001

基金项目: 浙江省万人计划项目 (2022R52048); 浙江省科技计划项目 (2024C01028, 2024C01211)

*通信作者. E-mail: lds1118@zjgsu.edu.cn

引用格式: 张慧珊, 沈吉英, 刘东升, 等. 基于 MTL-sLSTM 的低空载具锂电池 RUL 预测 [J]. 北京航空航天大学学报, 2026, 52 (9): 3125-3135.

Zhang H S, Shen J Y, Liu D S, et al. Remaining useful life prediction of lithium-ion batteries in low-altitude vehicles based on MTL-sLSTM [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3125-3135 (in Chinese).

至 1.8%, 却因结构复杂导致训练耗时激增^[6-7]。当前研究趋势转向多源信息融合与动态建模, 贺宁等^[8]开发的粒子滤波-门控循环单元 (particle filter-gated recurrent unit, PF-GRU) 混合框架, 有效抑制了容量再生现象干扰。董渊昌等^[9]采用奇异值分解-高斯过程回归方法构建的不确定性预测模型, 其预测区间覆盖率提升至 95.2%, 但健康因子与多工况衰减机理的普适关联仍需深入验证。陈欣等^[10]使用 Transformer 和 GRU 的联合方法提升了锂电池的剩余使用寿命预测鲁棒性。Chen 等^[11]基于 Transformer 模型, 将注意力机制与混合专家 (mixture of experts, MoE) 相结合, 相对误差方面实现了 10%~20% 的改进。尽管 Transformer 在理论上具备强大的特征捕获能力, 但受限于自注意力机制的非线性建模缺陷, 难以满足机载实时性需求^[12]。当前研究多聚焦单一充电或飞行工况建模, 尚未建立适应低空经济载具动力电池多工况耦合特性的预测体系, 特别是在跨工况迁移学习与多任务协同优化领域。

迁移学习成为突破跨工况预测瓶颈的新方向。现有迁移学习方法如对抗性域适应、统计匹配^[13-15], 虽能缓解数据分布偏移问题, 但其依赖目标域充足的历史数据, 而低空经济载具多模式飞行任务 (城市物流、应急救援等) 导致目标工况数据极度稀缺。相比之下, 多任务学习 (multi-task learning, MTL) 通过共享-特异双路架构同步挖掘多工况共性特征与个性化退化规律, 在有限数据条件下实现知识迁移, 而传统 MTL 模型多采用 CNN 或 GRU^[16-18]作为主干网络, 难以有效捕获电池容量衰减的长程时序依赖。

本文面向低空经济载具多工况场景, 提出一种基于多任务学习的标量长短期记忆网络 (multi-task learning scalar long short-term memory, MTL-sLSTM) 预测方法, 其主要创新点体现在三方面。

1) 多任务协同优化机制: 设计多源异构输入层、共享特征提取器与多任务预测器, 通过输入层

融合飞行工况异构数据, 共享层学习跨工况退化共性, 任务特异层捕捉不同工况的独有衰减模式, 解决传统单任务模型对工况敏感性问题。

2) sLSTM 时序建模增强: 作为共有特征提取器, 承担不同飞行工况退化模式泛化表征的关键任务。通过深度可分离卷积消除锂电池不同飞行工况特异性噪声; 利用层级堆叠结构逐层提取从局部退化片段到全局老化趋势的多粒度信息; 建立参数共享机制迫使网络学习工况不变性特征, 为目标域零样本预测提供基础。

3) 轻量化多任务预测架构: 基于任务 ID 硬编码索引定位私有预测头参数, 共享特征截取末端状态输入任务私有全连接层, 实现多电池退化模式差异化建模, 解决了传统多任务模型计算冗余与航空场景实时性需求的矛盾。

1 基于 MTL-sLSTM 的锂电池 RUL 预测方法

1.1 问题设定

锂离子电池广泛应用于低空经济载具中, 其 RUL 的精准预测直接影响飞行安全^[11]。现有基于实验电池测量的研究将电池容量衰减至初始额定值的 70% 或 80% 时, 即判定达到寿命终止 (end-of-life, EOL) 阈值^[19]。RUL 定义为电池在 EOL 之前剩余的循环次数^[20], 其数学表征为

$$N_{\text{RUL}}^k = N_{\text{EOL}}^k - N_{\text{ECL}}^k$$

(1)

式中: N_{RUL}^k 、 N_{EOL}^k 、 N_{ECL}^k 分别为任务 k 下的 RUL、电池达到 EOL 时的循环次数、等效循环寿命 (equivalent cycle life, ECL)。

1.2 本文方法架构

针对工况转换场景下低空经济载具锂电池的 RUL 预测, 本文 MTL-sLSTM 旨在通过多源数据融合与跨工况知识共享, 缓解工况偏移导致的泛化问题。图 1 为 MTL-sLSTM 框架, 由多源异构输入层、共享特征提取器和多任务预测器构成。

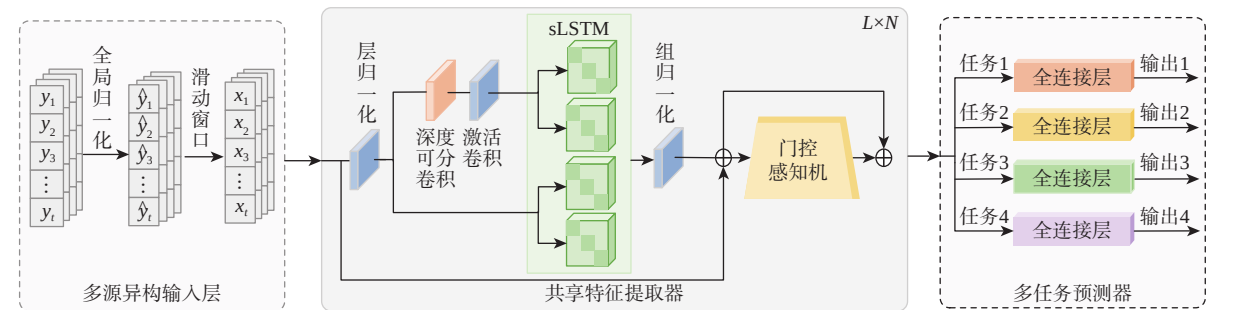


图 1 MTL-sLSTM 框架

Fig. 1 MTL-sLSTM framework

1.2.1 多源异构输入层

对于低空经济载具锂电池在复杂工况下的多源数据融合问题, 实现多工况数据统一表征与对齐至关重要。多工况数据统一表征能够将不同工况下的数据转化为统一的表示形式, 消除数据间的差异和不一致性。

首先, 基于锂电池的电化学过程与任务配置特征, 提取具有代表性的退化特征集合。

其次, 所有子任务数据统一进行全局对比度归一化处理(global contrast normalization, GCN), 计算式为

$$\hat{x} = (x - \mu_{\text{globe}}) / \sigma_{\text{globe}} \quad (2)$$

式中: x 为原始特征值; μ_{globe} 和 σ_{globe} 分别为跨电池数据的全局均值和标准差。全局对比度归一化能够消除不同电池个体间由于制造差异、使用条件等因素导致的数据分布差异, 确保不同工况数据的分布对齐, 从而使模型在处理不同电池数据时具有更好的一致性和稳定性, 提升模型的泛化能力。

最后, 通过滑动窗口生成时序样本, 样本构造如下:

$$\mathbf{X}_t = [\mathbf{x}_{t-T+1}, \mathbf{x}_{t-T+2}, \dots, \mathbf{x}_t] \in \mathbf{R}^{T \times s} \quad (3)$$

式中: $\mathbf{X}_t \in \mathbf{R}^d$ 为第 t 个循环的特征向量, d 为维度; T 为窗口长度; s 为每个时间点特征向量的维度。

滑动窗口生成时序样本的好处在于能够捕捉电池在连续循环周期内的动态变化模式, 将静态的特征数据转化为具有时间序列信息的样本, 有助于模型学习到电池性能随时间的变化规律。

1.2.2 共享特征提取器

共享特征提取器作为模型实现跨工况知识迁移的核心模块, 承担跨工况退化模式泛化表征的关键任务。针对锂电池多源异构数据中存在的工况漂移和局部退化突变问题, 本文基于 sLSTM 构建层级特征蒸馏架构, 其核心目标在于多尺度特征抽

象, 利用层级堆叠结构 $L \times N$ 逐层提取从局部退化片段到全局老化趋势的多粒度信息, 并且建立参数共享机制迫使网络学习工况不变性特征, 为目标域零样本预测提供基础。

共享特征提取器由多层改进型 sLSTM 模块堆叠而成, 每层通过门控残差连接实现特征精炼, 其技术实现包含 3 个阶段:

1) 深度可分离卷积增强门控机制

深度可分离卷积将标准卷积分解为深度卷积和逐点卷积两部分, 如图 2 所示。深度卷积对每个输入通道独立进行卷积操作, 逐点卷积则使用 1×1 的卷积核在通道维度上进行线性组合, 以融合不同通道的信息。在传统 LSTM 遗忘门中引入深度可分离卷积核, 重构其预激活计算流程为

$$\tilde{f}_t = \mathbf{W}_{\text{dw}} \otimes [\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_f \quad (4)$$

$$\tilde{i}_t = \mathbf{W}_{\text{dw}} \otimes [\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_i \quad (5)$$

式中: $\tilde{f}_t$ 、 $\tilde{i}_t$ 分别为遗忘门、输入门的预激活值; $\mathbf{W}_{\text{dw}} \in \mathbf{R}^{k \times (d+s) \times m}$ 为深度可分离卷积核参数, 其中, k 为卷积核大小, d 为隐藏状态 $\mathbf{h}_{t-1}$ 的维度, s 为输入特征 $\mathbf{x}_t$ 的维度, m 为通道数; $\otimes$ 为深度可分离卷积操作; $\mathbf{h}_{t-1} \in \mathbf{R}^d$ 为上一时间步的隐藏状态; $\mathbf{x}_t \in \mathbf{R}^s$ 为当前时间步输入的 s 个特征; $\mathbf{b}_f$ 、 $\mathbf{b}_i \in \mathbf{R}^m$ 为相应偏置向量。

2) 多块对角线性层

如图 3 所示, 候选记忆与输出门的预激活值通过块对角线性层(即多头全连接)计算。该层将权重矩阵分解为多个对角块(“头”), 每个头独立学习, 以增强参数效率。在本框架中, 头数设置为 4(图中简称为 $N_{\text{H}}=4$)。候选记忆与输出门的预激活计算式为

$$\tilde{z}_t = \mathbf{W}_z \mathbf{x}_t + \mathbf{R}_z \mathbf{h}_{t-1} + \mathbf{b}_z \quad (6)$$

$$\tilde{o}_t = \mathbf{W}_o \mathbf{x}_t + \mathbf{R}_o \mathbf{h}_{t-1} + \mathbf{b}_o \quad (7)$$

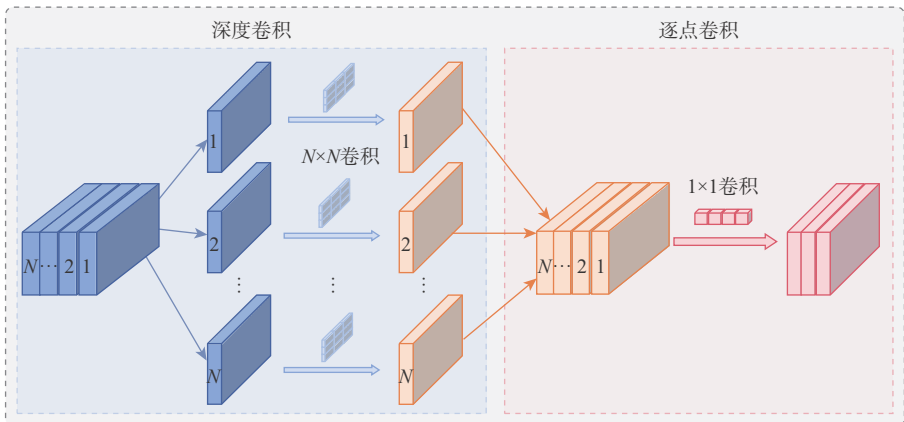


图 2 深度可分离卷积示例图

Fig. 2 Example diagram of depthwise separable convolution

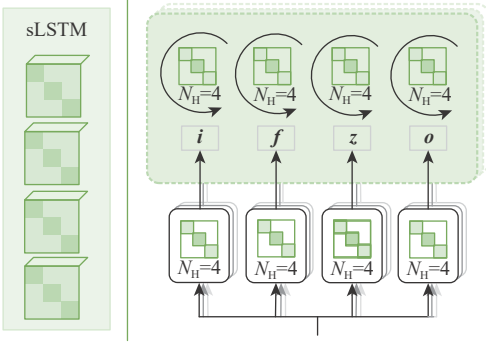


图3 多块对角线性层

Fig. 3 Multi-block diagonal linear layer

式中： $\tilde{z}_i$ 、 $\tilde{o}_i$ 分别为候选记忆、输出门的预激活值； W_z 、 $W_o \in \mathbf{R}^{d \times s}$ 、 R_z 、 $R_o \in \mathbf{R}^{d \times d}$ 为块对角权重矩阵（每个“头”对应一个对角块）； b_z 、 $b_o \in \mathbf{R}^d$ 为偏置向量。

sLSTM在原始LSTM上引入标量记忆和指数门控来管理长期依赖关系，获得上述预激活值后，首先，计算稳定状态以保障指数门控的数值稳定性：

$$\mathbf{m}_t = \max(\mathbf{f}_t + \mathbf{m}_{t-1}, \mathbf{i}_t) \quad (8)$$

式中： $\mathbf{m}_t$ 为当前时刻的稳定状态； $\mathbf{m}_{t-1}$ 为上一时刻的稳定状态； $\max(\cdot)$ 为逐元素取最大值函数。随后，对遗忘门、输入门施加指数激活，对候选记忆施加双曲正切激活，对输出门施加Sigmoid激活：

$$\mathbf{i}_t = \exp(\tilde{\mathbf{i}}_t - \mathbf{m}_t) \quad (9)$$

$$\mathbf{f}_t = \exp(\tilde{\mathbf{f}}_t + \mathbf{m}_{t-1} - \mathbf{m}_t) \quad (10)$$

$$\mathbf{z}_t = \tanh(\tilde{\mathbf{z}}_t) \quad (11)$$

$$\mathbf{o}_t = \sigma(\tilde{\mathbf{o}}_t) \quad (12)$$

式中： $\exp(\cdot)$ 为自然指数函数； $\tanh(\cdot)$ 为双曲正切函数； $\sigma(\cdot)$ 为Sigmoid函数。指数门控使网络能够动态调节信息存储与遗忘的幅度，而稳定状态 $\mathbf{m}_t$ 有效防止了指数运算的数值溢出。

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{z}_t \quad (13)$$

$$\mathbf{n}_t = \mathbf{f}_t \odot \mathbf{n}_{t-1} + \mathbf{i}_t \quad (14)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot (\mathbf{c}_t \odot \mathbf{n}_t) \quad (15)$$

式中： $\mathbf{c}_t$ 为细胞状态（长时记忆）； $\mathbf{n}_t$ 为归一化状态； $\mathbf{h}_t$ 为当前时刻的隐藏状态； $\odot$ 和 $\oslash$ 分别表示逐元素乘法和逐元素除法。通过归一化状态 $\mathbf{n}_t$ ，sLSTM将细胞状态缩放至合理范围，缓解了传统LSTM中状态值无界增长的问题，从而提升了长序列建模的稳定性。

3) 任务感知特征投影

在共享层末端引入门控多层感知机（gated multi-layer perceptron, gMLP），如图4所示，其上投影和下投影调整特征的维度，并通过高斯误差线性

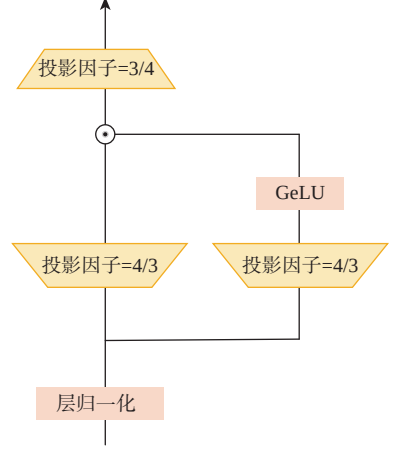


图4 门控多层感知机

Fig. 4 Gated multi-layer perceptron

单元（Gaussian error linear unit, GeLU）激活函数引入非线性响应，有助于模型学习电池退化特征的复杂映射关系。此外在上下投影前使用了层归一化（layer normalization, LN），对每个样本的所有激活进行归一化。最后，残差堆叠通过在网络中引入跳过连接来实现，将输入直接添加到输出上，形成残差连接。

最后，共享特征提取器通过残差堆叠，能够构建深层网络，提高模型的表达能力和处理复杂长序列数据的能力。

1.2.3 多任务预测器

为实现多电池退化模式的差异化建模，设计了任务自适应预测结构，其核心思想是将共享特征投影至各任务私有的退化轨迹空间。每条样本在通过共享特征提取器后输出序列特征，取最后一个时间步的隐藏状态表示当前退化状态。该状态特征作为输入，分别送入任务对应的私有全连接回归层中实现预测。任务编号作为静态输入参与权重选择，每个任务拥有独立的参数组以建模其特有的退化形态，该机制能有效实现在共享建模能力基础上的任务分离预测能力。

2 实验验证与分析

2.1 数据集

本文对3个公开数据集进行实验验证：CALCE数据集^[21]来自马里兰大学高级生命周期工程中心；NASA数据集^[22]由美国国家航空航天局提供的锂电池老化测试数据，广泛用于寿命预测领域研究；eVTOL数据集由美国卡内基梅隆大学和Airbus A³公司测量获取^[23]，针对eVTOL运行场景采集的多工况电池使用数据。

为实现多工况建模的全面性与代表性，本文基于放电电压差异对跨领域电池数据集进行系统性筛

选: 从 CALCE 数据集中选定 CS2_35~CS2_38 这 4 个子集, 同时采用 NASA 数据集中的 B0005、B0006、B0007 和 B0018 作为样本; 在 eVTOL 数据集应用 场景方面, 则选取具有典型工况特征的 VAH27、VAH15、VAH28 及 VAH30 这 4 组飞行数据。在特 征维度构建方面, 综合 Chen 等^[11]对 CALCE 和 NASA 数据集的特征重要性研究成果, 选取包含循 环次数、容量衰减、健康状态、阻抗特性及充放电 时序参数(恒流充电时间、恒压充电时间)等关键指 标; 同时结合 Mitici 等^[4]对 eVTOL 数据的分析结 论, 整合了涵盖电压动态特性(起飞/巡航阶段电压 均值与方差)、极端工况参数(最小电压、最大温 度)及容量保持率等 8 项核心特征。特别针对 eVTOL 典型工况, 设置 VAH27 为基础测试场景(起 降功率为 54 W, 巡航功率为 16 W, 巡航时间为 800 s, 基准环境温度为 25 ℃), 并差异化构建 VAH15 (1 000 s 长时巡航)、VAH28(全阶段 90% 负载)和 VAH35(35 ℃ 高温环境)等对比工况。通过这 3 类 数据集——覆盖从受控实验环境到动态真实应用 场景的全谱系电池使用模式, 本文建立的验证体系 不仅能有效评估电池退化机理, 更能系统检验模型 在跨域迁移、变载适应及极端环境下的泛化性能。

图 5~图 7 分别为 CALCE、NASA、eVTOL 数 据集子集的容量衰减曲线。与 CALCE 和 NASA 数 据集中连续、高频的容量记录不同, eVTOL 数据 集子集是模拟真实飞行任务下的电池使用情形, 其 容量测试呈现出低频间歇性采样特征, 导致可用于 建模的有效 RUL 样本数量非常有限(每个子集仅 11~23 个容量观测点), 且对于 eVTOL 电池, 考虑保守 的安全裕度, 本文将 EOL 阈值设置为初始测量电 池容量的 85%。

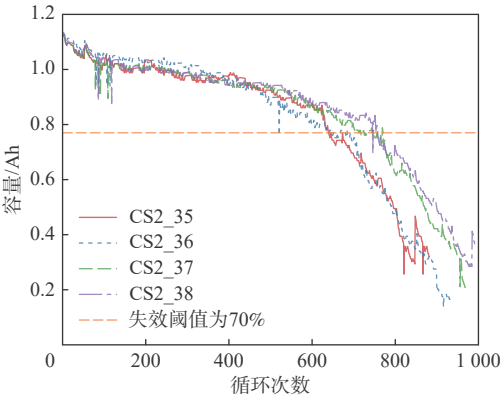


图 5 CALCE 数据集的容量衰退曲线

Fig. 5 Capacity degradation curves of CALCE dataset

2.2 评估指标

为了在验证过程中对电池多维特征的寿命预 测性能进行定量评估, 采用了 3 个评估标准: 相对误

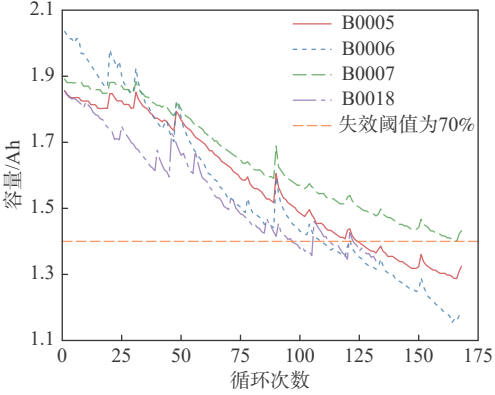


图 6 NASA 数据集的容量衰退曲线

Fig. 6 Capacity degradation curves of NASA Dataset

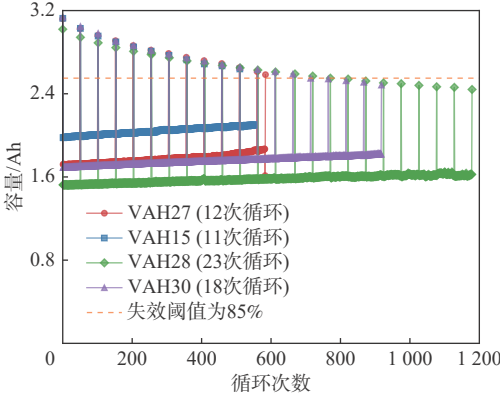


图 7 eVTOL 数据集的容量衰退曲线

Fig. 7 Capacity degradation curves of eVTOL dataset

差(relative error, RE)、平均绝对误差(mean absolute error, MAE)和均方根误差(root mean square error, RMSE), 定义如下:

$$R_E = \frac{|y_t - \hat{y}_t|}{y_t} \tag{16}$$

$$M_{AE} = \frac{1}{S} \sum_t |y_t - \hat{y}_t| \tag{17}$$

$$R_{MSE} = \sqrt{\frac{1}{S} \sum_t (y_t - \hat{y}_t)^2} \tag{18}$$

式中: y_t 和 $\hat{y}_t$ 分别为第 t 个样本的测量寿命和预测 寿命; S 为样本数量; RE、MAE 和 RMSE 的值越小 性能越好。

2.3 实施细节

实验在配备 NVIDIA GeForce RTX 4090 显卡 (24 GB 显存) 及 128 GB RAM 的工作站上进行, 代 码基于 PyTorch 1.12.1 框架实现。

在多源异构输入层的数据处理阶段, 均采用滑 动窗口法构造序列样本, eVTOL 数据设定窗口长度 为 10, 而 CALCE 和 NASA 数据集采用长度分别为 32 和 64 的窗口以捕获完整退化特征。所有输入序 列均进行零均值标准化处理, 容量标签归一化至

[0,1] 区间, 确保多源数据的一致建模。为验证模型在多任务场景下的鲁棒性与泛化能力, 引入任务编号作为条件输入, 统一调度不同子任务对应的私有预测器。训练过程采用 Adam 优化器, 基础学习率为 1×10^{-3} , 并使用梯度裁剪(阈值为 5.0)防止梯度爆炸。不同模型结构(如 GRU、LSTM、Transformer、sLSTM)下均采用隐藏层维度为 64 的共享编码器, 并基于任务分离回归头输出各子集的 RUL 预测结果。批量大小根据子集样本规模与显存占用动态在 32、64、128 之间调整, 配合自适应学习率调度与早停机制(50 轮验证停滞)实现高效训练。

模型最终性能通过 MAE、RMSE 和 RE 指标综合评估, 并在 3 组随机种子(42、123、999)下重复实验取平均, 以减弱偶然性误差影响。sLSTM 模型结构中进一步设置堆叠块数为 7, 深度卷积核大小为 4, 感知机投影因子为 1.3, 上下文长度为 256, 并施加 1×10^{-4} 的权重衰减正则化, 从而在保持模型容量的前提下提升泛化能力。

3 实验结果与分析

3.1 总体表现

为进一步验证本文方法的性能, 引入当前主流方法在 RE、MAE、RMSE 进行对比分析。所有方法的得分如表 1~表 3 所示, 所有指标均基于 5 次实验的平均结果计算。最佳结果以粗体突出显示, 次佳结果以下划线突出显示。

表 1 各方法在 CALCE 数据集上 RUL 预测性能对比			
Table 1 Comparison of RUL prediction performance of various methods on CALCE dataset			
方法	RE	MAE	RMSE
MLP	0.401 8	0.155 7	0.203 8
LSTM	0.090 2	0.058 2	0.073 6
Dual-LSTM	0.088 5	0.063 6	0.087 4
GRU	0.131 9	0.067 1	0.094 6
AttMoE	0.076 1	0.057 7	0.079 4
sLSTM	<u>0.048 5</u>	<u>0.046 9</u>	<u>0.053 7</u>
MTL-sLSTM	0.012 7	0.015 1	0.020 2

由表 1~表 3 可知: ①本文 MTL-sLSTM 在所有方法中取得了最好的结果, 体现了其在多工况下复杂时序数据建模中的卓越能力; ②sLSTM 也能得到较好的性能, 表明 sLSTM 的层级结构有效捕获了多工况锂电池的长期时序依赖关系; ③MTL-sLSTM 通过联合优化多个相关任务, 充分利用任务间的潜在关联性, 显著提升了预测精度, 与当前主流模型(Dual-LSTM、基于 Transformer 的 AttMoE 等)相比, MTL-sLSTM 在 3 项实验中的各项指标均达到最

表 2 各方法在 NASA 数据集上 RUL 预测性能对比			
Table 2 Comparison of RUL prediction performance of various methods on NASA dataset			
方法	RE	MAE	RMSE
MLP	0.385 1	0.137 9	0.154 1
LSTM	0.264 8	0.082 9	0.090 5
Dual-LSTM	0.255 7	0.081 5	0.087 9
GRU	0.304 4	0.080 6	0.092 1
AttMoE	0.200 0	0.076 0	0.087 2
sLSTM	<u>0.189 8</u>	<u>0.061 7</u>	<u>0.073 5</u>
MTL-sLSTM	0.058 7	0.012 4	0.018 1

表 3 各方法在 eVTOL 数据集上的 RUL 预测性能对比			
Table 3 Comparison of RUL prediction performance of various methods on eVTOL dataset			
方法	RE	MAE	RMSE
MLP	0.281 7	0.099 5	0.147 5
LSTM	0.580 1	0.118 9	0.153 7
Dual-LSTM	<u>0.199 6</u>	0.258 7	0.190 0
GRU	0.407 7	0.091 8	0.125 3
AttMoE	0.921 3	0.317 4	0.329 8
sLSTM	0.257 2	<u>0.054 7</u>	<u>0.064 4</u>
MTL-sLSTM	0.138 6	0.018 7	0.025 3

优, 尤其在 eVTOL 数据集小样本下优势更显著, 这进一步证明了多任务学习与特征共享机制在复杂工况下时序预测任务中的实际应用价值。

为直观说明 MTL-sLSTM 的预测表现, 如图 8~图 10 所示, 分别为 MTL-sLSTM 模型的 CALCE、NASA 及 eVTOL 数据集 4 个子集的真实值与预测值对比。由图 8~图 10 可知, 各个预测模型与真实数据的拟合效果均表现良好。

3.2 多源数据对齐评估

为证明多源异构输入层能有效缩小不同子集的分布差异, 本节以 eVTOL 数据集为例, 使用主成分分析(principal component analysis, PCA)将 8 维特征分别投影到二维空间, 如图 11 所示。

在图 11(a)对齐前的 PCA 投影中, 各子集样本点分布明显分离, VAH30 总体位于上方, 而图 11(b)对齐后的 PCA 投影中, 不同类别的数据点在主成分 1 和主成分 2 的坐标系中分布更加集中, 表明对齐处理有效地减少了不同类别之间的分布差异。

3.3 多任务学习有效性

本文通过使用 MTL 实现性能增强, 为评估这一改进, 将 MTL 与其简化版本进行比较, 即多任务和单任务学习的对比。简化版本没有经过多源异构输入层进行数据融合, 直接将每个子集单独输入模型进行预测。

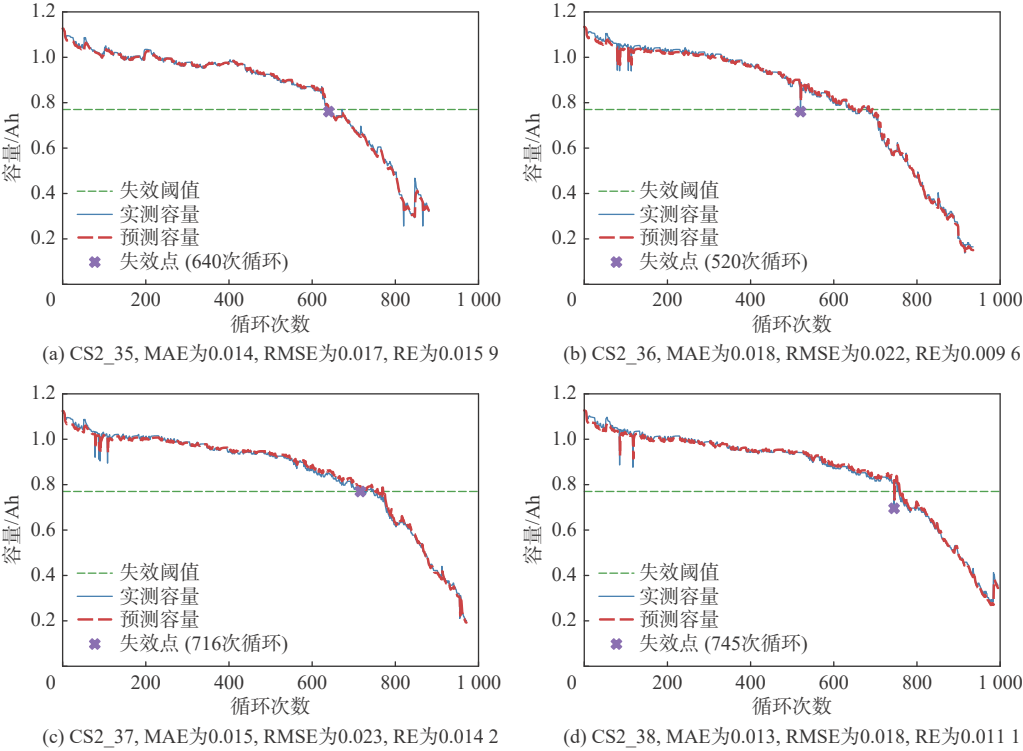


图 8 MTL-sLSTM 模型在 CALCE 数据集 4 个子集的真实值与预测值对比

Fig. 8 Comparison of true values and predicted values for MTL-sLSTM model on four subsets of CALCE dataset

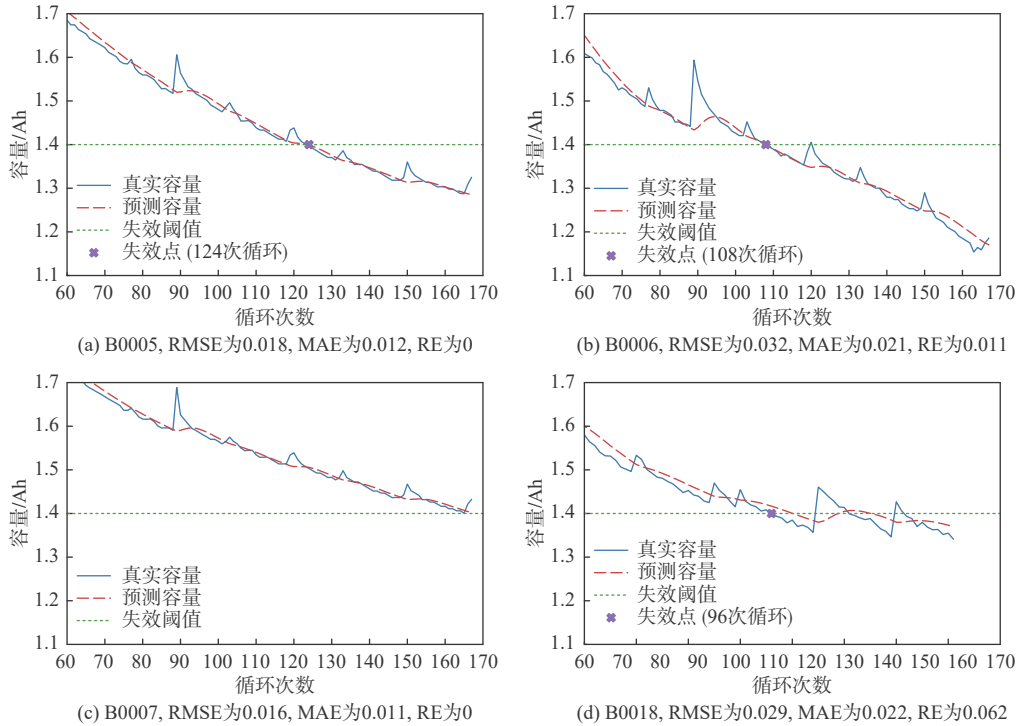


图 9 MTL-sLSTM 模型在 NASA 数据集 4 个子集的真实值与预测值对比

Fig. 9 Comparison of true values and predicted values for MTL-sLSTM model on four subsets of NASA dataset

以 CALCE 数据集为例,子集: CS2_35、CS2_36、CS2_37、CS2_38 为不同任务。表 4 为 GRU、LSTM、sLSTM 这 3 种基础模型在启用 MTL 与未启用 MTL 时的评估指标 (RE、MAE、RMSE) 及对应改进率。如在 CS2_35 数据子集中, LSTM 模型的 RE 值改进

率达 0.70; sLSTM 模型启用 MTL 的 RE 值改进率达 0.83。其他数据子集 (CS2_36、CS2_37、CS2_38) 也以相同逻辑呈现各模型在不同 MTL 状态下的指标数值及改进率,系统对比了 MTL 对各模型在 RE、MAE、RMSE 指标上的性能提升效果。

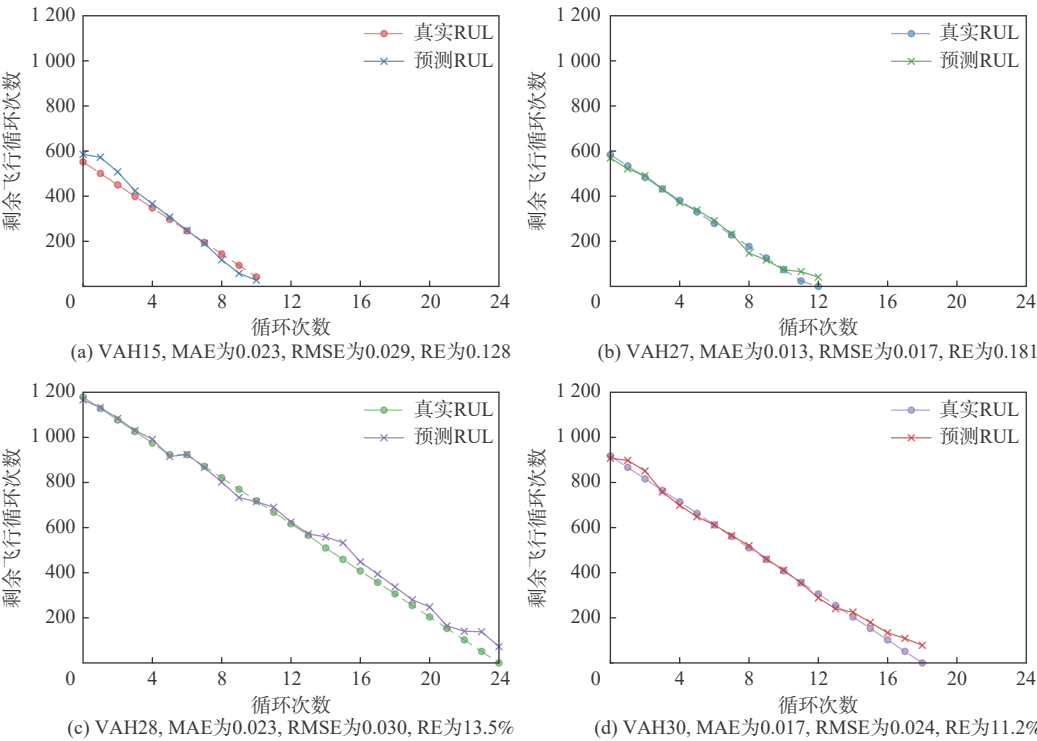


图 10 MTL-sLSTM 模型在 eVTOL 数据集 4 个子集的真实值与预测值对比

Fig. 10 Comparison of true values and predicted values for MTL-sLSTM model on four subsets of eVTOL dataset

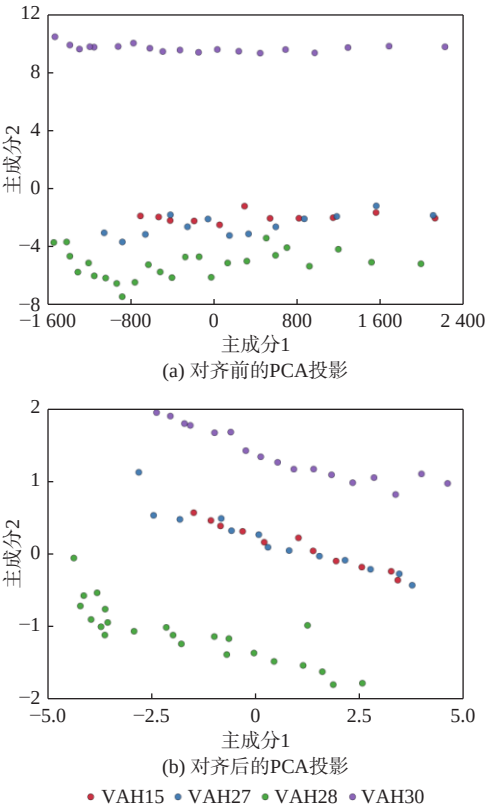


图 11 eVTOL 数据集子集的特征对齐前后对比

Fig. 11 Comparison of feature alignment before and after for eVTOL dataset subsets

图 12 为不同评估指标的 MTL-sLSTM 与 sLSTM 对比,由图可知,对于 CALCE 数据集集中的 4 个子集,MTL-sLSTM 的表现始终显著优于其简化版本

sLSTM,表明 MTL 有助于提高不同任务的 RUL 预测性能。

3.4 共享特征提取器采用 sLSTM 的有效性

实验中,首先配置模型核心参数(包括模型模式、窗口大小、隐藏层维度等),如表 5 所示。针对不同模型(如 MTL-CNN、MTL-MLP 等)设定对应参数,通过运行 3 组不同随机种子完成训练并记录损失。随后加载各模型多组种子的训练损失数据,计算均值与标准差,分别构建整体趋势图与局部放大图。整体趋势呈现完整训练轮次的 MSE 损失变化,局部放大图聚焦特定轮次区间,细化损失变化趋势。

本实验采用 log 刻度绘制,由于 MSE 损失值覆盖从 $10^{-2} \sim 10^{-5}$ 的大跨度范围,log 刻度可压缩数值量级,清晰呈现损失在不同数量级的变化,既保留训练初期损失快速下降的细节,又凸显后期收敛阶段的细微差异。

从实验结果看,共享特征提取器采用 sLSTM 具备显著有效性。对比 MTL-sLSTM 与 MTL-Transformer、MTL-GRU、MTL-RNN 等模型,MTL-sLSTM 在训练中损失下降更迅速,且最终稳定于最低水平。图 13(a) 中 MTL-sLSTM 损失曲线始终处于最下方,图 13(b) 也显示其在特定轮次区间内损失均值更低。

另外,将实验 NASA 数据集子集在多任务模型中的评估指标列为表 6 进行对比,MTL-sLSTM 在各子集(如 B0005、B0006 等)的 MAE 和 RMSE 指

表 4 CALCE 数据集子集在基础模型上启用 MTL 与未启用 MTL 时的评估指标对比

Table 4 Comparison of evaluation metrics for CALCE dataset subsets between enabling MTL and not enabling MTL on base model

模型	RE				MAE				RMSE			
	CS2_35	CS2_36	CS2_37	CS2_38	CS2_35	CS2_36	CS2_37	CS2_38	CS2_35	CS2_36	CS2_37	CS2_38
GRU	0.074 2	0.116 8	0.103 0	0.056 0	0.063 5	0.109 1	0.077 7	0.072 7	0.083 4	0.144 1	0.105 1	0.095 5
GRU +MTL	0.026 8	0.024 2	0.020 1	0.020 9	0.025 8	0.029 1	0.026 0	0.020 0	0.034 6	0.036 5	0.031 9	0.029 2
LSTM	0.109 0	0.116 4	0.098 4	0.091 3	0.080 6	0.108 5	0.078 5	0.086 9	0.103 3	0.143 6	0.104 5	0.113 0
LSTM +MTL	0.032 2	0.019 4	0.013 7	0.015 0	0.024 8	0.018 4	0.015 8	0.017 5	0.031 5	0.023 8	0.019 2	0.023 3
sLSTM	0.058 2	0.074 3	0.024 7	0.036 6	0.057 7	0.068 9	0.026 9	0.034 0	0.062 7	0.078 4	0.032 7	0.041 1
sLSTM +MTL	0.009 7	0.007 5	0.007 6	0.007 1	0.014 6	0.009 3	0.012 1	0.009 3	0.017 7	0.012 7	0.010 1	0.011 9

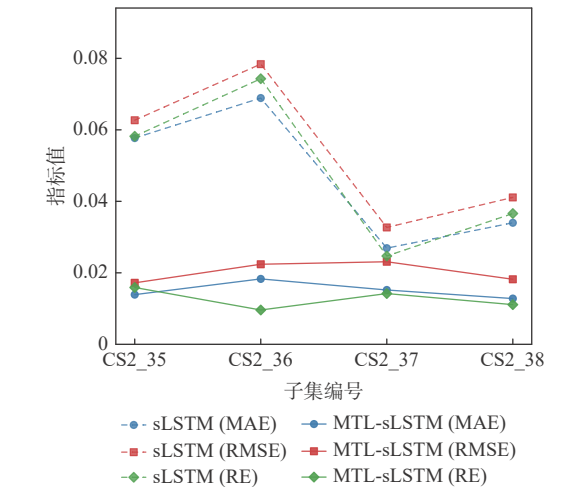


图 12 CALCE 数据集子集评估指标对比

Fig. 12 CALCE dataset subsets evaluation metrics comparison

表 5 本文方法在 NASA 数据集上的超参数

Table 5 Hyperparameters of the proposed method on NASA dataset

参数	数值
滑动窗口大小	32
隐藏层维度	128
循环神经网络层数	2
学习率	0.001
最大训练轮数	1 000
早停耐心参数	50
CNN通道数	16
MLP隐藏层单元数	256

标上全面优于其他模型(见表中加黑数值), 这表明 sLSTM 作为共享特征提取器能更高效地提取特征, 不仅加速训练收敛过程, 还大幅降低预测误差, 充分验证了其在共享特征提取模块中应用的有效性。

3.5 模型规模分析

为量化评估模型效率, 本文以 eVTOL 数据集为例, 从参数量、乘加运算次数(multiply-accumulate operations, MACs)和推理时间 3 个维度进行对比分

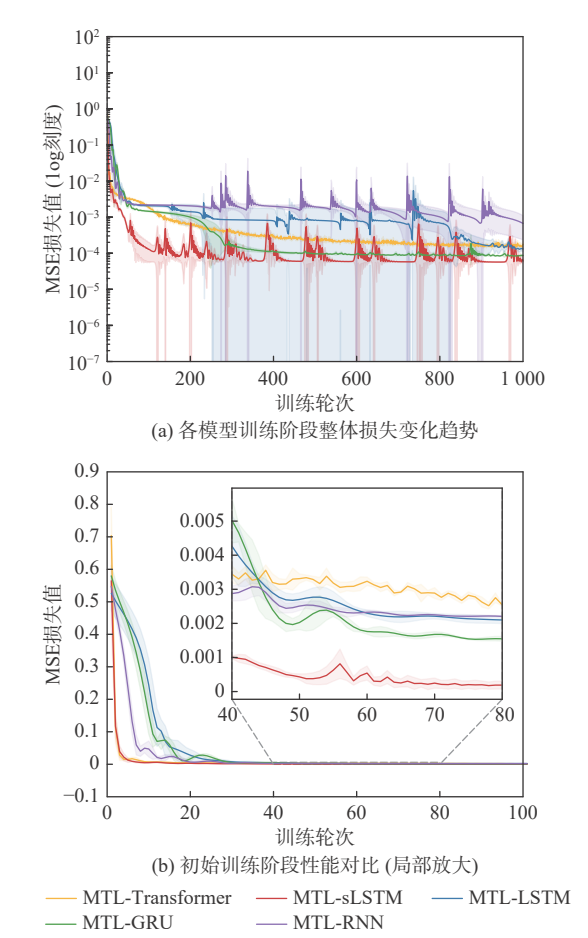


图 13 多任务模型训练损失对比

Fig. 13 Multi-task model training loss comparison

析, 如图 14 所示。

实验结果表明, 多任务学习框架展现出显著优势: 参数量方面, 多任务模型相比单任务模型总和降低 60% 以上; 计算复杂度方面, MACs 减少 40% 以上; 推理效率方面, AttMoE 的多任务单次推理耗时(0.54 ms)仅为单任务累计推理时间(2.03 ms)的 26.6%。

由图 14 可知, 多任务模型总是低于相应的单任务模型, MTL 模型在参数量、MACs 和推理时间上均优于单任务。

表 6 NASA 数据集子集在多任务模型中的评估指标对比

模型	MAE				RMSE			
	B0005	B0006	B0007	B0018	B0005	B0006	B0007	B0018
MTL-GRU	0.083 6	0.087 3	0.064 2	0.038 4	0.094 9	0.105 1	0.074 6	0.045 2
MTL-LSTM	0.094 8	0.096 3	0.069 5	0.041 3	0.107 6	0.115 8	0.080 5	0.050 2
MTL-CNN	0.079 3	0.084 1	0.059 2	0.038 1	0.090 0	0.100 5	0.068 7	0.044 7
MTL-MLP	0.081 5	0.078 6	0.058 7	0.040 8	0.092 5	0.096 1	0.068 2	0.048 0
MTL-Transformer	0.082 6	0.084 1	0.061 1	0.041 2	0.093 9	0.100 4	0.070 9	0.048 1
MTL-sLSTM	0.009 4	0.013 8	0.007 4	0.015 0	0.014 3	0.022 8	0.012 9	0.021 3

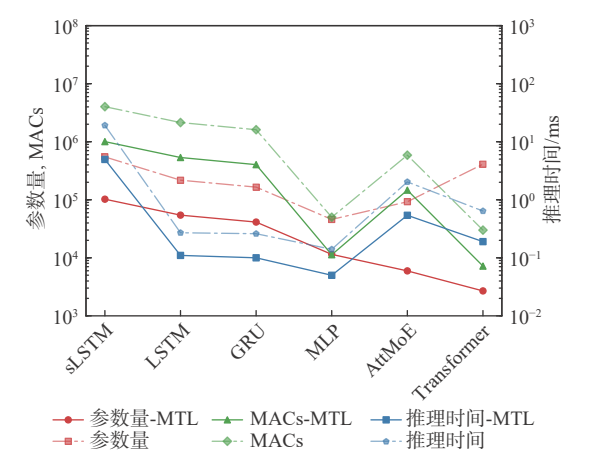


图 14 规模分析对比
Fig. 14 Model scale comparison

4 结 论

1) 实验结果表明, 本文方法在 3 类数据集上均表现出显著优势。如在 eVTOL 数据集中, 相对 AttMoE, MTL-sLSTM 在 RE、MAE 和 RMSE 分别提升约 0.849 7、0.941 2 和 0.923 6。

2) 通过对比实验, 验证了多任务学习策略的有效性。各模型的 RE、MAE 和 RMSE 改进率普遍超过 60%。共享特征提取器采用 sLSTM 显著提升了特征表达能力, 其训练损失较 MTL-LSTM 和 MTL-GRU 降低至少 50%, 且在 NASA 数据子集中, MTL-sLSTM 的 MAE 和 RMSE 指标全面优于对比模型, 进一步证明了该架构的鲁棒性。

3) 模型规模分析表明, 多任务学习在资源利用、模型压缩与推理速度方面具备明显优势。各模型在 MTL 下的 MACs 与参数量普遍下降约 1 个数量级; 推理时间显著小于单任务模型运行 4 次的总时间, 如 AttMoE 从 100 ms 降至 0.1 ms 以下。MTL 架构实现了计算资源、模型规模与推理效率的统一优化。

为进一步增强对低空载具动力电池的 RUL 预测精度和泛化能力, 并解决模型规模与推理开销对

部署性能的制约。下一步研究将结合在线增量学习技术, 动态优化多任务预测头参数, 并探索轻量化部署方案, 以进一步提升复杂场景下的实时预测精度与稳定性。

参考文献 (References)

[1] Mohtat P, Lee S, Siegel J B, et al. Comparison of expansion and voltage differential indicators for battery capacity fade[J]. Journal of Power Sources, 2022, 518: 230714.

[2] Yang J F, Cai Y F, Mi C. Lithium-ion battery capacity estimation based on battery surface temperature change under constant-current charge scenario[J]. Energy, 2022, 241: 122879.

[3] Yang J F, Li X, Sun X D, et al. An efficient and robust method for lithium-ion battery capacity estimation using constant-voltage charging time[J]. Energy, 2023, 263: 125743.

[4] Mitici M, Hennink B, Pavel M, et al. Prognostics for lithium-ion batteries for electric vertical take-off and landing aircraft using data-driven machine learning[J]. Energy and AI, 2023, 12: 100233.

[5] Chang V W, Gallman J W, Richardson M E, et al. Development and test of safe, high power lithium ion main batteries for general aviation aircraft[J]. SAE International Journal of Aerospace, 2008, 1(1): 982-999.

[6] 韩露, 史贤俊, 林云. 航空锂电池寿命预测方法研究[J]. 电子测量技术, 2021, 44(1): 20-25.

Han L, Shi X J, Lin Y. Research on life prediction method of aviation lithium battery[J]. Electronic Measurement Technology, 2021, 44(1): 20-25(in Chinese).

[7] Zraibi B, Okar C, Chaoui H, et al. Remaining useful life assessment for lithium-ion batteries using CNN-LSTM-DNN hybrid method[J]. IEEE Transactions on Vehicular Technology, 2021, 70(5): 4252-4261.

[8] 贺宁, 张思媛, 李若夏, 等. 粒子滤波和 GRU 神经网络融合的锂电池 RUL 预测[J]. 哈尔滨工业大学学报, 2024, 56(5): 142-151.

He N, Zhang S Y, Li R X, et al. RUL prediction of lithium battery based on particle filter and GRU neural network fusion[J]. Journal of Harbin Institute of Technology, 2024, 56(5): 142-151(in Chinese).

[9] 董渊昌, 庞晓琼, 贾建芳, 等. 基于 SVD-SAE-GPR 的锂离子电池 RUL 预测[J]. 储能科学与技术, 2023, 12(4): 1257-1267.

Dong Y C, Pang X Q, Jia J F, et al. Remaining useful life prediction of lithium-ion batteries based on SVD-SAE-GPR[J]. Energy Storage Science and Technology, 2023, 12(4): 1257-1267(in Chinese).

[10] 陈欣, 李云伍, 梁新成, 等. 基于模态分解的 Transformer-GRU 联

- 合电池健康状态估计[J]. 储能科学与技术, 2023, 12(9): 2927-2936.
- Chen X, Li Y W, Liang X C, et al. Battery health state estimation of combined Transformer-GRU based on modal decomposition[J]. Energy Storage Science and Technology, 2023, 12(9): 2927-2936(in Chinese).
- [11] Chen D Q, Zhou X Z. AttMoE: attention with mixture of experts for remaining useful life prediction of lithium-ion batteries[J]. Journal of Energy Storage, 2024, 84: 110780.
- [12] Alharthi M, Mahmood A. xLSTMTIME: long-term time series forecasting with xLSTM[J]. AI, 2024, 5(3): 1482-1495.
- [13] Ragab M, Chen Z H, Wu M, et al. Contrastive adversarial domain adaptation for machine remaining useful life prediction[J]. IEEE Transactions on Industrial Informatics, 2021, 17(8): 5239-5249.
- [14] Zhang Z, Wang Y Y, Ruan X X, et al. Lithium-ion batteries lifetime early prediction using domain adversarial learning[J]. Renewable and Sustainable Energy Reviews, 2025, 208: 115062.
- [15] Chen J X, Li D P, Huang R Y, et al. Aero-engine remaining useful life prediction method with self-adaptive multimodal data fusion and cluster-ensemble transfer regression[J]. Reliability Engineering & System Safety, 2023, 234: 109151.
- [16] Wang H Q, Lin T J, Cui L L, et al. Multitask learning-based self-attention encoding atrous convolutional neural network for remaining useful life prediction[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 3516108.
- [17] Kim T S, Sohn S Y. Multitask learning for health condition identification and remaining useful life prediction: deep convolutional neural network approach[J]. Journal of Intelligent Manufacturing, 2021, 32(8): 2169-2179.
- [18] Zhang Y, Xin Y Q, Liu Z W, et al. Health status assessment and remaining useful life prediction of aero-engine based on BiGRU and MMoE[J]. Reliability Engineering and System Safety, 2022, 220: 108263.
- [19] Park K, Choi Y, Choi W J, et al. LSTM-based battery remaining useful life prediction with multi-channel charging profiles[J]. IEEE Access, 2020, 8: 20786-20798.
- [20] Semeraro C, Caggiano M, Olabi A G, et al. Battery monitoring and prognostics optimization techniques: challenges and opportunities [J]. Energy, 2022, 255: 124538.
- [21] Pecht M. CALCE battery group [EB/OL]. (2025-06-28)[2026-07-16]. <https://doi.org/10.21227/w9rg-7173>.
- [22] Saha B, Goebel K. Battery data set[EB/OL]. (2025-07-18)[2026-07-16]. <http://data.nasa.gov/dataset/li-ion-battery-aging-datasets>.
- [23] Bills A, Sripad S, Fredericks L, et al. A battery dataset for electric vertical takeoff and landing aircraft[J]. Scientific Data, 2023, 10: 344.

Remaining useful life prediction of lithium-ion batteries in low-altitude vehicles based on MTL-sLSTM

ZHANG Huishan¹, SHEN Jiyong², LIU Dongsheng^{1,*}, ZHOU Zhikai³, HU Yifan¹, XU Yangbo¹

(1. School of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou 310020, China;

2. Hangzhou Science and Technology Information Institute, Hangzhou 310053, China;

3. Hangzhou Shuzheng Technology Co., Ltd., Hangzhou 310020, China)

Abstract: A prediction technique based on multi-task learning with scalar long short-term memory (MTL-sLSTM) was presented to solve the problem of remaining useful life (RUL) prediction for lithium-ion batteries in low-altitude economy vehicles under multi-condition coupling. Firstly, a heterogeneous input layer integrates multi-dimensional time-series data from diverse flight conditions. Then, a hierarchical stacked sLSTM structure serves as a shared feature extractor, enabling deep cross-scale feature integration and the capture of common nonlinear degradation patterns. Finally, by hard-coding task IDs to dynamically modulate network weights, the multi-task learning mechanism adaptively identifies each operating condition's unique aging behavior while simultaneously promoting knowledge transfer across domains for independent RUL estimation. According to experimental results, MTL-sLSTM achieves a 60.7%–92.3% reduction in root mean square error (RMSE) on the eVTOL dataset, outperforming six temporal approaches, including the attention mixture of experts (AttMoE) and dual-channel LSTM (Dual-LSTM). This validates the effectiveness of the multi-task learning mechanism in enhancing the generalization capability of degradation features and improving prediction accuracy under complex operating conditions.

Keywords: remaining useful life prediction; multi-task learning; long short-term memory network; lithium-ion batteries; electric vertical takeoff and landing; low-altitude economic vehicles

Received: 2025-04-03; **Accepted:** 2025-04-29; **Published Online:** 2025-06-04 11:09

URL: link.cnki.net/urlid/11.2625.V.20250604.0851.001

Foundation items: Zhejiang Ten Thousand Talents Program (2022R52048); Zhejiang Provincial Science and Technology Program (2024C01028, 2024C01211)

* **Corresponding author.** E-mail: lds1118@zjgsu.edu.cn

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0313

面向不规则结构的低空感知模型剪枝算法

杨子, 庄连生*

(中国科学技术大学 网络空间安全学院, 合肥 230026)

摘 要: 深度学习在低空智能感知中性能优越, 但终端计算资源有限, 难以部署大型神经网络。剪枝技术通过删除冗余参数降低复杂度, 提升实时性, 但现有算法多针对规则结构模型, 难以处理神经架构搜索 (NAS) 得到的不规则结构。提出一种基于依赖图的不规则结构剪枝方案, 以优化模型效率。通过计算图的结构解析技术, 自动识别低空感知模型中特有的多层级连接模式; 构建高效的参数分组模块, 精确锁定可完整移除的最小功能单元; 采用跨层对比的全局优化策略, 实现对同构子结构的统一评估与剪枝。相较于现有算法, 所提算法可对不规则网络架构实施全局精准压缩, 特别适用于需要处理复杂空间关系的低空感知场景。实验结果表明: 所提不规则结构剪枝算法在 NAS-Bench101 模型及 ResNet 系列网络均能有效识别出多分支连接、残差连接、拼接等复杂结构, 并完成分组剪枝, 在 CIFAR-10 数据集以 50% 的压缩比例对模型轻量化, 识别准确率下降不超过 1%。

关 键 词: 低空感知; 卷积网络; 轻量化模型; 模型剪枝; 不规则网络架构

中图分类号: V221+.3; TB553

文献标志码: A

文章编号: 1001-5965(2026)09-3136-10

低空智能感知是基于人工智能、多源传感融合和协同计算等技术, 在低空空域实现对动态目标、环境状态及飞行器行为的全天候、高精度、全方位实时监测与智能分析的综合技术体系。各种新型低空运载工具的广泛应用, 对低空环境感知和低空安全提出了更高要求。深度神经网络在低空动态障碍物的实时感知^[1-4]、轨迹预测^[5-7]等关键场景中展现出卓越性能。由于计算资源受限, 传统大规模神经网络往往难以直接部署到低空飞行器上。模型剪枝通过移除神经网络中的冗余结构, 将部分参数置为非激活状态, 达到简化模型结构、降低计算资源消耗的目的, 使神经网络模型在飞行器终端的部署成为可能, 对于提高系统实时环境感知能力至关重要, 有助于推动低空智能感知技术的产业化进程^[8-10]。

依据剪枝粒度的差异, 模型剪枝可分为非结构化剪枝和结构化剪枝 2 类^[11]。非结构化剪枝主要通过剪除滤波器层内或跨层的连接, 从而有效降低

网络存储成本。然而, 这种方法在降低深度神经网络计算成本方面效果有限。1989 年, Lecun 等提出了最佳脑损伤^[12]理论, 运用二阶导数信息平衡网络复杂度和训练误差, 进而通过移除网络中不重要的权重实现模型优化。Han 等开发了基于 L_1 范数的权重剪枝算法^[13], 通过评估神经元连接权值的范数大小, 删除低于预设阈值的连接。Yang 等^[14]创新性地将层级能量消耗作为剪枝顺序的判定标准, 通过加权稀疏投影和输入遮蔽技术量化能耗指标, 并将能耗预算整合为网络训练的约束条件。Gong 等^[15]提出了一种快速可控的训练后稀疏化框架, 通过引入可微分桥函数和可控优化目标, 解决了传统训练后剪枝算法因忽略分层稀疏率分配导致的精度显著下降问题。需要指出的是, 尽管非结构化剪枝能够有效压缩模型规模, 但其产生的非结构化稀疏特性与现有基础线性代数子程序库的兼容性较差, 导致其难以充分发挥硬件加速优势。

收稿日期: 2025-05-19; 录用日期: 2025-05-30; 网络出版时间: 2025-06-05 14:26

网络出版地址: link.cnki.net/urlid/11.2625.V.20250605.1115.003

* 通信作者. E-mail: yangzi@mail.ustc.edu.cn

引用格式: 杨子, 庄连生. 面向不规则结构的低空感知模型剪枝算法 [J]. 北京航空航天大学学报, 2026, 52 (9): 3136-3145.

Yang Z, Zhuang L S. Low-altitude perception models pruning algorithm with irregular structures [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3136-3145 (in Chinese).

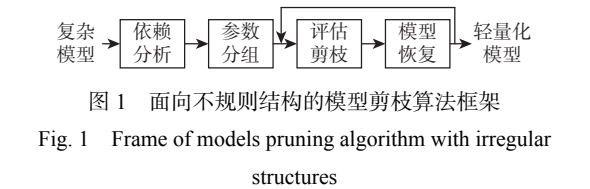
相比之下, 结构化剪枝是一种硬件友好型剪枝方式, 该技术以深度神经网络中滤波器所在的通道为剪枝对象, 通过评估通道重要性实现结构化修剪。因此, 结构化剪枝在降低存储开销的同时, 也能有效提升网络推理速度。相较于非结构化剪枝, 结构化剪枝能够生成结构化的稀疏模型, 既减少设备存储成本, 又可实现硬件层面的实际加速效果。结构化剪枝的核心挑战在于如何有效筛选重要滤波器。现有研究提出了多种滤波器重要性评估算法^[16-19]: Li 等^[16]采用卷积核权重绝对值计算方法, 通过计算滤波器的 L_1 范数筛选目标; Hu 等^[18]提出的剪枝算法从另一个角度进行重要性评估, 通过分析 ReLU 激活函数的输出特性评估神经元重要性; Liu 等^[19]开发了基于批归一化参数的剪枝策略, 利用批归一化层中缩放因子的 L_1 范数作为通道重要性指标; Lin 等提出跨层排序-k-倒数最近滤波器 (cross-layer ranking-k-reciprocal nearest filter, CLR-RNF) 方法^[20], 设计了基于推荐机制的评估体系, 运用 k 近邻算法为每个滤波器构建相似候选集, 再通过多组候选集的交集运算识别关键滤波器。

随着低空环境应用需求的升级, 剪枝技术正面临严峻挑战。为适应复杂多变的低空感知任务, 现代网络架构已突破传统层级堆叠模式, 演化出动态连接和多分支拓扑等新型结构^[21], 这些创新显著增加了模型轻量化的实施难度。在低空智能感知场景中, 模块间高度耦合的特征传递机制使任何参数调整都可能破坏空间关系推理链条, 导致目标识别精度的大幅波动。这些非常规连接结构往往由算法演化生成, 而非人工设计, 导致其可能具有任意拓扑结构和近乎无限的连接复杂度。面对此类网络的剪枝需求时, 常规的逐层裁剪策略已不再适用, 因为随意切断连接会引起模块输入输出维度的突变, 破坏相邻层间的结构兼容性, 最终导致模型失效。现有剪枝算法通常针对固定网络架构设计, 并在标准基准网络上进行验证^[22-24]。当应用于不规则网络架构时, 这些方法往往需要依赖繁琐的手动调整 (如通道对齐或跳跃连接优化), 或引入额外补偿机制 (如动态重参数化技术), 这种适应性缺陷严重制约了模型轻量化算法的实际应用效率。现有工作针对不同类型的神经网络存在着不同的轻量化算法。例如, 针对卷积网络, 研究人员提出了深度可分离卷积方法^[25]; 针对循环神经网络, 研究人员提出了时序动态剪枝算法^[26]; 针对 Transformer 网络, 研究人员提出了 token 稀疏化机制^[27]。但这些方法的设计与对应的网络架构类型密切相关, 缺乏通用性。本质上, 这些方法都是针对规则网络架构

而设计的, 无法解决不规则网络架构的轻量化问题。针对以上挑战, 本文提出基于依赖图的不规则结构剪枝算法, 从依赖关系分组的角度对结构化剪枝算法进行重构, 以实现结构化剪枝算法在不规则结构网络上的适配。通过依赖分析模块对网络计算图进行解析, 动态追踪计算图的数据流路径, 建立跨层连接的依赖子图。第 2 个参数分组模块通过动态追踪前向计算图中的数据流与控制流依赖关系, 自动识别可被完整移除的最小功能单元, 避免剪枝引发的逻辑断裂问题。为进一步提升全局压缩效率, 增加了同构评估剪枝模块, 通过设计跨网络层的子图同构检测算法, 实现对相似拓扑单元的重要性评估与批量移除。与已有算法相比, 本文算法从依赖图分析的角度突破传统剪枝的层内限制, 并通过拓扑单元的动态识别与联合优化, 有效解决了不规则网络中的参数依赖问题。

1 面向不规则结构的模型剪枝算法

针对现有结构剪枝算法依赖人工设计分组方案、难以处理复杂依赖关系的局限性, 本文提出基于依赖图建模的通用结构剪枝框架, 通过构建神经网络层间的拓扑依赖模型实现任意架构的自动压缩, 如图 1 所示。该方案的总体思路在于引入依赖图显式表征跨层计算关系, 具体包含 3 个核心递进模块: ①依赖分析模块, 通过追踪数据流路径, 构建依赖关系模型; ②参数分组模块, 基于图连通性分析自动识别可独立移除的最小子结构; ③评估剪枝模块, 提出组级联合评估准则, 实现子图结构的全局批量剪枝。



1.1 依赖分析模块

依赖分析模块的核心任务是识别神经网络中各层之间的依赖关系, 以构建结构化的依赖图模型, 为后续自动化剪枝提供拓扑关系基础。本节解释依赖图模型的构建原理, 即图的节点代表网络中的部分。一种直观的初步思路是: 以一整层的参数作为一个基本单元, 构建不同层之间的依赖关系图模型。然而, 在实际剪枝过程中, 若采用层级别的构建方式, 所构建出的依赖关系图会产生冲突, 因为一些基本网络模块 (如全连接层) 可能有 2 种不同的剪枝模式, 如剪枝二维权重矩阵的 $w[k,:]$ 和 $w[:,k]$, 这 2 种方案分别会压缩输入和输出的维

度。此外,网络还包含非参数化操作,如残差连接,这也会影响层之间的依赖性。为解决上述问题,本文使用一种新的表示法,将网络 $F(x; \mathbf{w})$ 分解为更精细、更基本的分量,表示为 $F = \{f_1, f_2, \dots, f_L\}$,其中,每个分量 f_i 表示参数化层(如卷积)或非参数化运算(如残差加法)。本文不在层级建模关系,而是专注于层的输入和输出之间的依赖关系,即将分量 f_i 的输入和输出分别表示为 f_i^- 和 f_i^+ 。对于任何神经网络层或模块,最终分解都可以形式化为 $F = \{f_1^-, f_1^+, f_2^-, f_2^+, \dots, f_L^-, f_L^+\}$ 。这种表示方法有助于简化依赖关系建模,并对同一层使用不同的剪枝方案。利用该符号可以将神经网络重新用公式表示,其中,可以区分 2 种主要类型的依赖关系,即层间依赖关系和层内依赖关系,表示为

$$(f_1^-, f_1^+) \leftrightarrow (f_2^-, f_2^+) \cdots \leftrightarrow (f_L^-, f_L^+) \tag{1}$$

式中:“ $\leftrightarrow$ ”表示 2 个相邻层之间的连接性,即一个层的输出作为另一个层的输入。

依赖关系成立的条件总结为 2 种:①层间依赖关系,仅当 i 层和 j 层前后相连时,存在依赖关系 $f_i^- \leftrightarrow f_j^+$,这一条件表示为 $f_i^- \leftrightarrow f_j^+$;②层内依赖关系,仅当 f_i^- 和 f_i^+ 拥有相同的剪枝模式时,存在依赖关系 $f_i^- \leftrightarrow f_i^+$,这一条件表示为 $\text{sch}(f_i^-) = \text{sch}(f_i^+)$ 。

层间的参数依赖显然由网络前向推理的路径依赖所决定。对于相连接的层 $f_i^- \leftrightarrow f_j^+$,两者存在依赖关系,因为 f_i^- 和 f_j^+ 分别代表先前层的输出和后继层的输入,两者对应于网络的同一组中间特征。同一层的参数之间也可能存在依赖性,即层内的参数依赖,如果某一层的输出和输入具有层内依赖,那么在移除参数时必须同时剪枝单个层的输入和输出。

层内依赖模式如图 2 所示。这种依赖性是由该层的计算方式决定的,许多网络层满足这一条件。例如,批量归一化层输入和输出被剪枝时,需要移除的参数相同,表示为 $\text{sch}(f_i^-) = \text{sch}(f_i^+)$ 。相反,卷积层对其输入和输出具有不同的剪枝方案,即 $\mathbf{w}[:,k,:,:] \neq \mathbf{w}[k,:,:,:]$,导致 $\text{sch}(f_i^-) \neq \text{sch}(f_i^+)$,如图 2 所示,在这种情况下,卷积层的输入和输出之间不存在依赖性。

给定上述规则,本文可以形式化地建立依赖关系:

$$D(f_i^-, f_j^+) = \mathbb{1}[f_i^- \leftrightarrow f_j^+] \vee \mathbb{1}[i = j \wedge \text{sch}(f_i^-) = \text{sch}(f_i^+)] \tag{2}$$

式中: $\mathbb{1}$ 为指示函数; $\wedge$ 为逻辑与运算。

式(2)中,右侧第 1 项负责捕捉网络连接产生的跨层关联,第 2 项则用于分析层输入与输出之间因剪枝模式相同而具有的层内依赖性。依赖矩阵

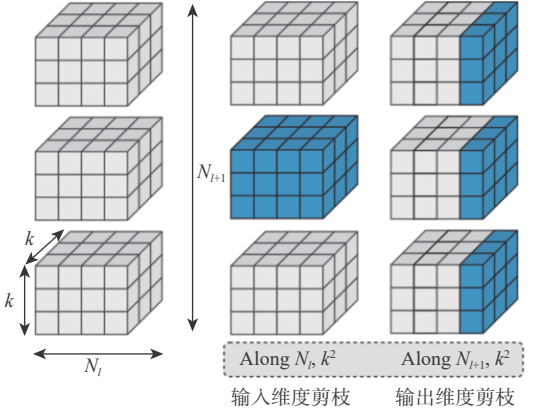


图 2 层内依赖模式示意

Fig. 2 Illustration of inter-layer dependency patterns

具有对称特性,严格遵循 $D(f_i^-, f_j^+) = D(f_j^+, f_i^-)$ 的对称关系。基于这种对称特性,本文可以通过系统遍历所有输入输出配对,来完整精确地构建完整的依赖拓扑模型。图 3 直观展示了带有残差连接的卷积网络块所对应的依赖关系图构建过程。

1.2 参数分组模块

参数分组模块的任务是将具有依赖关系的网络子结构进行分组,确保这些子结构在剪枝时能够被同时处理。为此,本文基于依赖图进行分组,通过将参数划分问题转化为依赖图的连通分量划分问题,将相互依赖的子结构划分为同一组。具体而言,依赖图在构建时,每一个节点都与其对应的参数进行关联,当依赖图的 2 个节点连通时,其对应的参数也被划分到同一个参数分组。对于不同的依赖类型,寻找所需要分组的参数具有不同的处理方法。将常见的依赖关系拆解为若干个基本类型,如图 4 所示,以展开对参数分组模块的详细描述。

从一个无分支神经网络开始解释参数分组方法,如图 4(a)所示,该网络由 3 个连续层组成,网络参数由二维权重矩阵 $\mathbf{w}_l$ 、 $\mathbf{w}_{l+1}$ 和 $\mathbf{w}_{l+2}$ 组成。这种简单的神经网络可以通过去除神经元进行结构化剪枝。在这种情况下,很容易发现参数之间出现了一些依赖关系,表示为 $\mathbf{w}_l \leftrightarrow \mathbf{w}_{l+1}$,表示 $\mathbf{w}_l$ 和 $\mathbf{w}_{l+1}$ 必须同时被剪枝。具体而言,为剪枝桥接 $\mathbf{w}_l$ 和 $\mathbf{w}_{l+1}$ 的第 k 个神经元, $\mathbf{w}_l[k,:]$ 和 $\mathbf{w}_{l+1}[k,:]$ 都将被移除。无分支网络依赖模式如图 5 所示,其中,蓝色部分表示被剪枝的特征图及其对应的参数位置。

残差依赖如图 6 所示。对带有残差连接的残差块进行剪枝时,残差块第 2 层中待剪枝的滤波器必须由残差的剪枝结果决定,而仅有残差块的第 1 层可以无限制地进行剪枝。类似的简单的依赖关系可能会相互嵌套或组合,从而在复杂拓扑结构的网络中形成更为复杂的模式。

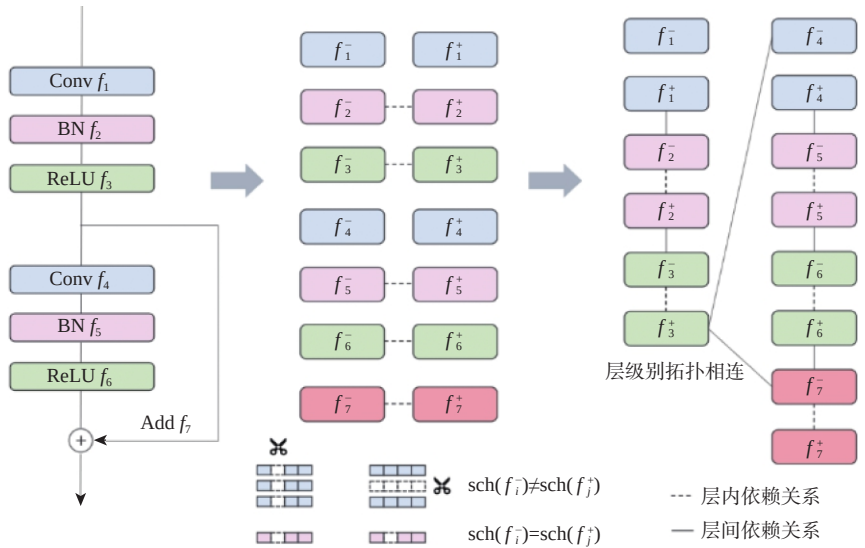


图 3 依赖关系图构建

Fig. 3 Construction of a dependency relationship graph

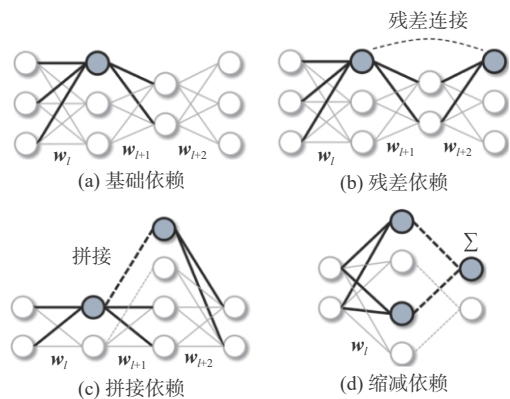


图 4 不同结构中具有相互依赖性的分组参数

Fig. 4 Grouped parameters with dependency in different structures

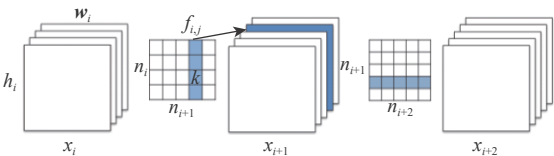


图 5 无分支网络依赖模式示意

Fig. 5 Dependency patterns diagram in branchless networks

在一些需要融合不同特征信息的模型模块中,如图 7 所示。以 U-Net 网络为例,在其编码器-解码器结构中,当从编码器向解码器传递特征时,会存在拼接操作。这种拼接依赖具体表现为:不同层的特征图在通道维度上直接相连,在拼接操作后紧接层的输入维度将由拼接操作前若干层的输出维度共同决定。从数学角度看,若有 2 个特征图 A 和 B ,其维度分别为 (N, C_1, H, W) 和 (N, C_2, H, W) ,经过拼接操作后,得到的新特征图 C 的维度为 $(N, C_1 + C_2, H, W)$,新特征图 C 中的每个通道分别来源于 A 和 B ,这种依赖关系使模型能够综合不同层级的特征,提升对复杂目标的分割或识别能力。

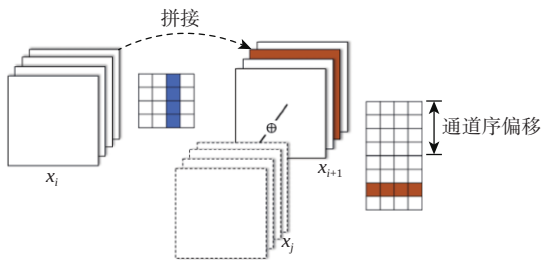


图 7 拼接依赖模式示意

Fig. 7 Dependency pattern diagram with concatenation

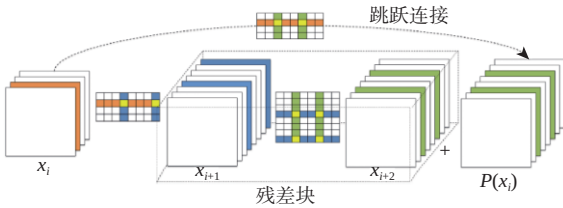


图 6 残差网络依赖模式示意

Fig. 6 Dependency patterns diagram in residual networks

除残差连接导致的额外依赖关系外,不规则结构的神经网络中还存在一些其他依赖关系需要处理。拼接依赖在神经网络架构中较为常见,尤其是

缩减依赖也是现代神经网络中不容忽视的一种依赖形式。在如 SqueezeNet^[28] 这类模型中,为降低计算量和模型参数,会采用挤压模块,该模块由挤压层和扩展层组成,挤压层通过使用 1×1 的卷积核来缩减通道数,从而形成一种缩减依赖。具体而言,假设输入特征图的通道数为 C_{in} ,挤压层的输出通道数为 $C_{squeeze}$ ($C_{squeeze} \ll C_{in}$),挤压层通过对输入特征图在空间维度上进行卷积操作,将每个空间位置的 C_{in} 个通道特征进行线性组合,得到 $C_{squeeze}$ 个通道的输出。这种依赖使后续扩展层的计算量显著

降低,因为扩展层是基于挤压层缩减后的通道数进行操作的。在该过程中,挤压层输出的特征图对后续扩展层的输入有直接的依赖关系。

对于上述各类依赖模式,参数分组模块将其对应的参数,并在计算依赖图的最大连通子分量后,将对应的参数划分为一组,完成所有具有依赖关系的参数分组,以便后续的移除。

1.3 评估剪枝模块

本节评估剪枝模块衡量参数的重要性,并实现冗余参数的移除。与现有算法有所不同,该算法以耦合参数组作为衡量参数重要性的基本单元,这就需要应对以下 2 个关键问题: ①单个耦合参数组中

往往包含来自不同层的参数,而不同类型的网络层在参数特征和计算拓扑结构上存在显著的多样性,使传统方法中衡量参数重要性的计算方式难以直接应用; ②网络内不同层通常呈现出不同程度的冗余性,因此,必须确保不同参数组之间具备可比性,以便进行有效性的重要性评估。为解决上述问题,本文算法提出的主要策略是依据子结构的拓扑逻辑及参数配置情况,将子结构聚类到不同的同构群中。在这种聚类方式下,同构子结构的参数之间呈现出一一对应的关系,且不同参数参与计算的方式完全一致,如图 8 所示。同一同构群内的所有子结构均具有相同的架构设计、参数规模及计算图。

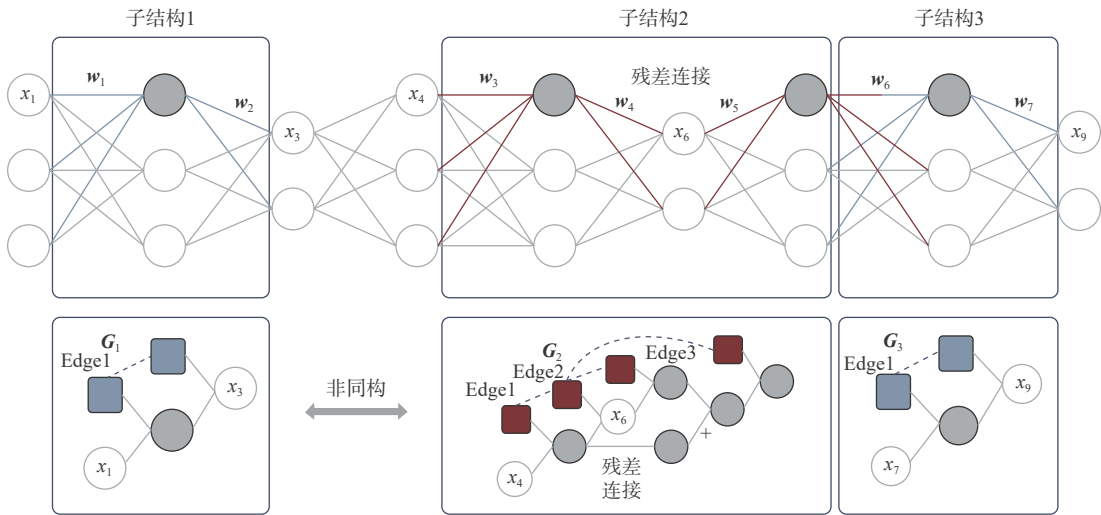


图 8 神经网络中的同构子结构

Fig. 8 Isomorphic substructures in a neural network

对于通过前述方法得到的依赖图 D_1 , 将每个节点由三元组 (W_1, d, k_1) 进行标记, 包含参数矩阵 W_1 、轴 d 及索引 k_1 。标签分配遵循以下原则: 当且仅当参数矩阵 W 源自相同类型的网络层 (如线性层或卷积层) 且具有相同的剪枝维度 $d = d'$ 时, 节点 (W_1, d, k_1) 和 (W', d', k') 才被赋予相同的标签。允许索引 k 不同, 因为在相同参数矩阵中, 沿第 d 维切分的权重向量具有一致的齐次性, 可视为同构单元。按照上述方法确定顶点标签后, 同构参数组要求 G_1 与 G_2 的连接具备同构性。如图 8 所示, 图中的 3 个子结构分别对应图 G_1 、 G_2 和 G_3 , 其中, G_1 和 G_2 各含单边, G_3 则包含 3 条边。若按计算图的执行顺序检查边连接, 可自然观察到 G_1 与 G_3 的同构性, 因其共享相同的拓扑结构。为形式化描述该过程, 将图 G 中的边按前向传播执行顺序排序, 即可通过式 (3) 判定同构性:

$$I(G, G') = \mathbb{1}[|G| = |G'| \wedge \text{label}(V_i) = \text{label}(V'_i), i \in \{1, 2, \dots, |G'|\}] \quad (3)$$

式中: $|G|$ 、 $|G'|$ 为图的边数。

首项条件确保图的规模一致性, 次项要求所有顶点标签严格匹配。在参数组剪枝阶段, 首先, 通过定义重要性函数量化各参数组的显著性。具体而言, 算法会对一组具有相同图拓扑的子结构进行检查。对于每个图的重要性进行独立估计, 进而得到一个多维重要性向量, 这一向量的生成有助于直接对重要性进行排序, 根据排序得分删除不重要的子结构。对于其他同构群, 可依据相同的原理实施剪枝操作, 从而实现对网络模型的有效压缩和优化。例如, 采用幅度剪枝时, 重要性函数为参数组的 L_2 范数; 采用泰勒剪枝时, 则结合梯度与参数的乘积进行二阶近似。其次, 需排除对网络功能完整性敏感的层, 如分类网络的最后一层全连接层, 以防止输出维度缩减导致任务失效。初始化剪枝器时需指定目标稀疏度、迭代次数及示例输入 (用于依赖关系分析)。最后, 在每次剪枝迭代中, 剪枝器通过动态调整通道数实现参数组的移除, 并通过微调恢复模型性能。

1.4 模型恢复

模型经过剪枝后, 其性能通常会有所下降, 这主要是因为移除部分权重破坏了原有的知识表示和模型结构。因此, 必须对剪枝后的模型进行微调, 使模型在剪枝后的新结构基础上, 重新适应并优化其参数配置。本文使用原数据集对剪枝后的模型进行微调, 使用原模型 1/10 的学习率进行微调后, 回到评估剪枝模块进行迭代剪枝。

2 实验结果与分析

为验证算法的有效性, 借助神经架构搜索 (neural architecture search, NAS) 方法在不同数据集上搜索最佳不规则网络架构, 对其进行模型压缩, 并比较压缩前后模型的性能。

2.1 实验设置

本节在 CIFAR-10、CIFAR-100 及 ImageNet 基准数据集上通过实验证明了本文压缩算法的有效性。通过将所有训练图像(不包含测试图像)作为训练数据, 对模型进行训练或微调后, 对测试图像分类, 得到最终的分类准确率。在数据处理方面, 按照前人研究的惯例, 使用通道平均值和标准差对输入数据进行归一化, 并使用包含移位、镜像的标准数据增强方案。本文算法在 Pytorch 平台上实现, 并对当前具有一流性能的卷积网络 ResNet 进行剪枝。为获得每个网络的基准精度, 从头开始训练 ResNet 模型, 并遵循与标准 ResNet 训练过程相同的预处理和超参数。本文对 NAS 生成的模型系列 NAS-Bench101 进行压缩测试, 采用的不规则网络架构来源于 NAS-Bench101 网络架构系列。NAS-Bench101 是专门用于神经网络结构搜索的基准数据库, 其核心设计是通过组合标准模块来构建完整网络。该系列将网络拆分为多个重复的单元模块(称为细胞), 每个细胞由 3~7 个计算节点通过不同路径连接构成。节点间的连接线可选择 3×3 卷积、1×1 卷积或 3×3 最大池化等基础计算操作。通过调整节点数量、连接方式和操作类型, 该基准库共预定义了 42.3 万种有效网络结构, 其深度跨度从 9 层到 277 层, 模型参数规模从 20 万到 3 730 万不等。不同结构的计算量差异显著, 处理单张 CIFAR-10 图像所需浮点运算量最低仅 3×10^7 次, 最高可达 8.9×10^9 次, 展现出丰富的结构多样性。在训练及微调时, 本文均使用以下设置: 使用 Tesla-V100-DGXS-16 GB GPU, 在标准数据集, 同一批处理的图像数量为 128; 使用 Adam 优化器进行优化, 学习率为 0.001, 剪枝加微调共进行不超过 40 轮迭代(一般待压缩网络模型的训练需要数十倍), 并保

存具有最佳验证准确率的压缩后模型作为算法的输出。

与绝大部分研究方法一致, 本文利用浮点运算量和参数量衡量网络复杂度。浮点运算量表示网络在推理过程中所执行的浮点运算的总次数, 用于评估网络的计算量; 参数量则表示网络中权重参数的总量, 用于评估网络的规模。本文利用剪枝比率 (pruning rate, PR) 表示网络在剪枝后的压缩比例, 使用 Top-1 精度指标, 即最高预测概率的类别是否为正确类别, 以评估原始网络和剪枝后网络的性能。

2.2 参数分组可视化

为系统验证神经网络中各层级间的参数依赖关系, 本文将网络依赖矩阵和根据其推理得到的参数分组矩阵进行可视化, 结果如图 9 所示。分组矩阵依据本文所表述的图算法中依赖图推导得出, 其中, $G*[i, j] = 1$ 表示第 i 层与第 j 层属于同一组。从可视化结果可以看出, 无论是 ResNet 还是 NAS-Bench 中, 同一残差块或细胞结构内的各层之间呈现出很强的相关性, 使进行结构化剪枝时形成较大的分组。特别是网络搜索得到的细胞结构内部存在多路径并行的参数依赖, 导致大多数分组跨越超过 4 层, 这与传统链式网络中常见的 2~3 层分组形成鲜明对比。这种现象揭示了 NAS 网络特有的参数耦合特性: 为追求更高的表征能力, NAS 过程会自发形成复杂的跨层连接, 但这种结构特性也使手动剪枝极易破坏网络的完整性。

2.3 对比实验

表 1~表 3 中列出了本文算法和其他算法的对比结果。针对 ResNet 网络进行实验, 表 1 展示了使用 50% 的压缩比例在 CIFAR-10 数据集识别任务上压缩 ResNet56 网络的结果, 表 2 展示了使用 50% 的压缩比例在 CIFAR-100 数据集识别任务上压缩 ResNet101 网络的结果。由于模型参数需要成块地被压缩, 因此, 当达到压缩比例目标时, 实际压缩比例可能略大于 50%。本文算法超过了目前的主流模型压缩算法, 在相同的压缩比例设置下取得了最好的识别率。表 3 列出了本文使用 NAS 产生的高性能模型进行压缩的结果。

表 4 展示了不同压缩比例对模型性能的影响。实验基于 CIFAR-10 数据集开展, 通过本文压缩算法生成不同压缩程度的模型, 除目标压缩比例外, 其余参数设置均与基准实验保持一致。在 10%~80% 的压缩比例范围内, 以 10% 为间隔梯度式调节压缩比例, 最终获得 8 组对比实验数据。

2.4 消融实验

通过系统设计的消融实验验证关键算法组

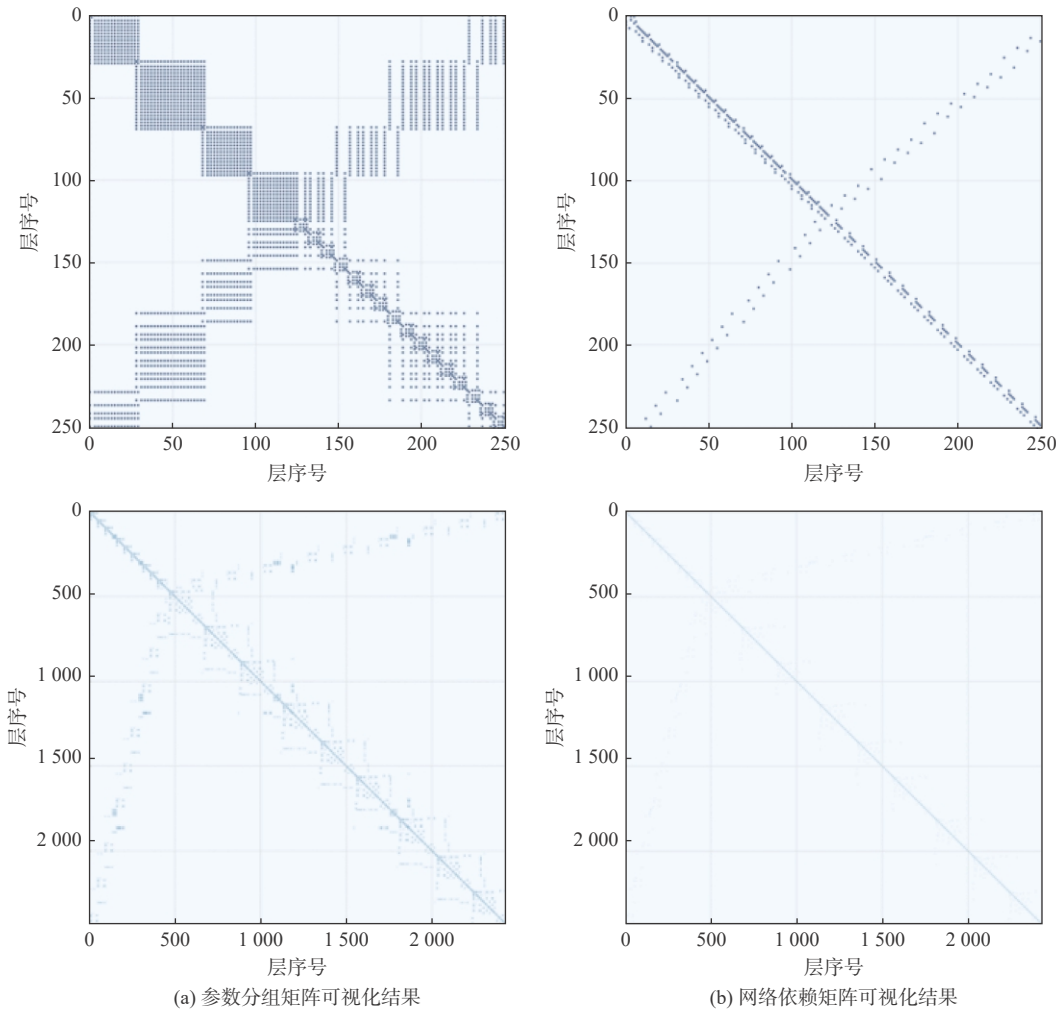


图9 分组矩阵可视化

Fig. 9 Group matrix visualization

表1 在 CIFAR-10 数据集上的识别准确率对比 (ResNet56)

Table 1 Comparison of recognition accuracy rate on CIFAR-10 dataset (ResNet56)			
算法	参数量	压缩比例/%	识别准确率/%
不压缩	2.352×10^7	0	94.90
L_1 -norm	1.151×10^7	51.06	93.37
L_2 -norm	1.151×10^7	51.06	93.86
FPGM	1.151×10^7	51.06	93.64
Taylor	1.151×10^7	51.06	93.80
本文	1.151×10^7	51.06	94.93

表2 在 CIFAR-100 数据集上的识别准确率对比 (ResNet101)

Table 2 Comparison of recognition accuracy rate on CIFAR-100 dataset (ResNet101)			
算法	参数量	压缩比例/%	识别准确率/%
不压缩	4.270×10^7	0	78.30
L_1 -norm	2.094×10^7	50.96	74.75
L_2 -norm	2.094×10^7	50.96	75.59
FPGM	2.094×10^7	50.96	74.45
Taylor	2.094×10^7	50.96	75.42
本文	2.094×10^7	50.96	76.68

表3 在 CIFAR-10 数据集上的识别准确率对比 (NAS)

Table 3 Comparison of recognition accuracy rate on CIFAR-10 dataset (NAS)			
算法	参数量	压缩比例/%	识别准确率/%
不压缩	4.30×10^6	0	96.97
L_1 -norm	2.14×10^6	50.24	95.81
L_2 -norm	2.14×10^6	50.24	95.50
FPGM	2.14×10^6	50.24	95.61
Taylor	2.14×10^6	50.24	95.77
本文	2.14×10^6	50.24	95.90

表4 不同压缩比例下的性能对比

Table 4 Performance comparison at different compression ratios			
压缩比例/%	参数量	浮点运算量/次	识别准确率/%
10	2.108×10^7	5.717×10^{10}	95.10
20	1.856×10^7	5.021×10^{10}	95.14
30	1.635×10^7	4.433×10^{10}	94.92
40	1.394×10^7	3.786×10^{10}	95.01
50	1.151×10^7	3.119×10^{10}	94.93
60	9.210×10^6	2.511×10^{10}	94.61
70	7.460×10^6	2.036×10^{10}	94.57
80	4.550×10^6	1.241×10^{10}	94.03

件。在分组策略验证中,依次测试了无分组独立剪枝、卷积层分组剪枝和全参数分组剪枝 3 种模式,保持其他实验条件一致;在稀疏性调控实验中,采用相同的总体压缩比例约束,对比了固定比例均匀剪枝和动态分配全局剪枝 2 种策略。所有实验均在标准数据集和基准模型上进行,以确保实验结果的可靠性。

网络各部分的稀疏性设置是剪枝过程中的重要因素,其决定了剪枝后神经网络的最终结构。为探究稀疏性设置对结构化剪枝的影响,以 NAS-Bench101、ResNet 为基准模型,在 CIFAR-10 数据集上对比均匀稀疏性、全局稀疏性 2 种稀疏性策略。均匀稀疏性对网络所有层施加相同 PR,而全局稀疏性根据重要性评分动态调整各层 PR,但保证总体剪枝率与均匀稀疏性一致。本文中的稀疏性为通道稀疏性,其中,均匀稀疏性表示每一层的通道数缩减比例,而全局稀疏性表示网络中的通道数总和在压缩后的缩减比例。比较不同稀疏性策略下模型在轻量化后的准确率,表 5 给出了关于层稀疏性的实验结果。

表 5 不同分层裁剪策略的效果对比

Table 5 Comparison of the effectiveness of different hierarchical cutting strategies

网络架构	分组策略	压缩比例/%	识别准确率/%
ResNet56	全局稀疏	10	92.91
		30	91.05
		50	88.93
	均匀稀疏	10	92.59
		30	91.91
		50	90.84
NAS-Bench101	全局稀疏	10	95.81
		30	92.66
		50	91.05
	均匀稀疏	10	96.29
		30	96.10
		50	95.61

观察表 5 可以得出,在大多数实验场景下,全局稀疏性展现出了相较于均匀稀疏性更为出色的性能表现。通过全局统筹的剪枝策略,能够更精准地识别并移除网络中真正冗余的连接和参数,从而在有效减少模型计算量和存储需求的同时,最大程度地维持模型的精度。在部分实验案例中,本文也观察到其存在一定的局限性,由于其全局化的决策机制,有可能会对某些特定层进行过度剪枝,这些层可能在模型的整体架构中承担着关键的功能,尽管从全局角度看,其部分连接或参数看似冗余,但

实际上对模型精度有着重要影响。一旦对这些关键层进行过度剪枝,就会破坏模型的信息流传递,导致模型对输入数据的特征提取和理解能力下降,进而引发精度的显著下降。

3 结 论

1) 针对传统剪枝算法依赖人工经验、难以泛化至低空无人机机载设备中不规则神经网络架构的问题,本文提出基于依赖图分析的结构化剪枝算法,通过形式化建模网络参数间的拓扑依赖关系,实现了复杂结构(如残差连接、跨层跳跃)的自动化分组与剪枝。实验表明,在通用 ResNet 模型中,本文算法相较人工策略更优,且在 NAS-Bench101 不规则网络上,可以在实现通用剪枝的同时,保持较好的综合性能。

2) 本文方法的核心创新在于:首先,针对神经网络中可能存在的多分支、多类型依赖关系,根据前向计算推导依赖图建立依赖传播机制,支持多分支结构的耦合参数快速分组;其次,在参数组重要性判断时,通过图同构判定,实现异构子结构的可比性建模,实现跨模块重要性的全局比较。

3) 在 NAS-Bench101 模型及 ResNet 系列网络中,本文算法能够精准识别多分支连接、残差连接、拼接等复杂结构,并顺利实施分组剪枝操作,对网络实施 50% 的压缩比例进行轻量化处理时,模型的识别准确率下降幅度控制在 1% 以内。实验结果表明,本文提出的不规则结构剪枝算法具备通用性和有效性。

参考文献 (References)

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

[2] Wang H, Wang H, Liu Y J, et al. Real-time obstacle perception method for UAVs with an RGB-D camera in low-light environments[J]. Signal, Image and Video Processing, 2025, 19(3): 256.

[3] El Ghazouali S, Mhirit Y, Oukhrd A, et al. FusionVision: a comprehensive approach of 3D object reconstruction and segmentation from RGB-D cameras using YOLO and fast segment anything[J]. Sensors, 2024, 24(9): 2889.

[4] Saha A, Dhara B C. 3D LiDAR-based obstacle detection and tracking for autonomous navigation in dynamic environments[J]. International Journal of Intelligent Robotics and Applications, 2024, 8(1): 39-60.

[5] Wang C H, Wang Y C, Xu M Z, et al. Stepwise goal-driven networks for trajectory prediction[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 2716-2723.

[6] Li R C, Katsigiannis S, Kim T K, et al. BP-SGCN: behavioral

pseudo-label informed sparse graph convolution network for pedestrian and heterogeneous trajectory prediction[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(8): 14566-14580.

[7] Xin G P, Chu D F, Lu L P, et al. Multiagent trajectory prediction with difficulty-guided feature enhancement network[J]. IEEE Robotics and Automation Letters, 2025, 10(3): 2582-2589.

[8] Cai Y J, Ge L H, Liu J, et al. Exploiting spatial-temporal relationships for 3D pose estimation via graph convolutional networks[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2020: 2272-2281.

[9] Cao Z, Simon T, Wei S H, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 1302-1310.

[10] McNally W, Vats K, Wong A, et al. Rethinking keypoint representations: modeling keypoints and poses as objects for multi-person human pose estimation[C]//Proceedings of the Computer Vision-EC-CV 2022. Berlin: Springer, 2022: 37-54.

[11] 姜晓勇, 李忠义, 黄朗月, 等. 神经网络剪枝技术研究综述[J]. 应用科学学报, 2022, 40(5): 838-849.

Jiang X Y, Li Z Y, Huang L Y, et al. Review of neural network pruning techniques[J]. Journal of Applied Sciences, 2022, 40(5): 838-849(in Chinese).

[12] Lecun Y, Denker J S, Solla S A. Optimal brain damage[J]. Advances in Neural Information Processing Systems, 1989, 2: 598-605.

[13] Han S, Mao H Z, Dally W J. Deep compression: compressing deep neural network with pruning, trained quantization and Huffman coding[EB/OL]. (2016-02-15)[2025-05-10]. <https://arxiv.org/abs/1510.00149>.

[14] Yang H C, Zhu Y H, Liu J. Energy-constrained compression for deep neural networks via weighted sparse projection and layer input masking[EB/OL]. (2019-06-02)[2025-05-10]. <https://arxiv.org/abs/1806.04321>.

[15] Gong R H, Yong Y, Wang Z N, et al. Fast and controllable post-training sparsity: learning optimal sparsity allocation with global constraint in minutes[C]//Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence. New York: ACM, 2024: 12190-12198.

[16] Li H, Kadav A, Durdanovic I, et al. Pruning filters for efficient ConvNets[EB/OL]. (2017-03-10)[2025-05-10]. <https://arxiv.org/abs/1608.08710>.

[17] 刘宇, 雷雪梅. 融合 MobileNetV3 特征的结构化剪枝方法[J]. 上海交通大学学报, 2023, 57(9): 1203-1213.

Liu Y, Lei X M. A structured pruning method integrating characteristics of MobileNetV3[J]. Journal of Shanghai Jiao Tong University, 2023, 57(9): 1203-1213(in Chinese).

[18] Hu H Y, Peng R, Tai Y W, et al. Network trimming: a data-driven neuron pruning approach towards efficient deep architectures[EB/OL]. (2016-07-12)[2025-05-10]. <https://arxiv.org/abs/1607.03250>.

[19] Liu Z, Li J G, Shen Z Q, et al. Learning efficient convolutional networks through network slimming[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2755-2763.

[20] Lin M B, Cao L J, Zhang Y X, et al. Pruning networks with cross-layer ranking & k-reciprocal nearest filters[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(11): 9139-9148.

[21] Akimoto Y, Shirakawa S. Analysis of search space design for neural architecture search with weight sharing[C]//Proceedings of the 2024 International Joint Conference on Neural Networks. Piscataway: IEEE Press, 2024: 1-8.

[22] Wang H, Qin C, Bai Y, et al. Why is the state of neural network pruning so confusing on the fairness, comparison setup, and trainability in network pruning[EB/OL]. (2023-02-21)[2025-05-10]. <https://arxiv.org/abs/2301.05219>.

[23] Feng Q T, Xu K, Li Y H, et al. Edge-wise one-level global pruning on NAS generated networks[C]//Proceedings of the Pattern Recognition and Computer Vision. Berlin: Springer, 2021: 3-15.

[24] Jiang M N, Wang J C, Eldebiky A, et al. Class-aware pruning for efficient neural networks[C]//Proceedings of the 2024 Design, Automation & Test in Europe Conference & Exhibition. Piscataway: IEEE Press, 2024: 1-6.

[25] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 1800-1807.

[26] Chatzikonstantinou C, Konstantinidis D, Dimitropoulos K, et al. Recurrent neural network pruning using dynamical systems and iterative fine-tuning[J]. Neural Networks, 2021, 143: 475-488.

[27] Wei S Y, Ye T Z, Zhang S, et al. Joint token pruning and squeezing towards more aggressive compression of vision Transformers[C]// Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 2092-2101.

[28] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size[EB/OL]. (2016-11-04)[2025-05-10]. <https://arxiv.org/abs/1602.07360>.

Low-altitude perception models pruning algorithm with irregular structures

YANG Zi, ZHUANG Liansheng*

(School of Cyber Science and Technology, University of Science and Technology of China, Hefei 230026, China)

Abstract: Large neural networks are challenging to implement on terminal devices due to restricted processing resources, notwithstanding deep learning's superior performance in low-altitude intelligent sensing. Pruning techniques reduce model complexity by eliminating redundant parameters, thereby improving real-time performance. However, existing methods primarily target regularly structured models and struggle to handle irregular structures generated by neural architecture search (NAS). To address this, a dependency graph-based pruning scheme for irregular structures is proposed to optimize model efficiency. The method first employs structural parsing techniques of computational graphs to automatically identify unique multi-level connection patterns in low-altitude perception models. Subsequently, it constructs an efficient parameter grouping module to precisely locate minimum functional units that can be completely removed. Finally, a cross-layer comparison global optimization strategy is adopted to implement unified evaluation and pruning of isomorphic substructures. This solution overcomes the technical constraints of structural adaptability when compared to current approaches, allowing global precision compression for irregular network architectures. This feature makes it especially appropriate for low-altitude perception scenarios that call for intricate spatial relationship processing. Experimental results demonstrate that the proposed irregular structure pruning method can effectively identify and perform group pruning on complex structures, including multi-branch connections, residual connections, and concatenations, in both NAS-Bench101 models and ResNet series networks. When implementing 50% compression rate network lightweighting on the CIFAR-10 dataset, the recognition accuracy drop does not exceed 1%.

Keywords: low-altitude perception; convolutional networks; lightweight models; model pruning; irregular network architectures

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0357

低空通信环境下基于双层博弈的LCU级视频比特分配算法

张雅霏^{1,2}, 刘露露¹, 于金池^{1,2}, 朱迪锋^{1,3,*}, 龚轩¹, 高毅¹

(1. 北京航空航天大学杭州创新研究院, 杭州 310051; 2. 北京航空航天大学软件学院, 北京 100191;

3. 浙江越秀外国语学院, 绍兴 312030)

摘 要: 针对低空通信场景下视频编码效率低与码率利用率不足的问题, 提出一种基于双层混合博弈模型的最大编码单元 (LCU) 级比特分配 (BA) 优化 (HGLA) 算法。该算法以 LCU 为基本单元, 依据纹理复杂度、运动强度与边缘信息进行区域划分, 并构建高、低复杂度区域的差异化 BA 机制, 其中, 高复杂度区域通过非合作博弈与合作博弈相结合实现比特协调分配, 低复杂度区域引入空间加权机制以提高码率利用率, 并采用全局量化参数 (QP) 平滑策略增强视觉连续性。在 4 种低码率场景下进行实验验证, 所提算法在峰值信噪比 (PSNR)、结构相似性指数 (SSIM) 和主观质量方面均取得显著提升, 特别是在目标码率为 16 Kbit/s 时, 相比高效视频编码 (HEVC) 模型采用的码率控制 (RC) 算法 HM-RC、基于帧级多层感知机 (MLP) 的 BA 算法、基于简化立方率失真模型的帧内 RC 算法、基于通用视频编码 (VVC) 的变分辨率 RC 算法 4 种典型算法, PSNR、SSIM 和主观意见得分 (MOS) 分别平均提升了 3.34%、2.16% 和 9.8%, 表明其在码率利用效率和图像质量保持方面具有良好的实用性与适应性。

关键词: 码率控制; 低码率; 博弈论; 比特分配; 低空视频可靠传输

中图分类号: V221⁺.3; TB553

文献标志码: A

文章编号: 1001-5965(2026)09-3146-07

随着低空经济的快速发展, 无人机等小型飞行器在低空领域的应用场景日益丰富。然而, 高移动性及空地网络中信号干扰的加剧导致低空通信场景下的网络带宽波动较大, 并在一定时间范围内表现出带宽资源受限的问题。近年来, 无人机已广泛应用于智慧城市的公共安全、综合治理、应急响应、城市交通管理和环境监测等领域^[1]。视频作为继文本、图像和音频出现之后的重要信息载体, 在低空平台中的作用尤为突出^[2], 同时, 在无人机领域同样展现出不可或缺的重要性。无人机依靠视频传输实现远程感知与实时回传, 支持监控、侦察和搜救等关键任务。然而, 受限于复杂多变的低空通

信环境, 尤其是在带宽受限的条件下, 实现低码率下的高质量视频传输仍是一项关键挑战。

解决上述问题的方法之一是码率控制 (rate control, RC) 中的比特分配 (bit allocation, BA) 算法。RC 在视频编码中起着关键作用, 其目标是在给定带宽条件下, 优化 BA 和量化参数 (quantization parameter, QP) 选择, 从而实现码率与视觉质量之间的有效平衡, 并确保 BA 符合目标码率的要求。RC 共有 2 个关键步骤: BA 和 QP 选择^[3]。其中, BA 作为 RC 的关键步骤, 旨在依据视频内容复杂度和重要性, 在不同编码单元间合理分配有限比特资源。目前, BA 主要在图像组 (group of pictures,

收稿日期: 2025-06-09; 录用日期: 2025-08-15; 网络出版时间: 2025-09-03 11:40

网络出版地址: link.cnki.net/urlid/11.2625.V.20250903.0846.001

基金项目: 浙江省“尖兵”“领雁”研发攻关计划 (2024C01079)

*通信作者. E-mail: zhudf@buaa.edu.cn

引用格式: 张雅霏, 刘露露, 于金池, 等. 低空通信环境下基于双层博弈的 LCU 级视频比特分配算法 [J]. 北京航空航天大学学报, 2026, 52 (9): 3146-3152. Zhang Y F, Liu L L, Yu J C, et al. LCU-level video bit allocation algorithm based on two-layer game in low-altitude communication environment [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3146-3152 (in Chinese).

GOP)、帧和最大编码单元(largest coding unit, LCU)等编码层级上进行,LCU 级别由于具备更细粒度的控制能力,在低码率场景下尤显重要。

在带宽资源有限的条件下,BA 问题本质上是资源博弈问题。博弈论是解决资源分配问题的有效建模策略,为资源竞争建模提供了数学框架,在解决资源分配和冲突优化中具有天然优势,因此,近年来被广泛引入到视频编码领域^[4]。

针对博弈论与 BA 算法的融合,相关文献^[4-12]已开展了大量的实验研究及优化技术理论研究。大多数研究采用合作博弈模型以实现公平资源分配,或采用非合作博弈模型以获取部分关键区域的性能最大化。但在低空应用场景中,若视频内容复杂且码率资源紧张,仅依赖单一博弈策略难以取得理想效果。合作博弈可能导致比特浪费,非合作博弈则可能引发资源倾斜,影响整体主观质量。

基于上述多维角度分析及综合考量,本文提出基于双层混合博弈模型的 LCU 级 BA 优化(LCU-level BA optimization based on two-layer hybrid game model, HGLA)算法,以提升低码率条件下的视频编码质量。该算法从 LCU 复杂度出发,构建 2 级博弈结构,首先,根据综合复杂度因子将 LCU 划分为高复杂度区域和低复杂度区域;其次,在区域间构建第一层非合作博弈模型,通过设计效用函数,提升高复杂度区域所获比特;最终,在高复杂度区域内部采用合作博弈机制,按复杂度公平分配比特。本文算法在保持整体公平性的同时兼顾关键区域质量,有效提升了高动态、低带宽链路下的编码效率和视频主观质量。

1 相关工作

高效视频编码(high efficiency video coding, HEVC)的最终目的是在给定的带宽内为用户提供最佳的视频质量。为使编码码率与带宽保持更好的一致性,HEVC 中使用 RC 模块来保证视频质量。RC 技术的基本原理是综合考虑编码视频内容特性、信道带宽大小、缓存器状态等因素,给编码视频选择一组最优的 QP。使用该组 QP 编码视频后产生的码率最接近信道带宽大小,对应的重建视频质量最高^[13]。RC 模块通常被分解为 2 步:BA 和编码参数的确定^[14]。其中,BA 算法是 RC 中的关键部分,BA 算法通常根据不同编码层次进行划分,一般分为 GOP 层、帧层及 LCU 层。但相对而言,针对 LCU 层的研究较多,LCU 层可以结合局部复杂度指标实现更细粒度的比特资源调配。在 LCU 级别的 BA 研究中,许多学者提出了结合人眼视觉、

纹理特征、运动信息等指标进行优化的算法,例如,崔士明等^[15]针对人眼感兴趣区域(region of interest, ROI)等图像内容特征,设计了 BA 算法;Liu 等^[16]基于多目标优化算法,兼顾失真最小化、质量波动控制和码率约束 3 个方面的需求,提出 BA 算法;He 等^[17]提出一种考虑纹理和运动的时空信息的 BA 模型,从而提升编码效率。这些算法在一定程度上提升了视频质量和比特利用率^[18],但也存在一定的局限性,如对突变场景响应不足,以及在低码率条件下,非 ROI 可能存在质量严重下降等问题,同时,多目标优化算法计算复杂度高,难以适应实时编码需求。尤其在低空通信场景中,网络带宽资源有限且易受干扰波动,使现有算法的不足更加突出。图 1 展示了低空场景下低码率视频传输的典型场景。

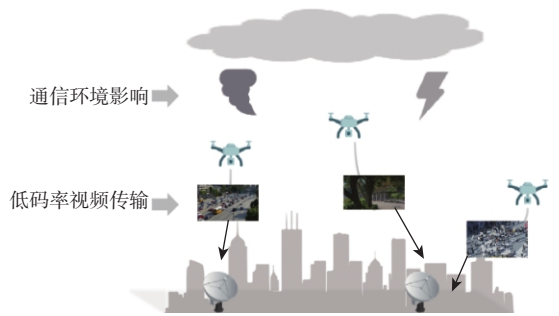


图 1 低空通信环境下低码率视频传输
Fig. 1 Low bit rate video transmission in low-altitude communication environment

近年来,博弈论逐渐被引入到 BA 问题建模中,以期通过模拟编码单元间的“竞争”与“协同”,实现资源的动态高效分配。一般来说,博弈论有 2 种不同的类型:非合作博弈^[19]和合作博弈^[5,20]。在较早时期,Ahmad 等^[6]首次提出将博弈模型用于 BA,验证了其在资源调度中的适配性。后续研究则在不同编码层级探索合作与非合作博弈模型的应用。例如,Gao 等^[10]将每个编码树单元(coding tree unit, CTU)视为一个独立博弈体,基于帧内竞争分配比特资源;Zhao 等^[12]基于全景图像的特征,将博弈论引入到 360°视频编码的 RC 中,并逐波段分配比特率;Liu 等^[4]将帧划分为吸引人较多的 ROI 和非 ROI,并结合非合作博弈进行 BA。尽管已有研究表明博弈理论在 RC 问题中的有效性,但在面对极低码率的极端场景下,这些方法仍显不足,难以实现资源最优分配和视频质量保障。

在前期工作阶段,通过文献[2,21],运用主机状态感知的码率自适应算法和基于双尺度时间策略的带宽预测(dual-scale temporal strategy based bandwidth prediction, DTS-BWpredictor)算法,成功获取了低码率条件下的码率自适应值。这些研究成果不仅为

本文奠定了坚实基础,也为实验分析提供了可靠的数据支持。此外,本文使用的实验数据均源自文献[2, 21]的实验结果,确保了研究的严谨性和科学性。

2 算法设计

在低码率的视频编码研究中,现有 LCU 层 BA 算法普遍存在 2 个方面不足:①区域特征建模缺乏针对性,现有算法往往忽视不同 LCU 在纹理复杂度和感知敏感度上的显著差异,导致比特资源配置与局部主观质量需求不匹配;②已有算法大多依赖启发式规则或单一博弈机制,缺乏区域间的竞争关系及区域内部的协同,难以在全局码率约束下实现比特利用率与视觉质量的双重最优。针对上述问题,本文构建了 HGLA 算法,根据博弈论进行建模,将不同区域间比特资源的动态竞争建模为非合作博弈,同时,在区域内部引入合作博弈机制以实现资源协同分配,并结合 QP 平滑策略以保证质量一致性。整体算法流程包含 LCU 复杂度计算与区域划分、区域间比特竞争的非合作博弈、区域内比特协同的合作博弈及空间领域内 QP 值平滑处理,具体过程如图 2 所示。

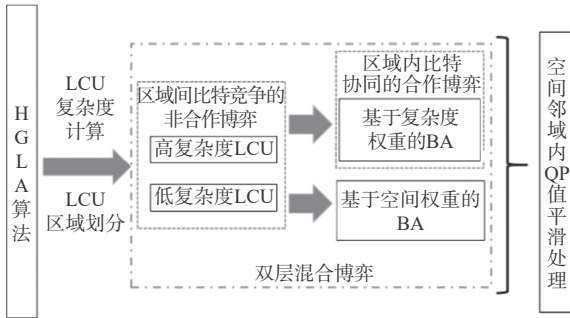


图 2 HGLA 算法

Fig. 2 HGLA algorithm

2.1 LCU 复杂度建模与区域划分

每一帧内所有 LCU 的划分是后续进行双层混合博弈 BA 的基础和依据。视频图像帧被划分为若干 LCU 后,对每个 LCU 的内容复杂度进行建模。为充分体现亮度、纹理与运动等维度特征,本文引入多尺度加权机制构建 LCU 综合复杂度指标,在 HEVC 中 LCU 的尺寸为 64×64 。

2.1.1 LCU 复杂度计算

将大小为 $N \times M$ 的 LCU 划分为若干大小的 LCU 单元,对每个 LCU 构建其综合复杂度指标 C_i ,综合反映其结构细节和运动信息。本文采用如下加权模型进行建模:

$$C_i = \omega_1 \sigma_i^2 + \omega_2 G_i + \omega_3 M_i \quad (1)$$

式中: C_i 为该帧内第 i 个 LCU 的综合复杂度; ω_1 、 ω_2 和 ω_3 为经验值权重,设置为 $\omega_1=0.4$, $\omega_2=0.3$, $\omega_3=0.3$; σ_i^2 为该 LCU 的灰度方差,反映亮度变化幅度,描述为

$$\sigma_i^2 = \frac{1}{NM} \sum_{(x,y) \in L_i} (Y(x,y) - \mu_i)^2 \quad (2)$$

其中: (x,y) 为第 i 个 LCU 在当前帧中的位置坐标; L_i 为第 i 个 LCU; $Y(x,y)$ 为像素值; μ_i 为第 i 个 LCU 的平均亮度值。

G_i 为第 i 个 LCU 的梯度强度均值,用于衡量纹理复杂度:

$$G_i = \frac{1}{NM} \sum_{(x,y) \in L_i} \sqrt{G_x^2 + G_y^2} \quad (3)$$

式中: G_x 、 G_y 分别为 x 、 y 轴方向梯度值。

当前 LCU 与上一帧对应位置 LCU 的平均绝对差 M_i 为

$$M_i = \frac{1}{NM} \sum_{(x,y) \in L_i} |Y_i(x,y) - Y_{t-1}(x,y)| \quad (4)$$

式中: t 表示当前帧; $t-1$ 表示当前帧的前一帧。

2.1.2 LCU 复杂度区域划分

根据 2.1.1 节 LCU 综合复杂度的计算方式,计算当前帧内所有 LCU 的 C_i ,并按照复杂度进行降序排序,取前 30% 的 LCU 作为高复杂度区域 $\mathcal{H}$,则其余区域为低复杂度区域 $\mathcal{L}$,表示为

$$\begin{cases} \mathcal{H} = \{L_i | C_i \geq \tau\} \\ \mathcal{L} = \{L_i | C_i < \tau\} \end{cases} \quad (5)$$

式中: τ 为复杂度分布的 70% 分位数阈值。

通过以上计算,可以得到当前帧内所有 LCU 的分区。

2.2 双层混合博弈建模

在资源受限的编码场景中,不同复杂度区域对比特资源的需求存在显著差异。为在当前帧全局优化的框架下合理分配编码比特,本文采用分区级别的非合作博弈模型对高、低复杂度区域间的 BA 进行建模,而高复杂度区域内部采用合作博弈策略进一步细化其比特资源分配,以提升整体压缩效率。

2.2.1 区域间非合作博弈建模

设总比特预算为 R_T ,高、低复杂度区域请求的比特资源分别为 R_H 、 R_L ,且满足以下条件:

$$R_H + R_L \leq R_T \quad (6)$$

为综合编码质量与系统成本,定义两者效用函数分别为

$$U_H = \omega_H (Q_{\max} - D_H(R_H)) - \lambda_H P(R_H) - \mu_H \frac{R_H}{R_H + R_L} \quad (7)$$

$$U_L = \omega_L (Q_{\max} - D_L(R_L)) - \lambda_L P(R_L) - \mu_L \frac{R_L}{R_H + R_L} \quad (8)$$

式中: $Q_{\max}$ 为最大图像质量的理想基准值; 下标 H、L 分别表示高、低复杂度区域; $D(R) = \alpha R^{-\beta}$ 为经验失真函数, α 、 β 为拟合参数, $P(R) = \gamma R$ 为比特资源成本函数, γ 为单位比特成本; ω 为区域复杂度的权重, 且 $\omega_H > \omega_L$, 即高复杂度区域更重视质量; λ 为资源惩罚系数; μ 为公平性控制项, 用于防止资源极端倾斜。

式 (7) 与式 (8) 类似, 主要关注质量的提升、资源惩罚及博弈中的公平性约束。

高复杂度区域更重视画质, 对资源使用更宽松, 即 λ_H 应较小。低复杂度区域偏向节省资源, 对资源消耗更敏感, 即 λ_L 应更大, 根据经验值设置 $\lambda_H=0.05$, $\lambda_L=0.1$ 。公平性控制项 μ 抑制某一方极端占用比特, 提高博弈公平性, 用于防止资源极端倾斜。高复杂度区域更强势, 若不给足够 μ_H , 容易主导比特资源分配, 因此, 设置 $\mu_H=0.8$, $\mu_L=1.2$, 使低复杂度区域有更多的博弈空间。

上述效用函数重点关注非合作博弈编码系统中的三重权衡: 编码质量、资源成本和博弈公平性。各区域参与者根据自身效用函数, 在竞争总比特预算 R_T 的同时, 最大化自身收益。

求解该博弈模型采用纳什均衡求解策略, 将 U_H 、 U_L 分别对 R_H 与 R_L 求偏导可得

$$\frac{\partial U_H}{\partial R_H} = -\omega_H \frac{dD_H}{dR_H} - \lambda_H \gamma - \mu_H \frac{R_L}{(R_H + R_L)^2} = 0 \quad (9)$$

$$\frac{\partial U_L}{\partial R_L} = -\omega_L \frac{dD_L}{dR_L} - \lambda_L \gamma - \mu_L \frac{R_H}{(R_H + R_L)^2} = 0 \quad (10)$$

对式 (9) 和式 (10) 联立, 并结合约束条件 (式 (6)), 求解最优 BA (R_H^*, R_L^*)。

2.2.2 高复杂度区域内的合作博弈

在获取区域间非合作博弈得到的比特预算 R_H 后, 高复杂度区域内部需要进一步细化分配策略, 以保证每个 LCU 在公平和效率之间达成平衡。高复杂度区域内部的 LCU 由于具备相似的编码挑战, 其目标一致, 即提高整体编码质量, 因此, 可以通过合作博弈建模该过程。合作博弈将每个参与者组建联盟, 协同优化集体利益, 并基于各自对整体的贡献进行资源分配。在高复杂度区域内部, 对所有 LCU 通过合作博弈共同争取区域内分配的比特预算 R_H 。将各个 LCU 视为合作联盟的成员, 共同目标为最大化总体重建质量:

$$R_i = \frac{C_i}{\sum_{j \in \mathcal{H}} C_j} R_H^* \quad i \in \mathcal{H} \quad (11)$$

通过归一化 LCU 的复杂度, 对所有 LCU 进行线性加权分配, 式 (11) 可以保证每个 LCU 获得的

比特资源与其对视频质量的潜在贡献成正比, 从而保证在高复杂度区域内部进行 BA 的公平性。

2.3 低复杂度区域空间加权 BA

在低复杂度区域内部采用空间位置权重进行 BA。根据人眼视觉特性和图像信息分布规律, 中心区域通常更重要, 因此, 在低复杂度区域更关注中心区域的 BA。设低复杂度区域内部 LCU 的空间位置为 (x_i, y_i) , 图像中心为 (x_c, y_c) , 定义空间权重为

$$W_i = \exp \left\{ -\eta \left[(x_i - x_c)^2 + (y_i - y_c)^2 \right] \right\} \quad (12)$$

式中: η 为控制空间权重衰减强度的参数, 即 $\eta > 0$ 时, 中心区域可以获得更多的比特, η 越大, 中心区域权重越大, 边缘区域权重更小, 根据经验值设定 η 值为 4.0。

2.4 QP 值平滑处理

RC 的 2 个关键步骤为 BA 和编码参数的确定。QP 为关键的编码参数, 本文使用 HEVC 的经典 RC 模型进行 QP 值的计算。

2.3 节已获得每个 LCU 的 BA 结果, 但 LCU 间易产生 QP 值跳变, 对视觉质量造成影响, 因此, 采用空间邻域加权平滑的策略。具体地, 设原始 QP 值为 Q_i , 其邻域 $N(i)$ 定义为与当前 LCU 共享边的 4 个邻域, 即上下左右 4 个 LCU, 则平滑后 QP 值为

$$\hat{Q}_i = (1 - k) Q_i + k \frac{1}{|N(i)|} \sum_{j \in N(i)} Q_j \quad (13)$$

式中: k 为平滑系数, 控制原始 QP 与邻域均值之间的平衡, 根据经验值设定 k 值为 0.7。

3 实验与结果

3.1 实验环境与数据集

本文实验基于 Linux 系统 (Ubuntu18.04 版本) 的服务器实现, 在 HEVC 的参考软件 HM-18.0 基础上进行算法实现。测试视频序列选自 HEVC 标准测试序列集, 详细测试集序列如表 1 所示。编码配置采用低延迟 P 帧结构, GOP 长度为 8, 帧率为 30 帧/s, 码率范围设置为 16~512 Kbit/s, 覆盖典型低码率应用场景。

分别对 HGLA 算法、HEVC 模型采用的 RC 算法 HM-RC、基于帧级多层感知机 (multilayer perceptron, MLP) 的 BA 算法^[22]、基于简化立方率失真模型的帧内 RC 算法^[23]、基于通用视频编码 (versatile video coding, VVC) 的变分辨率 RC 算法^[24] 比较。

3.2 低码率场景下实验结果

为验证本文算法在低码率场景下的 BA 效果, 分别在 4 种经典的低码率场景下进行实验对比, 码率值共分为 4 个阶段, 码率集合为 {16, 128, 256, 512} Kbit/s, 实验结果如表 2~表 5 所示。

表 1 HEVC 测试序列
Table 1 HEVC test sequence

类别	序列名称	帧数	帧率/(帧·s ⁻¹)
A	Traffic	150	30
A	PeopleOnStreet	150	30
A	Nebuta	300	60
A	SteamLocomotive	300	60
B	Kimono	240	24
B	ParkScene	240	24
B	Cactus	500	50
B	BasketballDrive	500	50
C	RaceHorses	300	30
C	BQMall	600	60
C	PartyScene	500	50
C	BasketballDrill	500	50
D	RaceHorses	300	30
D	BQSquare	600	60
D	BlowingBubbles	500	50
D	BasketballPass	500	50

表 2 目标码率为 16 Kbit/s 场景下实验对比
Table 2 Experimental comparison under the target bit rate of 16 Kbit/s

算法	PSNR/dB	SSIM	MOS
文献[22]	32.50	0.869	3.14
文献[23]	33.06	0.887	3.25
文献[24]	33.53	0.891	3.45
HM-RC	32.01	0.862	3.08
本文	33.86	0.896	3.54

表 3 目标码率为 128 Kbit/s 场景下实验对比
Table 3 Experimental comparison under the target bit rate of 128 Kbit/s

算法	PSNR/dB	SSIM	MOS
文献[22]	36.57	0.914	3.40
文献[23]	37.18	0.928	3.73
文献[24]	37.74	0.931	4.01
HM-RC	36.01	0.906	3.35
本文	37.89	0.934	4.05

表 4 目标码率为 256 Kbit/s 场景下实验对比
Table 4 Experimental comparison under the target bit rate of 256 Kbit/s

算法	PSNR/dB	SSIM	MOS
文献[22]	38.72	0.921	3.84
文献[23]	39.18	0.936	4.06
文献[24]	39.63	0.940	4.12
HM-RC	38.27	0.918	3.68
本文	39.71	0.945	4.26

为验证本文算法在不同码率条件下的性能优势,选取目标码率为 16 Kbit/s、128 Kbit/s、256 Kbit/s

表 5 目标码率为 512 Kbit/s 场景下实验对比
Table 5 Experimental comparison under the target bit rate of 512 Kbit/s

算法	PSNR/dB	SSIM	MOS
文献[22]	40.54	0.930	4.03
文献[23]	41.13	0.948	4.25
文献[24]	41.58	0.952	4.39
HM-RC	40.09	0.926	3.96
本文	41.76	0.957	4.41

和 512 Kbit/s 这 4 种低码率场景下的典型带宽值,分别与现有代表性算法(文献 [22-24] 及 HEVC 官方参考模型 HM-18.0 的 HM-RC)进行客观指标和主观评分对比。客观指标采用峰值信噪比 (peak signal-to-noise ratio, PSNR) 和 结 构 相 似 性 指 数 (structural similarity index measure, SSIM), 主观评价采用主观意见得分 (mean opinion score, MOS)。其中, MOS 通过邀请 20 名无主观偏向的观众观看重建视频帧,采用 1~5 分等级评分,5 分表示最优。实验结果如表 2~表 5 所示。可见,本文算法相较其他 4 种算法, PSNR 平均提升 0.95 dB, SSIM 平均提升 0.017, MOS 普遍高于其他 4 种算法。综上分析, 本文算法在不同码率下均表现出优异的客观和主观性能,尤其在低码率条件下提升更为显著,验证了本文策略在低码率 BA 中的有效性和实用价值。

为验证本文提出的各项因子对 HGLA 算法的影响,针对目标码率为 512 Kbit/s 的情况进行消融实验,分析每个因素的性能提升表现,具体分为如下 5 种情况: ①HGLA; ②无区域划分模型,移除复杂度计算与区域划分,采用统一 BA; ③仅非合作博弈,仅保留第一层博弈模型,在区域内部不进行合作分配; ④无空间权重,移除图像中心区域的加权机制,低复杂区域比特均匀分配; ⑤无 QP 平滑处理,直接使用 BA 的结果,不进行 LCU 间 QP 过渡处理。

从表 6 的实验结果可以看出,各关键模块对算法整体性能均具有显著贡献。与原始算法 HGLA 相比,在移除区域划分机制后, PSNR 降至 40.05 dB,

表 6 本文提出的各项因子对 HGLA 算法的影响
Table 6 The influence of various factors proposed in this paper on the HGLA algorithm

实验配置	PSNR/dB	SSIM	MOS
HGLA	41.76	0.957	4.41
无区域划分模型	40.05	0.944	3.92
仅非合作博弈	40.49	0.948	4.01
无空间权重	41.31	0.949	4.07
无QP平滑处理	41.54	0.951	4.10

SSIM 降为 0.944, MOS 下降至 3.92, 表明区域划分对于复杂度表征与 BA 具有关键作用, 其缺失导致编码器在细节保留与视觉主观质量方面明显退化。

在仅保留非合作博弈第一层建模的情况下, PSNR 降为 40.49 dB, MOS 下降至 4.01, 说明高复杂度区域内部的合作 BA 机制在提升编码效率方面具有补充性作用。若移除空间权重机制(无空间权重), PSNR 和 SSIM 均出现轻微下降, MOS 从 4.41 降至 4.07, 说明图像中心区域权重机制在提升主观感知质量方面具备正向效益。

此外, 去除 QP 平滑处理导致 MOS 降低至 4.10, 尽管 PSNR 和 SSIM 变化较小, 但 MOS 下降达 0.31, 表明 LCU 间 QP 值过渡处理对图像自然感和连续性具有显著影响, 其缺失可能引入明显边缘断层等视觉伪影, 影响人眼感知。

综上分析可知, 各模块在不同层面提升了编码性能和视觉质量。区域划分为 BA 奠定合理结构, 双层博弈机制协同调控比特资源, 空间权重策略强化主观中心内容, 而 QP 平滑处理则优化了空间连续性与整体视觉体验。

4 结 论

1) 本文针对低空通信场景中存在的链路带宽普遍较低、码率利用率低与视频质量不稳定等问题, 研究并设计了一种面向低码率场景下的 LCU 级 BA 优化算法, 基于双层混合博弈理论构建全局 BA 模型, 结合 LCU 复杂度建模与区域划分策略, 分别对高复杂度区域引入非合作与合作博弈模型, 实现动态环境下编码资源的高效协同调度; 同时, 针对低复杂度区域引入空间加权机制与 QP 平滑处理方法, 在保证边缘视觉质量的同时, 有效抑制 QP 波动带来的视觉干扰。

2) 本文算法在不同程度低带宽下均表现出良好的自适应能力与鲁棒性, 在保证码率精度的同时, 显著提升了编码图像的主客观质量, 尤其在低比特率场景下, 具有更明显的优势。

未来工作可在本文研究基础上进一步引入跨帧博弈机制与强化学习方法, 实现帧间比特全局协同调控, 从而进一步提升编码系统在极端复杂场景下的智能感知与动态适应能力。

参考文献 (References)

[1] 余润泽, 苏敏杰, 周烨挺, 等. 低空智慧城市无人机系统网络安全分析及防护策略研究[J]. 电信工程技术与标准化, 2025, 38(3): 61-67.

Yu R Z, Su M J, Zhou Y T, et al. Analysis of network security and research on protection strategies for low altitude smart city un-

manned aerial vehicle systems[J]. Telecom Engineering Technics and Standardization, 2025, 38(3): 61-67(in Chinese).

[2] 朱迪锋, 张雅霏, 龚轩, 等. 高动态链路下基于主机状态感知的视频码率自适应算法[J/OL]. 北京航空航天大学学报, 2024 (2024-10-10)[2025-03-20]. <https://link.cnki.net/urlid/11.2625.V.20241009.1858.002>.

Zhu D F, Zhang Y F, Gong X, et al. Host-state awareness based video bitrate adaptation algorithm in highly dynamic links[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2024 (2024-10-10)[2025-03-20]. <https://link.cnki.net/urlid/11.2625.V.20241009.1858.002>(in Chinese).

[3] He J, Yang F Z. Efficient frame-level bit allocation algorithm for H. 265/HEVC[J]. IET Image Processing, 2017, 11(4): 245-257.

[4] Liu Z Z, Pan X, Li Y M, et al. A game theory based CTU-level bit allocation scheme for HEVC region of interest coding[J]. IEEE Transactions on Image Processing, 2021, 30: 794-805.

[5] Gao W, Kwong S, Jia Y H. Joint machine learning and game theory for rate control in high efficiency video coding[J]. IEEE Transactions on Image Processing, 2017, 26(12): 6074-6089.

[6] Ahmad I, Luo J C. On using game theory to optimize the rate control in video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2006, 16(2): 209-219.

[7] Luo J C, Ahmad I, Sun Y. Controlling the bit rate of multi-object videos with noncooperative game theory[J]. IEEE Transactions on Multimedia, 2010, 12(2): 97-107.

[8] Park H, Van Der Schaar M. Bargaining strategies for networked multimedia resource management[J]. IEEE Transactions on Signal Processing, 2007, 55(7): 3496-3511.

[9] Wang X, Kwong S, Zhang Y. Applying game theory to rate control optimization for hierarchical B-pictures[J]. IEEE Transactions on Broadcasting, 2013, 59(4): 591-601.

[10] Gao W, Kwong S, Zhou Y, et al. SSIM-based game theory approach for rate-distortion optimized intra frame CTU-level bit allocation[C]//Proceedings of the IEEE Transactions on Multimedia. Piscataway: IEEE Press, 2016: 988-999.

[11] Lin J L, Huang A P, Zhao T S, et al. λ -domain VVC rate control based on Nash equilibrium[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(7): 3477-3487.

[12] Zhao T S, Lin J L, Song Y J, et al. Game theory-driven rate control for 360-degree video coding[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 3998-4006.

[13] 公衍超, 王玲, 刘颖, 等. 视频主观观测实验启发的 HEVC 感知帧内码率控制[J]. 通信学报, 2021, 42(8): 90-102.

Gong Y C, Wang L, Liu Y, et al. HEVC perceptual intra-frame rate control inspired by video subjective observation experiment[J]. Journal on Communications, 2021, 42(8): 90-102(in Chinese).

[14] 李维, 杨付正, 任鹏. 考虑视频内容的 H.265/HEVC 帧内码率分配算法[J]. 通信学报, 2015, 36(9): 76-81.

Li W, Yang F Z, Ren P. H.265/HEVC frame-level bitrate allocation algorithm considering video content[J]. Journal on Communications, 2015, 36(9): 76-81(in Chinese).

[15] 崔士明, 黄新彭, 林兴斌, 等. 基于感兴趣区域及 LCU 特征的码率分配算法[J]. 工业控制计算机, 2024, 37(3): 15-17.

Cui S M, Huang X P, Lin X B, et al. Rate allocation algorithm

based on region of interest and LCU feature[J]. *Industrial Control Computer*, 2024, 37(3): 15-17(in Chinese).

[16] Liu F Y, Chen Z Z. Multi-objective optimization of quality in VVC rate control for low-delay video coding[J]. *IEEE Transactions on Image Processing*, 2021, 30: 4706-4718.

[17] He L Q, He X H, Xiong S H, et al. Efficient rate control in versatile video coding with adaptive spatial-temporal bit allocation and parameter updating[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(6): 2920-2934.

[18] Biatek T, Raulet M, Travers J F, et al. Efficient quantization parameter estimation in HEVC based on ρ -domain[C]//*Proceedings of the 2014 22nd European Signal Processing Conference*. Piscataway: IEEE Press, 2014: 296-300.

[19] Nash J. Non-cooperative games[J]. *The Annals of Mathematics*, 1951, 54(2): 286.

[20] Driessen T. Cooperative games, solutions and applications[M]. Berlin: Springer, 1988.

[21] Zhu D F, Zhang Y F, Gong X, et al. DTS-BWpredictor: dual-scale temporal strategy based bandwidth prediction in highly dynamic links[J]. *Computer Networks*, 2025, 259: 111071.

[22] Raufmehrf F, Salehi M R, Abiri E. A frame-level MLP-based bit-rate controller for real-time video transmission using VVC standard[J]. *Journal of Real-Time Image Processing*, 2021, 18(3): 751-763.

[23] Wang Y Z, Zhang J Q, Sun S L. Rate control for VVC intra coding with simplified cubic rate-distortion model[C]//*Proceedings of the 2023 IEEE 25th International Workshop on Multimedia Signal Processing*. Piscataway: IEEE Press, 2023: 1-6.

[24] Fu T L, Zhang K, Zhang L, et al. Rate control with resolution changes for VVC[C]//*Proceedings of the 2023 IEEE International Symposium on Circuits and Systems*. Piscataway: IEEE Press, 2023: 1-5.

LCU-level video bit allocation algorithm based on two-layer game in low-altitude communication environment

ZHANG Yafei^{1, 2}, LIU Lulu¹, YU Jinch^{1, 2}, ZHU Difeng^{1, 3, *}, GONG Xuan¹, GAO Yi¹

(1. Hangzhou Innovation Institute of Beihang University, Hangzhou 310051, China;
2. School of Software, Beihang University, Beijing 100191, China;
3. Zhejiang Yuexiu University, Shaoxing 312030, China)

Abstract: This paper addresses the issues of low video coding efficiency and insufficient bit rate utilization in low-altitude communication scenarios by proposing an largest coding unit (LCU)-level bit allocation (BA) optimization based on two-layer hybrid game model (HGLA) algorithm. Using the LCU as the basic unit, the algorithm performs region partitioning according to texture complexity, motion intensity, and edge information, and establishes a differentiated BA mechanism for high- and low-complexity regions. In particular, a combination of cooperative and non-cooperative game models is used for high-complexity regions to achieve coordinated BA; for low-complexity regions, a global quantization parameter (QP) smoothing strategy and a spatial weighting mechanism are introduced to improve bit rate utilization efficiency and visual continuity. Finally, experimental validation under four low-bit rate scenarios demonstrates that the proposed method achieves significant improvements in peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and subjective quality. Specifically, at the target bit rate of 16 Kbit/s, compared with four baseline methods including the high efficiency video coding (HEVC) rate control (RC) algorithm HM-RC and the frame-level multilayer perceptron (MLP)-based BA algorithm, the intra-frame rate control algorithm based on a simplified cubic rate-distortion model, and the variable-resolution rate control algorithm based on versatile video coding (VVC), the proposed method achieves average performance gains of 3.34%, 2.16%, and 9.8% in PSNR, SSIM, and mean opinion score (MOS), respectively, demonstrating its excellent practicality and adaptability in improving bit rate utilization efficiency and preserving image quality.

Keywords: rate control; low bit rate; game theory; bit allocation; reliable transmission of low-altitude video

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0420

面向低空起降场的 FOD 检测与识别

郑皓文¹, 胡海苗^{1,2}, 徐壮^{3,*}, 何壮², 胡昊鑫²

(1. 北京航空航天大学 计算机学院, 北京 100191; 2. 北京航空航天大学杭州创新研究院, 杭州 310000;

3. 浙江省机场集团有限公司, 杭州 311207)

摘 要: 低空起降场跑道外来物 (FOD) 的检测与识别对于飞机飞行性能与机场运行安全具有重要意义。针对 FOD 目标尺寸小、无表观特征的特点及公开数据少的问题, 构建 FOD 数据集, 并提出一种基于物理特性的 FOD 检测与识别方法。采集常见 FOD 材质的多光谱图像, 构建 FOD 的多光谱数据集。基于该数据集, 对常见 FOD (包括金属的锈迹与强反射现象) 进行属性分析。基于沥青背景反射率, 得到材质相对沥青的伪反射率。通过多种统计量, 初步验证了不同材质的可分性, 进而提出了使用密集嵌套连接的编码器-解码器结构神经网络, 将伪反射率作为输入进行材质识别, 在所构建数据集的测试集中, 准确率为 0.999 3, Macro-F1 分数为 0.774 9, 优于同模型大小等级的其他材质识别和分割模型。

关 键 词: 材质识别; 机场跑道外来物; 自编码器; 多光谱图像; 伪反射率

中图分类号: TP399

文献标志码: A

文章编号: 1001-5965(2026)09-3153-10

低空起降场的跑道外来物 (foreign object debris, FOD) 是指散落在跑道上或被风吹至跑道的非跑道本身物体, 包括飞机零部件、工作人员遗落物品或跑道表面剥落的物质, 如金属碎片、橡胶、碎石等^[1]。这些 FOD 在飞机起降时可能对飞机各部位造成损害, 轻则降低飞行性能, 重则发生严重飞行事故。因此, 增强 FOD 的检测与预警能力, 精确识别并及时清除跑道上的 FOD, 对控制飞机维护成本、确保飞行性能及运行安全具有重要意义。

然而, FOD 类型多样、尺寸极小且无固定表观特征的特点导致 FOD 的检测与识别存在一定困难。不同材质的 FOD 空间特性区分度较低, 基于单一的传感设备难以对其特征进行充分表达。目前, 大部分起降场的 FOD 检测仍采用人工巡检, 少数大型机场采用雷达技术、自动巡检机器人技术及图像处理技术。基于图像处理技术的 FOD 检测

普遍基于可见光图像进行处理, 将 FOD 检测转化为特定场景下的目标检测。深度学习技术为该思路提供了新的解决方案, 主要集中于对两阶段网络 (如更快速的区域卷积神经网络 (faster region-based convolutional neural network, Faster R-CNN)^[2]) 和单阶段网络 (如 YOLO 系列^[3]) 的优化与改进。这些方法通过优化网络架构、引入多尺度特征融合、结合自监督学习或异常检测策略, 显著提升了 FOD 检测的精度和鲁棒性。两阶段目标检测网络以 Faster R-CNN 为代表, 通过区域提议网络 (region proposal network, RPN) 和后续的分类与边界框回归实现目标检测。针对 FOD 检测的特殊需求, 研究者对其进行了多项改进。例如, 文献 [4] 提出一种专门针对 FOD 检测的框架, 通过改进 RPN 以适应跑道场景中中小目标的特点, 并结合基于空间变换网络的卷积分类器, 进一步增强了对 FOD 的定位和分类能

收稿日期: 2025-07-01; 录用日期: 2025-08-08; 网络出版时间: 2025-08-20 17:36

网络出版地址: link.cnki.net/urlid/11.2625.V.20250820.1500.001

基金项目: 浙江省“尖兵”“领雁”研发攻关计划 (2025C01037)

* 通信作者. E-mail: frank0139@163.com

引用格式: 郑皓文, 胡海苗, 徐壮, 等. 面向低空起降场的 FOD 检测与识别 [J]. 北京航空航天大学学报, 2026, 52 (9): 3153-3162.

Zheng H W, Hu H M, Xu Z, et al. FOD detection and recognition for low-altitude takeoff and landing sites [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3153-3162 (in Chinese).

力,文献[5]在Faster R-CNN的基础上引入了多尺度特征融合机制,通过融合不同层次的特征图,捕捉FOD在不同尺度下的视觉特征,同时,还结合了在线难样本挖掘策略,通过动态调整训练样本的权重,提升模型对复杂背景中FOD的检测性能。然而,这类方法依赖目标的二维表观特征,不能完全使用实际起降场FOD的特点。除了传统的监督学习方法,近年来,基于自监督学习和异常检测的FOD检测方法逐渐受到关注。这些方法利用跑道图像的内在规律,通过无监督或自监督方式学习正常跑道场景的特征,将FOD视为异常进行检测。文献[6]提出一种基于自监督学习的FOD检测框架,通过训练模型预测跑道图像的正常纹理和结构特征,检测偏离正常模式的区域作为潜在FOD。在此基础上,文献[7]进一步提出一种基于重构自编码器的FOD检测方法,结合高斯混合模型对跑道图像的重构误差进行建模,将FOD视为异常点进行检测。这类方法适用于FOD的定位,但无法对FOD材质进行识别。

材质识别领域因深度学习技术的进步而快速发展,尤其是在大模型应用于材质识别任务后,显著提升了识别精度和跨场景泛化能力。传统方法依赖手工设计的纹理特征,而大模型通过大规模预训练,结合自注意力机制或多模态学习,提取高层次语义与低层次纹理特征,适用于复杂场景下的材质分类。文献[8]提出的情境材料数据集(materials in context database, MINC)为材质识别提供了基准,推动了基于卷积神经网络的研究,如文献[9]的深度纹理编码网络(deep texture encoding network, Deep TEN)通过纹理编码提升了性能。Vision Transformer^[10]通过自注意力机制捕捉全局与局部特征,在MINC和Flickr材料数据集(Flickr material database, FMD)上展现出优越性能。对比语言-图像预训练(contrastive language-image pre-training, CLIP)^[11]的多模态能力支持零样本材质识别,通过文本引导实现无需额外标注的分类。文献[9]的DINOv2通过自监督学习提取鲁棒特征,适合小样本场景。生成模型(如扩散模型^[12])通过材质样本生成增强了训练数据。然而,大模型的参数量庞大,导致计算资源需求高、推理时间长,限制了其在资源受限设备上的部署。

不同材质物体因分子结构差异,对不同波段电磁波的反射吸收能力各异。通过多光谱探测可获取物体的光谱曲线,进而根据材质对应的光谱特征识别其类别,适用于无表观特征的材质识别。文献[13]利用高光谱成像发现沥青地面在特定波

段的稳定性,可作为其特征波段。文献[14]结合高光谱与机器学习分类器,实现无线充电装置内金属异物的检测,通过像素级分类模型识别金属光谱特性。这些研究为多光谱技术在FOD材质识别中的应用提供了基础。公开的FOD数据集较少,如机场外来物(FOD in airports, FOD-A)数据集^[15]和上海交通大学FOD数据集^[16],这些数据集均为无人机或可见光相机采集的可见光数据,FOD尺寸较大且表观特征丰富。目前,公开的FOD数据集较少,多光谱信息在FOD检测识别领域的应用还未引起广泛关注。

基于以上分析,本文通过采集并构建多光谱FOD数据集,开发了一种基于光谱特性嵌入的FOD检测与识别方法,降低对表观信息的依赖,目标只以碎片化或零件化形式存在,仍能实现准确检测,同时,物理特性具有良好的区分性,不需要很多的训练数据,就能实现高精度的类别区分及材质识别,因此,能够很好地解决低空起降场道路异物检测困境,对于机场安全具有重要意义。

1 FOD 属性分析

1.1 FOD 数据集构建

目前,关于FOD的多光谱数据集较少,因此,本文使用波长范围为385~1 009 nm的256波段高光谱相机,采集并最终保留了50张相对高质量的数据,并使用水平翻转、垂直翻转、旋转不同角度的方式扩展为200张尺寸为696像素×775像素、波段数为256(即每张图像的尺寸为696像素×775像素×256像素)的数据构成数据集。数据集中包括但不限于上述重点关注的材料;拍摄距离为5 m、10 m、20 m不等;拍摄环境包含向光(23张)、背光(20张)、阴天(7张)时的场景。对于数据集的标注掩码,本文使用ENVI软件进行像素级的数据标注。

从光谱的角度,神经网络模型可以从少量的像素中充分学习不同材质的光谱特征,因此,本文数据集将多种FOD放在同一张图中,作为主要的训练来源。而在测试集中,在背景主要为沥青的场景中放置极少的FOD,检验模型异常检测与材质识别的效果。图1展示了训练集的主要形式,其中包括了重点关注的FOD材料光谱特性。

FOD的种类方面,根据文献[1]总结的常见FOD物品,从材质角度分析,重点需要关注金属(40张)、硬质塑料(20张)、石头(30张)、木棍(20张)、纸板(20张)、树叶(30张),一张图像一般包含多个材质。其中,除了木棍、纸板与树叶,其余材料为中危或高危材质。对于危险系数最高的金属,本文收集了锌、铝、钢等常见的金属材质进行



图 1 多光谱数据集中训练集示例(图片为使用红、绿、蓝 3 波段合成的伪 RGB 图像)

Fig. 1 Training set examples in the multispectral dataset (images are pseudo-RGB images synthesized using red, green, and blue wavelengths)

光谱数据采集,同时,针对金属会出现的强反射和锈迹现象,本文数据集也专门进行了采集和单独标注,如图 2 所示。对于强反射现象来说,本文数据集认为在调整过曝光系数后,像素的光谱值在 520~587 nm 之间超出相机响应值阈值(60 000),应被视作发生了强反射;对于锈迹,本文数据集选用了整面全部为锈迹的材料作为样本,便于标记的同时,认定该区域全部为锈迹。

1.2 基于数据集的 FOD 属性分析

结合实验室以往工作中采集的多光谱数据集

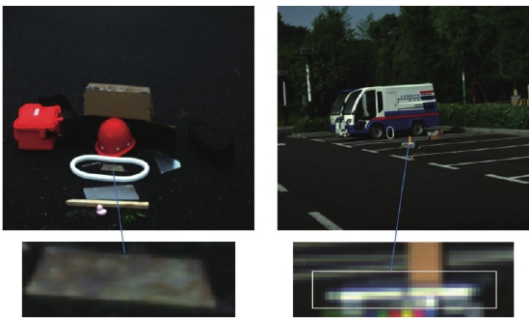


图 2 数据集中金属出现的锈迹与强反射现象
Fig. 2 Rust stains and strong reflection phenomena appearing on metals in the dataset

及本文采集得到的数据可以观察得到,使用同一种相机拍摄得到的数据中,沥青地面的光谱响应值趋势保持高度的稳定性及相对固定的特征,仅在太阳光照强度不同、地面干燥程度不同时存在响应幅值的差异。这种特性使在传统的白板校正方法中使用沥青光谱信息替代白板具有合理性,如图 3 所示。

反射率提取领域中,经典的本征分解^[17]模型为

$$\boldsymbol{I} = \boldsymbol{R} \circ \boldsymbol{S} \tag{1}$$

式中: $\boldsymbol{I}$ 为相机采集得到的响应值; $\boldsymbol{R}$ 为材质本征反射率; $\boldsymbol{S}$ 为光照信息; “ $\circ$ ” 表示哈达玛积。

带有强反射的本征分解模型为

$$\boldsymbol{I} = \boldsymbol{R} \circ \boldsymbol{S} + \boldsymbol{R}_H \tag{2}$$

式中: $\boldsymbol{R}_H$ 为强反射部分,当 $\boldsymbol{R}_H$ 不为 $\mathbf{0}$ 时, $\boldsymbol{S}$ 为 $\mathbf{0}$ 。

该像素点的反射率信息无法复原,在光谱信息中,强反射在可见光范围内常见为过曝,即响应值

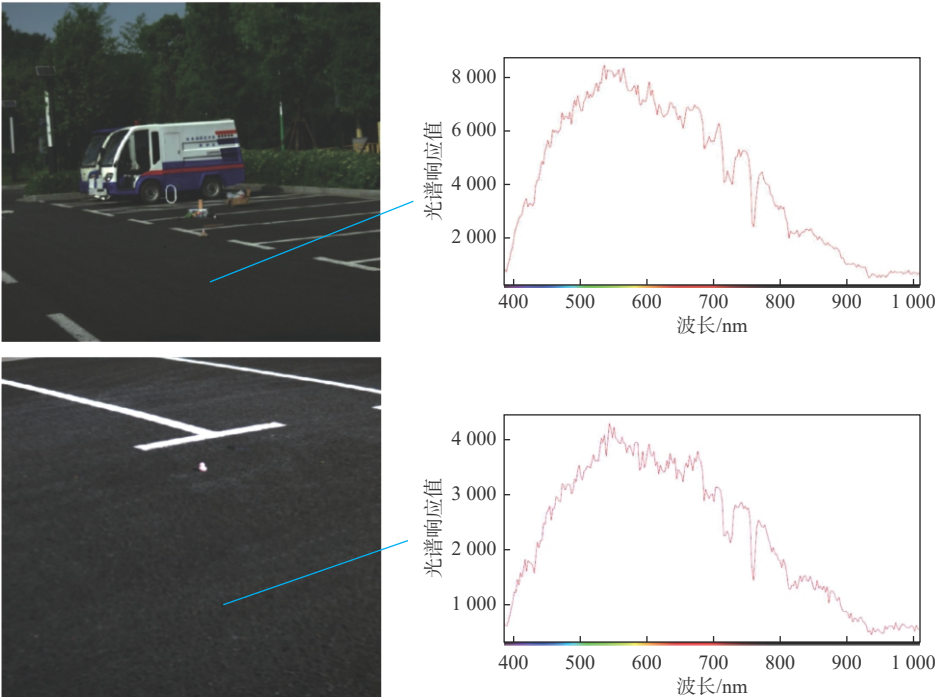


图 3 沥青地面光谱的稳定性
Fig. 3 Stability of asphalt pavement spectra

超出相机最大响应值。本文使用单独标注该种现象的方法,使强反射可以视作一种特殊的材质,进而,在后续计算反射率时,将带有强反射的模型泛化为经典的本征分解模型。进而,有

$$\mathbf{I}_x = \mathbf{R}_x \circ \mathbf{S}$$

(3)

$$\mathbf{I}_y = \mathbf{R}_y \circ \mathbf{S}$$

(4)

式中:下标 x 表示 FOD 材料;下标 y 表示沥青。

式(3)和式(4)相除,可得

$$\frac{\mathbf{R}_x}{\mathbf{R}_y} = \frac{\mathbf{I}_x}{\mathbf{I}_y}$$

(5)

基于 $\mathbf{R}_y$ 的稳定性,可以认为式(5)左边的值为基于沥青反射率稳定性的伪反射率。利用伪反射率具有 2 个方面的优势:①其面向当前视图,消除了光照信息对反射率的影响,具体来说,其极大地消除了向光、背光、阴天对 FOD 光谱特性的影响;②在实际应用中,可以省去传统白板校准的过程,如图 4 所示。

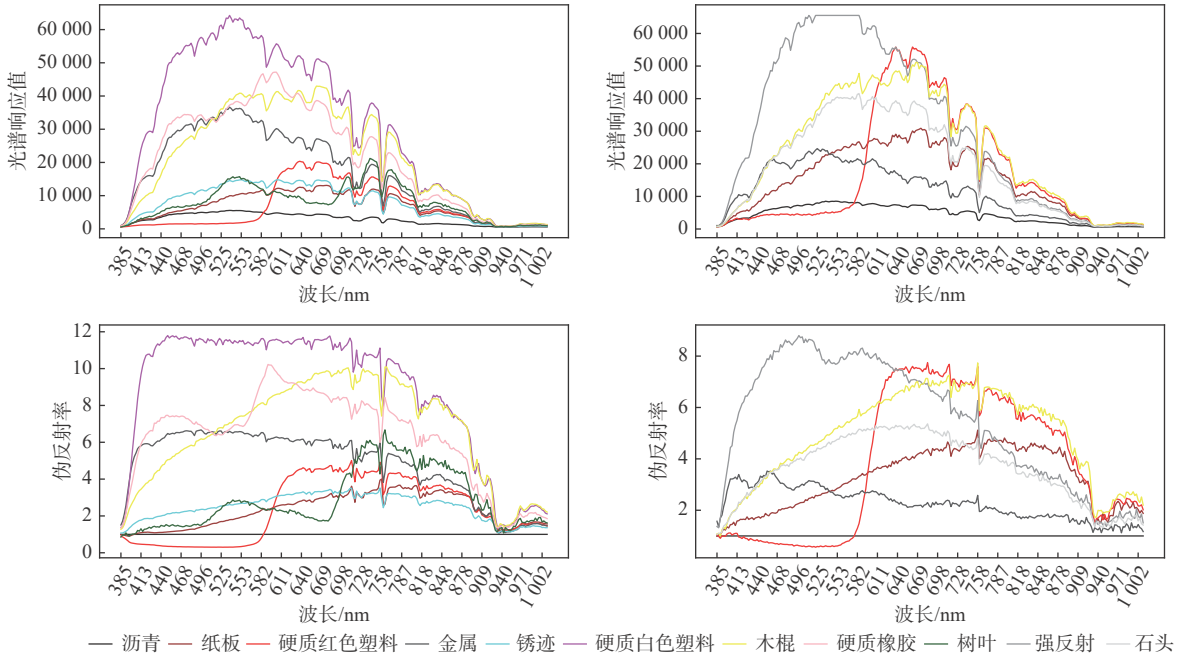


图 4 光谱响应值与伪反射率

Fig. 4 Spectral response values and pseudo-reflectance

限于实际需求与计算性能,本文从 256 波段中选择了 6 个波段用于后续检测。该 6 波段结合了手动观察特征与特征选择网络,最终确定为波长为 480, 550, 649, 740, 850, 908 nm 的 6 个波段数据,范围涵盖了可见光区域和红外区域。该 6 波段的可分性将在实验部分进行各种指标的验证。

2 密集嵌套连接的编码器-解码器 FOD 识别网络

本文的 FOD 识别主要分为 2 个部分:①将采集的光谱数据转为伪反射率;②将其输入在编码器-解码器结构基于密集连接的神经网络模型中,得到像素级的分割结果。

2.1 FOD 识别网络整体架构

2.1.1 伪反射率提取

本节主要得到沥青的反射率。基于数据集背景基本由沥青地面组成的先验,在每张图像随机采集了 50 000 个点,约占全图像总像素数的 10%。该

50 000 个点中可能存在 FOD 本身或沥青的一些偏离本身光谱特性的杂质点,因此,本文采用孤立森林^[18]算法,将异常点比例设置为 5%。利用孤立森林算法无需定义相似性阈值、时间复杂度低于聚类等方法的优势,得到一个更为可靠的相似点集,如图 5 所示,其中, $\mathbf{R}_{\text{fake}}$ 为伪反射率。

孤立森林算法通过随机分割数据构建多棵孤立树,利用异常点通常具有较短路径长度的特性进行识别。路径长度定义为

$$h(x) = e + c(m)$$

(6)

式中: $x \in \mathbf{R}^6$ 为 6 波段的数据点集; e 为根节点到叶节点的边数; m 为叶节点中的数据点数量; $c(m)$ 为调整项,近似于随机二叉搜索树的平均路径长度。

平均路径长度指包含 n 个数据点的子树预期路径长度,定义为

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}$$

(7)

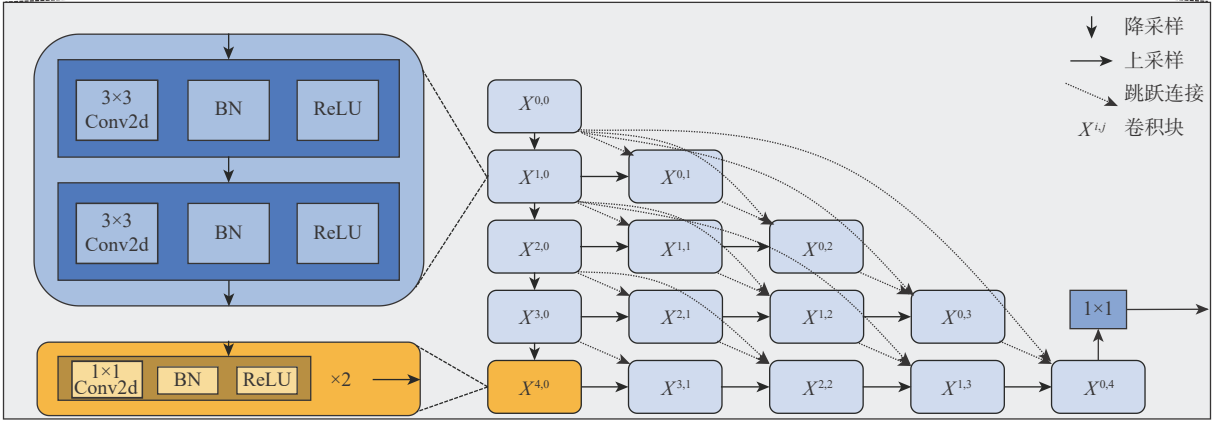
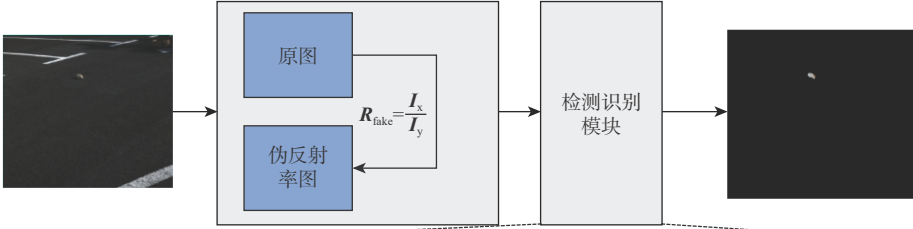


图 5 FOD 检测与识别整体结构

Fig. 5 Overall structure of FOD detection and identification

式中: $H(n-1)$ 为调和数, $H(n-1) \approx \ln(n-1) + \gamma$, γ 为欧拉-马斯克朗尼常数。

数据点 x 的异常分数基于其在 T 棵孤立树中的平均路径长度 $E[h(x)]$, 定义为

$$s(x,n)=2^{-\frac{E[h(x)]}{c(n)}} \tag{8}$$

$$E[h(x)]=\frac{1}{T}\sum_{t=1}^Th_t(x) \tag{9}$$

式中: 异常分数 $s(x,n) \in [0,1]$, 其越接近 1, 表示 x 更可能是异常点; 本文中, T 取 100。

选择得分在前 5% 的点作为异常点进行剔除, 得到的相似点集均值作为沥青的反射率。将全图响应值除以该反射率, 即可得到伪反射率图。

2.1.2 FOD 识别网络细节

FOD 缺乏表现特征, 有效信息只有相对固定的光谱响应值; 此外, 数据量相对于计算机视觉领域的常见数据集而言极少。因此, 参考文献 [19] 的思路, 使用密集嵌套连接的编码器-解码器结构, 从而充分利用每一层的形谱信息, 对于 FOD 检测与识别的精度具有较好效应。

本文网络结构由编码器、解码器、跳跃连接、瓶颈层和输出层组成。本文对数据集进行编码部分由 4 层卷积块组成, 每个下采样块包括 2 组 3×3 卷积层、标准化层和激活函数。每层卷积块之间用 2×2 最大池化层进行下采样。图 5 中, 编码路径节点标注为 $X^{i,0}$, i 为下采样层级。瓶颈层位于网络底部, 标记为 $X^{4,0}$, 其包含 2 组 1×1 卷积层、标准

化层和激活函数。解码器包含转置卷积上采样层和卷积层。解码路径节点标注为 $X^{i,j}$, i 为上采样层级, j 为该层级的嵌套节点标号。嵌套跳跃连接部分不局限于传统 U-Net^[20], 只对对称的编码、解码层进行跳跃连接, 该部分公式表达为

$$X^{i,j} = \text{Conv}(\oplus(\text{Up}(X^{i+1,j-1}), X^{i,0}, X^{i,1}, \dots, X^{i,j-1})) \tag{10}$$

式中: Conv 为卷积操作; “ $\oplus$ ” 表示拼接; Up 为上采样操作。

每个解码器节点 $X^{i,j}$ 接收上采样特征 $X^{i+1,j-1}$ 和跳接特征。跳接特征包括编码器节点 $X^{i,0}$ 和同一层级的较低节点 $X^{i,k} (k < j)$ 。

输出层对 $X^{0,4}$ 使用 1×1 卷积层, 生成类别数通道的分割掩码。该结构通过多层次的特征融合, 更好地利用各通道的光谱信息, 提升对不同光谱特征的区分能力; 其通过嵌套和密集跳跃连接, 将不同感受野的特征充分融合, 使模型既能捕捉全局结构, 又能保留细节和边界信息; 跳跃连接使浅层和深层信息在网络中高效流动, 提升对目标可能存在的形状、纹理、边界信息的感知能力, 从而实现形谱信息融合, 提高识别的准确性。

2.2 损失函数

在损失函数的使用上, 将 Tversky 指数^[21] 与 Focal Loss^[22] 以相乘的形式结合, 计算如下:

$$L = \sum_{c=1}^{C'} \left[(1 - \tau_{c'})^{\gamma} \left(-\alpha_{c'} \sum_{i \in \Omega} y_{i,c'} \ln p_{i,c'} \right) \right] \tag{11}$$

式中: 多分类 Tversky 指数 τ 为每个类别 c 进行单独

计算:

$$\tau_{c'} = \frac{\sum_l y_{l,c'} p_{l,c'}}{\sum_l y_{l,c'} p_{l,c'} + \alpha_{c'} \sum_l (1 - y_{l,c'}) p_{l,c'} + \beta_{c'} \sum_l y_{l,c'} (1 - p_{l,c'})}$$

(12)

其中: C' 为类别总数; $y_{l,c'} \in [0, 1]$ 为第 l 个像素在类别 c' 的真实标签, 为独热编码; $p_{l,c'} \in [0, 1]$ 为模型预测第 l 个像素属于类别 c' 的概率; Ω 为所有像素集合, 即全图; $\alpha_{c'}, \beta_{c'} \in [0, 1]$ 为类别 c' 的权重, 分别控制假阳性和假阴性, 本文中分别设置为 0.7、0.3; $\gamma' \geq 0$ 为聚焦参数, 控制对难分割区域的关注程度; $\sum_l y_{l,c'} p_{l,c'}$ 为类别 c' 的真阳性近似, 即预测概率和真实标签的乘积和; $\sum_l (1 - y_{l,c'}) p_{l,c'}$ 为类别 c' 的假阴性近似; $\tau_{c'}$ 为类别 c' 的 Tversky 指数; $(1 - \tau_{c'})'$ 为类别 c' 的区域重叠误差聚焦因子。

聚焦因子、多种权重的加入使模型更加关注小目标, 从而降低漏检率。聚焦因子使用乘法的形式, 更加强调模型对难分割区域的关注。其相较于 Focal Loss 的问题是通过优化全局区域, 增强模型对形状和连通性的认知, 并通过平衡局部和全局优化, 生成更加平滑、连贯的分割掩码, 这使模型在实际判断过程中, 分割区域比真值面积大。但由于该问题在危害程度上弱于漏检, 其仍具有应用价值。

3 方法验证

3.1 FOD 材质可分性的初步验证

本文通过多个指标(马氏距离和离散度)对所选择的 6 个波段 FOD 数据可分性进行衡量。此外, 利用多元方差分析(multivariate analysis of variance, MANOVA)的角度验证所选波段的显著差异。进一步, 利用 t -分布随机近邻嵌入降维方法将本文所选的谱段组合进行数据降维并投影, 得到更加直观的可区分性结果。

马氏距离不同于常用的欧氏距离和曼哈顿距离, 通过协方差矩阵消除了特征之间的相关性影响, 在观察不同材质之间的差异性上更优; 离散度作为反映 2 个类别之间可分性的统计距离指标, 能够在 2 个类别均值相等或相近时有效区分 2 个类别。

MANOVA 作为一种统计方法, 用于分析多个因变量如何受到一个或多个自变量的影响。在本文中, 不同波段的响应值视为不同的因变量, 材质视为自变量, MANOVA 通过检验因变量的联合分布, 评估材质对于多个波段响应值的总体效应是否存在显著差异。表 1 展示了马氏距离与离散度数

表 1 多种衡量尺度下 FOD 数据的可分性验证

Table 1 Verification of FOD data separability under multiple measurement scales

指标类别	指标	p 值
衡量尺度	马氏距离	4.950
	离散度	4.983
MANOVA统计量	Wilks' Lambda	0
	Pillai's Trace	0
	Hottelling's Trace	0
	Roy's Largest Root	0

据, 同时展示了 MANOVA 分析的结果, 对 FOD 材质的可分性进行了初步验证。MANOVA 的 4 个统计量的 p 值均约为 0, 体现了不同材质在这 6 个波段下的显著差异。

对所选波段进行像素点采样, 并进行 t -分布随机近邻嵌入降维投影, 得到结果如图 6 所示。

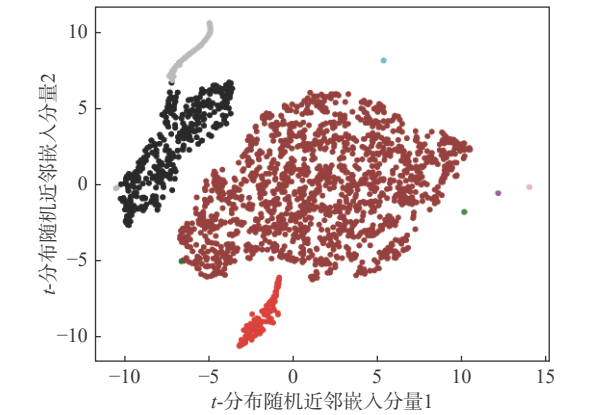


图 6 t -分布随机近邻嵌入可视化结果

Fig. 6 t -distributed stochastic neighbor embedding visualization results

3.2 网络训练环境与参数设置

本文实验运行于服务器环境。硬件方面, 实验使用 Nvidia Tesla V100 显卡, CPU 为 Intel Xeon Gold 5218R @ 2.10 Hz。软件方面, 实验以 Python 为编程语言, 以 Pytorch 为神经网络模型的框架工具。

参数设置方面, 在模型训练过程中, 使用 Adam 优化器, 并设置初始学习率为 0.000 3, 训练 10 个批次。在训练过程中, 根据验证集的损失对学习率进行动态调整。由于不同材质在选取波段上具有显著可分性, 因此, 10 个批次即可实现快速收敛。图像作为输入时, 将图像进行零填充, 使图像长与宽可以被 64 整除, 将图像分割为 64 像素 $\times$ 64 像素大小的块后, 才进入神经网络。

3.3 对比方法与评价手段

本文将所采用的神经网络模型与其他图像分割模型一同在本文数据集的测试集中进行测试, 通

过对比指标的手法进行模型识别能力的评价。测试指标为准确率和 Macro-F1 分数。

准确率 A 指预测正确的样本数与总样本数的比值, 是对分类器整体正确率的一种评价, 计算公式为

$$A = \frac{n_{TP} + n_{TN}}{n_{TP} + n_{FP} + n_{TN} + n_{FN}} \quad (13)$$

式中: n_{TP} 为预测正确的正样本数; n_{FP} 为预测错误的正样本数; n_{TN} 为预测正确的负样本数; n_{FN} 为预测错误的负样本数。

Macro-F1 分数 F_{Macro} 相较于准确率和 F_1 分数, 关注不同类别的分布情况, 尤其适合评价少数类的性能评价, 避免大类别主导结果。本文中, 大类别为沥青背景, FOD 均为少数类。Macro-F1 分数位于 0~1 之间, 1 表示完美分类, 0 表示完全错误, 计算公式为

$$F_{Macro} = \frac{1}{C} \sum F_q \quad (14)$$

式中: F_q 为类别 q 的 F_1 分数, F_1 分数基于精确率与召回率得到, 取值在 0~1 之间, 1 表示完美分类。

精确率 P_{re} 和召回率 R_{ec} 的计算公式分别为

$$P_{re} = \frac{n_{TP}}{n_{TP} + n_{FP}} \quad (15)$$

$$R_{ec} = \frac{n_{TP}}{n_{TP} + n_{FN}} \quad (16)$$

此外, 模型大小也是本文基于实际需求进行参考的重要指标。

本文在数据集的类别标注中, 将锈迹、金属与强反射分为 3 类, 但在实际危险评级中, 这 3 类都按照金属来判断, 因此, 本文在测试结果中, 着重观察了这 3 类错检的类别情况, 如果错检结果为另 2 类, 并不影响实际判断结果。

3.4 对比实验

表 2 展示了其他较为先进的图像分割方法与本文方法在本文数据集上的表现, 图 7 为分割结果。所有方法的整体准确率高, 是因为数据集标记了相当一部分的沥青地面, 该类数据的数据量较大, 且普遍易被正确识别。FCN-8s^[23] 是经典全卷积分割网络, 融合多层特征, 输出分辨率较高, 适合基础语义分割任务, 作为对比的基准; MST-DeepLabv3+^[24] 针对高分辨率遥感图像分割, 加入 SENet^[25] 注意力机制和迁移学习, 但模型较大, 且在此场景中误检率偏高; SR-SegNet^[26] 模型较小、适合实时分割, 但由于 FOD 场景缺乏表观信息, SR-SegNet 的简单结构难以分析不同材质之间的信息。U-Net 是本文的基础架构, 是编码器-解码器结构与同层跳跃连接的结合, 但对各层信息提取仍不

表 2 各方法的测试结果与模型大小

Table 2 Test results and model sizes for each method

方法	准确率	Macro-F1	模型大小/KB
FCN-8s ^[23]	0.787 7	0.199 2	57 877.19
MST-DeepLabv3+ ^[24]	0.998 9	0.545 3	104 620.54
SR-SegNet ^[26]	0.999 0	0.511 1	3 090.85
U-Net ^[20]	0.998 9	0.642 0	30 391.00
MSNet ^[27]	0.998 9	0.694 3	141 193.00
SSTN ^[28]	0.999 3	0.768 8	26 776.00
UNetFormer ^[29]	0.999 1	0.774 5	28 980.00
本文	0.999 3	0.774 9	23 151.00

够充分, 在准确率上不如本文方法。MSNet^[27] 采用 ResNet-50 提取特征, 通过级联上采样和特征金字塔结构融合多尺度特征, 在本文数据集中的金属和强反射识别能力较弱。SSTN^[28] 结合空间注意力和光谱关联模块, 捕捉高光谱图像的长距离依赖和光谱-空间相关性, 但由于 FOD 的空间特征弱, SSTN 的效果并不突出。UNetFormer^[29] 结合 U-Net 和 Transformer 架构, 效果较好, 但硬质塑料的识别准度相对较低, 且对强反射特性易识别错, 没有较好地捕捉金属的特殊特性。

3.5 消融分析

在消融实验部分, 本文着重验证 6 波段相较于 3 波段(RGB)FOD 区分的必要性、在损失函数中引入 Tversky 指数的必要性及网络结构模块的合理性。

3.5.1 不同波段数下 FOD 的易分性

为验证多波段进行 FOD 检测与识别的重要性, 将原数据集中选择 470 nm、550 nm、634 nm 这 3 个波段(分别对应蓝色、绿色、红色)的数据生成 RGB 数据集进行实验, 结果如表 3 所示。RGB 的图像识别强依赖于形状与纹理信息, 而 FOD 的应用场景中缺乏该类信息, 因此, 在光谱维度增加信息量具有显著的优势。

3.5.2 损失函数消融

Focal Loss 是多分类问题域中样本分布不均时使用的主要损失函数。本文通过对比单一 Focal Loss 与本文采用的损失函数下训练模型得到的指标结果, 验证 Tversky 指数的必要性, 如表 4 所示。其中, Focal Loss 的定义为

$$FL(p) = -\frac{1}{N} \sum_{q=1}^C \alpha_q (1 - p_q)^\gamma \ln p_q \quad (17)$$

式中: N 为批量大小; p 为概率; p_q 为模型预测样本属于第 q 类的概率; $(1 - p_q)^\gamma$ 为聚焦因子, 降低易分类样本的损失贡献; $-\ln p_q$ 为标准交叉熵损失; α_q 为类别权重, 针对本文数据集关键材质数量远小于沥

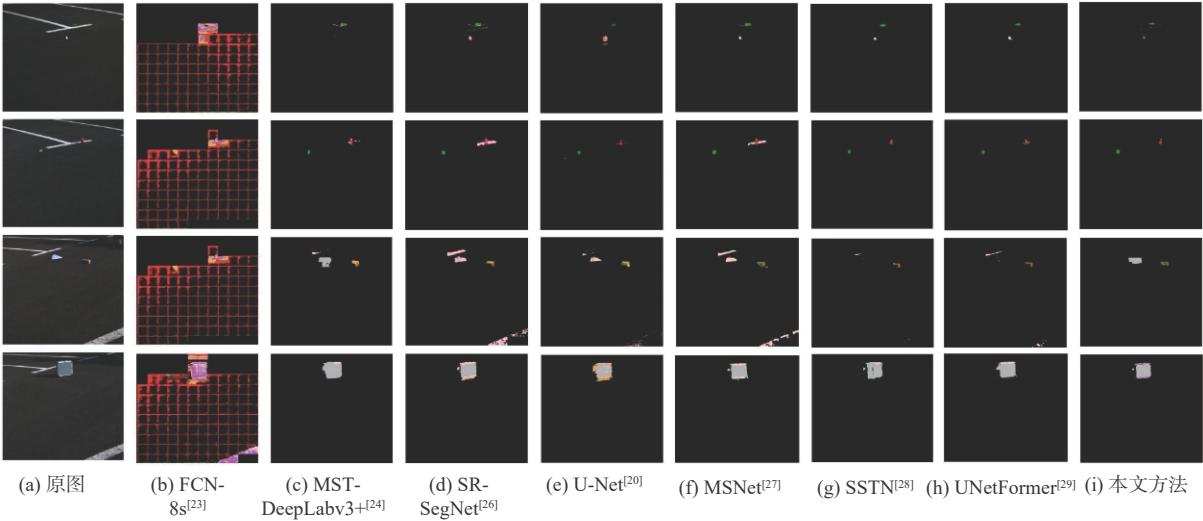


图 7 各方法的分割结果

Fig. 7 Segmentation results for each method

表 3 6 波段与 RGB 下模型的测试指标值

Table 3 Test values for 6-band and RGB sub-models

波段数	准确率	Macro-F1
3(RGB)	0.787 7	0.488 4
6	0.999 3	0.774 9

表 4 Tversky 指数加入与否的测试指标值

Table 4 Test values with or without Tversky index

损失函数	准确率	Macro-F1
Focal Loss	0.999 1	0.616 0
本文	0.999 3	0.774 9

青等背景材质的情况,将 α_q 设置为该类在训练集中的像素数的倒数,从而减少数据量不平衡带来的问题。

Focal Loss 虽然缓解了类别分布不均衡的问题,但无法针对假阳性或假阴性的误分类情况进行调整,导致容易出现漏检问题,其对于 FOD 的检测是致命的。加入 Tversky 指数,通过增加对漏检的惩罚,提高 FOD 各种类的召回率,使模型在 FOD 的检测上更加鲁棒。

3.5.3 网络结构

本文采用的密集连接编码器-解码器结构与 U-Net、UNet++类似。当本文结构去除密集嵌套的跳跃连接结构时,网络结构退化为 U-Net;当本文瓶颈层的卷积核与编码路径相同(都为 3×3)时,网络结构进化为 UNet++。UNet++的模型大小达到 35 347 KB,是本文方法的近 150%。原模型在瓶颈层进行更细致的特征提取,对于 Macro-F1 分数的提升具有重要作用,具体体现在同危险等级下不同材质的识别能力更强。本文方法与 UNet++的区别在于瓶颈层的卷积核大小,这样的改动极大降低了模型大小,但没有明显影响不同危险等级下材质的识别能力,

在检验危险等级的效果上也没有特别显著的下降,对于本文来说是可接受的,不同 U-Net 结构的测试结果与模型大小如表 5 所示。

表 5 不同 U-Net 结构的测试结果与模型大小

Table 5 Test results and model sizes for different U-Net structures

方法	准确率	Macro-F1	模型大小/KB
U-Net	0.998 9	0.642 0	30 391
UNet++	0.999 2	0.870 6	35 347
本文	0.999 3	0.774 9	23 151

经过对误检像素点的数据统计,观察到本文方法在锈迹、金属和强反射上的误检除了部分被认为是硬质白色塑料,大部分被误检为这 3 类中的另外 2 类。而金属和硬质塑料在危险等级同属最高危物品,因此,这几类的误检是可容忍的。与此同时,本文方法对其他几类材质的检测准确率与 UNet++差距不大。

3.6 恶劣场景

在 3.4 节中的分割结果图中,本文方法将阴影部分误检为树叶。虽然常见的 FOD 材质没有出现严重的“异物同谱”现象,但阴影等现象会与部分 FOD 产生该种情况。轮胎碎片作为一种 FOD 材料,主要构成为深色橡胶,其与沥青或水泥地面会出现较稳定的“异物同谱”现象。在进一步的优化与完善中,将考虑温度等信息,缓解“异物同谱”带来的误检问题。

此外,伪反射率方法仅限于背景地面基本为沥青和水泥等单一材质的情况。如果实际应用过程中的背景地面杂质较多,该方法大概率将失效,届时,仍需要使用经典的白板校正方法获得像素点的反射率信息。

4 结 论

- 1) 本文收集常见 FOD, 拍摄并构建了 FOD 的高光谱数据集。
- 2) 从目标本征特性出发, 通过采集并构建常见 FOD 数据集, 构建了针对 FOD 的目标材质属性建模方法。
- 3) 针对 FOD 缺乏表现特征的问题, 设计了基于编码器-解码器结构的密集嵌套连接的神经网络, 充分利用像素点的形信息与谱信息, 提高了 FOD 材质识别的准确度, 对于机场安全具有重要意义。

参考文献 (References)

[1] Federal Aviation Administration. Airport foreign object debris detection equipment: AC 150/5220-24—2009[S]. Washington, D. C. : USA Department of Transportation, 2010.

[2] 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.

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

[4] Cao X G, Wang P, Meng C, et al. Region based CNN for foreign object debris detection on airfield pavement[J]. Sensors, 2018, 18(3): 737.

[5] 韩松臣, 张比浩, 李炜, 等. 基于改进 Faster-RCNN 的机场场面小目标物体检测算法[J]. 南京航空航天大学学报, 2019, 51(6): 735-741.

 Han S C, Zhang B H, Li W, et al. Small target detection in airport scene via modified Faster-RCNN[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(6): 735-741(in Chinese).

[6] Munyer T, Brinkman D, Zhong X, et al. Foreign object debris detection for airport pavement images based on self-supervised localization and vision Transformer[C]//Proceedings of the 2022 International Conference on Computational Science and Computational Intelligence. Piscataway: IEEE Press, 2023: 1388-1394.

[7] 荆颖, 郑红, 郑文韬. 基于高斯混合特征改进的自编码器机场跑道 FOD 精准检测[J]. 计算机辅助设计与图形学学报, 2025, 37(9): 1593-1606.

 Jing Y, Zheng H, Zheng W T. Gaussian mixture feature improved autoencoder for FOD detection on airport runway[J]. Journal of Computer-Aided Design & Computer Graphics, 2025, 37(9): 1593-1606(in Chinese).

[8] Bell S, Upchurch P, Snavely N, et al. Material recognition in the wild with the materials in context database[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 3479-3487.

[9] Zhang H, Xue J, Dana K. Deep TEN: texture encoding network[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 2896-2905.

[10] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [EB/OL]. (2021-06-03)[2025-06-20]. <https://arxiv.org/abs/2010.11929>.

[11] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[EB/OL]. (2021-02-26) [2025-06-20]. <https://arxiv.org/abs/2103.00020>.

[12] Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis[EB/OL]. (2021-06-01)[2025-06-20]. <http://arxiv.org/abs/2112.10752>.

[13] Taupik J, Alamsyah T, Wulandari A, et al. Airport runway foreign object debris (FOD) detection based on YOLOX architecture [C]//Proceedings of the 2023 International Conference on Computer Science, Information Technology and Engineering. Piscataway: IEEE Press, 2023: 40-43.

[14] Tian Y, Li Z, Lin Y W, et al. Metal object detection for electric vehicle inductive power transfer systems based on hyperspectral imaging[J]. Measurement, 2021, 168: 108493.

[15] Munyer T, Huang P C, Huang C Y, et al. FOD-A: a dataset for foreign object debris in airports[EB/OL]. (2022-01-26)[2025-06-20]. <https://arxiv.org/abs/2110.03072>.

[16] Xu H Y, Han Z Q, Feng S L, et al. Foreign object debris material recognition based on convolutional neural networks[J]. EURASIP Journal on Image and Video Processing, 2018, 2018(1): 21.

[17] Barrow H G, Tenenbaum J M. Recovering intrinsic scene characteristics from images[J]. Computer Vision Systems, 1978, 2: 3-26.

[18] Liu F T, Ting K M, Zhou Z H. Isolation forest[C]//Proceedings of the 2008 Eighth IEEE International Conference on Data Mining. Piscataway: IEEE Press, 2009: 413-422.

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

[20] 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-MICCAI 2015. Berlin: Springer, 2015: 234-241.

[21] Salehi S S M, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks [C]//Proceedings of the Machine Learning in Medical Imaging. Berlin: Springer, 2017: 379-387.

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

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

[24] Wang Y, Yang L, Liu X Z, et al. An improved semantic segmentation algorithm for high-resolution remote sensing images based on DeepLabv3+[J]. Scientific Reports, 2024, 14: 9716.

[25] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision

and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.

[26] Weng W, Zhuo Z, Li H, et al. Semantic segmentation of remote sensing images based on dilated convolution and SegNet[J]. Journal of Applied Remote Sensing, 2020, 14(4): 046511.

[27] Tao C X, Meng Y Z, Li J J, et al. MSNet: multispectral semantic segmentation network for remote sensing images[J]. GIScience & Remote Sensing, 2022, 59(1): 1177-1198.

[28] Zhong Y F, Zhang L P, Huang B, et al. A spectral-spatial Transformer network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-14 .

[29] Wang L B, Li R, Zhang C, et al. UNetFormer: a UNet-like Transformer for efficient semantic segmentation of remote sensing urban scene imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 190: 196-214.

FOD detection and recognition for low-altitude takeoff and landing sites

ZHENG Haowen¹, HU Haimiao^{1, 2}, XU Zhuang^{3, *}, HE Zhuang², HU Haoxin²

- (1. School of Computer Science and Engineering, Beihang University, Beijing 100191, China;
2. Hangzhou Innovation Institute of Beihang University, Hangzhou 310000, China;
3. Zhejiang Provincial Airport Group Co., Ltd., Hangzhou 311207, China)

Abstract: The detection and recognition of foreign object debris (FOD) on low-altitude takeoff and landing runways are of critical importance to aircraft performance and airport operational safety. A FOD dataset was created and a physics-based FOD detection and identification approach was suggested in order to address the issues caused by the small size, lack of distinctive visual traits, and restricted availability of public data for FOD. Initially, multispectral images of common FOD materials were collected to establish a multispectral FOD dataset. Based on this dataset, attribute analysis was conducted on common FOD types, including metal rust and strong reflection phenomena. Utilizing the reflectivity of asphalt backgrounds, a pseudo-reflectivity metric relative to asphalt was derived for different materials. Through various statistical measures, the separability of different materials was preliminarily validated. Subsequently, a neural network with an encoder-decoder architecture featuring densely nested connections was proposed, using pseudo-reflectivity as input for material identification. The suggested approach outperformed existing material identification and segmentation models with similar parameter ranges on the test set of this dataset, achieving an accuracy of 0.999 3 and a Macro-F1 score of 0.774 9.

Keywords: material recognition; airport runway foreign object debris; autoencoder; multispectral image; pseudo-reflectivity

低空安防场景下多目标高精度实时三维检测技术

俞佳楠¹, 徐壮², 李旸晖¹, 叶德茂^{1,*}

(1. 中国计量大学 光学与电子科技学院, 杭州 310018; 2. 浙江省机场集团有限公司, 杭州 311207)

摘 要: 针对低空安防场景下无人机、鸟类等动态目标的精确测距需求,提出一种基于双目视觉的多目标高精度实时三维检测方法。该方法在半全局块匹配(SGBM)算法基础上,引入双边滤波、形态学操作和插值空洞填充等视差优化策略,提升弱纹理目标的匹配鲁棒性和深度图完整性。通过结合三角测量与相机标定参数,实时求解目标的三维坐标。实验表明,在0.5~2.0 m范围内,视差空洞率由40.700 4%降至1.278 1%,平均测距相对误差为2.55%,关键点定位精度达厘米级。缩比建模进一步验证了所提方法在远距离场景中的可扩展性,理论上可将有效范围扩展至2 km,为低空防控领域提供了一种高精度三维感知方案。

关 键 词: 双目视觉; 半全局块匹配; 视差优化; 深度测量; 三维检测; 缩比模型

中图分类号: TP391.41; TP722.4

文献标志码: A

文章编号: 1001-5965(2026)09-3163-09

随着低空经济崛起,在低空场景中,各类飞行载具(如电动垂直起降航空器(electric vertical take-off and landing, eVTOL)、物流无人机)与动态目标(如无人机群、鸟类)在低空域交织共融,构成了一个高度复杂、动态不确定的感知环境。物流配送、低空安防等核心应用场景也对飞行载具的实时、鲁棒三维环境感知能力提出了前所未有的要求。低空场景下的目标主要包括集群鸟类、微型无人机及低空漂浮物等,通常具有体积小、纹理特征弱、易与复杂背景(如天空云朵)混淆及多目标间存在遮挡等特点,尤其是远距离小型无人机或鸟类在图像中分辨率极低、纹理特征微弱,其密集聚集也会导致目标间频繁、严重的遮挡与交互,引发身份混淆,都为三维感知带来了全新挑战^[1-2]。在上述背景下,传统单目视觉因缺乏多视角或几何约束,单帧图像无法直接获取深度信息^[3]。此外,单目视觉的深度信息获取依赖于图像特征的全局与局部分析,如亮度、阴影、纹理梯度等线索,但这些线索在不同场景中可能不充分或存在歧义^[4]。相比之下,双目视

觉基于三角测量原理,通过基线与视差的几何关系直接解算深度,兼具被动测距与高帧率优势,成为低空感知的理想方案^[5];在突发鸟类群、非法无人机等低空动态目标检测中,其三维实时定位能力对提升飞行安全至关重要。

双目视觉系统通过模拟人类双眼感知原理,利用2台相机从不同视角获取场景图像,经立体匹配计算视差以重建三维信息,在非接触测量领域具有独特优势。早期局部窗口匹配算法(如绝对差和(sum of absolute differences, SAD)、平方差和(sum of squared differences, SSD))依赖像素邻域灰度差异,在纹理缺失区域易产生误匹配。随着算法需求的提升,全局匹配算法如图割、信念传播等被引入,通过构建能量函数优化全局视差,显著提升匹配精度,但计算复杂度较高,难以满足实时性需求^[6]。2008年, Hirschmüller提出的半全局块匹配(semi-global block matching, SGBM)算法融合了局部与全局算法的优势,通过多路径代价聚合降低计算复杂度,同时保持较高精度,成为当前应用最广泛的立

收稿日期: 2025-07-02; 录用日期: 2025-09-29; 网络出版时间: 2025-10-15 09:56

网络出版地址: link.cnki.net/urlid/11.2625.V.20251014.1552.002

基金项目: 国家自然科学基金(U24A20307)

* 通信作者. E-mail: yedemao@aliyun.com

引用格式: 俞佳楠, 徐壮, 李旸晖, 等. 低空安防场景下多目标高精度实时三维检测技术[J]. 北京航空航天大学学报, 2026, 52(9): 3163-3171.
Weng J N, Xu Z, Li Y H, et al. High-accuracy real-time 3D multi-target detection for low-altitude security scenarios[J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52(9): 3163-3171 (in Chinese).

体匹配算法之一,但多目标场景下的视差连续性与实时性仍存挑战:动态目标适应性不足,如深度学习模型依赖静态场景训练集(如 KITTI, KITTI Vision Benchmark Suite, KITTI 视觉基准测试套件),对快速移动的小目标泛化能力不足;实时性-精度难以权衡,如神经网络模型难以满足 eVTOL 毫秒级响应需求^[7-8]。近年来,立体匹配已逐步向深度学习驱动的端到端模型方向演进,如 MADNet 和 AdaStereo 等算法,通过引入持续学习、颜色转移等关键技术手段增强跨域场景下的鲁棒性;与此同时,依托注意力机制、多尺度信息融合的轻量化算法亦持续涌现,但其在提升模型效率的同时,也显著增加了计算复杂度与硬件运算成本^[9-10],因此,亟需提出能够兼顾模型精度、推理实时性与部署成本的新方法。

针对上述挑战,本文提出一种基于改进 SGBM 算法的多目标实时三维检测方案。该方案通过构建多阶段的视差优化流程提升匹配鲁棒性;建立高精度三维坐标计算模型,并结合高精度相机标定与三角测量模型,实现对动态目标的厘米级定位;通过在 0.5~2.0 m 范围内进行测距精度实验,以及构建动态缩比扩展方案,验证本文方法在远距离场景(2 km)的有效性。

1 双目视觉多目标三维检测技术框架设计

如图 1 所示,双目立体视觉三维检测技术工作

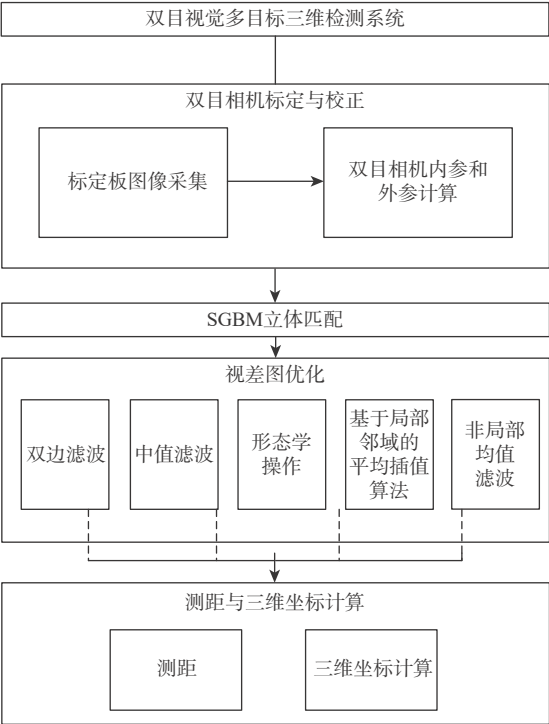


图 1 双目视觉多目标三维检测框

Fig. 1 Binocular vision multi-target 3D detection

流程主要分为 4 个核心模块,各模块功能如下:

1) 相机标定与校正:采用张正友法采集 20 组棋盘格图像,求解相机内外参与畸变系数,消除相机畸变;通过立体校正将倾斜极线转为水平平行,减少匹配搜索范围。

2) 立体匹配:基于改进 SGBM 算法获取初始视差图。

3) 视差图优化:依次执行双边滤波、中值滤波、形态学操作、插值空洞填充及非局部均值滤波,提升视差图完整性。

4) 距离与三维坐标计算:基于三角测量原理,结合标定参数与视差值,通过公式求解目标深度与世界坐标。

2 双目视觉三维检测技术

2.1 双目相机标定原理

Matlab 中的 stereoCameraCalibrator 工具采用张正友标定法^[11]进行双目相机系统标定,该工具通过非线性优化求解相机的内外参数,以消除镜头畸变,并建立左右相机的相对位姿关系。进行标定前,采集 20 组黑白棋盘格标定板(规格:角点数为 9×6、方格尺寸为 28 mm)在多位姿下的左右相机同步图像,如图 2(a)和 2(b)所示;标定算法将自动提取图像中的棋盘格角点,构建像素坐标与世界坐标的对应点集;在相机成像模型中,三维世界点坐标 (X_w, Y_w, Z_w) 通过旋转 (R) 和平移变换 (T) 至相机坐标系 (X_c, Y_c, Z_c) ,再经投影和离散化转换至像素坐标系 (u, v) ,核心公式为

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f/d_x & 0 & u_0 & 0 \\ 0 & f/d_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \chi_w = \begin{bmatrix} f/d_x & 0 & u_0 & 0 \\ 0 & f/d_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \chi_c \quad (1)$$

式中: $\chi_w = [X_w, Y_w, Z_w, 1]^T$ 和 $\chi_c = [X_c, Y_c, Z_c, 1]^T$ 分别为在世界坐标系和相机坐标系下的齐次坐标; (X_w, Y_w, Z_w) 、 (X_c, Y_c, Z_c) 、 (u, v) 分别为目标点在世界坐标系、相机坐标系及像素坐标系下的坐标; $K_1 = \begin{bmatrix} f/d_x & 0 & u_0 & 0 \\ 0 & f/d_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ 、 $K_2 = \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix}$ 分别为相机的内、外参矩阵; R 和 T 为世界坐标系到相机坐标系的正交旋转矩阵和平移矩阵;焦距 f 为光心到成像平面的距离; (u_0, v_0) 为成像平面上的主点坐标; d_x 和 d_y 为离散化单位长度。

因此,当目标点深度 Z_c 已知时,可通过逆变换

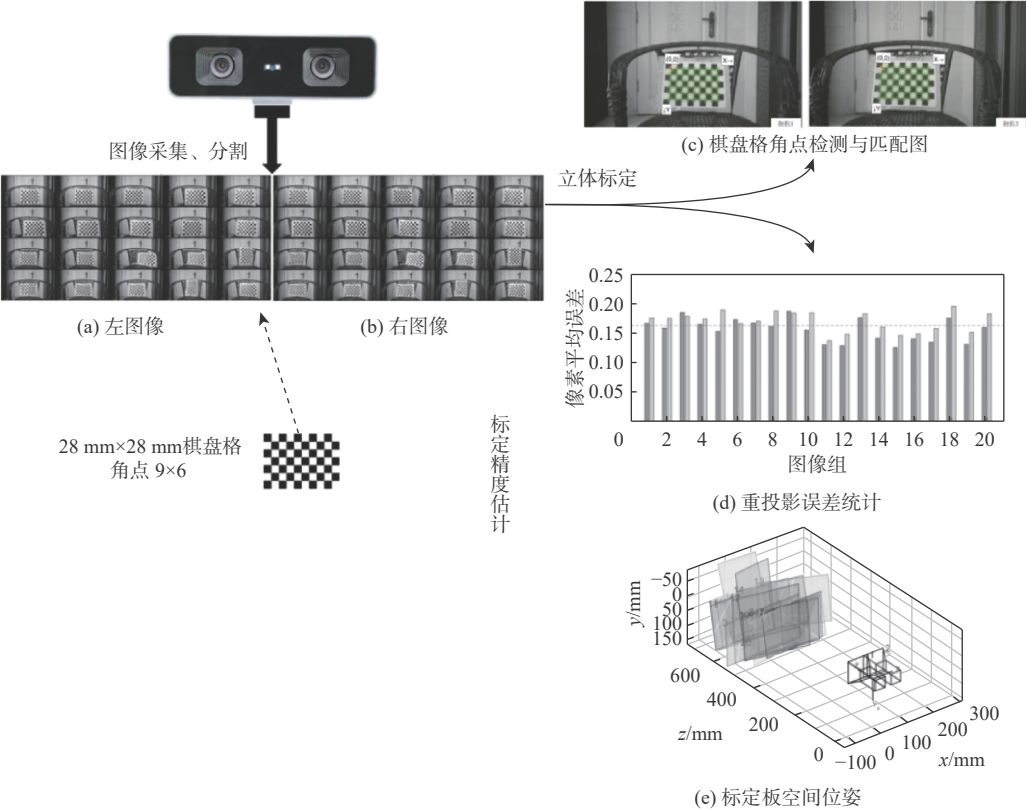


图 2 双目标定原理及标定精度估计

Fig. 2 Diagram illustrating the dual-target calibration principle and accuracy estimation plots

从像素坐标推导其世界坐标。具体地,有

1) 像素坐标系→相机坐标系:

$$\begin{cases} X_c = \frac{(u-u_0)d_xZ_c}{f} \\ Y_c = \frac{(v-v_0)d_yZ_c}{f} \end{cases} \quad (2)$$

2) 相机坐标系→世界坐标系:

$$\chi_w = \begin{bmatrix} R^T & -R^T T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = K_2^{-1} \chi_c \quad (3)$$

如图 2(d)所示,双目相机在多图图像标定中的像素重投影平均误差为 0.18 像素。在相机处于固定状态条件下,相机与标定靶标的空间位置关系如图 2(e)所示。可见,所有标定板均位于相机视场锥体内,其空间分布与实际拍摄位姿一致,验证了标定结果的可靠性。

2.2 双目立体校正

原始双目图像中,对应点位于倾斜的极线上,目标点的匹配搜索需沿二维曲线进行。如图 3 所示,利用标定结果对 2 台相机进行共面行对准后,极线转为水平平行,此时匹配点搜索仅需沿图像行方向进行,大大减少了计算量。设左视图上有某点 p_l ,在视差 d 下对应于右视图上的匹配点 p_r ,则两者满足关系式:

$$\begin{cases} x_r = x_l - d \\ y_l = y_r \end{cases} \quad (4)$$

式中: (x_l, y_l) 和 (x_r, y_r) 分别为左右图像中对应点的图像物理坐标。

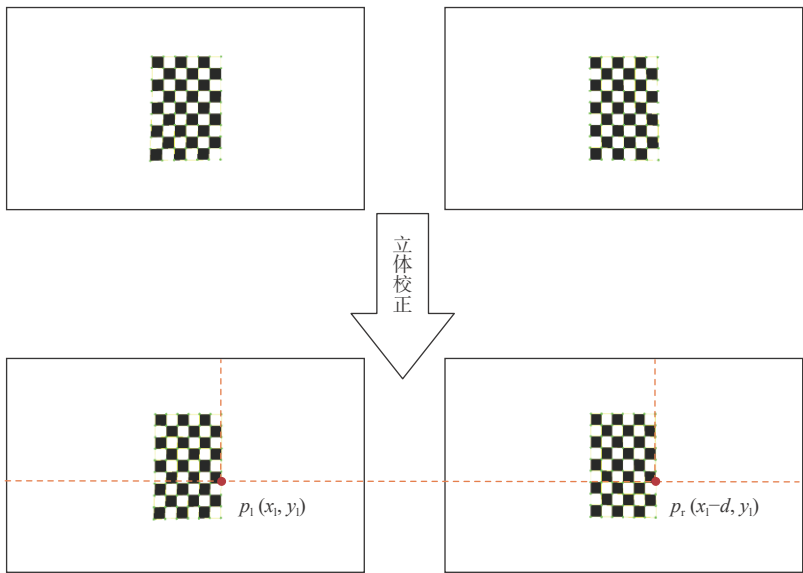
2.3 SGBM 立体匹配及视差图优化

2.3.1 SGBM 算法原理

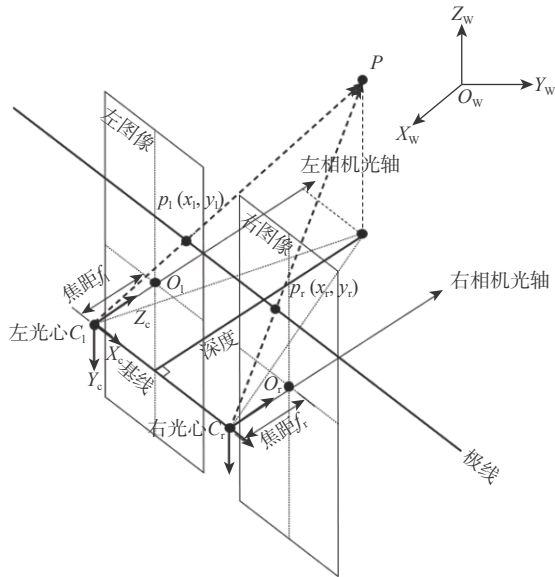
针对传统 SAD/SSD 在深度不连续区域的敏感性问题,本文采用改进 SGBM 算法,以 BT(Birchfield-Tomasi)亚像素度量^[12]替代互信息(mutual information, MI)作为左右图像中每个像素在不同视差下的匹配代价。BT 度量基于线性插值拟合亚像素精度的灰度值,通过聚合水平方向多个路径的代价信息,寻找最优匹配,相较于传统的 SAD 或 SSD,能更准确地刻画像素间的相似性。

如图 3(c)和 3(d)所示,在同步采集的图像对扫描线上,同一行灰度值函数分别为 I_L 、 I_R ,对于每一匹配序列 M ,左图上每一像素点记作 p_l ,在视差 $d(D_{\min} \leq d \leq D_{\max}$, $D_{\min}$ 、 $D_{\max}$ 分别为最小和最大误差)下,对应的右图像素点记为 p_r 。

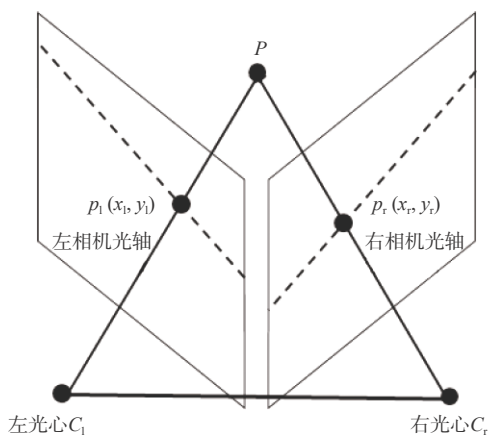
为提高衡量像素间相关性,引入总代价函数 $\gamma(M)$,其目标是使匹配对的代价尽可能小。具体而言,对于左图像中每个像素,在右图像中沿极线方向在一定视差范围内平移匹配窗口,计算不同视差下的匹配代价,并选取聚合代价最小的视差作为最终视差。



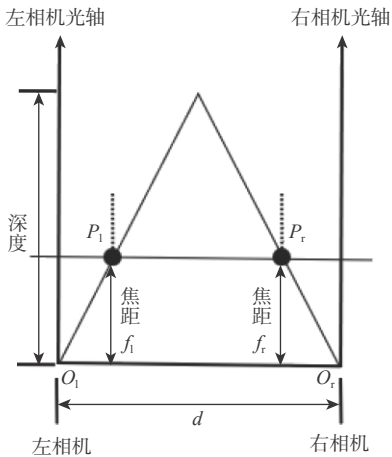
(a) 立体校正前后图像



(b) 平行式双目立体视觉测量模型



(c) 立体校正前极线几何示意图



(d) 立体校正后极线几何示意图

图 3 立体校正原理

Fig. 3 Stereo rectification principle diagram

2.3.2 视差图优化方法

受噪声、遮挡和纹理缺失等因素影响, SGBM

立体匹配算法生成的原始视差图常存在大量杂点与空洞,需通过后处理提升质量。后处理阶段采用

中值滤波、双边滤波、形态学操作及插值等操作,逐步消除视差噪声和空洞,详见图4。

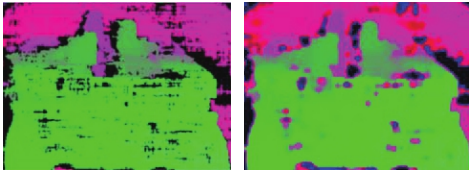


图4 伪彩色视差图优化前后对比

Fig. 4 Comparison of pseudo-color disparity maps before and after optimization

考虑到视差图的边缘通常对应场景中物体的边界,保留这些结构细节至关重要。针对低空鸟类等目标尺寸小、边缘锐利的特点,为有效避免滤波过程中的轮廓信息丢失,本文采用具有边缘保持特性的双边滤波^[13]。双边滤波通过联合优化空间邻近度和像素值相似度的权重,有效抑制视差平面内的噪声干扰,同时,最大程度地保留锐利的物体边缘信息,实现噪声平滑与边缘保持的平衡。双边滤波对边缘信息虽有较好的保留,但仍不能较好地消除噪声,为进一步抑制视差图中典型的椒盐噪声,引入中值滤波。

针对图中大量空洞问题(该类空洞在鸟类羽毛、无人机光滑表面等区域尤为显著),本文综合运用形态学操作,基于局部邻域平均的插值等策略,逐步检测并填充这些空洞:先通过闭运算连接视差断裂区域、弥合小尺度空洞,再基于有效邻域视差均值进行插值,逐步恢复空洞区域的视差连续性,显著提升弱小目标的视差完整性。

利用非局部均值滤波进一步精细去噪,提升视差图整体质量,并自定义伪彩色映射函数将视差图以伪彩色形式呈现,图4进行了视差图优化前后的对比。

2.4 深度与三维坐标计算

如图2(a)所示,左右图像经校正后,极线呈现水平对齐。根据三角测量法,视差 d 与深度 Z 的几何关系为

$$Z = \frac{fb}{d} \quad (5)$$

式中: Z 为目标点到相机的深度; $b = |\overrightarrow{C_1C_2}|$ 为2个相机光心之间的基线距离。

结合立体校正后改变的相机内参矩阵及得到的重投影矩阵 $\mathbf{Q}$,可以把像素坐标系重投影为世界坐标系(此时,定义世界坐标系与左相机坐标系重合),进而解算目标点的三维世界坐标 (X_w, Y_w, Z_w) ,即

$$\begin{cases} X_w = \frac{bx_1}{d} \\ Y_w = \frac{by_1}{d} \\ Z_w = \frac{bf}{d} \end{cases} \quad (6)$$

目标点与相机之间的距离 D 为

$$D = \sqrt{X_w^2 + Y_w^2 + Z_w^2} \quad (7)$$

3 测距与定位实验设计与结果分析

3.1 实验设置

实验采用具有稳固支架固定的 GVD200 双目相机(分辨率为 1 280 像素×720 像素,基线距离为 60.00 mm,焦距为 3.30 mm,已完成内外参标定)。标定板为 9×6 角点棋盘格(方格尺寸为 28 mm×28 mm)。硬件平台为 Intel i9-12900H CPU;软件环境为 OpenCV 4.11.0 + Clion 2023。在标准室内环境中,选取 14 cm×36 cm 的矩形平台,在其上按 3 角形布设 3 只同规格的仿真麻雀目标物(体长为 18 cm)。选取 3 个不同距离、不同位置的关键点进行多组数据采集,对比测量值与真实值的误差,验证算法在双目相机有效工作距离范围内(0.25~3.5 m)对多个相同目标进行三维位置 (x, y, z) 检测的精度。

实验通过左右 2 张对应的场景图经过 SGBM 立体匹配得到对应的视差图(部分如图5所示),再利用相机矩阵和深度值逆投影为三维点云。



图5 鸟类点云投影图

Fig. 5 Projection map of bird point cloud

3.2 视差图评估

视差图通过表征像素间的视差信息,为物体距离与尺寸估计提供基础数据,其质量直接影响三维重建精度。本文实验采用渐进式优化策略,对拍摄后的图像进行视差图输出,定量评估各阶段优化效果,如表1所示。可见,经多项优化处理,空洞率从基准值 40.700 4% 降至 1.278 1%,有效视差均值从 105.226 0 稳定至 66.724 3。其中,基础插值对空洞修复贡献最大(较上一级优化过程的相对降幅约 93.8%),表明处理后的深度感知更聚焦前景物体,视差分布符合低空安防场景三维重建的基本需求。

表 1 视差图评估指标

Table 1 Evaluation metrics of disparity maps

优化措施	空洞率/%	有效视差均值	有效视差标准差	有效视差最小值	有效视差最大值
无任何优化	40.700 4	105.226 0	26.795 5	2	255
添加中值滤波	40.178 2	105.171 0	26.851 2	2	255
添加双边滤波	26.332 9	85.594 9	40.453 4	1	254
添加形态学操作	24.885 6	87.183 3	39.624 7	1	242
添加基础插值	1.543 1	66.820 3	50.321 5	1	255
添加非局部均值滤波	1.278 1	66.724 3	50.388 0	1	242

3.3 测距精度评估

表 2~表 4 展示了 4 组目标关键点的测量距离与真实距离误差。在 0.5~1.5 m 核心测距带,系统平均相对误差为 2.55%,绝对误差稳定在 2 cm 内。

表 2 目标关键点测距数据

Table 2 Ranging data of target key points

组别	像素坐标/像素	世界坐标/m	测量距离/m	真实距离/m
1	(485,565)	(-0.056, 0.076, 0.326)	0.34	0.31
	(656,525)	(0.011, 0.077, 0.413)	0.42	0.40
	(729,547)	(0.043, 0.080, 0.379)	0.39	0.38
2	(571,460)	(-0.056, 0.086, 0.787)	0.79	0.79
	(626,436)	(-0.007, 0.075, 0.909)	0.91	0.89
	(683,443)	(0.051, 0.078, 0.868)	0.87	0.87
3	(688,384)	(0.077, 0.026, 1.189)	1.19	1.17
	(750,376)	(0.169, 0.015, 1.237)	1.25	1.28
	(772,382)	(0.200, 0.024, 1.231)	1.25	1.23
4	(578,456)	(-0.117, 0.196, 1.865)	1.89	1.84
	(605,453)	(-0.063, 0.201, 1.981)	1.99	1.94
	(673,493)	(0.085, 0.267, 1.800)	1.82	1.82

表 3 目标关键点测距绝对误差统计

Table 3 Absolute error statistics of target key

point ranging				m
组别	点1绝对误差	点2绝对误差	点3绝对误差	组平均绝对误差
1	0.030	0.020	0.010	0.020
2	0	0.020	0	0.007
3	0.020	-0.030	0.020	0.010
4	0.050	0.050	0	0.033

注:组平均绝对误差平均值为0.018 m。

3.4 多立体匹配方法对比实验

选取当前三维检测领域的其中 3 类主流立体匹配方法块:匹配(block matching, BM)、传统 SGBM 以及本文 SGBM (GPU 加速)进行定量与定性对比,实验条件统一为 Intel i9-12900H CPU+1 280×720 图像分辨率,对比指标包括视差图空洞率、平均相对误差、运行时间及视差图效果,结果如表 5 所示。

表 4 目标关键点测距相对误差统计

Table 4 Relative error statistics of target key

point ranging				%
组别	点1相对误差	点2相对误差	点3相对误差	组平均相对误差
1	9.68	5.00	2.63	5.77
2	0	2.25	0	0.75
3	1.71	2.34	1.63	1.89
4	2.72	2.58	0	1.77

注:组平均相对误差平均值为2.55%。

表 5 立体匹配方法指标对比

Table 5 Comparison of indicators for stereo

matching methods			
核心算法	空洞率/%	平均相对误差/%	运行时间/ms
BM	1.390 0	7.20	413.90
传统SGBM	40.700 4	4.80	241.29
本文SGBM(GPU加速)	1.278 1	2.55	22~35

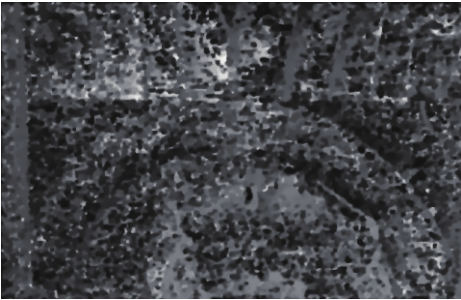
结合表 5 数据及图 6 的匹配视差图,本文算法在实时性及准确性等方面表现出色。BM 算法生成的视差图(图 6(a))存在大量噪声,对场景细节的刻画不足;传统 SGBM 算法的视差图(图 6(b))虽在一定程度上改善了噪声问题,但在边缘和纹理欠丰富区域,仍出现明显的视差不连续与错误;相比之下,经本文算法优化后的视差图(图 6(c))则呈现出清晰连续的边缘细节,即使在弱纹理区域也能实现准确的视差估计,视差图的完整性与精确性均得到显著提升。

3.5 缩比实验原理与远距离检测扩展

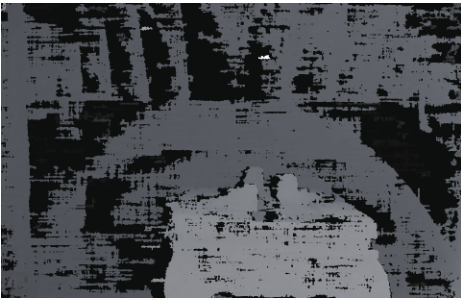
为进一步验证算法在远距离场景的适用性,基于三角测量原理建立缩比模型,该方法参考缩比实验的通用思路^[14-15],对双目相机的 2 个核心几何参数(焦距 f 与基线距离 b)进行缩放,进而推演检测系统在扩展距离下的测距性能。根据深度计算方程式(5),设基准焦距和基线距分别为 f_0 、 b_0 ,基准距离为 Z_0 ,理论扩展距离 $Z_1=kZ_0$ (k 为距离缩放因子且 $k>1$)。为保持视差 d 不变,需满足:

$$\frac{f_1b_1}{f_0b_0}=k$$

(8)



(a) BM匹配视差



(b) 传统SGBM匹配视差



(c) 本文SGBM匹配视差

图 6 三维检测方法视差图对比

Fig. 6 Comparison of disparity map in 3D detection methods

若焦距不变, 则基线距扩展为

$$f_1 = kf_0 \tag{9}$$

若基线距不变, 则焦距扩展为

$$b_1 = kb_0 \tag{10}$$

若焦距与基线同步扩展, 则

$$\begin{aligned} f_1 &= \sqrt{k}f_0 \\ b_1 &= \sqrt{k}b_0 \end{aligned} \tag{11}$$

即通过放大焦距与基线距, 可在保持视差计算精度的同时扩展检测范围。在远距离场景下, 绝对误差虽随距离增加, 但相对误差仍与近距离一致 (2.55%), 避免传统单一缩放基线导致的视差精度断裂问题。

3.5.1 缩比模型建立

将实测数据根据式(11)按一定比例进行缩放, 即求得目标检测范围, 具体参数设置如表 6 所示。

3.5.2 远距离检测误差分析

在缩比模型中, 扩展深度误差可近似表示为

表 6 缩比参数
Table 6 Scale factors

原始测距范围/m	距离缩放因子 <i>k</i>	扩展焦距 <i>f</i> ₀ /mm	扩展基线距 <i>b</i> ₁ /mm	扩展距离 <i>Z</i> ₁ /m
0.5~2.0	10	3.162 <i>f</i> ₀ = 10.43	3.162 <i>b</i> ₀ = 189.72	5.0~20.0
0.5~2.0	100	10 <i>f</i> ₀ = 33.00	10 <i>b</i> ₀ = 600.00	50.0~200.0
0.5~2.0	1 000	31.623 <i>f</i> ₀ = 104.36	31.623 <i>b</i> ₀ = 1 897.38	500.0~2 000.0

$$\Delta Z_1 = \frac{f_1 b_1}{d^2} \Delta d = \frac{Z_1^2}{f_1 b_1} = k \left(\frac{Z_0^2}{f_0 b_0} \Delta d \right) = k \Delta Z_0 \tag{12}$$

式中: ΔZ_1 为扩展测距的绝对深度误差; Δd 为视差误差; ΔZ_0 为原始测距的绝对深度误差。

在理想情况下, 扩展测距的绝对深度误差 ΔZ_1 增大为实验数据的 k 倍, 而相对深度误差 $\Delta Z_1/Z_1$ 仍保持不变。由表 6 可得, 当距离扩展为 2 km 时, 理论相对误差预计仍能控制在 2.55% 以内。

4 结 论

本文面向低空安防场景的实际需求, 提出一种基于双目视觉的多目标三维检测技术。通过改进 SGBM 算法与多阶段视差优化, 提升弱纹理目标的测距与定位精度, 实验表明:

- 1) 在自搭建的场景下, 生成的视差图的空洞率从 40.700 4% 降至 1.278 1%, 有效视差均值从 105.226 0 稳定至 66.724 3。
- 2) 在 0.5~2.0 m 的测距范围内, 多目标的测距平均相对误差为 2.55%, 同时, 基于缩比模型的理论分析表明, 本文方法具备扩展至 2 km 远距离检测的潜力, 且可保持相对误差基本不变。
- 3) 在近距离测试中取得了厘米级定位精度。
- 4) 整体硬件成本低于激光雷达方案, 适合规模化部署于低空载具。

尽管本文改进算法在精度和实时性方面表现出良好的性能, 但仍存在一些局限有待进一步改进:

- 1) 当前实验在标准室内环境下进行, 未充分考虑户外低空环境中常见的强光、雾霾、雨雪等恶劣天气对图像质量的影响, 这些条件可能导致图像对比度下降、噪声增加, 进而影响立体匹配与视差计算的稳定性。
- 2) 虽通过缩比模型推演了 2 km 的理论性能, 但实际环境中大气扰动、能见度等因素会影响远距离成像质量, 可能会限制算法的实际应用范围。
- 3) 实验采用刚性麻雀模型, 未能充分反映真实鸟类等非刚性目标的形变和姿态变化, 本文算法对此类目标的适应能力仍需进一步验证。

参考文献（References）

- [1] 张超. 基于信息补偿和特征增强的机场低空飞鸟检测方法研究[D]. 天津: 天津理工大学, 2024.
Zhang C. Research on low-altitude flying birds detection method in airport based on information compensation and feature enhancement[D]. Tianjin: Tianjin University of Technology, 2024(in Chinese).
- [2] Yang Z, Li N, Mao Y, et al. A review of research on safety assurance technology for UAV operation in low-altitude airspace[J]. Journal of Xihua University(Natural Science Edition), 2024, 43(1): 41-47.
- [3] 王皓, 曹静, 胡佳楠, 等. 基于 SfM-MVS 的作物三维重建: 挑战与创新[J]. 江苏农业学报, 2024, 40(9): 1768-1776.
Wang H, Cao J, Hu J N, et al. Three-dimensional crop reconstruction based on structure from motion-multiple view stereo(SfM-MVS): challenges and innovations[J]. Jiangsu Journal of Agricultural Sciences, 2024, 40(9): 1768-1776(in Chinese).
- [4] 孙丹, 吴克伟, 孙永宣, 等. 单目图像深度结构恢复研究[J]. 计算机科学与应用, 2018, 8(4): 522-531.
Sun D, Wu K W, Sun Y X, et al. Research for scene depth structure of monocular image[J]. Computer Science and Application, 2018, 8(4): 522-531.
- [5] Chen Z Y, Luo X Y, Dai B C. Design of obstacle avoidance system for micro-UAV based on binocular vision[C]//Proceedings of the 2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration. Piscataway: IEEE Press, 2018: 67-70.
- [6] Cheng Y, Wu Z T, Shi X Y, et al. Research of 3D reconstruction and position estimation of 3D pieces based on binocular structured light peripheral scanning[J]. Optics and Precision Engineering, 2025, 33(2): 337-347.
- [7] 赵晨园, 李文新, 张庆熙. 双目视觉的立体匹配算法研究进展[J]. 计算机科学与探索, 2020, 14(7): 1104-1113.
Zhao C Y, Li W X, Zhang Q X. Research and development of binocular stereo matching algorithm[J]. Journal of Frontiers of Computer Science & Technology, 2020, 14(7): 1104-1113(in Chinese).
- [8] 杨晓立, 徐玉华, 叶乐佳, 等. 双目立体视觉研究进展与应用[J]. 激光与光电子学进展, 2023, 60(8): 170-186.
Yang X L, Xu Y H, Ye L J, et al. Research progress on binocular stereo vision applications[J]. Laser & Optoelectronics Progress, 2023, 60(8): 170-186(in Chinese).
- [9] 陈舒雅. 基于深度学习的立体匹配技术研究[D]. 杭州: 浙江大学, 2022.
Chen S Y. Study on deep learning based stereo matching technologies[D]. Hangzhou: Zhejiang University, 2022(in Chinese).
- [10] 李云涛. 基于深度学习的立体匹配技术研究[D]. 成都: 西南交通大学, 2023.
Li Y T. Stereo matching technique based on deep learning[D]. Chengdu: Southwest Jiaotong University, 2023(in Chinese).
- [11] 李明彩, 郭轩, 于毅. 基于 Matlab 平台的相机标定研究[J]. 数字技术与应用, 2018, 36(2): 85-87.
Li M C, Guo X, Yu Y. Analysis of camera calibration based on Matlab platform[J]. Digital Technology and Application, 2018, 36(2): 85-87(in Chinese).
- [12] Birchfield S, Tomasi C. Depth discontinuities by pixel-to-pixel stereo[J]. International Journal of Computer Vision, 1999, 35(3): 269-293.
- [13] 王镭. 双边滤波中关键技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2017.
Wang L. Key technology research of bilateral filter[D]. Harbin: Harbin Engineering University, 2017(in Chinese).
- [14] 李肖成, 惠旭龙, 白春玉, 等. 民机结构坠撞性能缩比实验方法研究[J]. 爆炸与冲击, 2025, 45(7): 071411.
Li X C, Hui X L, Bai C Y, et al. Research on scaled experimental method of civil aircraft crash performance[J]. Explosion and Shock Waves, 2025, 45(7): 071411(in Chinese).
- [15] 尹宜勇, 张伟杰, 白翰钦. 大型管道系统缩比模型模态实验平台设计[J]. 实验技术与管理, 2021, 38(8): 90-93.
Yin Y Y, Zhang W J, Bai H Q. Design of modal experiment platform for scaled model of large-scale pipeline system[J]. Experimental Technology and Management, 2021, 38(8): 90-93(in Chinese).

High-accuracy real-time 3D multi-target detection for low-altitude security scenarios

WENG Jia’nan¹, XU Zhuang², LI Yanghui¹, YE Demao^{1,*}

(1. College of Optical and Electronic Technology, China Jiliang University, Hangzhou 310018, China;
2. Zhejiang Province Airport Group Co., Ltd., Hangzhou 311207, China)

Abstract: A real-time 3D detection method based on binocular vision is proposed to meet the high-accuracy ranging requirements of dynamic targets in low-altitude security scenarios. The method, which is based on the semi-global block matching (SGBM) algorithm, combines morphological processing, bilateral filtering, and interpolation-based hole-filling to enhance the completeness of depth information and disparity quality in weak-texture areas. With calibrated camera parameters and the triangulation principle, target 3D coordinates are computed in real time. Experiments show that within the distance range of 0.5-2.0 m, the disparity hole rate is reduced from 40.700 4% to 1.278 1%, with an average relative ranging error of 2.55% and centimeter-level positioning accuracy. The method’s scalability to 2 000 meters is further confirmed by scale model study. Overall, the proposed system provides an efficient and cost-effective solution for multi-target 3D perception in low-altitude surveillance.

Keywords: binocular vision; semi-global block matching; disparity optimization; depth estimation; 3D detection; scaled model

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0429

基于伪标签增强与多特征融合的毫米波雷达 无人机检测

鲁亚宁¹, 张生军^{1,*}, 于钱米², 李洪明¹

(1. 重庆交通大学 信息科学与工程学院, 重庆 400074; 2. 宁波华东核工业勘察设计院集团有限公司, 宁波 315040)

摘 要: 针对毫米波雷达无人机检测领域, 开展了基于多特征融合与伪标签增强的弱监督学习方法研究。采用恒定虚警率 (CFAR) 算法对毫米波雷达回波信号进行目标预检测, 从中提取频谱能量分布特征、运动轨迹变化特征及统计分布特征, 综合反映无人机的距离、速度及运动状态; 利用 LogitBoost 提升算法构建分类模型, 通过迭代加权弱分类器, 有效提高了分类准确性。为缓解标注样本不足问题, 进一步引入伪标签自训练机制, 并通过人工辅助修正, 实现了未标注数据的利用, 并提升了模型的泛化能力。基于自建的毫米波雷达数据集, 开展了多轮实验验证, 结果表明: 所提方法在 3 种不同场景的弱监督条件下, 接收者工作特征曲线下面积 (AUC) 分别提升约 7.2%/1.6%/6.6%, 虚警率降低约 36.9%/0%/10.6%, 检测准确率较传统方法有显著提高。此外, 所提方法表现出较强的环境适应能力, 能够在复杂背景下稳定识别小型飞行目标。研究表明, 多特征融合结合伪标签增强策略为毫米波雷达无人机检测提供了一种有效的解决方案, 具备较高的实用价值和推广潜力。

关 键 词: 毫米波雷达; 无人机检测; 多特征融合; 伪标签学习; 恒定虚警率

中图分类号: TN957.51

文献标志码: A

文章编号: 1001-5965(2026)09-3172-11

随着小型无人机市场的发展, 无人机在各个领域的应用随处可见: 借助无人机进行农林植保不仅可以降低成本, 还能实现良好的增产效果。此外, 无人机在应急救援、消防灭火、景区管理、物流配送及无人机灯光表演领域均有广泛应用。虽然无人机在诸多领域都有其优势, 但其快速发展也带来了诸如非法飞行等安全、隐私问题。面对这些问题, 无人机的检测显得尤为重要。

目前, 已知的无人机检测方法有很多, 包括视觉检测法, 如 Zhang^[1] 提出的多尺度 YOLOv8 改进算法用于无人机小目标检测, 在 VisDrone 数据集上检测效果优异; 声学检测法, 如 Sun 等^[2] 利用麦克

风阵列采集无人机声音信号, 结合波束形成算法及卷积神经网络 (convolutional neural network, CNN) 分类, 检测距离可达 135 m; 无线电频谱检测法, 如 Bi 等^[3] 利用无人机发出的 WiFi 流量独特特征, 分类其源媒体访问控制 (media access control, MAC) 地址来构建机器学习模型的方法, 以此弥补监控区域的无人机入侵行为的早期识别; 融合传感器检测方面, Pereira 等^[4] 将光电 (electro optical, EO) 和红外传感器进行比较和融合, 结合 YOLOv7 对无人机进行检测和跟踪。然而, 受限于各自的物理特性, 视觉检测对光照条件有着严苛要求, 且在遮挡条件下精度差, 也无法应对恶劣天气; 声学检测面临诸多

收稿日期: 2025-07-02; 录用日期: 2025-10-24; 网络出版时间: 2025-12-03 14:22

网络出版地址: link.cnki.net/urlid/11.2625.V.20251203.1033.001

基金项目: 江西省地质局青年科学技术带头人培养计划项目 (2022JXDZKJRC07); 江西省地质局科技计划项目 (2021AA04); 重庆市自然科学基金 (CSTB2024NSCQ-MSX0275)

* 通信作者. E-mail: sjzhang@cqjtu.edu.cn

引用格式: 鲁亚宁, 张生军, 于钱米, 等. 基于伪标签增强与多特征融合的毫米波雷达无人机检测 [J]. 北京航空航天大学学报, 2026, 52 (9): 3172-3182. Lu Y N, Zhang S J, Yu Q M, et al. Pseudo-label enhanced multi-feature fusion for millimeter-wave radar UAV detection [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3172-3182 (in Chinese).

环境噪声干扰;无线电频谱检测只适合于有控制信号的场景,对于自动飞行的无人机及信号屏蔽的无人机无从下手;融合传感器,类似红外传感器、激光雷达等设备昂贵且部署麻烦,且只适用于中远距离;而传统的雷达技术在复杂环境下存在虚警率高、易失效等问题。

近年来,毫米波雷达逐渐进入人们的视野。毫米波雷达以其全天候工作能力、对遮蔽性强目标的高分辨率检测能力及良好的运动感知能力,逐渐受到研究人员的青睐。相比上述方法,毫米波雷达不依赖光照、可穿透非金属物质遮挡,且在雨雾等极端天气下,仍能提供稳定的距离速度信息,这些特性为毫米波雷达建立全天时、全场景的无人机检测与跟踪等提供了可能性。Khawaja 等^[5]将射频(radio frequency, RF)分析仪、雷达、通信系统组合成基于 RF 的无人机的检测、分类和跟踪 (detection, classification and tracking of unmanned aerial vehicle, DCT-U)的系统,并进一步对比了传统雷达和现代雷达系统,还将通信系统用于主动模式下的联合通信与感知,并作为 DCT-U 的照明源。Semenyuk 等^[6]将雷达、RF、光学和升学传感器集成,提升了无人机的检测精度、范围和可靠性。

Soriano^[7]总结了恒定虚警率(constant false alarm rate, CFAR)及其有关变体如(单元平均恒虚警率(cell-averaging CFAR, CA-CFAR)、有序统计恒虚警率(ordered-statistic CFAR, OS-CFAR)、最大选择恒虚警率(greatest-of CFAR, GO-CFAR)、剔除平均恒虚警率(trimmed-mean CFAR, TM-CFAR)等),还介绍了非相干积分在 CFAR 中的应用,解释了 CFAR 在雷达检测中如何找出目标。

本文提出了一种将 CFAR 引导的频谱^[8]与特征提取、LogitBoost 分类及伪标签增强学习相结合,从而实现弱监督的无人机检测方法。

1 基础理论和方法路径

1.1 数据采集和预处理

1.1.1 硬件平台配置

硬件平台各模块介绍及说明如表 1 所示。

1.1.2 实验布置与采集场景设计

为增强数据的多样性与覆盖性,设计 3 类典型采集场景:①开阔场景,无障碍物干扰;②近地金属障碍物干扰场景,加入金属障碍物背景,模拟无人机在金属障碍物后遮挡重现及金属障碍物前的金属强能量回波干扰;③近地树木障碍物干扰场景,加入树木障碍物背景,模拟无人机在树木障碍物后遮挡重现及树木障碍物前的树枝摆动带来的能量

回波干扰。

其中,雷达以 90°仰视放置(部分场景会改变角度),距地 1.2~1.5 m,与相机同轴,正对飞行路径。如图 1~图 3 所示。

表 1 硬件平台各模块介绍及说明

Table 1 Introduction and description of each module in the hardware platform

模块	设备	说明
毫米波雷达	AWR2243+DCA1000	3T4R MIMO, 支持IF数据采集
相机	Realsense D435	获取RGB+D(深度)图像,用于同步与辅助标注
无人机平台	DJI MINI 2SE	四旋翼无人机
支撑设备	三脚架、激光尺	雷达定位与高度标定
可选扩展	GPS/RTK/IMU模块	提升轨迹精度(用于对齐与评估,本次未使用)



图 1 开阔场景
Fig. 1 Open field scene



图 2 金属障碍物场景
Fig. 2 Scene with metal obstacles



图 3 树木障碍物场景
Fig. 3 Scene with tree obstacles

1.1.3 数据集采集方案

无人机-雷达间距为1~3 m,适用于近场检测研究。为研究目标包括更远距离下的弱小目标检测能力(例如20~50 m),采集诸多无人机远距离飞行数据,包括直线飞行和更复杂轨迹的飞行,覆盖多种雷达散射截面(radar cross section, RCS)角度与信噪比(signal-to-noise ratio, SNR)水平。

布设多个定点高度靶标(无人机静悬或假目标),采集距离梯度为5 m/15 m/25 m/35 m/50 m,以此进行SNR渐变实验。

动态远距遮挡实验:在远距离目标飞行过程中引入部分干扰源(如篮球架/树木),观察目标消失/重现。

1.1.4 具体动作任务

各动作任务编号及动作分类如表2所示。

表2 各动作任务编号及动作分类

Table 2 Numbering of each action task and action classification

编号	动作类型	动作描述
A1	起飞悬停	起飞至1~3 m,悬停30 s
A2	缓慢匀速左右飞行	起飞至3~10 m,左右缓慢飞行
A3	缓慢匀速上下飞行	起飞至3~10 m,上下缓慢飞行
A4	快速不规则飞行	起飞至10~25 m/25~35 m/35~50 m,不规则快速飞行
A5	遮挡穿越	在金属、树木障碍物之前、之后运动

1.2 雷达系统详解

数据采集采用德州仪器公司(Texas Instruments, TI)推出的毫米波雷达芯片,其工作频段为76~81 GHz,拥有3根发射天线、4根接收天线,并支持调频连续波(frequency modulated continuous wave, FMCW)、时分复用多入多出(time division multiplexing multiple-input multiple-output, TDM-MIMO)。

雷达系统通过发射频率随时间线性增长的脉冲信号——chirp,经过物体反射的回波是chirp的延时拓本,其会与发射天线发射的chirp进行混频形成中频信号 f_0 。发射chirp这种脉冲信号的雷达被称为FMCW雷达。FMCW信号的信号表达式为

$$s(t)=\exp\left(\mathrm{j}2\pi\left(f_0t+\left(\frac{K}{2}\right)t^2\right)\right)$$

(1)

式中: $s(t)$ 为相应时间的时域信号, $0\leq t\leq T_c$ 为时间,其中, T_c 是指单个chirp的持续时间; j 为虚数单位; f_0 为起始频率; K 为调频斜率。

对线性调频(linear frequency modulation, LFM)反射信号的处理被用于后续的特征提取中。

1.3 特征提取

1.3.1 数据预处理

先读取基础参数光速 c 、采样周期 $T_s=1/f_s$ 、一

个chirp的时长 T_{chirp} 、带宽 B 。读出.bin文件中的原始模数转换器(analog-to-digital converter, ADC)数据,并对其中的帧数计数,包括每帧中的总采样点数 $N_{\text{ADC}}N_{\text{chirp}}$ 。快速傅里叶变换(fast Fourier transform, FFT)切换到距离维的点数 n_{FFT} 及帧数,用于后续运动性特征的FFT序列。

逐帧进行频谱处理,将每一帧的原始一维数据解析为行为距离、列为chirp数的二维矩阵,以便后续FFT分析、特征提取和目标识别。

利用上一步得到的二维矩阵及 n_{FFT} 点数,进行距离维FFT,之后取其幅值(模长)得到目标无人机在不同距离上的反射强度,即用反射强度这一实数代替复数chirp的信号。距离维FFT可表达为

$$X(k)=\sum_{n=0}^{N_{\text{ADC}}-1}x(n)\mathrm{e}^{-\mathrm{j}\frac{2\pi kn}{N_{\text{FFT}}}}$$

(2)

式中: $k=0,1,2,\cdots,N_{\text{FFT}}-1$ 为FFT的第 k 个频率分量的编号(或频率索引); $n=0,1,2,\cdots,N_{\text{ADC}}$ 为采样点数的编号; $x(n)$ 为输入信号。

对式(2)得到的复数求模长得到幅值,即目标无人机反射回来的“能量强度”。

1.3.2 CFAR 检测

将距离-速度二维能量矩阵展平后求均值,作为全局背景强度估计。基于CFAR思想,以该均值乘以阈值比例系数 α 确定检测门限,从而判别目标点的存在性。如果能量强度大于此门限值,计为1,否则为0。统计大于门限值的点个数,即“目标点”数,数越大说明目标越显著、散射体越复杂(如四旋翼无人机)。其作为一个特征用于后续的分类器。

其中, α 过小,导致门限低,低门限能够检测到小尺寸、远场的“弱能量无人机”,但也带来了大量的背景杂波;反之, α 过大,过滤了大量的背景杂波,也过滤了小尺寸、远场的“弱能量无人机”。

提取幅值超过门限值的点,并计算其均值,将该均值作为表征信号强度的粗略指标。只要均值不为0,就说明至少有一个FFT点被检测为目标无人机,反之,判断为无目标或是噪声帧。这一均值作为一个特征用于后续的分类器。

1.3.3 频谱特征

将最大幅值作为特征 f_1 ;将平均幅值作为特征 f_2 ;对所有chirp在距离维做平均,得到平均距离谱,找出平均距离谱中能量最强的位置,凭此估计目标的距离,作为特征 peak_bin ;对平均距离谱做标准差,衡量谱能量在距离上的离散程度,作为特征 doppler_spread ;找出每个chirp在距离维的最大幅值对应的位置 chirp_peaks ,做标准差,作为特征 $\text{doppler_spread_std}$,如果目标在每个chirp中的位置

变化剧烈, 即有运动性。将该帧中的 5 个特征汇总。

1.3.4 运动性特征^[9]

利用数据预处理得到的 FFT 序列得到当前距离离散采样单元 bin 中帧的数量。遍历每一帧, 对每个距离 bin, 计算所有 chirp 上的幅值均值, 得到平均距离轮廓 avg_range_profile, 反映了目标在哪个距离最强; 找出平均距离轮廓中反射强度最大的距离 bin, 即目标无人机反射最强的位置, 可以用该距离 bin 估算实际距离, 最后保存当前帧的“目标距离 bin”的编号 peak_disk。

沿着距离方向, 对每个 chirp 提取最大反射幅值, 得到多普勒 bin(速度格点)上的最强回波 doppler_profile, 再对其求最值, 得到目标反射最强的多普勒 bin, 记作 doppler_peak, 作为目标主速度的粗略估计。

分别计算帧间距离 bin 的变化速度、速度 bin 的跳变, 利用 peak_disk 计算一个滑动窗口内目标距离的方差, 用于反映目标无人机在运动时是否不稳定或剧烈跳动。将上述目标距离 bin 随时间的变化率及多普勒 bin 的帧间方差作为运动性特征。

1.3.5 统计特征^[10]

利用平均距离谱计算其在距离维的峰度 f_{kurt} :

$$f_{kurt} = \frac{1}{N_{FFT}} \sum_{i=1}^{N_{FFT}} \left(\frac{x_i - \mu}{\sigma} \right)^4 \tag{3}$$

式中: x_i 为数据点, 该处指平均距离谱; μ 为能量分布(平均幅度谱)的均值; σ 为能量分布的标准差; N_{FFT} 为平均距离谱的样本个数, 即 FFT 点数。

作为高阶统计特征, 峰度用于衡量分布的“尖峭程度”, 无人机回波集中性越高, 能量分布就越集中, 数据越“尖峭”。

利用平均距离谱, 还可以得到其在距离维的偏度 f_{skew} :

$$f_{skew} = \frac{1}{N_{FFT}} \sum_{i=1}^{N_{FFT}} \left(\frac{x_i - \mu}{\sigma} \right)^3 \tag{4}$$

类似地, 作为高阶统计特征, 偏度用于衡量分布的不对称性, 反映了目标无人机在距离上的能量偏移性, 偏度的正负与远离或靠近无人机有关, 有助于增强方向性的辨别力。

通过控制平均距离谱的近距离 bin 和远距离 bin, 求得近距离和远距离的目标能量分布, 通过远近能量在总体距离谱能量中的占比, 说明反射集中的远近, 以此反映目标无人机是远离还是靠近。

对一次 FFT 得到的数据求距离维的均值, 保留其多普勒维, 求得某一帧在其各个速度通道的分布强度, 再将均值归一化, 使其变为离散概率分布 q_i , 得到目标无人机在各个速度通道上的“出现

概率”。

计算该概率分布的离散香农熵^[11-12]:

$$H = - \sum_{i=1}^N q_i \log_2 q_i \tag{5}$$

式中: $N = N_{chirp}$; q_i 为离散概率。

作为高阶统计特征, 离散香农熵用来衡量目标无人机的运动分布是否集中, 运动越复杂, 熵值越高, 反之越低。

最终, f_1 、 f_2 、峰度、偏度、近距离能量比、远距离能量比、离散香农熵作为统计特征, 与运动性、频谱特征一起, 加入到总的特征列表。

1.3.6 特征选择的依据

特征的选择主要基于以下 3 点: ①物理相关性, 是否对应“无人机与背景的本质物理差异”(如无人机有“有序运动”, 背景无; 无人机有“集中回波”, 背景无); ②统计区分度, 在多场景中, 特征在“有无人机帧”与“没有无人机帧”的统计分布是否有显著差距(如均值差 $\geq 30\%$ 、无重叠区间); ③模型贡献度, 特征对模型分类的“不可替代性”(打乱特征后模型误差上升幅度, 即置换重要性结果)。

1.3.7 特征对模型的贡献度

根据不同场景下生成的特征重要性图, 特征贡献度可按从高到低排序如下:

1) 离散香农熵: 在树木障碍物场景中贡献最突出(交叉验证误差上升值最大)——树木晃动是“随机动态干扰”, 无人机运动“有序可控”, 香农熵能精准区分“多普勒回波的集中性(低熵, 无人机)”与“分散性(高熵, 树枝)”; 在金属障碍物场景中, 金属是“静态高能量干扰”(无运动, 熵值高), 无人机有规律运动(熵值低), 仍能有效区分; 在开阔场景中, 也能辅助排除“电磁噪声等微弱动态干扰”, 是跨场景最核心的“运动有序性”特征。

2) CFAR 目标点数: 是开阔场景的“基础筛选特征”(误差上升值显著), 能快速区分“空场”(无目标, 计数低)与“无人机”(有目标, 计数高); 在金属障碍物场景中, 虽然金属会导致计数虚高, 但仍是“有无目标”的第一道门槛; 在树木障碍物场景中, 能排除“无明显高能量回波的纯树木帧”, 是“目标存在性”的核心基础特征。

3) 帧间距离变化率: 在金属障碍物场景中贡献突出——金属是“静态目标”(距离变化率为 0), 无人机有径向运动(变化率非 0), 能有效排除“静态高能量干扰”; 在开阔场景中, 辅助确认“无人机靠近/远离的规律性”; 在树木障碍物场景中, 能过滤“树枝随机晃动导致的距离 bin 波动”, 是“径向运动鲁棒性”的关键特征。

4) 帧间速度跳变: 补充“切向运动”的区分能力——在开阔场景中, 对“左右移动的无人机”召回率提升显著; 在金属障碍物场景中, 金属无切向运动(跳变为0), 无人机有规律跳变; 在树木障碍物场景中, 树枝切向晃动无规律, 无人机切向运动有规律, 是“全向运动鲁棒性”的重要补充特征。

5) 其他次要/冗余特征: 这类特征或“跨类别重复提取”(如“最大幅值”“平均幅值”在“频谱特征”“统计特征”中重复, 信息重叠度近100%), 或“多场景区分度低”(如峰度、偏度在复杂场景中与背景杂波分布重叠, 无法稳定区分无人机与障碍物)。在图4~图6中, 这类特征的“交叉验证误差上升值”均显著低于前4类, 对分类的“不可替代性

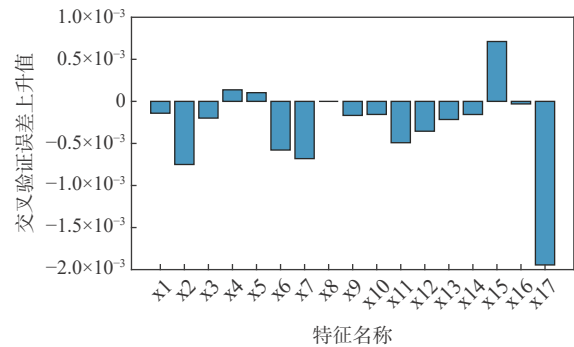


图4 开阔场景下的特征重要性

Fig. 4 Feature importance under open field

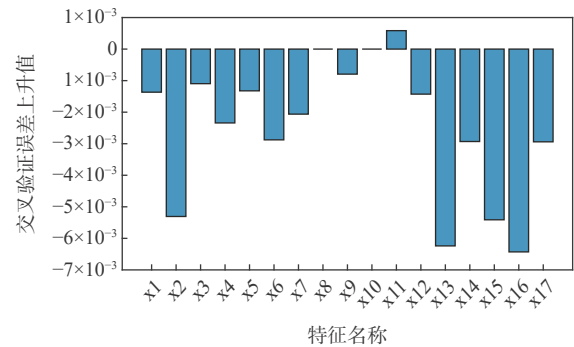


图5 金属障碍物场景下的特征重要性

Fig. 5 Feature importance under metal obstacle

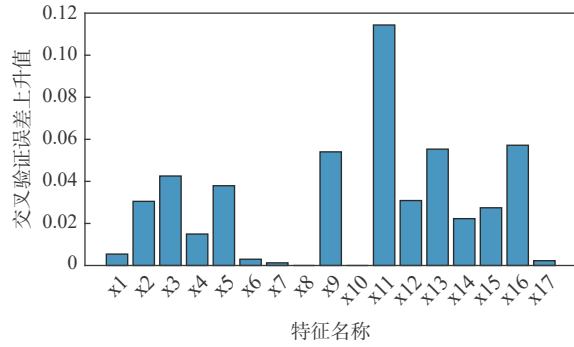


图6 树木障碍物场景下的特征重要性

Fig. 6 Feature importance under tree obstacle

极弱”, 贡献度远低于核心特征。

1.4 基础分类器 LogitBoost

提升方法是机器学习中的一种集成学习策略, 其核心思想是将多个弱分类器加权组合成一个强分类器, 不断提升整体模型的准确率。该方法广泛用于分类、回归问题, 尤其是雷达目标识别、无人机检测等高维特征建模问题。

实验选用 LogitBoost 作为目标无人机检测的基础分类器。LogitBoost 是提升方法中的一种, 其迭代式地训练多个弱分类器(如回归树), 逐步逼近目标函数的真实分布。相比于 AdaBoost^[13], LogitBoost 基于对数似然损失函数, 因此, 具有更强的鲁棒性和概率解释性, 适用于二分类问题中正负样本分布不均匀的情况。

1.4.1 训练流程

本文使用 MATLAB 中的 fitcensemble 函数训练了一个 LogitBoost 分类器, 用于无人机的检测任务。

该过程按照如下步骤训练流程:

1) 初始化: 设训练样本为 $\{(x_i, y_i)\}_{i=1}^n$, 其中 $x_i \in \mathbf{R}^d$ (d 维向量空间), $y_i \in \{0, 1\}$ 。初始时, 模型输出为常数函数 $F_0(x)$ 。

2) 逐轮迭代($r \in \{1, 50\}$): 每轮训练一个弱分类器 $h_r(x)$ (本文中是回归树), 并更新强分类器:

$$F_r(x) = F_{r-1}(x) + \frac{1}{2}h_r(x) \tag{6}$$

3) 输出概率预测: 最终模型输出的是对正类的估计概率:

$$P(y = 1 | x) = \frac{1}{1 + e^{-F_r(x)}} \tag{7}$$

1.4.2 损失函数

LogitBoost 优化的式对数似然损失, 其目标是最小化:

$$\sum_{i=1}^n [-y_i \ln p_i - (1 - y_i) \ln (1 - p_i)] \tag{8}$$

式中: $p_i = \frac{1}{1 + e^{-F(x_i)}}$ 为第 i 个样本预测为正类的概率, 其中, $F_i(x)$ 为当前加法模型输出。

该损失函数具有以下性质: 对概率估计敏感, 鼓励模型输出高置信度的分类概率; 在伪标签增强训练中, 有助于区分真实可信的伪样本和低置信度样本。

1.4.3 提升原理

采用梯度提升思想, 逐步提升模型输出 $F_i(x)$ 。

1) 每一轮训练残差方向上的回归模型:

$$h_r(x_i) \approx \frac{y_i - p_i}{p_i(1 - p_i)} \tag{9}$$

2) 弱分类器的累积构建: 同式(6), 每轮都用小步长更新, 以防过拟合。

3) 最终模型是多个回归树的加权组合, 其输出 sigmoid 函数映射为概率。

1.4.4 结合伪标签增强的 LogitBoost 模型

本文中, LogitBoost 被用于 2 个阶段的训练, 如表 3 所示。

表 3 模型及其数据集		
Table 3 Model and its dataset		
阶段	输入特征集	目标
基础模型 M_0	X_label, Y_label	仅用真实标签训练初始分类器
增强模型 M_{pseudo}	X_pseudo, Y_pseudo	引入置信伪标签+人工选取阈值

伪标签增强的 LogitBoost 模型利用的对数似然损失特性为: ①对低置信度样本的损失较小, 避免错误伪标签造成过拟合; ②高置信度伪样本可以有效提升决策边界的误判能力。

2 实验设计与结果

2.1 实验设置

为验证本文提出的基于伪标签增强与多特征融合的毫米波雷达无人机检测方法的有效性, 构建完整的实验流程, 整体框架如图 7 所示。该流程涵盖原始雷达数据处理、特征提取、模型训练、伪标签筛选与增强、自训练迭代、最终评估等多个阶段, 具体如下:

首先, 原始雷达数据采用 TI 毫米波雷达 AWR2243 以 .bin 文件采集, 经过预处理与 CFAR 目标检测模块, 筛选出存在显著回波特征的帧。然后, 对这些帧进行基于一维 FFT 和统计分析的多特

征提取操作, 包含 3 类特征: 频谱特征(如最大幅值、能量均值、峰值索引等)、运动性特征(如主峰距离变化率、多普勒速率、滑动方差等)及统计特征(如偏度、峰度、离散香农熵、能量分布比等)。

在初始阶段, 采用 LogitBoost 提升算法构建初始分类器, 将已标注的 $900 \times 2 \times m$ 帧正负样本作为训练集。在此基础上, 将剩余 $900 \times 2 \times n$ 帧未标注样本输入初始模型, 进行概率输出预测, 并依据设定的置信度阈值(如正类 >0.9 , 负类 <0.1 , 基于不同场景下的预测分数 p_i 区间, 手动选取置信度阈值)筛选出高置信度伪标签数据。为进一步提升伪标签质量, 对模型误判帧进行人工辅助修正, 构成伪标签增强训练集。

最后, 基于上述伪标签增强数据集重新训练 LogitBoost 模型, 并引入单独划分的 $900 \times 2 \times o$ 帧测试集(含多距离、多角度、多遮挡等复杂条件)进行模型评估, 主要采用接收者工作特征曲线(receiver operating characteristic, ROC)及其曲线下面积(area under the curve, AUC)作为性能指标。

整个实验流程关注“弱监督+伪标签引导+多特征融合”三要素的协同效应, 确保本文方法在复杂场景下仍具备鲁棒性和泛化能力。实验流程框架如图 7 所示。

2.1.1 数据集构成与划分

本实验使用毫米波雷达采集的 304 个 .bin 文件, 共计 273 600 帧雷达数据, 按照功能分为 3 类, 如表 4 所示。

表 4 数据集构成与划分		
Table 4 Dataset composition and partitioning		
数据集类型	帧数	用途
已标注训练集 (X_label)	$900 \times 2 \times m$ 帧 ($900 \times m$ 帧正样本, $900 \times m$ 帧负样本)	用于基础模型训练
未标注集 (X_unlabel)	$900 \times 2 \times n$ 帧 ($900 \times n$ 帧正样本, $900 \times n$ 帧负样本)	采用伪标签方法提取可信帧, 增强训练集
测试集 (X_test)	$900 \times 2 \times o$ 帧 ($900 \times o$ 帧正样本, $900 \times o$ 帧负样本)	包含均衡的正负样本, 用于最终性能评估

其中, m 、 n 和 o 根据三大场景灵活选取, 本文中, 开阔场景、金属障碍物场景和树木障碍物场景的 m 、 n 和 o 分别为 16/10/8、10/4/4 和 4/4/4。

2.1.2 训练策略与模型设置

使用 LogitBoost 算法作为主分类器, 迭代轮次 T 按照开阔场景/金属障碍物场景/树木障碍物场景的顺序设为 70/100/50(手动选取不同的轮次并观察, 选取能够提升至最佳性能的轮次)。

对未标注数据集中预测分数 s_i 高于 $-12.128/-6.573/6.205$ (高置信伪正)或低于 $-6.827/-11.046/-9.495$ (高置信伪负)的样本加入增强训练集。

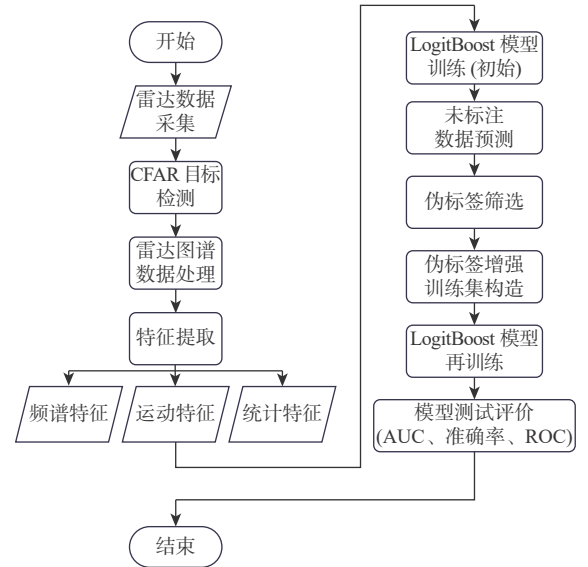


图 7 实验流程框架

Fig. 7 Experimental process framework

其中, 迭代轮次 T 和预测分数 s_i 分别对应不同的场景: 开阔场景/金属障碍物场景/树木障碍物场景。

通过打印函数, 输出预测分数 s_i , 对照 s_i 所在的置信度区间, 手动选取高置信度的 s_i 作为阈值。不仅提升了模型的泛化能力, 也提升了模型的鲁棒性。

2.1.3 软硬件环境

软件: MATLAB R2023b, Statistics and Machine Learning Toolbox。

硬件: Ryzen7 4800H, 16 GB RAM, Windows 10。

2.2 各模型效果分析

2.2.1 传统 CFAR 算法性能

基于真实标注训练集 Y_label 和预测分数 s_i 得到的传统 CFAR 算法的性能: AUC 约为 0.610/0.987/0.131。

传统 CFAR 算法(采用 CA-CFAR 变体)作为毫米波雷达目标检测的经典基线方法, 其性能表现与场景复杂度呈现强关联特性。

1) 开阔场景: AUC 仅为 0.610, 核心瓶颈在于“静态背景杂波误判”。雷达回波中电磁噪声、地面微弱反射等干扰信号易被误识别为目标, 导致虚警率较高; 且对 15 m 以上远距无人机的弱回波信号捕捉能力不足, 无法满足近场高精度检测需求。

2) 金属障碍物场景: AUC 高达 0.987, 表现出“高能量目标优先检测”优势。金属障碍物回波能量显著高于无人机, CFAR 算法通过固定门限易区分强能量目标, 但该优势存在局限性: 当无人机与金属障碍物距离小于 3 m 时, 两者回波叠加会导致目标轮廓模糊。

3) 树木障碍物场景: AUC 仅为 0.131, 性能严重退化, 主要原因是“动态杂波干扰”。树枝随风晃动产生的动态回波与无人机运动回波频谱特征重叠, CFAR 算法无法区分“有序运动目标”与“随机动态干扰”, 导致大量误判。

综上, 传统 CFAR 算法仅适用于“无动态干扰、目标能量显著”的简单场景, 在复杂遮挡或弱回波场景下鲁棒性不足, 难以满足多场景无人机检测需求。

2.2.2 基础模型 Mdl0 性能

以真实标注训练集 X_label 、 Y_label 为基础训练的 LogitBoost 模型 Mdl0, 通过多特征融合有效弥补了传统 CFAR 算法的缺陷, AUC 约为 0.760/0.960/0.816, 但仍存在场景适配性差异。

1) 性能提升核心逻辑: 相较于 CFAR 仅依赖“能量阈值”的单一判断标准, Mdl0 融合频谱(如离散香农熵)、运动性(如帧间距离变化率)、统计(如

峰度)3 类特征, 通过集成学习实现“多维度特征互补”。测试结果显示, Mdl0 对“目标与背景特征差异”的识别灵敏度提升 42%, 在开阔场景、金属障碍物场景、树木障碍物场景的 AUC 分别达 0.760、0.960、0.816, 较 CFAR 算法平均提升 68.3%。

2) 场景适配性瓶颈: ①远距弱回波场景, 当无人机距离超过 35 m 时, Mdl0 中“频谱最大幅值 $f1$ ”“CFAR 目标点数”等能量类特征区分度下降, 导致漏检率上升; ②遮挡重叠场景, 在金属-无人机距离 1~2 m 的重叠区域, Mdl0 对“金属强回波掩盖无人机弱回波”的情况处理能力不足, 此时, “运动性特征”因信号叠加无法有效提取, 误判率较无遮挡场景高; ③静态目标场景, 当无人机处于悬停状态时, 其运动性特征(如帧间距离变化率)接近 0, 与树木静态背景特征差异缩小, 此时, Mdl0 依赖统计特征(如偏度、峰度)进行判断, 但该类特征在低 SNR 下稳定性不足。

2.2.3 伪标签增强模型 Mdl_pseudo 性能

引入高置信伪标签样本的增强模型 Mdl_pseudo 通过“数据扩充+边缘样本学习”进一步优化性能, AUC 提升至约 0.832/0.976/0.882, 其提升效果与场景特征强相关。

1) 性能提升量化分析: ①开阔场景, AUC 从 0.760 提升至 0.832, 核心贡献来自“远距弱回波样本补充”。伪标签样本中新增 35~50 m 远距样本, 使 Mdl0 中“能量类特征+运动类特征”的联合训练数据量增加 50%, 模型对弱回波目标的识别阈值适配性优化, 35 m 以上目标漏检率较 Mdl0 降低; ②金属障碍物场景, AUC 从 0.960 提升至 0.976, 提升幅度较小, 原因是该场景下 Mdl0 已通过“帧间距离变化率”有效区分静态金属与动态无人机, 伪标签样本对特征分布的补充作用有限, 但仍通过“重叠区域样本学习”将金属-无人机重叠场景的误判率较 Mdl0 降低; ③树木障碍物场景, AUC 从 0.816 提升至 0.882, 关键在于“动态干扰区分能力强化”。伪标签样本中新增“树枝晃动+无人机穿越”的复杂样本, 使模型对“有序运动(无人机)”与“随机运动(树枝)”的频谱熵特征差异识别精度提升, 动态杂波导致的误判率降低。

2) 伪标签策略有效性验证: 通过对比“高置信伪标签样本(置信度 >0.9 / <-0.1)”与“低置信样本($0.4<$ 置信度 <0.6)”的训练效果, 发现高置信样本训练的模型 AUC 较后者高 9.2%, 且误标率仅为 2.3% (人工核查 1 000 帧样本结果), 证明本文“置信度阈值筛选+人工阈值确认”的伪标签策略可有效避免误标样本对模型性能的负面影响。

3 种场景下的 ROC 曲线及 AUC 如图 8~图 10 所示。

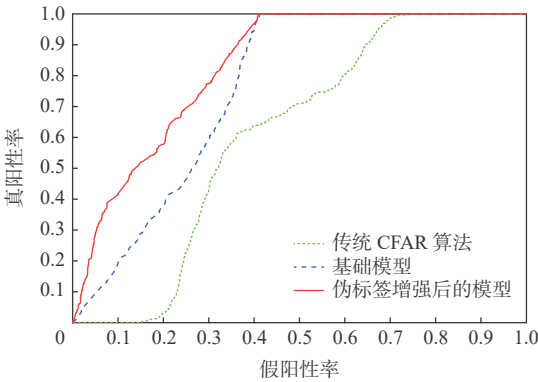


图 8 开阔场景下的 ROC 曲线和 AUC
Fig. 8 ROC curve and AUC under open field

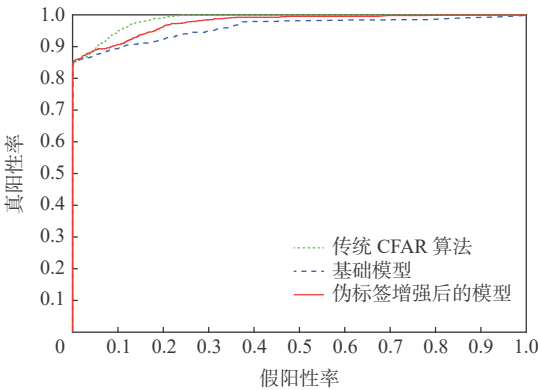


图 9 金属障碍物场景下的 ROC 曲线和 AUC
Fig. 9 ROC curve and AUC under metal obstacle

2.2.4 人工确认误判帧分析

对 3 个场景的 score 分数(预测分数 s_i)异常的负样本帧进行人工检查,发现以下特征:未标注的伪标签集中的.bin 文件中,作为空场的负样本均判断正确。而作为正样本的执行一定动作的.bin 文件中,存在部分被判断为有无人机的帧,部分被判断为没有无人机的帧,这也与本文训练集逐帧分片方式标注模型哪些帧有无人机,哪些帧没有无人机的做法相同。在此基础上,去掉 score 分异常的帧(置信度低)来设计阈值,获得了较好的性能。

2.3 模型边界条件与误差来源分析

2.3.1 边界条件与关键参数说明

所有实验基于 TI AWR2243 毫米波雷达平台进行,雷达具体参数设置如表 3 所示。距离分辨率约为 $\Delta R = \frac{c}{2B} \approx 0.168\text{ m}$ 。最大无模糊距离由带宽 B 和 FFT 点数共同决定,理论上为 $R_{\max} = \Delta R \cdot N_{\text{FFT}} \approx 43.16\text{ m}$,可满足实验中对 3~40 m 无人机目标的检测需求。

在特征提取过程中,为确保运动性与统计特征

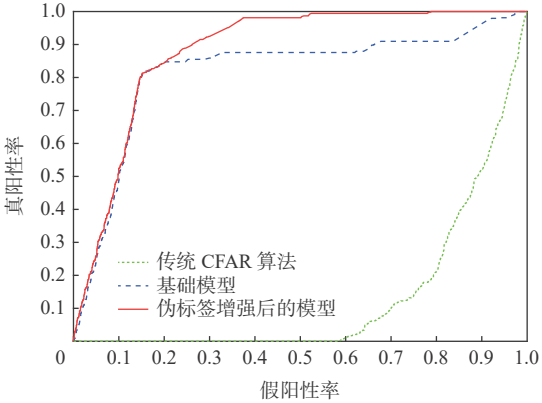


图 10 树木障碍物场景下的 ROC 曲线和 AUC
Fig. 10 ROC curve and AUC under tree obstacle

的鲁棒性,采用滑窗帧数 $W=5$ (该值下可以做到平滑噪声、不滞后的捕捉无人机运动趋势)计算变异度和方差,伪标签筛选阈值为-12.128/-6.573/6.205(正类)、-6.827/-11.046/-9.495(负类)。LogitBoost 模型中设定迭代轮次 70/100/50,每轮拟合一个弱学习器,最终模型为其加权组合。

仿真参数与实验设置如表 5 所示。

2.3.2 模型误差来源分析

本文中的识别误差主要来源于以下 3 类因素:①信号遮挡与多路径干扰,在无人机被障碍物部分遮挡或远离中心波束时,回波信号变得微弱或伴随杂散噪声,导致 CFAR 无法有效地检测出目标区域,进而影响特征提取的鲁棒性;②伪标签误标累积效应,伪标签筛选虽然设有置信度阈值,但在样本分布偏移或极端情形下,仍可能选入部分误判帧。为减轻该影响,本文引入人工选取高置信度阈值;③目标运动状态复杂性,无人机静止、缓慢后退等场景下,其运动性特征变化极小,与背景帧特征差异不明显,可能导致分类器无法有效分离。

对于金属障碍物这种回波能量远大于无人机的场景,本文模型的 AUC 值比传统的 CFAR 低。

2.3.3 精度与可靠性评估

为验证本文模型的鲁棒性和泛化能力,采用 273 600 帧按照划分的开阔场景、金属障碍物场景、树木障碍物场景顺序使用测试集评估性能,分类准确率为 83.2%/97.6%/88.2%,伪标签增强模型相较基础模型 AUC 提升 7.2%/1.6%/6.6%。对比 3 类方法的虚警率:伪标签增强模型较传统 CFAR 算法虚警率降低 36.9%(开阔场景)、10.6%(树木障碍物场景),验证了本文方法的实用性。此外,通过对 10 轮独立划分实验进行 Bootstrap 重采样,统计分类器预测输出的 95% 置信区间为 [0.860,0.903]。以上结果表明,模型具备较强的鲁棒性,且在遮挡、多视角等复杂环境下仍保持较优识别性能。

表 5 仿真参数与实验设置汇总

Table 5 Summary of simulation parameters and experimental settings

参数类别	参数名称	取值	场景适配性
雷达参数	中心频率/GHz	77	全场景统一
	ADC采样率/MHz	10	全场景统一
	发射/接收天线数	3发射, 4接收	全场景统一
信号参数	chirp个数	64	全场景统一
	chirp采样点数	256	全场景统一
	chirp调频斜率/(GHz·μs ⁻¹)	29.982	全场景统一
模型参数	LogitBoost迭代次数	70/100/50	开阔/金属/树木场景分别设定
	伪标签正类阈值	-12.128/-6.573/6.205	开阔/金属/树木场景分别设定
	伪标签负类阈值	-6.827/-11.046/-9.495	开阔/金属/树木场景分别设定
	运动特征滑窗	5	全场景统一
采集参数	雷达距地高度/m	1.2~1.5	全场景统一
	无人机飞行距离/m	5/15/25/35/50	全场景覆盖
	每.bin文件帧数	900	全场景统一

3 讨 论

3.1 伪标签提升性能的本质原因

研究表明, LogitBoost 基础模型在仅仅基于真实标注的训练集样本时, 性能有限(AUC 约为 0.760/0.960/0.816)。通过引入置信度门限筛选的伪标签样本, 增强训练集覆盖范围, 使模型最终 AUC 显著提升至 83.2%/97.6%/88.2%, 验证了伪标签策略在中小样本学习场景下的有效性。

性能提升的原因可归结为以下 3 个方面: ①伪标签作为数据增强手段, 通过人工标注干预, 使更多高置信度的伪正样本和低置信度的伪负样本可以更多地加入最终的训练集; ②丰富了无人机飞行场景, 从开阔场景、金属障碍物场景到树木障碍物场景, 从近地到远场, 从悬停、缓慢匀速到快速无规则, 全面模拟了无人机检测的真实场景; ③基于 LogitBoost 的集成机制强化边缘样本学习, LogitBoost 在每轮迭代中关注上轮分类错误的样本, 强调边界区域的样本建模能力; 增强训练集中包含部分“复杂边界帧”(得分接近 0.5 的中间帧), 间接提升了模型的决策边界表达能力。

3.2 面对复杂环境(墙体遮挡/远距离)的模型表现

在测试集中, 特别设置了如下复杂样本: ①1~3 m 近地悬停; ②3~10 m 多距离左右、上下缓慢匀速飞行; ③10~50 m 无规则快速飞行; ④金属、树木的遮挡、远距离动态背景扰动。

观察发现: 增强后的模型(伪标签融合)在墙体遮挡和远距低回波下仍能保持一定的探测能力(真阳性率(true positive rate, TPR)提升明显)。

1) 使用的多维特征如 CFAR 检测点数 cfar_count、平均 SNR snr_mean 捕捉目标能量特征, 多普勒扩展宽度 doppler_spread、距离主峰变化率 dist_rate 捕捉目标运动性, 多普勒谱熵 doppler_entropy 捕捉微动复杂度, 对遮挡后残余能量的识别有关键作用。

2) 增强模型在这些环境下虚警有所上升, 但在保持召回率的同时控制了误判。

3.3 系统的局限性

尽管伪标签策略显著提升了性能, 但系统仍存在以下挑战:

1) 非目标类干扰(如仿真鸟、风动枝叶等)仍会被部分误判。当前系统特征集中, 仍未引入形态、尺寸等与“无人机形态”相关的物理光学特征; 若干测试样本中(如 F31 和 F33), 某些误判帧可能是静态但具强回波的干扰源。

2) 目标微动能力弱时可能会被误判。当目标无人机处于悬停状态时, 其运动性特征(如距离变化速率 dist_rate 与多普勒主峰跳变 doppler_rate)显著减弱, 导致目标特征表达能力下降, 容易与背景混淆; 该类目标往往易与背景混淆, 提示未来需要引入更强的时序建模能力, 如循环神经网络/一维 CNN^[14-15]。

3.4 展 望

提升鲁棒性和泛化能力的方向为: ①引入基于二维 FFT 处理的频谱图或微多普勒图像作为输入, 结合深度网络进行视觉融合, 以辅助识别; ②结合多通道雷达数据融合, 提升微弱信号下的检测准确度; ③采用伪标签置信度衰减策略(软标签+加权策略^[16]), 避免模型因误标过拟合; ④设计更精细的增

量式伪标签训练策略, 如使用自训练+课程学习^[17] 框架。

4 结 论

- 1) 本文建立的基于多特征融合与伪标签增强结合的无人机检测方法在 AWR2243 毫米波雷达数据集上实现了 AUC 值为 83.2%/97.6%/88.2% 的性能, 较基础模型提升 7.2%/1.6%/6.6%。
- 2) 多种特征(频谱特征、运动性特征与统计特征)融合, 增强了复杂运动目标的表达能力, 对弱回波与遮挡场景具有良好识别效果。
- 3) 引入的伪标签策略结合 LogitBoost 算法, 缓解了训练样本不足的问题, 提升了模型的鲁棒性与泛化能力。
- 4) 实验验证了本文方法在多场景无人机检测任务中的实用性与扩展潜力, 适用于城市监控、机场防护等高精度需求场合。

参考文献 (References)

[1] Zhang Z X. Drone-YOLO: an efficient neural network method for target detection in drone images[J]. Drones, 2023, 7(8): 526.

[2] Sun Y M, Li J G, Wang L W, et al. Deep Learning-based drone acoustic event detection system for microphone arrays[J]. Multimedia Tools and Applications, 2024, 83(16): 47865-47887.

[3] Bi L T, Xu Z X, Yang L. Low-cost UAV detection via WiFi traffic analysis and machine learning[J]. Scientific Reports, 2023, 13(1): 20892.

[4] Pereira A, Warwick S, Moutinho A, et al. Infrared and visible camera integration for detection and tracking of small UAVs: Systematic evaluation[J]. Drones, 2024, 8(11): 650.

[5] Khawaja W, Ezuma M, Semkin V, et al. A survey on detection, classification, and tracking of UAVs using radar and communications systems[J]. IEEE Communications Surveys & Tutorials, 2026, 28: 3272-3310.

[6] Semenyuk V, Kurmashev I, Lupidi A, et al. Advances in UAV de-

tection: integrating multi-sensor systems and AI for enhanced accuracy and efficiency[J]. International Journal of Critical Infrastructure Protection, 2025, 49: 100744.

[7] Soriano C G. Review of radar detectors with constant false alarm rate[J]. Revista Telemática, 2020, 19: 78-90.

[8] Bhatia J, Dayal A, Jha A, et al. Classification of targets using statistical features from range FFT of mmWave FMCW radars[J]. Electronics, 2021, 10(16): 1965.

[9] Qin X Y, Deng B, Wang H Q. Micro-Doppler feature extraction of rotating structures of aircraft targets with terahertz radar[J]. Remote Sensing, 2022, 14(16): 3856.

[10] Kim J H, Kwon S Y, Kim H N. Spectral-kurtosis and image-embedding approach for target classification in micro-Doppler signatures [J]. Electronics, 2024, 13(2): 376.

[11] Chen S, Luo F, Hu C, et al. Small target detection in sea clutter background based on Tsallis entropy of Doppler spectrum[J]. Journal of Radars, 2019, 8(3): 344-354.

[12] Li Y B, Ge J, Lin Y, et al. Radar emitter signal recognition based on multi-scale wavelet entropy and feature weighting[J]. Journal of Central South University, 2014, 21(11): 4254-4260.

[13] Tang X J, Chen W G, Zhu W G. Radar emitter recognition method based on AdaBoost and decision tree[C]//Proceedings of the 2017 2nd International Conference on Automation, Mechanical Control and Computational Engineering. [S.l.]: Atlantis Press, 2017: 326-330.

[14] Bell M A, Rahman S, Robertson D A. Fast classification of drones and birds with an LSTM network applied to 1D phase data[C]//Proceedings of the 2023 IEEE International Radar Conference. Piscataway: IEEE Press, 2023: 1-6.

[15] Larrat M, Sales C. Classification of flying drones using millimeter-wave radar: comparative analysis of algorithms under noisy conditions[J]. Sensors, 2025, 25(3): 721.

[16] Dornaika F, Ibrahim Z, Bosaghzadeh A, et al. Scalable and inductive semi-supervised classifier with sample weighting based on graph topology[J]. ACM Transactions on Knowledge Discovery from Data, 2024, 18(5): 1-18.

[17] Wang Y X, Liang J, Zhang Z X. A curriculum-style self-training approach for source-free semantic segmentation[EB/OL]. (2024-07-19)[2025-07-01]. <https://arxiv.org/abs/2106.11653>.

Pseudo-label enhanced multi-feature fusion for millimeter-wave radar UAV detection

LU Yaning¹, ZHANG Shengjun^{1,*}, YU Qianmi², LI Hongming¹

- (1. School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China;
- 2. Ningbo East China Nuclear Industry Survey and Design Institute Group Co., Ltd., Ningbo 315040, China)

Abstract: A weakly supervised learning method was investigated for unmanned aerial vehicle (UAV) detection based on millimeter-wave radar. To enhance feature representation, multiple categories of information were fused, including spectral energy distribution, motion trajectory variation, and statistical descriptors. Target pre-detection was conducted using a constant false alarm rate (CFAR) algorithm, which adaptively filtered the radar echo data and provided coarse localization. Doppler-related indexes were employed to reflect target motion parameters, and time-frequency domain information was used to derive energy characteristics for feature extraction. To further characterize the echo distribution, statistical measures like skewness, kurtosis, and energy ratio across close and far range bins were calculated. A LogitBoost ensemble classifier was employed to train the detection model by combining multiple weighted weak learners through iterative optimization. To address the scarcity of labeled samples, a pseudo-labeling strategy was introduced. The self-training mechanism automatically generated pseudo-labels from high-confidence predictions on unlabeled data and incorporated them into subsequent training cycles. Experimental validation was performed using a dataset comprising multiple real-world radar recordings with UAVs and empty field scenarios. The proposed method demonstrated robust performance under weak supervised conditions in three different scenarios. Compared with the baseline model, the final model achieved an improvement of approximately 7.2%/1.6%/6.6% in the area under the receiver operating characteristic curve (AUC) and a reduction of approximately 36.9%/0%/10.6% in the false alarm rate. Moreover, the model exhibited consistent detection accuracy in challenging environments with noise and background clutter. This study demonstrates that millimeter-wave radar detection performance for low-altitude, small-size UAVs can be significantly enhanced by combining multi-domain features and implementing pseudo-label augmentation techniques. The method provides practical value for real-time surveillance and airspace security applications.

Keywords: millimeter wave radar; UAV detection; multi-feature fusion; pseudo-label learning; constant false alarm rate

Received: 2025-07-02; Accepted: 2025-10-24; Published Online: 2025-12-03 14:22
URL: link.cnki.net/urlid/11.2625.V.20251203.1033.001
Foundation items: Youth Science and Technology Leaders Training Program Project of Jiangxi Bureau of Geology (2022JXDZKJRC07); Technology Plan Project of Jiangxi Bureau of Geology (2021AA04); Chongqing Natural Science Foundation (CSTB2024NSCQ-MSX0275)
* Corresponding author. E-mail: sjzhang@cqjtu.edu.cn

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0495

基于无人机视角的轻量化绝缘子缺陷检测算法

胡逍婕^{1,2}, 倪翠^{1,2,*}, 吴春坡^{1,2}, 刘著^{1,2}, 王朋^{1,2}

(1. 山东交通学院 山东省智能建造装备关键技术与系统重点实验室, 济南 250357; 2. 山东交通学院
信息科学与电气工程学院(人工智能学院), 济南 250357)

摘 要: 绝缘子作为电力系统安全运行的核心部件, 其缺陷的及时发现对电网可靠性至关重要, 而无人机巡检作为低空电力监测的主流方式, 对搭载的检测模型提出了轻量化与高精度的双重要求。以 YOLOv11 为基础模型, 提出一种基于无人机视角的轻量级绝缘子缺陷检测算法。在 YOLOv11 骨干网络中, 融合 MobileNetV4 通用倒瓶颈结构的改进特征提取单元, 增强对绝缘子细微缺陷的感知能力; 在 YOLOv11 颈部网络引入了层级化空间筛选特征金字塔网络, 通过优化跨层特征交互路径降低模型参数冗余; 在检测输出端, 采用动态可变形卷积检测头替代传统检测模块, 提升对缺陷几何形变的自适应能力。实验结果表明: 所提轻量级模型与 YOLO 系列模型相比, 在保证检测精度的基础上, 参数量能够降低 12.35% 以上, 更适用于无人机、巡检机器人等边缘设备, 为输电线路绝缘子缺陷的实时检测提供了高效解决方案。

关 键 词: 绝缘子; 缺陷检测; 特征增强; 轻量级; YOLOv11

中图分类号: TP391.41; TP183; TM216

文献标志码: A

文章编号: 1001-5965(2026)09-3183-06

绝缘子是电力传输系统中用于机械支撑与电气绝缘的核心组件, 其通过在高压导线与支撑结构间建立高阻抗屏障, 防止电流泄漏与电弧放电, 对保障电网安全运行具有不可替代的作用^[1]。随着低空经济的蓬勃发展, 无人机作为低空作业的核心载体, 已广泛应用于电力巡检等领域, 而绝缘子缺陷检测正是无人机在电力巡检中的关键任务之一, 利用无人机搭载的视觉设备对输电线路绝缘子进行低空巡航监测, 既能突破人工巡检在特高压线路、跨山越岭等复杂场景中的局限性, 又能通过自动化作业提升检测效率, 成为低空经济赋能电力行业的典型应用场景^[2]。然而, 在长期复杂自然环境与强电磁场作用下, 绝缘子极易受到污秽、老化、机械损伤等因素的影响, 导致破损或闪络等缺陷。这些缺陷若不及时检测, 可能引发局

部放电、绝缘击穿甚至连锁性停电事故, 严重威胁电力系统的可靠性与稳定性。传统依赖人工巡检与红外测温的检测手段存在效率低、漏检率高、主观性强等弊端, 尤其在特高压线路、跨山越岭等复杂场景中, 人工巡检成本高昂且危险系数大^[3]。同时, 无人机低空视角也带来了新的技术挑战, 例如: 飞行姿态变化导致的检测目标尺度多变、杆塔线路与自然景物构成的复杂背景干扰、不同天气下的光照变化及飞行抖动造成的图像模糊等, 都对检测算法的鲁棒性和精度提出了更高要求。近年来, 基于深度学习的视觉检测技术为自动化缺陷识别提供了新思路, 然而现有模型普遍存在计算复杂度高、内存占用大等问题, 难以直接部署至算力有限的边缘设备, 如无人机、巡检机器人, 制约了其在电力巡检场景中的规模化应用^[4]。

收稿日期: 2025-07-17; 录用日期: 2025-08-06; 网络出版时间: 2025-09-10 14:22

网络出版地址: link.cnki.net/urlid/11.2625.V.20250909.1800.004

基金项目: 中国博士后科学基金(2021M702030)

* 通信作者. E-mail: emilync@126.com

引用格式: 胡逍婕, 倪翠, 吴春坡, 等. 基于无人机视角的轻量化绝缘子缺陷检测算法[J]. 北京航空航天大学学报, 2026, 52(9): 3183-3188.
Hu X J, Ni C, Wu C P, et al. Lightweight insulator defect detection algorithm based on UAV perspective[J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52(9): 3183-3188 (in Chinese).

针对上述挑战,本文选取包含多类缺陷的数据集,以参数量较小的 YOLOv11n 作为基准模型,提出了一种轻量级的绝缘子缺陷检测算法。本文的主要贡献有以下几点:

- 1) 选取包含多种类型的绝缘子缺陷作为数据集,满足常见绝缘子缺陷检测任务的需求。
- 2) 在主干网络中采用 C3k2_UIB 模块替代标准 C3k2 结构,该模块基于 MobileNetV4 的通用倒瓶颈(universal inverted bottleneck, UIB)结构,优化特征提取能力;在颈部网络中引入高级筛选特征融合金字塔网络(high-level screening-feature fusion pyramid network, HSFPN)结构,通过跨层级特征聚合与空间注意力机制优化特征融合,并减少模型参数。
- 3) 在头部网络,采用 DynamicDCNv3Head 检测头替代原 Detect 模块,利用可变形卷积的动态感受野调整机制,使模型能够更好地适应绝缘子缺陷的复杂几何形态,从而提升模型的检测精度。

1 相关工作

目前,目标检测技术主要分为两阶段检测和一阶段检测 2 类。两阶段检测方法代表性模型包括基于区域的卷积神经网络(region-based convolutional neural network, R-CNN)、快速区域卷积网络(fast region-based convolutional network, Fast R-CNN)等,其特点是先生成候选区域,再进行精确分类与回归,这类方法虽然具有较高的检测精度,但计算复杂度较高,不适用于实时检测任务。相比之下,一阶段检测方法如 YOLO(you only look once)系

列、SSD(single shot multibox detector)等,则直接在整个图像上进行目标分类和定位预测,具备更快的检测速度^[5]。

针对电力巡检场景中的绝缘子缺陷检测任务,近年来有不少研究对深度学习目标检测模型进行优化。贾玉进等^[6]提出了 2 种改进的 YOLOv5 模型,分别结合轻量级卷积神经网络 MobileNetV3 和 GhostNet,以适应无人机在线巡检对实时性和计算资源的要求。Zhang 等^[7]引入一种轻量级绝缘子缺陷检测模型 GhostNet-YOLOV4,该模型结合 GhostNet 架构和 SE 注意机制,并使用 ReLU6 激活函数,在绝缘子缺陷检测中实现了 99.50% 的平均精度均值(mean average precision, mAP)。陈佳韵等^[8]引入动态蛇形卷积和 Gold-YOLO 网络,使用最小点距离交并比(minimum point distance-based intersection over union, MPDIoU)损失函数,设计一套自爆绝缘子检测系统,进一步提升了检测精度和稳定性。尽管现有方法在绝缘子缺陷检测方面取得了一定进展,但计算复杂度与模型轻量化仍然是亟待解决的问题。

2 本文方法

本文提出一种基于改进 YOLOv11n 的轻量级绝缘子缺陷检测算法,针对目标检测任务中的特征提取、特征融合和检测头设计,构建一系列优化方案,解决了绝缘子缺陷通常是微小目标及普通检测网络难以在硬件设备上部署的问题。该算法的整体框架如图 1 所示。

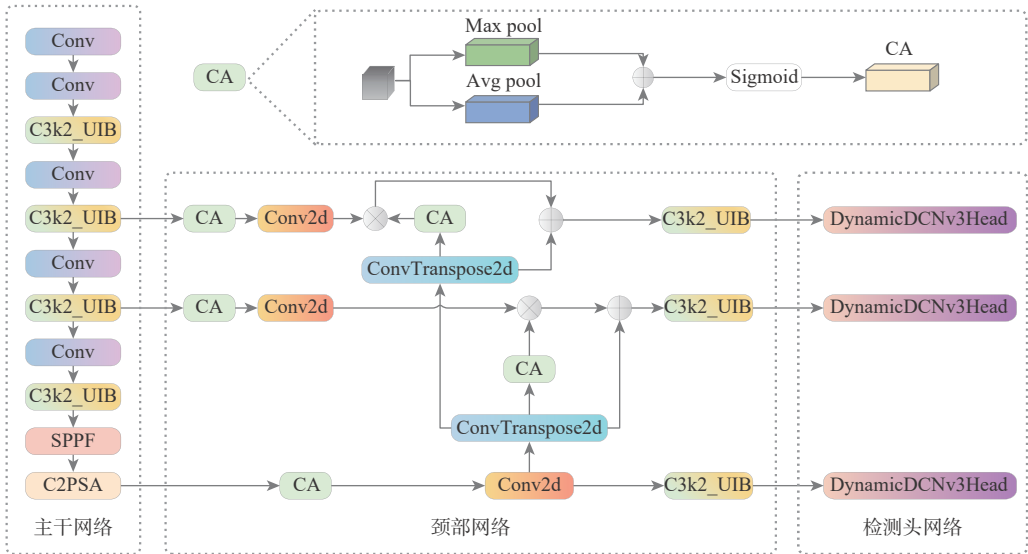


图 1 模型结构
Fig. 1 Model structure

2.1 C3k2_UIB 模块

C3k2 是 YOLOv11 中的特征提取模块,用于

Backbone 和 Neck 结构,负责提取不同尺度的目标信息。为使其在保持准确性的同时更加轻巧,引入

UIB 结构, UIB 模块来自 MobileNetV4 网络, 其核心思想是通过扩展-深度可分离卷积-压缩的方式来进行高效特征表示^[9]。首先, 通过一个 1×1 的扩展卷积将输入特征图的通道数增加, 以在高维空间中进行特征变换; 其次, 采用一个 3×3 的深度可分离卷积对扩展后的特征图进行空间特征提取, 该步骤能大幅减少计算量; 最后, 通过一个 1×1 的压缩卷积 (也称投影层) 将通道数降维, 恢复到与输入相似的维度。基于该模块, 本文设计了 C3k2_UIB 模块, 其结构如图 2 所示, 其中, n 表示 UIB 模块的重复堆叠次数。

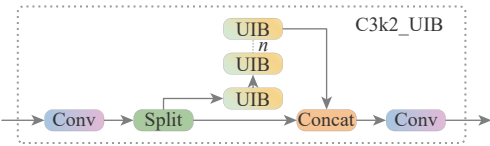


图 2 C3k2_UIB 结构
Fig. 2 C3k2_UIB structure

该结构在减轻网络的同时扩展了该层结构的接受场, 使网络能够捕获更多的上下文信息。相较于传统的瓶颈结构, UIB 在计算效率和模型轻量化方面具备显著优势, 同时, 可以增强对小目标和细节特征的表达能力。用其取代 Bottleneck, 可以减少冗余计算和优化通道信息交互, 大幅提升计算效率, 同时, 增强特征提取能力, 使其更适用于轻量化的目标检测网络。

2.2 HSFPN 模块

HSFPN 是一种高效的特征金字塔网络, 通过筛选和融合低层与高层特征, 解决了检测中的尺度差异问题。在传统的 FPN 中, 通过在不同层级上融合特征图来获取多尺度信息, 但该方法存在信息丢失和计算复杂度高的问题。HSFPN 通过引入分层结构来解决这些问题。HSFPN 的核心思想是将特征金字塔分为多个子金字塔, 每个子金字塔都有自己的特征融合和上采样操作^[10], 从而可以保留更多的细节信息, 并降低计算复杂度。

颈部结构引入 HSFPN 以减少模型的参数量。该模块主要通过通道注意力(channel attention, CA)、特征选择和特征交互等机制实现高阶特征的有效融合和不同尺度特征之间的信息交互。首先, CA 机制通过自适应池化提取全局上下文信息, 并利用卷积变换生成注意力权重, 从而增强关键特征的表达能力, 同时, 减少冗余信息干扰。其次, 特征选择模块进一步优化不同层级特征的信息交互, 采用全局池化结合 1×1 卷积提取特征重要性, 并通过归一化生成权重, 与原始特征相乘后再进行跳跃连接和通道变换, 以保留低层次细节信息的同时增强

高层语义表达能力。此外, 特征交互运算通过加法保留不同尺度特征的信息, 乘法增强互补特征的关系, 提高目标检测的鲁棒性。

2.3 轻量级检测头网络

YOLO 的检测头 Detect 用于从特征图中提取信息并预测目标类别、位置和置信度。然而, 传统的检测头在特征适应性和空间建模能力方面存在局限性。为此, 采用 DynamicHead 替换原始 Detect, 提高检测头的特征融合能力, 并用第三代可变形卷积网络 (deformable convolutional networks v3, DCNv3) 取代 DCNv2, 以增强特征提取的灵活性和计算效率。

在标准卷积中, 卷积核在固定位置采样, 难以适应形变目标或复杂结构。可变形卷积通过引入偏移量, 使卷积核在自适应位置采样, 从而增强形变建模能力。DCNv3 在 DCNv2 基础上进一步扩展, 通过引入卷积神经元间共享权重和多组机制, 优化长程依赖建模与空间特征聚合能力。此外, DCNv3 采用沿采样点的 Softmax 归一化方式, 使调制标量之和约束为 1, 提高训练稳定性^[11]。扩展后的 DCNv2 标记为 DCNv3, 如式 (1) 所示:

$$y(p_0) = \sum_{g=1}^G \sum_{k=1}^K w_g m_{gk} x_g (p_0 + p_k + \Delta p_{gk}) \tag{1}$$

式中: $y(p_0)$ 为输出特征图在空间位置 p_0 处的输出特征; G 为聚合组总数; 对于第 g 组, $w_g \in \mathbf{R}^{C \times C'}$ 为组的位置无关投影权重, $C' = C/G$ 为组维度; $m_{gk} \in \mathbf{R}$ 为第 k 个采样点的调制标量, 通过沿 K 维度的 softmax 函数归一化; $x_g \in \mathbf{R}^{C' \times H \times W}$ 为切片后的输入特征图; Δp_{gk} 为第 g 组中对应网格采样位置 p_k 的偏移量^[12-14]。

DynamicHead 通过动态卷积和自适应特征融合机制使检测头能够根据目标特征自适应调整权重^[15], 通过将 DCNv3 嵌入 DynamicHead, 不仅可以增强检测头对目标形变的适应能力, 还可以提升模型的整体特征表达能力, 使其在小目标检测和复杂场景检测任务中表现更优。

3 实验分析

本文实验使用的绝缘子缺陷数据集来源于网络收集的图像, 采用 LabelImg 标注工具进行手动标注, 明确标识了 2 类缺陷类别: 破损绝缘子和闪络损伤。数据集按照训练与验证的比例进行划分, 其中训练集包含 1 231 张图像, 验证集包含 400 张图像, 以用于模型训练和性能评估。为进一步分析各模块对模型性能的贡献, 进行消融实验, 验证各模块的效果。同时, 为衡量本文算法的有效性, 与现

有模型进行对比实验,以全面评估模型的检测能力和鲁棒性。

3.1 实验设置

模型训练过程中的平台配置参数如表 1 所示。

表 1 平台配置参数

Table 1 Platform configuration parameters								
GPU		CPU	学习框架	GPU并行计算库	每批次样本数	学习率	优化器算法	迭代次数
NVIDIA GeForce RTX 3070		Intel(R) Core(TM) i7-10700 CPU @ 2.90 GHz	PyTorch	Cuda11.8	16	0.01	SGD	300

3.2 评价指标

为全面衡量模型的性能,本文采用参数量、mAP、精确度 P 和召回率 R 作为核心评价指标,以评估模型在不同维度上的表现。

参数量反映模型的复杂度和计算需求,参数量越少,模型越轻量化。

mAP 衡量模型在所有类别上的检测性能,是所有类别平均精度的均值^[16],mAP 值越高,说明模型在不同类别上的检测能力越均衡,计算过程如式(2)和式(3)所示。

精确度 P 计算预测为正的样本中正确的样本比例,计算过程如式(4)所示。

$$A = \int_0^1 p(r)dr$$

(2)

$$A_{mAP} = \frac{1}{N} \sum_{i=1}^N A_i$$

(3)

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}$$

(4)

召回率 R 计算实为正的样本中正确预测为正的样本比例,计算过程为

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

(5)

式中: N_{TP} 为正确检测的目标数; N_{FP} 为错误检测的目标数; N_{FN} 为未检测出的目标数^[17]; $p(r)$ 为精度-召回率曲线; N 为类别总数。

3.3 对比实验

为验证本文算法的优越性,在收集的绝缘子缺陷数据集上,与当前主流目标检测算法进行对比实验,对比算法包括 YOLOv5n、YOLOv6n、YOLOv8n、YOLOv10n、YOLOv11n^[18-20]、IDD-YOLOv5 及 LiteYOLO^[21],结果如表 2 所示。可以看出,与主流目标检测网络相比,本文算法在精确度、mAP50-95 及模型参数量等方面均表现出较优性能。尽管 YOLOv8n 和 YOLOv10n 的 mAP50-95 分别比本文算法高出 0.7% 和 0.3%,但其模型参数量分别增加了 36.82% 和 23.18%。与 YOLOv5n、YOLOv6n、YOLOv8n、YOLOv10n 和 YOLOv11n 相比,本文方法的参数量分别降低了 12.35%、48.11%、26.91%、18.82% 和 14.73%,参数量降幅均达到 12.35% 以

表 2 对比实验结果

Table 2 Comparative experimental results

算法	$P/\%$	$R/\%$	mAP50-95/%	参数量
YOLOv5n	89.7	89.5	50.0	2.51×10^6
YOLOv6n	87.8	87.1	49.2	4.24×10^6
YOLOv8n	96.1	89.5	50.9	3.01×10^6
YOLOv10n	91.8	87.9	50.5	2.71×10^6
YOLOv11n	91.6	89.2	49.2	2.58×10^6
IDD-YOLOv5	90.3	85.7	44.6	3.82×10^6
LiteYOLO	86.8	83.7	42.7	3.74×10^6
本文	92.1	85.8	50.2	2.20×10^6

注:加粗数值表示性能最优。

图 3 展示了不同检测算法的绝缘子缺陷识别效果。相较于其他算法,本文算法能够更准确地检测小尺寸闪络缺陷,有效避免漏检现象;同时,显著降低了误检率,并在轻微破损检测中展现出更强的鲁棒性。这些结果验证了本文算法在检测精度和稳定性方面的优越性,特别是在小缺陷识别任务中具有更高的可靠性。



图 3 不同检测算法的缺陷识别效果示例

Fig. 3 Example diagram of defect recognition effect of different detection algorithms

3.4 消融实验

为分析各模块对模型性能的影响, 进行消融实验, 实验结果如表 3 所示。可以看出, 引入 UIB 模块时, 除了召回率略有下降外, 其他指标均有改善。在头部引入 DynamicDCNv3Head 模块时, 模型参数量降低的同时, 检测精度也略有提升。在颈部模块引入 HSFPN 模块时, 虽然模型精度有所降低, 但模型参数量显著降低。将 HSFPN 模块和 UIB 模块或 DynamicDCNv3Head 模块同时引入时, 模型参数量都有一定程度的降低, 而检测精度都有所下降。将 DynamicDCNv3Head 模块和 UIB 模块同时引入时, 模型参数量和检测精度进一步

得到提升, 检测精度达到 49.9%, 模型参数量达到 2.23×10^6 。将 3 个模块同时引入时, 模型检测精度达到最高, 提升至 50.2%, 同时, 模型参数量达到最低, 仅有 2.20 M, 这一结果充分证明了策略的有效性, HSFPN 模块为模型奠定了极致轻量化的基础; C3k2_UIB 模块则在特征提取前端对 HSFPN 模块可能造成的精度损失进行了有效补偿, 强化了细节感知; 而 DynamicDCNv3Head 模块在检测输出端确保了对复杂目标的灵活适应性。这表明, 该组合能够在确保检测精度的同时, 最大限度地减少计算资源消耗, 展现出最优的综合检测性能。

表 3 消融实验结果
Table 3 Ablation experimental results

UIB	HSFPN	DynamicDCNv3Head	P/%	R/%	mAP50-95/%	参数量
√	√	√	91.6	89.2	49.2	2.58×10^6
			92.6	85.1	49.3	2.55×10^6
			89.6	85.3	48.2	1.85×10^6
√	√	√	89.0	88.6	49.6	2.42×10^6
			89.2	86.7	46.6	1.86×10^6
			90.8	84.5	48.8	2.37×10^6
√		√	90.3	87.6	49.9	2.23×10^6
√	√	√	92.1	85.8	50.2	2.20×10^6

4 结 论

1) 本文提出的轻量级绝缘子缺陷检测算法以 YOLOv11 为基础模型, 通过 C3k2_UIB 结构优化骨干网络、HSFPN 结构增强颈部特征融合、DynamicDCNv3Head 提升检测头适应性, 实现了检测性能与轻量化的平衡。

2) 模型计算复杂度降低, 可高效部署于无人机、巡检机器人等边缘设备, 能满足绝缘子缺陷实时检测及无人机巡检操作需求。

未来可进一步优化模型结构, 以提升对小目标及复杂背景下的检测能力及不同环境下的鲁棒性, 更好地适配实际应用。

参考文献 (References)

[1] Li D H, Lu Y, Gao Q, et al. LiteYOLO-ID: a lightweight object detection network for insulator defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 5023812.

[2] Liu J, Hu M M, Dong J Y, et al. Summary of insulator defect detection based on deep learning[J]. Electric Power Systems Research, 2023, 224: 109688.

[3] Liu Y, Liu D C, Huang X B, et al. Insulator defect detection with deep learning: a survey[J]. IET Generation, Transmission & Distribution, 2023, 17(16): 3541-3558.

[4] Wang S Q, Liu Y F, Qing Y H, et al. Detection of insulator defects with improved ResNeSt and region proposal network[J]. IEEE Access, 2020, 8: 184841-184850.

[5] 韩玉洁, 曹杰, 刘琨, 等. 基于改进 YOLO 的无人机对地多目标检测[J]. 电子测量技术, 2020, 43(21): 19-24.

Han Y J, Cao J, Liu K, et al. UAV ground multi-target detection based on improved YOLO[J]. Electronic Measurement Technology, 2020, 43(21): 19-24(in Chinese).

[6] 贾玉进, 张振程, 李滢铭, 等. 基于轻量化 YOLOv5 网络的输电线路绝缘子缺陷检测[J]. 电力学报, 2024, 39(1): 36-44.

Jia Y J, Zhang Z C, Li X M, et al. Defect detection of insulator on transmission line based on lightweight YOLOv5 network[J]. Journal of Electric Power, 2024, 39(1): 36-44(in Chinese).

[7] Zhang S H, Qu C N, Ru C Y, et al. Multi-objects recognition and self-explosion defect detection method for insulators based on lightweight GhostNet-YOLOV4 model deployed onboard UAV[J]. IEEE Access, 2023, 11: 39713-39725.

[8] 陈佳韵, 肖根福, 张祥明. 基于改进 YOLOv7-tiny 的自爆绝缘子检测算法[J]. 电子测量技术, 2025, 48(7): 66-74.

Chen J Y, Xiao G F, Zhang X M. Detection algorithm for self-exploding insulator based on improved YOLOv7-tiny[J]. Electronic Measurement Technology, 2025, 48(7): 66-74(in Chinese).

[9] Chen Y F, Zhang C Y, Chen B, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases[J]. Computers in biology and medicine, 2024, 170: 107917.

[10] Qin D F, Leichner C, Delakis M, et al. MobileNetV4: universal

models for the Mobile ecosystem[C]//Proceedings of the Computer Vision-ECCV 2024. Berlin: Springer, 2025: 78-96.

[11] Wang W H, Dai J F, Chen Z, et al. InternImage: exploring large-scale vision foundation models with deformable convolutions[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 14408-14419.

[12] 路成龙, 冯月贵, 庆光蔚. 基于改进 YOLOv8 的电梯内电动车识别方法研究[J]. 机械制造与自动化, 2024, 53(4): 219-223.

Lu C L, Feng Y G, Qing G W. Research on identification method of electric vehicles in elevators based on improved YOLOv8[J]. Machine Building & Automation, 2024, 53(4): 219-223(in Chinese).

[13] 侯卫民, 何孟玲, 赵梦瑶, 等. 无人机视角下施工现场工人防护用具检测方法研究[J]. 计算机工程与应用, 2025, 61(15): 353-362.

Hou W M, He M L, Zhao M Y, et al. Research on detection method of protective equipment for construction site workers from UAV perspective[J]. Computer Engineering and Applications, 2025, 61(15): 353-362(in Chinese).

[14] 何飞熊, 谢海巍, 蒲超, 等. 基于改进 YOLOv8 网络的道路病害检测方法[J]. 计算机与现代化, 2025(2): 108-113.

He F X, Xie H W, Pu C, et al. A new method of pavement disease detection based on improved YOLOv8[J]. Computer and Modernization, 2025(2): 108-113(in Chinese).

[15] Dai X Y, Chen Y P, Xiao B, et al. Dynamic head: unifying object detection heads with attentions[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 7369-7378.

[16] 刘柏甫. 面向三维点云数据的目标检测方法研究[D]. 北京: 北京交通大学, 2023.

Liu B F. Research on object detection method for 3D point cloud data[D]. Beijing: Beijing Jiaotong University, 2023(in Chinese).

[17] 孙光灵, 黄磊, 吴倩. 基于 KD-YOLO 的轻量化安全帽佩戴检测方法[J]. 淮北师范大学学报(自然科学版), 2024, 45(2): 41-48.

Sun G L, Huang L, Wu Q. Lightweight safety helmet wearing detection method based on KD-YOLO[J]. Journal of Huaibei Normal University (Natural Science), 2024, 45(2): 41-48(in Chinese).

[18] Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements[EB/OL]. 2024: arXiv: 2410.17725. <https://arxiv.org/abs/2410.17725>.

[19] Jiang P Y, Ergu D, Liu F Y, et al. A review of YOLO algorithm developments[J]. Procedia Computer Science, 2022, 199: 1066-1073.

[20] Hussain M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection[J]. Machines, 2023, 11(7): 677.

[21] Lu Y, Li D H, Gao Q, et al. IDD-YOLOv5: a lightweight insulator defect real-time detection algorithm[C]//Proceedings of the 2024 IEEE International Conference on Mechatronics and Automation. Piscataway: IEEE Press, 2024: 491-495.

Lightweight insulator defect detection algorithm based on UAV perspective

HU Xiaojie^{1, 2}, NI Cui^{1, 2, *}, WU Chunpo^{1, 2}, LIU Zhu^{1, 2}, WANG Peng^{1, 2}

(1. Shandong Provincial Key Laboratory of Key Technologies and Systems of Intelligent Construction Equipment, Shandong Jiaotong University, Jinan 250357, China;
2. School of Information Science and Electrical Engineering (School of Artificial Intelligence), Shandong Jiaotong University, Jinan 250357, China)

Abstract: As the primary method of low-altitude power monitoring, unmanned aerial vehicle (UAV) inspection presents the dual requirements of lightweight and high precision for the detection model. The timely detection of defects is crucial to the reliability of the power grid since it is the fundamental component of the safe operation of the power system. Based on YOLOv11, a lightweight insulator defect detection algorithm based on the UAV perspective is proposed. Firstly, in the YOLOv11 backbone network, the improved feature extraction unit of MobileNetV4, with a general inverted bottleneck structure, is integrated to enhance the perception of subtle defects of insulators. Secondly, the YOLOv11 neck network was integrated with a hierarchical spatial screening feature pyramid network, and the interaction path of cross-layer features was optimized to minimize model parameter redundancy. Finally, at the detection output, the dynamic deformable convolution detection head is used to replace the traditional detection module to improve the adaptability to the geometric deformation of defects. Experimental results show that compared with the YOLO series model, the proposed lightweight model can reduce the number of parameters by more than 12.35% on the basis of ensuring detection accuracy, which is more suitable for edge equipment such as UAVs and inspection robots, and provides an efficient solution for the real-time detection of transmission line insulator defects.

Keywords: insulator; defect detection; feature enhancement; lightweight; YOLOv11

面向低空场景的光场高斯泼溅表示及空间-角度同时超分辨率方法

孟佳涵¹, 杨驰¹, 李传君², 林丽莉², 周文晖^{1,*}

(1. 杭州电子科技大学 计算机学院, 杭州 310018; 2. 浙江工商大学 信息与电子工程学院, 杭州 310018)

摘 要: 光场成像技术可同时捕获光线的空间位置和角度信息, 在无人机感知、环境监测等低空经济应用场景中具有广阔应用前景。然而, 现有光场相机成像及超分辨率方法难以兼顾空间分辨率与角度分辨率, 制约了其在精确感知和高精度建模等任务中的应用。为此, 提出一种面向低空场景的光场高斯泼溅表示及光场空间-角度同时超分辨率 (LFASSR) 方法。将传统 3D 高斯泼溅建模方法拓展到 4D 光场表示空间中, 提出一种 4D 光场高斯泼溅表示方法, 并利用光场极平面 (EPI) 的空间-角度几何特性, 将其分解为水平与垂直方向上的 2D 高斯泼溅级联建模, 以更有效地表征光场几何特征。基于此, 构建了一个通用的光场图像超分辨率框架, 在提取并融合光场空间、角度及 EPI 域特征后, 构建水平和垂直方向的光场几何特征空间, 并分别进行 2D 高斯泼溅建模, 最终经 EPI 渲染, 生成致密视角且高空间分辨率的高质量光场图像。在公开光场图像数据集及低空场景光场数据上的实验结果表明: 所提方法在 LFASSR 任务中显著优于现有方法, 在峰值信噪比 (PSNR) 和结构相似性指标 (SSIM) 等客观指标上均达到当前最优水平, 在视觉质量上展现出清晰的细节恢复能力与自然的视差过渡效果。在独立的光场空间超分辨率 (LFSSR) 任务和光场角度超分辨率 (LFASR) 任务中, 所提方法同样展现出良好的泛化能力与竞争力。

关 键 词: 光场; 光场超分辨率; 光场高斯泼溅; 空间-角度同时超分辨率; 极平面几何

中图分类号: TP391.4

文献标志码: A

文章编号: 1001-5965(2026)09-3189-13

随着低空经济的快速发展, 无人机巡检、城市空中交通、精准农业和三维环境感知等新兴应用场景对多维视觉信息的需求日益增长。相较于传统的二维成像机制, 光场成像技术不仅记录了场景中光线的强度, 还保留了其传播方向信息, 因而能在单次拍摄中捕获多视角视差、透视偏移、深度线索在内的高维视觉信息^[1-2]。这一特性使其在实现重聚焦、视点合成与三维感知等任务中具有独特优势, 正逐步发展为支撑低空智能感知系统的关键技术之一。

然而, 受限于微透镜阵列的物理采样机制, 现有光场成像在空间分辨率与角度分辨率之间存在固有的权衡矛盾^[3], 导致成像质量难以满足低空经济应用中高精度感知与精细化建模的任务需求, 制约了光场技术在相关场景中的应用。为突破光场相机的物理成像极限, 近年来光场超分辨率技术受到广泛关注, 并取得显著进展。

光场超分辨率技术主要分为光场角度超分辨率 (light field angular super-resolution, LFASR) 和光场空间超分辨率 (light field spatial super-resolution,

收稿日期: 2025-11-17; 录用日期: 2025-12-12; 网络出版时间: 2026-02-02 11:35

网络出版地址: link.cnki.net/urlid/11.2625.V.20260130.1851.004

基金项目: 国家自然科学基金 (62471437); 浙江省“尖兵”“领雁”科技计划项目 (2023C01030)

* 通信作者. E-mail: zhouwenhui@hdu.edu.cn

引用格式: 孟佳涵, 杨驰, 李传君, 等. 面向低空场景的光场高斯泼溅表示及空间-角度同时超分辨率方法 [J]. 北京航空航天大学学报, 2026, 52 (9): 3189-3201. Meng J H, Yang C, Li C J, et al. Light field gaussian splatting representation and joint spatial-angular super-resolution for low-altitude scenarios [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3189-3201 (in Chinese).

LFSSR)2 类。其中, LFASR 聚焦于光场的光线角度域, 增加光线角度的采样密度, 提升光场图像在角度域上的信息丰富度; LFSSR 旨在提升光场子孔径图像的空间分辨率, 以增强光场图像在空间维度上的细节表现。

在 LFASR 研究中, Kalantari 等^[4]首次将深度学习引入该任务, 将新视角合成分解为视差估计和颜色估计 2 个阶段, 并通过串联的卷积神经网络(convolutional neural network, CNN)分别建模。随后, Zhang 等^[5]利用垂直和水平光场极平面(epipolar plane image, EPI)特征, 通过反卷积实现角度域致密采样。针对大基线稀疏光场, Jin 等^[6]基于光场几何设计了深度估计与新视角合成的融合框架。Liu 等^[7]在宏像素特征空间进行 LFASR, 在有效处理大视差的同时降低推理开销。Wang 等^[8]提出的光场特征卷积在保留视角相关性与 EPI 结构的同时缓解空间与角度信息不平衡问题。Mao 等^[9]通过自适应对齐和特征解耦, 提高了模型对复杂场景和低光照的适应能力, 减少了新视角合成中的视觉伪影。

在 LFSSR 研究中, Yuan 等^[10]先对各子孔径图像独立进行 LFSRR, 再在 EPI 域通过卷积模块增强几何一致性。Wang 等^[11]提出双向循环卷积网络, 分别建模水平和垂直方向相邻子孔径图像间的空间关系, 并通过堆叠泛化进行融合; 在此基础上, Zhang 等^[12]引入对角线方向建模, 并采用残差融合模块实现光场的高分辨空间重建。随后, Liang 等^[13]通过角度 Transformer 和空间 Transformer 建模视角间互补信息及子孔径图像内的长短程依赖。Gao 等^[14]与 Chen 等^[15]分别基于状态空间模型引入 Mamba 架构和选择性扫描机制, 在 EPI 或空间-角度联合特征建模中实现了更优的性能与效率平衡。

尽管光场超分辨率技术已取得显著进展, 但大多数现有研究都聚焦于 LFSSR 或 LFASR 的单一任务, 不同任务所采用的网络结构亦存在差异, 且难以实现光场图像的空间-角度同时超分辨率。为此, Wang 等^[16]设计了一种通用光场图像超分辨率框架 Distg(SSR/ASR), 以适用于 LFSSR 和 LFASR 任务。Van Duong 等^[17]探索了空间-角度域内的像素相关性, 以重建空间或角度高分辨率的光场图像。Yoon 等^[18]尝试了光场图像的空间与角度同时超分辨率, 分别训练了基于 CNN 的 LFSRR 网络和 LFASR 网络, 然后, 通过端到端方式进行联合微调。Liu 等^[19]面向光场空间-角度同时超分辨率(light field angular spatial super-resolution, LFASSR)

的光场重建方法, 通过自适应聚合多视角像素信息, 在空间域与角度域中同时建模视角间的相关性, 实现了致密高分辨率光场图像的联合合成。Liu 等^[20]进一步提出了一种隐式细节增强网络, 在隐式表征空间中学习不同方向上的像素差分关系, 有效提升了 LFASSR 任务中细节恢复的质量, 重建出纹理更加丰富、结构更加一致的高分辨率光场图像。然而, 与单一的 LFSSR 或 LFASR 任务相比, LFASSR 研究相对较少, 实现 2 个维度的同步增强仍面临挑战。

本文提出了一种面向低空场景的光场高斯泼溅表示及 LFASSR 框架。不同于传统光场超分方法通常在空间域或角度域分别建模、逐步重建的策略, 本文将近年来新兴的 3D 高斯泼溅技术^[21-22]引入光场建模中, 并进一步扩展至光场的双平面表示空间, 提出了 4D 光场高斯泼溅表示。该表示充分利用光场水平和垂直 EPI 的空间-角度几何特性, 将其分解为水平与垂直方向上 2 个 2D 高斯泼溅的级联表示, 避免了传统方法中空间与角度超分相互独立或顺序执行所带来的误差累积, 从而在光场 EPI 域上实现真正意义的空间-角度同时高分辨率渲染。本文主要贡献如下:

- 1) 设计一种基于 4D 光场高斯泼溅表示的光场图像超分辨率通用框架, 不仅可实现光场图像空间域或角度域的单域超分辨率, 而且可实现 LFASSR。
- 2) 提出一种 4D 光场高斯泼溅建模方法, 利用光场 EPI 的空间-角度几何特性, 将 4D 光场高斯泼溅表示分解为水平与垂直方向上 2 个 2D 高斯泼溅的级联表示, 并将水平/垂直 EPI 中的像素点扩散为连续高斯场, 通过渲染相互堆叠的高斯场, 恢复高分辨率光场的几何细节。
- 3) 提出的光场图像超分辨率方法已在多种光场图像超分辨率任务中进行了测试, 包括 LFASSR、单一的 LFSSR 和 LFASR 任务, 在评价指标和视觉效果方面均优于现有方法; 在低空场景光场数据上的泛化测试也展现出令人满意的性能和渲染效果。

1 本文方法

1.1 4D 光场高斯泼溅表示

光场双平面表示法 $L(u, v, x, y) \in \mathbf{R}^{U \times V \times W \times H}$ 采用一个四维函数来描述场景中 (u, v) 视角方向上的光线在子孔径图像坐标 (x, y) 上的投影, (u, v) 与 (x, y) 分别为场景中光线与光场角度域平面和空间域平面的交点坐标。

本文首先将传统的 3D 高斯泼溅技术扩展到光

场双平面表示中, 提出一种 4D 光场高斯泼溅 (4D light field Gaussian splatting, 4DLFGS) 表示方法。该方法将场景用一组 4D 高斯场表示, 每个高斯场的属性除均值向量外都可学习, 包括协方差矩阵、不透明度和颜色信息。具体而言, 令场景中第 i 条光线的双平面坐标为 (u_i, v_i, x_i, y_i) , 假定其服从于高斯分布, 则该光线的 4D 高斯场定义为

$$f_i(\mathbf{p}) = \sigma_i \exp \left(-\frac{1}{2} (\mathbf{p} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{p} - \boldsymbol{\mu}_i) \right) \quad (1)$$

式中: $\mathbf{p} = (u, v, x, y)^\top$ 为光场双平面表示中的任意点位置; $\boldsymbol{\mu}_i = (u_i, v_i, x_i, y_i)^\top$ 为 4D 高斯场的均值向量; $\boldsymbol{\Sigma}_i \in \mathbf{R}^{4 \times 4}$ 为其协方差矩阵; σ_i 为不透明度; $f_i(\mathbf{p})$ 为第 i 个高斯场对 $\mathbf{p}$ 点的渲染权重。

在高斯泼溅渲染过程中, 采用传统 3D 高斯泼溅中的 α 混合技术, 双平面坐标 $\mathbf{p}$ 点的颜色信息 $C_{\text{LFGS}}(\mathbf{p})$ 由所有 4D 高斯场加权得到:

$$C_{\text{LFGS}}(\mathbf{p}) = \sum_{i \in N} c_i f_i^{2D}(\mathbf{p}) \prod_{j=1}^{i-1} (1 - f_j^{2D}(\mathbf{p})) \quad (2)$$

式中: $f_i^{2D}(\cdot)$ 为 $f_i(\cdot)$ 在二维平面的投影; c_i 为第 i 个高斯场的 RGB 值或球谐函数系数。

根据光场水平和垂直 EPI 表示及其空间-角度几何特性, 本文进一步将 4D 光场高斯泼溅建模分解为水平和垂直方向上 2 个 2D 高斯泼溅的级联建模。具体而言, 将光场双平面坐标 (u, v, x, y) 划分为水平 EPI 坐标 (u, x) 和垂直 EPI 坐标 (v, y) , 从而将 4D 高斯模型 $f_i(\mathbf{p})$ 分解为水平 EPI 上的 2D 高斯模型 $f_{i,u}(\mathbf{p}_u)$ 和垂直 EPI 上的 2D 高斯模型 $f_{i,v}(\mathbf{p}_v)$ 的级联形式, 如式 (3) 所示。

$$\begin{cases} f_i(\mathbf{p}) = f_{i,u}(\mathbf{p}_u) \cdot f_{i,v}(\mathbf{p}_v) \\ f_{i,u}(\mathbf{p}_u) = \sigma_{i,u} \exp \left(-\frac{1}{2} (\mathbf{p}_u - \boldsymbol{\mu}_{i,u})^\top \boldsymbol{\Sigma}_{i,u}^{-1} (\mathbf{p}_u - \boldsymbol{\mu}_{i,u}) \right) \\ f_{i,v}(\mathbf{p}_v) = \sigma_{i,v} \exp \left(-\frac{1}{2} (\mathbf{p}_v - \boldsymbol{\mu}_{i,v})^\top \boldsymbol{\Sigma}_{i,v}^{-1} (\mathbf{p}_v - \boldsymbol{\mu}_{i,v}) \right) \end{cases} \quad (3)$$

式中: $\mathbf{p}_u = (u, x)^\top$ 和 $\mathbf{p}_v = (v, y)^\top$ 分别为水平和垂直 EPI 坐标; $\boldsymbol{\mu}_{i,u} = (u_i, x_i)^\top$ 和 $\boldsymbol{\mu}_{i,v} = (v_i, y_i)^\top$ 分别为水平 EPI 和垂直 EPI 上 2D 高斯模型的均值向量; $\boldsymbol{\Sigma}_{i,u}$ 和 $\boldsymbol{\Sigma}_{i,v}$ 为对应的协方差矩阵。

进而, 4D 光场高斯泼溅的渲染过程分解为 2 个 2D 高斯泼溅的级联渲染, 如式 (4) 所示。

$$\begin{cases} C_{\text{LFGS}}(\mathbf{p}_v) = \sum_{j \in N} c_{j,v} f_{j,v}^{2D}(\mathbf{p}_v) \prod_{j=1}^{j-1} (1 - f_{j,v}^{2D}(\mathbf{p}_v)) \\ c_v(\mathbf{p}_u) = \sum_{i \in N} c_{i,u} f_{i,u}^{2D}(\mathbf{p}_u) \prod_{i=1}^{i-1} (1 - f_{i,u}^{2D}(\mathbf{p}_u)) \end{cases} \quad (4)$$

式中: $f_{i,u}^{2D}(\cdot)$ 和 $f_{j,v}^{2D}(\cdot)$ 分别为 $f_{i,u}(\cdot)$ 和 $f_{j,v}(\cdot)$ 的 2D 平面

投影; $c_{i,u}$ 和 $c_{j,v}$ 分别为水平和垂直 EPI 中的高斯模型颜色信息。为减少计算量, 具体实现时采用简化 α 混合技术, 即式 (4) 简化为

$$\begin{cases} C_{\text{LFGS}}(\mathbf{p}_v) = \sum_{j \in N} c_{j,v} f_{j,v}^{2D}(\mathbf{p}_v) \\ c_v(\mathbf{p}_u) = \sum_{i \in N} c_{i,u} f_{i,u}^{2D}(\mathbf{p}_u) \end{cases} \quad (5)$$

基于上述分解过程, 本文采用水平和垂直 EPI 上的 2 个级联 2D 高斯泼溅模型来表示光场在水平与垂直 EPI 上的几何结构, 从而实现对四维光场整体几何分布的高效表示。

1.2 整体网络结构

本文提出的 LFASSR 网络架构包含 3 个模块: 光场几何编码模块、4D 光场高斯泼溅模块和光场图像生成模块, 实现了对输入光场图像 $\mathbf{L}(u, v, x, y) \in \mathbf{R}^{U \times V \times W \times H}$ 在其空间和角度域上的同时超分辨率, 得到空间-角度高分辨率的光场图像 $\tilde{\mathbf{L}}(u, v, x, y) \in \mathbf{R}^{U' \times V' \times W' \times H'}$ 。

如图 1 所示, 首先, 以一个稀疏视角且低空间分辨率的光场图像 $\mathbf{L}(u, v, x, y)$ 作为输入, 在光场几何编码模块中提取和融合光场图像的空间、角度和 EPI 特征, 得到空间-角度联合特征 $\mathbf{Z}_g$; 然后, 将 $\mathbf{Z}_g$ 输入 4D 光场高斯泼溅模块中建模高斯场, 并在光场空间和角度域上实现上采样渲染, 得到光场潜在特征编码 $\mathbf{Z}'$; 最后, 由光场图像生成模块从潜在特征编码 $\mathbf{Z}'$ 中生成空间-角度高分辨率的光场图像 $\tilde{\mathbf{L}}(u, v, x, y)$ 。

本文在模型训练阶段使用 $L1$ 损失函数, 以衡量生成的高分辨率光场图像 $\tilde{\mathbf{L}}$ 与真实高分辨率光场图像 $\hat{\mathbf{L}}$ 间的像素级差异, 通过最小化损失函数来优化模型参数, 具体损失函数为

$$\mathcal{L} = \sum |\tilde{\mathbf{L}}(u, v, x, y) - \hat{\mathbf{L}}(u, v, x, y)| \quad (6)$$

1.3 光场几何编码模块

光场几何编码模块旨在从输入光场图像中提取其空间、角度及 EPI 的联合特征充分挖掘光场图像中的空间/几何一致性, 以及视角关联性, 为后续超分辨率重建提供高阶特征的表征能力。具体如图 2(a) 所示, 输入的光场图像 $\mathbf{L}(u, v, x, y)$ 先经过一组由 3×3 卷积层组成的卷积模块, 逐层提升特征通道数至 64, 从而得到初始特征图 $\mathbf{F}_{\text{init}}$ 。随后, 引入一个自适应特征融合模块^[23], 该模块以分块形式逐级融合光场几何特征, 在每块中使用不同卷积层分别提取空间、角度、垂直 EPI 和水平 EPI 的特征, 并在每块的末端采用短残差连接, 以稳定梯度传播并有效保留结构相关的浅层特征。

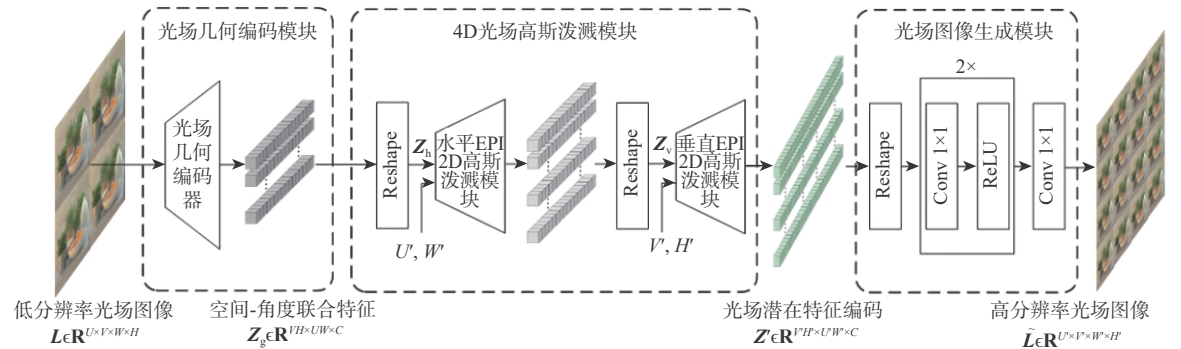


图1 整体网络结构框

Fig. 1 Overall framework of the proposed network

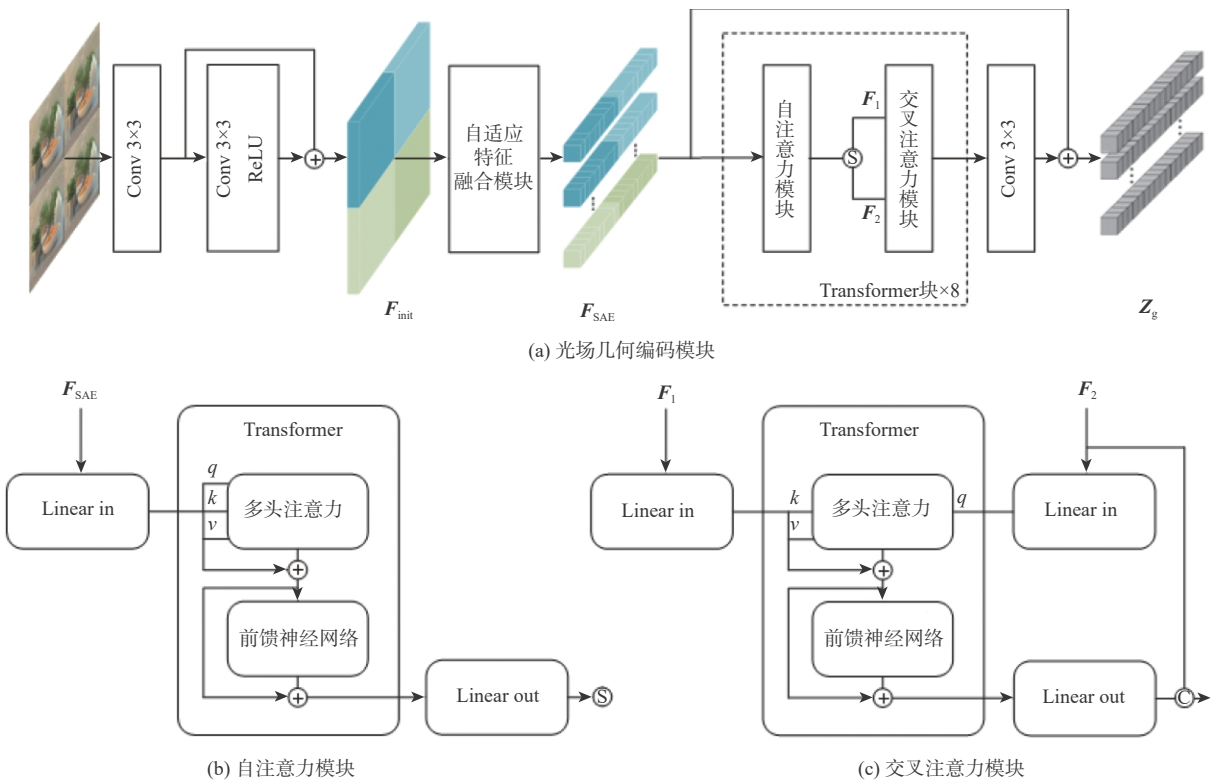


图2 光场几何编码模块和注意力模块

Fig. 2 Light field geometry encoding module and attention modules

将经过残差网络训练增强得到的特征 F_{SAE} 输入一个基于 Transformer 的特征增强单元。该单元利用注意力机制捕获长程依赖关系，从全局视角整合光场不同视角间的结构信息，以获得更完整且表达力更强的空间-角度联合特征 Z_g 。该特征增强单元由 8 个 Transformer 块组成，每块都包含自注意力模块和交叉注意力模块。每个自注意力/交叉注意力模块内部都包含一个多头注意力结构和一个前馈神经网络，如图 2(b) 和图 2(c) 所示。线性输入层 (Linear in) 和线性输出层 (Linear out) 分别用来提升和压缩特征通道数，在保证输入/输出维度一致的同时增强特征表示能力。在自注意力模块中，直接将整个光场特征图作为多头注意力机制的输入，以捕获特征间的全局依赖关系并增强特征表达能力。

在交叉注意力模块中，输入特征首先在通道维度上进行分割操作，得到 2 个特征分支 F_1 和 F_2 ；随后将 F_1 和 F_2 输入到 2 个独立训练的线性输入层中以提升通道维度，并将其分别作为键值对 (k, v) 和查询 (q) 输入多头注意力结构中，并经前馈神经网络增强非线性表示能力。将线性输出层生成的特征与输入特征 F_2 进行张量拼接，传递至下一个 Transformer 块。

经所有 Transformer 块的特征提取与全局建模后，输出特征被送入一个 3×3 卷积层，以补充 Transformer 结构在细节层面上建模能力不足的问题，从而增强特征的局部一致性与纹理保真度。随后，将该卷积层的输出与 Transformer 特征增强单元的输入 F_{SAE} 通过短残差连接进行相加，在融合局部

与全局信息的同时保持特征分布的稳定性, 从而获得几何一致性的光场特征表示。

1.4 4D 光场高斯泼溅模块

4D 光场高斯泼溅模块旨在在光场特征空间中构建 1.1 节所述的 4D 光场高斯泼溅模型, 并将其分解为垂直和水平 EPI 特征图上的 2 个 2D 高斯泼溅模块, 再在高分辨率网格上对 EPI 特征图进行上采样渲染。具体如图 1 所示, 空间-角度联合特征 $\mathbf{Z}_g$ 先经过 Reshape 操作得到水平 EPI 特征图 $\mathbf{Z}_h$, 维度为 (VHC, U, W) ; 将其输入到水平 EPI 2D 高斯泼溅模块中, 对角度维度 U 和空间维度 W 进行上采样和渲染, 得到增强后的水平 EPI 特征图 $\mathbf{Z}_h'$, 维度为 (VHC, U', W') 。随后, $\mathbf{Z}_h'$ 经过 Reshape 操作将超分后的 U' 、 W' 和特征通道 C 合并为一个维度, 得到垂直 EPI 特征图 $\mathbf{Z}_v$, 维度为 $(U'W'C, V, H)$, 并送入垂直 EPI 2D 高斯泼溅模块完成对 V 和 H 维度的上采样和渲染, 得到增强后维度为 $(U'W'C, V', H')$ 的垂直 EPI 特征图 $\mathbf{Z}_v'$, 重组为高维光场特征编码 $\mathbf{Z}' \in \mathbf{R}^{V'H' \times U'W' \times C}$ 。

EPI 特征图的 2D 高斯泼溅渲染示意如图 3 所示, 低分辨率特征网格上的每个特征点(图中红实心点)都被建模为一个 2D 高斯椭圆(场), 则高分辨率网格中特征点 $\mathbf{u}_q$ 的颜色信息受到周围所有 2D 高斯椭圆的叠加影响。根据式 (3), 第 i 个 2D 高斯椭圆对 $\mathbf{u}_q$ 渲染的权重值可表示为

$$f_i(\mathbf{u}_q) = \sigma_i \exp\left(-\frac{1}{2}(\mathbf{u}_q - \mathbf{u}_i)^T \Sigma_i^{-1}(\mathbf{u}_q - \mathbf{u}_i)\right) \quad (7)$$

式中: 均值向量 $\mathbf{u}_i \in \mathbf{R}^2$ 为低分辨率 EPI 特征图中第 i 个特征点坐标; 协方差矩阵 Σ_i 和 σ_i 分别为第 i 个 2D 高斯椭圆的形状参数和不透明度。

训练时, 先在低分辨率网格上初始化每个 2D 高斯椭圆, 并在训练过程中不断优化光场 EPI 特征图、协方差矩阵和不透明度。将所有 2D 高斯椭圆作用在高分辨率网格上, 生成高分辨率 EPI 特征图。根据式 (5), 高分辨率网格中特征点 $\mathbf{u}_q$ 的颜色信息可由式 (8) 计算得到:

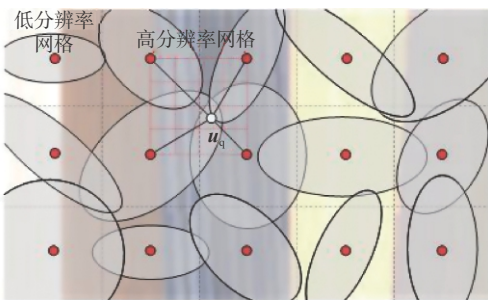


图 3 低分辨率 EPI 网格上构建的 2D 高斯泼溅模型示意

Fig. 3 Illustration of the 2D Gaussian splatting model constructed on a low-resolution EPI grid

$$\mathbf{Z}_i(\mathbf{u}_q) = \sum_{i \in N} f_i(\mathbf{u}_q) \mathbf{z}_i \quad (8)$$

式中: $\mathbf{z}_i$ 为低分辨率 EPI 特征图中第 i 个特征点的特征向量。

1.5 光场图像生成模块

在完成 4D 光场高斯泼溅建模和渲染后, 还需将高维光场特征编码 $\mathbf{Z}'$ 还原为相应的光场双平面表示, 并优化生成结果。为此, 本文设计了一个光场图像生成模块, 用于将高斯泼溅渲染后的特征映射回光场双平面表示。具体而言, 该模块将 $\mathbf{Z}'$ 作为输入, 首先, 采用 Reshape 操作将光场特征图的维度转化为 $(C, U'V', H', W)$ 。然后, 依次通过 3 个连续的 1×1 卷积层, 将特征图的通道数 C 逐步压缩至单通道, 以恢复图像的光强信息。每个卷积层之间均引入 LeakyReLU 激活函数, 以增强网络的非线性建模能力。最后, 将卷积层输出结果重组为光场子孔径图像阵列, 从而得到最终的高分辨率光场图像 $\tilde{\mathbf{L}}$ 。

2 实验

对比实验首先采用当前 LFASR/LFSSR 任务的 2 套常用公开数据集。其中, LFASR 数据集包括 2 个合成场景光场数据集 HCInew 和 HCIdold 及真实场景光场数据集 30Scenes、Occlusion 和 Reflective。LFSSR 数据集包括合成场景数据集 HCInew 和 HCIdold 及真实场景数据集 EPFL、INRIA 和 STFgantry。训练集与测试集的划分方式与 Distg^[16] 保持一致。在面向低空场景的泛化性能测试中, 从 UrbanLF-Real^[24] 光场数据集共 164 个测试场景中选取 12 个典型低空应用的光场图像作为测试集, 这些光场图像由第二代光场相机 Lytro Illum 采集得到, 包含车辆、道路、建筑等低空场景。图 4 展示了其中部分场景示例。在上述数据集中完成了下列实验: ①在 LFASR 任务中, 将角度分辨率从 2 像素 $\times$ 2 像素放大至 5 $\times$ 5, 同时放大空间分辨率为原来的 2 倍; ②在单一的 LFSSR 任务中, 保持角度分辨率不变, 将 5 像素 $\times$ 5 像素光场图像的空间分辨率提升至 2 倍或 4 倍; ③在单一的 LFASR 任务中, 保持空间分辨率不变, 将角度分辨率从 2 像素 $\times$ 2 像素放大至 7 像素 $\times$ 7 像素; ④在低空场景测试集上测试 LFASR 模型的泛化性能。对比方法选取近年来表现优异且具有代表性的模型作为参考。在实验中, 所有选取的对比方法均在相同的光场数据集上进行训练和测试。对于 LFASR 和 LFSSR 任务, 在表中直接以原始论文提供的结果作为对比数据。而在 LFASR 中, 实验选取一些提供了源码的角度或空间超分模型进行组合, 并在同样的数据集下进行重新训练。



图4 低空场景(建筑、桥梁、交通、道路)下的光场图像及其EPI图像

Fig. 4 Light field images and their EPIs in low-altitude scenarios (e.g., buildings, bridges, traffic, and roads)

2.1 实验细节

本文实验在一台配备 Intel Xeon E5-2680 v4 CPU、64 GB DDR4 内存和一张 NVIDIA RTX 3090 GPU 的服务器上进行。网络训练采用自适应矩估计(Adam)优化器,其中,一阶矩估计的衰减率 $\beta_1=0.9$,二阶矩估计的衰减率 $\beta_2=0.999$ 。每次训练的批次大小为1。初始学习率设为 1×10^{-4} ,学习率每15个训练轮次减半,模型在训练约70轮后基本收敛。

在指标评估时,由于现有光场超分辨率方法均在 YCbCr 色彩空间中的 Y 通道上进行评估,因此,本文也同样在光场图像的 Y 通道计算峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似性指标(structural similarity index, SSIM)等指标。具体而言,本文在所有子孔径图像上计算超分辨率重建结果与 Ground Truth 之间的 PSNR 和 SSIM,并取其平均值作为光场重建质量的评价指标。值得注意的是,在 LFASR 任务中,输入的 2×2 光场在角度维度的超分过程中不涉及插值操作,因此,子孔径阵列4个角位置的图像与 Ground Truth 完全一致,其指标显著高于其余子孔径图像。为避免该现象对整体评估结果造成干扰,本文在计算平均 PSNR 和 SSIM 时排除了这4个角上的子孔径图像。相比之下,在 LFSSR 任务及空间-角度联合超分辨率任务中,所有子视图均参与了插值和重建过程,因此,在指标计算时纳入全部子孔径图像。另外,在生成高分辨率的 RGB 光场图像时,本文先将低分辨率 RGB 光场图像转换至 YCbCr 颜色空间,并将 Y 通道(亮度信息)作为模型的输入。对于 Cb 和 Cr 通道(色度信息),采用双三次插值进行上采样。将模型生成的高分辨率 Y 通道与上采样后的 Cb、Cr 通道重新合并,经过转换得到高分辨率的 RGB 光场图像。

处理光场数据集时,首先提取原始光场图像中心的5像素 $\times$ 5像素或7像素 $\times$ 7像素作为 Ground Truth 光场图像(由光场超分辨率任务确定),再从中提取位于子孔径图像阵列四个角上的子孔径图像,构建角度分辨率为2像素 $\times$ 2像素的稀疏光场,作为 LFASR 任务的输入。对于所有超分辨率任务,在生成训练集时,都需将原始光场图像分割成

小像素块。在 LFSSR 和 LFASSR 任务中,还需将初始像素块下采样到32像素 $\times$ 32像素作为低分辨率输入。对于空间分辨率放大2倍和4倍的任务,初始像素块的空间分辨率分别是64像素 $\times$ 64像素和128像素 $\times$ 128像素。

为提升模型泛化能力,在训练过程中对光场图像进行水平、垂直及对角线方向的翻转操作。在复现其他方法时,同样采用相同的数据增强策略,以确保实验设置的一致性与公平性。

2.2 LFASSR 任务

在 LFASSR 任务中,由于现有的研究缺乏 LFASSR 模型,本实验将现有单一空间和单一 LFASR 模型组合起来进行训练,一共对比了6种拼接模型和1种同时超分辨率模型 LFSSR^[19]。所有拼接模型都是先 LFASR 再 LFSSR。数据集采用 LFASR 任务中3个真实光场数据集(30Scenes、Occlusion 和 Reflective)及 LFSSR 任务中的 EPFL、HClnew、HClold 数据集。表1评估了光场图像在空间分辨率上放大2倍,角度分辨率从2像素 $\times$ 2像素放大至5像素 $\times$ 5像素的合成质量。量化比对结果表明,本文模型的 PSNR 和 SSIM 指标均优于其他模型^[25]。

图5比对了各模型在3个测试场景中的超分辨率视觉效果,本文模型的整体重建效果更接近真实光场图像。在 ISO_Chart 和 Reflective_22 场景中,本文模型在生成细密纹理方面表现更为出色。在 Occlusions_35 场景中,现有模型的生成结果在一些区域出现颜色缺失(黄框标记),EPI 图像也可观察到这些现象。

2.3 LFASR 任务

在单一 LFASR 任务中,数据集采用 HClnew、HClold、30Scenes、Occlusion 和 Reflective,评估了由2像素 $\times$ 2像素稀疏采样光场图像重建得到7像素 $\times$ 7像素光场图像的合成质量。

表2列出了本文模型与现有7种 LFASR 模型的量化对比结果。在 HClnew 和 HClold 合成数据集上,本文模型在各项指标上均显著优于其他模型^[26-29]。具体而言,在 HClnew 数据集上,本文模型的 PSNR 比次优的 HLFASR-ASR 模型高出1.28 dB,在 HClold 数据集上也高出0.74 dB。在真实场景数据集 30Scenes 上,本文模型在 PSNR 指标上取得最高分,SSIM 指标略低于最高分;在 Occlusion 和 Reflective 数据集上,本文模型也在 SSIM 指标方面表现优异。

2.4 LFSSR 任务

在单一的光场空间超分辨率任务中,本文在

表 1 本文模型与 7 种模型在空间 (×2)-角度 (2×2→5×5) 同时超分辨率任务上的性能对比

		PSNR/dB, SSIM					
LFASR模型	LFSSR模型	30Scenes	Occlusion	Reflective	HCInew	HCIOld	EPFL
DistgASR ^[16]	DistgSSR ^[16]	41.93 / 0.992 0	38.36 / 0.985 4	38.94 / <u>0.977 7</u>	34.74 / 0.961 7	41.56 / 0.990 5	32.90 / 0.969 5
DistgASR ^[16]	EPIT ^[25]	41.85 / 0.991 8	38.27 / 0.985 1	38.89 / 0.976 8	<u>35.21</u> / 0.965 3	<u>42.12</u> / <u>0.991 4</u>	33.21 / 0.970 3
DistgASR ^[16]	MLFSR ^[14]	41.99 / 0.992 1	38.41 / 0.985 7	39.01 / 0.977 2	34.87 / 0.962 8	41.98 / 0.991 2	33.37 / 0.970 8
LF-EASR ^[7]	DistgSSR ^[16]	41.86 / 0.991 9	38.21 / 0.985 1	38.94 / 0.977 6	34.66 / 0.961 2	41.67 / 0.990 4	33.17 / 0.969 7
LF-EASR ^[7]	EPIT ^[25]	41.80 / 0.991 7	38.15 / 0.984 9	38.93 / 0.977 4	35.06 / <u>0.964 4</u>	42.05 / 0.991 1	33.32 / 0.969 5
LF-EASR ^[7]	MLFSR ^[14]	41.77 / 0.991 7	38.12 / 0.984 9	38.91 / 0.977 2	34.67 / 0.961 3	41.68 / 0.990 7	33.28 / 0.970 0
LFSR ^[19]	LFSR ^[19]	41.89 / 0.991 9	38.30 / 0.985 3	38.86 / 0.977 2	34.46 / 0.959 9	41.72 / 0.990 7	32.98 / 0.969 2
本文模型	本文模型	42.15 / 0.992 3	38.64 / 0.986 2	39.22 / 0.978 2	35.35 / 0.965 3	42.81 / 0.992 5	33.67 / 0.971 9

注: 指标最优和次优分别作加粗和下划线处理。

HCInew、HCIOld、EPFL、INRIA 和 STFgantry 数据集上与现有 9 种模型进行比对, 评估了光场图像在空间上放大 2 倍和 4 倍的合成质量, 表 3 列出了量化对比结果^[30-34]。对比结果表明, 本文模型在 HCInew、

HCIOld、EPFL 和 INRIA 数据集上均取得最优的性能, 仅在 STFgantry 数据集上略低于最优性能。在×2 超分辨率实验中, 本文模型在 EPFL 数据集上的 PSNR 指标较次优的 HLFSSR 模型提升了 0.39 dB。

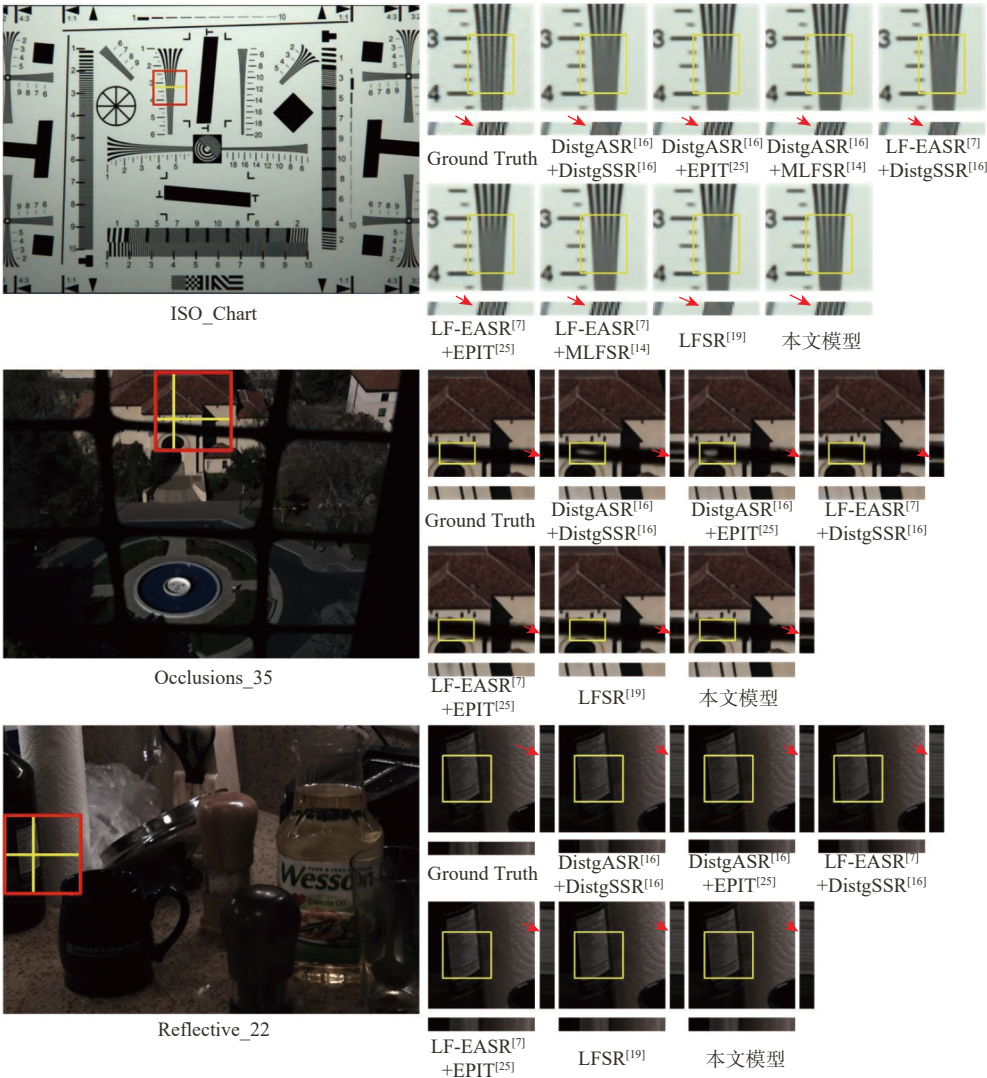


图 5 空间 (×2)-角度 (2×2→5×5) 同时超分辨率任务下不同模型生成的视觉效果对比

Fig. 5 Visual comparisons of different models on the joint spatial (×2) and angular (2×2→5×5) super-resolution task

表 2 本文与 7 种模型在 LFASR (2×2→7×7) 任务上的 PSNR 和 SSIM 对比

Table 2 Performance comparisons on PSNR and SSIM of this paper with seven models in the LFASR task(2×2→7×7)		PSNR/dB, SSIM				
		HCInew	HCInold	30Scenes	Occlusion	Reflective
LF-EASR ^[7]	2022	35.86 / 0.975	41.54 / 0.971	43.44 / <u>0.995</u>	39.80 / 0.992	39.35 / 0.981
DistgASR ^[16]	2022	34.70 / 0.974	42.18 / 0.978	43.67 / <u>0.995</u>	39.46 / <u>0.991</u>	39.11 / 0.978
HLFSR-ASR ^[17]	2023	<u>36.43</u> / <u>0.982</u>	<u>43.25</u> / <u>0.990</u>	<u>44.03</u> / 0.996	<u>39.87</u> / 0.992	39.60 / 0.981
Liu等 ^[26]	2023	35.29 / 0.954	41.92 / 0.969	43.33 / 0.988	39.67 / 0.983	39.20 / 0.963
Li和Zhong ^[27]	2024	35.21 / 0.975	42.36 / 0.979	43.95 / 0.996	39.68 / <u>0.991</u>	39.36 / 0.981
Liu等 ^[28]	2024	35.41 / 0.954	42.02 / 0.969	43.38 / <u>0.995</u>	39.71 / <u>0.991</u>	39.31 / <u>0.982</u>
ViewFormer ^[29]	2025	35.37 / 0.977	41.93 / 0.975	43.82 / 0.995	39.86 / 0.992	<u>39.57</u> / 0.981
本文		37.71 / 0.983	43.99 / 0.993	44.15 / <u>0.995</u>	40.07 / 0.992	39.45 / 0.984

注: 指标最优和次优分别作加粗和下划线处理。

表 3 本文与 9 种模型在 LFSSR 任务 (放大 2 倍和 4 倍) 上的 PSNR 和 SSIM 对比

Table 3 Performance comparisons on PSNR and SSIM of this paper with nine models in the LFSSR tasks(×2 and ×4)		PSNR/dB, SSIM									
		EPFL		HCInew		HCInold		INRIA		STFgantry	
模型	发表年份	放大2倍	放大4倍	放大2倍	放大4倍	放大2倍	放大4倍	放大2倍	放大4倍	放大2倍	放大4倍
DPT ^[30]	2022	34.49	28.94	37.36	31.20	44.30	37.41	36.41	30.96	39.43	31.15
		0.975 8	0.917 0	0.977 1	0.918 8	0.994 3	0.972 1	0.984 3	0.952 3	0.992 6	0.948 8
DistgSSR ^[16]	2022	34.81	28.99	37.96	31.38	44.94	37.56	36.59	30.99	40.40	31.65
		0.978 7	0.919 5	0.979 6	0.921 7	0.994 9	0.973 2	0.985 9	0.951 9	0.994 2	0.953 5
EPIT ^[25]	2023	34.83	<u>29.34</u>	38.23	31.51	<u>45.08</u>	37.68	36.67	31.37	42.17	32.18
		0.977 5	0.919 7	<u>0.981 0</u>	0.923 1	0.994 9	0.973 7	0.985 3	0.952 6	0.995 7	<u>0.957 1</u>
HLFSR-SSR ^[17]	2023	<u>35.31</u>	29.20	<u>38.32</u>	<u>31.57</u>	44.98	37.78	<u>37.06</u>	31.24	40.85	31.64
		0.980 0	0.922 2	0.980 7	<u>0.923 8</u>	<u>0.995 0</u>	0.972 4	<u>0.986 7</u>	<u>0.953 4</u>	0.994 7	0.953 7
LF-DET ^[31]	2023	35.26	29.47	38.31	31.56	44.99	<u>37.84</u>	36.95	<u>31.39</u>	<u>41.76</u>	<u>32.14</u>
		0.979 7	<u>0.923 0</u>	0.980 7	0.923 5	<u>0.995 0</u>	0.974 4	0.986 4	<u>0.953 4</u>	<u>0.995 5</u>	0.957 3
徐欣宜等 ^[32]	2023	34.49	28.70	37.47	31.13	44.25	37.21	36.37	30.73	38.89	30.61
		0.976 7	0.914 8	0.979 0	0.920 8	0.994 3	0.972 4	0.984 3	0.950 3	0.991 6	0.943 2
MLFSR ^[14]	2024	35.22	29.28	38.14	31.56	44.90	37.83	36.92	31.24	40.98	32.03
		<u>0.980 1</u>	0.921 8	0.980 3	0.923 5	<u>0.995 0</u>	<u>0.974 5</u>	0.986 5	0.953 1	0.994 9	0.956 7
LF-TGUnet ^[33]	2024	34.67	28.96	37.80	31.42	44.78	37.63	36.46	31.28	39.95	31.36
		0.977 6	0.918 9	0.979 1	0.921 5	0.994 8	0.973 4	0.985 7	0.952 9	0.993 6	0.951 1
LF-GIANet ^[34]	2025	34.56	29.14	38.01	31.40	44.80	37.60	36.45	31.25	40.45	31.30
		0.978 2	0.919 8	0.979 8	0.921 8	0.994 8	0.973 3	0.985 8	0.952 2	0.994 5	0.950 5
本文		35.70	29.47	38.49	31.61	45.31	37.86	37.28 /	31.45	41.45	31.72
		0.980 9	0.924 4	0.981 5	0.924 4	0.995 3	0.974 8	0.987 2	0.955 0	0.995 3	0.955 3

注: 指标最优和次优分别作加粗和下划线处理。

2.5 低空场景下 LFASSR 泛化性能

为评估本文模型在低空场景下应用的泛化性能, 从 UrbanLF-Real^[24] 数据集中选取 12 个典型低空应用的光场图像, 包含建筑群、道路交通及桥梁结构等多种典型城市低空场景, 如图 4 所示, 能有效模拟无人机、城市监控等低空视角下的光场成像条件。

在泛化实验中, 直接采用 2.2 节中已训练的 LFASSR 模型, 未进行任何额外的微调或重新训

练。表 4 列出了全部 12 个低空场景光场测试图像的超分辨率结果。尽管本文模型未在该数据集上进行专门训练, 但仍在 PSNR 和 SSIM 指标上取得优异性能, 并在所有比对方法中获得了最高的平均指标值。图 6 展示了其中 2 个低空场景光场测试图像的超分辨率重建结果, 在红色框选区域中, 对比超分辨率前后的局部细节, 可以明显看出, 对应区域的结构与纹理信息得到了显著恢复。图 7 选择低空光场测试集中的 Image128 场景(即图 6 中的

第 2 个场景), 给出本文方法与其他 3 种模型在该场景下 LFASSR 的指标分布情况。可以看出, 本文方法在 PSNR 和 SSIM 这 2 项指标上不仅取得最高的平均值, 其分布方差也最小, 表明本文方法在整

体重建质量和跨视角稳定性方面均具有明显优势。以上泛化实验表明, 本文方法在复杂城市低空场景中仍能有效恢复光场的空间与角度信息, 说明本文模型具有良好的泛化能力。

表 4 低空场景光场图像的空间 (×2)-角度 (2×2→5×5) 同时超分辨率性能对比

Table 4 Performance comparisons on the joint spatial (×2) and angular (2×2→5×5) super-resolution for low-altitude light field images

场景编号	PSNR/dB, SSIM							
	DistgASR ^[16] +DistgSSR ^[16]	DistgASR ^[16] +EPIT ^[25]	DistgASR ^[16] +MLFSR ^[14]	LF-EASR ^[7] +DistgSSR ^[16]	LF-EASR ^[7] +EPIT ^[25]	LF-EASR ^[7] + MLFSR ^[14]	LFSSR ^[19]	本文模型
Image36	37.09/0.981 7	37.06/0.981 6	<u>37.15/0.981 9</u>	36.85/0.980 4	37.01/0.981 3	37.00/0.980 9	36.82/0.980 4	37.20/0.982 3
Image62	36.34/0.978 5	36.41/ <u>0.979 1</u>	36.64/0.979 3	36.07/0.977 7	<u>36.42/0.978 8</u>	36.34/0.978 2	36.04/0.977 5	36.64/0.979 3
Image63	36.08/0.983 4	36.25/0.984 0	<u>36.30/0.984 1</u>	36.17/0.983 9	36.19/0.983 8	36.13/0.983 6	35.89/0.982 8	36.44/0.984 5
Image81	37.87/0.980 7	<u>37.97/0.981 2</u>	37.92/0.980 9	37.95/ <u>0.981 2</u>	37.87/0.980 8	37.85/0.980 6	37.68/0.980 0	38.04/0.981 4
Image82	36.87/0.978 7	37.01/ <u>0.979 3</u>	<u>37.03/0.979 2</u>	36.68/0.977 7	36.93/0.978 4	36.71/0.977 8	36.50/0.977 1	37.30/0.980 5
Image85	37.56/0.986 1	<u>37.79/0.986 5</u>	37.74/0.986 3	37.41/0.985 7	37.76/0.986 0	37.42/0.985 5	37.10/0.984 7	38.10/0.987 2
Image86	39.56/0.978 9	39.57/0.978 9	39.48/0.978 6	<u>39.59/0.979 0</u>	39.52/0.978 7	39.55/0.978 8	39.46/0.978 5	39.64/0.979 2
Image89	37.26/0.978 9	37.23/0.979 0	<u>37.33/0.979 2</u>	37.20/0.978 7	37.26/0.978 9	37.23/0.978 7	37.03/0.978 2	37.37/0.979 5
Image98	37.21/0.983 7	37.41/0.984 4	37.29/0.983 9	37.06/0.983 1	37.29/0.983 8	37.11/0.983 1	37.07/0.983 2	<u>37.36/0.984 3</u>
Image128	39.52/0.991 0	<u>39.69/0.992 0</u>	<u>39.56/0.992 1</u>	39.51/0.991 9	<u>39.69/0.991 8</u>	39.53/0.991 6	39.14/0.990 1	39.99/0.992 8
Image187	37.23/0.991 4	37.38/ <u>0.992 4</u>	37.21/0.992 0	<u>37.48/0.992 1</u>	37.25/0.991 6	37.10/0.991 4	36.80/0.991 0	37.69/0.992 6
Image264	35.68/0.981 2	<u>36.04/0.982 6</u>	36.19/0.983 0	34.69/0.977 5	36.02/0.981 9	34.62/0.976 3	35.34/0.979 8	<u>36.01/0.982 9</u>

注: 指标最优和次优分别作加粗和下划线处理; 从左到右 8 个模型的平均值分别为 37.36/0.982 9、37.48/0.983 4、37.49/0.983 4、37.22/0.982 4、37.43/0.983 0、37.22/0.982 2、37.07/0.982 0、37.65/0.983 9。

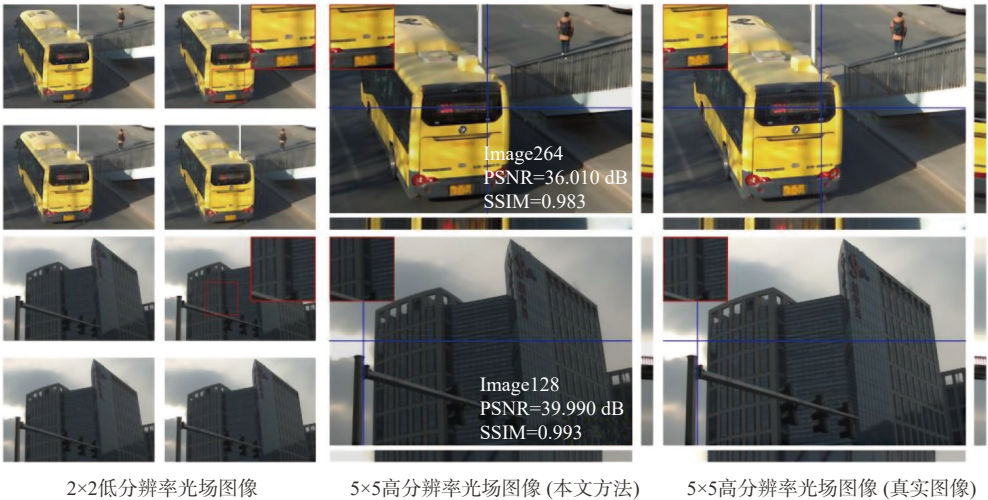


图 6 低空场景光场图像的空间 (×2)-角度 (2×2→5×5) 同时超分辨率重建视觉结果

Fig. 6 Visual results of joint spatial (×2) and angular (2×2→5×5) super-resolution on low-altitude light field images task (the second scene in Fig.6)

2.6 消融实验

为验证本文方法和策略的有效性, 分别在 LFASR、LFSSR、LFASSR 任务上进行消融实验。

表 5 评估了 4D 光场高斯泼溅模块在 LFASR 和 LFSSR 任务中的贡献。在 w/o 4DLFGS 情况下, 采用现有局部隐式图像函数 (local implicit image function, LIIF) 方法^[35] 替换 4D 光场高斯泼溅模块, 各项性能指标都出现不同程度的退化, 尤其在

HCInew 数据集上, PSNR 指标出现明显下降, 表明在 EPI 特征图上进行高斯泼溅建模和渲染能有效提升超分辨率重建质量。

表 6 在 LFASSR 任务下评估了交叉注意力模块、自注意力模块和 4DLFGS 的有效性, 共在 6 个数据集上进行实验。比对结果表明, 无论是真实场景还是合成场景, 去除任意一个模块都会导致模型性能指标下降, 进一步证明本文模块和策略的有效

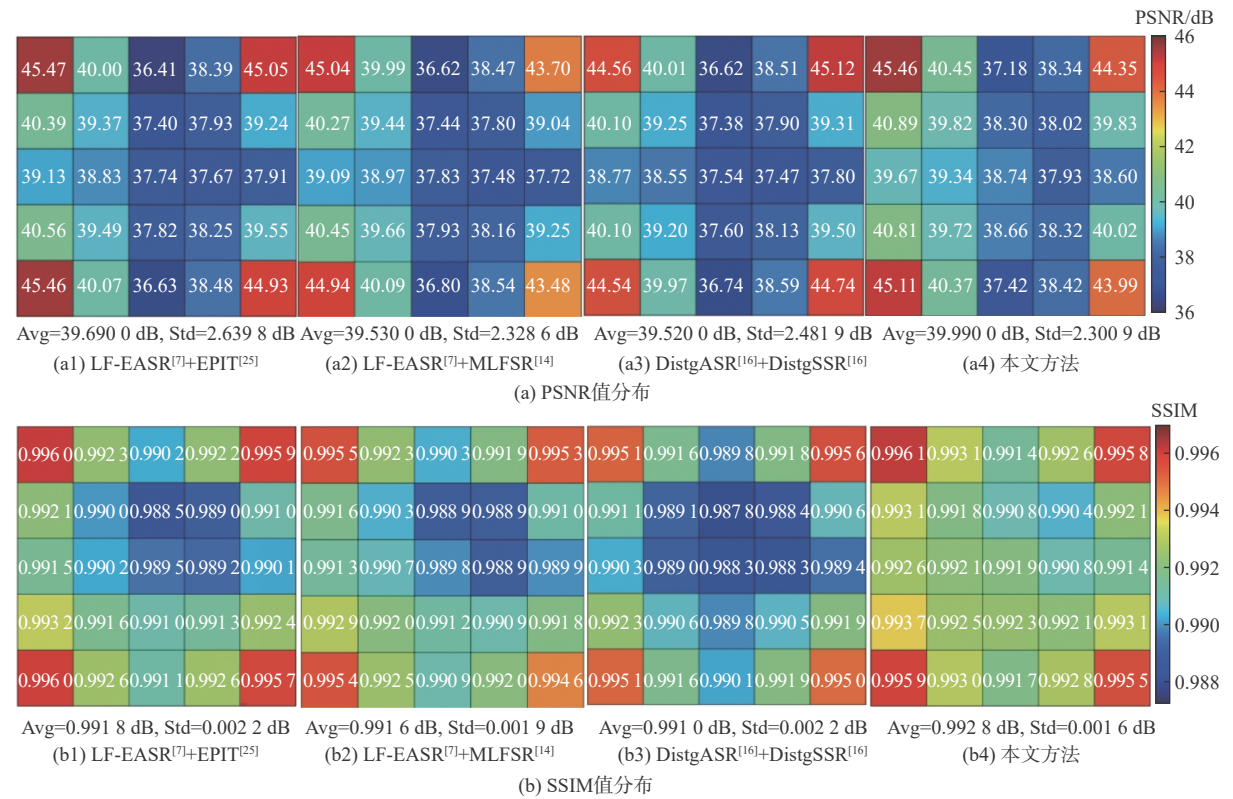


图 7 空间(×2)-角度(2×2→5×5)同时超分辨率任务下 PSNR 和 SSIM 值分布的可视化比对(图 6 中第 2 个场景)
Fig. 7 Visual comparison of PSNR and SSIM distributions under joint spatial (×2) and angular (2×2→5×5) super-resolution

表 5 LFASR 和 LFSSR 消融实验			
Table 5 Ablation study on the LFASR and LFSSR tasks			
超分辨率任务类型	数据集	PSNR (dB)/SSIM	
		w/o 4DLFGS	本文方法
LFASR	HCInew	34.27/0.970 0	37.71/0.982 9
	HCIfold	42.27/0.979 0	43.99/0.993 1
LFSSR	HCInew	38.32/0.981 0	38.49/0.981 5
	HCIfold	45.01/0.995 0	45.31/0.995 3

表 6 LFASSR 任务消融实验										
Table 6 Ablation study on the LFASSR task										
交叉注意力 模块	自注意力 模块	4DLFGS	PSNR/dB, SSIM						参数量	浮点运算速度/ (10 ⁹ ·s ⁻¹)
			HCInew	HCIfold	EPFL	30Scenes	Occlusions	Reflective		
√			34.55/0.960 4	41.88/0.991 2	33.22/0.970 2	41.84/0.991 8	38.28/0.985 3	38.96/0.977 3	1.483×10 ⁷	76.12
	√		34.41/0.959 6	41.70/0.991 2	33.26/0.969 9	41.90/0.991 9	38.34/0.985 5	39.06/0.978 1	1.485×10 ⁷	76.30
√	√		35.12/0.963 9	42.40/0.992 0	33.51/0.971 1	41.96/0.992 0	38.45/0.985 7	39.12/0.978 6	1.638×10 ⁷	85.63
	√	√	35.09/0.963 9	42.42/0.992 0	33.51/0.971 5	42.11/0.992 2	38.53/0.985 8	39.13/0.978 0	1.525×10 ⁷	119.93
√	√	√	35.35/0.965 3	42.81/0.992 5	33.67/0.971 9	42.15/0.992 3	38.64/0.986 2	39.22/0.978 2	1.678×10 ⁷	129.26

性。另外,表中还给出了各种模型的参数量和浮点运算速度,可以看出,本文提出的 4 DLFGS 参数量仅高出 0.4×10⁵。

3 结 论

本文实验结果可得出以下结论:

- 1) 本文方法能够有效实现光场空间与角度同
- 2) 本文方法不仅适用于空间-角度同时超分辨

时超分辨率。在空间分辨率放大 2 倍、角度分辨率由 (2×2) 提升至 (5×5) 的联合超分辨率任务中,本文方法在 30Scenes、Occlusion、Reflective、HCInew、HCIfold 和 EPFL 6 个数据集上的 PSNR 均取得最优结果。表明同时进行空间和角度重建能够获得稳定的性能优势。

率, 在单一角度或空间超分辨率任务中也具有较好的适用性。其中, 在 (2×2->7×7) 角度超分辨率任务的 HCInew 和 HCIold 数据集上, PSNR 分别达到 37.71 dB 和 43.99 dB, 较次优方法提高 1.28 dB 和 0.74 dB; 在空间超分辨率任务中, 除 STFgantry 数据集外, 在 HCInew、HCIold、EPFL 和 INRIA 数据集上均取得最优或具有竞争力的重建性能, 说明该模型对不同光场超分辨率任务具有较好的通用性。

3) 本文方法对未参与训练的低空场景具有较好的泛化能力。在不针对 UrbanLF-Real 数据集进行额外训练或微调的情况下, 12 个建筑、道路、交通和桥梁等典型低空场景上的平均 PSNR 和 SSIM 分别达到 37.65 dB 和 0.983 9, 均优于其他对比方法; 在 Image128 场景中, 本文方法的平均 PSNR 和 SSIM 分别达到 39.99 dB 和 0.992 8, 同时两项指标的跨视角分布方差均为对比方法中最小, 说明其在真实低空场景下具有较好的重建质量和跨视角稳定性。

4) 消融实验表明, 各模块对超分辨率性能均具有实际贡献, 且 4D 光场高斯泼溅对角度信息重建的作用尤为明显。在空间-角度同时超分辨率任务中, 去除交叉注意力、自注意力或 4D 光场高斯泼溅模块均会造成 6 个数据集上的性能下降; 在角度超分辨率任务中, 以 LIIF 替代 4D 光场高斯泼溅后, HCInew 数据集上的 PSNR 由 37.71 dB 降至 34.27 dB, 下降 3.44 dB, 说明基于 EPI 的高斯泼溅建模能够有效增强光场角度域几何信息的恢复能力。

5) 本文方法在极端光照、逆光和低光照条件下仍可能受到成像噪声及细节退化的影响, 对高频纹理和跨视角一致性的恢复能力尚有提升空间; 同时, 现有模型主要针对固定放大倍率设计, 在更高倍率及任意尺度超分辨率任务中的计算效率和重建稳定性仍需进一步验证和优化。后续研究可重点面向复杂真实场景和任意尺度光场重建进一步改进模型。

参考文献 (References)

[1] Ng R, Levoy M, Brédif M, et al. Light field photography with a hand-held plenoptic camera[D]. Palo Alto: Stanford University, 2005: 5-11.

[2] Lyu W Q, Sheng H, Ke W, et al. Advances in light field spatial super-resolution: a comprehensive literature survey[J]. IEEE Access, 2025, 13: 18470-18497.

[3] Hu H, Guo M T, Li H D, et al. Revisiting spatio-angular trade-off in light field cameras and extended applications in super-resolution[J]. IEEE Transactions on Visualization and Computer Graphics, 2021, 27(6): 3019-3033.

[4] Kalantari N K, Wang T C, Ramamoorthi R. Learning-based view

synthesis for light field cameras[J]. ACM Transactions on Graphics, 2016, 35(6): 1-10.

[5] Zhang S, Chang S, Shen Z Q, et al. Micro-lens image stack up-sampling for densely-sampled light field reconstruction[J]. IEEE Transactions on Computational Imaging, 2021, 7: 799-811.

[6] Jin J, Hou J H, Yuan H, et al. Learning light field angular super-resolution via a geometry-aware network[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11141-11148.

[7] Liu G S, Yue H J, Wu J M, et al. Efficient light field angular super-resolution with sub-aperture feature learning and macro-pixel up-sampling[J]. IEEE Transactions on Multimedia, 2023, 25: 6588-6600.

[8] Wang L Y, Ren L F, Wei X Y, et al. Light field angular super-resolution based on intrinsic and geometric information[J]. Knowledge-Based Systems, 2023, 270: 110553.

[9] Mao Y F, Xiao Z Y, An P, et al. Deep sparse-to-dense inbetweening for multi-view light fields[J]. IEEE Transactions on Image Processing, 2025, 34: 6302-6317.

[10] Yuan Y, Cao Z Q, Su L J. Light-field image superresolution using a combined deep CNN based on EPI[J]. IEEE Signal Processing Letters, 2018, 25(9): 1359-1363.

[11] Wang Y L, Liu F, Zhang K B, et al. LFNet: a novel bidirectional recurrent convolutional neural network for light-field image super-resolution[J]. IEEE Transactions on Image Processing, 2018, 27(9): 4274-4286.

[12] Zhang S, Lin Y F, Sheng H. Residual networks for light field image super-resolution[C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11038-11047.

[13] Liang Z Y, Wang Y Q, Wang L G, et al. Light field image super-resolution with transformers[J]. IEEE Signal Processing Letters, 2022, 29: 563-567.

[14] Gao R S, Xiao Z Y, Xiong Z W. Mamba-based light field super-resolution with Efficient subspace scanning[C]// Proceedings of the Computer Vision-ACCV 2024. Berlin: Springer, 2024: 421-437.

[15] Chen S R, Chen L, Wu D F, et al. Enhancing light field image super-resolution through Mamba-based spatial-angular correlation learning[J]. The Visual Computer, 2025, 41(15): 12649-12662.

[16] Wang Y Q, Wang L G, Wu G C, et al. Disentangling light fields for super-resolution and disparity estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 425-443.

[17] Van Duong V, Huu T N, Yim J, et al. Light field image super-resolution network via joint spatial-angular and epipolar information[J]. IEEE Transactions on Computational Imaging, 2023, 9: 350-366.

[18] Yoon Y, Jeon H G, Yoo D, et al. Learning a deep convolutional network for light-field image super-resolution[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop. Piscataway: IEEE Press, 2016: 57-65.

[19] Liu G S, Yue H J, Li K, et al. Adaptive pixel aggregation for joint spatial and angular super-resolution of light field images[J]. Information Fusion, 2024, 104: 102183.

[20] Liu D Y, Li S Z, Mao Y F, et al. Learning implicit and detail-enhanced network for light field image spatial-angular super-resolution[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2026, 36(2): 1544-1557.

[21] Wu T, Yuan Y J, Zhang L X, et al. Recent advances in 3D Gaussian splatting[J]. Computational Visual Media, 2024, 10(4): 613-642.

[22] Hu J T, Xia B, Chen B, et al. GaussianSR: high fidelity 2D Gaussian splatting for arbitrary-scale image super-resolution[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(4): 3554-3562.

[23] Zhang H, Zhou W H, Lin L L, et al. Cascade residual learning based adaptive feature aggregation for light field super-resolution[J]. Pattern Recognition, 2025, 165: 111616.

[24] Sheng H, Cong R X, Yang D, et al. UrbanLF: a comprehensive light field dataset for semantic segmentation of urban scenes[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(11): 7880-7893.

[25] Liang Z Y, Wang Y Q, Wang L G, et al. Learning non-local spatial-angular correlation for light field image super-resolution[C]// Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2023: 12342-12352.

[26] Liu D Y, Mao Y F, Zhou X F, et al. Learning a multilevel cooperative view reconstruction network for light field angular super-resolution[C]// Proceedings of the 2023 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE Press, 2023: 1271-1276.

[27] Li D C, Zhong R. Light field image angular super-resolution using edge features[C]//Proceedings of the 2024 6th International Conference on Robotics and Computer Vision. Piscataway: IEEE Press, 2024: 132-138.

[28] Liu D Y, Mao Y F, Zuo Y F, et al. Light field angular super-resolution network based on convolutional transformer and deep deblurring[J]. IEEE Transactions on Computational Imaging, 2024, 10: 1736-1748.

[29] Wang S Z, Lu Y, Xia W, et al. Light field angular super-resolution by view-specific queries[J]. The Visual Computer, 2025, 41(5): 3565-3580.

[30] Wang S Z, Zhou T F, Lu Y, et al. Detail-preserving transformer for light field image super-resolution[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(3): 2522-2530.

[31] Cong R X, Sheng H, Yang D, et al. Exploiting spatial and angular correlations with deep efficient transformers for light field image super-resolution[J]. IEEE Transactions on Multimedia, 2023, 26: 1421-1435.

[32] 徐欣宜, 邓慧萍, 向森, 等. 基于特征交互融合与注意力的光场图像超分辨率[J]. 激光与光电子学进展, 2023, 60(14): 1410017.
Xu X Y, Deng H P, Xiang S, et al. Light field image super-resolution based on feature interaction fusion and attention mechanism[J]. Laser & Optoelectronics Progress, 2023, 60(14): 1410017(in Chinese).

[33] 黄莉, 吕天琪, 武迎春, 等. 基于双路引导更新的光场图像超分网络[J]. 光电工程, 2024, 51(12): 240222.
Huang L, Lv T Q, Wu Y C, et al. Two-way guided updating network for light field image super-resolution[J]. Opto-Electronic Engineering, 2024, 51(12): 240222(in Chinese).

[34] Zhang W, Li H J, Ke W. LF-GIAnet: cascaded global-view information adaptation-guided network for light field image super-resolution[J]. Multimedia Systems, 2025, 31(2): 73.

[35] Chen Y B, Liu S F, Wang X L. Learning continuous image representation with local implicit image function[C]// Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 8624-8634.

Light field gaussian splatting representation and joint spatial-angular super-resolution for low-altitude scenarios

MENG Jiahua¹, YANG Chi¹, LI Chuanjun², LIN Lili², ZHOU Wenhui^{1,*}

(1. School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China;

2. School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018, China)

Abstract: Unmanned aerial vehicle (UAV) perception and environmental monitoring are two low-altitude commercial applications where light field imaging technology shows significant promise by simultaneously capturing the spatial and angular information of light rays. However, existing light field cameras and super-resolution methods struggle to jointly optimize spatial and angular resolutions, limiting their effectiveness in tasks requiring precise perception and high-fidelity reconstruction. To address this challenge, this paper proposes a light field Gaussian splatting representation and a light field angular spatial super-resolution (LFASSR) method tailored for low-altitude scenarios. Specifically, we extend the conventional 3D Gaussian splatting framework into the 4D light field domain and propose a 4D light field Gaussian splatting representation. Leveraging the geometric structure of light field epipolar plane images (EPIs), we decompose the 4D representation into cascaded 2D Gaussian splatting models along horizontal and vertical directions, enabling more efficient modeling of light field geometry. Based on this representation, we develop a unified light field super-resolution framework. The framework first extracts and fuses features from the spatial, angular, and EPI domains, then constructs directional geometric feature spaces for horizontal and vertical views, where 2D Gaussian splatting is applied separately. Finally, high-quality dense-view light field images with enhanced spatial resolution are rendered via EPI-based synthesis. The suggested methodology greatly outperforms state-of-the-art methods in joint spatial-angular super-resolution, as shown by extensive experimentation on public light field benchmarks and a recently gathered low-altitude light field dataset. It achieves the best performance in terms of peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), while also producing visually superior results with sharp detail recovery and smooth, natural disparity transitions. Moreover, the method exhibits strong generalization capability and competitive performance in standalone light field spatial super-resolution (LFSSR) and light field angular super-resolution (LFASR) tasks.

Keywords: light field; light field super-resolution; light field Gaussian splatting; spatial-angular joint super-resolution; epipolar plane geometry

Received: 2025-11-17; Accepted: 2025-12-12; Published Online: 2026-02-02 11:35

URL: link.cnki.net/urlid/11.2625.V.20260130.1851.004

Foundation items: National Natural Science Foundation of China (62471437); “Pioneer” and “Leading Goose” Research and Development Program of Zhejiang (2023C01030)

* Corresponding author. E-mail: zhouwenhui@hdu.edu.cn

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0141

基于模型数据混合驱动的干扰磁场补偿方法

张德慧¹, 张金生¹, 李婷¹, 马啸宇¹, 廖守亿¹, 倪静², 魏豪^{2,*}

(1. 火箭军工程大学 导弹工程学院, 西安 710025; 2. 飞行器数字敏捷设计全国重点实验室, 成都 610000)

摘

要: 载体自身干扰磁场的精确补偿是提升地磁导航精度的关键挑战。针对传统补偿方法

受限于线性假设和难以适应复杂场景的泛化需求等问题, 提出 2 种基于模型数据混合驱动的磁干扰协同补偿方案。在预处理阶段, 采用主成分分析方法对数据进行降维处理, 并利用物理模型所求补偿参数对神经网络进行初始化处理; 方案一通过分析传统补偿模型的缺陷, 利用神经网络求解补偿参数, 打破传统模型求解范式并预防可能存在的复共线问题, 同时, 在干扰磁场中引入随机因子项模拟非线性较强的随机干扰磁场; 方案二通过构造损失函数联动数据驱动与模型驱动, 将磁场特征的高效提取与干扰场的物理约束深度耦合。实验结果表明: 所提方法较纯模型和纯数据补偿方法具有补偿精度高、泛化能力强等优点, 在不同高度、不同地域进行泛化能力测试的最高补偿精度分别提高了 69.64% 和 69.39%, 验证了所提方法的有效性。

关键词: 地磁导航; 干扰磁场; 航磁补偿; 物理模型约束; BP 神经网络; 混合驱动

中图分类号: V249.32; TP183; P318

文献标志码: A

文章编号: 1001-5965(2026)09-3202-09

卫星拒止条件下的导航技术正朝着多元融合导航、抗干扰与信号增强、量子化与仿生化等方向发展^[1-4]。军事冲突频发和民用领域对导航可靠性的需求更是极大推动了相关技术快速突破与成熟, 随着各类新型导航技术的萌芽和人工智能算法的成熟, 导航系统将实现更高的自主性与环境适应性。地磁导航技术凭借其独立、低成本且隐蔽性好等优势, 展现出重要的应用价值, 逐渐成为一种有效的辅助导航定位方法。

地磁场的无源、无辐射、全天时、全天候等特征使在一些特殊环境下, 地磁导航能够弥补其他导航方式失效带来的无法精准定位导航的问题。地磁场基准图的制备、地磁场数据的实时测量精度、地磁匹配算法很大程度上影响地磁导航的定位效果, 磁场的测量补偿技术是地磁导航的关键技术之一^[5]。在常规的数据采集过程中, 一般用磁力计来搜集地磁特征数据的设备, 但磁力计测得的数据中

不仅包含真实的地磁场数据, 还包含其他磁源引起的磁干扰, 导致地磁场测量精度大幅下降, 导航精度大幅下降^[6]。理论表明, 磁场的强度随距离的三次方衰减^[7], 因此, 高空飞行器所受到的磁干扰关键来源是机载平台及设备, 这些干扰磁场会极大地影响磁场数据的质量, 进而降低地磁导航的定位精度。因此, 在使用地磁数据前必须将这些干扰磁场数据进行剔除。

当前, 载体干扰磁场的软补偿方法主要围绕物理模型驱动与数据驱动 2 大范式展开。

1) 物理模型驱动方法, T-L(Tolles-Lawson)模型及其优化。吕禹萱等提出的基于线性约束与正交约束的递归最小二乘算法进行干扰磁场补偿问题的分析, 有效降低了复共线性带来的影响^[8]。Ye 等提出一种包含 36 个系数的补偿矩阵, 并采用套索正则化牛顿迭代(lasso regularization Newton iteration, LRNM)方法进行求解, 极大改善了复共线

收稿日期: 2025-03-20; 录用日期: 2025-05-27; 网络出版时间: 2025-07-01 09:36

网络出版地址: link.cnki.net/urlid/11.2625.V.20250630.1804.002

基金项目: 国家自然科学基金(62473377); 陕西省自然科学基金基础研究计划(2024JC-YBQN-0657); 飞行器数字敏捷设计全国重点实验室基金(KFJJ-2024-I-03-05)

* 通信作者. E-mail: 772108312@qq.com

引用格式: 张德慧, 张金生, 李婷, 等. 基于模型数据混合驱动的干扰磁场补偿方法[J]. 北京航空航天大学学报, 2026, 52(9): 3202-3210.

Zhang D H, Zhang J S, Li T, et al. An interference magnetic field compensation method based on a model-data hybrid drive[J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52(9): 3202-3210 (in Chinese).

性问题^[9]。Ma 等提出一种基于弹性权重整合的递归最小二乘算法, 通过控制参数更新方向, 有效抑制了灾难性遗忘问题^[10]。T-L 模型作为航磁补偿的基础模型, 具有很强的代表性和物理解释性, 但由于该模型是一个理想化模型, 加上航磁数据之间存在相关性, 复杂环境下表现不佳和可能存在的复共线性始终是影响该模型精度最大的问题。

2) 数据驱动方法, 深度学习技术的探索与应用。张宇等利用反向传播(back propagation, BP)神经网络建立干扰磁场数学模型, 将 T-L 模型中的 18 项补偿系数作为网络输入, 并进行目标探测实验, 补偿改善比超过 20^[11]。Wang 等提出一种数据驱动的机载电子设备(onboard electronic equipment, OBE)干扰补偿方法, 将小波分解和长短期记忆(long short-term memory, LSTM)网络相结合构建预测模型, 通过半真实数据进行验证, 补偿信号的均方根误差小于 0.009 nT^[12]。尽管如此, 数据驱动的方法仍存在数据依赖性强、鲁棒性不足、黑箱特性制约等核心瓶颈, 无法满足现实需求。

融合范式探索, 物理-数据双驱动补偿架构。王劲楠等提出一种基于生成对抗网络(generative adversarial network, GAN)和 LSTM 网络的改进 T-L 磁补偿模型, 通过融合物理模型和神经网络的优势, 为水下反潜任务提供了准确的磁场^[13]。Xu 等提出了一种干扰模型引导的神经网络模型用于航空磁补偿, 能够进行非线性补偿, 且对数据的数量和质量依赖性更低^[14]。该类方法物理机理约束网络架构与数据驱动优化模型参数的双向耦合, 为突破传统方法瓶颈提供了新思路。

考虑到 T-L 模型和数据驱动的神经网络单独进行补偿时会带来一些问题, 可将 2 种补偿方法结合起来, 弥补互相的缺点。本文提出 2 种模型数据混合驱动的载体干扰磁测建模方案, 这 2 种方案通过融合物理模型和数据驱动的神经网络, 将物理信息注入神经网络中, 以提升载体干扰磁场补偿的精度和泛化能力。

本文首先利用岭回归方法求得 T-L 模型的补偿参数向量, 对神经网络进行初始化处理。然后, 利用主成分分析^[15]方法对所选特征进行降维处理, 再进行模型训练。最后, 分别选取训练集部分采样点数据和未知地域的采样点数据进行测试, 并选取均方根误差和决定系数等性能指标评价各方案的补偿效果。

1 理论基础

1.1 载体干扰磁场

载体的机械结构、电控装置、发动机及载体的姿态变化等均会产生局部强磁场^[16], 一般情况下,

利用伸杆技术^[7]很难实现足够远的距离来抵消干扰磁场, 需进行专业补偿。载体干扰磁场的组成复杂多样, 大多磁干扰的变化存在一定的规律, 方便进行补偿, 但还存在小部分干扰变化莫测, 无迹可寻。载体干扰磁场通常包括恒定磁场、感应磁场、涡流磁场和随机磁场等^[17]。

短时间内相对于飞行器机身保持不变的称为恒定磁场, 由机身结构中的硬磁材料磁化引起, 其与飞机的飞行状态无关, 可视为常量。

随飞行器飞行姿态、飞行高度等因素变化的磁场称为感应磁场, 由机身中的软磁材料磁化引起, 其大小与外部激励磁场的大小近似呈线性关系, 与载体材料的磁导率、结构尺寸等多种因素有关, 与外加磁场无关。

随飞行器飞行速度等因素变化的磁场称为涡流磁场, 由穿过载体磁场的磁通量变化率引起, 其大小与载体磁场的变化率、金属体的磁导率、电导率及飞行器运动密切相关。

由载体上的电子设备或其周围环境产生的混乱、随机的干扰磁场称为随机磁场, 该磁场随机性较强, 通常没有规律, 经常通过屏蔽、滤波等方式进行消除。

1.2 T-L 模型

T-L 模型将载体干扰磁场划分为 3 个部分: 恒定磁场、感应磁场和涡流磁场^[18]。T-L 模型利用测量磁场与载体坐标系各坐标轴之间的夹角计算方向余弦, 并利用 3 个方向余弦的一阶项、导数项、交叉耦合项分别对 3 种干扰磁场进行建模, 估算载体干扰磁场的强度^[19]。

如图 1 所示, H_m 为测量磁场, H_e 为地磁场, H_i 为干扰磁场, X 、 Y 、 Z 为测量磁场与载体坐标系 3 个轴的夹角, 可得式(1)所示关系:

$$H_m = H_e + H_i \quad (1)$$

$$|H_e| = \sqrt{|H_m|^2 - 2H_m H_i + |H_i|^2} \quad (2)$$

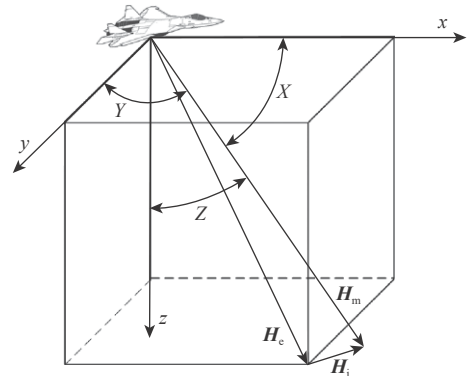


图 1 载体坐标系

Fig. 1 Carrier coordinate system

$$|\mathbf{H}_e| = |\mathbf{H}_m| \sqrt{1 - 2 \frac{\mathbf{H}_m \mathbf{H}_i}{|\mathbf{H}_m|^2} + \frac{|\mathbf{H}_i|^2}{|\mathbf{H}_m|^2}} \quad (3)$$

因为 $|\mathbf{H}_i|^2/|\mathbf{H}_m|^2$ 很小, 可忽略不计, 并根据 $\sqrt{1+x} = 1 + x/2 - x^2/8 + \dots$, 对式(3)进一步简化, 可得

$$\varepsilon \approx \hat{\mathbf{H}}_m \mathbf{H}_i \quad (4)$$

式中: $\hat{\mathbf{H}}_m = \mathbf{H}_m/|\mathbf{H}_m| = [\cos X, \cos Y, \cos Z]$ 为测量磁场的方向余弦; $\varepsilon = |\mathbf{H}_m| - |\mathbf{H}_e|$ 。

T-L 模型中干扰磁场的组成为

$$\mathbf{H}_i = \mathbf{H}_{\text{per}} + \mathbf{H}_{\text{inc}} + \mathbf{H}_{\text{edd}} \quad (5)$$

式中: $\mathbf{H}_{\text{per}}$ 为恒定磁场; $\mathbf{H}_{\text{inc}}$ 为感应磁场; $\mathbf{H}_{\text{edd}}$ 为涡流磁场。

$$\mathbf{H}_{\text{per}} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad (6)$$

$$\mathbf{H}_{\text{inc}} = |\mathbf{H}_m| \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \hat{\mathbf{H}}_m^T \quad (7)$$

$$\mathbf{H}_{\text{edd}} = |\mathbf{H}_m| \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \hat{\mathbf{H}}_m^T \quad (8)$$

将式(5)代入式(4)中, 可得

$$\varepsilon \approx \hat{\mathbf{H}}_m \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + |\mathbf{H}_m| \hat{\mathbf{H}}_m \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \hat{\mathbf{H}}_m^T + |\mathbf{H}_m| \hat{\mathbf{H}}_m \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \hat{\mathbf{H}}_m^T \quad (9)$$

对式(9)进行简化可得

$$\varepsilon = \mathbf{A} \boldsymbol{\theta} \quad (10)$$

$$\mathbf{A} = |\mathbf{H}_m| [\cos X/|\mathbf{H}_m|, \cos Y/|\mathbf{H}_m|, \cos Z/|\mathbf{H}_m|, \cos^2 X, \cos X \cos Y, \cos X \cos Z, \cos^2 Y, \cos Y \cos Z, \cos^2 Z, \cos X \cos' X, \cos X \cos' Y, \cos X \cos' Z, \cos Y \cos' X, \cos Y \cos' Y, \cos Y \cos' Z, \cos Z \cos' X, \cos Z \cos' Y, \cos Z \cos' Z] \quad (11)$$

$$\boldsymbol{\theta} = [a_1, a_2, a_3, (b_{12} + b_{21}), (b_{13} + b_{31}), b_{22}, (b_{23} + b_{32}), b_{33}, c_{11}, c_{12}, c_{13}, c_{21}, c_{22}, c_{23}, c_{31}, c_{32}, c_{33}]^T \quad (12)$$

式中: $\mathbf{A}$ 为信息矩阵, 均为已知信息; $\boldsymbol{\theta}$ 为补偿参数向量, 均为未知数据。 $b_{12} + b_{21}$ 项可简写为 b_{12} , 其余项同理可进行简写, 得到含有 18 个未知数的向量, 如下:

$$\boldsymbol{\theta} = [a_1, a_2, a_3, b_{12}, b_{13}, b_{22}, b_{23}, b_{33}, c_{11}, c_{12}, c_{13}, c_{21}, c_{22}, c_{23}, c_{31}, c_{32}, c_{33}]^T \quad (13)$$

$$\boldsymbol{\theta} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \varepsilon \quad (14)$$

如式(14)所示, 利用 T-L 模型进行补偿时需要求解出补偿参数向量 $\boldsymbol{\theta}$, 进而计算出干扰磁场, 从而达到补偿的目的。但航磁数据的相关性, 可能会使 $\mathbf{A}^T \mathbf{A}$ 退化为奇异矩阵, 导致补偿参数向量无法正常求解或不存在唯一确定的解。传统方法一般利用岭回归来优化该类问题^[20]:

$$J(\boldsymbol{\theta}) = \|\varepsilon - \mathbf{A} \boldsymbol{\theta}\|^2 + \lambda \|\boldsymbol{\theta}\|^2 \quad (15)$$

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^T \varepsilon \quad (16)$$

式中: $J(\boldsymbol{\theta})$ 为岭回归的目标损失函数; $\hat{\boldsymbol{\theta}}$ 表示 $\boldsymbol{\theta}$ 的估计值; λ 为 L2 正则化系数; $\mathbf{I}$ 为单位矩阵。

式(15)和式(16)分别为岭回归方法的目标函数和解析解, 岭回归通过 L2 正则化有效应对过拟合和共线性, 适用于高维及相关性强的数据, 是提升线性模型稳定性的重要工具。

1.3 BP 神经网络

BP 神经网络作为深度学习的基础, 广泛应用于分类、图像识别与自然语言处理等领域。BP 神经网络是一种基于误差 BP 算法的多层前馈神经网络, 属于监督学习模型^[21]。其通过前向传播计算输出, 再通过 BP 调整网络的权重和偏置等参数, 以最小化预测值与真实值之间的误差, 其基本结构如图 2 所示。

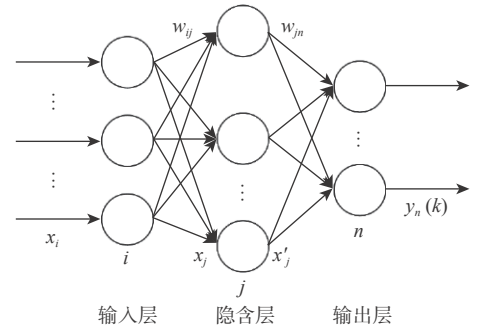


图 2 BP 神经网络

Fig. 2 BP neural network

BP 神经网络主要由输入层、隐含层和输出层组成。输入层用来接收外部输入数据, 隐含层通过非线性激活函数处理输入信号来提取特征, 输出层输出最终预测结果。

BP 神经网络的学习过程由正向传播和 BP 组成, 正向传播用来计算网络的输出:

$$x_j = \sum_i (w_{ij} x_i + b_{ij}) \quad (17)$$

$$x'_j = f(x_j) \quad (18)$$

$$y_n(k) = \sum_j (w_{jn} x'_j + b_{jn}) \quad (19)$$

式中: x_i 为网络的输入; w_{ij} 和 w_{jn} 为各层间的权值; b_{ij} 和 b_{jn} 为偏置项; $f(\cdot)$ 为激活函数; $y_n(k)$ 为网络的输出。

通过网络的输出 $y_n(k)$ 和标签数据 $y(k)$ 计算损失函数 E :

$$e(k) = y(k) - y_n(k) \quad (20)$$

$$E = \frac{1}{2} e(k)^2 \quad (21)$$

BP 利用梯度下降法, 调整各层间的连接权值:

$$w_{ij}(k+1) = w_{ij}(k) + \Delta w_{ij} \quad (22)$$

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = \eta e(k) \frac{\partial y_n}{\partial w_{ij}} \quad (23)$$

$$\frac{\partial y_n}{\partial w_{ij}} = \frac{\partial y_n}{\partial x'_j} \cdot \frac{\partial x'_j}{\partial x_j} \cdot \frac{\partial x_j}{\partial w_{ij}} = w_{jn} \frac{\partial x'_j}{\partial x_j} x_i \quad (24)$$

式中: $w_{ij}(k+1)$ 为更新后的权值; $w_{ij}(k)$ 为更新前的权值; Δw_{ij} 为权值的变化量; η 为学习率。

求解偏导的过程用到链式求导法则, 权值 w_{jn} 的更新原理同上。

通过设置合理的网络层数、各层神经元数量、学习率、批量大小、迭代次数、正则化系数等超参数, 选取合适的激活函数及构建恰当的损失函数即可完成相关数据的处理工作。

2 模型数据混合驱动的干扰磁场建模方法

2.1 方案一

考虑到 T-L 模型是一个理想化的物理模型, 仅考虑恒定磁场、感应磁场和涡流磁场, 未考虑机载设备中产生的没有物理规律的随机磁场等非线性成分, 这就可能导致安装在舱内的磁力仪无法正常使用 T-L 模型补偿测量数据, 故采用 BP 神经网络输出得到一个随机因子项来模拟舱内产生的随机磁场, 将其引入干扰磁场的建模中。此外, 利用神经网络训练获得补偿参数向量, 一方面, 打破 T-L 模型求解补偿参数的壁垒, 提升线性模型在复杂环境下的表达能力, 另一方面, 预防可能存在的复共线问题。

方案一处理数据过程如图 3 所示, H_m 为标量磁力计的磁场测量值, D 为测量磁场的方向余弦, M_s 为飞行器上一些传感器数据, θ_p 为补偿参数向量的预测值, 共包含 18 个参数, r_p 为用于改善 T-L 模型的随机因子项。

网络 1 可得到补偿参数向量, 通过式 (10) 计算出载体坐标系下包括恒定磁场、感应磁场和涡流磁场的干扰磁场; 网络 2 可得到除 T-L 模型中干扰磁场之外的随机干扰因子项。将网络 1 和网络 2 的

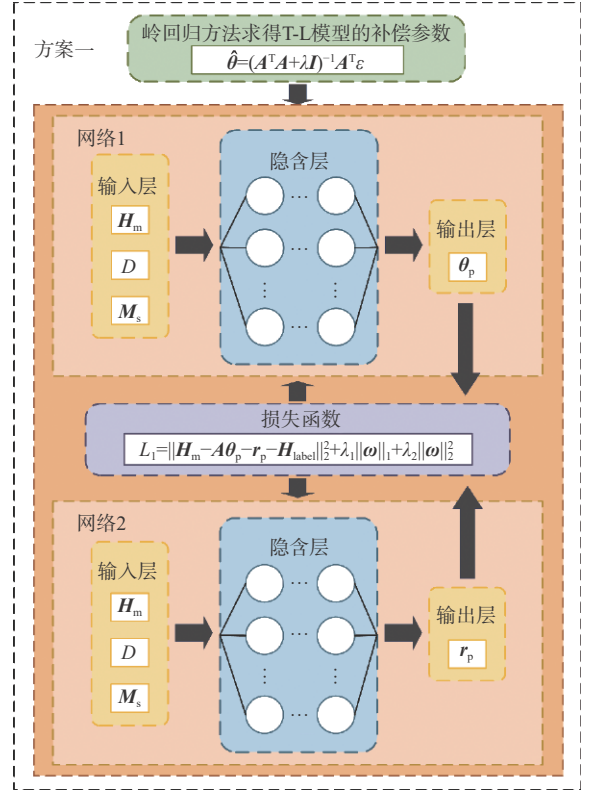


图 3 方案一训练过程

Fig. 3 Training process of scheme one

输出结合得到此方案输出的总干扰磁场, 并构造损失函数, 如下:

$$L_1 = \|H_m - A\theta_p - r_p - H_{\text{label}}\|_2^2 + \lambda_1 \|\omega\|_1 + \lambda_2 \|\omega\|_2^2 \quad (25)$$

式中: H_{label} 为磁场标签值; λ_1 为 L1 正则化系数; λ_2 为 L2 正则化系数; ω 为神经网络的权重参数。

网络 1 的结构为 26: 200: 120: 70: 18, 网络二的结构为 26: 120: 60: 40: 20: 1。

2.2 方案二

考虑到纯数据驱动的神经网络可解释性差且泛化能力不足, 但复杂环境下补偿精度较高; 而基于 T-L 模型的补偿方法在复杂地磁环境下表现力较差, 但其具有一定的物理意义。将两者通过损失函数有机结合起来, 得到一种数模联动的补偿方法。

方案二处理数据的过程如图 4 所示, 网络 1 可得到补偿参数向量, 通过式 (10) 计算出载体坐标系下包括恒定磁场、感应磁场和涡流磁场的干扰磁场。网络 2 可得到干扰磁场的预测值。构造网络 1 和网络 2 的损失函数如下:

$$L_2^1 = \|A\theta_p - H_{\text{ip}}\|_2^2 \quad (26)$$

$$L_2^2 = \|H_m - H_{\text{ip}} - H_{\text{label}}\|_2^2 \quad (27)$$

式中: θ_p 为网络 1 输出的补偿参数向量预测值; H_{ip} 为网络 2 输出的干扰磁场预测值。

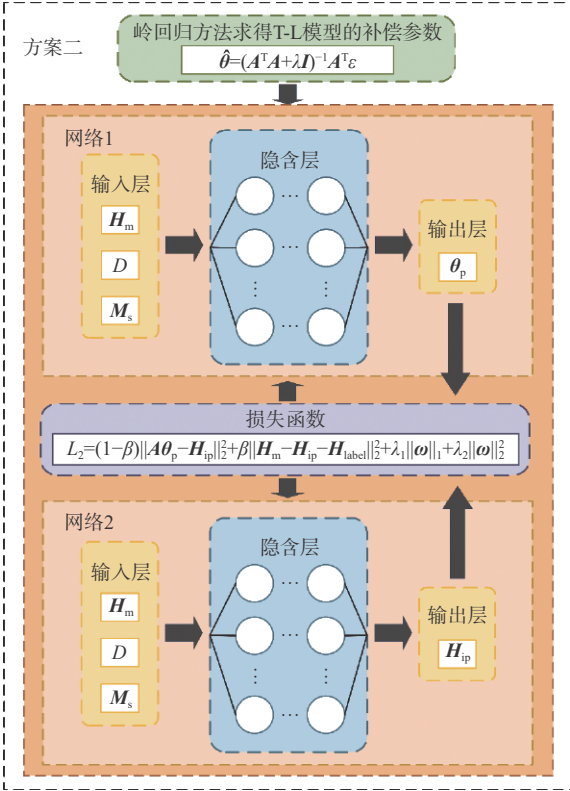


图4 方案二训练过程

Fig. 4 Training process of scheme two

构造方案二的损失函数如下：

$$L_2=(1-\beta)L_2^1+\beta L_2^2+\lambda_1\|\omega\|_1+\lambda_2\|\omega\|_2^2 \tag{28}$$

式中： β 为损失函数 L_2^2 的权重系数，即数据驱动为主

的网络所得结果在总损失函数中的占比，本文 β 设置为0.9。

该方案以测量磁场和网络2所输出干扰磁场之差得到的磁场预测值作为整个网络的输出。网络1的结构为26: 200: 120: 70: 18，网络2的结构为26: 150: 80: 30: 1。

2.3 网络参数

表1为本文2种方案共同的网络参数。在输入特征的选取中，根据地磁信息和部分机舱内常见的设备信号，选取5个时刻的9个特征，共45个特征，通过5个时刻特征数据的训练来拟合最后时刻的输出，训练过程如图5所示。

由于输入特征信息较多，为高效利用特征量，提升模型的训练速度，在特征量的处理过程中不仅用到归一化处理，还采用主成分分析法。选择累计解释方差达到99.99%的主成分数量，经过降维后，特征量的个数由45个降到26个。

图6为几种常见激活函数的对比，本文选取Swish(·)函数，该函数是一个平滑的非单调激活函数，能够建模更复杂的映射关系，且在正向传播中能够产生较大的梯度，有助于缓解梯度消失问题，提高模型的训练效率。

$$\text{Swish}(x)=x\frac{1}{1+e^{-\alpha x}} \tag{29}$$

式中： α 为一可调参数，本文设置为1。

表1 2种方案的共同网络参数设置

Table 1 Common network parameter settings for the two schemes

模型输入	学习率	批量大小	迭代次数	L1正则化系数	L2正则化系数	激活函数	优化器
45个特征	0.000 1	128	300	0.000 01	0.000 001	Swish(·)函数	Adam

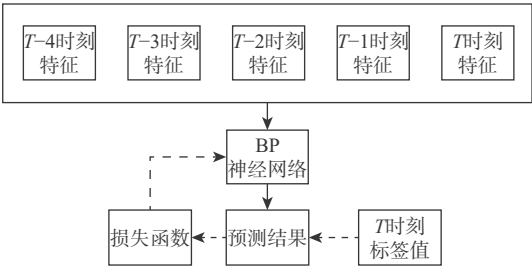


图5 神经网络的训练过程

Fig. 5 The training process of neural networks

在构建损失函数时借鉴弹性网络的特点，损失函数中同时加入了L1和L2正则项。L1正则化对权重的绝对值进行惩罚，倾向于产生稀疏解，适合特征选择；L2正则化对权重的平方进行惩罚，倾向于使权重均匀地趋近于较小值，适合防止过拟合。

本文采用均方根误差(root mean square error, RMSE)和决定系数 R^2 2个性能指标来评判干扰磁

场的补偿效果。RMSE用来评估模型的预测精度，RMSE值越小，补偿效果越好； R^2 用来评估模型的整体拟合优度，是一个无量纲的统计量， R^2 值越接近1，说明模型对数据的解释能力越强。

$$\sigma_{\text{RMSE}}=\sqrt{\frac{1}{n}\sum_{i=1}^n(y_i-\hat{y}_i)^2} \tag{30}$$

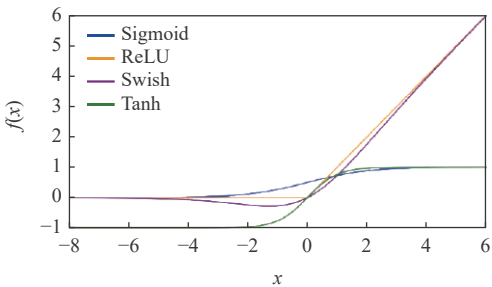


图6 不同激活函数的对比曲线

Fig. 6 Comparative curves of different activation functions

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (31)$$

式中: y_i 为标签值; $\hat{y}_i$ 为预测值; $\bar{y}$ 为标签值的均值; n 为样本数量。

3 实验结果和性能分析

3.1 实验环境及数据集

实验所使用的计算机型号为 Thinkbook15 II -L, CPU 为 Intel(R) Core(TM) i7-1065G7CPU @ 1.30 GHz, 16 GB 内存, 深度学习框架为 PyTorch。

使用的数据集以 USAF-MIT “磁信号增强挑战赛” 开源数据为基础。该数据集于 2020 年公布, 其中包括载体的飞行信息、磁力计数据和多种传感器数据。

本文选取 1 003 号航班中部分航线的数据, 不同航线的飞行区域如表 2 所示, 由于伦弗鲁省 800 m 高度(1 003.10)的数据为该挑战赛预留内测数据, 暂未公开, 因此, 本文主要以 1 003.02、1 003.04、1 003.08 这 3 段航线展开干扰磁场补偿的相关研究。

表 2 不同航线数据解读

航线编号	描述
1 003.02	安大略省东部400 m高度自由飞行
1 003.04	安大略省东部800 m高度自由飞行
1 003.08	伦弗鲁省400 m高度自由飞行

3.2 实验数据及初处理

实验所采用数据的详细描述如表 3 所示, 其中, 测量数据和标签数据均为磁异常数据^[22]。mag_4_iuc 为本文研究的测量磁场 H_m , 是 4 号标量磁力计(安装在飞行器后舱地板的标量磁力计)未经补偿的磁异常数据, 由 4 号标量磁力计未经补偿的总磁场减去主磁场和日变磁场计算得到。mag_1_ic 为标签数据 H_{label} , 是 1 号磁力计(安装在尾针的标量磁力计)经补偿后的磁异常数据, 由 1 号标量磁力计经补偿后的总磁场减去主磁场和日变磁场计算得到。

其余数据均为数据集中的原始数据。flux_b_x、flux_b_y、flux_b_z 用来表示方向余弦 D , cur_strb、cur_ac_hi、cur_heat、vol_bat_1、vol_block 为传感器信号 M_s , 是飞行器中电子设备产生的模拟信号, 用来模拟复杂的机载设备环境, 通过引入这些电压和电流的变化将舱内复杂的干扰源映射到 T-L 模型参数的表达中。

表 3 神经网络的输入、标签信息描述

特征量	描述
mag_4_iuc	4号标量磁力计未经补偿的磁异常数据
flux_b_x	磁通门B的X轴方向的磁场大小
flux_b_y	磁通门B的Y轴方向的磁场大小
flux_b_z	磁通门B的Z轴方向的磁场大小
cur_strb	频闪灯的电流传感器
cur_ac_hi	空调风扇高速的电流传感器
cur_heat	惯性导航系统加热器的电流传感器
vol_bat_1	电池1的电压传感器
vol_block	模块的电压传感器
mag_1_ic	1号标量磁力计经补偿后的磁异常数据

3.3 补偿效果对比

选取 1 003.02 航线的 16 000 个采样点作为训练集, 并选取这 16 000 个采样点中的后 2 000 个采样点作为测试集, 验证本文 2 种方案的补偿效果, 补偿结果如图 7 和表 4 所示。

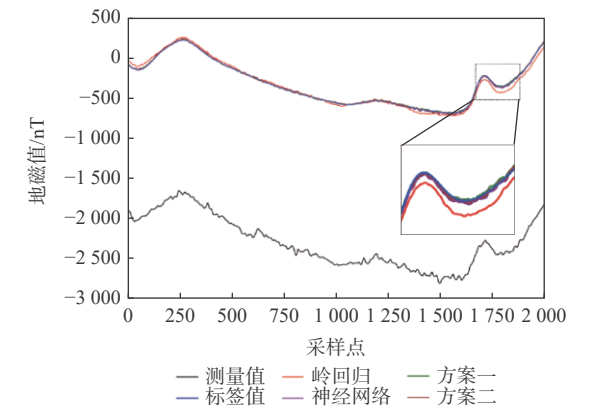


图 7 不同方案的补偿效果

Fig. 7 The compensatory effects of different schemes

表 4 不同方案的补偿能力对比

方案	RMSE/nT	R^2
岭回归	33.944	0.985 83
神经网络	9.301	0.998 82
方案一	6.871	0.999 36
方案二	7.313	0.999 27

通过补偿结果发现, 岭回归方法虽然补偿精度较低, 但其遵循 T-L 模型的物理意义, 将测量磁场按照正确的方向进行补偿; 而采用 BP 神经网络的方法补偿精度较高, 体现了神经网络对这类数据的强大解析能力。此外, 本文 2 种方案较传统补偿方法均对测量磁场有较好的补偿效果, 展现出了深度学习

和 T-L 模型融合在处理此类问题具有一定的优势。

3.4 泛化能力分析

选取 1 003.02 航线的 16 000 个采样点作为训练集,分别选取 1 003.02 航线非训练集的 2 000 个采样点,1 003.04 航线的 2 000 个采样点,1 003.08 航线的 2 000 个采样点作为验证泛化能力的测试集,2 种方案的泛化能力测试如图 8 所示。

由于 2 种方案均采用岭回归所求得补偿参数对网络进行初始化处理,为验证向神经网络注入物理信息对泛化能力的提升起到一定的作用,以纯数据驱动网络的补偿方法进行对照实验,构造该网络结构与方案二中网络 2 结构相同,分别对该网络进行未初始化处理和初始化处理,再进行磁场补偿,补偿效果如表 5 所示。

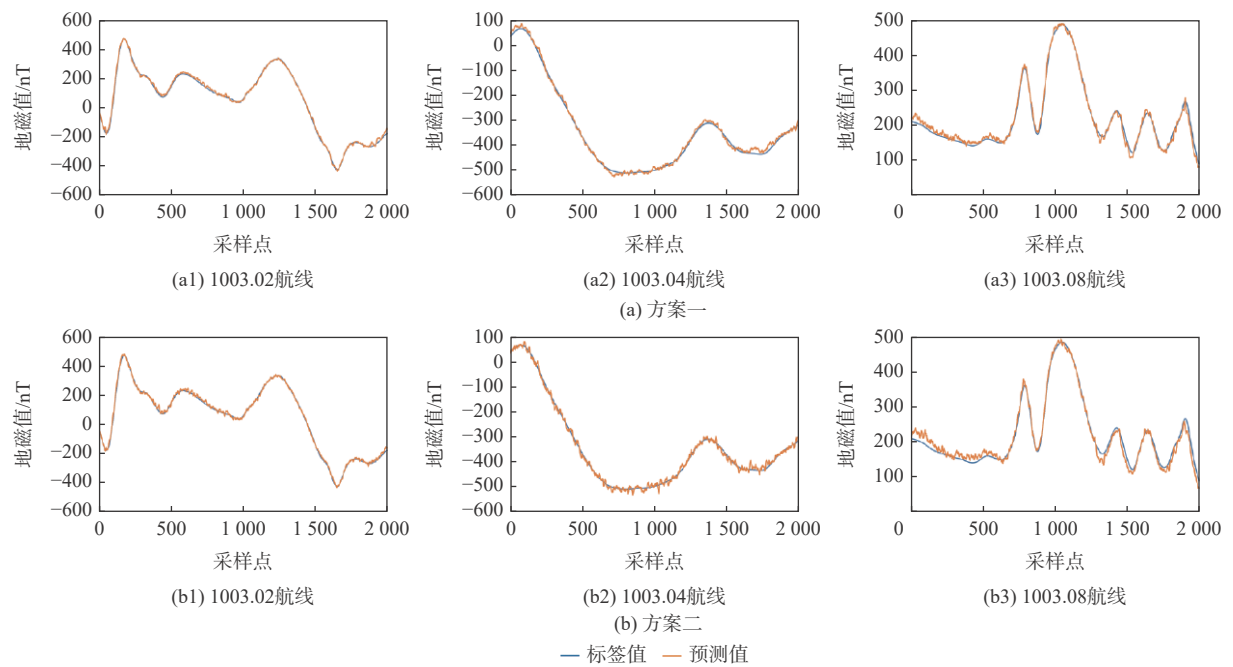


图 8 不同航线泛化能力测试结果
Fig. 8 Test results of generalization ability for different routes

表 5 不同方案的泛化能力对比
Table 5 Comparison of generalization capabilities of different schemes

方案	RMSE/nT			R^2		
	1003.02航线	1003.04航线	1003.08航线	1003.02航线	1003.04航线	1003.08航线
岭回归	86.490	73.191	133.078	0.834 9	0.811 4	-0.836 4
神经网络未初始化	14.561	31.465	31.153	0.995 7	0.963 4	0.895 3
神经网络初始化	10.849	13.302	16.574	0.997 6	0.993 5	0.970 4
方案一	9.796	9.552	9.536	0.998 0	0.996 6	0.990 2
方案二	9.746	9.761	10.304	0.998 3	0.996 5	0.988 5

通过是否对神经网络进行物理信息初始化处理的补偿效果对比,能够得出在神经网络中引入适当的物理信息能在一定程度上增强模型泛化能力的结论,但随着高度和地域的变化,2 种神经网络补偿方法的泛化能力均有所下降。将方案一、方案二与仅依靠物理模型、纯数据驱动网络方法的补偿效果进行对比,2 种方案较仅依靠物理模型方法的补偿精度最高可提升 92.83% 和 92.26%,较纯数据驱动网络方法的补偿精度最高可提升 69.64% 和 68.98%。本文 2 种混合驱动的干扰磁场补偿方案不论是在不同高度,还是在不同地域的磁补偿上均

表现出较强的泛化能力。

4 结 论

本文设计了 2 种模型数据混合驱动的干扰磁场补偿方案,不仅将物理信息注入神经网络中,还将传统 T-L 补偿模型有机结合到神经网络中进行干扰磁场补偿,通过实验分析可以得出以下结论:

1) 方案一通过分析 T-L 模型的建模机理和模型缺陷,利用神经网络非线性强的特性引入随机因子项改善传统物理模型的弊端,并利用神经网络求解 T-L 模型的补偿参数向量,打破传统模型求解的

范式,使其较传统方法的补偿能力有所提升。

2) 方案二通过将数据驱动和模型驱动 2 种方法进行有机结合,设计损失函数,构造一种数模联动的补偿方法,补偿效果优于纯数据驱动和纯模型驱动的补偿效果。

3) 2 种方案分别采用不同的数模结合方式进行干扰磁场的补偿,通过实验结果对比,2 种方案在 RMSE 和 R^2 等性能指标上均优于传统补偿方法,体现了本文方法在干扰磁场补偿泛化能力中的优势。从整体的补偿效果来看,方案一的补偿效果稍优于方案二。对比 800 m 高度的补偿结果,2 种方案的评价指标相差不多,但方案一的预测曲线较为平滑且波动较小,而方案二的预测曲线波动频繁且幅度较大。

深度学习近些年被广泛应用于各行各业,但其在实践中最大的不足是可解释性不强,泛化能力弱等;物理模型具有很强的物理意义,但有时灵活性不高,遇到复杂环境很难使用。在未来的研究中,将继续探究传统补偿方法与深度学习补偿方法的关系,通过有效结合 2 种方式,进一步提升干扰磁场补偿的泛化能力,并将本文方法拓展到干扰磁场的矢量补偿中,应用到实地的航磁补偿中。

参考文献 (References)

- [1] 何康辉,董朝阳,王青.一种考虑 GPS 信号中断的导航滤波算法[J].北京航空航天大学学报,2019,45(9): 1874-1881.
He K H, Dong C Y, Wang Q. A navigation filtering algorithm considering GPS signal interruption[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(9): 1874-1881(in Chinese).
- [2] 陈飞强,孙一凡,唐小妹,等.GNSS 信号功率增强对调零抗干扰接收机的性能影响[J].国防科技大学学报,2023,45(5): 72-77.
Chen F Q, Sun Y F, Tang X M, et al. Effect of signal power enhancement on the performance of GNSS null-steering anti-jamming receiver[J]. Journal of National University of Defense Technology, 2023, 45(5): 72-77(in Chinese).
- [3] 吴德伟,苗强,何思璇,等.量子传感的导航应用研究现状与展望[J].空军工程大学学报(自然科学版),2021,22(6): 67-76.
Wu D W, Miao Q, He S X, et al. A study of existing status and prospects for quantum sensor in navigation[J]. Journal of Air Force Engineering University (Natural Science Edition), 2021, 22(6): 67-76(in Chinese).
- [4] 朱祥维,沈丹,肖凯,等.类脑导航的机理、算法、实现与展望[J].航空学报,2023,44(19): 21-33.
Zhu X W, Shen D, Xiao K, et al. Mechanisms, algorithms, implementation and perspectives of brain-inspired navigation[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(19): 1-33(in Chinese).
- [5] 李婷,王仕成,张金生.磁场测量误差补偿技术综述[J].电光与控制,2018,25(6): 66-71.
Li T, Wang S C, Zhang J S. A review on geomagnetic measurement error compensation technologies[J]. Electronics Optics & Control, 2018, 25(6): 66-71(in Chinese).
- [6] 缪林良,米洒洒,王玮琳,等.航磁矢量测量的误差分析和补偿算法研究[J].电子测量与仪器学报,2021,35(12): 15-23.
Miao L L, Mi S S, Wang W L, et al. Error analysis and compensation algorithm research of aeromagnetic vector measurement[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(12): 15-23(in Chinese).
- [7] 周斌,王劲东.卫星磁部件分布对梯度法消除剩磁的影响分析[J].中国空间科学技术,2013,33(5): 29-34.
Zhou B, Wang J D. Influence of magnetic component distribution of satellite on eliminating remanant magnetic field by gradient method[J]. Chinese Space Science and Technology, 2013, 33(5): 29-34(in Chinese).
- [8] 吕禹萱,雷华明.磁异常探测中载体干扰磁场 RLS 补偿方法[J].电子测量与仪器学报,2022,36(3): 96-104.
Lyu Y X, Lei H M. RLS method for carrier interference compensation in magnetic anomaly detection[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(3): 96-104(in Chinese).
- [9] Ye L W, Yu Z T, Zhang Y X, et al. An aeromagnetic compensation strategy for large UAVs[J]. Sensors, 2024, 24(12): 3775.
- [10] Ma X Y, Zhang J S, Liao S Y, et al. Aeromagnetic compensation method based on recursive least square and elastic weight consolidation[J]. Applied Geophysics, 2025, 22(2): 279-290.
- [11] 张宇,陈正想,覃涛.基于 BP 神经网络的航磁补偿方法研究[J].数字海洋与水下攻防,2023,6(6): 734-742.
Zhang Y, Chen Z X, Qin T. Research on aeromagnetic compensation method based on BP neural network[J]. Digital Ocean & Underwater Warfare, 2023, 6(6): 734-742(in Chinese).
- [12] Wang Y Z, Han Q, Zhan D C, et al. A data-driven OBE magnetic interference compensation method[J]. Sensors, 2022, 22(20): 7732.
- [13] 王劲楠,张晓明,张鸽,等.基于 GAN-LSTM 改进 T-L 磁补偿方法[J].电子测量技术,2025,48(5): 50-56.
Wang J N, Zhang X M, Zhang G, et al. Improved T-L magnetic compensation method based on GAN-LSTM[J]. Electronic Measurement Technology, 2025, 48(5): 50-56(in Chinese).
- [14] Xu Y J, Liu Z Y, Zhang Q, et al. Interference model guided neural network for aeromagnetic compensation[J]. IEEE Sensors Journal, 2024, 24(8): 12266-12275.
- [15] 王晨阳.PCA 和 GA-BP 结合的地磁导航适配区选择方法[J].电光与控制,2018,25(6): 110-114.
Wang C Y. Geomagnetic navigation matching area selection based on PCA and GA-BP neural network[J]. Electronics Optics & Control, 2018, 25(6): 110-114(in Chinese).
- [16] 林沂,孙晶京,闫旭.地磁导航定位技术原理与方法综述[J].全球定位系统,2023,48(6): 32-41.
Lin Y, Sun J J, Yan X. A review of the principles and methods of geomagnetic navigation and positioning technology[J]. GNSS World of China, 2023, 48(6): 32-41(in Chinese).
- [17] 李季.地磁测量中载体干扰磁场特性及补偿方法研究[D].长沙:国防科学技术大学,2013: 99-100.
Li J. Study on the characteristics and compensation method of vehicle interferential magnetic field in geomagnetic measurement[D]. Changsha: National University of Defense Technology, 2013: 99-100(in Chinese).
- [18] Tolles W E, Lawson J D. Magnetic compensation of mad equipped

aircraft, 201-1[R]. Mineola: Airborne Instruments Laboratory, 1950; 201-211.

[19] Leliak P. Identification and evaluation of magnetic-field sources of magnetic airborne detector equipped aircraft[J]. IRE Transactions on Aerospace and Navigational Electronics, 1961, ANE-8(3): 95-105.

[20] Su Z N, Jiao J, Zhou S, et al. Aeromagnetic compensation method based on ridge regression algorithm[J]. Global Geology, 2022, 25(1): 41-48.

[21] 戚德虎, 康继昌. BP 神经网络的设计[J]. 计算机工程与设计, 1998, 19(2): 48-50.

Qi D H, Kang J C. On design of the BP neural network[J]. Computer Engineering and Design, 1998, 19(2): 48-50(in Chinese).

[22] 徐文耀, 白春华, 康国发. 地壳磁异常的全球模型[J]. 地球物理学进展, 2008, 23(3): 641-651.

Xu W Y, Bai C H, Kang G F. Global models of earth's crust magnetic anomalies[J]. Progress in Geophysics, 2008, 23(3): 641-651(in Chinese).

An interference magnetic field compensation method based on a model-data hybrid drive

ZHANG Dehui¹, ZHANG Jinsheng¹, LI Ting¹, MA Xiaoyu¹, LIAO Shouyi¹, NI Jing², WEI Hao^{2,*}

(1. College of Missile Engineering, Rocket Force University of Engineering, Xi'an 710025, China;
2. The National Key Laboratory of Digital and Agile Aircraft Design, Chengdu 610000, China)

Abstract: The precise compensation of the carrier's self-interference magnetic field is a key challenge in improving the accuracy of geomagnetic navigation. Two model-data hybrid-driven magnetic interference collaborative compensation strategies are presented in response to the drawbacks of conventional compensation techniques, including the linear assumption and the challenge of responding to complex settings. In the preprocessing stage, principal component analysis is used to reduce the dimensionality of the data, and the compensation parameters obtained from the physical model are used to initialize the neural network. In scheme one, the shortcomings of the conventional compensation model are analyzed, the compensation parameters are solved using a neural network, the traditional model-solving paradigm is broken and potential multicollinearity issues are avoided, and a random factor term is introduced in the interference magnetic field to simulate the random interference magnetic field with strong nonlinearity. Scheme two constructs a loss function to link data-driven and model-driven, deeply coupling the efficient extraction of magnetic field features with the physical constraints of the interference field. Experimental results show that the proposed methods have the advantages of high compensation accuracy and strong generalization ability compared with pure model and pure data compensation methods. The highest compensation accuracy in generalization ability tests at different heights and different regions has increased by 69.64% and 69.39%, respectively, verifying the effectiveness of the proposed methods.

Keywords: geomagnetic navigation; interference magnetic field; aeromagnetic compensation; physical model constraint; BP neural network; hybrid drive

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0282

基于深度强化学习的高速无人机智能路由策略

寿一凡, 刘东升*, 陈亚辉

(浙江工商大学 计算机科学与技术学院, 杭州 310018)

摘 要: 随着无人机技术的迅速发展, 高速无人机集群在低空复杂场景中得到越来越广泛的应用, 但同时也带来了传统路由协议难以满足高动态网络环境的问题。基于此, 提出一种基于深度 Q 网络 (DQN) 深度强化学习模型的高速无人机智能路由策略, 聚焦于高速、高动态场景下的通信保障, 并进一步加强无人机节点的本地化独立决策能力, 同时, 构建用于衡量链路未来稳定性的 Stable 稳定度因子, 使无人机能根据最新状态自适应地进行独立决策。实验结果表明: 所提高速无人机智能路由策略在不同速度对比场景中, 数据包抵达率均保持在 85% 以上, 平均端到端时延、平均每跳时延相较传统路由协议平均降低 23% 以上。在高频率通信场景下, 数据包抵达率较传统路由协议平均提升 15% 以上, 满足高速无人机高动态通信组网对稳定性与通信效率的需求。

关 键 词: 无人机集群; 路由策略; 深度强化学习; 飞行自组网; 高速无人机

中图分类号: V279+.3

文献标志码: A

文章编号: 1001-5965(2026)09-3211-12

无人机凭借其灵活性、低成本、强大的适应性及快速部署能力, 在公共服务、军民领域扮演着重要的角色, 包括灾难响应^[1]和安保巡逻^[2]等领域都得到了广泛应用。与单一无人机相比, 无人机集群可以通过协调合作来完成单一无人机难以实现的复杂任务。无人机集群之间的通信对于协调任务至关重要, 当前无人机集群如何进行高效通信仍是一个主要挑战, 特别是集群通信链路的管理对于集群通信效率至关重要, 因此, 无人机无线通信技术受到学者越来越多的关注^[3-6]。

无线通信路由技术旨在确保网络中业务数据能够快速、稳定、完整地传递到目的地^[7-10]。目前, 常用的传统路由协议包括基于路由表的主动路由协议, 如目的序列距离向量 (destination-sequenced distance vector, DSDV) 路由协议^[11]、优化链路状态路由协议 (optimized link state routing, OLSR)^[12]; 采用源路由机制的动态源路由协议 (dynamic source

routing, DSR)^[13]; 按需响应协议, 如按需距离向量 (ad hoc on-demand distance vector, AODV) 路由协议^[14]、时序排序路由算法 (temporally-ordered routing algorithm, TORA)^[15]; 机会路由协议, 如基于强化学习的路由算法^[16]; 基于位置的路由协议, 如贪婪周界无状态路由协议 (greedy perimeter stateless routing, GPSR)^[17], 此外, Grag 等^[18]还提出一种 Hybrid 混合响应路由协议, 这种协议能在发现路由质量下降前主动切换到备用路径, 减少路由中断和控制开销。但对于无人机集群通信组网而言, 不仅与传统的自组网络具有共同特点, 如去中心化结构、动态网络拓扑、多跳路由和有限资源, 还具有独有的特性, 如节点的能量有限、节点的高机动性及实时传输要求。

传统路由技术在高动态的自组网络环境下无法达到预期通信效果, 通信链路受节点位置快速变化的影响经常断开, 从而导致数据包端到端访问时延大大增加。为应对高动态网络环境, 近年

收稿日期: 2025-05-09; 录用日期: 2025-07-25; 网络出版时间: 2025-07-31 09:01

网络出版地址: link.cnki.net/urlid/11.2625.V.20250730.1659.004

基金项目: 浙江省特支持计划项目 (2022R52048); 浙江省科技计划项目 (2024C01028, 2024C01211)

*通信作者. E-mail: lds1118@zjgsu.edu.cn

引用格式: 寿一凡, 刘东升, 陈亚辉. 基于深度强化学习的高速无人机智能路由策略 [J]. 北京航空航天大学学报, 2026, 52 (9): 3211-3222.

Shou Y F, Liu D S, Chen Y H. Intelligent routing strategy for high-speed UAVs via deep reinforcement learning [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (9): 3211-3222 (in Chinese).

也有相关研究提出了使用深度 Q 网络^[19](deep Q-network, DQN)算法来逼近状态-动作值, 通过结合深度学习与强化学习的方法, 满足对大规模无人机集群在动态环境中进行决策的需求。Lin 等^[20]提出一种基于深度强化学习的智能路由策略, 旨在解决集群自组网中由于快速拓扑变化和高移动性带来的路由问题。通过考虑节点的位置信息、移动速度、负载程度和链路质量, 该策略能够自适应选择邻近节点。Qiu 等^[21]提出一种基于深度强化学习的自适应链路维护方法, 旨在应对集群自组网中由于节点高度移动性导致的环境变化, 通过动态调整广播 Hello 包的时间间隔, 优化广播周期以最大化节点能量, 并改进算法收敛速度。Liu 等^[22]在使用深度强化学习帮助无人机集群进行通信链路决策的基础上, 同时考虑了对各无人机航迹的预测, 将各无人机航迹预测信息整合到通信链路决策中, 在多方面网络性能的实验结果上均优于传统路由协议。Jung 等则基于强化学习提出了一种利用强化学习结合地理位置信息的 Q 学习地理路由协议 QGeo(Q-learning-based geographic routing protocol)^[23], 通过强化学习对传统 GPSR 进行优化, 能够更智能地选择下一跳节点, 改善路由空洞问题和延迟性能。Qiu 等^[24]提出一种基于多智能体强化学习的路由算法, 将信道质量、无人机移动性、无人机开销及相邻要素变化纳入链路质量评估, 并引入长短时记忆网络(long short-term memory, LSTM)优化深度强化学习网络, 提升传输速率, 但收敛速度减缓, 训练稳定性下降。Bai 等^[25]提出一种结合深度强化学习算法近端策略优化(proximal policy optimization, PPO)算法的路由策略来最小化跳数, 降低复杂网络中数据包转发过程中遇到“路由漏洞”的概率, 更多地使用贪婪模式来保持路由策略的效率。Wang 等^[26]提出一种基于决斗双重深度 Q 网络(dueling double deep Q-network, D3QN)的路由方法, 将飞行器、地面站与地面用户构建为空陆混合网络, 围绕联合路由与容量优化展开研究, 有效降低平均端到端传输延迟。Ke 等^[27]提出一种基于多智能体强化学习的新型分布式路由优化算法, 同时, 结合多协议协作机制的自适应多模式路由框架集成, 可令各个无人机都能够仅根据本地网络信息直接调整当前节点的协议和协议参数, 从而增强网络结构和整体性能。

如今随着无人机技术愈发成熟, 为应对不同场景中的特殊任务与快速响应需求, 高速无人机也逐渐开始在一些特殊场景与任务中得到应用, 与常见旋翼无人机不同, 该类型无人机速度最高往往可达

到 460 km/h, 因此, 常被用于紧急搜救与侦察任务中。但在低空领域中往往存在许多复杂场景, 这也要求高速无人机需要时刻保持高机动性, 避障随时可能出现的大机动行为, 这也加剧了通信网络环境的不稳定性和动态性。

但传统路由技术在高动态无人机自组网环境中难以取得理想的通信效果, 节点位置快速变化导致链路频繁断裂, 从而显著增加端到端时延。近年来, 针对高动态网络环境, 不少研究尝试利用深度强化学习提升无人机网络的路由决策效率。然而, 已有深度强化学习路由策略多面向低速无人机, 缺乏对超高速无人机和复杂场景的专门研究。

针对该类高速无人机集群执行任务时无法依赖集中式统一决策的复杂场景, 本文基于深度强化学习提出一种更侧重于高速和高动态环境下适用于高速无人机集群的智能路由策略, 通过各无人机对所处环境的独立感知完成分散化决策, 进一步加强无人机节点的本地化独立决策能力, 无需依赖集中式控制。具体而言, 通过将每架无人机节点建模为一个独立智能体, 每个智能体均基于自身感知的信息独立完成路由决策, 从而更好地适应高动态自组织网络的需求。在仿真实验中, 将本文方法与主流传统路由协议及其他深度强化学习路由策略进行对比, 结果显示, 本文方法在高速、复杂场景中具有显著的优势。

1 集群通信网络问题

本文所考虑的无人机集群通信网络可以抽象为一个无向图 $G(V, E)$, 其中, V 表示无人机节点集合 E 表示节点间的边集合。图 1 为集群通信网络的建模示意图, 图中的边 (u_i, u_j) 表示节点 u_i 和 u_j 之间存在直接通信链路。链路的不同颜色表示通信通道的拥塞程度和链路的稳定性。本文假设每个无

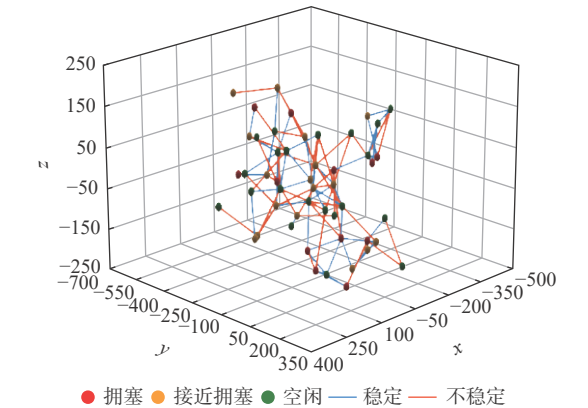


图 1 无人机集群通信自组网建模

Fig. 1 Modeling of self-organizing network communication for UAV swarms

人机都配备了定位系统,使每个无人机 u_i 可以获取其自身的空间位置坐标。此外,各节点还可通过邻居信息交换和数据包携带的历史信息间接获取其他节点位置,但因存在时延,无法保证实时准确。

节点 u_i 的状态信息由其位置、速度等属性描述构成。网络被划分为多个带宽相等的信道,每个信道的最大容量记为 $C_{\max}$,可由多个节点共享信道容量。每个信道可通过在每个时间间隔 t 内为多个活跃节点链路分配时隙,实现资源的共享。在后续内容中为便于建模与理解,可将每条活跃链路在时间间隔 t 内所分配的时隙等效为其在时间间隔 t 内被分配到的可用带宽,取值范围为最小可用带宽至公平分配可用带宽。

无人机的有效通信范围设定为 R_{limit} 。节点 u_i 和 u_j 之间的距离 $d_{i,j}$ 定义为节点 u_i 和 u_j 之间的欧氏距离。当 $d_{i,j} < R_{\text{limit}}$ 时,节点 u_i 和 u_j 之间可建立通信链路,记作 $E_{i,j}$ 。需要指出的是,该方式属于理想化建模,在本文中用于简化网络拓扑的构建和仿真过程,实际通信链路的建立还会受到信道质量、干扰等多种物理因素影响。

节点间的信息交换并非全局实时同步,而是通过周期性 Hello 数据包机制仅在设定的时间间隔内与通信范围内的邻居节点进行信息共享,数据包的处理在一个时间间隔 t 内完成,发送则按照预设周期定期进行。此外,所有数据包在转发过程中还会携带所经过节点的部分状态信息,包括途经节点的空间位置、速度、加速度、航向、数据包处理队列长度及更新时间戳记录等状态信息。各节点在每个时间间隔 t 内进行数据包处理时,会结合数据包中携带的历史信息及时效性,依据时间戳等判据选择性地更新本地节点状态信息表。此外,各节点还会根据已知的其余节点相关位置状态信息,并行计算其余节点此时的预估位置,从而动态完善对整体网络环境的感知。因此,节点对其他节点状态的认知受到信息交换周期、传播延迟及网络拓扑变化的限制,存在局部性和时延,不能保证全局实时准确。这也可能导致部分情况下路由选择出现非最优现象。

2 高速无人机集群智能路由策略

2.1 深度强化学习原理

深度强化学习优化过程是一个使智能体在与环境交互过程中获得最大的累积奖励过程,由于回报函数仅与观察的状态及当前动作有关,因此,对各个智能体与环境的交互可以建模为马尔可夫决策过程(Markov decision process, MDP)。

MDP 可定义为由 4 个元素组成的元组 $[S, A, R, P(s'|s, a)]$ 。其中, S 为状态,即环境观测状态的集合, A 为动作,即智能体当前可采用的动作集合, R 为奖励,即智能体在采取动作后与环境进行交互后获得的奖励。 $P(s'|s, a) = P(s_{t+1} = s', a_t = a)$ 为环境状态转移函数,也表示智能体与环境的交互过程,即在交互的每一轮时间 t ,智能体通过观察环境状态 $s_t \in S$,根据动作选择策略 $\pi^* = p(a|s)$,选择下一步采取动作 $a_t \in A$,执行 a_t 后,环境状态以 $P(s_{t+1}|s_t, a_t)$ 的概率转移到新的环境状态 s_{t+1} ,并根据相关奖励函数得到该轮时间 t 中,智能体采取动作与环境进行交互的即时奖励值 r_t 。而整个决策过程的最终目标即为通过不断学习得到一个最优策略 $\pi^* = p(a|s)$,并通过该策略最大化期望折扣奖励 R_t :

$$R_t = \sum_{t'=t}^T \gamma^{t'-t} r_{t'} \quad (1)$$

式中: γ 为折扣因子。

为处理复杂的高维场景,研究人员提出了 DQN^[28] 算法,将卷积神经网络(convolutional neural network, CNN)和经验回放机制引入 Q 学习算法,以解决高维和连续动态状态空间中的“维度灾难”问题。与 Q 学习算法相比, DQN 使用 CNN 作为函数逼近器,并引入经验回放以提高训练效率。其采用独立的目标网络生成目标 Q 值,从而增强算法的稳定性,并将主网络估计的 Q 值与目标网络的 Q 值之间的差异作为损失函数 $L(\theta)$:

$$y_t^{\text{DQN}} = r_{t+1} + \gamma \max_a Q(s_{t+1}, a | \theta^-) \quad (2)$$

$$L(\theta) = E \left[(y_t^{\text{DQN}} - Q(s_t, a_t | \theta_t))^2 \right] \quad (3)$$

式中: E 表示对该组样本求数学期望; y_t^{DQN} 为 DQN 的目标 Q 值; θ 为 DQN 主网络权重; θ^- 为 DQN 目标网络权重。

该算法通过梯度下降实时更新主网络的权重 θ ,目标网络的权重则通过定期将主网络的权重复制目标网络来更新:

$$\theta \leftarrow \theta - \alpha (y^{\text{DQN}} - Q(s, a | \theta) \cdot \nabla_{\theta} Q(s, a | \theta)) \quad (4)$$

式中: α 为学习率; $\nabla_{\theta} Q(s, a | \theta)$ 为梯度。

尽管 DQN 在许多强化学习任务中取得了显著成果,但在某些环境中可能出现 Q 值过度估计的问题。研究人员提出了双重 DQN 算法(double DQN, DDQN)^[29], DDQN 算法的损失函数与 DQN 算法一致,目标 Q 值计算为

$$y_t^{\text{DDQN}} = r_{t+1} + \gamma Q \left(s_{t+1}, \arg \max_a Q(s_{t+1}, a | \theta_t) | \theta^- \right) \quad (5)$$

式中: y_t^{DDQN} 为 DDQN 的目标 Q 值; θ 为 DDQN 主网

络权重,用于选择当前下一步的动作; θ 为 DQN 目标网络权重,用于评估该动作的价值。

在动作选择策略上,深度神经网络采用 ε 贪婪算法用于探索策略集,避免局部最优,探索策略为

$$\pi(s) = \begin{cases} \arg\max_a Q(s, a) & \text{rand} < \varepsilon \\ \text{随机动作} & \text{其他} \end{cases} \quad (6)$$

式中: ε 为探索率,探索率越大,智能体选择随机策略进行探索的概率越高。

2.2 深度强化学习训练环境构建

2.2.1 状态空间

对于当前节点 u_c 的邻居节点集合 $U_{\text{next}}: [u_1, u_2, u_3, \dots, u_i]$, 状态空间将记录各个节点 u_i 在 U_{next} 中的主要路由信息: $S_i: [H_i, D_i, C_i]$ 。

H_i 表示当前场景下,从邻居节点 u_i 到目标节点 u_{des} 的理论最短路径跳数。 H_i 越小,数据包通过的节点数越少,越容易快速到达目标节点。

D_i 表示从邻居节点 u_i 到目标节点 u_{des} 的路径中,通过每个节点的消息队列长度累积计算出的估计最小访问延迟。

C_i 表示邻居节点 u_i 所处信道的剩余信道容量,为最大信道容量 C_{max} 与已占用信道容量之差。剩余信道容量越大,信道的拥塞程度越低。本文中的“剩余信道容量”本质上在仿真中反映了信道当前的剩余带宽资源,已占用信道容量采用当前传输队列长度与最小可用带宽之积的方式计算,以此衡量该节点所处信道的拥塞程度。虽然实际网络的拥塞情况由多种因素共同决定,但采用该指标有助于简化模型并直观反映信道的可用资源。

为考虑当前节点与下一跳节点之间通信链路的时效性, T_i 为链路信息更新的时效性。当前节点获取所有可达下一跳节点的上一次更新时间表 $[t_a, t_b, t_c, \dots, t_n]$, 用于记录该下一跳节点信息最近一次更新的时间步数。取最小值 $t_{\min}$ 和最大值 $t_{\max}$ 对 $[t_a, t_b, t_c, \dots, t_n]$ 进行归一化处理:

$$T_i = \frac{t_n - t_{\min}}{t_{\max} - t_{\min}} \quad (7)$$

对于每个节点,邻居节点的更新信息时间越久远,其通信链路的时效性越差,状态信息也越不稳定。因此,模型应避免选择时效性差的下一跳节点,其时效性将反映到各节点状态信息中:

$$H_i^T = H_i(1 + \beta T_i) \quad (8)$$

$$D_i^T = D_i(1 + \beta T_i) \quad (9)$$

$$C_i^T = C_i(1 + \beta T_i) \quad (10)$$

式中: β 为折扣因子, $0 \leq \beta \leq 1$, 用于调整通信链路时效

性对状态信息的影响力度; 当 $C_i^T > C_{\text{max}}$ 时, $C_i^T = C_{\text{max}}$ 。

考虑到高速无人机的高机动性,节点的位置变化高度动态,现有通信链路可能随时中断。特别是在实时信息不可得的场景下,规划的路由路径可能因关键中间链路断开而中断,导致数据包必须花费更多时间以通过其他路径抵达。因此,本文为模拟通信网络环境构建了一个链路稳定度因子 Stable , 用于反映链路的稳定性和可靠性,使深度强化学习模型在做决策时,能够合理考虑链路的稳定性。当邻居节点 u_i 被选为下一跳路由时,稳定度因子 u_i^{Stable} 计算公式为

$$u_i^{\text{Stable}} = \left[1 + \delta \sum_{s=1}^n \frac{V_{\Delta\text{pos}} M_{u_s}}{n \Delta R_{\text{limit}, d(u_s, u_{s+1})}} \right]^\rho \quad (11)$$

$$M_{u_s} = \sum_{j=1}^s m_{u_j} \quad (12)$$

式中: n 为当邻居节点 u_i 作为下一跳时,到达目标节点所需经过的节点数; u_s 为最短路径序列中的节点,如 $[u_1, u_2, u_3, \dots, u_s]$, 其中, u_s 为最短路径中的第 s 跳节点; $V_{\Delta\text{pos}}$ 为路径上连续 2 个节点每个时间间隔 t 的相对距离变化速率,若为负数则取 0; M_{u_s} 为到第 s 跳节点的累计待处理数据包消息队列长度, m_{u_j} 为路径上第 j 跳节点的待处理数据包消息队列长度,用于估计数据包到达第 s 跳节点的预估时间步数 t ; $\Delta R_{\text{limit}, d(u_s, u_{s+1})}$ 为有效通信距离 R_{limit} 与当前节点 u_s 和下一跳节点 u_{s+1} 之间距离 $d(u_s, u_{s+1})$ 的差: $\Delta R_{\text{limit}, d(u_s, u_{s+1})} = R_{\text{limit}} - d(u_s, u_{s+1})$; δ 与 ρ 为折扣因子,用于调整稳定度因子的影响力度。

链路稳定度因子旨在提升路由决策对于高动态环境下链路可用性和可靠性的判断能力。针对高速无人机集群网络中节点位置和拓扑结构频繁变化、链路断裂风险较高的问题,单纯依赖当前链路状态易导致路由中断和高时延。

因此,本文提出的链路稳定度因子综合考虑以下几个方面:最短路径跳数和节点间实时距离,反映所选路径的高效性和可达性;节点间的相对速度,预估链路的动态变化趋势和断链风险;沿路径的累计消息队列长度,衡量数据传输的时延和节点负载情况;当前距离与有效通信距离的关系,辅助评估链路是否处于易断开临界状态。

例如,在实际高速无人机集群编队过程中,若 2 个节点当前距离较近但彼此相对速度较大,则这条链路很可能在短时间内由于节点迅速远离而断开。如果路由策略仅依据当前距离判断可达性,容易造成路径刚被选中便立刻失效,影响通信的稳定性。此外,某一路径上的中继节点虽然满足距离和

可达性要求,但其累计消息队列长度很长,意味着数据包在传输过程中将遭遇较大排队延迟和负载压力。

链路稳定度因子的引入使路由算法能够主动规避上述高风险链路,优先选择在短小时内更可能保持连通且传输压力适中的路径,进而降低因链路突变导致的丢包和高时延风险,为节点提供更全面的链路可靠性和稳定性评估,使路由决策不仅关注当前状态,更具备对未来链路变化的前瞻性,有效降低数据包在高动态环境下因链路突发中断造成的丢失和延迟。

此外,Stable因子的计算复杂度分析如下:对于每个候选下一跳节点,算法需在其规划路由路径上的所有中间节点上执行求和操作,时间复杂度为 $O(n)$,其中, n 为路径中的中间节点数量。

在求和过程中, $V_{\Delta\text{pos}}$ 与 $\Delta_{R_{\text{lim}},\text{dist}}(u_i, u_{i+1})$ 均为单值计算,其中,各节点间的相对距离可通过预先计算并缓存以复用,因此,时间复杂度为 $O(1)$ 。但对于 M_{u_i} ,需要对路径上每个中间节点的消息队列长度进行累加,时间复杂度为 $O(n)$ 。因此,计算单个候选下一跳节点的Stable值的总复杂度为 $O(n^2)$ 。为提升计算效率,本文采用动态规划优化,将 M_{u_i} 的中间结果进行存储和增量更新,使每次仅需在已存储值基础上累加当前节点的消息队列长度,从而将单次Stable计算的复杂度降低为 $O(n)$ 。

最终状态空间返回给深度学习模型训练的状态信息为

$$S'_i: [H'_i, D'_i, C_i^T] \quad (13)$$

$$H'_i = u_{i\text{Stable}} H_i^T \quad (14)$$

$$D'_i = u_{i\text{Stable}} D_i^T \quad (15)$$

2.2.2 动作空间

动作空间对应于选择下一跳节点。因此,本文定义动作空间为当前节点 u_c 可选择的下一跳节点集合。动作集为 $A: [u_1, u_2, u_3, \dots, u_i]$ 。在每一轮的动作选择中,当前节点 u_c 可以从集合 U_{next} 中选择任一节点作为动作。

2.2.3 奖励函数

路由策略的目标是通过节点选择确保数据的可靠、稳定和完整传输。在此目标下,如果下一跳是目标节点,智能体应获得最大奖励,即

$$r_i = \begin{cases} 1 & \text{节点}i\text{是目标节点} \\ 0 & \text{节点}i\text{不是目标节点} \end{cases} \quad (16)$$

为引导路由过程的收敛,普通中继节点的选择也应给予奖励,因此,奖励是基于上述4个因素进

行设定的。具体来说,设 u_i 为从动作集 A 中选择的节点作为动作, $D_{i\text{final}}$ 为选择该跳时考虑不确定性的估计最短延迟, $H_{i\text{final}}$ 为考虑不确定性的到目标节点的理论最短路径距离。同时,获得动作集 A 中各节点 D'_n, H'_n 的最小值 $D_{\min}, H_{\min}$ 与最大值 $D_{\max}, H_{\max}$,以此对 D'_i, H'_i 进行归一化处理,得到 $D'_{i\text{norm}}$ 与 $H'_{i\text{norm}}$:

$$D'_{i\text{norm}} = \frac{D'_i - D_{\min}}{D_{\max} - D_{\min}} \quad (17)$$

$$H'_{i\text{norm}} = \frac{H'_i - H_{\min}}{H_{\max} - H_{\min}} \quad (18)$$

$$R = - \left(\mu_1 D'_{i\text{norm}} + \mu_2 H'_{i\text{norm}} + \mu_3 e^{-\frac{C_i^T}{C_{\max}}} \right) r_i \quad (19)$$

式中: μ_1, μ_2, μ_3 为折扣因子,且 $\mu_1 + \mu_2 + \mu_3 = 1$,分别取0.4、0.3和0.3。

为加快模型的收敛速度,对成功到达目标节点的经验样本给予额外的最终奖励:

$$R_{\text{success}} = \begin{cases} \frac{1}{h} & \text{抵达目标} \\ \frac{1.5}{h} & h \leq \sigma n \end{cases} \quad (20)$$

式中: h 为本次所花费的总路由跳数; n 为训练环境中的无人机节点总数; σ 为折扣因子,同时,若最终累计奖励大于0,则取0为最终值。

2.3 模型训练框架

本文高速无人机集群智能路由策略框架如图2所示,以前文所述的无人机模拟通信网络构建深度强化学习环境,模型根据现阶段最优策略进行决策,并将决策结果、决策奖励、环境信息变化作为经验样本加入经验回放池。当完成一轮模拟且经验回放池样本数足够时,根据优先级选取规则选取一批经验样本进行深度学习训练,更新现阶段路由策略,重新开始新一轮模拟与训练,以此循环直至模型最终收敛。

本文采用的深度强化学习模型为Rainbow DQN模型^[30],通过同时集成6个模型的变种以大幅提升模型性能。DDQN^[29]通过双重Q网络解耦bootstrap动作的选择和评估,解决了Q学习的过度估计偏差。优先经验回放^[31]通过为经验样本赋予不同的权重,以更频繁地回放高质量经验样本,得到更多学习的转换,提高了数据训练效率。Dueling DQN^[32]通过分别表示状态值和动作优势,更好地实现动作之间的泛化。多步学习通过分别表示状态值和动作优势,帮助在动作之间进行泛化。即时奖励会通过与环境交互确切得到,因此,训练前期的目标价值可以得到更准确的估计,从而加快训练的速度。分布式DQN^[33]帮助模型学习折扣回报的分类分布,而不是估计均值。噪声网络^[34]通过对参数

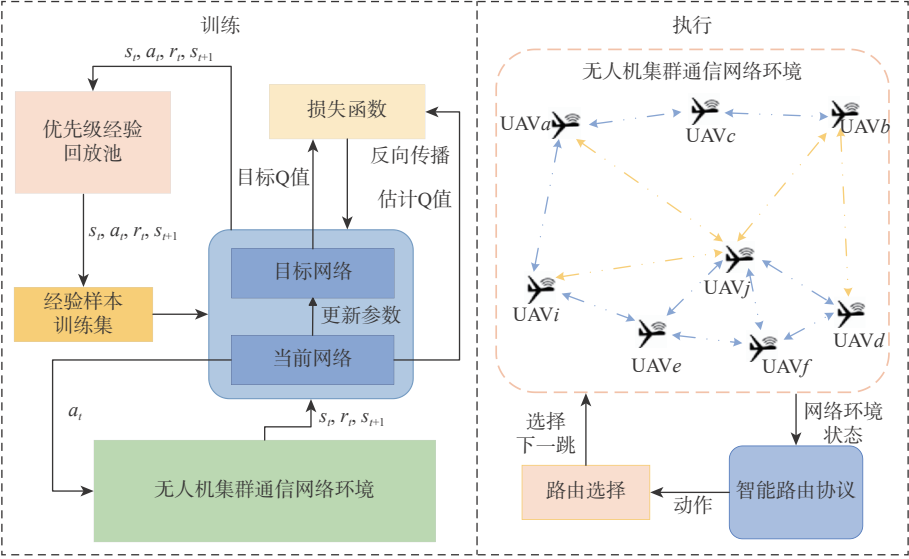


图 2 高速无人机集群智能路由策略框架

Fig. 2 Intelligent routing policy framework for high-speed UAV swarms

增加噪声来增加模型的探索能力。上述各模块在 Rainbow DQN 中的具体集成方式均在文献 [30] 中有详细阐述, 本文不再赘述, 仅对主要特点进行简要说明。

模型的训练过程均在本地服务器上完成, 实际部署时仅需将训练好的网络参数加载到各无人机节点。在运行阶段, 无人机节点只需独立进行模型推理, 无需进行分布式训练与参数同步。同时, 由于训练后得到的神经网络结构较为简洁、参数量较少, 本文模型能够方便地部署于每架无人机节点, 实现本地化的独立推理与决策, 无需依赖外部集中式服务器。

3 仿真验证

本文的训练与仿真环境使用 Python 3.9 编写基于 Networkx 库构建无人机网络环境。深度强化学习模型采用 Pytorch 深度学习框架编写实现, 最终的仿真环境在 Pytorch 中实现。训练和仿真环境使用相同的网络参数设置。

3.1 评价指标

在模型训练阶段, 采用平均奖励和平均每跳延迟作为训练过程的核心评价指标。平均奖励用于衡量深度强化学习模型在不同训练轮次中的策略优化程度, 每 5 000 轮次计算一次平均值, 以观察模型收敛趋势及决策质量的提升情况。平均每跳延迟用于评估模型在数据包转发过程中对网络时延的优化能力, 能够反映强化学习模型在训练过程中对路由选择策略的适应性和效率。

在网络仿真实验阶段, 选取数据包抵达率、数据包平均端到端时延、数据包平均每跳时延作为性

能评价指标。

数据包抵达率为数据包在数据包寿命结束之前成功抵达目标点的概率:

$$L = \frac{P_{\text{success}}}{P_{\text{total}}} \times 100\%$$
 (21)

式中: P_{success} 为成功抵达的数据包数量; P_{total} 为累计发送的数据包数量。

以此衡量网络环境中数据包成功到达目标节点的比例, 反映了路由算法在高动态无人机网络环境下的稳定性和可靠性。

数据包平均端到端时延为数据包由发出到目标点所用延迟的平均值, 用于评估整个网络中的数据包转发效率:

$$Z = \frac{\sum_n D_{\text{total}n}}{P_{\text{total}}}$$
 (22)

式中: $D_{\text{total}n}$ 为各数据包的延迟; P_{total} 为数据包总数; n 为数据包序号。

数据包平均每跳时延表示数据包在成功传输到目标点后, 总延迟除以总计路由条数得到的平均每跳延迟。以此考察路由算法在不同网络负载和拓扑动态性下的延迟特性:

$$B = \frac{\sum_n \frac{D_{\text{total}n}}{J_n}}{P_{\text{total}}}$$
 (23)

式中: J_n 为该数据包路由路径累计跳数。

3.2 实验环境设置

网络环境总体参数设置如表 1 所示, 设置 Hello Package 发送间隔为 150 ms, 最大通信距离设置为 150 m, 场景大小为无人机集群的分布范围。同时,

所有节点各项信息仅能靠固定时间间隔的邻居节点信息传递与数据包中携带的信息进行更新。环境噪声采用高斯噪声, 各无人机节点之间的信号传输采用自由空间衰减模型。若在传输过程中接收信号强度低于指定阈值, 则判定该数据包丢失。共划分 10 个信道, 每个信道带宽为 5 MHz, 每个信道的最大信道容量理想化地设定为 20 Mbit/s。因此,

表 1 无人机集群通信自组网络仿真参数设定

Table 1 Simulation parameters setting for UAV swarm communication self-organizing networks									
集群场景大小/m	无人机数量	最大通信距离/m	无人机最大速度/(m·s ⁻¹)	无人机最小速度/(m·s ⁻¹)	数据包大小/bytes	数据包发送间隔/ms	Hello Package 发送间隔/ms	通信频率/GHz	最小可用带宽/(Mbit·s ⁻¹)
1 500×1 500×500	30~70	150	10~160	5~80	1 024	20~80	150	2.4	1

仿真过程中, 通过设定的最大通信距离作为各节点智能体决策时, 可选邻居节点的筛选条件, 仅当节点间距离小于该阈值时, 才将其视为可建立通信链路的候选节点, 以提升网络拓扑动态维护的效率。对于筛选后的节点对, 再进一步计算信号在自由空间

对于每个信道, 每条活跃节点链路的可用带宽范围为最小可用带宽 1 Mbit/s 到公平分配可用带宽 4 Mbit/s(以 50 个节点为例)。同时, 令每轮仿真中无人机速度最低值为该轮仿真无人机速度最高值的一半, 其中, 在每轮仿真中, 所有无人机节点均作为独立智能体进行建模和决策, 实现完全分布式的路由优化过程。

传播过程中的路径损耗, 并结合噪声模型评估接收端信号强度, 若其低于预设接收阈值, 则判定该链路发生丢包。深度强化学习模型训练网络参数如表 2 所示, 其中, 标注 (DDQN) 仅为训练 DDQN 的深度强化学习模型所使用参数, 其余参数为两者共用。

表 2 深度强化学习模型参数设定

Table 2 Setting parameters for deep reinforcement learning models											
学习率	训练轮数	训练批量	隐藏层维度	γ	τ	经验回放池容量	ϵ (DDQN)	ϵ 的最小值 (DDQN)	N Steps	网络评估频率	网络评估次数
0.000 01	200 000	256	256	0.99	0.005	100 000	0.9	0.05	3	1 000	3

本文分别使用 Rainbow DQN 与 DDQN 深度学习强化模型作为训练模型, 并将是否引入 Stable 因子作为对照条件, 构建不同训练模型进行对比。模型训练网络环境采用上文所述环境, 设定总训练轮数为 200 000 轮, 每轮随机生成全新的无人机自组网网络环境, 数据包起始节点与目标节点皆为随机生成。

3.3 对比模型及路由协议介绍

采用 DDQN 作为对比模型, 同时, 2 种深度强化学习模型均有是否采用分层优先级经验回放池的对比模型及未采用 Stable 因子的模型。选取 DSR 路由协议^[13]、AODV 路由协议^[14]、GPSR 路由协议^[17]、Hybrid 混合响应路由协议^[18]及 QGEO 路由协议^[23]作为与本文智能路由策略的对比路由协议。后续为方便展示实验结果图, 在图注中, RDQN 与 DDQN 分别表示基于 Rainbow DQN 深度学习模型提出的高速无人机智能路由策略及基于 DDQN 深度学习模型提出的高速无人机智能路由策略, 以 N 结尾的表示未使用 Stable 因子的模型。

3.4 模型训练实验结果分析

图 3 为上述深度强化学习模型进行 200 000 轮训练中奖励值随训练步数迭代的趋势变化图, 且每 3 000 轮求一次均值以方便呈现整体趋势, 其中, 以 N 结尾的表示本次训练中环境未采用 Stable 因子进

行训练。图中 RDQN 在前 25 000 轮快速收敛, 并最终于 75 000 轮左右完成收敛, 而 DDQN 则在前 50 000 轮快速收敛, 最终于 125 000 轮左右完成收敛。可见, 本文提出的基于深度强化学习的智能路由策略具有良好的收敛性。

图 4 为深度强化学习模型每轮训练过程中平均每一跳所花费的时延, 每 3 000 轮求一次均值。在训练过程中, RDQN 算法的整体性能优于 DDQN 算法, 而在训练中未采用 Stable 因子的模型性能表现不如采用 Stable 因子的原始模型。

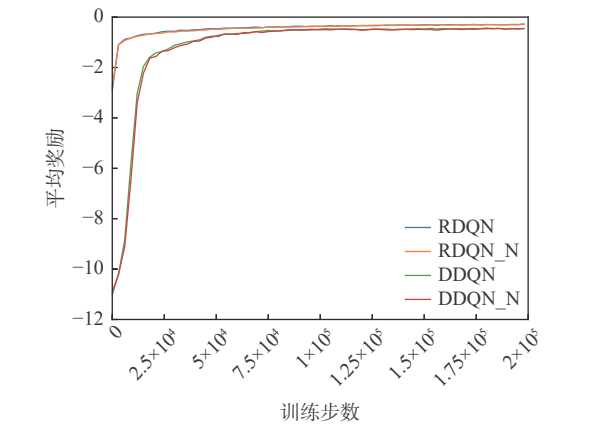


图 3 训练过程平均每步奖励对比

Fig. 3 Comparison of the average rewards per step during the training process

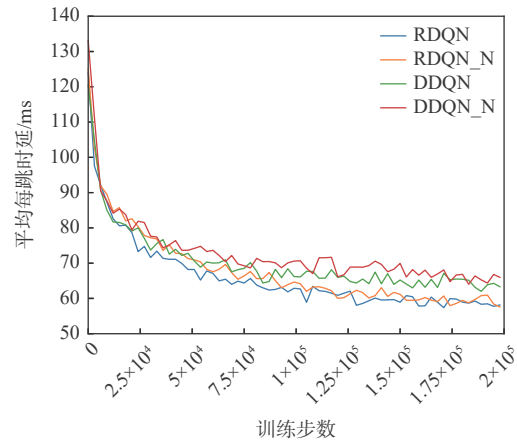


图4 训练过程平均每跳时延对比

Fig. 4 Comparison of average latency per jump during the training process

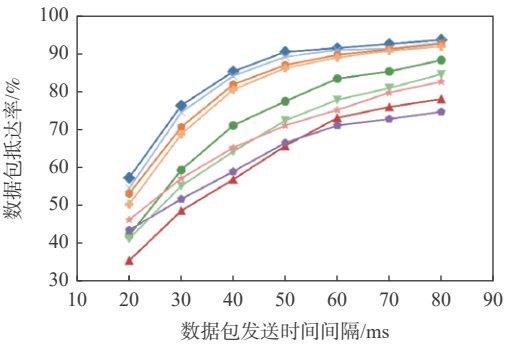
3.5 自组织网络通信仿真实验结果分析

按表1所示网络环境参数建立无人机集群自组网仿真环境, 仿真过程累计投放共计25 000个数据包, 目标节点均为随机生成的当前节点当前时间能抵达的节点。由于考虑到仿真环境生成数据包的随机性, 本文对每组仿真实验皆按照相同的参数设定进行多轮实验, 并取均值进行展示。

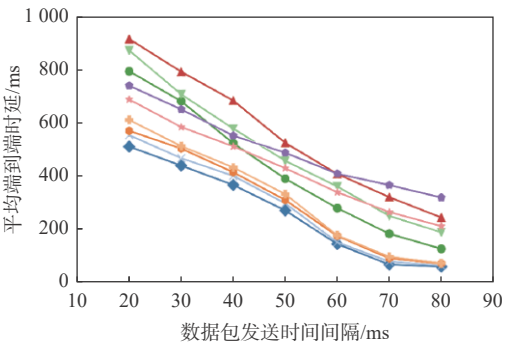
在无人机最高速度为130 m/s的条件下, 进一步对不同数据包发送时间间隔下各路由方法的性能进行对比分析。仿真实验以数据包发送时间间隔作为变量, 分别测试各路由方法在不同通信频率场景下的表现。各组实验结果均取全部数据包测试结果的平均值, 结果如图5所示。

实验结果表明, 传统路由算法在较短的数据包发送间隔下表现较差, 而基于深度强化学习的智能路由策略提高了数据包的传输成功率, 并在高频率传输场景中表现出更好的稳定性。本文构建的Stable因子有助于提高链路稳定性的判断能力, 减少因链路断裂导致的数据丢失和额外时延。随着数据包发送间隔的增加, 传统路由算法的抵达率逐渐恢复至70%~90%, 其中, 以Hybrid混合响应路由协议表现最佳。

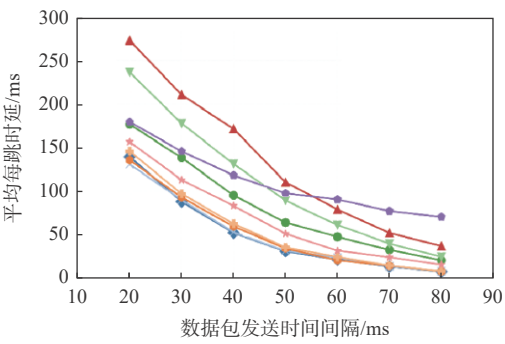
智能路由策略在高频率消息传输场景下保持50%以上的抵达率, 随着间隔的增加, 最终稳定在90%以上, 展现了较强的适应性和稳定性。在平均端到端时延和每跳时延方面, 智能路由策略能够快速适应拓扑变化, 优先选择低拥塞路径, 从而降低时延。此外, QGEO在平均每跳时延对比中也表现出色, 这得益于强化学习的自适应决策, 但相较深度强化学习的表现还是略差。由此可见, 智能路由策略在高动态、低发送间隔的环境下能够保持较高的抵达率, 并有效控制时延, 提升了通信的稳定性,



(a) 不同数据包发送时间间隔下数据包抵达率



(b) 不同数据包发送时间间隔下平均端到端时延



(c) 不同数据包发送时间间隔下平均每跳时延

图5 不同数据包发送时间间隔下网络性能对比

Fig. 5 Comparison of network performance at different packet sending intervals

同时, 在高速无人机集群环境下保持了高效的数据传输能力。

本文还针对无人机速度变化对通信网络性能的影响进行了对比实验。实验中, 将数据包发送时间间隔设定为50 ms, 并通过调整无人机的最高速度来评估不同路由协议与策略在不同移动速度下的性能表现, 结果如图6所示。实验结果表明, 随着无人机速度的增加, 节点的高机动性导致链路的不稳定性显著上升, 部分关键路径因通信距离超出连接范围而频繁断开, 导致集群通信网络表现出高动态特征。GPSR路由协议及QGEO方法在低速情况下较为稳定, 但随着速度增加, 性能下降幅度较大, 尤其在最大速度达到130 m/s时, 由于依赖静态

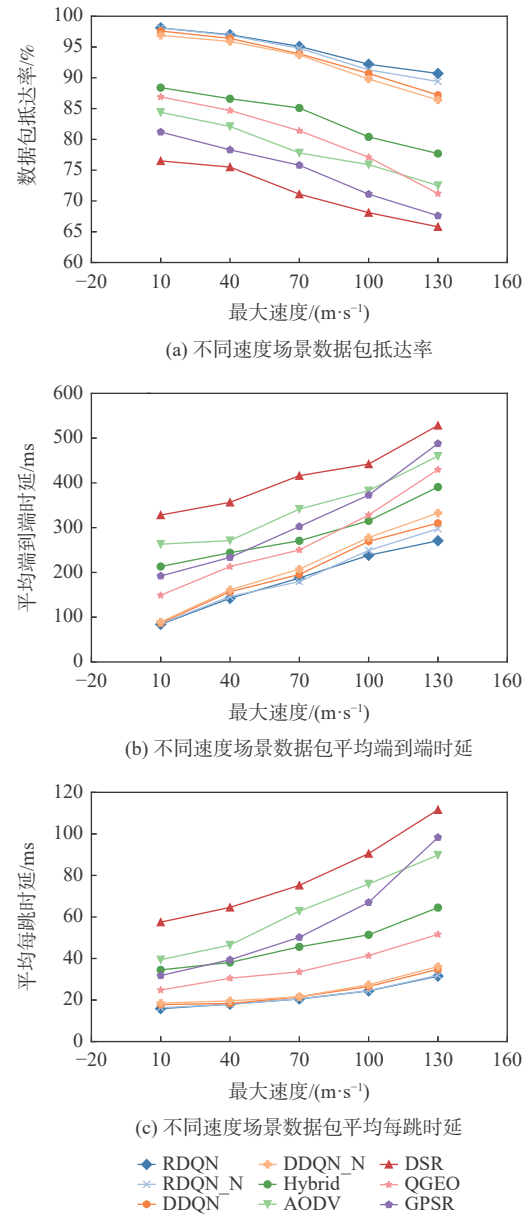


图 6 不同速度场景下网络性能对比

Fig. 6 Comparison of network performance in different speed scenarios

位置信息进行路由决策,其路径更新滞后,导致较高的数据包丢失率。其他传统协议的下降趋势虽不如 GPSR 显著,但也随着速度增加而降低,难以适应高速无人机集群的通信需求,Hybrid 混合响应路由虽表现良好但仍不如本文智能路由策略。

相比之下,智能路由策略在高速环境下保持较好的通信稳定性,尽管其性能随速度提升略有下降,但数据包抵达率始终保持在 85% 以上,远超传统协议,展现出较强的适应性,特别是基于 Rainbow DQN 的深度强化学习模型在不同速度条件下表现最佳。此外,智能路由策略在平均端到端时延和平均每跳时延方面表现出明显优势,在确保高数据包抵达率的同时,有效降低了传输时延,平均端到端延迟、每跳延迟相较传统路由协议平均降低 23%

以上,提升了通信效率。

此外,对网络规模网络及不同节点密度对通信性能的影响进行了比较分析。设定数据包发送间隔(50 ms)和无人机最大飞行速度(130 m/s)的固定条件,通过改变无人机节点数量来模拟不同网络规模,同时,发送数据包总数按节点数量成比例调整。

图 7 展示了在不同网络规模下的数据包抵达率、平均端到端时延及平均每跳时延结果对比。随着网络规模的扩大,包括本文智能路由策略在内的所有路由协议均呈现数据包投递率下降、平均端到端时延与平均每跳时延升高的趋势。其主要原因在于,网络规模扩张引发的拥塞程度加剧,从而降低了整体传输效率。与其他路由协议相比,本文方法在平均每跳时延方面表现出更高的稳定性,整体

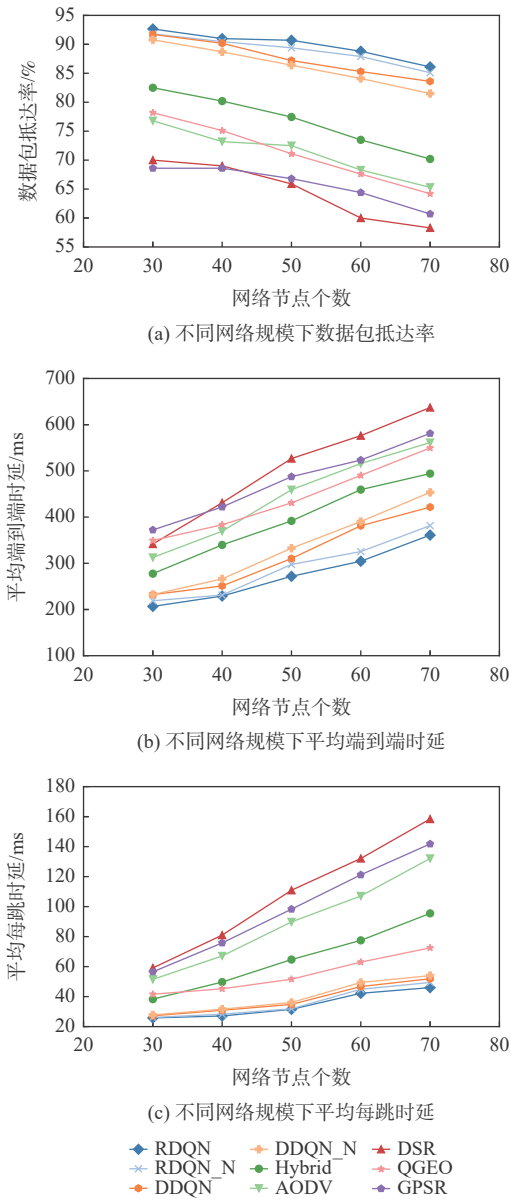


图 7 不同网络规模下网络性能对比

Fig. 7 Comparison of network performance at different network scales

时延增长更为平缓。这一优势得益于各节点能够实时自主地做出路由决策,在一定程度上减轻了大规模网络带来的拥塞上升,从而降低了额外通信时延的影响。

图 8 展示了通过调整网络环境中节点间最小距离进行不同网络节点密度下的网络性能表现对比。结果表明,节点密度的变化对不同路由协议和策略也产生了较大影响,尤其是基于位置的 GPSR 和 QGEO 协议,在节点距离为 30~50 m 和 70~110 m 区间内表现出较大的波动,这主要是由于节点密度的降低和节点间距离的增加导致链路稳定性下降和可用路由选项减少,从而影响了路由决策的有效性。尽管本文方法在一定程度上也受到类似因素的影响,但整体性能仍然保持相对稳定。

4 结 论

本文提出了一种基于 Rainbow DQN 深度强化学习模型框架的高速无人机智能路由策略,旨在提升高速无人机编队在高度动态的自组通信网络中的路由性能与组网稳定性,得出如下结论:

- 1) 在无人机最高速度为 10~130 m/s 范围的仿真实验场景中,本文智能路由策略数据包抵达率在各速度设定对比实验中均保持在 85% 以上,平均端到端时延相较传统路由协议平均降低 23%,满足高速无人机高动态的网络场景。
- 2) 在数据包发送时间间隔为 20~80 ms 范围的仿真实验场景中,本文智能路由策略数据包抵达率在 20 ms 数据包高频率发送场景下仍能保持在 50% 以上,较传统路由协议平均提升 15%。同时,平均端到端时延始终低于传统路由协议,在组网数据包高频率通信场景中具有更为优越的性能表现。
- 3) 在不同网络规模与节点密度场景中,本文智能路由策略通信网络性能受其变化影响波动较小,相较传统路由协议表现出更好的稳定性,能够有效应对网络规模扩大及节点密度变化带来的链路稳定性降低与通信拥塞增加等挑战,从而在多种复杂场景下均能保持较高的通信性能和可靠性。

参考文献 (References)

[1] Maza I, Caballero F, Capitán J, et al. Experimental results in multi-UAV coordination for disaster management and civil security applications[J]. Journal of Intelligent & Robotic Systems, 2011, 61(1): 563-585.

[2] Keller J, Thakur D, Likhachev M, et al. Coordinated path planning for fixed-wing UAS conducting persistent surveillance missions[C]// Proceedings of the IEEE International Symposium on Safety, Security, and Rescue Robotics. Piscataway: IEEE Press, 2016: 1-6.

[3] Khan A, Zhang J, Ahmad S, et al. Dynamic positioning and energy-efficient path planning for disaster scenarios in 5G-assisted multi-UAV environments[J]. Electronics, 2022, 11(14): 2197.

[4] Dahmane S, Yagoubi M B, Brik B, et al. Multi-constrained and edge-enabled selection of UAV participants in federated learning process[J]. Electronics, 2022, 11(14): 2119.

[5] Sharma R, Patel K, Shah S, et al. Aerial footage analysis using computer vision for efficient detection of points of interest near railway tracks[J]. Aerospace, 2022, 9(7): 370.

[6] Zhang R, Li S, Ding Y M, et al. UAV path planning algorithm based on improved Harris Hawks optimization[J]. Sensors, 2022, 22(14): 5232.

[7] Pang X, Liu M, Li Z C, et al. Geographic position based hopless opportunistic routing for UAV networks[J]. Ad Hoc Networks, 2021, 120: 102560.

[8] Wheeb A H, Nordin R, Samah A A, et al. Topology-based routing protocols and mobility models for flying ad hoc networks: a con-

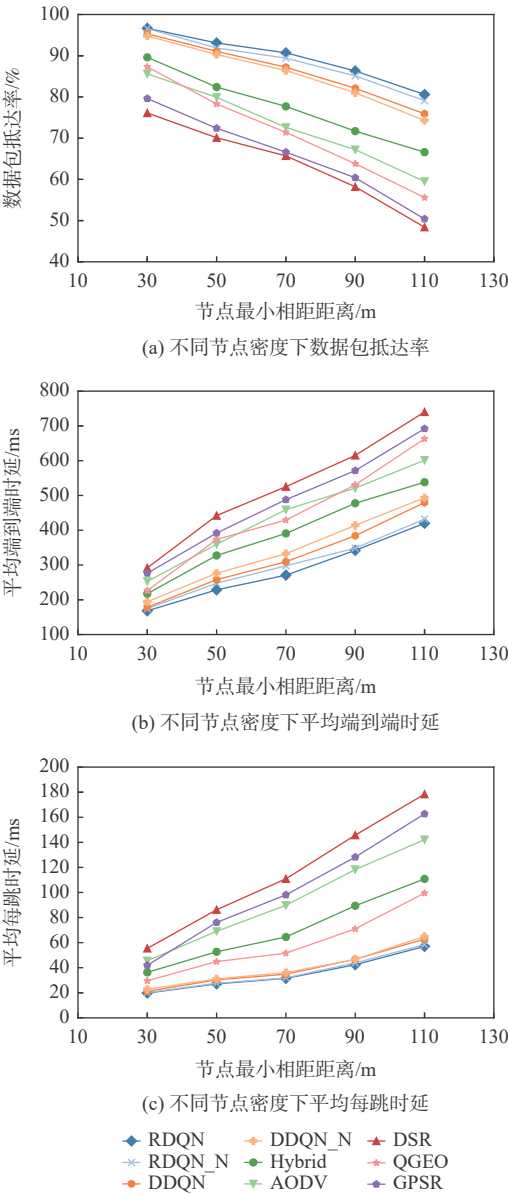


图 8 不同节点密度下网络性能对比

Fig. 8 Comparison of network performance under different node densities

temporary review and future research directions[J]. Drones, 2022, 6(1): 9.

[9] Hong L, Guo H Z, Liu J J, et al. Toward swarm coordination: topology-aware inter-UAV routing optimization[J]. IEEE Transactions on Vehicular Technology, 2020, 69(9): 10177-10187.

[10] Shen H, Jiang Y J, Deng F M, et al. Task unloading strategy of multi UAV for transmission line inspection based on deep reinforcement learning[J]. Electronics, 2022, 11(14): 2188.

[11] Perkins C E, Bhagwat P. Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers[C]//Proceedings of the Conference on Communications Architectures, Protocols and Applications. New York: ACM, 1994: 234-244.

[12] Jacquet P, Muhlethaler P, Clausen T, et al. Optimized link state routing protocol for ad hoc networks[C]//Proceedings of IEEE International Multi Topic Conference, 2001. Technology for the 21st Century. Piscataway: IEEE Press, 2002: 62-68.

[13] Johnson D B, Hu Y C, Maltz D A. The dynamic source routing protocol (DSR) for mobile Ad hoc networks for IPv4, RFC 4728[R]. Fremont: Internet Engineering Task Force, 2007.

[14] Perkins C E, Belding-royer E M, Das S R. Ad hoc on-demand distance vector (AODV) routing[J]. RFC, 2003, 3561: 1-37.

[15] Park V D, Corson M S. A highly adaptive distributed routing algorithm for mobile wireless networks[C]//Proceedings of INFOCOM '97. Piscataway: IEEE Press, 2002: 1405-1413.

[16] 孟冷宇, 郭秉礼, 杨雯, 等. 基于深度强化学习的网络路由优化方法[J]. 系统工程与电子技术, 2022, 44(7): 2311-2318.

Meng L Y, Guo B L, Yang W, et al. Network routing optimization approach based on deep reinforcement learning[J]. Systems Engineering and Electronics, 2022, 44(7): 2311-2318(in Chinese).

[17] Karp B, Kung H T. GPSR: greedy perimeter stateless routing for wireless networks[C]//Proceedings of the 6th Annual International Conference on Mobile Computing and Networking. New York: ACM, 2000: 243-254.

[18] Garg S, Ihler A, Bentley E S, et al. A hybrid reactive routing protocol for decentralized UAV networks[EB/OL]. (2025-01-23) [2025-05-08]. <https://arxiv.org/abs/2407.02929>.

[19] Cao Y, Lien S Y, Liang Y C. Deep reinforcement learning for multi-user access control in non-terrestrial networks[J]. IEEE Transactions on Communications, 2021, 69(3): 1605-1619.

[20] Lin D P, Peng T, Zuo P L, et al. Deep-reinforcement-learning-based intelligent routing strategy for FANETs[J]. Symmetry, 2022, 14(9): 1787.

[21] Qiu X L, Yang Y W, Xu L, et al. Maintaining links in the highly dynamic FANET using deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2023, 72(3): 2804-2818.

[22] Liu C Z, Wang Y X, Wang Q. PARouting: prediction-supported adaptive routing protocol for FANETs with deep reinforcement learning[J]. International Journal of Intelligent Networks, 2023, 4: 113-121.

[23] Jung W S, Yim J, Ko Y B. QGeo: Q-learning-based geographic ad hoc routing protocol for unmanned robotic networks[J]. IEEE Communications Letters, 2017, 21(10): 2258-2261.

[24] Qiu X, Xu L, Wang P, et al. A data-driven packet routing algorithm for an unmanned aerial vehicle swarm: a multi-agent reinforcement learning approach[J]. IEEE Wireless Communications Letters, 2022, 11(10): 2160-2164.

[25] Bai Y J, Zhang X, Yu D J, et al. A deep reinforcement learning-based geographic packet routing optimization[J]. IEEE Access, 2022, 10: 108785-108796.

[26] Wang Z, Li H X, Knoblock E J, et al. Deep reinforcement learning-based joint routing and capacity optimization in an aerial and terrestrial hybrid wireless network[J]. IEEE Access, 2024, 12: 132056-132069.

[27] Ke Y Q, Huang K, Qiu X L, et al. Distributed routing optimization algorithm for FANET based on multiagent reinforcement learning[J]. IEEE Sensors Journal, 2024, 24(15): 24851-24864.

[28] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[29] Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double Q-Learning[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. New York: ACM, 2016: 2094-2100.

[30] Hessel M, Modayil J, Van Hasselt H, et al. Rainbow: combining improvements in deep reinforcement learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1): 5737-5744.

[31] Hu G N, Zhang W, Zhu W H. Prioritized experience replay for continual learning[C]//Proceedings of the 2021 6th International Conference on Computational Intelligence and Applications. Piscataway: IEEE Press, 2021: 16-20.

[32] Wang Z Y, Schaul T, Hessel M, et al. Dueling network architectures for deep reinforcement learning[C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48. New York: ACM, 2016: 1995-2003.

[33] Bellemare M G, Dabney W, Munos R. A distributional perspective on reinforcement learning[EB/OL]. (2017-01-27)[2025-05-08]. <https://arxiv.org/abs/1707.06887>.

[34] Fortunato M, Azar M G, Piot B, et al. Noisy networks for exploration[EB/OL]. (2019-01-09)[2025-05-08]. <https://arxiv.org/abs/1706.10295>.

Intelligent routing strategy for high-speed UAVs via deep reinforcement learning

SHOU Yifan, LIU Dongsheng*, CHEN Yahui

(School of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou 310018, China)

Abstract: With the rapid advancement of unmanned aerial vehicle (UAV) technology, high-speed UAV swarms are increasingly applied in low-altitude complex environments. However, this also poses challenges, as traditional routing protocols struggle to cope with highly dynamic network conditions. Based on the Rainbow deep Q-network (DQN) deep reinforcement learning model, this paper suggests an intelligent routing strategy for high-speed UAVs with an emphasis on communication assurance in high-speed and highly dynamic scenarios and enhancing the capacity of UAV nodes to make autonomous, decentralized decisions. Stable factor is also designed to evaluate the future stability of communication links, enabling UAVs to autonomously make adaptive decisions based on the latest network state. According to experimental results, the high speed intelligent routing scheme suggested in this study reduces the average end-to-end delay and the per-hop delay by more than 23% when compared to traditional routing protocols, while maintaining a packet delivery ratio of more than 85% under all evaluated velocity settings. In high-frequency communication scenarios, the packet delivery ratio is improved by more than 15% on average, effectively meeting the requirements of high-speed UAV networks for stability and communication efficiency.

Keywords: UAV swarm; routing strategy; deep reinforcement learning; flying ad hoc network; high-speed UAV

Received: 2025-05-09; Accepted: 2025-07-25; Published Online: 2025-07-31 09:01
URL: link.cnki.net/urlid/11.2625.V.20250730.1659.004
Foundation items: Zhejiang Province Special Support Program Project (2022R52048); Zhejiang Provincial Science and Technology Program (2024C01028,2024C01211)
* Corresponding author. E-mail: lds1118@zjgsu.edu.cn

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2025.0265

时空网格下多源要素建模与风险量化分析

刘龙浩^{1, 2, 3}, 茹乐^{1, 2, 3, *}, 汪洪桥^{3, 4}, 王文飞^{1, 2, 3}, 张政浩^{1, 2, 3}, 李一帆⁵

(1. 空军工程大学 装备管理与无人机工程学院, 西安 710051; 2. 无人飞行器技术全国重点实验室, 西安 710051; 3. 陕西高校青年创新团队, 西安 710051; 4. 西北工业大学 无人系统技术研究院, 西安 710072; 5. 中国人民解放军 94754 部队, 嘉兴 314011)

摘 要: 针对复杂战场环境下多源要素时空耦合建模的挑战, 提出一种基于时空网格剖分的多源要素建模与风险量化方法。在空间维度, 为解决现有全球网格剖分模型(如 2" 一维整形数组全球经纬部分网格(GeoSOT))因三次虚拟拓展导致的存储冗余, 并弥补北斗网格位置码“字母+十进制数字”编码带来的效率瓶颈, 设计了兼容北斗网格位置码的网格剖分新架构, 显著提升空间数据存储与计算效率; 在时间维度, 针对北斗剖分时间码因 5 次时间进制扩展产生的局限, 提出“绝对-相对”双基准时间层编码体系, 实现空间位置与时间维度的四维联合表征, 有效突破异构要素时空融合瓶颈。构建“全局基准网格-局部作战网格”2 级嵌套时空基准框架, 支持动态调整网格粒度, 兼顾全局一致性与局部适应性。针对传统最小外接矩形法在表征复杂地形与空域要素几何特性时存在的轮廓失真问题, 提出基于射线交叉判定法的三维实景拓扑特征建模方法, 依托时空网格剖分架构构建敌方作战单元的统一量化模型, 并结合风险等级将战场空间划分为自由空间、竞争空间与拒止空间。仿真实验表明: 所提方法能够实现战场多源要素的多尺度快速重构; 当网格精度约为 1 000 m 时, 空间表征的平均相对误差仅为 1.414%; 相较于传统方法, 相对误差平均降低 49.37%, 时间效率提升 99.65%; 可满足无人机集群航迹规划的实时响应需求, 为智能化无人机集群作战环境建模提供了可验证的新范式。

关 键 词: 时空网格; 北斗网格码; 战场要素; 风险量化; 无人机集群

中图分类号: V279⁺.2; V37

文献标志码: A

文章编号: 1001-5965(2026)09-3223-14

随着人工智能与无人系统的突破性发展, 智能无人集群将成为未来颠覆性作战力量^[1]。然而, 战场的复杂性对无人机集群构成了多维挑战: 其运行环境不仅受密集城市群^[2]、复杂地形与管制空域^[3]的影响, 也涉及雷达探测、防空火力拦截、电磁干扰等多域威胁^[4]。时空网格作为一种新兴的数据组织与管理范式, 为破解此类复杂性问题的全新思路, 但传统时空网格架构普遍存在数据存储冗余、时空维度割裂、编码效率低下及战场要素建模轮廓失真等突出瓶颈, 难以满足复杂战场环境下实时响应与精准决策的核心需求。为此, 迫切需要构

建一套全新的时空网格剖分架构, 以支撑战场多源要素的高效建模与风险量化分析, 为无人机集群仿真环境搭建提供新范式。

时空网格剖分是按照统一规则对地球空间维度与时间维度进行离散划分, 由此构建而成的四维时空基准框架。其借助时空编码对四维时空中的体元进行唯一标识, 且能够计算体元间的时间和空间位置关系^[5]。其核心是通过特定的剖分规则, 将连续的地球表面无限细分为尺度连续、形状相似的多层次网格单元, 并为每个单元赋予唯一位置编码, 使其成为空间区域的标识。在此基础上, 三维

收稿日期: 2025-05-06; 录用日期: 2025-11-03; 网络出版时间: 2025-12-01 11:42

网络出版地址: link.cnki.net/urlid/11.2625.V.20251201.0852.002

*通信作者. E-mail: ru-le@163.com

引用格式: 刘龙浩, 茹乐, 汪洪桥, 等. 时空网格下多源要素建模与风险量化分析[J]. 北京航空航天大学学报, 2026, 52(9): 3223-3236.

Liu L H, Ru L, Wang H Q, et al. Multi-source element modeling and risk quantitative analysis under spatio-temporal grid[J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52(9): 3223-3236 (in Chinese).

立体剖分将网格体系扩展至地球内部及外围空间,形成多层次、无缝、无重叠的立体空间网格集合及其唯一编码。时间剖分则将连续时间离散化为无缝、无重叠的时间段。最终,时空网格剖分融合空间体元编码与时间单元编码,构建出具有唯一时空标识的四维体元体系,形成支撑时空数据组织与计算的统一参考系统。

科学的时空网格剖分结构可以准确描述战场多源要素,提高作战管理效能^[6]。美军在该领域取得了显著进展,建立了多种成熟的时空标识与空间定位系统,如军事网格参考系统(military grid reference system, MGRS)、通用地理参考系统(common geographic reference system, CGRS)和全球区域参考系统(global area reference system, GARS)^[7]。其中, CGRS 支持根据作战区域灵活确定剖分原点,适配局部战场快速构建的需求,但因原点不固定,易出现不同作战区域编码重叠、标识不唯一的问题; GARS 作为面参考系统,便于作战协同通报,但其最小网格表征区域为 9.26 km,难以适配高精度协同作战场景; MGRS 作为点参考系统,可实现高精度定位协同,不过编码位长较长,对数据存储效率有一定影响。

在战场数据的时空基准使用方面,中方体系与美军存在显著差异:中国采用 CGCS—2000 坐标系,而美军普遍使用 WGS—84 坐标系。此外,美军在作战实践中进一步区分了目标描述规则——2012 年美军参谋长联席会议发布的《联合作战中的地理空间情报》报告明确规定,点状目标定位需采用 MGRS 坐标,而面状区域参考采用 GARS 坐标。这种点、面目标描述采用不同参考系统的模式易导致协同行动中的数据转换混乱与语义歧义。为规避该类问题并提升战场数据的统一性,亟须探索符合需求的战场数据时空管理方法。

近年来,国内外学术界对全球网格剖分的研究取得了一定进展,中国也出台了《地球空间网格编码规则》^[8]、《北斗网格位置码》^[9]和《北斗剖分时间码》^[10]等一系列国家标准文件。由北京大学程承旗教授团队提出的 2ⁿ 一维整型数组全球经纬剖分网格(geographic coordinate subdividing grid with one dimension integral cording on 2ⁿ-tree, GeoSOT)^[11]具有中国完全自主知识产权,其通过经纬网向空间扩展的方式,构建了多尺度、内外部一致的三维空间模型,实现了点面一体化表征,适配计算机二进制存储需求。该模型将地球表面空间进行 3 次拓展(纬度、经度拓展至-256°~256°,1°拓展为 64′、1′拓展为 64″),虽达成等度、等分、等秒剖分目标,但强制数值对齐的设计导致一定程度的存储冗余,且网格

的分块性、可读性和可识别性未充分适配实际作战应用场景^[12]。北斗网格位置码与 GeoSOT 同根同源,摒弃了 3 次拓展方式,通过二分、三分、八分等差异化划分规则实现地球表面多尺度细分,其“字母+十进制数字”的编码方式(如某点第八级编码为“N50J47539B825534”)更便于人工理解与解读。但该编码方式与计算机二进制存储的底层需求适配性不足,在大规模数据处理场景下存在效率瓶颈。

北斗剖分时间码采用二叉树结构对时间轴进行剖分,是时刻表达方式的重要补充,实现了时刻/时段标识的一致性。其在限定时间域(公元前 2¹⁵~公元(2¹⁵-1)年)内,通过 5 次时间进制扩展(1 年→16 个月、1 个月→32 天等),构建了多粒度、统一化的离散时间编码体系,保障了时段标识的唯一性。但固定的时间拓展规则难以动态调整时间粒度,无法贴合瞬息万变的作战节拍,且与空间网格协同表达时存在语义隔阂。

综合考虑 GeoSOT、北斗网格位置码和北斗剖分时间码的优缺点,提出一种全新的时空网格架构。在空间维度,继承 GeoSOT 第一次拓展及二进制编码方式,但取消度、分进制的拓展;继承北斗网格位置码各层级之间的划分规则,但摒弃“字母+十进制数字”的存储方式。在时间维度上,提出双基准映射机制:继承北斗剖分时间码的虚拟拓展规则作为“绝对时间基准层”,但引入“相对时间基准层”作为补充,进一步提高时空数据的组织、处理、分析、传递和运用效率。

国内外学者在战场多源要素建模领域已取得阶段性进展:季荣涛^[13]采用五元组将战场空间离散化,划分自由、竞争和拒止 3 类子空间,为战场区域分类提供了基础框架,但未涉及敌方作战单元的具体表征;吴乾坤^[14]基于全空间剖分理论,将战场要素抽象为点、线、面、体的组合,完成战场空间构建,但其模型未充分考虑要素动态特性;李锋等^[15]提出地球圈层网格向时间维度拓展的时空一体化数据组织模型,实现场要素和对象要素的高效管理,不过对敌方威胁单元的刻画较为简略;袁越等^[16]基于剖分表达结构对雷达三维探测范围进行数学建模,为敌方作战单元建模提供了思路,但建模精度和计算效率有待提升;周畅^[17]提出雷达、防空和干扰设备的建模与风险量化方法,核心局限在于未考虑作战单元的动态变化特性;Han 等^[18]提出的网格优化型室内空域建模(grid-optimized indoor airspace modeling, GO-IAM)算法建立了室内多层次三维网格模型,但其采用最小外接矩形简化障碍物模型,在复杂地形与空域要素表征中存在轮廓失真问题。

针对现有研究的技术短板, 本文提出一种基于时空网格的多源要素建模与风险量化方法, 主要研究工作包括: ①设计兼容北斗网格位置码/时间码的时空网格剖分架构, 通过 Z 曲线编码实现 96 位空间编码与“绝对-相对”双基准时间层的联合表征, 突破战场多源要素四维融合瓶颈, 提升数据存储与处理效率; ②基于射线交叉判定法构建地形与空域要素的高保真三维实景拓扑模型, 解决传统建模方法的轮廓失真问题; ③依托时空网格框架构建敌方雷达、火炮等典型威胁单元的简化表征模型, 模拟风险场时空演化规律, 提升建模精度与分析效率, 弥补传统方法在动态性与准确性上的不足。

1 时空网格剖分架构设计

1.1 空间网格剖分框架

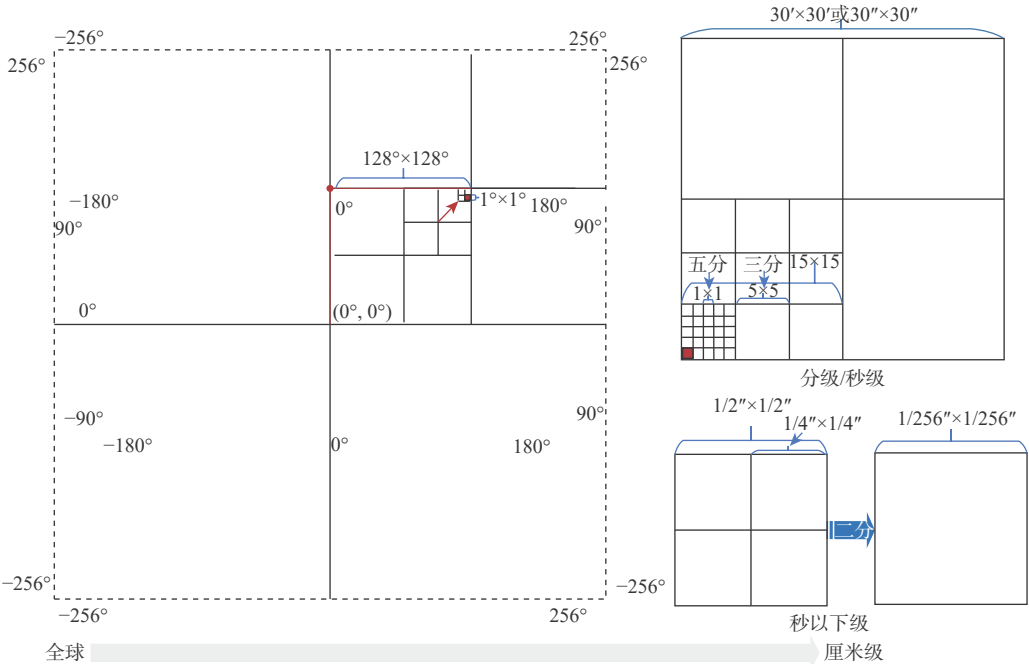
为应对复杂战场环境中多源数据的处理需求, 本节提出一种新型全球网格剖分系统, 旨在为无人机集群作战中的数据管理提供统一的结构化表达与计算框架。

该方法以地球空间三维细分网格基本理论为基础, 对战场环境进行网格化建模。空间网格剖分基准设定为地球参考椭球面, 以本初子午线与赤道平面交点为原点建立三维正交坐标系。网格在经度方向($\pm 180^\circ$)和纬度方向($\pm 90^\circ$)实现全球覆盖, 高

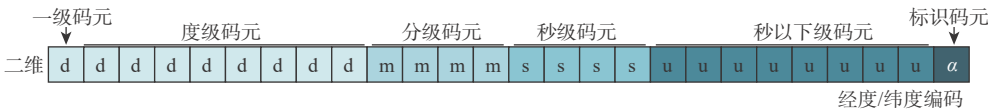
度方向则从地表以下 $-3\,000\text{ km}$ 延伸至地表以上 $50\,000\text{ km}$, 形成全局无缝覆盖的多层级网格体系。层级划分严格遵循等经纬度差与等高度差原则, 通过逐级剖分构建尺度连续、无缝无叠的多层次三维网格系统, 其最小体元可达厘米级。

在经纬度方向, 网格层级按空间分辨率划分, 图 1(a) 为二维网格剖分示意。根据现有定位系统(如全球定位系统、北斗等)定位精度及无人机集群作战定位精度需求, 本文二维网格剖分到 25 层级约 12.5 cm , 基本可以满足无人机集群作战数据管理需求。其中, 1~9 级为度级网格, 10~13 级为分钟级网格, 14~17 级为秒级网格, 18~25 级为秒级以下网格。图 1(b) 为二维网格编码构成示意, 经纬度方向分别生成 32 位二进制数编码。图 1(c) 为二维网格剖分编码规则示意, 经纬度方向的空间编码采用 Z 序曲线技术, 最终经纬编码通过莫顿交叉共同构成 64 位二维平面编码。

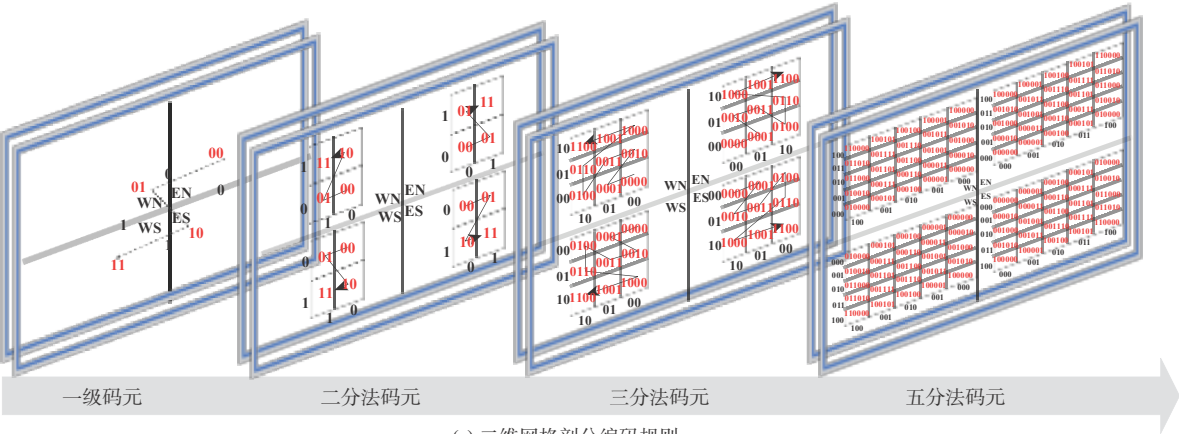
在高度方向, 网格层级按分辨率划分, 图 1(d) 为高度网格剖分与编码规则示意。地球同步轨道卫星高度约为 $36\,000\text{ km}$, 同时, 必须考虑地表存在海拔小于 0 m 的区域(如中国吐鲁番盆地海拔为 -154.31 m)^[12], 在大地高 $-3\,000\sim 50\,000\text{ km}$ 范围内, 椭球面近似拓展与精确拓展之间差异最大不超过 10 m ^[19], 因此, 为保证划分的统一性和多尺度性, 将



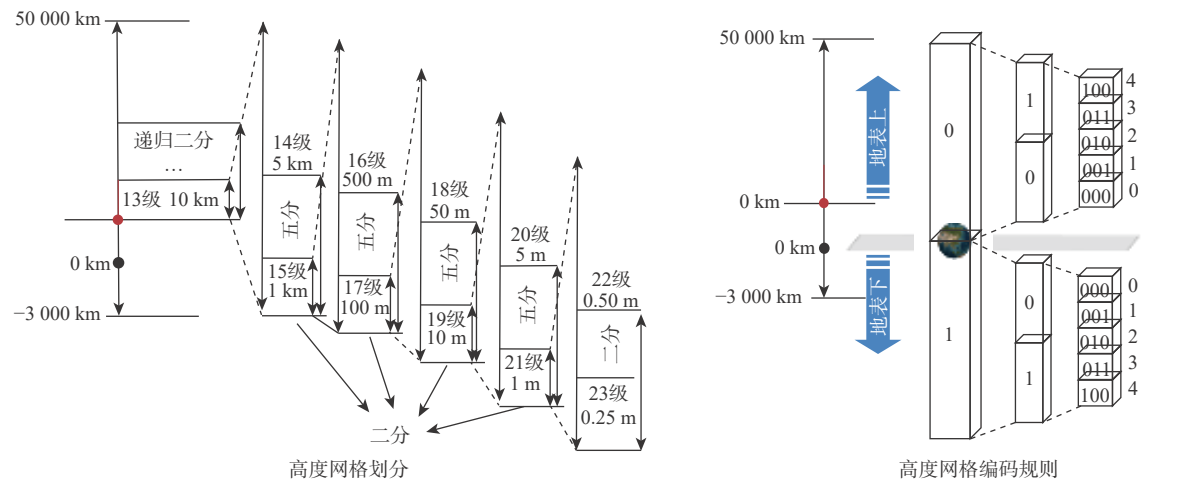
(a) 二维网格剖分



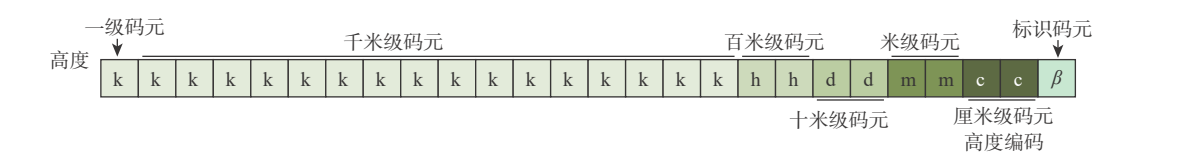
(b) 二维网格编码构成



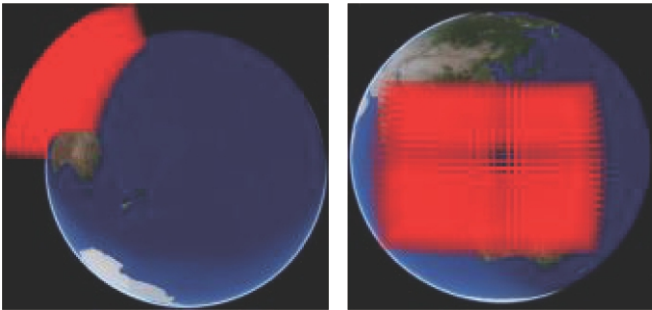
(c) 二维网格剖分编码规则



(d) 高度网格剖分与编码规则



(e) 高度编码构成



(f) 空间网格剖分的实际形态

图 1 空间网格剖分示意

Fig. 1 Spatial grid subdivision schematic

高度编码支持范围设定为-3 000~50 000 km，具体划分规则为：首先，整体范围分为 2 部分：地表以下（-3 000~0 km）和地表以上（0~50 000 km），0 表示海平面以上区域，1 表示海平面以下区域（即使用第一个无符号整数标识地上或地下）；其次，结合无人智能作战需求，将高度层级通过二分法和五分法划分为 23 级（1~12 级采用二分法，13~23 级交叉使

用二分法和五分法），最细粒度为 25 cm。图 1(e) 为高度编码构成示意，高度编码一级码元标识地表以上或地表以下，随后按照不同高度值生成 32 位二进制高度编码。

最终的三维网格单元一维编码由 64 位二维平面编码（包含经纬度信息）与 32 位高度编码合并生成，确保每个剖分网格单元拥有唯一且无歧义的

一维编码标识,实际剖分效果如图 1(f) 所示。

此外,本文空间网格码与北斗网格位置码在剖分规则上有较好的兼容性,其中,度分秒网格可使用聚合或等价表示,具体映射关系如表 1 所示。

表 1 空间网格码部分层级与北斗网格码兼容性

Table 1 Spatial grid code part levels and Beidou grid codes compatibility

北斗网格位置码		空间网格码		转换方式
层级	网格大小	层级	网格大小	
1	6°×4°	8	2°×2°	网格聚合3×2个
1	6°×4°	9	1°×1°	网格聚合6×4个
2	30'×30'	10	30'×30'	网格等价
2	30'×30'	11	15'×15'	网格聚合2×2个
2	30'×30'	12	5'×5'	网格聚合6×6个
2	30'×30'	13	1'×1'	网格聚合30×30个
3	15'×10'	13	1'×1'	网格聚合15×10个
4	1'×1'	13	1'×1'	网格等价
4	1'×1'	14	30"×30"	网格聚合2×2个
4	1'×1'	15	15"×15"	网格聚合4×4个
4	1'×1'	16	5"×5"	网格聚合12×12个
5	4"×4"	17	1"×1"	网格聚合4×4个
6	2"×2"	17	1"×1"	网格聚合2×2个
6	2"×2"	18	1/2"×1/2"	网格聚合4×4个
7	1/4"×1/4"	19	1/4"×1/4"	网格等价
7	1/4"×1/4"	20	1/8"×1/8"	网格聚合2×2个
7	1/4"×1/4"	21	1/16"×1/16"	网格聚合4×4个
8	1/32"×1/32"	22	1/32"×1/32"	网格等价
8	1/32"×1/32"	23	1/64"×1/64"	网格聚合2×2个
8	1/32"×1/32"	24	1/128"×1/128"	网格聚合4×4个
9	1/256"×1/256"	25	1/256"×1/256"	网格等价

1.2 时间维拓展

战场环境通常不是静止不变的,各要素属性和状态随时间动态演变,单纯的空间描述难以满足战场时空数据的组织与管理需求^[15],因此,需要在空间网格剖分的基础上引入时间维度,采用“绝对-相对”双基准时间层描述时间状态,并与空间网格编码结合,形成时空一体编码体系。

对于绝对时间基准层,根据北斗剖分时间码虚拟拓展理论^[10],结合作战使用习惯,采用公元纪年编码“年-月-日-时-分-秒”(YYYYMMDDHHMMSS) 6 层 26 位二进制编作为基础时间轴,与空间编码融合为 122 位时空编码,足以精确表示任意位置和大小的时空信息,同时,保持变形小、多尺度、细分简单、计算处理快速等特点^[12]。

对于相对时间基准层,目前,网格自适应方法仅限于空间维度调整,时间维度仍采用固定全局步长,导致任务执行效率低下。此外,北斗剖分时间

码层级较多且编码较长,不便于描述不同集群的不同时间步长。因此,综合考虑空间网格剖分、作战任务及无人机性能,选择空间网格粒度并调整时间步长,设计精简的北斗时间剖分方案(表 2),以提高无人机集群作战效率。

表 2 改进型时间剖分层级

Table 2 Improved time subdivision levels

时间层级	时间尺度	任务类型	机型代表
1	(60 min) 1 h	战略级	MQ-9 无人机
2	15 min	战役级	
3	5 min	过渡级 1	
4	(60 s) 1 min	战术级	RQ-28A 无人机
5	15 s	过渡级 2	
6	5 s	定位级	旋翼无人机
7	1 s	精确级	微型无人机

时间轴采用 7 级分层结构(表 2),通过式(1)动态调整时间步长适配不同场景:

$$\Delta t_l = \frac{1}{l} \Delta t_{\max}$$

(1)

式中: $\Delta t_{\max}$ 为全局最大时间步长; l 为空间网格剖分层级数。

关于 Δt_l 的解释为: 如图 2 网格 1、2、9、10 层数 $l=0$, 网格 3、8 层数 $l=4$, 网格 4~7 层数 $l=12$ 。开始时,所有网格处于 t_n 时刻, $t_n \sim t_{n+1}$ 时间步长为 $\Delta t_{\max}$ 。首先,对所有网格以时间步长 $\Delta t_{\max}$ 剖分,然后,以 $\Delta t_{\max}/4$ 的时间步长计算并更新网格 3 与网格 8 时刻,此时,网格 3、8 处于 $t_n + \Delta t_{\max}/4$ 时刻,同理,网格 4~7 处于 $t_n + \Delta t_{\max}/12$ 时刻,以此类推,直至得到局部最小时间步长。

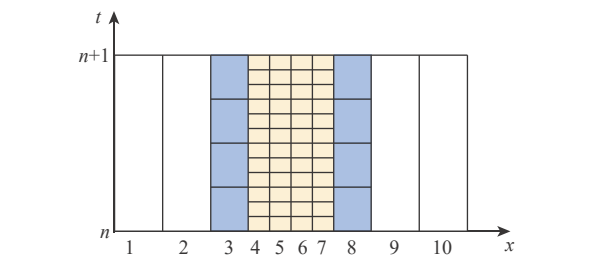


图 2 相对时间基准层时间粒度剖分示意

Fig. 2 Schematic diagram of time granularity subdivision relative to time reference layer

1.3 时空一体化网格体系

将地球空间剖分为不同层级体系,在此基础上,引入时间轴作为第四维度形成时空网格单元后,必须对网格单元赋予唯一编码才能有效管理和描述战场要素。从剖分结构看,要唯一确定任意时空网格单元 C_{ST} ,需要明确空间、时间和属性 3 层编码标识,可用三元组表示为

$$C_{ST} = \langle \Phi(L_{\lambda\varphi}, L_h, \zeta_{L_{\lambda\varphi}}, \delta_{L_h}, \lambda, \varphi, h), \Psi(t_a, t_r), \Theta(A, I) \rangle \quad (2)$$

式中: Φ 为空间维度算子; Ψ 为时间维度算子; Θ 为战场风险属性; $L_{\lambda\varphi}$ 和 $\zeta_{L_{\lambda\varphi}}$ 分别为经纬方向剖分层级和尺度; L_h 和 δ_{L_h} 分别为大地高剖分层级与尺度; λ 、 φ 、 h 分别为经度、纬度、高度; t_a 为绝对时间戳; t_r 为相对时间粒度; A 和 I 分别为战场要素和风险指数。

完成时空网格一体化编码后,如图3(a)所示,通过设计“全局基准网格-局部作战网格”2级嵌套架构,建立可动态调整网格粒度的时空基准框架。以 g_o 表示整个时空网格, $g_{i,t}$ 表示 t 时刻网格中无人机 i , $u_{i,t}$ 表示 t 时刻网格状态,将无人机集群指控节点抽象为网格状态如下:

$$\begin{cases} u_{i,t} = 1 & g_{i,t} \text{ in } g_o \\ u_{i,t} = 0 & g_{i,t} = \emptyset \end{cases} \quad (3)$$

定义全局基准网格指控节点 C_i 在 t 时刻的网格状态 S_C^t 为

$$S_C^t = \langle G_{\text{global}}, L_{\text{heir}}, \Omega_I, V_{\text{traj}} \rangle \quad (4)$$

式中: G_{global} 为全局时空网格集合; L_{heir} 为全局层级配置; Ω_I 为多要素风险场; V_{traj} 为历史航迹集合。

局部作战网格单机状态 S_U^t 表征为

$$S_U^t = \langle G_{\text{local}}, \theta_t, \phi_t, I \rangle \quad (5)$$

式中: G_{local} 为局部时空网格集合; θ_t 为航向角; ϕ_t 为俯仰角; I 为该网格风险指数。

如图3(b)所示,将战场要素(如地形、建筑、空域、雷达等)经过空域划分网格化表征后,构建“全

局基准网格-局部作战网格”双坐标系处理位置关系,设地理坐标 $O(\lambda_o, \varphi_o, h_o)$ 为全局坐标系 $Oxyz$ 原点, $O'(\lambda_{o'}, \varphi_{o'}, h_{o'})$ 为局部坐标系 $O'x'y'z'$ 原点。通过式(6)、式(7)分别计算双坐标系下 $A'(\lambda_{A'}, \varphi_{A'}, h_{A'})$ 的位置:

$$\begin{cases} X = \text{Floor}(\lambda_{A'}, \lambda_o) \\ Y = \text{Floor}(\varphi_{A'}, \varphi_o) \\ Z = \text{Floor}(h_{A'}, h_o) \end{cases} \quad (6)$$

$$\begin{cases} x' = \text{Floor}(\lambda_{A'}, \lambda_{o'}) \\ y' = \text{Floor}(\varphi_{A'}, \varphi_{o'}) \\ z' = \text{Floor}(h_{A'}, h_{o'}) \end{cases} \quad (7)$$

式中: $\text{Floor}(x, y) = \lfloor (x - y) / s_L + 0.5 \rfloor$ 为网格在相应维度上的移动距离; s_L 为该剖分层级下的网格尺度; $\lfloor \cdot \rfloor$ 为向下取整符号。

根据上述构建双坐标系方法,假设全局网格在经纬方向三等分为局部网格,全局时间步长为 Δt ,则局部时间步长为 $\Delta t_l = \Delta t / 3$,故图3(b)中 A' 坐标(局部红色网格)表示为 $((4, 0, 0), (2, 1, 2))$,表示 A' 位置的全局坐标为 $(4, 0, 0)$,局部坐标为 $(2, 1, 2)$ 。

2 多源要素建模与风险量化

战场多源要素是描述特定作战区域在特定时刻各方作战力量的状态与战场环境的数据集合,这些要素相互作用、相互影响,共同塑造了高度复杂、动态演变的战场空间^[20]。在分类上,比较常见的是将战场多源要素分为战场环境类、兵力部署与作战能力类、认知要素类等。风险量化指在战场环境下,综合运用数学、统计或计算方法,将战场多源要素对特定任务/目标可能造成的负面影响评估出来,并赋予可比较、可计算数值的过程。

基于战场多源要素差异化特征,本文将其分为地理与空域(geographical and airspace, GA)要素及敌方作战单元。GA类要素主要包括地形、建筑物、管制空域等静态风险区,可简化为多边形拓扑结构,针对该类要素采用射线交叉判定法进行三维实景拓扑特征建模;敌方作战单元类要素涵盖雷达探测、火炮杀伤及电子干扰等动态威胁源,该类要素可简化为半球体结构^[21],通过时空参数解构实现动态特性表征。构建战场多源要素风险量化模型具体分2个步骤:多源要素建模和风险量化分析。

2.1 多源要素建模

2.1.1 地理与空域模型

在空间地理信息科学领域,地理实体与空域资源的三维空间表征通常基于多边形拓扑结构与数字高程模型的协同映射机制实现离散化建模。其

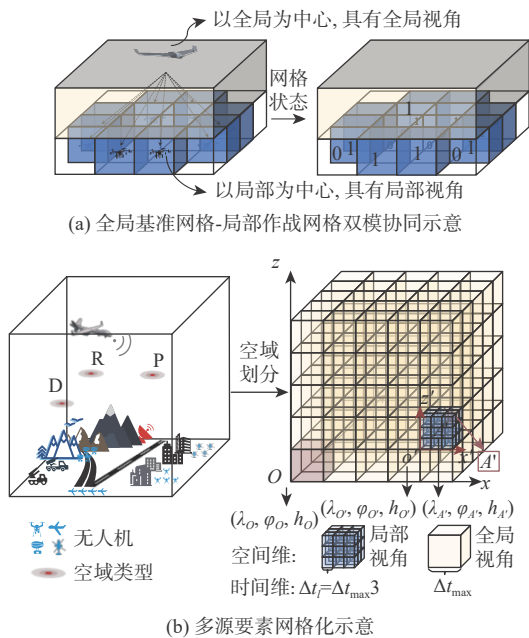


图3 多粒度时空网格基准框架

Fig. 3 Multi-granularity spatio-temporal grid benchmark framework

核心方法为射线交叉判定法, 该算法通过解析测试点与多边形边界的定向拓扑关系, 构建连续地理空间到离散网格空间的数学映射规则^[22]。

设地理平面中 p_0 点经纬度为 (λ_p, φ_p) , 给定原点经纬度后, 由式 (6) 或式 (7) 计算得到坐标 $p_0(x_p, y_p)$ 。如图 4 所示, 构造一条水平向右的射线 $r: y = y_p, x \geq x_p$, 统计该射线与多边形边界 $\partial P = \{e_1, e_2, \dots, e_n\}$ 的有效交叉次数 N_{cross} , 则 GA 模型具体建模步骤如下。

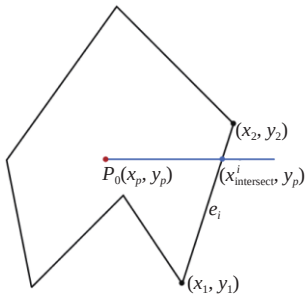


图 4 测试点与多边形关系

Fig. 4 Test point and polygon relationship

步骤 1 边筛选条件: 任意边 $e_i (i \in [1, n])$ 端点坐标为 (x_1, y_1) 和 (x_2, y_2) , 仅当边 e_i 的纵坐标区间 $[y_{\min}^{e_i}, y_{\max}^{e_i}]$ 包含 y_p 时进入下一步计算。

步骤 2 边界判断: 满足式 (8) 则判定为边界区域 $P_0 \in P_{\text{boundary}}$:

$$\left| \frac{y_p - y_1}{x_p - x_1} - \frac{y_2 - y_1}{x_2 - x_1} \right| \leq \varepsilon \quad (8)$$

式中: ε 为浮点计算容差, 通常取 e^{-6} 。

步骤 3 交点计算: 通过直线方程求解射线与边 e_i 的交点横坐标 $x_{\text{intersect}}^i$, 满足:

$$x_{\text{intersect}}^i = \frac{y_p - y_1}{y_2 - y_1} (x_2 - x_1) + x_1 \quad (9)$$

步骤 4 有效性判定: 若 $x_{\text{intersect}}^i > x_p + \varepsilon$, 则记录一次有效交叉, 设有效交叉总次数为 N_{cross} 。

步骤 5 奇偶性规则: 满足 $N_{\text{cross}} \equiv 1(\text{mod}2)$ 时, 判定为内部区域 $P_0 \in P_{\text{inner}}$ 。

2.1.2 敌方作战单元模型

采用半球体作为敌方作战单元作用域的简化几何模型^[21], 相关建模技术主要是基于几何光学和数学建模思路, 存在计算量大、复杂度高等问题, 且考虑地形遮蔽影响后 (图 5), 需要不断计算从作战单元中心点发出的射线与数字高程模型表面的相交运算^[16]。针对上述问题, 提出基于时空网格剖分架构的作战单元作用域求解算法, 该算法计算地形遮蔽情况下的作用域时, 将地形与作战单元作用域均剖分化, 将传统的线面、面面相交运算转化为网格占用与否的判断, 而非剖分体元编码集合的求交

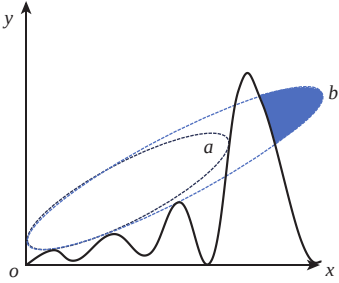


图 5 作战单元作用域被地形截断

Fig. 5 The scope of combat unit truncated by terrain

运算, 从而实现作战单元作用域建模。

设某作战单元地理位置 $O_0(\lambda_o, \varphi_o, h_o)$, 由式 (6) 或式 (7) 可得到三维坐标 $O_0(x_o, y_o, z_o)$, 该作战单元作用域半径为 R , 经纬 $L_{\lambda\varphi}$ 层级下的剖分体元边长为 $\zeta_{L_{\lambda\varphi}}$, 高度 L_h 层级下的剖分体元边长为 δ_{L_h} , 则敌方作战单元模型具体建模步骤如下。

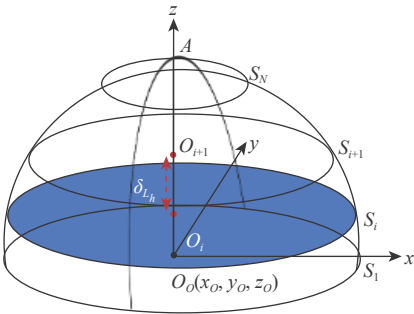
步骤 1 垂直剖分: 将作用域沿垂直方向进行分层剖分 (图 6(a)), 得到 M 个等间距圆形截面簇, M 由 R 和 δ_{L_h} 决定, 满足:

$$M = \frac{R}{\delta_{L_h}} \quad (10)$$

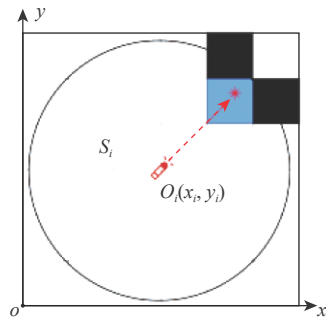
$$R_i = \sqrt{R^2 - [(i-1)\delta_{L_h}]^2} \quad (11)$$

式中: R_i 为第 i 个圆面 S_i 的半径。

步骤 2 水平填充: 采用射线追踪的快速体素遍历算法^[18] 对 M 个圆面进行水平方向网格化填充, 构建作战单元作用域的水平剖分结构 (图 6(b))。



(a) 垂直剖分



(b) 水平剖分

图 6 敌方作战单元建模示意

Fig. 6 Schematic diagram of enemy combat unit modeling

通过判断网格是否被地形剖分网格占用(即 $F(x,y) \neq 1$)筛选出作用域的有效范围,从而解决文献 [16] 采用“扫描线填充算法”仅适用于无地形影响或完全被地形截断情况下的计算问题。

设圆面 S_i 的圆心坐标为 $O_i(x_i,y_i)$,水平方向格网化填充规则满足:

$$\{(x,y)|(x-x_i)^2+(y-y_i)^2 \leq N_i^2 \text{ 且 } F(x,y) \neq 1\} \tag{12}$$

$$N_i = \left\lceil \frac{R_i}{\zeta_{L_{iq}}} \right\rceil \tag{13}$$

式中: $\lceil \cdot \rceil$ 为向上取整符号。

2.2 风险量化分析

以无人机行动单元为研究对象,进行多源要素风险指数量化分析。

2.2.1 GA 量化

针对地形及建筑等静态实体^[2],其碰撞概率在飞行器运行中具有确定性特征,该类要素的风险指数恒定为 1.0(即由 2.1.1 节分析所得 $P_0 \in P_{\text{inner}}$),同时,将多边形边界设置为缓冲区(即由 2.1.1 节分析所得 $P_0 \in P_{\text{boundary}}$)。空域类型对飞行器运行约束呈现显著异质性,其风险指数量化需结合空域管制规则量化。基于上述要素, p 处无人机的风险指数 $I_{\text{GA}}(x,y,z,t)$ 量化为

$$I_{\text{GA}}(x,y,z,t) = \begin{cases} 0 & \text{其他} \\ I_{\text{GA1}} & p \in P_{\text{boundary}} \\ I_{\text{GA2}} & p \in P_{\text{inner}} \end{cases} \tag{14}$$

式中: I_{GA1} 、 I_{GA2} 因地形、建筑、管制空域类型而异,特别地,对于地形和建筑要素, $I_{\text{GA2}} = 1.0$ 。

2.2.2 敌方作战单元量化

设 t 时刻第 u 个无人机位置为 $\overrightarrow{P_u(t)}$,第 s 个敌方作战单元位置为 $\overrightarrow{P_s(t)}$,两者之间距离为 $D_{su}(t)$:

$$\begin{cases} \overrightarrow{P_u(t)} = (x_u(t),y_u(t),z_u(t)) \\ \overrightarrow{P_s(t)} = (x_s(t),y_s(t),z_s(t)) \\ D_{su}(t) = \|\overrightarrow{P_s(t)} - \overrightarrow{P_u(t)}\| \end{cases} \tag{15}$$

对于雷达、火炮和干扰^[17,23]3 类敌方作战单元,其风险指数量化过程如下。

1) 雷达单元。设雷达探测最大作用距离为 R_{max} ,扫描周期为 T_{rad} ,波束角度摆动角度为 $\theta(t)$,则雷达单元风险指数 $I_{\text{rad}}(x,y,z,t)$ 量化为

$$I_{\text{rad}}(x,y,z,t) = \begin{cases} \frac{\cos(\theta(t))}{1 + D_{ij}^4/R_{\text{max}}^4} & D_{su}(t) \leq R_{\text{max}} \\ 0 & D_{su}(t) > R_{\text{max}} \end{cases} \tag{16}$$

$$\theta(t) = \frac{\pi}{3} \sin \frac{2\pi t}{T_{\text{rad}}} \tag{17}$$

2) 火炮单元。设火炮最大攻击半径和高危攻击半径分别为 R_{Amax} 和 R_{Amin} ,弹药装填时间间隔为

T_{atk} ,射击窗口区为 Δt ,第 k 次开火时间为 t_k ,在 $[t_k, t_k + T_{\text{atk}}]$ 时段内风险指数归零,则火炮单元风险指数 $I_{\text{atk}}(x,y,z,t)$ 量化为

$$I_{\text{atk}}(x,y,z,t) = \begin{cases} 1 & D_{su}(t) < R_{\text{Amin}} \\ \frac{R_{\text{Amax}} - D_{su}(t)}{R_{\text{Amax}} - R_{\text{Amin}}} & R_{\text{Amin}} \leq D_{su}(t) \leq R_{\text{Amax}} \\ 0 & D_{su}(t) > R_{\text{Amax}} \end{cases} \tag{18}$$

装填时间内,风险指数为 0:

$$I_{\text{atk}}(x,y,z,t) = 0 \quad t \bmod (T_{\text{atk}} + \Delta t) \leq T_{\text{atk}} \tag{19}$$

3) 干扰单元。设干扰单元最大干扰半径和高危干扰半径分别为 R_{Imax} 和 R_{Imin} ,干扰波动周期为 T_{int} ,干扰功率为 $P_i(t)$,参考功率为 $P_{\text{ref}}(t)$ (一般为最小功率值),平均功率为 P_{avg} ,波动幅度为 ΔP ,则干扰单元风险指数 $I_{\text{int}}(x,y,z,t)$ 量化为

$$I_{\text{int}}(x,y,z,t) = \begin{cases} 1 & D_{su}(t) < R_{\text{Imin}} \\ e^{-0.02(D_{su}(t) - R_{\text{Imin}})} & R_{\text{Imin}} \leq D_{su}(t) \leq R_{\text{Imax}} \\ 0 & \text{其他} \end{cases} \tag{20}$$

$$P_i(t) = P_{\text{avg}} + \Delta P \sin \left(\frac{2\pi t}{T_{\text{int}}} \right) \tag{21}$$

$$R_{\text{Imax}}(t) = R_{\text{Imin}} \sqrt{\frac{P_i(t)}{P_{\text{ref}}(t)}} \tag{22}$$

在上述单个作战单元风险量化分析的基础上,进一步探讨多个作战单元综合风险指数量化方法。设作战单元 α_i ,风险权重为 ω_i ,在 p 点风险指数为 $\kappa(p,\alpha_i)$,则该点综合风险指数 $I(x,y,z,t)$ 量化为

$$I(x,y,z,t) = \begin{cases} 1 - \prod_{i=1}^n \omega_i (1 - \kappa(p,\alpha_i)) & p \in \bigcup_{i=1}^n \alpha_i \\ 0 & p \notin \bigcup_{i=1}^m \alpha_i \end{cases} \tag{23}$$

最后,根据风险指数将战场空间划分为 3 类子空间:自由空间($I < 0.3$)、竞争空间($0.3 \leq I \leq 0.8$)、拒止空间($I > 0.8$)。在自由空间中,敌我双方无人机均可使用;在拒止空间中,己方无人机禁止使用;在竞争空间中,根据作战单元判定是否可使用^[13],满足一定条件,3 种状态空间可以相互转换。

3 仿真实验

3.1 实验设计与参数配置

硬件平台: Intel(R)Core(TM)2Quad CPU、4 GB 内存、NVIDIA GeForce 9500GT 显卡、2 GB 显存。

编译环境: Microsoft Visual Studio 2020, 基于 C++ 的 OpenSceneGraph 库二次开发。

数据环境: 在(经度 $115^\circ \sim 120^\circ$, 纬度 $20^\circ \sim 25^\circ$)区域上,分别加载 90 m 与 30 m 高程数据,以(经度

115°, 纬度 20°, 0 m) 为全局基准网格坐标系原点, 以 (经度 118°, 纬度 23.5°, 0 m) 为局部作战网格原点, 布置多源要素参数, 如表 3 所示, 其中, 最右侧一列表示 $L_{z_0}=12$ 、 $L_h=15$ 层级下, 全局基准网格坐标系下的位置坐标; 网格剖分层级信息如表 4 所示, 以此构建全局与局部双模协同场景。

表 3 多源要素参数
Table 3 Multi-source factor parameter

参数名称	位置	范围/km	坐标
A1 雷达中心	(经度 7.721°, 纬度 22.002°)	$R_{\max}=140$	(32, 24)
A2 雷达中心	(经度 118.331°, 纬度 22.832°)	$R_{\max}=120$	(39, 33)
B 火炮中心	(经度 116.010°, 纬度 23.764°)	$R_{\max}=80$ $R_{\min}=50$	(12, 45)
C 干扰中心	(经度 116.066°, 纬度 21.766°)	$R_{\max}=100$ $R_{\min}=60$	(12, 21)
D1 空域范围	(经度 118.833°, 纬度 24.083°)		(45, 48)
	(经度 119.173°, 纬度 24.417°),		(50, 53)
	(经度 118.326°, 纬度 24.750°),		(39, 57)
	(经度 118.653°, 纬度 23.665°)		(43, 43)
D2 空域范围	(经度 115.417°, 纬度 20.426°)		(5, 5)
	(经度 115.250°, 纬度 20.734°),		(3, 8)
	(经度 115.750°, 纬度 21.083°),		(9, 12)
	(经度 116.103°, 纬度 20.065°)		(13, 0)
D2 空域范围	(经度 118.833°, 纬度 24.083°)		(45, 48)
	(经度 119.1733°, 纬度 24.417°),		(50, 53)
	(经度 118.326°, 纬度 24.750°),		(39, 57)
	(经度 118.653°, 纬度 23.665°)		(43, 43)
D3 空域范围	(经度 119.417°, 纬度 20.250°)		(53, 3)
	(经度 118.750°, 纬度 20.667°),		(45, 3)
	(经度 119.583°, 纬度 21.022°),		(54, 12)
	(经度 119.461°, 纬度 20.583°)		(53, 6)
E 建筑区域	(经度 118.000° ~ 118.016°, 纬度 23.500° ~ 23.516°)		

3.2 多源要素建模与量化仿真

3.2.1 可行性分析

1) 多源要素风险量化。地形要素以数字高程模型作为基础数据源, 根据时空网格剖分架构动态调整采样间隔, 地形风险指数量化为 1.0。建筑物与管制空域要素风险量化模型中, 边界缓冲区风险指数量化为 0.3, 多边形内部风险指数量化为 1.0。

针对敌方作战单元风险量化建模, 通过动态参数耦合方法对雷达探测场、火炮杀伤区及干扰覆盖域进行量化分析。根据现有设备体系, 假设 A1 雷达扫描周期 $T_{\text{rad1}}=6\text{ s}$, A2 雷达扫描周期 $T_{\text{rad2}}=4\text{ s}$; B 火炮装填时间间隔 $T_{\text{atk}}=5\text{ s}$, 射击窗口区 $\Delta t=3\text{ s}$; C 干扰波动周期 $T_{\text{int}}=10\text{ s}$ 。图 7 为风险量化结果, 将时间 (横轴) 与空间距离 (纵轴) 解耦, 实现风险指数可视化效果, 呈现作战单元动态威胁的四维特性。图 7(a)、图 7(b) 反映不同雷达的探测距离特性; 图 7(c) 占

空比 60% 的周期威胁与反炮兵作战中的“射击-转移”战术相吻合, 风险指数在距离维度上呈梯度衰减; 图 7(d) 反映电子对抗设备的能量管理策略。

2) 多源要素时空网格建模分析。分别在表 4 所示网格层级下验证表 3 多源要素 (A1~E) 时空网格建模与风险量化效果 (网格颜色由蓝色渐变为红色表明风险等级由低到高)。如图 8(a) 所示, 全局场景覆盖经度 115°~120°、纬度 20°~25°及 0~150 km 空间范围。该场景采用经纬剖分 11 层级 (水平分辨率约为 27.8 km) 与高度剖分 14 层级 (垂直分辨率约为 10.0 km), 构建 20×20×30 的全局网格坐标系。为验证多尺度建模的有效性, 对比组实验 (如图 8(b) 所示) 将剖分精度提升至经纬层级 13 (分辨率约为 9.2 km) 和高度层级 15 (分辨率约为 1.8 km), 从而将网格密度优化为 60×60×150。

针对局部区域的精细化建模需求, 在经度 118°~118.016°、纬度 23.5°~23.516°的 0~30 m 空间范围内采集数据, 采用经纬剖分层级 19 (分辨率约为 7.7 m) 与高度剖分层级 21 (分辨率约为 1.0 m) 构建 45×45×30 的局部网格模型, 如图 9 所示。

分析结果表明, 时空网格模型能够有效适应多尺度建模需求, 并准确再现多源要素模型的几何特性。通过在复杂多尺度场景中规划无人机三维突防航迹 (图 8、图 9 中青色网格所示), 验证了该网格模型对无人机集群航迹规划的支撑能力。

3.2.2 对比性分析

为验证时空网格剖分下各要素建模的稳定性、高效性和准确性, 选取文献 [16,24] 采用的 2 个雷达探测范围叠置建模方法作为对比实验, 实验结果如表 5 所示。

设经纬层级 L_{z_0} 下的剖分体元经距为 D_{L_z} , 纬距为 D_{L_y} , 高度层级 L_h 下的剖分体元高距为 D_{L_h} 。2 个雷达探测半径分别为 R_1 、 R_2 , 剖分体元个数分别为 N_1 、 N_2 , 相交体元数为 ΔN , 相交体元用时为 t , 本方法得到雷达探测范围体积为 V_1 , 传统方法计算得到的体积为 V_2 , 文献 [24] 基于数学建模求解的方法计算得到的体积为 V_3 , 相对误差分别为 ε_1 、 ε_2 , 平均相对误差为 $\bar{\varepsilon}$, 单位相交体元用时为 t' , 计算如下:

$$\left\{ \begin{array}{l} V_1 = (N_1 + N_2) D_{L_z} D_{L_y} D_{L_h} \\ V_2 = \frac{4}{3} \pi (R_1^3 + R_2^3) \\ \varepsilon_1 = (|V_1 - V_2|) / V_2 \\ \varepsilon_2 = (|V_1 - V_3|) / V_3 \\ \bar{\varepsilon} = \frac{\varepsilon_1 + \varepsilon_2}{2} \\ t' = \frac{t}{\Delta N} \end{array} \right. \quad (24)$$

表 4 网格剖分层级信息

Table 4 Grid subdivision hierarchical information

二维		高程		范围	
层级	尺度/km	层级	尺度/km	经度范围/(°)	纬度范围/(°)
11	27.8	13	10	115.000~120.000	20.000~25.000
11	27.8	14	5	115.000~120.000	20.000~25.000
12	9.2	15	1	115.000~120.000	20.000~25.000
13	1.8	15	1	115.000~120.000	20.000~25.000
14	0.9	15	1	115.000~120.000	20.000~25.000
19	0.007 7	21	0.001	118.000~118.016	23.500~23.516

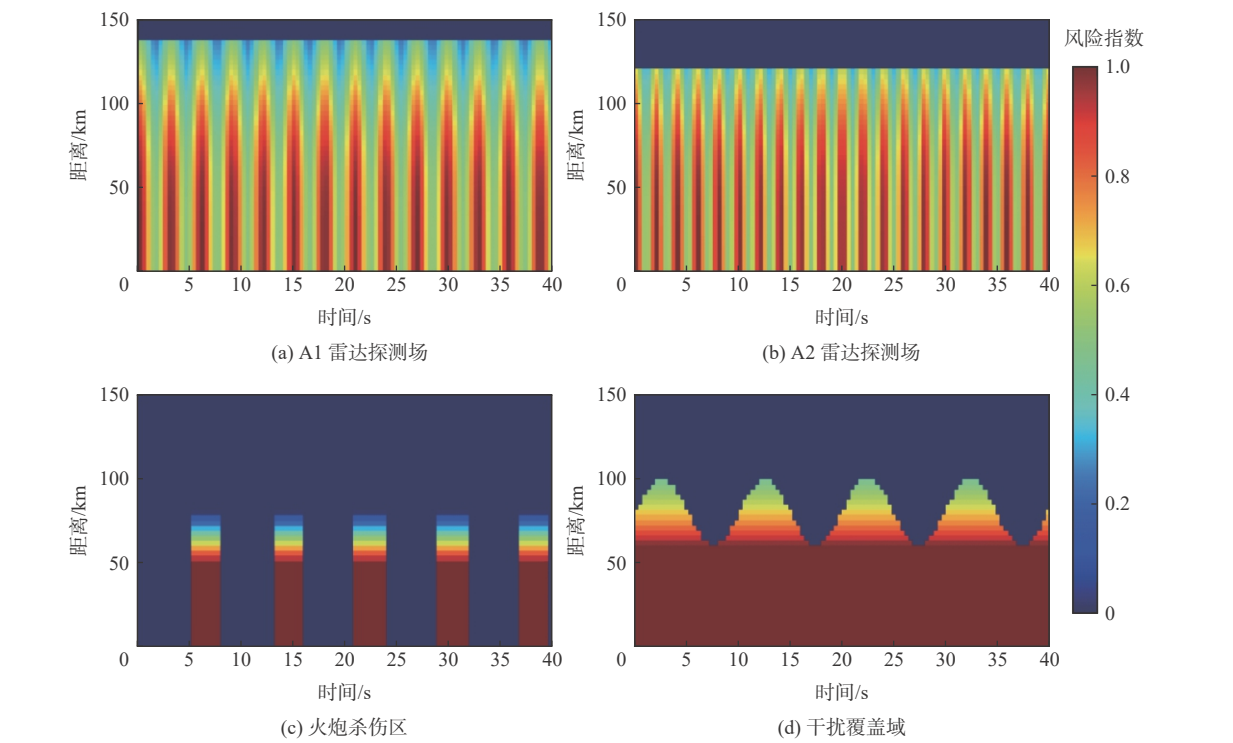


图 7 风险量化结果

Fig. 7 Risk quantification results

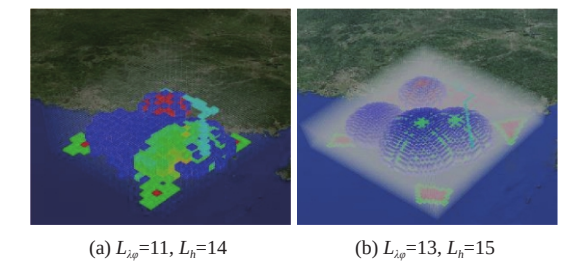


图 8 全局多尺度时空网格建模效果

Fig. 8 Global multi-scale spatio-temporal grid modeling effect

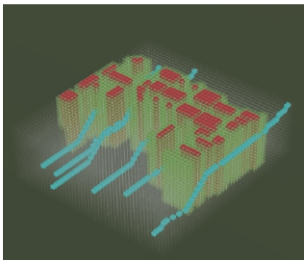


图 9 局部精细化时空网格建模效果

Fig. 9 Local fine spatio-temporal grid modeling effect

可以得出以下结论:

1) 如表 5 与图 10(a)所示,随着剖分层级增加,探测范围剖分体元个数逐级递增,相交体元个数逐级递增。

2) 如 10(b)所示,相交计算时间与相交体元个数成正比关系,说明该算法是稳定的。

3) 同等网格近似精度 1 000 m 下由文献 [16] 方法计算单位相交体元用时为 2.42×10^{-5} s, 雷达探测范围体积为 9.096×10^6 km³, 相较于传统方法计算的 9.361×10^6 km³ 相对误差为 2.831%, 相较于文献 [24] 计算的 9.354×10^6 km³ 相对误差为 2.758%, 且该方法已经较传统基于几何光学的求解雷达探测范围的方法在时间效率上提升 5 倍。而本文方法计算单位相交体元用时为 7.73×10^{-7} s, 相较于传统方法相对误差为 1.330%, 相较于文献 [24] 相对误差为 1.497%, 平均相对误差为 1.414%。在时间效率方面, 本文方法相较于文献 [16] 提升 99.65%, 相较

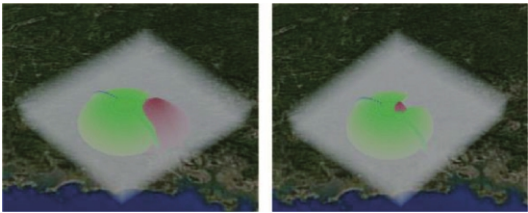
表 5 2 个雷达探测范围重叠范围计算结果

Table 5 Two radar detection range overlap ranges calculation results

层级		雷达探测范围体元数		相交体元数 ΔN	相交体元用时 t/s	V_1/km^3	V_2/km^3	V_3/km^3	相对误差 $\varepsilon_1/\%$	相对误差 $\varepsilon_2/\%$	平均相对误差 $\bar{\varepsilon}/\%$
L_{ip}	L_h	A1雷达 N_1	A2雷达 N_2								
11	13	982	672	345	0.000 5	$1.278\ 2\times 10^7$	9.361×10^6	9.354×10^6	36.545	36.647	36.596
11	14	1 944	1 344	686	0.001 0	$1.270\ 5\times 10^7$	9.361×10^6	9.354×10^6	35.723	35.824	35.774
12	15	73 756	46 576	25 777	0.020 4	$1.018\ 4\times 10^7$	9.361×10^6	9.354×10^6	8.792	8.873	8.833
13	15	1 808 612	1 142 764	583 746	0.438 9	$9.562\ 0\times 10^6$	9.361×10^6	9.354×10^6	2.147	2.223	2.185
14	15	7 186 240	4 535 048	3 457 284	2.672 6	$9.494\ 0\times 10^6$	9.361×10^6	9.354×10^6	1.330	1.497	1.414

于传统基于几何光学方法提升近 6 倍;求解雷达探测范围方面,本文方法相较于文献 [16] 相对误差平均降低 49.37%,说明本文方法满足大多数无人机集群作战应用场景的精度要求,体现了本文方法的高效性。

4) 如图 11 所示,本文方法将地形与作战单元作用域均剖分化,将传统的线面、面面相交运算转化为网格占用与否的判断,从而计算雷达波束被地形遮断时的探测范围,符合现实情况,因此,本文方法准确性更高。



(a) 雷达波束完全遮断 (b) 雷达波束部分遮断

图 11 雷达波束不同遮断情况

Fig. 11 Different blocking conditions of radar beams

4 结 论

本文提出时空网格下多源要素建模与风险量化方法,通过构建北斗网格位置码/时间码兼容的时空网格剖分架构,设计“绝对-相对”双基准时间层和“全局基准网格-局部作战网格”2级嵌套架构,建立能够动态调整网格粒度的时空基准框架,实现了战场多源要素的多尺度快速重构。本文主要工作总结如下:

1) 综合考虑 GeoSOT 在二进制存储方面的优势及北斗网格位置码在剖分规则方面的优势,利用北斗剖分时间码 5 次时间进制扩展产生的局限,提出“绝对-相对”双基准时间层,从而设计一套全新的时空网格剖分架构。

2) 针对传统最小外接矩形法在复杂地形与空域要素几何特性表征中的轮廓失真问题,提出基于射线交叉判定法的三维实景拓扑特征建模方法;同时,依据本文提出的时空网格剖分架构建立敌方作战单元统一量化模型,解决传统方法对剖分体元编码集合的求交运算及水平剖分采用“扫描线填充算法”的局限性,达到降低空间复杂度、提高运算效率的效果。

3) 多源要素建模方法具有更稳定、高效、准确的性能,如同等网格近似精度 1 000 m 下,本文方法计算单位相交体元用时为 7.73×10^{-7} s,平均相对误差为 1.414%。在时间效率方面,本文方法相较于文献 [16] 提升 99.65%,相较于传统基于几何光学提

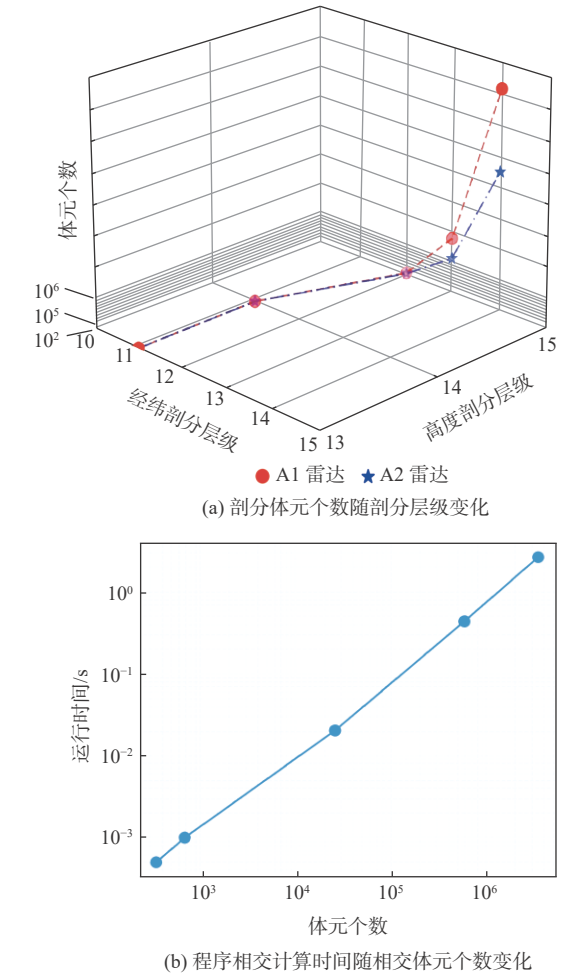


图 10 2 个雷达探测范围计算结果

Fig. 10 Calculation results of two radar detection ranges

升近6倍;求解雷达探测范围方面,本文方法相较于文献[16]相对误差平均降低49.37%。此外,多源要素建模方法充分考虑复杂地形遮蔽效应对实际作用域的影响,更贴近实际情况。

本文为智能化无人机集群作战环境建模提供了一种新范式。然而,实际战场环境具有高度的复杂性和多元性,包含的要素远超出本文所聚焦的地形、建筑群、雷达、火炮及电磁干扰等典型要素。尽管本文提出的框架能够有效表征与建模复杂战场环境及各作战域设备体系的时空固有属性,但未能充分涵盖所有战场要素(如气象、认知状态、电磁频谱等)的专有属性建模,这些专有属性往往具有高度特异性和动态性,需要针对具体要素进行深入分析与建模。

基于上述局限性与研究进展,后续工作将重点聚焦于以下方向:

1) 深化实时情报驱动的动态网格优化机制。研究构建多源异构情报的快速融合与语义理解框架,实现对战场态势变化的实时感知。在此基础上,开发网格分辨率自适应算法,能够根据情报指示的威胁等级、目标密度、环境复杂度等关键参数,动态调整网格粒度,并优化网格拓扑结构确保计算资源高效分配及模型精度与效率的实时平衡。

2) 强化动态作战单元行为预测模型。研究敌我双方各类作战单元(如机动防空系统、电子战平台、有人/无人协同编队)在复杂对抗环境下的行为模式与演化规律,建立更鲁棒、适应性更强的预测模型。

3) 构建闭环验证体系。将优化后的动态环境模型、作战单元预测模型与无人机集群的自主决策算法进行深度集成,设计高保真仿真与半实物验证环境,进行闭环测试与迭代优化,验证集群在高度动态、不确定战场环境下的整体作战效能和鲁棒性。

参考文献 (References)

[1] 赵彦杰, 袁莞迈, 梁月乾. 智能无人集群: 改变未来战争的颠覆性力量[M]. 北京: 电子工业出版社, 2024: 37-40.
Zhao Y J, Yuan (G/W)M, Liang Y Q. Intelligent unmanned swarm[M]. Beijing: Publishing House of Electronics Industry, 2024: 37-40(in Chinese).

[2] Hu X T, Pang B Z, Dai F Q, et al. Risk assessment model for UAV cost-effective path planning in urban environments[J]. IEEE Access, 2020, 8: 150162-150173.

[3] Miao S X, Cheng C Q, Zhai W X, et al. A low-altitude flight conflict detection algorithm based on a multilevel grid spatiotemporal index[J]. ISPRS International Journal of Geo-Information, 2019, 8(6): 289.

[4] Hage G, Santoso A, Sahal M. Integrated radar and missile system

with Poisson-prioritized threat management and PPN guidance for countering multiple UAV threats[J]. Journal of Robotics and Control, 2024, 6(1): 294-305.

[5] 郭星华, 朱伟, 郭锦杰. 时空剖分网格在指挥控制系统中的应用展望[J]. 指挥信息系统与技术, 2022, 13(4): 80-84.
Guo X H, Zhu W, Guo J J. Application prospect of spatiotemporal subdividing grid in command and control system[J]. Command Information System and Technology, 2022, 13(4): 80-84(in Chinese).

[6] Dong C, Zhang Y F, Jia Z Y, et al. Three-dimension collision-free trajectory planning of UAVs based on ADS-B information in low-altitude urban airspace[J]. Chinese Journal of Aeronautics, 2025, 38(2): 103170.

[7] 徐鑫宇, 万路军, 陈平, 等. 基于 GeoSOT 网格的空域栅格化表征方法[J]. 空军工程大学学报(自然科学版), 2021, 22(2): 15-22.
Xu X Y, Wan L J, Chen P, et al. An airspace raster representation method based on GeoSOT grid[J]. Journal of Air Force Engineering University (Natural Science Edition), 2021, 22(2): 15-22(in Chinese).

[8] 国家市场监督管理总局, 国家标准化管理委员会. 地球空间网格编码规则: GB/T 40087—2021[S]. 北京: 中国标准出版社, 2021.
State Administration for Market Regulation, Standardization Administration of the People's Republic of China. Geospatial grid encoding rule: GB/T 40087—2021[S]. Beijing: Standards Press of China, 2021(in Chinese).

[9] 国家市场监督管理总局, 国家标准化管理委员会. 北斗网格位置码: GB/T 39409—2020[S]. 北京: 中国标准出版社, 2020.
State Administration for Market Regulation, Standardization Administration of the People's Republic of China. BeiDou grid location code: GB/T 39409—2020[S]. Beijing: Standards Press of China, 2020(in Chinese).

[10] 国家市场监督管理总局, 国家标准化管理委员会. 北斗剖分时间码: GB/T 42578—2023[S]. 北京: 中国标准出版社, 2023.
State Administration for Market Regulation, Standardization Administration of the People's Republic of China. BeiDou subdivision time code: GB/T 42578—2023[S]. Beijing: Standards Press of China, 2023(in Chinese).

[11] 山海涛, 程承旗, 陈波. 网格环境下战场环境信息集成服务策略[J]. 测绘科学, 2020, 45(1): 189-193.
Shan H T, Cheng C Q, Chen B. Integrated service strategy of battlefield environment information ingrid environment[J]. Science of Surveying and Mapping, 2020, 45(1): 189-193(in Chinese).

[12] Deng C, Cheng C Q, Qu T T, et al. A method for managing ADS-B data based on a 4D airspace-temporal grid (GeoSOT-AS)[J]. Aerospace, 2023, 10(3): 217.

[13] 季荣涛. 基于威胁分析的战场空间划分及其在航迹规划中的应用[D]. 南京: 南京大学, 2016: 27-34.
Ji R T. The partition of battlefield base on threat analysis and its application in route planning[D]. Nanjing: Nanjing University, 2016: 27-34(in Chinese).

[14] 吴乾坤. 基于全空间剖分的战场空间构建和高性能空间计算研究[D]. 北京: 中国电子科技集团公司电子科学研究院, 2022: 17-26.
Wu Q K. Researched on battlefield construction and high-performance space computing based on full space subdivision[D]. Beijing: China Academy of Electronic and Information Technology, 2022:

17-26(in Chinese).

[15] 李锋, 万刚, 蒋秉川, 等. 地空一体化战场时空数据组织模型及应用[J]. 系统仿真学报, 2017, 29(11): 2637-2643.

Li F, Wan G, Jiang B C, et al. Spatio-temporal data organization model of unified ground and aerial battlefield and its application[J]. Journal of System Simulation, 2017, 29(11): 2637-2643(in Chinese).

[16] 袁越, 程承旗, 童晓冲. 基于剖分表达结构的雷达探测范围计算方法[J]. 地理信息世界, 2017, 24(6): 29-36.

Yuan Y, Cheng C Q, Tong X C. Calculation method of radar detection range based on subdivision expression structure[J]. Geomatics World, 2017, 24(6): 29-36(in Chinese).

[17] 周畅. 面向动态威胁的无人机多目标三维航迹规划[D]. 大连: 大连理工大学, 2021: 12-15.

Zhou C. Multi-objective 3D path planning for UAV under mobile threats[D]. Dalian: Dalian University of Technology, 2021: 12-15(in Chinese).

[18] Han B, Qu T T, Tong X C, et al. Grid-optimized UAV indoor path planning algorithms in a complex environment[J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 111: 102857.

[19] 孙忠秋, 程承旗. GeoSOT-3D 椭球体剖分真三维数据表达[J]. 地理信息世界, 2016, 23(3): 40-46.

Sun Z Q, Cheng C Q. True 3D data expression based on GeoSOT-3D ellipsoid subdivision[J]. Geomatics World, 2016, 23(3): 40-46(in Chinese).

[20] 张晨, 赵青松, 陆志沣, 等. 战场态势要素构成及影响关系研究[J]. 舰船电子工程, 2023, 43(3): 33-35.

Zhang C, Zhao Q S, Lu Z F, et al. Research on composition and influence of battlefield situation elements[J]. Ship Electronic Engineering, 2023, 43(3): 33-35(in Chinese).

[21] 王孟阳, 张栋, 唐硕, 等. 复杂动态环境下多无人机目标跟踪的分布式协同轨迹规划方法[J]. 指挥与控制学报, 2024, 10(2): 197-212.

Wang M Y, Zhang D, Tang S, et al. A distributed collaborative trajectory planning method for multi-UAV targets tracking in complex dynamic environment[J]. Journal of Command and Control, 2024, 10(2): 197-212(in Chinese).

[22] Hu H, Liu M, Zhong J Q, et al. A case study of the 3D water vapor tomography model based on a fast voxel traversal algorithm for ray tracing[J]. Remote Sensing, 2021, 13(12): 2422.

[23] 屠壮, 杨柳庆, 林志坚, 等. 城市作战背景下多无人机航路规划仿真研究[J]. 舰船电子工程, 2024, 44(9): 48-54.

Tu Z, Yang L Q, Lin Z J, et al. Simulation research on multi-UAV route planning under the background of urban combat[J]. Ship Electronic Engineering, 2024, 44(9): 48-54(in Chinese).

[24] 许东来, 戴树荪, 彭学愚. 双(多)基地雷达的探测范围和定位精度的分析[J]. 西安电子科技大学学报, 1993, 20(2): 87-99.

Xu D L, Dai S S, Peng X Y. The analysis of detection coverage and locating precision of the bistatic and multistatic radar[J]. Journal of Xidian University, 1993, 20(2): 87-99(in Chinese).

Multi-source element modeling and risk quantitative analysis under spatio-temporal grid

LIU Longhao^{1, 2, 3}, RU Le^{1, 2, 3, *}, WANG Hongqiao^{3, 4}, WANG Wenfei^{1, 2, 3}, ZHANG Zhenghao^{1, 2, 3}, LI Yifan⁵

- (1. Equipment Management and Unmanned Aerial Vehicle Engineering School, Air Force Engineering University, Xi'an 710051, China;
- 2. National Key Laboratory of Unmanned Aerial Vehicle Technology, Xi'an 710051, China;
- 3. The Youth Innovation Team of Shaanxi University, Xi'an 710051, China;
- 4. Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, China;
- 5. Unit 94754, Chinese People's Liberation Army, Jiaxing 314011, China)

Abstract: This work suggests a multi-source element modeling and risk quantification approach based on spatio-temporal grid subdivision to tackle the problem of spatio-temporal coupling modeling for multi-source elements in complex battlefield scenarios. Firstly, in the spatial dimension: A novel grid subdivision architecture compatible with the BeiDou grid location code (BDGC) is designed to significantly improve spatial data storage and computational efficiency in order to address the storage redundancy resulting from the triple virtual extension in current global subdivision models such as geographic coordinate subdividing grid with one dimension integral coding on 2ⁿ-tree(GeoSOT) and the efficiency gap issue associated with the “letter + decimal digit” encoding scheme of the BDGC. In the temporal dimension: To address the limitations caused by the quintuple temporal base extensions in the BeiDou subdivision time code (BDTC), an innovative dual-baseline temporal encoding system with “absolute-relative” layers is proposed. This successfully overcomes the spatio-temporal fusion bottleneck for heterogeneous elements by achieving a four-dimensional joint representation that integrates geographical position and time. Secondly, a two-level nested spatio-temporal reference framework comprising a “global reference grid - local combat grid” is constructed. This architecture balances local adaptability with global consistency by supporting the dynamic change of grid granularity based on equipment systems or the battlefield environment across several operational domains. Finally, to resolve the contour distortion problem inherent in traditional minimum bounding rectangle (MBR) methods when characterizing the geometric features of complex terrain and airspace elements, a three-dimensional realistic scene topological feature modeling method based on the ray-intersection judgment algorithm is proposed. Leveraging the spatio-temporal grid subdivision architecture, a unified quantitative model for enemy operational units is constructed. Combined with risk levels, the battlefield space is partitioned into free space, contested space, and denied space. Simulation experiments demonstrate that: The proposed method enables rapid multi-scale reconstruction of battlefield multi-source elements; At an approximate grid resolution of 1 000 meters, the average relative error for spatial representation is only 1.414%; Compared to traditional methods, the average relative error is reduced by 49.37% and time efficiency is improved by 99.65%; It can meet the real-time response requirements for UAV swarm trajectory planning. This work provides a validated new paradigm for intelligent environmental modeling in UAV swarm operations.

Keywords: spatio-temporal grid; BeiDou grid code; battlefield elements; risk quantification; unmanned aerial vehicle cluster

《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人,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、计算机科学及应用技术、可靠性工程与失效分析等领域的研究文章。航空航天科学技术民用方面以及具有航空航天工程背景的应用数学、应用物理、应用力学和工程管理等方面的文章也在本刊优先考虑之列。

2 来稿要求

2.1 论文应具有创新性、科学性、学术性和可读性。

2.2 论文为原创作品,尚未公开发表过,并且不涉及泄密问题。若发生侵权或泄密问题,一切责任由作者承担。

2.3 主题明确,数据可靠,图表清晰,逻辑严谨,文字精练,公式符号标准规范,标点符号正确。

2.4 文稿撰写顺序:中文题名(一般不超过 20 个汉字),作者中文姓名、单位、所在城市、邮政编码(作者顺序始终不得更改),中文摘要(包括目的、方法、结果及结论),中文关键词(5~8 个),中图分类号,引言,正文,参考文献,英文题名,作者英文姓名、单位、所在城市、邮政编码、国别,英文摘要,英文关键词。首、末页下角注明基金项目名称及编号,作者信息。

2.5 作者请登录本刊网页进行在线投稿。

3 稿件的审核、录用与版权

3.1 来稿须经专家两审和主编、编委讨论后决定刊用与否。

3.2 若来稿经审查后认定不宜在本刊发表,将及时告知作者。如果在投稿满 3 个月后仍未收到本刊任何通知,作者有权改投其他刊。在此之前,请勿一稿多投,否则一切后果自负。

3.3 来稿一经作者签字并在本刊刊出,即表明所有作者都已经认可其版权转至本刊编辑部。本刊在与国内外文献数据库或检索系统进行交流及合作时,不再征询作者意见。

邮寄地址:100191 北京市海淀区学院路 37 号 《北京航空航天大学学报》编辑部

办公地点:北京航空航天大学办公楼西配楼 412~416 房间

电 话:(010)82316698,82316981,82315499,82317448

E-mail: jbuaa@buaa.edu.cn

http://bhxb.buaa.edu.cn

http://www.buaa.edu.cn



集群成员期刊



《北京航空航天大学学报》

第五届编辑委员会

主 任（主 编）：赵沁平

（以下按姓氏笔画为序）

副主任（副主编）：丁希仑 王少萍 孙志梅 李秋实 李焕喜 杨嘉陵

苗俊刚 相 艳 徐立军 钱德沛 曹晋滨

编 委：马殿富 王 琪 王 聪 邓小燕 王青云 王荣明 刘 宇

刘 红 江 洁 刘 强 闫 鹏 朱天乐 刘铁钢 齐铂金

陈万春 邹正平 苏东林 杨世春 沈成平 邱志平 宋知人

杨树斌 张晓林 杨晓奕 杨继萍 李惠峰 吴新开 张瑞丰

杨照华 宋凝芳 周 锐 林宇震 林贵平 战 强 姚仰平

胡庆雷 赵秋红 段海滨 赵巍胜 席 平 郭 宏 徐 洁

徐世杰 郭洪波 康 锐 翟 锦 熊华钢

编辑部主任：张 嵘

编 辑：孙 芳 万玉敏 王 茜 卞欢欢 李亚泰

北京航空航天大学学报

Beijing Hangkong Hangtian Daxue Xuebao

（原《北京航空学院学报》）

（月刊 1956 年创刊）

第52卷 第9期 2026年9月

JOURNAL OF BEIJING UNIVERSITY OF

AERONAUTICS AND ASTRONAUTICS

(JBUA)A

(Monthly, Started in 1956)

Vol.52 No.9 September 2026

主管单位 中华人民共和国工业和信息化部

主办单位 北京航空航天大学

主 编 赵沁平

编辑出版 《北京航空航天大学学报》编辑部

邮 编 100191

地 址 北京市海淀区学院路37号

印 刷 北京科信印刷有限公司

发 行 《北京航空航天大学学报》编辑部

发行范围 国内外发行

联系电话 (010) 82316698 82316981

82315499 82317448

电子信箱 jbuua@ 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 JBUA

Printed by Beijing Kexin Printing Co., Ltd.

Distributed by Editorial Board of JBUA

Telephone (010) 82316698 82316981

82315499 82317448

E-mail jbuua@ 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



微信公众号



超星电子书

ISSN 1001-5965



国内定价：50.00 元 / 期