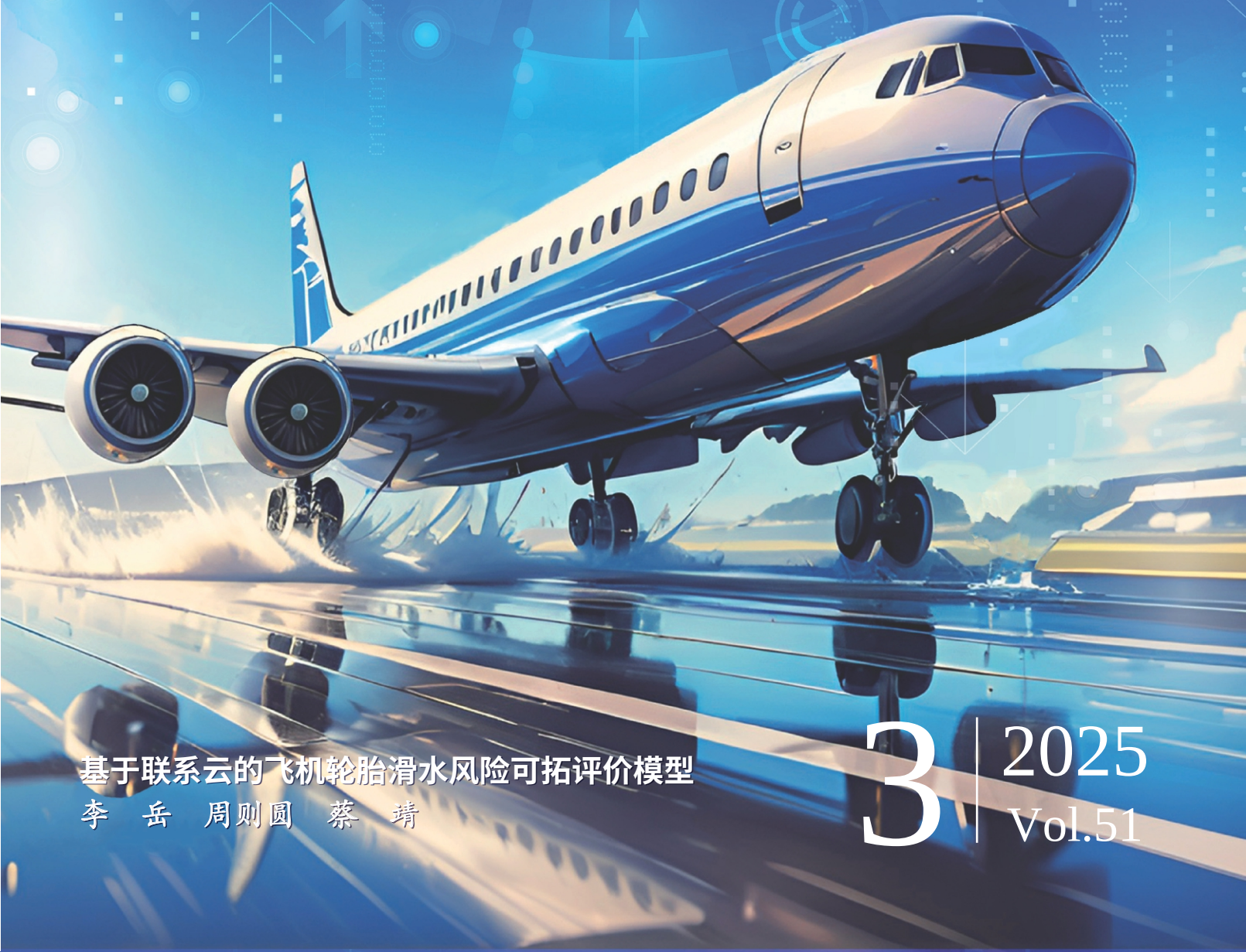
主管单位：中华人民共和国工业和信息化部
主办单位：北京航空航天大学

ISSN 1001-5965
CN 11-2625/V

北京航空航天大學学报

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS



基于联系云的飞机轮胎滑水风险可拓评价模型

李　岳　周则圆　蔡　靖

3 | 2025
Vol.51

北京航空航天大学学报

第 51 卷　第 3 期　（总第 385 期）　2025 年 3 月

目　　次

基于联系云的飞机轮胎滑水风险可拓评价模型 …… 李岳，周则圆，蔡靖（705）

融合 Mobile Vit 和倒置门控编解码的视网膜血管分割算法 ……
…… 梁礼明，阳渊，朱晨锟，何安军，吴健（712）

改进 5G-R 自适应高速铁路越区切换算法 …… 陈永，康婕，陶瑄（724）

基于双路径监督的遥感图像语义分割网络 … 刘春娟，乔泽，闫浩文，吴小所，王嘉伟，辛钰强（732）

基于等边 Bennett 机构的变形翼机构设计与分析 ……
…… 田大可，张珺威，金路，刘荣强，雷宏强，崔锡赫（742）

含局部缺陷的角接触球轴承时变位移激励及动力学建模 ……
…… 雷春丽，宋瑞哲，樊高峰，刘凯，薛伟，李建华（753）

基于改进 YOLOv5s 的动态视觉 SLAM 算法 …… 蒋畅江，刘朋，舒鹏（763）

超声速进气道出口封闭的非定常流动特性研究 …… 温玉芬，张炜群，郝思思（772）

基于邻域点特征分割与对应的航空发动机叶片损伤点云配准 …… 蔡舒妤，郝峰伟，史涛（784）

基于 MC-GERT 的舰载机出动回收作业流程时间不确定性分析 ……
…… 郭放，韩维，刘玉杰，苏析超，白天，刘纯（795）

城市低空场景下无人机运行对地风险量化评估 …… 陈艺君，余莎莎，张学军（806）

充气式返回舱迎风面真实气体效应 …… 何青松，孙素蓉（816）

固化应力松弛对复材模具型面精度的影响 …… 肖遥，李勇，李东升，王磊，江超（824）

基于语义关系的单视图隐式场三维重建技术 …… 徐岗，靳炯超，唐泽皓，冷彪（833）

矩形微织构浮环气膜密封特性的方向效应 …… 王世鹏，丁雪兴，力宁，丁俊华，张兰霞（845）

丝网印刷石墨烯标签天线的不确定性分析 …… 李凯，沈志刚，张晓静（857）

离子推力器外壳非预期放电抑制及热特性 …… 李建鹏，靳伍银，赵以德，代鹏，张兴民（865）

THS 功率回流问题分析及其解决方案 …… 刘学武，徐向阳，黄河，赵江灵，李凯峰，董鹏（874）

超临界压力碳氢燃料在增材制造通道中流动传热的实验研究 ……
…… 蔡磊，肖颖，韩怀志，于瑞天，罗文（881）

融合注意力和多尺度特征的航空发动机缺陷检测 ……
…… 赵崇林，朱江，胡永进，李祖泽，王鹏举，谢涛（892）

多约束下升力式再入飞行器可达域计算方法 …… 冉云霆，泮斌峰（904）

多谱自适应小波和盲源分离耦合的生理信号降噪方法 … 王振宇，向泽锐，支锦亦，丁铁成，邹瑞（910）

新型复合式构型直升机气动干扰特性分析 …… 梁家辉，张夏阳，招启军（922）

基于位置约束与注意力的低空无人机障碍物检测方法 …… 唐友军，缪存孝，张贺，李玉峰，叶文（933）
基于学习算法的结构大变形预测及气动弹性分析 …… 陈乔，安朝，谢长川，杨超（943）
基于监督对比学习的无线电引信干扰识别方法 …… 钱鹏飞，秦高林，陈齐乐，郝新红（953）
轴压铝合金蜂窝圆筒整体稳定性研究 …… 魏家睿，吴琼，吴伟，李群智，时红伟，张虔（962）
基于深浅层特征融合的孪生网络目标跟踪算法 …… 王子康，姚文进，薛尚捷，司婷波（973）
单离子通道神经元电路的动力学分析方法 …… 杨梓悦，吴静，李乐桐（985）
基于改进 NSGA-Ⅱ算法的不规则集装器装载优化 …… 文淦，袁雷峰，王旭东，毕军，王永兴（992）
前倾转变体无人机固定时间轨迹跟踪控制 …… 窦磊，李新凯，张宏立，王昊，杨加秀（1005）
基于多轮次分布式拍卖的异构多任务分配算法 …… 吕晔，周锐，李兴，刘志恒，邸斌（1018）
基于注意力机制的 GRU-IKF 场面滑行轨迹预测模型 …… 刘雨生，汤新民，任宣铭（1028）

期刊基本参数：CN 11-2625/V * 1956 * m * A4 * 332 * zh * P * ￥50.00 * 400 * 33 * 2025-03

（**编　辑**　张　嵘　孙　芳　万玉敏　卞欢欢　王　茜　李亚泰）

JOURNAL OF BEIJING UNIVERSITY OF
AERONAUTICS AND ASTRONAUTICS

Vol. 51 No. 3 (Sum 385) March 2025

CONTENTS

Extensible evaluation model of aircraft tire hydroplaning risk based on connection cloud
…… LI Yue, ZHOU Zeyuan, CAI Jing (705)

Fusion of Mobile Vit and inverted gated codec retinal vessel segmentation algorithm
…… LIANG Liming, YANG Yuan, ZHU Chenkun, HE Anjun, WU Jian (712)

An improved 5G-R adaptive high-speed railway handover algorithm
…… CHEN Yong, KANG Jie, TAO Xuan (724)

Semantic segmentation network of remote sensing images based on dual path supervision
…… LIU Chunjuan, QIAO Ze, YAN Haowen, WU Xiaosuo, WANG Jiawei, XIN Yuqiang (732)

Design and analysis of morphing wing mechanism based on equilateral Bennett mechanism
…… TIAN Dake, ZHANG Junwei, JIN Lu, LIU Rongqiang, LEI Hongqiang, CUI Xihe (742)

Time-varying displacement excitation and dynamic modeling of local defects in angular contact ball bearings
…… LEI Chunli, SONG Ruizhe, FAN Gaofeng, LIU Kai, XUE Wei, LI Jianhua (753)

Dynamic visual SLAM algorithm based on improved YOLOv5s
…… JIANG Changjiang, LIU Peng, SHU Peng (763)

Investigation on unsteady flow characteristics of a supersonic inlet with exit blocked
…… WEN Yufen, ZHANG Weiqun, HAO Sisi (772)

Partition based on features of neighborhood points and corresponding point cloud registration of aero-engine damaged blade
…… CAI Shuyu, HAO Fengwei, SHI Tao (784)

Time uncertainty analysis on cyclic operation procedures of carrier aircraft based on MC-GERT
…… GUO Fang, HAN Wei, LIU Yujie, SU Xichao, BAI Tian, LIU Chun (795)

Ground risk quantitative assessment for UAV operations in urban low-altitude scenarios
…… CHEN Yijun, YU Shasha, ZHANG Xuejun (806)

Real gas effect of inflatable reentry decelerator on windward side
…… HE Qingsong, SUN Surong (816)

Influence of curing stress relaxation on profile accuracy of composites tools
…… XIAO Yao, LI Yong, LI Dongsheng, WANG Lei, JIANG Chao (824)

Semantic part based single-view implicit field for 3D shape reconstruction technology
…… XU Gang, JIN Jiongchao, TANG Zehao, LENG Biao (833)

Orientation effect on sealing characteristics of rectangular micro-textured floating ring gas film
…… WANG Shipeng, DING Xuexing, LI Ning, DING Junhua, ZHANG Lanxia (845)

Study on uncertainties of graphene tag antenna by screen printing
…… LI Kai, SHEN Zhigang, ZHANG Xiaojing (857)

Unexpected electric breakdown control and thermal characteristics of ion thruster shell
…… LI Jianpeng, JIN Wuyin, ZHAO Yide, DAI Peng, ZHANG Xingmin (865)

Analysis of power reflux problem and its solution of THS
…… LIU Xuewu, XU Xiangyang, HUANG He, ZHAO Jiangling, LI Kaifeng, DONG Peng (874)

Experimental study on flow and heat transfer of hydrocarbon fuels in additive manufacturing channels at supercritical pressure
…… CAI Lei, XIAO Ying, HAN Huaizhi, YU Ruitian, LUO Wen (881)

Aero-engine defect detection by fusing attention and multi-scale features
…… ZHAO Chonglin, ZHU Jiang, HU Yongjin, LI Zuze, WANG Pengju, XIE Tao (892)

A method for calculating reachability regions of lift re-entry vehicles with multiple constraints
…… RAN Yunting, PAN Binfeng (904)

Physiological signal denoising method based on multi-spectrum adaptive wavelet and blind source separation
…… WANG Zhenyu, XIANG Zerui, ZHI Jinyi, DING Tiecheng, ZOU Rui (910)

Aerodynamic interference characteristics of a new compound configuration helicopter
…… LIANG Jiahui，ZHANG Xiayang，ZHAO Qijun （922）

A low-altitude UAV obstacle detection method based on position constraint and attention mechanism
…… TANG Youjun，MIAO Cunxiao，ZHANG He，LI Yufeng，YE Wen （933）

Large deformation prediction and geometric nonlinear aeroelastic analysis based on machine learning algorithm
…… CHEN Qiao，AN Chao，XIE Changchuan，YANG Chao （943）

A recognition method of radio fuze signal based on supervised contrastive learning
…… QIAN Pengfei，QIN Gaolin，CHEN Qile，HAO Xinhong （953）

Study on global stability of aluminum alloy honeycomb cylinder under axial compression
…… WEI Jiarui，WU Qiong，WU Wei，LI Qunzhi，SHI Hongwei，ZHANG Yu （962）

Siamese network object tracking algorithm based on deep and shallow feature fusion
…… WANG Zikang，YAO Wenjin，XUE Shangjie，SI Tingbo （973）

Dynamic analysis method of single ion channel neuron circuit
…… YANG Ziyue，WU Jing，LI Letong （985）

Loading optimization of irregular unit load device based on improved NSGA-Ⅱ algorithm
…… WEN Gan，YUAN Leifeng，WANG Xudong，BI Jun，WANG Yongxing （992）

Fixed time trajectory tracking control of forward-tilting morphing aerospace vehicle
…… DOU Lei，LI Xinkai，ZHANG Hongli，WANG Hao，YANG Jiaxiu （1005）

A heterogeneous multi-task assignment algorithm based on multi-round distributed auction
…… LYU Ye，ZHOU Rui，LI Xing，LIU Zhiheng，DI Bin （1018）

Prediction of aircraft surface trajectory based on the GRU-IKF model with attention mechanism
…… LIU Yusheng，TANG Xinmin，REN Xuanming （1028）

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0136

基于联系云的飞机轮胎滑水风险可拓评价模型

李岳[1]，周则圆[1]，蔡靖[1, 2, *]

(1. 中国民航大学 交通科学与工程学院，天津 300300；2. 民航机场智能建造与工业化工程技术研究中心，天津 300456)

摘 要：针对飞机轮胎滑水问题影响因素多、评价指标具备随机模糊与离散性特征问题，基于联系云可拓理论建立滑水风险评价模型，定量描述滑水评价指标在不同分类等级之间的转换态势。根据评价指标分级标准确定联系云数字特征，生成有限区间内联系云，由云相关度构建联系云可拓矩阵，结合变权权重综合评定最终风险等级，表征待评价物元与风险等级之间的动态联系；依托飞机轮胎滑水流固耦合仿真获得案例分析数据，弥补经典滑水试验工况变量条件少的不足。分析结果表明：传统正态云模型与可拓联系云模型对滑水事件样本1和样本3评价结论一致；对于样本2，由所提方法计算滑水风险等级为Ⅲ级，同等参数条件下风险控制标准更严格，上述样本风险评价置信因子均小于0.01，评价可信程度较高；所提分析模型为多重不相容指标的随机模糊及不确定性分析提供了可能，克服了正态云模型无法有效模拟指标在有限区间内分布的缺陷。

关 键 词：轮胎滑水；联系云；可拓理论；流固耦合；风险评价
中图分类号：U416.217; V351.11
文献标志码：A **文章编号**：1001-5965(2025)03-0705-07

据统计，近10年内由飞机轮胎滑水(aircraft tire hydroplaning, ATH)引起的各类冲偏出跑道事故共计101起，对社会公众安全产生显著负面影响[1]。相较全球年均6 900万架次起降量而言，ATH事件虽然发生几率较低，但关联损失严重，针对此类航空安全问题的研究一直备受重视。

依照ATH形成机理与防范方法，以往研究可分为2个阶段。第1阶段以美国国家航空航天局(National Aeronautics and Space Administration, NASA)提出的滑水速度公式为代表，以临界滑水速度作为ATH发生判据，将轮胎磨损[2]、道面平整度[3]、附着系数[4]、道面刻槽[5]、轮组构型[5]、机身侧滑角[6]与侧风扰动[7]等因素纳入分析考量，拓宽了NASA公式的适用条件。参照英国Bristol大学轮胎溅水喷流试验，文献[8]以道面积水附加阻力作为ATH补充判据，该指标在滑水过后迅速降低并持续影响飞机滑行姿态。上述研究的不足之处是将ATH视为确定性事件，与事故统计特征不符合。第2阶段将公路车辆滑水分析理论引入航空领域，提出基于全概率的ATH风险量化公式[6]，覆盖飞机高速滑行全过程，避免仅以最大速度瞬时为最不利分析工况。文献[9]将机载快速存取记录器(quick access record, QAR)数据引入滑水风险分析，以表征特定机型在特定道面条件下的起降特征，促进ATH研究从规律化向精细化方向转变，克服第1阶段“二元”判定结论较为粗糙的不足(即滑水或不滑水)。

近年来，云模型在解决空管运行保障、飞机重着陆风险等复杂航空随机模糊问题中有广泛的应用[10-13]。胡宇祺等[14]探讨了基于正态云模型的飞机滑水安全评价方法，提取单因素云模型数字特征，以综合隶属度指标表征滑水风险等级。在此基础上，本文建立了基于联系云的ATH风险可拓评

收稿日期：2023-03-23；**录用日期**：2023-05-29；**网络出版时间**：2023-06-15 11:19
网络出版地址：link.cnki.net/urlid/11.2625.V.20230614.1459.002
基金项目：国家自然科学基金(52472369)；中央高校基本科研业务费专项资金(3122019103)；民航机场智能建造与工业化工程技术研究中心开放课题(MHJGKFKT-01,MHJGKFKT-04)

* **通信作者**. E-mail：jcai@cauc.edu.cn

引用格式：李岳，周则圆，蔡靖. 基于联系云的飞机轮胎滑水风险可拓评价模型[J]. 北京航空航天大学学报，2025，51(3)：705-711.
LI Y，ZHOU Z Y，CAI J. Extensible evaluation model of aircraft tire hydroplaning risk based on connection cloud[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：705-711 (in Chinese).

价模型，依据ATH指标分级标准确定联系云数字特征，生成有限区间内联系云，由云相关度构建联系云可拓矩阵，结合变权权重综合评定ATH风险等级。开展ATH流固耦合仿真分析，丰富研究案例数据，为ATH风险评价提供新的参考。

1 联系云可拓理论

1.1 联系云基本原理

云模型以概率论和模糊数学为基础，以自然语言值为切入点，构建定性定量转换模型，是解决多层次多指标评价问题的分析工具[15-18]。

对于简单随机变量，假定其在理想无限区间内服从正态分布，采用正态云模型描述数字特征，这与常见滑水评价指标（飞机轴载等级、地面滑行速度、道面积水厚度等）的有限区间分布形式存在差异。对此，汪明武教授等提出了联系云模型，可在有限区间内生成非对称分布云模型，为改进ATH风险评价提供了理论支撑[19]。

假定将ATH风险划分 p 个等级（i=1,2,⋯,p），评价指标 q 个（j=1,2,⋯,q），评价指标 j 等级为 i 的云由左右2朵非对称有限区间分布云构成，以期望值 E_x^i 为中间分界点。云滴是云模型的基本单元，单个云滴隶属度 μ_i 计算式为

$$\mu_i = \left[1 - \left(\frac{x_i - E_x^i}{\alpha_i'}\right)^2\right]^{\lambda_i} \tag{1}$$

$$\alpha_i' = E_{\mathrm{n}}^{i'}\sqrt{2\lambda_i + 3} \tag{2}$$

$$E_x^i = \frac{L_{\max}^i + L_{\min}^i}{2} \tag{3}$$

$$E_{\mathrm{n}}^i = \frac{\alpha_i}{\sqrt{2\lambda_i + 3}} \tag{4}$$

$$H_{\mathrm{e}}^i = k \tag{5}$$

式中：$\mu_i \in [0,1]$；E_{n}^i 和 H_{e}^i 分别为等级 i 左右半朵联系云的熵和超熵，E_{n}^i 越大，云模型分布越离散，H_{e}^i 表征随机性与模糊性相关度；α_i、α_i' 分别为修正前、后半区间长度；对左半区间，$x_i \in \left[E_x^i - \alpha_i', E_x^i\right]$，对右半区间，$x_i \in \left[E_x^i, E_x^i + \alpha_i'\right]$；$E_{\mathrm{n}}^{i'} \sim \mathrm{Normrnd}\left(E_{\mathrm{n}}^i, H_{\mathrm{e}}^{i2}\right)$；$L_{\max}^i$、$L_{\min}^i$ 分别为等级 i 区间上、下限；常数 k 反映云漂浮模糊阈值，本文中取0.01，k 值越小，表明联系云模型稳定性越好[20]；λ_i 为 α_i' 对应分布密度函数阶数，由等级 i 区间长度、α_i 和 α_i' 三者共同决定，计算式为

$$\lambda_i = \frac{\lg 0.5}{\lg\left[1 - \left(\frac{l_i - E_x^i}{\alpha_i}\right)^2\right]} \tag{6}$$

其中：l_i 为 $L_{\max}^i$ 或 $L_{\min}^i$。

联系云模型的基本数字特征如图1所示。图中：$\alpha_{\mathrm{left}} = 30$，$\alpha_{\mathrm{right}} = 32$，$E_x = 35$，$H_{\mathrm{e}} = 0.01$，$N$=1 000。可以看出，依据“$3E_{\mathrm{n}}$ 规则”，传统正态云模型云滴主要分布在 $[E_x - 3E_{\mathrm{n}}, E_x + 3E_{\mathrm{n}}]$ 范围内，客观上忽视了分类等级的模糊性和被测指标分布区间有限性。而联系云模型由左右2朵非对称独立云组成，图1中评价指标 j 与等级 i 存在3种集对关系。当实测指标值位于等级 i 区间内时，为同一关系，位于相邻等级 i+1 或 i−1 时，为差异关系，位于相隔等级时，符合对立关系。半区间长度 α_{left} 和 α_{right} 由同一关系和差异关系所包含的区间构成。

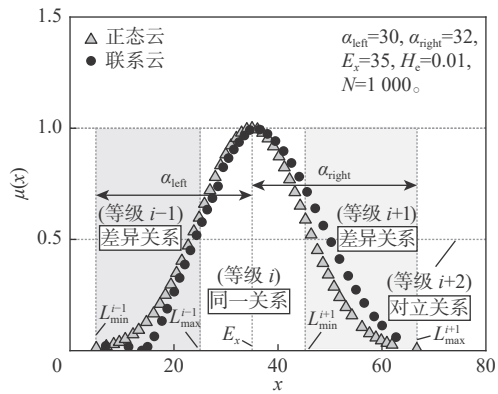


图1 联系云模型数字特征

Fig. 1 Numerical characteristic of connection cloud model

1.2 可拓理论物元表达

可拓理论是研究元素内部结构与事物可变性的不确定性分析方法[20-22]，利用物元可拓性，通过有序三元组表达事物的量与质，公式为

$$\boldsymbol{R} = \begin{bmatrix} N & C_1 & V_1 \\ N & C_2 & V_2 \\ \vdots & \vdots & \vdots \\ N & C_q & V_q \end{bmatrix} \tag{7}$$

式中：物元 $\boldsymbol{R}$ 由事物名称 N、事物特征 C 和事物特征量值 V 共同表示。特征量值 V 为确定数值或数值区间。

为构建可还原实际ATH复杂风险量化分析系统，将联系云引入可拓理论物元表达，以联系云模型数字特征替换事物特征量值 V，基于联系云的可拓物元形式如下：

$$\boldsymbol{R}_i = \begin{bmatrix} N & C_1 & \left(E_x^1, E_{\mathrm{n}}^1, H_{\mathrm{e}}^1, \alpha_1, \lambda_1\right) \\ N & C_2 & \left(E_x^2, E_{\mathrm{n}}^2, H_{\mathrm{e}}^2, \alpha_2, \lambda_2\right) \\ \vdots & \vdots & \vdots \\ N & C_q & \left(E_x^q, E_{\mathrm{n}}^q, H_{\mathrm{e}}^q, \alpha_q, \lambda_q\right) \end{bmatrix} \tag{8}$$

式中：$\left(E_x^j, E_{\mathrm{n}}^j, H_{\mathrm{e}}^j, \alpha_j, \lambda_j\right)$ 为 N 关于评价指标 C_j 的联系云表达。通过联系云解决ATH风险等级划分与边界处理问题，利用可拓理论定量分析事物特征 C_j 间不相容问题，实现有限区间分布评价指标的随

机模糊分析。

2　飞机轮胎滑水评价模型

2.1　基本评价流程

图 2 为基于联系云可拓理论的 ATH 风险评价基本流程。

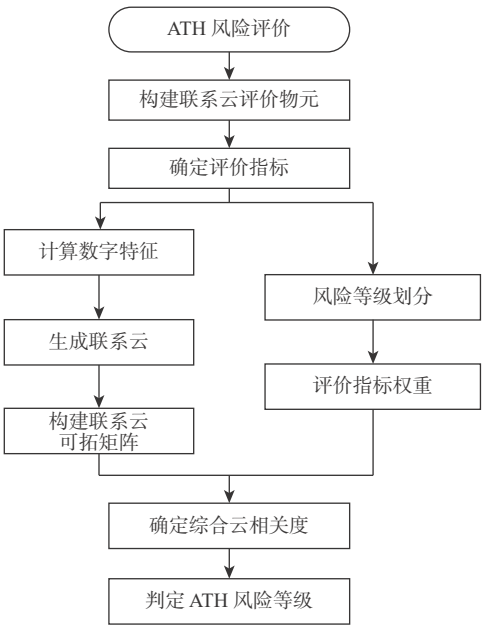


图 2　基于联系云的 ATH 可拓评价流程

Fig. 2　Extension evaluation process of ATH based on connection cloud

步骤 1　ATH 受飞机性能、道面状况及污染物分布因素共同影响, 根据待评价物元选定评价指标 C_j, 确定各风险等级左右半区间长度 α_{left} 和 α_{right}, 求解联系云数字特征 $(E_x, E_{\text{n}}, H_{\text{e}}, \alpha, \lambda)$, 生成有限区间等级联系云。

步骤 2　根据实测样本数据计算该样本隶属于各风险等级的云相关度, 构建基于联系云的可拓矩阵。

步骤 3　综合评价指标权重和联系云可拓矩阵, 确定当前样本综合云相关度, 最终判定待评价物元的 ATH 风险等级。

2.2　可拓矩阵构建

构建可拓矩阵是确定 ATH 风险等级的基础。对于分类标准区间值随风险等级增大而单调增大的指标(如地面滑行速度、道面积水厚度等), 等级 i 左右半区间长度计算式为

$$\begin{cases} \alpha_{i\text{left}} = E_x^i - L_{\min}^{i-1} \\ \alpha_{i\text{right}} = L_{\max}^{i+1} - E_x^i \end{cases} \tag{9}$$

相反的, 对于分类标准区间值随等级增大而单调减小的指标(如道面摩擦系数、飞机轴载等级等), 等级 i 的左右半区间长度计算式为

$$\begin{cases} \alpha_{i\text{left}} = E_x^i - L_{\min}^{i+1} \\ \alpha_{i\text{right}} = L_{\max}^{i-1} - E_x^i \end{cases} \tag{10}$$

按 2.1 节步骤 1 分别计算左右半区间数字特征, 得出 ATH 风险等级界限的标准联系云; 生成以等级 i 区间熵值 E_{n}^i 为均值、以超熵 H_{e}^i 为标准差的正态随机数 $E_{\text{n}}^{i\prime}$。假定待评价 ATH 样本指标值为 x_i, 由式 (1) 计算其对应隶属度, (x_i, μ_i) 共同构成一个云滴, 重复上述步骤直至生成由 n 个云滴组成的联系云。μ_i 表征了指标值 x_i 关于等级 i 的云隶属度, μ_{ji} 表示事物指标与等级 i 云表示的事物指标之间的关联度, 称为云相关度。由云相关度组成的联系云可拓矩阵 $\boldsymbol{S}$ 形式如下:

$$\boldsymbol{S} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2p} \\ \vdots & \vdots & & \vdots \\ \mu_{q1} & \mu_{q2} & \cdots & \mu_{qp} \end{bmatrix} \tag{11}$$

参考以往研究经验[23-25], 确定本文 ATH 评价指标 q 为 5 个, 滑水风险分类等级 p 为 4 级。

2.3　评价指标与等级确定

近年来, 耦合欧拉-拉格朗日(coupled Eulerian-Lagrangian, CEL)算法支撑复杂 ATH 仿真分析快速发展, 弥补了经典轮胎滑水试验工况较简单、考察变量少等不足, 可为 ATH 风险分析提供案例数据。此类仿真方式与分析精度经相关研究反复验证并普遍采用, 详细建模及模型验证流程参见文献[26-27], 典型轮胎滑水 CEL 模型形式如图 3 所示。

图 3 中, 滚动飞机轮胎与道面积水相对高速运动并产生冲击变形, 轮胎-道面-积水层体系相互作用直至滑水临界状态, 可根据分析需要调节飞机动力滑行、道面表面性能及积水层分布等参数条件。在选择评价指标时, 既要考虑其对 ATH 系统各部分代表性, 也应考虑机场运行实践中数据获取的难易程度与可靠性。综合相关研究成果与行业标准, 本文选取的 5 项 ATH 风险评价指标包括滑行速度、积水厚度、摩擦系数、单轮轴载和刻槽深度, 减

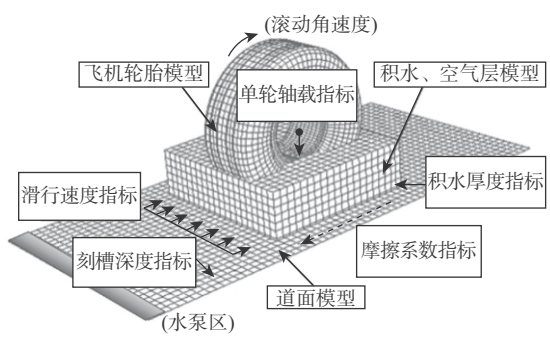


图 3　典型轮胎滑水 CEL 算法仿真模型

Fig. 3　Typical simulation model of tire hydroplaning based on coupled Eulerian-Lagrangian algorithm

少人为主观评价对ATH分析的不利影响。

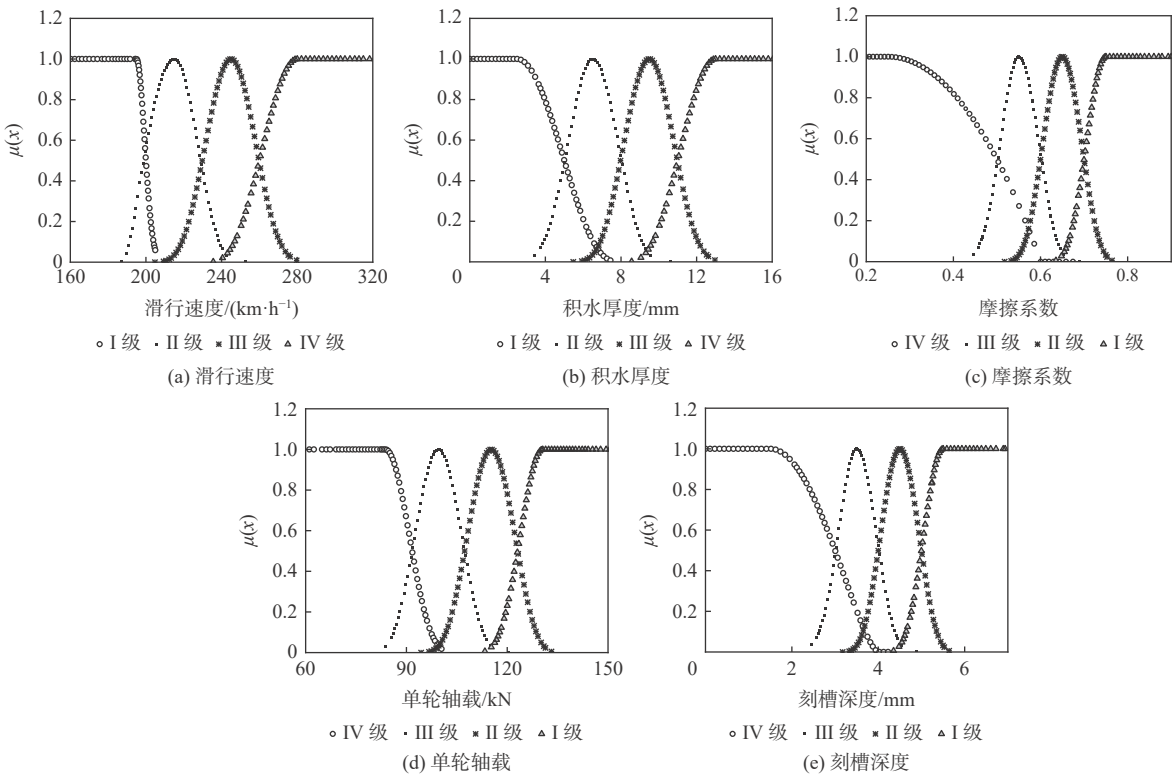
采用文献[9]提出滑水风险指标(hydroplaning risk index, HRI)进行滑水风险分级,该指标基于QAR数据特征分析提出,可量化地面滑行速度降至安全区间范围前,高滑水风险“窗口期”在飞机起降全过程中的占比。将ATH风险划分为可忽略(Ⅰ级)、可接受(Ⅱ级)、可接受需整改(Ⅲ级)和不可接受(Ⅳ级)共4个等级。通过前期仿真分析,确定上述ATH风险评价指标分级标准,限于文章篇幅不再展开,评价指标分类标准如表1所示。

图4中,4朵联系云依次对应风险等级Ⅰ~Ⅳ。Ⅱ和Ⅲ等级分布与图1中样式一致;由于指标在有限范围取值,Ⅰ和Ⅳ等级外侧半区间云模型不“下落”,边界隶属度为固定值1,较以往正态云分布特征更合理。本文选取5项评价指标均为定量指标,各指标不同风险等级标准云之间存在交叉重叠,可还原ATH风险评价模糊随机特性。

表1 ATH风险等级评价指标分类标准

Table 1 Classification standards of evaluation indexes for ATH risk grades

风险等级	滑行速度/(km·h^{-1})	积水厚度/mm	摩擦系数	单轮轴载/kN	刻槽深度/mm
Ⅰ	≤200	≤5	>0.7	>123.1	>5
Ⅱ	≤230	≤8	≤0.7	≤123.1	≤5
Ⅲ	≤260	≤11	≤0.6	≤107.3	≤4
Ⅳ	>260	>11	≤0.6	≤91.7	≤3



(a) 滑行速度 (b) 积水厚度 (c) 摩擦系数

(d) 单轮轴载 (e) 刻槽深度

图4 评价指标隶属ATH风险等级标准联系云

Fig. 4 Standard connection clouds of evaluation index under different membership of ATH risk grades

依托联系云可拓矩阵$\boldsymbol{S}$及权重向量$\boldsymbol{W}$,两者相乘得到综合云相关度向量$\boldsymbol{A}$,ATH风险等级评价模型如下:

$$a_i(N)=\max\{a_i(N)|i=1,2,3,4\} \tag{12}$$

$$r=\sum_{i=1}^{4}a_ig_i\Big/\sum_{i=1}^{4}a_i \tag{13}$$

式中:a_i为$\boldsymbol{A}$的最大分量值,表征当前待评价样本N关于等级i的综合云相关度;g_i为4个等级得分分值;r为加权平均法计算评判分数。

可通过多次计算减弱可拓矩阵$\boldsymbol{S}$中μ_{ji}的随机性,公式如下:

$$\begin{cases}E_{rx}=\dfrac{r_1(x)+r_2(x)+\cdots+r_n(x)}{n}\\ E_{rn}=\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(r_i(x)-E_{rx})^2}\\ \theta=E_{rn}/E_{rx}\end{cases} \tag{14}$$

式中:置信因子θ表示ATH评价结果分散程度与可信度,为多次计算期望值E_{rx}与熵值E_{rn}之比;重复运算次数n=300~500次;通常取$\theta\leqslant0.01$作为控制标准。

3　实例应用

为验证联系云可拓评价理论的合理性和比较优势, 以某多雨地区 4E 等级机场为例开展滑水风险等级评价应用分析。该机场道面因长期运行导致表面性能下降, 湿滑起降条件下, 轮胎滑水事件频发导致 ATH 问题突出, 3 起典型滑水事故征候跑道运行参数如表 2 所示。

选用惩罚性状态变权向量, 权重向量 $\boldsymbol{W}$ 与变权理论结合, 依据飞机滑跑状态和外部环境动态调整。对偏离正常运行的某些指标值, 根据偏离严重程度给予惩罚性调整, 使评估结果与实际运行状态趋近一致。变权向量 $\boldsymbol{W}$ 计算结果如下:

$$\begin{cases} \boldsymbol{W}_1 = [0.066\,5,\ 0.436\,9,\ 0,\ 1\,477,\ 0.265\,3,\ 0.083\,6] \\ \boldsymbol{W}_2 = [0.198\,7,\ 0.206\,7,\ 0,\ 2\,044,\ 0.189\,0,\ 0.201\,2] \\ \boldsymbol{W}_3 = [0.185\,3,\ 0.131\,7,\ 0,\ 2\,316,\ 0.235\,6,\ 0.215\,8] \end{cases} \tag{15}$$

由式 (12) 确定各样本的评判等级, 结果可信程度在表 3 中列出。以样本 1 滑行速度为例 (x_0 = 220 km/h), 说明对应Ⅰ～Ⅳ等级联系概率的计算过程。按式 (1) 计算隶属度 $\mu_{1,11}$=0, $\mu_{1,21}$=0.919 2, $\mu_{1,31}$=0.122 1, $\mu_{1,41}$=0, 此时样本 1 滑行速度指标对等级Ⅱ隶属程度最大, 其次为等级Ⅲ, 隶属等级Ⅰ或等级Ⅳ可能性为 0。同理, 对其他 4 项指标求解各等级云相关度。进而结合评价指标权重向量求得隶属不同等级的综合云相关度, 并按最大综合云相关度原则判定样本风险等级。

表 2　滑水事故征候跑道运行参数组合

Table 2　Parameter combination of operating runway of hydroplaning symptoms

样本	滑行速度/(km·h^{-1})	积水厚度/mm	单轮轴载/kN	刻槽深度/mm	摩擦系数
1	220	3	138.8	6	0.76
2	250	5	100.4	5	0.66
3	280	8	76.5	2	0.60

表 3　ATH 风险安全性评价结果及对比

Table 3　Safety evaluation results of ATH risks and comparison

样本	综合云相关度				置信因子	稳定性等级与起降条件		
	Ⅰ	Ⅱ	Ⅲ	Ⅳ		本文方法	正态云法	规范判定
1	0.922 9	0.068 7	0.008 2	0	0.003 9	Ⅰ	Ⅰ	允许起降
2	0.212 7	0.441 8	0.390 3	0.037 6	0.001 2	Ⅱ	Ⅲ	允许起降
3	0.000 7	0.171 3	0.288 5	0.546 2	0.008 3	Ⅳ	Ⅳ	允许起降

表 3 中, 3 组样本评价结果的置信因子 θ 均小于 0.01, 表明采用本文建议的联系云可拓理论计算 ATH 风险等级结果分散性小, 可信程度高。与传统正态云模型评价结果相比, 两者对样本 1 和样本 3 的评价等级结论一致, 分别为Ⅰ级和Ⅳ级; 对于样本 2, 采用本文计算理论得出 ATH 风险等级为Ⅲ级, 风险控制标准更为严格。依托集对关系构建 ATH 联系云可拓评价模型, 为多重不相容指标的随机模糊及不确定性分析提供了可能, 克服了正态云模型不能有效模拟评价指标在有限区间内分布的缺陷, 使之更符合指标分类和信息获取的实际特征。

4　结　论

1) 将联系云模型引入可拓理论物元表达, 实现 ATH 风险评价基于语言值的定性概念和数值表示之间的相互映射, 拓展了传统正态云模型的应用方式。

2) 联系云模型在有限区间内生成非对称分布云模型, 更加符合 ATH 评价指标分布特征; 依托飞机轮胎滑水流固耦合仿真, 可确定评价指标分级标准。

3) 实例应用表明, 联系云可拓评价模型可统一定量描述等级模糊转换态势和指标的确定与不确定性, 对 ATH 风险控制更为严格, 且置信因子小于 0.01, 符合可信度要求, 可为同类研究提供参考。

参考文献 (References)

[1] 赵安家, 孙丽莹, 孟哲理. 飞机轮胎滑水与预防控制措施研究综述[J]. 飞机设计, 2015, 35(5): 46-51.
ZHAO A J, SUN L Y, MENG Z L. A search for mechanism and preventability measure of the aircraft tire hydroplaning[J]. Aircraft Design, 2015, 35(5): 46-51(in Chinese).

[2] 蔡靖, 许诤. 沟槽磨损对飞机轮胎滑水影响仿真分析[J]. 中国民航大学学报, 2020, 38(2): 38-43.
CAI J, XU Z. Simulation analysis on influence of groove abrasion on aircraft hydroplaning[J]. Journal of Civil Aviation University of China, 2020, 38(2): 38-43(in Chinese).

[3] 许诤. 考虑道面平整度的飞机轮胎滑水安全问题研究[D]. 天津: 中国民航大学, 2019: 10-23.
XU Z. Study on water skiing safety of aircraft tires considering pavement smoothness[D]. Tianjin: Civil Aviation University of China, 2019: 10-23(in Chinese).

[4] 张恒. 轮胎与湿滑道面相互作用下的飞机滑水行为研究[D]. 天津: 中国民航大学, 2018: 62-66.

ZHANG H. Study on aircraft water skiing behavior under the interaction between tire and wet road surface[D]. Tianjin: Civil Aviation University of China, 2018: 62-66(in Chinese).

[5] 朱兴一, 庞亚凤, 杨健, 等. 湿滑条件下基于真实纹理道面的机轮着陆滑水行为解析[J]. 中国公路学报, 2020, 33(10): 159-170.
ZHU X Y, PANG Y F, YANG J, et al. Analysis on the hydroplaning of aircraft tire under real texture pavement conditions[J]. China Journal of Highway and Transport, 2020, 33(10): 159-170(in Chinese).

[6] 刘芳兵. 湿滑跑道飞机侧风着陆滑行安全研究[D]. 天津: 中国民航大学, 2020: 45-47.
LIU F B. Study on the safety of aircraft landing and taxiing in crosswind on wet runway[D]. Tianjin: Civil Aviation University of China, 2020: 45-47(in Chinese).

[7] TREMBLAY L, METIVET M, MEUNIER F, et al. Method and system for aircraft sideslip guidance: US11054437[P]. 2021-07-06.

[8] 李岳, 宗辉杭, 蔡靖, 等. 飞机轮组滑水行为与道面积水附加阻力[J]. 北京航空航天大学学报, 2023, 49(5): 1099-1107.
LI Y, ZONG H H, CAI J, et al. Hydroplaning behavior of aircraft wheel group and additional resistance due to accumulated water on pavement[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(5): 1099-1107(in Chinese).

[9] 李岳, 胡宇祺, 蔡靖, 等. 湿滑道面飞机着陆滑水风险量化分析[J]. 南京航空航天大学学报, 2022, 54(6): 1138-1144.
LI Y, HU Y Q, CAI J, et al. Quantification analysis of hydroplaning risks of aircraft landing on wet pavement[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2022, 54(6): 1138-1144 (in Chinese).

[10] FWA T F, PASINDU H R, ONG G P. Critical rut depth for pavement maintenance based on vehicle skidding and hydroplaning consideration[J]. Journal of Transportation Engineering, 2012, 138(4): 423-429.

[11] 张兆宁, 石峰. 基于组合赋权云模型的塔台管制系统运行安全评估[J]. 安全与环境学报, 2024, 24(4): 1254-1265.
ZHANG Z N, SHI F. Operational safety assessment of tower control system based on combined weighted cloud model[J]. Journal of Safety and Environment, 2024, 24(4): 1254-1265(in Chinese).

[12] 史佳辉, 徐吉辉, 陈玉金, 等. 基于交互作用矩阵多维云模型的飞机重着陆风险评估方法研究[J]. 系统工程与电子技术, 2021, 43(10): 3026-3032.
SHI J H, XU J H, CHEN Y J, et al. Research on risk assessment method of aircraft heavy landing based on interaction matrix-multidimensional cloud model[J]. Systems Engineering and Electronics, 2021, 43(10): 3026-3032(in Chinese).

[13] 唐家文, 董兵, 王超峰. 基于云模型的空管安全运行保障能力评价[J]. 航空工程进展, 2021, 12(4): 59-67.
TANG J W, DONG B, WANG C F. Evaluation on safe operation support ability of air traffic management based on cloud model[J]. Advances in Aeronautical Science and Engineering, 2021, 12(4): 59-67(in Chinese).

[14] 李岳, 胡宇祺, 蔡靖, 等. 基于变权重-正态云模型的飞机轮胎滑水风险研究[J]. 北京航空航天大学学报, 2023, 49(9): 2299-2305.
LI Y, HU Y Q, CAI J, et al. Hydroplaning risk of aircraft tire based on variable weight theory-normal cloud model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(9): 2299-2305 (in Chinese).

[15] 周义蛟, 郭基联, 周舟. 基于云模型与组合赋权法的飞机保障性评估研究[C]//2017年首届航空保障设备发展论坛. 北京: 中国航空航天工具协会, 中国航空学会航空维修工程专业分会, 2017: 142-148.
ZHOU Y J, GUO J L, ZHOU Z. A study on military aircraft supportability assessment based on cloud model and game theory[C]// Proceedings of the 2017 1st Aviation Support Equipment Development Forum. Beijing: China Aerospace Tools Association, Aviation Maintenance Engineering Branch of the Chinese Aerospace Society, 2017: 142-148(in Chinese).

[16] 叶琼, 李绍稳, 张友华, 等. 云模型及应用综述[J]. 计算机工程与设计, 2011, 32(12): 4198-4201.
YE Q, LI S W, ZHANG Y H, et al. Cloud model and application overview[J]. Computer Engineering and Design, 2011, 32(12): 4198-4201(in Chinese).

[17] 李德毅, 刘常昱. 论正态云模型的普适性[J]. 中国工程科学, 2004, 6(8): 28-34.
LI D Y, LIU C Y. Study on the universality of the normal cloud model[J]. Engineering Science, 2004, 6(8): 28-34(in Chinese).

[18] 刘俊杰, 张瑞瑞, 叶英豪, 等. 基于云模型的航空器地面滑行错误事件风险分析[J]. 中国民航飞行学院学报, 2022, 33(5): 51-56.
LIU J J, ZHANG R R, YE Y H, et al. Risk analysis of aircraft ground taxiing error event based on cloud model[J]. Journal of Civil Aviation Flight University of China, 2022, 33(5): 51-56(in Chinese).

[19] 汪明武, 王霄, 龙静云, 等. 基于多维联系正态云模型的泥石流危险性评价[J]. 应用基础与工程科学学报, 2021, 29(2): 368-375.
WANG M W, WANG X, LONG J Y, et al. Risk assessment of debris flow based on multidimensional connection normal cloud model[J]. Journal of Basic Science and Engineering, 2021, 29(2): 368-375(in Chinese).

[20] 汪明武, 朱其坤, 赵奎元, 等. 基于有限区间联系云的围岩稳定性评价模型[J]. 岩土力学, 2016, 37(增刊 1): 140-144.
WANG M W, ZHU Q K, ZHAO K Y, et al. Stability evaluation model of surrounding rock based on limited interval connection cloud[J]. Rock and Soil Mechanics, 2016, 37(Sup 1): 140-144(in Chinese).

[21] 马丽叶, 张涛, 卢志刚, 等. 基于变权可拓云模型的区域综合能源系统综合评价[J]. 电工技术学报, 2022, 37(11): 2789-2799.
MA L Y, ZHANG T, LU Z G, et al. Comprehensive evaluation of regional integrated energy system based on variable weight extension cloud model[J]. Transactions of China Electrotechnical Society, 2022, 37(11): 2789-2799(in Chinese).

[22] 蔡文, 杨春燕, 何斌. 可拓学基础理论研究的新进展[J]. 中国工程科学, 2003, 5(2): 80-87.
CAI W, YANG C Y, HE B. New development of the basic theory of extenics[J]. Engineering Science, 2003, 5(2): 80-87(in Chinese).

[23] 关晓吉. 基于可拓联系云模型的隧道塌方风险等级评价方法[J]. 中国安全生产科学技术, 2018, 14(11): 186-192.
GUAN X J. Evaluation method on risk grade of tunnel collapse based on extension connection cloud model[J]. Journal of Safety Science and Technology, 2018, 14(11): 186-192(in Chinese).

[24] 牛亚东, 张思祥, 田广军, 等. 机场跑道摩擦系数影响因素研究[J]. 应用力学学报, 2021, 38(2): 715-720.
NIU Y D, ZHANG S X, TIAN G J, et al. Research on influencing factors of friction coefficient in airport runway[J]. Chinese Journal

of Applied Mechanics, 2021, 38(2): 715-720(in Chinese).

[25] 王迎超, 靖洪文, 张强, 等. 基于正态云模型的深埋地下工程岩爆烈度分级预测研究[J]. 岩土力学, 2015, 36(4): 1189-1194.
WANG Y C, JING H W, ZHANG Q, et al. A normal cloud model-based study of grading prediction of rockburst intensity in deep underground engineering[J]. Rock and Soil Mechanics, 2015, 36(4): 1189-1194(in Chinese).

[26] 宗一鸣. 湿滑道面条件下轮胎力学行为与飞机着陆安全问题研究[D]. 天津: 中国民航大学, 2017: 19-30.
ZONG Y M. Research on tire mechanical behavior and aircraft landing safety under wet road surface conditions[D]. Tianjin: Civil Aviation University of China, 2017: 19-30(in Chinese).

[27] OH C W, KIM T W, JEONG H Y, et al. Hydroplaning simulation for a straight-grooved tire by using FDM, FEM and an asymptotic method[J]. Journal of Mechanical Science and Technology, 2008, 22(1): 34-40.

Extensible evaluation model of aircraft tire hydroplaning risk based on connection cloud

LI Yue[1], ZHOU Zeyuan[1], CAI Jing[1, 2, *]

(1. College of Transportation Science and Engineering, Civil Aviation University of China, Tianjin 300300, China;
2. Intelligent Construction and Industrialization Engineering Technology Research Center of Civil Aviation Airport, Tianjin 300456, China)

Abstract: Since aircraft tire hydroplaning can be influenced by several factors, and the characteristic of evaluation indexes can be described as fuzzy, random, and discrete, a hydroplaning risk evaluation model based on extension theory and connection cloud was established, so as to quantify the transformation of hydroplaning evaluation indexes among different classification levels. The numerical characteristics of the connection cloud were calculated according to the leveling criteria of evaluation indexes, and the connection cloud within a limited range was generated. The extensible matrix of the connection cloud was built by using certainty degrees. In this way, the final risk level could be obtained on the basis of variable weights, which demonstrated the dynamic connection between elements to be evaluated and risk level. The case analysis data was obtained by the fluid-solid coupling simulation of aircraft tire hydroplaning to make up for the lack of variable conditions in the classic hydroplaning test. The analysis results show that the evaluation conclusions of sample 1 and sample 3 are consistent based on the traditional normal cloud model and extensible connection cloud model. The hydroplaning risk level of sample 2 is given as Ⅲ by using the proposed model in this paper. Therefore, the risk control is considered more restrict under the same parameter condition. The confidence factor of the above sample risk assessment is less than 0.01, and the credibility of the evaluation results is high. The proposed model in this paper provides an alternative method for random-fuzzy and uncertainty analysis involving multiple incompatibility indexes. Hence, the defect of the normal cloud model in simulating the distribution of evaluation indexes within a limited range can be overcome.

Keywords: tire hydroplaning; connection cloud; extension theory; fluid-solid coupling; risk evaluation

Received: 2023-03-23; **Accepted**: 2023-05-29; **Published Online**: 2023-06-15 11:19

URL: link.cnki.net/urlid/11.2625.V.20230614.1459.002

Foundation items: National Natural Science Foundation of China (52472369); The Fundamental Research Funds for the Central Universities (3122019103); Open Project of the Intelligent Construction and Industrialization Engineering Technology Research Center of Civil Aviation Airport (MHJGKFKT-01,MHJGKFKT-04)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0088

融合 Mobile Vit 和倒置门控编解码的视网膜血管分割算法

梁礼明，阳渊，朱晨锟，何安军，吴健*
(江西理工大学 电气工程与自动化学院，赣州 341000)

摘 要：针对视网膜血管分割时存在背景噪声干扰、边界纹理模糊和微细血管提取难等问题，提出一种融合 Mobile Vit 和倒置门控编解码的视网膜血管分割算法 (FMVG-Net)。改进 Mobile Vit 模块，在编码部分实现双联合特征提取；利用多谱注意力模块，从频域维度减少图像特征信息缺失，精确分割血管前景像素；提出特征自适应融合模块，建立血管纹理上下文依赖关系，提高血管分割灵敏度；优化编解码结构，设计倒置门控编解码模块，进一步捕获空间信息与深层语义信息，提高视网膜血管图像分割精度。在公共数据集 DRIVE、STARE 和 CHASE_DB1 上对所提算法进行实验，特异性分别为 0.986 3、0.989 7 和 0.987 3，准确度分别为 0.970 9、0.975 4 和 0.976 0，敏感度分别为 0.810 9、0.801 0 和 0.807 9。仿真实验证明，所提网络对视网膜血管分割具有较好的分割效果，为眼科疾病的诊断提供了新窗口。

关 键 词：视网膜血管；Mobile Vit 模块；离散余弦变换；倒置门控编解码模块；特征自适应融合

中图分类号：TP391

文献标志码：A **文章编号**：1001-5965(2025)03-0712-12

视网膜血管的定期检测对高血压、糖尿病和动脉硬化等疾病的诊断、筛查和治疗具有重要作用，及时检测异常血管能帮助病患尽早发现疾病并阻止疾病的恶化。由于视网膜血管拓扑结构复杂多变，血管粗细度、弯曲度和分支形态等特征难以预测，对血管分割难度较大。因此，迫切需要一种快速且精准的视网膜血管自动分割算法[1-2]。

目前，相关领域的学者已经提出了诸多分割算法，主要分为无监督学习和有监督学习 2 类。其中，无监督学习的分割算法提取血管的固有拓扑结构特征，无需专家标记的血管信息，包括血管追踪[3]、匹配滤波[4] 和水平集[5] 等。由于无监督学习的分割算法缺乏监督信息，鲁棒性不高且难以精确分割血管，无法满足实际临床的辅助需求。而有监督学习的分割算法主要通过提取视网膜血管结构特征，并依据专家标签来训练分类模型，如支持向量机 (support vector machine, SVM)[3] 和随机森林[6] 等。以上算法在特征人工提取时受主观因素干扰较大，因而无法自适应血管的形态几何变换等问题。

近年来，基于有监督学习的深度学习分割算法在相关领域备受关注，其对高级语义特征的提取更加充分、对血管前景像素点的识别更加精准，并能不断自动学习改进，大幅提升了模型的稳健性和眼底视网膜血管分割的精度。在深度学习分割算法中，Ronneberger 等[7] 基于全卷积网络提出具有编解码结构的 U-Net 算法，通过对称的编码解码结构实

收稿日期：2023-02-28；录用日期：2023-06-16；网络出版时间：2023-06-29 17:47
网络出版地址：link.cnki.net/urlid/11.2625.V.20230629.1317.001
基金项目：国家自然科学基金 (51365017,61463018)；江西省自然科学基金面上项目 (20192BAB205084)；江西省教育厅科学技术研究重点项目 (GJJ170491,GJJ2200848)；江西省研究生创新专项资金 (YC2022-S676)
* 通信作者. E-mail：yangy27980218@163.com

引用格式：梁礼明，阳渊，朱晨锟，等. 融合 Mobile Vit 和倒置门控编解码的视网膜血管分割算法 [J]. 北京航空航天大学学报，2025，51（3）：712-723. LIANG L M，YANG Y，ZHU C K，et al. Fusion of Mobile Vit and inverted gated codec retinal vessel segmentation algorithm [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（3）：712-723（in Chinese）.

现对医学图像的理想分割。学者们对 U-Net 算法进行了一系列的改进优化, 提出了诸多基于 U-Net 算法的改进算法[8-10], 这些算法虽然在视网膜血管分割上取得了不错的分割效果, 但在分割时仍存在模型感受野不足、血管局部特征缺失和微细血管纹理结构模糊等问题。Transformer 模型[11]的出现给视觉领域带来了生机, 其最先应用于自然语言处理(natural language processing, NLP)领域, 随后视觉 Transformer(vision Transformer, Vit)模型[12]在一定程度上解决了高级语义提取难的问题, 但 Transformer 的弊端明显, 参数量大, 对算力要求高, 模型训练困难。Jiang 等[13]基于多尺度 Transformer 和位置注意力(mulei-scale Transformer-position attention, MTPA), 联合 Transformer 和卷积神经网络(convolutional neural network, CNN)提出 MTPA_Unet 视网膜血管分割算法, 虽然取得了比较理想的分割结果, 但仍存在少许非血管渗出物和末端微小血管分割断裂等问题。综上所述, 以上算法在分割领域虽然取得了较好的研究成果, 但仍存在模型特征固化、对眼底视网膜图像背景噪声处理稳健性不高和尺度与拓扑形态结构特征提取不健全的问题。

针对上述算法中存在的挑战, 本文提出一种融合 Mobile Vit 和倒置门控编解码的视网膜血管分割算法(fusion of Mobile Vit and inverted gated codec retinal vessel segmentation algorithm, FMVG-Net)应用于视网膜血管图像分割, 主要工作为: ①在编解码器端设计倒置门控编解码(inverted gated codec, IGC)模块, 促进对视网膜主干血管和分支血管像素的动态提取; ②改进 Mobile Vit(improved Mobile Vit, IMV)模块, 与倒置门控编码构成联合双编码结构, 建立局部与全局语义的竞争依赖关系, 从而提升网络对视网膜血管的提取和分割定位能力; ③提出特征自适应融合(feature adaptive fusion, FAF)模块,置于网络底端, 实现对多尺度血管语义的跨维度捕捉, 提高深层网络的稳健性; ④引入多谱注意力模块, 通过对图像频域进行离散余弦解析的方式, 实现对血管不相关噪声的剔除和抑制背景像素的干扰。使用联合加权损失函数抑制正负样本分布不均的问题, 弱化因微细血管区域像素与全图像素占比率过低而引起的误分割。

1　本文算法

1.1　算法设计

针对视网膜血管边缘分割模糊、非血管噪声干扰和微细血管提取难等问题, 本文提出融合 Mobile Vit 和倒置门控编解码算法, 算法框架如图 1 所示, 主要由图像扩充及预处理、纹理信息双提取、特征编码融合和血管像素解码 4 部分组成。图像扩充及预处理部分, 先对数据集图像以 0.6 的概率随机翻转扩充, 再将扩充后的图像做深度预处理来增强血管前景与背景的对比度。纹理信息双提取部分, 在网络上端, 由 IMV 模块和 IGC 模块组成, 建立血管纹理信息和脉络拓扑结构的依赖关系。特征编码融合部分, 主要对编码器部分与纹理信息双提取部分的纹理特征进行编码整合, 抑制冗余信息。血管像素解码部分, 主要对编码信息自适应整合, 并对血管像素进行恢复, 最终生成视网膜血管分割效果图。

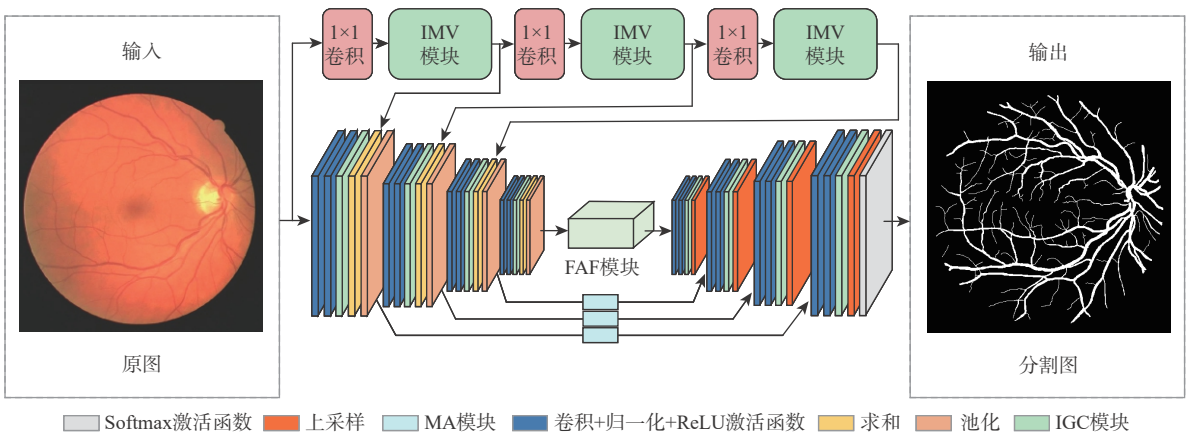


图 1　本文网络结构

Fig. 1　Network structure of this paper

1.2　改进的 Mobile Vit 模块

深度学习算法在很大程度上解决了图像特征信息提取难的问题, 但对小目标任务而言, 还有提升空间。MobileViT[12]模型是较为轻量化的视觉 Vit 模型, 能较好地处理因参数量过大而导致的模型训练困难的问题。

针对视网膜血管微细纹理提取难的问题, 本文设计 IMV 模块, 如图 2 所示。IMV 模块由局部学

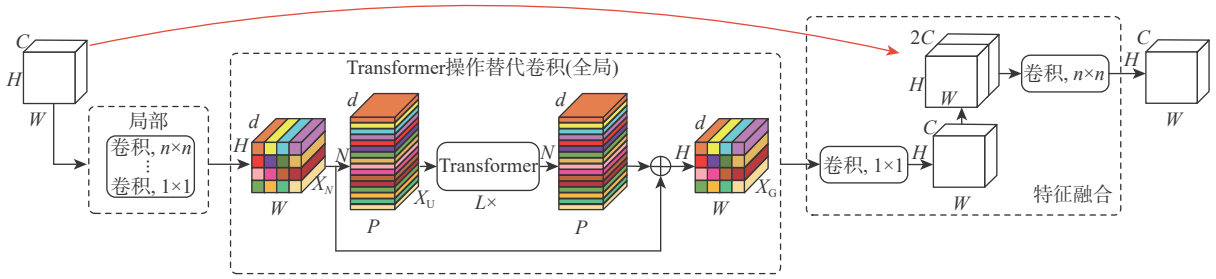


图 2 改进的 Mobile Vit 模块

Fig. 2 Improved Mobile Vit module

习块、全局学习块和特征融合 3 部分构成。首先，特征图通过局部学习块做局部特征建模。然后，经卷积核大小为 $n\times n$ 的卷积层，做卷积核为 1×1 的卷积操作，进行通道调整。在全局学习块部分，为学习更多空间感应偏差，首先，将局部特征 $X_N\in\mathbf{R}^{W\times H\times d}$（$W$ 为宽，H 为高，d 为长）展开为 N 个不同的重叠斑纹块 $X_U\in\mathbf{R}^{P\times N\times d}$，进行 Transformer 全局建模，在数学上表示为

$$X_T(p)=\text{Transformer}(X_U(p)) \qquad p\in[1,P] \tag{1}$$

式中：X_T 为全局特征块。

然后，通过通道融合复原操作建立局部学习块与全局学习块的上下文依赖关系，即

$$X_G=\xi(X_N,X_U) \tag{2}$$

式中：X_N 为输入局部特征块；X_U 为展开局部特征块；X_G 为融合特征块；ξ 表示特征加权融合操作。

最后，经多尺度卷积对通道进行特征还原，通过特征融合结构实现对视网膜血管纹理特征的提取。

1.3 倒置门控编解码模块

在视网膜血管分割任务中，能否提取到更多的准确血管像素对分割的结果非常重要。U-Net 算法通过简单编码解码的结构实现血管分割，不可避免地会出现血管像素误分割现象，因此，有必要设计一种能降低血管误分割率的模块。Ruan 等[14] 提出了一种用于轻量级医学图像分割的倒置外部注意力（inverted external attention, IEA）块，如图 3 所示，该模块可增强局部目标特征的关注度，但随着网络的加深，其容易出现梯度爆炸和网络退化等问题。

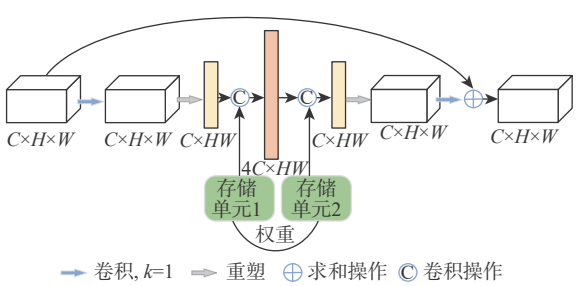


图 3 倒置外部注意力块[14]

Fig. 3 Invert external attention module[14]

He 等[15-16] 提出残差结构解决了该问题。

为增强原始编解码结构对视网膜血管分割的精确度，受文献 [14-16] 启发，本文设计了一种 IGC 模块，如图 4 所示。

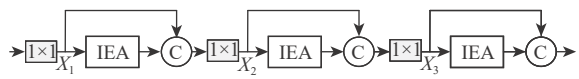


图 4 倒置门控编解码模块

Fig. 4 Inverted gated codec module

IGC 模块主要由 3 个 IEA 模块和门控结构单元所组成。IEA 模块（见图 3）原理如下：给定输入 $X\in\mathbf{R}^{C\times H\times W}$，先经维度重塑变为 $X\in\mathbf{R}^{C\times HW}$（$C$ 为通道数目），再通过内存权重存储单元 1（MemoryUnit1）进行 4 倍特征展平，得到维度为 $\mathbf{R}^{4C\times HW}$ 的特征细节图，与内存权重存储单元 2（MemoryUnit2）进行权重共享与重塑，将特征维度恢复为 $\mathbf{R}^{C\times H\times W}$，最终进行特征融合操作。其中，2 个存储单元均为 1×1 的卷积操作，该单元能将输入映射到更高维度的空间，使存储单元更全面地描述整体特征信息。IGC 模块输入特征先通过 1×1 卷积完成通道校正，再经过 IEA 模块增强对血管微观纹理的关注度，经门控单元进行特征聚合，完成对血管特征的粗识别，最终通过 3 次上述操作实现对血管纹理的精确析出，具体为

$$X_{i+1}=X_iX_{Ai} \qquad i\in\{1,2,3\} \tag{3}$$

式中：X_i 为 IGC 模块中每个基本单元节点的输出；X_{Ai} 为 IGC 模块每个单元中共享注意权重。

1.4 多谱注意力模块

注意力机制在计算机视觉领域取得了较大成功，通常使用通道注意力和空间注意力机制来捕获输入信息，但由于池化操作的简单性，使其难以捕捉到复杂的输入信息。Qin 等[17] 提出频率域通道注意力网络，从频域上加强对目标特征的关注度。为剔除不相干的图像特征对目标特征的影响，提高眼底血管分割的精度，基于上述模型优化的基础，在连接处引入多谱注意力 (multispectral attention, MA) 模块，其内部结构如图 5 所示。该模块主要从频域角度去增强注意力度，给定输入 $X\in\mathbf{R}^{C\times H\times W}$，沿通道维度将输入 X 分裂成 n 个部分，表示为 $X_i\in\mathbf{R}^{C'\times H\times W}$，

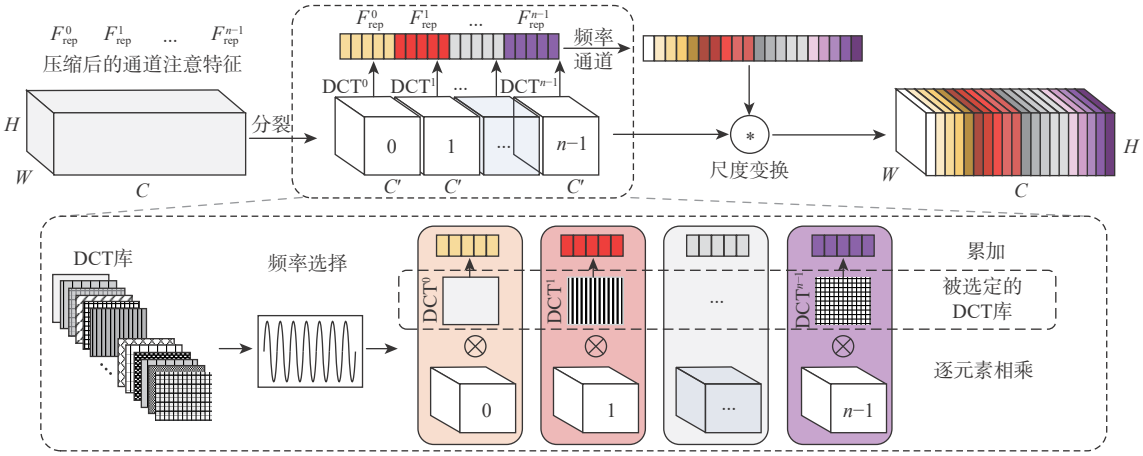


图 5　多谱注意力模块

Fig. 5　Multispectral attention module

$i\in\{0,1,2,\cdots,n-1\}$, 其中, $C'=C/n$(C能被n整除), 每个部分分配相应的离散余弦变换(discrete cosine transform, DCT)频率分量, DCT 分量被压缩成通道注意特征, 在数学上表示为

$$F_{\text{rep}}^{i}=D(X_i)\qquad i\in\{0,1,2,\cdots,n-1\}\tag{4}$$

式中: F_{rep}^{i}为压缩后的通道注意特征; D表示离散余弦变换操作。

压缩后的通道注意特征进行串联融合, 得到融合多谱特征矢量$\boldsymbol{F}_{\text{Fuse}}$, 具体为

$$\boldsymbol{F}_{\text{Fuse}}=\sigma\left(\left[F_{\text{rep}}^{0},F_{\text{rep}}^{1},\cdots,F_{\text{rep}}^{n-1}\right]\right)\tag{5}$$

式中: $\boldsymbol{F}_{\text{Fuse}}$为融合多谱特征矢量; σ表示特征融合操作。

通过特征尺度变换, 还原成原来特征尺度大小, 达到剔除不相干图像特征对目标特征和提高血管分割准确度的目的。

1.5　特征自适应融合模块

为捕获更多视网膜多尺度上下文血管像素, Huang 等[18]提出密集卷积网络, 受密集卷积思想影响, Wang 等[9]提出密集 U 网络实现对视网膜血管特征的获取。Liao 等[19]提出 HarDBlock 模块实现对多尺度目标特征的获取。受文献 [9,18-19] 启发, 本文提出一种 FAF 模块, 其结构如图 6 所示。

该模块由 4 个特征融合块组成, 首先, 通过 1×1 卷积、标准化和 ReLU 激活操作进行特征歧义校正; 然后, 使用卷积核为 3 的膨胀卷积对语义进行扩张, 扩张后的语义信息再特征聚合。特征歧义部分与扩张块部分两两相互反馈, 特征融合块之间采用密集连接的思想搭建上下文依赖关系, 最终实现对视网膜多尺度语义信息的捕捉。

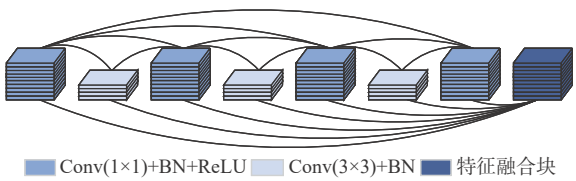


图 6　特征自适应融合模块

Fig. 6　Feature adaptive fusion module

2　实验结果分析

2.1　实验环境与参数设置

本文实验基于 64 位 Windows11 操作系统, 使用 Python 编程语言在 Pytorch 框架中搭建, 主要硬件条件如下: 处理器(CPU)AMD Ryzen 7 5800H with Radeon Graphics, 显卡 NVIDIA GeForce RTX 3060 运行内存 16 GB。

本文模型训练使用 Adam 优化器, 训练过程采取早停机制, 若连续 5 轮无最佳权重保存, 则停止训练, 迭代轮次为 50, 批次大小为 6, 学习率设置为 0.001, 采用余弦退火法对学习率进行衰减, 即

$$l_{\text{r}}=l_{\min}+\frac{1}{2}(l_{\max}-l_{\min})\left(1+\cos\left(\frac{T_{\text{cur}}}{T_{\max}}\right)\right)\tag{6}$$

式中: l_{r}为学习率; $l_{\max}$为最大学习率; $l_{\min}$为最小学习率; T_{cur}为当前迭代次数; $T_{\max}$为最大迭代次数。

2.2　实验数据集

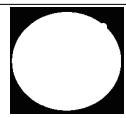
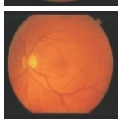
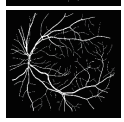
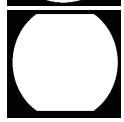
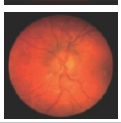
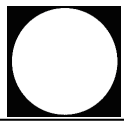
本文采用视网膜眼底图像公共数据集 DRIVE[20]、STARE[21] 和 CHASE_DB1[22] 进行训练与测试。3 个彩色眼底视网膜数据集的血管图像信息如表 1 所示。

DRIVE 数据集包含 40 张 tif 格式的彩色眼底视网膜图像, 分辨率尺寸为 565 ×584, 其中, 训练集与测试集均包含 20 张图像; STARE 数据集包含 20 张 ppm 格式的彩色眼底视网膜图像, 分辨率为 605×700, 由于该数据集没有设置训练集与测试集, 对该数据集采用五折交叉验证的方法进行实验; CHASE_DB1 数据集包含 28 张 jpg 格式的彩色眼底视网膜图像, 分辨率尺寸为 999×960, 将前 20 张彩色样本图像作训练集, 后 8 张彩色样本图像作测

试集。另外，以上3个数据集每张图像均包含2位专家手工标注的金标准图像，本文将第1位专家的金标准图作为实验的参考标准。

表1 彩色眼底视网膜数据集血管图像信息

Table 1 Vascular image information of color fundus retinal dataset

数据集	原图	金标准图	掩码图
DRIVE			
STARE			
CHASE_DB1			

2.3 数据集预处理与扩充

眼底视网膜图像在收集时存在噪声干扰、畸变、光照不足等现象，使图像分割时容易出现局部特征提取不足、微细血管分割模糊及血管断裂，因此，本文对数据集做如下处理，如图7所示。

1）三通道加权融合(three-channel weighted fusion,TWF)。通过对视网膜图像进行RGB通道分离发现，绿色通道的特征对比度优于其他2个通道，故对RGB三通道做赋权值加权处理，数学上表示为

$$I = C_r W_r + C_g W_g + C_b W_b \tag{7}$$

式中：I为加权融合后的图像；C_r、C_g、C_b分别为红、绿、蓝三通道的原始值；W_r、W_g、W_b分别为红、绿、蓝三通道的加权分量。

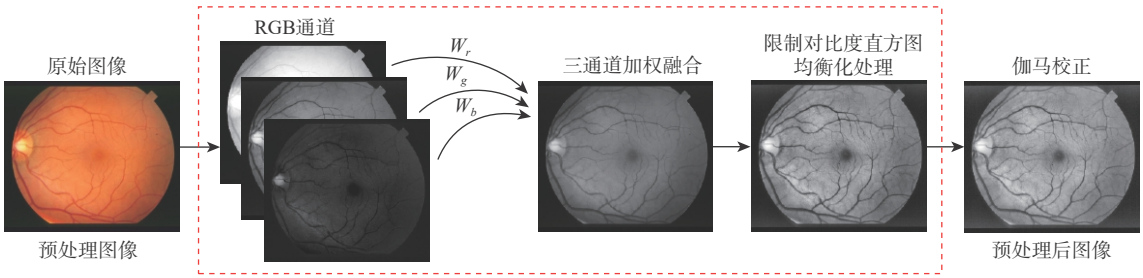


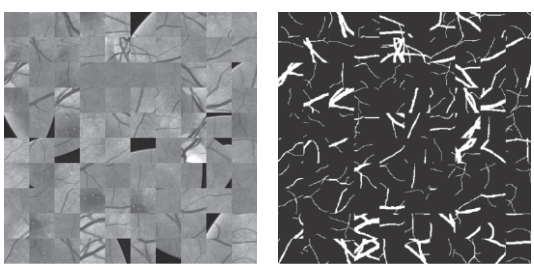
图7 图像预处理流程

Fig. 7 Image preprocessing flow

2）限制对比度直方图均衡化(contrast limited adaptive histogram equalization, CLAHE)处理。提升前景背景之间的细节对比度。

3）伽马校正。对图像亮度整体提升，增强眼底图像可视度，降低非血管伪影干扰。

4）数据集扩充。由于数据集容量少，容易出现过拟合现象，对数据集以0.6的概率做随机旋转扩充，采用64×64的滑动窗口将数据集中专家图像和训练图像随机分割成10 000张局部特征块，如图8所示。



(a) 原始局部特征块 (b) 金标准局部特征块

图8 眼底局部特征块切片

Fig. 8 Section of fundus local feature block

2.4 评价指标

为实现对视网膜血管分割结果的定量评估与分析，本文用准确度A_{cc}、特异性S_{pe}、敏感度S_{en}、F_1和受试者工作特征曲线AUC等指标进行对比分析，具体如式(8)～式(11)所示。其中，准确度表示被正确分类的像素点占全部像素点的比例，特异性表示非血管像素被正确分类的比例，敏感度表示血管像素被正确分类的比例。

$$S_{pe} = \frac{T_N}{T_N + F_P} \tag{8}$$

$$S_{en} = \frac{T_P}{T_P + F_N} \tag{9}$$

$$F_1 = \frac{2T_P}{2T_P + F_P + F_N} \tag{10}$$

$$A_{cc} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{11}$$

式中：F_P和F_N、T_P和T_N分别为与专家分割图不一致的血管像素和背景像素、与专家分割图一致的血管像素和背景像素。

2.5 联合加权损失函数

视网膜血管分割任务中，常用交叉熵损失函数评估模型性能和优化模型，但该函数不能比较理想地处理像素的分布不平衡问题，函数表达式如式(12)所示。焦点损失函数[23]能优化前景像素和背景像素关系，缓解类别间不平衡的问题，如式(13)所

示。因此, 在原常用损失函数基础上引入超参数α和β, 构成联合加权损失函数, 定义如式(14)所示。

$$L_{\mathrm{Cro}}(y_i,p_i)=-\frac{1}{N}\sum_{i\in\psi}\left[y_i\ln p_i+(1-y_i)\ln(1-p_i)\right] \quad (12)$$

$$L_{\mathrm{Foc}}(p_i)=-\xi_i(1-p_i)^{\gamma}\lg p_i \quad (13)$$

$$L_{\mathrm{JW}}=\alpha L_{\mathrm{Cro}}+\beta L_{\mathrm{Foc}} \quad (14)$$

式中: ψ为像素总和; y_i和p_i分别为像素i的真实标签和预测标签; ξ_i和γ为调节因子; L_{Cro}和L_{Foc}分别为交叉熵损失函数和焦点损失函数; α和β为加权调节因子, 取值分别为 0.7 和 0.3, 取值对比如表 2 所示。

表 2　加权调节因子不同取值在 DRIVE 数据集上的对比

Table 2　Comparison of different values of weighted regulator on DRIVE dataset

α	β	AUC	准确度	F_1	敏感度	特异性
0.1	0.9	0.987 9	0.970 8	0.827 5	0.798 9	0.987 3
0.2	0.8	0.987 3	0.970 6	0.823 7	0.783 3	0.988 6
0.3	0.7	0.987 1	0.970 6	0.823 8	0.785 1	0.988 4
0.4	0.6	0.986 7	0.969 8	0.817 1	0.771 2	0.988 8
0.5	0.5	0.988 1	0.970 9	0.827 9	0.798 5	0.987 5
0.6	0.4	0.988 0	0.971 0	0.828 6	0.801 1	0.987 3
0.7	0.3	0.988 0	0.970 9	0.830 0	0.810 9	0.986 3
0.8	0.2	0.987 4	0.970 2	0.824 0	0.795 8	0.987 0
0.9	0.1	0.987 9	0.970 9	0.827 2	0.794 3	0.987 9

2.6　与不同算法血管分割结果客观分析

为验证本文算法的优越性, 将算法 U-Net[7]、Attention U-Net[8]、Dense U-Net[9]、AA-Net[10] 和本文算法置于同一实验环境进行对比, 实验结果如表 3 所示。

表 3 表示在数据集 DRIVE、STARE 和 CHASE_DB1 上不同算法之间的性能对比。可知, 在 DRIVE 数据集中, 本文算法除特异性指标外, 其他性能指标均达到最优值, 说明本文算法在 DRIVE 数据集中整体分割性能优越。在 STARE 数据集中, AA-Net 的敏感度指标达到最优, Attention U-Net 的特异性指标达到最优, 其他指标的最优值均为本文算法, 说明本文算法对视网膜血管的准确感知能力、微细血管分割方面比较稳健。在 CHASE_DB1 数据集中, AA-Net 的敏感度指标达到最优, 其他指标的最优值均为本文算法, 说明本文算法性能提升明显、分割效果理想。

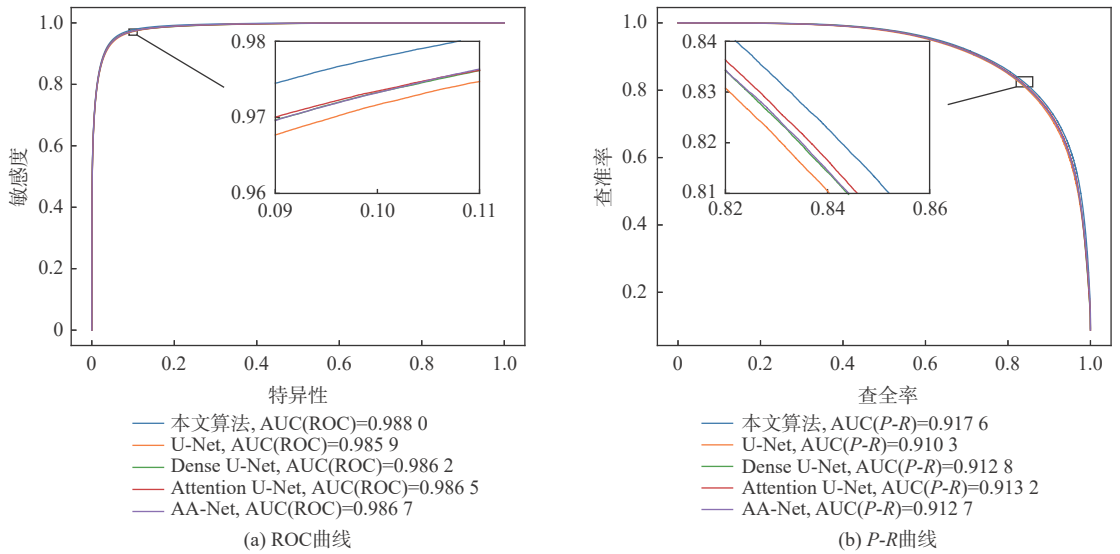
图 9 和图 10 为数据集 DRIVE 和 CHASE_DB1 上 ROC 曲线与查准率-查全率(P-R)曲线的对比。AUC 值范围为 0～1, AUC 值越大, 表示模型的预测能力越强, 而 AUC 值越小, 表示预测模型的错误分类概率越高。观察图 9 和图 10 可知, 本文算法的 AUC(ROC)和 AUC(PR)值均明显优于其他对比算法。在图 9 中, 本文算法的 AUC(ROC)和 AUC(P-R)值较其他算法平均高出 0.001 7 和 0.005 4; 在图 10 中, 本文算法的 AUC(ROC)和 AUC(P-R)值较其他算法平均高出 0.001 9 和 0.013 0。说明本文算法能有效提升对微细血管分割的准确率, 减少血管误分割现象。

图 11 给出数据集 DRIVE、STARE 和 CHASE_DB1 上本文算法与文献 [7-10] 算法视网膜血管分割的效果。其中, 第 1 行表示 DRIVE 数据集中健康眼底图像, 第 2 行表示 DRIVE 数据集中病变眼底图像, 第 3 行表示 STARE 数据集中健康眼底图

表 3　不同算法血管分割性能对比

Table 3　Performance comparison of different algorithms for blood vessel segmentation

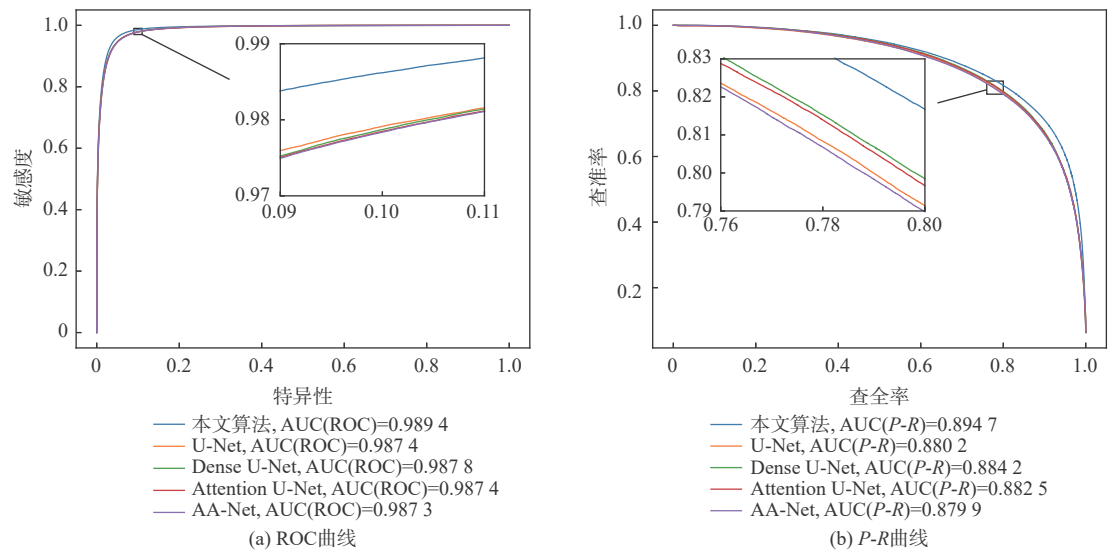
数据集	算法	AUC(ROC)	AUC(P-R)	F_1	准确度	敏感度	特异性
DRIVE	U-Net[7]	0.985 9	0.910 3	0.822 8	0.970 1	0.793 2	0.987 1
	Attention U-Net[8]	0.986 5	0.913 2	0.825 1	0.970 5	0.795 3	0.987 3
	Dense U-Net[9]	0.986 2	0.912 8	0.825 1	0.970 4	0.797 9	0.986 9
	AA-Net[10]	0.986 7	0.912 7	0.824 2	0.970 3	0.794 8	0.987 2
	本文	0.988 0	0.917 6	0.830 0	0.970 9	0.810 9	0.986 3
STARE	U-Net[7]	0.989 3	0.917 0	0.826 5	0.974 9	0.786 2	0.990 4
	Attention U-Net[8]	0.989 1	0.915 1	0.822 0	0.974 5	0.774 3	0.991 0
	Dense U-Net[9]	0.989 1	0.915 8	0.827 2	0.974 9	0.792 7	0.989 8
	AA-Net[10]	0.990 1	0.917 4	0.827 6	0.974 6	0.801 8	0.988 8
	本文	0.990 7	0.922 0	0.831 8	0.975 4	0.801 0	0.989 7
CHASE_DB1	U-Net[7]	0.987 4	0.880 2	0.794 5	0.974 4	0.784 6	0.987 2
	Attention U-Net[8]	0.987 4	0.882 5	0.798 0	0.974 6	0.795 0	0.986 7
	Dense U-Net[9]	0.987 8	0.884 2	0.799 4	0.974 5	0.805 0	0.985 9
	AA-Net[10]	0.987 3	0.879 9	0.794 9	0.973 2	0.824 6	0.983 2
	本文	0.989 4	0.894 7	0.809 3	0.976 0	0.807 9	0.987 3



(a) ROC曲线 (b) *P-R*曲线

图 9 DRIVE 数据集上 ROC 曲线和 *PR* 曲线对比

Fig. 9 Comparison of ROC curves and *PR* curves on DRIVE dataset



(a) ROC曲线 (b) *P-R*曲线

图 10 CHASE_DB1 数据集上 ROC 曲线和 *PR* 曲线对比

Fig. 10 Comparison of ROC curves and *PR* curves on CHASE_DB1 dataset

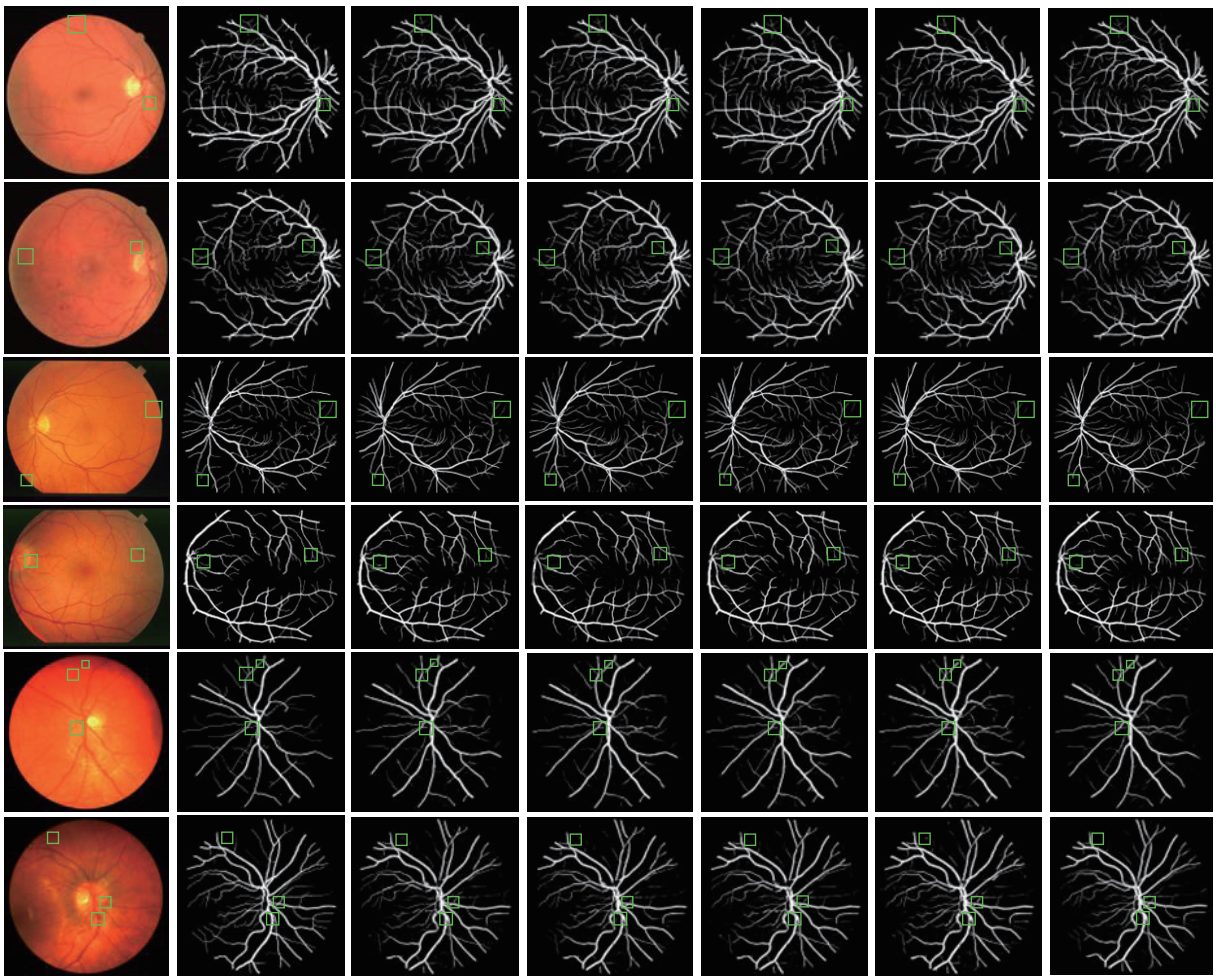
像，第 4 行表示 STARE 数据集中病变眼底图像，第 5 行表示 CHASE_DB1 数据集中健康眼底图像，第 6 行表示 CHASE_DB1 数据集中病变眼底图像。

从图 11 第 1、2 行 DRIVE 数据集眼底图像中可以看出，本文算法、Dense U-Net、Attention U-Net 和 AA-Net 分割的血管数目较多，Attention U-Net 和 AA-Net 分割的血管数目相比本文算法和 Dense U-Net 相对较少，本文算法与 Dense U-Net 分割的血管数目虽多，但 Dense U-Net 将一定量的病理信息误分割成了血管信息，因此，本文算法在视网膜血管分割时具有更好的稳健性。从图 11 第 3、4 行 STARE 数据集眼底图像中可以看出，本文算法与 AA-Net 分割的微细血管数量比其他算法多，由此说明本文算法与 AA-Net 的血管分割能较强，但 AA-Net 分割的血管准确性欠佳，说明该算法在处理非血管背景与病理干扰信息方面不如本文算法。从图 11 第 5、6 行 CHASE_DB1 数据集眼底图像中可以看出，本文算法对血管的准确分割层面显著优于其他对比算法。在微细血管分割数目层面，自适应聚合特征信息算法最多，但其在视盘位置处血管分割稳健性较弱，说明 AA-Net 的抗干扰能力较弱。

2.7 与不同算法血管分割结果主观分析

为更细致地分析分割结果，图 12 展示了不同算法视网膜血管分割结果的细节对比。图 12 中，第 1、2、3 行分别为图 11 第 1、3、5 行的细节放大图。

观察图 12 可知，在第 1 行的细节图中，图 12(d)～图 12(g)均出现了不同程度的微细血管分割模糊和



(a) 原图　(b) 第1位专家金标准图　(c) U-Net[7]　(d) Attention U-Net[8]　(e) Dense U-Net[9]　(f) AA-Net[10]　(g) 本文算法

图 11　不同算法视网膜血管分割结果

Fig. 11　Results of retinal blood vessel segmentation by different algorithms

断裂的问题, 而图 12(h)细节图同金标准细节图像基本一致; 在第 2 行的细节图中, 图 12(d)～图 12(g)出现了边界模糊、非血管背景像素分辨不明和微细血管断裂的问题, 而图 12(h)细节图中非血管背景像素分明, 误分割率低; 在第 3 行的细节图中, 图 12(d)～图 12(g)存在噪声干扰和主血管断裂的问题, 而图 12(h)细节图整体分割流畅。

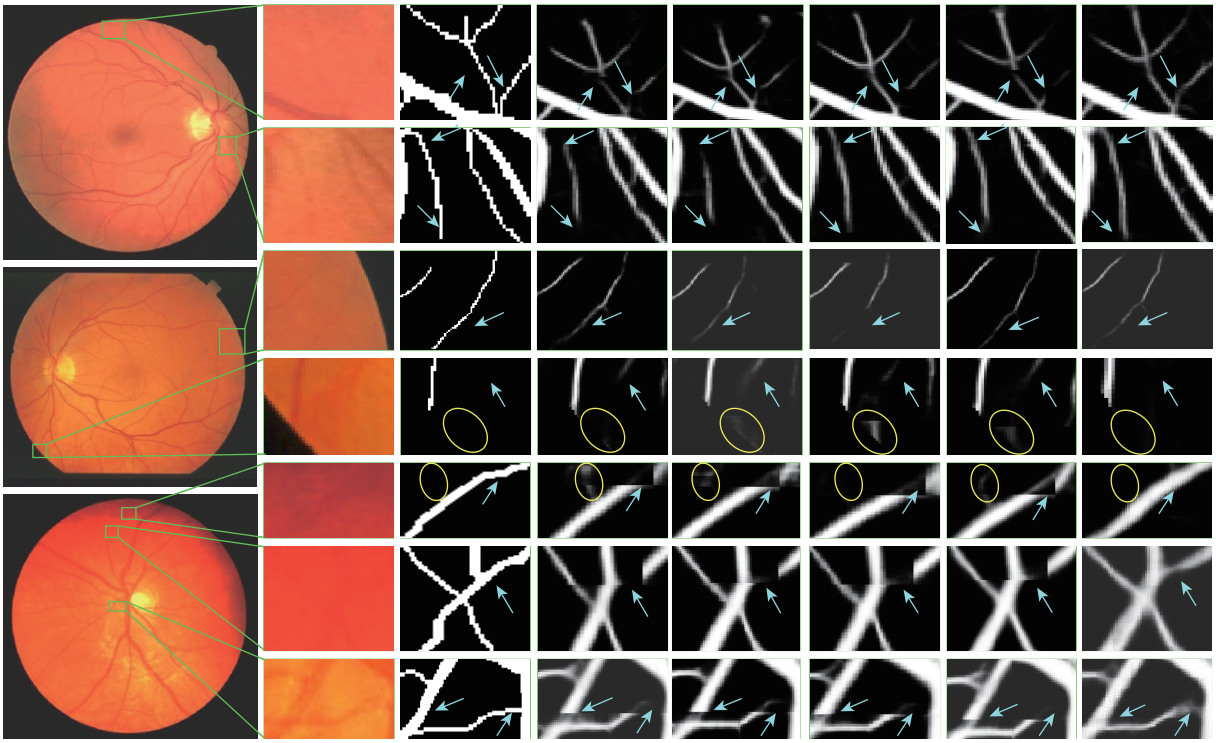
实验表明, 本文算法在处理复杂血管和薄弱血管结构时具有更理想的分割性能, 其分割结果优于其他对比算法。

2.8　与近年来其他算法性能对比

为进一步体现本文算法在视网膜血管分割任务上性能的稳健性, 表 4～表 6[24-34] 分别列出本文算法与近年其他算法在数据集 DRIVE、STARE 和 CHASE_DB1 上的性能指标对比, 数据均来自原文。

观察表 4 知, 在 DRIVE 数据集的指标对比中, 文献 [28] 算法特异性为最优值, 但其他指标均低于本文算法。文献 [32] 算法的敏感度为 0.849 1, 为该项的最优值, 但准确度和 AUC 指标均明显低于本文算法。另外, 本文算法具有极高的准确度和 AUC 值, 分别为 0.970 9 和 0.988 0, 明显高出近年的其他算法。观察表 5 可知, 在 STARE 数据集的指标对比中, 文献 [28] 算法特异性值为 0.994 5, 到达该项的最优值, 比本文算法高 0.004 8, 但本文算法的准确度和 AUC 值显著高于文献 [28] 算法, 分别高出 0.011 3 和 0.028 7, 说明本文算法在血管的正确分类层面明显优于文献 [28] 算法。文献 [32] 算法的敏感度为 0.857 3, 是该项指标的最优值, 但其准确度、特异性和 AUC 的值均低于本文算法, 说明本文算法能更精确地分割血管, 整体性能优越。观察表 6 可知, 在 CHASE_DB1 数据集的指标对比中, 敏感度最优值为文献 [32] 算法, 虽然其在敏感度方面显著优于本文算法, 但本文算法除敏感度指标外, 其他指标均为各项的最优值, 说明本文算法在视网膜血管分割方面具有较强的可靠性。

综上, 本文算法具有较高的特异性、准确度和 AUC, 说明其泛化能力较强, 也体现在视网膜血管分割任务上的性能稳健。通过与近年来其他算法



(a) 原图 (b) 原图细节 (c) 第1位专家金标准图 (d) U-Net[7] (e) Attention U-Net[8] (f) Dense U-Net[9] (g) AA-Net[10] (h) 本文算法

图 12 不同算法视网膜血管分割结果细节对比

Fig. 12 Detail comparison of retinal vessel segmentation results by different algorithms

表 4 与近年来不同算法在 DRIVE 数据集上的性能指标对比

Table 4 Compared with performance indexes of different algorithms on DRIVE dataset in recent years

算法	年份	AUC	准确度	敏感度	特异性
文献[24]	2019	0.975 0	0.953 8	0.763 1	0.982 0
文献[25]	2020	0.982 3	0.958 1	0.799 1	0.981 3
文献[26]	2021	0.967 8	0.948 0	0.735 2	0.977 5
文献[27]	2021	0.987 1	0.969 2	0.825 8	0.982 9
文献[28]	2021	0.974 8	0.951 2	0.806 0	0.986 9
文献[29]	2021	0.983 7	0.969 7	0.828 9	0.983 8
文献[30]	2023	0.982 0	0.958 0	0.813 9	0.981 8
文献[31]	2022	0.975 0	0.951 8	0.758 0	0.980 4
文献[32]	2023	0.985 0	0.965 9	0.849 1	0.977 4
文献[33]	2022		0.961 0	0.812 5	0.976 3
文献[34]	2023	0.980 7	0.956 8	0.805 4	0.978 9
本文算法	2023	0.988 0	0.970 9	0.810 9	0.986 3

表 5 与近年来不同算法在 STARE 数据集上的性能指标对比

Table 5 Compared with performance indicators of different algorithms in STARE dataset in recent years

算法	年份	AUC	准确度	敏感度	特异性
文献[24]	2019	0.983 3	0.963 8	0.773 5	0.985 7
文献[25]	2020	0.988 1	0.967 3	0.818 6	0.984 4
文献[26]	2021	0.968 6	0.954 8	0.726 5	0.975 9
文献[28]	2021	0.962 0	0.964 1	0.823 0	0.994 5
文献[29]	2021	0.987 7	0.973 6	0.820 7	0.983 9
文献[30]	2023	0.990 2	0.971 2	0.797 2	0.977 9
文献[31]	2022	0.989 3	0.968 3	0.774 7	0.991 0
文献[32]	2023	0.987 3	0.971 9	0.857 3	0.981 3
文献[33]	2022		0.958 6	0.807 8	0.972 1
文献[34]	2023	0.985 0	0.964 8	0.839 7	0.979 5
本文算法	2023	0.990 7	0.975 4	0.801 0	0.989 7

的性能对比，表明本文算法能精确地分割出更多的血管像素，鲁棒性更强。

2.9 模型消融实验

表 7 给出了本文模型与其他不同模型的参数量和浮点运算数量的对比。可知，本文算法在浮点运算数量与其他算法在同一水平的情况下，其模型参数量为 3.236×10^6，远低于 U-Net、AttentionU-Net 和 DenseU-Net 算法。说明在浮点运算数量为同一水平时，本文算法拥有更低的模型参数量，是一个低参数量模型。

为验证本文算法所提模型的有效性和合理性，在 DRIVE 数据集上进行消融实验，测试的 AUC 值、F_1 值、准确度、敏感度和特异性如表 8 所示。其中，M1 表示 U-Net，M2 表示在 M1 中 U 结构底部融入 FAF 模块，M3 表示在 M2 中编解码部分设计 IGC 模块，M4 表示在 M3 中融合 IMV 模块，M5 表

表 6　与近年来不同算法在 CHASE_DB1 数据集上的性能指标对比

Table 6　Compared with performance indicators of different algorithms in CHASE_DB1 dataset in recent years

算法	年份	AUC	准确度	敏感度	特异性
文献[24]	2019	0.977 6	0.960 7	0.764 1	0.980 6
文献[25]	2020	0.987 1	0.967 0	0.823 9	0.981 3
文献[26]	2021	0.968 1	0.945 2	0.727 9	0.965 8
文献[27]	2021	0.988 4	0.974 5	0.822 7	0.985 3
文献[29]	2021	0.986 7	0.974 4	0.836 5	0.983 9
文献[30]	2023	0.986 7	0.966 9	0.819 4	0.981 7
文献[32]	2023	0.986 9	0.973 1	0.860 7	0.980 6
文献[33]	2022		0.957 8	0.801 2	0.973 0
文献[34]	2023	0.983 6	0.963 5	0.824 0	0.977 5
本文算法	2023	0.989 4	0.976 0	0.807 9	0.987 3

表 7　模型参数量与浮点运算数量对比

Table 7　Comparison of number of model parameters and number of floating-point operations

模型	参数量	浮点运算量/GFLOPs
U-Net[7]	34.526×10^6	4.087
Attention U-Net[8]	34.877×10^6	4.156
Dense U-Net[9]	7.553×10^6	9.479
AA-Net[10]	4.508×10^6	2.657
本文算法	3.236×10^6	4.623

表 8　模型消融实验

Table 8　Model ablation experiment

方法	AUC	F_1	准确度	敏感度	特异性
M1	0.985 9	0.822 8	0.970 1	0.793 2	0.987 1
M2	0.987 0	0.826 5	0.970 6	0.799 6	0.987 0
M3	0.987 7	0.826 5	0.970 9	0.791 3	0.988 1
M4	0.987 6	0.827 0	0.970 7	0.799 3	0.987 2
M5	0.988 0	0.830 0	0.970 9	0.810 9	0.986 3

示在 M4 的连接处加入 MA 模块, 构成本文算法。

由表 8 可知, 传统 U-Net 算法对小数据的医学图像具有较强的可靠性, 在模型结构未改进前, 其准确度和特异性已经达到 0.970 1 和 0.987 1, 但存在 AUC 值、F_1 值和敏感度较低的现象, 于是在传统 U-Net 算法的基础上进行改进。通过在 M1 的底部加入 FAF 模块, F_1 值和敏感度显著提升, 说明 FAF 模块对血管尺度信息和形态结构具有较强的捕捉能力。在 M2 的基础上设计 IGC 编解码模块, 使得特异性、AUC 值和准确度有所提升, 其中准确度达到该项最优, 说明 IGC 编解码结构可有效恢复图像的细节信息, 有助于保持目标结构特征的良好拓扑结构, 进而实现视网膜血管的有效分割。M3 的准确度和特异性虽然都已经达到最优值, 但敏感度指标下降明显, 于是在 M3 的编码部分设计 IMV 模块, 实现视网膜血管特征的双提取, 从 M4 可知, 敏感度提升 0.008 0, 达到 0.799 3。在 M4 的连结部分加入 MA 模块, AUC、F_1 值、敏感度和准确度均达到各项最优值。

本文将上述结构进行合理融合, 尽可能发挥各模块的优势, 使得 AUC 值、F_1 值、准确度、敏感度和特异性分别达到 0.988 0、0.830 0、0.970 9、0.810 9 和 0.986 3, 说明本文算法整体有效且合理。

3　结　论

针对视网膜血管图像分割时存在边界分割模糊、背景渗出物干扰和微小血管分割难等问题, 本文提出融合 Mobile Vit 和倒置门控编解码的视网膜血管分割算法。

1) 对眼底视网膜图像进行对比度增强处理, 抑制背景渗出物干扰, 再扩张图像数量以避免模型训练过拟合。

2) 利用改进的 Mobile Vit 模块和倒置门控编解码模块构建双联合编码与解码结构, 促进对视网膜血管特征的提取和重建, 多谱注意力模块充分建立编解码之间的频域逻辑, 特征自适应融合模块挖掘深层语义信息, 最大限度减少信息流丢失, 动态搭建上下文特征依赖关联。

3) 联合加权损失函数抑制前景与背景像素分布不平衡引起的分割误差, 因而使得分割效果更为理想。

本文算法在数据集 DRIVE、STARE 和 CHASE_DB1 数据集上具有较高的 AUC 值和分割准确度, 表明算法对视网膜血管的预测与分割准确度高, 对眼科疾病的预防和诊断具有一定应用价值。

参考文献 (References)

[1] 郑宇祥, 郝鹏翼, 吴冬恩, 等. 结合多层特征及空间信息蒸馏的医学影像分割[J]. 北京航空航天大学学报, 2022, 48(8): 1409-1417.
ZHENG Y X, HAO P Y, WU D E, et al. Medical image segmentation based on multi-layer features and spatial information distillation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(8): 1409-1417(in Chinese).

[2] 梁礼明, 余洁, 陈鑫, 等. 跨级融合门控自适应网络用于视网膜血管分割[J]. 北京航空航天大学学报, 2024, 50(4): 1097-1109.
LIANG L M, YU J, CHEN X, et al. Cross-level fusion gated adaptive network for retinal vessel segmentation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(4): 1097-1109 (in Chinese).

[3] BEKKERS E, DUITS R, BERENDSCHOT T, et al. A multi-orientation analysis approach to retinal vessel tracking[J]. Journal of Mathematical Imaging and Vision, 2014, 49(3): 583-610.

[4] LI Q, YOU J, ZHANG D. Vessel segmentation and width estimation in retinal images using multiscale production of matched filter responses[J]. Expert Systems with Applications, 2012, 39(9): 7600-7610.

[5] 梁礼明, 黄朝林, 石霏, 等. 融合形状先验的水平集眼底图像血管分割[J]. 计算机学报, 2018, 41(7): 1678-1692.
LIANG L M, HUANG C L, SHI F, et al. Retinal vessel segmentation using level set combined with shape priori[J]. Chinese Journal of Computers, 2018, 41(7): 1678-1692(in Chinese).

[6] ZHANG J, CHEN Y, BEKKERS E, et al. Retinal vessel delineation using a brain-inspired wavelet transform and random forest[J]. Pattern Recognition, 2017, 69: 107-123.

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

[8] OKTAY O, SCHLEMPER J, LE FOLGOC L, et al. Attention U-Net: learning where to look for the pancreas[EB/OL]. (2018-05-20)[2023-02-01]. https://arxiv.org/abs/1804.03999.

[9] WANG C, ZHAO Z Y, REN Q Q, et al. Dense U-Net based on patch-based learning for retinal vessel segmentation[J]. Entropy, 2019, 21(2): 168.

[10] LIANG L M, FENG J, ZHOU L S, et al. U-shaped retinal vessel segmentation based on adaptive aggregation of feature information[J]. Interdisciplinary Sciences: Computational Life Sciences, 2022, 14(2): 623-637.

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

[12] MEHTA S, RASTEGARI M. MobileViT: light-weight, general-purpose, and mobile-friendly vision transformer[EB/OL]. (2022-03-04)[2023-02-01]. https://arxiv.org/abs/2110.02178?context=cs.

[13] JIANG Y, LIANG J, CHENG T T, et al. MTPA_Unet: multi-scale Transformer-position attention retinal vessel segmentation network joint Transformer and CNN[J]. Sensors, 2022, 22(12): 4592.

[14] RUAN J C, XIANG S C, XIE M Y, et al. MALUNet: a multi-attention and light-weight UNet for skin lesion segmentation[C]//Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. Piscataway: IEEE Press, 2022: 1150-1156.

[15] 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.

[16] HE K M, ZHANG X Y, REN S Q, et al. Identity mappings in deep residual networks[C]//Proceedings of the14th European Conference. Berlin: Springer, 2016: 630-645.

[17] QIN Z Q, ZHANG P Y, WU F, et al. FcaNet: frequency channel attention networks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 783-792.

[18] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 4700-4708.

[19] LIAO T Y, YANG C H, LO Y, et al. HarDNet-DFUS: an enhanced harmonically-connected network for diabetic foot ulcer image segmentation and colonoscopy polyp segmentation[EB/OL]. (2022-09-15)[2023-02-01]. https://arxiv.org/abs/2209.07313.

[20] STAAL J, ABRÀMOFF M D, NIEMEIJER M, et al. Ridge-based vessel segmentation in color images of the retina[J]. IEEE Transactions on Medical Imaging, 2004, 23(4): 501-509.

[21] HOOVER A, KOUZNETSOVA V, GOLDBAUM M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response[J]. IEEE Transactions on Medical Imaging, 2000, 19(3): 203-210.

[22] OWEN C G, RUDNICKA A R, MULLEN R, et al. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program[J]. Investigative Ophthalmology & Visual Science, 2009, 50(5): 2004-2010.

[23] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2980-2988.

[24] YAN Z Q, YANG X, CHENG K T. A three-stage deep learning model for accurate retinal vessel segmentation[J]. IEEE Journal of Biomedical and Health Informatics, 2019, 23(4): 1427-1436.

[25] WANG D Y, HAYTHAM A, POTTENBURGH J, et al. Hard attention net for automatic retinal vessel segmentation[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(12): 3384-3396.

[26] TCHINDA B S, TCHIOTSOP D, NOUBOM M, et al. Retinal blood vessels segmentation using classical edge detection filters and the neural network[J]. Informatics in Medicine Unlocked, 2021, 23: 100521.

[27] HU X L, WANG L J, CHENG S L, et al. HDC-Net: a hierarchical dilation convolutional network for retinal vessel segmentation[J]. PLoSOne, 2021, 16(9): e0257013.

[28] WANG C, ZHAO Z Y, YU Y. Fine retinal vessel segmentation by combining Nest U-net and patch-learning[J]. Soft Computing, 2021, 25(7): 5519-5532.

[29] WU H S, WANG W, ZHONG J F, et al. SCS-net: a scale and context sensitive network for retinal vessel segmentation[J]. Medical Image Analysis, 2021, 70: 102025.

[30] LI D Y, PENG L X, PENG S H, et al. Retinal vessel segmentation by using AFNet[J]. The Visual Computer, 2023, 39(5): 1929-1941.

[31] XU X M, WANG Y X, LIANG Y, et al. Retinal vessel automatic segmentation using SegNet[J]. Computational and Mathematical Methods in Medicine, 2022, 2022: 3117455.

[32] AURANGZEB K, ALHARTHI R S, HAIDER S I, et al. An efficient and light weight deep learning model for accurate retinal vessels segmentation[J]. IEEE Access, 2023, 11: 23107-23118.

[33] KHAN T M, KHAN M A U, REHMAN N U, et al. Width-wise vessel bifurcation for improved retinal vessel segmentation[J]. Biomedical Signal Processing and Control, 2022, 71: 103169.

[34] 梁礼明, 冯骏, 彭仁杰, 等. 融合多标签损失与双注意力的U型视网膜血管分割[J]. 计算机辅助设计与图形学学报, 2023, 25(1): 75-86.
LIANG L M, FENG J, PENG R J. et al. U-shaped retinal vessel segmentation combining multi-label loss and dual attention[J]. Journal of Computer-Aided Design & Computer Graphics, 2023, 25(1): 75-86(in Chinese) .

Fusion of Mobile Vit and inverted gated codec retinal vessel segmentation algorithm

LIANG Liming，YANG Yuan，ZHU Chenkun，HE Anjun，WU Jian*

(School of Electrical Engineering and Automation，Jiangxi University of Science and Technology，Ganzhou 341000，China)

Abstract: An algorithm based on Mobile Vit and inverted gated codec is proposed for retinal vessel segmentation (FMVG-Net), aiming to tackle issues such background noise interference, boundary texture blurring, and challenging extraction of microvascular areas. First, we improve the Mobile Vit module, and realize the double joint feature extraction cleverly in the coding part. Subsequently, we use the multispectral attention module. The module reduces the missing image feature information from the frequency domain dimension, so as to accurately segment the foreground pixel of the vessel. Next, we propose a feature adaptive fusion module to establish the context dependence of vascular texture and improve the sensitivity of vascular segmentation. In order to enhance the precision of retinal vascular picture segmentation, lastly, we have implemented an inverted gated codec module and optimized the codec structure to further collect deep and spatial semantic information. Experiments are performed on the DRIVE, STARE, and CHASE_DB1 datasets,whereby the obtained specificity are 0.986 3, 0.989 7, and 0.987 3, respectively; the accuracies are 0.970 9, 0.975 4, and 0.976 0, respectively; the sensitivity is 0.810 9, 0.801 0, and 0.807 9, respectively.Simulation experiments reveal that this paper demonstrates a superior segmentation effect on eye lesions images, which opens new avenues for the diagnosis of eye diseases.

Keywords: retina vessel; Mobile Vit module; discrete cosine transform; inverted gated codec module; feature adaptive fusion

Received：2023-02-28；**Accepted**：2023-06-16；**Published Online**：2023-06-29 17：47

URL：link.cnki.net/urlid/11.2625.V.20230629.1317.001

Foundation items：National Natural Science Foundation of China (51365017,61463018); General Project of Jiangxi Provincial Natural Science Foundation (20192BAB205084); Science and Technology Research Key Project of Jiangxi Provincial Department of Education (GJJ170491,GJJ2200848); Jiangxi Provincial Graduate Innovation Special Fund (YC2022-S676)

* **Corresponding author**. E-mail：yangy27980218@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0148

改进 5G-R 自适应高速铁路越区切换算法

陈永*，康婕，陶瑄
(兰州交通大学 电子与信息工程学院，兰州 730070)

摘 要：在高速行车条件下，越区切换作为未来高速铁路 5G-R 通信的关键技术，对于保障行车安全至关重要。下一代高速铁路 5G-R 无线通信系统越区切换算法采用固定切换参数，但当列车高速运行时，极易受到多普勒效应影响，导致切换成功率低，基于此，提出了一种考虑多普勒频移影响的改进 5G-R 自适应高速铁路越区切换算法。分析多普勒频移对切换成功率的影响，得到多普勒频移与切换成功率的关系函数；提出考虑多普勒频移影响的越区切换动态函数，设计余弦、余切、余割 3 种函数对切换迟滞门限及触发时延自适应调整；在不同多普勒频移及不同高铁场景下进行切换成功率的量化比较分析。研究结果表明：所提算法可有效调高切换成功率，在高架桥和山区场景下，余弦、余切、余割 3 种函数的切换成功率均优于对比算法，且满足中国无线通信系统切换成功率服务质量 (QoS) 大于 99.5% 的要求。研究结果为下一代高速铁路 5G-R 无线通信系统演进提供了理论参考依据。

关 键 词：越区切换算法；5G-R；多普勒频移；动态函数；自适应切换
中图分类号：TP301.6; TP391.9
文献标志码：A **文章编号**：1001-5965(2025)03-0724-08

目前，中国高速铁路采用铁路全球移动通信系统 (global system for mobile communications-railway, GSM-R) 作为无线通信系统，承载着大量列车的安全控制业务。但随着高速铁路的快速发展，GSM-R 窄带通信系统难以满足高速铁路高承载量、高可靠性及业务多样化的需求。随着 5G 技术上升为国家战略，中国铁路无线通信系统将逐步从 GSM-R 演进到 5G-R[1]。5G-R 具有高带宽、高速度、大容量、低时延等优点。然而，在列车高速运行环境下，列车与发射基站相对运动，发射信号频率极易偏移，出现多普勒频移现象，造成通信质量恶化，导致越区切换成功率下降，严重时将影响列车运行控制系统的安全[2]。因此，高速行车条件下，如何实现快速切换以提高列车越区切换成功率成为目前研究的热点问题。

现有文献针对越区切换展开了大量研究。Qian 等[3] 提出了一种基于模糊逻辑控制的切换判决算法，选取接收信号强度、相对运动角度等参数进行优化，但存在参数主观性难以量化的问题，影响最终切换性能。Davaasambuu 等[4] 提出了考虑列车速度和单元切换性能指标(切换失败率、乒乓切换率)的迟滞门限自适应算法，但存在高速环境下触发时延为定值的问题。Duan 等[5] 提出了一种基于列车运行速度和运动方向的快速切换算法，但迟滞门限为定值，难以适应动态铁路的运行场景。Angri 等[6] 提出了一种基于可用带宽和小区信号强度的切换算法，但存在切换迟滞门限固定的问题。Chen 等[7] 提出了一种基于椭圆函数的联合判决算法来提高切换成功率，但存在未考虑多普勒频移对越区切换影响的问题。陈永刚等[8] 提出了一种通过速度函

收稿日期：2023-03-28；**录用日期**：2023-05-19；**网络出版时间**：2023-06-28 15:55
网络出版地址：link.cnki.net/urlid/11.2625.V.20230628.0939.001
基金项目：国家自然科学基金 (62462043,61963023)；兰州交通大学重点研发项目 (ZDYF2304)
* **通信作者**. E-mail：edukeylab@126.com

引用格式：陈永，康婕，陶瑄. 改进 5G-R 自适应高速铁路越区切换算法 [J]. 北京航空航天大学学报，2025，51（3）：724-731.
CHEN Y，KANG J，TAO X. An improved 5G-R adaptive high-speed railway handover algorithm [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（3）：724-731（in Chinese）.

数对迟滞门限和延迟触发时延动态调整的切换算法，但未考虑高速环境下多普勒频移对越区切换的影响。

综上所述，现有大部分越区切换算法采用固定阈值进行切换，且未考虑高速行车条件下多普勒效应对越区切换的影响。针对上述问题，本文提出了一种基于多普勒频移影响的 5G-R 自适应高速铁路越区切换算法。主要工作有：①通过量化分析得到了多普勒频移与切换成功率之间的函数关系。②提出了考虑多普勒频移因素的改进 5G-R 自适应越区切换算法，设计了动态函数自适应调整切换阈值，实现对高速铁路越区切换的自适应控制。③对不同多普勒频移及不同高铁场景下越区切换性能进行了比较分析，实验结果表明，本文算法切换成功率均能满足中国高速铁路无线通信系统服务质量 (quality of service, QoS) 对于越区切换要求功率大于 99.5% 的要求。

1　多普勒频移对越区切换影响性分析

在 5G-R 越区切换时，一般选择 A3 事件作为切换算法。A3 切换算法在进行切换判决时，当待切换目标基站小区的参考信号接收功率 (reference signal received power, RSRP) 与当前服务基站小区的 RSRP 的差值超过迟滞门限时，触发 A3 事件，若列车在触发时延 (time-to-trigger, TTT) 内都满足上述条件，则进行越区切换[9]。传统 A3 切换算法使用固定迟滞门限和固定触发时延[10]。然而，在高速移动场景下，多普勒频移会造成信道的时间选择性衰落，影响通信质量，导致切换性能降低[11]。当列车速度为 v，基站发射的信号到达列车载终端时，受到的多普勒频移表示如下：

$$f_{\mathrm{d}}=\frac{f_{\mathrm{c}}}{c}v\cos\theta \tag{1}$$

式中：f_{d}为多普勒频移，Hz；f_{c}为载波频率，Hz；c为光速，取值为 3×10^{8} m/s；θ为信号传输方向与列车运行方向之间的夹角，(°)；v为列车速度，m/s。

列车运行过程中受到的多普勒频移示意图如图 1 所示。图中：列车距源基站间的水平距离为 x，

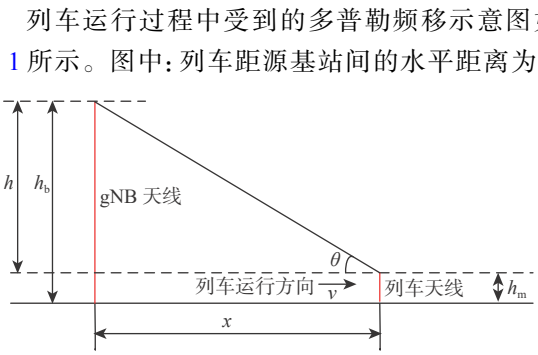


图 1　多普勒频移参数示意图

Fig. 1　Schematic of Doppler shift parameters

m；基站天线高度为h_{b}，m；列车天线高度为h_{m}，m。

根据 3GPP 规定[9]，使用高铁车顶天线，包括 4 组天线辐射单元，每组天线辐射单元为由 2 个对称阵子组成的二元阵，每个对称阵子的轴向与高铁的行进方向垂直，且 2 个对称阵子在垂直高铁的行进方向上成直线排布。天线间的高度差$h=h_{\mathrm{b}}-h_{\mathrm{m}}$，则$\theta$可以求得

$$\theta=\arctan(h/x) \tag{2}$$

将式 (2) 代入式 (1)，得

$$f_{\mathrm{d}}=\frac{f_{\mathrm{c}}}{c}v\cos(\arctan(h/x)) \tag{3}$$

由于光速 c、天线高度差 h 为定值，多普勒频移仅与列车速度v和源基站距离 x 有关。在式 (3) 中，由于基站天线高度差h为较小定值，随着列车远离源基站 x 值逐渐增大，$\arctan(h/x)$快速变小，此时$\cos(\arctan(h/x))$的值趋近于 1，式 (3) 化简为

$$f_{\mathrm{d}}=\frac{f_{\mathrm{c}}}{c}v \tag{4}$$

从式 (4) 可以得出，列车速度与多普勒频移成正比关系。在高速行车条件下，多普勒频移会引起子载波间干扰，导致越区切换的性能下降[12]。为判断切换性能，以切换成功率作为评价指标，切换成功率的计算公式如下[13]：

$$P_{\mathrm{s}}(x)=P_{\mathrm{handover}}(x)P_{\mathrm{exe}}(x) \tag{5}$$

式中：P_{handover}为切换触发概率；P_{exe}为切换执行率。

列车在位置x处的切换执行率P_{exe}的计算公式如下[8]：

$$P_{\mathrm{exe}}(x)=\frac{1}{vt_{\mathrm{exe}}}\int_{x}^{x+vt_{\mathrm{exe}}}P(R(b,x)\geqslant T)\mathrm{d}x \tag{6}$$

式中：t_{exe}为切换执行时间，ms；$R(b,x)$为列车在位置x处，目标基站 b 的信号强度，dBm；T为信号强度阈值，dBm。

将式 (4) 代入式 (6) 中，可得到多普勒频移和切换执行率$P_{\mathrm{exe}}(x)$之间的关系为

$$P_{\mathrm{exe}}(x)=\frac{f_{\mathrm{c}}}{f_{\mathrm{d}}ct_{\mathrm{exe}}}\int_{x}^{x+\frac{f_{\mathrm{d}}c}{f_{\mathrm{c}}}t_{\mathrm{exe}}}P(R(b,x)\geqslant T)\mathrm{d}x \tag{7}$$

将式 (7) 代到式 (5) 中，即可得到多普勒频移f_{d}与切换成功率P_{s}之间的关系，如下：

$$P_{\mathrm{s}}(x)=P_{\mathrm{handover}}(x)\frac{f_{\mathrm{c}}}{f_{\mathrm{d}}ct_{\mathrm{exe}}}\int_{x}^{x+\frac{f_{\mathrm{d}}c}{f_{\mathrm{c}}}t_{\mathrm{exe}}}P(R(b,x)\geqslant T)\mathrm{d}x \tag{8}$$

2　本文算法

在分析得到多普勒频移与越区切换成功率定量关系函数的基础上，针对在高速移动的环境下，越区切换频繁，而传统 A3 切换算法使用固定迟滞

门限和触发时延，且未考虑多普勒频移对切换成功率的影响，本文提出了基于多普勒频移影响的5G-R自适应越区切换算法，设计了考虑多普勒影响的动态函数，来动态调整迟滞门限和触发时延，以达到动态自适应切换的目的。本文算法主要步骤包括：①切换测量；②多普勒频移动态函数设计；③切换判决；④切换执行。

2.1 切换测量

测量基站小区RSRP时，列车在位置x距离源基站a和目标基站b的距离为

$$d_a = \sqrt{x^2 + h^2} \tag{9}$$

$$d_b = \sqrt{(d_{ab} - x)^2 + h^2} \tag{10}$$

式中：d_a为列车距离源基站a的距离，m；d_b为列车距离目标基站b的距离，m；d_{ab}为相邻2个基站之间的距离，m。

列车在位置x处接收到源基站a和目标基站b的信号强度分别为

$$R(a,x) = P_t - P_L(b,d_a) - \varepsilon(0,\sigma_a) \tag{11}$$

$$R(b,x) = P_t - P_L(b,d_b) - \varepsilon(0,\sigma_b) \tag{12}$$

式中：$R(a,x)$和$R(b,x)$分别为列车在位置x处接收到的源基站a和目标基站b的信号强度，dBm；P_t为基站发射功率，dBm；$\varepsilon(0,\sigma)$为满足均值为0的高斯分布阴影衰落，dB；$P_L(a,d_a)$为信号从源基站a到列车的路径损耗，dB；$P_L(b,d_b)$为信号从目标基站b到列车的路径损耗，dB。

一般情况下，路径损耗采用COST-231Hata模型表示。COST-231Hata模型是在Okumura-Hata模型的基础上，对载波频率进行扩展，并将其适用于信号传播距离为1～20 km、基站天线有效高度为30～200 m、移动台天线有效高度为1～10 m的场景。高铁5G专网基站一般采用沿铁路线状覆盖的定向天线，基站间距小，距离铁轨较近，其基站发射机和列车接收机之间距离、基站天线和移动台天线高度等参数符合COST-231Hata模型的要求，COST-231Hata模型能很好地适应高铁场景[14]。路径损耗模型的计算公式为

$$P_L = A + B\lg x + C \tag{13}$$

式中：

$$A = 46.3 + 33.9\lg f_c - 13.82\lg h_b - h_x \tag{14}$$

$$B = 44.9 - 6.55\lg h_b \tag{15}$$

$$C = 0 \tag{16}$$

$$(h_x) = (1.1\lg f_c - 0.7)h_m - (1.56\lg f_c - 0.8) \tag{17}$$

式中：A为覆盖场景路径损耗，dB；B为环境传播损耗，dB；C为场景校正因子，dB；h_x为移动台天线高度修正因子，dB。

2.2 考虑多普勒频移的动态函数

在完成2.1节切换测量后，进行越区切换动态函数的设计。在高速运行场景下，多普勒频移与列车速度成正比关系。当多普勒频移增大，即列车速度增大时，应减小阈值以促进切换；多普勒频移减小时，应增大阈值避免不必要的切换[8]。根据这一规律，在动态函数设计时，应选取在一定区间内呈单调递减趋势的函数来满足上述规律。因此，在动态函数设计时，分别采用三角函数中的余弦、余切、余割函数，从而达到动态调整迟滞门限和触发时延的目的。余弦、余切、余割3种函数曲线如图2所示。

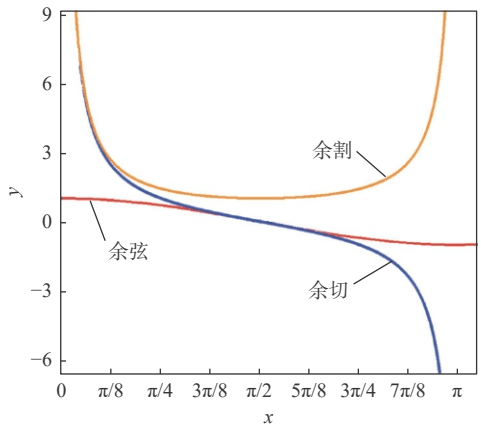


图2 动态函数曲线

Fig. 2 Dynamic function curves

图2中，余弦、余切、余割函数在[π/8,π/2]区间内均呈现出单调递减的趋势，该区间内能够符合上述高速列车迟滞门限和触发时延变化要求的规律。为使余弦、余切、余割函数在越区切换性能比较时具有可比性，需要将门限值设定在相同的范围之内。一般情况下，将迟滞门限的取值范围设置为1～5 dB，使迟滞门限可满足高速铁路无线通信系统规范的规定[15]。当多普勒频移为1 200 Hz时，迟滞门限的下限设置为1 dB，而当多普勒频移为0 Hz时，迟滞门限的上限设置为5 dB。因此，动态函数在定义域[π/8,π/2]内需要满足：① $\lim\limits_{x\to\pi/8} f(x) = 5$；② $\lim\limits_{x\to\pi/2} f(x) = 1$。按照以上要求，将余弦、余切、余割函数的上下限进行统一，得到迟滞门限动态函数参数取值，如表1所示。

同理，对触发时延进行动态调整。依据3GPP规范[9]定义，将最小触发时延和最大触发时延设置为100 ms和480 ms[16]。进行参数约束后，得到触发时延相关控制参数，如表2所示。

表 1　迟滞门限参数设置

Table 1　Hysteresis threshold parameter setting

函数名称	函数表达式	函数参数		
		j	k	z
余弦函数	$H_s = j\cos\left(f_d k + \frac{\pi}{8}\right) + z$	4/cos(π/8)	cos(π/8)/400	1
余切函数	$H_s = j\cot\left(f_d k + \frac{\pi}{8}\right) + z$	4/cot(π/8)	cot(π/8)/400	1
余割函数	$H_s = j\left(\csc\left(f_d k + \frac{\pi}{8}\right) - 1\right) + z$	4/(csc(π/8)−1)	csc(π/8)/400	1

表 2　触发时延参数设置

Table 2　Time-to-trigger parameter setting

函数名称	函数表达式	函数参数		
		j	k	z
余弦函数	$T_s = j\cos\left(f_d k + \frac{\pi}{8}\right) + z$	380/cos(π/8)	cos(π/8)/400	100
余切函数	$T_s = j\cot\left(f_d k + \frac{\pi}{8}\right) + z$	380/cot(π/8)	cot(π/8)/400	100
余割函数	$T_s = j\left(\csc\left(f_d k + \frac{\pi}{8}\right) - 1\right) + z$	380/(csc(π/8)−1)	csc(π/8)/400	100

2.3　切换判决

在设计完越区切换动态函数后，进行切换判决。切换判决时，需要满足以下 2 个条件：①目标基站小区的 RSRP 与当前基站小区的 RSRP 的差值应超过迟滞门限；②在触发时延内均满足条件①。因此，切换判决计算式 (18) 设计时，由条件①和条件②共同构成。在式 (18) 中，首先，对于条件①判断是否满足 A3 切换判决条件，即列车接收的目标基站 RSRP 与当前服务基站 RSRP 的差值是否大于迟滞门限，当满足判决条件，计算事件切换阈值比较概率 P_1。其次，对于条件②，判断列车在以速度 v 运行触发时延后，是否仍满足判决条件①，满足时计算维持触发时延事件概率 P_2；若不满足，则无法触发切换判决，不发生越区切换。因此，可得到越区切换触发概率 P_{handover}，计算公式如下[7]：

$$P_{\text{handover}}(x) = P_1(R(b,x) - R(a,x) \geqslant H_s) \cdot P_2(R(b, x - vT_s) - R(a, x + vT_s) \geqslant H_s) \tag{18}$$

式中：

$$P_1(R(b,x) - R(a,x) \geqslant H_s) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_b^2}} \cdot Q\left(\frac{H_s + P_L(b,d_b) - P_L(a,d_a) + \varepsilon_0}{\sigma_a}\right) \cdot \exp\left(-\frac{\varepsilon_0^2}{2\sigma_b^2}\right) \mathrm{d}\varepsilon_0 \tag{19}$$

$$P_2(R(b, x - vT_s) - R(a, x + vT_s) \geqslant H_s) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_b^2}} Q(q) \exp\left(-\frac{\varepsilon_0^2}{2\sigma_b^2}\right) \mathrm{d}\varepsilon_0 \tag{20}$$

其中：$q = \dfrac{H_s + P_L(b, d_b - vT_s) - P_L(a, d_a + vT_s) + \varepsilon_0}{\sigma_a}$；$\varepsilon_0$ 为阴影衰落噪声；$Q(x)$ 为正态分布右尾函数，定义为

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t \tag{21}$$

2.4　切换执行

当满足切换触发条件后，进入执行切换过程。此时，目标基站接收到的平均信号强度需大于信号强度阈值 T，则通过式 (7) 进行切换执行。其中，式 (7) 中与目标基站小区 RSRP 比较计算如下：

$$P(R(b,x) \geqslant T) = 1 - P(P_t - P_L(b,d_b) - \varepsilon(0,\sigma_b) < T) = 1 - Q\left(\frac{P_t + P_L(b,d_b) - T}{\sigma_b}\right) \tag{22}$$

当满足上述切换条件后，列车断开与现有小区基站的连接，建立与目标小区基站的通信，从而完成越区切换。

2.5　切换成功率评价

完成 2.3 节和 2.4 节越区切换过程后，对切换性能进行比较，采用切换成功率作为评价指标[7-8]。由式 (5) 可知，切换成功率 P_s 由切换触发概率 P_{handover} 和切换执行率 P_{exe} 这 2 部分组成，将式 (18) 和式 (22) 代入式 (5)，得到列车在位置 x 触发切换的切换成功率为

$$P_s(x) = P_{\text{handover}}(x) P_{\text{exe}}(x + vT_s) = P(R(b,x) - R(a,x) \geqslant H_s) \cdot P(R(b, x - vT_s) - R(a, x + vT_s) \geqslant H_s) \cdot \frac{f_c}{f_d c t_{\text{exe}}} \int_{x+vT_s}^{x + \frac{f_d c}{f_c}(t_{\text{exe}} + T_s)} P(R(b, x + vT_s) \geqslant T) \mathrm{d}x \tag{23}$$

3　仿真结果分析

3.1　仿真参数配置

为验证本文算法的有效性，将其与传统 A3 切换算法、文献 [8] 基于速度的高速铁路越区切换算法和文献 [17] 基于模糊逻辑的切换算法进行比较

分析。使用MATLAB软件对切换进行仿真分析，对比实验配置相同，相关参数如表3所示[7-8,16]。

表3 仿真参数设置[7-8,16]

Table 3 Simulation parameter setting[7-8,16]

参数	数值
载波频率f_c/MHz	2 600
基站发射功率P_t/dBm	86
相邻基站间的距离d_{ab}/m	3 000
基站天线高度h_b/m	30
列车天线高度h_m/m	3.5
阴影衰落标准差σ/dB	8
切换执行时间t_{exe}/ms	100
信号强度阈值T/dBm	−58

3.2 仿真结果

本节比较不同切换算法对越区切换迟滞门限的调整性能。传统A3切换算法、文献[8]算法、文献[17]算法与本文算法的比较结果如图3所示。可以看出，由于传统A3切换算法使用固定切换阈值，迟滞门限不随多普勒频移的变化而变化，无法做到自适应调整。文献[8]算法考虑了速度因素，采用速度动态函数对切换迟滞门限有一定的调节作用，较传统A3切换算法有一定的改进。但在同一多普勒频移条件下，文献[8]算法的迟滞门限值高于本文提出的余弦、余切、余割函数的迟滞门限值，说明在同等条件下，文献[8]算法需要较高的切换迟滞门限值，无法以较小代价响应到切换迟滞门限，其对于切换迟滞门限阈值的调节能力有限。文献[17]将信噪比、小区负载和列车速度作为参数输入模糊逻辑算法中，实现了对迟滞门限值的动态调整，但该算法选择隶属函数指标权重时存在主观不唯一的缺点，导致算法稳定性较差，出现了迟滞门限抖动的问题。

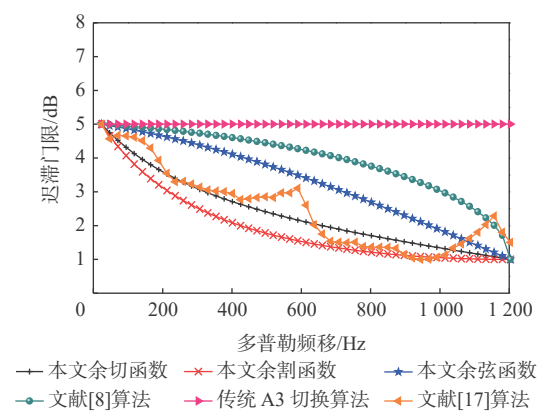


图3 不同切换算法的迟滞门限

Fig. 3 Hysteresis threshold of different handover algorithms

此外，在图3中可以得出，当多普勒频移大小一定时，余割函数对于切换迟滞门限值调整的性能优于余切函数和余弦函数。原因为：不同动态函数自适应调整迟滞门限时，函数自身的衰减变化速率不同[18]。以多普勒频移[0, 600] Hz区间为例，余弦函数衰减了40%，余切函数衰减了72.5%，而余割函数衰减速率最大，衰减了87.5%。因此，在相同多普勒频移下，衰减率越大，越能更好地响应多普勒频移变化，余割函数能够更加敏捷地响应多普勒频移变化，其迟滞门限切换阈值最小，切换速率最快。

在对迟滞门限性能调整分析的基础上，对触发时延的动态调整性能进行比较分析，结果如图4所示。可以看出，传统A3切换算法同样无法完成对于切换时延的调整，其触发时延为固定值。文献[8]算法、文献[17]算法和本文算法触发时延都随着多普勒频移的增大而呈现减小的趋势，其原因为：当列车速度增大时，减小触发时延可以促进列车及时完成切换，以提高切换性能[15]。上述比较算法和本文算法均较好地体现了这一要求。在相同多普勒频移大小下，本文算法与文献[17]算法较文献[8]算法切换触发时延更小，算法实时性更好。但文献[17]算法在触发时延调节时存在时延值波动的问题，其算法稳定性较本文算法较差。

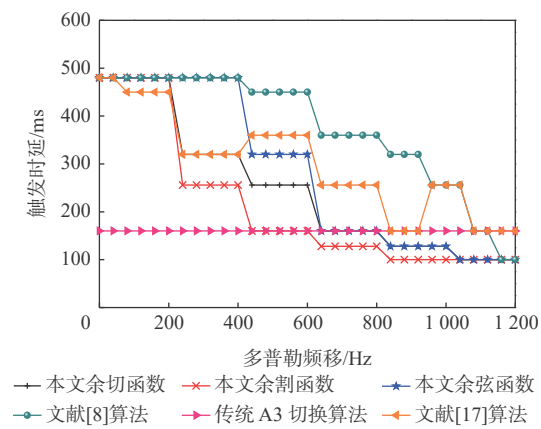


图4 不同切换算法的触发时延

Fig. 4 Time-to-trigger of different handover algorithms

对不同切换算法的切换成功率进行比较实验，如图5所示。可知，本文算法通过考虑多普勒频移因素来动态调整切换阈值，因而能够针对不同的多普勒频移做出相应的自适应切换调整，当多普勒频移增大时，相应地减小切换阈值，避免了判决条件过高而无法触发切换的问题。本文算法余弦、余切和余割函数较文献[8]算法和传统A3切换算法，能够更有效地提高切换成功率，且余割函数的切换成功率在不同多普勒频移下均高于其他算法，性能最优。此外，文献[17]算法对输入数据进行简单模糊处理后，导致切换成功率控制的精确度降低，继而同样出现了切换成功率振荡的问题，同时，其切换

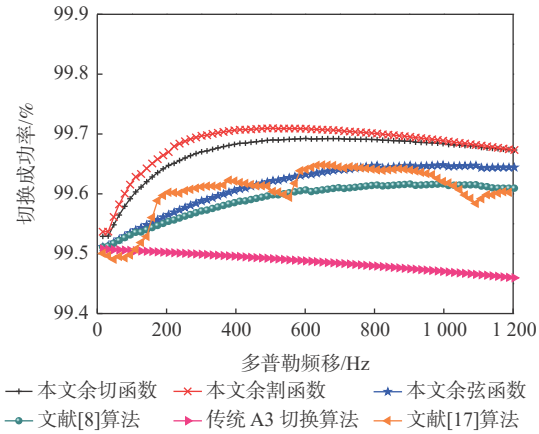


图 5　不同多普勒频移下切换成功率的比较

Fig. 5　Comparison of handover success rates under different Doppler shifts

成功率稳定性较差，难以有效形成对于高速铁路越区切换的稳定控制。

在分析完切换成功率后，对不同高铁运行场景的切换成功率进行比较。高速铁路运行环境复杂，运行线路大多面临山区、高架桥等特殊运行环境，而不同的高铁运行场景会给无线通信带来较大的影响。以高架桥和山区典型场景为例，对 350 km/h 高速行车条件下的切换成功率进行比较，其中，高架桥和山区的阴影衰落标准差分别取 4.1 dB 和 3.3 dB[19]，比较结果如表 4 所示。可知，文献 [8] 算法和文献 [17] 算法在高架桥和山区场景下的切换成功率优于传统 A3 切换算法，这是因为：传统 A3 切换算法使用固定切换阈值，无法根据多普勒频移自适应调整迟滞门限和触发时延，导致切换成功率低，且在山区场景下，传统 A3 切换算法的切换成功率低于 99.5%，无法满足中国高速铁路无线通信系统 QoS 的要求。而文献 [8] 算法和文献 [17] 算法分别采用椭圆函数和模糊逻辑进行切换阈值调整，较传统 A3 切换算法在切换性能上有了一定的提升。但文献 [8] 算法和文献 [17] 算法的切换成功率均低于本文所提 3 种函数的切换成功率，这是因为：文献 [8] 算法仅考虑了速度因素，采用速度动态函数对切换迟滞门限有一定的调节作用，但对切换阈值的调节性有限，无法适应多普勒快速变化场景。而文献 [17] 算法采用模糊逻辑进行切换阈值调整，但隶属函数指标权重存在主观性不足，导致算法稳定性较差，继而影响切换性能。而本文算法通过余弦、余切、余割函数，实现切换判决迟滞门限和触发时延的及时调整，提高了切换成功率，证明了本文算法根据多普勒频移大小动态调整切换阈值的有效性。

此外，采用本文余弦、余切、余割函数进行越区切换时，余割函数在不同场景下的切换成功率均最高。这是因为：3 种函数衰减变化速率不同，其中，余割函数衰减速率最大，因而在动态调整阈值时，能够更加敏捷地响应多普勒频移变化，其切换成功率最高。从表 4 还可以看出，在相同切换算法中，山区场景的切换成功率均低于高架桥场景。这是因为：高架桥场景中，信号传播路径上不存在大的遮挡物，而在山区场景下，一般会有一定的遮挡物，如树木、房屋、山丘等，导致信道存在一定的非直射信号分量。综上可知，在相同多普勒频移条件下，高架桥场景下的切换成功率要高于山区场景，该结论也符合铁路无线通信的现场实际规律，从而进一步验证了本文算法的有效性。

在实际切换场景中，由于列车的高移动性，基站发送机和列车接收机的快速相对位移，导致接收机信号载波频率和本振频率之间存在误差，影响切换性能。因此，在式 (23) 计算越区切换成功率的基础上，引入多普勒频移估计误差，得到多普勒频移估计误差对本文算法的影响，结果如表 5 所示。

由表 5 可知，在相同多普勒频移条件下，本文算法未加入多普勒频移估计误差前切换成功率均大于加入多普勒频移估计误差扰动后切换成功率，即多普勒频移误差扰动时会降低本文算法的切换性能。这是因为：在高速行车条件下，多普勒频移估计误差会对通信时延、数据传输率等通信性能产生严重影响，继而影响越区切换性能[20]。此外，从表 5 可以发现，在不同多普勒频移条件下，本文算法在加入多普勒频移扰动误差后切换成功率和未加入多普勒频移误差前切换成功率均能满足中国高速铁路无线通信系统 QoS 对于越区切换要求功率大于 99.5% 的要求[21]，说明本文算法考虑了多普勒频移影响因素后，具有较强的鲁棒性。

进一步量化分析多普勒频移估计误差对本文 3 种函数切换的定量影响，采用切换成功率相对误

表 4　不同场景下切换成功率的比较

Table 4　Comparison of handover success rates in different scenarios

场景	阴影衰落标准差/dB	列车距源基站距离/km	切换成功率/%					
			传统A3切换算法	文献[8]算法	文献[17]算法	本文余弦函数	本文余切函数	本文余割函数
高架桥	4.1	2.4	99.56	99.71	99.78	99.79	99.83	99.84
山区	3.3	2.34	99.43	99.69	99.71	99.72	99.73	99.74

表5 多普勒频移估计误差切换成功率比较

Table 5 Comparison of handover success rates of Doppler shift estimation errors

多普勒频移/Hz	未加入误差前切换成功率/%			加入误差扰动后切换成功率/%		
	本文余切函数	本文余割函数	本文余弦函数	本文余切函数	本文余割函数	本文余弦函数
200	99.647 5	99.669 8	99.565 2	99.645 6	99.660 1	99.564 9
400	99.682 7	99.706 3	99.605 4	99.681 9	99.704 4	99.604 6
600	99.692 5	99.708 8	99.632 6	99.691 7	99.708 4	99.631 7
800	99.690 7	99.700 8	99.647 4	99.690 2	99.700 5	99.644 7
1 000	99.684 0	99.688 9	99.648 1	99.683 3	99.680 7	99.646 4

差进行比较，结果如图6所示。可知，随着多普勒频移估计误差扰动的增加，切换成功率相对误差逐步增大，说明误差越大，对切换成功率的干扰也越大，会降低切换成功率。此外，从表5可以看出，在不同多普勒频移估计误差扰动下，本文余割函数的切换成功率相对误差最小，这是因为：余割函数衰减速率最大，随着多普勒频移扰动的增大，较余弦、余切函数能更及时地调整迟滞门限和触发时延，切换性能更优。该结论与图3仿真结果一致，即余割函数能够更加敏捷地响应多普勒频移变化，更适合高速铁路越区切换场景的需求。

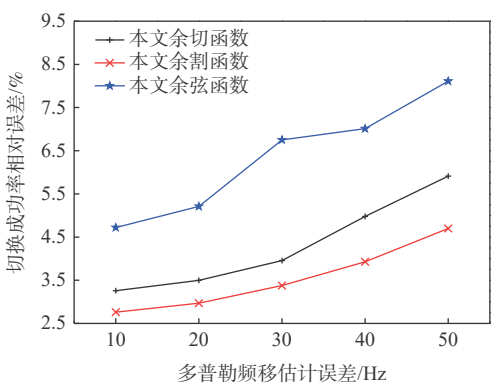


图6 多普勒频移估计误差扰动对本文算法的影响

Fig. 6 Influence of Doppler shift estimation error disturbance on the proposed algorithm

4 结 论

本文提出了一种考虑多普勒频移影响因素的5G-R自适应高速铁路越区切换算法，结论如下：

1）在高速环境下，多普勒频移对高速列车越区切换成功率有较大影响。

2）能够有效提高高速铁路越区切换成功率，且满足中国铁路无线通信系统切换成功率QoS大于99.5%的要求。

3）在高架桥、山区等场景下，本文算法均优于比较算法，且基于余割函数的切换算法性能最优，取得了更高的切换成功率。

研究结果可为下一代高速铁路5G-R系统演进提供一定的理论参考依据。

参考文献（References）

[1] 陈永, 詹芝贤, 张薇. 基于MPC的车车通信时间同步方法[J]. 中国铁道科学, 2022, 43(5): 111-122.
CHEN Y, ZHAN Z X, ZHANG W. Time synchronization method for train-to-train communication based on model predictive control [J]. China Railway Science, 2022, 43(5): 111-122(in Chinese).

[2] WANG Z G, WANG Z H, ZHU F S, et al. Joint Doppler shift and channel estimation for high-speed railway wireless communications in tunnel scenarios[J]. Wireless Communications and Mobile Computing, 2022, 2022: 6804412.

[3] QIAN P H, SHANG T, GAO Y, et al. Research on dynamic handover decision algorithm based on fuzzy logic control in mobile FSO networks[J]. Photonic Network Communications, 2021, 41(2): 136-147.

[4] DAVAASAMBUU B, YU K P, SATO T. Self-optimization of handover parameters for long-term evolution with dual wireless mobile relay nodes[J]. Future Internet, 2015, 7(2): 196-213.

[5] DUAN B F, LI C R, XIE J L, et al. Fast handover algorithm based on location and weight in 5G-R wireless communications for high-speed railways[J]. Sensors, 2021, 21(9): 3100.

[6] ANGRI I, NAJID A, MAHFOUDI M. Available bandwidth and RSRP based handover algorithm for LTE/LTE-Advanced networks tested in LTE-Sim simulator[J]. International Journal of Electronics and Telecommunications, 2019, 65(1): 85-93.

[7] CHEN Y, NIU K Y, WANG Z. Adaptive handover algorithm for LTE-R system in high-speed railway scenario[J]. IEEE Access, 2021, 9: 59540-59547.

[8] 陈永刚, 李德威, 张彩珍. 一种基于速度的LTE-R越区切换优化算法[J]. 铁道学报, 2017, 39(7): 67-72.
CHEN Y G, LI D W, ZHANG C Z. A speed-based cross-zone handoff optimization algorithm of LTE-R[J]. Journal of the China Railway Society, 2017, 39(7): 67-72(in Chinese).

[9] 3GPP.Technical specification group services and system aspects; system architecture for the 5G system (5GS); stage2 (Release17): TS23.501[S].Valbonne: 3GPP Organizational Partners, 2021.

[10] EL BANNA R, ELATTAR H M, ABOU EL-DAHAB M M. Fast adaptive handover using fuzzy logic for 5G communications on high speed trains[C]//Proceedings of the 16th International Conference on Telecommunications. Piscataway: IEEE Press, 2021: 10-17.

[11] CHEN Z, DAN L. Fast fading channel estimation for OFDM systems with complexity reduction[J]. Chinese Journal of Electronics,

2021, 30(6): 1173-1177.

[12] YANG Y C, BAI Z Q, PANG K, et al. Design and analysis of spatial modulation based orthogonal time frequency space system[J]. China Communications, 2021, 18(8): 209-223.

[13] ZHAO J H, LIU Y Y, WANG C Y, et al. High-speed based adaptive beamforming handover scheme in LTE-R[J]. IET Communications, 2018, 12(10): 1215-1222.

[14] 李岸宁. 铁路 5G 专网无线通信信道特征分析与建模[D]. 北京: 中国铁道科学研究院, 2022.
LI A N. Characteristic analysis and modeling of wireless communication channels in railway 5G dedicated network[D]. Beijing: China Academy of Railway Sciences, 2022(in Chinese).

[15] LIU Q Y, KWONG C F, WEI S, et al. Reinforcement learning-based joint self-optimisation method for the fuzzy logic handover algorithm in 5G HetNets[J]. Neural Computing and Applications, 2023, 35(10): 7297-7313.

[16] 3GPP. Radio resource control(RRC): TS 36.331 v16.6. 0[S]. Sophia Antipolis Cedex: ETSI, 20211.

[17] ALHAMMADI A, HASSAN W H, EL-SALEH A A, et al. Intelligent coordinated self-optimizing handover scheme for 4G/5G heterogeneous networks[J]. ICT Express, 2023, 9(2): 276-281.

[18] 田蜜. 地空宽带通信越区切换算法研究[D]. 重庆: 重庆大学, 2019.
TIAN M. Research on handoff algorithm of ground-to-air broadband communication[D]. Chongqing: Chongqing University, 2019 (in Chinese).

[19] 韩丽. 复杂场景与超高速铁路移动通信系统的快速切换算法研究[D]. 北京: 北京交通大学, 2021.
HAN L. Research on fast handover algorithm for composite scene and ultra-high speed railway mobile communication system[D]. Beijing: Beijing Jiaotong University, 2021(in Chinese).

[20] WANG Z F, WEN X M, LU Z M, et al. Random access optimization for initial access and seamless handover for 5G-satellite network[J]. Computer Networks, 2022, 214: 109176.

[21] ZAKIA I, KURNIAWAN A, ISKANDAR. Impact of Doppler shift error on least-squares MIMO channel estimation for high-speed railway[J]. IET Communications, 2020, 14(2): 206-218.

An improved 5G-R adaptive high-speed railway handover algorithm

CHEN Yong*，KANG Jie，TAO Xuan

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

Abstract: Under high-speed driving conditions, over-area handover, as a key technology for future 5G-R communication of high-speed railways, is crucial for ensuring driving safety. The next-generation 5G-R wireless communication system of high-speed railways adopts fixed handover parameters, but when the train is running at high speed, it is highly susceptible to the Doppler effect, resulting in low handover success. To address this issue, an improved 5G-R adaptive high-speed railway handover algorithm that took into account the influence of the Doppler shift was proposed. First, the influence of the Doppler shift on the handover success rate was analyzed, and the relationship function between Doppler shift and handover success rate was obtained. Then, the dynamic function of handover over the area considering the influence of Doppler shift was proposed, and three functions, namely cosine, cotangent, and cosecant, were designed to adjust the handover hysteresis threshold and time-to-trigger adaptively. Finally, a quantitative comparison analysis of the handover success rate was carried out for different Doppler shift sizes and different high-speed railway scenarios. The results show that the proposed method can effectively improve the handover success rate over the area, and the handover success rate of the cosine, cosecant, and cosecant functions in the viaduct and mountain areas is better than the comparison algorithm and meets the requirement of quality of service (QoS) higher than 99.5% for the handover success rate of China's wireless communication system. The research results provide a certain theoretical reference for the evolution of the 5G-R system for next-generation high-speed railways.

Keywords: handover algorithm; 5G-R; Doppler shift; dynamic function; adaptive handover

Received：2023-03-28；**Accepted**：2023-05-19；**Published Online**：2023-06-28 15：55

URL：link.cnki.net/urlid/11.2625.V.20230628.0939.001

Foundation items：National Natural Science Foundation of China (62462043,61963023); Key Research and Development Project of Lanzhou Jiaotong University (ZDYF2304)

* **Corresponding author**. E-mail：edukeylab@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0155

基于双路径监督的遥感图像语义分割网络

刘春娟[1]，乔泽[1]，闫浩文[2]，吴小所[1,*]，王嘉伟[1]，辛钰强[1]

(1. 兰州交通大学 电子与信息工程学院，兰州 730070；2. 兰州交通大学 测绘与地理信息学院，兰州 730070)

摘 要：为解决遥感图像语义分割任务中目标物体边界分类模糊的问题，提出双路径监督与注意力筛选网络。引入可监督的边界提取模块来增加边界信息通道，提高边界信息在语义分割中的权重，增强对目标物体边界像素的注意力；引入注意力筛选模块，通过注意力图筛选出浅层网络中的空间细节信息和深层网络中的抽象语义信息，舍弃网络中的冗余信息，防止过拟合。双路径监督与注意力筛选网络在Potsdam数据集和Jiage数据集上的平均交并比分别为85.44%和86.07%，比次优网络MagNet和SAPNet分别提升了1.24%和1.28%、1.54%和1.27%。实验结果表明，所提网络能更精准地分割目标物体的边界。

关 键 词：遥感图像；语义分割；可监督；边界信息；注意力筛选

中图分类号：TP751.1; V19

文献标志码：A **文章编号**：1001-5965(2025)03-0732-10

语义分割是一项基础的计算机视觉任务，其目标是给输入图像的每个像素分配一个语义标签[1]。近年来，不断发展的深度学习提高了语义分割性能，促进了计算机视觉领域中与高分辨率遥感图像语义分割相关研究的发展，如灾害监测[2]、城市规划[3-4]、道路提取[5-6]。然而，高分辨率遥感图像含有许多较为复杂的地物信息，对其进行的语义分割存在分割精度低、分割边界模糊等问题。

随着深度学习的发展，Long等[7]提出全卷积网络(fully convolutional network, FCN)，在语义分割问题中推广使用端对端的卷积神经网络(convolutional neural network, CNN)，并取得了令人瞩目的成就。Ronneberger等[8]提出了U-Net网络，该网络包含一套完整的编解码结构，通过跳跃连接将编码结构中的信息拼接到解码结构，广泛应用于语义分割领域。对于解决语义分割中普遍存在的、因相同类别物体间差异过大或不同类别物体间差异过小而导致的分类错误的问题，一种方法是多尺度特征融合[9-12]。例如，DeepLabv2[13]和DeepLabv3[14]通过空洞空间金字塔池化模块来增大感受野，提取上下文信息，通过不同膨胀率的膨胀卷积得到不同感受野的特征图。金字塔场景解析网络(pyramid scene parsing network, PSPNet)[15]采用多尺度金字塔池化模块扩大感受野，提取上下文信息，通过不同的池化窗口大小得到不同感受野的特征图。另一种方法是通过注意力机制得到像素间的依赖关系[16-19]。例如，双注意力网络(dual attention network, DANet)[20]采用空间注意力模块和通道注意力模块捕捉空间和通道维度中的全局特征依赖关系。交叉注意力网络(criss-cross attention network, CCNet)[21]采用十字交叉注意力模块捕捉各像素间的长程依赖关系，并减少自注意力机制的计算量。

为解决类内差异大、类间差异小导致目标物体边缘分割精度低的问题，本文提出双路径监督与注意力筛选网络，用于遥感图像语义分割任务。在训练过程中，该网络引入双路径监督机制，

收稿日期：2023-03-31；**录用日期**：2023-05-26；**网络出版时间**：2023-06-30 15:51
网络出版地址：link.cnki.net/urlid/11.2625.V.20230630.1501.004
基金项目：国家重点研发计划(2022YFB3903604)；甘肃省自然科学基金(21JR7RA289)；甘肃省重点研发计划(20YF8GA035)
*通信作者. E-mail：43452740@qq.com

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

即在训练过程中增加一路边界信息监督路径, 获得较为准确的边界信息, 增加网络对目标物体边界信息的关注度。同时, 引入注意力筛选模块, 提取浅层细节信息和深层语义信息, 补充下采样丢失的细节信息。

1　原理与方法

1.1　网络整体结构

双路径监督与注意力筛选网络结构如图 1 所示。在训练过程中, 首先, 将遥感图像输入到边界提取模块和注意力筛选模块中, 分别得到输入图像的边界信息和语义信息。然后, 将边界提取模块输出的包含边界信息的特征图与从标签提取出的真实类别间的边界特征图进行对比, 并进行反向传播, 使边界提取模块输出的边界信息与真实类别输出的边界信息近似相等。最后, 将边界提取模块的输出结果与注意力筛选模块的输出结果进行通道融合。经过 1×1 卷积后, 进行通道压缩, 将得到的结果与标签进行对比, 更新权值, 提高语义分割的分割精度。

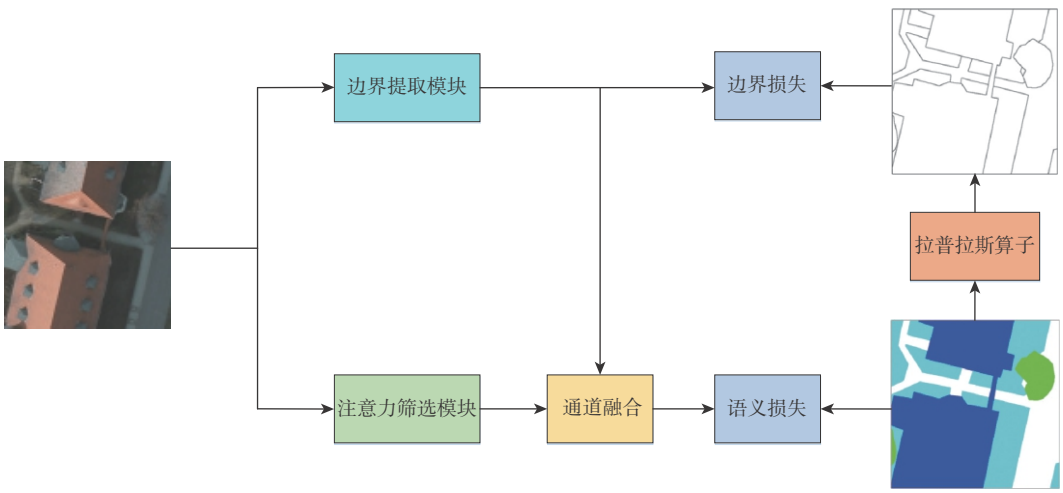


图 1　双路径监督与注意力筛选网络整体结构

Fig. 1　Overall structure of a dual path supervision and attention filtering network

1.2　可监督的边界提取模块

传统的语义分割网络存在不同类别之间分类错误的问题, 特别是在相似类别的边界处更容易出现分类错误, 导致分割精度难以提升。因此, 提出一种可监督的边界提取模块, 通过浅层的卷积神经网络提取出边界信息, 再利用传统的图像处理方法得到标签中所包含的准确边界信息, 通过提取出的准确边界信息对边界提取模块的输出结果进行监督训练。

在语义分割任务中, 数据集包含人工标注的真实标签。在真实标签中, 不同类别用不同颜色代替, 因此, 真实标签中含有准确的边界信息。通过数字图像处理中的边界提取方法得到真实的边界信息, 再对网络结构中的边界提取模块进行监督训练, 使边界提取模块能够提取到更准确的边界信息。为提取标签中的真实边界信息, 采用拉普拉斯算子与真实标签进行卷积运算的方法。拉普拉斯算子如下:

$$\boldsymbol{C}=\begin{bmatrix}-1 & -1 & -1\\ -1 & 8 & -1\\ -1 & -1 & -1\end{bmatrix} \tag{1}$$

首先, 将标签与拉普拉斯卷积核进行卷积运算, 对得到的特征矩阵进行归一化处理, 小于零的值设置为 1, 大于零的值设置为 0, 即设边界为黑色, 其他为白色, 得到真实的边界信息。然后, 将遥感图像输入到边界提取模块中, 该模块采用 3 层结构, 每层包含 2 个 3×3 的卷积层和 1 个最大池化层。解码部分采用双线性插值上采样来恢复到与输入图像相同的尺寸。最后, 计算边界提取模块得到的特征图与真实边界特征图之间的损失, 并通过反向传播对边界提取模块中的参数进行更新, 以确保从边界提取模块得到的特征图与真实的边界特征图近似相等。

在边界提取模块输出的特征图中, 目标物体边界的像素值为 0, 其他区域的像素值为 1。将边界提取模块输出的边界信息特征图与注意力筛选模块输出的语义信息特征图进行通道拼接, 并经过 1×1 卷积后进行通道压缩, 使得边界区域的像素值仍为 0, 其他区域的像素值不变。因此, 突出了边界信息, 增强了对边界信息的注意力。

1.3　注意力筛选模块

在语义分割网络中, 浅层网络包含更多的空间细节信息, 而深层网络包含更多的抽象语义信

息[22]。浅层的空间细节信息对于语义分割任务至关重要，因此，需要在网络中融合细节信息。理论上，分割精度会随着融合信息的增加而提高。实际上，在融合大量信息的过程中，会融合很多冗余信息，导致分割结果变差。因此，引入注意力筛选模块来解决上述问题，该模块既可以融合细节信息与语义信息，又可以舍弃对结果无益的参数。注意力筛选模块的结构如图2所示。图中：up表示上采样操作，F表示拼接操作。

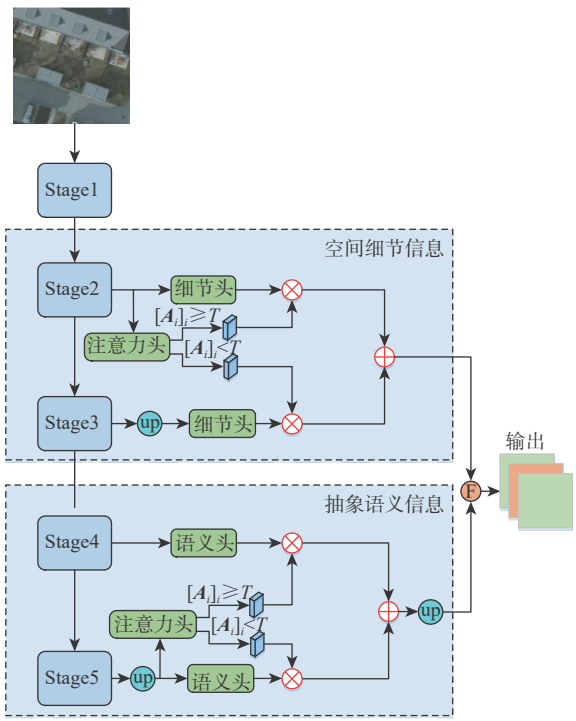


图2 注意力筛选模块的结构

Fig. 2 Structure of attention filtering module

注意力筛选模块的设计思路是：从浅层网络和深层网络中分别筛选出空间细节信息和抽象语义信息，舍弃网络中的冗余信息，并将空间细节信息和抽象语义信息进行通道拼接。如图2所示，该网络用VGG16[23]作为骨干网络，原始图像输入到VGG16骨干网络后，Stage2、Stage3层输出特征图$\boldsymbol{n}_1$、$\boldsymbol{n}_2$，Stage4、Stage5层上采样后输出特征图$\boldsymbol{n}_3$、$\boldsymbol{n}_4$。对于浅层网络，通过$\boldsymbol{n}_1$得到的注意力图将$\boldsymbol{n}_1$和$\boldsymbol{n}_2$这2层的空间细节信息进行互补，在Stage2层中保留从Stage2层下采样到Stage3层过程中丢失的细节信息，在Stage3层中保留浅层的空间细节信息。对于深层网络，通过$\boldsymbol{n}_4$得到的注意力图将$\boldsymbol{n}_3$和$\boldsymbol{n}_4$这2层的抽象语义信息进行互补，在Stage5层保留最深层的语义信息，在Stage4层中保留因下采样导致Stage5层丢失的信息。将Stage2和Stage3层筛选后的信息融合生成包含空间细节信息的特征图，再将Stage4和Stage5层筛选后的信息融合生成包含抽象语义信息的特征图，并将得到的特征图进行通道拼接。

Stage2层输出的特征图$\boldsymbol{n}_1$通过细节头和注意力头分别得到空间细节信息和注意力图，Stage3输出的特征图$\boldsymbol{n}_2$通过细节头得到该层包含的空间细节信息。通过由$\boldsymbol{n}_1$生成的注意力图筛选出不同网络深度的空间细节信息，并将这些信息融合成完整的特征图。Stage4层输出的特征图$\boldsymbol{n}_3$通过语义头得到抽象语义信息，Stage5层输出的特征图$\boldsymbol{n}_4$通过语义头和注意力头分别得到该层包含的抽象语义信息和注意力图。通过由$\boldsymbol{n}_4$生成的注意力图筛选出不同网络深度的抽象语义信息，并将这些信息融合成完整的特征图。

细节头、语义头和注意力头都由2个block块和1个1×1卷积组成。其中，每个block块由1个3×3的卷积层、1个批标准化层和1个ReLU激活层组成，注意力头在1×1卷积后利用Sigmoid激活函数来生成注意力图。通过注意力头得到的注意力图如式(2)所示，通过细节头得到的包含空间细节信息的特征图如式(3)所示，通过语义头得到的包含抽象语义信息的特征图如式(4)所示。

$$\boldsymbol{A}_i = \mathrm{Atte}(\boldsymbol{n}_i) \qquad i = 1,4 \tag{2}$$

$$\boldsymbol{D}_i = \mathrm{Det}(\boldsymbol{n}_i) \qquad i = 1,2 \tag{3}$$

$$\boldsymbol{S}_i = \mathrm{Seg}(\boldsymbol{n}_i) \qquad i = 3,4 \tag{4}$$

在注意力图的基础上，通过动态更新阈值的方式对来自不同网络深度的空间细节信息进行筛选，得到2个布尔类型的矩阵张量，如式(5)和式(6)所示，称为布尔张量。

$$\left[\boldsymbol{b}_{A_i}^{\mathrm{T}}\right]_i = \begin{cases} 1 & [\boldsymbol{A}_i]_i \geqslant T, i = 1,4 \\ 0 & [\boldsymbol{A}_i]_i < T, i = 1,4 \end{cases} \tag{5}$$

$$\left[\boldsymbol{b}_{A_i}^{\mathrm{F}}\right]_i = \begin{cases} 0 & [\boldsymbol{A}_i]_i \geqslant T, i = 1,4 \\ 1 & [\boldsymbol{A}_i]_i < T, i = 1,4 \end{cases} \tag{6}$$

在布尔张量中每个位置上，满足条件的值被设置为1(True)，其余部分被设置为0(False)。相应地，式(5)表示大于或等于阈值T的位置被设置为1，剩余位置被设置为0；式(6)表示小于阈值T的位置被设置为1，剩余位置被设置为0。在训练过程中，阈值T是动态更新的，在初始状态，阈值T被统一设置为0.5，根据网络的损失值不断对阈值T进行更新，直到训练结束。

布尔张量筛选后得到2个尺度相同且包含空间细节信息的特征图和2个尺度相同且包含抽象语义信息的特征图，如式(7)～式(10)所示。分别融合空间细节信息特征图和抽象语义信息特征图，如式(11)和式(12)所示。将融合后的抽象语义信

息特征图进行上采样, 与得到的空间细节信息特征图进行通道拼接, 如式(13)所示。

$$\boldsymbol{O}_{\boldsymbol{n}_1}=\boldsymbol{b}_{A_1}^{\mathrm{T}}\mathrm{Det}(\boldsymbol{n}_1) \tag{7}$$

$$\boldsymbol{O}_{\boldsymbol{n}_2}=\boldsymbol{b}_{A_1}^{\mathrm{F}}\mathrm{Det}\left(\mathrm{up}(\boldsymbol{n}_2)\right) \tag{8}$$

$$\boldsymbol{O}_{\boldsymbol{n}_3}=\boldsymbol{b}_{A_4}^{\mathrm{T}}\mathrm{Seg}(\boldsymbol{n}_3) \tag{9}$$

$$\boldsymbol{O}_{\boldsymbol{n}_4}=\boldsymbol{b}_{A_4}^{\mathrm{F}}\mathrm{Seg}\left(\mathrm{up}(\boldsymbol{n}_4)\right) \tag{10}$$

$$\boldsymbol{O}_{\boldsymbol{D}}=\boldsymbol{O}_{\boldsymbol{n}_1}+\boldsymbol{O}_{\boldsymbol{n}_2} \tag{11}$$

$$\boldsymbol{O}_{\boldsymbol{S}}=\boldsymbol{O}_{\boldsymbol{n}_3}+\boldsymbol{O}_{\boldsymbol{n}_4} \tag{12}$$

$$\boldsymbol{O}=F\left(\boldsymbol{O}_{\boldsymbol{D}},\mathrm{up}(\boldsymbol{O}_{\boldsymbol{S}})\right) \tag{13}$$

2　数据集及评估指标

2.1　数 据 集

1) Potsdam 数据集。该数据集包括在德国勃兰登堡州首府上空拍摄的数字正射影像图和数字表面影像图, 地面分辨率为 5 cm。数据集包含 38 张高分辨率的遥感图像, 每张图像的尺寸为 6 000×6 000。这些图像均具有 6 类地物类别, 包括建筑物、地面、低植被、树、汽车和背景, 其中, 建筑物用深蓝色表示, 地面用白色表示, 低植被用浅蓝色表示, 树用绿色表示, 汽车用黄色表示, 背景用红色表示。实验中, 因为服务器硬件资源的限制, 所以将数据集中的遥感图像裁剪成 2 304 张大小为 512×512 的图像。1 612 张图像作为训练集, 346 张图像作为验证集, 346 张图像作为测试集。

2) Jiage 数据集。该数据集数据由一架无人机在华南某地区收集, 包括 4 个中等分辨率的遥感影像及相应的地面真实标签。这些图像均具有 5 类地物类别, 包括水、植被、道路、建筑物和背景, 其中, 水用黄色表示, 植被用红色表示, 道路用蓝色表示, 建筑物用绿色表示, 背景用黑色表示。每个图像的分辨率为亚米级别, 光谱为可见光带。Jiage 数据集使用了与 Potsdam 数据集相同的裁剪方式, 将这些中等分辨率图像及相应的地面真实标签随机裁剪成大小为 512×512 的图像。由于数据集较小, 使用了常用的数据增强方法, 如平移、旋转、增加噪声、仿射变换等, 共得到 3 173 张小分辨率图像, 将其中 2 390 张图像作为训练集, 400 张图像作为验证集, 383 张图像作为测试集。

2.2　评估指标

实验使用平均交并比(mean intersection over union, mIoU)、F_1 得分和像素精度(pixel accuracy, PA)3 个指标来评估双路径监督与注意力筛选网络的优越性。

$$P_{\mathrm{A}}=\frac{\sum_{i=1}^{n}T_{Pi}}{\sum_{i=1}^{n}(T_{Pi}+F_{Pi})} \tag{14}$$

$$F_1=2\frac{PR}{P+R} \tag{15}$$

$$M_{\mathrm{IoU}}=\frac{T_P}{T_P+F_P+F_N} \tag{16}$$

$$R=\frac{T_P}{T_P+F_N} \tag{17}$$

$$P=\frac{T_P}{T_P+F_P} \tag{18}$$

式中: P 为精确度; R 为召回率; T_P、F_P 和 F_N 分别为真阳性、假阳性和假阴性的数量。

3　实验设计及结果

双路径监督与注意力筛选网络是在 PyTorch 深度学习框架的基础上实现的, 所有实验都在 64 位 Windows10 系统上且配备英特尔至强 R 处理器 E5-2650 v4(2.20 GHz)、80 GB 内存(RAM)、Nvidia GeForce GTX 1080 Ti(11 GB)的服务器上进行。软件环境和编译器环境分别为 Python3.7 和 PyCharm。使用小批次的随机梯度下降法(stochastic gradient descent, SGD), 批次大小为 4, 动量为 0.9, 权重衰减为 0.000 1, 设置初始学习率为 0.001 来训练模型, 并且模型训练 500 个轮次。本文采用“poly”的学习率衰减策略来保证训练过程, 其表达式为

$$l=l_{\mathrm{initial}}\left(1-\frac{R_{\mathrm{epoch}}}{R_{\mathrm{max_epoch}}}\right)^{0.9} \tag{19}$$

式中: l 为当前学习率; l_{initial} 为初始学习率; R_{epoch} 为当前的训练轮数; $R_{\mathrm{max_epoch}}$ 为最大的训练轮数。为保证实验数据之间的可比性, 所有实验环境和参数设置均相同。

整个实验过程包括消融研究和对比实验。将骨干网络为 VGG16 的深度卷积编码-解码(deep convolutional encoding-decoding, DCED)网络作为基准网络, 分别添加边界提取模块(DCED add boundary extraction module, DCED-BEM)和注意力筛选模块(DCED add attention filtering module, DCED-AFM)进行实验, 再将 2 个模块同时加入基准网络(DCED-BEM-AFM)进行实验, 定性定量分析本文方法的有效性。

3.1　Potsdam 数据集消融实验结果展示与分析

表 1 展示了 Potsdam 数据集上的消融实验结果, 评价标准包括各模型的 mIoU、PA、F_1 得分及各类别的 IoU。

表1 Potsdam数据集消融实验结果

Table 1 Results of ablation experiments on Potsdam dataset %

网络模型	IoU						F_1	mIoU	PA
	背景	汽车	地面	树	低植被	建筑物			
DCED	54.57	76.05	79.37	72.02	74.97	89.02	84.85	74.33	86.29
DCED-BEM	80.70	79.07	85.60	78.41	80.39	89.89	91.02	82.34	91.21
DCED-AFM	81.57	79.48	86.38	81.26	82.28	89.82	91.78	83.47	92.09
DCED-BEM-AFM	84.93	79.51	87.37	84.33	85.07	91.41	92.89	85.44	93.36

DCED-BEM的消融实验如表1所示。与DCED相比,DCED-BEM的mIoU、PA、F_1分别增加了8.01%、4.92%、6.17%。特别是背景、地面、树和低植被的IoU提升尤为明显,分别提升了26.13%、6.23%、6.39%和5.42%。在深度卷积网络中,随着网络深度的加深,提取出的特征信息更偏向于抽象语义信息,导致丢失大量的边界细节信息。传统语义分割网络大多采用单监督机制,同时训练语义信息和边界细节信息,使语义信息与边界细节信息之间相互影响,导致分割精度下降。因此,加入可监督的边界提取模块不仅可以补充丢失的边界信息,还可以将训练语义信息与训练边界细节信息相隔离,减少语义信息训练对边界细节信息训练的影响。从图3中的第4列可以看出,与DCED相比,DCED-BEM能更好地提取物体的边界信息,提高语义分割精确度。

DCED-AFM的消融实验如表1所示。与DCED相比,DCED-AFM的mIoU、PA、F_1分别增加了9.14%、5.80%、6.93%。与DCED-BEM相比,DCED-AFM的mIoU、PA、F_1分别增加了1.13%、0.88%、0.76%。语义分割网络中,浅层网络包含更多的空间细节信息,深层网络包含更多的抽象语义信息。经典的语义分割网络为防止浅层空间细节信息的丢失,通常将浅层的空间细节信息与深层的抽象语义信息进行融合,但语义分割性能提升并不明显,这是因为:网络在融合大量信息的过程中,会将很多冗余的信息融合进网络中。因此,加入注意力筛选模块不仅可以融合浅层的空间细节信息和深层的抽象语义信息,还可以筛选出对分割精度提升有帮助的信息,舍弃网络中导致分割精度下降的冗余信息。从图3第5列可以看出,DCED-AFM可以很

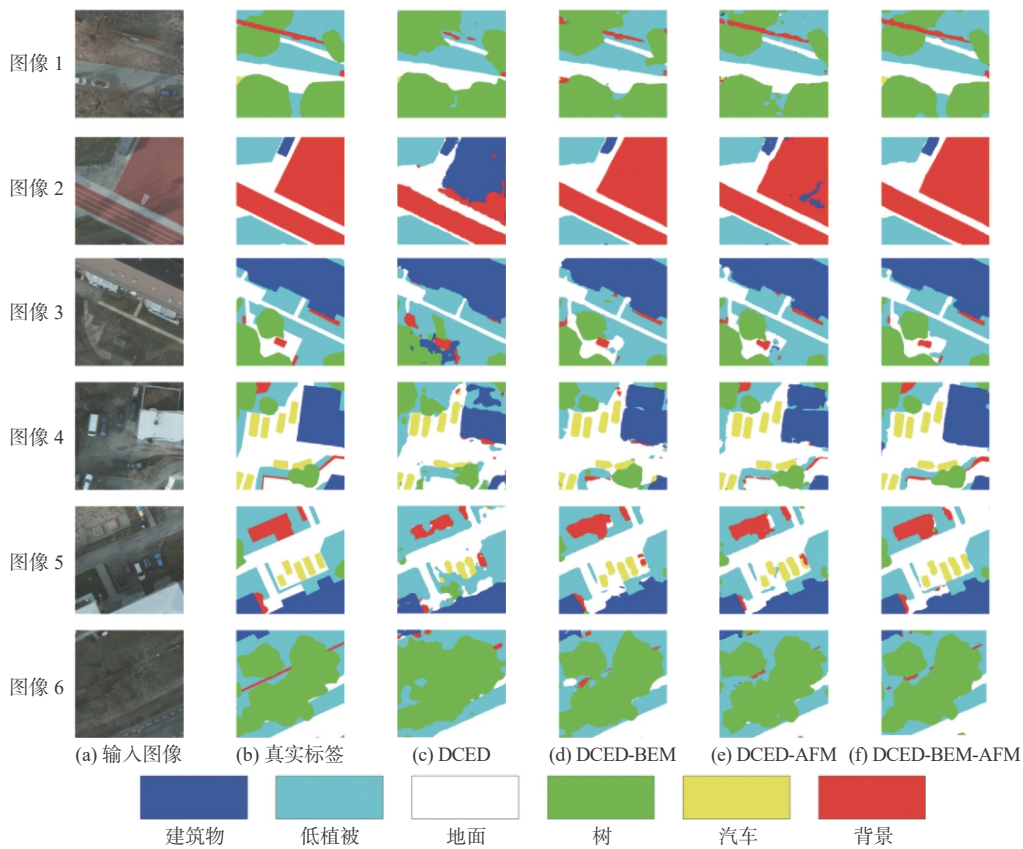


图3 Potsdam数据集上消融实验局部视觉对比结果

Fig. 3 Local visual comparison results of ablation experiments on Potsdam dataset

好地提取目标物体的空间细节信息, 与 DCED-BEM 相比, DCED-AFM 对于图像中容易忽略的小物体有更好的分割效果, 但对于规则的大物体会存在部分区域分类错误的现象。因此, 注意力筛选模块可以很好地融合浅层的空间细节信息和深层的抽象语义信息, 并在整体性能上略优于边界提取模块。

DCED-BEM-AFM 的消融实验如表 1 所示。与 DCED-BEM 和 DCED-AFM 相比, DCED-BEM-AFM 的分割效果有了再次提升, mIoU 达到了 85.44%, PA 达到了 93.36%, F_1 达到了 92.89%。从图 3 中的第 6 列可以看出, DCED-BEM-AFM 不仅对于图像中容易忽略的小物体有更好的分割效果, 而且对于规则的大物体分割效果也有提升。因此, DCED-BEM-AFM 能更好地提取边界信息和浅层的空间细节信息, 处理背景复杂的遥感图像更为擅长。

3.2　Jiage 数据集消融实验结果展示与分析

表 2 显示了 Jiage 数据集上的消融实验结果, 评价标准包括各模型的 mIoU、PA、F_1 得分及各类别的交并比。

DCED-BEM 的消融实验如表 2 所示。与 DCED 相比, DCED-BEM 的 mIoU、PA、F_1 分别增加了 7.60%、2.29%、5.53%, 特别是背景、道路、水的 mIoU 提升尤为明显, 分别提升了 5.35%、20.42%、5.41%。从图 4 中的第 4 列可以看出, 与 DCED 相比, DCED-BEM 能更好地提取物体的边界信息, 提高语义分割精确度。

DCED-AFM 的消融实验如表 2 所示。与 DCED-相比, DCED-AFM 的 mIoU、PA、F_1 上分别增加了 9.10%、2.87%、6.51%。与 DCED-BEM 相比, DCED-AFM 的 mIoU、PA、F_1 分别增加了 1.50%、0.58%、

表 2　Jiage 数据集消融实验结果

Table 2　Results of ablation experiments on Jiage dataset　%

网络模型	IoU					F_1	mIoU	PA
	背景	植被	道路	水	建筑物			
DCED	77.04	92.84	43.91	83.91	78.56	84.71	75.25	91.89
DCED-BEM	82.39	95.34	64.33	89.32	82.85	90.24	82.85	94.18
DCED-AFM	83.55	95.94	68.05	90.60	83.63	91.22	84.35	94.76
DCED-BEM-AFM	83.84	96.25	73.51	92.21	84.52	92.32	86.07	95.15

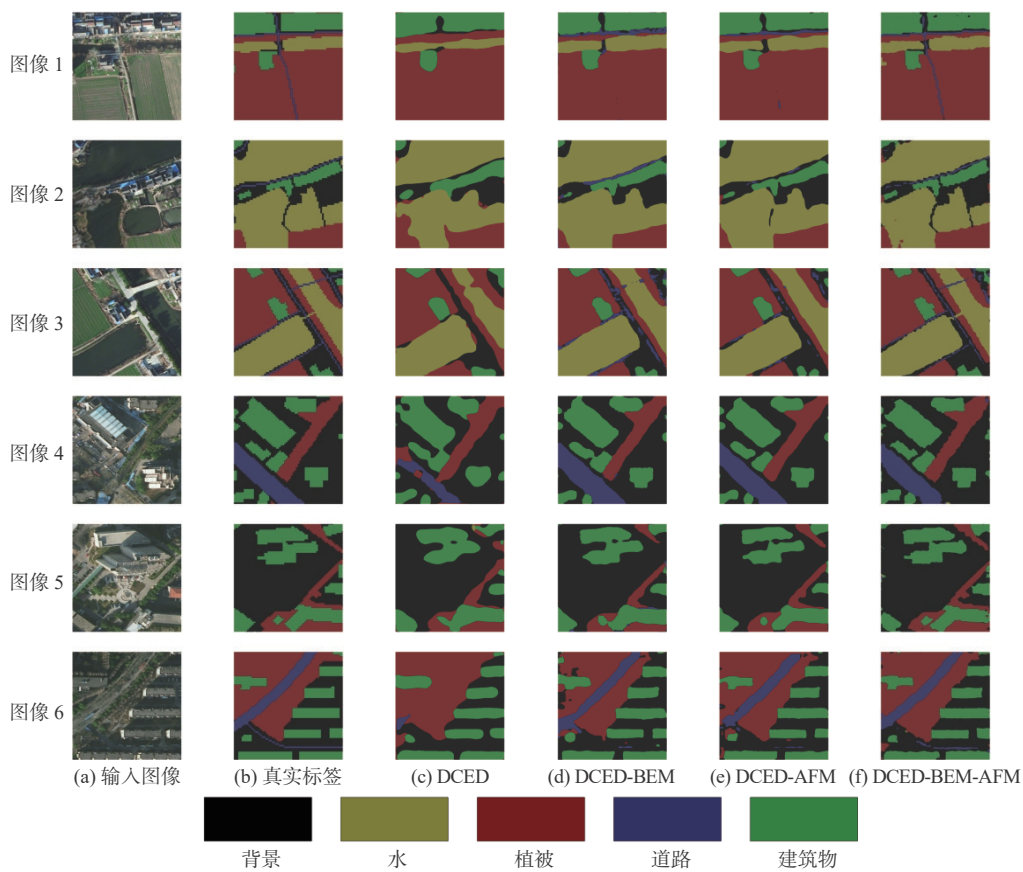


图 4　Jiage 数据集上消融实验局部视觉对比结果

Fig. 4　Local visual comparison results of ablation experiments on Jiage dataset

0.98%。从图4中的第5列可以看出，DCED-AFM可以很好地融合浅层的空间细节信息和深层的抽象语义信息，并且在整体性能上略优于边界提取模块。

DCED-BEM-AFM的消融实验如表2所示。与DCED-BEM和DCED-AFM相比，DCED-BEM-AFM的分割效果有了再次提升，mIoU达到了86.07%，PA达到了95.15%，F_1达到了92.32%。从图4中的第6列可以看出，增加可监督的边界提取模块和注意力筛选模块后，语义分割性能又有明显的提升。

3.3 Potsdam数据集对比实验结果展示与分析

表3展示了Potsdam数据集上的对比实验结果。与SegNet[24]、PSPNet[15]、DeepLabv3[14]、全局卷积网络(global convolutional network, GCN)[25]、期望最大化注意力网络(expectation-maximization attention networks, EMANet)[26]、CCNet[21]、多尺度最大池化密集网络(multiscale max-pooling dense network, DMAU-Net)[27]、协同注意力感知神经网络(synergistic attention perception neural network, SAPNet)[28]和MagNet[29]相比，DCED-BEM-AFM的mIoU分别增加了14.54%、9.27%、8.69%、7.26%、3.64%、3.05%、2.18%、1.28%、1.24%，得到了更好的效果。其中，EMANet引入期望最大化注意力机制，通过期望最大化计算出一组基，并在这组基上进行注意力机制的运算，减少了计算复杂度。DMAU-Net通过基于密集连接的多尺度池化模块提高了网络的特征提取能力，并引入高效的通道注意力机制增强有效特征图，抑制冗余信息。SAPNet引入协同注意力模块构建上下文亲和性特征图，并通过协同注意力感知模块融合空间和通道中的细节信息。MagNet是一个多尺度的网络框架，通过在多个放大级别上观察图像来解决局部的模糊性。

表3 Potsdam数据集上对比实验结果

Table 3 Comparison of experimental results on Potsdam dataset

模型	IoU/%						mIoU/%	参数量	预测时间/s
	背景	汽车	地面	树	低植被	建筑物			
SegNet[24]	69.49	59.85	83.44	52.97	79.26	80.36	70.90	15.62×10^6	0.12
PSPNet[15]	78.33	65.84	86.78	56.21	81.55	88.32	76.17	30.95×10^6	0.34
DeepLabv3[14]	78.86	67.57	85.63	60.38	80.57	87.51	76.75	59.24×10^6	0.41
GCN[25]	75.12	73.15	85.36	67.32	82.85	85.27	78.18	25.91×10^6	0.15
EMANet[26]	77.40	75.60	85.60	80.70	82.10	89.30	81.78	34.72×10^6	0.12
CCNet[21]	76.39	78.79	87.60	79.62	82.24	89.71	82.39	60.11×10^6	0.23
DMAU-Net[27]	80.69	76.54	86.76	80.87	82.15	92.55	83.26	36.42×10^6	0.19
SAPNet[28]	82.30	73.59	87.27	85.38	84.65	91.75	84.16	63.83×10^6	0.37
MagNet[29]	79.54	82.09	88.67	79.85	83.00	92.07	84.20	51.67×10^6	0.32
DCED-BEM-AFM	84.93	79.51	87.37	84.33	85.07	91.41	85.44	56.57×10^6	0.24

从表3可以看出，与DMAU-Net和MagNet相比，DCED-BEM-AFM在背景、树和低植被类别的IoU上具有明显提升。与建筑物相比，背景、树和低植被具有边界不规则的特性，这几类物体的边界很难被神经网络所识别，说明DCED-BEM-AFM对边界复杂物体的分割更具有优势。与SAPNet相比，DCED-BEM-AFM在汽车类别的IoU提升了5.92%。相较于建筑物，汽车的尺度较小，说明DCED-BEM-AFM对小尺度物体的分割效果相对较好。在网络的参数量上进行对比可以看出，与SAPNet相比，DCED-BEM-AFM的参数量下降了7.26×10^6。与MagNet相比，DCED-BEM-AFM的参数量上升了4.90×10^6，虽然DCED-BEM-AFM的参数量略有上升，但相较于SAPNet和MagNet的预测时间，DCED-BEM-AFM分别减少了0.13 s和0.08 s。由于SAPNet和MagNet都引入自注意力机制增强特征提取，增加了网络的计算复杂度，导致网络预测时间增加。

图5展示了DCED-BEM-AFM和其他语义分割模型在Potsdam数据集上的定性比较。其中，有差异的地方用蓝色标注，直观显示了真实标签与预测之间的差异，蓝色部分越少，说明分割效果越好。由图5可以看出，DCED-BEM-AFM在Potsdam数据集上的总体表现优于其他模型，尤其是对背景、树和低植被的分类效果最好。因此，DCED-BEM-AFM能更好地提取边界复杂类别的边界信息，提高分割精度，比其他模型表现的更好。

3.4 Jiage数据集对比实验结果展示与分析

表4显示了在Jiage数据集上进行的对比实验结果。与SegNet[24]、PSPNet[15]、DeepLabv3[14]、GCN[25]、EMANet[26]、CCNet[21]、DMAU-Net[27]、MagNet[29]和SAPNet[28]相比，DCED-BEM-AFM的mIoU分别增加了15.64%、7.01%、6.08%、5.75%、3.70%、2.84%、

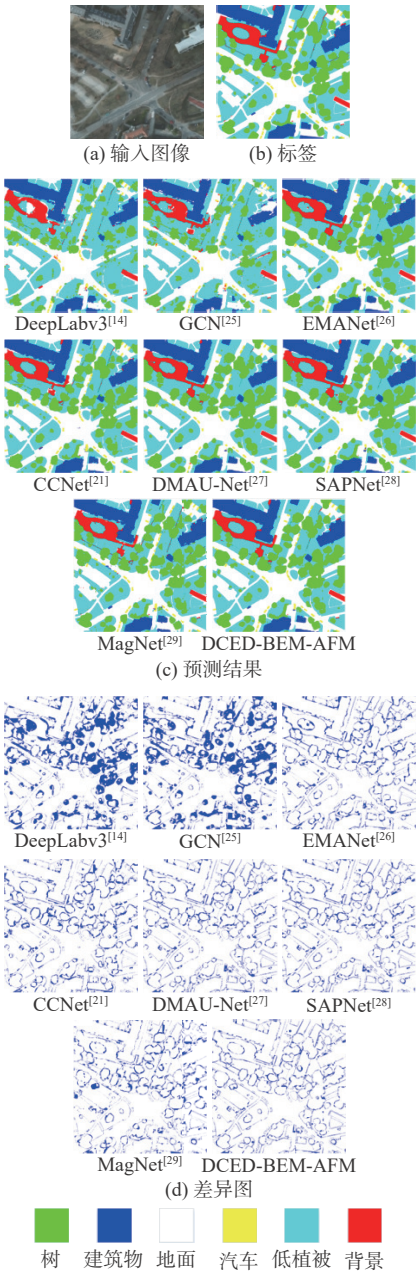


图 5　Potsdam 数据集上与 7 种网络的定性比较

Fig. 5　Qualitative comparison with seven networks on Potsdam dataset

2.60%、1.54% 和 1.27%, 得到了更好的效果。

从表 4 可以看出, 与 DMAU-Net 相比, DCED-BEM-AFM 在道路、水和建筑物类别的 IoU 上提升了 5.72%、2.44% 和 2.32%。与 MagNet 和 SAPNet 相比, DCED-BEM-AFM 在道路类别上的交并比分别提升了 3.04% 和 5.14%。说明 DCED-BEM-AFM 可以更好地分割小尺度物体和边界复杂的物体。

图 6 定性展示了 DCED-BEM-AFM 和其他语义分割模型的对比结果。实验结果表明, DCED-BEM-AFM 能够有效提高分割性能, 并且比其他模型表现得更好, 特别是在道路和背景类别的分割上具有很大提升。

表 4　Jiage 数据集上对比实验结果

Table 4　Comparison of experimental results on Jiage dataset

模型	IoU/%					mIoU/%
	背景	植被	道路	水	建筑物	
SegNet[24]	61.42	91.44	45.42	87.27	66.58	70.43
PSPNet[15]	79.08	96.25	48.81	89.91	81.27	79.06
DeepLabv3[14]	80.83	95.27	56.51	88.67	78.66	79.99
GCN[25]	75.77	95.76	58.94	90.11	81.00	80.32
EMANet[26]	81.93	95.13	63.88	88.37	82.52	82.37
CCNet[21]	81.29	95.30	67.06	90.86	81.64	83.23
DMAU-Net[27]	82.02	95.55	67.79	89.77	82.20	83.47
MagNet[29]	82.37	95.70	70.47	91.31	82.78	84.53
SAPNet[28]	83.88	96.07	68.37	91.10	84.58	84.80
DCED-BEM-AFM	83.84	96.25	73.51	92.21	84.52	86.07

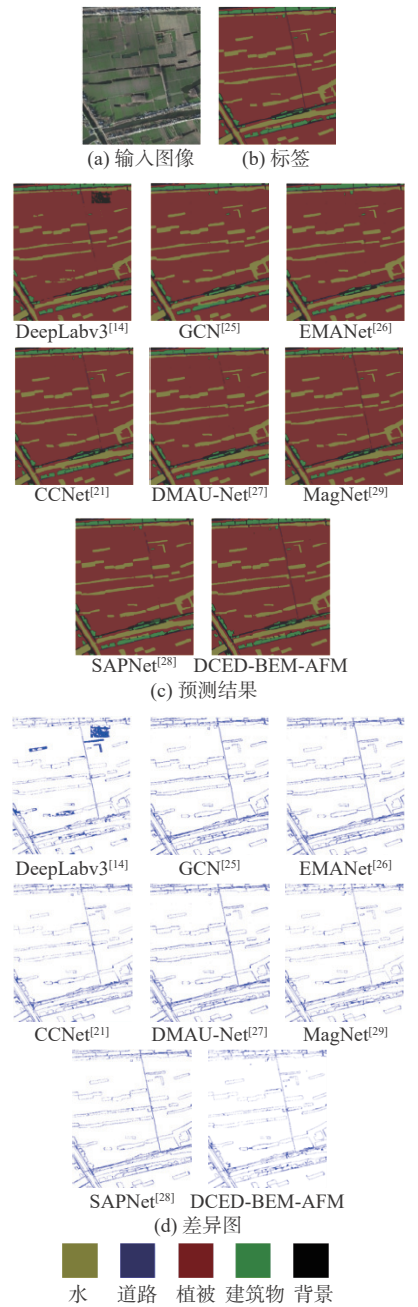


图 6　Jiage 数据集上与 7 种网络的定性比较

Fig. 6　Qualitative comparison with seven networks on Jiage dataset

4 结 论

本文提出一种双路径监督与注意力筛选网络增强对目标物体边界的分割能力，通过实验分析得出以下结论：

1）DCED-BEM-AFM可以提升目标物体边界的分割效果，对边界复杂类别有更好的分割能力。

2）DCED-BEM-AFM可以有效地提取浅层的空间细节信息，提升对小尺度物体的分割效果。

3）通过在Potsdam和Jiage数据集上的对比实验结果显示，DCED-BEM-AFM的分割效果优于9种经典的网络模型，大幅提高了对边界复杂物体的分割精度。

后续将探索遥感图像中因目标物体之间巨大尺度差异导致小尺度目标物体分割精度降低的问题，同时探索优化网络的方法，减少网络参数量，提高预测速度。

参考文献（References）

[1] FAN M, LAI S, HUAN G J, et al. Rethinking BiseNet for real-time semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 9716-9725.

[2] LATHA P, RAJ T. Advanced deep learning method for aerial image segmentation of landscape changes in pre-and post-disaster scenarios[J]. Global NEST Journal, 2022, 24(3): 544-552.

[3] GUO Z L, SHENGOKU H, WU G M, et al. Semantic segmentation for urban planning maps based on U-Net[C]//Proceedings of the IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2018: 6187-6190.

[4] FAN R Y, LI F P, HAN W, et al. Fine-scale urban informal settlements mapping by fusing remote sensing images and building data via a Transformer-based multimodal fusion network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5630316.

[5] ZHU Q Q, ZHANG Y N, WANG L Z, et al. A global context-aware and batch-independent network for road extraction from VHR satellite imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 175: 353-365.

[6] ZHOU M T, SUI H G, CHEN S X, et al. BT-RoadNet: a boundary and topologically-aware neural network for road extraction from high-resolution remote sensing imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 168: 288-306.

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

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

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

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

[11] LIN D, SHEN D G, SHEN S T, et al. ZigZagNet: fusing top-down and bottom-up context for object segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 7482-7491.

[12] 吴泽康, 赵姗, 李宏伟, 等. 遥感图像语义分割空间全局上下文信息网络[J]. 浙江大学学报 (工学版), 2022, 56(4): 795-802.
WU Z K, ZHAO S, LI H W, et al. Spatial global context information network for semantic segmentation of remote sensing image[J]. Journal of Zhejiang University (Engineering Science), 2022, 56(4): 795-802(in Chinese).

[13] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[14] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking-atrous convolution for semantic image segmentation[EB/OL]. (2017-06-17)[2023-03-01]. https://arxiv.org/abs/1706.05587.

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

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

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

[18] ZHOU Z, ZHOU Y, WANG D L, et al. Self-attention feature fusion network for semantic segmentation[J]. Neurocomputing, 2021, 453: 50-59.

[19] 谭大宁, 刘瑜, 姚力波, 等. 基于视觉注意力机制的多源遥感图像语义分割[J]. 信号处理, 2022, 38(6): 1180-1191.
TAN D N, LIU Y, YAO L B, et al. Semantic segmentation of multi-source remote sensing images based on visual attention mechanism[J]. Journal of Signal Processing, 2022, 38(6): 1180-1191(in Chinese).

[20] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 3141-3149.

[21] HUANG Z L, WANG X G, HUANG L C, et al. CCNet: criss-cross attention for semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 603-612.

[22] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 936-944.

[23] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2023-03-

01]. https://arxiv.org/abs/1409.1556.

[24] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[25] PENG C, ZHANG X Y, YU G, et al. Large kernel matters—improve semantic segmentation by global convolutional network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 1743-1751.

[26] LI X, ZHONG Z S, WU J L, et al. Expectation-maximization attention networks for semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 9166-9175.

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

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

[29] HUYNH C, TRAN A T, LUU K, et al. Progressive semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 16750-16759.

Semantic segmentation network of remote sensing images based on dual path supervision

LIU Chunjuan[1]，QIAO Ze[1]，YAN Haowen[2]，WU Xiaosuo[1,*]，WANG Jiawei[1]，XIN Yuqiang[1]

(1. School of Electronic and Information Engineering，Lanzhou Jiaotong University，Lanzhou 730070，China；
2. Faculty of Geomatics，Lanzhou Jiaotong University，Lanzhou 730070，China)

Abstract：A dual path supervision and attention filtering network was proposed to solve the problem of blurry boundary classification of target objects in semantic segmentation tasks of remote sensing images. A supervised boundary extraction module was introduced to increase the channel of boundary information, improve the weight of boundary information in semantic segmentation, and enhance attention to the boundary pixels of the target object. The attention filtering module was introduced to filter out spatial details in shallow networks and abstract semantic information in deep networks through attention maps, discarding redundant information in the network to prevent overfitting. The mean intersection over union of the dual path supervision and attention filtering network on the Potsdam dataset and the Jiage dataset was 85.44% and 86.07% respectively, which increased by 1.24%, 1.28% and 1.54%, 1.27% compared with the suboptimal network MagNet and SAPNet. Experimental results show that the proposed network can more accurately segment the boundaries of target objects.

Keywords：remote sensing images；semantic segmentation；supervision；boundary information；attention filtering

Received：2023-03-31；**Accepted**：2023-05-26；**Published Online**：2023-06-30 15：51

URL：link.cnki.net/urlid/11.2625.V.20230630.1501.004

Foundation items：National Key Research and Development Program of China (2022YFB3903604); Natural Science Foundation of Gansu Province (21JR7RA289); Key Research and Development Projects of Gansu Province (20YF8GA035)

* **Corresponding author**. E-mail：43452740@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0139

基于等边 Bennett 机构的变形翼机构设计与分析

田大可[1]，张珺威[1]，金路[2,*]，刘荣强[3]，雷宏强[1]，崔锡赫[1]

(1. 沈阳建筑大学 机械工程学院，沈阳 110168；2. 沈阳建筑大学 土木工程学院，沈阳 110168；
3. 哈尔滨工业大学 宇航空间机构全国重点实验室，哈尔滨 150001)

摘 要：变体飞行器是一种通过变形机构改变自身气动外形，以适应多种飞行环境与任务要求的新概念飞行器，在航空航天、军事侦察等领域具有重要的应用价值，是未来飞行器研究的前沿与热点。针对变体飞行器广空域、宽速域等新的发展需求，基于等边 Bennett 机构提出一种具有空间多回路闭链特征的展向弯曲变形翼机构。研究等边 Bennett 机构的几何特性，提出基于等边 Bennett 机构的多回路闭环变形翼机构设计方案；采用约束螺旋求解法对变形翼机构自由度进行求解，基于 D-H 坐标变换法建立运动学模型；建立机构三维模型及虚拟样机，并进行运动学仿真验证，制作样机并进行实验。研究结果表明：所提的变形翼机构仅需一个动力源即可驱动机构运动，结构简单、模块化率高；能够实现各零部件间的准确连接及预期的展向弯曲变形动作。研究成果为新型变体飞行器变形翼的基础研究及工程应用提供了借鉴与参考。

关 键 词：变体飞行器；变形翼；模块化结构；空间过约束机构；运动学分析
中图分类号：V224
文献标志码：A **文章编号**：1001-5965(2025)03-0742-11

变体飞行器是随着机构学、空气动力学、材料学等多学科快速发展、交叉融合而产生的一类具有高机动性与高灵活性的新概念飞行器。与传统固定气动外形的飞行器相比，变体飞行器可借助变形机构的运动改变自身的气动外形，从而满足宽飞行包线、高性能、低功耗等新的任务需求[1-2]。按照变形部位的不同，变体飞行器主要包括机身变形和机翼变形 2 种形式。其中，机翼变形包括变展长、变后掠角、变展向弯度及弦向变形等多种形式[3-4]。机翼变形改变飞行器气动外形的幅度大、形式灵活多样，是实现快速自适应多任务飞行的重要途径之一，受到国际航空宇航领域研究机构和研究人员的广泛关注[5-6]。

变形翼机构是实现机翼变形的重要前提和基础。近年来，国内外诸多学者针对变形翼机构开展了广泛而深入的研究，并取得了较为丰硕的研究成果。Neal[7] 提出了一种可根据不同任务进行自适应变形的变体飞行器，通过蜗轮蜗杆机构实现变翼展、变后掠角和尾翼的变形，从而显著改变展弦比、平面形状面积和静态边距，并通过不对称机翼变形进行有效的侧倾控制，实现快速机动。NextGen 公司[8-9] 研发了 MFX-1 变体无人机，该无人机通过平面连杆机构实现机翼的变展长和变后掠角，并使用分布式驱动和柔性蒙皮实现机翼面积与后掠角同步改变。Manzo 等[10] 设计了一种变体飞行器机翼，通过分段式变形机构实现展向变弯度。Lazos[11] 设计了一种展向弯曲变形翼，该机翼可利用特殊设计的单自由度连杆机构、形状记忆合金丝及拉胀结构等实现机翼展向弯曲变形。Wang 等[12] 设计了一种基于 Sarrus 连杆机构概念的单自由度变展长机翼结构并

收稿日期：2023-03-23；录用日期：2023-04-21；网络出版时间：2023-05-12 11:39
网络出版地址：link.cnki.net/urlid/11.2625.V.20230511.1636.004
基金项目：国家自然科学基金联合基金 (U2341237)；宇航空间机构全国重点实验室开放课题；辽宁省教育厅基本科研项目 (LJ222410153096，JYTMS20231592)
*通信作者. E-mail：jinlu@sjzu.edu.cn

引用格式：田大可，张珺威，金路，等. 基于等边 Bennett 机构的变形翼机构设计与分析 [J]. 北京航空航天大学学报，2025，51（3）：742-752. TIAN D K，ZHANG J W，JIN L，et al. Design and analysis of morphing wing mechanism based on equilateral Bennett mechanism [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（3）：742-752（in Chinese）.

进行有限元仿真, 通过仿真结果发现, Sarrus 连杆机构在应用于变展长机翼结构情况下存在应力集中问题。Grant 等[13] 从飞鸟的翅膀得到灵感, 提出了一种多关节可变后掠翼飞行器, 该变形翼结构采用多组连杆和关节连接, 内外 2 层机翼独立驱动, 每个关节可转角度为±30°, 可通过各关节独立驱动实现不同机翼段独立转动, 实现空中快速姿态变换。Moosavian 等[14] 提出了一种模块化驱动变形机翼结构, 通过模块化放置的并联机构对机翼 6 个空间自由度进行可控运动, 以实现机翼扭转变弯度、变后掠角和变展长等多种变形控制。陈钱等[15] 提出了一种能够提升升阻比和降低阻力的变展长/变后掠组合体机翼构型, 通过绕固定点旋转的前缘, 实现“可控剪切”的平行四边形机构与绕固定点旋转的后缘共同作用实现剪切变后掠。王云飞等[16] 提出了一种平行连杆式变形翼, 该变形翼使用平行连杆机构与分布式驱动器配合, 可实现机翼变弦长、变后掠、变面积、变展弦比。肖洪等[17] 提出了一种基于四面体单元的变形翼骨架设计方案, 使用多个驱动器与四面体单元机构组合, 通过不同驱动配置可实现展向弯曲变形、扭转变形、变后掠角等多种变形。

近年来, 变形翼始终保持着较高的研究热度, 飞行器正在向着跨空域、跨速域、高空高速巡航的趋势发展, 对飞行器的高效平稳变形提出更高的要求[18-20]。而空间机构可实现从二维平面到三维曲面的运动变化, 为未来变体飞行器机构创新设计提供了一个新的研究思路。通过将空间机构和平面机构组合设计成多回路闭链空间机构, 实现机构-结构一体化设计, 是一种新颖的变体飞行器机构设计方案。

鉴于此, 本文提出一种基于多回路闭链空间机构的模块化变形翼机构, 该机构可以实现机翼沿展向弯曲变形。首先, 分析了机构几何约束条件, 并基于约束条件提出具有多回路闭链特征的变形翼机构; 其次, 基于螺旋理论对机构进行了自由度分析, 利用机器人学理论对变形翼机构进行运动学建模; 最后, 通过建立虚拟样机完成对机构的运动学分析, 并研制了原理样机, 通过变形实验验证了变形翼机构整体方案的可行性。

1　变形翼机构方案设计

1.1　变形机构原理

机翼展向弯曲变形可视为机翼形状由二维平面向三维曲面的变形, 而 Bennett 机构作为一种经典空间过约束机构, 其构件运动为三维运动, 故选择 Bennett 机构作为变形翼机构基础单元是实现变形翼机构模块化、一体化设计的新途径。本文选用等边 Bennett 机构作为变形翼机构的基础空间机构回路。

等边 Bennett 机构示意图如图 1 所示。其中, 杆 AB、杆 BC、杆 CD、杆 DA 为 Bennett 机构的 4 个连杆; ϕ_a 为杆 AD 和杆 AB 之间的夹角, 根据 Bennett 机构几何关系, ϕ_a 也为杆 BC 和杆 CD 之间的夹角; ϕ_b 为杆 CD 和杆 AD 之间的夹角, 也为杆 AB 和杆 BC 之间的夹角; α_a 和 α_b 分别为关节 D 相对于关节 C 和关节 A 相对于关节 D 的扭角。

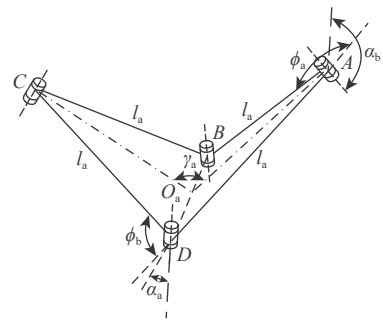


图 1　等边 Bennett 机构示意图

Fig. 1　Schematic of equilateral Bennett mechanism

γ_a 为面 BCD 和面 ABD 之间的二面角 $\angle AO_aC$, 即 Bennett 机构变化角度, 表示为

$$\cos\gamma_a=\frac{1+\cos\phi_a+\cos\phi_a}{\cos\phi_a-1} \tag{1}$$

对于等边 Bennett 机构, 各角度关系如下[21]:

$$\tan\left(\frac{\phi_a}{2}\right)\tan\left(\frac{\phi_b}{2}\right)=\frac{\sin\left(\dfrac{\alpha_a+\alpha_b}{2}\right)}{\sin\left(\dfrac{\alpha_a-\alpha_b}{2}\right)}=-\frac{1}{\cos\alpha_a} \tag{2}$$

将式 (1)、式 (2) 联立, 可得机构变化角度 γ_a、相邻连杆扭角 α_a 和相邻连杆间角度 ϕ_a 三者关系为

$$\gamma_a=\arccos\left(\frac{1+\cos\phi_a+\dfrac{2\left(\cos^2\alpha_a\tan^2\left(\dfrac{\phi_a}{2}\right)-1\right)}{1+\cos^2\alpha_a\tan^2\left(\dfrac{\phi_a}{2}\right)}}{\cos\phi_a-1}\right) \tag{3}$$

在不同相邻关节扭角 α_a 下, 求解相邻连杆间角度 ϕ_a 与机构变化角度 γ_a 之间的关系, 如图 2 所示。可以看出, 在固定相邻连杆扭角 α_a 的情况下, 相邻连杆间角度 ϕ_a 与相对应的等边 Bennett 机构变化角度 γ_a 相同。

由于等边 Bennett 机构可以实现二面角的改变, 采用模块化思路, 通过由多个等边 Bennett 机构组合成的多回路闭链空间机构进行沿机翼弦向分布, 实现机翼沿展向变弯度的变形结构。其中, 沿弦向分布的相邻等边 Bennett 机构回路相邻连杆扭角为 α_1 和 $\pi-\alpha_1$, 可实现沿弦向的多个回路联动; 在

展向上，为方便安装翼肋和承受升力载荷，采用平面四杆机构回路和等边 Bennett 机构回路交错放置的方式进行组合，组合后的多回路闭链空间机构如图 3 所示。

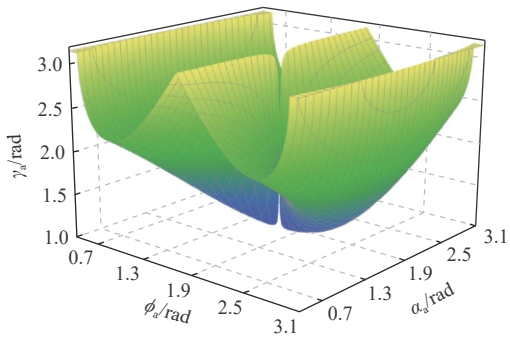


图 2 等边 Bennett 机构参数关系

Fig. 2 Parameter relationship of equilateral Bennett mechanism

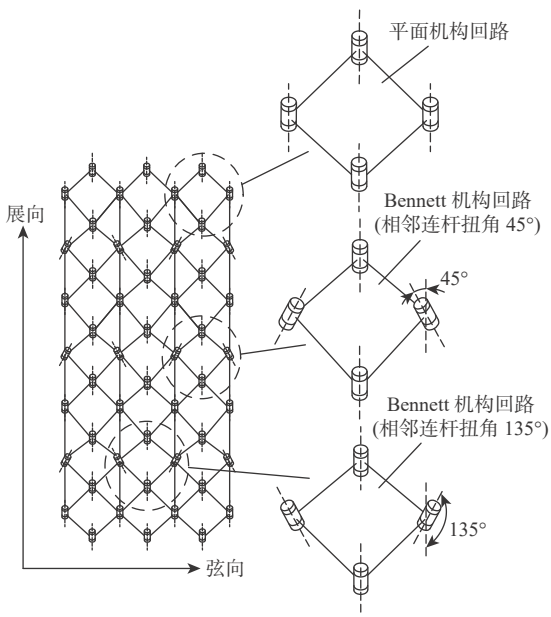


图 3 变形翼机构示意图

Fig. 3 Schematic of morphing wing mechanism

1.2 变形翼结构设计

以等边 Bennett 机构和平面机构作为变形机构回路，采用模块化设计思想，基于等边 Bennett 机构的几何特性，对多个变形机构回路进行组合，同时重新设计组合处的连接结构，进而实现模块化多回路闭链变形翼结构设计。

1.2.1 变形机构回路设计

各变形机构回路如图 4 所示，具体设计过程如下：

1）将每个回路中的杆件进行重新设计，将平面机构回路中杆件设计为 L 形构件，得到变形机构回路 1，变形机构回路 1 为 4 个相同组件组成的平面四杆机构，由构件 1 顺次连接组合而成。

2）将 Bennett 机构回路中的各杆件设计为由一斜板和一直板组成的构件，得到变形机构回路 2 和变形机构回路 3，其中，变形机构回路 2 由构件 2 和构件 3 交错连接组合而成。

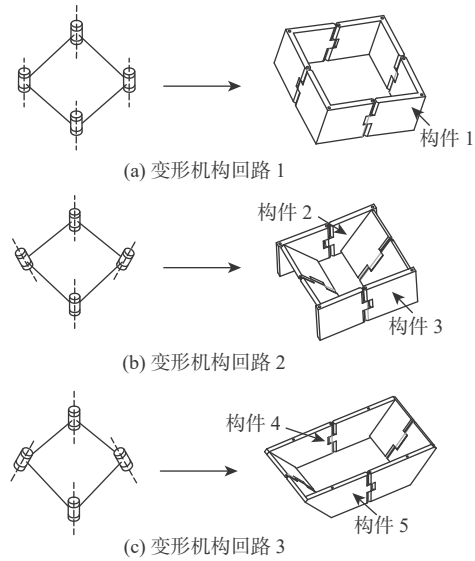


图 4 3 种变形机构回路结构

Fig. 4 Loop structures of three kinds of morphing mechanism

3）变形机构回路 3 的结构与变形机构回路 2 相似，区别在于：旋转斜置的 2 个转动副角度不同，变形机构回路 3 由构件 4 和构件 5 交错连接组合而成。

1.2.2 变形翼机构组网设计

各回路在机构中的分布如图 5 所示。

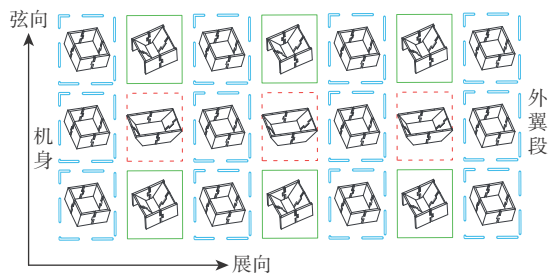


图 5 变形机构回路位置分布

Fig. 5 Location distribution of morphing mechanism loop

为实现各回路联动，对相邻回路的共用杆件进行重新设计：

1）2 个构件 1 组合成 1 个 T 型组件，即外边组件；单一的构件 1 命名为外角组件。

2）构件 1 和构件 3 组合成三接头组件 1；构件 1 和构件 2 组合成三接头组件 2。

3）构件 2、构件 5 与 2 个构件 1 组成四接头组件 1；构件 3、构件 4 与 2 个构件 1 组成四接头组件 2。

由以上规则，可得图 6 所示的各组件在机构中分布情况。

对各组件进行组合设计，将与斜置铰链相连接的边设计为华伦桁架，其余各边设计为矩形框架结构，设计后各零件及装配后的变形翼机构如图 7 所示。

由于变形机构回路 2 和变形机构回路 3 均为空

间机构, 以变形机构回路 2 为例, 阐述变形翼机构变形原理, 如图 8 所示。定义 I、J、K、L 这 4 个点为变形机构回路 2 各转动副轴线与该回路所包含零件的上表面交点, 定义 2 个平面 P_1、P_2, 其中, P_1 由 I、J、L 这 3 个点确定, 平面 P_2 由 J、K、L 这 3 个点确定。在变形机构回路 2 中, 2 个相邻组件间夹角发生变化时, P_1、P_2 这 2 个平面之间的夹角 γ'改变, 由于变形翼机构中存在多列空间机构, 通过多列空间机构实现多个位置的二面角 γ'改变, 进而实现变形翼机构沿机翼展向弯曲变形。

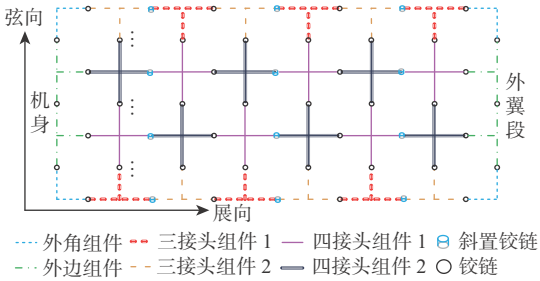


图 6　各组件位置分布

Fig. 6　Distribution of each component location

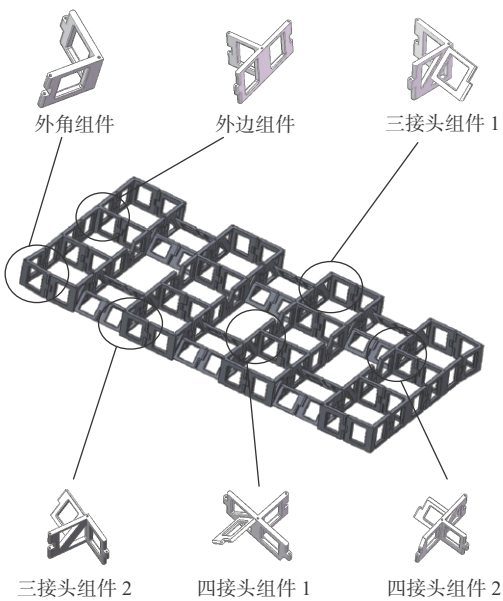


图 7　变形翼机构三维模型

Fig. 7　3D model of morphing wing mechanism

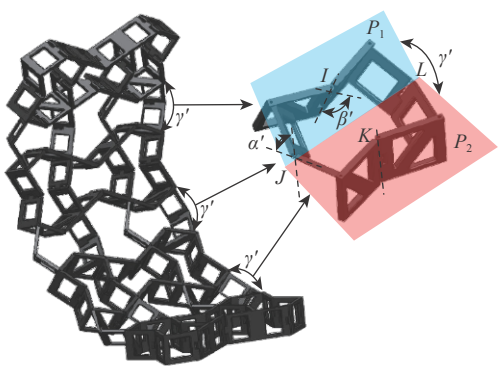


图 8　变形翼机构原理示意图

Fig. 8　Principle of morphing wing mechanism

为实现变体飞行器机翼的气动外形设计, 在变形翼机构中的平面机构段加装支撑翼肋, 如图 9 所示, 为后续变形翼气动外形设计和柔性蒙皮加装提供设计与研究基础。展向变弯度变体飞行器结构如图 10 所示。通过变形翼机构中的空间机构回路, 可实现变体飞行器沿展向向上或向下弯曲, 以达到更优的气动外形, 满足多任务飞行需求。

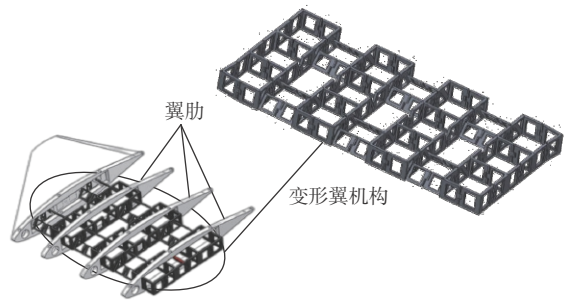
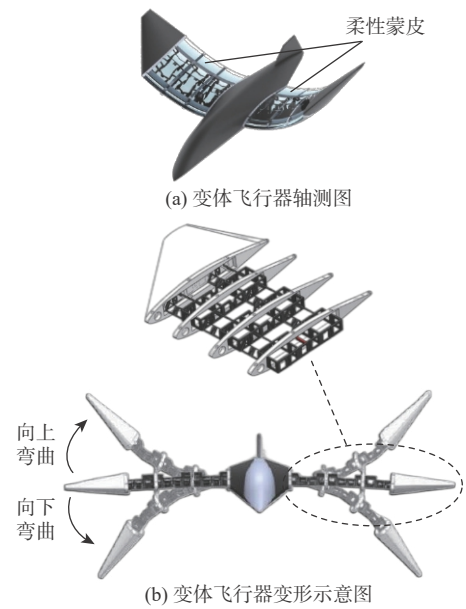


图 9　变形翼结构示意图 (包含翼肋)

Fig. 9　Schematic of morphing wing structure (including wing rib)



(a) 变体飞行器轴测图

(b) 变体飞行器变形示意图

图 10　变体飞行器结构示意图

Fig. 10　Schematic of morphing aircraft structure

2　变形机构运动特性分析

2.1　机构自由度分析

使用约束螺旋求解法对变形机构进行自由度分析, 先针对变形机构回路进行分析, 基于螺旋理论计算机构中的公共约束, 机构的自由度 M 可由修正的 Grübler-Kutzbach 公式 (G-K 公式) 表述为[22]

$$M = d(n-g-1)+\sum_{i=1}^{g} f_i + v \tag{4}$$

式中: d 为机构的阶数; n 为包括机架的构件数目; g 为运动副的数目; f_i 为第 i 个运动副提供的自由度

数；v 为并联冗余约束数。

机构的阶数 d 由机构公共约束数 λ 决定：

$$d = 6 - \lambda \tag{5}$$

变形机构回路 2 和变形机构回路 3 均为 Bennett 机构，图 11 为等边 Bennett 机构约束螺旋参数示意图，EG=2m_1，FH=2m_2，EG、FH 之间距离为 n_1，EG 与 FH 的夹角为 π/2，因此，该等效 Bennett 机构的运动螺旋系的 4 个螺旋可以求得为

$$\begin{cases} S_E = \{m_1, n_1, 0; 0, 0, -m_1^2\} \\ S_F = \{m_2, 0, n_1; 0, -m_2^2 - n_1^2, 0\} \\ S_G = \{-m_1, n_1, 0; 0, 0, -m_1^2\} \\ S_H = \{-m_2, 0, n_1; 0, -m_2^2 - n_1^2, 0\} \end{cases} \tag{6}$$

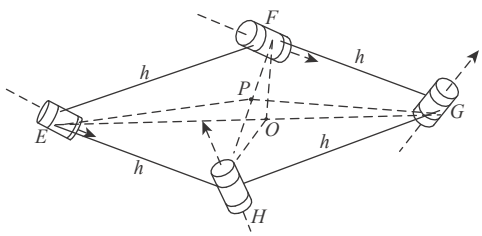


图 11 等边 Bennett 机构约束螺旋参数

Fig. 11 Constrained spiral parameters of equilateral Bennett mechanism

从螺旋表达式 (6) 可推导出

$$h(S_E - S_G) = h(S_F - S_H) \tag{7}$$

故上述 4 个螺旋线性相关，通过进一步推导可得到该运动螺旋系的秩为 3，因此，该运动螺旋系存在 3 个反螺旋，λ=3。

由于变形机构回路 3 与变形机构回路 2 类似，均为等边 Bennett 机构，变形机构回路 3 的螺旋及公共约束求解与变形机构回路 2 类似，在此不再赘述。

变形机构回路 1 为平面机构，故其公共约束为 λ=3。因此

$$d = 6 - \lambda = 3 \tag{8}$$

在变形机构中，所有组件连接均采用铰链连接，组件间运动副均为转动副，整体机构为多回路闭链空间机构，f_i=1，变形机构无冗余约束，故 v=0，应用修正的 G-K 公式对变形机构的自由度进行求解：

$$M = 3 \times (36 - 52 - 1) + 52 + 0 = 1 \tag{9}$$

由式 (9) 可见，变形机构为单自由度机构，通过一个驱动器即可实现展向弯曲变形。

2.2 变形机构运动学建模

对变形翼机构进行分析，机翼变形过程为分段变形，每段通过 Bennett 机构回路进行一次二面角改变，故先将变形翼机构划分为 4 组机构，如图 12 所示。

a 组由平面机构回路组成，由自由度分析可知，变形机构自由度为 1，因此，在 a 组中间的平面机构回路设计同步驱动机构，使其带动其他回路同步变

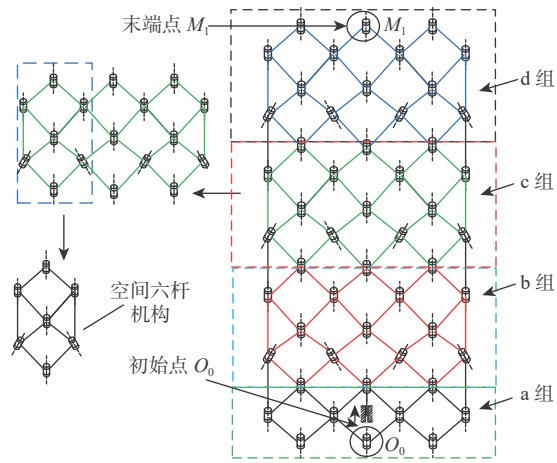


图 12 变形翼机构模块组合示意图

Fig. 12 Schematic of morphing wing mechanism module combination

形，a 组与 b 组包含回路组合如图 13 所示。令平面机构回路杆长为 l，其中，杆 1 和杆 6 为三副杆，杆 1 和杆 6 上位于平面机构回路和回路 1 之间角度为 θ，平面机构回路相邻两连杆角度分别为 β 和 π−β，在机翼平直状态下 β=π/2，在以速度 v_1 驱动下，β 表示为

$$\beta = 2\arccos\left(\frac{\sqrt{2}l + v_1 t}{2l}\right) \tag{10}$$

回路 1 中，相邻连杆角度分别为 ϕ_1、ϕ_2，ϕ_2 可由 β 与 θ 表述为

$$\phi_2 = 2\theta + \beta - \pi \tag{11}$$

故 $\{O_1\}$坐标系对于 $\{O_0\}$坐标系的变换矩阵为

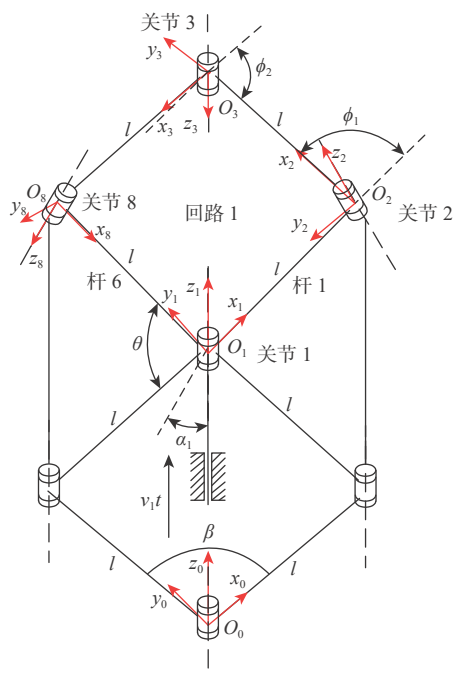


图 13 带有同步驱动机构的平面机构回路示意图

Fig. 13 Schematic of planar mechanism loop with synchronous drive mechanism

$$
{}_1^0\boldsymbol{T}=\begin{bmatrix}
\cos\left(\dfrac{\pi/2-\beta}{2}\right) & -\sin\left(\dfrac{\pi/2-\beta}{2}\right) & 0 & \dfrac{2l+\sqrt{2}v_1t}{2}\\
\sin\left(\dfrac{\pi/2-\beta}{2}\right) & \cos\left(\dfrac{\pi/2-\beta}{2}\right) & 0 & \dfrac{2l+\sqrt{2}v_1t}{2}\\
0 & 0 & 1 & 0\\
0 & 0 & 0 & 1
\end{bmatrix} \tag{12}
$$

除 a 组外, b、c、d 各组均包含 3 个空间 4R 机构和 3 个平面 4R 机构。在一组变形机构中, 通过空间 4R 机构的运动可以实现角度改变, 对一组变形机构进行划分, 将其划分为多个空间六杆机构, 该空间六杆机构由一个平面 4R 机构和一个空间 4R 机构通过共用 2 个杆件组合而成, 针对该空间六杆机构进行分析。

对平面 4R 机构和空间 4R 机构组合成的六杆机构建立 D-H 坐标系, 其中, 空间 4R 机构为回路 1, 平面 4R 机构为回路 2, 空间六杆机构示意图和 D-H 坐标系如图 14 所示, 杆 2 和杆 5 为三副杆, 杆 2 和杆 5 上位于回路 1 和回路 2 之间角度均为 θ。

使用 D-H 法对空间六杆机构进行运动学分析, 结合图 14 建立该六杆机构的 D-H 参数表, 如表 1 所示。其中, Q_i 为等边 Bennett 机构杆长, ϕ_i 为相邻连杆间夹角, α_i 为相邻关节扭角, d_i 为相邻转动关节距离。

回路 1 中, ϕ_1 与 ϕ_2 关系为

$$
\tan\left(\frac{\phi_1}{2}\right)\tan\left(\frac{\phi_2}{2}\right)=\frac{\sin\left(\dfrac{\alpha_1+\alpha_2}{2}\right)}{\sin\left(\dfrac{\alpha_1-\alpha_2}{2}\right)}=-\frac{1}{\cos\alpha_1} \tag{13}
$$

式中: $\alpha_2=\pi-\alpha_1$。

关节 2 对于关节 1 的变换矩阵为

$$
{}_2^1\boldsymbol{T}=\begin{bmatrix}
\cos\phi_1 & -\sin\phi_1 & 0 & l\\
\sin\phi_1\cos\alpha_1 & \cos\phi_1\cos\alpha_1 & -\sin\alpha_1 & 0\\
\sin\phi_1\sin\alpha_1 & \cos\phi_1\sin\alpha_1 & \cos\alpha_1 & 0\\
0 & 0 & 0 & 1
\end{bmatrix} \tag{14}
$$

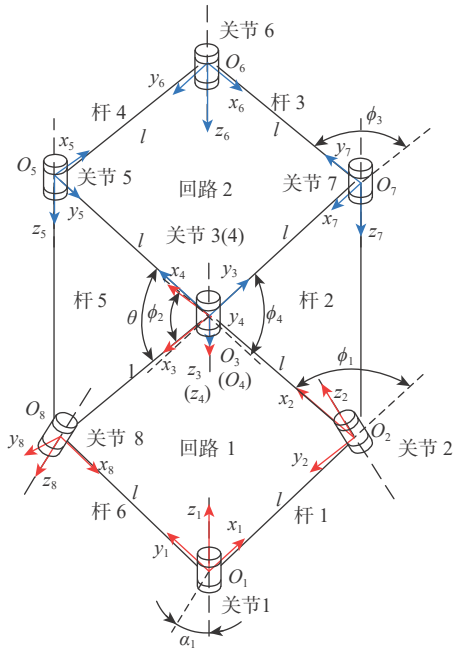


图 14　空间六杆机构 D-H 坐标系

Fig. 14　D-H coordinate system of space six-bar mechanism

表 1　空间六杆机构 D-H 参数

Table 1　D-H parameters of space six-bar mechanism

i	a_i	α_i	d_i	ϕ_i
1	l	α_i	0	ϕ_1
2	l	$\pi-\alpha_i$	0	ϕ_2
3	l	0	0	ϕ_3
4	l	0	0	ϕ_4

关节 3 对于关节 2 的变换矩阵为

$$
{}_3^2\boldsymbol{T}=\begin{bmatrix}
\cos\phi_2 & -\sin\phi_2 & 0 & l\\
-\sin\phi_2\cos\alpha_1 & -\cos\phi_2\cos\alpha_1 & -\sin\alpha_1 & 0\\
\sin\phi_2\sin\alpha_1 & \cos\phi_2\sin\alpha_1 & -\cos\alpha_1 & 0\\
0 & 0 & 0 & 1
\end{bmatrix} \tag{15}
$$

故关节 3 对于关节 1 的变换矩阵为

$$
{}_3^1\boldsymbol{T}={}_2^1\boldsymbol{T}{}_3^2\boldsymbol{T}=\begin{bmatrix}
\cos\phi_1\cos\phi_2+\sin\phi_1\sin\phi_2\cos\alpha_1 & -\cos\phi_1\sin\phi_2+\sin\phi_1\cos\phi_2\cos\alpha_1 & \sin\phi_1\sin\alpha_1 & l\cos\phi_1+l\\
\sin\phi_1\cos\phi_2\cos\alpha_1-\cos\phi_1\sin\phi_2\cos^2\alpha_1-\sin\phi_1\sin^2\alpha_1 & -\sin\phi_1\sin\phi_2\cos\alpha_1-\cos\phi_1\cos\phi_2\cos^2\alpha_1-\cos\phi_2\sin^2\alpha_1 & -\cos\phi_1\sin\alpha_1\cos\alpha_1-\sin\alpha_1\cos\alpha_1 & l\sin\phi_1\cos\alpha_1\\
\sin\phi_1\cos\phi_2\sin\alpha_1-\cos\phi_1\sin\phi_2\cos\alpha_1\sin\alpha_1-\cos\phi_1\sin\phi_1\sin\alpha_1 & -\sin\phi_1\sin\phi_2\sin\alpha_1-\cos\phi_1\cos\phi_2\cos\alpha_1\sin\alpha_1-\cos\phi_2\sin\alpha_1\cos\alpha_1 & -\cos\phi_1\sin^2\alpha_1-\cos^2\alpha_1 & l\sin\phi_1\sin\alpha_1\\
0 & 0 & 0 & 1
\end{bmatrix} \tag{16}
$$

其中, 变换矩阵的第 4 列去掉第 4 行代表末端点在固定坐标系中的坐标, 由此可得 O_3 点在固定坐标系 $\{O_1\}$ 的坐标为

$$
{}^{O_1}\boldsymbol{P}_{O_3}=\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}l\cos\phi_1+l\\l\sin\phi_1\cos\alpha_1\\l\sin\phi_1\sin\alpha_1\end{bmatrix} \tag{17}
$$

$\{O_4\}$坐标系对于$\{O_3\}$坐标系的变换矩阵为

$$ {}^{3}_{4}\boldsymbol{T}=\begin{bmatrix}\cos\theta & -\sin\theta & 0 & 0\\ \sin\theta & \cos\theta & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{18}$$

对于该空间六杆机构，回路1为等边Bennett机构，其输出角为ϕ_2，则回路2中的输入角为

$$\phi_3=2\theta-\phi_2 \tag{19}$$

将杆4和杆5之间的转动副去掉，则关节5对于关节4的变换矩阵为

$$ {}^{4}_{5}\boldsymbol{T}=\begin{bmatrix}\cos\phi_3 & -\sin\phi_3 & 0 & l\\ \sin\phi_3 & \cos\phi_3 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{20}$$

关节6对于关节5的变换矩阵为

$$ {}^{5}_{6}\boldsymbol{T}=\begin{bmatrix}\cos\phi_4 & -\sin\phi_4 & 0 & l\\ \sin\phi_4 & \cos\phi_4 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{21}$$

ϕ_3、ϕ_4关系为

$$\phi_3+\phi_4=\pi \tag{22}$$

故关节6对于关节4的变换矩阵为

$$ {}^{4}_{6}\boldsymbol{T}={}^{4}_{5}\boldsymbol{T}{}^{5}_{6}\boldsymbol{T}=\begin{bmatrix}\cos\phi_3\cos\phi_4-\sin\phi_3\sin\phi_4 & -\cos\phi_3\sin\phi_4-\sin\phi_3\cos\phi_4 & 0 & l\cos\phi_3+l\\ \sin\phi_3\cos\phi_4+\cos\phi_3\sin\phi_4 & -\sin\phi_3\sin\phi_4+\cos\phi_3\cos\phi_4 & 0 & l\sin\phi_3\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{23}$$

综上，得到回路2中$\{O_6\}$坐标系在以$\{O_4\}$坐标系为固定坐标系下的坐标为

$$ {}^{O_4}\boldsymbol{P}_{O_6}=\begin{bmatrix}x\\ y\\ z\end{bmatrix}=\begin{bmatrix}l\cos\phi_3+l\\ l\sin\phi_3\\ 0\end{bmatrix} \tag{24}$$

若以$\{O_1\}$坐标系为固定坐标系，则$\{O_6\}$坐标系方位可由如下矩阵求得：

$$ {}^{1}_{6}\boldsymbol{T}={}^{1}_{3}\boldsymbol{T}{}^{3}_{4}\boldsymbol{T}{}^{4}_{6}\boldsymbol{T} \tag{25}$$

综上，完成了由空间机构回路和平面机构回路所构成的空间六杆机构的运动学分析。由图12可知，在沿展向分布有b、c、d这3组机构，图15为b、c这2组机构中的空间六杆机构组合示意图，规定b组中所包含的空间六杆机构变换矩阵为$ {}^{1b}_{6b}\boldsymbol{T}$，c组中所包含的空间六杆机构变换矩阵为$ {}^{1c}_{6c}\boldsymbol{T}$，由b组空间六杆机构到c组空间六杆机构需进行一次坐标转换：

$$ {}^{6b}_{1c}\boldsymbol{T}=\begin{bmatrix}\cos\theta & -\sin\theta & 0 & 0\\ \sin\theta & -\cos\theta & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{26}$$

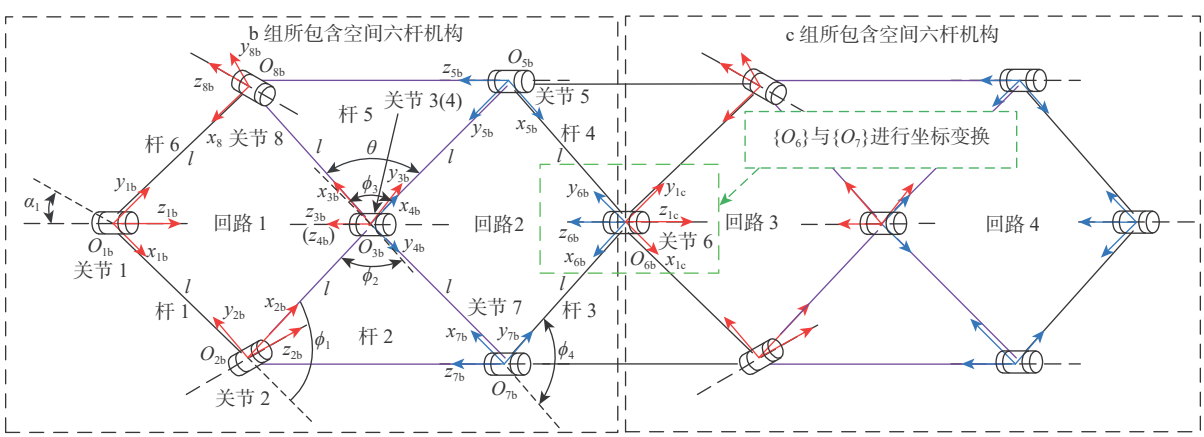


图15 2组空间六杆机构组合示意图

Fig. 15 Combination of two groups of space six-bar mechanism

由于b、c、d这3组机构形式相同，规定d组中所包含空间六杆机构的变换矩阵为$ {}^{1d}_{6d}\boldsymbol{T}$，由c组所属空间六杆机构到d组所属空间六杆机构亦进行坐标变换：

$$ {}^{6c}_{1d}\boldsymbol{T}={}^{6b}_{1c}\boldsymbol{T}=\begin{bmatrix}\cos\theta & -\sin\theta & 0 & 0\\ -\sin\theta & -\cos\theta & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{27}$$

由此，可得到末端点M_1所在坐标系$\{M_1\}$到固

定点 O_0 所在坐标系 $\{O_0\}$ 的相对位置的变换矩阵为

$$ {}^{O_0}_{M_1}\boldsymbol{T} = {}^{0}_{1}\boldsymbol{T}\,{}^{1b}_{6b}\boldsymbol{T}\,{}^{6b}_{1c}\boldsymbol{T}\,{}^{1c}_{6c}\boldsymbol{T}\,{}^{6c}_{1d}\boldsymbol{T}\,{}^{1d}_{6d}\boldsymbol{T} \tag{28} $$

综上, 完成了对变形翼机构沿展向的运动学分析, 同时该运动学分析方法适用于更多模块拓扑情况下的运动学分析, 具有较好的通用性和拓展性。

3　机构仿真分析与样机验证

3.1　机构仿真分析

变形翼在飞行状态下进行变形, 变形的运动学特性对其整体空气动力学特性影响显著, 因此, 对变形翼机构的运动学分析具有重要意义。

由于该变形翼机构通过等边 Bennett 机构回路实现展向弯曲变形, 通过三维实体建模, 并将模型导入到 ADAMS/View 中, 如图 16 所示。对变形翼机构中各铰链设置转动副, 图 16 右端中间转动副设置驱动。

在变形翼机构中, 变形机构回路 1 在弦向组成一列, 用于与翼肋装配, 进而带动变形翼变形, 因此, 选取变形机构回路 1 的中间回路沿升力方向的最高点作为特征点, 在该机构中共选取 4 个特征点, 图 17 为各特征点所在位置, 图 18 为机构向下弯曲变形过程中末端点 (即特征点 4) 运动轨迹。

应用 MATLAB 对运动学模型进行仿真, 将 MATLAB 运动学模型中的位移数据点与 Adams 虚拟样机中测量得到的位移数据点进行对比, 2 组数据算法不同, 因此, 可进行对比验证, 如图 19 所示。可以看到, 2 组数据趋势一致, 且曲线高度相似, 验证了运动学模型的正确性。对 2 个外边组件之间施加 10 (°)/s 角速度, 机翼在变形过程中沿升力方向向上变形, 变形过程如图 20 所示。对变形翼机构的各特征点进行位移、加速度和角加速度测量。变形翼机构中各特征点的升力方向位移如图 21 所示, 各特征点速度和加速度如图 22 和图 23 所示, 各特征点所属零件的角加速度如图 24 所示, 由于特征点 1 仅在平面内运动, 测量特征点 2、3、4 的位移、速度、加速度及其所属零件角加速度。

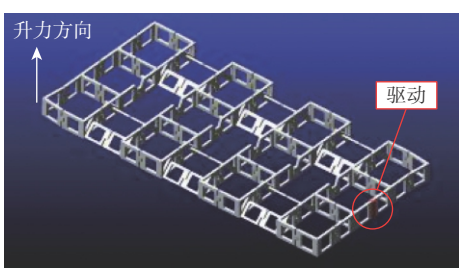


图 16　变形翼机构虚拟样机

Fig. 16　Virtual prototype of morphing wing mechanism

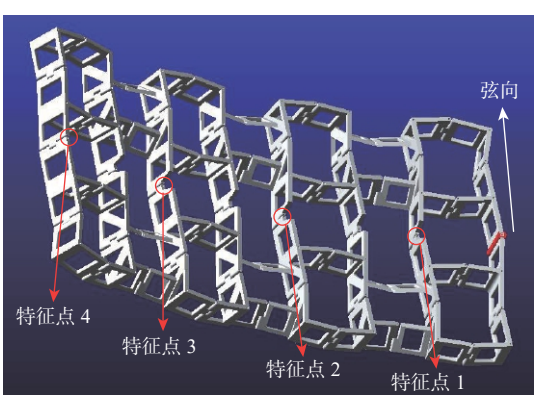


图 17　各特征点位置

Fig. 17　Location of each feature point

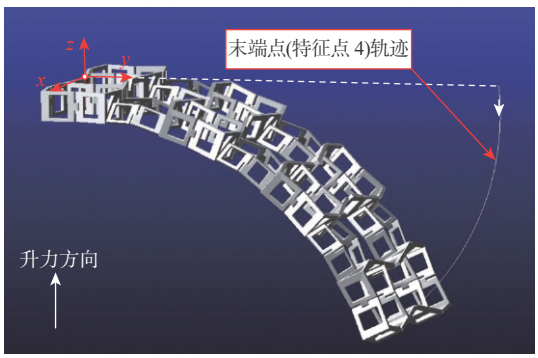


图 18　虚拟样机末端点运动轨迹

Fig. 18　End point movement trajectory of virtual prototype

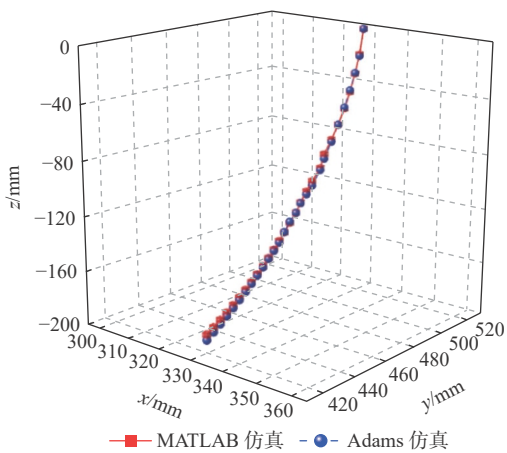


图 19　末端点运动轨迹仿真对比

Fig. 19　Simulation comparison of end point movement trajectories

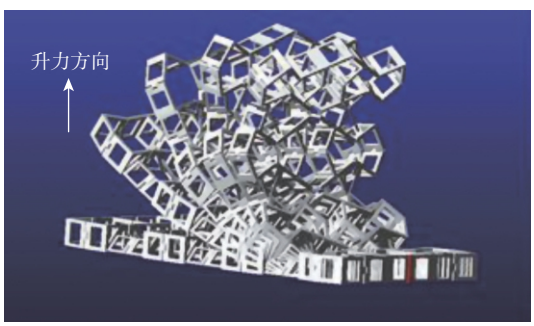


图 20　变形翼机构变形过程

Fig. 20　Deformation process of morphing wing mechanism

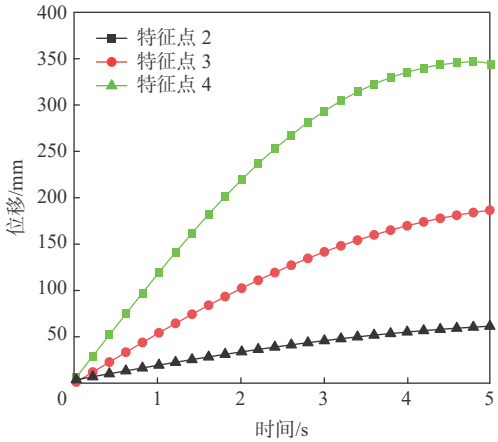


图 21 向上弯曲变形升力方向位移曲线

Fig. 21 Lift direction displacement curves in upward bending

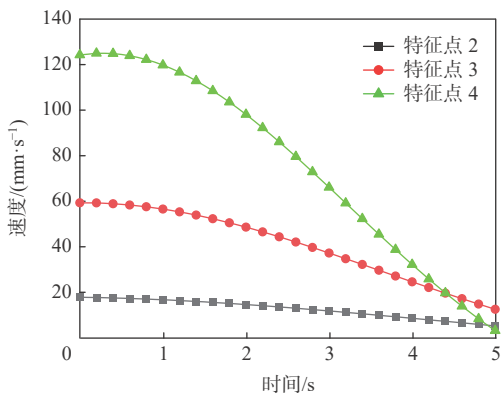


图 22 特征点速度曲线

Fig. 22 Velocity curves of feature points

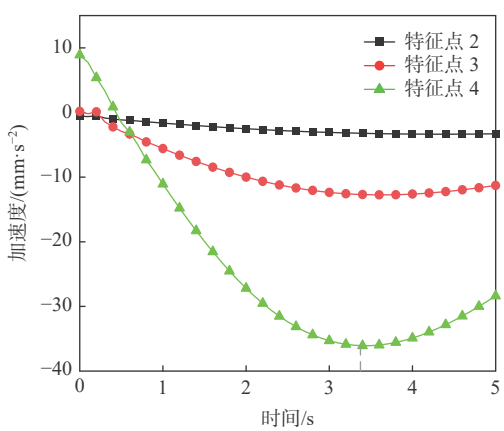


图 23 特征点加速度曲线

Fig. 23 Acceleration curves of feature points

对图 19～图 24 进行分析，可以发现：

1）由图 19、图 20 可以看出，机构可沿升力方向向上/下进行弯曲变形；同时由图 21 可以看出，特征点 4 升力方向位移达到最大后减小，与预期变形翼机构最大变形角度超过 90°变形相符合。

2）由图 23、图 24 可见，特征点 2、3、4 加速度及其所属零件的角加速度趋势相同，均为先增大再减小，但各特征点加速度及各特征点所属零件的角加速度变化幅度逐一增大，这是由于多回路闭链机构中，上一回路的输出角度作为下一回路的输入角度，各回路运动相对于翼根沿展向叠加。

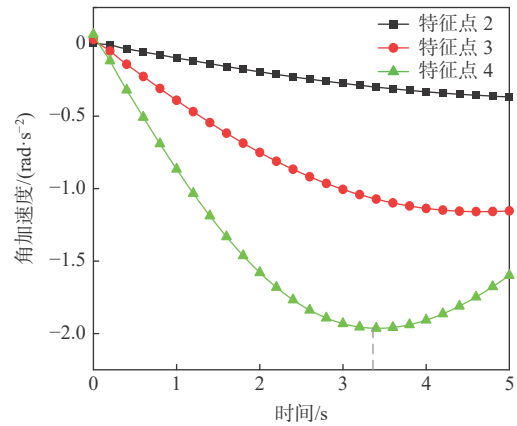


图 24 特征点角加速度曲线

Fig. 24 Angular acceleration curves of feature points

3）通过图 23、图 24 可以看出，在 3.2 s 前加速度和角加速度变化较快，因此在后续驱动设计中，需要抑制加速度快速变化带来的冲击。

4）进一步对图 21～图 24 进行分析，可以看到，各图的运动规律曲线整体呈平滑状态，说明变形翼机构整体变形流畅性较好、同步性较高。

3.2 样机研制与功能验证

根据以上设计方案和仿真分析，设计并研制了变形翼机构原理样机，每个变形机构回路边长 45 mm，铰链直径 2 mm，各零件壁厚 4 mm。样机采用 3D 打印制作，原理样机各零件如图 25 所示。

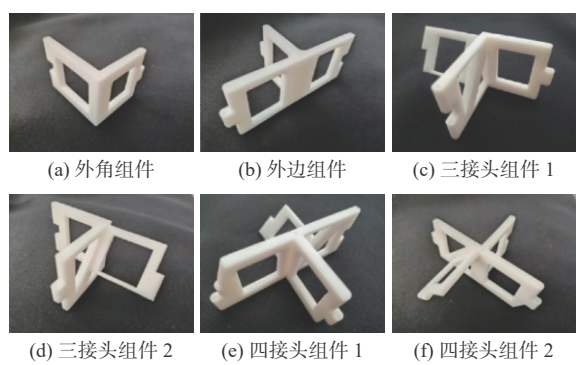


图 25 变形翼机构原理样机零件

Fig. 25 Parts of morphing wing mechanism prototype

变形翼原理样机如图 26 所示，其中，铰链采用 304 钢销轴进行配合；该样机采用直线电机作为驱动器，在设置电机不同位移量的情况下，该变形翼机构可带动支撑翼肋实现机翼弯曲变形；翼肋选用轻木制作，采用 NACA0012 翼型，使用螺栓将翼肋与变形翼机构连接装配。

原理样机变形实验如图 27 所示，对该样机进

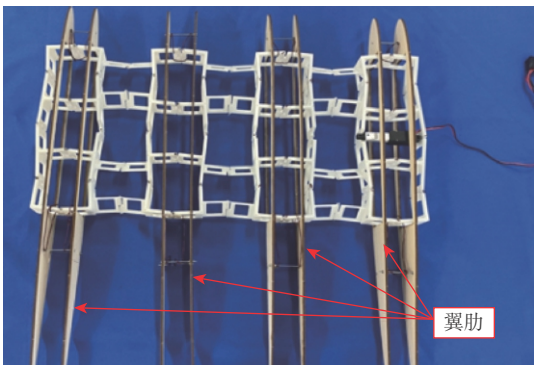


图 26　变形翼机构原理样机

Fig. 26　Morphing wing mechanism prototype

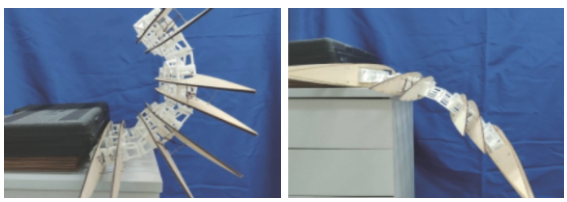


图 27　变形翼机构原理样机变形实验

Fig. 27　Deformation experiment of morphing wing mechanism prototype

行机翼变形实验。通过实验可看到, 变形过程与仿真结果一致, 通过向上/下展向弯曲验证了机构能够带动翼肋有效实现变形翼沿展向向上/下弯曲变形。

4　结　论

本文针对变体飞行器发展需求, 对变形翼机构进行了创新设计与研究, 开展方案设计、运动特性分析、仿真分析及实验验证, 通过研究可以得到以下主要结论:

1) 提出了一种基于多回路闭链空间机构的变形翼机构, 该机构可实现机翼沿展向弯曲变形, 进而带动翼肋实现变体飞行器在飞行状态中变形, 为变形翼机构设计提供一种参考方案, 同时也为多回路闭链空间机构在更多领域的应用提供了新思路。

2) 进行了变形翼机构运动特性分析, 基于螺旋理论, 求得了多回路闭链空间变形机构自由度为 1, 表明该机构通过单一驱动即可实现机翼沿展向弯曲变形。

3) 通过运动学仿真分析发现, 各特征点运动趋势一致; 同时, 对加速度曲线和角加速度曲线分析可以看到, 在 0～3.2 s 间加速度和角加速度变化较快, 机构可能存在由于速度波动引起展开平稳性下降的隐患, 该分析为后续驱动设计和运动规划研究提供了参考。

4) 研制了变形翼机构原理样机, 并进行了变形功能验证, 变形过程无干涉、卡死等现象, 验证了机构设计的正确性。该研究成果对于丰富变体飞行器机构基础理论及促进工程应用等具有较为重要的指导意义和参考价值。

参考文献（References）

[1] 周文雅, 张宗宇, 王晓明, 等. 机翼中小尺度主动变形研究进展及关键技术[J]. 机械工程学报, 2021, 57(2): 121-138.
ZHOU W Y, ZHANG Z Y, WANG X M, et al. Research progress and key techniques of active morphing wing at medium and small scales[J]. Journal of Mechanical Engineering, 2021, 57(2): 121-138 (in Chinese).

[2] AJAJ R M, PARANCHEERIVILAKKATHIL M S, AMOOZGAR M, et al. Recent developments in the aeroelasticity of morphing aircraft[J]. Progress in Aerospace Sciences, 2021, 120: 100682.

[3] CHU L L, LI Q, GU F, et al. Design, modeling, and control of morphing aircraft: a review[J]. Chinese Journal of Aeronautics, 2022, 35(5): 220-246.

[4] KAO J Y, CLARK D L, WHITE T, et al. Conceptual multidisciplinary design and optimization of morphing aircraft: AIAA-2019-0175[R]. Reston: AIAA, 2019.

[5] GUO J B, ZHAO C J, SONG Z G. Discussion on research status and key technologies of morphing aircraft[C]//Proceedings of the International Conference on Mechatronics Technology and Aerospace Engineering. Bristol: IOP Publishing, 2021: 24-26.

[6] 冉茂鹏, 王成才, 刘华华, 等. 变体飞行器控制技术发展现状与展望[J]. 航空学报, 2022, 43(10): 527449.
RAN M P, WANG C C, LIU H H, et al. Research status and future development of morphing aircraft control technology[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(10): 527449(in Chinese).

[7] NEAL A D. Design, development, and analysis of a morphing aircraft model for wind tunnel experimentation[D]. Blacksburg: Virginia Polytechnic Institute, 2006.

[8] FLANAGAN J, STRUTZENBERG R, MYERS R, et al. Development and flight testing of a morphing aircraft, the NextGen MFX-1: AIAA-2007-1707[R]. Reston: AIAA, 2007.

[9] YOUNG T M, HIRST M. Innovation in aeronautics[M]. [S.l.]: Woodhead Publishing, 2012: 37-55.

[10] MANZO J, GARCIA E. Demonstration of an in situ morphing hyperelliptical cambered span wing mechanism[J]. Smart Materials and Structures, 2010, 19(2): 025012.

[11] LAZOS B S. Biologically inspired fixed-wing configuration studies[J]. Journal of Aircraft, 2005, 42(5): 1089-1098.

[12] WANG Q, CHEN Y, TANG H. Mechanism design for aircraft morphing wing: AIAA-2012-1608[R]. Reston: AIAA, 2012.

[13] GRANT D T, ABDULRAHIM M, LIND R. Flight dynamics of a morphing aircraft utilizing independent multiple-joint wing sweep [J]. International Journal of Micro Air Vehicles, 2010, 2(2): 91-106.

[14] MOOSAVIAN A, XI F F, HASHEMI S M. Design and motion control of fully variable morphing wings[J]. Journal of Aircraft, 2013, 50(4): 1189-1201.

[15] 陈钱, 白鹏, 尹维龙, 等. 飞机外翼段大尺度剪切式变后掠设计与分析[J]. 空气动力学学报, 2013, 31(1): 40-46.
CHEN Q, BAI P, YIN W L, et al. Design and analysis of a variable-sweep morphing aircraft with outboard wing section large-scale shearing[J]. Acta Aerodynamica Sinica, 2013, 31(1): 40-46(in Chi-

nese).

[16] 王云飞, 肖洪, 杨广, 等. 平行连杆式变形翼结构设计及分布式驱动配置[J]. 哈尔滨工业大学学报, 2022, 54(1): 65-72.
WANG Y F, XIAO H, YANG G, et al. Structure design and distributed actuators configuration of a parallel linkage morphing wing[J]. Journal of Harbin Institute of Technology, 2022, 54(1): 65-72(in Chinese).

[17] 肖洪, 郭宏伟, 张蒂, 等. 一种基于四面体单元的变形翼骨架设计与分析[J]. 航空学报, 2022, 43(7): 425391.
XIAO H, GUO H W, ZHANG D, et al. Design and analysis of morphing wing skeleton based on tetrahedral element[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(7): 425391(in Chinese).

[18] JACOB J, SMITH S. Design limitations of deployable wings for small low altitude UAVs: AIAA-2009-745[R]. Reston: AIAA, 2009.

[19] PERKINS D, REED J, HAVENS E. Morphing wing structures for loitering air vehicles: AIAA-2004-1888[R]. Reston: AIAA, 2004.

[20] KOTA S. Twistable wings take flight[J]. IEEE Spectrum, 2016, 53: 28-33.

[21] SONG X K, GUO H W, LIU S J, et al. Cable-truss hybrid double-layer deployable mechanical network constructed of Bennett linkages and planar symmetric four-bar linkages[J]. Mechanism and Machine Theory, 2019, 133: 459-480.

[22] 黄真, 赵永生, 赵铁石. 高等空间机构学[M]. 北京: 高等教育出版社, 2006: 121-137.
HUANG Z, ZHAO Y S, ZHAO T S. Advanced spatial mechanism [M]. Beijing: Higher Education Press, 2006: 121-137(in Chinese).

Design and analysis of morphing wing mechanism based on equilateral Bennett mechanism

TIAN Dake[1], ZHANG Junwei[1], JIN Lu[2,*], LIU Rongqiang[3], LEI Hongqiang[1], CUI Xihe[1]

(1. School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang 110168, China;
2. School of Civil Engineering, Shenyang Jianzhu University, Shenyang 110168, China;
3. National Key Laboratory of Aerospace Mechanism, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Morphing aircraft is a new concept vehicle that changes its aerodynamic shape by a morphing mechanism to adapt to various flight environments and mission requirements, which has important application value in aerospace, military reconnoitre, and other fields and is the frontier and hot spot of future vehicle research. In response to the new development needs of wide airspace and wide speed domain of the morphing aircraft, a spanwise bending morphing wing mechanism with a multi-closed-loop space mechanism based on the equilateral Bennett mechanism was proposed. Firstly, the geometric characteristics of the equilateral Bennett mechanism were studied, and the design scheme of the multi-closed-loop morphing wing mechanism based on the equilateral Bennett mechanism was proposed. Secondly, the constrained spiral solution method was used to solve the degrees of freedom of the morphing wing mechanism, and the kinematic model was established based on the D-H coordinate transformation method. Then, the 3D model and virtual prototype of the mechanism were established, and the kinematic simulation was carried out to verify the above model. The prototype was made, and the experiment was carried out. The research results show that the proposed morphing wing mechanism needs only one power source to drive the mechanism movement, with a simple structure and high modularity rate, and the proposed morphing wing mechanism can realize the accurate connection between the parts and the expected spanwise bending action. The research results provide a reference for the basic research and engineering application of morphing wings for new types of morphing aircraft.

Keywords: morphing aircraft; morphing wing; modular structure; space over-constrained mechanism; kinematic analysis

Received: 2023-03-23; **Accepted**: 2023-04-21; **Published Online**: 2023-05-12 11:39

URL: link.cnki.net/urlid/11.2625.V.20230511.1636.004

Foundation items: Joint Funds of the National Natural Science Foundation of China (U2341237); Open Project of National Key Laboratory of Aerospace Mechanism; Foundation of Educational Department of Liaoning Province (LJ222410153096,JYTMS20231592)

* **Corresponding author**. E-mail: jinlu@sjzu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0165

含局部缺陷的角接触球轴承时变位移激励及动力学建模

雷春丽[1]，宋瑞哲[1,*]，樊高峰[2]，刘凯[1]，薛伟[1]，李建华[1]

(1. 兰州理工大学 机电工程学院，兰州 730050；2. 云南文山铝业有限公司，文山 663000)

摘 要：角接触球轴承在长时间工作的情况下，会产生故障损伤从而影响系统正常运行。以外圈具有局部缺陷的角接触球轴承为研究对象，提出不同局部缺陷轮廓的判别方法，建立角接触球轴承局部缺陷时变位移激励广义表征模型，研究局部缺陷演化过程及其位移激励机理。在此基础上，考虑轴承缺陷引起的时变位移对动力学特性的影响，基于赫兹接触理论，建立角接触球轴承故障动力学模型，并通过实验验证了所建模型的正确性。分析结果表明：矩形局部缺陷最终会演化成梯形形态；不同缺陷形貌诱发的位移激励变化趋势是不同的；与局部缺陷的长度相比，宽度对位移激励的影响更大。研究结果对轴承优化设计和故障诊断提供了理论基础。

关 键 词：角接触球轴承；局部缺陷；时变位移激励；缺陷演变；动力学建模

中图分类号：TH133.3

文献标志码：A **文章编号**：1001-5965(2025)03-0753-10

滚动轴承是旋转机械中必不可少的零部件，其运行状况直接影响着整机的性能[1]，同时，其动力学性能对旋转机械的动态特性有着不可或缺的作用[2]。角接触球轴承是一种主要用于载荷较轻的高速旋转场合的滚动轴承。角接触球轴承在运转情况下会产生裂纹、凹坑、胶合等局部缺陷，会影响轴承内部部件的运转；存在局部缺陷使得轴承各部件间的接触与位移关系变得更为复杂，当缺陷处与轴承其他元件接触时，会产生冲击，引起系统谐振，导致振动增大。因此，分析角接触球轴承在产生故障损伤情况下所产生的时变位移激励机理，并对其的振动响应进行分析具有重要的现实意义。

对于具有局部缺陷的滚动轴承，国内外学者做了大量深入的研究。曹宏瑞等[3]将额外间隙引入轴承局部损伤故障中，计算了滚子和滚道之间的趋近量。Patel 和 Upadhyay[4]建立了一个由内圈和外圈各有一个缺陷构成的复合缺陷圆柱滚子轴承-转子系统的9自由度模型，采用 Newton-Rophson 方法求解非线性接触力。徐东等[5]分析了滚动轴承运转过程中承压滚子数的变化规律，用分段函数描述轴承滚子的位置及其是否处于承载区，使用缺陷冲击函数来描述轴承运转过程。并建立了3种单表面故障的滚动轴承非线性动力学方程。吕延军等[6]改进了求解轴承系统非线性动力学响应的方法，运用该方法对转子-轴承系统的动力学响应进行了求解。Cheng 等[7]在轴承拟静态模型的基础上，引入了局部缺陷深度和周向变化范围的局部轮廓函数，建立了考虑局部缺陷的轴承拟静态分析模型。马辉等[8]将轴承滚道上的剥落定义为圆周矩形，针对内外圈故障，研究了滚子经过故障区域边缘单元的动态参数，对滚子滚过内外圈局部缺陷的过程进行了分析。常斌全等[9]将滚动体与缺陷之间的接触情况划分为进入事件、冲击事件、离开事件、载荷补偿事件，并假设缺陷边缘形貌为圆弧状，考虑缺

收稿日期：2023-04-06；**录用日期**：2023-05-29；**网络出版时间**：2023-06-07 13:11
网络出版地址：link.cnki.net/urlid/11.2625.V.20230606.1354.001
基金项目：国家自然科学基金 (51465035)；甘肃省自然科学基金 (20JR5RA466)
*通信作者. E-mail：610306728@qq.com

引用格式：雷春丽，宋瑞哲，樊高峰，等. 含局部缺陷的角接触球轴承时变位移激励及动力学建模 [J]. 北京航空航天大学学报，2025，51（3）：753-762. LEI C L，SONG R Z，FAN G F，et al. Time-varying displacement excitation and dynamic modeling of local defects in angular contact ball bearings [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（3）：753-762（in Chinese）.

陷和轴承变刚度振动引起的时变位移对动力学响应特性的影响，研究了不同缺陷尺寸和不同转速下轴承系统振动加速度幅值的变化规律，将滚动轴承中的滚动体简化为一个并联的考虑质量的弹簧-阻尼系统，建立内圈表面局部缺陷的滚动轴承两自由度动力学模型。李志农等[10]把滚动体的局部点蚀故障简化成矩形缺口，考虑滚道变形量及变形量渐变的规律，提出一个改进的圆弧形缺口模型，建立了一个滚动体故障动力学模型，在此基础上仿真出了滚动体局部单点故障时振动的速度和加速度脉冲信号。Patel 等[11]考虑轴承内圈和外圈表面有单个和多个缺陷的情况，建立了深沟球轴承动力学模型，采用龙格-库塔法得到控制方程的耦合解。Jafari 等[12]对外圈存在缺陷的角接触球轴承进行了仿真和实验分析，并对无外圈缺陷和有外圈缺陷的轴承进行建模和求解，同时，模拟了不同转速和轴向预载条件下的振动响应。王凯等[13]建立了复合故障深沟球轴承四自由度动力学模型，研究复合故障对振动响应的影响。周泽文等[14]针对内外圈各有一处剥落缺陷的情况，将剥落缺陷简化为矩形，考虑碰撞缺陷边缘产生的冲击力造成的时变接触力激励和滚动体通过矩形缺陷时的时变位移激励，仿真出不同负载、不同缺陷尺寸时的振动响应特性。Astekar 等[15]将局部缺陷看作是一个半圆形，分析了接触区域的压力分布情况。Cui 等[16]基于赫兹接触理论和轴承运动学、动力学的相关知识，对类似于三角形、梯形的缺陷进行了简要分析，并对内外圈和滚动体分别发生局部故障时的情况进行了仿真信号模拟和分析。刘静[17]考虑轴承滚子的凸度，计算轴承滚道间的接触刚度，并考虑了不同表面形貌特征下圆柱滚子轴承局部故障的时变位移激励模型，并用滚子与缺陷之间的接触点来划分不同的局部缺陷形状，分析了轴承内外圈不同波纹度对轴承振动响应特征的影响。丁为民等[18]探究了缺陷形貌与轴承内部载荷分布的关系，并对修正的缺陷深度进行了分析。王震等[19]提出了不规则轴承故障的表征方法，将其引入单转子-轴承系统动力学模型，建立了轴承内外圈非规则故障模型，并对含故障轴承转子系统的振动响应进行了分析。

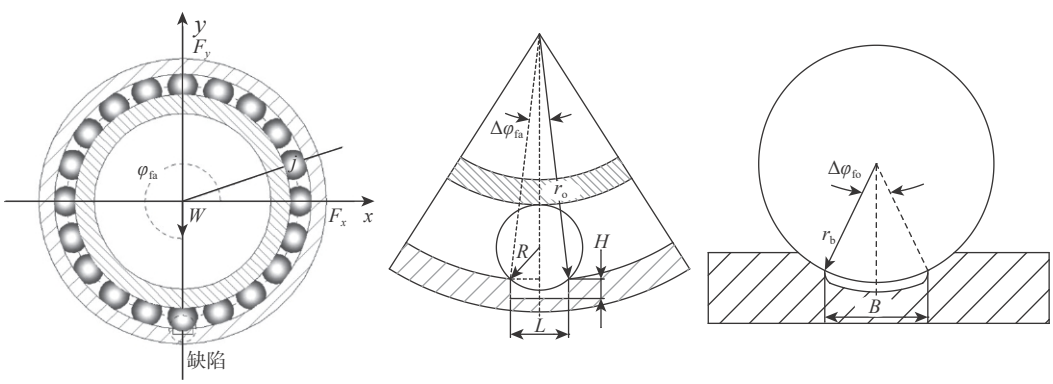
以上研究大多数将局部缺陷简化为矩形形状，而矩形难以准确描述不同类型的局部缺陷表面轮廓的复杂特征，同时以上研究鲜有对不同轮廓的缺陷和演化过程进行深入研究。本文以外圈含有局部缺陷的角接触球轴承为研究对象，建立角接触球轴承不同局部缺陷时变位移激励模型，研究局部缺陷演化过程，并探讨不同局部缺陷及其演化过程中的时变位移激励规律。在此基础上，建立具有外圈缺陷角接触球轴承动力学模型，研究了球轴承存在局部故障时的振动特性。

1 局部缺陷时变位移激励模型

1.1 轴承外圈局部缺陷轮廓描述

采用长度 L 和宽度 B 来描述其形貌，如图 1 所示。缺陷中心所在的方位角为 φ_{fa}，$\Delta\varphi_{\mathrm{fa}}$ 为环向跨度角，$\Delta\varphi_{\mathrm{fo}}$ 为轴向跨度角，R 为滚动体半径，r_{o} 为外滚道沟底环向半径，r_{b} 为外沟道曲率半径，W 为外加载荷，H 为缺陷深度。

当滚动体经过局部缺陷时，其不会瞬间陷入和滚出该缺陷内部，而是一个缓慢变化的过程。因此，滚动体进入故障区域时其深度变化是一个时变的位移函数，基于此，本文将提出不同缺陷轮廓下的时变位移激励函数。



(a) 整体模型　(b) 缺陷环向尺寸　(c) 缺陷轴向尺寸

图 1　轴承局部缺陷轮廓描述

Fig. 1　Description of local defect contour in bearing

1.2 局部缺陷时变位移激励广义表征模型

本文以环向缺陷轮廓为参考，提出了不同局部缺陷形貌的判别方法，即采用角度 α、β 作为 3 种缺陷的区分条件。图 2 为局部缺陷轮廓类别划分简图。当 α=90°、β=90°时为矩形缺陷；当 α>90°、β<90°时为三角形缺陷；当 α>90°、β>90°时为梯形缺陷。

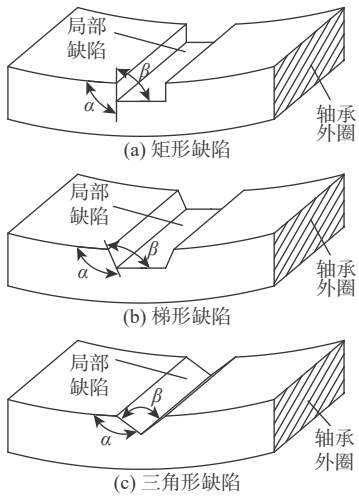


图 2　局部缺陷轮廓类别划分简图

Fig. 2　Diagram of classification of local defect contour categories

则局部缺陷时变位移激励广义表达式为

$$H'=\begin{cases}H_1 & \alpha=90°,\beta=90°\\ H_2 & \alpha>90°,\beta<90°\\ H_3 & \alpha>90°,\beta>90°\end{cases} \tag{1}$$

式中：H_1 为矩形缺陷时变位移激励函数；H_2 为三角形缺陷时变位移激励函数；H_3 为梯形缺陷时变位移激励函数。

1.2.1　矩形缺陷时变位移激励模型

将缺陷定义为矩形，缺陷深度为 H，采用长度 L 和宽度 B 来描述其轮廓。由于缺陷尺寸比较微小，为便于研究，本文将缺陷所在外圈处简化成长方体，其外圈局部缺陷轮廓表征如图 3 所示。

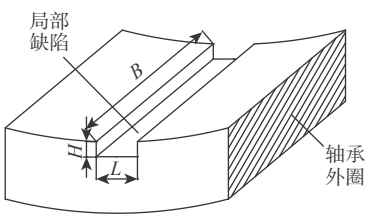


图 3　轴承外圈矩形缺陷轮廓表征

Fig. 3　Characterization of rectangular defect contour in outer ring of bearing

长度 L 和宽度 B 分别与环向、轴向跨度角的关系式为

$$L=2r_{\mathrm{o}}\sin\Delta\varphi_{\mathrm{fa}} \tag{2}$$

$$B=2r_{\mathrm{b}}\sin\Delta\varphi_{\mathrm{fo}} \tag{3}$$

当保持架位于初始位置时，第 j 个滚动体的方位角为

$$\varphi_j=\omega_{\mathrm{c}}t+\frac{2\pi(j-1)}{N}\qquad j=1\sim Z \tag{4}$$

式中：ω_{c} 为保持架转速；Z 为滚动体个数。

图 1(b) 中滚动体与缺陷接触深度可表示为[18]

$$d_1(\varphi)=\begin{cases}\min\left\{H,R-\sqrt{[R^2-0.25r_{\mathrm{o}}^2(\varphi_j-\varphi_{\mathrm{fa}}+0.5\Delta\varphi_{\mathrm{fa}})]}\right\}\\ \qquad\varphi_{\mathrm{fa}}-\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}<\varphi_j<\varphi_{\mathrm{fa}}\\ \min\left\{H,R-\sqrt{[R^2-0.25r_{\mathrm{o}}^2(\varphi_{\mathrm{fa}}-\varphi_j+0.5\Delta\varphi_{\mathrm{fa}})]}\right\}\\ \qquad\varphi_{\mathrm{fa}}<\varphi_j<\varphi_{\mathrm{fa}}+\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}\end{cases} \tag{5}$$

图 1(c) 中滚动体与缺陷接触深度可表示为

$$d_2(\varphi)=r_{\mathrm{b}}+\sqrt{{r_{\mathrm{b}}}^2-(r_{\mathrm{b}}\sin(0.5\Delta\varphi_{\mathrm{fo}}))^2}-\sqrt{R^2-(r_{\mathrm{b}}\sin(0.5\Delta\varphi_{\mathrm{fo}}))^2} \tag{6}$$

由式(5)和式(6)可推导出矩形缺陷的时变位移激励表达式为

$$H_1=\begin{cases}\min(H,d_1(\varphi),d_2(\varphi))\\ \qquad\varphi_{\mathrm{fa}}-\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}<\varphi_j<\varphi_{\mathrm{fa}}+\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}\\ 0\qquad\text{其他}\end{cases} \tag{7}$$

1.2.2　三角形缺陷时变位移激励模型

将缺陷定义为三角形，缺陷深度为 H，采用长度 L 和宽度 B 来描述其轮廓，其外圈局部缺陷轮廓表征如图 4 所示。

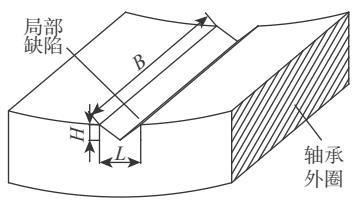


图 4　轴承外圈三角形缺陷轮廓表征

Fig. 4　Characterization of triangular defect contour in outer ring of bearing

根据轴承几何参数，本文定义变量 $M=\dfrac{0.02d_{\mathrm{b}}}{B}$，$N=\dfrac{L}{B}$，$\Delta\varphi_{\mathrm{fa}}$ 为环向跨度角，d_{b} 滚动体直径，则三角形缺陷的时变位移激励表达式为

$$H_2=\{R-[R^2-(0.5MN)^2]^{0.5}\}\arcsin\left(\frac{\pi}{\Delta\varphi_{\mathrm{fa}}}\varphi\right)\quad\varphi\in[0,\Delta\varphi_{\mathrm{fa}}] \tag{8}$$

1.2.3　梯形缺陷时变位移激励模型

将缺陷定义为梯形，缺陷深度为 H，采用长度 L 和宽度 B 来描述其轮廓，其外圈局部缺陷轮廓表征如图 5 所示。

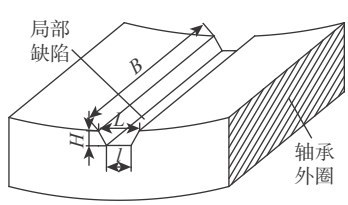


图 5　轴承外圈梯形缺陷轮廓表征

Fig. 5　Characterization of trapezoidal defect contour in outer ring of bearing

与缺陷轮廓缺陷为三角形相似，梯形缺陷定义变量 $M=\dfrac{0.02d_{\mathrm{b}}}{B}$，$N=\dfrac{L}{B}$，$\Delta\varphi_{\mathrm{fa}}$ 为环向跨度角，则梯形缺陷的时变位移激励表达式为

$$H_3=\begin{cases}\left\{R-\left[R^2-(0.5MN)^2\right]^{0.5}\right\}\sin\left(\dfrac{0.5\pi}{\varphi_1}\right) & 0\leqslant\varphi<\varphi_1\\ R-\left[R^2-(0.5MN)^2\right]^{0.5} & \varphi_1\leqslant\varphi<\varphi_2\\ \left\{R-\left[R^2-(0.5MN)^2\right]^{0.5}\right\}\sin\left(\dfrac{0.5\pi}{\varphi_1}+\dfrac{2}{3}\pi\right) & \varphi_2\leqslant\varphi<\Delta\varphi_{\mathrm{fa}}\end{cases} \tag{9}$$

式中：φ_1为进入缺陷时，滚动体接触缺陷前边沿至刚接触底部所跨越的角度，$\varphi_1=\Delta\varphi_{\mathrm{fa}}(L-l)/2L$；$\varphi_2$为滚出缺陷时，滚动体刚退出接触底部至与缺陷后边沿接触所跨越的角度，$\varphi_2=\Delta\varphi_{\mathrm{fa}}(L+l)/2L$。

2 局部缺陷轮廓演化过程

2.1 矩形缺陷情况下的演化过程

图 6 为矩形缺陷的演化过程，外圈存在矩形缺陷时，缺陷环向视图如图 6(a) 所示。滚动体在进入和滚出缺陷边缘的过程中会产生冲击力，使缺陷边缘发生磨损，其磨损部分可视为三角形缺陷，该情况可看作是矩形缺陷和三角形缺陷的复合缺陷，称为复合缺陷 1，如图 6(b) 所示。

复合缺陷 1 的时变位移激励表达式为

$$H_1'=\begin{cases}H_2 & \varphi_{\mathrm{fa}}-\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}-\dfrac{1}{2}\Delta\varphi_{\mathrm{fa1}}<\varphi_j<\varphi_{\mathrm{fa}}-\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}\\ H_1 & \varphi_{\mathrm{fa}}-\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}<\varphi_j<\varphi_{\mathrm{fa}}+\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}\\ H_2 & \varphi_{\mathrm{fa}}+\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}<\varphi_j<\varphi_{\mathrm{fa}}+\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}+\dfrac{1}{2}\Delta\varphi_{\mathrm{fa1}}\end{cases} \tag{10}$$

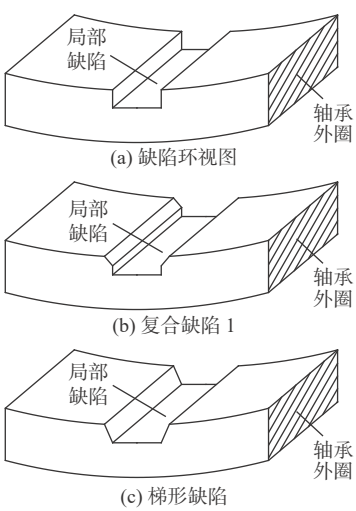


图 6 矩形缺陷的演化过程

Fig. 6 Evolution process of rectangular defect

式中：$\Delta\varphi_{\mathrm{fa1}}$ 为三角形缺陷跨度角。

当外圈存在复合缺陷 1 时，滚动体进入和滚出缺陷边缘时，会进一步使缺陷边缘发生磨损，磨损量达到一定程度后便形成梯形缺陷，如图 6(c) 所示。该情况可看作是梯形缺陷，时变位移激励表达式如式 (10) 所示。

2.2 梯形缺陷情况下的演化过程

图 7 为梯形缺陷的演化过程，外圈存在梯形缺陷时，缺陷环向视图如图 7(a) 所示。滚动体在进入和滚出缺陷边缘的过程中会产生冲击力，使缺陷边缘发生磨损，其磨损部分可视为三角形缺陷，该情况可看作是梯形缺陷和三角形缺陷的复合缺陷，称为复合缺陷 2，如图 7(b) 所示。

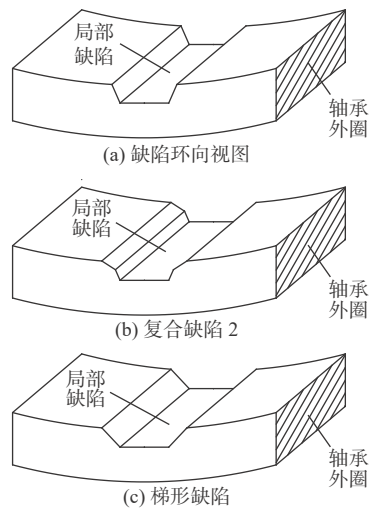


图 7 梯形缺陷的演化过程

Fig. 7 Evolution process of trapezoidal defect

复合缺陷 2 的时变位移激励表达式为

$$H_2'=\begin{cases}H_2 & \varphi_{\mathrm{fa}}-\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}-\dfrac{1}{2}\Delta\varphi_{\mathrm{fa1}}\dfrac{1}{2}\Delta\varphi_{\mathrm{fa2}}<\varphi_j<\varphi_{\mathrm{fa}}-\\ & \dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}-\dfrac{1}{2}\Delta\varphi_{\mathrm{fa1}}\\ H_3 & \varphi_{\mathrm{fa}}-\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}-\dfrac{1}{2}\Delta\varphi_{\mathrm{fa1}}<\varphi_j<\varphi_{\mathrm{fa}}+\\ & \dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}+\dfrac{1}{2}\Delta\varphi_{\mathrm{fa1}}\\ H_2 & \varphi_{\mathrm{fa}}+\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}+\Delta\varphi_{\mathrm{fa1}}<\varphi_j<\varphi_{\mathrm{fa}}+\dfrac{1}{2}\Delta\varphi_{\mathrm{fa}}+\\ & \dfrac{1}{2}\Delta\varphi_{\mathrm{fa1}}+\dfrac{1}{2}\Delta\varphi_{\mathrm{fa2}}\end{cases} \tag{11}$$

式中：$\Delta\varphi_{\mathrm{fa2}}$ 为梯形缺陷跨度角。

当外圈存在复合缺陷 2 时，滚动体进入和滚出缺陷边缘时，会使边缘发生磨损，磨损量达到一定程度后便形成梯形缺陷，如图 7(c) 所示。该情况可看作是梯形缺陷，时变位移激励表达式如式 (11) 所示。

3　具有外圈缺陷角接触球轴承动力学建模与模型验证

3.1　具有外圈缺陷角接触球轴承动力学模型

考虑到具有外圈局部缺陷时轴承接触关系复杂等问题, 本节将轴承系统模型进行简化, 轴承简化模型如图 8 所示[18]。其中, I 为外圈缺陷, k_1 和 k_2 分别为在水平和竖直方向的刚度系数, c_1 和 c_2 分别为在水平和竖直方向的阻尼系数。

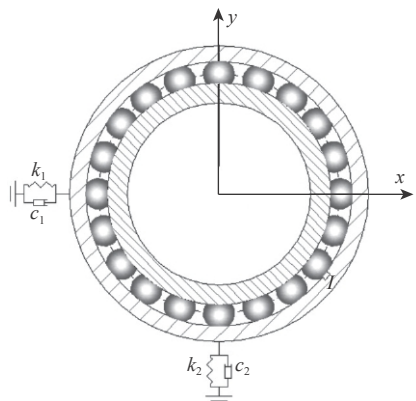


图 8　角接触球轴承外圈故障简化模型[18]

Fig. 8　Simplified model for fault in outer ring of angular contact ball bearing[18]

对于正常的滚道与滚动体点接触而言, 其载荷位移关系式为

$$Q_{\mathrm{c}}=K_{\mathrm{p}}\delta^{\frac{3}{2}} \tag{12}$$

式中: Q_{c} 为接触载荷; K_{p} 为载荷变形系数; δ 为弹性变形量。

缺陷区域的接触载荷等于滚动体与无损滚道的接触载荷减去缺陷部分所卸载的载荷。因此, 缺陷区域处的接触载荷可表示为[16]

$$Q_{\mathrm{d}}=Q_{\mathrm{c}}-\iint_{D}\frac{3Q_{\mathrm{c}}}{2\pi ab}\left[1-\left(\frac{x}{a}\right)^{2}-\left(\frac{y}{b}\right)^{2}\right]^{\frac{1}{2}}\mathrm{d}x\mathrm{d}y \tag{13}$$

式中: a、b 分别为正常载荷在无缺陷滚道上所引起的正常接触椭圆的半长、半宽。式中的二重积分区域为缺陷部分所承受的载荷, 积分区域 D 为

$$\left\{(x,y)\left|x\in\left[-\frac{1}{2}r_{\mathrm{b}}\Delta\varphi_{\mathrm{fo}},\frac{1}{2}r_{\mathrm{b}}\Delta\varphi_{\mathrm{fo}}\right],\right.\right.$$
$$\left.\left.y\in\left[r_o\left(-\varphi_j+\varphi_{\mathrm{fa}}-\frac{1}{2}\Delta\varphi_{\mathrm{fa}}\right),r_o\left(-\varphi_j+\varphi_{\mathrm{fa}}+\frac{1}{2}\Delta\varphi_{\mathrm{fa}}\right)\right]\right\} \tag{14}$$

通过式(12)～式(14)推导可得出考虑缺陷的接触刚度为

$$K_{\mathrm{f}}(\varphi_j)=\lim_{\delta\to 0}\frac{Q_{\mathrm{d}}}{\delta} \tag{15}$$

轴承运转时滚动体同时与内、外滚道接触, 其接触变形等于滚动体与每一个滚道的接触变形之和, 将二者串联构成等效接触刚度, 其表达式为

$$K_{\mathrm{n}}(\varphi_j)=\left[1\Big/\left(\frac{1}{K_{\mathrm{i}}}\right)^{\frac{1}{n}}+\left(1/K_{\mathrm{f}}(\varphi_j)\right)^{\frac{1}{n}}\right]^{n} \tag{16}$$

式中: K_{i} 为滚动体与内滚道的接触刚度; $K_{\mathrm{f}}(\varphi_j)$ 为滚动体与具有缺陷的外滚道接触刚度, 对于球轴承, n=3/2。

在如图 1(a)所示的含缺陷角接触球轴承整体模型中, 轴承外圈存在单故障时, 第 j 个滚动体与滚道接触产生的总接触变形为

$$\delta_j=(x_1-x_2)\cos\varphi_j+(y_1-y_2)\sin\varphi_j-H'-c \tag{17}$$

式中: x_1 和 y_1 为轴的位置坐标; x_2 和 y_2 为轴承座的位置坐标; c 为滚动轴承径向游隙。则第 j 个滚动体与滚道产生的接触载荷为

$$Q_j=K_{\mathrm{n}}(\varphi_j)\delta_j^{\frac{3}{2}} \tag{18}$$

将接触载荷在 x、y 方向分解, 得到含缺陷滚动轴承载荷-位移非线性赫兹接触力方程组:

$$\begin{bmatrix}F_x\\F_y\end{bmatrix}=\sum_{j=1}^{Z}Q_j\begin{bmatrix}\cos\varphi_j\\\sin\varphi_j\end{bmatrix} \tag{19}$$

式中: F_x 和 F_y 分别为非线性轴承反力在 x 和 y 方向分量。

根据牛顿第二定律和拉格朗日方程, 建立了外圈含单点缺陷的滚动轴承动力学模型, 其四自由度简化动力学方程为

$$\begin{cases}m_1\ddot{x}_1+c_1\dot{x}_1+k_1x_1+F_x=W_x\\m_1\ddot{y}_1+c_2\dot{y}_1+k_2y_1+F_y=W_y\\m_2\ddot{x}_2+c_1\dot{x}_2+k_1x_2-F_x=0\\m_2\ddot{y}_2+c_2\dot{y}_2+k_2y_2-F_y=0\end{cases} \tag{20}$$

式中: W_x 和 W_y 分别为水平方向和竖直方向上滚动轴承系统的外加载荷; m_1 和 m_2 分别为内圈和外圈的质量。

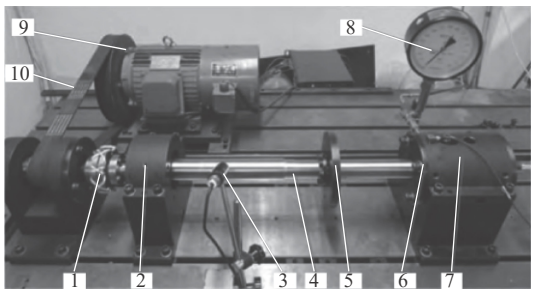
根据式(17)和式(20)轴承内部受力变形及非线性方程组, 利用仿真软件编写计算程序, 采用龙格-库塔法进行求解。

3.2　模型验证

本文采用文献 [20] 的实验结果来验证 3.1 节含外圈故障滚动轴承动力学模型的正确性。图 9 为文献 [20] 中的实验台。

选取文献 [20] 中型号为 HRB7206AC 的角接触球轴承, 设置缺陷尺寸 L 为 0.3 mm, 转速为 1 200 r/min, 预紧力为 600 N。通过计算, 得出缺陷轴承的振动响应, 如图 10 所示。

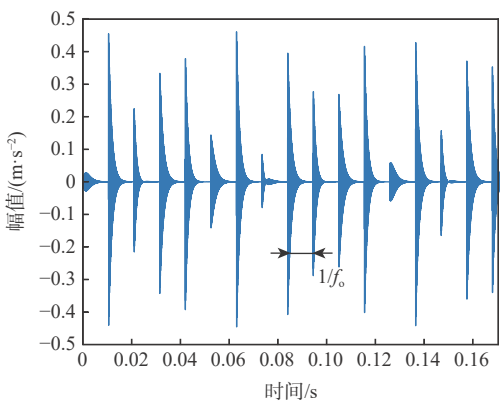
图 10 为外圈存在缺陷时本文动力学模型的仿真结果。本文动力学模型中的时域波形和文献 [20]



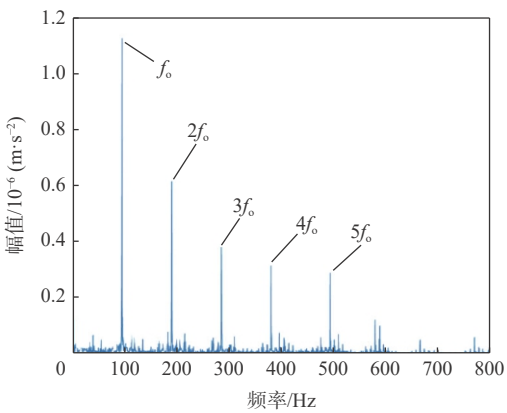
1. 表示柔性联轴器; 2. 表示左侧轴承座; 3. 表示转速传感器;
4. 表示主轴; 5. 表示质量盘; 6. 表示加速度传感器;
7. 表示右侧轴承座; 8. 表示预紧力显示液压表;
9. 表示驱动电机; 10. 表示传动带。

图 9 实验台[20]

Fig. 9 Experiment rig[20]



(a) 时域波形图



(b) 包络谱图

图 10 本文动力学模型外圈损伤振动响应

Fig. 10 Vibration response of fault in outer ring of the proposed dynamics model

中的结果基本一致, 均存在频率为 f_o 的冲击衰减信号; 包络谱图中以频率 f_o 及其倍频为主导成分。文献 [20] 中角接触球轴承的外圈故障特征频率为 f_o=97.48 Hz, 图 10 中, 外圈故障特征频率为 f_o=95.21 Hz, 误差为 2.3%。采用本文动力学模型得到的包络谱中外圈故障特征频率 f_o 及其倍频清晰可见, 与文献 [20] 中实验所得的结果十分吻合。由此充分说明了本文动力学模型的准确性。

4 结果分析

4.1 不同工况下的局部缺陷轮廓时变位移激励分析

本节对具有局部缺陷角接触球轴承的时变位移激励进行研究, 选取 H7013C 型角接触球轴承为研究对象, 其结构和材料参数如表 1 所示。

表 1 H7013C 轴承基本参数

Table 1 Basic parameters of H7013C bearing

内圈直径/mm	外圈直径/mm	滚动体直径/mm	滚动体个数	轴承宽度/mm	初始接触角/(°)	滚动体密度/(g·cm^{-3})
65	100	8.73	25	18	15	3.2

4.1.1 单一局部缺陷情况

1) 矩形缺陷情况下的时变位移激励分析

当轴承外圈存在矩形缺陷时, 滚动体直径与缺陷之间的大小关系将影响到滚动体是否能接触到缺陷底部, 因此, 要考虑缺陷宽度 B、缺陷长度 L 和滚动体直径 d_b 之间的关系。图 11 为滚动体与滚道之间的时变位移, 与文献 [20] 对比可知, 位移激励变化趋势基本一致, 表明了本文的位移激励变化规律是正确的。

由图 11 可知, 滚动体在进出缺陷时是一个渐变的过程。当 B 为定值, 改变缺陷长度 L, 随着 L 的增大, 滚动体进入缺陷的角位置在减小, 退出缺陷的角位置在增大, 陷入缺陷的位移激励逐渐增大。当 L 为定值, 改变缺陷宽度 B, 随着 B 的增大, 滚动体陷入缺陷的位移激励也在不断增大。当 B、L 为定值时, 改变 H, 随着 H 的增大滚动体陷入缺陷的

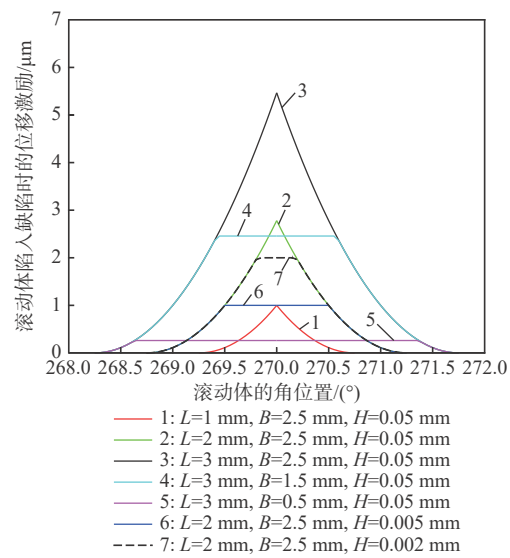


图 11 滚动体陷入矩形缺陷时的位移激励

Fig. 11 Displacement excitation of rolling element trapped in rectangular defect

位移激励也在不断增大。对比分别固定 B、L 值时的位移激励的变化情况, 可知缺陷宽度 B 在很大程度上影响了滚动体与滚道之间的位移激励。

2) 三角形缺陷情况下的时变位移激励分析

当轴承外圈存在三角形缺陷时, 可得滚动体与滚道之间的位移激励, 如图 12 所示。与图 11 进行对比, 位移激励的变化趋势是一致的。

此外, 由图 12 可知, 当 B 为定值, 随着 L 的增大, 滚动体陷入缺陷的位移激励逐渐增大。当 L 为定值, 随着 B 的增大, 滚动体陷入缺陷的位移激励也不断增大。

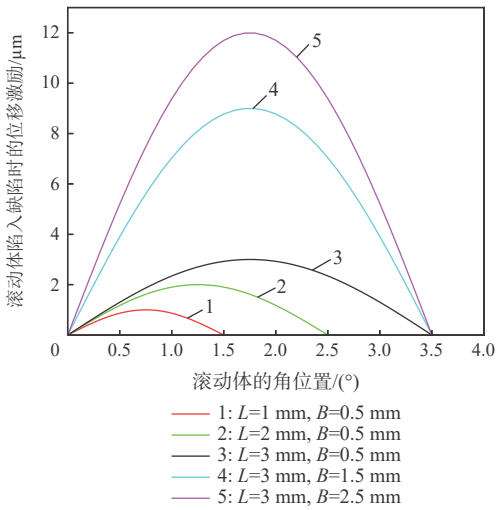


图 12　滚动体陷入三角形缺陷时的位移激励

Fig. 12　Displacement excitation of rolling element trapped in triangular defect

3) 梯形缺陷情况下的时变位移激励分析

当轴承外圈存在梯形缺陷时, 可得滚动体与滚道之间的位移激励, 如图 13 所示。与图 11 进行对比, 曲线趋势变化一致。

此外, 由图 13 可知, 水平线部分为滚动体与缺陷底部接触的部分。固定 B 的值, 随着 L 的增大滚动体陷入缺陷的位移激励逐渐增大。固定 L 的值, 随着 B 的增大滚动体陷入缺陷的位移激励也不断增大。

4.1.2　局部缺陷演化过程中的时变位移激励分析

1)矩形缺陷演化过程中的时变位移激励分析

设定矩形缺陷宽度为一个定值, 可得矩形缺陷演化过程中滚动体与滚道之间的位移激励变化, 如图 14 所示。

由图 14 中矩形缺陷曲线和复合缺陷曲线的变化可知, 矩形缺陷向复合缺陷 1 的演化过程中位移激励的变化规律: 当缺陷从矩形缺陷向复合缺陷 1 演变的过程中滚动体进入缺陷的角位置在减小, 退出缺陷的角位置在增大。同时, 滚动体陷入缺陷的最大位移激励是不变的, 改变的过程是在角位置区

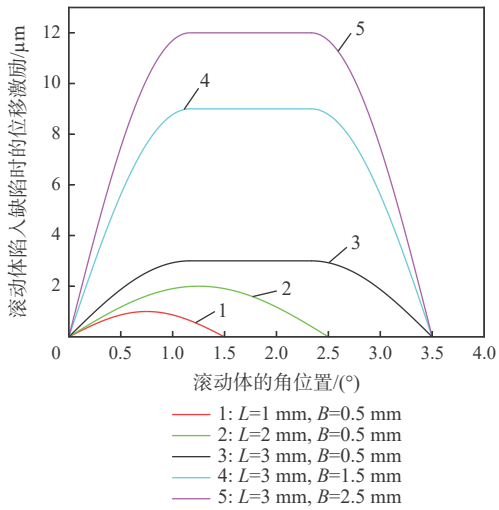


图 13　滚动体陷入梯形缺陷时的位移激励

Fig. 13　Displacement excitation of rolling element trapped in trapezoidal defect

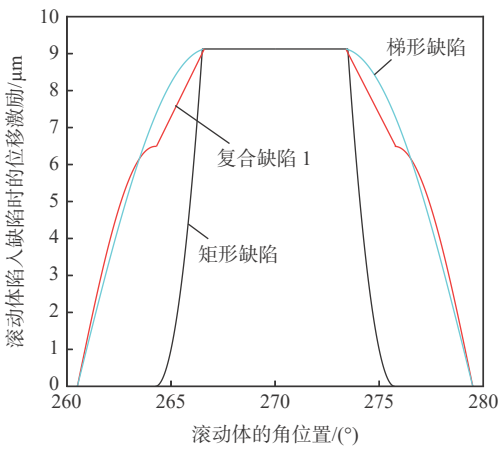


图 14　滚动体陷入矩形局部缺陷时的位移激励

Fig. 14　Displacement excitation of rolling element trapped in rectangular local defect

间 [260.2°, 266.6°] 和 [273.4°, 279.5°], 也就是滚动体落入和滚出缺陷的区间, 出现此情况是因为在缺陷边缘处发生了磨损。

由图 14 中复合缺陷曲线和梯形缺陷曲线的变化可知, 复合缺陷 1 向梯形缺陷的演化过程中位移激励的变化规律。此过程假设边缘未发生磨损, 仅考虑边缘内侧磨损量, 如图 6(b)～图 6(c) 的过程。同时, 可以看出当缺陷从复合缺陷 1 向梯形缺陷演变的过程中滚动体陷入缺陷的最大位移激励是不变的, 改变的过程是在角位置区间 [260.2°, 266.6°] 和 [273.4°, 279.5°], 也就是滚动体与缺陷边缘内侧接触的区间。

2) 梯形缺陷演化过程中的时变位移激励分析

设定梯形缺陷宽度为一个定值, 可得梯形缺陷演化过程中滚动体与滚道之间的位移激励变化, 如图 15 所示。

由图 15 中梯形缺陷曲线和复合缺陷 2 曲线的变化可知梯形缺陷向复合缺陷 2 的演化过程中位移激励的变化规律：当缺陷从梯形缺陷向复合缺陷 2 演变的过程中滚动体陷入缺陷的最大位移激励是不变的，改变的过程是在角位置区间 [256.75°, 266.6°] 和 [273.4°, 283.75°]，也就是滚动体落入和滚出缺陷的区间，出现此情况是因为在缺陷边缘处发生了磨损。

由图 15 中复合缺陷 2 曲线和梯形缺陷曲线的变化可知，复合缺陷 2 向梯形缺陷的演化过程中位移激励的变化规律。此过程假设边缘未发生磨损，仅考虑边缘内侧磨损量，如图 7(b)～图 7(c) 的过程。同时，可以看出当缺陷从复合缺陷 2 向梯形缺陷演变的过程中滚动体陷入缺陷的最大位移激励是不变的，改变的过程是在滚动体与缺陷边缘内侧接触的角位置区间：[256.75°, 266.6°] 和 [273.4°, 283.75°]。

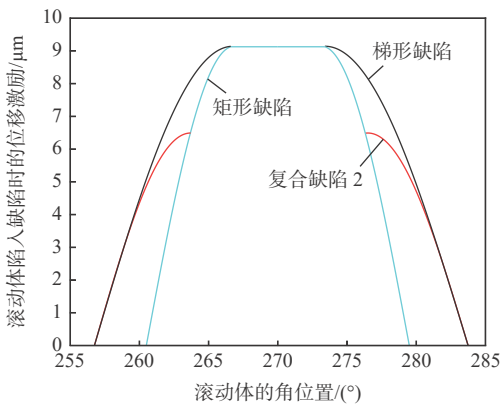


图 15 滚动体陷入梯形局部缺陷时的位移激励

Fig. 15 Displacement excitation of rolling element trapped in trapezoidal local defect

4.2 故障轴承的振动特性分析

本节对具有局部缺陷角接触球轴承的振动响应进行研究，对于外圈局部缺陷形貌为矩形时，设置外圈缺陷尺寸 L 为 0.3 mm，轴的转速为 1 200 r/min，预紧力为 600 N。将表 1 参数代入本文外圈缺陷角接触球轴承动力学模型中，得到如图 16 所示的外圈损伤振动响应图。

由滚动轴承故障特征理论[21] 可知，角接触球轴承外圈故障特征频率 f_o=224.25 Hz，其计算式为

$$f_o = \frac{1}{2}Zf_r\left(1-\frac{d_b}{d_m}\cos\alpha\right) \tag{21}$$

式中：f_r 为轴的转频；d_m 为滚动体节圆直径。

图 16 中，f_o=223.7 Hz，式 (21) 计算值与模型包络谱分析之间误差为 0.24%，再次验证了本文动力学模型的正确性。包络谱中外圈故障频率 f_o 及其倍频清晰可见。同时，结合时域波形图能反映出外滚道存在缺陷时轴承的振动响应，能够有效地仿真

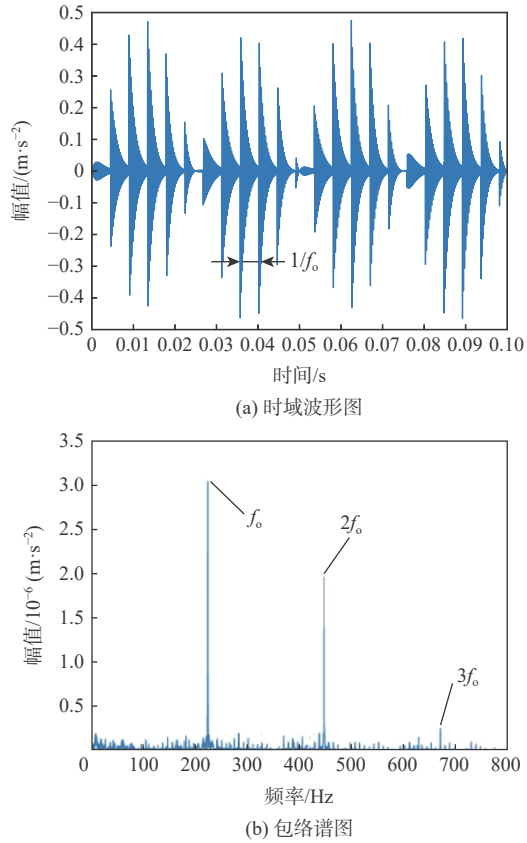


图 16 本文动力学模型外圈局部损伤振动响应

Fig. 16 Vibration response of local fault in outer ring of proposed dynamics model

出滚动体经过局部缺陷时的冲击脉冲性，这样可以有效地捕捉轴承故障在振动信号中显现的信息，避免设备重大故障的发生。三角形和梯形缺陷与矩形缺陷的振动响应计算过程类似。

5 结 论

本文建立了角接触球轴承时变位移激励模型，获得了不同工况下的局部缺陷时变位移激励规律，在此基础上，提出了角接触球轴承局部故障动力学模型，对外圈具有缺陷的轴承进行了振动响应分析，得到如下结论。

1）提出采用角度来划分 3 种局部缺陷轮廓的方法，通过比较，验证了本文方法的正确性。同时，定义了参数 M、N，并用该参数来描述三角形和梯形缺陷的时变位移激励模型，经过对比，表明本文动力学模型是正确的。

2）研究了局部缺陷轮廓分别为矩形、三角形和梯形的时变位移激励规律。由位移激励变化的最大值可以看出，缺陷的宽度与长度相比，缺陷宽度在很大程度上决定了滚动体与滚道之间位移激励的大小。同时，矩形缺陷的磨损变化是一个渐变的过程，其可以按照矩形缺陷到梯形缺陷的过程进

行演变。

3) 分析了具有局部缺陷的角接触球轴承的振动响应特性, 通过与已有文献数据进行对比, 结果吻合度较好, 验证了本文动力学模型的正确性。

后续将进一步揭示转速、载荷、润滑条件、缺陷程度等多种工况因素对缺陷轴承动力学响应的影响规律。

参考文献（References）

[1] 邓子豪, 李录平, 刘瑞, 等. 基于 SCADA 提取的风电机组偏航齿轮箱故障诊断方法研究[J]. 动力工程学报, 2021, 41(1): 43-50.
DENG Z H, LI L P, LIU R, et al. Research on fault diagnosis method of yaw gearbox of wind turbine based on SCADA extraction[J]. Journal of Power Engineering, 2021, 41(1): 43-50(in Chinese).

[2] 张建平, 唐五湘. 我国数控机床行业发展战略研究[J]. 企业经济, 2018, 37(9): 66-72.
ZHANG J P, TANG W X. Research on development strategy of china's CNC machine tool industry[J]. Enterprise Economy, 2018, 37(9): 66-72(in Chinese).

[3] 曹宏瑞, 景新, 苏帅鸣, 等. 中介轴承故障动力学建模与振动特征分析[J]. 机械工程学报, 2020, 56(21): 89-99.
CAO H R, JING X, SU S M, et al. Dynamic modeling and vibration analysis for inter-shaft bearing fault[J]. Journal of Mechanical Engineering, 2020, 56(21): 89-99(in Chinese).

[4] PATEL U K A, UPADHYAY S H. Nonlinear dynamic response of cylindrical roller bearing–rotor system with 9 degree of freedom model having a combined localized defect at inner–outer races of bearing[J]. Tribology Transactions, 2017, 60(2): 284-299.

[5] 徐东, 徐永成, 陈循, 等. 单表面故障的滚动轴承系统非线性动力学研究[J]. 机械工程学报, 2010, 46(21): 61-68.
XU D, XU Y C, CHEN X, et al. Study on nonlinear dynamics of rolling bearing system with single surface fault[J]. Journal of Mechanical Engineering, 2010, 46(21): 61-68(in Chinese).

[6] 吕延军, 曹治军, 黑棣, 等. 具有陀螺效应的转子-轴承系统的非线性动力学行为[J]. 西安理工大学学报, 2010, 26(1): 13-19.
LYU Y J, CAO Z J, HEI D, et al. Nonlinear dynamic behaviors of unsymmetrical rotor-bearing system with gyroscopic effect[J]. Journal of Xi'an University of Technology, 2010, 26(1): 13-19(in Chinese).

[7] CHENG H C, ZHANG Y M, LU W J, et al. Research on time-varying stiffness of bearing based on local defect and varying compliance coupling[J]. Measurement, 2019, 143: 155-179.

[8] 马辉, 李鸿飞, 俞昆, 等. 含局部故障的滚动轴承动力学建模及振动分析[J]. 东北大学学报 (自然科学版), 2020, 41(3): 343-348.
MA H, LI H F, YU K, et al. Dynamic modeling and vibration analysis of rolling bearing with local faults[J]. Journal of Northeast University (Natural Science Edition), 2020, 41(3): 343-348(in Chinese).

[9] 常斌全, 剡昌锋, 苑浩, 等. 多事件激励的滚动轴承动力学建模[J]. 振动与冲击, 2018, 37(17): 16-24.
CHANG B Q, YAN C F, YUAN H, et al. Dynamic modeling for rolling bearings under multi-event excitation[J]. Journal of Vibration and Shock, 2018, 37(17): 16-24(in Chinese).

[10] 李志农, 李云龙, 任帅, 等. 局部点蚀故障滚动体的滚动轴承动力学模型研究[J]. 振动工程学报, 2020, 33(3): 597-603.
LI Z N, LI Y L, REN S, et al. Study on dynamic model of rolling bearing with local pitting failure[J]. Journal of Vibration Engineering, 2020, 33(3): 597-603(in Chinese).

[11] PATEL V N, TANDON N, PANDEY R K. A dynamic model for vibration studies of deep groove ball bearings considering single and multiple defects in races[J]. Journal of Tribology, 2010, 132(4): 041101.

[12] JAFARI S M, ROHANI R, RAHI A. Experimental and numerical study of an angular contact ball bearing vibration response with spall defect on the outer race[J]. Archive of Applied Mechanics, 2020, 90(11): 2487-2511.

[13] 王凯, 剡昌锋, 王风涛, 等. 深沟球轴承复合故障动力学特征[J]. 哈尔滨工业大学学报, 2020, 52(1): 133-140.
WANG K, YAN C F, WANG F T, et al. Dynamic characteristics of compound fault in deep groove ball bearing[J]. Journal of Harbin Institute of Technology, 2020, 52(1): 133-140(in Chinese).

[14] 李志农, 周泽文, 胡莺庆, 等. 考虑冲击力的球轴承复合缺陷动力学建模[J]. 沈阳工业大学学报, 2022, 44(2): 173-179.
LI Z N, ZHOU Z W, HU P Q, et al. Dynamic modeling of composite defects of ball bearing with consideration of impact force[J]. Journal of Shenyang University of Technology, 2022, 44(2): 173-179(in Chinese).

[15] ASTEKAR A, SADEGHI F, STACHE L E. Surface defects effects on bearing dynamics[J]. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2010, 224(1): 25-35.

[16] CUI L L, CHEN X, CHEN S J. Dynamics modeling and analysis of local fault of rolling element bearing[J]. Advances in Mechanical Engineering, 2015, 7(1): 262351.

[17] 刘静. 滚动轴承缺陷非线性激励机理与建模研究[D]. 重庆: 重庆大学, 2014.
LIU J. Research on nonlinear excitation mechanism and modeling of rolling bearing defects[D]. Chongqing: Chongqing University, 2014(in Chinese).

[18] 丁为民, 潘帅航, 张执南. 含缺陷滚动轴承内部载荷分布[J]. 机械设计与研究, 2016, 32(4): 79-84.
DING W M, PAN S H, ZHANG Z N. Load distribution of rolling bearing with localized defect[J]. Machine Design & Research, 2016, 32(4): 79-84(in Chinese).

[19] 王震, 杨正伟, 何浩浩, 等. 非规则轴承故障的动力学建模与仿真[J]. 北京航空航天大学学报, 2021, 47(8): 1580-1593.
WANG Z, YANG Z W, HE H H, et al. Dynamic modeling and simulation of irregular bearing failure[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(8): 1580-1593(in Chinese).

[20] 周智, 朱永生, 闫柯, 等. 考虑缺陷因素的角接触球轴承七自由度动力学模型[J]. 计算机集成制造系统, 2014, 20(11): 2836-2842.
ZHOU Z, ZHU Y S, YAN K, et al. Seven degrees of freedom dynamic model of angular contact ball bearing with defects[J]. Computer Integrated Manufacturing Systems, 2014, 20(11): 2836-2842(in Chinese).

[21] PETERSEN D, HOWARD C, PRIME Z. Varying stiffness and load distributions in defective ball bearings: analytical formulation and application to defect size estimation[J]. Journal of Sound and Vibration, 2015, 337: 284-300.

Time-varying displacement excitation and dynamic modeling of local defects in angular contact ball bearings

LEI Chunli[1], SONG Ruizhe[1,*], FAN Gaofeng[2], LIU Kai[1], XUE Wei[1], LI Jianhua[1]

(1. School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, China;
2. Yunnan Wenshan Aluminum Co., Ltd., Wenshan 663000, China)

Abstract: Angular contact ball bearings will suffer from fault damage after working for a long time, thus affecting the normal operation of the system. In this paper, angular contact ball bearingswith local defects in the outer ring were taken as the research objects, and the discrimination methods of different local defect contours were proposed. A generalized representation model of time-varying displacement excitation of local defects in angular contact ball bearing was established, and the evolution process of local defects and its displacement excitation mechanism were studied. On this basis, the influence of time-varying displacement caused by bearing defects on dynamic characteristics was explored, and a fault dynamic model of angular contact ball bearings was established based on Hertz contact theory. The correctness of the model was verified by experiments. The results show that the rectangular local defects will eventually evolve into trapezoidal local defects. The changing trend of displacement excitation induced by different defect morphologies is different. Compared with the length of the local defect, the width has a greater influence on the displacement excitation. The research results provide a theoretical basis for bearing optimization design and fault diagnosis.

Keywords: angular contact ball bearings; local defect; time-varying displacement excitation; defect evolution; dynamic modeling

Received: 2023-04-06; **Accepted**: 2023-05-29; **Published Online**: 2023-06-07 13:11

URL: link.cnki.net/urlid/11.2625.V.20230606.1354.001

Foundation items: National Natural Science Foundation of China (51465035); Natural Science Foundation of Gansu Province (20JR5RA466)

* **Corresponding author**. E-mail: 610306728@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0154

基于改进YOLOv5s的动态视觉SLAM算法

蒋畅江*，刘朋，舒鹏
(重庆邮电大学 自动化学院，重庆 400065)

摘 要：针对室内动态场景中存在的动态目标会降低同步定位与地图构建(SLAM)系统的鲁棒性和相机定位精度问题，提出了一种基于目标检测网络的动态视觉SLAM算法。选择YOLOv5系列中深度和特征图宽度最小的YOLOv5s作为目标检测网络，并将其主干网络替换为PP-LCNet轻量级网络，在VOC2007+VOC2012数据集训练后，由实验结果可知，PP-LCNet-YOLOv5s模型较YOLOv5s模型网络参数量减少了41.89%，运行速度加快了39.13%。在视觉SLAM系统的跟踪线程中引入由改进的目标检测网络和稀疏光流法结合的并行线程，用于剔除动态特征点，仅利用静态特征点进行特征匹配和相机位姿估计。实验结果表明，所提算法在动态场景下的相机定位精度较ORB-SLAM3提升了92.38%。

关 键 词：同步定位与地图构建；目标检测；动态特征点剔除；定位精度；光流法
中图分类号：V221+.3; TB553
文献标志码：A **文章编号**：1001-5965(2025)03-0763-09

同步定位与地图构建(simultaneous localization and mapping, SLAM)指移动机器人在未知环境中行进时，实时构建环境地图和确定自身在地图中的位置。SLAM根据传感器不同主要分为2类：①搭载激光雷达的SLAM，该系统已趋于成熟，计算量小、测距准确，但搭载的激光雷达成本较高，且后期不易维护[1]，不常用于室内机器人；②搭载相机的视觉SLAM，具有成本低廉、性价比高的特点，并能获取丰富的环境信息，因此，成为了机器人研究领域备受关注的热点。

目前，视觉SLAM系统的研究发展迅速，根据视觉里程计使用的算法不同，可将视觉SLAM分为2类：基于特征匹配的视觉SLAM和基于直接法的视觉SLAM。Campos等[2]在ORB-SLAM2[3]的基础上于2020年发布的开源算法ORB-SLAM3是特征点法的代表，新增了惯性测量单元(inertial measurement unit, IMU)融合算法及对多端地图Atlas的维护，是第一个能够在所有算法阶段重用所有先前信息的系统；Gomez-Ojeda等[4]提出了PL-SLAM，利用关键点和线段的组合进行特征匹配，能在低纹理场景中稳定运行且定位精度良好。关于直接法，张有全等[5]提出了一种基于直接法和共视图优化的紧耦合视觉惯性SLAM系统，在跟踪线程利用IMU信息和基于稀疏图像对齐的直接法进行初始位姿估计，提高了系统的定位精度和实时性；Engel等[6]提出的大规模直接单目SLAM(large-scale direct monocular SLAM, LSD-SLAM)算法利用图像帧所有高对比度像素的光度一致性对图像帧进行对齐，估计像素的深度，建立三维点云地图。传统的视觉SLAM算法在静态场景中具有定位精度高、鲁棒性好的特点，但在动态场景下，由于运动物体的影响，难以取得较好的定位精度。因此，这些算法很难适用于复杂多变的动态场景。

针对动态场景，目前的动态视觉SLAM算法主要通过剔除动态特征点来提高系统的鲁棒性和相机位姿估计精度。动态目标的处理方法包括基于

收稿日期：2023-03-31；录用日期：2023-06-09；网络出版时间：2023-07-26 08:39
网络出版地址：link.cnki.net/urlid/11.2625.V.20230724.1959.003
基金项目：国家自然科学基金(62277008)
*通信作者. E-mail：jiangcj@cqupt.edu.cn

引用格式：蒋畅江，刘朋，舒鹏. 基于改进YOLOv5s的动态视觉SLAM算法[J]. 北京航空航天大学学报，2025，51(3)：763-771.
JIANG C J，LIU P，SHU P. Dynamic visual SLAM algorithm based on improved YOLOv5s[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：763-771 (in Chinese).

多视图几何的方法和基于深度学习的方法。在依赖多视图几何方法中，Engel等[7]提出的直接稀疏里程计(direct sparse odometry, DSO)算法，利用深度神经网络和几何约束实现了相机的位姿估计；Li等[8]结合轻量级神经网络和多视图几何来检测未知运动物体，提高系统的运行速度和鲁棒性；Long等[9]提出的基于金字塔场景解析网络的语义SLAM系统(pyramid scene parsing network-based SLAM, PSPNet-SLAM)，基于ORB-SLAM2系统集成了金字塔结构的语义线程和反向蚁群搜索策略的几何线程，提高了系统在动态场景下的定位精度，并增强了系统的实时性。

在依赖深度学习方法中，Bescos等[10]提出的DynaSLAM算法通过深度学习神经网络掩码区域卷积神经网络(mask region-based convolutional neural network, Mask R-CNN)[11]分割具有先验信息的动态目标，利用多视图几何分割潜在的动态特征点，仅利用静态特征点进行跟踪和建图；Han等[12]利用语义分割识别出图像的先验动态区域，利用静态特征点进行轨迹跟踪和建图；Yuan等[13]提出了基于ORB-SLAM2的基于语义与深度信息的视觉SLAM系统(visual SLAM based on semantic and depth information, SaD-SLAM)，利用语义信息和深度信息检测从运动目标中提取静态特征点，用于相机位姿估计；Yu等[14]提出的基于语义信息的动态视觉SLAM(semantic information based dynamic visual SLAM, DS-SLAM)算法，结合了SegNet[15]实时语义分割网络与移动一致性检测，降低了动态目标对系统的影响，在动态场景中，定位精度相比于ORB-SLAM2提高了一个量级，但语义分割比较耗时，无法满足实时性要求；Liu等[16]提出了基于语义分割的实时动态SLAM(real-time dynamic SLAM using semantic segmentation, RDS-SLAM)算法，基于ORB-SLAM3系统添加了语义跟踪线程和优化线程，使用数据关联算法剔除跟踪线程的异常值；Su等[17]提出的实时动态SLAM(real-time dynamic SLAM, RTD-SLAM)在系统的跟踪线程添加基于YOLOv5s的并行语义线程及光流模块，用于剔除动态特征点，提高了系统的定位精度及实时性；文献[18]基于ORB-SLAM3系统添加了语义分割线程，提高了动态场景下相机的定位精度，但会将潜在动态目标(如人拿着的书)的动态特征点检测为静态特征点，用于跟踪线程的特征匹配和相机位姿计算，导致系统定位精度下降。

综上，现有动态视觉SLAM系统虽然在一定程度上解决了动态目标对系统的影响，但仍存在定位精度低、实时性差的问题。基于此，本文选取YOLOv5s作为目标检测网络，将其主干网络替换为PP-LCNet网络，结合光流法，用于SLAM系统跟踪线程剔除不满足要求的特征点，仅利用处理后的特征点进行相机的位姿估计。在TUM数据集进行实验，并与ORB-SLAM3及其他动态SLAM算法进行对比，本文改进算法综合性能更优。

1 系统框架

1.1 改进YOLOv5s目标检测网络

YOLOv5是一种one-stage目标检测算法，用于对象识别和定位，由输入端、主干网络、颈部网络、头部网络4部分组成。YOLOv5相较之前的版本，每部分都作了性能优化。输入端使用Mosaic数据增强操作，能够对图像进行随机缩放、裁剪和排布操作；主干网络中采用Focus结构和CSP结构，减少了计算量；颈部网络部分采用CSP2，加强了网络特征融合能力；在头部网络采用CIOU_LOSS作为bounding box的损失函数。YOLOv5共有4个网络模型，分别为YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x，其中，YOLOv5s是深度最小、特征图宽度最小的网络模型，该模型运行速度最快、参数最少，但能够准确识别出图像帧中具有先验信息的动态目标和潜在动态目标，能够满足SLAM系统对目标检测网络的精度要求，因此，本文选择YOLOv5s网络作为目标检测网络。

为使YOLOv5s网络更加轻量化，选取PP-LCNet网络替换YOLOv5s的主干网络，改进后的结构示意图如图1所示。图中：Stem conv表示主干卷积层，h-swish表示硬Swish激活函数。PP-LCNet是一种轻量级CPU网络，可有效降低室内机器人对硬件设备的要求。为提取更明显的特征，在卷积核为5×5的卷积层添加SE模块，增强显著特征并抑制不重要特征，将SE模块中的Sigmoid函数替换为Hard-Sigmoid函数，使得计算量更小。使用交叉熵损失来优化网络，使网络更好地学习类别之间的差异性。

1.2 LK光流法

光流法通过比较2幅图像中对应像素的亮度值变化，来推断像素在图像中的运动方向和大小，可以应用于计算机视觉任务。光流法包括稠密光流法和稀疏光流法2种。由于本文仅需对提取的部分特征点进行判断，不用判断图像帧的每个像素，因此，选取稀疏光流法作为动态特征点的剔除方法，降低了计算量，提高系统实时性。

稀疏光流法是基于3个假设条件而成立：①像素在运动过程中的灰度值恒定；②时间连续或相邻

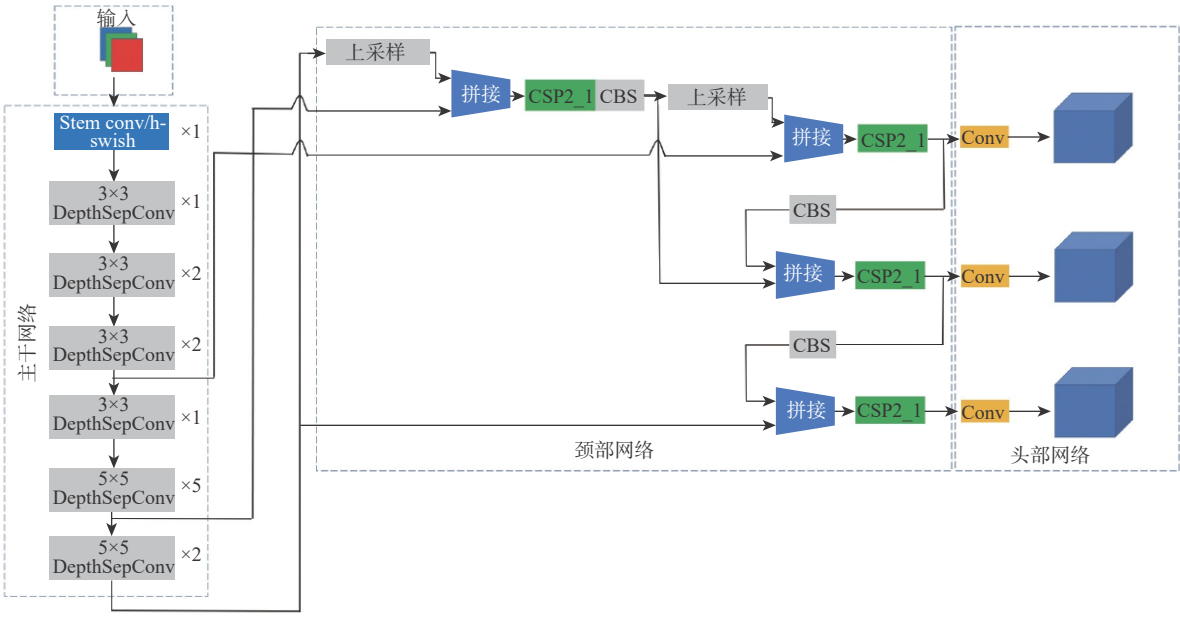


图 1　PP-LCNet-YOLOv5s 结构示意图

Fig. 1　Diagram of PP-LCNet-YOLOv5s structure

2 帧之间运动为“小运动”; ③相邻像素具有相似的运动。

根据假设①可得

$$I(x,y,t)=I(x+\mathrm{d}x,y+\mathrm{d}y,t+\mathrm{d}t) \tag{1}$$

式中, t和$t+\mathrm{d}t$分别为相邻图像帧的对应时间; $I(x,y,t)$和$I(x+\mathrm{d}x,y+\mathrm{d}y,t+\mathrm{d}t)$分别为同一像素在相邻 2 帧间的坐标。

将式(1)右端泰勒展开, 得

$$I(x,y,t)=I(x,y,t)+\frac{\partial I}{\partial x}\mathrm{d}x+\frac{\partial I}{\partial y}\mathrm{d}y+\frac{\partial I}{\partial t}\mathrm{d}t+\varepsilon \tag{2}$$

式中: ε为高阶无穷小项, 根据假设②, 该值可忽略不计。

将式(2)等号两边进行相消处理后再同时除以$\mathrm{d}t$, 可得

$$\frac{\partial I}{\partial x}\cdot\frac{\mathrm{d}x}{\mathrm{d}t}+\frac{\partial I}{\partial y}\cdot\frac{\mathrm{d}y}{\mathrm{d}t}+\frac{\partial I}{\partial t}=0 \tag{3}$$

令u、v分别为像素在x轴、y轴的运动速度, 可得

$$\begin{cases}u=\dfrac{\mathrm{d}x}{\mathrm{d}t}\\ v=\dfrac{\mathrm{d}y}{\mathrm{d}t}\end{cases} \tag{4}$$

$\partial I/\partial x$为图像在该点处x轴的梯度, $\partial I/\partial y$为图像在该点处y轴的梯度, 分别记为I_x、I_y, $\partial I/\partial t$为图像灰度对时间的变化量, 记为I_t, 代入式(3)并写成矩阵形式, 得

$$\begin{bmatrix}I_x & I_y\end{bmatrix}\begin{bmatrix}u\\ v\end{bmatrix}=-I_t \tag{5}$$

根据式(5)可知, 求取像素的速度矢量需引入额外约束。在 LK 光流中, 选取以特征点为中心的 3×3 窗口, 根据假设③, 内部 9 个像素具有相同的运动, 该方程转变为关于u、v的超定线性方程, 利用最小二乘法求解。

在实际运用场景中, 图像的静态部分由于相机的运动会产生光流矢量, 依据式(6)计算出图像静态部分的平均光流矢量。

$$\begin{bmatrix}U\\ V\end{bmatrix}=\frac{1}{N}\sum_{k=1}^{N}\begin{bmatrix}u_k\\ v_k\end{bmatrix} \tag{6}$$

式中: U、V分别为图像静态部分的像素在x轴、y轴的平均运动速度; u_k、v_k分别为图像静态部分计算的像素点在x轴、y轴的运动速度; N为计算的像素个数。

利用式(7)对特征点进行筛选, 判断特征点是否为动态特征点。

$$\sqrt{(u-U)^2+(v-V)^2}>l \tag{7}$$

式中: l为判断特征点是否为动态特征点的阈值, 一般为静态平均光流矢量的 2 倍。若大于该值, 则判定为动态特征点, 反之, 为静态特征点。

1.3　改进的 ORB-SLAM3 系统框架

ORB-SLAM3 系统在 ORB-SLAM2 系统和 ORB-SLAM-Ⅳ[19] 系统的基础上进行改进, 由 3 个线程组成: 跟踪、局部建图、回环检测。跟踪线程通过提取和匹配特征点来估算相机的位姿; 局部建图线程对跟踪线程得到的关键帧进行局部地图构建, 并进行优化; 回环检测线程检测地图是否出现回环, 根据检测结果进行地图融合, 以提高系统定位精度和鲁棒性。

ORB-SLAM3系统的优秀性能基于静态场景假设的基础之上，但在动态场景下，会受动态目标的影响，使得系统定位精度下降，甚至跟踪失败。针对这些问题，本文在该系统原有基础上添加了轻量级目标检测网络和LK光流法，用于提高系统的定位精度和鲁棒性。改进后系统结构如图2所示。主要在跟踪线程中加入改进后的轻量级目标检测网络和稀疏光流法，剔除动态特征点，只保留静态特征点，提高系统定位精度。图中：局部BA优化主要是对当前帧及其共视图中的关键帧进行联合优化，并且调整相机位姿和地图点的三维坐标，以最小化重投影误差，提高局部地图的精度；计算Sim3的作用是：通过计算Sim3变换，对齐不同地图的尺度，并在多地图模式下对齐和合并不同的子地图。

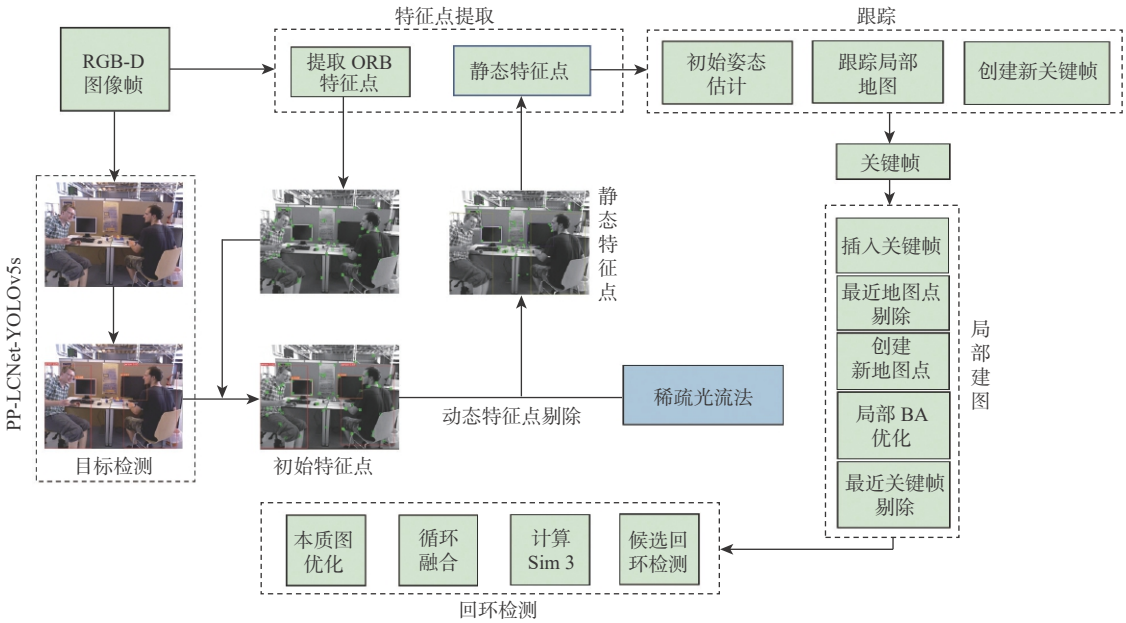


图2 改进的系统流程及框架

Fig. 2 Improved system process and framework

2 动态特征点剔除策略

动态特征点的剔除对于ORB-SLAM3系统在动态场景下能够稳定、准确的运行十分重要。但只基于改进的YOLOv5s目标检测网络，直接粗暴剔除动态目标锚框区域内的特征点，会将一些静态特征点剔除，导致SLAM系统在跟踪线程提取的特征点数量减少，从而影响定位精度，并且存在部分静态目标由于动态目标的运动而运动，如被人移动的书、椅子等，这些目标检测框区域的特征点也应作为动态特征点剔除。

现有的动态特征点剔除策略（如文献[12,17]）是将目标检测锚框内的特征点全部进行动态检测，该策略虽然在低动态序列下的定位精度会更好，但会影响系统的实时性。本文提出了一种动态特征点剔除策略，根据先验知识对目标进行分类，如人、动物、扫地机器人等会经常移动，可分为动态目标；书、水杯等可能会因为人的运动而运动，可分为潜在动态目标。表1为室内常见目标的分类。将位于动态目标锚框的特征点归类为动态特征点，位于潜在动态目标锚框的特征点归类为潜在动态特征点，剩余特征点归类为静态特征点。假设图像帧中动态特征点集合为Q，潜在动态特征点集合为P，静态特征点集合为S。以图3(a)为例，首先，在SLAM系统的跟踪线程提取特征点的同时，利用改进的YOLOv5s目标检测网络进行目标检测，将动态目标（见图中的2个人）锚框内的特征点放入集合Q，将潜在动态目标（见图中的电脑屏幕、椅子）锚框内的特征点放入集合P，其余特征点放入集合S。然后，利用LK光流法计算集合S、P中特征点的光流矢量，将集合S中特征点的光流矢量作为静态区域的光流矢量，代入式(6)，计算出平均光流矢量；将集合P中特征点的光流矢量代入式(7)进行判断，若小于设定的阈值，则说明该特征点为静态特征点，移

表1 室内目标的动态属性设定

Table 1 Dynamic property setting for indoor targets

目标	类别	目标	类别
人	动态目标	椅子	潜在动态目标
鸟	动态目标	tv	潜在动态目标
猫	动态目标	水杯	潜在动态目标
狗	动态目标	书	潜在动态目标
扫地机器人	动态目标	桌子	潜在动态目标



(a) 特征点提取　　(b) 动态特征点剔除结果

图 3　动态特征点剔除流程

Fig. 3　Flowchart for dynamic feature point elimination

入静态特征点集合S。最后, 保留集合S中的特征点用于特征匹配、相机的位姿估计, 其余特征点被剔除。

该策略没有将动态目标锚框和潜在动态目标锚框重合部分的特征点直接作为动态特征点而剔除, 而是利用 LK 光流法判断, 若光流矢量小于设定的阈值, 则说明未受动态目标影响, 可放入集合S, 增加了特征点匹配数量, 进一步提高了系统定位精度及鲁棒性。动态特征点剔除效果如图 3(b) 所示, 可以看到, 人身上的特征点被作为动态特征点剔除, tv 锚框部分和椅子锚框部分与人的锚框重合, 但 tv 部分的特征点因 LK 光流法检测为静态特征点而被保留, 而椅子部分的特征点因人的运动而移动被检测为动态特征点剔除。

3　实　验

3.1　基于轻量化网络改进的 YOLOv5s 实验

为验证 PP-LCNet-YOLOv5s 目标检测网络的性能, 将训练的超参数设置为: 初始学习率设置为 0.01, 使用 Adam 优化器, 优化器的动量设置为 0.937, 权重的衰减设置为 0.000 5, batch_size 设置为 4, 训练 500 个轮次, 并设定若精度在 50 个轮次内没有提升, 则停止训练, 边界框回归损失L_{box}的超参数设置为 0.05, 目标物体损失L_{obj}的超参数设置为 0.7, 分类损失L_{cls}的超参数设置为 0.3, iou 的阈值设置为 0.2。使用的数据集为 VOC2007+VOC2012, 其中包含 20 个种类, 有人、猫、tv 等种类的图像, 将数据集划分为训练集和验证集, 划分比例为 1∶1。

经训练后, 改进的 PP-LCNet-YOLOv5s 与 YOLOv5 各系列在 CPU 上的模型性能测试结果如表 2 所示。可知, PP-LCNet-YOLOv5s 较 YOLOv5s 网络参数减少了 41.89%, 且在 CPU 上运行速度加快了 39.13%; 在精度方面, 由于 PP-LCNet 网络的优势在于模型小、速度快, 但检测性能有所不足, 改进后的目标检测网络较 YOLOv5s, 精确率下降了 2.8%, mAP@0.5 下降了 2.6%, 牺牲了部分精度, 提高了系统的运行速度。

表 2　目标检测网络在 CPU 上的性能测试

Table 2　Performance test of object detection network on CPU

模型	算力要求	参数量	帧速率/(帧·ms^{-1})	精确率/%	召回率/%	mAP@0.5/%
PP-LCNet-YOLOv5s	8.7 GFLOPs	4.3×10^6	11.2	76.1	84.3	73.0
YOLOv5s	17.4 GFLOPs	7.4×10^6	18.4	78.9	87.5	75.6
YOLOv5m	37.4 GFLOPs	21.4×10^6	26.8	81.2	90.5	78.8
YOLOv5l	82.1 GFLOPs	47.1×10^6	43.9	83.2	91.1	80.7
YOLOv5x	141.8 GFLOPs	89.2×10^6	71.0	84.0	92.0	81.6

3.2　实验数据集

为评估算法性能, 实验使用了由德国慕尼黑工业大学提供的标准 RGB-D 数据集, 该数据集使用 Kinect 传感器捕获的室内场景同步 RGB 图像和深度图像, 常用于 SLAM 算法的评估。本文使用 TUM 数据集中 5 个序列进行测试, 包括 sitting_xyz、sitting_static、walking_halfsphere、walking_rpy 及walking_xyz, 其中, sitting_xyz 和 sitting-static 序列属于低动态场景, walking_halfsphere、walking-rpy 和 walking-xyz 序列属于高动态场景。实验中可能会出现系统在动态序列跟踪失败的情况, 为降低这种不确定性的影响, 将每个序列运行 10 次, 取各序列的平均值为实验数据。

实验使用的设备为联想笔记本电脑, CPU 型号为 R5-5600H, 主频为 3.3 GHz, 图形显卡为 NVIDIA Geforce GTX1650, 系统环境为 Ubuntu18.04。

3.3　策略比较分析

为比较基于本文提出的动态特征点剔除策略的 SLAM 算法与基于现有动态特征点剔除策略的 SLAM 算法的差异性, 分别选择低动态序列 sitting_static 和动态序列 walking_halfsphere 进行测试。本文选取绝对轨迹误差(absolute trajectory error, ATE)和相对姿态误差(relative pose error, RPE)作为评估指标。实验中, 计算均方根误差(root mean squared error, RMSE)、标准偏差(standard deviation,STD)表明算法的定位精度。测试结果如表 3 所示, 其中, 策略 1 是基于现有的动态特征点剔除策略的 SLAM 算法, 策略 2 是基于本文提出的动态特征点剔除策略的 SLAM 算法。

表3 不同策略下的定位精度和每帧耗时

Table 3 Positioning accuracy(unit m) and time consuming per frame(unit s) under different strategies

数据集	均方根误差/m		标准偏差/m		每帧耗时/s	
	策略1	策略2	策略1	策略2	策略1	策略2
sitting_staic	0.007 9	0.010 3	0.004 3	0.006 7	0.081 3	0.063 1
walking_halfsphere	0.019 6	0.020 4	0.009 8	0.010 7	0.078 1	0.066 4

由表3可知，策略2在低动态场景下，由于将先验信息为动态的静态目标（如坐着不动的人）锚框的特征点直接作为动态特征点剔除，导致定位精度下降，但由于需要动态检测的特征点数量减少，实时性能有所提高；在高动态场景下，策略2能够正确剔除动态特征点，因此，定位精度与策略1的定位精度相当，并且实时性较策略1平均每帧耗时少12 ms。在定位精度不影响系统正常运行时，实时性能也是一个重要的指标，综合考虑，本文策略更优。

3.4 实验结果分析

由于本文系统是在ORB-SLAM3系统的基础上进行改进的，将其与ORB-SLAM3系统进行比较，以量化本文算法在SLAM算法中的改进。

图4为ORB-SLAM3算法与改进算法在低动态场景及高动态场景下相机位姿估计轨迹与真实轨迹对比。由图4(a)、图4(b)可明显观察到，ORB-SLAM3算法在高动态场景下受动态目标的影响，导致相机轨迹估计与真实轨迹相差较大，定位精度较差，而改进算法利用目标检测网络和光流法将动态特征点剔除，仅利用静态特征点进行相机位姿估计，因此，估计的相机轨迹与真实轨迹基本一致，误差较小。从图4(c)、图4(d)可知，在低动态场景中，改进算法与ORB-SLAM3算法的相机位姿估计轨迹与真实轨迹基本一致，说明ORB-SLAM3算法在低动态场景下定位精度较高，改进算法效果不明显。

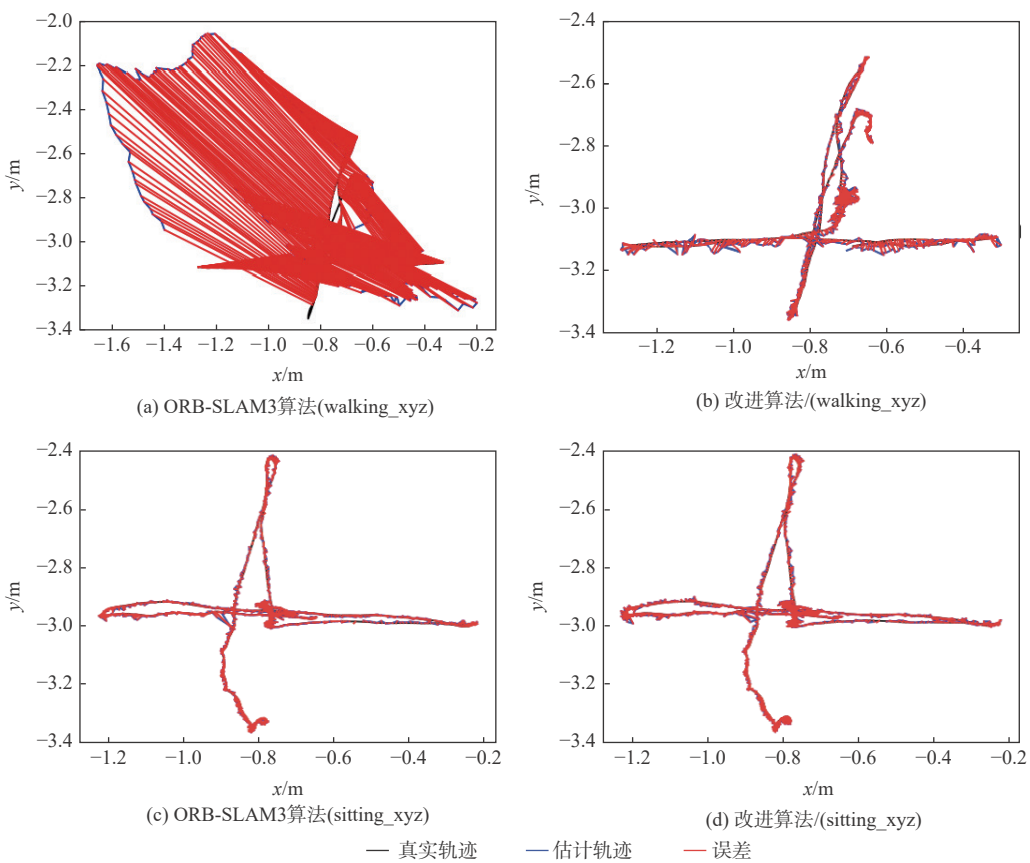


图4 估计轨迹与真实轨迹之间的差距

Fig. 4 Difference between estimated trajectory and real trajectory

图5为ORB-SLAM3算法与改进算法在sitting_xyz序列和walking_xyz序列的轨迹误差分布。可以发现，ORB-SLAM3算法在低动态场景下的轨迹误差很小，而在高动态场景下轨迹误差较大，为dm级别；改进算法不仅在低动态场景下轨迹误差小，且在高动态场景下误差为cm级别，较ORB-

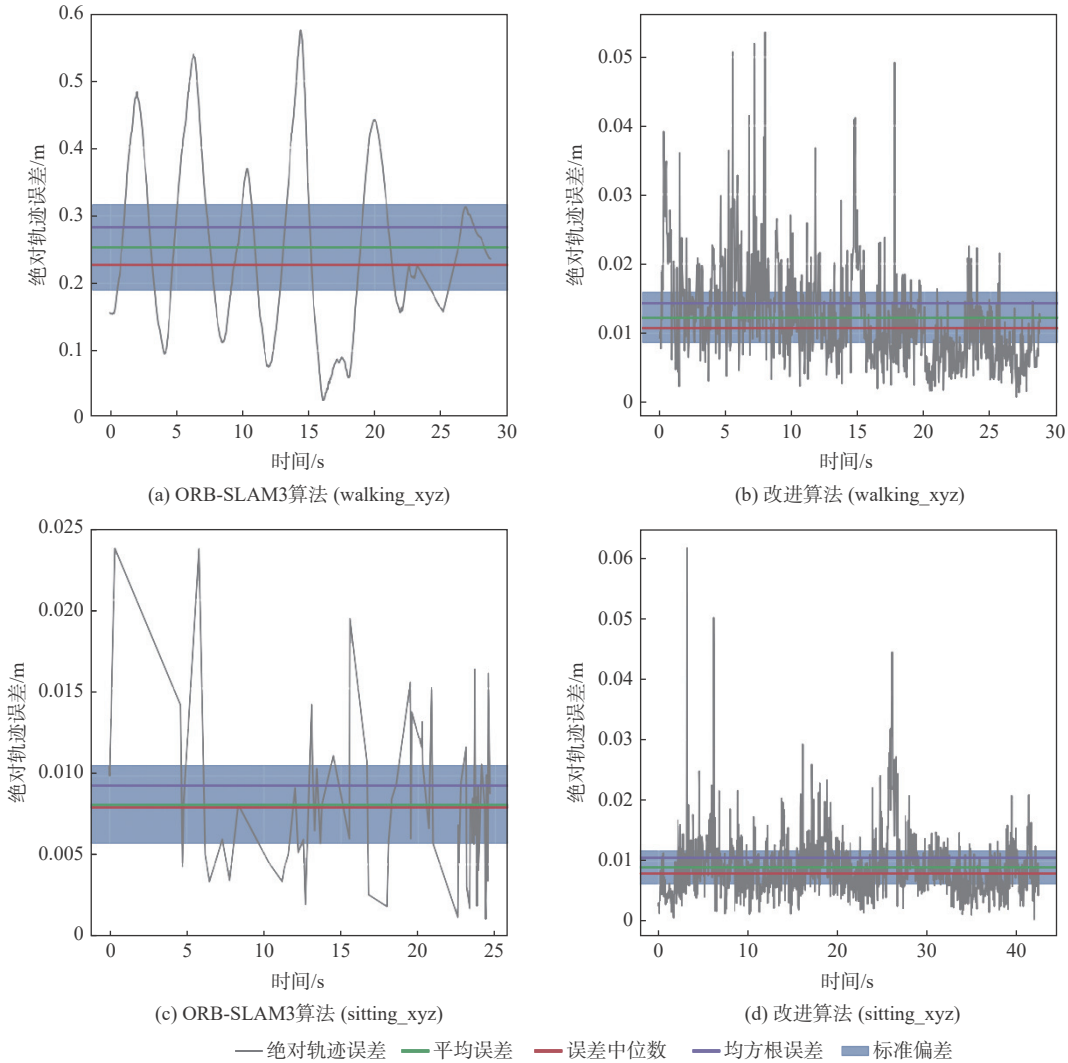


图 5　walking_xyz 和 sitting_xyz 序列的误差分布

Fig. 5　Error distribution of walking_xyz and sitting_xyz sequences

SLAM3 算法平均有一个数量级的提升, 相机位姿的定位精度有明显提升。

由表 4 及表 5 数据可知, 改进算法在低动态场景下, 由于将部分静态特征点检测为动态特征点剔除, 导致正确匹配的特征点数量减少, 系统定位精度有所降低, 但不影响系统在低动态场景下正常运行; 在高动态场景下, 相比 ORB-SLAM3 算法, 本文算法在 3 个动态数据集下绝对轨迹误差的均方根误差的平均提升率达 92.38%, 说明本文算法在高动态场景下的定位精度更好。

同时, 将改进算法与其他动态 SLAM 算法在动态数据集的运行结果进行对比, 包括 DynaSLAM 算法、RTD-SLAM 算法和 RDS-SLAM 算法, 比较数据为绝对轨迹路径下的均方根误差。为排除硬件设备等其他因素的干扰, 本文计算了相对提升率, 能直接说明改进算法较原 SLAM 算法在同一实验条件下的提升率, 计算公式为

$$\rho = \frac{\mu - \nu}{\mu} \times 100\% \tag{8}$$

式中: ρ为相对提升率; μ为 ORB-SLAM 算法的运行

表 4　ORB-SLAM3 算法与改进算法的绝对轨迹误差对比分析

Table 4　Comparison of ATE analysis between ORB-SLAM3 algorithm and improved algorithm　　m

序列	均方根误差		标准偏差		平均误差		误差中位数	
	ORB-SLAM3算法	改进算法	ORB-SLAM3算法	改进算法	ORB-SLAM3算法	改进算法	ORB-SLAM3算法	改进算法
sitting_xyz	0.009 2	0.011 5	0.005 3	0.007 5	0.007 9	0.008 9	0.007 3	0.007 9
sitting_staic	0.007 1	0.010 3	0.003 5	0.006 7	0.006 3	0.007 2	0.005 7	0.006 3
walking_rpy	0.357 1	0.030 2	0.171 7	0.016 8	0.298 3	0.022 5	0.284 6	0.020 3
walking_halfsphere	0.218 5	0.020 4	0.087 9	0.010 7	0.199 9	0.017 4	0.186 4	0.016 1
walking_xyz	0.283 1	0.014 3	0.126 3	0.007 4	0.253 4	0.012 2	0.227 8	0.010 7

表5 ORB-SLAM3与改进算法的相对姿态误差对比分析

Table 5 Comparison of RPE analysis between ORB-SLAM3 algorithm and improved algorithm m

序列	均方根误差		标准偏差		平均误差		误差中位数	
	ORB-SLAM3算法	改进算法	ORB-SLAM3算法	改进算法	ORB-SLAM3算法	改进算法	ORB-SLAM3算法	改进算法
sitting_xyz	0.009 6	0.011 9	0.006 1	0.006 4	0.008 4	0.010 0	0.007 5	0.008 4
sitting_staic	0.004 8	0.007 4	0.002 5	0.004 6	0.004 1	0.006 7	0.003 6	0.005 9
walking_rpy	0.374 1	0.031 9	0.331 6	0.021 9	0.421 3	0.022 3	0.411 3	0.021 1
walking_halfsphere	0.153 4	0.014 3	0.098 4	0.008 7	0.117 5	0.011 3	0.102 5	0.010 5
walking_xyz	0.149 7	0.012 1	0.125 7	0.007 1	0.137 1	0.009 8	0.122 4	0.008 7

结果；ν为改进算法的运行结果。

对比结果如表6所示。可知，最优结果分别出现在DynaSLAM算法、RTD-SLAM算法和本文算法，但是由于DynaSLAM算法使用了语义分割算法进行动态特征点剔除，在语义分割线程耗时较多，不能满足系统实时性要求，而RTD-SLAM算法与本文算法使用轻量化目标检测网络和光流法进行动态特征点剔除，耗时较少，但RTD-SLAM算法在光流模块部分将图像帧所有特征点进行动态检测且未对目标检测网络进行轻量化处理，耗时较本文算法更多。综上，本文算法性能更优。ORB-SLAM3、DynaSLAM、RTD-SLAM及改进算法测试TUM数据集的平均每帧耗时如表7所示。

表6 改进算法与其他动态SLAM算法的绝对轨迹误差对比分析

Table 6 Comparison of ATE analysis between improved algorithm and other dynamic SLAM algorithms %

序列	DynaSLAM	RTD-SLAM	RDS-SLAM	改进算法
walking_rpy	93.64	92.46	90.68	91.54
walking_halfsphere	88.79	90.13	84.46	90.66
walking_xyz	93.31	95.57	91.68	94.95

表7 改进算法和其他算法的跟踪线程耗时和每帧耗时

Table 7 Tracking thread time consumption and time consumption per frame of improved algorithm and other algorithms

算法	网络模型	跟踪线程耗时/ms	每帧消耗时间/ms
ORB-SLAM3			26
DynaSLAM	Mask R-CNN	200	230
RTD-SLAM	YOLOv5s	43	82
改进算法	PP-LCNet-YOLOv5s	31	65

4 结 论

1）本文使用轻量级网络PP-LCNet替换YOLOv5s的主干网络，使目标检测网络的模型参数减少41.89%，在CPU上运行推理速度提升39.13%。

2）提出一种在ORB-SLAM3系统的跟踪线程结合改进后的YOLOv5s网络和LK光流法的算法，在动态场景下较ORB-SLAM3算法，相机的位姿估计精度的均方根误差提升92.38%，相比其他动态SLAM算法，也有部分提升，且能满足实时性要求。

对于下一步工作：一是考虑使用包含更多种类的数据集对目标检测网络进行训练，提高网络泛化性及系统动态目标的识别率，使系统具有更强的适应性。二是考虑将光流法与SLAM系统计算的深度信息结合，提高光流法动态检测的精度和可靠性。

参考文献（References）

[1] 田野, 陈宏巍, 王法胜, 等. 室内移动机器人的SLAM算法综述[J]. 计算机科学, 2021, 48(9): 223-234.
TIAN Y, CHEN H W, WANG F S, et al. Overview of SLAM algorithms for mobile robots[J]. Computer Science, 2021, 48(9): 223-234(in Chinese).

[2] CAMPOS C, ELVIRA R, RODRÍGUEZ J J G, et al. ORB-SLAM3: an accurate open-source library for visual, visual-inertial, and multimap SLAM[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.

[3] MUR-ARTAL R, TARDÓS J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.

[4] GOMEZ-OJEDA R, MORENO F A, ZUÑIGA-NOËL D, et al. PL-SLAM: a stereo SLAM system through the combination of points and line segments[J]. IEEE Transactions on Robotics, 2019, 35(3): 734-746.

[5] 张有全, 祁宇明, 邓三鹏, 等. 直接法和共视图优化的视觉惯性SLAM系统研究[J]. 自动化与仪器仪表, 2022(5): 197-203.
ZHANG Y Q, QI Y M, DENG S P, et al. Research on visual-inertial SLAM system based on direct method and common view optimization[J]. Automation and Instrumentation, 2022(5): 197-203(in Chinese).

[6] ENGEL J, STÜCKLER J, CREMERS D. Large-scale direct SLAM with stereo cameras[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2015: 1935-1942.

[7] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611-625.

[8] LI H, JIANG Y H, FAN F Y, et al. RGB-D SLAM based on semantic and geometry for indoor dynamic environments[C]//Proceed-

ings of the China Automation Congress. Piscataway: IEEE Press, 2022: 5581-5586.

[9] LONG X D, ZHANG W W, ZHAO B. PSPNet-SLAM: a semantic SLAM detect dynamic object by pyramid scene parsing network[J]. IEEE Access, 2020, 8: 214685-214695.

[10] BESCOS B, FÁCIL J M, CIVERA J, et al. DynaSLAM: tracking, mapping, and inpainting in dynamic scenes[J]. IEEE Robotics and Automation Letters, 2018, 3(4): 4076-4083.

[11] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.

[12] HAN S Q, XI Z H. Dynamic scene semantics SLAM based on semantic segmentation[J]. IEEE Access, 2020, 8: 43563-43570.

[13] YUAN X, CHEN S. SaD-SLAM: a visual SLAM based on semantic and depth information[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2020: 4930-4935.

[14] YU C, LIU Z X, LIU X J, et al. DS-SLAM: a semantic visual SLAM towards dynamic environments[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2018: 1168-1174.

[15] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[16] LIU Y B, MIURA J. RDS-SLAM: real-time dynamic SLAM using semantic segmentation methods[J]. IEEE Access, 2021, 9: 23772-23785.

[17] SU P, LUO S Y, HUANG X C. Real-time dynamic SLAM algorithm based on deep learning[J]. IEEE Access, 2022, 10: 87754-87766.

[18] ZHANG H, YE M, SUN X. Robust indoor visual-inertial SLAM with pedestrian detection[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 68(10): 3897-3908.

[19] MUR-ARTAL R, TARDÓS J D. Visual-inertial monocular SLAM with map reuse[J]. IEEE Robotics and Automation Letters, 2017, 2(2): 796-803.

Dynamic visual SLAM algorithm based on improved YOLOv5s

JIANG Changjiang*, LIU Peng, SHU Peng

(School of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: A dynamic visual simultaneous localization and mapping (SLAM) algorithm based on an object detection network is proposed to address the robustness and camera localization accuracy issues caused by dynamic targets in indoor dynamic scenes. The lightweight network PP-LCNet replaces the YOLOv5 backbone network, and the YOLOv5s with the shortest depth and feature map width are chosen as the object detection network. After training on the VOC2007+VOC2012 dataset, experimental results show that the PP-LCNet-YOLOv5s model reduces the network parameters by 41.89% and improves the running speed by 39.13% compared to the YOLOv5s model. In order to eliminate dynamic feature points from the tracking thread of the visual SLAM system, a parallel thread that combines the enhanced object recognition network and sparse optical flow approach is implemented. Only static feature points are used for feature matching and camera position estimation. Experimental results show that the proposed algorithm improves the camera localization accuracy in dynamic scenes by 92.38% compared to ORB-SLAM3.

Keywords: simultaneous localization and mapping (SLAM); target detection; dynamic feature point elimination; positioning accuracy; optical flow approach

Received: 2023-03-31; **Accepted**: 2023-06-09; **Published Online**: 2023-07-26 08:39

URL: link.cnki.net/urlid/11.2625.V.20230724.1959.003

Foundation item: National Natural Science Foundation of China (62277008)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0142

超声速进气道出口封闭的非定常流动特性研究

温玉芬*，张炜群，郝思思
(中国航天科技集团公司 中国运载火箭技术研究院，北京 100076)

摘 要：冲压发动机转级过程中可能会出现进气道入口通流、出口封闭的盲腔流动现象，使得进气道发生大幅气流振荡，飞行器面临姿态失稳和结构破坏的风险。针对进气道的气流振荡问题，对一种超声速双侧布局进气道在出口封闭状态下的非定常流动特性开展风洞试验和数值仿真研究，获得了模型尺寸、来流马赫数及边界层吸除对进气道压强振荡特性的影响。结果表明：出口封闭时，进气道发生周期性的流动振荡现象，气流振荡频率与来流声速成正比，与进气道长度成反比，压强振荡峰值与来流总压接近，且随着来流马赫数的增大显著上升。采用的数值仿真方法较好地模拟了进气道出口封闭状态下的非定常流动，数值仿真结果与风洞试验结果吻合良好。非定常仿真结果进一步表明：在气流振荡过程中，进气道内收缩段的边界层吸除槽可以起到泄压作用，使得进气道头部斜激波系在回退过程中达到封口状态，捕获的流量系数较无边界层吸除槽时明显增大，从而导致了进气道压强振荡峰值上升49.47%，频率下降21.78%。

关 键 词：超声速进气道；振荡压强；边界层吸除；非定常；风洞试验
中图分类号：V221.3
文献标志码：A **文章编号**：1001-5965(2025)03-0772-12

整体式固体火箭冲压发动机无法在地面静止条件下起动工作，通常需要通过助推器加速到一定的速度后再进行转级接力。在发动机助推加速的过程中，为隔离发动机内高温高压燃气对进气道的干扰影响，防止进气道发生结构破坏，通常在进气道出口设置堵盖结构。与此同时，为避免进气道在助推加速阶段出现长时间的盲腔流动，从而导致飞行阻力增大和姿态控制困难的问题，工程上较常采用的措施是在进气道入口也设置堵盖结构。当飞行器到达接力点后，进气道前、后堵盖打开，发动机转级工作。然而，实际工程设计时，考虑到助推发动机工作的偏差、发动机后效影响等，为保证转级的可靠性，防止出现进气道出口堵盖过早打开从而导致火焰回卷及因燃烧室压力过高导致进气道无法起动工作的问题，通常在动作时序上按照进气道前、后堵盖顺序相继打开进行设计。这种设计方法可以避免进气道在助推加速阶段出现长时间盲腔流动问题，但仍不可避免地在转级过程中出现短时的这种流动现象。

当进气道入口堵盖打开，而出口堵盖仍然处于封闭状态时，进气道为一端开口、另一端封闭的盲腔。高速气流进入这种盲腔结构后，会导致进气道发生短时的大幅气流振荡，从而使得飞行器仍然在短时内面临姿态失稳和结构破坏的风险。至今为止，进气道的非定常流动振荡问题已经引起了国内外众多学者的关注，大部分工作集中于对进气道喘振特性[1-5]及激波振荡特性[6-9]的研究。其中，西北工业大学和空空导弹研究院分别对串联式冲压发动机和头部两侧进气方式的冲压发动机在助推段的压强振荡特性进行了研究[10-11]，获得了飞行马赫数及高度变化对压强振荡特性的影响；国防科技大学采用数值方法对进气道整流罩开启前后的流场

收稿日期：2023-03-24；录用日期：2023-07-26；网络出版时间：2023-08-07 17:54
网络出版地址：link.cnki.net/urlid/11.2625.V.20230807.1347.001
*通信作者. E-mail：wyf369963@163.com

引用格式：温玉芬，张炜群，郝思思. 超声速进气道出口封闭的非定常流动特性研究[J]. 北京航空航天大学学报，2025，51(3)：772-783.
WEN Y F，ZHANG W Q，HAO S S. Investigation on unsteady flow characteristics of a supersonic inlet with exit blocked[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：772-783 (in Chinese).

结构进行了分析[12]，结果表明，在冲压发动机助推段，当进气道进口通流、出口封闭时，其流动表现为整体式的非定常振荡现象。然而，这些研究均是针对无流动控制措施的进气道开展，对于引入边界层流动控制可能给这种整体式气流振荡特性带来的影响均未涉及。

对于宽包线工作的飞行器而言，通常要求进气道在宽速域范围内均具有较高的性能。被动流动控制措施，如边界层吸除槽[13-14]，由于其结构简单且无需引入额外的复杂控制装置，因此，在导弹上具有较大的应用潜力。然而，这种被动式的边界层吸除槽，由于无主动控制装置，因此，在导弹的整个工作过程中均与外部环境处于连通状态。进气道正常工作状态下，边界层吸除槽的作用和流动特性已被大量的研究验证[15-22]，但对于发动机转级阶段，当进气道处于盲腔流动状态下、内流道压力大幅非定常变化时，这种边界层吸除槽所起到的作用及其对进气道气流振荡特性的影响尚不清楚。

为此，本文针对一种采用整体式固体火箭冲压发动机作为动力装置的飞行器，设计了带边界层吸除槽的超声速双侧二元进气道，采用风洞试验和非定常数值仿真方法分析了冲压发动机转级过程中该种进气道的盲腔流动特性，并获取了模型尺寸、来流马赫数及边界层吸除槽对进气道盲腔气流振荡特性的影响规律。

1　研究对象

本文针对一种采用整体式固体火箭冲压发动机作为动力装置的飞行器开展研究，该飞行器采用双侧倒置式（压缩面朝下）二元进气道布局，进气道出口截面设置有堵盖结构，如图 1 所示。在发动机助推加速阶段，进气道出口堵盖处于封闭状态，可将发动机补燃室内的高温高压燃气与进气道相隔离。本文主要针对发动机转级过程进气道入口通流、出口封闭状态的流动特性开展研究。

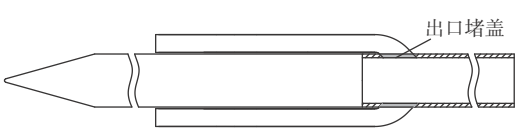
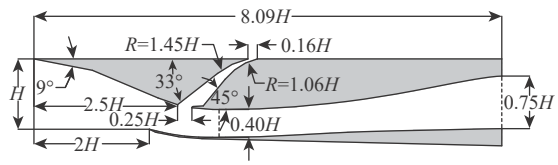


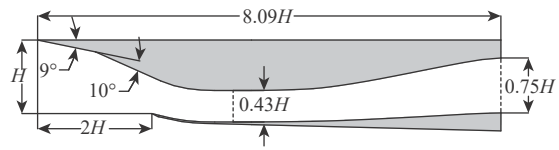
图 1　双侧进气道布局的飞行器示意图（俯视图）

Fig. 1　Schematic diagram of configuration of flying vehicle with twin-duct inlets (top view)

本文研究的飞行器进气道采用三波系混压形式，设计马赫数为 3.0，一、二级压缩楔角分别为 9°和 10°，内收缩比为 1.2（有边界层吸除槽时）。进气道压缩面前缘距离飞行器头部顶点的距离为 8.9H，其中，H 为唇口高度，单侧进气道宽度为 0.66H，进气道总长 L 为 8.09H。为对比边界层吸除槽对进气道出口封闭状态下的流动特性影响，本文分别设计了有边界层吸除槽和无边界层吸除槽的 2 种进气道方案，如图 2 所示。其中，边界层吸除槽开设在进气道内收缩段，吸除槽进口前缘距离进气道压缩面尖点的轴向距离为 2.5H，将该方案记为方案 1；无边界层吸除槽方案记为方案 2。由于方案 1 设计了边界层吸除槽，为保证进气道具有相同的自起动能力，其喉道高度由无边界层吸除槽的 0.43H 减小至 0.40H。



(a) 方案1，有边界层吸除槽



(b) 方案2，无边界层吸除槽

图 2　进气道模型示意图

Fig. 2　Schematic diagram of inlet model

2　风洞试验

对方案 1 进气道出口封闭状态下的气流振荡特性开展风洞试验研究。

2.1　试验设备

试验在中国航空工业空气动力研究院 FL-60 亚跨超三声速风洞开展，该风洞为直流引射式风洞，马赫数范围为 0.3～4.0，超声速试验时采用柔壁喷管获得不同的马赫数，试验段横截面尺寸为 1.2 m×1.2 m，风洞湍流度不大于 4‰。

风洞试验模型缩比为 1∶2.5，主要的试验参数如表 1 所示，其中，Ma 为来流马赫数，α 和 β 分别为攻角和侧滑角，P_0^* 和 T_0^* 分别为来流总压和总温。

表 1　风洞试验参数

Table 1　Parameters of wind tunnel test

Ma	α/(°)	β/(°)	P_0^*/Pa	T_0^*/K
3.0	0	0	390 000	300

2.2　动态压力测量

开展风洞试验时，采用动态压力测量系统对进气道沿程典型位置的壁面静压进行测量。如图 3 所示，在左、右侧进气道内流道上侧壁面分别布置了 4 个动态压力测点，各测点分别位于进气道入口截面、喉道截面、扩张段前段、扩张段出口截面，距

离进气道压缩面前缘的距离分别为 2.28H、3.35H、4.88H、8.09H。其中，左侧进气道各动态压力测点分别标注为 p_{L1}、p_{L2}、p_{L3}、p_{L4}，右侧进气道各动态压力测点分别标注为 p_{R1}、p_{R2}、p_{R3}、p_{R4}。试验中，采用美国 Kulite 公司的 XCQ-062-50A 型动态压力传感器进行测量，传感器的固有频率为 500 kHz，量程为 350 kPa，额定过载为 2 倍。动态压力信号使用德国的 GEN16T 系统进行采集，数据采样频率为 10 kHz，截止频率为 5 kHz，采集时间为风洞起动至风洞关车全过程。

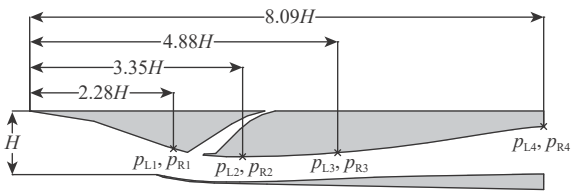


图 3 进气道动态压力测点位置示意图

Fig. 3 Schematic diagram of measuring point positions for dynamic pressure of inlet

2.3 试验结果分析

先对进气道在 Ma=3.0、出口封闭状态下的流动特性开展试验研究。图 4 为风洞试验获得的 Ma=3.0 状态下左侧进气道沿程不同位置的压强时间历程曲线，其中，各测点的压强以来流静压 p_0 进行了无量纲化处理，标示为 p/p_0；图 5 为左侧进气道 4 个测点动态压力的频谱分析结果，图中，横坐标 f 为频率，纵坐标为功率谱密度（power spectrum density, PSD）。由图可知，进气道沿程 4 个测点的压强均出现了周期性的振荡，且各阶振荡频率一致，基频均为 122 Hz。这表明，当进气道出口封闭时，内流道发生了整体式的强烈气流振荡，自进气道入口至出口，其压强振荡特性相似。然而，进气道沿程不同位置的压强振荡峰值表现出了一定的差异，沿流向压强峰值逐渐增大，以边界层吸除槽

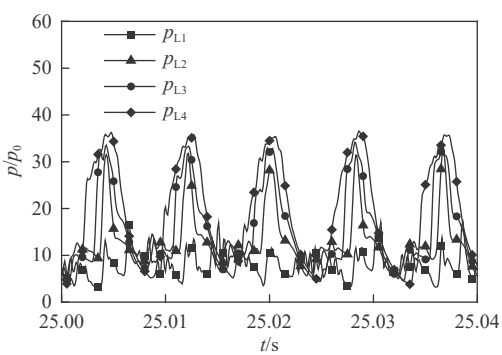


图 4 风洞试验获得的左侧进气道沿程不同位置的压强时间历程曲线（Ma=3.0）

Fig. 4 Time histories curves of pressure at different positions of left inlet in wind tunnel test (Ma = 3.0)

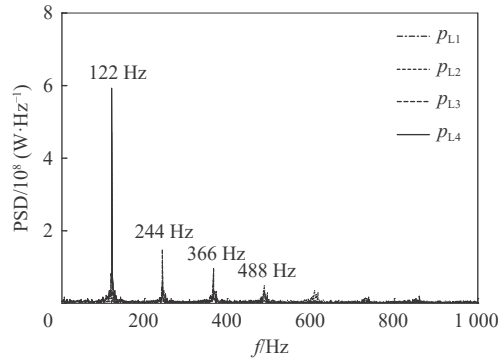


图 5 风洞试验获得的左侧进气道动态压力频谱特性（Ma=3.0）

Fig. 5 Spectral characteristics of dynamic pressure of left inlet in wind tunnel test (Ma = 3.0)

为分界面，其上、下游测点的压强变化最为显著：其中，边界层吸除槽下游的 p_{L2}～p_{L4} 测点压强峰值在 $30p_0$～$36p_0$ 范围内，为边界层吸除槽上游 p_{L1} 测点压强峰值（$15p_0$）的 2 倍以上。

由图 4 还可知，p_{L1} 测点的压强在一个周期内出现了多次波动，这主要是由于在气流振荡的过程中，进气道进口附近存在流动分离。随着激波振荡运动，p_{L1} 测点附近的分离包大小和位置随之发生变化，从而导致了其压强在一个周期内出现多次波动变化。这个现象在 3.4.1 节的数值仿真结果中将进一步进行分析说明。

为对比双侧布局进气道的左、右侧流道在这种强烈的非定常流动振荡过程中的特性差异，图 6 和图 7 分别给出了 Ma=3.0 时，左、右侧进气道扩张段出口位置 p_{L4} 和 p_{R4} 测点的压强时间历程曲线和频谱特性对比。由图可知，左、右侧进气道的压强振荡曲线规律一致，其压强振荡峰值较为接近，左、右侧分别为 $36.06p_0$ 和 $33.23p_0$，二者相差不到 8%。左、右侧进气道压强振荡的各阶频率一致，基频均为 122 Hz。这表明，在零侧滑状态下，当进气道出

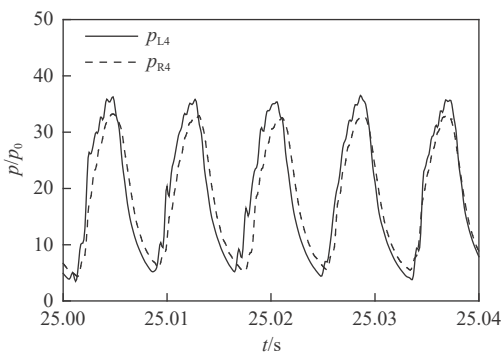


图 6 风洞试验获得的左、右侧进气道扩张段出口位置的压强时间历程曲线（Ma=3.0）

Fig. 6 Time histories curves of pressure at exit of diffuser between left and right inlets in wind tunnel test (Ma = 3.0)

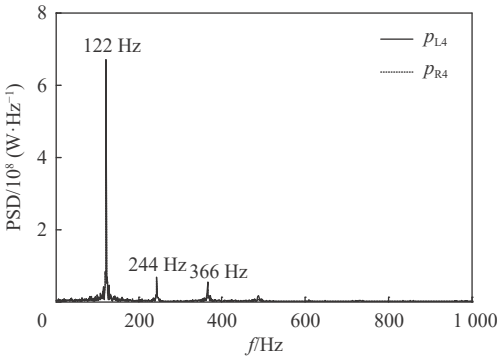


图 7　风洞试验获得的左、右侧进气道扩张段出口位置的动态压力频谱特性(*Ma*=3.0)

Fig. 7　Spectral characteristics of dynamic pressure at exit of diffuser between left and right inlets in wind tunnel test (*Ma* = 3.0)

口封闭时, 左、右两侧进气道的非定常气流振荡特性表现出了良好的一致性。

根据文献 [7], 进气道盲腔气流振荡频率可采用声学振荡频率计算式进行计算:

$$f=\frac{c}{4L}\left(1-\overline{Ma}^2\right) \tag{1}$$

式中: c 为来流声速, *Ma*=3.0 时风洞试验来流声速为 207.5 m/s; L 为进气道试验模型唇口至出口堵盖位置的长度, 该值为 0.377 m; $\overline{Ma}$为进气道内流道的平均马赫数。如图 8 所示, 通过两道斜激波和一道结尾正激波可计算进气道进口稍下游位置的马赫数 Ma_3, 根据风洞试验测量得到的气流振荡频率, 结合式(1)反算得到内流道的平均马赫数$\overline{Ma}$约为进气道进口马赫数 Ma_3 的 0.5 倍。

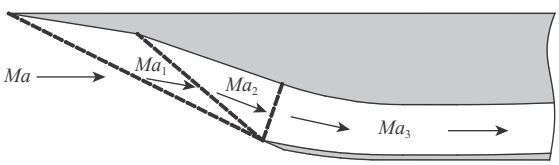


图 8　进气道流动参数变化示意图

Fig. 8　Schematic diagram of variation of inlet's flow parameters

综上所述, 当进气道出口封闭时, 内流道发生强烈的整体式非定常气流振荡, 沿程壁面压强呈现出稳定的周期性变化; 对于方案 1 内收缩段开设边界层吸除槽的进气道, 边界层吸除槽下游的压强振荡峰值为其上游的 2 倍以上; 结合风洞试验测量结果和声学振荡频率计算式可反算得到内流道的平均马赫数约为进气道进口马赫数的 0.5 倍。

3　数值仿真

对方案 1 和方案 2 的进气道在零姿态角、出口封闭状态下的流动特性开展非定常数值仿真研究, 并分析模型尺寸、马赫数及边界层吸除槽对进气道气流振荡特性的影响。

3.1　仿真网格

本文研究的进气道在超声速条件下工作, 但考虑到进气道出口封闭时存在气流振荡现象, 对进气道内外流场会产生影响, 因此, 在开展非定常仿真计算时选取进气道进口上、下游的部分外流区间、整个进气道内流道作为计算域。其中, 二维仿真的外流计算域自进气道进口向上、下游分别延伸 8H, 向外侧延伸 13H; 三维仿真计算模拟部件还包括飞行器前体和进气道边界层隔道。整个计算域采用结构化网格填充, 二维计算网格单元总数约为 12 万, 三维半模计算网格总数约为 570 万。考虑到超声速进气道内存在激波/边界层干扰, 且在进气道出口封闭的情况下, 进口附近会存在激波振荡及流动分离现象, 因此, 在近壁面处及进气道进口附近进行了网格加密。对于方案 1, 还对边界层吸除槽及其附近的网格进行了加密。方案 1 进气道的二维和三维计算网格分别如图 9 和图 10 所示。

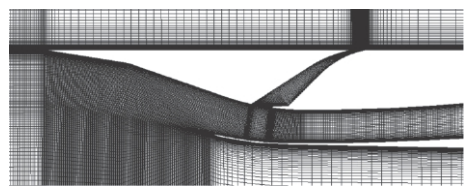
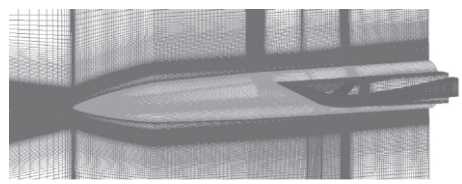


图 9　方案 1 进气道二维计算网格

Fig. 9　Grid of inlet in two-dimensional computation for option 1



(a) 整体网格



(b) 进气道局部网格

图 10　方案 1 进气道三维计算网格

Fig. 10　Grid of inlet in three-dimensional computation for option 1

3.2　仿真设置

对进气道盲腔流动开展非定常数值仿真计算, 使用基于有限体积法的雷诺平均 Navier-Stokes(N-S)方程求解器, 其中, 无黏对流通量采用 Roe 平均方式进行 MUSCL(monotonic upstream-centered scheme for conservation laws)插值, 黏性通量采用二阶迎风格式进行离散, 时间推进采用隐式格式, 湍流黏性系数采用 k-ε 标准模型, 近壁区采用壁面函数法处理。

本文数值计算工况及模拟参数如表 2 所示, 其

中，h为飞行高度，T为来流静温，ξ为模型缩比，计算工况1和工况2模拟的参数与风洞试验参数相同。数值计算的攻角和侧滑角均为0°，计算时间步长设置为1×10^{-5} s，子迭代步数为500。为获得流场内的参数变化规律，在进气道内流道上侧壁面分别布置了4个流动参数监测点，各监测点分别位于进气道入口截面、喉道截面、扩张段前段、扩张段出口截面，与风洞试验的动态压力测点位置一致，分别标注为p_1、p_2、p_3、p_4。

表2 数值计算工况及模拟参数

Table 2 Numerical calculation conditions and simulated parameters

工况	方案	h/km	Ma	p_0/Pa	T/K	ξ	维度
1	1		3.0	10 617.2	107.14	1∶2.5	三维
2	1		3.0	10 617.2	107.14	1∶2.5	二维
3	1	15	3.0	12 111.8	216.65	1∶1	二维
4	1	12	2.8	19 399.4	216.65	1∶1	二维
5	1	12	2.6	19 399.4	216.65	1∶1	二维
6	2	15	3.0	12 111.8	216.65	1∶1	二维

3.3 仿真与试验对比

首先对方案1进气道缩比模型(ξ=1∶2.5)在试验来流条件下分别开展三维和二维非定常数值仿真计算(见表2中的工况1和工况2)，将仿真与试验获得的进气道出口位置的压强进行对比，如图11所示，压强振荡峰值和频率对比结果如表3所示，其中，p_{max}为压强振荡峰值，以来流静压进行了无量纲化处理，Δf和$\Delta(p_{max}/p_0)$分别表示数值与试验获得的压强振荡频率和峰值相对误差。可以看出，当进气道出口封闭时，二维和三维数值仿真与风洞试验获得的进气道扩张段出口压强振荡曲线无论从规律上还是数值上均吻合良好，仿真与试验的压强振荡频率和峰值相对误差在3.5%以内。这表明，本文所采用的二维和三维非定常数值仿真方法均准确地模拟了进气道出口封闭情况下的内流动。

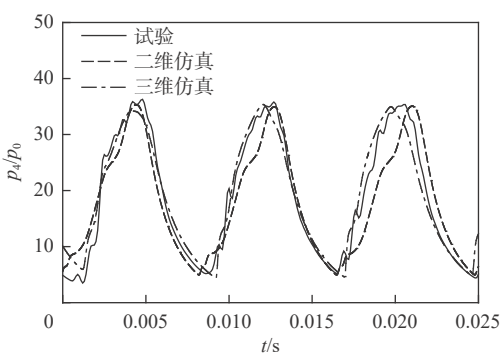


图11 方案1进气道扩张段出口压强试验值与仿真值对比

Fig. 11 Comparison of experimental and simulated pressure at exit of inlet's diffuser for option 1

表3 方案1进气道扩张段出口的压强振荡幅值和频率试验值与仿真值对比

Table 3 Comparison of experimental and simulated inlet's oscillating pressure amplitudes and frequencies at exit of diffuser for option 1

参数来源	f/Hz	p_{max}/p_0	Δf/%	$\Delta(p_{max}/p_0)$/%
试验	122.00	36.06		
二维仿真	120.47	35.40	−1.25	−1.83
三维仿真	121.99	34.87	−0.008	−3.30

由图11和表3还可知，二维和三维数值仿真的压强振荡曲线虽然略有差异，但其压强振荡峰值和频率吻合良好，这表明，零姿态角下，飞行器前体及三维效应对进气道的这种盲腔气流振荡的影响很小。这主要是由于：①进气道出口封闭时的盲腔流动为整体式的气流振荡问题，其主要以激波轴向振荡为主导，进气道高宽比受限所带来的三维效应对这种激波轴向运动的影响相对较弱；②进气道距离飞行器头部较远，与飞行器头部顶点的轴向距离为8.09倍的进气道进口高度，由表4进气道在h=15 km、Ma=3.0时的三维定常仿真结果可以看出，至进气道进口略上游位置，气流马赫数已经基本恢复至自由来流马赫数的水平，二者仅相差0.008，静温和总压与远前方自由来流仅相差0.35%和2.25%，即飞行器前体对进气道进口流动参数的影响很小。

表4 方案1进气道进口前方的气流参数与来流参数对比（h=15 km）

Table 4 Comparison of flow parameters and incoming flow parameters ahead of inlet entrance for option 1 (h = 15 km)

位置	Ma	P_0^*/Pa	T/K
自由来流	3.0	444 899.38	216.65
进气道进口前方	2.992	434 888.47	217.40

综上，零姿态角下，飞行器前体及三维效应对进气道出口封闭状态下的盲腔流动特性影响很小，尤其对这种气流振荡的幅值和频率影响非常小，即二维非定常仿真可以较准确地刻画进气道的这种整体式轴向气流振荡特性。考虑到计算效率问题，3.4节主要采用二维非定常仿真方法对进气道的盲腔气流振荡特性开展研究。

3.4 仿真结果分析

3.4.1 气流振荡特性分析

对方案1进气道1∶1全尺寸模型在出口封闭状态下开展非定常数值仿真，仿真计算的来流参数按照表2中的工况3进行设定。图12为仿真计算获得的方案1进气道不同位置监测点的压强时间历程曲线，图13为方案1进气道结尾激波串位置的时间历程曲线，其中，x_s为结尾激波串在进气道中

的轴向位置, 以进气道唇口高度 H 进行无量纲化处理, 标示为 x_s/H。由图 12 可知, 与风洞试验结果类似, 当进气道出口封闭时, 进气道沿程不同位置测点的压强均呈现出了周期性振荡现象, 且振荡平稳后, 压强峰值的沿程变化规律与风洞试验结果类似, 均是边界层吸除槽下游测点的压强峰值较高, 而边界层吸除槽上游测点的压强峰值较低。

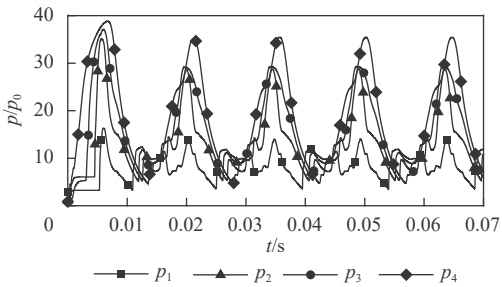


图 12　方案 1 进气道不同位置的压强时间历程曲线

Fig. 12　Time histories of pressure at different positions of inlet for option 1

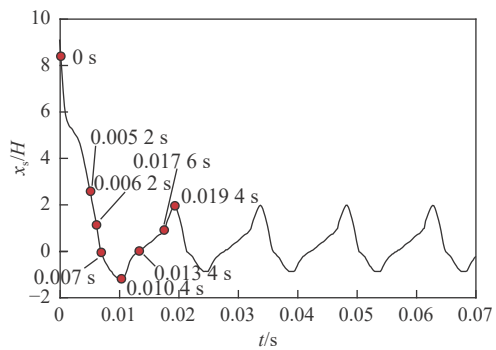


图 13　方案 1 进气道结尾激波串位置的时间历程曲线

Fig. 13　Time history of terminal shock train position of inlet for option 1

由图 12 还可知, 在第 1 个振荡周期内, 扩张段出口的压强先出现上升, 其他各测点压强均是先稳定在一较小值一段时间后才出现急剧上升, 位置越靠近上游, 出现压强剧烈增大的时刻越晚。这是由于自上游而来的气流在出口位置被截断, 由入口源源不断地流入进气道的空气在出口上游堆积压缩, 从而导致出口附近的气体密度和静压急剧增大。在该高压扰动的作用下, 进气道出口首先产生一道结尾正激波, 随着时间的推移, 结尾正激波向上游运动, 并与边界层相互作用, 衍变为结尾激波串结构, 具体见图 13 中 t=0～0.005 2 s 时间段。当结尾激波串位于某一监测点下游时, 由于波前为超声速流动, 因此, 该监测点的流动未受到扰动, 其压强保持不变; 当结尾激波串向上游传播越过某一监测点时, 该监测点附近的流动受到干扰, 且由于此时监测点位于结尾激波串的波后, 因此, 其压强急剧上升。正是由于结尾激波串自下游向上游传播, 因此, 监测点位置越靠近上游, 其压强出现上升的时刻就越晚。

由图 12 还可知, 对于同一个监测点, 第 1 个振荡周期的压强峰值较高, 自第 2 个振荡周期开始, 压强振荡趋于稳定, 峰值保持在一定值附近, 且较第 1 个振荡周期的压强峰值有所下降。图 14 为进气道沿程 4 个监测点第 1 个振荡周期(t_0～t_1)和第 2 个～第 5 个振荡周期(t_1～t_6)的平均压强峰值对比。可以看出, 沿程 4 个监测点第 2 个～第 5 个振荡周期的平均压强峰值均较第 1 个振荡周期的压强峰值降低, 下降幅度约为 9%～21%。图 15 为进气道的捕获流量系数φ变化曲线(图中流量系数正值表示进气, 气流自外部流向进气道内部; 负值表示排气, 气流自进气道内部流向外部), 可知, 在第 1 个振荡周期内, 进气道的最大捕获流量系数为 0.95, 而自第 2 个振荡周期开始, 进气道捕获流量系数下降, 约为 0.81。进气道捕获流量的下降, 使得压缩存储于内流道的最大气体质量下降, 从而导致了进气道内存储的压力能减小, 压强峰值降低。

此外, 由图 14 还可知, 与试验结果类似, 数值仿真的进气道沿程第 1 个测点 p_1 的压强峰值均较

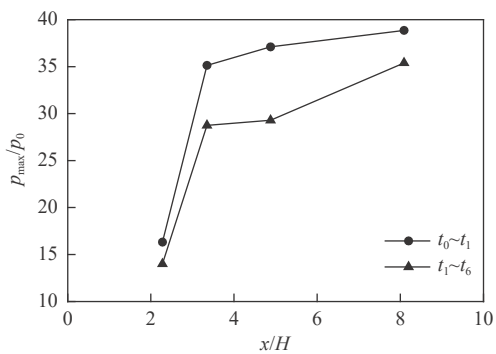


图 14　不同时间段内方案 1 进气道 4 个监测点的压强振荡峰值对比

Fig. 14　Comparison of oscillating pressure peaks at four monitoring points of inlet during different time ranges for option 1

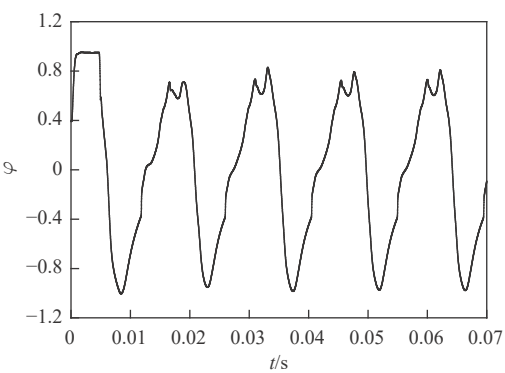


图 15　方案 1 进气道捕获流量系数的时间历程曲线

Fig. 15　Time history of captured flow coefficient of inlet for option 1

低，在 $15p_0$ 左右，第 2 个测点 p_2 的压强峰值急剧上升，其后自第 2 个测点～第 4 个测点（p_2～p_4），进气道的压强峰值变化较小，量值相对接近，约为 p_1 测点压强峰值的 2 倍。这是由于，当进气道出口封闭时，进气道发生了激波振荡现象，在激波回退的过程中，受边界层吸除槽的影响，结尾激波停在了进气道进口附近，此时 p_1 和 p_2 测点分别位于结尾激波的波前和波后，因此，出现了压强的急剧跳变；自 p_2 测点之后，受结尾激波的影响，进气道扩张段内为亚声速流动，气流速度变化较小，因此，压强梯度较小，进气道的压强振荡峰值变化减小；至 p_4 测点，因其位于出口堵盖附近，气流速度几乎为 0，因此，压强振荡峰值达到最大。

图 16 为方案 1 进气道入口截面的平均马赫数时间历程（马赫数为正值时表示进气道为顺流状态，马赫数为负值时表示进气道为倒流状态），图 17 为一个周期内典型时刻的进气道静压等值图谱和平均流线分布，图 17 中的时刻与图 16 中标示的时刻相对应。可知，一个周期内进气道的流动可分为以下 4 个过程。

1）顺流状态，激波前移（t=0.005 2～0.006 2 s）。在逆压梯度的作用下，结尾激波串前移，并与边界层相互作用，在进口附近形成了大规模的流动分离区，并诱导产生了分离激波（t=0.005 2 s）。随着时间的推移，进口分离区逐渐增大并前移，从而使得进气道头部斜激波向外偏斜，激波角增大，溢流增加，至 t=0.006 2 s 时，分离包几乎占据了进气道的整个口面，进气道完全不进气，入口气流速度下降为 0，前后逆压差下降。

2）倒流状态，激波前移（t=0.006 2～0.010 4 s）。由于进气道不再继续进气，因此，其逆压梯度下降。随着时间的推移，头部斜激波向上游运动并脱离压缩面尖点，形成了脱体弯曲激波。至 t=0.010 4 s 时，弯曲激波运动至最上游位置，进气道的倒流马赫数达到最大值，约为 1.5。

3）倒流状态，激波后移（t=0.010 4～0.013 4 s）。随着倒流的发生，内流道的压强下降，前后压强梯

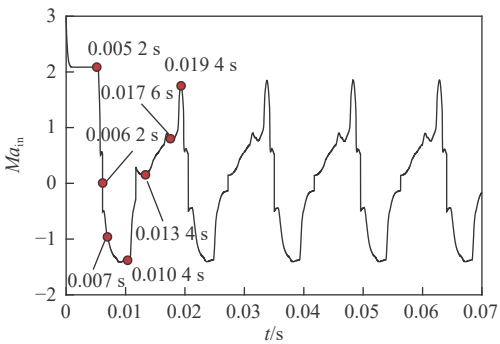


图 16 方案 1 进气道入口截面的平均马赫数时间历程

Fig. 16 Time history of average Mach number of inlet entrance for option 1

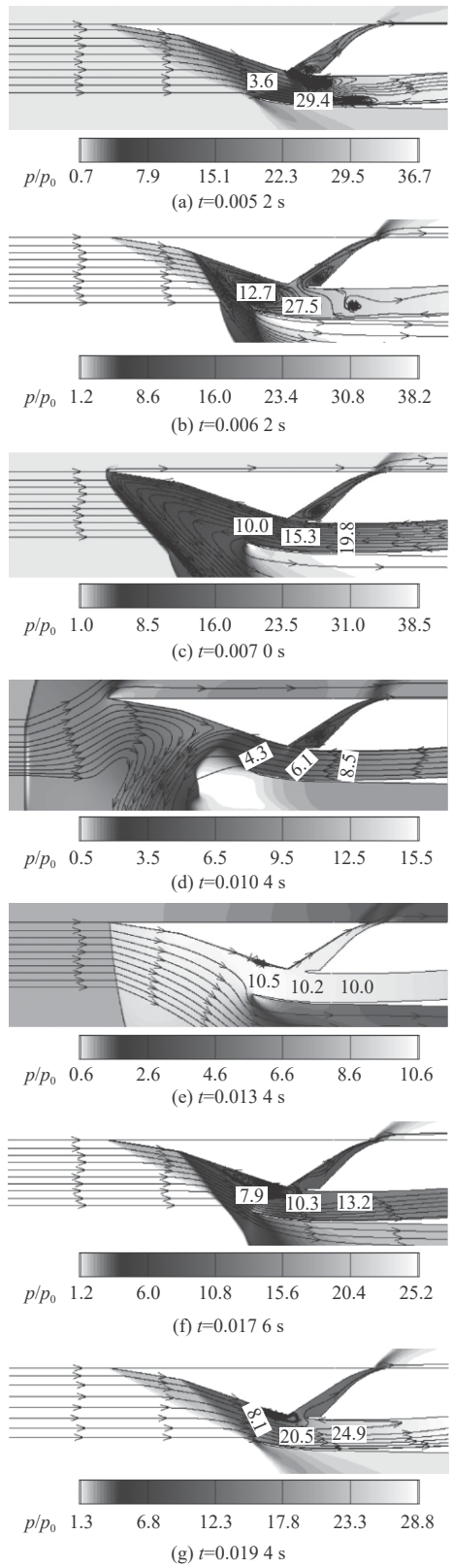


图 17 一个周期内典型时刻方案 1 进气道的静压等值图谱和平均流线分布

Fig. 17 Static pressure contour plots and average streamline distribution of inlet for option 1 at typical moments in one cycle

度变为顺压强梯度, 倒流逐渐减弱, 进气道头部弯曲激波开始向下游运动。至 t=0.013 4 s 时, 弯曲激波后移至进气道压缩面尖点附近, 进气道前后压差和入口马赫数均接近于 0, 进气道入口几乎无气流进出。

4) 顺流状态, 激波后移(t=0.013 4～0.019 4 s)。随着时间的推移, 进气道头部弯曲激波后移, 且受压缩面的影响逐渐衍变为 λ 形激波, 进气道开始重新进气, 与此同时, 在 λ 激波的根部附近形成了小面积局部分离区。在顺压强梯度的作用下, λ 激波逐渐向下游运动, 并衍变为斜激波, 激波角及流动损失减小, 唇口溢流减弱, 进气量增加, 内流道压强逐渐增大, 进气道前后为逆压强梯度。至 t=0.019 4 s 时, 进气道头部斜激波封口, 唇口溢流消失, 进气量达到最大, 进口气流马赫数达到峰值, 约为 1.8。

此后, 随着时间的推移, 进气道的头部波系发生周期性的往复运动, 内流道的压强呈现周期性的振荡现象。

虽然进气道头部斜激波系在回退的过程中仍然达到封口状态, 但由于激波回退过程中形成了流动分离, 并诱导产生了分离激波, 使得进气道的流动损失增大, 因此, 自第 2 个振荡周期开始, 进气道的压强振荡峰值较第 1 个振荡周期有所下降, 这与图 12 和图 14 的分析结果一致。

3.4.2　模型尺寸影响分析

表 5 为数值仿真获得的方案 1 进气道 1:1 全尺寸模型及 1:2.5 缩比模型的扩张段出口位置 p_4 监测点的压强振荡幅值和频率对比, 其中 1:1 全尺寸模型仿真的来流参数按照高度为 15 km 的大气条件设定, 对应于表 2 中的工况 3; 1:2.5 缩比模型仿真的来流参数按照风洞试验条件设定, 对应于表 2 中的工况 2。可以看出, 相同马赫数下, 不同尺寸模型的进气道压强振荡峰值基本一致, 相对偏差在 0.03% 以内, 但二者的频率出现了明显的差异: 其中 1:1 全尺寸模型的压强振荡频率明显要低, 仅为 69 Hz, 而 1:2.5 缩比模型的压强振荡频率为 120.47 Hz。根据式(1)可知, 相同来流马赫数下, 进气道流态相似, 内流道的平均马赫数 $\overline{Ma}$ 相同, 因此, 气流振荡的频率主要与来流声速及内流道的长度相关。由于数值仿真 2 个工况的来流声速及模型尺寸均存在区别, 因此, 二者的频率出现了明显差异。

表 5　不同尺寸模型进气道扩张段出口位置的压强振荡峰值和频率对比

Table 5　Comparison of inlet's oscillating pressure peaks and frequencies at exit of diffuser for models of different sizes

尺寸模型	f/Hz	p_{max}/p_0
1:1(工况3)	69.00	35.41
1:2.5(工况2)	120.47	35.40

根据式(1), 将数值仿真 1:2.5 缩比模型的气流振荡频率转换到 1:1 全尺寸模型:

$$f_1' = f_1 \xi \frac{c_2}{c_1} \tag{2}$$

式中: f_1 为数值仿真工况 2 缩比模型的气流振荡频率; f_1' 为将数值仿真缩比模型转换至与工况 3 相同模型尺度和来流条件下的气流振荡频率; c_1 和 c_2 分别为计算工况 2 和工况 3 的自由来流声速。根据式(2)计算得到工况 2 转换到与工况 3 相同模型尺度和来流条件下的压强振荡频率为 68.52 Hz, 与工况 3 的仿真结果 69 Hz 仅相差 0.70%。可见, 相同来流马赫数下, 模型尺寸不影响进气道盲腔气流振荡的幅值, 仅影响其振荡频率, 不同尺寸模型及来流声速下的压强振荡频率大小可通过模型缩比和来流声速进行转换。

3.4.3　来流马赫数影响分析

对方案 1 进气道在不同来流马赫数条件下开展数值仿真计算(见表 2 中的工况 3～工况 5), 图 18 为不同马赫数下进气道扩张段出口 p_4 监测点的静压时间历程曲线。可以看出, 不同来流马赫数下, 进气道扩张段出口的压强振荡规律相似, 频率相近, 但压强振荡峰值存在明显的差异, 随着马赫数的增大, 压强振荡峰值逐渐增大。Ma=2.6 和 Ma=3.0 条件下, 振荡平稳段扩张段出口的压强峰值分别为 18.98p_0 和 35.40p_0, 二者相差接近 1 倍。

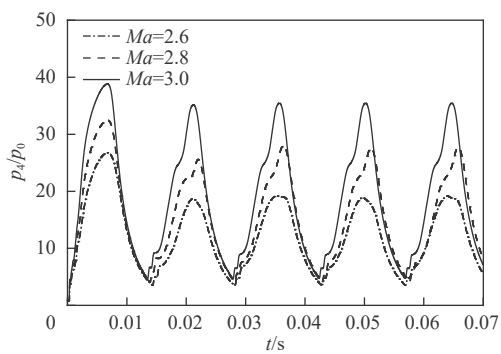


图 18　不同马赫数下, 方案 1 进气道 p_4 监测点的压强时间历程曲线

Fig. 18　Time histories curves of pressure at monitoring point p_4 for option 1 of different Mach numbers

根据式(1)计算不同来流马赫数下的压强振荡频率, 将其与仿真结果进行对比, 结果如表 6 所示。由于来流声速相同, 且不同来流马赫数下的内流道平均马赫数差异较小, 因此, 其气流振荡频率接近, 理论值与仿真值相差在 5% 以内。

3.4.4　边界层吸除影响分析

本文还进一步仿真对比了出口封闭时, 有边界

表 6 不同马赫数下进气道 p_4 监测点的压强振荡频率仿真值与理论值对比

Table 6 Comparison of simulated and theoretical frequencies of inlet's oscillating pressure at monitoring point p_4 for different Mach numbers

Ma	Ma_3	$\overline{Ma}$	c/(m·s^{-1})	理论f/Hz	仿真f/Hz	f相对误差/%
2.6	0.608	0.304	295.07	71.03	69.08	2.75
2.8	0.580	0.290	295.07	71.69	68.44	4.53
3.0	0.556	0.278	295.07	72.22	69.00	4.07

层吸除槽和无边界层吸除槽的进气道气流振荡特性差异，仿真计算状态对应于表 2 中的工况 3 和工况 6。图 19 为方案 1 和方案 2 进气道扩张段出口 p_4 监测点的静压时间历程曲线，表 7 为 2 种方案进气道扩张段出口 p_4 监测点的压强振荡峰值和频率对比。可知，2 种方案进气道的压强振荡变化规律相似，但其压强振荡峰值及频率存在明显的差异。其中，2 种方案在第 1 个振荡周期的压强峰值基本一致，但自第 2 个振荡周期开始，方案 1 进气道的压强振荡峰值较方案 2 高 49.47%。

以下从 2 种方案的流态及捕获流量差异对二者的压强振荡峰值差异进行分析。如图 20 所示，方案 2 进气道无边界层吸除槽，当结尾激波被推出进气道唇口时，气流总压损失增大，因此，头部斜激波在回退的过程中无法再次达到封口的状态，进气道存在较大的溢流。而方案 1 进气道开设了边界层吸除槽，当出口封闭时，内流道静压显著上升，由于内流道与外部大气环境存在明显的压差，因此，

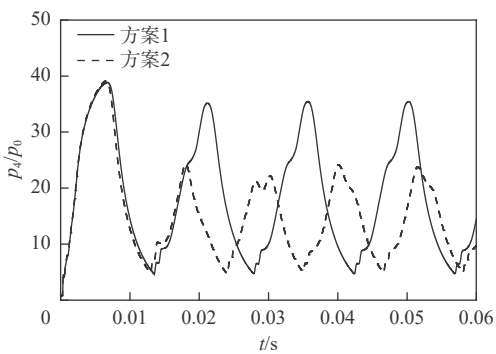


图 19 不同方案进气道 p_4 监测点的压强时间历程曲线

Fig. 19 Time histories of pressure at monitoring point p_4 of inlet for different models

表 7 不同方案进气道 p_4 监测点的压强振荡峰值和频率对比

Table 7 Comparison of inlet's oscillating pressure peaks and frequencies at monitoring point p_4 of inlet for different options

方案	f/Hz	p_{max}/p_0
方案1	69.00	35.41
方案2	88.21	23.69

部分高压气体通过边界层吸除槽进入外部大气环境，即进入进气道入口的气体分成两股：一股经过喉道继续流向进气道尾部，在进气道内流道堆积压缩，导致进气道内流道的静压持续上升；另一股通过边界层吸除槽进入外部大气环境。当结尾激波串传播至进口附近时，进口处流动分离区内的部分低能流通过边界层吸除槽排出内流道，从而导致了方案 1 流动分离区面积较方案 2 减小，进气道头部斜激波系的激波角减小，溢流减弱，进气道内部所能存储的最大气流量较方案 2 增大。

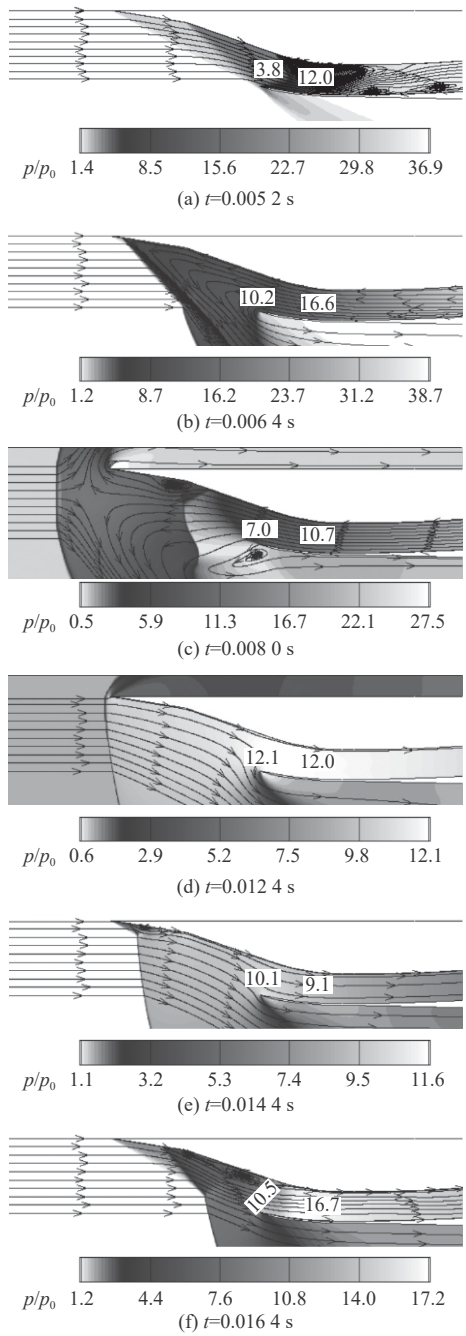


图 20 一个周期内典型时刻方案 2 进气道的静压等值图谱和平均流线分布

Fig. 20 Static pressure contour plots and average streamline distribution of inlet for option 2 at typical moments in one cycle

图 21 为 2 种方案第 2～第 5 个振荡周期的流量系数对比, 图中, 横轴表示时间, 以气流振荡周期 T_s 进行无量纲化。可知, 方案 1 进气道头部斜激波封口, 其进口流量系数最大值为 1, 经过边界层吸除槽后其流量系数下降, 喉道位置的流量系数最大为 0.81, 而方案 2 因进气道头部溢流严重, 其流量系数最大仅为 0.68, 即方案 1 进气道的喉道流量系数较方案 2 偏大约 19.12%。此外, 与方案 2 相比, 因方案 1 进气道头部斜激波封口, 激波角减小, 总压损失较小, 因此, 其进口截面的气流总压更高。受进气道内部存储流量及头部斜激波系损失的共同影响, 方案 1 进气道的压强振荡峰值较方案 2 偏大接近 50%。

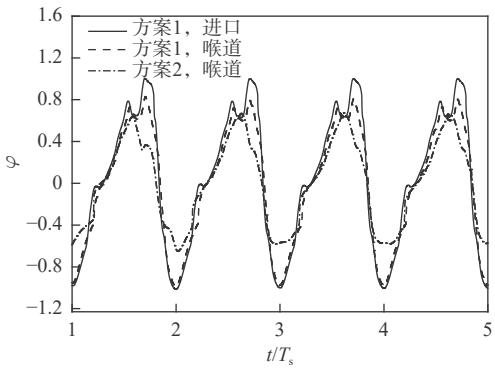


图 21　不同方案进气道的进口和喉道截面流量系数对比

Fig. 21　Comparison of flow coefficient on entrance and throat section of inlet for different options

正是由于方案 1 进气道内部所能存储的最大气流量较方案 2 要大, 因此, 在相同的来流马赫数和姿态角下, 方案 1 进气道的充气时间加长, 其压强振荡周期增大, 压强振荡峰值上升 49.47%, 频率相对方案 2 减小 21.78%。

综上所述, 当出口堵盖封闭时, 进气道内流道的压强振荡峰值随着来流马赫数的增大显著上升, 马赫数为 3.0 时的压强振荡峰值约为马赫数为 2.6 的 2 倍; 进气道的压强振荡峰值随着头部斜激波系向唇口靠近及进气道捕获流量的增大而增大; 在进气道内收缩段开设边界层吸除槽后, 进气道的压强振荡峰值上升 49.47%, 频率下降 21.78%, 实际工程使用时应注意边界层吸除槽对发动机转级过程中气流振荡特性的不利影响及由此带来的进气道结构承载变化问题。

4　结　论

本文采用风洞试验及数值仿真的方法, 对一种超声速混压式二元进气道在出口堵盖封闭条件下的非定常流动特性开展研究, 并分析了模型尺寸、来流马赫数及边界层吸除槽对进气道盲腔气流振荡特性的影响, 得到的结论主要如下。

1) 采用非定常仿真方法可以较好地模拟双侧二元进气道在零姿态角、出口堵盖封闭情况下的内流动, 数值仿真获得的进气道振荡压强与试验值无论从规律上还是数值上均吻合良好, 压强振荡频率及峰值的仿真结果与试验结果相对误差在 3.5% 以内。

2) 采用声学振荡频率计算式计算得到的进气道压强振荡频率与试验值非常接近, 相对误差在 5% 以内; 相同来流马赫数下, 不同模型尺寸及来流声速下的压强振荡频率可通过模型缩比和来流声速进行转换。

3) 随着来流马赫数的增大, 进气道的盲腔气流振荡压强峰值大幅上升, 马赫数为 3.0 时进气道出口的压强振荡峰值约为马赫数为 2.6 的 2 倍。

4) 在进气道内收缩段增加边界层吸除槽, 可以起到边界层吸除和泄压的作用, 使得进气道出口堵盖封闭时, 在激波回退的过程中头部斜激波系仍然可以达到封口状态, 进气道捕获流量上升, 从而导致了气流振荡过程中的压强峰值较无边界层吸除槽的进气道上升 49.47%, 频率下降 21.78%。

参考文献（References）

[1] NISHIZAWA U, KAMEDA M, WATANABE Y, et al. Computational simulation of shock oscillation around a supersonic air-intake[C]//Proceedings of the 36th AIAA Fluid Dynamics Conference and Exhibit. Reston: AIAA, 2006.

[2] NAKAYAMA T, SATO T, AKATSUKA M, et al. Investigation on shock oscillation phenomenon in a supersonic air inlet[C]//Proceedings of the 41st AIAA Fluid Dynamics Conference and Exhibit. Reston: AIAA, 2011.

[3] LU P J, JAIN L T. Numerical investigation of inlet buzz flow[J]. Journal of Propulsion and Power, 1998, 14(1): 90-100.

[4] FUJIWARA H, MURAKAMI A, WATANABE Y. Numerical analysis on shock oscillation of two-dimensional external compression intakes[C]//Proceedings of the 32nd AIAA Fluid Dynamics Conference and Exhibit. Reston: AIAA, 2002.

[5] 王玉峰, 杨宝娥. 超声速进气道喘振的机理研究[J]. 火箭推进, 2008, 34(1): 17-22.
WANG Y F, YANG B E. Study of the buzz mechanism of supersonic inlets[J]. Journal of Rocket Propulsion, 2008, 34(1): 17-22(in Chinese).

[6] OH J Y, MA F H, HSIEH S Y, et al. Interactions between shock and acoustic waves in a supersonic inlet diffuser[J]. Journal of Propulsion and Power, 2005, 21(3): 486-495.

[7] 刘占生, 张云峰, 田新. 冲压发动机超声速进气道流动自激振荡研究[J]. 航空动力学报, 2008, 23(9): 1595-1602.
LIU Z S, ZHANG Y F, TIAN X. Research on self-excited oscillation flows in inlet of ramjet[J]. Journal of Aerospace Power, 2008, 23(9): 1595-1602(in Chinese).

[8] SAUNDERS J, KEITH T. Results from computational analysis of a mixed compression supersonic inlet[C]//Proceedings of the 27th Joint Propulsion Conference. Reston: AIAA, 1991.

[9] SAJBEN M, BOGAR T J, KROUTIL J C. Experimental study of flows in a two-dimensional inlet model[J]. Journal of Propulsion and Power, 1985, 1(2): 109-117.

[10] 李强, 刘佩进, 李江, 等. 冲压发动机助推段压强振荡现象数值分析[J]. 推进技术, 2008, 29(6): 673-676.
LI Q, LIU P J, LI J, et al. Numerical simulation of boosting stage pressure oscillation in ramjet[J]. Journal of Propulsion Technology, 2008, 29(6): 673-676(in Chinese).

[11] 孙振华, 吴催生. 冲压发动机加速阶段进气道内动态特性[J]. 固体火箭技术, 2011, 34(3): 285-289.
SUN Z H, WU C S. Dynamic characteristics in ramjet inlet during acceleration phase[J]. Journal of Solid Rocket Technology, 2011, 34(3): 285-289(in Chinese).

[12] 白晓征, 刘君, 郭正, 等. 冲压发动机进气道压力振荡过程的数值研究[J]. 推进技术, 2008, 29(5): 562-565.
BAI X Z, LIU J, GUO Z, et al. Numerical simulation of pressure oscillation in ramjet inlet[J]. Journal of Propulsion Technology, 2008, 29(5): 562-565(in Chinese).

[13] 刘志伟, 马高建, 崔金平. 附面层吸除对二元混压式进气道起动影响分析[J]. 弹箭与制导学报, 2009, 29(3): 149-152.
LIU Z W, MA G J, CUI J P. Influence of boundary layers bleed sew to start[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2009, 29(3): 149-152(in Chinese).

[14] 贺永杰, 马高建, 刘志伟. 通过附面层泄除提高定几何混压式进气道性能的方法研究[J]. 航空兵器, 2010(2): 28-31.
HE Y J, MA G J, LIU Z W. Study on improving the performance of mixed-compression inlet with fixed-geometry through boundary-layer bleed[J]. Aero Weaponry, 2010(2): 28-31(in Chinese).

[15] HARLOFF G J, SMITH G E. Supersonic-inlet boundary-layer bleed flow[J]. AIAA Journal, 1996, 34(4): 778-785.

[16] SYBERG J, KONCSEK J L. Experimental evaluation of an analytically derived bleed system for a supersonic inlet[J]. Journal of Aircraft, 1976, 13(10): 792-797.

[17] FUJIMOTO A, NIWA N, SAWADA K. Numerical investigation of supersonic inlet with realistic bleed andbypass systems[J]. Journal of Propulsion and Power, 1992, 8(4): 857-861.

[18] HERRMANN D, BLEM S, GULHAN A. Experimental study of boundary-layer bleed impact on ramjet inlet performance[J]. Journal of Propulsion and Power, 2011, 27(6): 1186-1195.

[19] 翁小侪, 郭荣伟. 一种二元定几何混压式超声速进气道流场控制概念研究[J]. 航空动力学报, 2012, 27(11): 2492-2498.
WENG X C, GUO R W. Study of novel flow control concept for fix-geometry two-dimensional mix-compression supersonic inlet[J]. Journal of Aerospace Power, 2012, 27(11): 2492-2498(in Chinese).

[20] WAN D W, GUO R W. Experimental investigation of a fixed-geometry two-dimensional mixed-compression supersonic inlet with sweep-forward high- light and bleed slot in an inverted “X”-type layout[J]. Chinese Journal of Aeronautics, 2007, 20(4): 304-312.

[21] 麻肖妃, 谢旅荣, 郭荣伟. 双下侧布局带泄流腔二元进气道试验[J]. 航空动力学报, 2010, 25(8): 1818-1824.
MA X F, XIE L R, GUO R W. Investigation of two-dimensional supersonic twin inlet with slot-coupled cavity in 90° configuration at venter[J]. Journal of Aerospace Power, 2010, 25(8): 1818-1824(in Chinese).

[22] SCHÜLEIN E. Skin friction and heat flux measurements in shock/boundary layer interaction flows[J]. AIAA Journal, 2006, 44(8): 1732-1741.

Investigation on unsteady flow characteristics of a supersonic inlet with exit blocked

WEN Yufen*, ZHANG Weiqun, HAO Sisi

(China Academy of Launch Vehicle Technology, China Aerospace Science and Technology Corporation, Beijing 100076, China)

Abstract: Cavity flow oscillation phenomenon would occur during the transiton of ramjet with the inlet entrance unobstructed while its exit blocked. As the flow oscillating dramatically, the flight vehicle would face a crisis of instability attitude controlling and structural failure. Due to the problem of fluctuation flow, the unsteady flow characteristics of a supersonic twin-duct inlet with its exit blocked were studied by wind tunnel test and numerical simulation. The effects of the model scale, Mach number of the incoming flow, and boundary layer suction on the characteristics of oscillating pressure of the inlet were acquired. The results indicate that periodic oscillating flow is observed when the exit of the inlet is blocked. The frequency of the oscillating flow is positively correlated to the acoustic velocity of the incoming flow but inversely correlated to the length of the inlet. The oscillating pressure peak is found to approximate the total pressure value of the incoming flow, and it raises obviously with the increment of the Mach number of the incoming flow. The unsteady flow of the inlet with its exit blocked is approximately simulated by the numerical method adopted in this paper, and the numerical simulation result agrees well with that of the wind tunnel test. Furthermore, unsteady simulation results show that the inlet's pressure is relieved in the duration of flow oscillation by conducting the boundary layer suction method at the internal contracted region, which results in a shock-on-lip state during the backward movement of the shock wave system. A rise of captured flow coefficient is observed when compared to the inlet without boundary layer suction, leading to a 49.47% increase of amplitude peak while 21.78% descending of the frequency for the oscillating pressure.

Keywords: supersonic inlet; oscillating pressure; boundary layer suction; unsteady; wind tunnel test

Received: 2023-03-24; **Accepted**: 2023-07-26; **Published Online**: 2023-08-07 17:54

URL: link.cnki.net/urlid/11.2625.V.20230807.1347.001

* **Corresponding author**. E-mail: wyf369963@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0081

基于邻域点特征分割与对应的航空发动机叶片损伤点云配准

蔡舒妤[1,*]，郝峰伟[1]，史涛[2]

(1. 中国民航大学 航空工程学院，天津 300300；2. 湖南南方通用航空发动机有限公司，株洲 412002)

摘 要：为满足损伤压气机叶片点云配准的精度及效率要求，提出一种基于邻域点特征分割与对应的航空发动机叶片损伤点云精确配准算法。以协方差矩阵为基础，通过多步分割模型定义特征子块的分割，获取有效特征区域；根据局部曲率、点最大距离和点最大法向量夹角性质，构建稳定 n 维特征向量；引入迭代最近点理论，构建对应点、对应块间点到面的最小欧式距离，实现叶片损伤点云和叶片标准点云两模型位姿的精确修正；使用单位四元数算法完成对损伤叶片点云的精确配准。实验结果表明：所提算法能完成损伤压气机叶片点云模型表面的点云配准，显著提高了配准的效率和精度，并在多组航空发动机损伤叶片点云数据模型下，验证了基于邻域点特征分割与对应的点云配准的适应性和鲁棒性。

关 键 词：点云配准；特征提取；迭代最近点；损伤检测；压气机叶片
中图分类号：TP39; V232
文献标志码：A **文章编号**：1001-5965(2025)03-0784-11

压气机叶片作为航空发动机的重要核心构件，长期处于高温、高压的极端工况，极易造成诸如掉块、撕裂、压坑、叶尖卷曲等结构损伤[1]，对其结构损伤进行精确检测和精准评估是分析压气机叶片损伤规律、评估结构适航性、保障发动机安全运行的关键。

随着三维传感器工程的技术进步，三维点云模型以其直观、立体的成像方式成为结构损伤分析与检测的新兴媒介，基于点云模型的损伤检测技术也成为新兴的研究热点。在航空发动机离位检测中，由于点云模型本身具有完整几何信息及较高的表面精度，可以通过损伤模型与标准模型的两两配准技术，精确体现实体模型表面的微小曲率畸变，自动提取损伤区域的损伤参数，实现结构损伤的精确检测定位，为基于点云模型的损伤识别提供技术支持。然而，传统三维点云配准无法满足航空发动机压气机叶片的高精度需求，配准结果极易忽略压气机叶片上微小损伤区域，使得叶片表面损伤的高精度检测和评估难以实现，制约了基于三维点云模型在结构损伤分析中的技术发展[2]。因此，研究航空发动机压气机叶片点云模型的配准，保证点云配准的高精度尤为重要。

现阶段，针对损伤叶片点云模型的配准问题，相关学者已经提出了很多解决方案并进行了深入研究，其中，以叶片轮廓线特征数据为配准基准的配准[3-6]尤为高效，其主要利用一条或多条叶面曲线表示整体三维形状，但其非常依赖轮廓线的提取结果，易受叶片前、后缘轮廓曲线边缘等不可靠区域的影响。所以基于几何特征的点云配准的应用更为广泛，如提取点特征、线特征或曲率特征等。在针对叶片点云模型几何特征的研究中，Ding 等[7]基于自适应距离函数和 M 估计，建立了以迭代重加权最小二乘法为基础的自适应距离(iterative reweighted least squares-adaptive distance function,

收稿日期：2023-02-25；**录用日期**：2023-06-28；**网络出版时间**：2024-09-30 15:16
网络出版地址：link.cnki.net/urlid/11.2625.V.20240930.1430.001
*通信作者. E-mail：csy0313@163.com

引用格式：蔡舒妤，郝峰伟，史涛. 基于邻域点特征分割与对应的航空发动机叶片损伤点云配准 [J]. 北京航空航天大学学报，2025，51（3）：784-794. CAI S Y，HAO F W，SHI T. Partition based on features of neighborhood points and corresponding point cloud registration of aero-engine damaged blade[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（3）：784-794（in Chinese）.

IRLS-ADF)非线性优化模型, 提高了叶片配准精度。田吉委等[8]采用多尺度特征描述符与基于空间角度差分阈值的结合, 进一步提高了叶片配准精度。Xie 等[6]提出一种迭代方差最小匹配算法, 利用自定义的平均距离特征, 提高了对差异区域检测的灵敏度。但以点、线和面特征为基础的点云配准, 在进行几何推理时较少考虑点云模型的全局特征, 导致配准适用范围较窄, 同时损伤区域的几何变形也会干扰配准结果, 导致配准精度不高。

因此, 李聪波等[9]将对应点曲率特征与对应点距离特征结合, 实现损伤叶片点云区域的自动剔除, 保证了配准叶片点云对应点的准确性及配准的快速性。Ghorbani 等[10-11]利用叶片模型的局部邻域几何特征, 定义几何不相似函数, 迭代剔除损伤区域, 提高了配准精度, 并利用基于点局部邻域几何属性的对应搜索技术, 提出一种受损叶片配准法, 从配准结果中去除受损区域的影响。赵红壮等[12]采用分阶段对叶片的配准法, 采用自配准法对每部分进行配准后, 对相邻两部分采用改进的配准算法进行整体配准, 提高了配准效率。王文博等[13]以提取分块叶片点云特征点为核心、块状配准点云重叠率为约束, 在设定的数量阈值内对点云进行分块, 提高了运行效率。Li 等[14]提取了点云模型的 3 个不同位姿平面, 利用各个平面间的多个空间特征, 多次修正模型位姿, 实现损伤点云模型和标准点云模型两模型高精度全局配准。

针对模型位姿的优化, 目前应用较为广泛的精配准是 Schenker[15]提出的迭代最近点(iterative closest point, ICP)算法, 该算法对点云的位置进行变化以达到配准的目的, 但是对初始位置依赖严重, 极易陷入模型位姿局部最优。为解决该问题, Choi 等[16]提出用 k 维数(k-dimensional tree, KD-Tree)改进 ICP 算法, 通过加快搜索最近点的速度, 提高了配准效率。Yang 等[17]提出全局最优 Go-ICP 算法, 采用分支定界算法对 ICP 算法的目标函数进行了全局最优的优化。Bergström 和 Edlund[18]通过求解具有拟议约束点到平面距离的最小化问题, 提高了配准效率。王宾等[19]在精确配准阶段提出基于双向距离比例的 ICP 算法, 提高了配准精度。石峰源等[20]采用主成分分析的改进 ICP 算法, 缩短了配准时间。

在基于上述损伤叶片点云配准算法的探索中, 当前仍存在一些亟待解决的问题, 如: 现有配准对初始位置依赖严重, 极易造成局部最优, 造成配准精度低; 在特征提取阶段, 存在点特征提取不全, 进而配准适用范围较窄; 同时易受叶片损伤区域的影响, 获取的点云模型存在数据缺失, 从而干扰配准结果, 导致配准效率不高。本文以航空发动机压气机损伤叶片为研究对象, 提出一种基于邻域点特征分割与对应的损伤点云精确配准算法。首先, 以协方差矩阵为基础, 设计以多步分割的特征区域提取; 其次, 基于局部曲率、点最大距离和点最大法向量夹角等几何特征, 构建特征区域的稳定点 n 维特征向量, 获取正确的点特征对应关系; 最后, 引入迭代最近点理论, 通过迭代计算, 实现两模型位姿的精确修正, 提高配准的效率和精度。

1　基于多步分割的特征区域提取

结合航空发动机叶片外物损伤规律分析[1]可知, 压气机叶片损伤大多发生在叶面、叶根及叶根以上区域, 而底座榫头区域发生损伤的概率非常小, 相对其他区域在配准过程中更为可靠。为充分利用该可靠区域, 人工将叶片点云初步分割为叶面区域及底座榫头区域, 在基本排除损伤区域对叶片点云配准影响后, 以榫头区域模型为数据输入, 基于协方差矩阵计算, 引入区域生长法与主成分分析法将底座榫头区域进行二次分割, 提取特征明显区域的点集, 以提高配准效率和配准精度。

1.1　特征区域提取

设某点云为 $P\subset\mathbf{R}^3$, 存在 $p=\{\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_i, \cdots, \boldsymbol{p}_n\}$, 其中, $\boldsymbol{p}_i=[x_i, y_i, z_i]^{\mathrm{T}}$ 为第 i 个点的坐标, 在其半径 r 内构建球型邻域空间, 利用该邻域内的点集合 $N(\boldsymbol{p}_i)=\{\boldsymbol{p}_{ij}|\boldsymbol{p}_{ij}\in P, |\boldsymbol{p}_i\boldsymbol{p}_{ij}|<r\}$, 构建协方差矩阵 $\mathrm{Cov}(\boldsymbol{p}_i)$:

$$\begin{cases}\mathrm{Cov}(\boldsymbol{p}_i)=\dfrac{1}{k}\displaystyle\sum_{i=0}^{k}(\boldsymbol{p}_{ij}-\overline{\boldsymbol{p}})(\boldsymbol{p}_{ij}-\overline{\boldsymbol{p}})^{\mathrm{T}}\\ \overline{\boldsymbol{p}}=\dfrac{1}{k}\displaystyle\sum_{i=1}^{k}\boldsymbol{p}_{ij}\end{cases}\tag{1}$$

$$\mathrm{Cov}(\boldsymbol{p}_i)\boldsymbol{v}_{ij}=\lambda_{ij}\boldsymbol{v}_{ij}\qquad j\in\{0,1,2\}\tag{2}$$

$$\begin{cases}\sigma_i=\dfrac{\lambda_{i\min}}{\mathrm{sum}\{\lambda_{i0},\lambda_{i1},\lambda_{i2}\}}\\ \lambda_{i\min}=\min\{\lambda_{i0},\lambda_{i1},\lambda_{i2}\}\end{cases}\tag{3}$$

式中: $\overline{\boldsymbol{p}}$为邻域点 $\boldsymbol{p}_{ij}$ 的质心; λ_{ij} 为邻域点 $\boldsymbol{p}_{ij}$ 所对应协方差矩阵的特征值; $\boldsymbol{v}_{ij}$ 为特征值 λ_{ij} 所对应的特征向量; σ_i 为该任意点 $\boldsymbol{p}_i$ 的曲率, 用来估计局部区域平滑度。当特征值取最小时, 所对应的特征向量则可作为该点的法向量方向, 对该特征向量进行归一化后得该点的法向量, 构建协方差矩阵。

设某点云 P 上存在任意点 $\boldsymbol{p}_i$ 及其相邻点 $\boldsymbol{p}_{i+1}$, 通过协方差矩阵计算所得曲率为 σ_i、σ_{i+1}, 最小特征值所对应的特征向量为 $\boldsymbol{v}_i$、$\boldsymbol{v}_{i+1}$, 对其单位化处理后得对应法向量 $\boldsymbol{n}_i$、$\boldsymbol{n}_{i+1}$:

$$\begin{cases} \boldsymbol{n}_i = \dfrac{\boldsymbol{v}_i}{|\boldsymbol{v}_i|} \\ \boldsymbol{n}_{i+1} = \dfrac{\boldsymbol{v}_{i+1}}{|\boldsymbol{v}_{i+1}|} \end{cases} \tag{4}$$

由于模型曲面变化程度不同，因此，相邻两点法向量必有角度差异，且差值越大曲面变化越剧烈。计算两向量间夹角值，设定法向量夹角阈值为 ε_1，计算曲率变化量，设定曲率阈值为 ε_2，其表达式为

$$\begin{cases} \dfrac{\boldsymbol{n}_i \cdot \boldsymbol{n}_{i+1}}{|\boldsymbol{n}_i| \cdot |\boldsymbol{n}_{i+1}|} < \varepsilon_1 \\ |\sigma_i - \sigma_{i+1}| = \varepsilon_2 \end{cases} \tag{5}$$

为区分无特征点及有特征点，选择夹角阈值小于 ε_1=0.996，曲率阈值小于 ε_2=0.005[21] 范围内的点。对曲率进行排序后，设定其上曲率最小值点作为种子点，利用 k 维数搜索当前种子点的相邻点，当种子点法向量与相邻点法向量夹角大于 ε_1 且种子点曲率与相邻点曲率变化量大于 ε_2 时，将该点保留，完成特征点云的提取。

1.2 模型区域分割

为提高配准效率及精度，需对 1.1 节获取的特征点云进行区域分割，利用主成分分析法获得多个互不相交的特征子块，以满足后续特征计算需求。

对具有 N 个点的点云 P，其点云质心 $\overline{\boldsymbol{p}}=[x_o, y_o, z_o]^{\mathrm{T}}$：

$$\overline{\boldsymbol{p}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{p}_i = \left[\frac{1}{N}\sum_{i=1}^{N}x_i \quad \frac{1}{N}\sum_{i=1}^{N}y_i \quad \frac{1}{N}\sum_{i=1}^{N}z_i\right]^{\mathrm{T}} \tag{6}$$

求该点云的散度矩阵并进行特征分解：

$$S(\boldsymbol{p}) = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{p}_i - \overline{\boldsymbol{p}})(\boldsymbol{p}_i - \overline{\boldsymbol{p}})^{\mathrm{T}} \tag{7}$$

得到特征值及特征向量，并对特征值从大到小排序得 λ_u、λ_v、λ_w，其对应的特征向量为 $\boldsymbol{\eta}_u$、$\boldsymbol{\eta}_v$、$\boldsymbol{\eta}_w$。

具主成分分析原理可知，将 $\boldsymbol{\eta}_u=[x_u, y_u, z_u]^{\mathrm{T}}$ 定义为 u 轴方向，即第一主成分是三维点云方差最大方向；将 $\boldsymbol{\eta}_v=[x_v, y_v, z_v]^{\mathrm{T}}$ 定义为 v 轴方向，即第二主成分是与第一主成分正交方差最大方向；将 $\boldsymbol{\eta}_w=[x_w, y_w, z_w]^{\mathrm{T}}$ 定义为 w 轴方向，即第三主成分是与第一、第二主成分平面正交方差的方向。

以重心坐标为原点，特征向量为轴方向，构建模型重心点坐标系向量为 $\boldsymbol{L}_p$，其三维直角坐标为

$$\begin{cases} \boldsymbol{x}_p = (x_u, y_u, z_u) + (x_o, y_o, z_o) \\ \boldsymbol{y}_p = (x_v, y_v, z_v) + (x_o, y_o, z_o) \\ \boldsymbol{z}_p = (x_w, y_w, z_w) + (x_o, y_o, z_o) \end{cases} \tag{8}$$

获得点云数据点在其重心处的向量 $\boldsymbol{p}_i=[x_i, y_i, z_i]^{\mathrm{T}}$，并计算与该坐标系坐标轴的夹角余弦值：

$$\begin{cases} \cos\theta_x = (\boldsymbol{p}_i \cdot \boldsymbol{x}_p) / (|\boldsymbol{p}_i| \cdot |\boldsymbol{x}_p|) \\ \cos\theta_y = (\boldsymbol{p}_i \cdot \boldsymbol{y}_p) / (|\boldsymbol{p}_i| \cdot |\boldsymbol{y}_p|) \\ \cos\theta_z = (\boldsymbol{p}_i \cdot \boldsymbol{z}_p) / (|\boldsymbol{p}_i| \cdot |\boldsymbol{z}_p|) \end{cases} \tag{9}$$

依据 cos 值的正负对点云进行空间划分，若 $\cos\theta_x>0$、$\cos\theta_y>0$、$\cos\theta_z>0$，则该点为第Ⅰ象限内点；若 $\cos\theta_x<0$、$\cos\theta_y>0$、$\cos\theta_z>0$，则该点为第Ⅱ象限内点，以此类推，获得多块不相交点云子块。

通过第 1.1 节和第 1.2 节所述可以精确提取榫头的特征区域，并对特征区域进行子块分割，以满足后续精确配准需求。

2 应用特征对应块的点云配准

2.1 数据点 n 维特征向量构建

由于局部特征描述的局限性及单一特征描述的不稳定性，为使点云数据点具有区分度，更能提供准确的结构信息，本文从全局点云着手，选取点高斯曲率、点法向最大旋转角及点最大欧式距离等 3 个点特征，评估点云表面特征。

对给定点云 P，其上存在任意点 $\boldsymbol{p}_i=[x_i, y_i, z_i]$ 及其邻域点集合 $N(\boldsymbol{p}_i)=\{\boldsymbol{p}_{ij}|\boldsymbol{p}_{ij}\in P\}$，由式(1)～式(3)可得，该点云 $\boldsymbol{p}_i$ 的法向量为 $\boldsymbol{n}_i$。

1)点高斯曲率构建。利用相邻点的法向量估计单个点的高斯曲率，用来描述点云表面区域的平滑度，如图 1 所示。

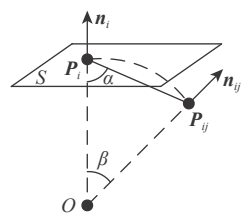


图 1 法曲率局部拟合示意图

Fig. 1 Schematic diagram of local fitting of normal curvature

设 $\boldsymbol{p}_{ij}$ 法向量为 $\boldsymbol{n}_{ij}$，则 $\boldsymbol{p}_i$ 相对于 $\boldsymbol{p}_{ij}$ 的法曲率为

$$k_m^i = -\frac{\sin\beta}{|\boldsymbol{p}_i\boldsymbol{p}_{ij}|\sin\alpha} \tag{10}$$

式中：α 为 $\boldsymbol{n}_i$ 与向量 $\boldsymbol{p}_i\boldsymbol{p}_{ij}$ 间的夹角；β 为向量 $\boldsymbol{n}_i$ 与向量 $\boldsymbol{n}_{ij}$ 间的夹角；$|\boldsymbol{p}_i\boldsymbol{p}_{ij}|$ 为 $\boldsymbol{p}_i$、$\boldsymbol{p}_{ij}$ 间的欧式距离。

利用欧拉公式，法曲率与主曲率有以下关系：

$$k_m^i = k_1\cos^2(\theta_i+\theta) + k_2\sin^2(\theta_i+\theta) \tag{11}$$

式中：i=1, 2, …, m；k_1 和 k_2 为主曲率，$(\theta_i+\theta)$ 为点 $\boldsymbol{p}_i$ 过 $\boldsymbol{p}_{ij}$ 的法截线的切线与主方向的夹角。

利用最小二乘求解主曲率，获得主曲率 k_1 和 k_2：

$$\min_{k_1,k_2,\theta}\sum_{i=0}^{m}\left(k_1\cos^2(\theta_i+\theta) + k_2\sin^2(\theta_i+\theta) - k_m^i\right)^2 \tag{12}$$

则第 1 个尺度-点高斯曲率为 $k_1\times k_2$：

$$\begin{cases} F_1(\boldsymbol{p}_i)=k_{\boldsymbol{p}_i} \\ k_{\boldsymbol{p}_i}=k_1\times k_2 \end{cases} \tag{13}$$

2)点法向最大旋转角构建。利用特征区域内点的法向量估计单个点的法向最大旋转角的角度值, 用来描述点云表面变化特征, 如图 2 所示。

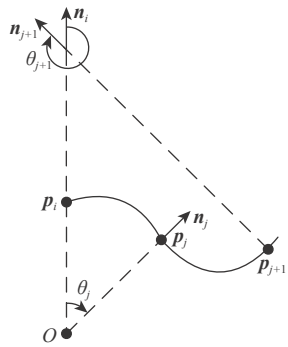


图 2　点法向最大旋转角示意图

Fig. 2　Schematic diagram of maximum rotation angle normal to a point

设特征区域内点 $\boldsymbol{p}_j$ 的法向量为 $\boldsymbol{n}_j$, 则任意点 $\boldsymbol{p}_i$ 相对于 $\boldsymbol{p}_j$ 的法向夹角为

$$\theta_j=\arccos\left(\boldsymbol{n}_i\cdot\boldsymbol{n}_j/|\boldsymbol{n}_i|\cdot|\boldsymbol{n}_j|\right) \tag{14}$$

但其结果在 (0～π) 内, 需要判定正负来确定角的方向, 统一所求角度值在 (0～2π), 则计算两向量叉乘:

$$\boldsymbol{n}_i\times\boldsymbol{n}_j=|\boldsymbol{n}_i|\cdot|\boldsymbol{n}_j|\cdot\sin\theta_j \tag{15}$$

若 $\boldsymbol{n}_i\times\boldsymbol{n}_j<0$, 那么 $\theta_j=2\pi-\theta_j$, 则第 2 个尺度-点法向量最大夹角为

$$\begin{cases} F_2(\boldsymbol{p}_i)=\theta_{\boldsymbol{p}_i} \\ \theta_{\boldsymbol{p}_i}=\max\left\{\theta_1,\theta_2,\cdots,\theta_j,\cdots,\theta_n\right\} \end{cases} \tag{16}$$

3) 点最大欧式距离构建。利用特征区域内点间距离估计单个点的最大欧式距离, 用来描述点云几何尺寸特征, 如图 3 所示。

设特征区域内点为 $\boldsymbol{p}_j=[x_j, y_j, z_j]$, 则与点 $\boldsymbol{p}_i$ 间的欧式距离 d_j 为

$$d_j=\sqrt{\left(x_i-x_j\right)^2+\left(y_i-y_j\right)^2+\left(z_i-z_j\right)^2} \tag{17}$$

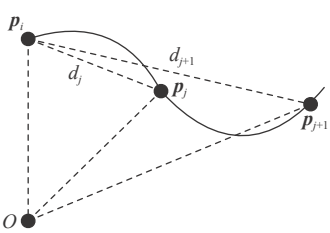


图 3　点最大欧氏距离示意图

Fig. 3　Schematic diagram of maximum Euclidean distance between points

取第 3 个尺度-点最大欧式距离为

$$\begin{cases} F_3(\boldsymbol{p}_i)=d_{\boldsymbol{p}_i} \\ d_{\boldsymbol{p}_i}=\max\left\{d_1,d_2,\cdots,d_j,\cdots,d_n\right\} \end{cases} \tag{18}$$

综上所述, 这 3 类特征不会随点云位姿的改变而改变, 具有旋转平移不变性, 因此, 构建该点 $\boldsymbol{p}_i$ 的 n 维特征向量为

$$\boldsymbol{F}(\boldsymbol{p}_i)=\left[F_1(\boldsymbol{p}_i),F_2(\boldsymbol{p}_i),F_3(\boldsymbol{p}_i)\right]^{\mathrm{T}}=\left[k_{\boldsymbol{p}_i},\theta_{\boldsymbol{p}_i},d_{\boldsymbol{p}_i}\right]^{\mathrm{T}} \tag{19}$$

2.2　点特征对应关系获取

为提升基于数据点 n 维特征向量配准的准确性及可靠性, 提出一种以特征对应点对为基础, 构建多组特征对应块关系的。

设待配准点云为 $P=\{\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_n\}$, 其所对应的 n 维特征向量集合为 $P_{\boldsymbol{F}}=\{\boldsymbol{F}(\boldsymbol{p}_1), \boldsymbol{F}(\boldsymbol{p}_2), \cdots, \boldsymbol{F}(\boldsymbol{p}_n)\}$; 目标点云为 $Q=\{\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_m\}$, 其所对应的 n 维特征向量集合为 $Q_{\boldsymbol{F}}=\{\boldsymbol{F}(\boldsymbol{q}_1), \boldsymbol{F}(\boldsymbol{q}_2), \cdots, \boldsymbol{F}(\boldsymbol{q}_m)\}$, 其中, $\boldsymbol{F}(\boldsymbol{q}_j)=[F_1(\boldsymbol{q}_j), F_2(\boldsymbol{q}_j), F_3(\boldsymbol{q}_j)]^{\mathrm{T}}$, n, m 分别为待配准点云、目标特征点云的点数, $\boldsymbol{p}_i$ 为待配准点云上的任意点, $\boldsymbol{q}_j$ 为目标特征点云上的任意点, $\boldsymbol{q}_i$ 为待配准点云任意点在目标特征点云上的对应点。

1)初步点对应关系构建。计算两点云中点 $\boldsymbol{p}_i$ 与点 $\boldsymbol{q}_j$ 特征向量间的欧式距离:

$$\begin{cases} d_j=\sqrt{\sum_{k=1}^{3}\left(F_k(\boldsymbol{p}_i)-F_k\left(\boldsymbol{q}_j\right)\right)^2} \\ D_{\boldsymbol{p}_i}=\left\{d_1,d_2,\cdots,d_j,\cdots,d_m\right\} \end{cases} \tag{20}$$

通过 $\min\{D_{\boldsymbol{p}_i}\}$获取该点 $\boldsymbol{p}_i$ 的最小特征向量欧式距离, 并利用最小欧式距离获取所对应的点 $\boldsymbol{q}_i$, 以此类推, 建立两点云块各点间的最近对应关系点对, 可表示为

$$k_1=\{(\boldsymbol{p}_i,\boldsymbol{q}_i)|\boldsymbol{p}_i\in P,\boldsymbol{q}_i\in Q,i=1,2,\cdots,N_i\} \tag{21}$$

2) 错误点对应关系删除。对 k_1 的最小距离集进行距离检测, 筛选距离过大的对应点对, 即错误匹配对应点对, 则删除此类匹配点, 以保证配准的准确性。

利用 K-Means 聚类法, 通过迭代计算, 最终保留最大的聚类集, 建立新的对应关系:

$$k_2=\left\{\left(\boldsymbol{p}_j,\boldsymbol{q}_j\right)\left|\boldsymbol{p}_j\in P,\boldsymbol{q}_j\in Q,j=1,2,\cdots,N_j\right.\right\}\quad N_j<N_i \tag{22}$$

3)特征块对应关系构建。为进一步提高配准精度, 计算各特征子块的对应点关系, 构建子块间的对应关系。则互不相交的特征子块分别为 $\{P_i\}_{i=1}^{M}$、$\{Q_i\}_{i=1}^{N}$, M、N 表示子块的划分数目, 则获取的对应块可表示为

$$K=\{(P_i,Q_i)|P_i\in\varphi_1,Q_i\in\varphi_2,i=1,2,\cdots,M_i\} \tag{23}$$

最终，通过对 K 使用单位四元数算法[22]，获取一系列对应子块旋转矩阵 $\boldsymbol{R}_i(i=1,2,\cdots,M)$和平移矩阵 $\boldsymbol{T}_i(i=1,2,\cdots,M)$，利用点云重叠度估计算法选出重叠度最高的旋转矩阵 $\boldsymbol{R}_i$ 和平移矩阵 $\boldsymbol{T}_i$，最终得到最优变换矩阵。

2.3 对应模型配准位姿修正

对于点云 Q，根据2.2节理论获取的旋转矩阵和平移矩阵，通过变换得到 $\boldsymbol{Q}'$，如式(24)所示，将 P 和 $\boldsymbol{Q}'$作为输入点云。

$$\boldsymbol{Q}' = \boldsymbol{R}_i Q + \boldsymbol{T}_i \tag{24}$$

为得到精度更高的配准结果，本文使用点到面ICP算法进行位姿修正，具体步骤如下：

步骤1 输入待配准点云 P 和 $\boldsymbol{Q}'$，设定终止阈值 η 和最大迭代次数(本文设定迭代次数为500)；

步骤2 对于 $\boldsymbol{Q}'$中的任意点，利用KD-Tree在 P 中寻找对应点所在微切平面的欧式距离最近点，得到对应点对，直到遍历 $\boldsymbol{Q}'$；

步骤3 用单位四元数算法计算对应点对的旋转矩阵 $\boldsymbol{R}_k$ 和平移矩阵 $\boldsymbol{T}_k$，然后计算目标函数：

$$E_k = \min\left\{\frac{1}{n}\sum_{i=1}^{n}\left\|\left[(\boldsymbol{R}_k\cdot\boldsymbol{p}_i+\boldsymbol{T}_k)-\boldsymbol{q}_i\right]\cdot\boldsymbol{n}_i\right\|^2\right\} \tag{25}$$

步骤4 将步骤3的旋转矩阵 $\boldsymbol{R}_k$ 和平移矩阵 $\boldsymbol{T}_k$ 应用于 $\boldsymbol{Q}'$，计算 $\boldsymbol{Q}'_{k+1}=\boldsymbol{R}_k\boldsymbol{Q}'_k+\boldsymbol{T}_k$，用 $\boldsymbol{Q}'_{k+1}$ 代替 $\boldsymbol{Q}'_k$；

步骤5 计算步骤3中相邻两目标函数的变化率，若 $E_{k+1}-E_k<\eta$ 或达到设定最大迭代次数时，终止迭代，否则转到步骤2。

3 叶片损伤点云配准流程

叶片损伤点云配准流程，如图4所示，基于点云模型配准的损伤精确检测步骤如下。

步骤1 点云特征区域提取。将待配准点云 P 及点云 Q 作为输入点云，通过式(4)和式(5)得到输入点云内各点法向量间的夹角 ε_1 及曲率 ε_2，当 $\varepsilon_1>0.996$ 且 $\varepsilon_2>0.005$ 时，则特征点被区分，特征区域被提取；

步骤2 特征点云子块分割。根据主成分分析法原理，通过式(6)和式(8)计算输入点云重心点处的向量 $\boldsymbol{p}_i=[x_i, y_i, z_i]^{\mathrm{T}}$ 及其重心坐标系向量 $\boldsymbol{L}_p$，并获取重心点向量与该坐标系各坐标轴向量间的夹角余弦值，若 $\cos\theta_x>0$、$\cos\theta_y>0$、$\cos\theta_z>0$，则该点为第Ⅰ象限内点，以此类推，获得多块不相交的特征点云子块。

步骤3 点 n 维特征向量计算。在某特征点云子块中，通过式(13)、式(16)和式(18)分别计算与各数据点对应的 n 维特征向量，构建稳定的数据点 n 维特征向量。

步骤4 对应特征点云子块获取。通过式(20)和式(21)，在剔除错误点对应关系后，构建新点对应关系，并基于该对应关系获取多组对应特征点云子块。

步骤5 最优变换矩阵获取。对获取的多组对应子块使用单位四元数算法，得到一系列对应子块间的旋转平移矩阵，再通过点云重叠度估计法筛选重叠度最高的最优变换矩阵。

步骤6 配准位姿修正。引入改进最近迭代算法，通过式(25)获取迭代目标函数的最优解，在最优变换矩阵的基础上，计算最终的变换矩阵，修正待配准点云的配准位姿，实现损伤模型的精确配准。

步骤7 对应点云拓扑关系构建。为进一步精确检测损伤叶片上的损伤区域的大小，提取损伤区域，引入八叉树索引结构，并基于该树拓扑结构的叶子节点步长 d_t，分别构建两点云的拓扑关系，依次关系构建两点云间的点对应关系。

步骤8 损伤区域提取。计算待配准点云 P 上的任意数据点 $\boldsymbol{p}_i$ 与其对应点云 Q 上的最邻近点 $\boldsymbol{q}_i$ 间的距离 d_s：

$$d_s = \sqrt{(x_p-x_q)^2+(y_p-y_q)^2+(z_p-z_q)^2} \tag{26}$$

当 $d_s>d_t$ 时，则该数据点为损伤数据点，该区域为损伤区域。

步骤9 损伤参数计算。以损伤区域的损伤边界为平面基准，计算该损伤区域的损伤参数，通过损伤跨度、损伤深度叶尖卷曲弦长占比等参数[23-24]来衡量损伤程度，最终可视化损伤叶片点云模型的损伤区域。

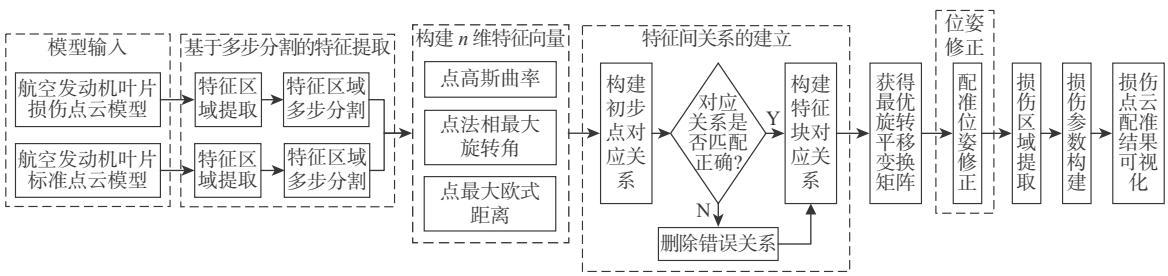


图4 叶片损伤点云配准流程

Fig. 4 Flow of point cloud registration of damaged blade

4　实例验证分析

为验证本文算法的健壮性与可行性, 本文算法在 Visual studio 2017 环境下, Intel Core i5-9400F 2.90 GHz CPU、8 GB 内存的 Windows10 64 位 PC 机上进行实验。实验采用 Handyscan 3D 结构光三维扫描仪获取的多组不同损伤、标准航空发动机压气机叶片点云模型作为数据输入。

4.1　数据点特征提取对比实验

为证明构造 n 维数据点特征向量的准确性及有效性, 本文实验采用扫描获取的两点云点量相同 RB-211 型航空发动机 2 级高压气机叶片点云模型作为数据输入, 将 n 维特征向量与点高斯曲率、点法向最大旋转角、点最大欧氏距离及各组合特征向量进行对比, 如图 5 和图 6 所示, 并讨论几种特征向量的计算效率及配准结果的影响。

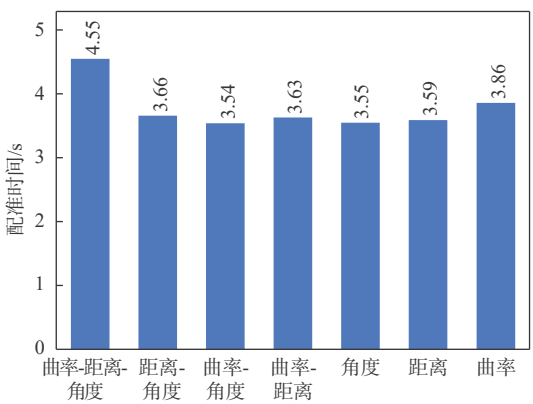


图 5　配准时间

Fig. 5　Registration time

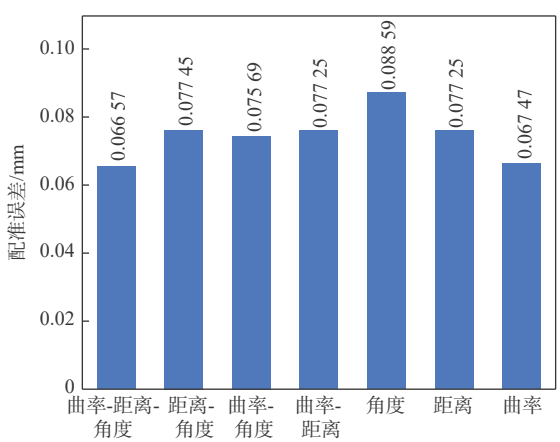


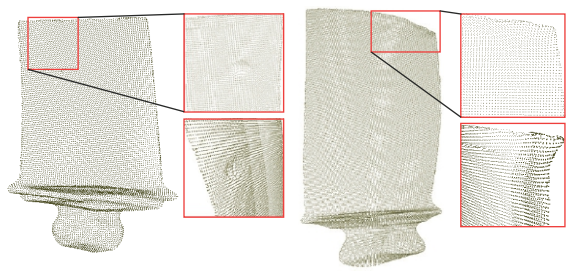
图 6　配准误差

Fig. 6　Registration error

由图 5 可知, 在相同模型的相同位姿状态下, 7 种描述符的粗配准时间均相差少于 1.01 s, 证明 3 个维度描述符的计算效率都很高, 本文提出的 n 维特征向量的时间计算复杂度低; 同时, 如图 6 所示, 每种描述符的配准误差相差 0.022 02 mm 以内, 而混合特征描述符的误差最小, 证明了特征混合的有效性。

4.2　实体损伤叶片配准比较

本文实验为进一步验证本文算法的有效性, 采用扫描仪获取的 CFM 56-7B 航空发动机 6 级、2 级的损伤压气机叶片及其标准压气机叶片点云模型作为数据输入, 叶片损伤位置及形状如图 7 所示, 为更好表述损伤位置, 右下角小图是通过旋转模型得到的精细位置。具体叶片点云模型参数及损伤数据如表 1 所示, 其中, 6 级叶片的损伤特征为叶面凹坑, 2 级叶片的损伤特征为叶尖卷曲。2 级叶片的损伤模型损伤的弦长占比为 42.04。



(a) 6 级叶片-凹坑损伤　　(b) 2 级叶片-叶尖卷曲损伤

图 7　损伤叶片

Fig. 7　Damaged blades

表 1　损伤叶片实验数据

Table 1　Experimental data of damaged blades

点云点量				损伤尺寸/mm			
6级叶片		2级叶片		6级叶片		2级叶片	
标准模型	损伤模型	标准模型	损伤模型	跨度	深度	跨度	
21 398	21 445	76 189	83 731	2.367	0.632	5.531	

为充分体现本文算法拥有较高的精度及配准效率, 本文实验设置 2 组损伤叶片配准, 在不同位姿(轴向旋转 45°、平移 $x_{50\ mm}$-$y_{50\ mm}$)、不同点云模型、不同损伤条件下进行实验。并与经典 ICP 算法、一阶改进最小二乘优化 ICP(levenberg marquardt and ICP, LM-ICP)算法及位姿调整 ICP (trimmed ICP, Tr-ICP) 算法、采样一致性正态分布(sample consensus initial alignment and norm aldistribution transform, SAC-IA NDT)算法、样本初始对齐二阶 ICP(简称 SAC-IA ICP-ICP2)改进算法和快速点云配准(graph for fast point cloud registration, GROR)算法和初始配准(简称 FPFH)算法进行对比, 配准结果及配准数据如图 8 和图 9 所示, 配准数据如表 2 和表 3 所示, 本文算法均保持在 10^{-5} 阶精度。评价指标为配准总耗时和配准误差 ξ_{RMSE}:

$$\xi_{\mathrm{RMSE}}=\sqrt{\frac{1}{N_{\mathrm{p}}}\sum_{i=1}^{N_{\mathrm{p}}}\left\|\left[(\boldsymbol{R}\boldsymbol{p}_i+\boldsymbol{T})-\boldsymbol{q}_i\right]\cdot\boldsymbol{n}_i\right\|^2}\tag{27}$$

由图 8 可知, 对于不同重叠位姿的压气机叶片点云模型, 除传统 ICP、LM-ICP、SAC-IA NDT 算法

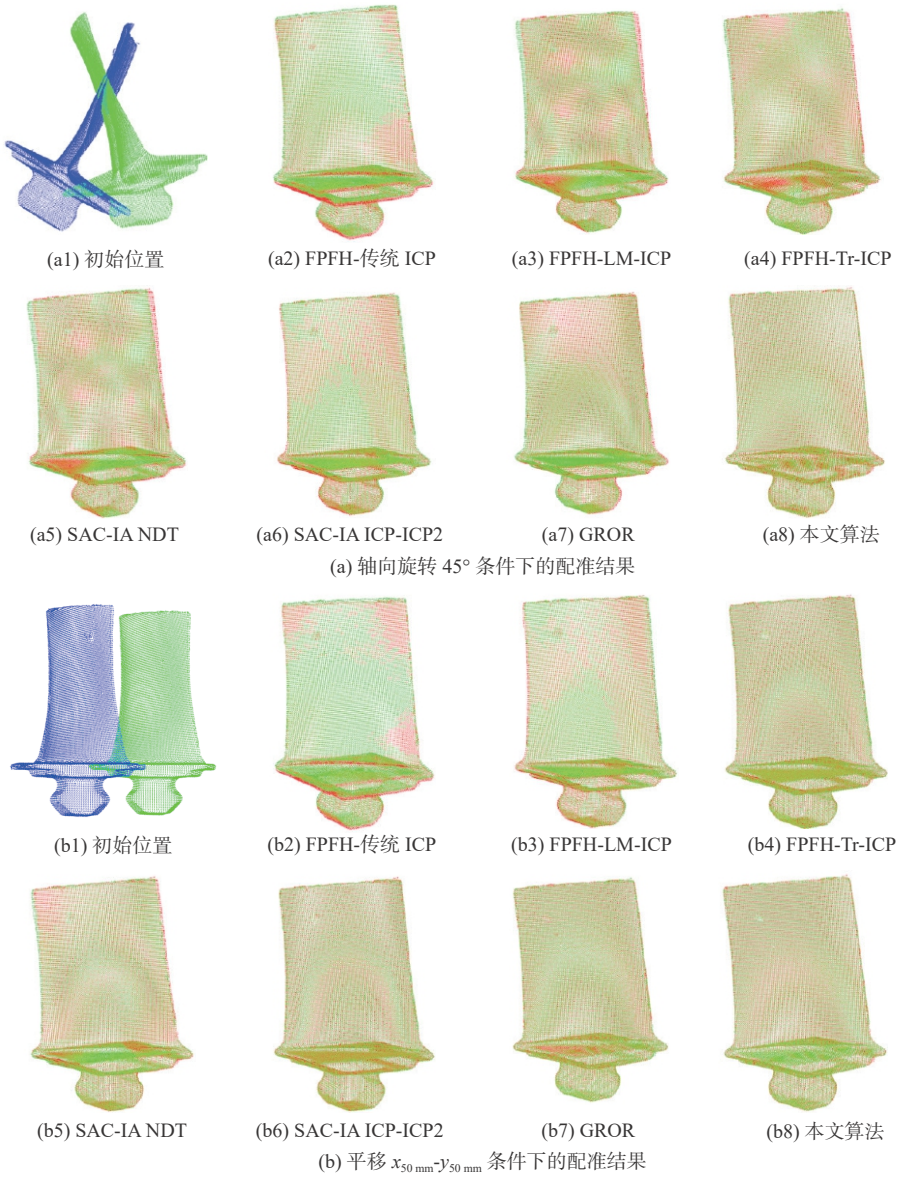


图 8 CFM 56-7B 发动机 6 级压气机叶片凹坑损伤配准效果

Fig. 8 Registration effect of pit damage of 6-stage compressor blade of CFM 56-7B engine

有较大的误差，其他几种算法配准效果较好，基于视觉效果，基本无模型位姿的偏差。据表 2 实验结果分析，传统算法的平均配准误差相对改进算法相差较大，平均配准误差降低 2～3 个量级。

由表 2 可知，在损伤检测方面，传统 ICP、LM-ICP、Tr-ICP 及 SAC-IA NDT 等算法，由于存在配准位姿偏移问题，导致无法精确检测到叶片的损伤区域，损伤检测结果不可信；而改进类算法配准位姿得到了准确对正，更符合损伤检测的精度要求，SAC-IA ICP-ICP2 算法的检测精度为 2.11×10^{-3}，GROR 算法的将测精度为 4.52×10^{-2}，本文算法的检测精度为 8.44×10^{-4}，可以实现对叶面平缓区域小损伤的较准确识别与提取，但因改进策略不同，SAC-IA ICP-ICP2 算法及 GROR 算法的配准效率相对较低，本文算法在效率上更占优势。

配准误差方面，相比于传统 ICP、LM-ICP、Tr-ICP、SAC-IA NDT 算法，本文算法的平均配准误差降低 2～3 个量级。

配准效率方面，相比于传统 ICP 算法，本文算法的配准时间平均降低 23.19% 和 32.96%；与 SAC-IA NDT 算法相比，本文算法的配准时间平均降低 41.18% 和 41.72%；与 LM-ICP 算法相比，本文算法的配准时间平均降低 77.47% 和 85.34%；与 Tr-ICP 算法相比，本文算法的配准时间平均降低 56.72% 和 60.72%；与 SAC-IA ICP-ICP2 算法相比，本文算法的配准时间平均降低 72.49% 和 86.99%；与 GROR 算法相比，本文算法的配准时间平均降低 43.11% 和 36.20%。

综合配准精度及配准效率，与 SAC-IA ICP-ICP2 算法和 GROR 算法相比，在保持配准精度的同时进一步提高了配准效率，在叶片的表面损伤检测应用中，本文算法相对更具优势。

根据图 9 和表 3，对于不同损伤、不同点量的压

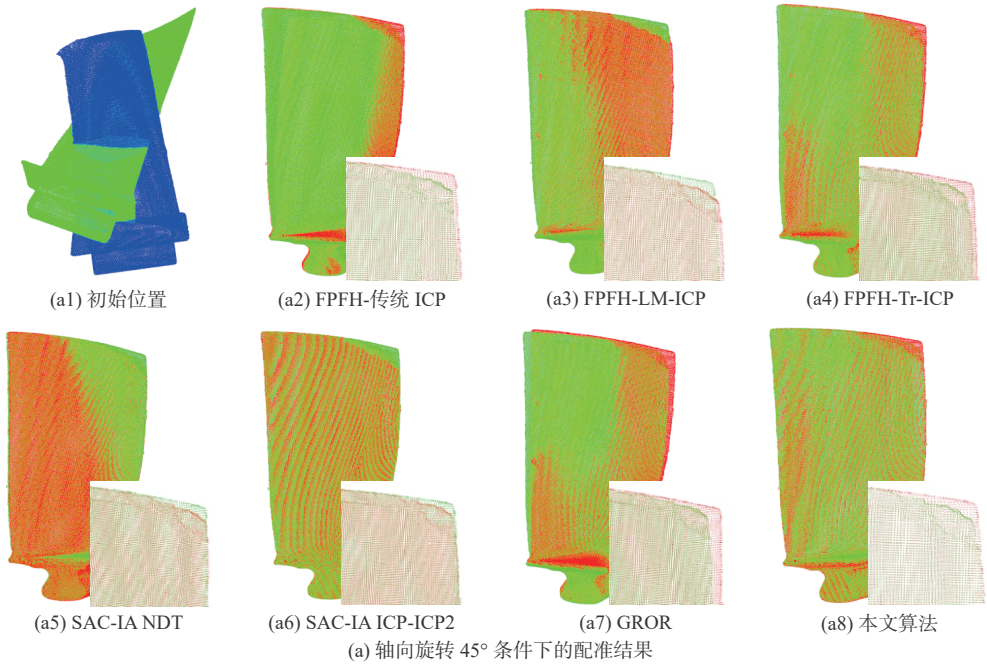


图 9　CFM 56-7B 发动机 2 级压气机叶尖弯曲损伤配准效果

Fig. 9　Registration effect of tip bending damage of 2-stage compressor of CFM 56-7B engine

表 2　凹坑损伤配准误差分析

Table 2　Error analysis of pit damage registration

算法	平均配准误差/mm		配准时间度量/s		损伤尺寸/mm			
	轴向旋转	平面位移	轴向旋转	平面位移	轴向旋转		平面位移	
					跨度	深度	跨度	深度
传统ICP	0.87×10^{-2}	0.89×10^{-2}	55.66	68.15	1.561	0.548	1.559	0.545
LM-ICP	0.93×10^{-2}	0.95×10^{-2}	189.74	311.57	1.491	0.527	1.490	0.527
Tr-ICP	0.85×10^{-2}	0.82×10^{-2}	98.78	116.32	1.494	0.490	1.498	0.492
SAC-IA NDT	0.88×10^{-2}	0.89×10^{-2}	72.68	78.41	1.551	0.537	1.551	0.539
SAC-IA ICP-ICP2	0.98×10^{-4}	0.89×10^{-5}	155.41	351.26	2.362	0.631	2.362	0.631
GROR	0.97×10^{-4}	0.98×10^{-5}	75.14	71.62	2.260	0.629	2.361	0.630
本文算法	0.89×10^{-5}	0.46×10^{-5}	42.75	45.69	2.363	0.631	2.365	0.632

表3 叶尖弯曲损伤配准误差分析

Table 3 Error analysis of tip bending damage registration

算法	平均配准误差/mm		配准时间度量/s		损伤尺寸/mm		占比	
	轴向旋转	平面位移	轴向旋转	平面位移	轴向旋转	平面位移	轴向旋转	平面位移
传统ICP	0.12×10^{-2}	0.15×10^{-2}	509.44	488.25	4.878	4.854	36.53	36.09
LM-ICP	0.126 5	0.161 3	571.62	566.05	4.701	4.625	33.84	32.49
Tr-ICP	0.251 6	0.263 6	572.68	504.06	4.505	4.578	30.85	30.87
SAC-IA NDT	0.56×10^{-3}	0.44×10^{-3}	3 040.76	3 544.07	4.981	5.054	38.27	38.39
SAC-IA ICP-ICP2	0.17×10^{-3}	0.16×10^{-3}	1 370.33	1 425.95	5.204	5.255	39.34	39.49
GROR	0.381 6	0.355 6	102.53	116.22	3.319	3.242	30.44	31.82
本文算法	0.13×10^{-5}	0.39×10^{-5}	260.44	347.25	5.529	5.530	41.94	42.00

气机叶片点云模型，在配准带有较大损伤的叶片模型时，除SAC-IA NDT、SAC-IA ICP-ICP2算法存在较少的配准错误，LM-ICP、Tr-ICP、GROR等算法都存在较大的配准误差。同时，传统ICP算法不受验证条件的影响，配准精度保持较小浮动，不受损伤变形大小的影响，但算法仍无法适用于叶片表面损伤检测。

由表3可知，在损伤检测方面，由于LM-ICP、Tr-ICP及GROR等算法存在较大的配准误差，导致此类算法的损伤检测精度急剧下降，无法精确检测到叶片的损伤区域，损伤检测结果不可信；SAC-IA NDT算法、SAC-IA ICI-ICP2算法的配准精度也受到较大影响，此类受较大点云变形区域影响，也无法保证损伤检测的精度；而本文算法的配准精度不受模型损伤区域的影响，同时配准位姿得到了准确对正，损伤检测精度为1.81×10^{-4}，可以实现对叶尖卷曲损伤的准确识别与提取，在损伤检测中具有较大优势，可以较好适用于损伤叶片表面的损伤检测。

配准误差方面，相比于传统ICP、SAC-IA NDT、SAC-IA ICP-ICP2等算法，本文算法的平均配准误差降低2～3个量级；相比于LM-ICP、Tr-ICP、GROR等算法，本文算法的平均配准误差降低5个量级；

配准效率方面，相比传统ICP算法，本文算法的配准时间平均降低48.87%和28.88%；与LM-ICP算法相比，配准时间平均降低54.44%和38.65%；与Tr-ICP算法相比，配准时间平均降低54.52%和31.10%；与SAC-IA NDT算法相比，配准时间平均降低91.43%和90.20%；与SAC-IA ICP-ICP2算法相比，配准时间降低80.99%和75.65%。

综合配准精度及配准效率，对于较大点云点量及较大损伤区域的点云模型实例验证，本文算法具有较大优势，由此证明，本文算法在模型较大形变下依旧具有配准普适性。

5 结 论

本文提出一种基于邻域点特征分割与对应的损伤点云精确配准算法，其具有良好的鲁棒性和普适性。

1）在配准适应性方面，本文算法可以配准较常见的损伤叶片点云模型，如叶面凹坑以及叶尖弯曲等；同时，配准结果不受压气机叶片损伤、叶片型号及叶片层级的影响，精度及效率均满足损伤检测的要求，为基于点云模型的损伤识别提供技术支持。

2）在配准效果方面，本文算法配准精度不受叶片损伤程度及点云点量的影响，具有较高鲁棒性，保持在10^{-5}阶精度；本文配准效率随点云点量的增长而增长，与其他算法相比，配准效率至少提升23.19%，运算效率在可接受范围内。

参考文献（References）

[1] 舒畅, 程铭, 许煜, 等. 航空发动机压气机叶片外物损伤规律研究[J]. 机械工程学报, 2019, 55(13): 87-94.
SHU C, CHENG M, XU Y, et al. Study on foreign object damage regular pattern of aero engine compressor blades[J]. Journal of Mechanical Engineering, 2019, 55(13): 87-94(in Chinese).

[2] 赵江洪, 孙铭悦, 王殷瑞, 等. 三维点云孔洞修复综述[J]. 测绘科学, 2021, 46(1): 114-123.
ZHAO J H, SUN M Y, WANG Y R, et al. Review of three-dimensional point cloud completion methods[J]. Science of Surveying and Mapping, 2021, 46(1): 114-123(in Chinese).

[3] PENG W X, WANG Y N, ZHANG H, et al. Stochastic joint alignment of multiple point clouds for profiled blades 3-D reconstruction[C]//Proceedings of the IEEE Transactions on Industrial Electronics. Piscataway: IEEE Press, 2022: 1682-1693.

[4] 蔺小军, 吴刚, 单秀峰, 等. 基于叶片截面线CMM测量数据的ICP配准改进算法[J]. 机械工程学报, 2020, 56(2): 1-8.
LIN X J, WU G, SHAN X F, et al. An improved ICP registration algorithm based on CMM measurement data of blade section line[J]. Journal of Mechanical Engineering, 2020, 56(2): 1-8(in Chinese).

[5] 刘峻峰, 何小妹, 黄翔, 等. ICP算法在叶型点云数据配准中的应用[J]. 航空制造技术, 2019, 62(12): 79-82.
LIU J F, HE X M, HUANG X, et al. Application of ICP algorithm in data registration of blade profile point cloud[J]. Aeronautical Manufacturing Technology, 2019, 62(12): 79-82(in Chinese).

[6] XIE H, LI W L, YIN Z P, et al. Variance-minimization iterative matching method for free-form surfaces—part I: theory and method[C]//Proceedings of the IEEE Transactions on Automation Science and Engineering. Piscataway: IEEE Press, 2019: 1181-1191.

[7] DING J, LIU Q, SUN P P. A robust registration algorithm of point clouds based on adaptive distance function for surface inspection[J]. Measurement Science and Technology, 2019, 30(7): 075003.

[8] 田吉委, 王耀南, 毛建旭, 等. 复杂曲面机器人三维测量的误配准筛除算法[J]. 仪器仪表学报, 2019, 40(3): 85-92.
TIAN J W, WANG Y N, MAO J X, et al. Misregistration screening algorithm for 3D measurement of complex surface robots[J]. Chinese Journal of Scientific Instrument, 2019, 40(3): 85-92(in Chinese).

[9] 李聪波, 肖卫洪, 杜彦斌, 等. 基于改进 ICP 算法的损伤零部件精确配准[J]. 计算机集成制造系统, 2016, 22(04): 1021-1028.
LI C B, XIAO W H, DU Y B, et al. Fine registration method for defective parts based on improved ICP algorithm[J]. Computer Integrated Manufacturing Systems, 2016, 22(04): 1021-1028(in Chinese).

[10] GHORBANI H, KHAMENEIFAR F. Accurate registration of point clouds of damaged aeroengine blades[J]. Journal of Manufacturing Science and Engineering, 2021, 143(3): 031012.

[11] GHORBANI H, KHAMENEIFAR F. Airfoil profile reconstruction from unorganized noisy point cloud data[J]. Journal of Computational Design and Engineering, 2021, 8(2): 740-755.

[12] 赵红壮, 刘瑾, 杨海马, 等. 面向航空损伤叶片点云的分阶段配准研究[J]. 激光技术, 2023, 47(2): 241-246.
ZHAO H Z, LIU J, YANG H M, et al. Research on staged registration of point clouds for aeronautical damaged blades[J]. Laser Technology, 2023, 47(2): 241-246(in Chinese).

[13] 王文博, 田茂义, 俞家勇, 等. 改进的迭代最近点点云配准[J]. 激光与光电子学进展, 2022, 59(2): 0215007.
WANG W B, TIAN M Y, YU J Y, et al. Improved iterative nearest point point cloud alignment method[J]. Laser & Optoelectronics Progress, 2022, 59(2): 0215007(in Chinese).

[14] LI R B, YUAN X P, GAN S, et al. A point cloud registration method based on dual quaternion description with point-linear feature constraints[J]. International Journal of Remote Sensing, 2022, 43(7): 2517-2537.

[15] SCHENKER P. Sensor fusion IV: control paradigms and data structures[C]//Proceedings of the Society of Photo-Optical Instrumentation Engineers. Boston: SPIE, 1992.

[16] CHOI W S, KIM Y S, OH S Y, et al. Fast iterative closest point framework for 3D LIDAR data in intelligent vehicle[C]//Proceedings of the IEEE Intelligent Vehicles Symposium. Piscataway: IEEE Press, 2012: 1029-1034.

[17] YANG J L, LI H D, CAMPBELL D, et al. Go-ICP: a globally optimal solution to 3D ICP point-set registration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(11): 2241-2254.

[18] BERGSTRÖM P, EDLUND O. Robust registration of surfaces using a refined iterative closest point algorithm with a trust region approach[J]. Numerical Algorithms, 2017, 74(3): 755-779.

[19] 王宾, 刘林, 侯榆青, 等. 应用改进迭代最近点的三维心脏点云配准[J]. 光学 精密工程, 2020, 28(2): 474-484.
WANG B, LIU L, HOU Y Q, et al. Three-dimensional cardiac point cloud registration by improved iterative closest point method[J]. Optics and Precision Engineering, 2020, 28(2): 474-484(in Chinese).

[20] 石峰源, 张春明, 姜丽辉, 等. 采用主成分分析的迭代最近点算法优化与验证[J]. 激光与光电子学进展, 2022, 59(22): 2211001.
SHI F Y, ZHANG C M, JIANG L H, et al. Optimization and verification of iterative closest point algorithm using principal component analysis[J]. Laser & Optoelectronics Progress, 2022, 59(22): 2211001(in Chinese).

[21] WANG J, HUO S L, LIU Y J, et al. Research of fast point cloud registration method in construction error analysis of hull blocks[J]. International Journal of Naval Architecture and Ocean Engineering, 2020, 12: 605-616.

[22] 李宇翔, 郭际明, 潘尚毅, 等. 一种基于 ISS-SHOT 特征的点云配准算法[J]. 测绘通报, 2020(4): 21-26.
LI Y X, GUO J M, PAN S Y, et al. A point cloud registration algorithm based on ISS-SHOT features[J]. Bulletin of Surveying and Mapping, 2020(4): 21-26(in Chinese).

[23] 张玉燕, 渠文涵, 郭全丽, 等. 基于 3D 点云平均曲率估算的损伤参数计算[J]. 计量学报, 2018, 39(5): 609-614.
ZHANG Y Y, QU W H, GUO Q L, et al. Damage parameter calculation based on 3D point cloud mean curvature estimation[J]. Acta Metrologica Sinica, 2018, 39(5): 609-614(in Chinese).

[24] 徐照, 李苏豪, 陈楠, 等. 基于点云的建筑物表面损伤三维重建与属性提取[J]. 土木工程与管理学报, 2018, 35(3): 28-33.
XU Z, LI S H, CHEN N, et al. 3D reconstruction and attribute extraction of building superficial defects based on point-cloud[J]. Journal of Civil Engineering and Management, 2018, 35(3): 28-33(in Chinese).

Partition based on features of neighborhood points and corresponding point cloud registration of aero-engine damaged blade

CAI Shuyu[1,*], HAO Fengwei[1], SHI Tao[2]

(1. College of Aeronautical Engineering, Civil Aviation University of China, Tianjin 300300, China;
2. Hunan General Aviation Engine Co., Ltd., Zhuzhou 412002, China)

Abstract: To satisfy the requirements on accuracy and efficiency of point cloud registration of damaged compressor blades a algorithm for partition based on features of neighborhood points and the corresponding accurate point cloud registration of aero-engine damaged blades was proposed. First of all, based on the covariance matrix, a multi-step partition model was employed to define the method to divide feature sub-blocks, and thus obtain effective feature regions. Secondly, a stable n-dimensional feature vector was constructed in accordance with the local curvature, the maximum distance between points, and the angle property of the maximum normal vector; then, by introducing the iterative closest point theory, the minimum Euclidean distance between the corresponding points and that from the point to the surface between the corresponding blocks were established. The accurate position correction of the two models was realized. Finally, the unit quaternion algorithm was used to complete the accurate point cloud registration of damaged blades. Experimental results show that the proposed algorithm can achieve point cloud registration on the surface of the point cloud model of damaged compressor blades, significantly improving the efficiency and accuracy of registration. Moreover, the advantages and robustness of the unit quaternion algorithm are verified through the point cloud database of multiple groups of aero-engine damaged blades.

Keywords: point cloud registration; feature extraction; iterative closest point; damage detection; compressor blade

Received: 2023-02-25; **Accepted**: 2023-06-28; **Published Online**: 2024-09-30 15:16
URL: link.cnki.net/urlid/11.2625.V.20240930.1430.001

* **Corresponding author**. E-mail: csy0313@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0129

基于MC-GERT的舰载机出动回收作业流程时间不确定性分析

郭放[1]，韩维[1]，刘玉杰[1,2]，苏析超[3,*]，白天[4]，刘纯[5]

(1. 海军航空大学 航空基础学院，烟台 264001；2. 军队政治工作研究院，北京 100071；3. 海军航空大学 航空作战勤务学院，烟台 264001；4. 中国船舶工业系统工程研究院，北京 100094；5. 洪都航空工业集团650所，南昌 330024)

摘 要：舰载机出动回收作业完成时间通常呈现较大的不确定性，直接影响航母任务规划人员对舰载机后续任务方案的制定。为确定舰载机出动回收作业流程完成时间的分布规律，辅助规划人员掌控舰载机任务执行进度，提出舰载机出动回收作业流程时间不确定性分析方法。基于舰载机出动回收作业流程，建立舰载机出动回收作业流程随机网络模型；提出以Laplace变换为传递函数的图示评审技术网络，对其传递函数特性进行研究，证明了舰载机出动回收作业流程随机网络模型可以运用信号流图理论进行求解，并引入蒙特卡罗方法用于求解随机网络中带有“与”型节点关系的并联支路简化及提升网络求解的运算效率。基于结合蒙特卡罗的图示评审技术（MC-GERT）方法，分别对舰载机飞行前准备作业、出动离场作业、波次出动离场作业及循环作业模式下的周期作业的完成时间的分布特性进行分析，验证了所提方法的可行性，表明该方法能够应用于对舰载机出动回收作业规划方案进行在线分析，为舰载机作业规划相关部门提供一定的方案执行预期。

关 键 词：舰载机；出动回收作业；MC-GERT；Laplace变换；蒙特卡罗；不确定性分析
中图分类号：V271.4; V355.2
文献标志码：A **文章编号**：1001-5965(2025)03-0795-11

航母是一个国家军事力量和水平的重要象征，其搭载的大量舰载机延伸了陆基飞机的打击范围，提高了远海作战能力，丰富了作战手段。航母的作战能力主要通过出动多种类型舰载机完成各类作战任务来体现，因而如何提升舰载机作战效能成为提高航母作战能力的一项关键课题。围绕这一课题，研究人员将舰载机出动回收作业流程分为甲板作业、出动离场、空中任务和回收着舰4个部分，并进行了较为充分的研究。

围绕如何制定高效可行的舰载机甲板保障作业规划方案和资源配置方案，合理规划舰载机甲板作业流程和资源分配，学者对舰载机甲板规划作业的研究先后经历了从单机规划[1-2]向多机规划[3]、从机务保障规划[4]向机务勤务保障协同规划[5-10]、从多站式规划[11-12]向集中式[13]和一站式规划[14]、从静态保障规划[8-14]向鲁棒保障规划[15-17]和动态保障规划[18-20]的发展与革新。

面向如何提高舰载机出动效能，当前研究主要聚焦在如何规划舰载机甲板滑行路径和优化舰载机出动离场作业流程2个方面。针对舰载机甲板滑行路径规划的研究进展，Wang等[21]进行了详细的综述。针对舰载机出动离场作业流程优化，Cui等[22]将出动离场规划问题分解为起飞位选择和各起飞位上的飞机排序两阶段决策过程，分别设计规划决策规则和自动规划方法；Liu等[23]基于舰载机出动离场作业流程建立混合流水车间调度问题模型，考

收稿日期：2023-03-15；录用日期：2023-04-23；网络出版时间：2023-06-28 19:22
网络出版地址：link.cnki.net/urlid/11.2625.V.20230628.1323.002
基金项目：国家自然科学基金(62003366)
*通信作者. E-mail：suxich@126.com

引用格式：郭放，韩维，刘玉杰，等. 基于MC-GERT的舰载机出动回收作业流程时间不确定性分析[J]. 北京航空航天大学学报，2025，51(3)：795-805. GUO F，HAN W，LIU Y J，et al. Time uncertainty analysis on cyclic operation procedures of carrier aircraft based on MC-GERT[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：795-805 (in Chinese).

虑了多机甲板滑行协同轨迹规划问题，利用文献 [24] 中的方法构建了舰载机出动离场滑行路径库，并设计双层遗传算法对问题求解；万兵等[25] 在建立混合流水车间调度模型的基础上，引入间隔变量和逻辑约束，建立了一种约束规划模型，设计一种多机出动离场规划启发式规则将多机规划问题转化为单机规划问题，并提出一种单机约束引导启发式搜索与约束规划二分法迭代算法对舰载机出动离场流程进行优化。

针对如何提高舰载机回收效能，规划返航舰载机着舰序列，Wu 等[26] 设计了一种基于蚁群优化算法的求解方法，并用来解决回收着舰调度问题，同时分析了 2 种复飞加入着舰航线的方法；刘玉杰等[27-28] 针对不同的回收着舰模式，分别提出一种改进人工蜂群算法和改进灰狼优化算法，并将其应用到舰载机回收着舰作业规划过程中，并分析了复飞对着舰完成时间的影响。

以上各项研究在一定程度上为提升舰载机作战效能提供了理论参考，能够为甲板调度人员提供一个切实可行的出动回收作业。但是，在舰载机飞行前准备作业和着舰作业中单项作业执行过程通常呈现一定的概率分布[15-17,29-30]，且甲板设备可能发生故障[31]，其次，舰载机执行空中任务时间较长且战时易受实时性因素影响（例如突发性敌情）。因此，舰载机出动回收作业流程存在较大的时间不确定性[31]，导致甲板调度人员执行当前研究得到的调度方案无法得到一个科学、合理的作业完成时间预期。

本文针对当前研究中舰载机出动回收作业流程存在的不足，进行了以下研究：①基于结合蒙特卡罗的图示评审技术 (Monte Carlo graphical evaluation and review technique, MC-GERT) 对舰载机出动回收作业流程中的时间不确定性进行分析，将舰载机出动回收作业流程建模为 GERT 模型。②针对基于矩母函数的 GERT 网络和以特征函数为传递函数的 GERT 网络的不足，定义了以 Laplace 变换为传递函数的 GERT 网络，并对其运算特征进行了分析。③给出了基于 MC-GERT 的舰载机出动回收作业流程时间不确定性分析流程。④通过案例仿真分析，研究了出动回收作业流程中舰载机周期作业完成时间的概率密度分布特性。本文研究能够为舰载机调度人员提供一个合理的效能预期，对于舰载机出动回收作业调度方案的分析评估具有一定的理论参考意义。

1 舰载机出动回收 GERT 模型

构建一个客观系统 GERT 模型的目的在于将系统用网络的形式表达，使得原系统得以简化，进而研究系统中各要素间的相互关系，得到客观系统的各种特性。本节介绍 GERT 基本理论，描述舰载机出动回收作业流程，并建立舰载机出动回收 GERT 模型。

1.1 GERT 基本理论

随机网络技术是系统工程一个重要分支，其是在具有随机性的环境条件和系统内部随机因素共同作用下对系统的一种网络化描述。GERT 是一种对随机网络中的随机变量进行分析的网络计划技术。自 Pritsker 和 Happ[32] 提出 GERT 以来，其在可靠性管理、项目管理、供应链管理、再制造管理等领域中得到了广泛应用，其研究进展详见文献 [33]。

GERT 网络基本要素由节点、概率分支和传递参数的分布等构成。网络中除去网络开始的源节点和网络结束的终节点外，其余任一节点均具有“输入”和“输出”两端，根据输入、输出端逻辑类型可以将 GERT 网络中的节点分为如表 1 所示的 6 种类型。

表 1 GERT 网络节点类别

Table 1 Taxonomy of nodes in GERT network

输出端	输入端		
	“异或”型	“或”型	“与”型
概率型			
肯定型			

网络中发生状态转移时，某一节点依照不同的概率转移到其他可能的节点，从而引出了不同的概率分支。该特性导致了网络运行过程的随机性。

网络状态转移过程中发生的传递关系表现为传递参数的变化。传递参数通常由时间、成本、规模或其他一些客观系统关键参数组成[34]。两节点之间的传递参数通常服从一定的概率密度分布。

1.2 舰载机出动回收作业流程

舰载机是航母编队的核心作战力量，其作战效能的发挥与舰载机性能和综合保障能力密切相关。舰载机性能在其列装之后基本固定，难以提升，因而提升舰载机综合保障能力是实际应用过程中提升航母编队作战效能的唯一途径。

舰载机综合保障能力主要体现在舰载机出动回收作业过程中的各个阶段。为提高舰载机作战效能，多个波次舰载机通常以循环作业模式交替执行空中任务，图 1 为一个两波次循环作业示意图。

图 1 中的单周期作业，主要包括飞行前准备、出动离场、空中任务、回收着舰 4 个阶段。飞行前

准备阶段，由机务勤务保障人员执行起飞前的检查维护、充气、加油、通电和挂弹等一系列保障工序并提供该过程中相应的供给性资源[9]，以确保舰载机具有可出动执行各类作战任务的能力。出动离场作业阶段，由调度人员安排合理的出动顺序，确定出动起飞位，明确舰载机从保障停机位的滑出时刻，确保出动舰载机高效离场。空中任务阶段，舰载机通常以编队形式执行各类任务。回收着舰阶段，调度人员安排舰载机着舰序列，确保着舰安全，并提高舰载机回收效率[27]。

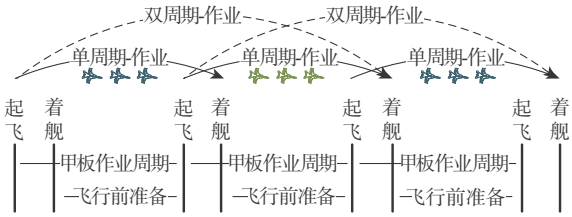


图 1　舰载机连续出动及其循环作业示意图

Fig. 1　Schematic diagram of continuous sortie and cyclic operations of carrier aircraft

1.3　舰载机出动回收 GERT 模型

舰载机出动回收作业流程，可以视为舰载机随着时间的推移在各类作业状态之间的转移过程，该过程可以通过有向 GERT 网络表示，网络中的节点表示舰载机作业状态，节点之间的概率分支表示状态之间的传递关系。

舰载机执行空中任务过程中可能发生战损情况。假设甲板上的舰载机刚好满足任务需求，战损舰载机需要经过维修后继续执行下一波次任务。结合如图 1 所示作业流程，将战损维修均摊到波次作业中的每一架舰载机，考虑编队战损概率、作业执行时间分布、资源转移时间分布、更换备份机、充氮不确定性、机械维修不确定性等不确定性因素，可以建立相应的两波次舰载机循环作业 GERT 模型如图 2 所示，其中，彩色箭头表示舰载机出动回收波次间的作业流程约束；X 取值正整数，表示某一作业周期，使得建立的模型可以表示循环作业中的任意阶段；①～㉒表示作业编号。

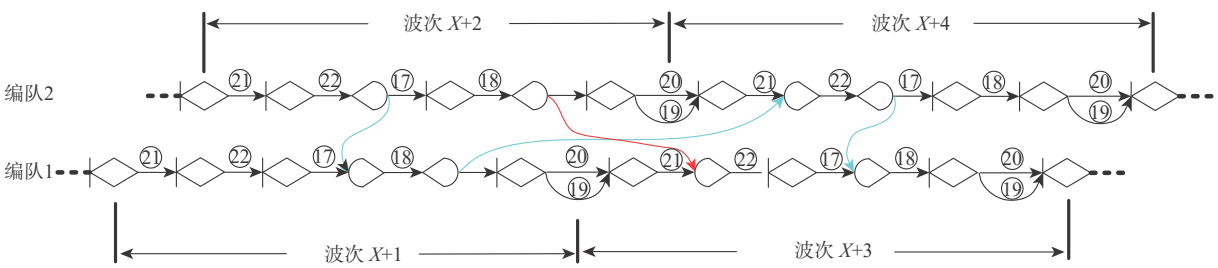


图 2　舰载机循环作业 GERT 模型

Fig. 2　GERT model for cyclic operations of carrier aircraft

对于图 2 中的飞行前准备作业，将机务准备中的作业流程表示为工序流，将勤务准备中的资源供给流程表示为资源流，建立如图 3 所示的飞行前准备作业 GERT 模型。对于有消耗性资源需求的作业 (例如加油、充氮等)，需要工序流和资源流同时到达该作业工位时，该作业才能实施保障，在 GERT 模型中其对应的汇入节点的输入端为“与”型。图中的虚工序以概率 1 执行，但不消耗资源和时间。

针对出动离场作业，可以建立如图 4 所示的出动离场作业 GERT 模型，其中，虚线箭头表示同一波次中不同舰载机离场作业流程约束。

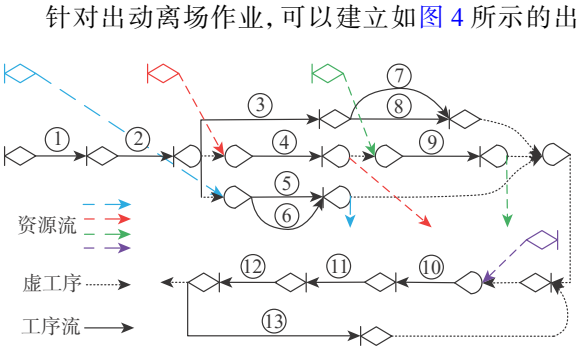


图 3　飞行前准备作业 GERT 模型

Fig. 3　GERT model of pre-flight preparations

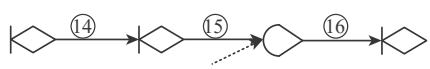


图 4　出动离场作业 GERT 模型

Fig. 4　GERT model of sortie and departure

2　以 Laplace 变换为传递函数的 GERT 网络

相对于基于矩母函数的 GERT 网络[33] 和以特征函数为传递函数的 GERT 网络[35]，以 Laplace 变换为传递函数的 GERT 网络可以通过 Laplace 逆变换直接求出对应概率密度函数。本节构造以 Laplace 变换为传递函数的 GERT 网络，并分析其求解特征。

2.1　传递函数定义

对于如图 5 所示的舰载机出动回收作业中的任一概率分支 (其中 p_{kl} 为节点 k 实现时概率分支 (k,l) 被执行的概率，t_{kl} 为执行分支 (k,l) 耗时)，定义 Laplace 变换：

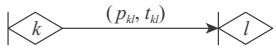


图5 通用概率分支

Fig. 5 Generic probabilistic branch

$$L(t_{kl})=\begin{cases}\int_{0^+}^{\infty}\mathrm{e}^{-st_{kl}}f(t_{kl})\mathrm{d}t_{kl} \\ \sum\mathrm{e}^{-st_{\mathrm{m}}}p \qquad t_{kl}=t_{\mathrm{m}}\end{cases} \tag{1}$$

式中：s为复变量；t_{kl}为实变量；t_{m}为t_{kl}可能的取值；$L(t_{kj})$为随机变量t_{kl}的Laplace变换，t_{kl}为连续分布时，其计算方式为等号右侧上方表达式，t_{kl}为离散分布时，其计算方式为等号右侧下方表达式；$f(t_{kl})$和$p(t=t_{kl})$分别为随机变量t_{kl}的概率密度分布函数和分布率。

根据数学期望的定义，式(1)可写为

$$L(t_{kl})=E(\mathrm{e}^{-st_{kl}}) \tag{2}$$

式中：E为期望。

因而，对于相互独立的随机变量序列$t_{ku}^1,t_{tu}^2,\cdots,t_{kt}^n$，其和的Laplace变换满足：

$$L\left(\sum_1^n t_{kl}^i\right)=E\left(\mathrm{e}^{-s\sum_1^n t_{kl}^i}\right)=\prod_1^n E\left(\mathrm{e}^{-st_{kl}^i}\right)=\prod_1^n L\left(t_{kl}^i\right) \tag{3}$$

若$t_{kl}^1,t_{kl}^2,\cdots,t_{kl}^n$均为连续随机变量，且对应的概率密度分布分别为$f_1\left(t_{kl}^1\right),f_2\left(t_{kl}^2\right),\cdots,f_n\left(t_{kl}^n\right)$，结合Laplace变换的卷积定理，由式(3)可以进一步得到

$$L\left(\sum_1^n t_{kl}^i\right)=L(f_1(t_{kl}^1)*f_2(t_{kl}^2)*\cdots*f_n(t_{kl}^n)) \tag{4}$$

式中：*为卷积符号。

进而，定义分支(k,l)的传递函数W_{kl}为

$$W_{kl}=p_{kl}L(t_{kl}) \tag{5}$$

2.2 传递函数特征

在第1节建立的舰载机出动回收作业流程GERT网络中，存在串联、并联和自环等结构，本节对其简化运算过程中的传递函数特性进行研究。

对于如图6(a)所示的相互独立的串联概率分支，其等效概率如图6(b)所示，其中$k_1\sim k_3$为节点编号。

根据概率分支逻辑关系，等效概率分支被执行概率和耗时分别为

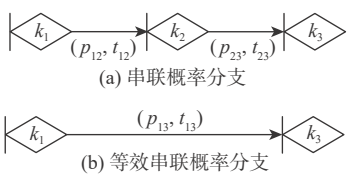


图6 串联概率分支简化

Fig. 6 Reduction of probabilistic branches in series

$$p_{13}=p_{12}p_{23} \tag{6}$$

$$t_{13}=t_{12}+t_{23} \tag{7}$$

式中：p_{12}、p_{23}和p_{13}分别为分支(1, 2)、(2, 3)和等效分支(1, 3)的发生概率；t_{12}、t_{23}、t_{13}分别表示执行对应分支作业耗时。结合式(3)～式(5)，可得等效概率分支的传递函数为

$$W_{13}=W_{12}W_{23} \tag{8}$$

对于并联概率分支，在1.3节构建的舰载机出动回收GERT模型中包括2种类型，即具有“异或”型关系的并联概率分支和具有“与”型关系的并联概率分支。

对于如图7所示的具有“异或”型关系的并联概率分支，结合异或型节点特性，等效概率分支被执行概率和耗时分别满足：

$$p_{45}=p_{4c5}+p_{4d5} \tag{9}$$

$$t_{45}=(p_{4c5}t_{4c5}+p_{4d5}t_{4d5})/p_{45} \tag{10}$$

式中：c和d用于区分节点k_4和k_5间的并联分支，4c5和4d5分别表示这2个分支。

结合Laplace变换的线性特性[36]，等效概率分支耗时分布的Laplace变换满足：

$$L(t_{45})=\frac{p_{4c5}}{p_{45}}L(t_{4c5})+\frac{p_{4d5}}{p_{45}}L(t_{4d5}) \tag{11}$$

进而，根据式(5)的定义，等效概率分支的传递函数满足：

$$W_{45}=W_{4c5}+W_{4d5} \tag{12}$$

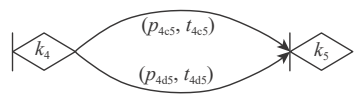


图7 具有“异或”型节点关系的并联分支

Fig. 7 Parallel branches with “exclusive or” relationship

对于如图8所示的具有“与”型关系的并联概率分支，若两概率分支相互独立，则对应的等效概率分支被执行概率和耗时分别满足：

$$p_{67}=p_{6e7}\cap p_{6g7}=p_{6e7}p_{6g7} \tag{13}$$

$$t_{67}=\max\left(t_{6e7},\ t_{6g7}\right) \tag{14}$$

式中：e和g用于区分节点k_6和k_7间的并联分支，6e7和6g7分别表示这2个分支。

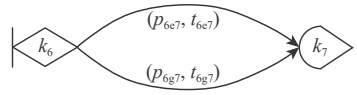


图8 具有“与”型节点关系的并联概率分支

Fig. 8 Probabilistic branches in parallel with “and” relationship

对于如图9(a)所示带自环结构的概率分支组合，结合串、并联结构关系，其中，自环结构可以等

效为如图 9(b) 所示的并联结构, 其共由 $n+1$ 个串联概率分支并联组成, 每个串联概率分支由 $n(n=0,1,\cdots,\infty)$ 个传递函数为 W_8 的概率分支串联后与传递函数为 W_{8h9} 的概率分支串联组成, 其中, 8h9 表示节点 k_8 和 k_9 间直接相连的分支, h 用于区分自环分支和等效分支。因此, 图 9(a) 对应的等效传递函数满足:

$$W_{89}=W_{8h9}+W_8W_{8h9}+\cdots+(W_8)^nW_{8h9} \tag{15}$$

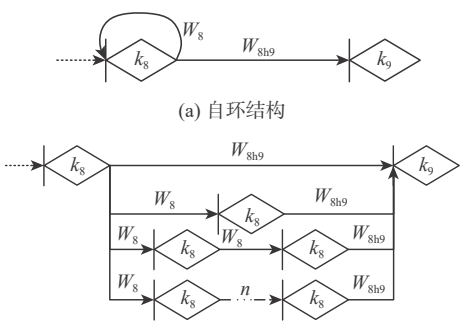
结合式 (1) 和式 (5) 中的定义并考虑到 $p_g\leqslant 1$, 可得

$$W_8=p_8L(t_8)\leqslant\int_{0^+}^{\infty}\mathrm{e}^{-st_8}f(t_8)\mathrm{d}t_8\leqslant \int_{0^+}^{\infty}\left|\mathrm{e}^{-st_8}\right|f(t_8)\mathrm{d}t_8\leqslant\mathrm{e}^{-Re(s)t_8}<1\quad Re(s)<1 \tag{16}$$

式中: $Re(s)$ 为复变量 s 的实部。进而, 结合式 (15) 可以进一步得到带自环结构的概率分支对应的等效概率分支传递函数满足:

$$W_{89}=\frac{W_{8h9}}{1-W_8} \tag{17}$$

综上所述, 舰载机出动回收 GERT 网络的简化过程满足信号流图串联结构、并联结构和自环结构运算理论, 因而可以采用基于梅森公式[32] 或基于矩阵求解[37-38] 的方式对舰载机出动回收 GERT 网络模型进行求解。



(a) 自环结构

(b) 等效自环结构

图 9　自环结构的简化

Fig. 9　Reduction of self-loop structure

3　舰载机出动回收 GERT 模型求解

根据实际舰载机出动回收作业过程中各项作业的完成人员基本不同, 可以假设舰载机出动回收作业中的各项作业相互独立。

以 Laplace 变换为传递函数的 GERT 网络的精确求解能够得到网络传递参数的精确概率密度分布函数, 在舰载机出动回收作业流程中可得到作业完成时间的概率密度分布函数。但是, 其运算过程较为耗时, 无法实现舰载机出动回收作业方案时间效能的实时分析。同时考虑到, 信号流图理论无法通过对具有“与”型关系的并联概率分支简化求解得到等效概率分支的概率密度分布函数, 因而在舰载机出动回收作业流程 GERT 网络分析求解过程中引入蒙特卡罗方法。在提出的 MC-GERT 中, 通过 GERT 网络传递函数的运算实现舰载机出动回收作业流程的简化, 借助蒙特卡罗方法实现“与”型节点并联概率分支的求解。

基于 MC-GERT 对舰载机出动回收作业流程时间不确定性进行分析, 具体的求解步骤包括:

步骤 1　统计舰载机出动回收过程中各项作业的完工时间数据, 通过多项式拟合各项作业的完成时间概率密度分布函数, 由式 (4) 所示的卷积定理和 Laplace 变换的线性性质, 可以推得经过 2.2 节串联、并联结构的运算, 其结果的分布函数仍为多项式形式;

步骤 2　计算各项作业的传递函数;

步骤 3　针对特定作业阶段, 结合传递函数的运算特点, 利用梅森公式[32] 对 GERT 网络进行最大程度简化, 并得到简化后 GERT 网络各模块的时间参数的概率密度分布函数;

步骤 4　由步骤 3 得到的阶段作业概率密度分布函数生成足够数量的随机数矩阵;

步骤 5　针对特定的舰载机出动回收作业流程调度方案, 构建对应的 GERT 网络模型;

步骤 6　通过阶段随机数矩阵的矩阵运算, 得到出动回收作业流程完成时间的随机数矩阵, 例如, 对于“与”型节点并联概率分支的求解通过取各分支的随机数矩阵的最大值实现;

步骤 7　通过随机数拟合得到作业流程完成时间概率密度分布曲线, 并分析舰载机出动回收作业流程完成时间的分布特征。

上述步骤中, 步骤 1～步骤 4 在不同的舰载机出动回收作业规划方案中具有通用性, 可以在线下完成, 有利于进一步提升不确定性分析的时效性。

4　实例分析

本节通过海上飞机拦截作战[39] 构建舰载机出动回收作业案例分析研究作业流程完成时间的分布特性。

4.1　案例构建

作战想定为蓝方舰载机对红方突防飞机进行海上拦截, 据此构建蓝方舰载机不同样式下的出动回收作业案例, 分别对其出动回收作业方案耗时进行时间不确定性分析。

案例 1　蓝方单次出动 3 架舰载机编队对红方飞机进行拦截。

案例 2 蓝方两波次连续出动 3 架舰载机对红方飞机进行拦截。

通过对收集到的出动回收过程中各项作业完工时间数据进行拟合，可以得到各项作业完工时间的概率密度分布，假设某次得到的作业完成时间数据如表 2 所示，其中，u_β为便于表示各保障作业完工时间的概率密度分布函数而定义分段函数：

$$u_\beta(t)=\begin{cases}1 & t\geqslant\beta\\ 0 & t<\beta\end{cases} \tag{18}$$

式中：β为区间 $[0,+\infty)$ 上的常实数。

根据表 2 中的数据，对图 4 中舰载机飞行前准备作业中的自环结构进行近似简化。结合如图 9(b)所示的并联结构，该自环等效结构最上方的 2 条并联支路发生概率的和已经达到 0.999 1，因而可较小误差取发生概率最大的 2 条概率支路近似自环结构。

表 2 舰载机出动回收流程作业数据

Table 2 Data of cyclic operations of carrier aircraft

作业编号	作业名称	概率	作业完成时间分布
①	牵引入位	1	$\frac{1}{12}[(u_2-t_6)(8.5-t)]$
②	系留	1	2
③	机械检查	1	$\frac{1}{2}(4u_6-u_4)(8-t)$
④	充氧	1	$(u_4-u_2)(t-4)+(u_2-u_6)(6-1)$
⑤	充氮	0.5	$(u_4-u_2)(t-4)+(u_2-u_6)(6-1)$
⑥		0.5	0
⑦	维修	0.1	$\frac{1}{14}(u_6-z_{t_0})$
⑧		0.9	0
⑨	加油	1	$\frac{1}{50}[(z_{20}-z_{20})(t-20)]$
⑩	挂载	1	$\frac{1}{4}(t_{10}-t_{21})(t-19)$
⑪	惯导对准	1	(u_4-u_0)
⑫	暖机自检	1	2
⑬	更换备份机	0.03	$(\mu_1-\mu_2)$
⑭	滑行入位	0.97	$\frac{1}{2}(u_1-t_4)$
⑮	准备位等待	1	$(u_{0.3}-u_{0.8})(8t-2.4)$
⑯	离场	1	0.2
⑰	编队空中作业	1	$N(120,6)$
⑱	编队着舰作业	1	$\frac{1}{18}(u_6-u_{ti})(12-t)$
⑲	编队战损维修	0.3	$(u_{15}-u_{20})(t-15)/50+$ $(u_{20}-u_{35})(35-t)/150$
⑳		0.7	0
㉑	编队飞行前准备作业	1	
㉒	编队出动离场	1	

4.2 案例结果及分析

该部分依次对单机出动、单次出动多架舰载机、波次出动多架舰载 3 种情况下的出动回收作业流程进行时间不确定性分析。

4.2.1 单机出动回收作业流程时间不确定性分析

先对舰载机出动回收作业流程中各阶段作业完成时间进行分析。对于如图 3 所示的舰载机飞行前准备作业 GERT 模型和如图 4 所示的舰载机出动离场作业 GERT 模型，可以得到单架舰载机飞行前准备作业、出动离场作业完成时间概率密度分布分别如图 10 和图 11 所示。

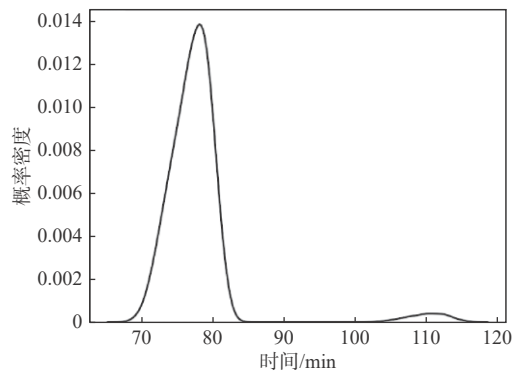


图 10 舰载机飞行前准备作业完成时间概率密度分布

Fig. 10 Probability density distribution of completion time for pre-flight preparations of carrier aircraft

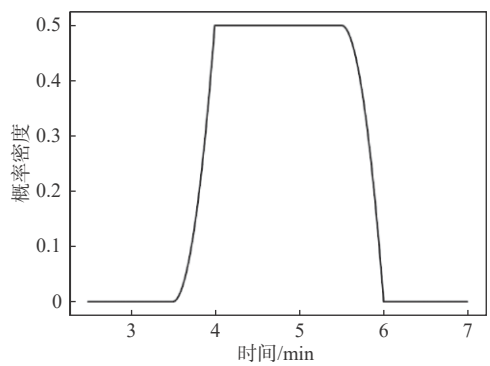


图 11 舰载机出动离场作业完成时间概率密度分布

Fig. 11 Probability density distribution of completion time for departure of carrier aircraft

由图 10 所示的舰载机飞行前准备作业完成时间概率密度分布，可以发现单机完成飞行前准备作业的完成时间主要分布在 2 个时间段，主要分布在前一个时间段并呈现正态分布形式，且后一个时间段的密度曲线下方面积约为 0.03，与表 2 中更换备份机的概率相等，因而可推断其主要是由更换备份机导致。

图 11 中单机出动离场作业完成时间对应的概率密度分布函数如式 (19) 所示。说明对于特定的作业流程，以 Laplace 变换为传递函数的 GERT 网络能够精确求出传递参数的概率密度分布，进而增加不确定评估的准确性。

$$f(t)=60.5\left[\left(0.033t^2-0.364t+0.992\right)u_6-\left(0.033t^2-0.364t+1\right)u_{5.5}-\left(0.033t^2-0.231t+0.397\right)u_4+\left(0.033t^2-0.231t+0.405\right)u_{3.5}\right] \quad (19)$$

4.2.2　单次出动多架舰载机时间不确定性分析

舰载机出动回收作业调度方案给出了各项作业之间的流程约束关系和资源匹配关系，结合单机出动流程，可以建立案例 1 中的单波次 3 架舰载机出动离场作业 GERT 模型，如图 12 所示。

根据图 12 所述的作业流程，可以得到案例 1 中舰载机出动离场完成时间概率密度分布函数如图 13 所示。由图可见，由于更换备份机作业的存在，单次出动多架舰载机完成时间概率密度分布仍分布在 2 个时间段。根据图中的概率密度分布，调度人员可以评估单次出动多架舰载机时出动方案能够按规划方案进行的可信度。

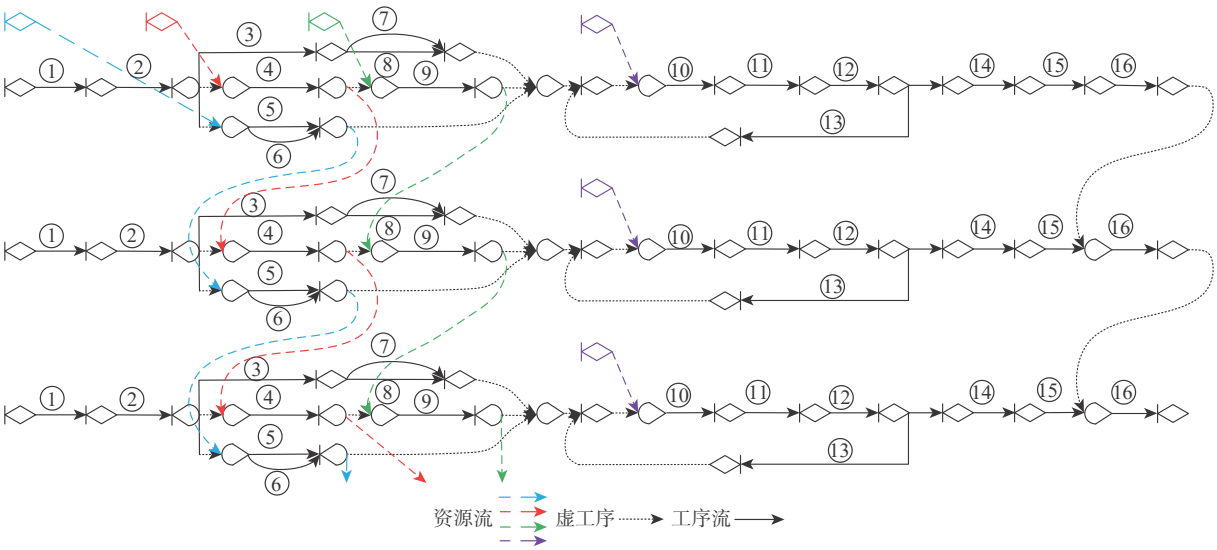


图 12　3 架舰载机出动离场作业 GERT 模型

Fig. 12　GERT model for launching three carrier aircraft

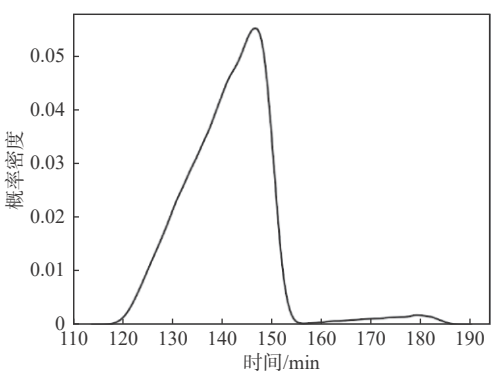


图 13　3 架舰载机出动离场作业完成时间概率密度分布

Fig. 13　Probability density distribution of completion time for launching three carrier aircraft

4.2.3　波次出动多架舰载机时间不确定性分析

进一步，结合单次出动多架舰载作业流程，可以计算得到案例 2 中两波次 3 架舰载机循环作业周期完成时间概率密度分布，为体现如图 1 所示的两波次舰载机出动离场、回收着舰之间的流程约束关系，计算图 2 中的波次 1 完成两周期出动回收作业完成时间，其概率密度分布函数如图 14 所示。

战损概率体现了蓝方舰载机的作战性能，为分析舰载机性能对出动回收作业完成时间的影响，以案例 2 中的编队战损概率 (记为 P_{19}) 为变量，得到不同战损概率条件下，波次 1 完成两周期作业耗时的概率密度分布如图 15 所示。

为表征周期作业完成时间概率密度分布曲线的特点，统计不同战损概率条件下波次 1 舰载机两周期作业完成时间对应的概率密度曲线的均值、标准差、峰度和偏度，如表 3 所示。

舰载机周期作业完成时间直接影响舰载机出动架次率。结合表 3，就两周期作业完成时间均值而言，随着舰载机性能的下降，即战损概率的增加，循环作业模式下周期作业完成时间均值呈线性上

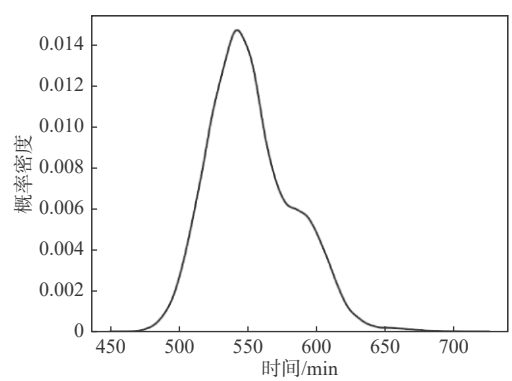


图 14　波次 1 完成两周期作业耗时概率密度分布

Fig. 14　Probability density distribution of time spent for completing two cyclic operations within wave 1

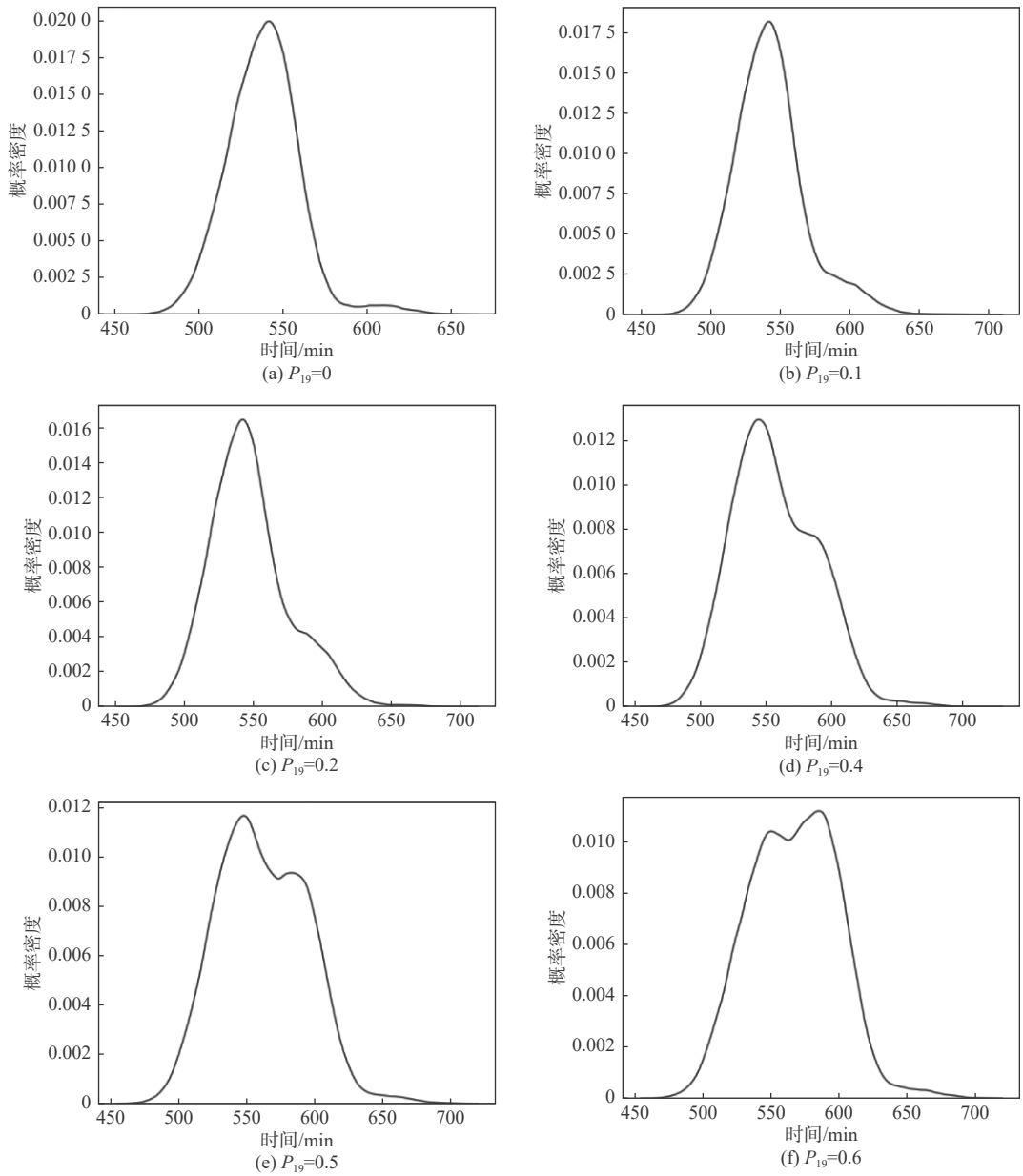


图 15 不同战损概率条件下波次 1 完成两周期作业耗时概率密度分布

Fig. 15 Probability density distribution of time spent for completing two cyclic operations within wave 1 under different battle damage probability conditions

升趋势，因而舰载机出动架次率随舰载机性能的下降而降低。

表 3 波次 1 两周期作业完成时间分布特性

Table 3 Distribution characteristics of completion time for completing two cyclic operations within wave 1

战损概率	均值/min	标准差/min	峰度	偏度
0.0	538	22.08	1.87	0.59
0.1	543	26.22	1.43	0.79
0.2	548	19.14	0.71	0.7
0.3	553	31.1	0.24	0.56
0.4	557	32.28	−0.01	0.41
0.5	562	32.78	−0.13	0.26
0.6	567	32.38	−0.06	0.13

峰度是表征概率密度分布曲线在平均值处峰值高低的特征参数，直观反映了峰部的尖度。由图 15 和表 3 可知，战损概率不大于 0.3 情况下，两周期作业完成时间分布均为厚尾分布（峰度>0）；战损概率大于 0.4 时，两周期作业完成时间分布为瘦尾分布（峰度<0）。从整体来看随着战损概率增大，两周期作业完成时间概率密度分布函数尖度越来越大，结合愈趋稳定的标准差数据，说明随着战损概率的增大，周期作业完成时间愈趋于集中分布在越来越大的均值附近。在实践过程中则意味着作业完成时间的确定性增大，随着战损概率的增大，调度人员能够对作业完成时间增大有一个明显的预期。

偏度是统计数据分布偏斜方向和程度的度

量。从表 3 来看, 各类战损概率情况下, 波次 1 两周期作业完成时间概率密度分布函数均为右偏分布(偏度>0)。就概率密度函数偏斜程度而言, 随着战损概率增大, 概率密度函数右偏程度先增大, 后减小。结合均值、标准差和峰度特征, 可以进一步说明, 随着战损概率的增大, 周期作业完成时间向均值附近移动, 增加了作业完成时间的确定性。

此外, 由图 15 可知, 随着战损概率的增大, 波次 1 完成两周期作业最大耗时明显增加, 说明了周期作业完成时间向越来越大的均值附近偏移的趋势是整体性的。

综合而言, 战损概率对于波次连续出动舰载机出动回收作业完成时间具有较为明显的影响。因此, 在舰载机实际应用过程中需要注重舰载机性能的定期检查与维护, 并加强高性能舰载机的研发与装配。

5　结　论

1) 本文结合 Laplace 变换的优势, 实现了任意概率密度分布概率分支作业时间参数的求解, 相对于现有研究中分支作业时间均为负指数分布的不足, 本文方法适用性更为广泛。

2) 本文方法可以实现 GERT 网络传递参数概率密度分布函数的求解, 而基于矩母函数为传递函数的 GERT 网络仅能求解传递参数均值、方差等数据特征, 无法实现概率密度分布函数求解; 以特征函数为传递函数的 GERT 网络中概率密度分布函数的求解较为繁琐。

3) 本文方法计算效率高, 本文的 MC-GERT 能够实现 GERT 网络中带“与”型节点关系的并联支路的求解, 能够实现在线分析舰载机出动回收作业方案。

为实现 GERT 网络中带“与”型节点关系的并联支路的简化求解, 根据“与”型节点的运算特点, 在今后的研究中, 一方面可以探索以累计概率分布函数为传递函数的 GERT 网络; 另一方面, 探索新型随机网络结构以实现参数的不确定性的准确分析也是一项具有挑战性的工作。

参考文献（References）

[1] 罗湘勇, 冯浩源, 潘星, 等. 面向流程的保障体系集成建模方法研究[J]. 系统工程与电子技术, 2013, 35(11): 2314-2319.
LUO X Y, FENG H Y, PAN X, et al. Integration modeling approach of process oriented support system of systems[J]. Systems Engineering and Electronics, 2013, 35(11): 2314-2319(in Chinese).

[2] 冯浩源, 吕卫民, 江式伟. 基于时间 Petri 网和模拟退火算法的单机机务准备流程优化[J]. 兵工自动化, 2013, 32(3): 90-93.
FENG H Y, LYU W M, JIANG S W. Task ready process optimization of single airplane based on timed petri net and simulated annealing algorithm[J]. Ordnance Industry Automation, 2013, 32(3): 90-93(in Chinese).

[3] 韩维, 苏析超, 陈俊锋. 舰载机多机一体化机务保障调度方法[J]. 系统工程与电子技术, 2015, 37(4): 809-816.
HAN W, SU X C, CHEN J F. Integrated maintenance support scheduling method of multi-carrier aircrafts[J]. Systems Engineering and Electronics, 2015, 37(4): 809-816(in Chinese).

[4] 韩庆田, 曹文静, 张毅. 舰载机机务保障资源配置研究[J]. 价值工程, 2012, 31(31): 311-312.
HAN Q T, CAO W J, ZHANG Y. Research on resource distribution of maintenance support for carrier aircraft[J]. Value Engineering, 2012, 31(31): 311-312(in Chinese).

[5] 苏析超, 韩维, 萧卫, 等. 基于 Memetic 算法的舰载机舰面一站式保障调度[J]. 系统工程与电子技术, 2016, 38(10): 2303-2309.
SU X C, HAN W, XIAO W, et al. Pit-stop support scheduling on deck of carrier plane based on Memetic algorithm[J]. Systems Engineering and Electronics, 2016, 38(10): 2303-2309(in Chinese).

[6] CUI R W, HAN W, SU X C, et al. A multi-objective hyper heuristic framework for integrated optimization of carrier-based aircraft flight deck operations scheduling and resource configuration[J]. Aerospace Science and Technology, 2020, 107: 106346.

[7] CUI R W, HAN W, SU X C, et al. A dual population multi-operator genetic algorithm for flight deck operations scheduling problem [J]. Journal of Systems Engineering and Electronics, 2021, 32(2): 331-346.

[8] 崔荣伟, 韩维, 苏析超, 等. 舰载机甲板机务勤务保障作业调度与资源配置集成优化[J]. 系统工程与电子技术, 2021, 43(7): 1884-1893.
CUI R W, HAN W, SU X C, et al. Integrated optimization of carrier based aircraft flight deck operations scheduling and resource configuration for pre-flight preparation stage[J]. Systems Engineering and Electronics, 2021, 43(7): 1884-1893(in Chinese).

[9] 苏析超, 韩维, 张勇, 等. 考虑人机匹配模式的舰载机甲板机务勤务保障调度算法[J]. 航空学报, 2018, 39(12): 214-232.
SU X C, HAN W, ZHANG Y, et al. Scheduling algorithm for maintenance and service support of carrier-based aircraft on flight deck with different man-aircraft matching patterns[J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(12): 214-232(in Chinese).

[10] 韩维, 崔荣伟, 苏析超, 等. 基于双种群模糊引力搜索算法的舰载机甲板作业调度[J]. 控制与决策, 2021, 36(11): 2751-2759.
HAN W, CUI R W, SU X C, et al. Flight deck operations scheduling based on dual population fuzzy gravitational search algorithm [J]. Control and Decision, 2021, 36(11): 2751-2759(in Chinese).

[11] HAN W, GUO F, SU X C. A reinforcement learning method for a hybrid flow-shop scheduling problem[J]. Algorithms, 2019, 12(11): 222.

[12] YU L F, ZHU C, SHI J M, et al. An extended flexible job shop scheduling model for flight deck scheduling with priority, parallel operations, and sequence flexibility[J]. Scientific Programming, 2017, 2017: 2463252.

[13] JIANG T T, SU X C, HAN W. Optimization of support scheduling on deck of carrier aircraft based on improved differential evolution algorithm[C]//Proceedings of the 3rd IEEE International Conference on Control Science and Systems Engineering. Piscataway: IEEE Press, 2017: 136-140.

[14] 苏析超, 伍恒, 崔荣伟, 等. 基于边际-人工蜂群算法的舰载机机群出动保障人员配置-调度联合优化方法[J]. 北京航空航天大学学报, 2020, 46(11): 2056-2068.
SU X C, WU H, CUI R W, et al. Joint optimization method for carrier-based aircraft fleet sortie support personnel configuration and scheduling based on marginal-ABC algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(11): 2056-2068(in Chinese).

[15] 万兵, 苏析超, 郭放, 等. 不确定性工时下甲板作业的前摄性鲁棒调度[J]. 航空学报, 2022, 43(12): 379-396.
WAN B, SU X C, GUO F, et al. Proactive robust scheduling of aircraft carrier flight deck operations with uncertain activity durations[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(12): 379-396(in Chinese).

[16] SU X C, HAN W, WU Y, et al. A proactive robust scheduling method for aircraft carrier flight deck operations with stochastic durations[J]. Complexity, 2018, 2018(1): 1-38.

[17] SU X C, HAN W, WU Y, et al. A robust scheduling optimization method for flight deck operations of aircraft carrier with ternary interval durations[J]. IEEE Access, 2018, 6: 69918-69936.

[18] 冯强, 曾声奎, 康锐. 基于 MAS 的舰载机动态调度模型[J]. 航空学报, 2009, 30(11): 2119-2125.
FENG Q, ZENG S K, KANG R. A MAS-based model for dynamic scheduling of carrier aircraft[J]. Acta Aeronautica et Astronautica Sinica, 2009, 30(11): 2119-2125(in Chinese).

[19] 范加利, 孟杨凯, 黄葵, 等. 基于禁忌算法的舰载机甲板作业动态调度优化算法[J]. 系统工程与电子技术, 2023, 45(10): 3172-3182.
FAN J L, MENG Y K, HUANG K. Aircraft deck operations dynamic scheduling optimization algorithm based on the tabu algorithm[J]. Systems Engineering and Electronics, 2023, 45(10): 3172-3182(in Chinese).

[20] YUAN P L, HAN W, SU X C, et al. A dynamic scheduling method for carrier aircraft support operation under uncertain conditions based on rolling horizon strategy[J]. Applied Sciences, 2018, 8(9): 1546.

[21] WANG X W, LIU J, SU X C, et al. A review on carrier aircraft dispatch path planning and control on deck[J]. Chinese Journal of Aeronautics, 2020, 33(12): 3039-3057.

[22] CUI J P, WU Y, SU X C, et al. A task allocation model for a team of aircraft launching on the carrier[J]. Mathematical Problems in Engineering, 2018, 2018: 7920806.

[23] LIU J, HAN W, LI J, et al. Integration design of sortie scheduling for carrier aircrafts based on hybrid flexible flowshop[J]. IEEE Systems Journal, 2020, 14(1): 1503-1511.

[24] LIU J, HAN W, WANG X W, et al. Research on cooperative trajectory planning and tracking problem for multiple carrier aircraft on the deck[J]. IEEE Systems Journal, 2020, 14(2): 3027-3038.

[25] 万兵, 韩维, 梁勇, 等. 舰载机出动离场调度优化算法[J]. 系统工程与电子技术, 2021, 43(12): 3624-3634.
WAN B, HAN W, LIANG Y, et al. Optimization algorithm of carrier-based aircraft sortie departure scheduling[J]. Systems Engineering and Electronics, 2021, 43(12): 3624-3634(in Chinese).

[26] WU Y, SUN L G, QU X J. A sequencing model for a team of aircraft landing on the carrier[J]. Aerospace Science and Technology, 2016, 54(7): 72-87.

[27] 刘玉杰, 万兵, 苏析超, 等. 基于 IABC 算法的舰载机着舰调度[J]. 控制与决策, 2022, 37(7): 1810-1818.
LIU Y J, WAN B, SU X C, et al. Scheduling of landing for carrier-based aircraft based on improved artificial bee colony algorithm[J]. Control and Decision, 2022, 37(7): 1810-1818(in Chinese).

[28] 刘玉杰, 韩维, 苏析超, 等. 基于改进灰狼优化算法的舰载机着舰调度[J]. 北京航空航天大学学报, 2024, 50(3): 803-813.
LIU Y J, HAN W, SU X C, et al. Research on carrier aircraft landing scheduling problem based on improved gray wolf optimization[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(3): 803-813(in Chinese).

[29] 夏宏青, 焦健, 褚嘉运, 等. 基于改进 GERT 的任务过程时间特性建模分析方法[J]. 北京航空航天大学学报, 2020, 46(11): 2140-2148.
XIA H Q, JIAO J, CHU J Y, et al. Improved GERT based time characteristic modeling and analysis method for task process[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(11): 2140-2148(in Chinese).

[30] 吴勇, 潘星, 康锐, 等. 基于图示评审技术的舰载机航空保障时间分析[J]. 兵工学报, 2013, 34(12): 1611-1615.
WU Y, PAN X, KANG R, et al. Aircraft support process time analysis based on graphic evaluation and review technique[J]. Acta Armamentarii, 2013, 34(12): 1611-1615(in Chinese).

[31] JEWELL A. Sortie generation capacity of embarked airwings[R]. Washington, D. C. : Center For Naval Analyses Alexandria Va, 1998.

[32] PRITSKER A A B, HAPP W W. GERT: graphical evaluation and review technique, part I. Fundamentals[J]. Journal of Industrial Engineering, 1966, 17(5): 267-274.

[33] GHAELI M R, SADI-NEZHAD S. Recent advances on graphical evaluation and review techniques[J]. Journal of Project Management, 2017, 2(3): 107-112.

[34] ELMAGHRABY S E. An algebra for the analysis of generalized activity networks[J]. Management Science, 1964, 10(3): 494-514.

[35] TAO L Y, WU D S, LIU S F, et al. Schedule risk analysis for new-product development: the GERT method extended by a characteristic function[J]. Reliability Engineering & System Safety, 2017, 167: 464-473.

[36] SCHIFF J L. The Laplace transform: theory and applications[M]. Berlin: Springer, 1999: 16-17.

[37] MARTÍNEZ LEÓN H C, FARRIS J A, LETENS G, et al. An analytical management framework for new product development processes featuring uncertain iterations[J]. Journal of Engineering and Technology Management, 2013, 30(1): 45-71.

[38] 陶良彦, 刘思峰, 方志耕, 等. 以特征函数为传递参数的 CF-GERT 及其矩阵法求解[J]. 系统工程理论与实践, 2018, 38(2): 509-521.
TAO L Y, LIU S F, FANG Z G, et al. CF-GERT model conveying characteristic function and its matrix solution[J]. Systems Engineering-Theory & Practice, 2018, 38(2): 509-521(in Chinese).

[39] 周琛, 尚柏林, 宋笔锋, 等. 基于作战环的航空武器装备体系贡献率评估[J]. 航空学报, 2022, 43(2): 309-320.
ZHOU C, SHANG B L, SONG B F, et al. Contribution evaluation of aviation armament system-of-systems based on operation loop[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(2): 309-320(in Chinese).

Time uncertainty analysis on cyclic operation procedures of carrier aircraft based on MC-GERT

GUO Fang[1]，HAN Wei[1]，LIU Yujie[1, 2]，SU Xichao[3, *]，BAI Tian[4]，LIU Chun[5]

(1. School of Basic Science for Aviation，Naval Aviation University，Yantai 264001，China；
2. Military Political Work Research Institute，Beijing 100071，China；
3. Aeronautical Operations College，Naval Aviation University，Yantai 264001，China；
4. Systems Engineering Research Institute，CSSC，Beijing 100094，China；
5. 650 Aircraft Design Institute of AVIC Hongdu，Nanchang 330024，China)

Abstract：The completion time of the cyclic operation of carrier aircraft usually presents significant uncertainty, which directly affects the development of subsequent mission plans for carrier aircraft by carrier mission planners. To determine the distribution pattern of the completion time of the cyclic operation of carrier aircraft and assist planners in tracking the progress of carrier aircraft mission execution, the time uncertainty analysis method for cyclic operations of carrier aircraft was proposed. Firstly, based on the operation process from dispatching to recycling, the stochastic network model of the cyclic operation of carrier aircraft was established. Secondly, a graphic evaluation and review technique network with Laplace transform as the transfer function was proposed, and the characteristics of its transfer function were studied. It was proven that the stochastic netw ork model of the cyclic operation of carrier aircraft could be solved by utilizing the signal flow graph theory. In addition, the Monte Carlo graphical evaluation and review technique（MC-GERT）method was introduced to simplify the parallel branches with the “AND” relationship in the stochastic network and improve the calculation efficiency of the network solution. By the proposed graphic evaluation and review technique method combined with Monte Carlo, finally, the distribution characteristics of the completion time of the pre-flight preparations, sortie and departure, single wave departure operation, and the periodic operation under the cyclic mode of carrier aircraft were analyzed respectively, which verified the feasibility of the proposed analysis method. The results show that the proposed method can be applied to the online analysis of the planning scheme of the cyclic operation of carrier aircraft and can provide certain scheme implementation expectations for relevant departments of carrier aircraft operation planning.

Keywords：carrier aircraft; cyclic operation; MC-GERT; Laplace transform; Monte Carlo; uncertainty analysis

Received：2023-03-15；**Accepted**：2023-04-23；**Published Online**：2023-06-28 19：22
URL：link.cnki.net/urlid/11.2625.V.20230628.1323.002
Foundation item：National Natural Science Foundation of China (62003366)
* **Corresponding author**. E-mail：suxich@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0244

城市低空场景下无人机运行对地风险量化评估

陈艺君[1,2]，余莎莎[1,3,*]，张学军[3,4]

(1. 西华大学 航空航天学院，成都 610039；2. 国家空管监视与通信系统工程技术研究中心，成都 610039；3. 北航（四川）西部国际创新港科技有限公司，成都 610039；4. 北京航空航天大学 电子信息工程学院，北京 100191)

摘 要：为保证无人机(UAV)在城市低空安全运行，构建UAV运行的地面风险量化评估模型，提出使用考虑多因素多层次的对地风险地图，来定义与UAV相关的对地风险，量化风险程度。分析城市低空典型特征和UAV坠地事故因果链，建立多因素影响下的地面风险量化评估模型；考虑UAV的下降行为和风参数的不确定性，计算UAV坠地的影响区域概率密度函数；通过结合城市区域地图、人口密度、遮蔽效应、障碍物和禁飞区多个图层，计算每个地理参考单元的风险值，生成大区域范围的地面风险地图。基于真实算例对所建模型进行验证分析，结果表明：不同城市区域的风险分布区别明显，风险程度与人口密度分布、地表遮蔽效果、地面障碍物高度分布、城市禁飞区的划设及UAV飞行高度密切相关；UAV在运行中应用地面风险地图，有利于识别所飞越区域的风险程度，避开高风险区域，提高在城市低空运行的安全性。

关 键 词：城市低空；地面风险；多影响因素；风险评估；风险地图；运行安全
中图分类号：V279; V355; X913.4
文献标志码：A **文章编号**：1001-5965(2025)03-0806-10

随着无人机(unmanned aerial vehicles, UAV)技术的突破和无人机系统(unmanned aircraft systems, UAS)的多样性开发，其在军事和民用领域都得到广泛应用。目前，中国UAV产业呈现高速发展态势，据统计[1]：相较于2022年，2023年UAV产业的年复合增长率约达13.8%，全国UAV企业已超过1.5万家，UAV注册数也已超过100万架。特别是，2024年3月，中国将“低空经济”写入政府工作报告，预示着国家对低空经济的鼓励发展，未来UAV在城市低空场景中的使用将显著增加。

在UAV被大规模投入应用之前，其运行给地面带来的安全风险得到了国内外空管部门和学者的关注，尤其对于人口密度较高的城市复杂环境，UAV的运行会给城市地面的公共安全带来巨大的安全隐患。UAV运行的地面风险是指UAS失效坠地后，对地面人员和地面基础设施的威胁和危害程度[2]。通过梳理近年来国内外对UAV运行地面风险的研究，发现UAV安全运行的风险研究方法主要包括定性分析和定量评估[3]2种。

在定性分析方面，无人系统规则制定联合体(joint authorities for rulemaking of unmanned system, JARUS)开发了一种基于风险意识的特定类型UAV运行风险评估(specific UAV operations risk assessment, SORA)方法和UAS管理方法，SORA对特定类别UAV运行有关的地面风险进行评估，并提供了一套完整的程序框架，可最终确定获得可飞许可的特定类型UAV作业风险是否已达到可接受的空域安全水平[4]。在此基础上，2018年，Denney等[5]提出了一种初步的、基于数学的风险评估方法，通过扩展领结图(bow tie diagram, BTD)完善安全概率

收稿日期：2024-04-23；**录用日期**：2024-07-12；**网络出版时间**：2024-08-06 10:20
网络出版地址：link.cnki.net/urlid/11.2625.V.20240805.1413.002
基金项目：国家空管监视与通信系统工程技术研究中心开放研究基金 (w222394)
* **通信作者**. E-mail：yushasha@mail.xhu.edu.cn

引用格式：陈艺君，余莎莎，张学军. 城市低空场景下无人机运行对地风险量化评估[J]. 北京航空航天大学学报，2025，51(3)：806-815. CHEN Y J，YU S S，ZHANG X J. Ground risk quantitative assessment for UAV operations in urban low-altitude scenarios[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：806-815 (in Chinese).

模型, 并利用贝叶斯框架扩展和增强初步概率形式。2021 年, Martinez 等[6] 扩展了 SORA 单个 UAS 的运行安全流程框架, 并提出了一种基于多无人机安全评估的综合流程。Wyszywacz[7] 在 2022 年提出了一种用于特定类型 UAS 的简易风险评估(easy risk assessment, ERA)方法, 通过一种相对 SORA 更简单的方式来确定 UAS 运行的威胁和来源, 提供相应的风险管理并确定风险容忍度水平。但是, SORA 仅向各国和研究者提供了定性的风险分析框架和目标安全水平, 缺乏有效风险的量化研究。在推进“低空经济”发展的道路上, 仅仅开展定性的风险分析是远远不够的。

定量评估[8] 是指在 UAV 作业之前或事故之后, 针对 UAV 潜在的或实际的事故发生概率及其相关影响后果的严重程度, 利用科学的计算方法和精确的模型进行量化估计。国内外部分学者在该研究方面提出了自己的方案和思考。王文涛[9] 和 Jiao[10] 等提出了根据 UAV 下降过程的动力学模型来推导其坠地影响区域, 进而量化地面风险值。其中, Jiao 等[10] 提出了一个基于深度学习的动态地面风险计算方法, 该方法利用卷积神经网络和深度学习联合模型预测地面人口密度时空分布特征, 通过建立深度学习和动力学模型相结合的动态模型预测地面撞击伤亡率, 量化地面风险, 为地面人员死亡风险量化研究提供了重要参考价值。韩鹏和张冰玉[11] 建立了坠地撞击点预测模型, 以分析不同航迹误差对坠地时间、地面撞击点位置和坠地动能的影响, 其对 UAV 地面撞击点位置分布特征和坠地动能的研究, 为 UAV 运行的地面风险评估方法的精确化提出了新思路。Shao[12] 通过飞行试验计算了空中交通管理系统环境下 UAV 进行物流配送的预期安全水平及其风险等级, 对 UAV 风险进行了有效评估, 并降低了运行风险, 证明了 UAV 运行中地面风险评估的重要性和必要性。此外, 文献 [13-14] 的研究团队近年来致力于研究 UAV 下降事件对地面风险的影响, 进一步证明了区域环境因素和 UAV 自身参数对地面风险评估的影响。新加坡在 2017 年正式提出服务城市环境的 UAV 交通管理系统(urban traffic management-UAV systems, u TM-UAS)运行概念[15], 随后, 新加坡民航局(Civil Aviation Authority of Singapore, CAAS)联合新加坡南洋理工大学致力于从空域利用、风险管理、功能系统 3 个需求点出发, 搭建 u TM-UAS 系统, 在城市风险管理需求中, 明确阐明, 建立适应城市环境需要的技术和风险管理以达到较精确的风险预警、风险管控及风险评估。2022 年, 新加坡南洋理工大学的 Pang 等[16] 提出了一种综合风险评估模型, 对第三方风险(third party risk, TPR)问题指标在建模和评估水平上均进行了扩展, 模型纳入噪声影响指标, 并证明了基于地面风险成本的路径规划可有效规避运行的高风险区域, 实现 UAV 的安全性运行, 地面区域特征分析和空域栅格化概念的提出给地面风险研究带来了新展望。

2020 年, 都灵理工大学研究团队的 Primatesta 等[17] 提出了使用风险地图来定义与 UAV 飞行作业相关的地面风险。同年, 胡莘婷和戴福青[18] 以简单的风险数学计算式构建风险量化评估模型, 也给出了风险地图的概念。随后, 其在文献 [19] 中基于成本-效益对城市环境下 UAV 运行风险进行了量化评估, 但是未考虑 UAV 失效下降运动中真实的地面撞击位置分布, 且城市人口密度分布特征也仅是用低分辨率的区域人口统计数据表示。曲欣宇等[20] 基于风险地图进行 UAV 安全路径规划研究, 但风险地图的生成未考虑城市低空环境特征。Primatesta 等[17] 在风险评估方法研究中, 考量了机场禁飞区域的划设范围对地面风险的影响。Li 等[21] 在对城市环境中 UAV 的风险量化研究中, 通过考量多个潜在风险源, 重点关注地面人员安全的潜在威胁。

基于研究现状, 国内外研究均在 UAV 的安全运行风险评估方法研究中取得了一些进展, 但是大多是基于风险评估方法、模型和基于风险成本的路径规划算法方面的创新, 对风险地图的生成多采用简单的矩阵定义, 未充分考虑城市地面典型特征和城市低空场景特征。而对城市地面和低空环境各影响因素量化的忽视, 以及对 UAV 失效坠地事件中不确定性参数分析的忽略, 将会导致评估风险与实际风险之间出现显著的误差。这种误差会影响生成的地面风险地图的精确度和可靠性, 使其在为飞行任务、低空空域管理及划设提供决策依据时存在潜在的缺陷, 而安全是未来空中交通系统必须坚守的底线。

基于上述分析, 本文对 UAV 城市低空运行场景下的对地风险展开研究。根据城市地面和低空环境特征、UAV 坠地事件的因果链分析和不确定性参数分析, 对城市地面风险的多影响因素进行量化, 分析各影响因素的对地风险作用机理, 以构建多影响因素下的地面风险量化评估模型和多图层独立模型, 并通过多图层融合生成地面风险地图。

1　地面风险量化评估模型

地面风险是指 UAV 低空飞越特定区域时, 由于系统空中失效而撞击地面, 对地面人口造成的伤

亡威胁和严重程度。类似民航领域的事故经典衡量方法，本文的风险值用每飞行小时的伤亡人数表示。

UAV失效撞击地面致人员伤亡的事件发生，经历了3个条件事件：①UAV任务过程系统失效或失控；②UAV撞击地面后与至少一人发生撞击事件；③被撞击者死亡事件。因此，UAV失效撞击地面事故的伤亡概率[21]可表达为

$$R(x,y)=\iint_{(x,y)\subset A_{\exp}}\lambda_{\text{event}}P_{\text{collision}}P_{\text{fatality}}\mathrm{d}x\mathrm{d}y \tag{1}$$

式中：R为伤亡概率，即风险值，定义为每飞行小时的伤亡人数；λ_{event}为UAV坠地事故发生率；$A_{\exp}$为UAV撞击地面后的影响区域范围大小，即杀伤面积，m^2；(x,y)为地面地理参考位置；P_{fatality}为撞击地面致人员伤亡的概率；$P_{\text{collision}}$为UAV坠毁于地面与至少一人发生碰撞的概率。

1.1 UAV失效坠地事件模型

UAV坠地的撞击点位置与无人机的失效下降行为密切相关[14]。在失效坠地阶段，UAV失去所有推力，将以弹道下降事件行为撞击地面。为精确求解UAV坠地撞击区域的概率密度函数（probability density function, PDF），缩小风险评估实际误差，需要根据UAV失效坠地事件行为，并考虑不确定性参数，模拟UAV坠落轨迹。图1为UAV下降过程的受力情况和非确定性风参数影响下的可能坠地轨迹。

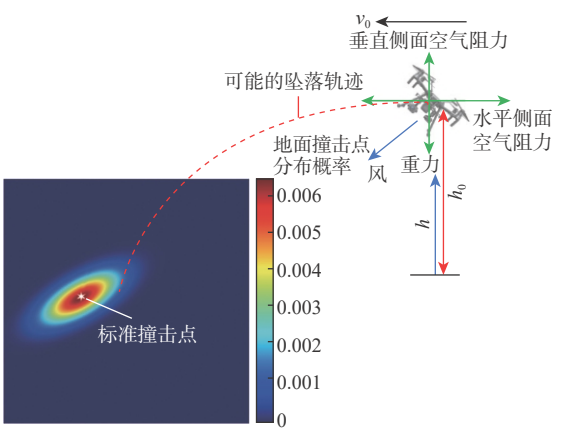


图1 UAV下降轨迹

Fig. 1 UAV descent trajectory

分析可知，UAV失效坠地轨迹受空气动力学和风运动的约束。

在空气动力学模型中，UAV的坠地运动由重力、阻力决定，根据牛顿第二定律，其坠落运动被分为水平和垂直2个方向的运动，表达式为

$$\begin{cases}\boldsymbol{F}_{\text{d}}=m\boldsymbol{a}\\ m\boldsymbol{a}_x=-\boldsymbol{D}_x\\ m\boldsymbol{a}_z=\boldsymbol{G}-\boldsymbol{D}_z\end{cases} \tag{2}$$

式中：$\boldsymbol{F}_{\text{d}}$和$\boldsymbol{a}$分别为UAV失效坠地运动的动力和加速度；$m$为UAV的质量，kg；$\boldsymbol{a}_x$和$\boldsymbol{a}_z$分别为水平和垂直方向的加速度，$\mathrm{m/s}^2$；$\boldsymbol{D}_x$和$\boldsymbol{D}_z$分别为UAV水平和垂直2个不同侧面的空气阻力；$\boldsymbol{G}$为重力。

在风运动模型中，UAV的运动仅受风力影响[17]，可以建模为

$$\boldsymbol{x}_{\text{wind}}=\int_0^{t_{\text{drop}}}\boldsymbol{v}_{\text{wind}}\mathrm{d}t \tag{3}$$

式中：$\boldsymbol{x}_{\text{wind}}$为UAV在风力作用下的位移；$\boldsymbol{v}_{\text{wind}}$为风速；$t_{\text{drop}}$为无人机坠地时间。

由此，无人机失效下降阶段的实际水平移动距离$\boldsymbol{x}_{\text{d}}$是$\boldsymbol{x}$和$\boldsymbol{x}_{\text{wind}}$运动状态的组合，可表达为

$$\boldsymbol{x}_{\text{d}}=\boldsymbol{x}+\boldsymbol{x}_{\text{wind}} \tag{4}$$

无人机的地面撞击点分布受航迹误差的影响[11]，在考虑航迹误差的影响下，UAV失效坠地的撞击点分布概率可以用$\boldsymbol{x}_{\text{d}}$的二维PDF（two-dimensional PDF, 2D PDF）建模。

图2为M210-RTK失效坠地事件的2D PDF。图中，红色区域表示UAV坠落概率最高，深蓝色区域表示撞击该区域的概率为0。

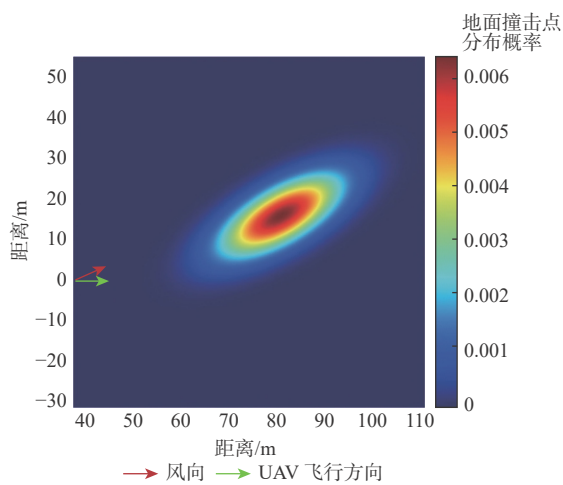


图2 M210-RTK下降事件的2D PDF

Fig. 2 2D PDF of M210-RTK descent event

1.2 撞击事件模型

与至少一人发生碰撞的概率$P_{\text{collision}}$，取决于遭受坠机影响的区域和撞击区域的人口密度[18]，计算式为

$$P_{\text{collision}}(x,y)=\sum_{(x,y)\in f_{\text{PDF}}}f_{\text{PDF}}\rho(x,y)A_{\exp} \tag{5}$$

式中：$\rho(x,y)$为在地面地理参考位置(x,y)处的人口密度，人/m^2；f_{PDF}为撞击区域的概率密度函数。式(5)中求和运算符考虑了概率影响区域内的所有地理参考单元及其对应位置单元在f_{PDF}下的概率。

UAV撞击地面的影响区域[22]（又称杀伤区域）形似胶囊状，如图3所示，可表示为

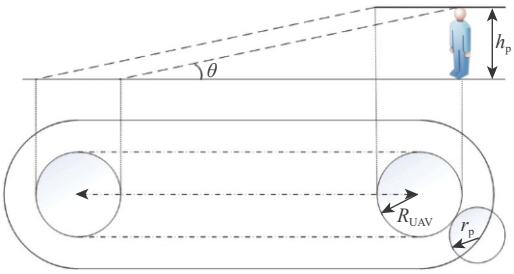


图 3　UAV 坠落事故影响区域示意图

Fig. 3　Schematic diagram of impact area in UAV fall accident

$$A_{exp}=2(r_p+R_{UAV})h_p\cot\theta+\pi(r_p+R_{UAV})^2 \quad (6)$$

式中: r_p 和 R_{UAV} 分别为被撞击人员的半径和坠落 UAV 的最大半径, m; θ 为 UAV 坠落角度; h_p 为地面被撞击人员的平均高度, m。

1.3　伤亡事件模型

地面人员被撞击伤亡的概率, 与地表树木、建筑物等其他障碍物的遮蔽效果和 UAV 撞击地面时的动能密切相关[22], 可表示为

$$P_{fatality}(x,y)=1\Big/\left[1+\sqrt{\frac{\alpha}{\beta}}\left(\frac{\beta}{E_{imp}(x,y)}\right)^{\frac{1}{4P_s(x,y)}}\right] \quad (7)$$

式中: $P_{fatality}(x, y)$ 为在地面地理参考位置 (x, y) 处的人员伤亡概率, (x, y) 在 UAV 撞击地面的 PDF 影响范围内; $E_{imp}(x, y)$ 和 $P_s(x, y)$ 分别为 UAV 撞击 (x, y) 区域时, 对应的撞击动能和遮蔽因子; α 为遮蔽系数为 0.5 时死亡率达到 50% 所需的撞击能量, J; β 为当 P_s 趋于 0 时发生死亡所需要的撞击能量阈值。

遮蔽因子是一个参数, 决定地面人群如何受障碍物遮挡, 避开 UAV 的撞击。由于遮蔽效果直接影响 UAV 撞击地面时的动能大小, 对遮蔽因子的精确量化非常重要。

2　地面风险地图

风险地图用于直观可视化地展示 UAV 飞行中给地面不同区域带来的风险等级分布。生成地面风险地图的步骤如图 4 所示。首先, 考虑 UAV 参数、风因素和城市低空环境特征, 映射 UAV 下降事件下的地面撞击位置 PDF 分布特征; 其次, 以城市人口密度分布和地面遮蔽效应作为输入, 求解地面风险量化评估模型, 得到人口伤亡风险层; 然后, 考虑 UAV 飞行高度和城市典型特征, 再次输入城市禁飞区、障碍物和特殊区域数据, 使用 MATLAB 软件进行多层数据融合; 最后, 生成地面风险地图。地面风险地图是基于地理数据栅格化的二维地图, 其中, 每个栅格单元都包含对应的风险值。

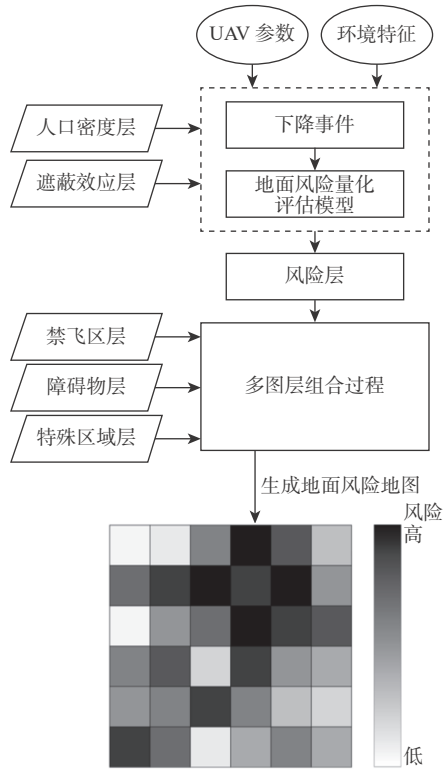


图 4　地面风险地图生成步骤

Fig. 4　Steps for generating ground risk map

2.1　多影响因素层分析

地面风险地图通过多图层组合生成, 其中, 每层都基于地理参考位置, 包含所需数据的地图, 这些图层具有与地面风险地图相同的维度。

2.1.1　人口密度层

人口密度层用于定义区域内人口密度的分布。人口密度数据是风险评估中的重要因素之一, 直接影响 UAV 撞击地面人员的概率, 因此, 具有适当精细分辨率的人口密度网格是降低风险评估误差的关键。依据中国民用航空局文件[23] 中提到的最佳空域栅格划分规范, 本文选取 100 m×100 m 精度的分辨率, 来估计区域人口密度, 并生成人口密度栅格图。对获取的世界人口数据、中国第 7 次人口普查数据使用 ArcGis 软件进行研究区域全区和分区处理, 以获得真实的百米量级人口密度分布数据, 该人口密度分布数据与 UAV 地面撞击杀伤面积的百米量级相匹配, 有助于风险的精确量化。

2.1.2　遮蔽效应层

遮蔽效应层用于定义地表障碍物和基础设施等对人群的庇护水平, 是用遮蔽因子来量化区域内其他对象的遮挡级别。如图 5 所示, 遮挡因子是 [0,1] 之间的实数。无遮蔽物遮挡时, 定义因子为 0; 依次定义被稀疏树木遮挡区域、被树丛或低矮建筑物庇护区域、高耸建筑物区和工业区的因子分别为 0.25、0.50、0.75 和 1.00。



图 5 区域遮蔽效应层

Fig. 5 Regional sheltering factor layer

2.1.3 禁飞区层

禁飞区层用于识别地图上出于安全性而划定的禁止各类飞行器在上空飞越的区域，由政府强制性执行，通常指城市机场区、军事区域及国家监管机构等重点防控目标区。国内禁止飞行区域划设面范围标准部分明确规定：城市机场和军用机场执行跑道端外 20 km 范围内的净空保护区域；军事管理区、监管机构等重点防控目标区，周边 100 m 范围内禁飞。

图 6 为成都市禁飞区示例，数据来源于大疆的安全飞行指引 GEO 系统数据库。禁飞区层是一个类似风险图的地图，也是一个采用 0 和 1 两类值表示的矩阵，记为 $\boldsymbol{F}$。$\boldsymbol{F}$ 接受地面地理参考位置 (x, y) 作为输入，并返回一个值，表示该位置是否位于禁飞区域。若返回 0 值，表示处于禁飞区域，禁止飞行；若返回 1，则表示位于禁止飞越的区域以外，允许飞行。



图 6 成都市禁飞区示例

Fig. 6 Example of no-fly zones around city of Chengdu

由于城市机场禁飞区范围较大，本文算例分析中的禁飞区图层将 2 个主要的城市中心广场定义为假想的禁飞区，在城市地图上被突出显示。

2.1.4 障碍物层

障碍物层用于定义 UAV 特定飞行高度下，被建筑物、树木、基础设施等障碍物占据的栅格单元区域。城市低空环境中，密集的高层障碍物给 UAV 的飞行带来了碰撞后的坠地风险。如图 7 所示，使用 MATLAB 软件仿真，以 6 km 范围内的建筑物分布特征为例，在 30～60 m 层高度下，影响 UAV 飞行的建筑障碍物多且密集；而相比之下，在 120 m 以上的高空层，影响 UAV 飞行的建筑障碍物则很少。

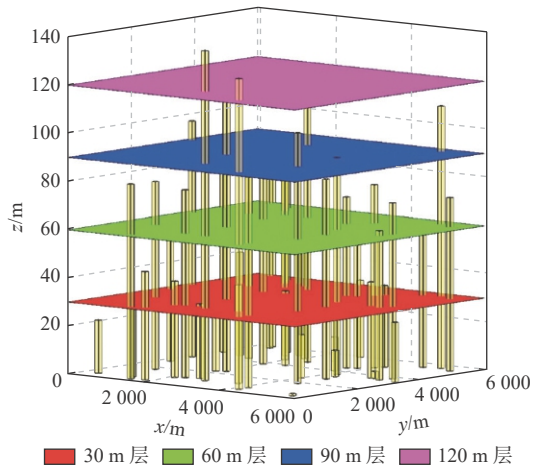


图 7 建筑物高度分布

Fig. 7 Building height distribution

本文中，障碍物层被看作一个与区域地图维度类似的矩阵 $\boldsymbol{B}$。每个元素单元都包含对应地理位置的障碍物垂直高度数据，如图 8 所示，使用世界开源地图 OpenStreetMap 提取城市三维模型，并基于 ArcGis 软件对数据进行处理，生成区域内障碍物分布图层。



图 8 障碍物层

Fig. 8 Obstacle layer

2.1.5 特殊区域层

特殊区域层用于定义城市大型枢纽车站、医院、学校和加油站等易引发二次事故的特殊区域。二次伤害的概念在文献 [23] 中被中国民用航空局

正式提出, 并给出了特殊区域风险程度的等级分类。特殊区域风险等级分类如表 1 所示。这些特殊区域若发生 UAV 撞击事故, 将引发二次交通事故和区域爆炸, 地面伤害将进一步扩大和加重。如图 9 所示, 以加油站特殊区域为例, 为避免 UAV 撞击地面发生二次伤害事故, 定义地面特殊区域的安全范围是一个以特殊区域为中心、UAV 下降的水平移动距离 $\boldsymbol{x}_{\mathrm{d}}$ 为半径的圆形或椭圆形地面区域及其空域范围。可以根据下降事件模型对其暴露区域半径进行评估, 因此, 地面特殊区域的暴露范围是动态的。

表 1　特殊区域风险等级分类

Table 1　Classification of risk levels in special areas

特殊区域	风险等级
发电厂、大型车站、易燃易爆仓库、易发生火灾和水污染的区域	高风险
大型加油站、中型车站、中型码头和港口区域	中风险
医院、学校等人群聚集区	低风险
其他正常区域	安全

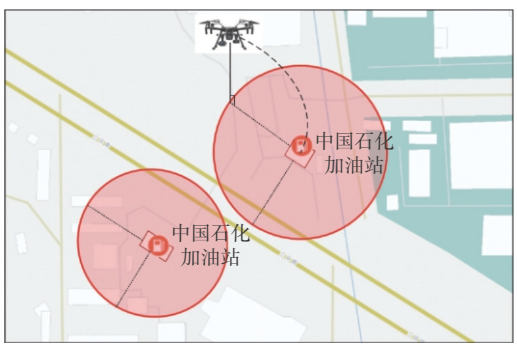


图 9　特殊区域的安全飞行范围

Fig. 9　Safe flying ranges for special areas

为充分验证不同风险级别的城市地面特殊区域对 UAV 安全飞行的潜在风险的作用机理, 本文算例分析中的特殊区域层分别选择了存在潜在高风险的大型车站、中风险的大型加油站及低风险的中小学区域 3 类级别潜在风险区域。

2.2　安全目标水平

由于风险管理的最终目标是安全目标水平 (target level of safety, TLS)[24], 民航领域将其定义为每飞行小时内事故发生的次数。同样, 在 UAV 运行风险评估程序中, JARUS[25] 将 UAV 每飞行小时造成的地面伤亡人数作为等效安全水平 (equivalent level of safety, ELS), 用于和 TLS 比较是否达到安全标准。

依据中国民用航空局航空器适航审定司发布的文件[26], TLS 取值为 10^{-6} 人/h, ELS 值与 TLS 值比较, 比值越高, 潜在风险越高, UAV 对地面人员伤害程度越高。风险等级评定指标如表 2 所示。

表 2　风险等级评定指标

Table 2　Risk levels rating indicators

风险值	风险等级
<1	高风险
$<10^{-4}$	中风险
$<10^{-5}$	低风险
$<10^{-6}$	安全或无风险

2.3　多图层的合并过程

基于地面风险量化评估模型进行地面人员伤亡风险评估之后, 将生成的人员伤亡风险图层与禁飞区层、障碍物层及特殊区域层合并, 获得最终的地面风险地图。该过程对地图中每个栅格单元的存储数据进行合并, 过程如下:

$$R(x,y)=\begin{cases}0 & \boldsymbol{F}(x,y)=0\cup\boldsymbol{B}(x,y)>h\\ r_{\boldsymbol{S}}(x,y) & r_{\boldsymbol{S}}(x,y)>r_{\boldsymbol{R}}(x,y)\\ R(x,y)\to r_{\boldsymbol{R}}(x,y) & \text{其他}\end{cases}\tag{8}$$

式中: $\boldsymbol{F}$ 和 $\boldsymbol{B}$ 分别为禁飞区和障碍物 2 个图层的函数表示, 0 表示由于障碍物和禁飞规定而禁止飞行; $r_{\boldsymbol{S}}(x,y)$ 和 $r_{\boldsymbol{R}}(x,y)$ 分别为特殊区域层 (x,y) 对应位置处的风险等级和人员伤亡风险层 (x,y) 对应位置处的风险等级; $\boldsymbol{r}$ 为与风险地图维度相同的风险等级矩阵, 风险值矩阵 $\boldsymbol{R}$ 与风险等级矩阵 $\boldsymbol{r}$ 可实现互相转化。

3　算例分析

3.1　参数设置

选取成都市某区规格 2 km×2 km 大小的城市区域作为本文算例分析场景。对该研究区域进行栅格化分, 如图 10 所示, 形成 20×20 的网格图, 每个栅格单元的边长为 100 m。

本文选择在城市应用中具有代表性的经典机型 M210-RTK 进行算例分析, 该机型相关参数设置如表 3 所示。

该研究区域的风参数分布数据是从开源气象数据库 rp5 收集得到的 3 月份实时数据, 风向和风速分布如图 11 所示。

3.2　生成地面风险地图

依据第 1 节所构建的地面风险量化评估模型, 结合城市 GIS 数据、人口密度分布数据及区域遮

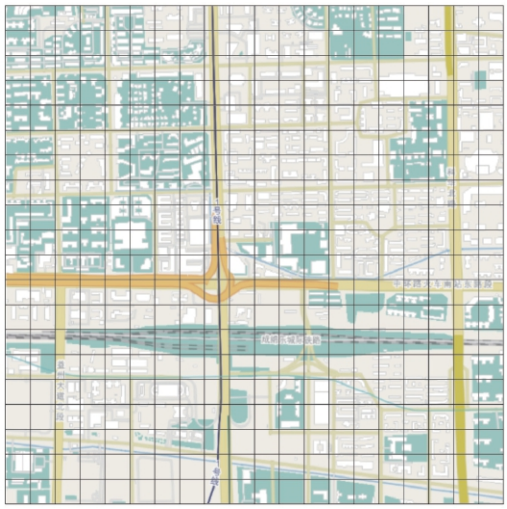


图 10 研究区域栅格化

Fig. 10 Rasterization of study area

表 3 模型参数

Table 3 Model parameters

λ_{event}	m/kg	v_0/(m·s^{-1})	h_0/m	g/(m·s^{-1})	空气阻力系数C_d
6.71×10^{-6}	6.14	20	N(100, 2)	9.8	0.4
$\lvert\boldsymbol{v}_{wind}\rvert$/(m·s^{-1})	r_p/m	R_{UAV}/m	h_p/m	$\lvert\boldsymbol{v}_{max}\rvert$/(m·s^{-1})	β/J
N(10, 2)	0.25	0.44	1.65	23	34

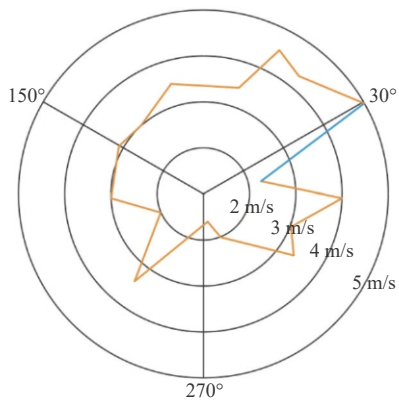


图 11 研究区域的风分布

Fig. 11 Wind distribution of study area

蔽效应分布特征，通过多影响因素融合，应用MATLAB软件仿真得到研究区域内的地面人口伤亡风险图，如图 12 所示。伤亡风险分布图显示了不同区域内 UAV 坠地给地面人群造成的潜在伤害程度，其中，该区域内的平均伤亡风险为 4.98×10^{-6} 人/h，与取值为 10^{-6} 人/h 的 TLS 值在同一数量级上，表示该区域内 100 m 高空飞行的无人机 M210-RTK 致地面人群死亡概率较低，且达到中国民用航空局规定的安全目标水平。

为降低风险评估误差，UAV 在城市低空场景下运行，还需要充分考虑城市低空特征和城市地面典

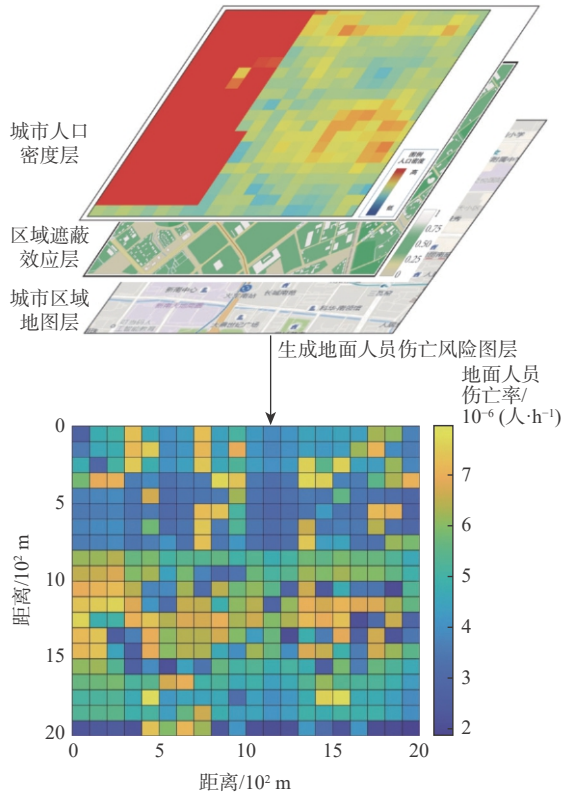


图 12 地面人员伤亡风险层生成过程框架

Fig. 12 Framework of the process of generating the ground fatality risk layer

型特征。将城市禁飞区、障碍物高度分布特征及城市地面易造成二次伤害的特殊区域分布情况纳入地面风险量化评估模型中，通过融合多图层（见式(8)），即可得到 UAV 在城市低空场景下运行的地面风险地图。地面风险地图生成过程及结果如图 13 和图 14 所示。

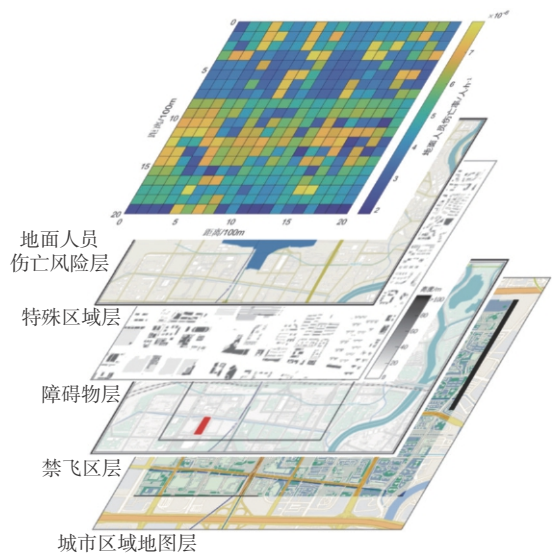


图 13 多层叠加式的地面风险地图生成过程

Fig. 13 Process of generating ground risk map with multiple layers

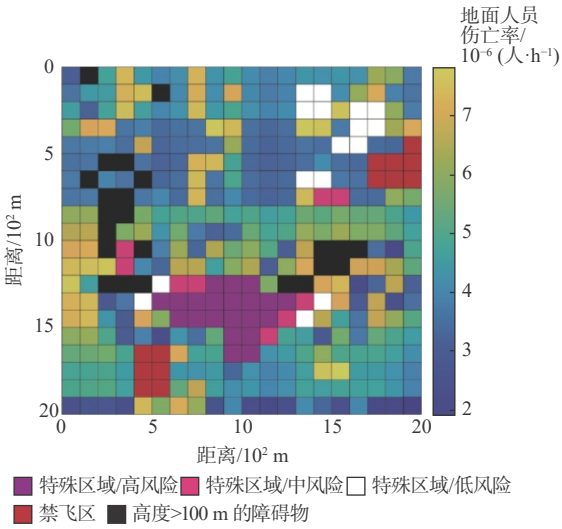


图 14　地面风险地图

Fig. 14　Ground risk map

4　结　论

1）UAV 在系统失效的坠地事件中，其坠落地面的概率密度分布对于确定危害地面人员的概率非常重要，这是通过将其坠地事件的 PDF 与研究区域的人口密度分布和遮蔽效应相关联起来实现的。UAV 从事故发生点到撞击地面的水平移动距离和最终地面撞击位置，不仅受 UAV 自身参数影响，还受不同区域内的实际风向和风速的影响。

2）UAV 坠落地面致人员伤亡的概率和危害程度，与城市区域内的人口密度分布特征、对应地理参考位置的地面遮蔽效应密切相关。城市地面人群越密集，UAV 坠地后的伤亡率越高；对应地理参考位置处的遮蔽因子越大，地面要素对人群的庇护能力越强，人员伤亡概率越低，危害程度越低。

3）UAV 在城市低空场景飞行时，其对地面的风险与城市地面典型特征（包括城市禁飞区、障碍物高度分布特征、易引起二次伤害的特殊区域）密切相关。对城市地面典型特征的考虑，将有助于降低 UAV 运行的地面风险评估误差，得到更精确、可靠和完整的地面风险地图。城市机场及其周边是常见的禁飞区域，不仅划设禁止飞行区域，还划设限高区范围，是需要 UAV 飞行操作严格执行的，UAV 的误入将带来极大的撞机风险；在被城市树木、建筑物、基础设施等占据的空域单元，UAV 为避免撞击风险，是需要规避不同飞行高度内的障碍物群的；UAV 在城市低空运行，若发生失效坠地事故，有撞击城市地面大型车站、加油站及人群密集的学校等区域的风险，将极易造成二次事故，扩大对地伤害，必须按照中国民用航空局标准，规范高、中、低、无风险区域，实现飞行规划中规避高风险区域，降低对地风险。

4）本文的无人机安全运行的地面风险地图生成方法，可以用于量化更大城市区域的地面风险，实现风险可视化，为飞行任务提供规避风险的决策支持。

安全是未来城市空中交通系统发展的前提和底线，对地风险评估为空域划设和航路规划提供着关键的科学依据。因此，针对无人机运行对地风险评估的研究需要持续深入。目前，地面风险地图多以静态形式展示，随着低空经济的发展，未来对地风险评估可以考虑到城市地面不同时段的人、车时空流动性和气象变化性，以得到多时段、变化气象条件下的动态风险分布地图。并且考虑依据不同机型坠地特点，尤其在伞降等风险缓解装置配备条件下，开展多机型对地风险量化评估研究。此外，还需要进一步延展到空中风险，以提高未来城市空中飞行安全。

参考文献（References）

[1] 陈志杰. 依法助推低空经济发展[EB/OL]. (2023-06-29)[2024-03-24]. https://www.gov.cn/zhengce/202306/content_6888958.htm.
CHEN Z J. Promoting low-altitude economic development in accordance with the law[EB/OL]. (2023-06-29)[2024-03-24]. https://www.gov.cn/zhengce/202306/content_6888958.htm(in Chinese).

[2] XIA C Y, YANG C R, XUE K, et al. A conflict risk analysis of MAV\UAV flight in shared airspace[J]. International Journal of Aerospace Engineering, 2021, 2021: 1692896.

[3] 韩鹏, 赵嶷飞, 刘宏. 无人机地面撞击风险评估体系构建及趋势展望[J]. 中国民航大学学报, 2021, 39(1): 40-47.
HAN P, ZHAO Y F, LIU H. Assessment system construction and trend foresight of UAV ground impact risk[J]. Journal of Civil Aviation University of China, 2021, 39(1): 40-47(in Chinese).

[4] MELNYK R, SCHRAGE D, VOLOVOI V, et al. A third-party casualty risk model for unmanned aircraft system operations[J]. Reliability Engineering & System Safety, 2014, 124: 105-116.

[5] DENNEY E, PAI G, JOHNSON M. Towards a rigorous basis for specific operations risk assessment of UAS[C]//Proceedings of the IEEE/AIAA 37th Digital Avionics Systems Conference. Piscataway: IEEE Press, 2018: 1-10.

[6] MARTINEZ C, SANCHEZ-CUEVAS P J, GERASIMOU S, et al. SORA methodology for multi-UAS airframe inspections in an airport[J]. Drones, 2021, 5(4): 141.

[7] WYSZYWACZ W. Easy risk assessment for unmanned aircraft systems: outline of the method[J]. Transactions on Aerospace Research, 2022, 22(3): 32-47.

[8] HAN P, YANG X Y, ZHAO Y F, et al. Quantitative ground risk assessment for urban logistical unmanned aerial vehicle (UAV) based on Bayesian network[J]. Sustainability, 2022, 14(9): 5733.

[9] 王文涛, 甘旭升, 吴亚荣, 等. 考虑不确定性的低空无人机运行风险评估方法[J]. 现代防御技术, 2022, 50(5): 14-21.
WANG W T, GAN X S, WU Y R, et al. Low-altitude UAV opera-

tion risk assessment method considering the uncertainty[J]. Modern Defence Technology, 2022, 50(5): 14-21(in Chinese).

[10] JIAO Q Y, LIU Y S, ZHENG Z G, et al. Ground risk assessment for unmanned aircraft systems based on dynamic model[J]. Drones, 2022, 6(11): 324.

[11] 韩鹏, 张冰玉. 航迹误差对无人机坠地伤人风险评估的影响[J]. 中国安全科学学报, 2021, 31(2): 106-111.
HAN P, ZHANG B Y. Effect of track error on safety risk assessment of UAV ground impact[J]. China Safety Science Journal, 2021, 31(2): 106-111(in Chinese).

[12] SHAO P C. Risk assessment for UAS logistic delivery under UAS traffic management environment[J]. Aerospace, 2020, 7(10): 140.

[13] LA COUR-HARBO A. Quantifying risk of ground impact fatalities for small unmanned aircraft[J]. Journal of Intelligent & Robotic Systems, 2019, 93(1): 367-384.

[14] LA COUR-HARBO A. Ground impact probability distribution for small unmanned aircraft in ballistic descent[C]//Proceedings of the International Conference on Unmanned Aircraft Systems. Piscataway: IEEE Press, 2020: 1442-1451.

[15] KIN H L.Framework for urban traffic management of unmanned aircraft system (u TM-UAS)[EB/OL]. (2017-09-22)[2024-03-28]. https://www.icao.int/Meetings/UAS2017/Documents/Kim%20Huat%2OLo_Singapore_UTM_%20Day%201.pdf.

[16] PANG B Z, HU X T, DAI W, et al. UAV path optimization with an integrated cost assessment model considering third-party risks in metropolitan environments[J]. Reliability Engineering & System Safety, 2022, 222: 108399.

[17] PRIMATESTA S, RIZZO A, LA COUR-HARBO A. Ground risk map for unmanned aircraft in urban environments[J]. Journal of Intelligent & Robotic Systems, 2020, 97: 489-509.

[18] 胡莘婷, 戴福青. 基于城区行人安全的无人机运行风险评估[J]. 中国安全科学学报, 2020, 30(8): 137-142.
HU X T, DAI F Q. Risk assessment model for UAV operation considering safety of ground pedestrians in urban areas[J]. China Safety Science Journal, 2020, 30(8): 137-142(in Chinese).

[19] 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.

[20] 曲欣宇, 叶博嘉, 王华中, 等. 基于风险地图的城市物流无人机路径规划研究[J]. 航空计算技术, 2024, 54(1): 87-91.
QU X Y, YE B J, WANG H Z, et al. Research on global path planning method for urban logistics UAV based on risk maps[J]. Aeronautical Computing Technique, 2024, 54(1): 87-91(in Chinese).

[21] LI Q Y, WU Q G, TU H Y, et al. Ground risk assessment for unmanned aircraft focusing on multiple risk sources in urban environments[J]. Processes, 2023, 11(2): 542.

[22] LIU Y, ZHANG X J, WANG Z, et al. Ground risk assessment of UAV operations based on horizontal distance estimation under uncertain conditions[J]. Mathematical Problems in Engineering, 2021, 2021: 3384870.

[23] 中国民用航空局航空器适航审定司. 城市场景轻小型无人驾驶航空器物流航线划设规范[EB/OL]. (2022-07-29)[2024-03-28]. http://www.caac.gov.cn/big5/www.caac.gov.cn/XXGK/XXGK/BZGF/HYBZ/202208/P020220811600885528858.pdf.
Aircraft Airworthiness Certification Department, Civil Aviation Administration of China. Route design specification of the light-small unmanned aircraft system for urban logistics[EB/OL]. (2022-07-29)[2024-03-28]. http://www.caac.gov.cn/big5/www.caac.gov.cn/XXGK/XXGK/BZGF/HYBZ/202208/P020220811600885528858.pdf (in Chinese).

[24] 朱元军, 李妍, 高子昂. 可接受风险水平下城市空域无人机路径规划方法研究综述[J]. 西华大学学报 (自然科学版), 2022, 41(1): 7-12.
ZHU Y J, LI Y, GAO Z A. A review of UAS route planning in urban airspace with acceptable risk levels[J]. Journal of Xihua University (Natural Science Edition), 2022, 41(1): 7-12(in Chinese).

[25] Joint Authorities for Rulemaking of Unmanned Systems. JARUS guidelines on specific operations risk assessment (SORA) V2.0 [EB/OL]. (2019-01-30)[2024-03-28]. http://jarus-rpas.org/wp-content/uploads/2024/06/SORA-v2.5-Main-Body-Release-JAR_doc_25.pdf.

[26] 中国民用航空局航空器适航审定司. 民用无人驾驶航空器系统适航审定分级分类和系统安全性分析指南[EB/OL]. (2022-11-21)[2024-03-28]. http://www.caac.gov.cn/XXGK/XXGK/GFXWJ/202302/P020230213505366854054.pdf.
Aircraft Airworthiness Certification Department, Civil Aviation Administration of China. Guidelines for civil unmanned aerial vehicle system airworthiness validation classification and system safety analysis [EB/OL]. (2022-11-21)[2024-03-28]. http://www.caac.gov.cn/XXGK/XXGK/GFXWJ/202302/P020230213505366854054.pdf. (in Chinese).

Ground risk quantitative assessment for UAV operations in urban low-altitude scenarios

CHEN Yijun[1, 2], YU Shasha[1, 3, *], ZHANG Xuejun[3, 4]

(1. School of Aeronautics and Astronautics, Xihua University, Chengdu 610039, China;
2. National Center of ATC Surveillance and Communication System Engineering Research, Chengdu 610039, China;
3. Beihang University International Center for Innovation in Western China, Chengdu 610039, China;
4. School of Electronic Information Engineering, Beihang University, Beijing 100191, China)

Abstract: A ground risk assessment model for unmanned aerial vehicle (UAV) operation was developed in order to guarantee the safe operation of UAVs in low-altitude urban areas. It was proposed that a ground risk map that considered multi-factors and multi-levels be used to define the ground risk associated with UAVs and quantify the degree of risk to the ground from flight operations over different areas. First, the typical characteristics of urban low-altitude and the causal chain of UAV crash-landing accidents were analyzed to establish a quantitative ground risk assessment model under the influence of multiple factors, and analyze in detail the mechanism of the influence of each factor on the ground risk. Second, to determine the chance of the danger happening at various locations, the probability density function of the impact region of a UAV crash-landing was computed while accounting for the UAV's descending behavior and the uncertainty of the wind parameter. Then, the risk value of each geographic reference unit was calculated by combining multiple layers such as urban area map, population density, shelter effect, obstacles and no-fly zones; then the ground risk map of a large regional scale was generated, which can intuitively show the risk level of different areas of the city. Finally, the proposed method was validated and analyzed based on real examples. The findings demonstrate that there are clear differences in the risk distribution across different urban areas, and that the risk level is strongly correlated with the distribution of population density, the impact of surface shelter, the distribution of ground-level obstacle heights, the designation of the city's no-fly zones, and the UAV's flight height. The application of ground-based risk maps in the operation of UAVs will help to identify the level of danger of the areas flown over, avoid high-risk areas, and enhance operational safety in urban low-altitude.

Keywords: urban low-altitude; ground risk; multiple factors; risk assessment; risk map; operational safety

Received: 2024-04-23; **Accepted**: 2024-07-12; **Published Online**: 2024-08-06 10:20

URL: link.cnki.net/urlid/11.2625.V.20240805.1413.002

Foundation items: Open Research Fund of National Center of ATC Surveillance and Communication System Engineering Research (w222394)

* **Corresponding author**. E-mail: yushasha@mail.xhu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0093

充气式返回舱迎风面真实气体效应

何青松[1, 2, *]，孙素蓉[3]

(1. 北京空间机电研究所，北京 100194；2. 中国航天科技集团有限公司 航天进入、减速与着陆技术实验室，北京 100084；
3. 北京航空航天大学 宇航学院，北京 100191)

摘 要：分别采用空气9组分化学非平衡模型和基于理想气体假设模型对充气式返回舱再入流场进行数值模拟，考察2种方法计算结果的差异，研究真实气体效应的表现形式，探究充气式返回舱真实气体效应与刚性返回舱真实气体效应存在差异的原因。研究结果表明：相比于理想气体假设，真实气体效应使激波位置更靠近壁面，激波后空气温度降低，壁面热流密度下降；在飞行高度为83 km处，激波后气体比热比高于1.4，空气发生解离反应，而在飞行高度为73 km处真实气体效应的作用较弱，激波后空气的比热比维持在1.4，空气仍以分子的形式存在；导致充气式返回舱与刚性返回舱在相同高度范围真实气体强弱差异的主要原因是充气式返回舱的阻重比相较于刚性返回舱更大，进入大气后速度下降更快，在相同高度速度更低。

关 键 词：充气式返回舱；真实气体效应；理想气体；比热比；数值模拟
中图分类号：V445.4
文献标志码：A **文章编号**：1001-5965(2025)03-0816-08

充气式返回舱是一种新型的再入返回装置，由柔性材料充气形成圆锥形阻力面，集合了传统飞行器的热防护系统、降落伞减速系统和着陆缓冲/水上漂浮系统，可以大幅度降低航天器回收系统的成本[1-2]。充气式返回舱主要由柔性材料制成，再入过程中靠柔性材料保持气动外形及提供热防护。因此，在高速气流的冲击下，充气式返回舱极易发生结构变形甚至破坏，再入过程的高温气流也给防热设计带来巨大挑战。

结构强度和防热的设计需要对充气式返回舱再入过程中的流场特性有深入的认识。与刚性再入飞行器类似，充气式返回舱高速再入过程中头部形成强激波，激波后空气温度急剧升高，激波后气体温度超过10 000 K，导致空气发生复杂的解离和电离反应，即真实气体效应。真实气体效应对再入飞行器有十分重要的影响，对阿波罗飞船的研究中发现飞行高度为53 km时，采用理想气体假设得到头部驻点温度约为60 000 K，而实际飞行中头部驻点温度约为10 000 K[3]。日本对HYFLEX高超声速再入飞行实验也证明真实气体效应在Ma>10以后影响飞行器的配平特性，飞行器实际配平角度超出地面预测值范围[4]。Holden等[5]对高速再入飞行器的研究表明真实气体效应会降低机翼后缘曲面上的压力，导致飞行器俯仰力矩和稳定性降低。对于充气式返回舱，目前虽然有较多对充气式返回舱气动特性的研究[6-9]，但在真实气体效应对充气式返回舱气动特性的影响认识还存在不足。

充气式返回舱作为一种新型的再入飞行器，其流场特性与传统的刚性飞行器虽然相似，但其本身特殊的组成及构型势必导致其流场特性有所不同。本文着重研究充气式返回舱的真实气体效应，求解连续流Navier-Stokes方程，不包含稀薄气体效应。分别采用空气9组分化学非平衡模型和理想气体假设模型对充气式返回舱的高速再入流场进

收稿日期：2023-03-01；**录用日期**：2023-06-05；**网络出版时间**：2023-06-30 14:47
网络出版地址：link.cnki.net/urlid/11.2625.V.20230630.1343.001
基金项目：国家自然科学基金 (12005014,12005010)
* **通信作者**. E-mail：heqs2021@sina.com

引用格式：何青松，孙素蓉. 充气式返回舱迎风面真实气体效应 [J]. 北京航空航天大学学报，2025，51（3）：816-823.
HE Q S，SUN S R. Real gas effect of inflatable reentry decelerator on windward side [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（3）：816-823（in Chinese）.

行数值模拟研究, 化学非平衡模型考虑了空气解离、电离及相应的复合反应, 表征了真实气体效应的作用。本文主要考察真实气体效应对充气返回舱流场激波、温度及热流密度等的影响, 为充气式返回舱的设计提供一定理论参考。

1　计算模型

本文采用日本研制的充气式返回舱 Titans 为研究对象, Titans 主要由充气环、蒙皮和舱体组成, 舱体半径为 0.4 m, 充气环半径为 0.075 m, 充气环围成的环形直径为 2.5 m, 蒙皮的半锥角为 70°, 其具体结构尺寸可参考文献 [10]。由于 Titans 为轴对称构型, 且本文的计算迎角均为 0°, 因此, 计算域可简化为二维轴对称, 本文采用的计算域如图 1 所示。

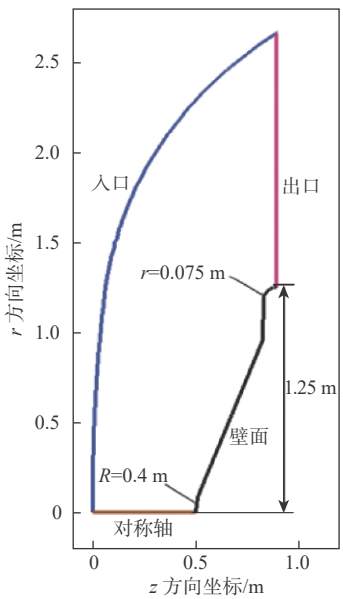


图 1　计算域

Fig. 1　Computational domain

本文采用的双温度化学非平衡模型将电子和重粒子看作 2 种不同的流体, 流场内部气体的流动为轴对称、稳态、层流流动。计算中采用二维轴对称可压缩流动控制方程:

$$\frac{\partial \boldsymbol{U}}{\partial t}+\frac{\partial \boldsymbol{E}}{\partial z}+\frac{\partial \boldsymbol{F}}{\partial r}+\boldsymbol{H}=\frac{\partial \boldsymbol{E}_{\mathrm{v}}}{\partial z}+\frac{\partial \boldsymbol{F}_{\mathrm{v}}}{\partial r}+\boldsymbol{H}_{\mathrm{v}}+\boldsymbol{S}_{\mathrm{c}} \tag{1}$$

式中: $\boldsymbol{U}$ 为守恒变量; $\boldsymbol{E}$、$\boldsymbol{F}$ 分别为无黏项在 z、r 方向上的分量; $\boldsymbol{E}_{\mathrm{v}}$、$\boldsymbol{F}_{\mathrm{v}}$ 分别为黏性项在 z、r 方向上的分量; $\boldsymbol{H}$ 和 $\boldsymbol{H}_{\mathrm{v}}$ 分别表示无黏性和黏性项在柱坐标系下的源项矢量; $\boldsymbol{S}_{\mathrm{c}}$ 为化学反应源项。

各矢量的表达式为

$$\boldsymbol{U}=\begin{bmatrix}\rho\\ \rho u\\ \rho v\\ \rho e_{\mathrm{t}}\\ e_{\mathrm{e}}e_{\mathrm{e}}\\ \rho f_i\end{bmatrix};\quad \boldsymbol{E}=\begin{bmatrix}\rho u\\ \rho uu+p\\ \rho uv\\ (\rho e_{\mathrm{t}}+p)u\\ (e_{\mathrm{e}}e_{\mathrm{e}}+p_{\mathrm{e}})u\\ \rho f_i u\end{bmatrix}$$

$$\boldsymbol{F}=\begin{bmatrix}\rho v\\ \rho uv\\ \rho vv+p\\ (\rho e_{\mathrm{t}}+p)v\\ (e_{\mathrm{e}}e_{\mathrm{e}}+p_{\mathrm{e}})v\\ \rho f_i v\end{bmatrix};\quad \boldsymbol{H}=\frac{1}{r}\begin{bmatrix}\rho v\\ \rho uv\\ \rho vv\\ (\rho e_{\mathrm{t}}+p)v\\ (e_{\mathrm{e}}e_{\mathrm{e}}+p_{\mathrm{e}})v\\ \rho f_i v\end{bmatrix}$$

$$\boldsymbol{E}_{\mathrm{v}}=\begin{bmatrix}0\\ \tau_{zz}\\ \tau_{zr}\\ u\tau_{zz}+v\tau_{zr}-q_z\\ -q_{ez}\\ -J_{iz}\end{bmatrix};\quad \boldsymbol{F}_{\mathrm{v}}=\begin{bmatrix}0\\ \tau_{rz}\\ \tau_{rr}\\ u\tau_{rz}+v\tau_{rr}-q_r\\ -q_{er}\\ -J_{ir}\end{bmatrix}$$

$$\boldsymbol{H}_{\mathrm{v}}=\frac{1}{r}\begin{bmatrix}0\\ \tau_{rz}\\ \tau_{rr}-\tau_{\theta\theta}\\ u\tau_{rz}+v\tau_{rr}-q_r\\ -q_{er}\\ -J_{ir}\end{bmatrix};\quad \boldsymbol{S}_{\mathrm{c}}=\begin{bmatrix}0\\ 0\\ 0\\ -E_{\mathrm{c}}\\ -E_{\mathrm{ce}}-Q_{\mathrm{ech}}\\ \dot{n}_i m_i\end{bmatrix}$$

式中: u、v 分别为速度在 z、r 方向的分量; E_{c} 为化学能损失; τ 为切应力; 下标 i 表示各个重粒子; 下标 e 和 h 分别表示电子和重粒子; 下标 t 表示电子和重粒子的总和; 下标 z 和 r 分别表示 z 和 r 的方向分量; p、e、ρ、T、h 和 f 分别为气体的压强、内能、密度、温度、比焓和质量分数; J_{iz} 和 J_{ir} 分别为组分 i 的轴向扩散通量分量和径向扩散通量分量。Q_{ech} 为电子与重粒子的能量交换, 表达式为

$$Q_{\mathrm{ech}}=3e_{\mathrm{e}}\sum_{i=1}^{n}\frac{v_{ei}}{m_i}k_{\mathrm{B}}(T_{\mathrm{e}}-T_{\mathrm{h}}) \tag{2}$$

式中: v_{ei} 为碰撞频率, 计算方法参考文献 [11]; k_{B} 为玻尔兹曼常数。

空气混合气体的黏性和热导率通过分别计算单个组分的黏性和热导率, 然后, 采用 Wilke[12] 半经验公式获得混合气体的黏性和热导率。

扩散通量由自洽有效扩散系数法获得[13-16], 计算式为

$$J_i=-\left(\frac{pm_iD_i}{RT_i}\right)G_i+y_i\sum_j\left(\frac{pm_jD_j}{RT_j}\right)G_j \tag{3}$$

式中: D_i 和 G_i 分别为组分 i 的有效扩散系数和扩散驱动力, 扩散驱动力由流场中的温度、压强和浓度梯度构成, 具体表达式参考文献 [16]。

本文考虑的空气组分包括氮分子(N_2)、氧分子(O_2)、氮原子(N)、氧原子(O)、氮氧分子(NO)、氮离子(N^+)、氧离子(O^+)、氮氧离子(NO^+)和电子(e)。

基于理想气体假设的仿真模型求解连续方程、动量方程及能量方程, 计算中涉及的热导率和黏性系数采用 Sutherland 公式, 该方法较为成熟, 应用也十分广泛, 本文不作细致介绍。

计算中, 在流场入口给定来流速度、密度和温

度，在出口所有参数外推，在壁面采用无滑移条件，壁面温度为300 K，在对称轴采用轴对称边界条件。

本文分别采用化学非平衡模型和理想气体假设模型，分别对飞行高度为73 km和83 km的流场特性进行研究，弹道参数可参考文献[10]。根据弹道参数可知，Titans的速度在飞行高度为73～83 km变化较为剧烈，由于真实气体效应与来流速度紧密相关，因此，选择73 km和83 km进行研究，便于比较真实气体效应影响的差异，计算工况的来流参数如表1所示。

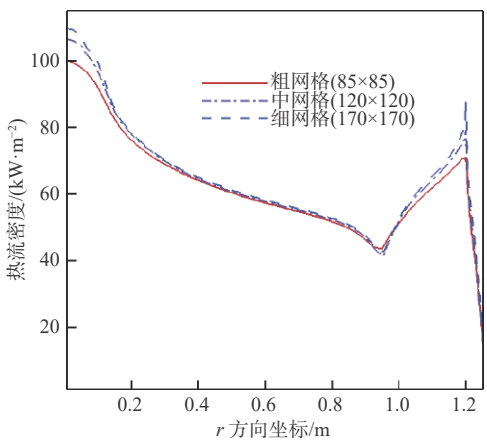
表1 计算工况的来流参数

Table 1 Inflow parameters of operating conditions

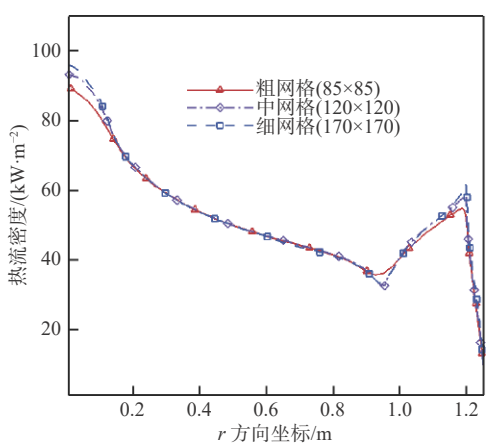
飞行高度/km	速度/(m·s^{-1})	密度/(kg·m^{-3})	温度/K
73	2 396	5.125×10^{-5}	211.5
83	5 081	1.105×10^{-5}	191.2

2 计算结果

本文分别采用粗(85×85)、中(120×120)、细(170×170)这3套网格对2种计算模型进行仿真计算，比较了飞行高度为83 km沿壁面的热流密度，如图2所示。由图可知，细网格和中网格在驻点和圆环处的峰值更加接近，粗网格相差较远。为保证计算准确性，同时也不大幅增加计算量，本文采用120×120的网格进行计算研究。



(a) 理想气体



(b) 化学非平衡

图2 网格无关性验证

Fig. 2 Grid independence test

为验证本文所采用的化学非平衡模型及计算程序的可靠性，分别采用文献[17]的半球形模型和文献[18]的圆柱体模型进行比较。半球形模型直径为76 cm，来流速度为5 732 m/s，温度为535 K，密度为0.001 6 kg/m^3，壁面热流密度的比较结果如图3所示。由图可知，本文计算结果与实验结果变化趋势一致，总体符合较好。文献[18]的圆柱体模型直径为90 mm，计算工况为来流速度为4 776 m/s，温度为694 K，压强为687 Pa。图4为壁面压强分布和热流密度的比较结果，由图可知，本文的仿真结果与试验结果符合较好。

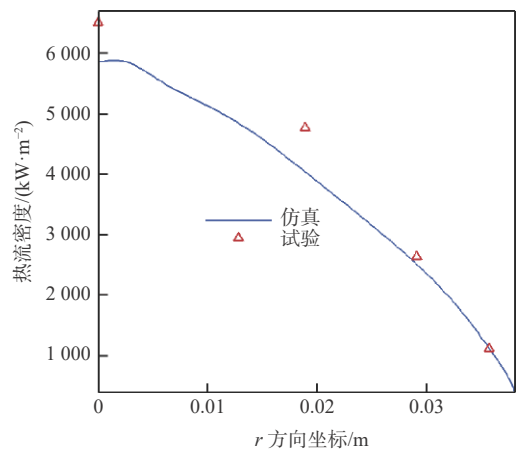
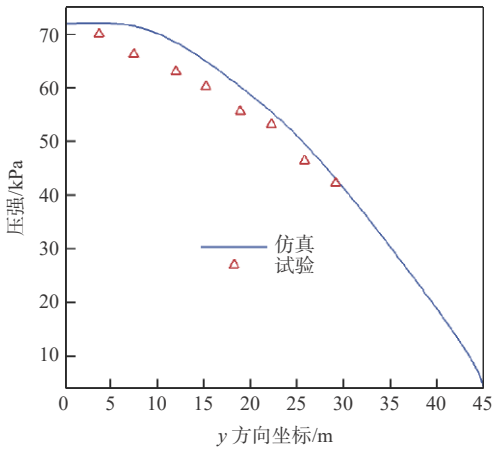


图3 半球模型壁面热流密度计算结果与实验结果比较

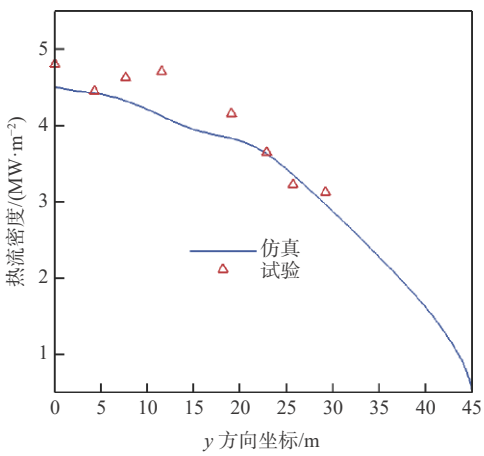
Fig. 3 Comparison between calculation and experimental results of heat flux densities on wall of hemisphere model

本文分别采用化学非平衡模型和理想气体假设模型对表1列出的工况进行研究，图5为流场中马赫数的分布情况。对比2种模型的计算结果发现，2种计算模型马赫数的分布规律一致，激波前马赫数最高，气体经过激波后急剧下降。在飞行高度为73 km处，化学非平衡模型和理想气体假设模型计算的激波位置与基本一致，而在飞行高度为83 km处，化学非平衡模型计算得到的激波位置更靠近物面。

图6为沿轴线温度分布，其中，化学非平衡模型给出的是重粒子温度，从图中看出，激波后气体温度急剧上升，采用理想气体假设获得的气体温度在激波后保持平稳，在壁面附近逐渐下降至300 K，采用化学非平衡模型获得的气体温度急剧上升后逐渐下降至壁面的300 K。值得注意的是，在飞行高度为73 km处，两者计算得到的沿轴线的气体温度基本一致。



(a) 壁面压强计算结果与实验结果比较



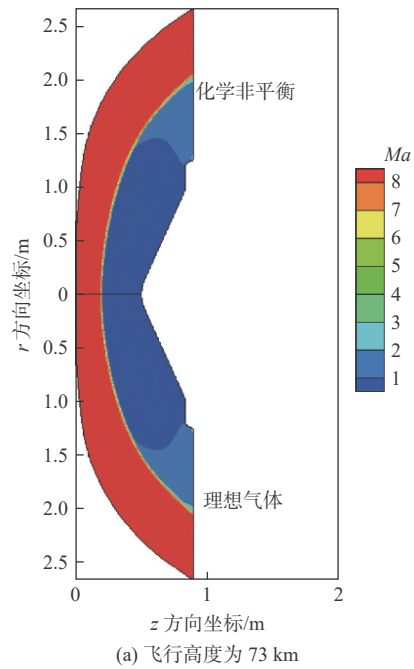
(b) 壁面热流密度计算结果与实验结果比较

图 4　圆柱体模型计算结果与实验结果比较

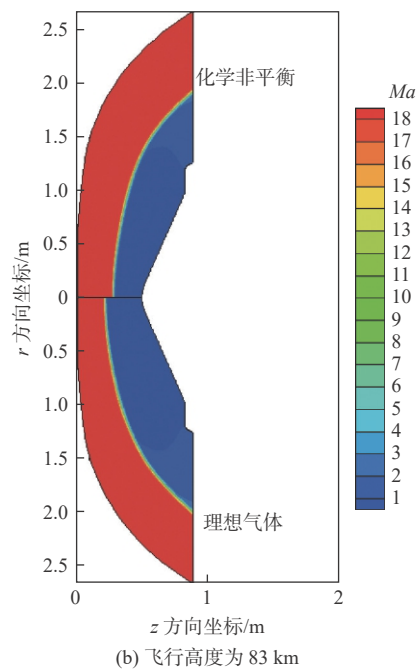
Fig. 4　Comparison between calculation and experimental results of cylinder model

导致图 6 温度差异的主要原因是真实气体效应的存在, 激波后的空气发生化学反应, 吸收了气体的热量, 导致气体温度下降[19]。在数值模拟中, 化学非平衡模型的能量方程不仅考虑了化学反应引起的能量变化, 还考虑了气体的导热及组分扩散等引起的能量变化等。图 7 为沿轴线气体的热导率, 由于高温空气导致气体解离电离的发生, 气体成分变得复杂, 气体的热导率也与理想气体不一致。其中, 最大热导率的差距在飞行高度为 73 km 处, 化学非平衡模型比理想气体假设高 29%, 而飞行高度为 83 km 处, 化学非平衡模型是理想气体假设的 2.31 倍。同样, 由于气体成分的变化, 气体组分扩散导致的能量变化也有较大区别, 这些因素是否也是导致温度差异的原因值得进一步研究。

图 8 为沿壁面热流密度分布情况, 由图可知, 飞行高度为 83 km 处壁面热流密度显著高于 73 km 处热流密度。另外, 在飞行高度为 83 km 处, 基于理想气体假设获得的热流密度显著高于化学非平衡模型的热流密度, 在驻点处两者计算得到的热流



(a) 飞行高度为 73 km



(b) 飞行高度为 83 km

图 5　流场中马赫数分布

Fig. 5　Mach number distribution in flow field

密度相差 14.3%。而在飞行高度为 73 km 处, 两者计算得到的热流密度十分接近, 在驻点处两者计算得到的热流密度相差仅为 4%。总体来讲, 基于理想气体假设高估了壁面热流密度。

由图 5、图 6 和图 8 可知, 真实气体效应对于充气式返回舱的影响规律与刚性返回舱并无差异。但同时也应该注意到在飞行高度为 73 km 处, 真实气体效应对充气式返回舱的影响较小。为进一步深入探究飞行高度为 73 km 和 83 km 真实气体效应的特性, 本文将讨论在这 2 个高度充气式返回舱真实气体效应的表现形式。

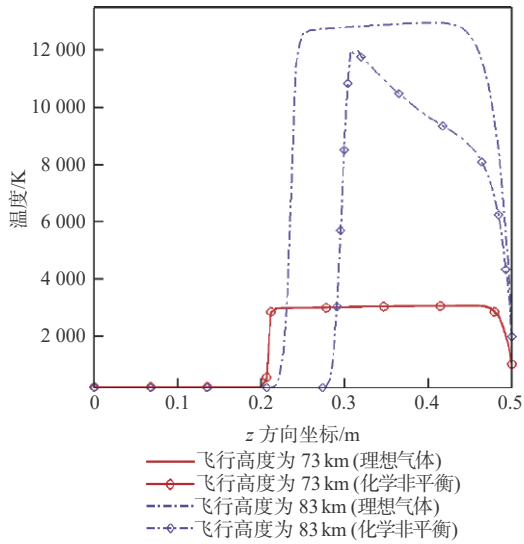


图 6 沿轴线温度分布

Fig. 6 Temperature distribution along axis

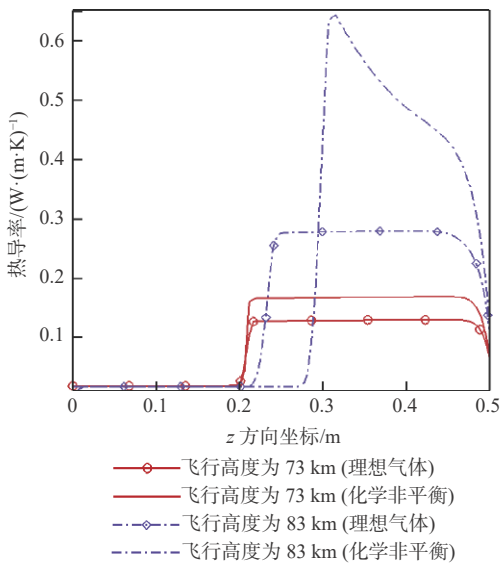


图 7 沿轴线热导率分布

Fig. 7 Thermal conductivity distribution along axis

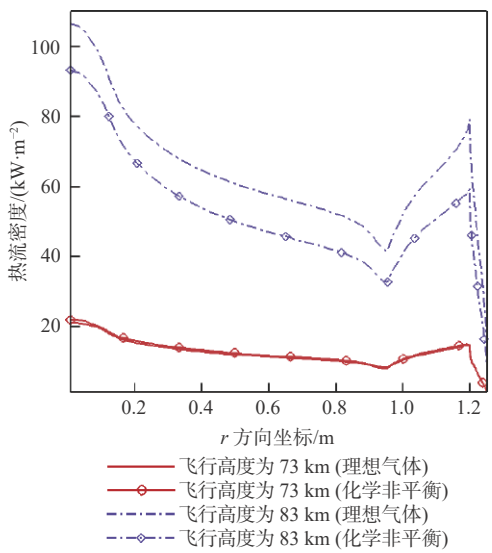


图 8 沿壁面热流密度分布

Fig. 8 Heat flux density distribution along wall

真实气体效应与来流马赫数、气体密度及焓值等属性相关，其表现形式主要有 2 个方面：①气体是非理想气体激波层内比热比的变化；②激波层内化学反应的发生导致气体组分的变化。第 1 种现象在马赫数相对较低的高超声速流动中也会发生（如马赫数为 6），主要是由于激波层内温度的升高导致气体比热容的变化。然而，这 2 个现象是连续的，即第 2 个效应跟随第 1 个效应[20]。

图 9 为采用化学非平衡模型时沿轴线的比热比，从图中看出激波前温度较低，空气比热比为 1.4。激波后随着气体温度升高，飞行高度为 83 km 的比热比逐渐升高，在靠近壁面附近逐渐下降，而飞行高度为 73 km 空气的比热比仍然维持在 1.4，说明在飞行高度为 73 km 处，周围气体十分接近理想气体。

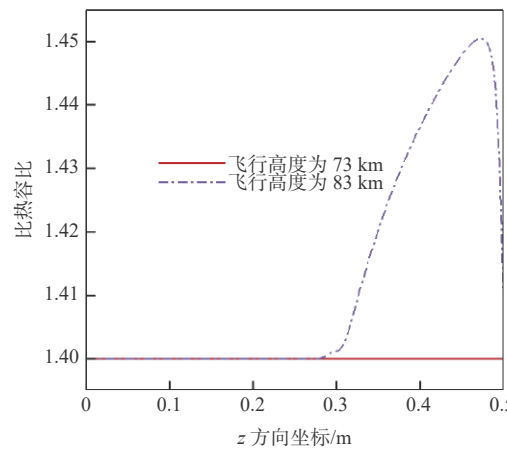
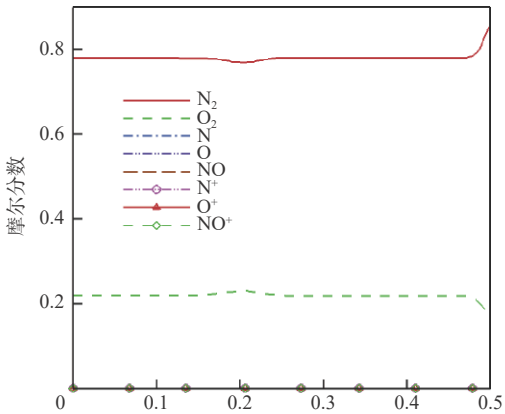


图 9 沿轴线比热比

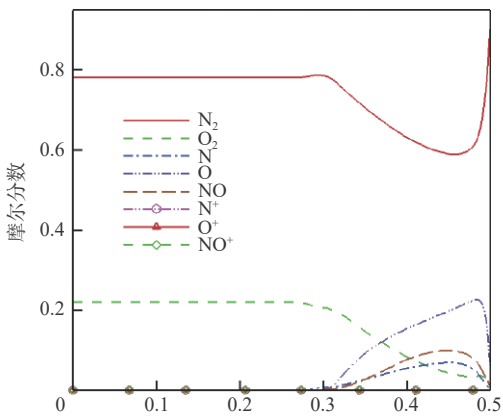
Fig. 9 Specific heat ratio along axis

图 10 为沿轴线各组分的摩尔分数，由图可知，2 个工况气体的电离作用都较弱，空气主要以中性气体的形式存在。在飞行高度为 73 km 处，沿轴线的成分主要是 N_2 和 O_2，说明气体的解离和电离反应都非常弱。在飞行高度为 83 km 处，激波后气体温度逐渐升高，N_2 和 O_2 开始发生解离反应，由于 O_2 的解离能更小，因此，O_2 解离更彻底，沿轴线 N_2 和 O_2 的最小摩尔分数分别为 0.6 和 0.03，O 的摩尔分数显著高于 N 的摩尔分数。由于 N 与 O 结合生成 NO，甚至 NO 的摩尔分数都高于 N 的摩尔分数。

总体来看，真实气体效应对充气式返回舱的流场有较大的影响，在飞行高度为 83 km 处真实气体效应的影响十分显著。但在飞行高度为 73 km 处，真实气体效应对流场基本无影响，在这个高度气体的化学反应已经非常弱，流场内的空气十分接近理想气体。这一点有别于传统的刚性返回舱，阿波罗飞船在飞行高度为 53 km 处真实气体效应仍然影



(a) 飞行高度为 73 km



(b) 飞行高度为 83 km

图 10　沿轴线各组分摩尔分数

Fig. 10　Mole fraction of each component along axis

响非常大[3]，航天飞机在飞行高度为 50～80 km 处真实气体效应对流场均有较大影响[21]，而充气式返回舱在飞行高度为 73 km 处真实气体效应的影响已经非常弱。文献 [22] 的研究表明随着飞行高度增加，气体变得更加稀薄，参与化学反应的粒子减少，返回舱壁面热流载荷受真实气体效应的影响减弱。由此可以推断，相对于刚性返回舱，真实气体效应对充气式返回舱的受热环境的影响整体上更小一些。因为真实气体效应对充气式返回舱的影响主要在较高的飞行高度，而在较高的飞行高度真实气体效应对返回舱热流载荷的影响相对较低高度更弱。

导致充气式返回舱与刚性返回舱存在这种差异的主要原因可能是充气式返回舱与刚性返回舱的飞行轨迹不同。图 11 为 Titans、双子座飞船和航天飞机飞行过程中飞行速度随飞行高度的变化[10,21,23]，由图可知，Titans 速度下降较快的区域在飞行高度为 60～95 km 的范围，而双子座飞船和航天飞机速度下降较快的区域在飞行高度为 20～70 km 的范围。在相同高度上，Titans 的飞行速度远小于双子

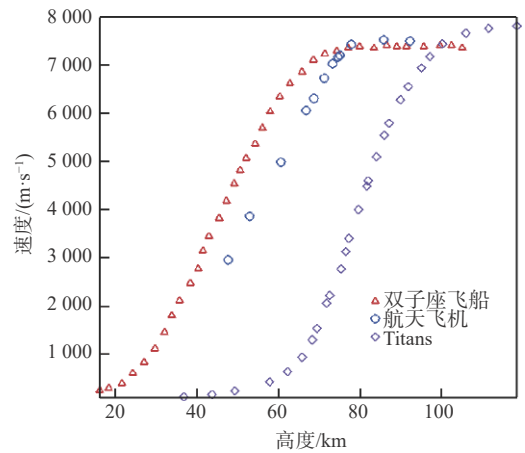


图 11　速度随高度的变化(双子座飞船、航天飞机、Titans)

Fig. 11　Variation of speed with altitude (Gemini, Space Shuttle, and Titans)

座飞船和航天飞机，激波对气体的压缩减弱，激波后气体温度更低，进而导致在相同高度充气式返回舱的真实气体效应更弱。而产生这种速度差异的原因是充气式返回舱主要由柔性材料制成，其阻重比相对于刚性返回舱更大，进入大气后速度下降更快，在相同高度速度更低。

3　结　论

本文分别采用化学非平衡模型和基于理想气体假设模型对充气式返回舱进行了数值模拟研究，比较了两者计算得到的充气式返回舱流场中激波位置、沿轴线气体温度和壁面热流等的差异。可得结论如下：

1）相比于理想气体假设模型的计算结果，化学非平衡模型计算得到的激波位置更靠近壁面；

2）激波后，基于理想气体假设得到的气体温度保持平稳，飞行高度为 83 km 时，化学非平衡模型计算得到的温度达到峰值后逐渐下降，在飞行高度为 73 km 处与理想气体假设模型得到的气体温度接近；

3）飞行高度越高，理想气体假设和真实气体效应得到的热流密度差异越大，理想气体假设高估了壁面热流密度；

4）激波后随着气体温度升高，在飞行高度为 83 km 处沿轴线的比热比逐渐升高，而在飞行高度为 73 km 处空气的比热比始终维持在 1.4；

5）流场中的电离作用较弱，空气主要以中性气体的形式存在，而在飞行高度为 73 km 处，解离和电离的作用都非常弱，空气主要以分子的形式存在；

6）充气式返回舱在飞行高度为 73 km 处真实气体效应的作用已经较弱，与刚性返回舱存在较大差异，导致这种差异的原因是充气式返回舱的阻重

比相比于刚性返回舱更大，进入大气后速度下降更快，在相同高度速度更低。

参考文献（References）

[1] FURUYA H, MURATA S, NAKAHARA M, et al. Concept of inflatable tensegrity for large space structures[C]// Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston: AIAA, 2006.

[2] 卫剑征, 谭惠丰, 王伟志, 等. 充气式再入减速器研究最新进展[J]. 宇航学报, 2013, 34(7): 881-890.
WEI J Z, TAN H F, WANG W Z, et al. New trends in inflatable reentry aeroshell[J]. Journal of Astronautics, 2013, 34(7): 881-890(in Chinese).

[3] HILLJE E R, SAVAGE R. Status of the aerodynamic characteristics of the Apollo entry configuration[C]// Proceedings of the Entry Vehicle Systems and Technology Meeting. Reston: AIAA, 1968.

[4] WATANABE S, ISHIMOTO S, YAMAMOTO Y. Aerodynamic characteristics evaluation of hypersonic flight experiment vehicle based on flight data[J]. Journal of Spacecraft and Rockets, 1997, 34(4): 464-470.

[5] HOLDEN M, WADHAMS T, MACLEAN M, et al. Experimental studies in LENS I and X to evaluate real gas effects on hypervelocity vehicle performance[C]//Proceedings of the 45th AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2007.

[6] GUO J H, LIN G P, BU X Q, et al. Effect of static shape deformation on aerodynamics and aerothermodynamics of hypersonic inflatable aerodynamic decelerator[J]. Acta Astronautica, 2017, 136: 421-433.

[7] TAKAHASHI Y, KOIKE T, OSHIMA N, et al. Aerothermodynamic analysis for deformed membrane of inflatable aeroshell in orbital reentry mission[J]. Aerospace Science and Technology, 2019, 92: 858-868.

[8] TAKAHASHI Y, HA D, OSHIMA N, et al. Aerodecelerator performance of flare-type membrane inflatable vehicle in suborbital reentry[J]. Journal of Spacecraft and Rockets, 2017, 54(5): 993-1004.

[9] MAZAHERI A. High-energy atmospheric reentry test aerothermodynamic analysis[J]. Journal of Spacecraft and Rockets, 2013, 50(2): 270-281.

[10] TAKAHASHI Y, YAMADA K. Aerodynamic heating of inflatable aeroshell in orbital reentry[J]. Acta Astronautica, 2018, 152: 437-448.

[11] MITCHNER M, KRUGER C. Partially ionized gases[M]. New York: Wiley, 1973.

[12] WILKE C R. A viscosity equation for gas mixtures[J]. Journal of Chemical Physics, 1950, 18(4): 517-519.

[13] RAMSHAW J D. Self-consistent effective binary diffusion in multicomponent gas mixtures[J]. Journal of Non-Equilibrium Thermodynamics, 1990, 15(3): 295-300.

[14] RAMSHAW J D, CHANG C H. Ambipolar diffusion in two-temperature multicomponent plasmas[J]. Plasma Chemistry and Plasma Processing, 1993, 13(3): 489-498.

[15] RAMSHAW J D. Hydrodynamic theory of multicomponent diffusion and thermal diffusion in multitemperature gas mixtures[J]. Journal of Non Equilibrium Thermodynamics, 1993, 18(2): 121-134.

[16] WANG H X, HE Q S, MURPHY A B, et al. Numerical simulation of nonequilibrium species diffusion in a low-power nitrogen-hydrogen arcjet thruster[J]. Plasma Chemistry and Plasma Processing, 2017, 37(3): 877-895.

[17] MACLEAN M, MARINEAU E, PARKER R, et al. Effect of surface catalysis on measured heat transfer in expansion tunnel facility[J]. Journal of Spacecraft and Rockets, 2013, 50(2): 470-475.

[18] HANNEMANN K, SCHRAMM J M, KARL S, et al. Cylinder shock layer density profiles measured in high enthalpy flows in HEG[C]// Proceedings of the 22nd AIAA Aerodynamic Measurement Technology and Ground Testing Conference . Reston: AIAA, 2002.

[19] 杨建龙, 刘猛, 阿嵘. 高超声速热化学非平衡对气动热环境影响[J]. 北京航空航天大学学报, 2017, 43(10): 2063-2072.
YANG J L, LIU M, A R. Influence of hypersonic thermo-chemical non-equilibrium on aerodynamic thermal environments[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(10): 2063-2072(in Chinese).

[20] PRABHU D, PAPADOPOULOS P, DAVIES C, et al. Shuttle orbiter contingency abort aerodynamics, II: real-gas effects and high angles of attack[C]// Proceedings of the Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2003.

[21] SHINN J, MOSS J, SIMMONDS A. Viscous-shock-layer heating analysis for the shuttle windward-symmetry plane with surface finite catalytic recombination rates[C]// Proceedings of the 3rd Joint Thermophysics, Fluids, Plasma and Heat Transfer Conference. Reston: AIAA, 1982.

[22] PEZZELLA G, VOTTA R. Finite rate of chemistry effects on the high altitude aerodynamics of an Apollo-shaped reentry capsule [C]// Proceedings of the 16th AIAA/DLR/DGLR International Space Planes and Hypersonic Systems and Technologies Conference. Reston: AIAA, 2009.

[23] YOUNG K A, OSGOOD C T. Preflight orbital and reentry trajectory data for Gemini VI[R]. Washington, D. C. : NASA, 1965: 1-114.

Real gas effect of inflatable reentry decelerator on windward side

HE Qingsong[1, 2, *], SUN Surong[3]

(1. Beijing Institute of Space Mechanics & Electricity, Beijing 100194, China;
2. Laboratory of Aerospace Entry, Descent and Landing Technology, China Aerospace Science and Technology Corporation, Beijing 100084, China;
3. School of Astronautics, Beihang University, Beijing 100191, China)

Abstract: In this paper, a 9-component chemical non-equilibrium model and a perfect gas model were used to study the reentry flow of an inflatable reentry decelerator through numerical simulation. The differences in the calculation results of the two models were investigated. The manifestation of the real gas effect was studied, and the reason for the difference between the real gas effect of inflatable reentry decelerator and that of rigid capsule was explored. The results show that compared to the perfect gas hypothesis, the shock wave position is closer to the wall in the real gas effect. After the shock wave, the air temperature decreases, and the wall heat flux decreases. At 83 km, the specific heat ratio of the gas after the shock wave is higher than 1.4, and the air undergoes a dissociation reaction. At 73 km, the real gas effect is very weak, and the specific heat ratio of the gas after the shock wave remains at 1.4. The air still exists in the form of molecules. The main reason for the difference in real gas strength between inflatable reentry decelerator and rigid capsule in the same altitude range is that the inflatable reentry decelerator has a larger resistance-weight ratio than the rigid capsule. Its speed drops faster after entering the atmosphere, and its speed is lower at the same altitude.

Keywords: inflatable reentry decelerator; real gas effect; perfect gas; specific heat ratio; numerical simulation

Received: 2023-03-01; **Accepted**: 2023-06-05; **Published Online**: 2023-06-30 14:47
URL: link.cnki.net/urlid/11.2625.V.20230630.1343.001
Foundation items: National Natural Science Foundation of China (12005014,12005010)
* **Corresponding author**. E-mail: heqs2021@sina.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0109

固化应力松弛对复材模具型面精度的影响

肖遥[1]，李勇[1]，李东升[1,*]，王磊[2]，江超[3]

(1. 北京航空航天大学 机械工程及自动化学院，北京 100191；2. 湖南航天环宇通信股份有限公司，长沙 410006；
3. 沈阳飞机工业（集团）有限公司，沈阳 110850)

摘 要：具有低温固化高温使用特点的复合材料工装预浸料，是制造航空航天高精度复合材料成型工装极具潜力的材料之一。为实现其低成本、高精度成型，模具用预浸料一般采用更为复杂的低温预固化和高温后固化组成的2步工艺。模具成型固化应力不仅影响模具初始型面精度，而且在模具热循环使用过程中还可能产生应力松弛，进一步引起模具型面精度变化，进而影响零件成型精度。本文针对碳纤维增强复合材料模具凹凸复杂型面特征，通过试验研究了固化成型阶段和热循环使用过程中模具型面精度变化规律及影响因素。试验结果表明：成型阶段的固化应力及其在使用过程中的应力松弛行为是造成模具型面精度变化的主要因素，且型面偏差随热循环使用周期的增加而逐渐趋于稳定；成型阶段，模具型面偏差主要来源于由热膨胀不匹配、树脂固化收缩及金属母模作用诱导的固化应力，而使用阶段，固化残余应力的松弛行为是引起模具型面偏差的主要原因。研究结果能够帮助理解复合材料模具结构稳定性变化，同时，对于降低复合材料模具制造成本、提高使用寿命，实现复杂高精度复合材料构件的精确制造具有重要意义。

关 键 词：复合材料模具；固化变形；残余应力；应力松弛；模具型面精度
中图分类号：V262.3; V258$^+$.3
文献标志码：A **文章编号**：1001-5965(2025)03-0824-09

由于材料特性和制造工艺影响，热固性树脂基复合材料在固化过程中不可避免地会产生应力，从而造成构件固化后的回弹或翘曲变形，导致界面开裂或分层失效，影响航空航天用先进复合材料结构的机械性能和使用寿命[1-2]。复合材料的固化变形主要来源于树脂-纤维热膨胀差异、树脂固化收缩及模具-零件相互作用，其中，模具-零件界面热膨胀不匹配产生的剪切应力是诱导复合材料产生固化变形的关键因素[3-4]。为克服铝、钢、铸铁等金属模具高热膨胀系数(coefficient of thermal expansion, CTE)造成的复合材料构件成型精度低的突出问题，大尺寸、高精度复合材料构件成型主要采用低CTE的殷钢模具。然而，殷钢密度大、热容高、加工性能差，应用范围有限[5]。由于复合材料模具具有轻质、高强、低热容的特点及CTE可设计的优点，能够显著降低模具-零件CTE不匹配引起的固化变形，是当前复合材料成型模具的重要发展方向[6-8]。

飞机复杂变截面类回转体复合材料进气道，其结构形式和型面特征复杂、尺寸大、精度要求高、成型难度大。采用复合材料模具是克服由殷钢模具重量大带来的自动铺丝过程中模具回转挠曲变形严重及脱模困难问题，保证成型精度和效率的有效解决途径。复合材料模具型面精度及其稳定性控制是成型复杂高精度复合材料构件的关键，也是设计、优化模具制造工艺参数的难点。复合材料成型过程中的固化应力不仅会诱导模具产生变形，造

收稿日期：2023-03-07；**录用日期**：2023-07-07；**网络出版时间**：2023-09-22 08:17
网络出版地址：link.cnki.net/urlid/11.2625.V.20230920.2014.002
基金项目：中央高校基本科研业务费专项资金 (YWF-23-L-1012)
* **通信作者**. E-mail：lidongs@buaa.edu.cn

引用格式：肖遥，李勇，李东升，等. 固化应力松弛对复材模具型面精度的影响[J]. 北京航空航天大学学报，2025，51(3)：824-832.
XIAO Y，LI Y，LI D S，et al. Influence of curing stress relaxation on profile accuracy of composites tools[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：824-832 (in Chinese).

成型面偏差, 而且在模具热循环使用过程中还会产生应力松弛, 引起模具型面变化, 影响型面精度的稳定性。复合材料的固化应力和变形受结构特征及工艺参数影响, 尤其是对具有复杂凹凸型面特征的结构, 其影响更为显著[9-10]。

针对复合材料结构的成型精度影响因素及规律, 众多学者进行了试验和仿真研究。Kappel[11] 针对复合材料平板件和曲率件的试验研究发现, 带角度试样的回弹变形比平板试样的翘曲变形高 80%。贾丽杰等[12] 通过仿真分析发现准各向同性无曲率层合板(平板试样)的残余应力呈对称分布; 而曲率层合板的残余应力呈非对称分布, 且翘曲变形随着曲率半径的减小而增大。Fiorina 等[13] 针对 A350 翼肋结构(凸凹翻边)的仿真分析结果显示, 0.5 m 长的翼肋的固化变形超过 1 mm。Dong[14] 针对复合材料加筋壁板进行的参数化分析表明, 筋条结构参数对复合材料壁板的变形有显著影响, 因为不同形式的复合材料结构性能, 如导热系数、CTE 及刚度等均存在差异, 并会进一步耦合工艺参数的影响。

不同于复合材料构件的单步固化成型, 为提高复合材料模具的成型精度, 采用工装织物预浸料铺贴并经由特殊的低温预固化与高温后固化组成的 2 步固化工艺成型。预固化阶段, 受金属母模和辅助材料(真空袋、透气毡等)约束的工装预浸料在低温条件下进行长时间固化以成型结构。此阶段较低的固化温度(323.15～353.15 K)可降低母模材料对其固化变形的影响[15-16]。后固化阶段, 将脱模后的已成形试验件以自由变形状态在高温下进行短时间固化, 使试验件达到所需的固化度。相较于复合材料构件级单步固化工艺, 模具特殊的 2 步固化工艺有助于降低复合材料固化残余应力和变形。然而, 已有研究表明: 复合材料在固化中和固化后均存在应力松弛行为, 该行为诱导的复合材料后续变形高达 30%[17-19], 但有关其作用机制和影响因素的研究却尚未见报道。随着航空航天领域对大型、高精度复合材料模具需求和要求的不断增加与提高, 亟需开展相关研究。

本文针对复杂 S 形变截面类回转体复合材料模具局部典型凹凸特征结构, 通过试验研究了型面精度在 2 步固化成型阶段及热循环使用过程中的影响因素与变化规律, 并进一步分析了由固化残余应力引起的应力松弛行为对模具型面稳定性的影响。

1　试验设计

1.1　试验材料

试验件采用 EasyComposites 公司生产的具有良好铺覆性、尺寸稳定性及抗冲击性能的复合材料斜纹织物工装预浸料, 牌号分别为 XT-200 和 XT-650, 预浸料的名义厚度分别为 0.232 mm 和 0.646 mm, 纤维体积分数分别为 48.49% 和 55.64%, 主要力学性能如表 1 所示。

表 1　复合材料工装预浸料基本力学性能

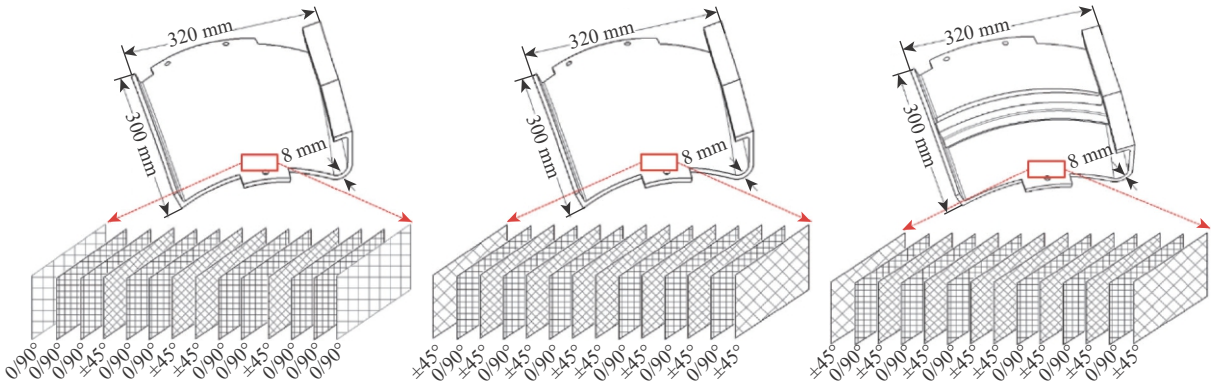
Table 1　Basic mechanical properties of tooling composite material prepreg

牌号	轴向弹性模量/GPa	面内剪切模量/GPa	面内泊松比
XT-200	56	3.66	0.04
XT-650	59.5	4.35	0.08

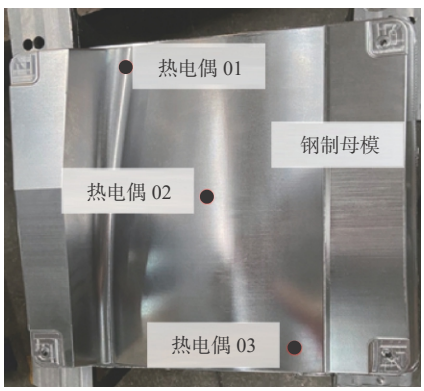
1.2　试验件制备与测试

图 1 为试验件制备与测量过程。试验件取自复杂 S 形变截面类回转体复合材料模具局部典型凹凸特征结构, 分别编号为 00 号、01 号、02 号, 其中, 00 号和 01 号试验件为蒙皮壁板结构, 壁板长为 300 mm, 宽为 320 mm, 厚为 8 mm, 02 号试验件为加筋壁板结构, 帽型加强筋长、宽、厚分别为 20 mm、40 mm、4 mm, 如图 1(a)所示。3 个试验件的蒙皮壁板分别按照 $[(0/90)_s/(0/90)/(0/90)/(\pm45)/(0/90)/(0/90)/(\pm45)]_s$(00 号)和 $[(\pm45)_s/(0/90)/(\pm45)/(0/90)/(\pm45)/(0/90)/(\pm45)]_s$(01 号和 02 号)顺序进行铺层, 02 号试验件的帽型加强筋的铺层顺序为 $[(0/90)_s/(0/90)/(\pm45)/(0/90)]_s$, 加强筋预先固化后和蒙皮在预固化阶段采用共胶接成型。对于所有的铺层结构, 第一层和最后一层均采用高树脂含量的 XT-200 作为表面层, 以提高表面光洁度, 中间采用高纤维含量的 XT-650 作为增强结构刚度的体积层。

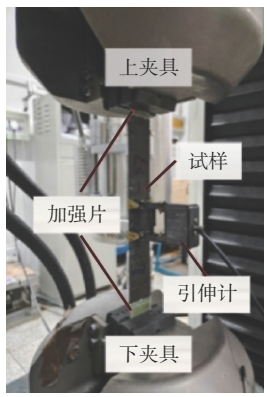
采用 Q235 钢作为所有试验件的母模材料。首先, 试验件铺贴前在母模成型面上涂刷 2 遍脱模剂, 每次间隔 20 min; 然后, 按上述铺层顺序依次进行预浸料铺贴。其间, 每铺贴 3 层进行一次压力为 0.1 MPa, 持续时间为 30 min 的真空预压实。采用真空袋-热压罐工艺和供应商推荐固化曲线(manufacture recommend curve of cure, MRCC)对试验件进行固化。MRCC 包括预固化和后固化 2 个阶段, 预固化时, 试验件受母模和辅助材料的约束, 固化压力为 0.5 MPa, 固化温度以 0.8 K/min 速率从室温升至 343.15 K, 保温 12 h 后以相同的速率降温至 312.15 K 后出罐并脱模。后固化阶段, 试验件静置在母模上, 处于无约束的自由变形状态, 固化压力为 0 MPa, 固化温度以 0.3 K/min 速率从室温升至 473.15 K, 保温 2 h 后, 再以相同速率降温至 323.15 K 后出罐。



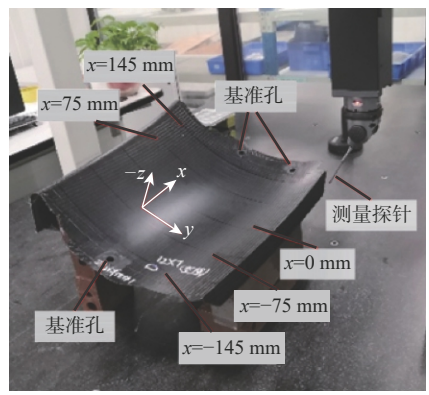
(a) 试验件结构与铺层示意图



(b) 热电偶分布



(c) 单向拉伸测试



(d) 三坐标测量

图 1 试验件制备与测量过程

Fig. 1 Manufacturing and measuring flowchart of tool composites part

固化过程中，在 Q235 母模和试验件之间嵌入热电偶进行温度场的测量，热电偶分布位置如图 1(b)所示。2 次固化完成后，测量单向铺层的随炉试件轴向拉伸模量，如图 1(c)所示。预固化和后固化结束后，分别采用三坐标测量仪(HEXAGON Explorer 08.12.06)对每个试验件的型面进行测量。测量线的分布位置如图 1(d)所示。

2 试验结果讨论

2.1 温度和模量试验结果

2.1.1 温度测量结果

预固化和后固化阶段的温度测量结果如图 2 所示。预固化升温阶段，由于升温速率较快，热电偶测得的温度滞后于热压罐输入的工艺温度。然而 3 个热电偶的温度差异很小，说明尽管温度滞后，但是温度场的分布仍然均匀。在保温初期，由于复合材料固化放热作用，热电偶测得实际峰值温度高出理论固化温度 3.4 K。30 min 后，复合材料放热效应减弱，热电偶温度与工艺温度的温差稳定在 1.6 K。降温阶段，热电偶降温速率同样滞后于工艺温度，如图 2(a)和图 2(b)所示。预固化后，工艺试验件固化度测试结果为 0.905，表明其固化反应趋于完全，因此，复合材料在后固化阶段的固化放热对温度测量结果的影响很小，热电偶温度的变化主要受热压罐热源影响。此外，后固化中没有母模和辅助材料约束，试验件能够直接与罐内气体进行强迫对流换热，加上更低的升温速率(0.3 K/min)，减小了热电偶测量温度与工艺温度的差异，两者最大温差为 1 K。预固化和后固化阶段，固化温度的热电偶测量值与输入值偏差均较小，最大为 3.4 K，表明试验件固化温度场分布均匀。

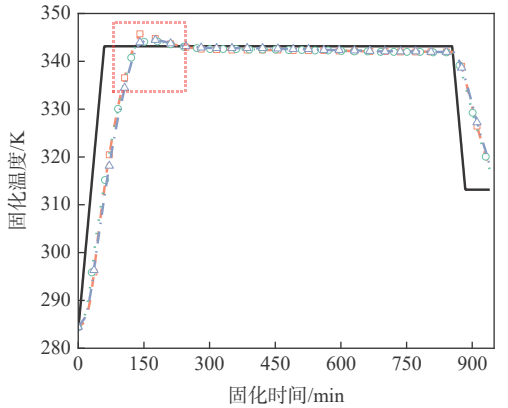
2.1.2 模量测试结果

随炉试样的单向拉伸试验结果与轴向弹性模量的线性拟合结果如图 3 所示，图中，δ 为应力，ε 为应变。5 个随炉试样弹性模量的线性拟合结果为(59.06±1.04) GPa，与供应商提供的数据(59.5 GPa)吻合良好，表明复合材料工装预浸料材料性能良好，满足试验要求。

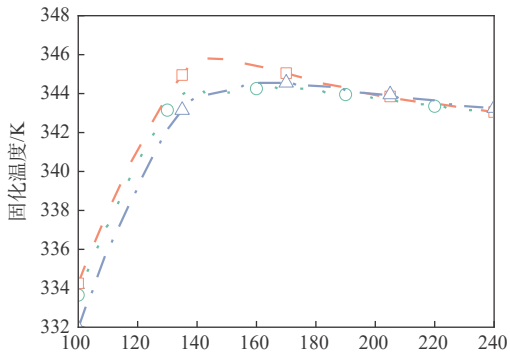
2.2 固化成型阶段的模具型面偏差

2.2.1 预固化阶段

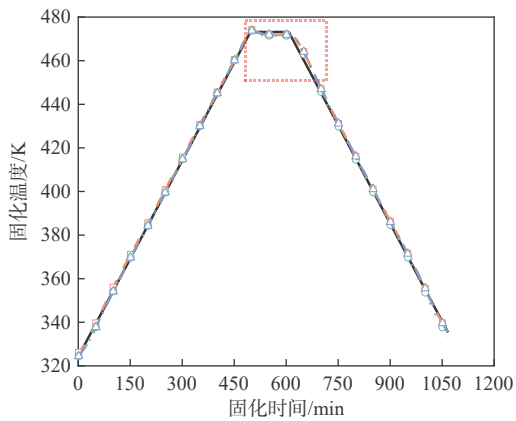
00 号、01 号、02 号试验件在 5 条测量线上(见图 1(d))的型面偏差平均值如图 4 所示。无加强筋的非准各向同性铺层 00 号试验件、无加强筋的准各向同性铺层 01 号试验件和有加强筋的准各向同性铺层 02 号试验件的型面偏差分别为 0.478 mm、0.462 mm 和 0.457 mm。在 x=−145 mm 和 x=−75 mm 位置，01 号试验件的型面偏差最大，分别为 0.451 mm 和 0.445 mm。在 x=0 mm 位置，3 个试验件的型面



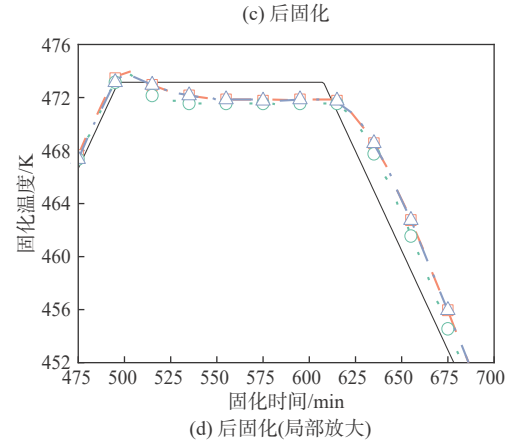
(a) 预固化



(b) 预固化(局部放大)



(c) 后固化



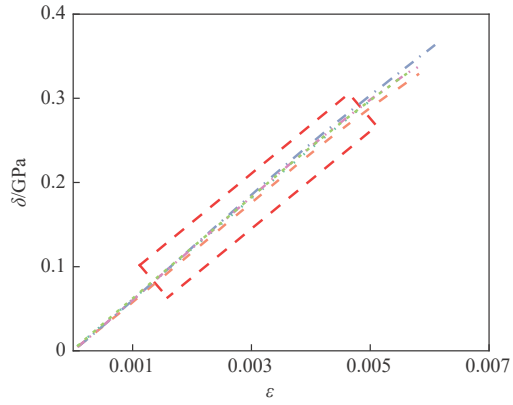
(d) 后固化(局部放大)



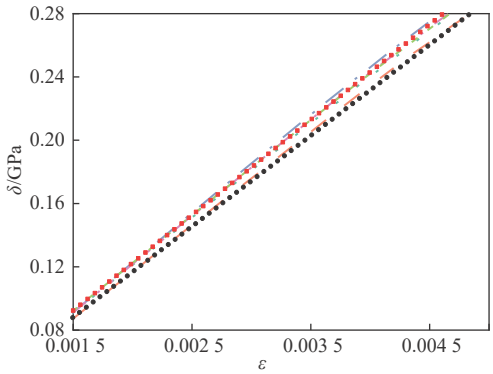
图 2　固化温度的热电偶测试结果

Fig. 2　Results of curing temperature measured by thermocouple

偏差均为 0.446 mm。在 x=75 mm 和 x=145 mm 位置, 00 号试验件的型面偏差最大, 分别为 0.510 mm



(a) 单向拉伸



(b) 单向拉伸 (局部放大)

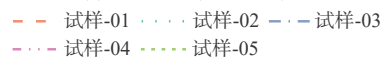


图 3　随炉试样的单向拉伸试验结果

Fig. 3　Results of unidirectional tensile test specimens

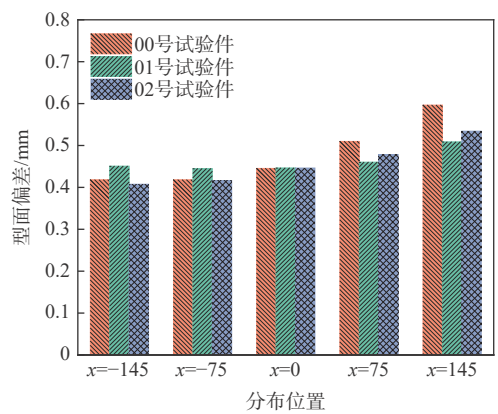


图 4　预固化阶段的型面偏差

Fig. 4　Profile deviation in pre-cure step

和 0.596 mm。3 个试验件型面偏差最大值均位于 x=145 mm 处, 分别为 0.596 mm、0.509 mm 和 0.534 mm。

预固化阶段, 复合材料试验件的固化度从 0 上升至 0.905, 树脂基体显著的材料性能变化造成热膨胀差异和固化收缩, 进而诱导产生固化变形。其中, 由树脂化学收缩诱导产生的残余应变可占总应变的 0.15%～0.48%[20-21]。此外, 尽管复合材料模具的 343.15 K 的预固化温度要显著低于复合材料构件 393.15～453.15 K 的工艺温度, 但由于辅助材料

及预固化压力、温度等因素的共同作用，金属母模会诱导试验件产生变形。而型面偏差大小随 x 的变化，可能是由于结构的不对称造成的。

2.2.2 后固化阶段

后固化后 00 号、01 号、02 号试验件在 5 条测量线上的型面偏差平均值如图 5 所示。无加强筋的非准各向同性铺层 00 号试验件的型面平均偏差最大，有加强筋的准各向同性铺层 02 号试验件的型面平均偏差最小。00 号、01 号、02 号试验件型面整体平均偏差分别为 0.634 mm、0.559 mm 与 0.315 mm。00 号和 01 号试验件的型面偏差分布情况与预固化阶段相似，随 x 的增大而增加，而 02 号试验件的型面偏差随 x 的增大，先减小后增大，并且在 x=0 mm 位置有最小值。

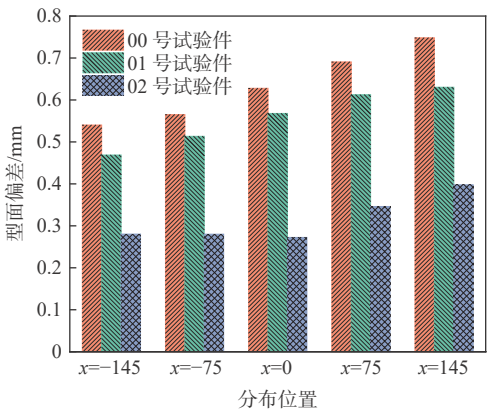


图 5 后固化阶段的型面偏差

Fig. 5 Profile deviation in post-cure step

后固化阶段，试验件固化度的 DSC 试验测量结果为 0.95，相比于预固化的 0.905，仅增加了约 5.6%，说明材料固化度变化很小，由此造成的热膨胀差异和树脂收缩对固化变形的贡献较小。因此，材料热膨胀差异、树脂固化收缩、模具-零件相互作用等预固化阶段的变形主要诱导因素在后固化阶段的影响有限。然而，后固化后，3 个试验件同样产生了显著的变形。其中，结构因素是造成固化变形的一个重要原因。如图 5 所示，非准各向同性铺层 00 号试验件在所有测量线上的型面偏差均大于准各向同性铺层 01 号试验件，说明采用准各向同性铺层能够减小复合材料的固化变形。其原因在于准各向同性铺层能够消除复合材料各向异性刚度矩阵中的耦合阵，避免诱导产生拉弯、弯扭等耦合变形，有助于减小型面偏差。而 02 号试验件由于有加强筋的存在，提高了结构刚度，因此，其型面偏差最小。

2.2.3 结构特征影响

00 号、01 号、02 号试验件的型面偏差累积变化量如图 6 所示，后固化后 00 号、01 号、02 号试验件的型面偏差依次减小。2.1.2 节的分析表明，铺层顺序和加强筋结构特征均对试验件的型面偏差有影响，然而两者影响的显著性尚不明确。为对比铺层顺序和加强筋特征对试验件型面偏差的影响情况，分别计算了 2 种特征在每条测量线上的型面偏差影响百分比，计算结果如图 7 所示。铺层顺序特征的型面偏差影响百分比最大为 20.1%，平均为 12.6%，而加强筋特征的型面偏差影响百分比最大为 28.5%，平均为 24.2%。加强筋特征对型面偏差的影响比铺层顺序平均高出 11.6%。结果分析表明相较于调整铺层顺序，采用加筋结构能够有效控制复合材料的固化变形[22-23]，尤其是对于需要采用两步固化工艺成型的复杂双曲率复合材料模具结构。

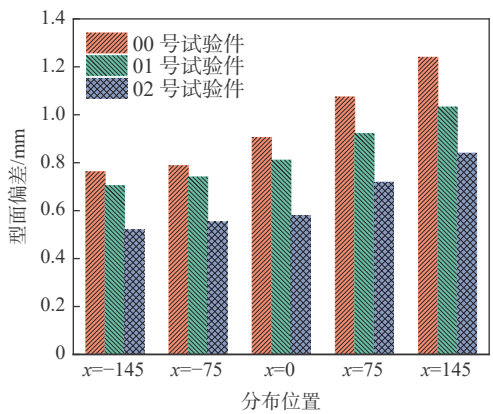


图 6 后固化阶段的累积型面偏差

Fig. 6 Accumulated profile deviation in post-cure step

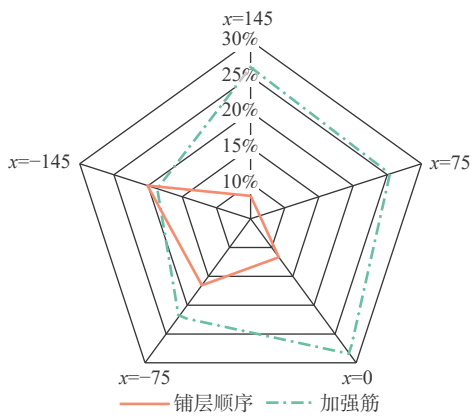


图 7 后固化阶段的型面偏差雷达图

Fig. 7 Radar map of profile deviation in post-cure

2.3 热循环使用阶段模具型面偏差

2.3.1 型面偏差的来源

2.2 节关于铺层顺序和加强筋结构特征对固化变形的影响分析结果表明固化变形对加强筋结构特征更为敏感，因此，对准各向同性铺层的 01 号无加强筋试验件和 02 号加强筋试验件的型面偏差进行进一步分析。01 号和 02 号试验件预、后固化后的型面偏差结果如图 8 所示。预固化后，2 个试验件的变形主要集中于右侧翻边位置，最大型面偏差

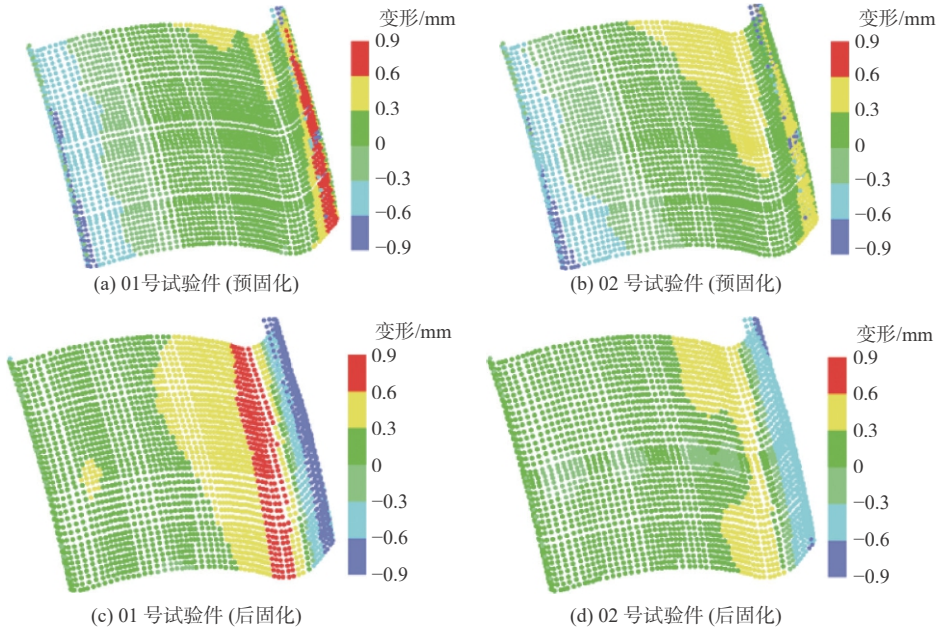


图 8　01 号和 02 号试验件的变形云图

Fig. 8　Cloud chart of deformation of No. 01 and No. 02 specimens in pre-cure and post-cure

分别为 1.08 mm 和 0.85 mm。后固化后, 01 号试验件的变形位置从右侧翻边下移, 集中于拐角位置, 最大型面偏差为 1.26 mm, 增加了约 16.7%。而 02 号试验件的变形位置同样发生了下移, 但是由于 x=0 mm 处的帽型加强筋的约束, 试验件壁板中部变形增加量有限, 最大型面偏差为 0.81 mm 减小了约 4.7%。试验结果显示试验件在后固化阶段均产生了显著的固化变形。

目前的研究主要将复合材料固化变形归结于材料热膨胀差异、树脂固化收缩、模具-零件相互作用等因素, 这对于解释预固化后测量到的型面偏差是准确的。然而, 后固化阶段测得的试验件型面偏差高于预固化阶段, 01 号试验件更是增加了 16.7%。预、后固化阶段的固化变形分布和大小差异仅依靠结构因素难以解释, 这表明试验件在后固化阶段产生的变形可能来自不同于预固化的其他诱因。

预固化阶段, 由于结构特征及金属母模影响, 纤维在拐角处发生屈曲或弯曲且树脂在该处堆积, 造成固化应力集中在拐角位置[10]。而后固化阶段, 试验件主要变形均产生于此拐角处, 如图 8 所示。因此, 拐角处集中的残余应力可能是导致后固化变形的潜在因素。由于材料、结构及工艺等因素影响, 热固性树脂基复合材料在固化后, 内部会产生显著的残余应力, 而在一定力或温度载荷作用下, 内部平衡的残余应力会重新分布引起外部变形也随之变化。在后固化高温作用下, 处于自由变形状态的试验件会发生应力松弛, 引起固化变形进一步变化, 从而造成模具型面精度降低。因此, 预固化阶段试验件产生的残余应力(集中于拐角处)的应力松弛行为, 是造成后固化阶段其型面变化的重要原因。

2.3.2　应力松弛影响

2.3.1 节的分析表明应力松弛可能是造成后固化模具型面偏差增大的重要原因。由于复合材料模具在使用过程中将经历多次高温热循环, 因此, 应力松弛行为将导致其形状在后续应用过程中不断变化, 降低结构稳定性, 不利于复合材料构件的精确成型。但由于材料内部应力水平有限, 其应力松弛特性将会随时间的增加而逐渐减弱, 并最终达到残余应力水平较低的稳定状态。基于此, 对复合材料模具而言, 通过增加热循环周期, 即延长成型时间, 有利于减小模具内部残余应力, 从而提高结构稳定性。因此, 对 01 号试验件进行第 3 次、第 4 次固化热循环并对其型面偏差进行分析, 固化工艺参数与测量基准和后固化保持一致。

试验件 4 次固化的型面偏差测量结果如图 9 所

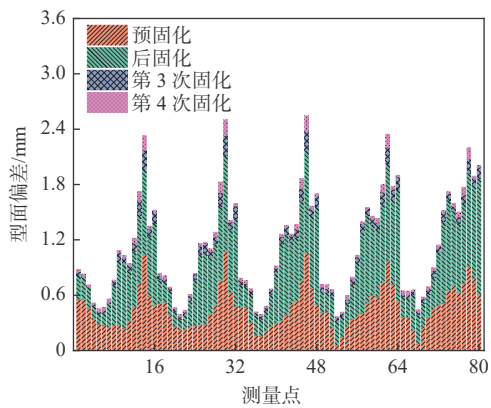
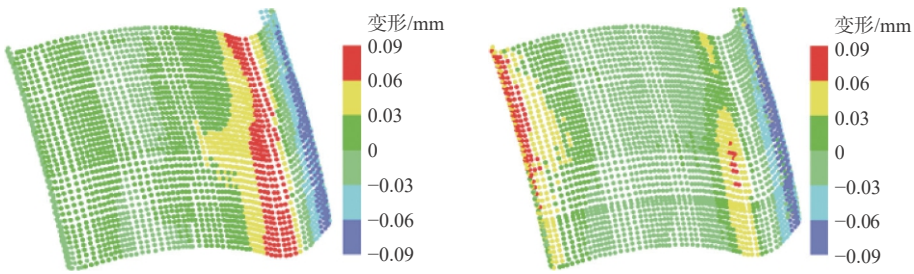


图 9　4 次固化后的 01 号试验件型面偏差

Fig. 9　Profile deviation of No. 01 specimen after four times cure steps

示。预固化和后固化后，试验件产生了较大的固化变形，整体型面偏差平均值分别为 0.46 mm 和 0.56 mm。而第 3 次和第 4 次固化热循环后的型面偏差平均值仅为 0.07 mm 和 0.03 mm，只有预固化和后固化阶段平均型面偏差的 13.7% 和 5.9%。

无约束自由变形状态下的试验件经过后固化后，型面发生了显著变化，平均偏差高出预固化约 20%，说明在后固化阶段，应力松弛行为会诱导内部残余应力重新分布，进而造成外部变形变化。而与后固化相比，试验件第 3 次固化热循环后的型面偏差最大值为 0.22 mm，且超过 98% 的数据点均位于±0.2 mm 内，如图 10 所示。第 4 次固化热循环后的最大型面偏差为 0.17 mm，且约 98% 的数据点均位于±0.15 mm 内，右侧拐角位置的型面偏差减小到±0.01 mm 内。结果表明：该处在预固化和后固化阶段累积的残余应力已经稳定到一个很低的水平，此时，左侧拐角较低的残余应力水平开始主导型面偏差的变化。



(a) 第 3 次固化　　(b) 第 4 次固化

图 10　01 号试验件型面偏差云图对比

Fig. 10　Comparison of profile deviation of No. 01 specimen

3　结　论

本文对采用特殊 2 步固化工艺成型的 S 形变截面类回转体复合材料模具局部典型特征结构试验件的型面偏差产生原因与变化规律进行了试验研究，在此基础上得出了以下结论。

1）试验件在预固化和后固化成型阶段的型面偏差平均值分别为 0.46 mm 和 0.56 mm。预固化阶段，试验件的型面偏差主要由材料热膨胀差异、树脂固化收缩及模具-零件相互作用造成的固化应力引起；而后固化阶段，预固化阶段产生的固化残余应力的松弛行为是引起模具型面偏差变化的主要原因。

2）试验件在第 3 次和第 4 次固化热循环后的型面偏差最大值分别为 0.22 mm 和 0.17 mm。试验件的型面偏差变化量随着热循环周期的增加而减小。通过增加热循环周期，利用复合材料应力松弛行为来降低模具在固化过程中积累的残余应力水平，有助于提高其型面精度的稳定性。

对于本文的复合材料模具特征结构试验件，经过第 4 次固化热循环后，固化成型阶段（预固化和后固化阶段）产生的残余应力的进一步变化量很小，满足使用要求。然而，对于尺寸更大、结构更复杂的复合材料模具，需要根据实际情况来选择热循环周期以降低结构内部的残余应力水平。

参考文献（References）

[1] HASSAN M H, OTHMAN A R, KAMARUDDIN S. A review on the manufacturing defects of complex-shaped laminate in aircraft composite structures[J]. The International Journal of Advanced Manufacturing Technology, 2017, 91(9): 4081-4094.

[2] 刘振东, 郑锡涛, 范雯静, 等. 固化残余应力对无人机复合材料机翼强度的影响[J]. 航空学报, 2022, 43(6): 526117.
LIU Z D, ZHENG X T, FAN W J, et al. Effect of process-induced residual stress on strength of UAV composite wing[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(6): 526117(in Chinese).

[3] LIAN J Y, XU Z B, RUAN X D. Analysis and control of cured deformation of fiber-reinforced thermosetting composites: a review[J]. Journal of Zhejiang University: Science A, 2019, 20(5): 311-333.

[4] 岳广全, 张博明, 戴福洪, 等. 固化过程中模具与复合材料构件相互作用分析[J]. 复合材料学报, 2010, 27(6): 167-171.
YUE G Q, ZHANG B M, DAI F H, et al. Interaction between mold and composite parts during curing process[J]. Acta Materiae Compositae Sinica, 2010, 27(6): 167-171(in Chinese).

[5] YANG Q D, WEI K, YANG X J, et al. Microstructures and unique low thermal expansion of Invar 36 alloy fabricated by selective laser melting[J]. Materials Characterization, 2020, 166: 110409.

[6] GRANKÄLL T, HALLANDER P, PETERSSON M, et al. The true shape of composite cure tools[J]. Journal of Manufacturing Processes, 2020, 59: 279-286.

[7] 肖遥, 李东升, 吉康, 等. 大型复合材料航空件固化成型模具技术研究与应用进展[J]. 复合材料学报, 2022, 39(3): 907-925.
XIAO Y, LI D S, JI K, et al. Research and application progress of curing tooling technology for large composite aeronautical components[J]. Acta Materiae Compositae Sinica, 2022, 39(3): 907-925(in Chinese).

[8] 魏灵航, 安鲁陵, 鲍益东, 等. 基于响应面的框架式复材成型模具轻量化设计[J]. 南京航空航天大学学报, 2020, 52(3): 422-429.
WEI L H, AN L L, BAO Y D, et al. Lightweight design of frame molding die for composite materials based on response surface[J].

Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(3): 422-429(in Chinese).

[9] ZHANG G M, WANG J H, NI A Q, et al. Process-induced deformation of L-shaped variable-stiffness composite structures during cure[J]. Composite Structures, 2019, 230: 111461.

[10] 王乾, 关志东, 王仁宇, 等. 结构参数对复合材料 V 型构件固化变形影响试验及解析分析[J]. 复合材料学报, 2017, 34(12): 2722-2733.
WANG Q, GUAN Z D, WANG R Y, et al. Experimental and analytic analysis of the structure parameters on process-induced distortions of V-shaped composite parts[J]. Acta Materiae Compositae Sinica, 2017, 34(12): 2722-2733(in Chinese).

[11] KAPPEL E. Forced-interaction and spring-in–relevant initiators of process-induced distortions in composite manufacturing[J]. Composite Structures, 2016, 140: 217-229.

[12] 贾丽杰, 叶金蕊, 刘卫平, 等. 结构因素对复合材料典型结构件固化变形影响[J]. 复合材料学报, 2013, 30(S1): 261-265.
JIA L J, YE J R, LIU W P, et al. Influence of structural factor on curing deformation of typical composite structural part[J]. Acta Materiae Compositae Sinica, 2013, 30(S1): 261-265 (in Chinese).

[13] FIORINA M, SEMAN A, CASTANIE B, et al. Spring-in prediction for carbon/epoxy aerospace composite structure[J]. Composite Structures, 2017, 168: 739-745.

[14] DONG C S. Process-induced deformation of composite T-stiffener structures[J]. Composite Structures, 2010, 92(7): 1614-1619.

[15] 王雪明, 谢富原. 碳纤维/双马树脂复合材料整体成型过程分层扩展行为实验研究[J]. 航空学报, 2021, 42(2): 424918.
WANG X M, XIE F Y. Experimental study on behavior of delamination propagation of carbon fiber/bismaleimide composites during integral forming process[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(2): 424918(in Chinese).

[16] 杨晓波, 湛利华, 蒋成标, 等. 模具材料对复合材料制件固化过程应变的影响分析[J]. 南京航空航天大学学报, 2018, 50(1): 24-29.
YANG X B, ZHAN L H, JIANG C B, et al. Influence of mould materials on curing process and deformation of composite part[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2018, 50(1): 24-29(in Chinese).

[17] LIU X D, GUAN Z D, WANG X D, et al. Study on cure-induced residual stresses and spring-in deformation of L-shaped composite laminates using a simplified constitutive model considering stress relaxation[J]. Composite Structures, 2021, 272: 114203.

[18] BENAVENTE M, MARCIN L, COURTOIS A, et al. Numerical analysis of viscoelastic process-induced residual distortions during manufacturing and post-curing[J]. Composites Part A: Applied Science and Manufacturing, 2018, 107: 205-216.

[19] BENAVENTE M, MARCIN L, COURTOIS A, et al. Viscoelastic distortion in asymmetric plates during post curing[J]. Composites Part A: Applied Science and Manufacturing, 2017, 103: 122-130.

[20] HU H X, LI S X, WANG J H, et al. Monitoring the gelation and effective chemical shrinkage of composite curing process with a novel FBG approach[J]. Composite Structures, 2017, 176: 187-194.

[21] GARSTKA T, ERSOY N, POTTER K D, et al. In situ measurements of through-the-thickness strains during processing of AS4/8552 composite[J]. Composites Part A: Applied Science and Manufacturing, 2007, 38(12): 2517-2526.

[22] SHAH D B, PATEL K M, PATEL A I, et al. Experimental investigation on spring-back deformation during autoclave curing of parabolic antenna reflectors[J]. Composites Part A: Applied Science and Manufacturing, 2018, 115: 134-146.

[23] FERNLUND G, POURSARTIP A, TWIGG G, et al. Residual stress, spring-in and warpage in autoclaved composite parts[C]// International Conference on Composite Materials. Vancouver: University of British Columbia, 2003: 1-10.

Influence of curing stress relaxation on profile accuracy of composites tools

XIAO Yao[1], LI Yong[1], LI Dongsheng[1,*], WANG Lei[2], JIANG Chao[3]

(1. School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China;
2. Hunan Aerospace Huanyu Communication Technology Co., Ltd., Changsha 410006, China;
3. AVIC Shenyang Aircraft Co., Ltd., Shenyang 110850, China)

Abstract: Tooling composites prepreg cured at low temperatures and used in high temperature is one of the most promising materials for forming composite tools to manufacture high-precision composite structures in the aerospace field. Tooling composites prepreg utilizes a specially designed two-step method that consists of low-temperature procuring and high-temperature post-curing in order to realize its low cost and high precision manufacture. The curing stress induced in composite tool manufacturing not only affects the accuracy of the initial tool profile, but may produce stress relaxation in the process of thermal recycling, which further causes the change of tool profile accuracy, and then affects the accuracy of components manufacturing. In this paper, the variation rules and influencing factors of profile accuracy of the complex profile characteristics of carbon fiber reinforced composite tools during solidification and thermal recycling were studied by experiments. The results show that the curing stress in the manufacturing process and the stress relaxation in the process of thermal recycling are the main factors that cause the variation of tool profile accuracy, and the profile deviation tends to be stable with the increase of the number of thermal recycling. Subsequent investigation reveals that the primary causes of the tool profile deviation during the manufacturing process are the interaction between the master mold and the composites tool, resin curing shrinkage, and thermal expansion mismatch-induced curing stress. In contrast, the primary cause of the mold surface deviation during the use stage is the relaxation behavior of the residual curing stress. The research results can help to understand the changes in the structural stability of composite tools. Meanwhile, it is of great significance to reduce the manufacturing cost of composite tools, improving the service life of composite tools and realizing the precise manufacturing of complex and high-precision composite components.

Keywords: composites tool; curing deformation; residual stresses; stresses relaxation; accuracy of tool profile

Received: 2023-03-07; **Accepted**: 2023-07-07; **Published Online**: 2023-09-22 08:17

URL: link.cnki.net/urlid/11.2625.V.20230920.2014.002

Foundation item: The Fundamental Research Funds for the Central Universities (YWF-23-L-1012)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0089

基于语义关系的单视图隐式场三维重建技术

徐岗[1,2]，靳炯超[1]，唐泽皓[1]，冷彪[1,*]

(1. 北京航空航天大学 计算机学院，北京 100191；2. 平安银行股份有限公司，深圳 518001)

摘 要：随着对深度学习的深入研究，基于隐式场的三维重建技术成为三维领域的研究热点，隐式场重建相比于显式重建得到了更好的重建结果，但仍存在缺少语义信息、重建结果缺失局部特征等缺点。针对现有的隐式场重建中存在的语义信息缺失问题，对隐式场重建算法提出改进，相比于直接由隐式解码器恢复出整体模型，改进方法选择先重建模型的各个语义零件，再将各个零件进行拼接得到整体模型，进而提高隐式场三维重建的完整度。基于改进方法，提出基于语义关系的多分支隐式重建网络，并在多个国际通用三维模型数据集中与现有隐式场重建算法在5个类别的模型上利用多项评价指标进行定量测试，取得了优于当前隐式场重建算法的重建结果。

关 键 词：三维重建；语义分割；隐式场；单视图重建；多分支网络
中图分类号：TP391.9
文献标志码：A **文章编号**：1001-5965(2025)03-0833-12

图形学是计算机科学中最活跃的分支之一。近年来，图形学在图形视频处理、工业建模、游戏制作、生物信息、医药医疗等各行各业都有着极其重要的作用。相比于二维图像，三维模型能提供更多信息，而三维重建是快速获得三维模型的重要手段，因此，如何更好地从二维图像中恢复物体的三维模型逐渐成为计算机视觉领域的主要研究方向。同时，研究者整合了许多大型三维模型数据集，这为三维重建技术的发展提供了良好的研究基础。

近年来，随着深度学习的发展，计算机对视觉信息的提取和学习能力进一步提升，三维重建生成的模型质量也大大提高。根据重建过程中是否使用深度学习，三维重建方法可以分为传统三维重建方法和基于深度学习的三维重建方法。基于深度学习的三维重建方法又可以进一步分为显式三维重建和隐式场三维重建，其中，显式三维重建利用输入的单张或多张二维视图直接生成重建的三维形状表示，同时根据最终生成的形状表示是体素、点云还是三角网格，其对应的网络结构也会相应地有所不同。而隐式场重建则利用输入的二维视图或采样的三维空间点，先由神经网络预测得到每个三维空间点对应的有符号距离函数值(signed distance functiom, SDF)，再利用提取零等值面的方法生成最后的三维模型，隐式场重建算法可以根据采样点数量的变化生成任意分辨率的重建三维形状表示结果。显式重建算法会受到几何表示结构不规则或空间消耗过大等问题影响，在研究和实际应用上有较大局限；隐式场重建算法使用神经网络学习三维模型的隐式场函数，从本质上更改了三维模型的表示形式，因此，不存在几何表示和空间消耗的问题。而且隐式场表示能够充分发挥神经网络的函数拟合能力，在深度学习技术的加持下，在各项任务上都能取得更好的效果。

然而，目前隐式重建技术还存在以下几个方面的局限[1-2]：①位于三维物体细节部分表面的三维采

收稿日期：2023-02-28；录用日期：2023-05-08；网络出版时间：2023-07-10 11:05
网络出版地址：link.cnki.net/urlid/11.2625.V.20230707.2037.001
基金项目：国家自然科学基金 (61972014)
*通信作者. E-mail：lengbiao@buaa.edu.cn

引用格式：徐岗，靳炯超，唐泽皓，等. 基于语义关系的单视图隐式场三维重建技术 [J]. 北京航空航天大学学报，2025，51（3）：833-844.
XU G，JIN J C，TANG Z H，et al. Semantic part based single-view implicit field for 3D shape reconstruction technology [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（3）：833-844（in Chinese）.

样点往往包含更重要的几何信息和结构特征，但现有隐式场重建算法的采样方式是均匀采样，对三维模型局部特征的采样较少，因此，三维模型的局部细节很容易被神经网络忽略，同时固定长度的特征编码也很难较好地表示物体的细节部分，从而使得重建结果精度产生缺失，针对细节缺失问题，本文提出的模型增加了解码器数量，减少了每个解码器的重建负担，使得每个解码器得到的结果精度更高；②二维视图不包含对于重建任务至关重要的物体几何信息，因此，网络只能通过图像信息提取物体几何信息，而且由二维特征到三维特征的转换过程也会产生一定的信息损失，这对网络的特征提取能力提出挑战，进一步加大了重建难度，这些困难在单张二维视图的重建中会更加明显，另外，单视图只能反映三维物体在一个角度的视觉呈现，这使得二维视图的图像特征在转化为有符号距离函数值时由于维度的转换和二维图片与三维世界的映射关系不明确而容易产生重建结果的部分缺失，针对单张二维视图的视角遮挡问题，本文让网络中的每个解码器承担零件层面的重建任务，迫使网络解码器更多地关注零件的几何特点，缓解了视角遮挡带来的重建模型特征缺失的问题；③在单视图三维重建过程中，容易出现由于输入图片视角过差导致最终重建结果不可用，利用模型的语义信息帮助网络进行重建可以有效解决这个问题，而目前多数隐式场重建算法直接从输入得到整体三维模型，重建过程中缺少对模型语义信息的利用。针对重建结果缺少语义的问题，本文设计了适合三维重建的语义模型分割方法，并基于分割后的零件设计网络架构，使得网络先重建生成模型的各个语义零件，根据零件的位置信息拼接成整体模型，使得重建模型带有一定的语义信息。

现有隐式场重建算法存在的局限启发了本文的研究开展和实验设计。本文的主要创新点如下。

1）提出多分支隐式重建网络，在解码器承担零件级重建任务的前提下对网络参数进行了轻量化，同时得到了较好的重建结果。

2）优化了解码器重建目标，使得每个解码器能关注模型的几何关系和局部特征，缓解了视角遮挡造成的重建模型特征缺失问题。

3）相比于现有隐式场重建算法通过全局特征得到重建结果及通过全局特征和局部特征重建整体模型后恢复模型细节，本文方法在仅利用轻量级卷积网络对输入图片提取全局特征的前提下，得到了重建精度较高的重建结果。

本文方法相较于目前已有的隐式场重建算法，在重建的同时关注了物体的语义信息，并将其有效利用，形成最终的重建三维形状。同时，在国际通用模型重建数据集 ShapeNet、PartNet 上与其他隐式场三维重建算法的对比实验结果表明，通过有效利用物体的语义信息有助于更好地生成三维形状表示，使模型的重建能力得到提升。

1 相关工作

根据重建过程中是否使用深度学习，三维重建方法可以分为传统三维重建方法和基于深度学习的三维重建方法。传统的三维重建方法主要利用物体的几何信息与二维三维投影的联系来进行三维重建，大多涉及数学计算；而基于深度学习的三维重建方法则利用神经网络模拟人类的认知思考过程，直接利用图片推理出对应的三维形状表示。更进一步，基于深度学习的三维重建方法又可以进一步分为显式三维重建和隐式场三维重建，其中，显式三维重建利用输入的单张或多张二维视图直接生成三维模型，同时根据最终生成的三维模型由体素、点云还是三角网格表示，其对应的网络结构也会相应地有所不同。

隐式场重建利用输入的二维视图或采样的三维空间点，先由神经网络预测得到每个三维空间点对应的有符号距离函数值（signed distance functiom, SDF），再利用提取零等值面的方法生成最后的三维模型，隐式场重建算法可以根据采样点数量的变化生成任意分辨率的重建三维形状表示结果。

1.1 传统三维重建方法

传统的三维重建方法利用同一个物体的不同二维视图之间的图像信息和三维物体的几何信息的关联，并且利用视差几何辅助运算求解三维模型信息。传统的三维重建方法主要包括阴影恢复形状法（shape from shading, SFS）、结构光法和运动法。

SFS 主要利用图片中物体的边缘阴影隐含的物体轮廓信息进行三维重建。SFS 最早由 Horn 在 1970 年提出[3]。但 SFS 方程要求唯一解需要额外约束条件[4]，在满足约束条件的情况下阴影恢复形状法仍然被局限于简单光照的重建场景。

结构光法通过对被测物体投射带有特征点的光线，然后利用投影得到的结构光图案中的失真信息恢复三维物体的表面形状[5]。陈荣等[6]设计了一种利用单色光三角函数条纹和伪随机点的结构光编码算法对不同光照条件下的模型进行了重建。

运动法全称为运动恢复形状法（shape from motion, SFM），运动法通过分析物体的二维视图序列，计算相机参数信息从而对三维物体进行重建。

Snavely 等[7]较早提出 SFM 的实现方法, 并且可以在使用几百和几千数量级的独立拍摄照片的实际场景中进行精度较高的重建。Lowe[8]提出一种应用较广的尺度不变特征变换 (scale invariant feature transform, SIFT), 利用 SIFT 把输入图像转换为一组局部特征, 使得 SFM 的重建精度大大提高。然而输入图片序列规模较大导致运动法在进行图片特征提取和匹配时计算量巨大, 所以运动法重建往往效率不高。

1.2　多视图重建方法

随着深度学习和计算机视觉技术的发展, 研究者开始把深度学习应用于三维重建领域。基于深度学习的三维重建方法首先用神经网络从输入的二维图像中提取特征向量和几何信息, 再将其转换为物体的三维信息, 重建指定的三维模型, 根据输入图片数量的不同, 这些三维重建方法又分为单视图三维重建和多视图三维重建。

多视图三维模型重建是利用多张视图进行三维模型重建任务, 与单视图重建任务相比, 其需要多张视图, 并且对拍摄角度等因素有一点要求。但多视图重建由于融合了多个角度的视图信息, 重建出的三维模型更加精确、全面。多视图重建还需要考虑多视图特征的融合方式, 避免视图融合过程中的信息丢失。

Choy 等[9]提出 3D-R2N2 网络, 该网络结构由二维编码器和三维解码器组成的标准神经网络模型进行多视图, 其中, 三维编码器继承自视觉几何组 (visual geometry group, VGG) 网络[10]。三维解码器由三维卷积、三维长短时记忆单元和三维反卷积代替二维神经网络部件(如二维卷积等), 使神经网络能处理三维特征。Wu 等[11]继续将已有神经网络组件拓展到体素上, 提出多个体素处理网络, 如三维生成对抗网络、三维可变自动编码器等, 极大地增强了网络的体素生成能力。Xie 等[12]设计了 Pix2-Vox 网络, 该网络结构根据各个视图分别进行粗略重建, 再使用上下文感知模块对各个重建结果进行评分, 通过加权得到融合多视图的三维特征。Wu 等[13]使用轮廓图、深度图、法向图等多种中间特征作为监督分阶段地预测三维模型, 首先, 通过二维卷积神经网络从输入图像预测这些中间特征, 然后, 基于这些中间特征提取三维模型, 这样能够减轻网络学习的负担, 从而便于重建网络的训练。Wang 等[14]将可视外壳的概念引入特征提取中, 提出三维模型特征生成的双分支网络结构。其中, 一个分支是常规的体素生成网络, 用于生成三维模型的粗略形状, 另一个分支利用 2 个卷积神经网络预测三维模型的轮廓图和输入图片的相机矩阵参数, 通过对 2 个信息组合得到三维模型的特征外壳, 再基于特征外壳和三维模型粗略形状生成精细的三维模型。

1.3　单视图重建方法

单视图重建是利用三维物体的一张二维渲染图恢复出物体的原本形状的三维重建方法。在只有一张渲染图的情况下, 在某些角度人眼甚至都无法确定三维模型的具体结构, 因此, 单视图重建网络的重建质量很大程度受到输入图片对应的模型视角的影响。这使得神经网络很难准确学习到三维形状分布的特征, 也给单视图重建增加了难度。

1.3.1　显式重建方法

Pontes 等[15]提出不需要图片标注和轮廓的网络框架, 该框架包含卷积自动编码器、多标签分类器和前馈网络。卷积自动编码器模块从图像中提取特征。随后特征向量被分别送入多标签分类器和前馈网络, 其中, 多标签分类器的作用是学习特征向量与不同的模型对应图像之间的映射, 从而预测出最接近该图像的三维模型。前馈网络通过稀疏线性组合参数和估计参数辅助三维模型重建。Jack 等[16]使用特征金字塔结构设计了一个轻量级卷积神经网络, 该网络通过学习如何从单张图像输入中选择基础形状并将基础形状进行形变, 最终生成最匹配输入图像的三维模型。该网络将输入图像映射到特征向量空间, 再将特征向量映射到每个单独基础形状的变形参数来完成上述过程, 还引入了显式熵惩罚和变形正则化, 并对不同的训练机制进行了实验。Li 等[17]提出一种使用二维卷积网络通过图片生成密集点云式三维模型的方法。首先, 在图像编码器产生输入的特征向量后, 结构生成器对该向量使用二维卷积运算生成若干多通道图像, 并用这些图像表示来自不同视点的三维形状表面上的点的坐标; 然后, 通过正交投影计算将所有结构生成器生成的视图融合到规范坐标空间中; 最后, 使用伪渲染器生成深度图, 这种方法提高了模型对于不同种类三维模型的重建能力。

Kong 等[18]基于已有面片特征数据库, 提出通过组合已有面片进行单视图特征提取的双阶段算法。首先, 算法根据输入图像估计出相机的拍摄姿态; 然后, 将标注好的关键点在二维平面和三维空间中进行关联, 从而在面片特征数据库中找到最匹配的面片模型作为基础; 最后, 算法根据基础模型进一步优化相机的拍摄姿态, 线性地组合近似面片模型到基础面片上, 从而得到准确的全局面片特征。因为这个算法涉及的环节过多, 有累计误差较大的问题。Groueix 等[19]提出 AtlasNet 算法, 该算

法建立在拓扑学假设之上，即二维平面块进行表面参数化可以转化得到大部分的三维面片，而三维面片模型可以通过组合三维面片得到，则面片特征提取可以参数化转换一组二维平面块实现。AtlasNet网络使用全连接网络拟合二维平面块到三维面片之间的映射关系，使用多个全连接网络组合，将多个二维平面块映射为不同的三维面片，再将其进行组合，恢复出完整的三维形状。

1.3.2 隐式场重建算法

Chen和Zhang[20]基于隐式场建模方法，提出隐式场网络重建(implicit network, IM-Net)算法。隐式场重建算法与显式重建方法的本质区别在于解码器网络，该网络输入一个特征向量和空间点坐标，特征向量指定输入的三维形状信息，空间点坐标指明重建位置。网络输出这组空间点是否在三维模型内部的二分类标签。IM-Net解码器由全连接网络实现，结构简单、易于实现。OccNet[21]和IM-Net的网络结构基本一致，二者同样采用隐式场建模，在图像编码器及网络训练策略上存在差异。Park等[2]设计了DeepSDF算法，DeepSDF没有采用传统的编码器、解码器结构，而是提出自动解码器结构和精巧的训练方式，其将解码器所需的特征向量初始化为随机向量，然后，跟随解码器参数一同进行训练。在测试时，冻结解码器参数，只优化特征向量，经过一定次数的迭代后，取得测试样例的特征向量进行重建。Wang等[22]率先将图像局部特征引入隐式场解码器。其使用额外的全连接神经网络学习输入图像的相机参数矩阵，该矩阵提供了三维坐标点到二维像素点的映射。然后，在隐式场重建的过程中，将空间点映射为平面点坐标，提取出平面点处的图像特征，输入隐式场解码器，从而增强重建每个点时解码器获得的局部信息，提升了隐式场解码器对三维空间的感知能力。Bechtold等[23]通过局部三维模型重建的方式实现了图像局部特征的引入。其将图像划分为块，对应着将三维模型也划分为区域。将图像块分别输入重建网络，监督算法重建对应的三维区域。这样的学习机制能够保证网络不能从高层语义生成三维模型，而只能根据图像的纹理信息重建三维模型。相当于三维模型的各个区域被放大之后再进行重建，所以能够一定程度上克服重建结果过于平滑的问题。在算法测试时，为解决三维区域交界处如何拼接的问题，算法令重建出的三维模型区域存在重叠，然后，使用高斯核函数对各个局部区域进行加权处理，组合形成最后的三维模型。Li和Zhang[1]将隐式场三维重建分成主体重建和细节重建2个环节，这2个环节分别用单独的隐式场函数进行学习。主体重建通过普隐式场重建网络接收全局特征，重建出相对平滑的三维模型形状。细节重建以预测模型局部细节与整体形状之间残差的方式进行，通过单独的神经网络预测三维点到像素点之间的映射，引入图像局部特征。为更充分地监督重建方法对模型纹理的学习，重建结果会利用可微分渲染技术对重建模型渲染得到二维图像，将这张图像与原图像素之间的均方误差作为重建损失函数，从而形成图像纹理到三维模型之间的直接监督。

2 本文方法

本文提出的多分支隐式重建网络结构如图1所示，输入图片先经过残差网络构成的图片编码器得到128维全局特征，全局特征分别通过不同的隐式解码器后得到模型的各个语义零件，根据零件之间的位置关系拼接得到最终模型。

2.1 隐式场解码器

基于隐式场的重建方法舍弃了直接恢复模型的三维反卷积网络解码器，转为解决三维空间中的每单位立方体是否位于三维物体内部的二分类问题，网络输出每个单位立方体的隐式场函数值，再利用提取零等值面的方式渲染得到最终的三维模型。由于单位立方体可以划分至任意小，隐式场重建可以呈现任意分辨率的重建结果，弥补了体素重建只能针对低分辨率场景的缺陷。因此，相比于显式重建，隐式场重建的结果相对精细和平滑。

隐式场这一概念来自SDF。某点的SDF的定义是：距离该点对应的单位立方体的最近曲面点。如果此点在曲面内，SDF为负数，否则为正数。SDF的函数值分布示意图如图2所示。

在实际重建过程中，空间点的SDF符号比数值更有价值。该点的SDF符号表征该点位于曲面内还是曲面外，隐式场解码器正是通过预测空间坐标的SDF符号将三维重建简化为典型的二分类问题。

因此，隐式场函数对SDF进行简化，其定义是一个在三维空间的连续函数，其接收一个点的坐标和输入图片的特征向量，输出该点位于三维模型内部还是外部的二分类标签。二分类问题可以通过多层感知机(multi layer perception, MLP)解决。当网络具有足够的隐藏层时，MLP就能拟合任意函数曲面，也就能够拟合出隐式场函数。隐式场函数的数学表示如下：

$$F(\boldsymbol{p})=\begin{cases}0 & \boldsymbol{p}\in S\\ 1 & \text{其他}\end{cases} \tag{1}$$

式中：$F(\cdot)$为隐式场函数；$\boldsymbol{p}$为空间中三维点坐标；S表示三维模型；$\boldsymbol{p}\in S$表示点在三维模型内部。基

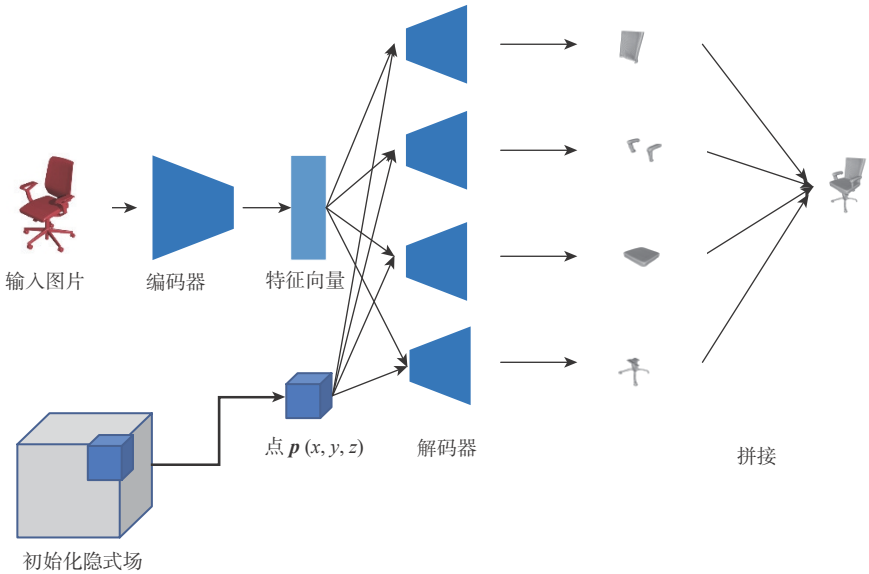


图 1　多分支隐式网络整体结构

Fig. 1　Architecture of multi branch implicit network's overall structure

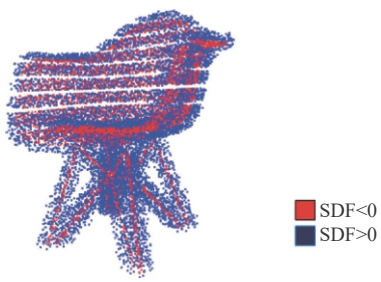


图 2　SDF 值分布示意图

Fig. 2　Visualization of the distribution of SDF values

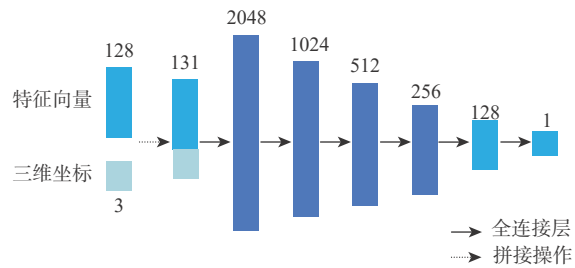


图 3　隐式场解码器结构

Fig. 3　Architecture of implicit decoder

于隐式场的三维模型重建就是寻找一组网络参数 θ, 使隐式场解码器能将点 $\boldsymbol{p}$ 映射到 $F(\boldsymbol{p})$。

为更好地拟合隐式场函数, 隐式场解码器的全连接层需要具备一定深度和参数量。本文采用 Leaky ReLU 函数作为隐式解码器的激活函数, 该函数可以避免神经单元在输出负值时无梯度反馈的现象, 加速网络训练。本文输入图像分辨率较小, 内容比较单纯, 不存在噪声, 因此, 本文选择较浅的卷积神经网络实现图像编码器, 可以保证特征向量含有足够的语义信息。隐式场解码器采用全连接网络实现, 本文使用的隐式场解码器具体参数如图 3 所示。

2.2　三维模型的分割方法

2.2.1　PartNet 的语义分割方法

确定对三维模型的分割方法是本文进行重建工作的基础。三维模型分割算法中, 主流算法通常在一个模型类别中事先确定模型包含的所有零件标签, 再针对这些固定的零件标签对模型进行分割。也有少数分割算法不对模型进行标签的分割, 而是利用递归神经网络, 依据零件的对称性和邻接关系对三维模型进行分割。后者的优势在于对于不同类别的三维模型, 不需要重新对模型进行标签。然而在本文中, 零件标签数量和网络直接相关, 零件数量确定后解码器数量也随之固定, 所以本文选择对一个类别中模型的所有零件进行标签, 然后, 针对固定标签对模型进行分割的方法对三维模型进行处理。

PartNet 认为语义分割中缺乏公认的经验准则来定义好的标签模板, 文献 [24] 总结了如下模型分割标准: ①定义的零件应该能被多数人清楚地识别; ②没有非必需的零件, 定义的零件应该尽可能复用; ③定义的零件在不同的模型类别中也应该尽可能复用; ④定义的零件应该有层次性, 包含粗略分割的零件和精细分割的零件; ⑤需要定义原子零件。

在上述标准的指导下, PartNet 为近 27 000 个模型进行了 3～7 层不等的分解, 并为每个模型生成零件用于树记录零件标签、文件名及所属父零件作为模型注释, 如图 4 所示。该分割标准只是 PartNet 数据集在制作时依照的规则, 并没有得到广泛认可, 从三维重建的角度出发, PartNet 对零件的分解

比较精细，直接套用这些标签会对重建任务造成如下困难：①标签数量增加会相应增加解码器的数量，使整体重建网络变得更加复杂，训练时间也会成倍上升；②部分标签对应的模型数量较少，这会导致相应的解码器接受过多的空样本，在解码器利用率减少的同时也会影响重建准确率。

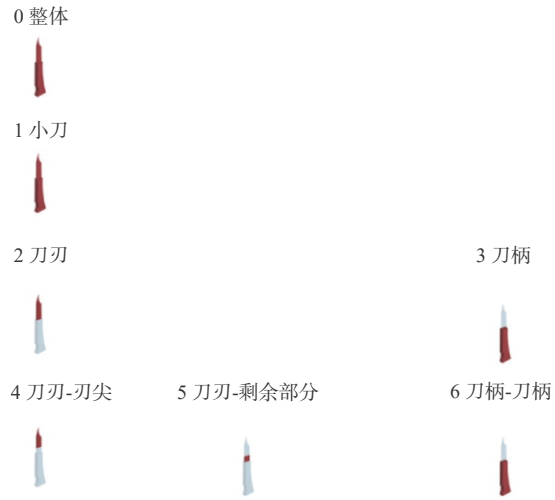


图4 PartNet定义的零件树

Fig. 4 Tree structure in PartNet

2.2.2 基于最大语义零件的分割方法

针对重建任务，本文设定了4条分割准则：

1）定义的零件可以覆盖整个模型；

2）零件对应的解码器空样本不应该超过50%；

3）零件之间没有重复部分；

4）不存在零件数更少的分割方法，且零件数量限制在4个以内。

前3条准则如式(2)～式(4)所示：

$$\sum P_i = A \tag{2}$$

$$\frac{\sum\limits_{p_i \in P_i, p_i=0} p_i}{\sum\limits_{p_i \in P_i} p_i} \leqslant 0.5 \tag{3}$$

$$P_i \cap P_j = \varnothing \qquad i \neq j \tag{4}$$

式中：A表示一个类别的模型；P_i表示模型类别下的一个语义标签；p_i表示标签下一个具体的零件模型。

根据这4条准则，本文提出基于最大语义零件的分割方法：在自顶而下的分割中，如果分割生成的模型(零件)已经具有语义，就不再对这部分模型进行分割，直到所有零件还原成原模型。其中，自顶而下的分割是指从完整的三维模型开始分割得到语义零件，零件具有语义是指这部分模型独立承担某项功能或被熟知并认可的零件。

本文基于提出的模型分割方法，事先为每类三维模型划分语义零件(以椅子为例，本文将椅子划分为靠背、基座、扶手、坐垫4个部分)，再将PartNet分割生成的语义零件重定向到本文分割的语义零件下并进行组合，在组合过程中，首先，算法读取PartNet生成的零件树文件，并根据零件树中的零件从属关系，生成本文确定的零件标签下包含的模型文件，然后，通过3.3节介绍的三维模型拼接算法将PartNet的分割结果组合成按本文方法分割生成的语义零件。

按照2种语义分割方法生成模型的可视化结果对比如图5所示，其中，左边模型按PartNet提出的方法进行分割，右边模型按照本文方法对模型进行分割，图中模型上每种颜色代表分割生成的一个语义零件。按照2种语义分割方法生成模型的平均零件数量对比如表1所示，表中较小的平均零件数量已加黑。



(a) PartNet分割方式 (b) 本文方法分割方式

图5 模型分割可视化(以椅子为例)

Fig. 5 Visualization of semantic segmentation (taking chairs as an example)

表1 语义分割后模型平均零件数量

Table 1 Average number of parts in the model after segmentation

模型类别	PartNet	本文方法
椅子	18.33	**3.36**
桌子	5.24	**2**
台灯	11.35	**2**
水槽	13.02	**2.66**
储物柜	9.51	**2**

以上数据表明本文方法大大减少了每个模型的组成零件数量，相比于利用PartNet中的分割方法，本文方法在三维重建方面具有以下应用优势：

1）生成的模型在具有语义信息的同时，使得重建网络尽可能简化。在隐式重建中，从特征向量恢复三维模型的工作由解码器承担，所以增加一个标签就需要对应增加一个解码器。PartNet将数据集中模型分解成3～230个不等的零件，直接套用PartNet的标签需要使用230个解码器，即使利用对称

关系进行部分解码器复用, 网络结构仍然复杂。而利用本文方法只需要 4 个隐式解码器, 大大简化了重建网络。

2) 增加了每个解码器的数据量, 缩短了训练时间。本文选用椅子、水槽、台灯、储物柜和桌子 5 种类别模型来进行重建工作。某些模型类别下还包含很多子类, 如椅子还可以细分成扶手椅、沙发、座椅等。每个子类别下的同一零件在结构上也有细微差别, 如有些扶手椅的腿部呈辐射对称, 有些椅子的扶手可以进一步拆解等。PartNet 在分割零件时并没有忽略这些细微的差别, 因为它是一个分割精度较高的数据集, 导致部分模型分割后零件数量升高。然而对模型进行过度分解会导致某些标签下的零件总数较少。在重建工作中, 这一标签下的可用数据也就相应较少, 进而影响重建精度。而不管有多少有用数据, 模型总数是不变的, 所以每个解码器的训练时间也是相同的(缺少零件的模型会用空白文件当作数据补齐, 因此, 所有解码器的训练数据总量一致)。因此, 基于最大语义零件的分割方法在提高了解码器的数据量的同时, 还缩短了整体网络训练所需的时间。

2.3　语义零件拼接

PartNet 数据集中的三维模型为 obj 格式, obj 文件内包含三维模型的顶点信息和面片信息, 其中, 顶点信息是每个顶点的坐标位置, 面片信息是组成三角形面片的顶点编号(以顶点信息在文件中的相对位置表征)。在一个 obj 文件中添加另一个 obj 文件中的顶点和面片信息即可实现 2 个三维模型的拼接。由于新添加的顶点信息在 obj 文件中的相对位置发生了变化, 所以在 obj 文件中添加面片信息时, 需要根据原文件中顶点个数在新加入的面片信息中对顶点编号进行更改。三维模型的拼接流程如图 6 所示, 图中, 偏移量指一个模型文件中的顶点信息数量, 在程序开始时偏移量初始化为 0, 每个模型读取完成后, 偏移量将增加这个模型的顶点数量。

在以往的隐式场重建算法中, 通常会将训练数据的三维模型进行中心化, 因为输入图片中所有模型都位于图片的中心, 如果训练数据中模型的位置是随机变化的, 网络无法从输入图片中学习到模型的实际位置, 导致网络损失上升。本文在对模型进行预处理时, 仅对整体模型进行中心化, 虽然模型在分解后各零件并不处于三维空间的中心, 但是每个零件在空间中的相对位置也因此得以保留, 这样可以直接利用该三维模型拼接算法对重建结果进行后处理, 省略对各零件进行平移的步骤。

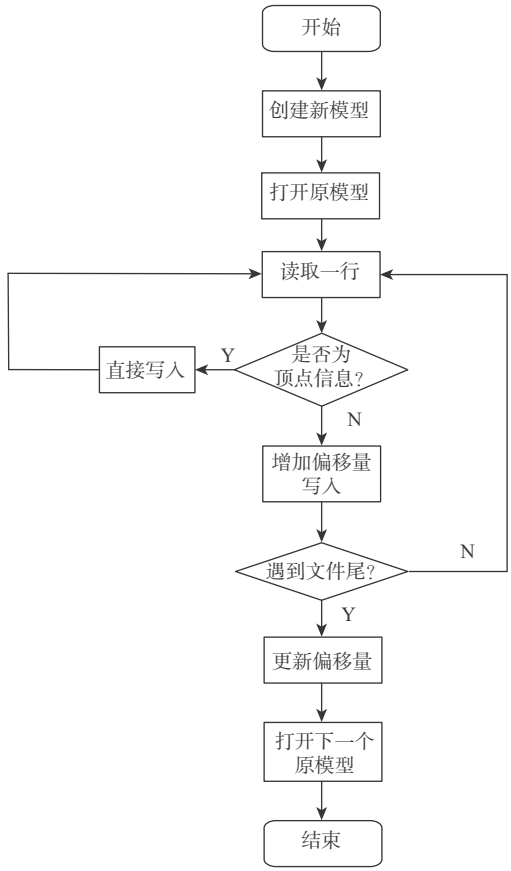


图 6　三维模型拼接算法流程

Fig. 6　3D model stitching algorithm process

2.4　重建网络结构

本文在网络设计上沿用了多数隐式场重建算法的图片编码器和隐式场解码器结构, 并根据训练数据的复杂程度对网络中各模块的隐藏层深度、参数大小及激活函数等设置进行了相应调整。相比于整体模型, 模型零件的几何结构会相对简单, 因此, 本文对重建网络的编码器和解码器结构都进行了简化以减少网络参数及训练时间。

本文选用残差网络作为图片编码器, 残差网络通过引入恒等映射, 在网络层数较深的情况下也能得到较好的收敛效果, 解决了梯度消失和梯度爆炸问题。由于本文输入图像较为简单, 用 7 个残差单元组成的图像编码器对输入图片进行特征提取, 减小了网络训练难度, 本文使用的残差单元如图 7 所示, 其中, 卷积核尺寸均为 3×3。

在 Chen 和 Zhang[20] 提出隐式场重建算法后, 多数算法均沿用其隐式解码器, 该解码器利用特征拼接算法在每个隐藏层输出之后以消除梯度消失现象, 并增加 MLP 获得的信息量。本文解码器仅用于重建模型零件, 训练难度较小, 因此, 本文对 Chen 和 Zhang[20] 提出的隐式解码器进行了调整, 将输出层的输出函数由 Leaky-ReLU 函数改为 Sigmoid 函数, 以此解决梯度消失问题。此外, 本文还取消了

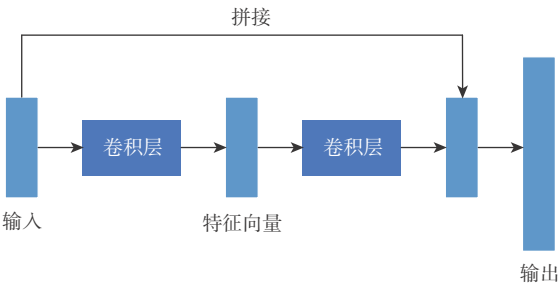


图7 残差单元结构

Fig. 7 Architecture of residual unit

隐藏层之间的拼接操作，并减少了2层线性隐藏层，得到较为轻量的隐式解码器。本文使用的解码器如图3所示。整体网络结构如图1所示，网络采用137×137分辨率图片作为模型输入，经过残差编码器后得到128维特征向量，隐式场解码器将该特征作为输入恢复模型语义零件，并通过拼接算法生成重建结果，整体重建过程如下：

$$\boldsymbol{z}=E(\boldsymbol{I})\oplus\boldsymbol{p}\qquad \boldsymbol{p}\in\boldsymbol{P} \tag{5}$$

$$i'=D(\boldsymbol{z}) \tag{6}$$

$$|\boldsymbol{P}'|=|\boldsymbol{P}[\boldsymbol{p}]|\leftarrow i' \tag{7}$$

$$a_l=M(\boldsymbol{P}')\qquad l\in L \tag{8}$$

$$A=\sum a_l\qquad l\in L \tag{9}$$

式中：$\boldsymbol{z}$表示输入隐式场解码器的特征向量；$E(\cdot)$表示编码器；$\boldsymbol{I}$表示输入图像；$\boldsymbol{p}$表示隐式场内任意坐标；$\boldsymbol{P}$表示初始隐式场；$\oplus$表示向量拼接操作；i'表示隐式场内某坐标更新后的隐式场函数值；$\boldsymbol{P}'$表示场值更新后的隐式场；$D(\cdot)$表示隐式场解码器；$M(\cdot)$表示对隐式场进行等值面提取；a_l表示三维模型的某一零件；L表示三维模型包含的所有零件标签；A表示组合后的三维模型。

3 实验及结果分析

本文利用PartNet中数据的80%对网络进行训练，并在PartNet的剩余数据及ShapeNet数据集上进行重建性能指标测试。

3.1 网络训练工作

由隐式场函数的表达式，将解码器拟合到隐式场函数，需要采用点值对作为数据进行训练。只有先对PartNet数据集中的obj文件进行体素化，才能从离散的体素模型中得到坐标——隐式场函数值的点值对。本文通过3.2节和3.3节提出的模型分割方法及零件拼接算法对PartNet提供的模型预处理生成供本文训练的语义零件。再将预处理后的语义零件进行体素化得到体素模型，并通过均匀采样的方式从离散的体素模型中得到坐标——隐式场函数值的点值对。由于卷积网络不能对缩放或平移旋转保持不变，因此，对三维模型采样时需要将模型限制在边长为1的正方体框内，再对模型进行采样，并保证模型的上端朝y轴正向。

在PartNet数据集中，部分模型并不包含类别下的所有语义零件，如某些椅子缺少扶手、某些储物柜没有门等，缺少的零件会用空体素替代，空体素在网络中表现为全零矩阵。

由于不同类型零件的几何结构存在差异，对于细长结构较多的零件，如椅子腿、水龙头等，本文在训练解码器时采用较小学习率以保证网络能较好地收敛。

本实验采取先训练隐式场解码器，再训练编码器的训练方式。在隐式场解码器训练过程中，先由三维卷积网络构成的编码器从体素模型提取128维特征向量，再利用隐式场解码器恢复出体素模型。隐式场解码器使用均方误差函数作为网络损失函数，表达式为

$$L_{\text{oss}}=\frac{\sum_S|F(i)-G(i)|^2}{N} \tag{10}$$

隐式场重建网络利用多种分辨率的训练数据对训练过程进行加速。在获得点值对的过程中，本文还对高分辨率的体素模型进行了下采样得到16^3、32^3、64^3这3种低分辨率的体素模型，在训练过程中，网络先利用16^3分辨率模型进行训练，在完成16^3分辨率体素重建任务的训练工作后，再对32^3分辨率体素重建任务进行训练，最后，对64^3分辨率的体素进行训练。该训练策略可以使损失函数收敛的速度提升到直接训练高分辨率体素重建网络的3倍。

在图像编码器的训练过程中，本文借鉴三维重建网络AtlasNet的训练方法，利用隐式场解码器训练阶段得到的三维编码器对体素模型进行特征提取，并利用这些特征向量作为图像编码器训练的真值。图像编码器也利用均方误差函数作为损失函数。

本文使用Adam优化器，学习率取0.000 05，Batch size取64。在16^3分辨率、32^3分辨率的解码器训练中，Epoch取200，在64^3分辨率的解码器训练中，Epoch取400，图像编码器训练Epoch取1 000。在图像编码器的训练过程中，每隔100个Epoch就将学习率衰减为原来的0.9倍，这样可以让网络在损失函数较小时仍然能进行收敛，提升网络训练效果。

在网络对输入图片重建得到各个分立零件后，

还需要对得到的零件进行后处理才能得到最终结果。由于本实验在对原数据进行预处理时没有改变顶点间的相对位置, 所以应用 3.3 节提出的模型拼接算法即可将各零件组合成整体模型。

3.2　重建性能指标评价和结果对比

本文选取了重建结果较好的单视图隐式场重建算法 IM-Net 及体素占用预测网络 (occupancy network, OCC-Net) 在 ShapeNet 和 PartNet 两大国际通用三维模型数据集上进行了重建精度定量比较。本实验主要通过 2 个指标评价重建性能, 分别是: ①交并比(intersection over union, IoU), 交并比是 2 个模型分布的交集区域和并集区域之比, 交并比越大, 说明 2 个模型越重合, 重建效果也就越好; ②钱伯距离 (Chamfer distance, CD), 钱伯距离表示重建模型点集和真值点集的平均最短点距离, 钱伯距离越小表示重建三维模型与真值三维模型之间差异越小, 即重建模型的效果更好。钱伯距离的计算式为

$$C_1 = \sum_{x_1 \in X_P} \min_{x_2 \in X_T} x_1 - x_2 \tag{11}$$

$$C_2 = \sum_{x_2 \in X_T} \min_{x_1 \in X_P} x_2 - x_1 \tag{12}$$

$$C_D = C_1 + C_2 \tag{13}$$

式中: X_P表示从重建模型表面采样的点集; X_T表示从真值模型表面采样的点集。

3.2.1　重建性能指标评价

本文在交并比的评价实验中, 首先, 对重建的模型进行体素化采样得到体素模型, 然后, 对体素模型比较交并比。

在钱伯距离评价实验中, 首先, 在重建模型和真值模型表面分别采样 4 096 个点作为计算钱伯距离的依据, 并存储为张量形式。实现钱伯距离的大致方法是在重建点集中遍历与真值点集的每个坐标的点距, 并将最小点距平方累加得到C_1, 再对真值进行相同的操作得到C_2。

表 2 为 3 个重建方法得到的 16^3 分辨率体素模型在 PartNet 数据集上的交并比结果, 表中较好的重建指标数值已加黑。表 3 为 3 个重建方法得到的模型在 PartNet 数据集上的钱伯距离结果, 表中较好的重建指标数值已加黑。其中, IM-Net 和 OCC-Net 利用 PartNet 中所有零件拼接得到的整体模型作为真值进行训练, 本文方法利用 PartNet 中所有零件拼接得到本文划分的 4 个零件标签下的模型作为真值进行训练。

由表 2 和表 3 可知, 在交并比和钱伯距离指标上, 本文方法比较 IM-Net、OCC-Net 均有较明显提升。这说明相比于 IM-Net 和 OCC-Net 等直接重建整体模型的隐式场重建算法, 本文提出的基于语义的多分支隐式场重建网络先重建各个语义零件, 再拼接得到整体模型的方法有效地让网络针对各个零件投入更多注意, 减小了重建误差。

表 2　重建交并比结果

Table 2　Result of IoU

模型类别	IM-Net	本文方法	OCC-Net
椅子	41.57	**46.56**	42.05
桌子	58.71	**62.37**	59.15
台灯	38.33	**39.12**	39.10
水槽	37.17	**39.15**	34.98
储物柜	46.16	**48.23**	45.88

表 3　重建钱伯距离结果

Table 3　Result of rebuilding CD

模型类别	IM-Net	本文方法	OCC-Net
椅子	0.973	**0.968**	0.981
桌子	0.835	**0.788**	0.812
台灯	1.150	**1.095**	1.152
水槽	1.205	**1.135**	1.213
储物柜	1.011	**0.983**	1.001

本文将 3 种重建方法的重建时间及显存开销进行了比较, 如表 4 所示, 表中较好的重建指标值已加黑。由于本文方法对网络参数进行了轻量化, 因此, 每个网络模块的运行时间以及显存都有所降低(由于网络中所有隐式场解码器并非同时运行, 显存占用相比其他网络有所减少), 但由于网络模块数量较多, 整体重建时间相比其他算法略有提高。

表 4　重建时间及显存消耗

Table 4　Rebuild time and memory comparison

算法名称	重建时间/s	显存开销/GB
本文方法	10.3	**2.9**
IM-Net	**5.6**	3.2
OCC-Net	6.2	3.5

3.2.2　重建结果可视化分析

重建结果可视化分析如图 8 所示。图 8(a)为实际模型的渲染图, 也是网络输入, 图 8(b)为本文方法重建得到的模型, 图 8(c)为 IM-Net 算法重建恢复得到的模型, 图 8(d)为 OCC-Net 算法重建得到的模型。本节挑选具有代表性的几个模型: 具有细杆结构的椅子、具有辐射对称基座的桌子、具有镂空结构的台灯、具有细长水龙头的水槽、具有薄板结构的储物柜, 这些几何特征在隐式场函数中会表现为正负符号快速变化的曲线, 可视化结果表

明，相比于其他隐式场重建算法，本文方法能够更好地拟合这种曲线特征，进而重建出具有更多几何特征的模型。



(a) 真值 (b) 本文方法 (c) IM-Net (d) OCC-Net

图 8 重建结果可视化分析

Fig. 8 Visualization and analysis of reconstruction result

本文方法解决了其他算法利用隐式场重建网络直接重建整体模型在以下几个方面存在的缺陷。

1）重建结果的局部特征恢复较好。相比于显式重建算法，隐式场重建算法可以重建出任意分辨率的、平滑的三维模型。但隐式场重建算法需要对体素模型采样获得点值对作为数据对解码器进行训练，而在对模型的局部特征采样过程中得到的采样点往往很少，这使得模型局部特征对网络收敛过程影响较小，所以传统隐式场重建算法重建结果缺少局部特征。本文方法在采样过程中直接对零件采样，提高了模型局部特征的采样点数，因此，网络可以较好地拟合变化速率较快的隐式场函数。从图 8 可以看出，本文方法的局部特征恢复能力明显优于其他 2 种算法，说明本文方法较好地恢复了模型的局部特征。

2）进一步提高了重建模型的平滑度。隐式场重建算法的重建结果相比于显式重建算法平滑，由于本文方法对训练数据进行了零件层面的采样，可以生成相比于其他隐式场重建算法更加平滑的重建结果。图 9 为本文方法重建生成的部分语义零件及完整模型，并将其与 IM-Net 或 OCC-Net(统称为其他算法)进行了对比。从图 9 看出，其他算法的重建结果在局部特征较复杂时，往往会得到不完整、较多凸起的重建结果，而本文方法得到的重建结果具有完整、平滑的复杂局部特征。



(a) 真值 (b) 语义零件 (c) 重建结果 (d) 其他算法

图 9 语义零件重建结果可视化

Fig. 9 Visualization of semantic part reconstruction results

3）缓解了单视图重建受到输入图片视角的影响。由于二维视图中存在被遮挡的部位，其他重建算法生成的结果存在模型不完整的情况，而本文方法通过零件层面的训练较好地保证了重建结果的完整性。图 9 中，其他隐式场重建算法得到的重建结果均由于输入图片的视角原因，存在不同程度的缺失、断裂、凸起；而本文方法成功地恢复了模型，并且在具有一定重建难度的三维模型中也能较好地恢复模型细节。

4 结 论

本文针对隐式场重建算法存在的受到二维图片视角影响较大，重建模型缺少语义信息，并且零件的局部细节存在损失等问题，提出基于语义关系的单视图隐式场三维重建技术，希望提升隐式场重建算法的重建精度和重建能力，本文工作可以总结如下：

1）提出一种基于最大语义的模型分割方法并基于大型三维模型数据集 PartNet 中的 obj 文件设计了零件拼接算法，在经过算法的预处理后，PartNet 中原本最大零件数高达 230 的模型被合成为只有 2～4 个零件的模型，大大减小了模型下的平均零件数。

2）基于最大语义的模型分割方法提出一种多分支隐式场重建网络，并根据训练数据复杂程度对网络结构进行轻量化。

相比于以往的隐式场重建网络，本文网络生成的模型在细节的完整性和模型的平滑程度都更好，并且网络结构更为简单。在一定程度上，证明了本文模型的语义信息可以帮助隐式场重建网络学习

三维模型的局部特征, 有效地解决了隐式场重建网络只关注整体结构, 不能较好地恢复局部细节的现象。

参考文献（References）

[1] LI M Y, ZHANG H. D2IM-Net: Learning detail disentangled implicit fields from single images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 10241-10250.

[2] PARK J J, FLORENCE P, STRAUB J, et al. DeepSDF: learning continuous signed distance functions for shape representation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 165-174.

[3] HORN B K P. Shape from shading: a method for obtaining the shape of a smooth opaque object from one view[M]. America: Project on Mathematics and Computation, 1970: 4-5.

[4] 包晓敏, 白晨. 单目三维重建技术综述[J]. 智能计算机与应用, 2020, 10(7): 49-52.
BAO X M, BAI C. Overview of monocular 3D reconstruction techniques[J]. Intelligent Computer and Applications, 2020, 10(7): 49-52 (in Chinese).

[5] GENG J. Structured-light 3D surface imaging: a tutorial[J]. Advances in Optics and Photonics, 2011, 3(2): 128-160.

[6] 陈荣, 许宏丽, 杨东学, 等. 一种基于空间编码结构光的稠密三维重建算法[J]. 西安电子科技大学学报, 2021, 48(6): 123-130.
CHEN R, XU H L, YANG D X, et al. A dense 3D reconstruction algorithm based on spatially encoded structured light[J]. Journal of Xidian University, 2021, 48(6): 123-130(in Chinese).

[7] SNAVELY N, SEITZ S M, SZELISKI R. Photo tourism: exploring photo collections in 3D[J]. ACM Transactions on Graphics, 2006, 25(3): 835-846.

[8] LOWE D G. Object recognition from local scale-invariant features[C]//Proceedings of the Seventh IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2002: 1150-1157.

[9] CHOY C B, XU D F, GWAK J, et al. 3D-R2N2: a unified approach for single and multi-view 3D object reconstruction[M]//Lecture Notes in Computer Science. Berlin: Springer International Publishing, 2016: 628-644.

[10] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2023-01-01]. https://arxiv.org/abs/1409.1556.

[11] WU J, ZHANG C, XUE T, et al. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: MIT Press, 2016: 82-90.

[12] XIE H Z, YAO H X, SUN X S, et al. Pix2Vox: context-aware 3D reconstruction from single and multi-view images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 2690-2698.

[13] WU J, WANG Y, XUE T, et al. Marrnet: 3D shape reconstruction via 2.5 d sketches[C]//Proceedings of the International Conference on Neural Information Processing Systems. New York: MIT Press, 2017: 171-190.

[14] WANG H Q, YANG J L, LIANG W, et al. Deep single-view 3D object reconstruction with visual hull embedding[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 8941-8948.

[15] PONTES J K, KONG C, SRIDHARAN S, et al. Image2Mesh: a learning framework for single image 3D reconstruction[M]//Lecture Notes in Computer Science. Berlin: Springer International Publishing, 2019: 365-381.

[16] JACK D, PONTES J K, SRIDHARAN S, et al. Learning free-form deformations for 3D object reconstruction[M]//Lecture Notes in Computer Science. Berlin: Springer International Publishing, 2019: 317-333.

[17] LI K J, PHAM T, ZHAN H Y, et al. Efficient dense point cloud object reconstruction using deformation vector fields[M]//Lecture Notes in Computer Science. Berlin: Springer International Publishing, 2018: 508-524.

[18] KONG C, LIN C H, LUCEY S. Using locally corresponding CAD models for dense 3D reconstructions from a single image[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 5603-5611.

[19] GROUEIX T, FISHER M, KIM V G, et al. A papier-Mache approach to learning 3D surface generation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 216-224.

[20] CHEN Z Q, ZHANG H. Learning implicit fields for generative shape modeling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 5932-5941.

[21] MESCHEDER L, OECHSLE M, NIEMEYER M, et al. Occupancy networks: learning 3D reconstruction in function space[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4455-4465.

[22] WANG T C, LIU M Y, ZHU J Y, et al. High-resolution image synthesis and semantic manipulation with conditional GANs[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8798-8807.

[23] BECHTOLD J, TATARCHENKO M, FISCHER V, et al. Fostering generalization in single-view 3D reconstruction by learning a hierarchy of local and global shape priors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 15875-15884.

[24] MO K C, ZHU S L, CHANG A X, et al. PartNet: a large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 909-918.

Semantic part based single-view implicit field for 3D shape reconstruction technology

XU Gang[1,2], JIN Jiongchao[1], TANG Zehao[1], LENG Biao[1,*]

(1. School of Computer Science and Engineering, Beihang University, Beijing 100191, China;
2. Ping An Bank Co., Ltd., Shenzhen 518001, China)

Abstract: With the development of deep learning techniques, learning implicit field for 3D shape reconstruction has become a heated topic, because implicit field can help networks learn a reasonable and sophisticated reconstruction model than explicit methods. However, there are still some challenges to be solved including lacking semantic information, local detail incompletions and so on. Thus, rather than recreating the entire model from a decoder directly, we first reconstruct the semantic components of a single model using an implicit filed structure based on semantic sections in our paper. Then we aggregate the reconstructed semantic parts together to get the final model. Finally, we test those results on the public 3D shape dataset PartNet and compare them to other cutting-edge single-view reconstruction approaches. It's obvious that using a semantic part-based implicit field can learn more reasonable shape representations for reconstruction.

Keywords: 3D shape reconstruction; semantic segmentation; implicit field; single view reconstruction; multi-branch network

Received: 2023-02-28; **Accepted**: 2023-05-08; **Published Online**: 2023-07-10 11:05

URL: link.cnki.net/urlid/11.2625.V.20230707.2037.001

Foundation item: National Natural Science Foundation of China (61972014)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0125

矩形微织构浮环气膜密封特性的方向效应

王世鹏[1]，丁雪兴[1,*]，力宁[2]，丁俊华[1]，张兰霞[1]

(1. 兰州理工大学 石油化工学院，兰州 730050；2. 中国航发湖南动力机械研究所，株洲 412002)

摘 要：为研究微织构的方向效应对浮环气膜密封性能参数的影响，选取矩形微织构孔为研究对象，基于气体润滑理论，建立该织构化流体动压润滑控制的理论分析模型并采用有限差分法对其进行求解，获得了密封间隙气膜的压力分布，以转速、压力、偏心率、织构深度为自变量，重点研究织构方向角对其密封性能参数影响的规律。结果表明：开设微织构的浮环密封浮升力随着转速、压力和偏心率的增大均可提高，但微织构孔深度的增大会导致浮升力下降。泄漏率随着转速、压力、偏心率和织构深度的增加而提高；但转速的升高对其泄漏率数值影响不大。气膜摩擦力随转速、压力和偏心率的变大而升高，随织构深度的增大而降低。矩形微织构在改善密封性能参数方面起着重要的作用，合理的选择微织构方向角，可有效地提高浮环气膜密封的密封性能。研究结果为提高织构化表面浮环密封性能设计提供了新思路。

关 键 词：浮环密封；矩形微织构；方向效应；密封性能；数值计算

中图分类号：TH117.2

文献标志码：A **文章编号**：1001-5965(2025)03-0845-12

浮环气膜密封是一种非接触式流体动密封，其优良的可靠性和高效的密封性对旋转机械设备的安全服役具有重要的意义[1-2]。随着旋转机械面临的工作环境高参数化，对密封系统的稳定性和使用寿命提出了更高的要求。表面微织构技术作为一种新兴的可实现润滑减摩的有效手段，通过改变流体动密封系统中轴套表面微观形貌，在改善润滑表面的流体动压效应和避免表面黏附等方面已取得了显著的效果，现已成为先进密封技术领域的研究热点[3-5]。

随着表面微造型技术的发展，表面织构已被证实是改善流体表面润滑和降低摩擦性能的有效手段。Singh 和 Awasthi[6] 借助有限元法分析了圆柱织构对动压滑动轴承系统静、动态性能的影响，发现合理地布置凹槽位置和参数，可显著提高轴承的承载能力。Shi 等[7-8] 考虑织构化表面的相邻织构的集体效应，考察了微沟槽和微凹坑对机械密封性能的影响，通过引入性能评价参数，获得直边纹理形状的性能优于圆边纹理形状。Yang 等[9] 研究了三角形凹坑对机械密封的影响，发现具有等腰直角三角形凹坑的表面纹理有利于密封面开启行为和泄漏限制。Adjemout 等[10-11] 使用的数值模型与质量守恒模型耦合求解雷诺方程，分析了织构的主要几何特征对机械密封流体动力润滑的影响，发现对称放置的三角形织构对于提高机械密封的流体动力性能更加显著。

陈传刚等[12] 运用静环表面开设微孔同时在动环表面开设微槽的微造型复合表面技术，获取了微孔密度与直径对气体密封性能的影响。程香平等[13-15] 采用有限元分析的方法研究菱形孔结构，发现在低速时由菱形微孔所引起的动压效应使得密封端面产生周向和径向的波状变形，此变形可降低密封端面对平整度和光洁度的要求，弥补了动压效应不足的缺陷。魏伟等[16-17] 针对单一的槽型和单一的微孔织构，提出一种新的组合织构类型。通过分析不同织构化端面的流场特性，获取了最佳密封综合性

收稿日期：2023-03-14；**录用日期**：2023-07-08；**网络出版时间**：2023-07-11 08:49
网络出版地址：link.cnki.net/urlid/11.2625.V.20230710.1709.004
基金项目：国家自然科学基金 (51565029)；宁波重大专项 (2020Z112)
* **通信作者**. E-mail：dingxxseal@126.com

引用格式：王世鹏，丁雪兴，力宁，等. 矩形微织构浮环气膜密封特性的方向效应 [J]. 北京航空航天大学学报，2025，51（3）：845-856.
WANG S P，DING X X，LI N，et al. Orientation effect on sealing characteristics of rectangular micro-textured floating ring gas film [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（3）：845-856（in Chinese）.

能的复合微孔织构化密封结构参数。吉华等[18]以开设了等边三角形微孔的端面机械密封为研究对象，结合流体楔效应理论，揭示了等边三角形微孔几何特性的改变对密封性能参数的影响。国内外对微织构在工程中的应用研究取得了一定的成果，但仍存在不足，例如：大多关于微孔织构的相关研究主要集中在气浮轴承和端面干气密封领域，对浮环气膜密封而言少有触及；研究成果中多注重织构形状，力图通过构建一些新型织构类型来改善其密封性能；关于微织构孔的方向效应鲜有报道。

综上所述，相较于槽型结构而言，织构表面可以更好地控制气膜润滑的流动，从而实现更可靠的润滑效果。考虑到方向性微织构的导流作用明显，为进一步探究矩形织构的方向角对浮环气膜密封性能的影响，本文基于气体润滑理论，以浮环密封为研究对象，建立表面织构化浮环密封润滑理论模型，以气膜浮升力、泄漏率、摩擦力和浮漏比为研究目标，分析了不同转速、压力、偏心率和织构深度等条件下对研究目标的影响趋势，重点研究不同织构方向角下矩形微织构孔对改善浮环密封性能参数的影响，以充分发掘微织构对提高浮环密封综合性能的潜力，为流体动密封的表面织构设计提供补充。

1 理论模型

1.1 工作原理

在轴套外表面开设微织构孔是增强浮环气膜密封动压效应的一种有效方式。当密封设备正常工作时，轴套随转轴高速偏转，高压侧的气体遵从泵吸效应带入织构孔内，介质气体在微孔边缘处不断被压缩，其周围区域产生明显的动压效应，使动环与静环之间形成一层厚度较薄，刚度较大的气膜，从而达到对主泄漏通道进行密封的效果[19]。同时，轴套与浮环之间采取偏心安装，形成气膜厚度不一的收敛型楔形间隙，进一步增强了密封装置气膜的刚度，增加了浮环气膜密封装置的密封效果[20-21]。

织构化浮环密封系统简化几何模型如图1所示。图中，O_1为转轴圆心，O_2为浮环圆心，e为偏心距，h_{min}为气膜最薄处，h_{max}为气膜最大处，P_{max}为最大气膜压力。

1.2 理论模型

基于流体动力学基本理论，对浮环气膜密封系统间的流场分析基于下列基本假设：

1）密封环表面光滑且为刚性，轴套表面与介质流体无相对滑移。

2）密封介质的黏度和密度在径向方向上不发生变化。

3）密封间隙内流体为牛顿型流体，流动状态为层流。

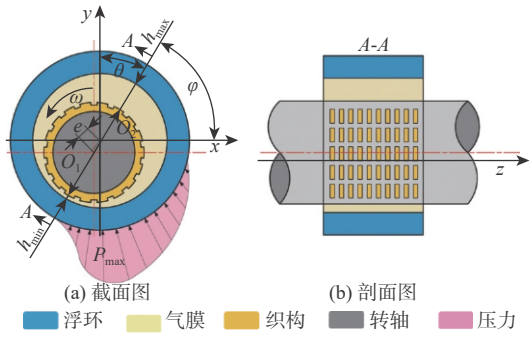


图1 织构化浮环密封简化几何模型

Fig. 1 Simplified geometric model of textured floating ring seal

4）同时忽略离心力、惯性力和振动对密封界面的影响。

在层流条件下，流体动力润滑的雷诺方程如下[22]：

$$\frac{1}{R^2}\cdot\frac{\partial}{\partial\theta}\left(\rho h^3\frac{\partial p}{\partial\theta}\right)+\frac{\partial}{\partial z}\left(\rho h^3\frac{\partial p}{\partial z}\right)=6\omega\mu\frac{\partial(\rho h)}{\partial\theta} \tag{1}$$

式中：R为轴套外半径，m；θ为圆周方向坐标，rad；p为气膜压力，MPa；h为气膜厚度，μm；μ为气体黏度，Pa·s；ρ为气体密度，kg/m^3；z为轴向坐标，m；ω为转动角速度，rad/s。

轴向介质黏度的变化在一定程度上会对气膜密封的性能产生影响，同一温度下的黏-压方程，通过NIST数据库查取介质压力范围为0～5.0 MPa的数据，经最小二乘法数据拟合获得多项式表达式，该例中拟合为5次多项式，黏度计算精度较高，即

$$\mu=f(p)=a_1+a_2p+a_3p^2+a_4p^3+a_5p^4+a_6p^5 \tag{2}$$

$$\begin{cases}a_1=1.839\times10^{-5}; & a_2=1.449\times10^{-13}\\ a_3=-9.445\times10^{-21}; & a_4=4.127\times10^{-26}\\ a_5=-4.247\times10^{-32}; & a_6=1.597\times10^{-38}\end{cases} \tag{3}$$

式中：多项式系数a_1、a_2、a_3、a_4、a_5、a_6为黏-压行为的系数。

浮环气膜密封微织构表面示意图如图2所示。其中，浮环气膜密封气膜厚度为

$$h_0=C(1+\varepsilon\cos\theta)+\Delta h(\theta,z) \tag{4}$$

矩形织构处气膜厚度为

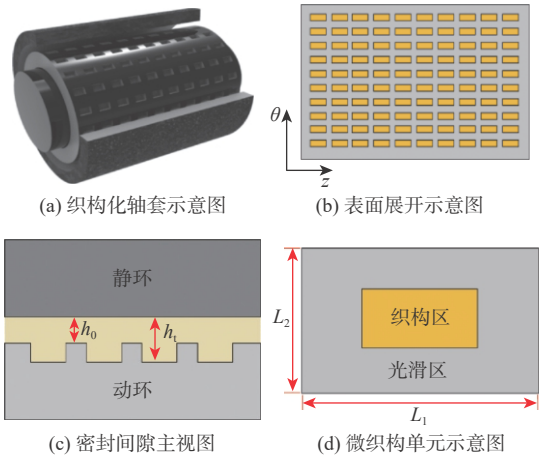
$$\Delta h(\theta,z)=\begin{cases}0 & (\theta,z)\in A_1\\ h_t & (\theta,z)\in A_2\end{cases} \tag{5}$$

式中：h_0为密封间隙，μm；h_t为矩形微织构深度，μm；A_1为无织构区域，m^2；A_2为有织构区域，m^2。

1.3 边界条件设定

密封系统的高压侧和低压侧的压力边界条件均采用强制边界条件，即

$$\begin{cases}z=0;p=P_i\\ z=1;p=P_o\end{cases} \tag{6}$$



(a) 织构化轴套示意图　(b) 表面展开示意图

(c) 密封间隙主视图　(d) 微织构单元示意图

图 2　浮环气膜密封微织构表面示意图

Fig. 2　Micro-textured surface of floating ring gas film seal

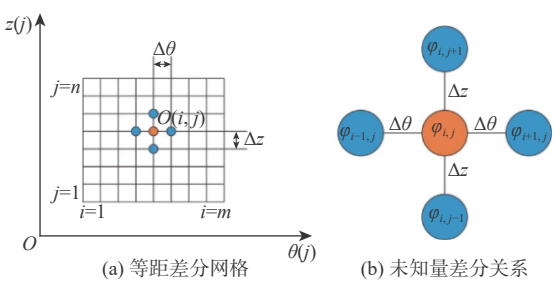
式中: P_i 为高压侧边界压力; P_o 为大气压力。

压力以 2π 为周期的周期性边界条件:

$$p(\theta,z)=p(\theta+2\pi,z) \tag{7}$$

1.4　网格划分

本文采用的求解方法为限差分法, 其求解的基本思想是将求解域分割成正交网格, 用差分近似代替微分, 使微分方程组变换成差分方程组, 从而对方程组进行求解[23]。一般而言, 为更加准确地获得特定压力的物理参数, 网格划分的疏密程度由计算精度和计算时间共同确定, 可在计算时间和计算精度之间得到较好的平衡[24-25]。图 3 为求解域网格划分示意图。Δz 为差分网格轴向方向步长, $\Delta z=L_z/(M-1)$; $\Delta\theta$ 为差分网格周向方向步长, $\Delta\theta=L_\theta/(M-1)$。



(a) 等距差分网格　(b) 未知量差分关系

图 3　计算域网格划分示意图

Fig. 3　Grid division in computational domain

1.5　求解精度的设定

一般而言, 绝对精度值因方程残差变化无法直接给出, 本文采用相对精度, 即利用 2 次迭代的压力载荷结果作为判断是否满足要求的准则, 通过控制迭代精度, 一般取 $[\alpha]=10^{-6}$。

$$\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|P_{i,j}^{(k+1)}-P_{i,j}^{(k)}\right|}{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|P_{i,j}^{(k+1)}\right|}\leqslant[\alpha] \tag{8}$$

式中: α 为相对误差限。

1.6　浮环密封稳态性能参数计算

通过对压力控制方程和膜厚控制方程的耦合求解, 获得浮环气膜密封的气膜压力分布, 进而计算出密封气膜特性参数。

浮环密封的气膜浮升力 F 的计算式为

$$\begin{cases}F_x=R\int_0^{\mathrm{L}}\int_0^{2\pi}P\cos\theta\mathrm{d}\theta\mathrm{d}z\\ F_y=R\int_0^{\mathrm{L}}\int_0^{2\pi}P\sin\theta\mathrm{d}\theta\mathrm{d}z\\ F=\left(F_x^2+F_y^2\right)^{1/2}\end{cases} \tag{9}$$

浮环密封的泄漏率 Q 的计算式为

$$Q=\int_0^{2\pi}\rho\left(-\frac{h^3}{12\mu}\cdot\frac{\partial p}{\partial z}\right)R\mathrm{d}\theta \tag{10}$$

浮环密封的气膜摩擦力 f 的计算式为

$$f=\int_0^{\mathrm{L}}\int_0^{2\pi}\left(-\frac{h}{2R}\cdot\frac{\partial p}{\partial\theta}+\frac{\mu\omega R}{h}\right)R\mathrm{d}\theta\mathrm{d}z \tag{11}$$

引入浮漏比 T_k, 用来表征浮环密封装置密封性能的优劣程度, 为织构化浮环气膜密封装置的工程应用选择提供一个标准。高浮漏比意味着该织构表面能够在保持有效润滑的同时减少润滑气膜的泄漏, 提高润滑效率和能源利用效率。浮漏比 T_k 的计算式为

$$T_k=\frac{F}{Q} \tag{12}$$

在整个求解过程中, 通过反复迭代压力, 即可得到收敛的气膜流场[26]。具体数值求解过程如图 4 所示。

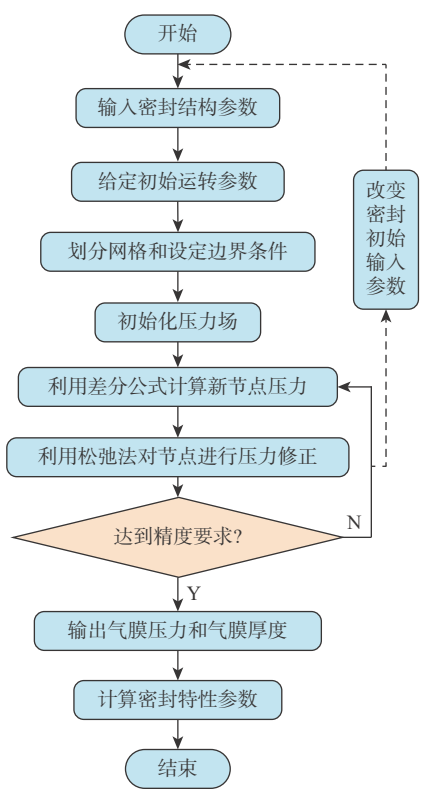


图 4　浮环气膜密封数值计算流程

Fig. 4　Flow of numerical calculation of floating ring gas seal

2 算例验证及结果分析

2.1 计算条件

基于气体动力润滑理论，为研究织构化表面的压力分布规律对浮环密封性能的影响。本文计算过程中采用如表 1 所示浮环密封的几何参数与工况参数，除被研究参数外，其他各参数均保持不变。

表 1 织构化浮环气密封的工况参数和几何参数

Table 1 Operating conditions and geometric parameters of textured floating ring gas seal

参数	数值
轴套外半径/mm	25
平均气膜厚度/μm	5
介质气体黏度/(Pa·s)	1.8×10^{-5}
偏心率	0.7
轴套长度/mm	52
出口压力/MPa	0.101 325
入口压力/MPa	0.7
转速/(r·min^{-1})	16 000
介质气体密度/(kg·m^{-3})	1.145 2
轴向织构间距/mm	3.36
周向织构间距/mm	2.9
织构长度/mm	12
织构宽度/mm	2
织构深度/μm	2
轴向织构个数	10
周向织构个数	10

2.2 网格无关性验证

网格尺寸的选取决定了密封性能的计算精确程度，所使用的网格应具有良好的质量，以便于精确模拟具有方向效应的织构在流场中的变化[27-28]。由于数值计算模型是基于织构化浮环密封微间隙内的气膜结构建立的，密封结构的特殊化导致气膜结构尺寸跨度极大，如微米级的织构深度和毫米级的轴向长度。选择织构方向角 β=0°的网格划分作网格无关性验证的研究对象。在保证边界层网格厚度不变的情况下以 50×50、100×100、100×150、150×100、150×150、150×200、200×150 和 200×200 共 8 种网格进行计算，以验证计算结果与网格的无关性，如图 5 所示。

由图 5 可知，随着网格数目增大，浮升力数值一直在增大，泄漏率的数值一直在减小。当网格数大于等于 150×150 时，气膜浮升力计算值变得稳定，而且网格数为 150×150 与网格数为 200×200 的数值差异小于 0.03%。对于泄漏率而言，当网格数为 150×150 时与 200×200 的数值差异小于 1.47%。

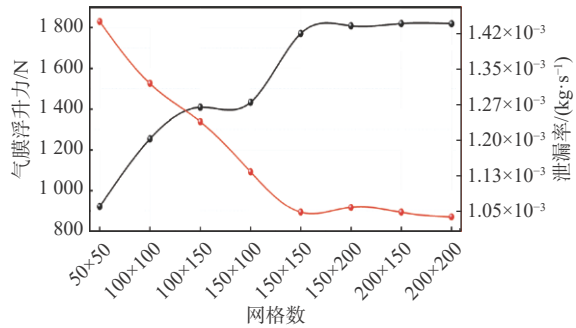


图 5 网格无关性验证

Fig. 5 Grid independence verification

当网格数增加时，计算时间显著增加，但结果基本无差别。在确保计算精度及尽量节省数值模拟计算所需时间的前提下，本文选用网格数为 150×150。

2.3 程序有效性验证

对于织构化浮环气膜密封而言，计算织构化浮环密封的偏位角随偏心率变化的规律，并将计算结果与文献 [29] 中的值进行比较，来验证所编写程序代码的有效性。引用文献 [29] 中的几何结构参数与工况参数进行计算，如表 2 所示。

根据表 2 中提供的参数，对本文所建立的数学模型的偏位角随偏心率变化的规律进行验证。如图 6 所示，随着浮环气膜密封的偏心率增加，偏位

表 2 文献 [29] 中工况参数和几何参数

Table 2 Operating conditions and geometric parameters in reference [29]

参数	数值
轴径半径/mm	17.5
织构深度/μm	2
转速/(r·min^{-1})	60 000
织构数量	22×11
轴承长度/mm	35
轴承半径间隙/μm	14
润滑气体黏度/(Pa.s)	1.932×10^{-5}
润滑气体密度/(kg·m^{-3})	1.161 4

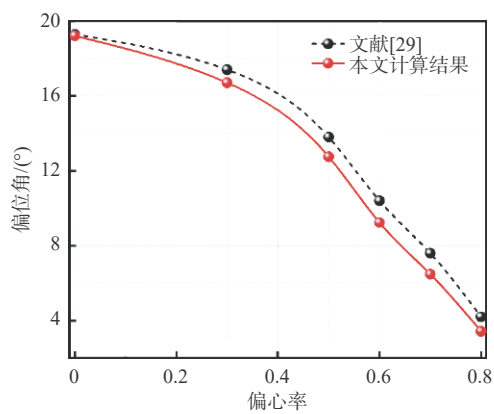


图 6 浮环气膜密封计算模型结果验证

Fig. 6 Verification of calculation model results of floating ring gas film seal

角的数值都在减小, 且文献 [29] 和数值计算的值差距逐渐增大, 但其数值差距均在 4% 内, 符合一般规律。综上所述, 本文计算结果与文献 [29] 具有较好的一致性, 其变化规律趋势基本相同, 因此, 所编写的程序代码是有效的。

3　结果与讨论

浮环气膜密封稳定运转后形成具有一定收敛度的润滑气膜, 此时矩形微织构加入后对浮环密封气膜的影响与平板摩擦副所产生的动压效应有所不同, 因此, 有必要对浮环气膜密封添加不同倾角微织构孔后的润滑密封性能进行研究。为方便研究过程中关于矩形织构旋转角度的区分, 规定矩形织构的长边平行于浮环进出口方向为转角度初始状态, 图 7 分别为织构方向角 β=0°、β=30°、β=45°、β=60°和 β=90°的矩形微织构气膜厚度和气膜压力的三维分布示意图。由图可知, 润滑气膜厚度收敛区域的正弦微沟槽尖角波峰处产生了峰值压力, 相邻微织构产生的动压润滑效应相互叠加且压力峰值沿轴向延伸, 气膜压力沿圆周方向周期性规律分布, 且收敛区域气膜压力增大更加明显, 沿轴向先增大后减小。产生这一结果的主要原因是轴套表面开设了矩形微织构孔, 介质气体流经开设微织构区域导致气膜厚度发生突变, 导致织构处出现气膜压力局部升高, 而沿发散间隙流动时不容易产生正压力, 即密封间隙发散的一侧不容易产生动压效应, 说明气膜厚度收敛区域的会聚形状有利于产生更高的局部气膜压力[30-31]。

3.1　转速对密封性能的影响

图 8 为密封性能参数随转速和织构方向角的变化曲线。不同织构方向角下气膜浮升力随转速的变化曲线如图 8(a) 所示, 随着转速的提高, 不同织构方向角下的矩形微织构气膜浮升力都显著提高。其中, 在相同转速下, 当 β=45°时, 气膜浮升力最大; 当 β=0°时, 气膜浮升力最小。这是由于介质气体进入密封间隙随开设矩形轮廓的微织构偏转时, 转速越高, 气膜分子的运动速度也越大, 但气膜分子运动受密封间隙的限制, 织构化表面所产生的动压效应越强, 导致气膜间隙内局部压力增大, 同时体现为气膜浮升力的增大。而且对规则排列的矩形微织构而言, 偏转一定的角度可提高气膜浮升力, 对气膜密封性能更加有益。

不同织构方向角下泄漏率随转速的变化曲线如图 8(b) 所示, 不同织构方向角下的矩形微织构浮环气膜密封的泄漏率随转速的增大整体提高。其中, β=0°和 β=90°的矩形织构是随转速增加, 泄漏率数值变化幅度较小(可以忽略), 而对于 β=30°、β=45°和 β=60°时的泄漏率数值而言, 泄漏率随转速的增大分别上升了 11.79%、11.15% 和 12.27%。呈现这种趋势的原因是: ①随着转速的增加, 密封间隙内介质气体受到微织构孔壁的阻碍而汇集使得动压效应显著增加, 从而在气膜最薄处出现了局部高压的现象, 压力梯度将导致密封间隙内的泄漏加剧; ②微织构孔的泵吸效应可将具有泄漏趋势的介质气体重新泵送回气膜密封区域。故在二者的协同作用下, 浮环气膜密封的泄漏率数值虽然在不断增加, 但整体增幅较小。

不同织构方向角下气膜摩擦力随转速的变化曲线如图 8(c) 所示, 随着转速的不断提高, 气膜摩擦力增大。而且, 转速越大, 不同织构方向角下的矩形微织构之间的摩擦力数值差异也逐渐增大。当转速相同, β=45°时, 气膜摩擦力最大, β=90°时, 气膜摩擦力最小, 且两者的数值差异最大为 12.33%。其原因在于介质气体在轴套表面织构内产生动压效应, 形成一层极薄的气膜, 当开设微织构孔的轴套高速偏转时, 由于牛顿黏性剪切力和气体间的黏滞力的存在导致气膜摩擦力的产生, 转速越大, 摩擦力越大。而微织构具有减阻的作用, 且介质气体黏度通常较小, 故浮环气膜密封的气膜摩擦力数值整体较低。

不同织构方向角下浮漏比随转速的变化曲线如图 8(d) 所示, 由图可知, 转速增大, 不同织构方向角下的微织构的浮漏比也随之增大, 而且其数值差异明显。其中, 织构方向角 β=0°的浮漏比数值最大, 而且转速越高, 其数值的增大幅度越大。其他 4 种织构方向角下的浮漏比数值均随转速的增大而提高, 其增大的幅度逐渐减小, β=90°的数值最小。开设微槽可以提高动压效应, 而且改变一定的织构方向角可以改善浮环密封的密封性能。在考虑满足工况的条件下, 工程中推荐使用 β=0°的矩形微织构排列方式以提高浮环气膜密封的密封特性参数。

3.2　压力对密封性能的影响

图 9 为密封性能参数随压力和织构方向角的变化曲线。不同织构方向角下气膜浮升力随压力的变化曲线如图 9(a) 所示, 气膜入口压力增大浮升力也随之增大。其中, 织构方向角 β=0°和 β=90°的气膜浮升力相差不大, 数值最大差异在 3.99%。其余的织构方向角、气膜浮升力相仿, 其数值差异在 0.9% 之内。归其原因为: 随着气膜入口压力的上升, 在楔形微间隙内气膜的流体静压效果增强, 气膜压差变大导致密封间隙内介质气体的泵送效应越明显; 而且润滑气膜受到来自密封气体压力的挤

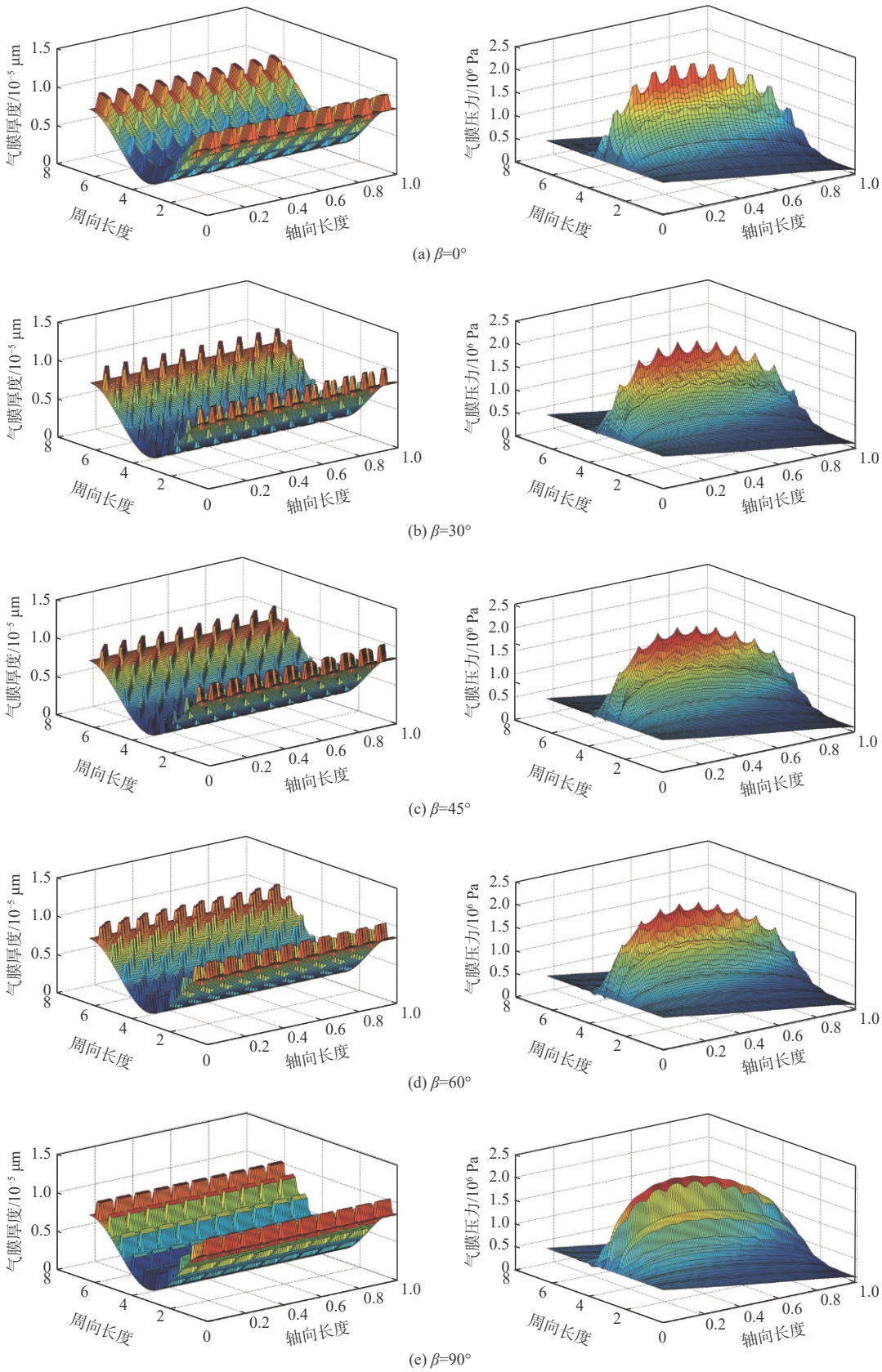


图 7 不同倾角矩形微织构的气膜厚度分布及其气膜压力分布示意图

Fig. 7 Thickness distribution and pressure distribution of rectangular micro-textured gas film with different angles

压，随着密封气膜压力的升高，气膜分子之间的距离越来越小，阻碍气膜厚度的降低，宏观下体现为气膜浮升力的持续增大。

不同织构方向角下泄漏率随压力的变化曲线如图 9(b) 所示，随着入口压力的提高不同织构方向角下微织构的泄漏率增大；其中，β=0°的泄漏率最

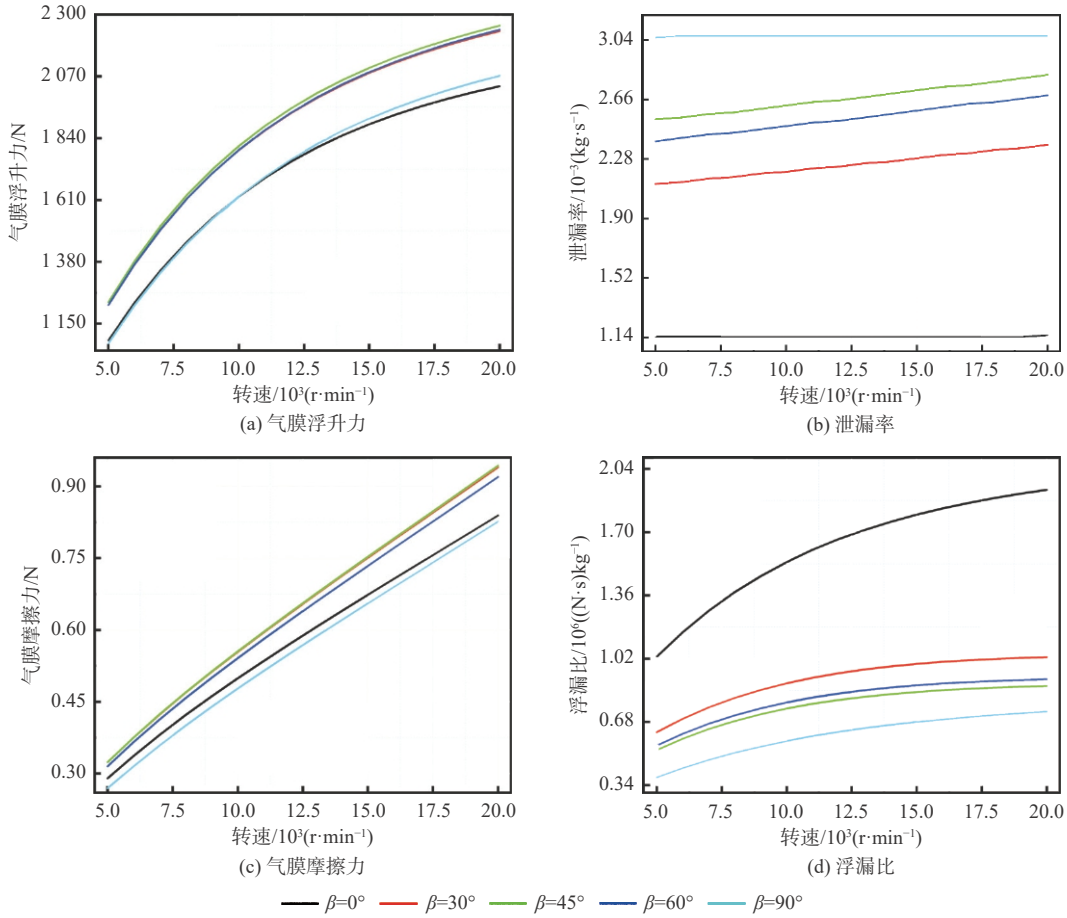


图 8　密封性能参数随转速和织构方向角度的变化曲线

Fig. 8　Variation curves of sealing performance parameters with rotational speed and texture direction angle

小, β=90°的泄漏率最大。当压力数值恒定时, 不同织构方向角的泄漏率数值差异随入口压力的提高而增大。说明动环偏转过程中, 当密封装置气源提供的气膜入口压力增大时, 高压侧和低压侧所产生的压差也逐渐增大, 静压效应增强, 致使密封副的浮升特性和气膜刚度增强。而转速保持恒定的情况下, 以密封压力形成的压力流占主导作用, 当压差增大时, 压力梯度的变大促使密封间隙内的介质气体流速增大, 加快了介质气体向空气侧流动, 进而导致泄漏率变大。

不同织构方向角下摩擦力随压力的变化曲线如图 9(c) 所示, 气膜摩擦力随压力的增大而增大, 摩擦力整体增大趋势从平缓变得陡峭。随着压力的提高, β=30°、β=45°和 β=60°下的气膜摩擦力数值差异基本保持不变, 而 β=0°和 β=90°下的气膜摩擦力数值差异逐渐增大, 整体数值变化较小。产生此现象的主要原因是在入口压力变大时, β=30°、β=45°和 β=60°所产生的动压效应相仿, 体现在气膜浮升力上数值差异较小, 而 β=0°和 β=90°所产生的动压效应相对较弱。因此, 随着密封压力的升高, 气体剪切力增大, 浮环气膜密封的密封副间的摩擦力增大。

不同织构方向角下浮漏比随压力的变化曲线如图 9(d) 所示, β=0°的浮漏比数值最大, 其他 4 种织构方向角下的浮漏比数值相仿。但从整体来看, 随着气膜入口压力的增大, 密封装置的浮漏比数值反而减小。故对于开设矩形微织构的浮环气膜密封而言, 通过对矩形微织构偏转一定的角度, 可以提高气膜动压效应。在工程应用选择时, 对频繁启停的机械设备而言, 应以浮升力最大为选择标准, 避免启、停阶段的碰磨, 推荐使用 β=45°; 当机械设备长期稳定运行时, 应以泄漏率最小为选择标准, 推荐使用 β=0°。

3.3　偏心率对密封性能的影响

图 10 为密封性能参数随偏心率和织构方向角的变化曲线。不同织构方向角下气膜浮升力随偏心率的变化曲线如图 10(a) 所示, 伴随偏心率的提高, 气膜浮升力的数值增长幅度不断增大。而且在相同的偏心率下, 不同织构方向角下的气膜浮升力相差不大, 都在 20.6% 以内。产生这种现象的主要原因是偏心率较低时, 浮环密封间隙内周向气膜厚度基本保持一致, 矩形微织构孔的泵吸效应无法发挥作用, 导致微织构的织构方向角变化对气膜浮升力没有显著的影响; 随着偏心率提高, 密封间隙内

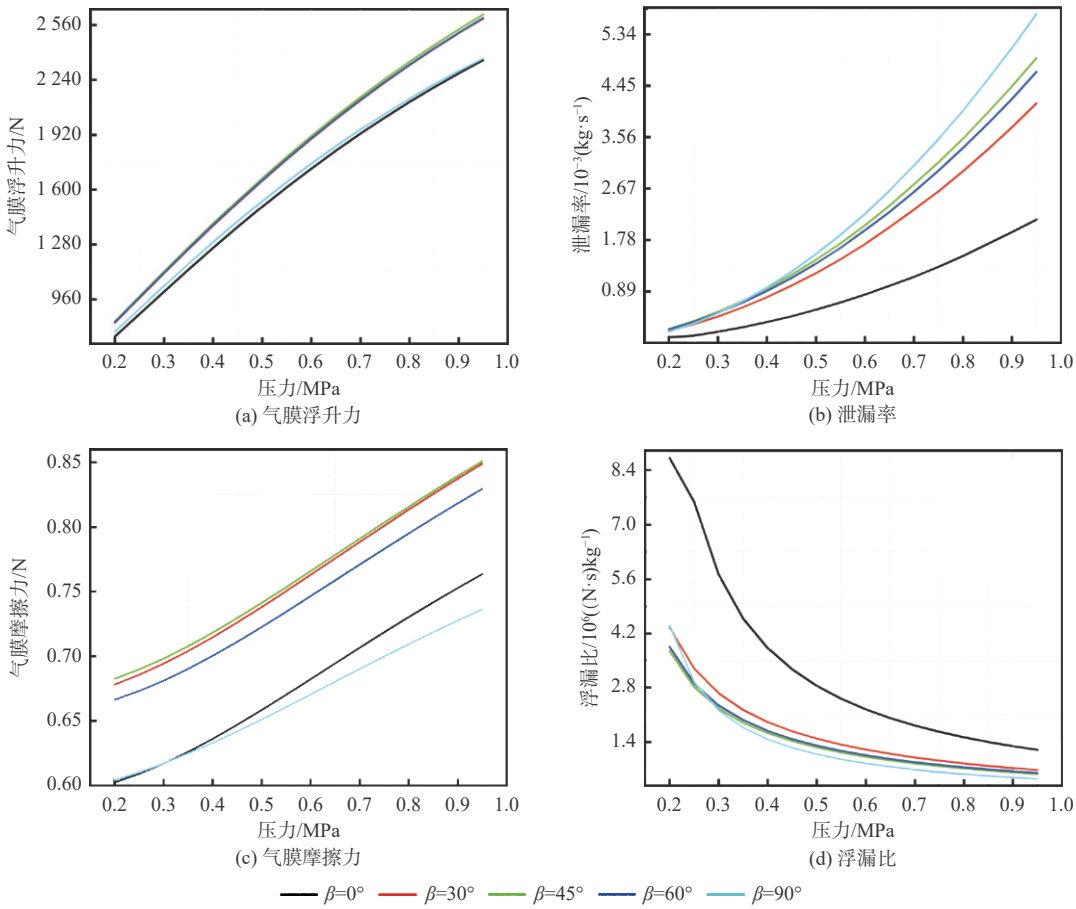


图 9 密封性能参数随压力和织构方向角度的变化曲线

Fig. 9 Variation curves of sealing performance parameters with pressure and texture direction angle

的周向气膜厚度出现分布不均匀现象，而且最小气膜厚度相比之前将会越薄，密封副间的润滑气膜受挤压效应和边界润滑条件的影响更大，导致气膜最薄处的动压效应增强；同时由于存在楔形效应，矩形微织构孔的动压效应也在气膜最薄处得到再次加强，所以宏观表现为气膜浮升力大幅度增大。说明了偏心率是制约微织构孔能否发挥动压作用的关键参数。

不同织构方向角下泄漏率随偏心率的变化曲线如图 10(b) 所示。当偏心率逐渐提高，不同织构方向角下的泄漏率都有明显的上升，而且随偏心率的提高其数值差异变大；其中，织构方向角 β=0°时泄漏率数值最小，增长趋势也较平缓，增值仅为 7.05×10^{-4} kg/s。而对于 β=30、β=45°、β=60°和 β=90°时的泄漏率数值而言，泄漏率随偏心率的提高分别上升了 2.14×10^{-3}, 2.35×10^{-3}, 2.30×10^{-3}, 1.82×10^{-3} kg/s。主要是由于偏心率逐渐增大，密封装置内的周向气膜厚度不均匀程度增加，在气膜最薄处出现了挤压效应，而且密封间隙内的介质气体在织构微孔域的孔壁处碰撞，由于壁面的阻挡，局部气膜压力快速增大，同时压力流占据气膜密封的主导地位，压差增大导致流速提高，在最大气膜厚度处的润滑气体流量增加，故泄漏率随之增大，密封装置的可靠性下降。

不同织构方向角下摩擦力随偏心率的变化曲线如图 10(c) 所示。偏心率的增加导致气膜摩擦力显著提升，但不同织构方向角下的气膜摩擦力数值随偏心率增加相差不大。而且相比较转速、压力和偏心率对气膜摩擦力的影响可以发现，偏心率对气膜摩擦力的影响最大。偏心率的提高使周向密封间隙内发生更大的压力梯度变化，可产生局部气膜高压。但偏心率是制约微织构孔能否发挥动压作用的关键参数，由于在轴套表面上开设了规则排列的矩形微织构孔，对轴向方向而言，气膜最小厚度随偏心率变大继续下降，虽然整体气膜刚度被大幅度的提高，但矩形微织构孔产生的动压效应进一步被削弱，故使得气膜摩擦力数值整体变大，但其数值大小比较相近。

不同织构方向角下浮漏比随偏心率的变化曲线如图 10(c) 所示。随着偏心率的提高，浮漏比的数值逐渐增大，而且矩形织构方向角对浮环密封的密封性能的影响增大，其中，织构方向角 β=0°的表现最好。当偏心率变大后，最薄气膜厚度会越小，密封间隙内的挤压效应和边界润滑效应增强，增强

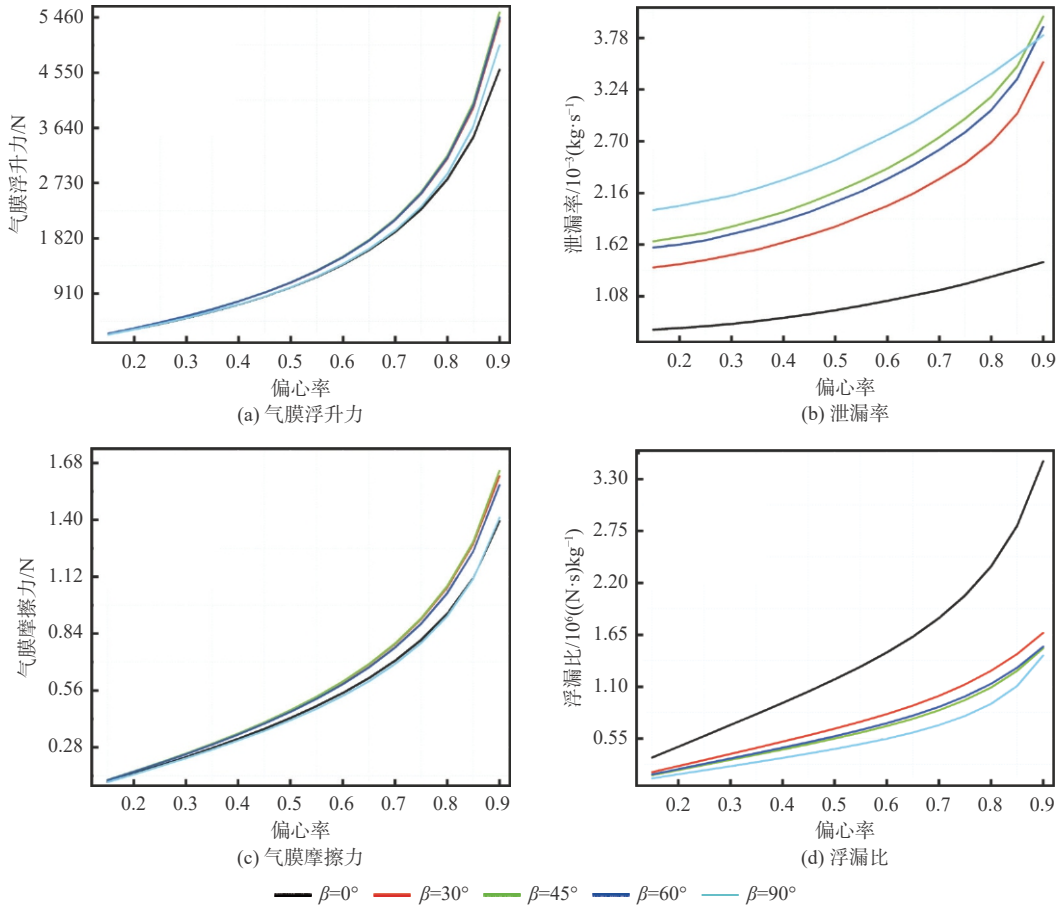


图 10　密封性能参数随偏心率和织构方向角度的变化曲线

Fig. 10　Variation curves of sealing parameters with eccentricity ratio and texture direction angle

了动压效应, 使得浮环气膜密封的浮升特性和稳定性提高, 这表明偏心率对提高密封性能整体有益。故进行密封设备安装时, 在不影响碰磨的情况下, 应尽量采取偏心安装, 以制造楔形间隙, 为产生动压效应制造有利条件。

3.4　织构深度对密封性能的影响

图 11 为密封性能参数随织构深度和织构方向角的变化曲线。不同织构方向角下气膜浮升力随织构深度的变化曲线如图 11(a) 所示, 随着微孔织构深度从 0.5 μm 提高到 4.25 μm, 不同织构方向角下密封装置的表面浮升力都在逐渐下降, 其中, β=0° 和 β=90°的浮升力下降程度最大。这是因为织构深度增加时, 平均气膜厚度也随之增加, 刚性表面条件下楔形间隙对介质气体的挤压效应减弱, 并且润滑气膜在运动过程中会在微孔织构的底部产生漩涡, 而且漩涡的产生使得上表面传递至润滑气膜的能量部分转化为织构内部的偏转能量, 削弱了织构的动压效应。因此, 过大的微织构深度使得润滑气膜厚增加, 气膜压力减小, 进而使得矩形微织构化浮环气膜密封的气膜浮升力不升反降。

不同织构方向角下泄漏率随织构深度的变化曲线如图 11(b) 所示, 微织构的深度变大, 泄漏率也随之变大。当微织构深度一定时, 织构方向角 β=0°的泄漏率与其他 4 种织构方向角的泄漏率数值差异较大。呈现这种趋势的主要原因是随着微孔织构深度的增加, 浮环与轴套之间的平均间隙也相应增大, 润滑介质泄漏通道变大, 加之浮环密封的偏心安装, 导致的密封装置间隙内周向气膜厚度分布不均匀, 造成气膜压力的波动变大。而且织构深度的增加会增加流场进气量, 从而轴端泄漏率也随之增加。不同织构方向角的矩形微织构孔所产生的动压效应有所差异, 故 5 种织构方向角的泄漏率数值差异随微织构深度增加而增大。

不同织构方向角下气膜摩擦力随织构深度的变化曲线如图 11(c) 所示, 其他工况参数不变时, 气膜摩擦力在矩形微孔织构深度从 0.5 μm 增加到 4.25 μm, 气膜摩擦力逐渐减小。并且不同织构方向角的矩形微织构之间的摩擦力数值差异也随织构深度增加而增大。气膜摩擦力的产生是由于牛顿的黏性剪切力和气体间的黏滞力导致, 当矩形微织构的深度逐渐增大时, 微织构的凹坑内部就会形成漩涡, 台阶效应也更加明显, 在有、无微孔的交接处气膜厚度发生突变, 压力梯度也相应增大。表面微织构减小了浮环密封楔形间隙内的黏性耗散, 漩涡

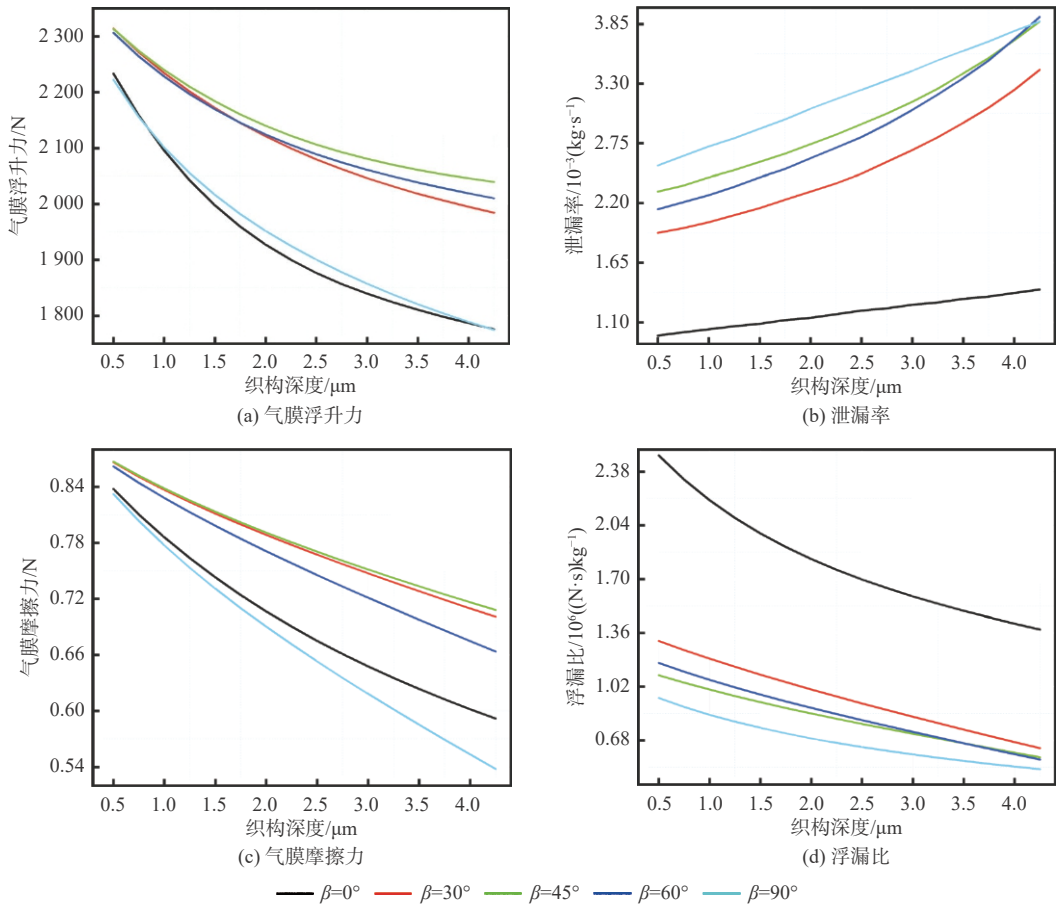


图 11 密封性能参数随结构深度和织构方向角度的变化曲线

Fig. 11 Variation curves of sealing parameters with texture depth and texture direction angle

效应会随着织构深度的增加而持续增强，从而导致气膜摩擦力下降。表明微织构在减小摩擦力方面有着显著的作用，微织构的深度越大体现得越明显。

不同织构方向角下浮漏比随织构深度的变化曲线如图 11(d) 所示，矩形微织构的深度变大，浮环密封综合性能逐渐降低。虽然 5 种织构方向角微织构的浮漏比数值都在下降，但 β=0°和微织构深度为 0.5 μm 的矩形微织构综合密封性能最佳。说明浮环气膜密封在服役的过程中，在沿运动方向，轴套表面的密封区域开设了矩形微织构孔，微孔的“富集效应”增强了介质气膜的流体动压效应；但并不是开设的微孔织构深度越深越好，为了使密封系统性能达到最佳，开设微孔的深度应合理选择，在满足气膜密封系统浮升力的同时也要考虑泄漏率不会显著增大。

4 结 论

1）浮升力影响：随着转速、压力和偏心率的增大均可以提高气膜浮升力，但微织构深度过大反而会降低浮升力。而且织构的方向角是影响密封润滑性能的重要因素，合理地选择织构方向角可提高气膜浮升力，提高动压效应能力的织构方向角排列为：45°>60°>30°>90°>0°。

2）泄漏率影响：随着压力、偏心率和织构深度的增加而迅速升高，但转速的升高对其泄漏率数值影响变化不大。织构方向角的微织构可改变密封间隙内润滑气体受到微织构孔壁的阻碍而汇集使得动压效应，进而导致泄漏率有所差异。减小泄漏率的织构方向角排列为：0°>30°>60°>45°>90°。

3）摩擦力影响：随着转速、压力和偏心率的升高而升高，随织构深度的增大而降低。微织构的凹坑内部就会形成漩涡，减小了楔形间隙内的黏性耗散，在减小摩擦力方面有着显著的作用。降低气膜摩擦力的织构方向角排列为：90°>0°>60°>30°>45°。

参考文献（References）

[1] XIA P, ZHANG G H, ZHAO J M, et al. Investigations on rotordynamic characteristics of a floating ring seal considering structural elasticity[C]//ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition. New York: ASME, 2017.

[2] XIA P, CHEN H, LIU Z S, et al. Analysis of whirling motion for the dynamic system of floating ring seal and rotor[J]. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2019, 233(8): 1221-1235.

[3] LI G Q, ZHANG Q, HUANG E L, et al. Leakage performance of

floating ring seal in cold/hot state for aero-engine[J]. Chinese Journal of Aeronautics, 2019, 32(9): 2085-2094.

[4] NAGAI K, KANEKO S, TAURA H, et al. Numerical and experimental analyses of dynamic characteristics for liquid annular seals with helical grooves in seal stator[J]. Journal of Tribology, 2018, 140(5): 052201.

[5] LU X L, ANDRÉS L S. Leakage and rotordynamic force coefficients of a three-wave (air in oil) wet annular seal: measurements and predictions[J]. Journal of Engineering for Gas Turbines and Power, 2019, 141(3): 032503.

[6] SINGH N, AWASTHI R K. Influence of dimple location and depth on the performance characteristics of the hydrodynamic journal bearing system[J]. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2020, 234(9): 1500-1513.

[7] SHI L P, ZHANG Y C, CHEN S, et al. Comparative research on gas seal performance textured with microgrooves and microdimples[J]. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 2019, 41(7): 280.

[8] SHI L P, WANG X Y, SU X, et al. Comparison of the load-carrying performance of mechanical gas seals textured with microgrooves and microdimples[J]. Journal of Tribology, 2016, 138(2): 021701.

[9] YANG X, PENG X D, MENG X K, et al. Thermo-elasto-hydrodynamic analysis of triangular textured mechanical face seals[J]. Journal of Zhejiang University: Science A, 2019, 20(11): 864-881.

[10] ADJEMOUT M, BRUNETIERE N, BOUYER J. Numerical analysis of the texture effect on the hydrodynamic performance of a mechanical seal[J]. Surface Topography: Metrology and Properties, 2015, 4(1): 014002.

[11] ADJEMOUT M, BRUNETIERE N, BOUYER J. Optimization of mesh density for numerical simulations of hydrodynamic lubrication considering textured surfaces[J]. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2015, 229(9): 1132-1144.

[12] 陈传刚, 丁雪兴, 陆俊杰, 等. 摩擦副界面微造型序列对气体密封性能的影响[J]. 化工学报, 2019, 70(3): 1016-1026.
CHEN C G, DING X X, LU J J, et al. Effect of friction pair interface micro-texture sequence on gas sealing performance[J]. CIESC Journal, 2019, 70(3): 1016-1026(in Chinese).

[13] 程香平, 孟祥铠, 彭旭东. 大菱形孔端面密封力变形及密封性能分析[J]. 化工学报, 2014, 65(8): 3089-3097.
CHENG X P, MENG X K, PENG X D. Analysis of surface deformation and seal performance of mechanical face seal with diamond macro-pores[J]. CIESC Journal, 2014, 65(8): 3089-3097(in Chinese).

[14] 程香平, 孟祥铠, 彭旭东, 等. 端面锥形参数对锥-孔组合型机械密封性能的影响[J]. 化工学报, 2016, 67(11): 4762-4770.
CHENG X P, MENG X K, PENG X D, et al. Effect of tapered face parameters on performance of cone-pore-combined mechanical seal[J]. CIESC Journal, 2016, 67(11): 4762-4770(in Chinese).

[15] 程香平, 张友亮, 康林萍, 等. 双向菱形孔织构端面密封性能研究[J]. 润滑与密封, 2020, 45(12): 73-80.
CHENG X P, ZHANG Y L, KANG L P, et al. Study on the performance of end face seal with bidirectional and double row diamond pores textured[J]. Lubrication Engineering, 2020, 45(12): 73-80 (in Chinese).

[16] 魏伟, 陈思, 章亦聪, 等. 三角形-椭圆复合微孔织构化机械端面密封性能的数值研究[J]. 流体机械, 2020, 48(8): 25-31.
WEI W, CHEN S, ZHANG Y C, et al. Numerical study on the sealing performance textured with multi-shape texture pattern combined with triangular and elliptical dimples[J]. Fluid Machinery, 2020, 48(8): 25-31(in Chinese).

[17] 魏伟, 时礼平, 俞凤桥, 等. 复合槽孔织构化端面机械密封性能的研究[J]. 安徽工业大学学报 (自然科学版), 2020, 37(1): 32-39.
WEI W, SHI L P, YU F Q, et al. A study of mechanical seal performance of textured end face with compound grooves and dimples[J]. Journal of Anhui University of Technology (Natural Science), 2020, 37(1): 32-39(in Chinese).

[18] 吉华, 王彦镐, 战琳月, 等. 等边三角形微孔端面机械密封多楔现象对性能的影响[J]. 工程科学与技术, 2019, 51(2): 168-175.
JI H, WANG Y H, ZHAN L Y, et al. Effect of multi-wedge on sealing performance in mechanical seal with equilateral triangular dimples[J]. Advanced Engineering Sciences, 2019, 51(2): 168-175 (in Chinese).

[19] 张然, 赵欢, 孙丹, 等. 螺旋槽对柱面气膜密封静力与动力特性影响机理研究[J]. 推进技术, 2023, 44(3): 87-96.
ZHANG R, ZHAO H, SUN D, et al. Influence mechanism of spiral groove on static and dynamic characteristics of cylinder gas film seal[J]. Journal of Propulsion Technology, 2023, 44(3): 87-96(in Chinese).

[20] LI Z G, FANG Z, LI J. Numerical investigation on the leakage and rotordynamic characteristics for three types of annular gas seals in wet gas conditions[J]. Journal of Engineering for Gas Turbines and Power, 2019, 141(3): 032504.

[21] WU D Z, JIANG X K, CHU N, et al. Numerical simulation on rotordynamic characteristics of annular seal under uniform and non-uniform flows[J]. Journal of Central South University, 2017, 24(8): 1889-1897.

[22] WANG X L, LIU M H, KAO-WALTER S, et al. Numerical evaluation of rotordynamic coefficients for compliant foil gas seal[J]. Applied Sciences, 2020, 10(11): 3828.

[23] KATSAROS K P, NIKOLAKOPOULOS P G. On the tilting-pad thrust bearings hydrodynamic lubrication under combined numerical and machine learning techniques[J]. Lubrication Science, 2021, 33(3): 153-170.

[24] 王国荣, 廖代胜, 钟林, 等. 沟槽形表面织构对柱塞密封副摩擦性能的影响[J]. 表面技术, 2019, 48(12): 165-173.
WANG G R, LIAO D S, ZHONG L, et al. Effect of groove-like surface texture on friction performance of plunger seal pair[J]. Surface Technology, 2019, 48(12): 165-173(in Chinese).

[25] 马彬鈅, 颜培, 余建杭, 等. 高压静密封表面微织构设计及其密封性能的有限元分析[J]. 表面技术, 2021, 50(8): 237-246.
MA B X, YAN P, YU J H, et al. Micro-texture design of high-pressure static sealing surface and finite element analysis of its sealing performance[J]. Surface Technology, 2021, 50(8): 237-246(in Chinese).

[26] 陆俊杰, 张炜, 谢方民, 等. 一种自适应柱状密封气膜特性分析[J]. 化工学报, 2020, 71(S1): 346-354.
LU J J, ZHANG W, XIE F M, et al. Analysis of gas film characteristics of an adaptive cylindrical seal[J]. CIESC Journal, 2020, 71(S1): 346-354(in Chinese).

[27] 纪敬虎, 邓智文, 陈天阳, 等. 局部织构无限长可倾瓦推力轴承流体动压润滑分析[J]. 表面技术, 2021, 50(2): 246-253.
JI J H, DENG Z W, CHEN T Y, et al. Analysis of hydrodynamic lubrication of partially textured infinitely long tilting pad thrust bearing[J]. Surface Technology, 2021, 50(2): 246-253(in Chinese).
[28] 王娟, 王隽, 田佳彬, 等. 表面织构水润滑聚合物轴承承载性能有限元分析[J]. 润滑与密封, 2021, 46(6): 90-96.
WANG J, WANG J, TIAN J B, et al. Finite element analysis of load-carrying property of surface texture water-lubricated polymer bearings[J]. Lubrication Engineering, 2021, 46(6): 90-96(in Chinese).
[29] 陈强. 表面微织构气体箔片轴承弹流润滑机理及特性研究[D]. 哈尔滨: 哈尔滨工业大学, 2017.
CHEN Q. Study on EHL mechanism and characteristics of surface micro-textured gas foil bearing[D]. Harbin: Harbin Institute of Technology, 2017(in Chinese).
[30] 于海武, 王晓雷, 孙造, 等. 圆柱形微凹坑表面织构对流体动压润滑性能的影响[J]. 南京航空航天大学学报, 2010, 42(2): 209-213.
YU H W, WANG X L, SUN Z, et al. Theoretical analysis on hydrodynamic lubrication of cylinder micro-dimple surface texture[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2010, 42(2): 209-213(in Chinese).
[31] 王学良, 刘美红, 熊忠汾, 等. 考虑表面粗糙度的柔性箔柱面气膜密封紊流特性分析[J]. 化工学报, 2022, 73(4): 1683-1694.
WANG X L, LIU M H, XIONG Z F, et al. Turbulence characteristics of compliant foil gas seal considering surface roughness[J]. CIESC Journal, 2022, 73(4): 1683-1694(in Chinese).

Orientation effect on sealing characteristics of rectangular micro-textured floating ring gas film

WANG Shipeng[1], DING Xuexing[1,*], LI Ning[2], DING Junhua[1], ZHANG Lanxia[1]

(1. School of Petrochemical Engineering, Lanzhou University of Technology, Lanzhou 730050, China;
2. AECC Hunan Aviation Powerplant Research Institute, Zhuzhou 412002, China)

Abstract: To investigate the orientation effect of micro-texture on the sealing performance parameters of the floating ring gas film, a rectangular micro-textured hole was selected as the study object. A theoretical analysis model of textured hydrodynamic lubrication control was developed based on gas lubrication theory, and the finite difference method was used to solve the model. The gas film pressure distribution in the sealing gap was obtained. With rotational speed, pressure, eccentricity, and texture depth as independent variables, this study primarily investigated the influence of texture direction angle on sealing performance parameters. The results indicate that the sealing uplift of the micro-textured floating ring increases with rotational speed, pressure, and eccentricity, while increasing micro-textured hole depth results in a decrease in the uplift. Furthermore, the leakage rate increases as rotational speed, pressure, eccentricity, and texture depth increase. However, an increase in rotational speed has little effect on the leakage rate value. The gas film friction also increases with the increase in rotational speed, pressure, and eccentricity and decreases as the texture depth improves. Rectangular micro-texture plays a crucial role in improving the sealing performance parameters. The rational selection of micro-texture direction angle can effectively enhance the sealing performance of the floating ring gas film. These study outcomes provide a novel idea for enhancing the sealing performance of the textured surface floating ring.

Keywords: floating ring seal; rectangular micro-texture; orientation effect; sealing performance; numerical calculation

Received: 2023-03-14; **Accepted**: 2023-07-08; **Published Online**: 2023-07-11 08:49
URL: link.cnki.net/urlid/11.2625.V.20230710.1709.004
Foundation items: National Natural Science Foundation of China (51565029); Major Special Projects in Ningbo (2020Z112)
* **Corresponding author**. E-mail: dingxxseal@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0159

丝网印刷石墨烯标签天线的不确定性分析

李凯[1,2]，沈志刚[1,2]，张晓静[1,2,*]

(1. 北京航空航天大学 航空科学与工程学院，北京 100191；2. 粉体技术研究开发北京市重点实验室，北京 100191)

摘 要：石墨烯标签天线因具有价格低廉和环境友好的优势而有望替代金属标签天线。然而，石墨烯标签天线在制备过程中受到丝网印刷工艺和石墨烯油墨的影响，不可避免地引入不确定性，使得实际阅读距离与理论值不符。为指导包含不确定性的石墨烯标签天线的设计，使天线阅读距离处于特定范围内，提出一种基于区间分析的算法。通过分析天线理论公式，得到不确定性参数；根据制备经验，为每个不确定性参数估计区间范围，使用区间分析算法（包括区间数学建模和顶点法）计算阅读距离的上下限；通过实验测量了石墨烯标签天线的阅读距离，验证了所提算法。结果表明：所提算法可以有效地计算石墨烯标签天线的阅读距离区间，有助于石墨烯标签天线的实际使用，并为制造过程中的不确定性分析提供指导。

关 键 词：石墨烯；标签天线；阅读距离；区间分析；不确定性

中图分类号：V214; TB12

文献标志码：A **文章编号**：1001-5965(2025)03-0857-08

射频识别技术最早出现于20世纪60年代[1]，由于其在访问管理、货物和人员跟踪、无线通信等方面的广泛应用，正变得越来越流行。目前，金属基纳米颗粒（如银和铜）通常用于制造标签天线[2-4]。然而，金属基纳米颗粒价格昂贵[5]。金属天线的生产通常涉及化学蚀刻以形成导电电路。化学蚀刻过程需要使用大量具有污染性的化学试剂[6]。因此，急需使用环保材料进行替代。

石墨烯是一种由具有sp2杂化轨道的碳原子组成的二维碳纳米材料[7]。由于其具有高导电性、经济性和环境友好性，石墨烯有望取代金属基纳米材料作为标签天线的制造材料[8]。石墨烯标签天线和金属标签天线之间最显著的区别是石墨烯天线的方阻不能忽略。为降低方阻，研究人员已做出许多努力。例如，Huang等[9]开发了一种滚动压缩方法，以降低无黏合剂石墨烯层压板的方阻。通过使用石墨烯纳米片，Leng等[10]制造了一种丝网印刷的弯折偶极子天线来用作传感器。Akbari等[11]设计了一种类似于环形天线的新型石墨烯偶极子天线。此外，Arapov等[12]研究了不同柔性基板上的石墨烯基偶极子天线的性能。Merilampi等[13]研究了导电油墨层的厚度对标签天线性能的影响。这些研究证明了石墨烯在标签天线领域的潜力。为加速设计过程并节省成本，指导石墨烯天线设计的理论模型是必不可少的。然而，石墨烯天线的制造过程中存在不确定性，这可能导致无法将实验室级样品转移到大规模商业产品中。由于不确定性，在大规模生产的情况下，从实验室中的几个样品获得的性能可能不容易再现。

石墨烯标签天线的制造过程产生了不同类型的不确定性参数。由于信息交换是标签天线的核心功能，因此，最关键的性能参数——阅读距离，是本文考虑的重点。石墨烯标签天线的阅读距离受石墨烯油墨的性质和丝网印刷精度的影响，产生了

收稿日期：2023-04-04；**录用日期**：2023-05-15；**网络出版时间**：2023-06-06 14:26
网络出版地址：link.cnki.net/urlid/11.2625.V.20230605.1445.001
基金项目：中央高校基本科研业务费专项资金 (YWF-22-L-1219)
* **通信作者**. E-mail：zhangxiaojing@buaa.edu.cn

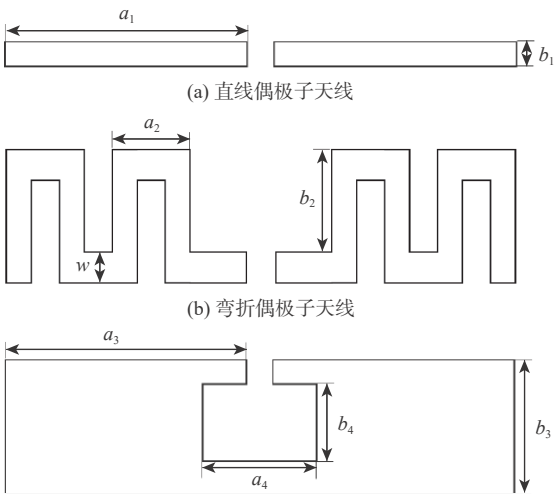
引用格式：李凯，沈志刚，张晓静. 丝网印刷石墨烯标签天线的不确定性分析 [J]. 北京航空航天大学学报，2025，51（3）：857-864.
LI K，SHEN Z G，ZHANG X J. Study on uncertainties of graphene tag antenna by screen printing[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（3）：857-864（in Chinese）.

不同的不确定性参数，如导电性、厚度、几何形状等。由于石墨烯油墨不可能绝对均匀，标签天线的导电性和厚度分布会不同。另外，印刷误差也很常见。因此，如果要设计石墨烯天线的理论模型，则该模型必须考虑不确定性，并制定相关的数学策略。然而，这种模型很少有人考虑到。

本文提出一种区间分析算法来计算石墨烯标签天线阅读距离的不确定性。区间数学分析算法[14-16]可以考虑所有不确定性参数，并可以获得包含阅读距离所有可能值的集合。区间分析是一种非概率不确定性分析算法，在样本数量较少的情况下具有优势。每个不确定参数被设定为一个区间数，计算得到的阅读距离是考虑所有区间数的情况下得到的，也是一个区间数。最后，通过实验验证了这种算法。

1 标签天线阅读距离的理论公式

分析理论阅读距离公式有助于找到可能影响阅读距离的参数。通常，直线偶极子天线、弯折偶极子天线和环形偶极子天线是标签天线中的3种典型形状，这3种偶极子天线如图1所示。



(a) 直线偶极子天线

(b) 弯折偶极子天线

(c) 环形偶极子天线

图1 3种偶极天线示意图

Fig. 1 Illustration of three types of dipole antennas

假设芯片接收的电磁波能量在转换过程中没有消耗，理论阅读距离可以定义为[17]

$$R=\frac{c}{4\pi f}\sqrt{\frac{P_{\mathrm{EIRP}}G_{\mathrm{ant}}(1-|\Gamma|)}{P_{\mathrm{Rsens}}}} \tag{1}$$

式中：c为光速；f为工作频率；P_{EIRP}为读取器设备的等效同位素辐射功率；G_{ant}为标签天线的增益；P_{Rsens}为微芯片的读取灵敏度；Γ为反射系数，可表示为

$$\Gamma=\frac{Z_{\mathrm{ant}}-Z_{\mathrm{chi}}^{*}}{Z_{\mathrm{ant}}+Z_{\mathrm{chi}}} \tag{2}$$

其中：Z_{ant}和Z_{chi}分别为天线和微芯片的阻抗；Z_{chi}^{*}为Z_{chi}的复共轭。

根据欧洲电信标准协会的EN300-220规范，P_{EIRP}最高可达3.28 W。读取灵敏度P_{Rsens}由芯片确定。Alien Higgs-3芯片被广泛使用，其灵敏度为−20 dBm，工作频率为860～960 MHz。一旦选择了芯片，其阻抗Z_{chi}就固定在某一频率上。根据式(2)，反射系数Γ仅受天线阻抗Z_{ant}的影响。只要获得天线增益G_{ant}和阻抗Z_{ant}，就可以计算理论阅读距离。

求解天线增益G_{ant}和阻抗Z_{ant}相当复杂，一方面，在解决过程中需要进行一些假设和近似；另一方面，近似解仅适用于某些特殊天线。因此，高频结构模拟器(high frequency structural simulator, HFSS)等仿真软件被大量使用。根据标签天线基本理论[18-21]，标签天线的外形、厚度及导电率都会影响标签天线的阻抗和增益。

2 石墨烯标签天线的不确定性参数

由于石墨烯天线的制备过程包括天线设计、制备油墨、印刷和干燥，导致大多数不确定性参数产生，因此，应首先介绍石墨烯标签天线的制备。

导电石墨烯油墨含有石墨烯、乙炔黑、分散剂、溶剂和黏合剂。石墨烯和乙炔黑作为导电材料。分散剂有助于导电材料溶解在溶剂中。黏合剂使石墨烯薄片形成连续薄膜。石墨烯油墨被丝网印刷在聚对苯二甲酸乙二醇酯(polyethylene terephthalate, PET)上以制造标签天线，使用孔径为0.25 mm的筛网。为使溶剂和分散剂挥发，天线需要在90℃下加热4 h。然后，用滚筒压缩天线以提高其导电性和均匀性。

石墨烯标签天线中的导电率和厚度因区域而异，因此，在计算阅读距离时，导电率和厚度成为不确定参数。此外，丝网印刷过程无法精确控制，这也带来了不确定性[22]。丝网印刷的孔径决定了石墨烯标签天线的尺寸精度。因此，印刷石墨烯标签天线的尺寸将偏离设计尺寸。此外，刮板对印刷效果有直接影响[23]。通常，刮板的角度大、速度快，会导致较少的油墨渗透到网中，穿过丝网网孔的油墨数量会影响标签天线的厚度。图2和图3分别为石墨烯弯折和石墨烯环形偶极子天线横截面的扫描电子显微镜(scanning electron microscope, SEM)图像，由于油墨的不均匀性和丝网印刷工艺，产生的石墨烯层的厚度不均匀。

根据本节分析，不确定性的来源可分为3类：加工误差、油墨均匀性和外部干扰，其对石墨烯标签天线的实际阅读距离有影响。本节导出的不确

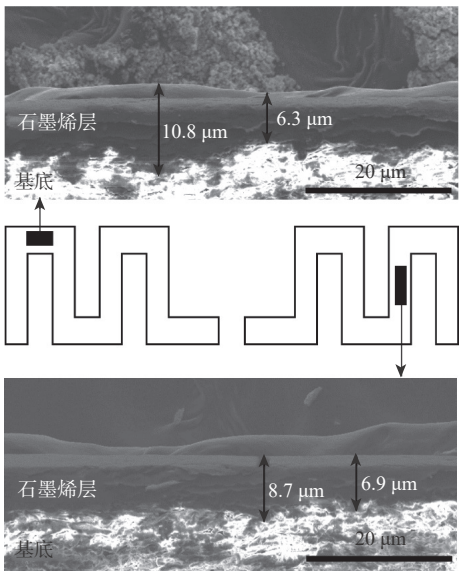


图 2　石墨烯弯折偶极子天线横截面的 SEM 图像

Fig. 2　Cross-section SEM images of graphene meandered-line dipole antenna

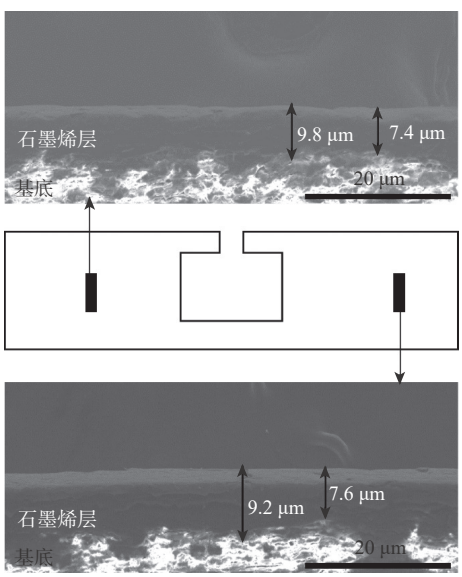


图 3　石墨烯环形偶极子天线横截面的 SEM 图像

Fig. 3　Cross-section SEM images of a graphene loop dipole antenna

定性参数是尺寸参数和电气性能参数, 如电导率。天线的参数值如表 1 所示, 表中各符号均为偶极子天线的不同位置尺寸, 其含义如图 1 所示。

表 1　3 种偶极天线的参数

Table 1　Parameters of the three types of dipole antennas　　mm

参数	数值	参数	数值
a_1	80	b_1	10
a_2	15	b_2	17
a_3	50	b_3	25
a_4	13	b_4	14
w	5	t	0.01

3　基于区间分析的阅读距离计算模型

本节提出一种新的基于区间分析的模型, 以计算考虑不确定性的石墨烯标签天线的阅读距离。使用的区间分析算法包括区间数学建模和区间顶点法[24-25]。计算模型包括灵敏度分析和阅读距离计算。

3.1　区间分析算法

区间分析算法假设在所有计算过程中, 所有不确定性参数都由未知但有界的集合描述[26]。这样, 可以获得包含石墨烯标签天线的阅读距离的所有可能值的集合。由未知但有界的参数组成的向量可以描述为

$$\begin{cases} \boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}} \\ x_i\in x_i^{\mathrm{I}}=\left[x_i^{\mathrm{l}},x_i^{\mathrm{u}}\right] \qquad i=1,2,\cdots,n \end{cases} \tag{3}$$

式中: $\boldsymbol{x}$ 为由 n 个元素组成的不确定性向量; 上标 I 表示未知但有界的集合, 也称为区间集合; 上标 u 和 l 分别代表不确定性参数的上限和下限。

不确定性向量 $\boldsymbol{x}_i^{\mathrm{c}}$ 的平均向量和半径向量 $\boldsymbol{x}_i^{\mathrm{r}}$ 定义为

$$\begin{cases} \boldsymbol{x}^{\mathrm{c}}=\left(\boldsymbol{x}_i^{\mathrm{c}}\right)=\dfrac{\boldsymbol{x}^{\mathrm{u}}+\boldsymbol{x}^{\mathrm{l}}}{2} \\ \boldsymbol{x}^{\mathrm{r}}=\left(\boldsymbol{x}_i^{\mathrm{r}}\right)=\dfrac{\boldsymbol{x}^{\mathrm{u}}-\boldsymbol{x}^{\mathrm{l}}}{2} \qquad i=1,2,\cdots,n \end{cases} \tag{4}$$

区间计算[27] 可以得到包含所有可能结果的集合。

3.2　基于区间分析的阅读距离计算模型

根据区间数学模型[28], 石墨烯标签天线的制造过程中存在不确定性参数, 包括尺寸参数和电气性能参数。由于考虑所有不确定性参数可能导致巨大的计算成本, 使用的灵敏度分析[29-30] 来选择对计算的阅读距离有主要影响的参数。

由于每个自变量都是区间集, 因此, 应先进行灵敏度分析。将间隔等分为 q 个部分, 并将每个分段点的值纳入计算中, 以获得如下灵敏度值:

$$\begin{cases} k_{jp}=\Delta R/\Delta x_{jp}=\left.\dfrac{R\left(x_{j(p+1)}\right)-R\left(x_{jp}\right)}{x_{j(p+1)}-x_{jp}}\right|_{x_{jp}=x_j^{\mathrm{l}}+\frac{p}{q}\left(x_j^{\mathrm{u}}-x_j^{\mathrm{l}}\right)} \\ p=1,2,\cdots,q \end{cases} \tag{5}$$

相应不确定性参数的最终灵敏度计算如下:

$$k_j=\max\left\{\left|k_{jp}\right|\right\} \qquad p=0,1,\cdots,q \tag{6}$$

在选择了重要的不确定性参数之后, 应用区间顶点法来实现计算的阅读距离的区间集。该方法要求自变量和因变量之间的函数关系应该是单调的, 或者函数关系具有有限极点。

在天线设计中，根据单调关系，将不确定性参数分为2部分。满足以下条件的不确定性参数组合称为第1部分：不确定性参数与阅读距离之间的函数关系类似于二次函数，存在阅读距离的最大值。与阅读距离单调的其他不确定性参数的组合被称为第2部分。

模型的实施过程步骤如下。

步骤1 确定所有不确定性参数并将其分为2部分。

1）创建包含所有不确定性参数的输入向量。假设元素的数量为N，则向量可以在数学上写为

$$\boldsymbol{P}=[\boldsymbol{P}(1)\quad \boldsymbol{P}(2)\quad \cdots\quad \boldsymbol{P}(N)] \tag{7}$$

2）确定不确定性参数与计算阅读距离之间的函数关系。应将任何选定的不确定性参数的间隔划分为m个部分，并通过模拟计算该参数取m+1个节点的值时的阅读距离。在此过程中，其他参数为平均值。如果存在计算阅读距离的极点，则应将所选参数放入第1部分。如果函数关系是单调的，则应将所选参数放入第2部分。假设第1部分和第2部分中的参数数量分别为N_1和N_2。

步骤2 当第1部分中的不确定性参数在其范围内变化时，计算阅读距离的最大值和最小值。

1）对于第1部分中的每个不确定性参数，将计算出的m+1值拟合成曲线，并选择最接近极点的3个节点，以建立二次函数。然后，找到这个二次函数的极点。该极点与区间边界一起构成向量$\boldsymbol{X}_j$，j=1, 2, N_1。$\boldsymbol{X}_j$中有3个元素，因此，当不确定性参数采用这3个元素时，阅读距离仅达到最大值或最小值。

2）计算阅读距离以获得最大值和最小值。在计算过程中，第1部分中$\boldsymbol{X}_j$参数将使用步骤3获得的向量中的每个元素，而第2部分的参数用平均值。显然将存在3^{N_1}个阅读距离值：

$$\begin{cases} R^{\mathrm{l1}}=\min(R_1,R_2,\cdots,R_{3^{N_1}}) \\ R^{\mathrm{u1}}=\max(R_1,R_2,\cdots,R_{3^{N_1}}) \end{cases} \tag{8}$$

当阅读距离达到最大值或最小值时，记录第1部分中的不确定性参数值。然后，将这些参数确定下来。

步骤3 当第2部分中不确定性参数在其范围内变化时，计算阅读距离的最大值和最小值。

1）构造第2部分中参数的间隔顶点阵列：

$$\begin{cases} \boldsymbol{M}_{\mathrm{vertex}}=\begin{bmatrix} 1+\beta_1 & \cdots & 1+\beta_{N_2} \\ \vdots & 1+\beta_j(-1)^{<(i-1)/2^{N_2-j}>} & \vdots \\ 1-\beta_1 & \cdots & 1-\beta_{N_2} \end{bmatrix} \\ i=1,2,\cdots,2^{N_2},\ j=1,2,\cdots,N_2 \end{cases} \tag{9}$$

式中：i和j分别为数组的行号和列号；$\boldsymbol{\beta}$=[β_1, β_2, ⋯, β_{N_1}]表示偏差因子，其定义为半径与向量$\boldsymbol{P}$中相应元素的平均值之比；<*>代表比*小的最大整数。

输入向量$\boldsymbol{P}_{\mathrm{v}}$被定义为

$$\begin{cases} \boldsymbol{P}_{\mathrm{v}}=\boldsymbol{M}_{\mathrm{vertex}}\boldsymbol{P}_{\mathrm{vertex}} \\ \boldsymbol{P}_{\mathrm{vertex}}=\mathrm{diag}(\boldsymbol{P}(1),\boldsymbol{P}(2),\cdots,\boldsymbol{P}(N_2)) \end{cases} \tag{10}$$

2）计算基于输入向量$\boldsymbol{P}_{\mathrm{v}}$的阅读距离，并获得最小值和最大值。在计算过程中，平均值用于第1部分中的参数。计算的阅读距离将有2^{N_2}个值。最小值和最大值选择为

$$\begin{cases} R^{\mathrm{l2}}=\min(R_1,R_2,\cdots,R_{2^{N_2}}) \\ R^{\mathrm{u2}}=\max(R_1,R_2,\cdots,R_{2^{N_2}}) \end{cases} \tag{11}$$

步骤4 计算阅读距离的上限和下限。

当用步骤2中第2步选择的第1部分中的参数和步骤3中第2步选择的第2部分中的最小参数计算阅读距离时，获得所计算的阅读距离的最小值。阅读距离的最大值以类似的方式实现。

为此，得到了区间集R的上下界，方程式的数量为$3^{N_1}+2^{N_2}$。因此，当输入参数的数量增加时，不确定性问题的计算时间比确定性问题的时间长得多。为减少计算时间，可以采用并行-串行组合顶点算法。

尽管基于区间的阅读距离计算方法提供了包含给定（有限）参数范围内的所有可能值的阅读距离集合，但集合可能过大。区间运算最明显的特征是区间扩张现象[31]，这可能是由保守估计引起的。导致区间过度扩张的另一个原因是相关性问题[32]，当区间参数在解析表达式中重复出现时，会出现相关性问题。依赖问题导致的高估可以通过符号简化来避免。本文石墨烯标签天线的阅读距离是通过软件计算的，因此，只能通过选择合适的间隔参数上限和下限来减少区间过度扩张，这将在未来的工作中进行研究。

4 数值计算和实验验证

本节通过使用区间分析算法进行计算，获得计算阅读距离的上限和下限。为验证区间分析算法，对天线的实际阅读距离进行实验测量。由于直线天线用处较小，所以未制备。此外，采用灵敏度分析来确定对理论阅读距离影响较小的不确定性参数，从而节省了计算资源。本节讨论每个不确定性参数与计算阅读距离之间的单调性。此外，利用选定的不确定性参数，通过顶点法计算阅读距离的边界。真实的阅读距离是在可以消除外部电磁干扰的消声室中进行实验测量的。

4.1　灵敏度分析

对每个不确定性参数进行灵敏度分析, 表 2 和表 3 分别为弯折线天线和环形天线的仿真结果。由于每个不确定性参数值的间隔长度不同, 因此, 增量是根据样本测量值和经验值确定的, 通过其增量可以测量不确定性参数对理论读数范围的影响。由表 2 可知, b_2 可以忽略, 因为其对阅读距离影响很小。相反, 厚度 t 和电导率 σ 对阅读距离有很大影响。由表 3 可知, 与电导率和厚度相比, 除厚度之外的尺寸参数对计算的阅读距离都有微弱的影响。因此, 在使用顶点方法的计算中, a_3、b_3、a_4、b_4 被忽略。

表 2　弯折偶极子天线不确定性参数的灵敏度分析

Table 2　Sensitivity analysis of uncertainty parameters for meandered-line dipole antenna

不确定参数的增量式	理论阅读距离增量/m
a_2+0.8 mm	0.619 5
b_2+0.8 mm	0.032 7
w+0.8 mm	1.314
t+6 μm	2.917 7
σ+6 000 S/m	0.893 9

表 3　环形偶极子天线不确定性参数的灵敏度分析

Table 3　Sensitivity analysis of uncertainty parameters for loop dipole antenna

不确定参数的增量式	理论阅读距离增量/m
a_3+0.8 mm	0.153 2
b_3+0.8 mm	0.184 0
a_4+0.8 mm	0.044 9
b_4+0.8 mm	0.043 5
t+6 mm	2.753 3
σ+6 000 S/m	1.347 8

4.2　阅读距离计算

首先, 分析不确定性参数与计算阅读距离之间的函数关系。如图 4 所示, 弯折天线的 a_2、w、t 和 σ 与计算的阅读距离正相关。如图 5 所示, 环形天线的 t 和 σ 与计算的阅读距离呈正相关。因此, 计算阅读距离时, 省略 3.2 节的步骤 2。

其次, 利用灵敏度分析选择的不确定性参数, 通过 HFSS 软件计算弯折偶极子天线和环形偶极子天线阅读距离的区间集。对于弯折偶极子天线, 当 a_2、w、t 和 σ 分别为 15.8 mm、5.8 mm、16 μm 和 34 571.4 S/m 时, 阅读距离最大值为 6.930 7 m; 当 a_2、w、t 和 σ 分别为 14.2 mm、4.2 mm、4 μm 和 22 571.4 S/m 时, 阅读距离最小值为 1.823 8 m。对于环形偶极子天线, 当 t 和 σ 分别为 16 μm 和 34 571.4 S/m 时, 阅读距离的最大值为 9.532 2 m; 当 t 和 σ 分别为 4 μm

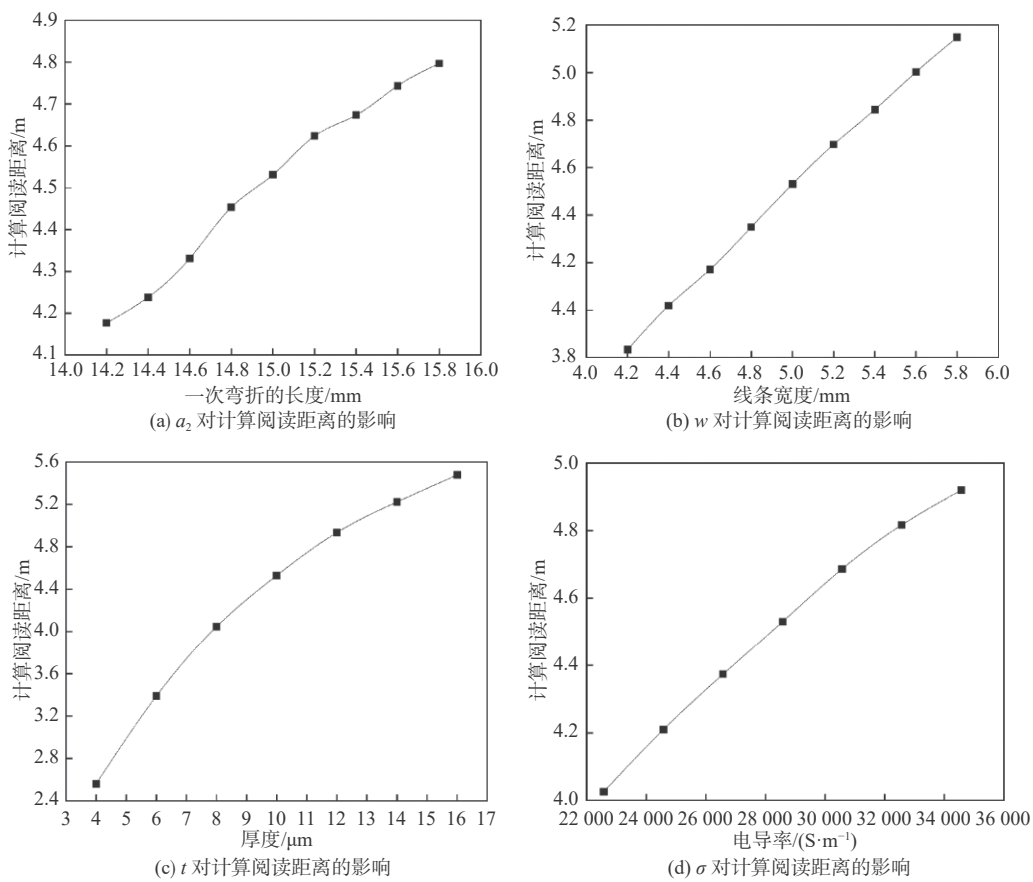
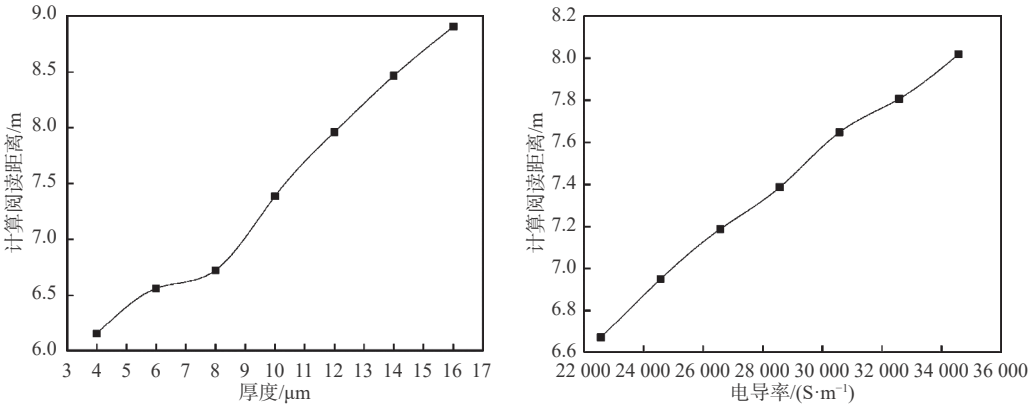


图 4　弯折偶极子天线在不同参数变化下的计算阅读距离

Fig. 4　Calculated read range for meandered-line dipole antenna with change of different parameters



(a) t 对计算阅读距离的影响　(b) σ 对计算阅读距离的影响

图 5　环形偶极子天线在不同参数变化下的计算阅读距离

Fig. 5　Calculated read range for loop dipole antenna with change of different parameters

和 22 571.4 S/m 时，阅读距离的最小值为 6.252 9 m。因此，弯折偶极子天线和环形偶极子天线的阅读距离集合分别为 R_{ms}=[1.823 8, 6.930 7] 和 R_{ls}=[6.252 9, 9.532 2]。

4.3　实验验证

丝网印刷石墨烯标签天线的过程如图 6 所示。丝网印刷所得的石墨烯标签天线如图 7 所示。天线的平均厚度为 10 μm，平均薄层电阻为 3.5 Ω/m^2。阅读距离的实验结果如图 8 所示。弯折偶极子天线的实际阅读距离的最大值和最小值分别为 5.23 m 和 3.01 m。环形偶极子天线实际阅读距离的最大值和最小值分别为 9.24 m 和 7.11 m。通过本文方法计算的阅读距离的集合包含了所有实验值。本文算法能够给出包含所有可能的实际结果的集合，

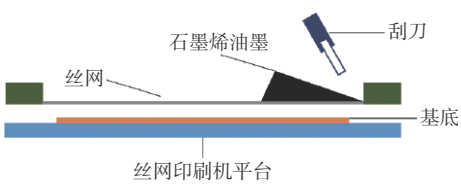


图 6　丝网印刷石墨烯标签天线的过程

Fig. 6　Fabrication of printed graphene tag antennas process

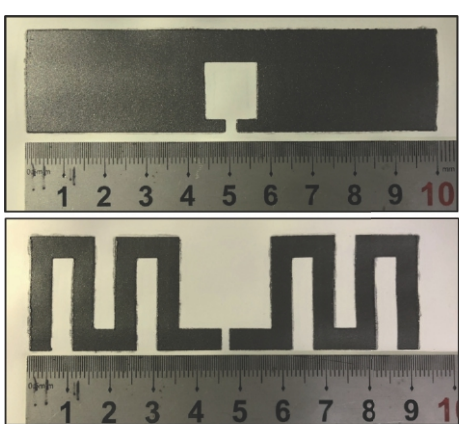
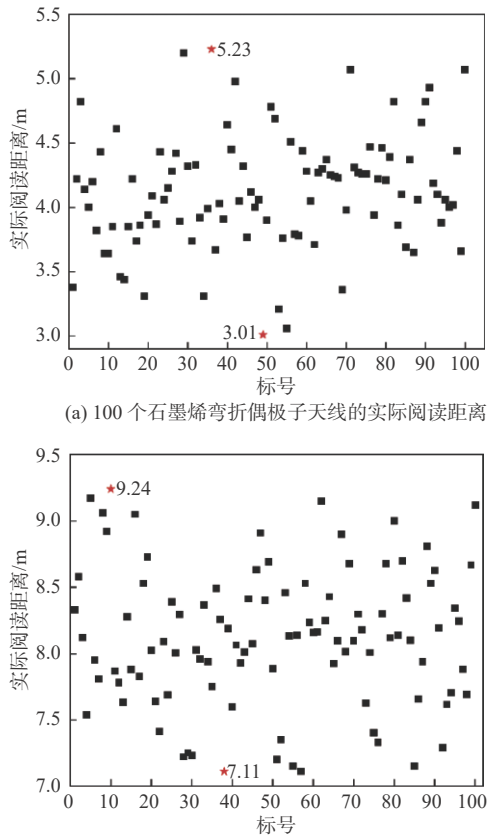


图 7　丝网印刷的石墨烯标签天线的实物图

Fig. 7　Physical image of screen-printed graphene tag antennae



(a) 100 个石墨烯弯折偶极子天线的实际阅读距离

(b) 100 个石墨烯环形偶极子天线的实际阅读距离

图 8　通过实验测试获得的实际阅读距离

Fig. 8　Real read range obtained by experimental testing

是一种有工程应用价值的方法。通过实验测量的实际阅读距离比计算的阅读距离保守得多。这可能是由于采用不确定性参数的顶点值进行计算，实际参数并不完全处于顶点值。结果表明，特别是对于弯折偶极子天线，区间边界值相差较大。因此，讨论阅读距离计算过程中的不确定性是有意义的。

5　结　论

1）根据灵敏度分析，弯折偶极子天线的阅读距离对线条宽度 w、厚度 t 灵敏度分别为 1.314 和 2.917 7。

这 2 个参数对阅读距离产生较大影响, 因此, 在设计和加工阶段需要重点控制其误差。

2) 根据灵敏度分析, 环形偶极子天线的厚度 t、电导率 σ 对阅读距离产生影响较大, 其灵敏度分别为 2.753 3 和 1.347 8, 因此, 这 2 个参数需要重点关注。

3) 根据不确定性分析, 弯折偶极子天线的阅读距离范围为 1.823 8～6.930 7 m, 环形偶极子天线的阅读距离范围为 6.252 9～9.532 2 m。

4) 经过实验, 弯折偶极子天线的实际阅读距离范围为 3.01～5.23 m, 环形偶极子天线实际阅读距离范围为 7.11～9.24 m, 均在理论分析范围内。

在未来的研究中, 将进一步提高石墨烯油墨性能, 改善印刷方式, 减少石墨烯标签天线的不确定性。此外, 也将进一步研究诸如喷墨打印等不同方式的不确定性算法。

参考文献（References）

[1] LANDT J. The history of RFID[J]. IEEE Potentials, 2005, 24: 8-11.

[2] KIM Y, LEE B, YANG S, et al. Use of copper ink for fabricating conductive electrodes and RFID antenna tags by screen printing[J]. Current Applied Physics, 2012, 12(2): 473-478.

[3] SALMERÓN J F, MOLINA-LOPEZ F, BRIAND D, et al. Properties and printability of inkjet and screen-printed silver patterns for RFID antennas[J]. Journal of Electronic Materials, 2014, 43(2): 604-617.

[4] WOLF F M, PERELAER J, STUMPF S, et al. Rapid low-pressure plasma sintering of inkjet-printed silver nanoparticles for RFID antennas[J]. Journal of Materials Research, 2013, 28(9): 1254-1261.

[5] LI Y, LU D, WONG C P. Electrical conductive adhesives with nanotechnologies [M]. Berlin: Springer, 2010.

[6] BJÖRNINEN T, MERILAMPI S, UKKONEN L, et al. The effect of fabrication method on passive UHF RFID tag performance[J]. International Journal of Antennas and Propagation, 2009, 2009: 920947.

[7] YI M, SHEN Z G. A review on mechanical exfoliation for the scalable production of graphene[J]. Journal of Materials Chemistry A, 2015, 3(22): 11700-11715.

[8] KOPYT P, SALSKI B, OLSZEWSKA-PLACHA M, et al. Graphene-based dipole antenna for a UHF RFID tag[J]. IEEE Transactions on Antennas and Propagation, 2016, 64(7): 2862-2868.

[9] HUANG X J, LENG T, ZHANG X, et al. Binder-free highly conductive graphene laminate for low cost printed radio frequency applications[J]. Applied Physics Letters, 2015, 106(20): 203105.

[10] LENG T, HUANG X J, CHANG K, et al. Graphene nanoflakes printed flexible meandered-line dipole antenna on paper substrate for low-cost RFID and sensing applications[J]. IEEE Antennas and Wireless Propagation Letters, 2016, 15: 1565-1568.

[11] AKBARI M, KHAN M W A, HASANI M, et al. Fabrication and characterization of graphene antenna for low-cost and environmentally friendly RFID tags[J]. IEEE Antennas and Wireless Propagation Letters, 2015, 15: 1569-1572.

[12] ARAPOV K, JAAKKOLA K, ERMOLOV V, et al. Graphene screen-printed radio-frequency identification devices on flexible substrates[J]. Physica Status Solidi (RRL) - Rapid Research Letters, 2016, 10(11): 812-818.

[13] MERILAMPI S L, BJÖRNINEN T, VUORIMÄKI A, et al. The effect of conductive ink layer thickness on the functioning of printed UHF RFID antennas[J]. Proceedings of the IEEE, 2010, 98(9): 1610-1619.

[14] JAULIN L, KIEFFER M, DIDRIT O, et al. Applied interval analysis[M]. Berlin: Springer, 2001.

[15] HICKEY T, JU Q, VAN EMDEN M H. Interval arithmetic: from principles to implementation[J]. Journal of the ACM, 2001, 48(5): 1038-1068.

[16] ALEFELD G, CLAUDIO D. The basic properties of interval arithmetic, its software realizations and some applications[J]. Computers & Structures, 1998, 67(1-3): 3-8.

[17] RAO K V S, NIKITIN P V, LAM S F. Antenna design for UHF RFID tags: a review and a practical application[J]. IEEE Transactions on Antennas and Propagation, 2005, 53(12): 3870-3876.

[18] KRAUS J D, MARHEFKA R J. Antennas for all applications [M]. New York: McGraw-Hill, 2002: 9-14.

[19] BALANIS C A. Antenna theory: analysis and design[J]. IEEE Antennas & Propagation Society Newsletter, 2003, 24(6): 28-29.

[20] 田川, 尹祖伟. 无源超高频标签天线工程设计案例教程[M]. 北京: 清华大学出版社, 2019: 20-25.
TIAN C, YIN Z W. Engineering design case studies of passive UHF tag antennas[M]. Beijing: Tsinghua University Press, 2019: 20-25(in Chinese).

[21] ENDO T, SUNAHARA Y, SATOH S, et al. Resonant frequency and radiation efficiency of meander line antennas[J]. Electronics and Communications in Japan (Part II: Electronics), 2000, 83(1): 52-58.

[22] BLAYO A, PINEAUX B. Printing processes and their potential for RFID printing[C]// Proceedings of the Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-Aware Services: Usages and Technologies. New York: ACM, 2005: 27-30.

[23] FADDOUL R, REVERDY-BRUAS N, BLAYO A. Formulation and screen printing of water based conductive flake silver pastes onto green ceramic tapes for electronic applications[J]. Materials Science and Engineering: B, 2012, 177(13): 1053-1066.

[24] HUSSEIN M T. A novel algorithm to compute all vertex matrices of an interval matrix: a computational approach[J]. International Journal of Computing & Information Sciences, 2005, 2(3): 137-142.

[25] CAI Z. Analysis of structures with uncertain parameters using interval method[J]. Mechanics Research Communications, 2012, 47(3): 24-31.

[26] MOORE R E. Interval analysis [M]. Englewood Cliffs, New Jersey: Prentice-Hall Inc, 1966: 8-13.

[27] HANSEN E. Interval arithmetic in matrix computations, part I[J]. Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis, 1965, 2(2): 308-320.

[28] CSENDES T. Interval analysis and verification of mathematical models [C]// Uncertainties in Environmental Modelling and Consequences for Policy Making. Berlin: Springer, 2009: 79-100.

[29] PARUGGIA M. Sensitivity analysis in practice: a guide to assessing scientific models[J]. Journal of the American Statistical Associ-

ation, 2006, 101(473): 398-399.
[30] SALTELLI A, TARANTOLA S, CAMPOLONGO F, et al. Sensitivity analysis practices: strategies for model-based inference.[J]. Reliability Engineering & System Safety, 2006, 91: 1109-1125.
[31] ROKNE J. Optimal computation of the Bernstein algorithm for the bound of an interval polynomial[J]. Computing, 1982, 28(3): 239-246.
[32] BERLEANT D, XIE L Z, ZHANG J Z. Statool: a tool for distribution envelope determination (DEnv), an interval-based algorithm for arithmetic on random variables[J]. Reliable Computing, 2003, 9(2): 91-108.

Study on uncertainties of graphene tag antenna by screen printing

LI Kai[1, 2], SHEN Zhigang[1, 2], ZHANG Xiaojing[1, 2, *]

(1. School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China;
2. Beijing Key Laboratory for Powder Technology Research & Development, Beijing 100191, China)

Abstract: Graphene tag antenna, owing to its low price and environmental friendliness, is a promising alternative to metal tag antenna. However, graphene antennas inevitably suffer from property uncertainties due to the manufacturing process of screen printing and the properties of graphene inks, making their real read range different from the theoretical one. A unique interval-based computation algorithm is proposed to guide the design of graphene antennas with uncertainties considered, as well as to make the read range particular. Firstly, uncertainty parameters are obtained by analyzing the theoretical formulas of antennas. Second, using an interval analysis approach that includes the interval mathematical modeling and the vertex algorithm, the upper and lower bounds of the read range are determined based on preparation experience, with each uncertainty parameter approximated as an interval set. In the end, experiments are carried out to measure the read range of graphene antennas to verify the proposed algorithm. Results show that the proposed algorithm can calculate the read range of graphene antennas effectively. Thus, this research facilitates the practical use of graphene antennas and provides guidance for uncertainty analysis during manufacture.

Keywords: graphene; tag antenna; read range; interval analysis; uncertainty

Received: 2023-04-04; **Accepted**: 2023-05-15; **Published Online**: 2023-06-06 14:26
URL: link.cnki.net/urlid/11.2625.V.20230605.1445.001
Foundation item: The Fundamental Research Funds for the Central Universities (YWF-22-L-1219)
* **Corresponding author**. E-mail: zhangxiaojing@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0162

离子推力器外壳非预期放电抑制及热特性

李建鹏[1,2]，靳伍银[1,*]，赵以德[2]，代鹏[2]，张兴民[2]

(1. 兰州理工大学 机电工程学院，兰州 730050；2. 兰州空间技术物理研究所 真空技术与物理重点实验室，兰州 730000)

摘 要：离子推力器非预期放电直接影响其工程应用可靠性，为降低推力器外壳与屏栅间的非预期放电频率，通过理论分析和试验的方法明确了非预期放电机制，综合考虑降低非预期放电频率和热设计开展了3种方案的外壳优化设计及热特性分析，搭建了外壳验证及热平衡试验系统。结果表明：高功率下非预期放电频率明显高于低功率，采用单面本色阳极化外壳方案后，5 kW工况下，推力器非预期放电频率从优化前6.90次/h降低到0.70次/h，3 kW工况下，从优化前的3.3次/h降低到0.20次/h，电连接器处测温最高温度为148.5 ℃，满足使用要求，不同外壳方案对热敏感性强的栅极组件温度影响较小，温差范围为5 ℃，推力器温度测试和热分析最大误差小于10 ℃。

关 键 词：离子推力器；外壳；非预期放电；抑制；热设计

中图分类号：TK124

文献标志码：A **文章编号**：1001-5965(2025)03-0865-09

离子电推进因其高比冲、长寿命、工作模式精确可调等优点被广泛应用于航天器姿态控制、轨道转移、深空探测推进、无拖曳控制等任务，是否应用电推进技术已经成为衡量航天器动力系统先进性的标志[1-3]。

随着各国对航天器在轨应用需求的不断增加，长寿命及高可靠已经成为目前世界各国研制离子推力器技术发展的共识，其中，非预期放电击穿现象在离子推力器工程应用中普遍存在，如果不有效抑制，存在降低离子电推进系统可靠性的风险，甚至造成任务失败。

离子推力器放电击穿属于工作中的非预期放电现象，若放电频率过高、能量过大，则可能会严重影响离子推力器工作稳定性和可靠性，国际上因击穿放电现象导致的离子推力器典型故障包括：美国NEXT、NSTAR推力器地面寿命试验因非预期放电导致放电室熄弧、离子推力器异常关机[4-11]，欧洲航天局水星探测离子推力器闪络中断[12-13]、日本隼鸟号深空探测器离子推力器击穿中断等[14-15]，当离子推力器产生持续的电弧时，甚至会造成推力器的整机失效。

Martinez等[16]研究了不同钼和碳基栅极的表面状态对非预期放电击穿的影响程度及影响机制。目前离子推力器非预期放电诱发因素及机制研究主要集中在电极表面粗糙化的场发射增强效应[17]，低气压环境诱发的气体放电[18-19]，等离子体鞘层，类真空弧等离子体源诱发的电击穿等[20]。

目前国外学者主要针对离子光学系统非预期放电及抑制措施进行了研究，对离子推力器外壳与内部组件非预期放电机理分析及消除措施研究鲜见公开报道。近期中国已启动小天体探测任务，探测任务需要实现近地小行星伴飞、附着、取样返回和主带彗星绕飞探测，需要多模式离子推力器作为主推进为小行星探测器从地球向主带彗星转移段提供推力，为小行星探测器与小行星交汇、主带彗星交汇提供制动力，其要求设计寿命达到36 000 h，

收稿日期：2023-04-07；**录用日期**：2023-07-15；**网络出版时间**：2023-08-30 17:30
网络出版地址：link.cnki.net/urlid/11.2625.V.20230830.1434.001
基金项目：国家自然科学基金(62201238)；甘肃省科技计划资助项目(22JR5RA789)
*通信作者. E-mail：ljplzjtedu@163.com

引用格式：李建鹏，靳伍银，赵以德，等. 离子推力器外壳非预期放电抑制及热特性[J]. 北京航空航天大学学报，2025，51(3)：865-873.
LI J P，JIN W Y，ZHAO Y D，et al. Unexpected electric breakdown control and thermal characteristics of ion thruster shell[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：865-873 (in Chinese).

远超国内首次在轨应用的实践二十卫星动力系统LIPS-300H双模式离子推力器21 000 h设计寿命[21-22]。

离子推力器在研制试验中发现非预期放电频次高于预期，通过非预期放电波形监测发现87%占比以上非预期放电发生在屏栅与外壳之间，如果不开展离子推力器外壳内表面非预期放电抑制研究，其频繁的放电会严重影响离子推力器工作稳定性，同时影响其配套单机电源处理单元(power processing unit, PPU)、控制单元稳定可靠输出和离子推力器长寿命的实现，对小行星探测多模式离子推力器而言，甚至影响其多模式的稳定调节。

本文针对离子推力器屏栅和外壳电位间非预期放电展开机理分析，综合考虑非预期放电抑制、热设计、减重等各方面开展不同方案的外壳优化设计，搭建了离子推力器外壳方案验证及热平衡试验平台，确定了离子推力器外壳最优方案，本文的研究为推力器非预期放电抑制、工作稳定性提升和长寿命实现提供了有力支撑，同时对推力器的热设计具有借鉴意义。

1 离子推力器外壳优化设计

1.1 非预期放电现象及原因分析

离子推力器基本原理是依靠气体放电、加速、引出离子产生推力。在大功率工况下，推力器放电室原初电子、离子和二次电子在放电室表面的能量沉积、溅射能量损失、等离子体对放电室表面的辐射共同构成推力器的主要热量来源，为降低推力器大功率工况下的温度，提高推力器的工作性能和可靠性，使推力器电连接器满足一级降额使用温度要求，对推力器铝合金材料柱段外壳的内外表面进行表面处理，目的是提升外壳表面发射率，增加推力器散热能力。通过铝合金材料阳极氧化表面处理的方法实现提升表面发射率的要求，阳极氧化是将铝或铝合金材料零件浸沉于酸性电解液中，在外电流作用下作为阳极，在零件表面形成与基底牢固结合的氧化膜层，这层氧化层具有绝缘性、表面高发射率、耐磨性等特殊性能。小天体离子推力器外壳零件的内表面和外表面开展了阳极氧化的表面处理，其中外观颜色呈现氧化膜层本来的颜色，下文简称双面本色阳极化处理，如图1所示。在离子推力器试验中，离子推力器各项关键参数正常，但试验阶段性检测中发现推力器非预期放电频次高于预期，通过放电波形监测发现多数非预期放电发生在屏栅与外壳之间，经推力器拆解发现柱段外壳上存在多处非预期放电痕迹，如图2所示。

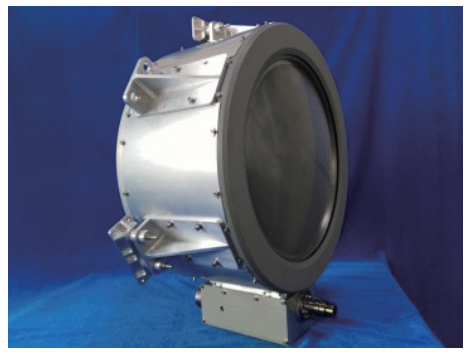


图1 离子推力器实物图

Fig. 1 Ion thruster photo

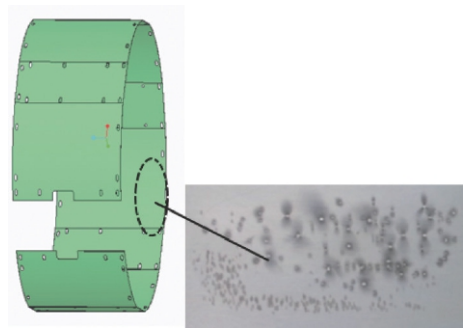


图2 离子推力器外壳非预期放电痕迹

Fig. 2 Traces of unexpected electric breakdown on ion thruster shell

分析原因是：离子推力器供电关系如图3所示，其中，屏栅、阳极、主阴极为高电位电极(一般在1 260 V以上)，加速栅(一般为−200 V)、外壳、中和器(一般近似零电位)为低电位电极。离子推力器外壳与供电高电压电极(屏栅或放电室阳极)之间存在高达1 450 V的电压，离子推力器放电室阳极和外壳通过真空距离绝缘，图4为非完全密封腔体，极少数放电室等离子体电子或中和器发射电子通过放电室或外壳进入外壳与放电室阳极间腔体中，这些电子在本色阳极化外壳绝缘表面积累，积累到一定程度时，在高电压作用下发生绝缘层击穿放电。

1.2 推力器外壳方案设计

基于1.1节非预期放电现象及原因分析可知，减少外壳非预期放电可以采取以下措施：①消除外壳与放电室阳极(屏栅电位)间气体、等离子体环境；②改变外壳表面阳极化处理后不导电状态，但外壳表面状态影响推力器散热能力，因此，要求外壳更改后温度裕度最小元器件电连接器处温度不超过150 ℃，对薄壁多孔结构栅极组件温差影响控制在10 ℃以内，提出如表1所示的3种方案。

方案1的设计思路：为减少电子在本色阳极化外壳绝缘表面积累，取消离子推力器外壳表面绝缘层，对外壳不做表面处理，保持2A12铝材料导电特

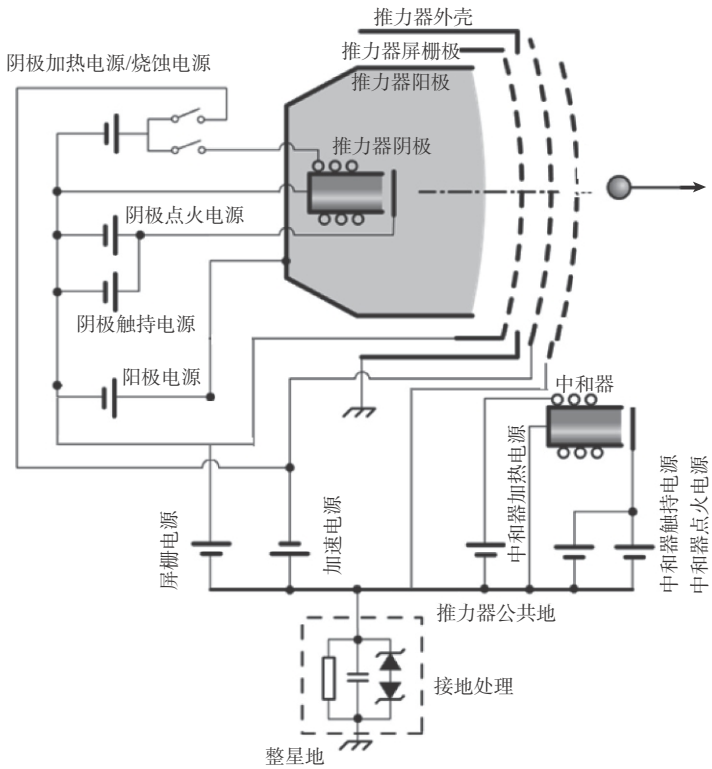


图 3　离子推力器供电

Fig. 3　Power supply for ion thruster

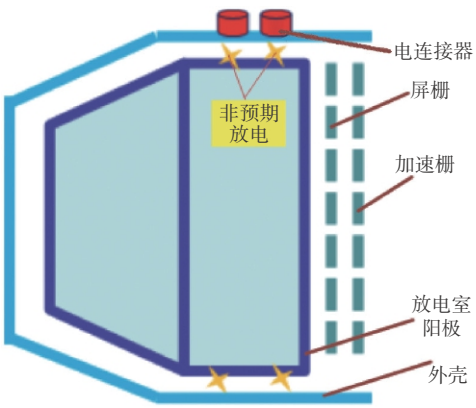


图 4　离子推力器外壳与放电室阳极位置

Fig. 4　Position between ion thruster shell and anode of discharge chamber

表 1　离子推力器不同外壳设计方案

Table 1　Different shell design schemes of ion thruster

方案	材料	表面处理	质量/kg	厚度/mm
原方案	2A12铝	双面本色阳极化	0.26	0.5
方案1	2A12铝	不处理	0.26	0.5
方案2	2A12铝	单面本色阳极化	0.26	0.5
方案3	不锈钢网	不处理	0.36	0.3

性。方案 2 的设计思路: 兼顾离子推力器散热能力, 对靠近放电室阳极一侧的外壳表面不做本色阳极化处理, 保持导电状态, 对外壳外表面做本色阳极化处理。方案 3 的设计思路: 采用 75 μm×51 μm×89 μm 规格不锈钢网保证了泄露到外壳与放电室阳极(屏栅电位)间气体、等离子体可以通过网格间隙通道流动, 降低了电荷积累风险, 同时, 网格结构提高了推力器散热能力。

2　不同外壳方案热特性分析

为研究不同外壳设计方案热设计对推力器影响, 尤其是对离子光学系统薄壁多孔结构栅片和热敏感元器件电连接器的影响, 基于商业软件 ANSYS 开展热特性仿真分析。

2.1　离子推力器热边界条件

离子推力器工作在太空环境, 热量由产热部件向其他部件传导, 最后, 辐射到外太空, 传热方式为热辐射和热传导。热传导通过分子间的相互作用实现热的传递, 傅里叶定律描述了热流密度与温度梯度成正比, 计算式为

$$\frac{Q}{A}=q=-k\nabla T=-k\left(\frac{\partial T}{\partial x}+\frac{\partial T}{\partial y}+\frac{\partial T}{\partial z}\right) \tag{1}$$

式中: Q 为传递的热量; k 为导热系数; T 为温度; A 为面积; q 为热流密度。

热辐射依靠物体表面对外发射可见和不可见的射线传递热量。满足斯特藩-玻尔兹曼定律:

$$\frac{Q}{A}=q=\varepsilon\sigma T^4 \tag{2}$$

式中: σ为斯特藩-玻尔兹曼常数, 等于 5.67×10^{-8} W/($m^2\cdot K^4$); ε为物体表面的发射率。

表 2 为离子推力器大推力模式下工作性能参

数，推力器工作期间的内部热源主要有中和器发射体、阴极发射体、放电室阳极筒、阴极触持极、屏栅、加速栅，基于文献[23]能量沉积数学模型计算得到推力器在不同工作模式下各部件相应位置内部热耗及发射体温度，如表3所示。材料发射率和吸收率是离子推力器散热特性的关键影响因素，表4为通过太阳能光谱反射率仪得到的不同外壳方案内外表面发射率、吸收率试验值，其测试原理是通过样品发射光和参照样品发射光之间的比较来测量样品的发射率，在测试时，光源发出的光照射在样品上，被测样品会吸收部分光线，其余光线会被反射，经光谱仪分析得到发射光谱，同时对参照样品发射率进行校准得到测试值。测量时制作不同外壳方案对应样品，样品尺寸为10 mm×10 mm×5 mm，在样品3个不同位置分别测量3次取平均值。

表2 离子推力器主要性能参数

Table 2 Main performance parameters of ion thruster

推力/mN	比冲/s	流率/(mg·s^{-1})	束电压/V	束电流/A	放电电压/V	放电电流/A	功率/kW
200	3 400	2.551	1 450	3.68	29.0	24.4	5

表3 离子推力器热分布

Table 3 Heat distribution of ion thruster

阳极热耗/W	主阴极触持热耗/W	屏栅热耗/W	加速栅热耗/W	主阴极发射体温度/℃	中和器触持热耗/W	中和器发射体温度/℃
494	29.8	133	29	1 776	32	1 400

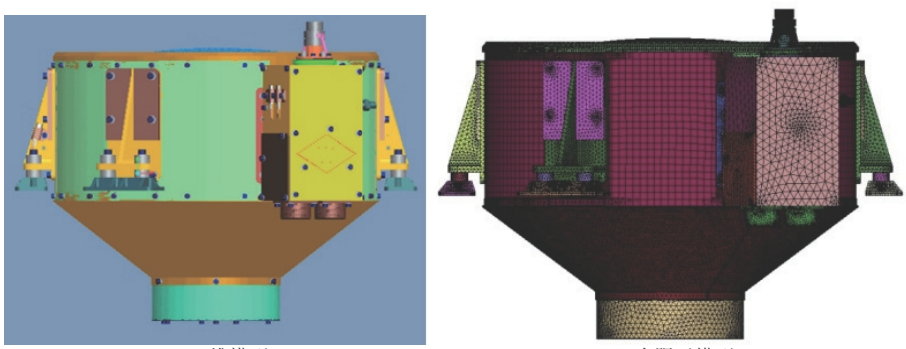
表4 离子推力器不同外壳方案发射率吸收率

Table 4 Emissivity and absorptivity for different shell schemes of ion thruster

方案	外表面发射率	内表面发射率	外表面吸收率	内表面吸收率
原方案	0.67	0.68	0.88	0.87
方案1	0.03	0.03	0.19	0.19
方案2	0.75	0.07	0.35	0.15
方案3	0.26	0.26	0.31	0.34

2.2 离子推力器有限元建模

本文建立了推力器模型，考虑了包括栅极组件、磁钢组件、放电室组件、空心阴极、中和器、外壳在内的1 000多个零件，其三维模型如图5所示。考虑到大部分部件结构复杂，在保证网格质量及计算精度的前提下模型简化主要包括栅极组件、约束部件(主要是标准件)及对热分析而言不必要的几何结构，对具有多孔、薄壁、间距小、形状结构复杂的栅极组件进行均匀化处理，保留拱高高度不变。在不改变约束面的基础上，对约束部件进行简化处理，保证部件之间的连接关系不变，其中，网格数为485 446，辐射面对数为44 473。



(a) 三维模型 (b) 有限元模型

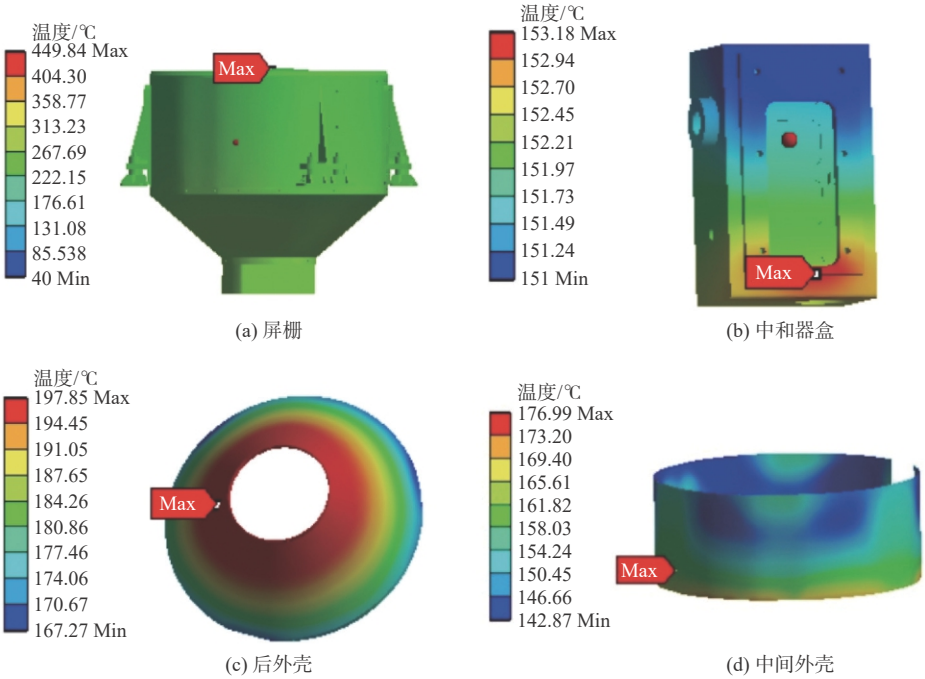
图5 离子推力器三维和有限元模型

Fig. 5 3D model and finite element model of ion thruster

2.3 热分析结果

在建模完成的推力器能量沉积部位加载相应热边界条件，在推力器内部、外部所有受真实固定约束处加载固定载荷，结果如图6所示，受篇幅所限，选择方案2外壳内表面单面本色阳极化方案热分析结果，仅给出个别部件温度分布。

表5为不同外壳方案下离子推力器关键部位热设计仿真结果，可以看出选择不同外壳方案对薄壁、多孔、热敏感性强的栅极温度变化影响很小，温差在5 ℃可接受范围内，不同外壳方案对离子推力器外壳相关部件温度影响较大，以热敏感元器件电连接器安装位置处最高温度为例，其热设计的优



(a) 屏栅　(b) 中和器盒

(c) 后外壳　(d) 中间外壳

图 6　离子推力器在 5 kW 工况下温度应力

Fig. 6　Temperature stress of ion thruster under 5 kW working condition

表 5　离子推力器在 5 kW 工况下热分析仿真结果

Table 5　Thermal analysis simulation results of ion thruster under 5 kW working condition　℃

方案	屏栅温度	加速栅温度	减速栅温度	中和器盒温度	后外壳温度	柱段外壳温度	支撑环温度	阳极温度
原方案	447.1	355.8	246.8	131.4	163.8	155.6	100.8	325.4
方案1	448.2	362.5	263.5	169.5	210.5	196.6	145.5	352.4
方案2	449.8	359.1	252.1	153.18	197.8	176.9	129.5	354.7
方案3	448.1	355.9	249.1	147.2	193.6	176.2	128.8	346.5

劣顺序为: 原方案双面本色阳极化外壳(131.4 ℃), 方案 3 不锈钢网外壳(147.2 ℃), 方案 2 单面本色阳极化外壳(153.18 ℃), 方案 1 铝本色外壳(169.5 ℃)。

3　试验设备及方法

外壳设计验证试验推力器为兰州空间技术物理研究所研制的多模式离子推力器原理样机, 推力器采用四极环型会切场放电室、石墨触持极六硼化镧空心阴极、变孔径凸面三栅极离子光学系统、石墨触持极六硼化镧中和器。试验组成如图 7 所示, 试验在真空技术与物理重点实验室 TS-7B 真空试验系统开展, TS-7B 主舱尺寸为: Φ4.5 m×10 m, 副舱尺寸为: Φ2.0 m×1.5 m, 真空抽气系统由 26 台低温泵主泵和其他螺杆泵组成, 对氙气有效抽速为 25 000 L/s, 氙气流率为 8.3 mg/s 下工作真空度优于 1×10^{-3} Pa, 供电系统调整率≤±0.04%, 工质流率控制精度为±2%。

真空系统配备有束流特性测试诊断设备、推力测量系统、供电和控制系统、供气及控制系统、非预期放电检测四通道示波器、测温传感器。外壳设计验证试验分别在 3 kW 和 5 kW 额定供气、供电参数下开展, 通过四通道示波器+电流探头(30 A)来监测非预期放电电流, 其中, 通道 1～通道 4 依次检测屏栅、加速栅、减速栅和外壳, 非预期放电触发设置源为屏栅, 触发电流设置为 5 A。外壳热设计验证试验共安装了 11 只 E 型热电偶, 热电偶测温精度为±1.5 ℃, 使用温度巡检仪同时对 11 个测温点持续监测。

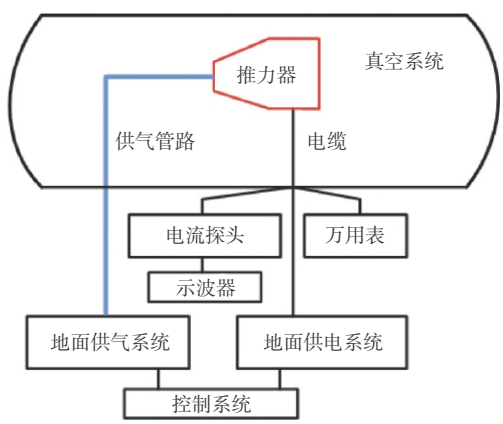


图 7　离子推力器试验系统结构

Fig. 7　Structure of ion thruster test system

4 结果与讨论

4.1 非预期放电验证试验

图8为不同外壳设计方案与放电频率的关系曲线，通过示波器+电流探头检测屏栅和外壳间电流，当电流值超过5 A时判定发生非预期放电击穿，放电波形如图9所示，分别统计60 h内不同外壳方案非预期放电击穿次数，求平均值得到非预期放电频率，可以看到外壳非预期放电，高功率比低功率工况放电频率高，分析原因是高真空条件下，外壳与屏栅（或阳极）电位之间非预期击穿放电初始诱发因素场发射电子产生的场发射电流I_{FE}和电场强度E存在以下关系：

$$\ln\frac{I_{FE}}{E^2}=\ln\left[\frac{1.54\times10^{-6}A_e10^{4.52\phi^{-0.5}}}{\phi}\right]-\frac{2.84\times10^{9}\phi^{1.5}}{E} \tag{3}$$

$$E=\frac{V_T}{l_g} \tag{4}$$

式中：ϕ为材料电子发射功函数；A_e为电子发射面积；V_T为总加速电压；l_E为电极间距。

高功率下电位差较大，高电位差使得电极间的电场强度增大，增加了真空击穿的概率，方案1取消外壳内外表面的阳极化处理后，3 kW工况下，放电频率从3.3次/h下降到0.2次/h，5 kW工况下，放电频率从6.9次/h下降到1.08次/h，说明外壳表面取消绝缘层后可以有效消除电荷在绝缘层积累，从而降低非预期放电频率。

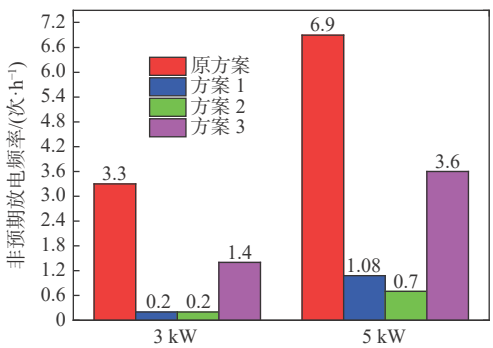


图8 不同外壳设计方案与放电频率的关系

Fig. 8 Relationship between different shell design schemes and electric breakdown frequency

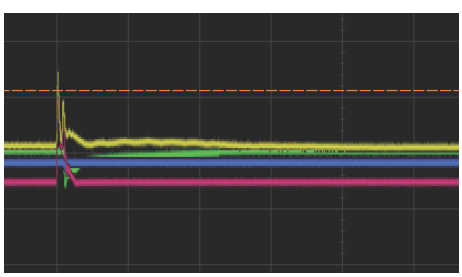


图9 外壳绝缘层击穿诱发非预期放电波形

Fig. 9 Unexpected electric breakdown waveform of shell insulation layer

方案2取消靠近推力器内部外壳表面处理后，低功率工况下，非预期放电频率与方案1相同，高功率工况下，从方案1的1.08次/h下降到0.7次/h，分析原因是外壳内外表面完全取消提高发射率的表面阳极化后，不能及时有效散发推力器放电室等离子体沉积的热量，使得外壳温度升高，增强的场发射和热发射的电子、等离子体将对阳极化处理形成的绝缘层充电，当电场增大到击穿电场阈值时产生电弧，电弧在外壳表面形成了斑点弧光，使得外壳表面出现非预期放电痕迹，同时阳极化表面处理镀层如果膜层不均匀，绝缘层与电极间存在间隙也会造成间隙内场强激增产生低气压击穿放电，因此，阳极化镀膜零件进行真空出气处理也是降低非预期放电频率的有效举措。

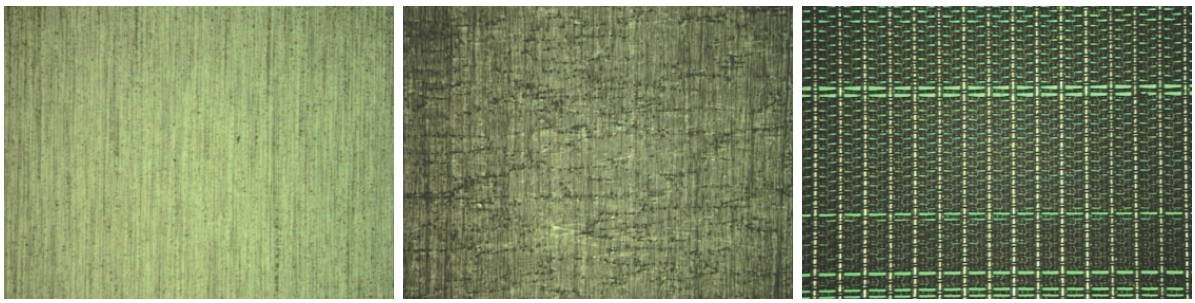
对比原方案和方案3可知，采用一定规格金属网制成的外壳可以降低外壳非预期放电频率，这主要是由于外壳金属网设计一定程度上保证了外壳与放电室阳极间的等离子体输运通道，同时外壳无绝缘层也规避了等离子体积累和绝缘击穿放电，但非预期放电频率下降占比较方案1、方案2小，主要原因是金属网外壳自身金属丝形态存在微凸或锐化结构，这种尖端结构容易增强局部电场强度，降低击穿电压，增大了外壳非预期放电频率。

试验后拆解检查外壳，除方案1存在非预期放电痕迹（见图2），其余方案未发现非预期放电痕迹，结合电流监测数据，说明采取优化方案后无大电流击穿发生，图10为64倍电镜下不同方案外壳试验后表面状态。

4.2 热设计验证试验

按照表1在同一台推力器上开展热平衡温度测试试验，不同外壳方案在5 kW模式下分别连续点火，电连接处的温度曲线如图11所示。

由图11可知，离子推力器在120～170 min达到热平衡，5 kW工况下，不同方案外壳推力器达到热平衡时电连接器处的最高温度如表6所示，前期热设计仿真结果表明在轨太阳辐照条件下推力器电连接器最高温升不超过10 ℃，电连接器设计最高使用温度为210 ℃，一级降额温度为50 ℃，因此，只要地面测试电连接器温度不超过150 ℃就能满足一级降额要求。即优化后方案2单面本色阳极化和方案3金属网方案满足热设计要求，同时测试温度试验值与仿真值误差在10 ℃以内，表明热仿真的有效性，为推力器热设计提供了有效参考，试验过程推力器各项指标正常，稳定工作。



(a) 方案 1　　(b) 方案 2　　(c) 方案 3

图 10　离子推力器不同外壳方案下 60 h 后外壳表面

Fig. 10　Shell surface of ion thruster in different schemes after 60 h

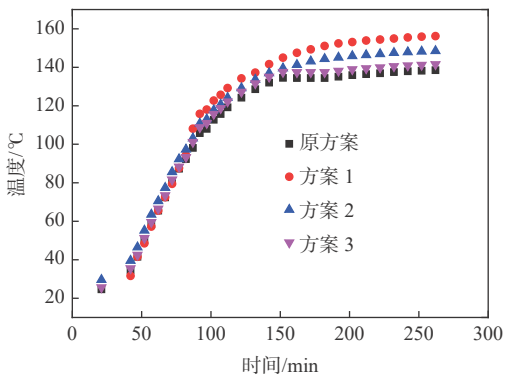


图 11　离子推力器不同外壳方案下电连接器处温度随时间变化

Fig. 11　Variation of temperature at electrical connector of ion thruster in different schemes with time

表 6　离子推力器电连接器热平衡后最高温度

Table 6　Maximum temperature of electrical connector of ion thruster after heat balance　℃

方案	实验温度	仿真温度
原方案	138.2	131.40
方案1	156.2	165.50
方案2	148.5	153.18
方案3	141.8	147.20

5　结　论

本文针对推力器外壳与屏栅电位非预期放电频率高, 放电能量大, 展开方案设计及试验对比研究, 最终选择非预期放电频率低, 满足电连接器使用温度, 重量轻的单面本色阳极化外壳设计方案, 研究得到如下结果或结论:

1) 明确了离子推力器外壳非预期放电机制, 外壳与放电室阳极间为非密封腔体, 放电室泄露的等离子体进入外壳与阳极间腔体中, 在高电压作用下, 本色阳极化外壳表面积累的电荷发生绝缘层击穿。

2) 提出 3 种外壳优化设计方案并开展对比试验研究, 试验发现高功率下非预期放电频率明显高于低功率, 采用单面本色阳极化外壳优化方案, 5 kW 工况下, 非预期放电频率从优化前的 6.9 次/h 降低到 0.70 次/h, 3 kW 工况下, 非预期放电频率从优化前的 3.3 次/h 降低到 0.20 次/h, 温度和外壳表面状态对非预期放电存在影响。

3) 开展了不同外壳方案下的热特性分析, 设计了外壳设计验证及热平衡试验系统, 试验表明离子推力器在 120～170 min 达到热平衡, 选择不同外壳方案对热敏感性强的栅极组件温度变化影响在 5 ℃ 内, 对离子推力器外壳相关部件温度影响较大, 优化后单面本色阳极化和金属网方案满足热设计要求, 同时温度测试试验值与仿真值误差在 10 ℃ 以内, 表明了热仿真的有效性。

参考文献（References）

[1] 李建鹏, 张天平, 赵以德, 等. 阳极电流和屏栅电压对 5kW 离子推力器性能的影响[J]. 推进技术, 2021, 42(6): 1435-1440.
LI J P, ZHANG T P, ZHAO Y D, et al. Effects of anode current and screen grid voltage on performance of 5kW ion thruster[J]. Journal of Propulsion Technology, 2021, 42(6): 1435-1440(in Chinese).

[2] GOEBEL D M, KATZ I. Fundamentals of electric propulsion ion and Hall thrusters[M]. Hoboken: Wiley, 2008.

[3] AHMED L N, CROFTON M W. Surface modification measurements in the T5 ion thruster plume[J]. Journal of Propulsion and Power, 1998, 14(3): 336-347.

[4] RAWLIN V, SOVEY J, HAMLEY J, et al. An ion propulsion system for NASA's deep space missions[C]// Space Technology Conference and Exposition. Reston: AIAA, 1999.

[5] PATTERSON M J, VERHEY T R. A 5kW xenon ion thruster lifetest[C]//International Electric Propulsion Conference Florida. Reston: AIAA, 1990.

[6] OHMICHI W, KUNINAKA H. Performance degradation of a spacecraft electron cyclotron resonance neutralizer and its mitigation[J]. Journal of Propulsion and Power, 2014, 30(5): 1368-1372.

[7] HERMAN D, SOULAS G, PATTERSON M. Performance evaluation of the prototype-model NEXT ion thruster[C]//Proceedings of the 43rd AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit. Reston: AIAA, 2007.

[8] MARTINEZ R A, WILLIAMS J D. Electric field breakdown characteristics of carbon-based ion optics[C]//Proceedings of the 29th

International Electric Propulsion Conference. Reston: AIAA, 2005.

[9] 李建鹏, 靳伍银, 赵以德, 等. 离子推力器外壳非预期放电抑制实验研究[J]. 推进技术, 2023, 44(4): 264-271.
LI J P, JIN W Y, ZHAO Y D, et al. Experimental study of unexpected breakdown reduction of ion thruster housings[J]. Journal of Propulsion Technology, 2023, 44(4): 264-271(in Chinese).

[10] SHASTRY R,HERMAN D A, SOULAS G C, et al. Status of NASA's Evolutionary Xenon Thruster(NEXT) long duration test as of 50 000 h and 900 kg throughput: IEPC 2013-121[R]. Washington, D.C.: NASA, 2013-10-06.

[11] VAN NOORD J L. Lifetime assessment of the next ion thruster[C]//Proceedings of the 43rd AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit. Reston: AIAA, 2007.

[12] WALLACE N, MUNDY D, FEARN D, et al. Evaluation of the performance of the T6 ion thruster[C]//Proceedings of the 35th Joint Propulsion Conference and Exhibit. Reston: AIAA, 1999.

[13] CLARK S, RANDALL P, LEWIS R, et al. Bepicolombo-solar electric propulsion system test and qualification approach[C]//Proceedings of the 36th International Electric Propulsion Conference. Reston: AIAA, 2019.

[14] KUNINAKA H, SHIMIZU Y, YAMADA T, et al. Flight report during two years on HAYABUSA explorer propelled by microwave discharge ion engines[C]//Proceedings of the 41st AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit. Reston: AIAA, 2005.

[15] KUNINAKA H, NISHIYAMA K, FUNAKI I, et al. Powered flight of HAYABUSA in deep space[C]//Proceedings of the 42nd AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit. Reston: AIAA, 2006.

[16] MARTINEZ R, WILLIAMS J, GOEBEL D. Electric field breakdown properties of materials used in ion optics systems[C]//Proceedings of the 42nd AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit. Reston: AIAA, 2006.

[17] SCHWIRZKE F R. Vacuum breakdown on metal surfaces[J]. IEEE Transactions on Plasma Science, 2002, 19(5): 690-696.

[18] SILI E, CAMBRONNE J P, KOLIATENE F. Temperature dependence of electrical breakdown mechanism on the left of the paschen minimum[J]. IEEE Transactions on Plasma Science, 2011, 39(11): 3173-3179.

[19] PEJOVIC M M, RISTIC G S, KARAMARKOVIC J P. Electrical breakdown in low pressure gases[J]. Journal of Physics D: Applied Physics, 2002, 35(10): R91-R103.

[20] ANDERS A. Breakdown of the high-voltage sheath in metal plasma immersion ion implantation[J]. Applied Physics Letters, 2000, 76(1): 28-30.

[21] 李宗良, 高俊, 刘国西, 等. 小行星探测电推进系统方案研究[J]. 深空探测学报, 2018, 5(4): 347-353.
LI Z L, GAO J, LIU G X, et al. Study on the electric propulsion system for asteroid detection[J]. Journal of Deep Space Exploration, 2018, 5(4): 347-353(in Chinese).

[22] 李建鹏, 靳伍银, 赵以德. 加速电压和阳极流率对离子推力器性能的影响[J]. 物理学报, 2022, 71(1): 172-179.
LI J P, JIN W Y, ZHAO Y D. Influence of acceleration grid voltage and anode flow rate on performance of ion thruster[J]. Acta Physica Sinica, 2022, 71(1): 172-179(in Chinese).

[23] 孙明明, 张天平, 陈娟娟, 等. LIPS-200 离子推力器热特性模拟分析研究[J]. 强激光与粒子束, 2014, 26(8): 252-257.
SUN M M, ZHANG T P, CHEN J J, et al. Thermal analysis of LIPS-200 ion thruster[J]. High Power Laser and Particle Beams, 2014, 26(8): 252-257(in Chinese).

Unexpected electric breakdown control and thermal characteristics of ion thruster shell

LI Jianpeng[1, 2], JIN Wuyin[1, *], ZHAO Yide[2], DAI Peng[2], ZHANG Xingmin[2]

(1. School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, China;
2. Science and Technology on Vacuum Technology and Physics Laboratory, Lanzhou Institute of Physics, Lanzhou 730000, China)

Abstract: Unexpected breakdowns of ion thrusters affect their reliability in engineering applications. In order to reduce the frequency of unexpected electric breakdown between shell and screen grid of the thruster, the unexpected breakdown mechanism was clarified through theoretical analysis and experimental methods. By considering the thermal design and the elimination of unexpected electric breakdowns, the shell optimization design and thermal characteristic analysis of three schemes were carried out. A shell scheme verification and heat balance test system was built. The results show that the frequency of unexpected electric breakdowns is significantly higher at high power than at low power. After the single-sided anodized shell on the inner surface is adopted, the frequency of unexpected electric breakdowns of the thruster is reduced from 6.90 times/h before optimization to 0.70 times/h under 5 kW working condition and from 3.3 times/h before optimization to 0.20 times/h under 3 kW working condition. The maximum temperature measured at the electrical connector is 148.5 ℃, meeting the application requirements. The effect of different shell schemes on the temperature change of the thermally sensitive assembly is small, with a temperature difference within 5 ℃, and the maximum error of the thruster temperature test and thermal analysis is less than 10 ℃.

Keywords: ion thruster; shell; unexpected electric breakdown; control; thermal design

Received: 2023-04-07; **Accepted**: 2023-07-15; **Published Online**: 2023-08-30 17:30

URL: link.cnki.net/urlid/11.2625.V.20230830.1434.001

Foundation items: National Natural Science Foundation of China (62201238); Science and Technology Department of Gansu Province Project (22JR5RA789)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0110

THS 功率回流问题分析及其解决方案

刘学武[1,2]，徐向阳[1]，黄河[2]，赵江灵[2]，李凯峰[1]，董鹏[1,*]

(1. 北京航空航天大学 交通科学与工程学院，北京 100191；2. 广州汽车集团股份有限公司汽车工程研究院，广州 511434)

摘 要：分析丰田混合动力系统（THS）在高速工况存在的功率回流现象及其原因，提出一种解决其功率回流问题的混合动力机电耦合系统方案，即在THS的基础上增加一个制动器，不仅消除了功率分流现象，而且增加了一个发动机直驱模式，降低了燃油消耗。仿真对比2种系统在定速工况的油耗，发现增加制动器后，系统的90 km/h等速油耗改善了7.7%，120 km/h等速油耗改善了3.9%。此外，对2种系统分别在2个等速工况下的控制策略进行分析。结果表明：改进后的系统在采用发动机直驱模式时，避免了THS的功率回流问题，提高了系统效率，改善了整车油耗。

关 键 词：混合动力；机电耦合系统；功率分流；等速油耗；发动机直驱模式；功率回流
中图分类号：U469.72
文献标志码：A **文章编号**：1001-5965(2025)03-0874-07

随着油耗法规、排放法规的日趋加严，各汽车企业都在大力开发混合动力汽车。机电耦合系统是混合动力汽车的关键零部件，直接影响了混合动力汽车的油耗和排放。各汽车企业都推出了各具特色的混合动力机电耦合产品，如丰田的THS、本田的i-MMD系统、通用的Voltec系统、广汽的G-MC系统、上汽的EDU系统、长城的柠檬系统、比亚迪的EHS、吉利的雷神系统、奇瑞的鲲鹏系统、东风的马赫系统、科力远的CHS等。开发一套先进的混合动力机电耦合系统，实现优异的动力性能与经济性能，从而提高整车的竞争力、实现更高的市场占有率，这已成为各车企的共同目标。

目前，混合动力汽车机电耦合系统方案主要分为2种类型：①以丰田的THS为代表的功率分流系统；②以本田i-MMD系统为代表的串并联系统。丰田的THS已迭代4代，THS不需要离合器，不需要液压模块，追求至简，以达到低损耗、小型化、轻量化、低成本的目的。但THS在高速低负荷的时候，却存在功率回流，这是THS的一个痛点，赵江灵等[1]对THS的功率回流问题进行了初步分析。张安伟等[2]针对THS的效率优化设计进行了总结，效率优化主要是从零部件出发。如果不改变THS的构型，就只能通过优化速比参数，减小功率回流的区域。如果想要从本质上解决这个痛点，就需要改变THS的构型。文献[3-10]对新型的功率分流的系统设计进行深入研究，提出行星齿轮的构型设计、参数匹配、能量管理等分析方法。

本文提出一种新型的混合动力机电耦合系统，在功率分流的基础上，增加了一个发动机直驱模式，可以在高速时采用发动机直驱模式，避免功率回流，改善整车的油耗。同时，该套混合动力机电耦合系统，也不需要液压模块，仅需要增加一个制动器。可以通过增加少量的成本，解决THS功率回流的问题。

1 功率回流问题分析及解决方案

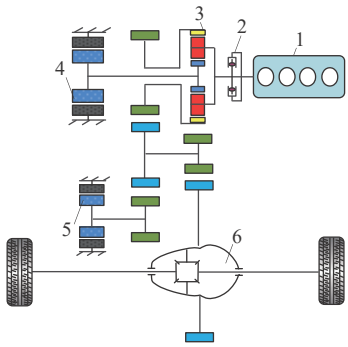
1.1 THS功率回流问题分析

THS原理如图1所示，THS一直坚持不使用离

收稿日期：2023-03-08；录用日期：2023-04-01；网络出版时间：2023-05-23 14:59
网络出版地址：link.cnki.net/urlid/11.2625.V.20230522.1526.001
基金项目：国家自然科学基金(52172352,52322216)
*通信作者. E-mail：dongpengbeihang@163.com

引用格式：刘学武，徐向阳，黄河，等. THS功率回流问题分析及其解决方案[J]. 北京航空航天大学学报，2025，51(3)：874-880.
LIU X W，XU X Y，HUANG H，et al. Analysis of power reflux problem and its solution of THS[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：874-880 (in Chinese).

合器等执行元件，结构简单，也取得了较好的节油效果。因无离合器，系统可以实现的工作模式少，系统只具备纯电动、功率分流和制动能量回收 3 种模式。



1. 发动机；2. 扭转减震器；3. 行星齿轮；4. 发电机；5. 驱动电机；6. 差速器。

图 1　THS 构型

Fig. 1　THS configuration

对于混合动力系统，电机根据不同的情况，有不同的作用。电机的工作状态与电机的旋转方向和扭矩方向有直接的关系。图 2 表明电机正扭矩或负扭矩与电机正向旋转或反向旋转进行不同组合时，电机的驱动和发电状态的不同对应关系。如果电机正向旋转，并施加负扭矩，则其处于发电状态；如果电机反向旋转，并施加负扭矩，则其处于驱动状态。

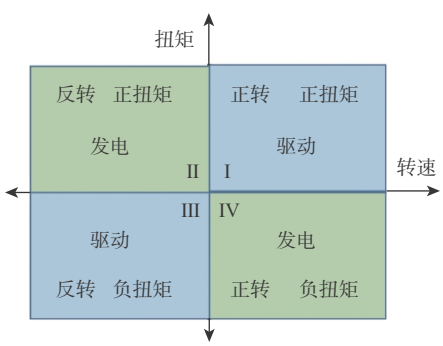


图 2　电机工作状态说明

Fig. 2　Description of the working state of the motor

THS 在中低速正常行驶时，系统的原理如图 3 所示，杠杆图如图 3(a)所示，功率流向如图 3(b)所示。此时，发电机是正向旋转、负扭矩，发电机处于发电状态。驱动电机处于驱动状态，发动机和驱动电机共同驱动车辆行驶。

通过发电机调速、驱动电机调扭，在中低速工况，让发动机工作在发动机效率较优的区间，从而实现较好的节油效果。行星排的扭矩分配关系如式(1)所示，利用行星排将发动机的扭矩分配给发电机和轮端。

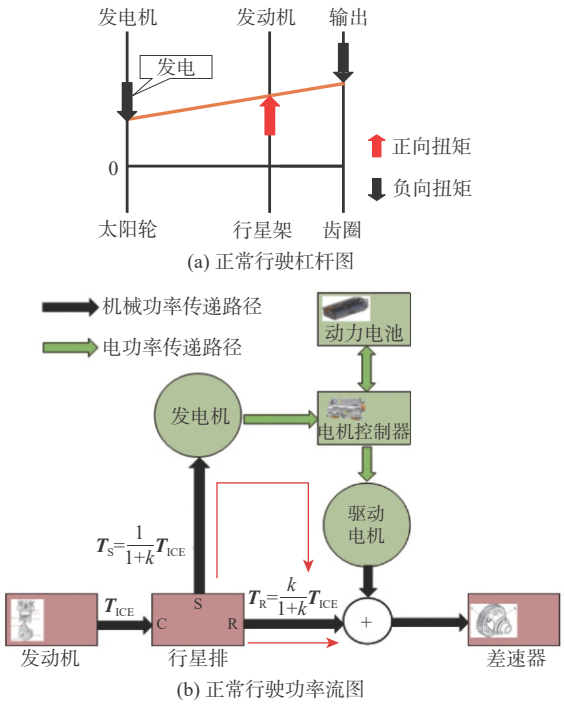


图 3　正常行驶原理

Fig. 3　Normal driving principle

$$\begin{cases} \boldsymbol{T}_C = \boldsymbol{T}_{ICE} \\ \boldsymbol{T}_S = \dfrac{1}{1+k}\boldsymbol{T}_{ICE} \\ \boldsymbol{T}_R = \dfrac{k}{1+k}\boldsymbol{T}_{ICE} \\ \boldsymbol{T}_C = \boldsymbol{T}_{ICE} = \boldsymbol{T}_S + \boldsymbol{T}_R \end{cases} \tag{1}$$

式中：$\boldsymbol{T}_C$ 为行星架的扭矩；$\boldsymbol{T}_{ICE}$ 为发动机的扭矩；$\boldsymbol{T}_S$ 为太阳轮的扭矩；$\boldsymbol{T}_R$ 为齿圈的扭矩；k 为行星排的特征参数。

正常行驶时的系统输出功率计算式为

$$P_{OUT} = P_R + P_{EM2} \tag{2}$$

式中：P_{OUT} 为轮端输出功率；P_R 为行星排输出功率；P_{EM2} 为驱动电机输出功率。

THS 在高速工况，功率回流时的原理如图 4 所示，杠杆图如图 4(a)所示，功率流向如图 4(b)所示。此时，发电机为平衡发动机的动力，让发动机工作在较优的工作点，发电机处于反向旋转、负扭矩，即驱动状态。此时驱动电机处于发电状态，通过驱动电机发电给发电机平衡发动机的动力，从而形成功率回流。

在功率回流时，行星排的扭矩关系如式(1)所示。但因为发电机转速变为负，所以发电机由发电状态变为驱动状态。系统输出功率计算式变为

$$P_{OUT} = P_R - P_{EM2} \tag{3}$$

功率回流的根本原因是发动机采用功率跟随控制，即轮端的功率主要来源于发动机。一直循环的那部分功率主要是为了行星齿轮系统的平衡。

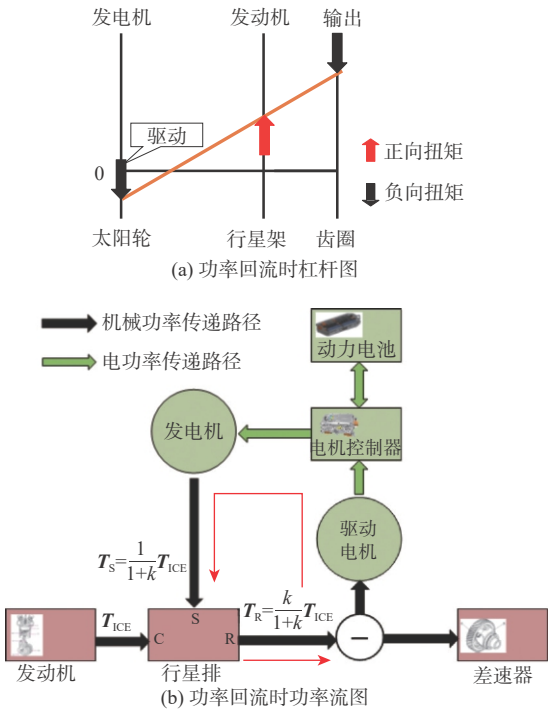


图 4 功率回流时的原理

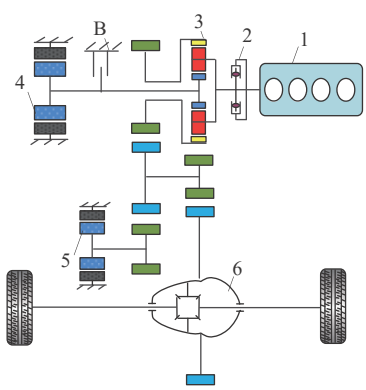
Fig. 4 Principle of power return

因为功率回流会降低系统的效率，增加整车的油耗，因此，有必要研究在高速低扭工况，增加一个新的工作模式，来解决功率回流的问题。

对于高速工况，发动机已经处于较好的工作区间，发动机直驱模式才是最优的工作模式，不需要通过发电机和驱动电机来调速和调扭，避免充电和放电的能量转换。

1.2 一种新型混合动力机电耦合方案

图 5 为一种新型混合动力机电耦合系统方案，该方案在 THS 的基础上增加了一个制动器 B，制动太阳轮，可以实现发动机直驱模式，可以解决高速低扭工况时功率回流的问题。该系统可以实现纯电动、功率分流、发动机直驱、制动能力回收和驻车发电等多个模式。



1. 发动机；2. 扭转减震器；3. 行星齿轮；
4. 发电机；5. 驱动电机；6. 差速器。

图 5 混合动力机电耦合系统构型

Fig. 5 Hybrid electromechanical coupling system configuration

混合动力机电耦合系统的模式控制逻辑如表 1 所示，其中，ICE 为发动机、EM1 为发电机、EM2 为驱动电机、B 为制动器。与 THS 相比，增加一个制动器，系统增加发动机直驱模式，在高速工况，通过制动器锁止太阳轮，发动机的动力通过行星架、齿圈，传递到中间轴和差速器，实现发动机直驱模式。

表 1 混动系统模式控制逻辑

Table 1 Hybrid system mode control logic

驱动模式	ICE	EM1	EM2	B
纯电动	停机	空转	驱动	
功率分流	驱动	发电	驱动	
发动机直驱	驱动	停机	空转	制动
制动回收	停机	空转	发电	
驻车发电	驱动	发电	停机	

2 仿真分析

2.1 仿真参数

为评估增加制动器后的改善效果，本文以某款 SUV 为搭载车型进行仿真分析，仿真参数如表 2 所示。

表 2 整车参数

Table 2 Vehicle parameters

整车整备质量/kg	半载质量/kg	轴距/mm	车胎滚动半径/m	迎风面积/m^2	风阻系数	阻力曲线/N	发动机
2 081.5	2 900	+300	0.37	2.88	0.355	$166.5+1.063v+0.047v^2$	2.0TM

2.2 仿真结果及分析

90 km/h 等速油耗与 120 km/h 等速油耗仿真结果如图 6 所示，THS 90 km/h 等速油耗为 6.51 L/100 km，120 km/h 等速油耗为 9.0 L/100 km。增加制动器后，90 km/h 等速油耗为 6.01 L/100 km，120 km/h 等速油耗为 8.65 L/100 km。

增加制动器后，等速油耗有所改善，节油率如图 7 所示。90 km/h 等速油耗改善了 7.7%，120 km/h 等速油耗改善了 3.9%。

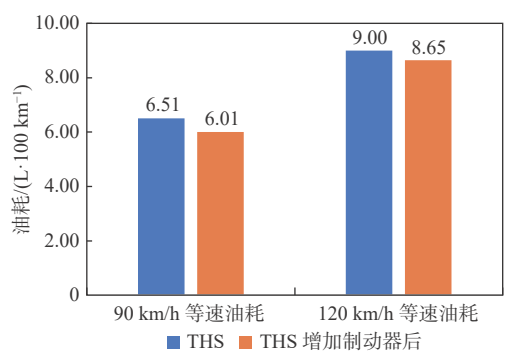


图 6 油耗仿真结果

Fig. 6 Fuel consumption simulation results

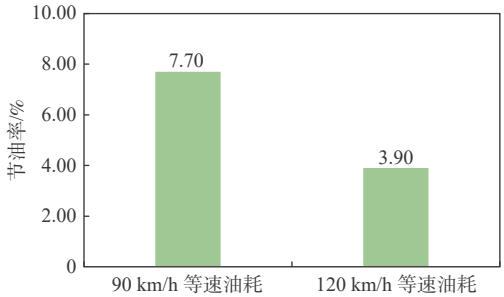


图 7　不同工况节油率

Fig. 7　Fuel saving rate under different working conditions

3　控制策略对比分析

THS 等速行驶时，处于功率分流模式，结果如表 3 所示。按轮端的功率需求，查发动机对应的最优转速点来控制发动机的工作点，控制策略计算式为

$$\begin{cases} \boldsymbol{T}_{\mathrm{ICE}} = \boldsymbol{T}_{\mathrm{ICE}}(\alpha, V, S_{\mathrm{OC}}) \\ \boldsymbol{T}_{\mathrm{EM1}} = \boldsymbol{T}_{\mathrm{EM1}}(\alpha, V, S_{\mathrm{OC}}) \\ \boldsymbol{T}_{\mathrm{EM2}} = \boldsymbol{T}_{\mathrm{EM2}}(\alpha, V, S_{\mathrm{OC}}) \end{cases} \tag{4}$$

式中：$\boldsymbol{T}_{\mathrm{EM1}}$ 为发电机的扭矩；$\boldsymbol{T}_{\mathrm{EM2}}$ 为驱动电机的扭矩；α 为油门开度；V 为车速；S_{OC} 为动力电池电量。

虽然发动机控制在较优的工作点，但是 90 km/h 等速行驶时有 10% 的发动机能量需要经过充电和放电的能量转化，120 km/h 等速行驶时有 20% 的发动机能量需要经过充电和放电的能量转化，这部分的电功率形成功率回流，系统的电损耗较高。

增加制动器后，3 个动力源的工作点如表 4 所示。由功率分流模式改为发动机直驱模式，直接驱动轮端，虽然发动机工作点劣化，90 km/h 等速行驶时热效率变差 2.4%，120 km/h 等速行驶时热效率变差 3.3%。但是 90 km/h 等速行驶时电驱动路径能量占比由 20% 变为 0，120 km/h 等速行驶时电驱动路径能量占比由 10% 变为 0。采用发动机直驱模式，电驱动路径能量占比降低，系统效率有所提高。

表 3　功率分流控制策略分析

Table 3　Power shunt control strategy analysis

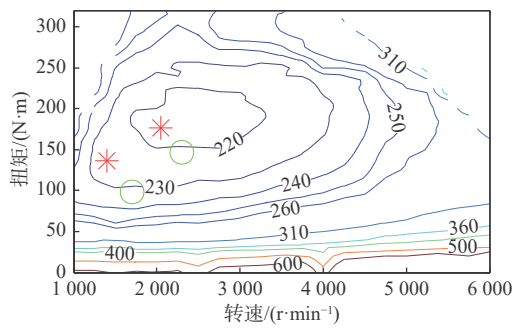
工况	动力源	工况点	备注
90 km/h等速	ICE	136 N·m 1 400 r/min 效率37.4%	“电驱动路径”能量占比约20%
	EM1	−37 N·m −1 100 r/min	
	EM2	−6 N·m 7 700 r/min	
120 km/h等速	ICE	176 N·m 2 050 r/min 效率39.2%	“电驱动路径”能量占比约10%
	EM1	−49 N·m −700 r/min	
	EM2	−6 N·m −10 300 r/min	

表 4　发动机直驱模式控制策略分析

Table 4　Analysis of engine direct drive mode control strategy

工况	动力源	工况点	备注
90 km/h等速	ICE	98 N·m 1 700 r/min 效率36.5%	热效率差2.4%
	EM1	停机	“电驱动路径”能量占比由20%变为0
	EM2	停机	
120 km/h等速	ICE	147 N·m 2 300 r/min 效率37.9%	热效率差3.3%
	EM1	停机	“电驱动路径”能量占比由10%变为0
	EM2	停机	

2 种不同系统的发动机的工作点对比如图 8 所示（等高线图数值代表燃油消耗率，单位：g/(kW·h)）。THS 使用功率分流模式，发动机的工作点对应的效率高于发动机直驱模式的效率。但直驱模式避免了功率回流的问题，提升了传动效率，总效率提升，油耗改善。



✳: THS 工作点　○: THS 增加制动器后的工作点

图 8　不同系统发动机的工作的对比

Fig. 8　Comparison of the work of engines of different systems

4　制动器实施方案

4.1　制动器的结构方案

基于 THS 增加一个锁止机构可以增加发动机直驱模式，在高速时采用发动机直驱模式，能有效避免功率回流问题，可以改善油耗 3.9%～7.7%。目前锁止机构多采用湿式离合器或制动器方案来实现锁止，且锁止机构的锁止和解锁需要液压系统控制。

如果想增加一个直驱模式，就需要增加一个制动器和一个液压模块。虽然增加发动机直驱模式可以改善油耗，但系统增加一个制动器和一个液压模块，总成本增量约上千元，系统成本增量大。

文献 [11-15] 提出离合器的优化设计方法和评价指标，本文提出一种新的制动器解决方案，如图 9 所示。

该制动器作动驱动由液压驱动，锁止后由棘轮

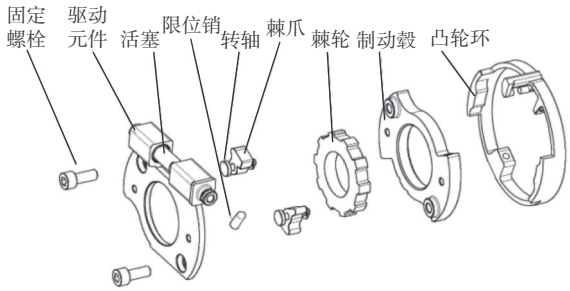


图 9 制动器方案

Fig. 9 Brake scheme

和棘爪所组成的机械结构锁止，可大幅度降低压力需求；同时该制动器无润滑需求，从而大幅简化液压控制系统，有效解决整机布置难题及成本增量的问题。该制动器由固定螺栓通过驱动原件及制动器毂与 THS 壳体固定，驱动原件上有 2 个液压活塞腔体，凸轮环装配与驱动原件与制动器毂中间浮动，与活塞杆相连接。活塞杆 2 个端部分别接入活塞腔，活塞杆左右运动可带动凸轮环转动，凸轮环内侧有凸轮，凸轮与棘爪上侧弧面线接触，从而通过凸轮环转动实现棘爪上下运动，棘爪上下运动可对棘轮实现锁止和分离，锁止及分离动作时，棘爪触点会沿着线段 1 进入，锁止与线段 2 处，分离时触点由线段 2 沿着线段 3 解出。如图 10 所示。

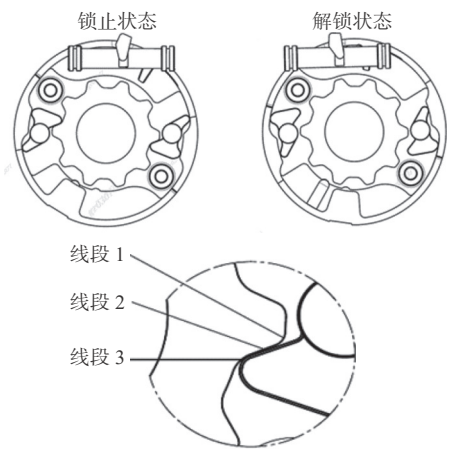


图 10 锁止和分离状态示意图

Fig. 10 Schematic diagram of lock and detach states

4.2 制动器的锁止与解锁控制策略

由于该制动器不需要液压力实现扭矩传递及冷却润滑，所以本文提出基于 THS 原有液压系统进行优化，实现液压缸的驱动。本文液压控制系统如图 11 所示，虚线框内为新增原件，框外为原有系统。

本文通过 THS 冷却润滑回路引出一个油口，外接一个电磁换向阀，实现制动器活塞左右运动的驱动。控制策略上修改为，当电机 2 需要电子油泵（electronic oil pump, EOP）提供冷却流量，同时制动器需求锁止或解锁时，电子油泵的输出流量优先满

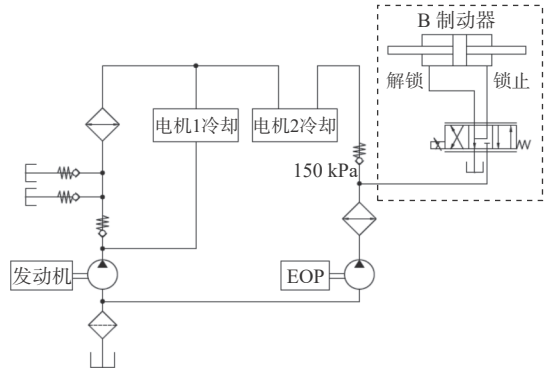


图 11 液压系统原理

Fig. 11 Hydraulic system principle

足制动器的控制。由于制动器锁止或解锁的时间很短，同时机械泵为电机 2 提供冷却流量，因此，对电机 2 的冷却影响不大，待制动器锁止或解锁之后可根据冷却需求通过提高电子油泵的转速增大电机 2 的冷却流量。

4.3 制动器的液压系统仿真

对制动器进行仿真计算，仿真结果如图 12 所示，锁止时间和解锁时间在 0.1 s 内，满足控制要求。

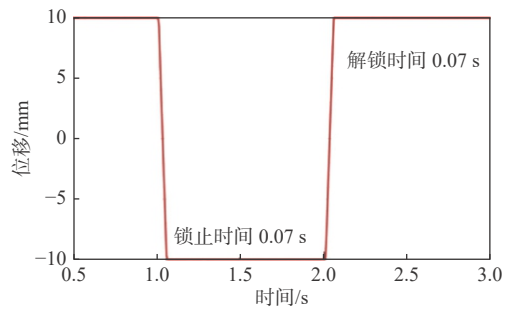


图 12 液压系统仿真

Fig. 12 Hydraulic system simulation

与湿式离合器或电磁离合器等相比，该制动器具有：制动扭矩大（嵌入式刚性传力）；体积紧凑；各个运动都是小幅摆动，无需轴向平移；无附加能耗损失，分离彻底（无拖曳损失，工作状态时无需保持动力，解除后彻底分离）；使用普通液压缸无需行程控制（低压、简单的开关控制，对油品的清洁度要求也较低），成本在百元以下。

此外，在 THS 上布置上述制动器，其外形尺寸可保持基本不变，可通过简单改变发电机轴及把原来的机械油泵去掉改为电子油泵实现，和 THS 有很多的互换通用性。

5 结 论

1）本文分析 THS 在高速工况存在功率回流的原因，提出一种新型的混合动力机电耦合系统方案，该方案在 THS 的基础上增加了一个发动机直驱模式。对比 2 种系统在定速工况的油耗，通过仿真

发现，增加制动器后，90 km/h 等速油耗改善了 7.7%，120 km/h 等速油耗改善了 3.9%。

2）分析 2 个系统在 2 种不同等速工况下的控制策略，新系统采用发动机直驱模式，避免了 THS 的功率回流的问题，提高了系统的效率，改善了整车的油耗。

3）本文提出一种新型制动器的结构方案，经过仿真计算，锁止和解锁时间满足控制要求。

参考文献（References）

[1] 赵江灵, 尚阳, 祖国强, 等. 单行星排式功率分流机电耦合传动方案[J]. 重庆理工大学学报 (自然科学), 2019, 33(12): 38-44.
ZHAO J L, SHANG Y, ZU G Q, et al. Single planetary gear set type of power split transmission concept analysis[J]. Journal of Chongqing University of Technology (Natural Science), 2019, 33(12): 38-44(in Chinese).

[2] 张安伟, 祁宏钟, 赵江灵, 等. 混合动力机电耦合系统效率优化分析[J]. 汽车零部件, 2021(11): 1-8.
ZHANG A W, QI H Z, ZHAO J L, et al. Efficiency optimization analysis of hybrid electromechanical coupling system[J]. Automobile Parts, 2021(11): 1-8(in Chinese).

[3] 王振, 崔亚辉, 许宝卉, 等. 双模式功率分流无级变速器的分析与设计[J]. 机械传动, 2020, 44(5): 49-58.
WANG Z, CUI Y H, XU B H, et al. Analysis and design of dual mode power split continuously variable transmission[J]. Journal of Mechanical Transmission, 2020, 44(5): 49-58(in Chinese).

[4] 李冲. 新型功率分流式动力系统能量管理研究[D]. 重庆: 重庆大学, 2015: 17-22.
LI C. Research on energy management of new power split power system[D]. Chongqing: Chongqing University, 2015: 17-22(in Chinese).

[5] 步曦, 杜爱民, 薛锋. 混合动力汽车用行星齿轮机构的理论研究与仿真分析[J]. 汽车工程, 2006, 28(9): 834-838.
BU X, DU A M, XUE F. Theoretical study and simulation on planetary gear set for hybrid electric vehicle[J]. Automotive Engineering, 2006, 28(9): 834-838(in Chinese).

[6] 伍国强, 秦大同, 胡建军, 等. 混合动力行星齿轮动力传动系统方案及参数匹配研究[J]. 机械设计, 2009, 26(6): 60-65.
WU G Q, QIN D T, HU J J, et al. Research on power transmission systematic scheme and matching of parameters of the hybrid power planetary gear train[J]. Journal of Machine Design, 2009, 26(6): 60-65(in Chinese).

[7] 朱元, 田光宇, 陈全世, 等. 行星齿轮结构的混合动力汽车的系统效率[J]. 汽车工程, 2004, 26(3): 260-265.
ZHU Y, TIAN G Y, CHEN Q S, et al. A research on the system efficiency of a hybrid electric vehicle with power-split powertrain[J]. Automotive Engineering, 2004, 26(3): 260-265(in Chinese).

[8] 蒋科军, 何仁, 束驰, 等. 混合动力汽车动力耦合技术综述[J]. 机械传动, 2015, 39(4): 175-181.
JIANG K J, HE R, SHU C, et al. Review of the power coupling technology for hybrid vehicle[J]. Journal of Mechanical Transmission, 2015, 39(4): 175-181(in Chinese).

[9] 杨亚联, 米娇, 胡晓松, 等. 混合动力汽车行星耦合传动系统的图论建模及动力学分析[J]. 汽车工程, 2015, 37(1): 9-15.
YANG Y L, MI J, HU X S, et al. Graph theory modeling and dynamics analysis on the coupled planetary transmission system of HEV[J]. Automotive Engineering, 2015, 37(1): 9-15(in Chinese).

[10] 张亿. 插电式混合动力汽车动力耦合系统设计与性能研究[D]. 重庆: 重庆大学, 2014: 13-26.
ZHANG Y. Design and performance research of power coupling system for plug-in hybrid electric vehicle[D]. Chongqing: Chongqing University, 2014: 13-26(in Chinese).

[11] 汤鹏翔, 刘艳芳, 王书翰, 等. 多片湿式离合器优化设计[J]. 北京航空航天大学学报, 2010, 36(3): 291-294.
TANG P X, LIU Y F, WANG S H, et al. Optimal design of clutch in automatic transmission[J]. Journal of Beijing University of Aeronautics and Astronautics, 2010, 36(3): 291-294(in Chinese).

[12] 高炳钊, 洪金龙, 陈虹. 汽车传动系统平顺性驾驶品质控制[J]. 控制理论与应用, 2017, 34(7): 849-866.
GAO B Z, HONG J L, CHEN H. Drivability control of automotive drivetrain[J]. Control Theory & Applications, 2017, 34(7): 849-866 (in Chinese).

[13] 侯祥颖, 张钰哲, 张红, 等. 超越离合器研究现状及发展趋势分析[J]. 机械传动, 2021, 45(12): 162-169.
HOU X Y, ZHANG Y Z, ZHANG H, et al. Research status and development tendency analysis of overrunning clutch[J]. Journal of Mechanical Transmission, 2021, 45(12): 162-169(in Chinese).

[14] 王成, 秦大同, 吴邦治, 等. 湿式离合器接合特性仿真与分析[J]. 重庆大学学报, 2020, 43(10): 38-51.
WANG C, QIN D T, WU B Z, et al. Simulation and analysis of wet clutch engagement characteristics[J]. Journal of Chongqing University, 2020, 43(10): 38-51(in Chinese).

[15] 张恒, 李和言, 昌和, 等. 湿式多片离合器摩擦转矩衰减特性分析[J]. 哈尔滨工业大学学报, 2018, 50(7): 94-102.
ZHANG H, LI H Y, CHANG H, et al. Experimental study on attenuation characteristics of friction torque transferred by the wet multi-disc clutch[J]. Journal of Harbin Institute of Technology, 2018, 50(7): 94-102(in Chinese).

Analysis of power reflux problem and its solution of THS

LIU Xuewu[1,2], XU Xiangyang[1], HUANG He[2], ZHAO Jiangling[2], LI Kaifeng[1], DONG Peng[1,*]

(1. School of Transportation Science and Engineering, Beihang University, Beijing 100191, China;
2. GAC Automotive Research & Development Center, Guangzhou 511434, China)

Abstract: After analyzing the occurrence and the reason behind the power reflux in toyota hybrid system (THS) at high speeds, a novel hybrid electromechanical coupling system scheme that can address the power reflux issue is put forth. The system not only solves the power reflux problem, but also adds an engine direct drive mode and reduces fuel consumption by adding a brake based on THS. By comparing the fuel consumption of the two systems under constant speed conditions, it is found through simulation that the fuel consumption at 90 km/h and 120 km/h can improve by 7.7% and 3.9% after adding brakes. Furthermore, an analysis is conducted on the control strategies employed by the two systems under the two constant speed conditions. The findings indicate that the enhanced system circumvents THS's power return issue, enhances system efficiency, and optimizes the vehicle's fuel consumption when it employs the engine direct drive mode.

Keywords: hybrid; electromechanical coupling system; power split; constant-speed fuel consumption; engine direct drive mode; power reflux

Received: 2023-03-08; **Accepted**: 2023-04-01; **Published Online**: 2023-05-23 14:59
URL: link.cnki.net/urlid/11.2625.V.20230522.1526.001
Foundation items: National Natural Science Foundation of China (52172352,52322216)
* **Corresponding author**. E-mail: dongpengbeihang@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0120

超临界压力碳氢燃料在增材制造通道中流动传热的实验研究

蔡磊[1]，肖颖[2]，韩怀志[1,*]，于瑞天[1]，罗文[1]

(1. 四川大学 化学工程学院，成都 610065；2. 中国航发四川燃气涡轮研究院 燃烧试验研究部，绵阳 621022)

摘 要：针对超临界压力正癸烷在增材制造（AM）粗糙通道与机械加工光滑通道中的流动和传热性能差异进行对比实验研究。采用粗糙度仪获得AM通道和光滑通道的表面平均粗糙度分别为11 μm和3 μm。考察不同系统压力（3 MPa、5 MPa和7 MPa）和壁面热流密度（0.29 MW/m^2、0.32 MW/m^{2}和0.35 MW/m^2）工况下，2种通道摩阻系数f和努塞尔特数Nu在不同流量下的变化规律。结果表明：在选定的流量范围内，AM通道的f和Nu皆大于光滑通道，AM通道的f是光滑通道的2.8～3.5倍，Nu是光滑通道的1.7～2.3倍；增大系统压力会使AM通道的f降低，但却使光滑通道增加，而Nu在AM和光滑通道皆降低；随着壁面热流密度增加，AM通道的f和Nu皆降低，而光滑通道皆增加；粗糙通道具有很好的综合传热性能，并且在小流量时效果更好，其综合传热系数为1.2～1.5；提出超临界碳氢燃料在增材制造通道中Nu和f的关联式。

关 键 词：粗糙度；增材制造；流动阻力；换热特性；实验研究

中图分类号：V231.1; TK172

文献标志码：A **文章编号**：1001-5965(2025)03-0881-11

各国都在竭力发展空天动力技术，目标是不断提高飞行器速度，但带来的问题是飞行器将承受极端的高温工作环境，这对其寿命和性能形成了威胁[1]。为解决这个问题，研究人员积极研究和开发各种创新技术。再生冷却技术是其中之一，该技术借助超临界碳氢燃料作为冷却剂，设计了诸多复杂的冷却结构[2-3]，以有效地降低热端部件的温度。而这些复杂的冷却结构，对制造业提出了先进且智能的需求。近年来，增材制造（additive manufacturing，AM）技术在空天动力领域中迅速发展，并逐渐成为促进空天动力装备制造性能提升的重要手段[4]。将AM技术应用到高速飞行器中，可为空天动力领域带来更加高效、节能和可靠的解决方案。

主动冷却技术一般使用超临界碳氢燃料，它比传统燃料更加高效环保，并且具有高温高压下的优越性能，可以有效提高飞行器的速度和推力。然而，超临界碳氢燃料在燃烧过程中会产生大量的高温气体，从而给飞行器的零部件造成过高的热负荷，因此，探索碳氢燃料的流动传热特性至关重要。Pu等[5]探究了航空煤油RP-3在超临界压力下层流和湍流之间的摩阻系数f与努塞尔特数Nu；Zhao等[6]研究了超临界压力下煤油的热传导行为，提出2个修正式，可以提高预测传热行为的精确性。此外，RP-3是一种混合燃料，过高的温度会让其出现堵塞通道现象；而正癸烷是RP-3航空煤油链烷烃中最具代表性的组分，重要理化特性参数都与实际航空煤油非常相近。因此，正癸烷能够还原实际航空煤油的物理与化学过程。黄世璋等[7]建

收稿日期：2023-03-10；录用日期：2023-04-04；网络出版时间：2023-05-05 08:41
网络出版地址：link.cnki.net/urlid/11.2625.V.20230504.1814.002
基金项目：国家自然科学基金(T2441001)；四川省燃气轮机燃烧工程技术研究中心专项基金(JMRH-0000-2024-00001)
* 通信作者. E-mail：hanhz@scu.edu.cn

引用格式：蔡磊，肖颖，韩怀志，等. 超临界压力碳氢燃料在增材制造通道中流动传热的实验研究[J]. 北京航空航天大学学报，2025，51(3)：881-891. CAI L，XIAO Y，HAN H Z，et al. Experimental study on flow and heat transfer of hydrocarbon fuels in additive manufacturing channels at supercritical pressure[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：881-891 (in Chinese).

立了一套模拟超临界正癸烷传热现象的算法。阮波[8]模拟了正癸烷的对流传热与裂解反应，发现正癸烷裂解吸热会起到冷却作用。张磊等[9]研究了多种工况下超临界正癸烷的传热特性，并获得了相应的传热关联式。超临界碳氢燃料在冷却通道中的流动传热特性已有一定规律认识，利用AM技术能够更好地适配飞行器热防护需求，使其具备更好的性能表现。

AM技术通过逐层堆积材料来制造零部件，其可以制造出精细的零件。而对于航空发动机而言，增材制造技术可以制造出更加复杂的一体化零部件，避免了需要进行焊接和铆接等装配工序，解决了连接件受热膨胀不一导致的密封问题。这项技术可以制造传统加工技术难以生产的复杂产品。Cormier等[10]借助AM技术设计了几种高度的圆锥紧凑阵列散热器。Huttunen等[11]借助AM技术设计了由拓扑优化的散热器，明显地降低了热阻。Alteneiji等[12]运用AM技术制造了基于数学理论得到的最小曲面紧凑型换热器，实现了从数值计算到实验测试的突破。这些研究表明AM技术突破了传统机械加工在流动换热中的局限性。需要注意的是，在技术水平限制下，AM制件的表面粗糙多数不可避免，对表面品质的控制是这项技术当前的研究重点之一。在增材制造过程中，就尺寸误差和表面粗糙度而言，建造方向对零件品质有很大影响[13]。由于材料的添加是逐层进行的，一般需要使用大量热量将材料熔化或烧结在一起，当材料颗粒的尺寸与建造层的厚度相同时，颗粒将填补连续层之间的空隙，从而影响实际表面粗糙度[14]。而且，复杂的零件不易进行后处理，确保表面光滑是一项挑战[15]。

从流动的角度来讲，一些研究表明粗糙度会增加流体扰动，从而促进热量传递。Shen等[16]研究水在矩形通道中的流动传热，发现摩阻系数和努塞尔特数与传统理论存在显著差异，他认为这是由粗糙度引起的；Huang等[17]也发现实验结果与理论的不一致。周红秀等[18]研究了表面粗糙度对流动的影响，发现粗糙度可以使壁面附近的流体湍动能增加。对粗糙度在流动传热方面的研究已有一定规律。Liu等[19]研究发现，f与Nu随表面粗糙度的增加而增加；Yuan等[20]的实验表明，粗糙的通道和宏观通道在流动方面表现不同，f随粗糙度增加而明显增加；李虹杨和郑赟[21]的研究发现湍流时粗糙度对壁面温度分布影响大，当粗糙度达到一定值时还会影响流态转变。文献[22-23]的研究中也发现了同样的现象。这些关于粗糙度的研究已有一定规律，即给通道引入粗糙度，既会带来传热的增强也会带来额外的流动损失。而计算机仿真模拟可以帮助解释实验现象。鲁聪达等[24]研究了具有不同形状的随机粗糙度壁面，结果表明流体在近壁面区域和主流区的流速差异较大，近壁面区域流动分离现象明显。Rovenskaya[25]的模拟发现粗糙元附近会出现明显的涡流，在粗糙峰处阻力损失增加；Rawool等[26]的模拟表明，粗糙峰会导致流体分离，从而增加阻力损失。这些针对粗糙表面研究得到的规律，有助于去探索增材制造的粗糙表面的流动传热特性。

目前，有关AM制件粗糙表面流动和传热性能的研究较少，针对一般粗糙表面的实验研究多数围绕水或空气常规流体展开。然而，AM制件的表面粗糙度与机加工表面的结构差别较大，且超临界压力碳氢燃料物性异于常规流体。本项目拟针对超临界压力正癸烷在AM粗糙通道与机械加工光滑通道中的流动和传热性能差异进行对比实验研究，获得增材制造通道的流动和传热特性，为燃料主动冷却技术在AM通道中的工程应用提供设计依据。

1 实验方法

1.1 实验系统

本文实验装置在四川大学燃烧动力学中心江安校区实验室建成。实验系统如图1所示。

氮气瓶用于在实验开始前清理系统；调节背压阀设置目标压力来控制整个系统；使用高压恒流柱塞泵输送一定体积流量的正癸烷，质量流量计获取质量流量数值；正癸烷进入预热段受电源加热升温，调整电源功率使其流入测试段时温度达到目标

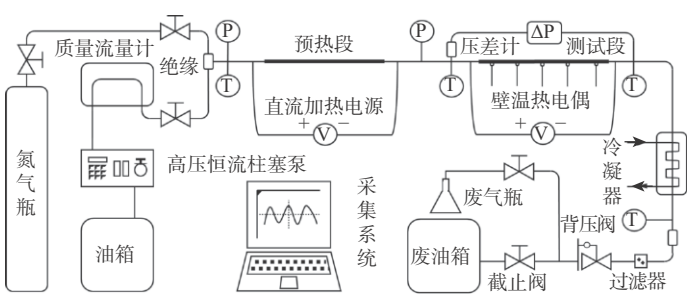


图1 实验系统示意图

Fig. 1 Schematic diagram of experiment setup

值; 测试段的电源作为热边界提供热流量目标值; 最后, 高温正癸烷流入冷凝器降温, 接着流入过滤器, 最终流入废油箱。

1.2　测试段与测点布置

测试段如图 2 所示, 由测试通道、直流电源、k 型热电偶和压差计组成。

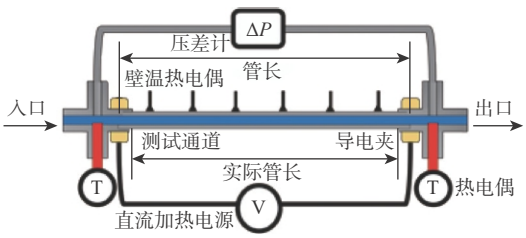


图 2　测试段示意图

Fig. 2　Schematic diagram of test section

2 种测试通道皆为内径为 2 mm、外径为 3 mm 和长度为 200 mm 的圆管。

为获得 2 种通道的传热性能差异, 需要获得测试通道的壁面温度。在测试段的实际长度为 170 mm, 内均匀布置 6 个测传感, 位置分别为 24.3, 48.6, 72.8, 97.1, 121.4, 145.7 mm。

1.3　测量及数据采集系统

实验中需要对测试段进出口压力和温度、测试段压力差、测试段壁面温度等数据实施监测并采集。

流量泵使用前需要校准, 测量一定时间内流出的正癸烷体积来标定校准, 同时称量质量用以校准流量计; 使用 k 型热电偶测量温度, 铠装热电偶测量通道内流体温度, 热电偶丝焊接在壁面上测量壁面温度, 利用加热的方法来校准热电偶, 特别是针对壁面热电偶的校准需要反复调试, 以确保其数据可靠; 使用 HONEYWELL 的 STA725/72S 压差变送器连接测试段两端来测量压力差 Δp; 实验中各测量数据通过电信号导入 HIOKI 的 LR8450 数据采集仪及 U8552 模块, 再接入电脑中配套的 LOGGER UTILITY 软件, 可以实现实时监测各数据并采集。

壁面粗糙度由 TR200 表面粗糙度仪测量获得, 其主要由测量探针获取粗糙度。测量方法示意如图 3 所示。为测量通道内壁面粗糙度, 对通道一端切削一部分, 以便测量探针可以触及内壁。如图 3(b)所示, 设定探针在粗糙表面滑动 5 mm 进行采样, 2 种通道分别测量 3 次并取平均值。结果表明: AM 通道的表面平均粗糙度为 11 μm, 光滑通道的表面平均粗糙度为 3 μm。

1.4　实验工况及参数定义

1.4.1　实验工况

对 2 种通道改变质量流量、系统压力和加热功率, 三组系统压力与两组加热功率, 每个变换对应

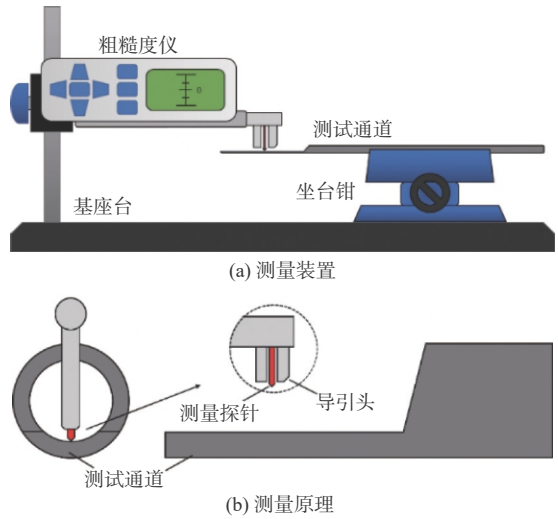


图 3　粗糙度测量示意图

Fig. 3　Schematic diagram of measuring roughness

一组流量变换, 共计 2×5×5=50 组。具体参数如下: 入口温度 T 设定值为 150 ℃; 2 种通道粗糙度分别为 11 μm 和 3 μm; 质量流速 m 选为 1.6, 1.7, 1.8, 1.9, 2.0 g/s; 系统压力 p 选取为 3, 5, 7 MPa; 壁面热流密度 q 设置为 0.29, 0.32, 0.35 MW/m^2。

1.4.2　计算参数

实验涉及摩阻系数 f、努塞尔特数 Nu 和综合传热系数 η。计算定义如下。

摩阻系数计算式为

$$f=\frac{d}{l}\cdot\frac{2\Delta p}{\rho u^2} \tag{1}$$

式中: d 为通道内径; l 为通道长度; Δp 为通道压力差; ρ 为工质密度; u 为流速, 计算式为

$$u=\frac{4m}{1\,000\rho\pi d^2} \tag{2}$$

其中: m 为质量流量。

努塞尔特数 Nu 计算式为

$$Nu=\frac{hd}{\lambda} \tag{3}$$

式中: λ 为工质导热系数; h 为换热系数, 计算式为

$$h=\frac{Q-Q_{\mathrm{loss}}}{\pi d(T_{\mathrm{w}}-T_{\mathrm{f}})} \tag{4}$$

其中: Q 为总热通量; Q_{loss} 为热通量损失; T_{w} 为平均壁面温度; T_{f} 为流体平均温度。

综合传热系数 η 计算式为

$$\eta=\frac{(Nu_{\mathrm{r}}/Nu_{\mathrm{s}})}{(f_{\mathrm{r}}/f_{\mathrm{s}})^{1/3}} \tag{5}$$

式中: Nu_{r} 为粗糙通道努塞尔特数; Nu_{s} 为光滑通道努塞尔特数; f_{r} 为粗糙通道摩阻系数; f_{s} 为光滑通道摩阻系数。

1.5 测量误差分析

1.5.1 直接测量误差

实验仪器本身存在的精度、灵敏度等方面的误差会对实验结果产生一定的影响。对于直接测量获得的参数，其相对误差计算式为

$$\varepsilon(d_a)=\frac{\delta(d_a)}{d_a} \tag{6}$$

式中：d_a 为测量的真实值；$\delta(d_a)$ 为真实值与测量值之差；$\varepsilon(d_a)$ 为测量误差。

直接测量的数据有电压、电流、压力差、压力、温度及质量流量。经计算后各参数误差如表1所示。

表1 直接测量误差

Table 1 Direct measurement error

仪器	直接测量参数	误差/%
直流电源	电压U	2.82
	电流I	1.64
压差计	压力差Δp	1.16
压力传感器	压力p	0.25
温度传感器	温度T	0.4
流量计	质量流量m	0.62

1.5.2 间接测量误差

实验操作中的不规范或不准确的操作，如读数不准、取样不精确等会导致实验误差。这些间接测量误差由误差传递理论计算获得。参数 D 分别由测量值 d_{a1}, d_{a2}, ⋯, d_{an} 获得，D 的误差 δ 可由式(7)计算获得。而本实验主要考察摩阻系数 f 与努塞尔特数 Nu。

$$\delta(D)=\sqrt{\sum_{i=1}^{n}\left[\frac{\partial D}{\partial d_{ai}}\delta(d_{ai})\right]^2} \tag{7}$$

实验中的各种物性参数皆由流体温度 T_f 与压力 p 决定，由 REFPROP 程序计算获得。

努塞尔特数 Nu 由对流换热系数 h 和物性热导率 k 决定：

$$\varepsilon(Nu)=\sqrt{\varepsilon(h)^2+\varepsilon(k)^2} \tag{8}$$

摩阻系数 f 由压力差 Δp、流速 u 和物性密度 ρ 决定：

$$\varepsilon(f)=\sqrt{\varepsilon(\Delta p)^2+\varepsilon(\rho)^2+\varepsilon(u)^2} \tag{9}$$

流速由质量流量 m 和物性密度 ρ 决定：

$$\varepsilon(u)=\sqrt{\varepsilon(m)^2+\varepsilon(\rho)^2} \tag{10}$$

内壁温度 T_i 由热量 Q 与外壁温度 T_o 通过导热微分方程计算获得：

$$\varepsilon(T_i)=\sqrt{\varepsilon(U)^2+\varepsilon(I)^2+\varepsilon(T_o)^2} \tag{11}$$

对流换热系数 h 由热量 Q、内壁温度 T_i 和流体温度 T_f 决定：

$$\varepsilon(h)=\sqrt{\varepsilon(Q)^2+\varepsilon(Q_{loss})^2+\varepsilon(T_i)^2+\varepsilon(T_f)^2} \tag{12}$$

经计算各误差如表2所示。

表2 间接测量误差

Table 2 Indirect measurement error

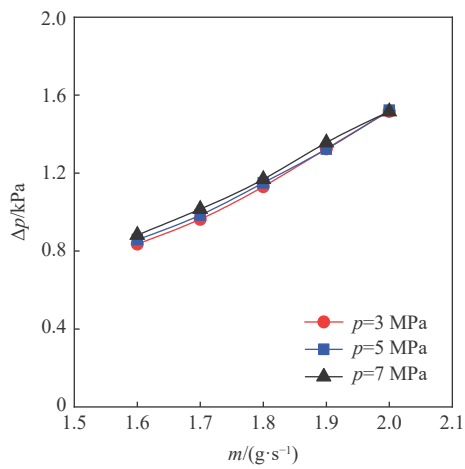
间接测量参数	误差/%
物性	0.47
流速	0.77
内壁温度T_i	3.28
对流换热系数h	6.72
摩阻系数f	1.46
努塞尔特数Nu	6.73

2 实验结果与分析

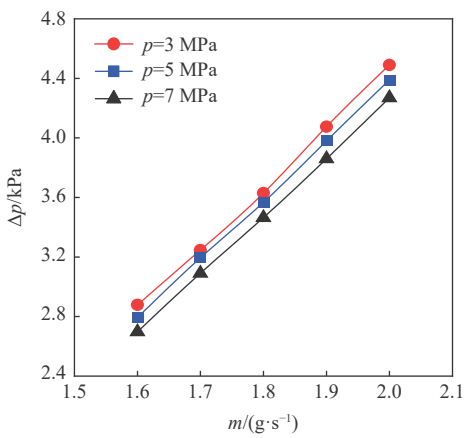
2.1 系统压力对流动和传热性能的影响

2.1.1 流动性能对比分析

壁面热流密度 q 为 0.29 MW/m^2 时，不同系统压力下光滑和粗糙通道的压降 Δp 如图4所示。2种通道的压力损失皆随着流量的增加而增加，光滑通



(a) 光滑通道



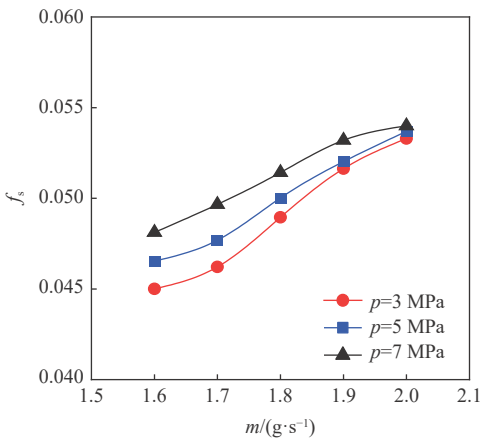
(b) 粗糙通道

图4 不同系统压力下的 Δp

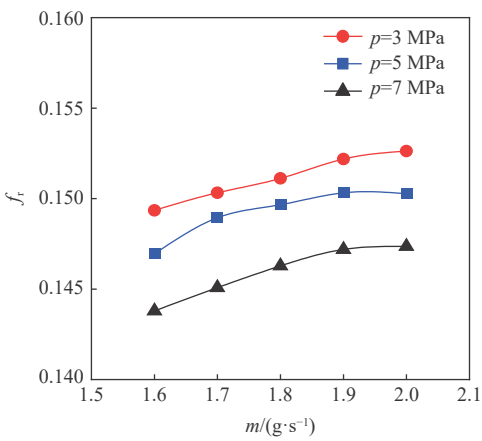
Fig. 4 Δp at various system pressure

道压力损失增加了 0.7 kPa, 而粗糙通道则增加了 1.6 kPa。粗糙通道的压力损失相比于光滑通道整体要高出 195%～205%。随着系统压力的增大, 光滑通道的阻力损失略有增加(最大提高 0.03 kPa), 而粗糙通道明显减小(最大减小 0.1 kPa)。光滑通道内摩擦和阻力还不够大, 增加系统压力会让摩擦阻力明显增加, 流动损失也会增加, 但光滑壁面对流体分子的扰动较小, 故其增加得不够明显。而在粗糙通道内, 虽然增加系统压力让流体黏性增加, 但流体撞击在高低不平的粗糙峰上会在一定程度抵消摩擦阻力, 从而减小流动损失。

壁面热流密度 q 为 0.29 MW/m^2 时, 不同系统压力下光滑和粗糙通道的摩阻系数 f 如图 5 所示。2 种通道的摩阻系数皆随流量增加而增加, 随着流量的增加, 光滑通道的摩阻系数 f_s 增加(最大增加 0.005), 粗糙通道的 f_r 增加较小(最小增加 0.003)。粗糙通道相比于光滑通道整体要高出 172%～231%。随着系统压力的增加, 粗糙通道的 f_r 最大降低了 2.6%, 而光滑通道的 f_s 规律与其压力损失 Δp 相似。摩擦阻力会随流量增大而增加, 而系统压力升高会引起密度升高, 根据质量守恒方程, 流速会相应的降低, 因此, 不同系统压力之间 f 的差距将会增大而清晰。



(a) 光滑通道



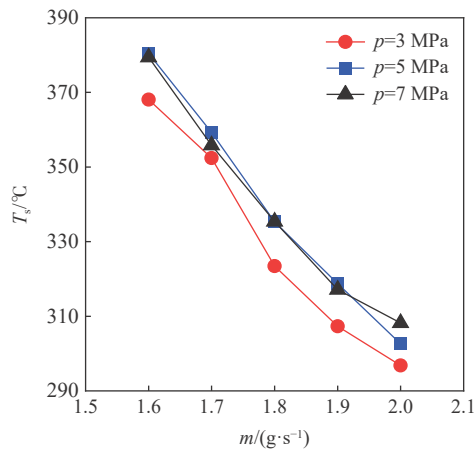
(b) 粗糙通道

图 5　不同系统压力下的摩擦系数

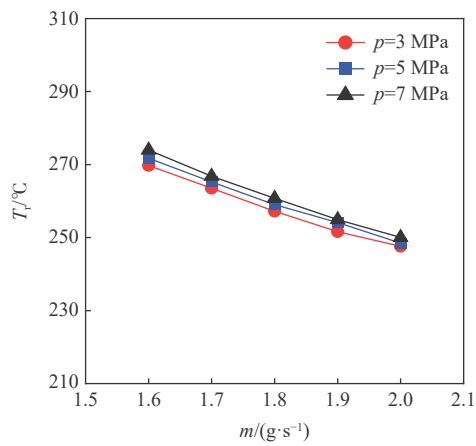
Fig. 5　Friction factor f change at various system pressure

2.1.2　传热性能对比分析

壁面热流密度 q 为 0.29 MW/m^2 时, 不同系统压力下光滑和粗糙通道的平均壁温 T_o 如图 6 所示。随着流量增加, 2 种通道的平均壁温降低, 粗糙通道的平均壁温 T_r 降低了 22 ℃, 而光滑通道的降低了 50 ℃。流量为 1.6 g/s 时, 2 种通道平均壁温相差 83 ℃。在粗糙通道的边界层内, 过高的粗糙峰形成了固体导热捷径, 会更直接传递主流区的热量, 并且流体在粗糙峰表面上的冲击也会促进热量传递。流量为 2.0 g/s 时, 2 种通道平均壁温差距缩小到 34 ℃, 粗糙通道的平均壁温 T_r 相比光滑通道整体要降低 8%～25%, 表明粗糙表面的冷却效果随流量的增大而变弱。这是因为随流量增加, 热量传递随之增多, 光滑通道仍有很大潜力来改善传热效果; 流量变化初始粗糙通道本身就已增强传热而降低了壁温, 随着流量提高, 粗糙通道冷却效果的提升便不再明显。随着系统压力提高, 光滑通道的平均壁温 T_s 增加明显(最大增加 11 ℃), 粗糙通道的



(a) 光滑通道



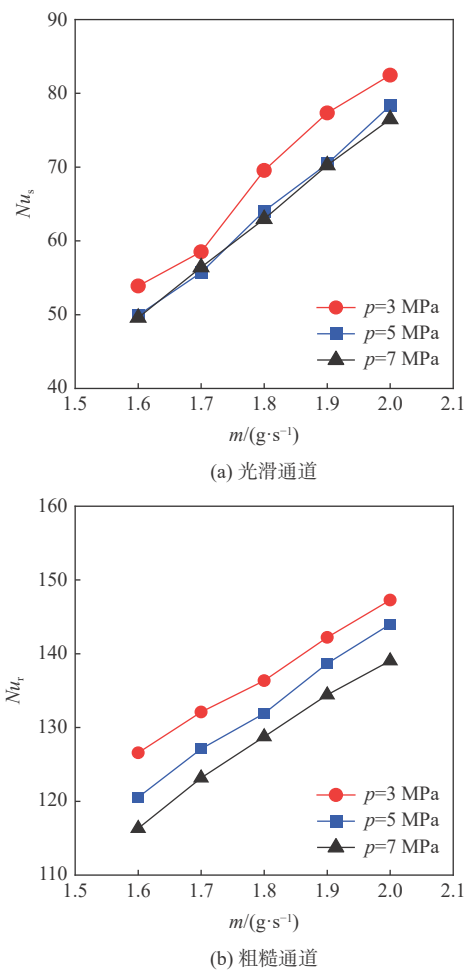
(b) 粗糙通道

图 6　不同系统压力下的平均壁温

Fig. 6　Average wall temperature at various system pressure

几乎不变(最大增加 4 ℃)。这是因为压力提高导致正癸烷黏性增加,阻碍光滑通道内流体的流动,从而降低传热效果,故壁温会增加;粗糙通道亦是如此。但不同系统压力间差距变化并不大,这也是由于其提升潜力不足而造成的。

壁面热流密度 q 为 0.29 MW/m^2 时,不同压力下光滑和粗糙通道的努塞尔特数 Nu 如图 7 所示。2 通道的努塞尔特数 Nu 数随流量增长,Nu_r 最大增加了 28,Nu_s 最大增加了 15,增长趋势相差不大。同摩阻系数的表现一样,粗糙表面带来明显的换热强化,粗糙通道的努塞尔特数 Nu_r 相比光滑通道的 Nu_s 整体要高出 115%～146%。随着系统压力提高,粗糙通道的努塞尔特数 Nu_r 降低(最大降低 4.7),光滑通道的 Nu_s 也是如此(最大降低 7)。光滑通道内流动损失随系统压力增加,流体与壁面之间的热量传递便会增强,这样壁面温度上升,导致通道内的流体温度也会上升,而这 2 种之间的差距并没有明显增大,因此 Nu_s 在不同系统压力之间的差距没有拉开。而在粗糙通道内,首先,流动损失随系统压力增大而减小,这是不利于流体与壁面之间进行热量传递的。其次,系统压力增加还会导致流体的黏性和密度增加,近壁面处的涡流便会消耗更多时间来脱离粗糙壁面,这也是不利于传热的。另外,流体温度上升还会使其导热率降低,在这些影响的作用之下,粗糙通道的传热效果会随系统压力变化的明显效果会比光滑通道的变化更明显。



(a) 光滑通道

(b) 粗糙通道

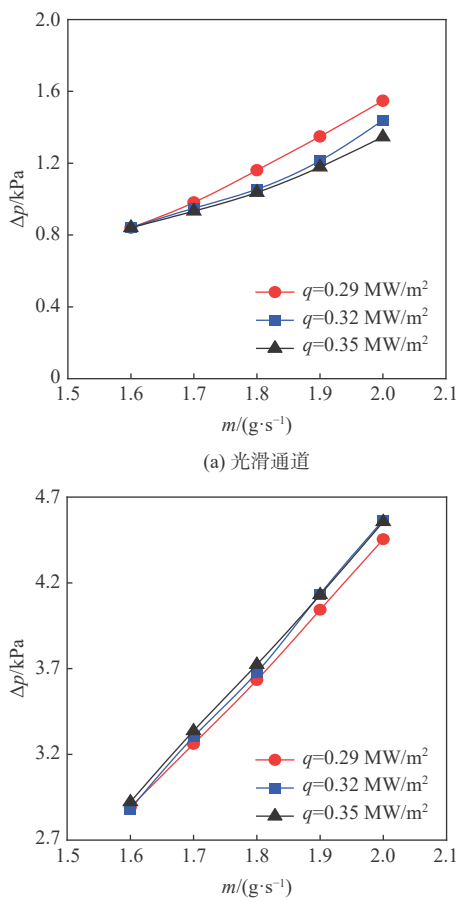
图 7 不同系统压力下的 Nu

Fig. 7 Nu at various system pressure

2.2 壁面热流密度对流动和传热性能的影响

2.2.1 流动性能对比分析

压力为 3 MPa 时,不同壁面热流密度下光滑和粗糙通道的压力损失如图 8 所示。在 1.6 g/s 小流量时,光滑通道不同加热功率间的压力损失几乎一致。随着热流密度提高,光滑通道的总压力损失减小(最大增加 0.13 kPa),而粗糙通道的总压力增加(最大增加 0.1 kPa),这是因为粗糙通道中额外增加的沿程阻力损失集中在壁面附近,粗糙通道本就比光滑通道的沿程阻力损失大,增大热流密度会让壁面温度提升,但会抑制这部分沿程压力的损失;而光滑通道的沿程阻力损失处于初步形成阶段,温度提升促使光滑通道的这部分阻力损失进一步增大。光滑通道内壁面粗糙度较小,在增加热流密度的情况下,导致了流体内部的温度上升,从而流体



(a) 光滑通道

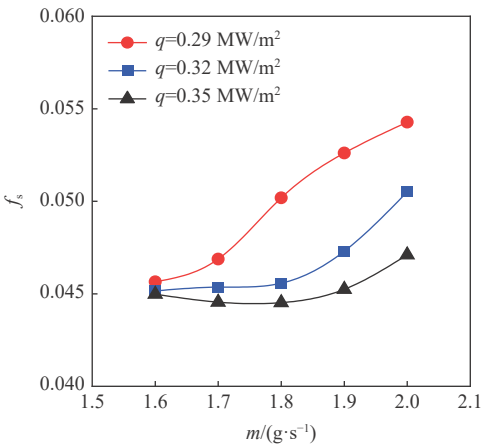
(b) 粗糙通道

图 8 不同热流密度下的 Δp

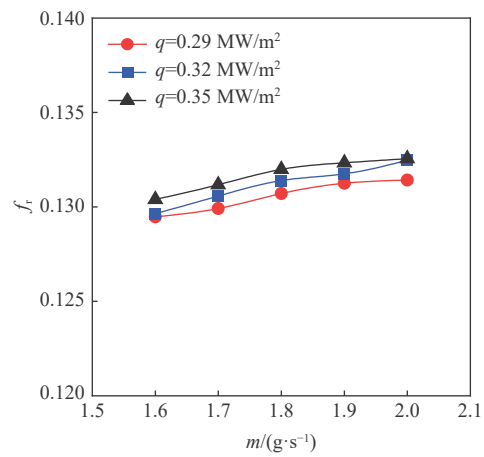
Fig. 8 Δp at various heat flux densities

黏性减小, 削弱了随流量增大的阻力。因此, 光滑通道的损失减少, 流动阻力降低。粗糙通道内流体分子与管道表面的摩擦阻力是主要的阻力, 壁面热流密度影响流体与管道表面的温度差和热传递强度, 虽然温度上升会使黏性减小, 但流体与粗糙壁面的阻力仍有增加, 粗糙通道内部流动已足够复杂, 增加壁面热流密度对流动影响不大。

压力为 3 MPa 时, 不同壁面热流密度下光滑和粗糙通道的摩阻系数如图 9 所示。摩阻系数 f 与压力差 Δp 成正比, 随着热流密度增加, 光滑通道的摩阻系数 f_s 降低明显(平均减小 5%), 而粗糙通道摩阻系数 f_r 增加幅度较弱(最大增加 2%)。光滑通道不同系统压力之间 f_s 的差距比粗糙通道的明显。流体温度升高导致流体黏性和密度降低, 光滑壁面对流体的摩擦作用小, 摩阻系数明显下降。而粗糙通道内, 流体温度升高导致其黏性降低, 会使流体微团受壁面摩擦作用减小, 摩阻系数便会轻微增加。



(a) 光滑通道



(b) 粗糙通道

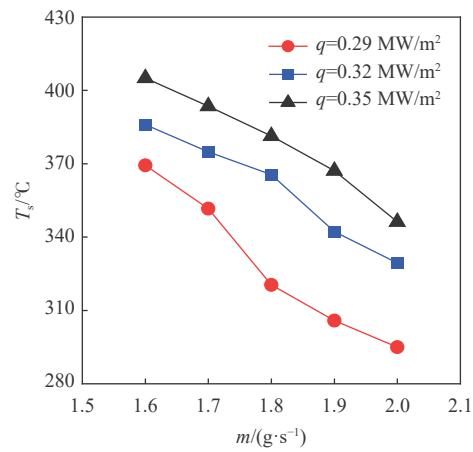
图 9　不同热流密度下的摩阻系数

Fig. 9　Friction factor at various heat flux densities

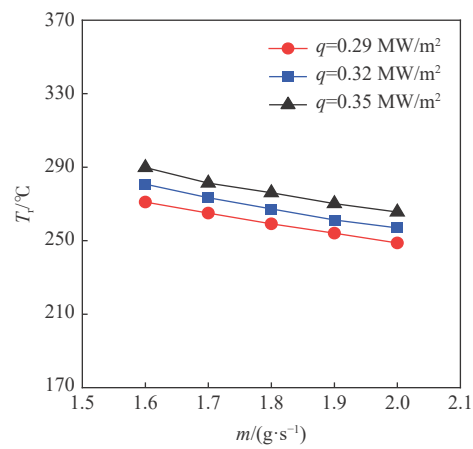
2.2.2　传热性能对比分析

压力为 3 MPa 时, 不同壁面热流密度下光滑和粗糙通道的平均壁温 T 如图 10 所示。随着热流密度提高, 两通道的平均壁温 T 有规律地增加, 光滑通道最大增加了 45 ℃, 而粗糙通道的最大增加 5 ℃。这是因为粗糙通道相比光滑通道本就强化了热量传递。首先, 粗糙通道内崎岖不平的粗糙峰直接接触到主流区的流体强化了热量传递; 其次, 流体冲撞粗糙峰之后会增加热流密度, 粗糙壁面则会传递更多热量给流体, 使壁面得到冷却从而温度上升较慢。而光滑通道内, 壁面与流体之间的热量传递没有粗糙通道的多, 故其壁面温度随热流密度增加有明显增加。

压力为 3 MPa 时, 不同壁面热流密度下的光滑和粗糙通道的努塞尔特数 Nu 如图 11 所示。随着热流密度增加, 光滑通道的 Nu_s 最大增加 13, 而粗糙通道的 Nu_r 最大增加 5.3。而热流密度对 Nu 的影响同其对壁温的影响相似, Nu_r 在各流量下随热流密度增加的步调相差无几。这是因为光滑通道与流体之间的传热由边界层内的纯流体导热主宰, 与粗糙通道内边界层含有粗糙峰固体导热的效果相差很大。并且, 壁面热流密度增加致使壁面附近流



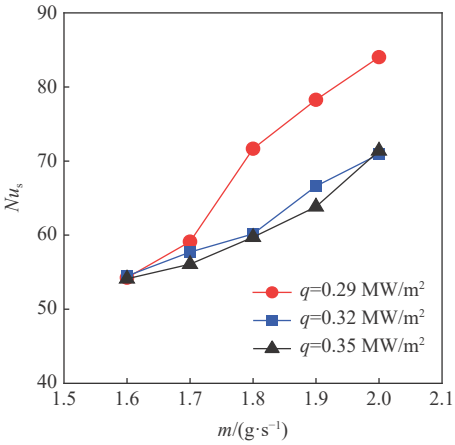
(a) 光滑通道



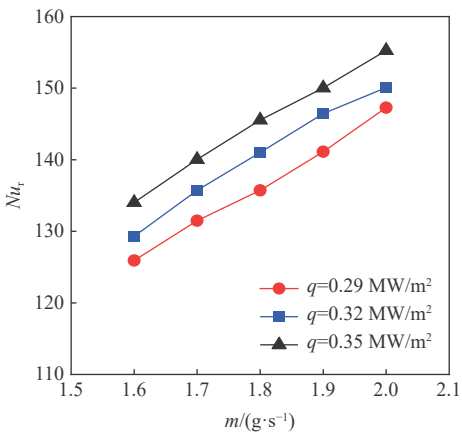
(b) 粗糙通道

图 10　不同热流密度下的平均壁温

Fig. 10　Average wall temperature at various heat flux densities



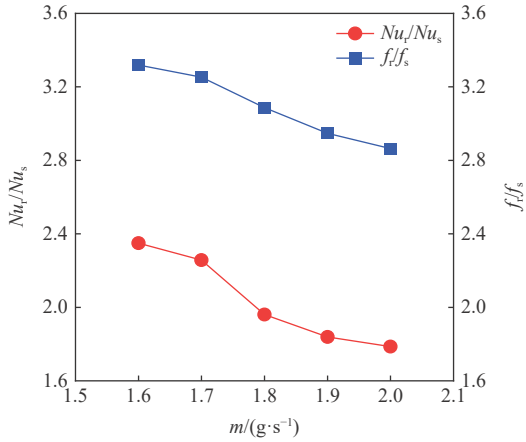
(a) 光滑通道



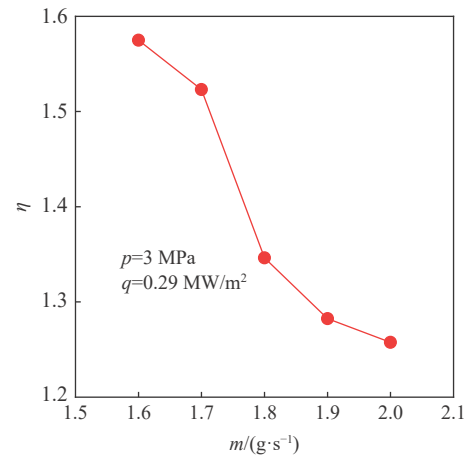
(b) 粗糙通道

图 11 不同热流密度下的 *Nu*

Fig. 11 *Nu* at various heat flux densities



(a) 流动与传热



(b) 综合传热系数 η

图 12 综合传热效果对比

Fig. 12 Comparison of integrated heat transfer characteristics

体密度降低，光滑通道内边界层内导热效率也会降低，因此，光滑通道的努塞尔特数随热流密度增加而减小。而在粗糙通道内，壁面热流密度的增加会通过粗糙峰直接与主流区的流体交换热量，并且，边界层内的流动比光滑通道的混乱得多，因为粗糙壁面会干扰流动，那些涡流与二次流会反复冲撞粗糙表面从而促进热量传递，流体温度上升导致其密度下降很有可能使冲撞更加活跃，传热效果便会随热流密度增加而增强。

2.3 综合效果

两通道摩阻系数 f 与努塞尔特数 *Nu* 的对比如图 12(a)所示。粗糙通道对比光滑通道虽然提高了2.3倍传热效果，但也带来了3.3倍的压力差。在流量变换初始粗糙通道这2种增强效果最为明显，但随着流量增加，光滑通道的提升空间比粗糙通道的大，故粗糙通道的强化作用随流量增加变弱了。图 12(b)表示粗糙通道的综合传热系数随流量变化，流量在1.6 g/s时，综合传热系数 η 为1.57，随流量增大而降低到1.26，这说明在选定的流量范围内由AM技术制造的粗糙通道的综合传热效果总是强于机加工光滑通道，特别是在小流量时的强化效果较为明显。

2.4 超临界正癸烷在AM通道内的摩阻系数与努塞尔特数经验关联式

本实验探究了系统压力与热流密度对AM通道与机械加工通道的流动和传热性能之间的关系。然而实际工况通常是复杂的，因为它受到多种因素的影响。为更好地理解实验数据并推导出有用的设计参考，需使用经验关联式来描述流动和传热性能与其他重要变量之间的关系，对于空天动力领域，则对热环境和动力亏损的评估更具有参考意义。因此，本研究对AM通道拟合了努塞尔特数和摩阻系数的经验关联式。

关于AM通道的努塞尔特数关联式为

$$Nu = 0.555\,8 \times Re_{in}^{0.819\,7} \times Pr^{-0.766\,8} \tag{13}$$

摩阻系数关联式为

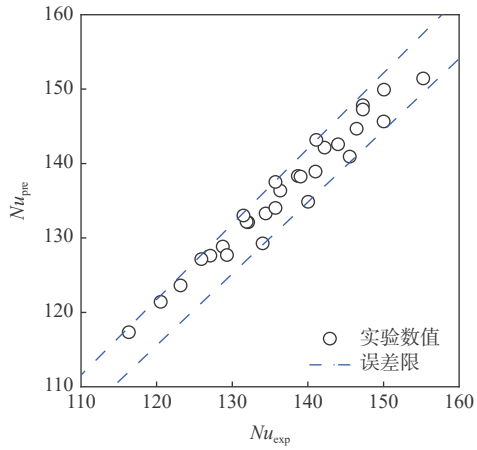
$$f = 0.075\,73 \times Re_{in}^{0.133\,2} \times Pr^{-0.325\,0} \tag{14}$$

式中：Re_{in} 为通道入口雷诺数；Pr 为流体工质的普朗特数。

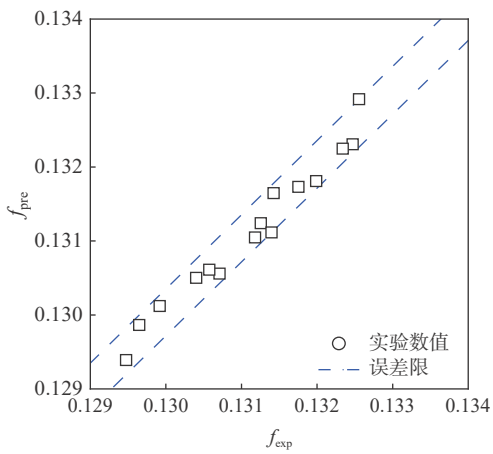
适用范围：AM通道粗糙度 Ra=11 μm、3 000<

Re_{in}<5 000、5<Pr<7。

图 13 为超临界碳氢燃料在 AM 通道内努塞尔特数和摩阻系数的实验数值与预测值之间的差异，预测值 Nu_{pre} 与实验值 Nu_{exp} 的差异在 5.2% 之内，预测值 f_{pre} 与实验值 f_{exp} 的差异在 0.3% 之内。



(a) 努塞尔特数



(b) 摩阻系数

图 13　超临界正癸烷在 AM 通道中努塞尔特数和摩阻系数预测值与实验值对比

Fig. 13　The Nu and f in predicted of supercritical n-decane in the AM channel compared with experimental data

3　结　论

本文通过实验方法对比了超临界压力下正癸烷在 AM 粗糙通道与普通光滑通道中的流动换热对比，得到了不同工况下 2 种通道的摩阻系数 f 与努塞尔特数 Nu，对各工况下流动换热特性规律分析，得到以下结论：

1）与光滑通道相比，粗糙通道的传热性能有所增强，努塞尔特数 Nu 最多增加了 146%，但带来额外沿程阻力损失，摩阻系数 f 最多增加了 230%。粗糙通道的综合传热性能随着流量的增加而降低，最高达到 1.57。

2）随着系统压力增加，可以显著减小粗糙通道中的沿程损失，摩阻系数 f 最多可降低 2.6%，Nu 值最多可降低 4.7%，表明增加系统压力对粗糙通道的流动和传热性能可起到抑制作用。相比之下，在光滑通道中增加系统压力会导致 f 增加。

3）随着壁面热流密度增加，粗糙通道的摩阻系数 f 增加幅度较弱（最大增加 2%），Nu 最多提升 4%，表明增加壁面热流密度对粗糙通道的流动和传热性能可起促进作用。

4）针对 AM 通道的努塞尔特数和摩阻系数关联式进行了拟合。

参考文献（References）

[1] NOURIN F N, AMANO R S. Review of gas turbine internal cooling improvement technology[J]. Journal of Energy Resources Technology, 2020, 143(8): 080801-080809.

[2] SINGH J, ALEJANDRO M C, NIGAM K D P. Process intensification for compact and micro heat exchangers through innovative technologies: a Review[J]. Industrial & Engineering Chemistry Research, 2019, 58(31): 13819-13847.

[3] HASAN M I, RAGEB A A, YAGHOUBI M, et al. Influence of channel geometry on the performance of a counter flow microchannel heat exchanger[J]. International Journal of Thermal Sciences, 2009, 48(8): 1607-1618.

[4] 刘洋, 周建平, 张晓天. 增材制造技术在载人航天工程中的应用与展望[J]. 北京航空航天大学学报, 2023, 49(1): 83-91.
LIU Y, ZHOU J P, ZHANG X T. Application and prospect of additive manufacturing technology in manned space engineering[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(1): 83-91(in Chinese).

[5] PU H, LI S, DONG M, et al. Convective heat transfer and flow resistance characteristics of supercritical pressure hydrocarbon fuel in a horizontal rectangular mini-channel[J]. Experimental Thermal and Fluid Science, 2019(108): 39-53.

[6] ZHAO W, SONG Z, LI H, et al. Research on heat transfer characteristics of kerosene at supercritical pressure in circular tubes[J]. Experimental Thermal and Fluid Science, 2018(96): 507-515.

[7] 黄世璋, 阮波, 高效伟, 等. 超临界压力下碳氢燃料裂解与流动传热模拟的快速算法[J]. 航空学报, 2018, 39(4): 121650.
HUANG S Z, RUAN B, GAO X W, et al. A fast algorithm for simulating hydrocarbon fuel heat transfer with endothermic pyrolysis under supercritical pressures[J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(4): 121650(in Chinese).

[8] 阮波. 超临界压力下正癸烷裂解吸热和对流传热现象的数值模拟研究[D]. 浙江: 浙江大学, 2013.
RUAN B. A numerical simulation study on the endothermic cracking and convective heat transfer phenomena of n-decane under supercritical pressure[D]. Zhejiang: Zhejiang University, 2013(in Chinses).

[9] 张磊, 乐嘉陵, 张若凌, 等. 超临界压力下湍流区碳氢燃料传热研究[J]. 推进技术, 2013, 34(2): 225-229.
ZHANG L, LE J L, ZHANG R L, et al. Heat transfer of hydrocarbon fuel in turbulent flow region under supercritical pressure[J]. Journal of Propulsion Technology, 2013, 34(2): 225-229(in

Chinese).

[10] CORMIER Y, DUPUIS P, FARJAM A, et al. Additive manufacturing of pyramidal pin fins: Height and fin density effects under forced convection[J]. International Journal of Heat and Mass Transfer, 2014(75): 235-244.

[11] HUTTUNEN E, NYKÄNEN M T, ALEXANDERSEN J. Material extrusion additive manufacturing and experimental testing of topology-optimised passive heat sinks using a thermally-conductive plastic filament[J]. Additive Manufacturing, 2022(59): 103123.

[12] ALTENEIJI M, ALI M, KHAN K A, et al. Heat transfer effectiveness characteristics maps for additively manufactured TPMS compact heat exchangers[J]. Energy Storage and Saving, 2022, 1(3): 9.

[13] DELGADO J, CIURANA J, RODRÍGUEZ C A. Influence of process parameters on part quality and mechanical properties for DMLS and SLM with iron-based materials[J]. The International Journal of Advanced Manufacturing Technology, 2012, 60(5): 601-610.

[14] STRANO G, HAO L, EVERSON R M, et al. Surface roughness analysis, modelling and prediction in selective laser melting[J]. Journal of Materials Processing Technology, 2013, 213(4): 589-597.

[15] TOWNSEND A, SENIN N, BLUNT L, et al. Surface texture metrology for metal additive manufacturing: a review[J]. Precision Engineering, 2016, 46: 34-47.

[16] SHEN S, XU J L, ZHOU J J, et al. Flow and heat transfer in microchannels with rough wall surface[J]. Energy Conversion and Management, 2006, 47(11): 1311-1325.

[17] HUANG K, WAN J W, CHEN C X, et al. Experimental investigation on friction factor in pipes with large roughness[J]. Experimental Thermal and Fluid Science, 2013, 50: 147-153.

[18] 周红秀, 张奥, 刘悦闻, 等. 壁面表面粗糙度对喷孔内空化流动特性的影响[J]. 内燃机与动力装置, 2022, 39(2): 7-13.
ZHOU H X, ZHANG A, LIU Y W, et al. The effect of wall surface roughness on cavitation flow characteristics in nozzle[J]. Internal Combustion Engine & PowerPlant, 2022, 39(2): 7-13(in Chinese).

[19] LIU Y, XU G, SUN J, et al. Investigation of the roughness effect on flow behavior and heat transfer characteristics in microchannels[J]. International Journal of Heat and Mass Transfer, 2015, 83: 11-20.

[20] YUAN Y, TAO Z, LI H, et al. Experimental investigation of surface roughness effects on flow behavior and heat transfer characteristics for circular microchannels[J]. Chinese Journal of Aeronautics, 2016, 29(6): 1575-1581.

[21] 李虹杨, 郑赟. 粗糙度对涡轮叶片流动转捩及传热特性的影响[J]. 北京航空航天大学学报, 2016, 42(10): 2038-2047.
LI H Y, ZHENG Y. Effect of surface roughness on flow transition and heat transfer of turbine blade[J]. Journal of Beijing University of Aeronautics and Astronsutics, 2016, 42(10): 2038-2047(in Chinese).

[22] NATRAJAN V K, CHRISTENSEN K T. The impact of surface roughness on flow through a rectangular microchannel from the laminar to turbulent regimes[J]. Microfluid Nanofluid, 2010(9): 95-121.

[23] 郝鹏飞, 姚朝晖, 何枫. 粗糙微管道内液体流动特性的实验研究[J]. 物理学报, 2007, 56(8): 5.
HAO P F, YAO C H, HE F. Experimental study of flow characteristics in rough microchannels[J]. Acta Physica Sinica, 2007, 56(8): 5(in Chinese).

[24] 鲁聪达, 薛浩, 吴化平, 等. 随机粗糙微通道内部流动与传质特性[J]. 中国机械工程, 2018, 29(12): 7.
LU C D, XUE H, WU H P, et al. Flow and mass transfer characteristics in random roughness surface microchannels[J]. China Mechanical Engineering, 2018, 29(12): 7(in Chinese).

[25] ROVENSKAYA O I. Numerical analysis of surface roughness effects on the Poiseuille flow caused by a small pressure drop[J]. International Journal of Heat and Mass Transfer, 2017, 110: 817-826.

[26] RAWOOL A S, MITRA S K, KANDLIKAR S G. Numerical simulation of flow through microchannels with designed roughness[J]. Microfluidics and Nanofluidics, 2006, 2(3): 215-221.

Experimental study on flow and heat transfer of hydrocarbon fuels in additive manufacturing channels at supercritical pressure

CAI Lei[1]，XIAO Ying[2]，HAN Huaizhi[1,*]，YU Ruitian[1]，LUO Wen[1]

(1. School of Chemical Engineering，Sichuan University，Chengdu 610065，China；
2. AECC Sichuan Gas Turbine Establishment，Mianyang 621022，China)

Abstract: This article presents a comparative experimental study on the flow and heat transfer performance differences between additive manufacturing (AM) rough channels and machined smooth channels for supercritical pressure n-decane. Using a roughness gauge, the surface average roughness of the smooth channel and AM channel were determined to be 3 μm and 11 μm, respectively. The study investigated variations in the friction factor (*f*) and Nusselt number (*Nu*) of the two channels under different flow rates, system pressures (3 MPa, 5 MPa, and 7 MPa), and wall heat flux densities (0.29 MW/m^2, 0.32 MW/m^2, and 0.35 MW/m^2). The results showed that, within the selected flow rate range, both *f* and *Nu* of the AM channel were higher than those of the smooth channel, with *f* being 2.8～3.5 times that of the smooth channel, and *Nu* being 1.7～2.3 times that of the smooth channel. Increasing the system pressure reduced *f* of the AM channel, but increased that of the smooth channel, while *Nu* decreased for both channels. As the wall heat flux density increased, both *f* and *Nu* of the AM channel decreased, while those of the smooth channel increased. The rough channel had a comprehensive heat transfer coefficient of 1.2～1.5, showing outstanding overall heat transfer performance and greater effectiveness at low flow rate. The correlations of the *Nu* and *f* of supercritical hydrocarbon fuel in additive manufacturing channel are proposed.

Keywords: roughness; additive manufacturing; friction loss; heat transfer characteristics; experimental study

Received：2023-03-10；**Accepted**：2023-04-04；**Published Online**：2023-05-05 08：41

URL：link.cnki.net/urlid/11.2625.V.20230504.1814.002

Foundation items：National Natural Science Foundation of China (T2441001)； Special Fund of Sichuan Provincial Gas Turbine Combustion Engineering Technology Research Center (JMRH-0000-2024-00001)

* **Corresponding author**. E-mail：hanhz@scu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0147

融合注意力和多尺度特征的航空发动机缺陷检测

赵崇林[1]，朱江[2]，胡永进[3]，李祖泽[3]，王鹏举[3]，谢涛[2,*]

(1. 昆明理工大学 信息工程与自动化学院，昆明 650500；2. 昆明理工大学 民航与航空学院，昆明 650500；
3. 东方航空技术有限公司云南分公司，昆明 650200)

摘　　要：航空发动机的结构完整性关乎飞行安全。目前基于孔探技术的航空发动机缺陷检测以人工操作为主。为提高检测精度和效率，提出一种融合注意力和多尺度特征的航空发动机缺陷智能检测算法，以辅助孔探工作。针对原始孔探图像中缺陷样本的类别不平衡问题，采用了一种基于几何变换和泊松图像编辑的多样本融合数据增强方法，丰富小样本图像并构建缺陷数据集。在基准网络YOLOv5中融入协调注意力模块(CA)，以强调缺陷特征的提取，增强网络对缺陷目标和复杂背景的区分。在颈部网络中构建加权双向特征金字塔结构(BiFPN)，以完成更高层次的特征融合，提升对多尺度目标的表达能力。将边界框回归损失函数定义为EIOU损失，实现对缺陷目标快速、准确的定位和识别。实验结果表明：所提算法检测缺陷的平均精确度达到了89.7%，较基准网络提升了6.3%，训练后的模型大小仅为14.0 M。因此，所提算法可以有效地检测航空发动机的主要缺陷。

关 键 词：航空发动机；缺陷检测；深度学习；YOLOv5；注意力机制
中图分类号：TP391.4
文献标志码：A　　**文章编号**：1001-5965(2025)03-0892-12

航空发动机作为飞行器的核心，为飞行器供应动力，其结构完整性将直接影响飞行安全。发动机叶片等航空发动机的核心部件，长期工作于恶劣环境，有造成发动机故障的隐患。因此，对航空发动机的定期检修、突发性排故检测和故障监控是保证飞行器适航的必要任务。目前对于航空发动机的检测主要是基于孔探技术，由专业人员观察和判断发动机内部的损伤状况[1-3]。虽然传统的孔探方法已经成为实际工作中不可或缺的检测手段，但受限于技术人员的经验和疲劳程度，检测排故的效率不高。因此，有必要发展一些其他的检测方法辅助孔探技术，以降低检测误差、保证飞行安全。

近年来，随着计算机硬件的发展，以卷积神经网络为代表的深度学习算法发展迅速。由于卷积神经网络良好的数据挖掘性和可移植性，已有学者开始尝试将深度学习方法应用到飞机机身和航空发动机的缺陷检测任务。Malekzadeh等[4]首次提出一种基于深度神经网络的飞机缺陷自动识别方法，使用拍摄到的飞机机身直视图作为数据集，利用在ImageNet上预训练的卷积神经网络作为特征提取器，支持向量机(support vector machine, SVM)作为分类器，实现了对飞机机身缺陷的检测。Kim和Lee[5]开发了一种航空发动机叶片损伤识别算法，首先，使用多种图像处理技术对发动机孔探图像进行预处理，从中选择疑似损伤区域，然后，通过卷积神经网络将疑似区域进一步分类为损伤区域或正常区域，实现了对发动机叶片缺陷的识别。Li等[6]研究了一种基于粗到细的深度学习框架以自动检测发动机叶片表面的缺陷，利用提出的粗分类器模块滤除大部分背景，细检测器模块定位和分类缺陷目标，对于5种发动机叶片的常见缺陷实现了较好的检测效果。李龙浦[7]针对航空发动机缺陷的特征提出改进的YOLOv3模型和Faster R-CNN模型，对比并选择了最优模型进行缺陷的识别和定位，同时实现了对掉块类缺陷的

收稿日期：2023-03-29；录用日期：2023-04-21；网络出版时间：2023-05-06 17:56
网络出版地址：link.cnki.net/urlid/11.2625.V.20230506.0932.001
*通信作者. E-mail：77183118@qq.com

引用格式：赵崇林，朱江，胡永进，等. 融合注意力和多尺度特征的航空发动机缺陷检测[J]. 北京航空航天大学学报，2025，51(3)：892-903. ZHAO C L，ZHU J，HU Y J，et al. Aero-engine defect detection by fusing attention and multi-scale features[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：892-903 (in Chinese).

尺寸测量。李彬等[8]在 YOLOv4 的基础上做出改进，所提算法将深层特征与浅层特征融合，同时去除颈部结构中自底向上的增强路径，并应用改进 Focal loss 作为分类损失函数，实现了对自建数据集中 4 类缺陷目标的检测。陈为和梁晨江[9]针对航空发动机缺陷目标小、检测难度大的问题提出改进的 SSD 模型，通过对原始网络的特征提取结构进行改进，提升了对航空发动机内部凸台缺陷的检测效果。樊玮等[10]提出基于 Mask R-CNN 的二分类到多分类的检测网络，首先，对航空发动机缺陷图像进行二分类并优化回归坐标；然后，进一步实现缺陷的多分类检测；最后，根据多分类检测的结果完成了对航空发动机缺陷的实例分割。

目前，在使用深度学习方法进行航空发动机缺陷检测的任务中，仍存在一些困难和挑战。如航空发动机原始孔探图像难以获得，即使收集到真实的孔探图像，其中缺陷样本少，大量无缺陷样本冗余且缺陷样本分布不平衡，导致数据集构建困难。同时，孔探图像中的缺陷目标尺度变化较大，小目标居多，背景区域的纹理复杂、易与缺陷目标混淆，在检测时可能出现漏检、误检的情况。在实际的孔探检测中，数据是以视频的形式传出供技术人员检查，对于匹配的深度学习算法，需具有较快的检测速度，满足实时性要求。综上所述，当前仍然缺乏针对缺陷图像的快速、准确的检测算法。

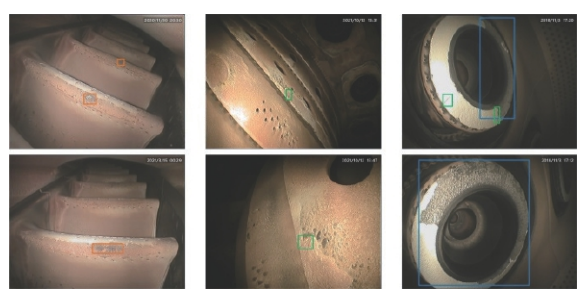
为解决上述问题，本文提出一种融合注意力和多尺度特征的航空发动机缺陷检测算法，以实现对氧化、裂纹和缺失类缺陷的有效识别。采用多样本融合的策略，利用几何变换和泊松图像编辑技术，扩充了小样本缺陷，平衡了数据集样本分布。在 YOLOv5 的主干网络中融合协调注意力机制以关注目标特征，为后续连接提供更多有用的特征信息，提升对真实缺陷目标的辨别能力。构建加权双向特征金字塔结构，实现了多尺度特征融合，尽可能多地聚合并传递提取到的特征，改善对多尺度目标尤其是小目标的检测效果。使用 EIOU 损失作为边界框回归损失函数，快速准确地定位缺陷目标。

1　数据预处理和数据集准备

本文收集的原始数据来源于当地某航空公司提供的 2018—2021 年 CFM56-7B 航空发动机真实的孔探图像，其中包括 343 张氧化、102 张裂纹、282 张缺失类缺陷和大量无缺陷的背景图像，图像的像素均为 752×576。缺陷主要存在于燃烧室、涡轮等热部件，也有部分裂纹和缺失缺陷产生于压气机部件。每张图像都包含至少一个缺陷目标，少数图像可能同时包含缺失和裂纹类目标。氧化、裂纹和缺失类缺陷的原始图像如图 1 所示。

在深度学习领域，卷积神经网络需要足够的训练样本来避免过拟合现象。我们注意到 3 类缺陷样本的数量不平衡，最多类和最少类的样本比例超过了 3∶1，这将导致网络对它们的表达效果差异较大，影响模型的整体检测能力。因此，在最终构建数据集之前，需要对小样本的数据做针对性的增强处理，使整体样本分布更均衡。

传统基于单样本的数据增强方法，如几何变换、颜色变换等，仅改变了图像的表现形式而没有实质地改变图像内容，并未达到真正意义上的丰富小样本数据，从而限制了卷积神经网络对小样本数据特征的学习和表示能力[11]。因此，本文采用多样本融合的数据增强方法，基于几何变换和泊松图像编辑技术，通过人工合成新样本来处理样本不平衡的问题。首先，对收集到的原始孔探图像进行初步筛选，在全部的缺陷图像中找到合适的小样本缺陷类别图像，在无缺陷的背景图像中选取可能发生该类缺陷的疑似背景图像，分别整理为缺陷图像集和背景图像集。然后，对小样本的缺陷目标进行裁剪，并运用几何变换手段如旋转、缩放等调整至合适的角度位置。最后，利用泊松图像编辑将缺陷目标和背景图像融合，使合成后图像更为接近真实采集的孔探图像。由于原始孔探图像中裂纹类缺陷所占的比例最低、缺失类次之，因此，通过多样本融合数据增强方法对这 2 类缺陷进行扩充，图 2 为小样本缺陷的数



(a) 氧化　(b) 裂纹　(c) 缺失

图 1　航空发动机的 3 类缺陷示例

Fig. 1　Three classes of defects in aero-engines

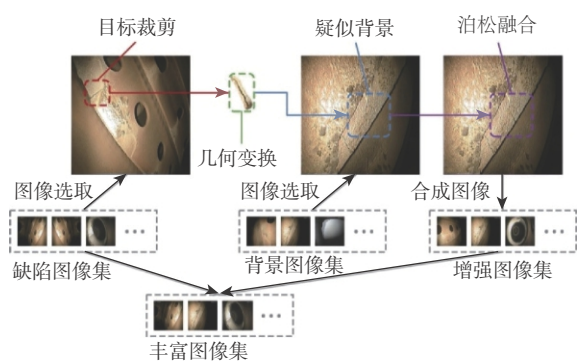
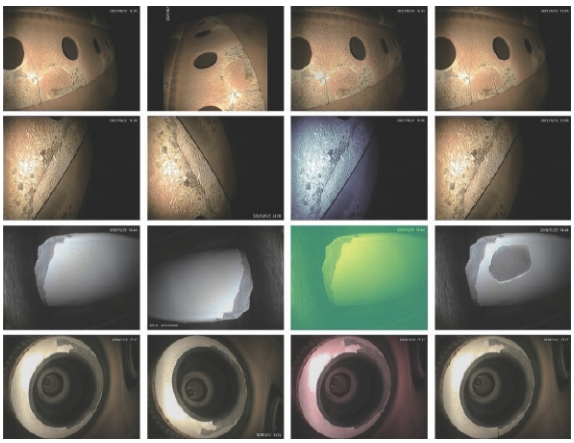


图 2　小样本缺陷的数据增强过程

Fig. 2　Data augmentation process for small sample defects

据增强过程。最终将裂纹类图像扩充至 306 张，缺失类图像 340 张，丰富后的数据集图像达到 989 张，按照 6:2:2 的比例划分为训练集、验证集和测试集。图 3 为应用不同数据增强方法后的直观化结果。



(a) 原始图像 (b) 几何变换 (c) 颜色变换 (d) 本文算法

图 3 应用不同数据增强方法的直观化结果

Fig. 3 Visualization results using different data augmentation methods

2 航空发动机缺陷检测算法

2.1 基准网络 YOLOv5

YOLOv5 是由 Glenn Jocher 提出的一阶段目标检测网络，融合了之前 YOLO 系列算法的优点，检测精度和速度都表现得更为强大。YOLOv5 模型有 4 种尺寸，分别是 s、m、l 和 x，它们的结构深度和宽度逐次增加。本文以 YOLOv5s 作为基准网络，原因是它更小更快。图 4 为 YOLOv5s 网络的整体框架和各个组件的详细结构，YOLOv5 主要由特征提取主干、颈部特征金字塔结构和预测头部 3 个部分组成，其中，颈部和头部也可以结合起来共同看作预测头部。

YOLOv5 的主干部分使用了跨阶段局部网络 (cross stage partial network, CSPNet)[12]，网络的学习能力得以增强，既能保持精确性又足够轻量化，主干网络主要由 (convolutional-batch normalization-SiLU) 模块、C3 模块和 (spatial pyramid pooling-fast) 模块组成。CBS 模块包括一个卷积层 (convolutional, Conv)、一个批量归一化层 (batch normalization, BN) 和一个 SiLU 激活函数，用于提取不同层次图像的信息特征。C3 模块包含了 3 个卷积层及多个残差组件，优化了梯度反向传播的路径，提升网络学习能力的同时降低了计算成本和内存开销。SPPF 模块借鉴了空间金字塔池化 (spatial pyramid pooling, SPP) 的思想[13]，增大感受野并提取出最重要的上下文特征，同时具有更快的检测速度。YOLOv5 的颈部结构由特征金字塔网络 (feature pyramid network, FPN)[14]

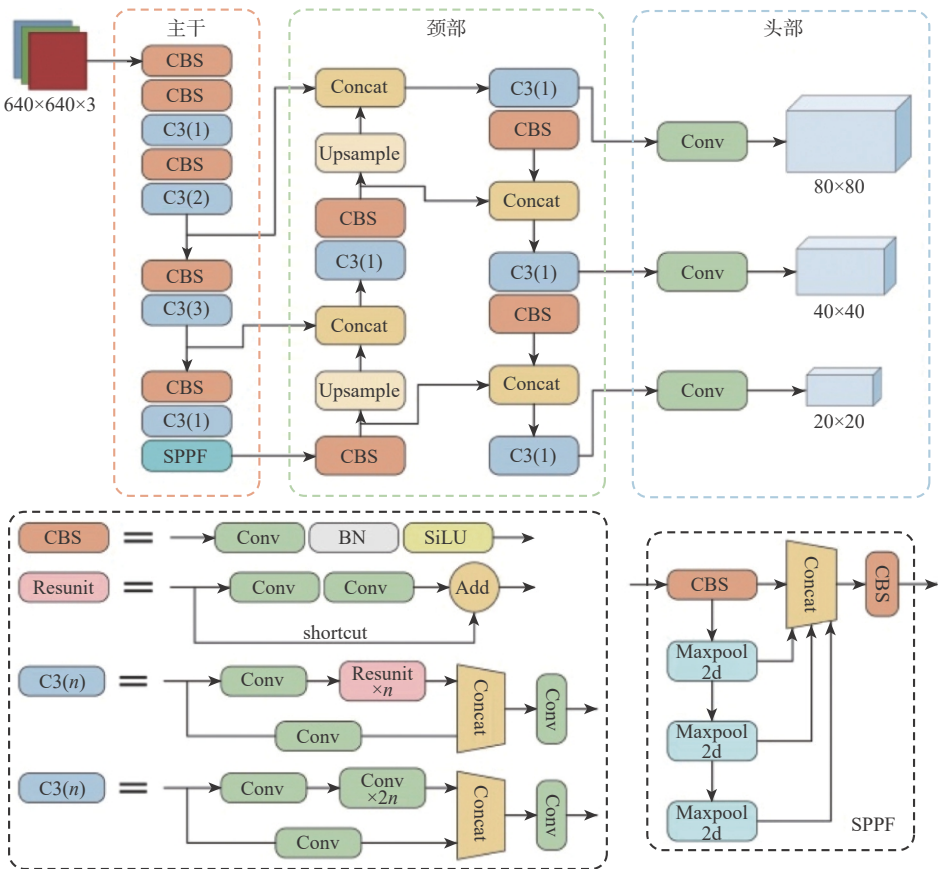


图 4 YOLOv5s 的网络结构

Fig. 4 Network structure of YOLOv5s

和路径聚合网络 (path aggregation network, PAN)[15] 组成, 在不同尺度的图像上生成不同尺度的特征, 能够实现对多尺度目标的检测。由于结合了 FPN 自顶向下传达的语义信息和 PAN 自底向上的位置信息, YOLOv5 网络具有很强的特征融合能力。头部用于完成最终的检测, 在 3 个大小不同的特征图上预测不同尺度的目标, 其中最大的特征图负责检测小目标。

2.2　本文算法

由于航空发动机内部结构复杂, 缺陷目标和背景容易混淆, 如裂纹类缺陷和纹理复杂的背景特征相似难以区分。大量缺陷(如氧化类和部分裂纹类缺陷)以小目标的形式存在, 不容易被检测。同时, 缺陷之间的尺寸大小不一, 即使是同类缺陷(如裂纹类缺陷和部分缺失类缺陷)也有较大的尺寸差异, 导致构建的数据集中存在着更为突出的目标多尺度变化问题。原始的 YOLOv5 无法很好地解决这些问题, 为此提出本文算法, 其整体框架如图 5 所示。首先, 在基准网络主干特征提取的第 3 次和第 4 次下采样层之后融入协调注意力模块, 由于网络的浅层中保留了足够的边缘特征, 这些边缘信息对于缺陷目标的识别十分重要[16], 因此, 利用注意力机制对相应的特征信息加以关注。然后, 在颈部构建单层加权双向特征金字塔结构, 对各特征层之间的连接赋予可学习权重参数, 在特征融合部分的第 3 次拼接操作中增加额外的路径, 将相同尺度的特征图进行拼接, 进而充分完成特征信息的聚合。最后, 将边界框回归损失函数设计为 EIOU 损失, 加快检测框的回归并且使目标定位更精确。

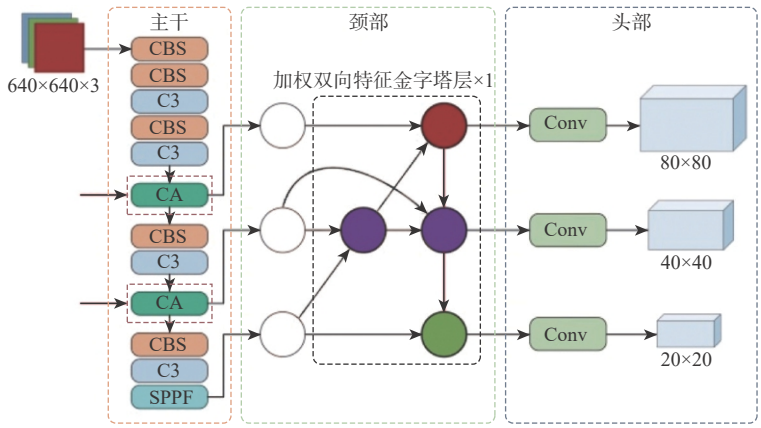


图 5　本文算法的整体框架

Fig. 5　Overall framework of the proposed algorithm

2.2.1　协调注意力模块

注意力机制对模型性能的提升已经被证明, 在计算机视觉任务中应用广泛。对于轻量级网络的研究表明, 通道注意力如挤压和激励 (squeeze and excitation, SE) 机制的应用是有效的[17], 可以建立起各个特征通道之间的关联, 使得各通道的特征响应能够被自动调整。然而以往的这些机制通常忽略了位置信息, 这对于空间选择性注意力图的生成十分重要[18]。考虑到航空发动机孔探图像的复杂背景和计算开销, 本文算法融合了一种高效轻量级的协调注意力机制, 同时考虑通道间信息的编码和位置关系, 以较低的计算量增强了网络对缺陷目标的特征表示, 强调有效特征的同时抑制了无关特征。图 6 为协调注意力模块的整体结构。

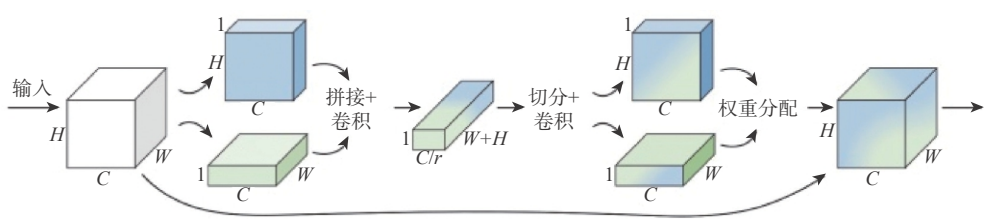


图 6　协调注意力模块的结构

Fig. 6　Structure of coordination attention module

协调注意力机制首先会沿水平方向和垂直方向, 对输入特征图分别进行一维特征编码, 获得图像宽、高方向上的注意力, 并且保留了精确的位置信息, 公式表示为

$$z_c^w(w)=\frac{1}{H}\sum_{0\leqslant i<H}x_c(i,w) \tag{1}$$

$$z_c^h(h)=\frac{1}{W}\sum_{0\leqslant j<W}x_c(h,j) \tag{2}$$

式中：$z_c^w(w)$表示宽度 w 处的第 c 通道的输出；H 为输入 X 的高度；$z_c^h(h)$表示高度为 h 的第 c 通道的输出；W 为输入 X 的宽度。

将水平和垂直方向的特征图拼接，此时 2 个方向的特征图均具有全局感受野，接着对它们作用一个 1×1 卷积变换函数 F_1，将通道数降低为 C/r。然后将经过 BN 处理的特征图送入非线性激活函数得到 $1\times(W+H)\times C/r$ 大小的特征图 f，通过较为简单的转换即可充分利用位置信息。表示为

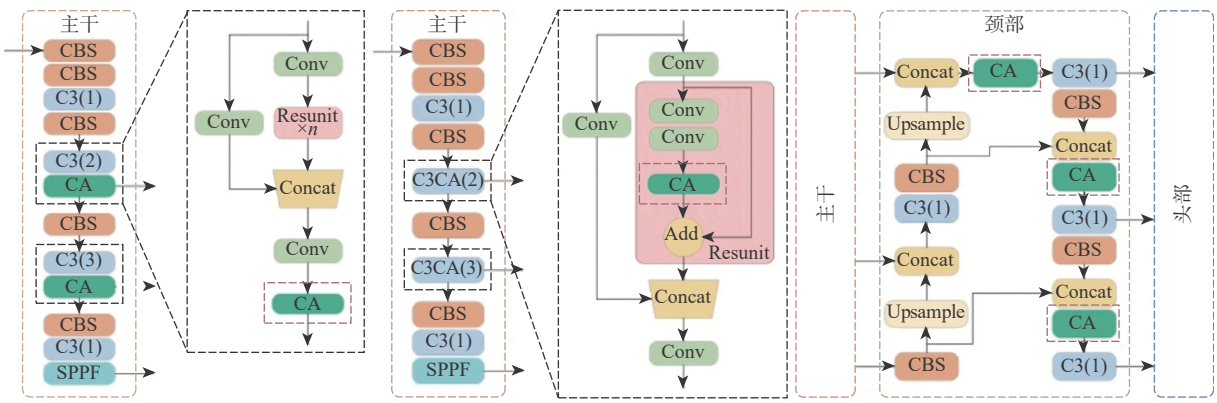
$$f=\delta\left(F_1\left(\left[z^h,z^w\right]\right)\right) \tag{3}$$

式中：f 包含 2 个方向的编码信息；δ 为非线性激活函数；[,,]表示沿空间维度的拼接操作。

将特征图 f 切分为二，得到独立张量 $\boldsymbol{f}^w\in\mathbf{R}^{C/r\times W}$ 和 $\boldsymbol{f}^h\in\mathbf{R}^{C/r\times H}$。然后，对 $\boldsymbol{f}^w$ 和 $\boldsymbol{f}^h$ 应用 2 个 1×1 卷积函数 F_w 和 F_h 复原通道数。经由 sigmoid 激活函数 σ 作用之后，特征图水平、垂直方向的注意力权重 g^w 和 g^h 即可得到。最终得到在 2 个方向上带有注意力权重的特征图 Y 可表示为

$$y_c(i,j)=x_c(i,j)\times g_c^w(i)\times g_c^h(j) \tag{4}$$

本文将协调注意力机制融入原始网络，更加关注有用的特征信息。根据放置位置的不同提出 3 种协调注意力模块的融合方案，位置 1、位置 2 和位置 3，如图 7 所示。位置 1 和位置 2 是将注意力机制融合于主干网络中，位置 3 是添加在颈部网络。其中，位置 1 是直接将协调注意力模块添加在第 3 次和第 4 次下采样之后的 C3 模块末端，位置 2 是添加到对应 C3 模块内部的残差组件中。



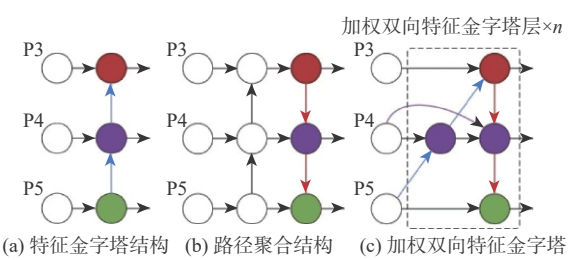
(a) 添加在主干网络的 C3 模块后　(b) 添加在 C3 模块内部残差组件中　(b) 添加在颈部网络中

图 7　协调注意力模块的不同融入位置

Fig. 7　Different integration positions of coordination attention module

2.2.2　加权双向特征金字塔结构

颈部网络是目标检测器承前启后的重要架构，它对主干部分提取到的特征进行聚合及合理利用，传递到预测头部以进行具体的分类和回归任务。在颈部构建特征金字塔是目标检测器常用的一种方式。传统的 FPN 结构本质上受到单向信息流的限制，因此，YOLOv5 的颈部网络采用了 PAN 架构，在 FPN 的后续又添加了一条路径，自底向上将高层的语义特征和低层的定位信息结合，进一步提高了特征提取和融合的能力。由于航空发动机缺陷本身尺度不一，检测时孔探镜头与缺陷之间又存在位置变化，导致传递出的孔探图像存在大量多尺度目标。因此，为更好地解决航空发动机孔探图像缺陷目标的多尺度检测，防止小目标信息特征的丢失，构建了一种更高层次的特征融合结构，即加权双向特征金字塔网络[19]。该结构以简单、快速的复合缩放方法融合多尺度特征，提升模型对缺陷在不同尺度下的检测精度。图 8 为特征金字塔网络、路径聚合网络和加权双向特征金字塔网络结构的示意图。



(a) 特征金字塔结构　(b) 路径聚合结构　(c) 加权双向特征金字塔

图 8　颈部网络的不同结构

Fig. 8　Different structures of neck network

与 YOLOv5 网络原始的特征金字塔和路径聚合网络结构相比，加权双向特征金字塔结构采用了双向跨尺度连接方法，并为各连接赋予权重，实现

高效的特征融合。从图 8(c) 可以看出, 为提高效率, 加权特征金字塔结构将只有一条输入边的节点移除 (如 P3 右边的节点), 并对相同级别的输入输出节点增加一条额外的边 (如 P4 增加的紫色边)。同时, 不同于特征金字塔和路径聚合网络, 加权双向特征金字塔网络采用快速归一化融合的方法为每个输入增加权重, 并让网络自适应学习不同特征之间的贡献差异, 公式表示为

$$O=\sum_i \frac{w_i}{\epsilon+\sum_j w_j} I_i \tag{5}$$

式中: w_i为一个可学习权重参数, ReLU 激活函数置于每个参数之后, 以保证$w_i \geqslant 0$; $\epsilon=0.000\,1$为一个很小的值以防止训练过程的振荡, 每个归一化权重的值控制在 0 和 1 之间。这种方式由于没有使用 softmax 操作, 效率更高。本文使用了单层加权双向特征金字塔结构改进原始的颈部网络, 提高了模型的训练效率, 提升了对多尺度目标的特征表示能力。

2.2.3　边界框回归损失函数

在深度学习领域, 损失函数是衡量网络输出与标签真值之间差别的依据, 即模型预测效果的好坏。YOLOv5 的整体损失函数计算包括 3 个方面: 边界框回归、分类和置信度损失, 其中, 前两方面损失由 BCE Loss 函数计算, 而边界框回归损失由 CIOU 损失函数计算。边界框回归是目标检测中的一种主流技术, 用于解决目标在图像中的定位问题并细化预测框的位置。基于 IOU 系列的损失函数是常用的边界框回归损失函数, 通过计算预测框和真实框之间的交集与并集之比作为损失值, 使预测框回归快速准确地收敛。随着 IOU 系列损失函数的发展, YOLOv5 中使用的 CIOU 损失函数兼顾了预测框和目标框的重叠面积、中心点之间的距离和长宽比, 公式如下:

$$L_{\mathrm{CIOU}}=1-\left(I-\frac{\rho^2(b,b^{\mathrm{gt}})}{c^2}-\alpha v\right) \tag{6}$$

式中: I为预测框和真实框之间的交并比; b和b^{gt}为预测框中点和目标框中点, $\rho(\cdot)=\|b-b^{\mathrm{gt}}\|_2$表示它们之间的距离, 由 2 个框的位置关系定义, 恰好覆盖它们的框为最小包围框, 其对角线长度为c。$v=\frac{4}{\pi^2}\left(\arctan\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}}-\arctan\frac{w}{h}\right)^2$反映了 2 个框之间的宽高比相似性, $\alpha=\frac{v}{(1-\mathrm{IOU})+v}$是$v$的影响因子。虽然 CIOU 在回归效果上提升明显, 但它依然存在一些问题, 对于式 (6) 中的v并没有很好的定义。当 2 个框的宽高之间满足$w=kw^{\mathrm{gt}}$和$h=kh^{\mathrm{gt}}$时, 有$v=0$, 这是不合理的。而且 CIOU 未能有效减少预测框和真实框的宽和高之间的差异。计算得到v关于w和h的梯度如下:

$$\frac{\partial v}{\partial w}=\frac{8}{\pi^2}\left(\arctan\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}}-\arctan\frac{w}{h}\right)^2\frac{h}{w^2+h^2} \tag{7}$$

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

可以发现$\frac{\partial v}{\partial w}=-\frac{h}{w}\cdot\frac{\partial v}{\partial h}$, $\frac{\partial v}{\partial w}$和$\frac{\partial v}{\partial h}$有相反的迹象, 当$w$和$h$有一个增加时, 另一个就会减小, 这减缓了 CIOU 的收敛速度。

为解决 CIOU 存在的问题, 实现航空发动机缺陷目标预测框回归的快速准, 在此任务中引入了 EIOU 损失函数, 其示意图如图 9 所示。EIOU 损失函数表示如下:

$$L_{\mathrm{EIOU}}=1-\left(I-\frac{\rho^2(b,b^{\mathrm{gt}})}{c^2}-\frac{\rho^2(w,w^{\mathrm{gt}})}{C_w^2}-\frac{\rho^2(h,h^{\mathrm{gt}})}{C_h^2}\right) \tag{9}$$

式中: I为预测框和真实框之间的交并比; b和b^{gt}为预测框中点和目标框中点; C_w为最小包围框的宽度; C_h为最小包围框的高度。EIOU 损失可以有效减小预测框和目标框(w,h)和$(w^{\mathrm{gt}},h^{\mathrm{gt}})$之间的差异, 加快收敛速度并得到更好的定位结果[20]。

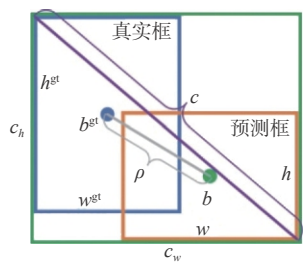


图 9　EIOU 损失的边界框回归示意图

Fig. 9　Bounding box regression for EIOU loss

本文算法融合了协调注意力机制、单层加权双向特征金字塔结构和 EIOU 损失, 根据协调注意力模块放置位置的不同区别为 3 种改进方案, 改进 1、改进 2 和改进 3。最终的改进后模型可以有效地检测航空发动机的主要缺陷, 能更好地区分缺陷目标和易混淆的背景, 对多尺度目标的检测能力更强, 尤其体现在对小目标的检测, 对于缺陷的定位和识别快速准确。具体的定量化结果和直观化效果在实验部分详细描述。

3　实验和结果分析

本节描述实验设计的流程并分析实验结果。首先, 介绍实验环境和评价指标。然后, 对改进模型和基线模型的性能进行比较, 并设计消融实验以评估最终的改进后模型, 分析各改进部分对基准网

络性能提升的贡献。还对采用不同方法的各个模块进行对比。最后，将本文模型与其他优秀的检测模型进行比较，证明本文算法在检测精度和速度上的优越性。

3.1 实验环境和评价指标

本文所有实验都是在一台配置有NVIDIA GeForce RTX 3 070 Ti GPU的计算机上进行的，在PyTorch框架下利用CUDA并行计算架构实现检测模型的构建。另外，Python和OpenCV用于实现数据增强。详细环境配置如表1所示。模型训练使用SGD优化器，初始学习率为0.01，最终学习率为0.000 1，动量为0.9，权值衰减为0.000 5，轮次为300，批样本数量为16，输入图像调整为640×640。

表1 环境配置细节

Table 1 Environment configuration details

配置	类型
Operating System	Windows 10
CPU	Intel Core i5-12400
GPU	NVIDIA GeForce RTX 3070 Ti, 8 GB
Pytorch	Version 1.10.1
CUDA	Version 11.3
cuDNN	Version 8.2.0
PyThon	Version 3.18.13
OpenCV	Version 4.6.0.66

在深度学习目标检测领域，对模型性能的评价主要集中于检测精确度、推理速度和模型大小，因此本文采用以下目标检测中常用的评价指标全方位衡量模型：①精确率*P*(precision)；②召回率*R*(recall)；③单一类别平均精确率(average precision, AP)；④整体平均精确率均值(mean average precision, mAP)；⑤检测帧率(frames per second, FPS)；⑥已训练的模型权重大小，单位为MB。

3.2 消融实验

首先，对改进模型进行评估，原始模型、改进1、改进2和改进3模型的mAP和损失曲线如图10所示。损失函数的值能够评估模型的预测效果好坏，通过损失曲线的变化可以呈现出网络训练的动态趋势。可以看出，在训练的过程中，3个改进模型的损失收敛速度更快并且损失值更小，随着训练轮次的增加而达到稳定收敛的状态，损失曲线没有发生明显的振荡现象，并且mAP也始终高于原始模型，因此，可以证明改进模型具有更好的检测精确度。同时可以发现，由于协调注意力模块放置位置的差异，3个改进模型的性能也有所不同，具体的检测结果如表2所示。通过对比可以看到改进2模型的整体表现最优，因此，本文将改进2模型作为最终的改进后模型。

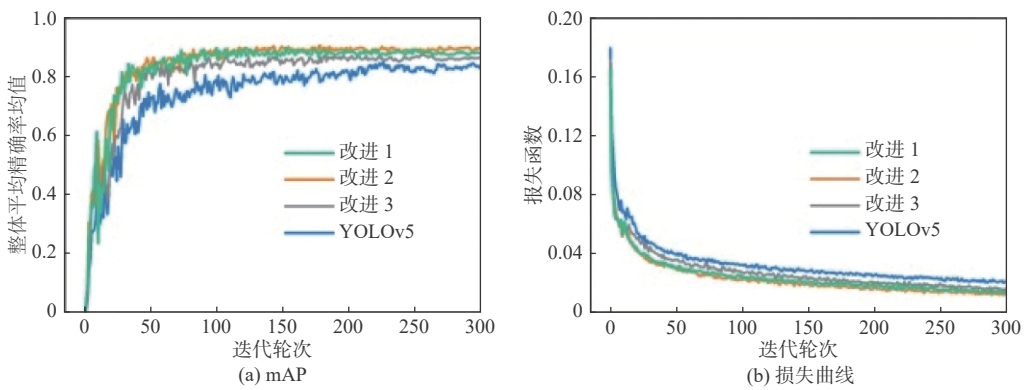


图10 改进后模型和原始模型的mAP和损失曲线比较

Fig. 10 Comparison of mAP and loss curves between improved model and original model

表2 改进后模型和原始模型的检测性能结果

Table 2 Test performance results of improved model and original model

模型	*P*/%	*R*/%	mAP/%
YOLOv5	79.6	86.5	83.4
改进1	86.8	89.3	89.1
改进2	87.9	89.6	89.7
改进3	85.5	88.8	88.0

为验证最终的改进后模型对于航空发动机缺陷检测的有效性，针对基准网络所提出的改进部分的必要性，评估各个模块对网络性能提升的贡献程度，设计了消融实验以定量地衡量数据增强、协调注意力模块、加权双向特征金字塔结构和EIOU损失函数带来的效果。基线模型由原始未经任何预处理的数据集训练得到，表3为消融实验的结果。可以看出，基线模型对航空发动机3类主要缺陷检测的mAP为83.4%，检测帧率达到了169.5帧/s。当对小样本缺陷进行数据增强，使缺陷样本分布更均衡之后，YOLOv5模型的mAP达到了86.1%，其对裂纹类缺陷的检测效果提升明显，AP增加了6.2%，缺失类缺陷的AP提升了2.1%。这证明本文

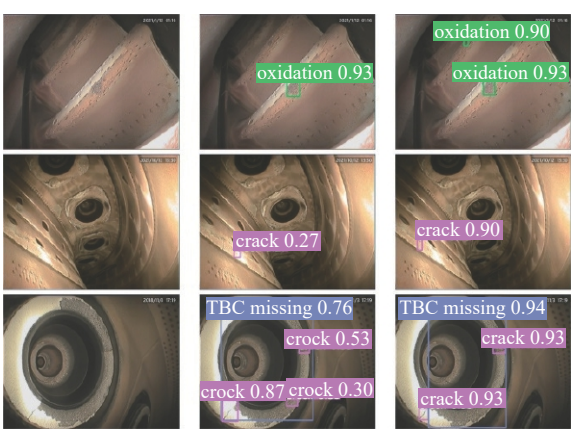
针对小样本的数据增强方法是有效的。在 YOLOv5 的主干部分融合了协调注意力模块后, 网络对缺陷目标的关注度和特征表示能力提高, 尤其体现在部分小目标的氧化类缺陷和容易与背景混淆的裂纹类缺陷, 此时的模型对这 2 类缺陷的 AP 分别提升了 2.3% 和 2.0%, 整体的 mAP 达到了 87.8%。由于并没有增加过多的计算开销, 因此, 检测帧率仍然达到了 166.7 帧/s。在此基础上, 构建了加权双向特征金字塔结构, 进一步聚合主干部分提取到的特征, 更高级别的跳层连接对于多尺度的缺陷的检测更为有利, 对于尺度变化较大的缺失类缺陷和部分裂纹类缺陷的检测结果得到了提升, 整体缺陷检测的 mAP 达到了 89.0%。由于增加了额外的参数量, 模型的计算开销变大, 此时检测帧率下降至 156.3 帧/s。最后, 将模型的边界框回归损失函数替换为 EIOU 损失, 对于缺陷的定位更快更准, 使模型的性能达到了最优。最终改进后模型的 mAP 达到了 89.7%, 相较于基线模型提高了 6.3%。对于 3 类缺陷检测的 AP 都达到了最高水平, 分别为 92.2%, 86.2% 和 90.8%。模型的检测速度不及基线模型, 但依然处于较高水平, 能够满足实时检测要求。改进后模型以牺牲较小检测速度的代价换来检测性能的较大提升。在航空发动机缺陷检测任务上检测精度是放在首位的, 因为它关乎着安全性。因此, 本文算法对于航空发动机主要缺陷的检测是有效的, 各部分改进是必要的。

表 3　消融实验结果

Table 3　Results of ablation experiment

数据增强	CA	BiFPN	EIOU	AP/%			mAP/%	检测帧率/(帧·s^{-1})
				氧化	裂纹	缺失		
				88.5	75.6	86.0	83.4	169.5
√				88.4	81.8	88.1	86.1	169.5
√	√			90.7	83.8	88.8	87.8	166.7
√	√	√		91.7	85.3	89.9	89.0	156.3
√	√	√	√	92.2	86.2	90.8	89.7	156.3

为直观地比较改进后模型和基线模型的实际检测效果, 图 11 展示了它们对 3 类缺陷的检测结果示例。可以看出, 改进后模型的实际检测效果比基线模型更好。在对第 1 行氧化类缺陷的检测中, 基线模型忽略了更小的目标, 改进后模型则准确地检测到该处缺陷。对于第 2 行裂纹类缺陷, 基线模型将背景误检为缺陷并漏检了真正的目标, 改进后模型纠正了这个错误, 正确地检测到了裂纹。对第 3 行同时存在缺失类和裂纹类缺陷的检测, 基线模型将缺失部分误检为裂纹, 并且对于缺陷的定位不够准确。然而改进后模型正确区分了 2 类缺陷, 对于缺陷的定位非常准确, 并且提升了检测的置信度。

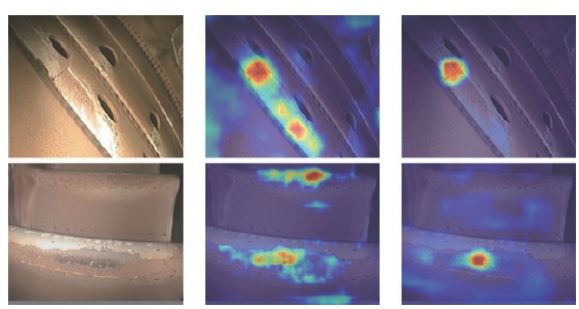


(a) 缺陷原图　(b) 基线模型效果　(c) 改进后模型效果

图 11　改进后模型和基线模型的实际检测效果

Fig. 11　Comparison of actual detection effects between improved model and baseline model

基线模型在检测时容易产生漏检和误检的情况, 尤其是漏检, 这在航空发动机的缺陷检测上是不可接受的。为此, 本文提出融合注意力和多尺度特征的处理方案, 最大限度地避免漏检和误检现象。为了清楚地看到本文算法对缺陷目标检测效果的提升, 使用 Grad-CAM 方法[21] 对基线模型和改进后模型进行热力图可视化, 如图 12 所示。Grad-CAM 是一种卷积神经网络可解释性的经典方法, 无须对模型进行改动就可生成热力图, 可直观展示



(a) 缺陷局部放大图　(b) 基线模型效果　(c) 改进后模型效果

图 12　改进后模型和基线模型的热力图可视化结果

Fig. 12　Visualization results of heat map of improved model and baseline model

图像各区域对类别分类的贡献程度,以热力图呈现出来的区域颜色分布即可做出判断,红色部分表示的贡献较大,蓝色部分贡献较小。可以看出,在对2个小目标的氧化和裂纹缺陷的检测中,改进后模型对缺陷部分的关注增加,对于目标和背景的辨别效果提升显著。

3.3 各改进部分使用不同方法的对比

对协调注意力机制的不同融合位置进行了实验,比较了在主干部分和颈部网络添加注意力机制对模型性能的影响,使用最终的数据增强后数据集进行测试。对基准网络不添加注意力机制,以及在位置1、位置2和位置3处融合协调注意力模块的准确性测试结果如表4所示。可以看到与不添加注意力机制的基准网络相比,在主干部分(位置1、位置2)融合协调注意力机制具有更好的检测结果,而在颈部网络(位置3)融入注意力机制效果不佳。由于处在较低层的特征图具有更小的感受野,主干网络中又保留了足够的局部和细节信息,这些条件有助于对小目标缺陷的检测,而颈部部分特征图的较大感受野和冗余的特征信息则不利于缺陷目标的识别。在位置2融合协调注意力机制的mAP达到了87.8%,比位置1的检测结果略好。协调注意力模块的数量同残差组件一起随着网络的加深而增加,多层堆叠的协调注意力模块对提取缺陷的特征起到了更好的效果。因此,在主干网络融合注意力机制有助于模型提高对于缺陷目标的表达能力。

表4 注意力机制不同融入位置的准确性结果

Table 4 Accuracy results of different integration positions of attention mechanism

位置	*P*/%	*R*/%	mAP/%
无注意力	82.7	86.7	86.1
位置1	85.0	87.9	87.3
位置2	85.4	88.1	87.8
位置3	82.5	87.2	85.8

对原始YOLOv5的颈部网络采用不同结构后的检测结果如表5所示。原始的颈部网络采用特征金字塔和路径聚合网络结构,本文构建了加权双向特征金字塔结构。对于这一部分的跳层连接还有一种常见的改进方法,即增加目标预测头,额外添加一个输出尺寸更大的特征图用于检测小目标。比较了这3种颈部结构对于航空发动机缺陷检测的结果。可以看到,采用加权双向特征金字塔结构的模型的mAP最高,达到了87.5%,比原始模型提高了1.4%。采用4个预测头对于缺陷的检测也有一些提升,但同时增加了额外的参数量,推理速度也有所下降。综合考虑,采用加权双向特征金字塔结构作为颈部部分进而预测缺陷目标是最佳选择。

表5 颈部网络不同结构的检测结果

Table 5 Detection results of different structures of neck network

结构	mAP/%	权重大小/MB	检测帧率/(帧·s^{-1})
FPN+PAN	86.1	13.7	169.5
4尺度预测头	87.2	14.6	143.2
BiFPN	87.5	13.9	162.6

表6为不同损失函数对基准网络检测结果的影响。原始的YOLOv5使用的边界框回归损失函数是CIOU损失,本文使用的是EIOU损失,作为比较的GIOU和DIOU损失都是IOU系列先进的损失函数。可以看到,使用EIOU损失的模型对3类缺陷的AP都达到了最优,mAP达到了87.3%,比原始使用CIOU的模型高了1.2%。证明EIOU损失可以更准确地定位和识别缺陷目标。同时本文EIOU损失的收敛更快速,事实证明EIOU损失对航空发动机缺陷的检测更具优势。通过对各个改进部分应用不同方法的实验对比,进一步说明本文算法在航空发动机缺陷检测方面的优越性。

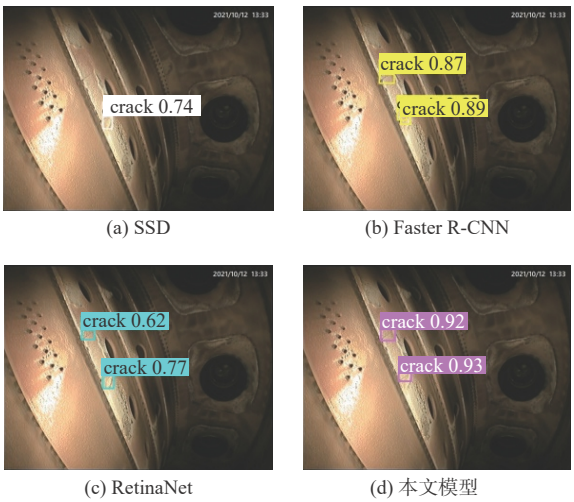
表6 不同损失函数的测试结果

Table 6 Test results of different loss functions

损失函数	AP/%			mAP/%
	氧化	裂纹	缺失	
GIOU	87.7	80.8	87.1	85.2
DIOU	88.6	81.1	87.2	85.7
CIOU	88.4	81.8	88.1	86.1
EIOU	89.5	83.6	88.8	87.3

3.4 对比实验

将最终的改进后模型与其他几种优秀的目标检测模型SSD[22]、Faster R-CNN[23]和RetinaNet[24]进行比较,以验证本文模型相较于同类模型的检测优势。对比实验中各模型在相同的实验环境下进行训练和测试,其中,SSD和Faster R-CNN使用的主干网络是VGG16,RetinaNet使用的主干网络为ResNet50。图13为各模型对小目标的裂纹类缺陷的检测效果。可以看出,本文模型的检测效果最好,精准地定位和识别出了图像中存在的2条裂纹,并取得了很高的置信度得分,证明本文针对性的改进方案对小目标的检测是有效的。SSD模型漏检了一个目标,同时对另外一个缺陷目标的定位不准确。Faster R-CNN模型将一条裂纹识别成了2条,没有完全区分开真实目标和相似背景。RetinaNet模型也没有精准的对缺陷定位,检测框没有恰当的包围裂纹目标。



(a) SSD　(b) Faster R-CNN

(c) RetinaNet　(d) 本文模型

图 13　不同模型的实际检测效果

Fig. 13　Comparison of actual detection effects of different models

各模型具体的测试结果如表 7 所示，以 mAP、检测帧率和已训练模型的权重大小作为评价指标来评估它们的检测精度、速度和模型大小。可以看到本文模型的 mAP 达到了 89.7%，优于其他模型，能够最精确地检测目标。检测帧率达到了 156.3 帧/s，检测速度仅逊于基线模型。在取得最高精度的同时，改进后模型的大小仅为 14.0 MB，证明本文模型足够轻量化并便于部署，可以应用于实际的检测场景。

表 7　不同模型的测试结果

Table 7　Test results of different models

模型	mAP/%	检测帧率/(帧·s^{-1})	权重大小/MB
SSD	78.0	119.3	91.6
Faster R-CNN	82.5	29.8	521.6
RetinaNet	83.6	45.1	139.1
YOLOv5	86.1	169.5	13.7
本文模型	89.7	156.3	14.0

4　结　论

本文针对现有深度学习算法在航空发动机缺陷检测任务中，存在缺陷样本分布不均衡、背景特征干扰性强和目标多尺度变化的问题展开了研究，得出以下结论：

1）为解决缺陷样本的类别不平衡问题，采用多样本融合的数据增强方法，基于几何变换和泊松图像编辑技术，有效丰富了小样本缺陷，平衡了数据集中的各类样本分布。

2）提出融合注意力和多尺度特征的检测算法，以实现对航空发动机缺陷的有效识别，本文算法在保证高精确度的同时兼顾了检测速度，轻量化方便部署，可以满足航空发动机主要缺陷的检测需要。

3）全面的实验验证表明，本文算法对于航空发动机缺陷的检测效果优于其他方法，虽然推理速度不及基线模型，但依然处于较高水平，在定量化结果和可视化效果上均达到预期。

本文算法在实际的航空发动机缺陷检测中具有应用前景，能够作为孔探检测的辅助技术手段，未来可以容纳更多缺陷类型并进一步优化，对于提高孔探工作的效率、降低误检率、减轻工作人员的疲劳强度等方面具有实用价值。

参考文献（References）

[1] PITKÄNEN J, HAKKARAINEN T, JESKANEN H, et al. NDT methods for revealing anomalies and defects in gas turbine blades[J]. Insight, 2001, 43: 601-604.

[2] ZOU F Q. Review of aero-engine defect detection technology[C]// Proceedings of the IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference. Piscataway: IEEE Press, 2020: 1524-1527.

[3] 旷可嘉. 深度学习及其在航空发动机缺陷检测中的应用研究[D]. 广州：华南理工大学, 2017.
KUANG K J. Research on deep learning and its application on the defects detection for aero engine[D]. Guangzhou: South China University of Technology, 2017(in Chinese).

[4] MALEKZADEH T, ABDOLLAHZADEH M, NEJATI H, et al. Aircraft fuselage defect detection using deep neural networks[J]. ArXiv e-Prints, 2017: arXiv: 1712.09213.

[5] KIM Y H, LEE J R. Videoscope-based inspection of turbofan engine blades using convolutional neural networks and image processing[J]. Structural Health Monitoring, 2019, 18(5-6): 2020-2039.

[6] LI D W, LI Y D, XIE Q, et al. Tiny defect detection in high-resolution aero-engine blade images *via* a coarse-to-fine framework[J]. IEEE Transactions on Instrumentation Measurement, 2021, 70: 3062175.

[7] 李龙浦. 基于孔探数据的航空发动机叶片损伤识别研究[D]. 天津：中国民航大学, 2020.
LI L P. Research on damage identification of aeroengine blades based on borescope data[D]. Tianjin: Civil Aviation University of China, 2020(in Chinese).

[8] 李彬, 汪诚, 丁相玉, 等. 改进 YOLOv4 的表面缺陷检测算法[J]. 北京航空航天大学学报, 2023, 49(3): 710-717.
LI B, WANG C, DING X Y, et al. Surface defect detection algorithm based on improved YOLOv4[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(3): 710-717(in Chinese).

[9] 陈为, 梁晨红. 基于改进 SSD 的航空发动机目标缺陷检测[J]. 控制工程, 2021, 28(12): 2329-2335.
CHEN W, LIANG C H. Aeroengine target defect detection based on improved SSD[J]. Control Engineering of China, 2021, 28(12): 2329-2335(in Chinese).

[10] 樊玮, 李晨炫, 邢艳, 等. 航空发动机损伤图像的二分类到多分类递进式检测网络[J]. 计算机应用, 2021, 41(8): 2352-2357.
FAN W, LI C X, XING Y, et al. Binary classification to multiple

classification progressive detection network for aero-engine damage images[J]. Journal of Computer Applications, 2021, 41(8): 2352-2357(in Chinese).

[11] HUANG R, DUAN B K, ZHANG Y X, et al. Prior-guided GAN-based interactive airplane engine damage image augmentation method[J]. Chinese Journal of Aeronautics, 2022, 35(10): 222-232.

[12] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2020: 1571-1580.

[13] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]//Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE Press, 2015: 1904-1916.

[14] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 936-944.

[15] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8759-8768.

[16] YING Z P, LIN Z T, WU Z Y, et al. A modified-YOLOv5s model for detection of wire braided hose defects[J]. Measurement, 2022, 190: 110683.

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

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

[19] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020.

[20] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506146-157.

[21] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks *via* gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 618-626.

[22] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[M]// Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016: 21-37.

[23] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE Press, 2017: 1137-1149.

[24] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the 16th IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2999-3007.

Aero-engine defect detection by fusing attention and multi-scale features

ZHAO Chonglin[1]，ZHU Jiang[2]，HU Yongjin[3]，LI Zuze[3]，WANG Pengju[3]，XIE Tao[2,*]

(1. Faculty of Information Engineering and Automation，Kunming University of Science and Technology，Kunming 650500，China；
2. Faculty of Civil Aviation and Aeronautics，Kunming University of Science and Technology，Kunming 650500，China；
3. Yunnan Branch of China Eastern Airlines Technology Company Limited，Kunming 650200，China)

Abstract：The structural integrity of the aero-engine is related to flight safety. Currently, the detection of aero-engine defects based on borescope technology is mainly manually operated. In order to improve detection accuracy and efficiency, an intelligent aero-engine defect detection algorithm by fusing attention and multi-scale features was proposed to assist the borescope detection. In view of the imbalanced distribution of defect classes in original borescope images, a multi-sample fusion data augmentation method based on geometric transformation and Poisson image editing was used to enrich the small sample images, and the defect dataset was constructed. The coordination attention (CA) mechanism was integrated into the baseline network YOLOv5 to emphasize the extraction of defect features and enhance the network's distinction between defect targets and complex backgrounds. The weighted bidirectional feature pyramid network structure (BiFPN) was constructed in the neck network to achieve a higher level of feature fusion and improve the expression ability for multi-scale targets. The bounding box regression loss function was defined as the efficient intersection over union (EIOU) loss. The fast and accurate location and recognition of defects were realized. The experimental results show that the average precision of the proposed algorithm in detecting defects is 89.7%, 6.3% higher than that of the baseline network. The size of the trained model is only 14.0 M. Therefore, the proposed algorithm can effectively detect the main defects of aero-engines.

Keywords：aero-engine；defect detection；deep learning；YOLOv5；attention mechanism

Received：2023-03-29；**Accepted**：2023-04-21；**Published Online**：2023-05-06 17：56
URL：link.cnki.net/urlid/11.2625.V.20230506.0932.001

* **Corresponding author**. E-mail：77183118@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0157

多约束下升力式再入飞行器可达域计算方法

冉云霆，泮斌峰*
(西北工业大学 航天学院，西安 710072)

摘 要：针对升力式再入飞行器再入过程中的热流率、过载和动压受约束问题，提出一种多约束条件下的再入可达域计算方法。基于虚目标法建立了不考虑状态不等式过程约束的再入可达域最优控制模型；基于无过程约束问题的最优控制设计预测校正方法，将状态不等式约束转换为控制不等式约束；以X-33再入飞行器为对象完成数值仿真验证。数值仿真结果表明：与传统的依赖于平衡滑翔假设的“软约束”方法相比，所提方法可实现“硬约束”，即使飞行器在短时间内进行大幅度机动，各项过程约束仍能严格满足。

关 键 词：可达域；再入飞行器；过程约束；平衡滑翔；间接法
中图分类号：V448.235
文献标志码：A **文章编号**：1001-5965(2025)03-0904-06

升力式再入飞行器是一种采用升力体气动布局的再入飞行器，具有较高的升阻比和较强的机动能力[1]。可达域是指飞行器在满足各种过程约束和终端约束的条件下所能到达区域的集合[2-3]。可达域反映了飞行器在再入过程中的机动范围和飞行能力，可为升力式再入飞行器再入轨迹的选择和优化提供支撑，对于确保再入飞行器的安全返回具有极其重要的作用[4-5]。

目前，可达域的计算方法主要包括椭圆近似法、遍历法等，其中：椭圆近似法将可达域近似为椭圆形，利用最近纵程、最远横程和最远纵程3个末端点计算椭圆参数形成边界[6-7]，该方法复杂程度低、求解速度快，容易实现，但可达域精度较低；遍历法主要通过数值优化方法求解不同纵程下对应的最远横程问题[8]，将所得到的若干轨迹末端点连接构成可达域边界。相较于椭圆近似法，遍历法具有较高的精度，国内外已在此方面展开了大量的研究。文献[9]提出一种基于高斯伪谱法的多阶段技术，可用于复杂过程约束下的再入轨迹求解问题。文献[10]提出一种改进的序列凸优化方法，以提高轨迹求解的快速收敛性。文献[11]将差分进化算法和倾侧角插值相结合，以实现带过程不等式约束的可达域快速求解。基于准平衡滑翔假设，文献[12]利用庞特里亚金极大值原理与准平衡滑翔假设[13](quasi-equilibrium glide condition, QEGC)将最远横程问题转化成双参数搜索问题，其结果具有更高的精度，但运算量大，效率较低。文献[14]提出虚目标法并结合最近距离问题的最优解计算可达域边界，证明了最近距离问题与最远横程问题的等价性，将对应的最优控制问题转换成有限范围内的单参数搜索问题，在确保结果精度的基础上极大减少了运算量。但文献[14]对过程约束的处理是基于平衡滑翔假设。平衡滑翔假设作为一种“软约束”，在飞行器再入过程中并不严格要求满足约束条件，因此，这种约束转换方式常常无法满足热流率、气动过载和动压等硬约束条件，不仅显著影响飞行性能，甚至可能造成飞行器结构的严重损伤和破坏。

本文提出一种多约束条件下再入可达域计算的改进方法。该方法不依赖于平衡滑翔假设，采用

收稿日期：2023-04-03；录用日期：2023-09-22；网络出版时间：2023-11-13 13:11
网络出版地址：link.cnki.net/urlid/11.2625.V.20231110.1011.001
*通信作者. E-mail：panbinfeng@nwpu.edu.cn

引用格式：冉云霆，泮斌峰. 多约束下升力式再入飞行器可达域计算方法[J]. 北京航空航天大学学报，2025，51(3)：904-909.
RAN Y T，PAN B F. A method for calculating reachability regions of lift re-entry vehicles with multiple constraints[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：904-909 (in Chinese).

预测校正方法将热流率、气动过载和动压等状态不等式约束转换为控制不等式约束, 并完成再入可达域的数值优化求解, 能确保硬约束条件的严格满足。

1　再入飞行器动力学模型

经过无量纲化处理的三自由度球形旋转地球坐标系下的升力式再入飞行器动力学模型如下:

$$\begin{cases}\dot{r}=V\sin\gamma\\ \dot{\theta}=V\cos\gamma\sin\psi/(r\cos\varphi)\\ \dot{\varphi}=V\cos\gamma\cos\psi/r\\ \dot{V}=-D-\sin\gamma/r^2+\Omega^2 r\cos\varphi(\sin\gamma\cos\varphi-\cos\gamma\sin\varphi\cos\psi)\\ \dot{\gamma}=(L\cos\sigma+(V^2-1/r)\cos\gamma/r+2\Omega V\cos\varphi\sin\psi+\\ \quad \Omega^2 r\cos\varphi(\cos\gamma\cos\varphi+\sin\gamma\cos\psi\sin\varphi))/V\\ \dot{\psi}=(L\sin\sigma/\cos\gamma+V^2/r\cos\gamma\sin\psi\tan\varphi-2\Omega V(\tan\gamma\cdot\\ \quad \cos\varphi\cos\psi-\sin\varphi)+\Omega^2 r/\cos\gamma\sin\psi\cos\psi\sin\varphi)/V\end{cases} \tag{1}$$

式中: r为无量纲地心距; θ和φ分别为飞行器的经度和纬度; V为飞行器相对地球的无量纲速度; γ为速度倾角; ψ为速度方向角, 以当地正北方向起顺时针旋转为正; Ω为无量纲地球自转角速度; σ为再入飞行器的倾斜角; L和D分别为无量纲气动升力和气动阻力, 其表达式为

$$\begin{cases}L=0.5\rho S_{\text{ref}}R_0C_LV^2/m\\ D=0.5\rho S_{\text{ref}}R_0C_DV^2/m\end{cases} \tag{2}$$

式中: m为飞行器质量; R_0为地球半径; S_{ref}为飞行器参考面积; ρ为大气密度, 采用随高度变化的标准大气模型, 其表达式为$\rho=\rho_0\text{e}^{-h/h_0}$; C_L和C_D分别为升力系数和阻力系数, 由迎角α和速度V决定, 迎角采用如下的剖面设计形式:

$$\alpha=\begin{cases}45^\circ & Ma>10\\ 45^\circ-0.612(Ma-10)^2(^\circ) & 10>Ma>2.5\end{cases} \tag{3}$$

考虑热流率、气动过载和动压约束:

$$\begin{cases}\dot{Q}=k_Q\sqrt{\rho/R_s}(VV_{\text{scale}})^{3.15}\leqslant\dot{Q}_{\max}\\ n=\sqrt{L^2+D^2}\leqslant n_{\max}\\ q=0.5\rho(VV_{\text{scale}})^2\leqslant q_{\max}\end{cases} \tag{4}$$

式中: $\dot{Q}$为热流率; R_s为飞行器驻点等效半径; k_Q为是与飞行器结构相关的热流系数; V_{scale}为速度无量纲化系数; n为过载; q为动压。

端点约束包括再入点的初始状态约束和滑翔段与终端能量管理段(terminal energy management, TEAM)交接点处的状态约束。再入点状态约束为全状态约束: $\boldsymbol{X}_0=[r_0,\theta_0,\varphi_0,V_0,\gamma_0,\psi_0]^{\text{T}}$。终端状态约束为部分状态约束, 一般指对终端高度、速度及位置的约束, 但考虑到升力再入式飞行器的高度与速度间存在的能量耦合关系, 通常将终端高度与速度的约束整合为对终端能量的约束:

$$|e_{\text{f}}|=\left|1/r_{\text{f}}-V_{\text{f}}^2/2\right|\leqslant\Delta e \tag{5}$$

2　再入可达域计算

2.1　不考虑过程约束的可达域

本文采用虚目标方法建立不考虑过程约束的可达域最优控制模型。如图 1 所示, 虚目标方法的基本思想是通过将可达域边界求解转换为虚目标的最近距离问题[14], 具有计算效率高、收敛性好等优势。

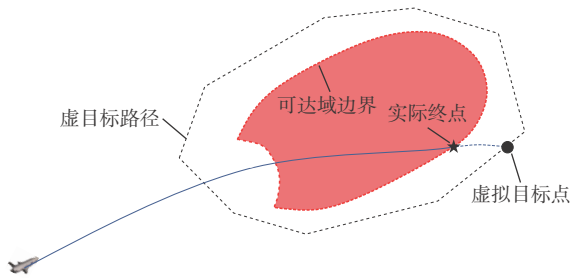


图 1　利用虚目标法求解可达域示意图

Fig. 1　Schematic diagram for solving reachability domain using virtual target method

最近距离问题要求当再入飞行器满足终端能量约束时与目标点的距离尽可能小, 结合球面距离公式, 可得最近距离问题的性能指标为

$$J=\min(-\cos s_{\text{f}})=\\ \min(-\sin\varphi_{\text{T}}\sin\varphi_{\text{f}}-\cos\varphi_{\text{T}}\cos\varphi_{\text{f}}\cos(\theta_{\text{T}}-\theta_{\text{f}})) \tag{6}$$

式中: $[\theta_{\text{T}},\varphi_{\text{T}}]$为目标点经纬度; $[\theta_{\text{f}},\varphi_{\text{f}}]$为飞行器实际终点经纬度。对应无过程约束最优控制问题的哈密顿函数为

$$H(\boldsymbol{X},u,\boldsymbol{p},t)=\boldsymbol{p}^{\text{T}}\boldsymbol{f}(\boldsymbol{X},u,t) \tag{7}$$

式中: $\boldsymbol{p}$为协态变量; u为控制变量; t为时间变量。将哈密顿函数对状态变量求导, 得到该问题的协态方程:

$$\dot{\boldsymbol{p}}=-\frac{\partial H}{\partial\boldsymbol{X}} \tag{8}$$

其与状态方程(式 (1))共同组成正则方程。根据庞特里亚金极小值原理, 若存在最优控制$u^*(t)$使得性能指标取极值, 则在对应的最优轨迹上哈密顿函数也应取极小值, 即

$$H(\boldsymbol{X}^*,\boldsymbol{p}^*,u^*,t)=\min H(\boldsymbol{X},\boldsymbol{p},u,t) \tag{9}$$

由于飞行器再入过程具有很强的非线性, 即使在不考虑过程约束的情况下, 也难以解析求解上述最优控制问题。因此, 在再入飞行器经过初始下降

段并进入平衡滑翔段后，引入平衡滑翔假设，飞行器的速度倾角很小且变化率近似为零($\dot{\gamma}=0$)，同时忽略地球的自转效应，式(1)中速度倾角的动力学约束项可被化简为

$$L\cos\sigma-\left(1/r-V^2\right)/r=0 \tag{10}$$

经过推导，可以得到最优控制[14]：

$$\begin{aligned}\tan\sigma=&\\&(1-1/V^2)(\sin\kappa\sin\varphi-\sin\kappa\tan\varphi_{\mathrm{T}}\cos\theta\cos\varphi+\\&\cos\kappa\sin\theta\cos\varphi)(\sin\kappa\cos\varphi\sin\psi-\\&\sin\kappa\tan\varphi_{\mathrm{T}}(\sin\theta\cos\psi-\cos\theta\sin\psi\sin\varphi)-\\&\cos\kappa(\sin\theta\sin\psi\sin\varphi+\cos\theta\cos\psi))\end{aligned} \tag{11}$$

式中：$\kappa\in[-\pi,\pi]$为待定参数。对应最优控制问题的横截条件为

$$\begin{cases}p_{\theta_{\mathrm{f}}}=\sin\kappa=\cos\varphi_{\mathrm{T}}\cos\varphi_{\mathrm{f}}\sin(\theta_{\mathrm{T}}-\theta_{\mathrm{f}})\\p_{\varphi_{\mathrm{f}}}=-\sin\kappa\tan\varphi_{\mathrm{T}}\sin\theta_{\mathrm{f}}-\cos\kappa\cos\theta_{\mathrm{f}}=\sin\varphi_{\mathrm{T}}\cos\varphi_{\mathrm{f}}-\\\qquad\cos\varphi_{\mathrm{T}}\sin\varphi_{\mathrm{f}}\cos(\theta_{\mathrm{T}}-\theta_{\mathrm{f}})\\p_{\psi_{\mathrm{f}}}=-\sin\kappa\sin\varphi_{\mathrm{f}}-\cos\varphi_{\mathrm{f}}(-\sin\kappa\tan\varphi_{\mathrm{T}}\cos\theta_{\mathrm{f}}+\\\qquad\cos\kappa\sin\theta_{\mathrm{f}})=0\end{cases} \tag{12}$$

综上，对于给定的初始条件和终端条件，无过程约束条件下最近距离问题的最优控制将仅由一个待定参数κ决定。因此，求解最近距离问题转变为：在$[-\pi,\pi]$范围内寻找一个合适的参数κ，进而确定最优控制，使得飞行器到达终端时需满足终端约束条件。

2.2 考虑过程约束的可达域计算

考虑过程约束的可达域计算通常采用基于平衡滑翔假设的处理方法[14-15]，即将热流率、气动过载和动压约束转换为对应的最低高度$r_{\min}$及对应的最大升力$L_{\max}$，并令式(10)中$r\to1$，从而得到对应的最大倾斜角$\sigma_{\max}$：

$$\sigma_{\max}=\left(1-V^2\right)/L_{\max} \tag{13}$$

由于平衡滑翔假设并不是强约束，在飞行器进行大幅度机动时，倾斜角长时间处于较大值，用于保持高度的升力部分将不足以维持平衡滑翔状态。此时，这种基于平衡滑翔假设的约束方式极有可能失效，这对飞行器结构和性能将产生巨大影响，甚至可能导致飞行器解体。

因此，本文提出一种基于预测校正的可达域计算方法，能够确保飞行器满足热流、过载和动压等硬约束条件。该方法的基本思想是将不考虑过程约束的可达域计算最优控制作为预测控制，借鉴文献[16]中对一阶非线性状态约束的处理方法，本文基于二阶泰勒级数将状态不等式约束转换为控制不等式约束，并将其作为校正手段，从而可以进一步确定满足过程约束的最优控制。

由于热流率、气动过载和动压均为状态变量$\boldsymbol{X}(t)$和时间t的函数，为便于描述，将过程约束记为

$$\boldsymbol{g}(\boldsymbol{X}(t),t)\geqslant0 \tag{14}$$

对$\boldsymbol{g}$在t时刻进行泰勒级数展开，取其前两阶并忽略高阶项，即

$$\begin{aligned}\boldsymbol{g}(\boldsymbol{X}(t),t+\mathrm{d}t)\approx&\boldsymbol{g}(\boldsymbol{X}(t),t)+\boldsymbol{g}'(\boldsymbol{X}(t),t)\mathrm{d}t+\\&\frac{1}{2}\boldsymbol{g}''(\boldsymbol{X}(t),t)\mathrm{d}t^2\end{aligned} \tag{15}$$

式中：过程约束$\boldsymbol{g}$对时间的一、二阶导数涉及状态变量$\boldsymbol{X}$对时间的求导，而根据式(2)和式(4)，此处的状态变量$\boldsymbol{X}$仅包含r和V，即$\boldsymbol{g}(r(t),V(t))$。结合式(1)可知，这组状态变量的一阶导数中不含控制变量σ，但其二阶导数中包含了$\dot{\gamma}$和$\dot{\psi}$，这两项中存在含σ的项$\cos\sigma$和$\sin\sigma$。将式(15)代入式(14)，并分离出$\boldsymbol{g}''(\boldsymbol{X}(\sigma,t),t)$：

$$\boldsymbol{g}''(\boldsymbol{X}(\sigma,t),t)\geqslant-2\frac{\boldsymbol{g}(\boldsymbol{X}(t),t)+\boldsymbol{g}'(\boldsymbol{X}(t),t)\mathrm{d}t}{\mathrm{d}t^2} \tag{16}$$

在针对具体不同的约束时，将$\boldsymbol{g}''(\boldsymbol{X}(\sigma,t),t)$中含$\cos\sigma$和$\sin\sigma$的项进一步分离即可完成过程约束向控制约束的转换。事实上，这种不等式约束转换方法并不依赖于平衡滑翔等其他假设，理应具有强约束的性质，可使飞行器严格满足各项过程约束。

由于热流率、气动过载和动压均为r和V的函数，因此，对应的控制约束具有高度的相似性。其在二阶泰勒级数下的形式为

$$\begin{cases}\dot{Q}(r,V,t+\mathrm{d}t)\approx\dot{Q}(r,V,t)+\\\qquad\dot{Q}'(\boldsymbol{X}(t),t)\mathrm{d}t+\dfrac{1}{2}\dot{Q}''(\boldsymbol{X}(t),t)\mathrm{d}t^2\leqslant\dot{Q}_{\max}\\n(r,V,t+\mathrm{d}t)\approx n(r,V,t)+\\\qquad n'(\boldsymbol{X}(t),t)\mathrm{d}t+\dfrac{1}{2}n''(\boldsymbol{X}(t),t)\mathrm{d}t^2\leqslant n_{\max}\\q(r,V,t+\mathrm{d}t)\approx q(r,V,t)+\\\qquad q'(X(t),t)\mathrm{d}t+\dfrac{1}{2}q''(\boldsymbol{X}(t),t)\mathrm{d}t^2\leqslant q_{\max}\end{cases} \tag{17}$$

结合再入飞行器的动力学方程式(1)，可推导出各个约束一、二阶导数的具体形式，并代入式(17)，分离出包含倾斜角σ的项：

$$F\sin(\sigma+G)\leqslant E \tag{18}$$

式中：E,F,G均为再入飞行器状态量的函数。将式(18)进一步写为对σ的范围约束：

$$\sigma\begin{cases}\leqslant\arcsin(E/F)-G & F>0\\\geqslant\arcsin(E/F)-G & F<0\end{cases} \tag{19}$$

基于预测校正的约束处理方法成功将过程约束转换为控制约束。如图2所示，式(17)对$t+\mathrm{d}t$时刻的飞行器状态进行预测，判定过程约束是否被满

足, 并根据式 (19) 对倾斜角进行校正。由于预测是基于二阶泰勒级数进行的, dt的大小将直接影响到约束效果, dt过大会导致预测结果不准确, 过小则可能会超出飞行器控制能力极限, 从而引起约束失效。热流率、气动过载和动压均依赖于飞行器所处的高度和速度。当倾斜角较大时, 升力主要被用于机动转弯, 此时飞行器高度下降, 大气密度随之增加, 同时速度也在变大, 热流率、气动过载和动压将快速增加; 若倾斜角变小, 升力主要被用于抬升飞行高度, 大气密度和飞行器速度将停止快速上升, 两者的叠加效应会使得热流率、气动过载和动压迅速得到控制。鉴于这 3 个约束指标对倾斜角的调整具有较短的反应周期, 因此, dt可选择较小的固定值。

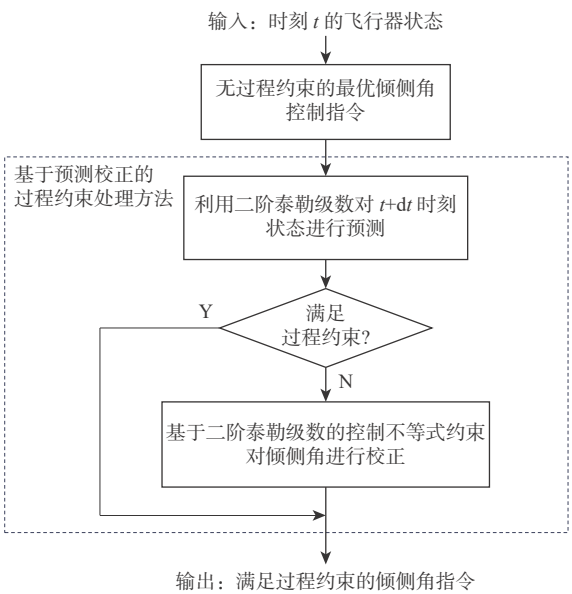


图 2　基于预测校正的约束处理算法流程

Fig. 2　Algorithm flowchart of constraint processing based on predictive correction

综上, 通过基于二阶泰勒级数的状态不等式约束成功将再入过程约束转换成再入控制约束, 结合由虚目标法得到的无过程约束最优控制律, 可得到含复杂过程约束的可达域求解方法。假设无过程约束最优控制律给出的倾斜角指令为σ^*, 由过程约束针对再入飞行器当前时刻的状态给出的倾斜角极值为$\sigma_{\min}$和$\sigma_{\max}$, 则倾斜角取值原则如下:

$$\sigma=\begin{cases}\sigma_{\min} & \sigma^*\leqslant\sigma_{\min}\\ \sigma^* & \sigma_{\min}<\sigma^*<\sigma_{\max}\\ \sigma_{\max} & \sigma^*\geqslant\sigma_{\max}\end{cases}\tag{20}$$

3　数值仿真验证与分析

本文选取水平着陆且可重复使用的升力再入式飞行器 X-33 为对象进行数值仿真验证。飞行器的质量为 37 362.9 kg, 气动参考面积为 149.388 m^2。各项过程约束选取如下: 热流率约束$\dot{Q}_{\max}$ = 850 kW/m^2, 气动过载约束$n_{\max}=2.5g$, 动压约束$q_{\max}$ = 15 000 kPa。飞行器再入初始状态如表 1 所示。终端高度和速度分别为$h_{\mathrm{f}}=30.427$ km, $V_{\mathrm{f}}=908.15$ m/s。飞行器最大倾斜角为 90°。

表 1　再入飞行器初始状态

Table 1　Initial state of re-entry vehicle

高度/km	速度/(m·s^{-1})	经度/(°)	纬度/(°)	速度倾角/(°)	速度方向角/(°)
79.5	6 100.0	243.1	−18.3	−1.0	38.3

数值仿真平台采用 C++, 计算机核心参数为 i9-13900k, 主频为 3.00 GHz。采用基于平衡滑翔假设的约束处理方法时, 平均每条轨迹耗时 0.043 136 s, 而采用本文方法时平均每条轨迹需 0.046 654 s, 仅增加 8%。图 3 为 X-33 在施加过程约束前后的可达域范围。由图可知, 在施加过程约束后, 可达域前端并无明显变化, 而后端有了较大范围的减小。不难分析出在到达可达域前端的轨迹中, 倾斜角普遍具有较小的值, 因为小倾斜角可使飞行器的升力尽可能用于保持高度, 从而让飞行器飞过更远的距离。这种情况下飞行器拥有更低的高度下降速率, 其热流率、气动过载和动压都远低于约束值, 因此, 施加过程约束并不会对可达域前端的相关轨迹产生影响。为接近可达域后侧对应的虚目标点, 飞行器需要进行大幅度机动, 这会导致飞行器高度快速下降, 从而使得过程约束边界很快被突破, 所以在施加过程约束后, 可达域的后半部分出现了较大变化。

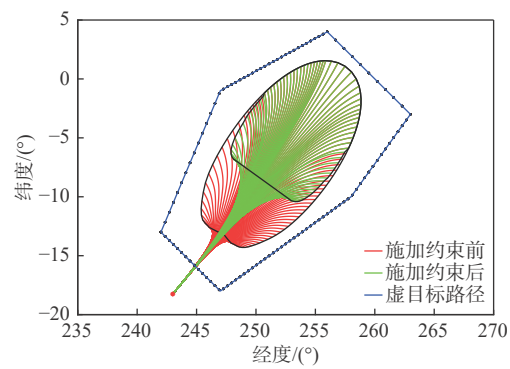


图 3　X-33 在施加约束前后的可达域预测结果

Fig. 3　Prediction results of reachability regions of X-33 before and after applying constraints

如图 4 所示, 热流率、气动过载和动压约束共同构成了再入走廊的下边界, 在飞行器再入的初期和末期, 热流率和动压约束分别起主要约束作用, 而在再入的绝大部分时间中, 气动过载约束起主要约束作用。此外, 受到过程约束边界影响的轨迹均具有更大的跳跃幅度, 其恰好对应了可达域两侧大

幅度机动的轨迹，而这些轨迹上方的轨迹簇高度下降速率较慢，基本处于平衡滑翔约束曲线周围，且全程从未接近约束下边界，这与到达可达域前方的轨迹相对应。

如图5所示，基于平衡滑翔假设和泰勒级数的约束形式对比结果，前者并没有使飞行器严格遵循各项过程约束，而本文方法可使过程约束在所有情况下均得到严格满足。由图4可知，接近再入走廊下边界的轨迹往往都偏离了平衡滑翔约束曲线，此时飞行器与平衡滑翔状态偏差较大，因此，导致了这种约束形式的失效，这与2.2节中的分析是一致的。与此同时，图5所示各过程约束的变化历程又说明了本文方法的有效性，这是因为基于泰勒级数的近似预测精度并不会受到飞行器状态和轨迹变化的影响。图6为再入轨迹倾斜角曲线，各个约束对倾斜角的限制也较为平和，并没有超出最大值的限制。

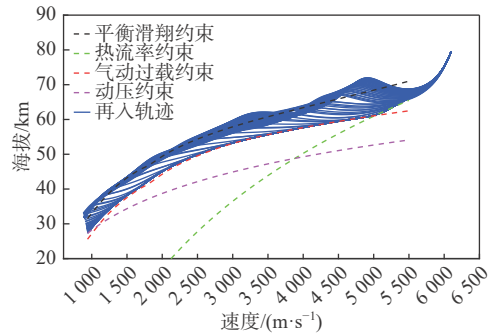
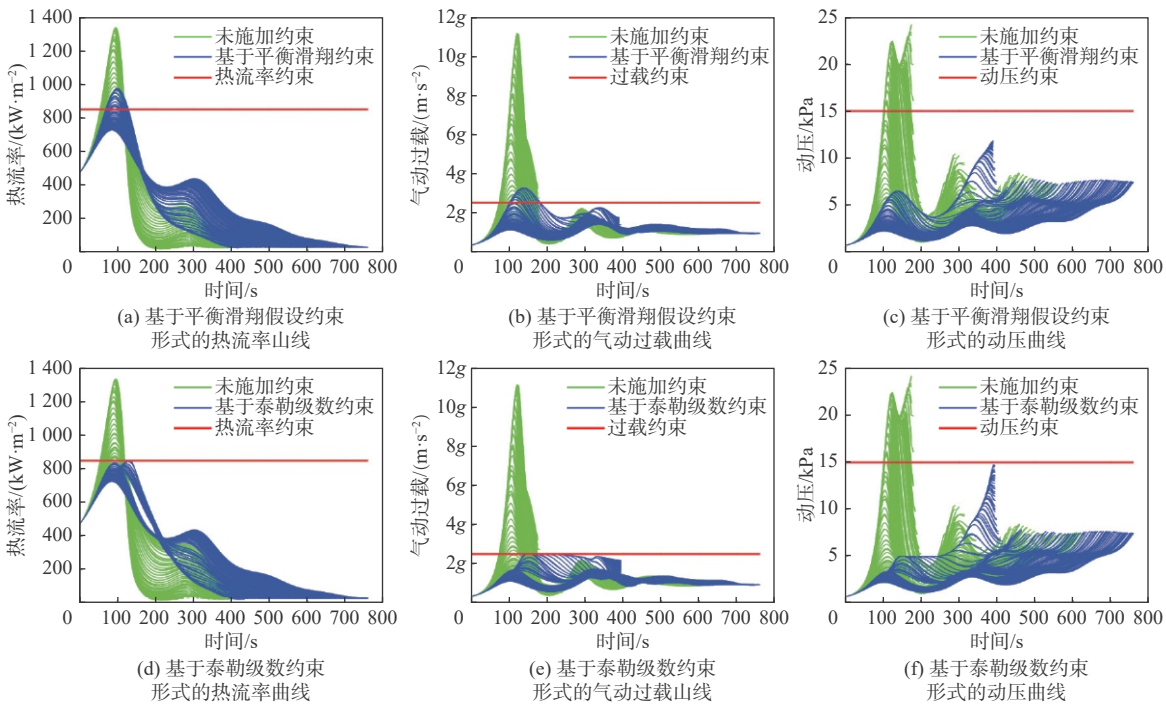


图4 再入轨迹的高度-速度曲线

Fig. 4 Altitude velocity curve of re-entry trajectory



(a) 基于平衡滑翔假设约束形式的热流率山线 (b) 基于平衡滑翔假设约束形式的气动过载曲线 (c) 基于平衡滑翔假设约束形式的动压曲线

(d) 基于泰勒级数约束形式的热流率曲线 (e) 基于泰勒级数约束形式的气动过载山线 (f) 基于泰勒级数约束形式的动压曲线

图5 基于平衡滑翔和泰勒级数约束形式的热流率、过载和动压曲线

Fig. 5 Heat flow rate, overload, and dynamic pressure curves based on quasi-equilibrium glide conditions and Taylor series constraint forms

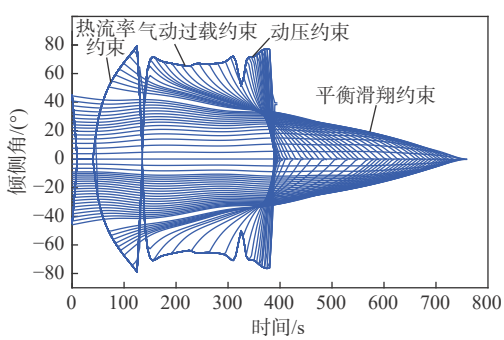


图6 再入轨迹的倾斜角曲线

Fig. 6 Inclination angle curve of re-entry entrant trajectory

4 结 论

本文提出一种基于预测校正的多约束下升力式再入飞行器可达域计算方法，研究分析和仿真结果表明：

1）本文方法易于实现，能够快速计算生成升力式再入飞行器可达域边界；

2）本文方法不依赖于平衡滑翔假设，能够严格满足热流率、气动过载和动压约束等硬约束条件；

3）本文方法具有一定的通用性，还可进一步应用于禁飞区约束和内点约束等其他过程约束，具有

广泛的应用前景。

参考文献（References）

[1] GUO H C, HUANG H B, YANG Y. Reentry trajectory planning and guidance method with no-fly zone constraints[C]//Proceedings of the China Automation Congress. Piscataway: IEEE Press, 2021: 4633-4638.

[2] SARAF A, LEAVITT J, FERCH M, et al. Landing footprint computation for entry vehicles[C]//Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit. Reston: AIAA, 2004.

[3] FU S N, LU T Y, YIN J, et al. Rapid algorithm for generating entry landing footprints satisfying the No-fly zone constraint[J]. International Journal of Aerospace Engineering, 2021, 2021: 8827377.

[4] PENG Q, GUO J G, YANG S J. Genetic algorithm-based re-entry trajectory optimization and footprint determination for lifting body spacecraft[M]// Lecture Notes in Electrical Engineering. Berlin: Springer, 2021: 1369-1379.

[5] ZHOU C J, LI M J, SHAO L, et al. An improved predictor-corrector guidance algorithm for reentry glide vehicle based on fast landing points position prediction[M]// Lecture Notes in Electrical Engineering. Berlin: Springer, 2023: 639-649.

[6] 吴楠, 王锋, 赵敏, 等. 高超声速滑翔再入飞行器的可达区快速预测[J]. 国防科技大学学报, 2021, 43(1): 1-6.
WU N, WANG F, ZHAO M, et al. Fast prediction for footprint of hypersonic glide reentry vehicle[J]. Journal of National University of Defense Technology, 2021, 43(1): 1-6(in Chinese).

[7] 冯必鸣, 聂万胜, 李柯. 再入飞行器可达区域近似算法及地面覆盖研究[J]. 航天控制, 2012, 30(6): 43-49.
FENG B M, NIE W S, LI K. Research on closest-aproach of footprint and coverage for reentry vehicle[J]. Aerospace Control, 2012, 30(6): 43-49(in Chinese).

[8] VINH N . Optimal trajectories in atmospheric flight[J]. Space Mankinds Fourth Environment, 1982: 449-468.

[9] ZHAO J, ZHOU R. Reentry trajectory optimization for hypersonic vehicle satisfying complex constraints[J]. Chinese Journal of Aeronautics, 2013, 26(6): 1544-1553.

[10] PEI P, FAN S P, WANG W, et al. Online reentry trajectory optimization using modified sequential convex programming for hypersonic vehicle[J]. IEEE Access, 2021, 9: 23511-23525.

[11] 蔺君, 何英姿, 黄盘兴. 基于差分进化算法的再入可达域快速计算[J]. 中国空间科学技术, 2020, 40(4): 54-60.
LIN J, HE Y Z, HUANG P X. Fast reentry landing footprint calculation using differential evolution algorithm[J]. Chinese Space Science and Technology, 2020, 40(4): 54-60(in Chinese).

[12] 肖翔, 杨业, 郭涛, 等. 升力式再入飞行器两种可达区域计算方法的探讨[J]. 航天控制, 2015, 33(2): 39-43.
XIAO X, YANG Y, GUO T, et al. Discussion of two algorithms for generating landing footprint for lifting reentry vehicles[J]. Aerospace Control, 2015, 33(2): 39-43(in Chinese).

[13] LU P. Asymptotic analysis of quasi-equilibrium glide in lifting entry flight[J]. Journal of Guidance, Control, and Dynamics, 2006, 29(3): 662-670.

[14] LU P, XUE S B. Rapid generation of accurate entry landing footprints[J]. Journal of Guidance, Control, and Dynamics, 2010, 33(3): 756-767.

[15] 李惠峰, 谢陵. 基于预测校正方法的 RLV 再入制导律设计[J]. 北京航空航天大学学报, 2009, 35(11): 1344-1348.
LI H F, XIE L. Reentry guidance law design for RLV based on predictor-corrector method[J]. Journal of Beijing University of Aeronautics and Astronautics, 2009, 35(11): 1344-1348(in Chinese).

[16] LU P. Non-linear systems with control and state constraints[J]. Optimal Control Applications and Methods, 1997, 18(5): 313-326.

A method for calculating reachability regions of lift re-entry vehicles with multiple constraints

RAN Yunting，PAN Binfeng*

(School of Astronautics，Northwestern Polytechnical University，Xi'an 710072，China)

Abstract: In view of the heat flow, overload, and dynamic pressure constraint problems during the reentry process of the lift reentry vehicle, a calculation method of the reentry reachability region under multiple constraints is designed. Using the virtual target method, an optimal control model for the reentry reachability region is developed without the state inequality restrictions. A predictive correction method is designed based on the optimal control without process constraints, which converts state inequality constraints into control inequality constraints. The numerical simulation of the X-33 is completed. The results of the numerical simulation demonstrate that, in contrast to the conventional "soft constraint" method, which depends on the quasi-equilibrium glide conditions, the suggested method is capable of achieving "hard constraint," meaning that all process constraints can still be strictly satisfied even in the event that the aircraft makes large maneuvers quickly.

Keywords: reachable region; reentry vehicle; process constraints; balanced glide; indirect method

Received：2023-04-03；**Accepted**：2023-09-22；**Published Online**：2023-11-13 13：11

URL：link.cnki.net/urlid/11.2625.V.20231110.1011.001

*** Corresponding author**. E-mail：panbinfeng@nwpu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0179

多谱自适应小波和盲源分离耦合的生理信号降噪方法

王振宇[1,2]，向泽锐[1,3,*]，支锦亦[1,3]，丁铁成[1]，邹瑞[1]

(1. 西南交通大学 设计艺术学院，成都 610031；2. 同济大学 设计创意学院，上海 200092；
3. 西南交通大学 人机环境系统设计研究所，成都 611730)

摘 要：为提高生理信号的质量和可靠性，将盲源分离和小波阈值方法进行耦合研究，提出了多谱自适应小波信号增强方法并与改进的盲源分离方法相结合进行降噪处理。为评估所提方法的有效性，使用小波变换中软阈值、硬阈值、自适应阈值3种方法计算信噪比（SNR）和均方根误差（RMSE）。结果表明：所提方法在软阈值下具有较强的适用性，增强后的信号软阈值相比硬阈值，SNR提升约44.2%，RMSE下降约28.8%，处理时间减少约1.4%。软阈值相比自适应阈值，SNR提升约706%，RMSE下降约16.7%，处理时间减少约3.0%。为对比软阈值下各参数差异，使用软阈值对原始信号、加噪信号和增强信号进行对比分析及归一化处理。结果显示增强后的信号具有较好的SNR、较低的RMSE和较短的处理时间，软阈值下增强后的信号与原始信号相比，SNR提升约0.12%，RMSE下降约2.5%，处理时间减少约3.9%，进一步验证了所提方法的有效性，并提高了信号质量。

关 键 词：多谱自适应小波；盲源分离；小波变换；降噪方法；生理信号
中图分类号：TP391.41
文献标志码：A **文章编号**：1001-5965(2025)03-0910-12

生理信号是从生物体内部或外部获取并反映生命活动的信号[1]。通过对生理信号的采集、分析和处理，以获得有关人体健康、疾病诊断、治疗效果等方面的信息，在医学诊断和生物信号处理领域具有广泛的应用[2-3]。然而，由于采集过程中存在肌肉运动、电源干扰等多种不可避免的干扰因素，极易导致生理信号中存在较多噪声，从而限制了其可靠性和准确性。因此，降噪方法成为解决该问题的有效途径。在众多的降噪方法中，盲源分离和小波阈值方法是2种被广泛应用的方法[4-5]。盲源分离(blind source separation, BSS)是一种通过对多个混合信号进行分离，使信号彼此独立的信号分离方法[6]，其适用于多个混合信号之间不存在先验知识的情况，可以有效去除信号中的噪声和干扰。在生理信号降噪中，盲源分离方法被广泛应用于脑电图(electroencephalogram, EEG)数据、心电图(electrocardiogram, ECG)数据、肌电图(electromyogram, EMG)等领域[7-8]。使用独立成分分析(independent component analysis, ICA)算法进行盲源分离，可以将原始信号分离成多个独立成分，从而最小化噪声和干扰对原始信号的影响。小波阈值方法是一种基于小波变换的信号降噪方法，通过对小波系数进行阈值处理，将信号中的噪声和干扰去除[9]。

文献[10]提出了一种基于盲源分离和小波阈值的脑电信号降噪方法，该方法使用ICA技术完成盲源分离，将原始数据分解为独立的源信号，对每

收稿日期：2023-04-14；**录用日期**：2023-06-16；**网络出版时间**：2023-07-19 11:01
网络出版地址：link.cnki.net/urlid/11.2625.V.20230718.1851.003
基金项目：教育部2022年第二批产学合作协同育人项目(220705329291641)；西南交通大学新型交叉学科培育基金(YG2022006)
* **通信作者**. E-mail：xiangzerui@163.com

引用格式：王振宇，向泽锐，支锦亦，等. 多谱自适应小波和盲源分离耦合的生理信号降噪方法[J]. 北京航空航天大学学报，2025，51(3)：910-921. WANG Z Y，XIANG Z R，ZHI J Y，et al. Physiological signal denoising method based on multi-spectrum adaptive wavelet and blind source separation[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：910-921 (in Chinese).

个源信号使用小波阈值去噪技术进行降噪。实验结果表明, 该方法可以有效提高脑电信号的信噪比, 提高信号质量和可靠性。文献 [11] 提出了一种结合盲源分离和小波阈值去噪技术的心电信号降噪方法。该方法使用盲源分离技术解决多通道心电信号中的相互干扰问题, 使用小波阈值去噪技术减小噪声干扰, 实验结果表明, 该方法可以有效提高心电信号的信噪比和减小噪声干扰。文献 [12] 提出了一种基于盲源分离和小波去噪技术的多通道骨骼肌电信号处理方法, 该方法使用 ICA 完成盲源分离, 将原始数据分解为独立的源信号, 对每个源信号使用小波去噪技术处理, 通过实验验证了该方法可以有效提高骨骼肌电信号的信噪比和减小噪声干扰。文献 [13] 通过卷积的方法对多通道表面肌电图进行盲源分离, 提出了分解多通道侵入性和非侵入性 EMG 信号的框架, 从而研究了大量同时活动的电极单元行为。综上所述, 盲源分离和小波阈值去噪技术是常用的信号处理方法, 能够有效提高生物电信号的信噪比和减小噪声干扰。

在实际工程应用中, 生理信号存在延迟混合和工频干扰等问题, 这些问题是不可避免的[14-15]。为解决上述问题, 提出了一种欠确定盲源分离算法 (unsupervised source counting algorithm, USCA)。与传统空间协方差分析 (spatial covariance analysis, SCA) 方法不同, USCA 采用稀疏表示和凸优化原理相结合的算法, 可以从多个混叠信号中恢复原始信号, 实现信号的分离和提取。为增强信号质量, 提出了一种改进的多谱自适应小波去噪信号增强 (multispectral adaptive wavelet denoising, MAWD) 方法, 将该方法应用于生理信号降噪处理, 以便提高信号的可靠性和准确性。

1　方法及流程

1.1　方　法

1.1.1　小波阈值

小波阈值是基于小波变换 (wavelet transform, WT) 来实现信号去噪的方法。该方法于 1993 年提出[16-17]。小波阈值的目的是消除信号中的噪声, 同时保留信号中有用的信息。其处理的基本思想是将信号分解为不同尺度的小波系数, 然后根据预设的阈值大小对小波系数进行裁剪或压缩。具体而言, 若某一小波系数的绝对值小于阈值, 则将该系数置 0, 否则保留原值。通过这样的处理, 可以滤除信号中的噪声, 并使信号更加平滑, 从而达到去噪的目的。小波阈值在信号处理、图像处理、语音处理等领域都有广泛的应用。该方法可以有效地消除噪声对信号质量的影响, 同时保留信号中的有用信息, 对于一些需要高信噪比的应用场景尤为重要。WT 的推导步骤如下。

步骤 1　设原始信号为 $x(t)$, t 为连续时间。小波变换后得到一组小波系数 $C_{j,k}$, 其中, j 为尺度, k 为离散时间, k=1, 2, ⋯, N, N 为总采样个数。设阈值为 T, 用于控制小波系数的阈值大小。具体的阈值选择方法可以根据具体应用和需求来确定, 例如硬阈值、软阈值或自适应阈值选择方法。则 WT 可表示为

$$\hat{C}_{j,k}=\begin{cases} C_{j,k} & |C_{j,k}|\geqslant T \\ 0 & |C_{j,k}|<T \end{cases} \tag{1}$$

式中: $\hat{C}_{j,k}$ 为处理后的小波系数。

步骤 2　将处理后的小波系数进行逆变换, 得到去噪后的信号 $y(t)$ 为

$$y(t)=\sum_{j=1}^{J}\sum_{k}\hat{C}_{j,k}\psi_{j,k}(t) \tag{2}$$

式中: $\psi_{j,k}(t)$ 为小波函数。

步骤 3　在实际应用中, 可通过最大信息熵准则来确定阈值大小:

$$T=\sigma\sqrt{2\lg N_{\mathrm{w}}} \tag{3}$$

式中: σ 为所有小波系数的标准差; N_{w} 为小波系数的总数。式 (3) 是通过最大信息熵准则得出的, 可以确保处理后的信号具有最大的信噪比。

WT 作为一种信号降噪方法, 具有保留有用信息、精度高、适用范围广、实时性强等优势, 但在阈值的选择、信号失真、多尺度分解等问题中存在缺陷。

1.1.2　多谱自适应小波

为弥补 WT 的缺陷, 本文提出 MAWD, 用于提高信号的质量和可靠性。其原理是基于小波变换的分析, 通过对信号进行分解和重构来去除噪声并增强信号。该方法的目的是提高信号的清晰度和可靠性, 使其更易于分析和使用[18]。通过去除噪声并增强信号, 可以有效地提高信号的信噪比, 从而减少误判和误差。MAWD 方法相对于传统小波去噪算法具有更强的自适应性、更全面的频谱分析, 提高了图像可视性, 能够更加精确地选择阈值, 适应不同信号的特征, 因此, 在信号增强方面效果更好。MAWD 方法计算步骤如下。

步骤 1　添加加性噪声和滤波。添加一个已知的加性噪声 (随机高斯噪声) 信号到原始信号中, 并对整个信号进行采样, 以使其具有 8 kHz 的通用采样率:

$$y(n)=x(n)+w(n) \tag{4}$$

式中：n为离散时间，用于表示信号和噪声的离散样本点；$y(n)$为含噪声的信号；$x(n)$为原始信号；$w(n)$为加性噪声；滤波可以采用有限脉冲响应低通滤波器（finite impulse response low-pass filter, FIR）或无限脉冲响应滤波器（infinite impulse response low-pass filter, IIR）等方法。

步骤 2 频谱估计及噪声估计。使用 Sine Tapers 窗函数对信号进行多普勒频谱估计（multiple signal classification Doppler spectral estimation, MTS）以估计信号的功率谱密度（power spectral density, PSD）和噪声水平，有助于选择合适的阈值来过滤噪声。正弦锥形窗函数 $w(n)$ 的计算为

$$w(n) = e(t,k) \qquad 1 \leqslant k \leqslant K \tag{5}$$

$$g(t,k) = e(t,k)/\mathrm{sqrt}(\mathrm{sum}(e(t,k)^2)) \tag{6}$$

式中：K为正弦锥形窗的宽度参数；k为窗口函数的离散时间，即窗口函数在离散时间点上的取值；$e(t,k)$为 Slepian 序列；$g(t,k)$为正弦锥形窗函数；sqrt 表示平方根函数；sum 表示求和函数。Slepian 序列可以通过求解 Toeplitz 特征值问题计算得到。噪声估计可以采用多锥窗谱估计方法 (multi-taper method, MTM) 方法。

步骤 3 定义信号噪声相关矩阵 $\boldsymbol{N}$ 和信号相关矩阵 $\boldsymbol{R}$。

$$\begin{cases} N(j,k) = E(d(j)d^*(k)) \\ R(j,k) = E(x(j)x^*(k)) \end{cases} \tag{7}$$

其中：$N(j,k)$表示信号噪声相关矩阵中第 j 行第 k 列的元素，$R(j,k)$为信号相关矩阵的第 j 行第 k 列元素；$d^*(k)$和 $x^*(k)$分别为 $d(k)$和 $x(k)$的复共轭；E为期望值；$d(j)$为去噪后的信号的第 j 个分量；$x(j)$为原始信号的第 j 个分量。

步骤 4 计算 MTM 的噪声谱：

$$P_{\mathrm{noise}(f)} = (1/M)\cdot \mathrm{sum}\left\{N(j,k)\exp(-2pif|j-k|/K)\right\} \tag{8}$$

式中：M为序列长度；f为频率；p为常数（通常为 1）。

步骤 5 绘制标准差。通过分析 MTM，可以计算出信号的标准差，以及各子带的相对噪声水平。为进一步调整阈值参数，以优化去噪效果。构建 σ_t 随时间 t 的变化关系：

$$\sigma_t = \mathrm{std}(y(t-n+1:t)) \tag{9}$$

式中：std 为标准差函数；n为窗口的大小；σ表示标准差。

步骤 6 PSD 估计。使用 MTS 方法估计信号的 PSD，并检查其频谱特征，以确定某一频段受到噪声干扰较大。有助于选择合适的小波基函数和阈值。采用 MTM 进行功率谱密度估计，定义 MTM 估计的功率谱密度 $P_{\mathrm{MT}}(f)$：

$$P_{\mathrm{MT}}(f) = (1/K)\cdot \mathrm{sum}|d_k(f)|^2 \tag{10}$$

式中：K为正弦锥形窗的宽度参数；$d_k(f)$为使用第 k 个正弦锥形窗估计得到的信号频域分量。

步骤 7 离散小波变换 (discrete wavelet transform, DWT) + 阈值处理 + 维纳滤波（wiener filter, WF）。使用 DWT 和自适应阈值方法来去除噪声。使用阈值方法，可以选择合适的阈值来保留信号的主要特征，并消除噪声，使用 WF 进一步提高信噪比。对信号进行 DWT，然后再进行阈值化处理和 WF，如下：

$$d = W(x) \tag{11}$$

$$d_{\mathrm{h}} = \mathrm{HardThreshold}(d,T) \tag{12}$$

$$d_{\mathrm{wiener}} = \mathrm{WF}(d_{\mathrm{h}},\sigma_n^2/\sigma_x^2) \tag{13}$$

式中：d为 DWT 分解得到的系数集合；T为阈值；σ_n为噪声方差；σ_x为信号方差；W为 DWT 变换函数；HardThreshold 为硬阈值。

步骤 8 重构信号。使用反离散小波变换 (inverse discrete wavelet transform, IDWT) 进行信号重构：

$$x_{\mathrm{denoised}} = \mathrm{IDWT}(d_{\mathrm{wiener}}) \tag{14}$$

式中：d_{wiener}为经过维纳滤波器处理得到的小波系数；x_{denoised}为经过 WF 处理后的去噪信号。

步骤 9 对去噪后的信号进行评估，计算信噪比等指标：

$$f_{\mathrm{denoised}}(j) = \begin{cases} f(j)-t_j & f(j) > t_j \\ 0 & -t_j \leqslant f(j) \leqslant t_j \\ f(j)+t_j & f(j) < -t_j \end{cases} \tag{15}$$

式中：$f(j)$为信号在第 j 个小波系数上的值；t_j为阈值，可以根据信噪比和其他参数进行自适应选择。

步骤 10 重复去噪过程。如需多次重复去噪，可调整参数，如阈值和小波函数，以获得更好的去噪效果。同时，也应该对去噪结果进行交叉验证，以确保去噪结果的稳健性。

另外，可以通过多级小波分解和重构，得到更加精细的频域信息和更好的噪声去除效果，如下：

$$d_{j,k} = \langle f,\psi_{j,k}\rangle = \int_{-\infty}^{+\infty} f(x)\psi_{j,k}(x)\mathrm{d}x \tag{16}$$

$$f(x)' = \sum_{j,k} d_{j,k}\phi_{j,k}(x) \tag{17}$$

式中：$\psi_{j,k}$为小波基函数；$\phi_{j,k}$为对应的重构基函数；通过对小波系数 $d_{j,k}$进行阈值处理，即可实现噪声去除；$f(x)$为原始信号；$f(x)'$为重构信号。

1.1.3 盲源分离

盲源分离是一种信号处理方法，最早在 1987 年

提出, 旨在将混合的信号分离成多个独立的源信号[19-20]。盲源分离是指在不知道混合信号成分的情况下, 从多个混合信号中分离出独立的信号成分。其基本原理是利用各信号成分的统计特性, 在多个观测通道上进行协同处理, 通过一定的算法将目标信号分离出来。在降噪应用中, 盲源分离可以帮助去除混合信号中的噪声成分, 从而提高信号质量。对于生理信号而言, 盲源分离能够去除信号叠加、干扰等因素造成的噪声, 得到更加纯净的信号成分。在信号处理领域, 常用的盲源分离技术包括独立成分分析(independent component analysis, ICA)、主成分分析(principal component analysis, PCA)等。相较于其他方法, ICA 被广泛认为是一种更优秀的降噪技术, 因为其可以更准确地估计信号的独立成分, 并且适用于非高斯、非线性、非平稳等信号类型[21]。然而, 尽管 BSS 方法在信号处理、图像处理等领域有着广泛的应用, 但是其也存在一些缺陷, 例如算法复杂度高、知识先验不足、数据冗余等问题。

1.1.4　改进的盲源分离

为解决 BBS 的缺陷, 本文提出了 USCA 方法。该方法采用了稀疏表示和凸优化原理相结合的思路, 旨在从多个混叠信号中恢复原始信号, 实现信号的分离和提取。传统的盲源分离方法通常采用线性分离模型来分离信号, 但无法处理非线性、延迟混合等情况, 在某些场景下会导致失效。为解决以上问题, USCA 方法利用稀疏性假设和低秩矩阵分解的思想, 将盲源分离问题转化为一个凸优化问题, 并通过分解矩阵求解模型, 以实现盲源分离。相对于传统方法, USCA 方法能够更好地处理非线性、延迟混合等情况, 并且能够适应高维数据和大规模数据的处理。

改进的盲源分离 USCA 方法的目的是提高盲源分离的准确性和稳定性, 以应对更加复杂的实际环境。与传统的 BSS 方法相比, 改进的盲源分离 USCA 方法在处理复杂场景时更加出色, 并且能够自适应调整优化参数以获得更准确的结果, 同时具有更高的计算效率和更加坚实的理论基础。USCA 方法的推导公式及步骤如下:

步骤 1　构建 n 个混合信号, 每个混合信号可以表示为 $\boldsymbol{x}=\boldsymbol{As}$, 其中, $\boldsymbol{A}$ 为混合矩阵, $\boldsymbol{s}$ 为原始信号矩阵。

步骤 2　根据稀疏性假设和低秩矩阵分解的思想, 将原始信号矩阵 $\boldsymbol{s}$ 分解为 $\boldsymbol{s}=\boldsymbol{LZ}$, 其中, $\boldsymbol{L}$ 为一个低秩矩阵, $\boldsymbol{Z}$ 为一个稀疏矩阵。

步骤 3　将 $\boldsymbol{x}=\boldsymbol{As}$ 带入 $\boldsymbol{s}=\boldsymbol{LZ}$ 中, 得到 $\boldsymbol{x}=\boldsymbol{ALZ}$, 将 $\boldsymbol{x}$ 中的每个元素用向量形式表示, 即可得到 $|x_1|\leqslant\sqrt{\lambda_1},|x_2|\leqslant\sqrt{\lambda_2},\cdots,|x_n|\leqslant\sqrt{\lambda_n}$, 其中, λ_i 为矩阵 $\boldsymbol{AA}^{\mathrm{T}}$ 的第 i 个特征值。

步骤 4　将盲源分离问题转化为下列凸优化问题:

$$\begin{cases}\min\left(\|\boldsymbol{x}\|_1\right)\\ \quad\boldsymbol{x}=\boldsymbol{ALZ}\\ \quad\|\boldsymbol{x}\|_1=\sum\limits_{i=1}^{n}|x_i|\end{cases}\tag{18}$$

其中: $\|\boldsymbol{x}\|_1$ 为向量 $\boldsymbol{x}$ 的 L_1 范数。

步骤 5　使用多重乘子交替方向法 (alternating direction method of multipliers, ADMM) 求解盲源分离问题[22]。ADMM 是一种常用的求解带有等式和不等式约束的凸优化问题的方法, 其目的是根据观测数据分离出多个源信号, 同时满足线性混合模型和非负性约束。ADMM 通过将原始问题转化为包含 2 个变量的问题, 并引入拉格朗日乘子来构建增广拉格朗日函数, 从而解决盲源分离问题。ADMM 将增广拉格朗日函数分解成 2 个子问题, 并交替迭代求解, 直至收敛。在每次迭代中, ADMM 算法需要更新原变量、拉格朗日乘子和新变量的值, 以便逐步逼近最优解。

具体可将 BSS 问题可以表述为

$$\begin{cases}\min\limits_{\boldsymbol{S}}\sum\limits_{i=1}^{b}\psi_i(s_i)+\lambda\sum\limits_{j=1}^{m}\left\|\boldsymbol{a}_j^{\mathrm{T}}s\right\|_2^2\\ \quad\text{s.t.}\quad s_i\geqslant 0,\quad i=1,2,\cdots,b\end{cases}\tag{19}$$

式中: $\boldsymbol{S}=[s_1,s_2,\cdots,s_n]\in\mathbf{R}^{k\times n}$ 为需要求解的源信号矩阵; $\boldsymbol{a}_j\in\mathbf{R}^k$ 为混合系数矩阵; $\psi_i(\cdot)$ 为每个源信号进行的非负性惩罚项; λ 为控制约束的超参数; b 为矩阵 $\boldsymbol{S}$ 中的列数; m 为约束条件中的系数向量的数量; t 为时间序列。

步骤 6　为使用 ADMM 求解盲源分离问题, 需要将原始问题转化为增广拉格朗日函数的形式, 推导如下:

$$L_\rho(\boldsymbol{S},\boldsymbol{U},\boldsymbol{V})=\sum_{i=1}^{n}\psi_i(s_i)+\lambda\sum_{j=1}^{m}\left\|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{s}\right\|_2^2+\frac{\rho}{2}\|\boldsymbol{S}-\boldsymbol{U}+\boldsymbol{V}\|_F^2-\frac{\rho}{2}\|\boldsymbol{U}-\boldsymbol{V}\|_F^2\tag{20}$$

式中: $\boldsymbol{U}\in\mathbf{R}^{k\times n}$ 和 $\boldsymbol{V}\in\mathbf{R}^{k\times n}$ 为拉格朗日乘子矩阵; ρ 为惩罚参数; L 为损失函数; L_ρ 为关于变量 $\boldsymbol{S}$、$\boldsymbol{U}$ 和 $\boldsymbol{V}$ 的目标函数; 参数 λ 是用于调节式(20)等号右侧第 2 项的权重因子; F 为矩阵之间的弗罗贝尼乌斯范数(Frobenius norm)。

为简化 ADMM, 将增广拉格朗日函数分解成 2 个子问题, 分别对 $\boldsymbol{S}$ 和 $(\boldsymbol{U},\boldsymbol{V})$ 求解:

子问题 1: 固定 $(\boldsymbol{U},\boldsymbol{V})$, 求解 $\boldsymbol{S}$, 如下:

$$\min_{S}\sum_{i=1}^{n}\psi_i(s_i)+\lambda\sum_{j=1}^{m}\left\|\boldsymbol{a}_j^{\mathrm{T}}\boldsymbol{s}\right\|_2^2+\frac{\rho}{2}\|\boldsymbol{S}-\boldsymbol{U}+\boldsymbol{V}\|_F^2 \tag{21}$$

使用交替方向乘子 (alternating lagrangian multiplier, ALM) 法或 ADMM 等方法求解上述问题。

子问题 2: 固定 $\boldsymbol{S}$, 求解 ($\boldsymbol{U}$, $\boldsymbol{V}$), 如下:

$$\min_{U,V}\frac{\rho}{2}\|\boldsymbol{S}-\boldsymbol{U}+\boldsymbol{V}\|_F^2-\frac{\rho}{2}\|\boldsymbol{U}-\boldsymbol{V}\|_F^2 \tag{22}$$

由于该子问题的解析比较简单, 可直接通过式(23)来更新 $\boldsymbol{U}$ 和 $\boldsymbol{V}$。

$$U_{d+1}=\frac{S_{d+1}+V_d-Y_d}{2}V_{d+1}=\max(0,U_{d+1}+(Y_d-S_{d+1})) \tag{23}$$

式中: Y_d 为拉格朗日乘子; d为迭代次数。值得注意的是, 在更新 $\boldsymbol{S}$ 和 ($\boldsymbol{U}$,$\boldsymbol{V}$) 的过程中, 需要对一些变量进行投影或修剪操作, 以满足非负性约束。此外, 在实际应用中, 还需要根据具体的问题调整超参数 ρ和λ来获得更好的结果。

通过本节推导, USCA 能够更加准确地解决盲源分离问题, 具有良好的稳定性和实用性。

1.2 流 程

1.2.1 耦合原理

盲源分离和小波阈值法可以在实际应用中结合使用, 以达到更好的去噪效果[23]。在信号或图像处理中, 盲源分离方法可以用于分离混合信号或图像, 得到独立成分后, 可对每个成分进行小波阈值去噪处理, 得到更干净、更清晰的信号或图像。相反地, 也可以使用小波变换增强信号或图像, 在信号或图像增强后, 再使用盲源分离方法, 分离出更加明显和有效的成分。同时, 盲源分离和小波阈值法在一些生物医学领域的应用也非常广泛[24-25]。盲源分离和小波阈值方法的耦合原理在于, 盲源分离可以将混合信号分解成不同的独立成分, 而每个独立成分可能包含有噪声。小波阈值法则可以通过去除每个独立成分中的高频噪声来达到去噪的目的, 改进的 MAWD 方法可先增强原始信号的信噪比, 其次使用盲源分离进行分解。因此, 耦合这 2 种方法可以进一步提高信号的去噪效果及信号质量。

1.2.2 耦合流程

由图 1 可知, BSS 和 WT 耦合的生理信号降噪流程分为以下 6 步。

步骤 1 数据采集。生理信号数据是通过生物电极等生物传感器采集到的, 如脑电图、心电图、肌电图等。

步骤 2 预处理。生理信号在采集和传输过程中会受到各种干扰, 需要进行预处理, 如滤波、去噪等操作, 以提高信号质量。

步骤 3 MAWD 方法信号增强。通过自适应加权矩阵对信号进行变换, 使信号能够更好地适应盲源分离算法。

步骤 4 盲源分离。盲源分离是将混合的信号拆解成多个独立的源信号的过程, 常用的方法包括 ICA、PCA 等。

步骤 5 信号合并。将降噪后的独立源信号合并, 构建多通道混合数据集, 并计算信号间的最大允许相关性(maximum allowable correlation, MAC), 进而选择最优的信号源, 得到重构后数据集。

步骤 6 评估验证。通过对比降噪前后的

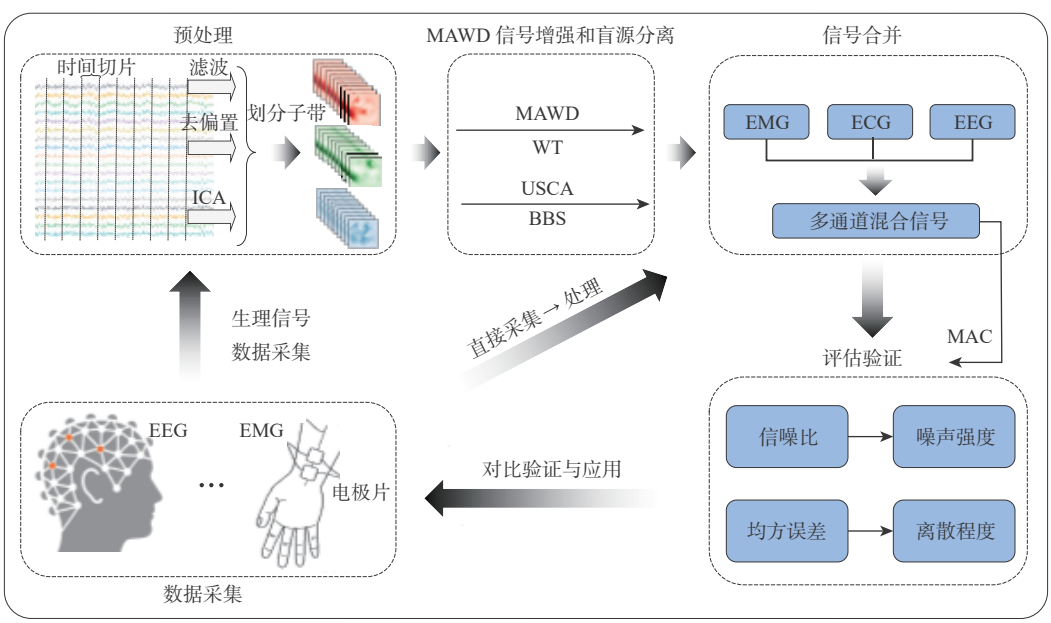


图 1 盲源分离和小波阈值耦合的生理信号降噪流程

Fig. 1 Flowchart for physiological signal denoising using blind source separation and wavelet threshold coupling

信号质量, 如信噪比、均方误差等指标, 评估降噪效果。

以上流程对于不同类型的生理信号数据可能会有所不同, 但是基本思路一致, 即通过预处理、增强、分离等方法降低信号中的噪声干扰, 提高信号质量。

2　本文方法验证

为验证本文方法的有效性, 使用 EMG、ECG、EEG 这 3 种类型生理数据构建混合信号, 使用多谱自适应小波进行数据增强, 使用改进 USCA 进行数据重构。EMG、ECG 和 EEG 是 3 种不同类型的生理信号。其中, EMG 信号是由肌肉收缩引起的电信号, 通常用于检测肌肉活动和运动控制; ECG 信号是由心脏肌肉收缩引起的电信号, 通常用于识别心脏节律和心血管疾病; EEG 信号是由神经元的电活动引起的电信号, 通常用于研究大脑功能和诊断神经系统疾病。混合 EMG、ECG 和 EEG 信号可以模拟人体内多种生理信号同时存在的情况, 以便更好地研究生理信号之间的相互影响和交互作用。混合信号的特点包括频率不同、强度不同和时域特性不同, 其中 EMG 信号占据主导地位, 并产生主要的干扰。混合信号可以通过模拟不同类型的生理信号并将其组合起来, 或者从真实的生理信号中提取多个成分并将它们组合起来产生, 例如在脑机接口研究中, 可以使用 EEG 信号和 EMG 信号构建混合信号以模拟人体运动时的大脑活动情况。

2.1　数据预处理

预处理是数据分析中至关重要的一步, 被广泛应用于 EMG、ECG 和 EEG 这 3 种类型的生理数据。预处理包括信号滤波、信号放大、信号降采样和信号对齐等步骤。这些步骤都是为去除噪声和提取信号, 并使数据适合后续分析。EMG 数据需要进行肌电信号的检测和分割, 以获取有价值的特定动作的 EMG 信号[26]。而 ECG 数据需要进行心拍检测和 R 波检测, 以便确定心率和心律失常情况[27]。EEG 数据则需要使用高通和低通滤波器来去除头皮电噪声和眼电运动噪声, 并使用 ICA 等技术来分离不同脑区的信号, 并注意时间延迟问题, 以确保信号的准确性[28]。因此, 虽然这 3 种类型的生理数据有些相同的预处理步骤, 但其也有各自独特的方法。

2.2　信号增强

使用多谱自适应小波处理预处理后的数据可以有效增强数据。多谱自适应小波处理是一种信号处理方法, 可以将不同频率的信号分离并进行处理[29]。在数据处理中, 利用多谱自适应小波处理可以将原始数据分解为不同频率的子信号, 然后对每个子信号进行独立处理。这样做可以更好地保留信号的特征, 提高数据的可读性和可分析性。同时, 多谱自适应小波处理还可以去除信号中的噪声和干扰, 进一步提高数据质量。

在原始信号增强中, 先涉及将信号分成长度为 N=1 024 个样本的较小帧。对 50% 的重叠帧进行重叠, 这意味着每个连续帧的起始点移动了半个长度(512 个样本)。使用重叠帧的原因是确保在帧的边界处不会丢失任何信息, 这很重要, 因为信号可以在任何时候发生突然变化, 如果帧在这种点结束, 则边界处的信息可能不足以准确表示信号。一旦帧被重叠, 增强过程就可以开始了。增强技术可以因应用而异, 但通常涉及滤波、去噪或其他方法来提高信号的质量或清晰度(见图 2)。

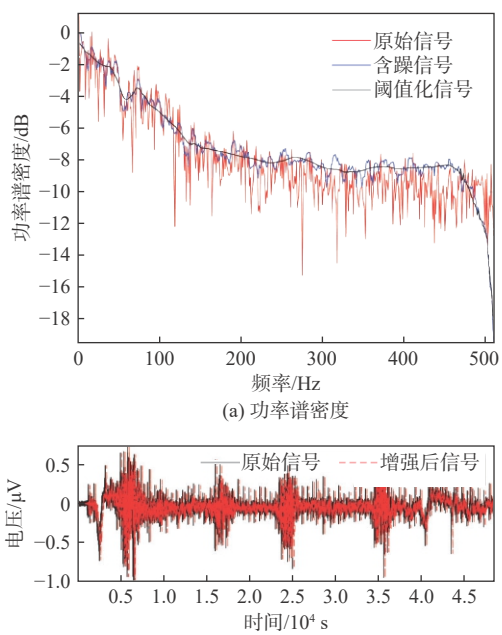


图 2　数据增强对比 (EMG 信号为例)

Fig. 2　Comparison of data augmentation for EMG signals

2.3　信号合并

为验证本文方法的可行性和收敛性, 构建 EMG、ECG 和 EEG 混合生理信号, 以模拟真实生理信号的复杂性和多样性, 从而提高盲源分离算法的鲁棒性和适用性[30-31]。为清晰对比融合的信号, 绘制混合信号的时频图如图 3。

2.4　盲源分离

2.4.1　不同类型信号源

为检验不同类型信号的差异性, 进而选择合适的信号源。使用改进的 USCA 方法进行信号分离

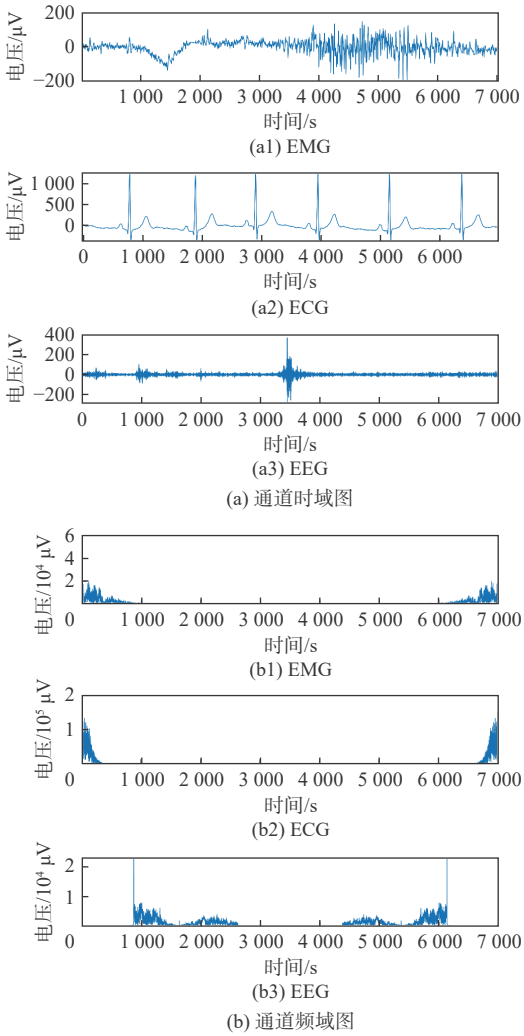


图3 信号合并时频

Fig. 3 Time-frequency of signal mixing

和提取，从而获得独立的EMG、ECG和EEG这3种原始信号。因信号之间的独立性，需调整参数，设置参数窗口长度为4 000/4+1，峰值检测阈值为0.005，以确保ECG模态响应能够被准确地分离出来。由于ECG模态响应振幅值较低，需要更低的阈值才能检测到。为清晰地表征分离信号的特征，绘制分离后信号的时频图如图4所示。

由图4可知，EMG和EEG数据能有效分离，由于EMG信号的微弱性和易受干扰性，EMG信号的振幅通常比ECG和EEG信号要小得多，被认为是一种微弱的电生理信号，因此，产生了多模态数据；而ECG信号是一种具有明显周期性的生物电信号，ECG信号的周期性源于心脏肌肉的收缩和舒张。每个周期就对应于心脏的一个完整的心跳，即心脏每次跳动时产生一个ECG波形，ECG信号的基本周期为一心跳周期，通常为0.5～1 s左右。相比于ECG信号来说，EEG和EMG信号的周期性在时间尺度上更加灵活和复杂。因此，尽管这些信号都具有一定的周期性特征，但其对应的时间尺度、周期长度和信号频率等方面的特征均有所不同。因此，需要更低的阈值才能检测出。

为获得ECG信号源，以分离出ECG信号频率范围的信号成分。需要手动设置参数，设置窗口长度为4 000/4+1，峰值检测阈值为0.000 1。通过参数的调整，可以有效地分离出ECG信号源。然而，由于不同的生理信号特征互不相同，需要根据实际情况调整参数才能完全分离信号源。另外，因阈值较低，可能会导致多余的信号源被分离出来，从而影响信号的分类效果，容易引起误差。基于多次实验，峰值检测阈值设置为[0.000 1, 0.001]区间内，方可分离出更多的源，以提高分离效果及准确性。

2.4.2 同类型信号源

为清晰对比同类信号的分离效果，以EMG信号为例展开研究，使用三通道的EMG数据进行盲源分离（见图5）。设置参数窗口长度为1 000/4+1，峰值检测阈值为0.001，使用盲源分离USCA方法进行分析并绘图时频图。

为评估信号间的相关性或相似性度量，进而选择合适的重构信号，绘制MAC图。MAC图通常是通过计算2个信号之间的互相关函数并将结果归一化来生成的，从而获得信号之间的相关性系数。在MAC图中，x轴和y轴表示2个信号的偏移量，z轴表示其之间的相关性系数。通过查找MAC图中的最大峰值，可以确定2个信号之间的最佳匹配点。具体而言，MAC用于比较2个信号或者是2个时间序列之间的相似程度，以确定其是否可以被成功地匹配或识别。MAC表达式为

$$\mathrm{MAC}(\tau) = \max\left\{|\mathrm{Rxy}(\tau')|\right\} \qquad \tau' \neq \tau \tag{24}$$

式中：Rxy(τ)为2个信号$x(t)$和$y(t)$在时间τ时的互相关函数。

式(24)表示在除去时间偏移τ的情况下，信号$x(t)$和$y(t)$之间的最大相关系数。通过计算不同时间偏移τ下的MAC值，可以绘制出MAC图来评估2个信号之间的相关性或相似性程度，进而确定其是否可以被匹配或识别。

图6可知，(1,1)信号源的相关性最高，达到0.76，(1,2)信号源的MAC高于(1,3)，在选择重构信号时应选择MAC较高的信号源。

2.5 评估验证

为评估信号质量，使用信噪比(signal-to-noise ratio, SNR)、均方根误差(root mean square error, RMSE)指标，通过定量地评价去噪算法的效果。其目的在于评估生理信号去噪的效果，判断去噪算法的优

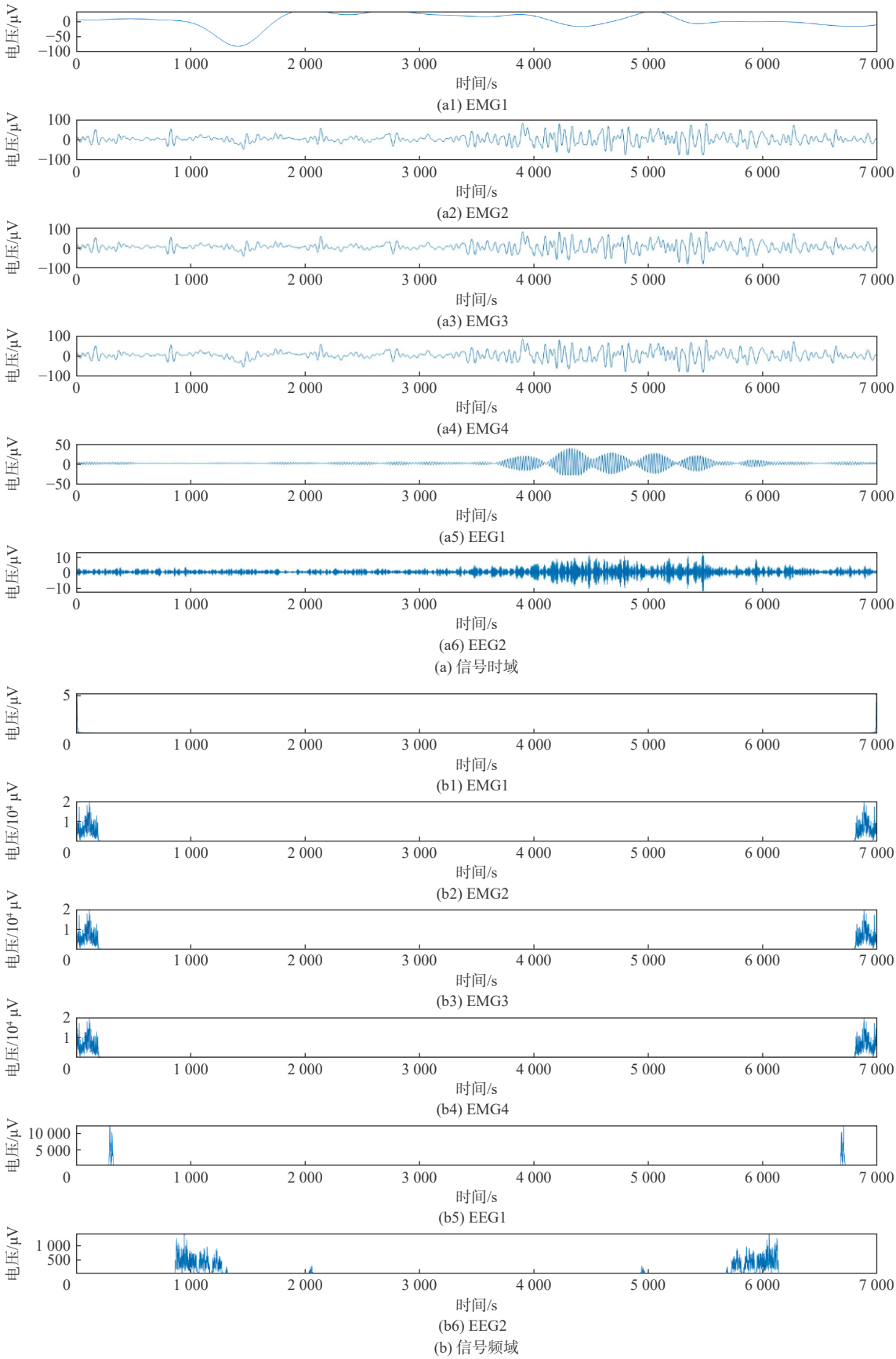


图 4　盲源分离对比

Fig. 4　Comparison of blind source separation

劣。SNR 是指信号的强度与噪声的强度之比, 通常使用分贝(dB)来表示。SNR 表达式为

$$V_{\mathrm{SNR}} = 10\lg(P_{\mathrm{s}}/P_{\mathrm{n}}) \tag{25}$$

式中, P_{s} 表示信号的功率; P_{n} 表示噪声的功率。

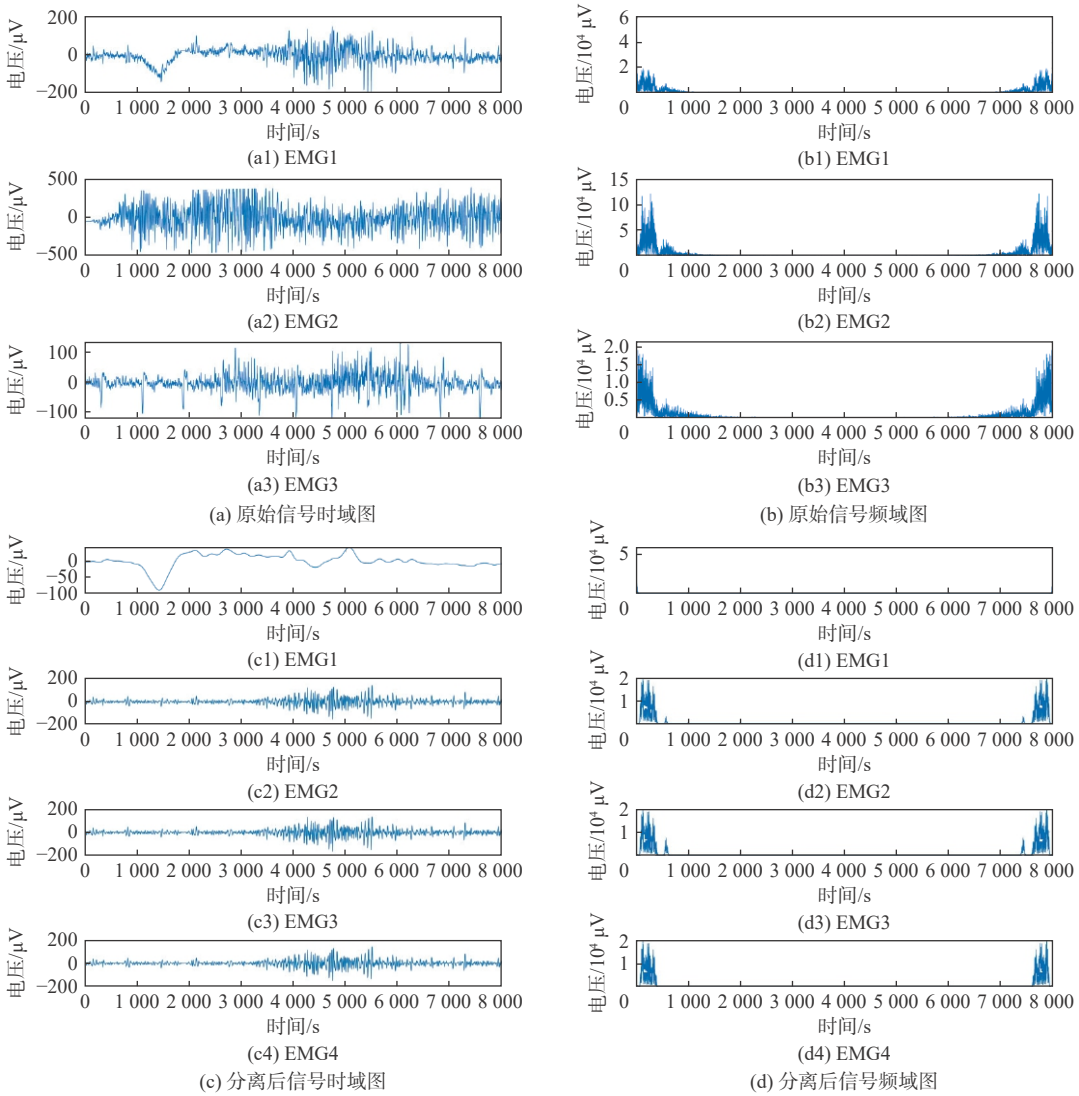


图 5 盲源分离时频

Fig. 5 Time-frequency of blind source separation

RMSE 是用于评估预测模型或估计量的误差大小的指标，通常用于比较预测值和实际观测值之间的差异。RMSE 表达式为

$$e_{\mathrm{RMSE}} = \mathrm{sqrt}\left(1/b \cdot \sum (y_i - x_i)^2\right) \tag{26}$$

式中：b 表示样本数量；y_i 表示实际观测值；x_i 表示预测值。

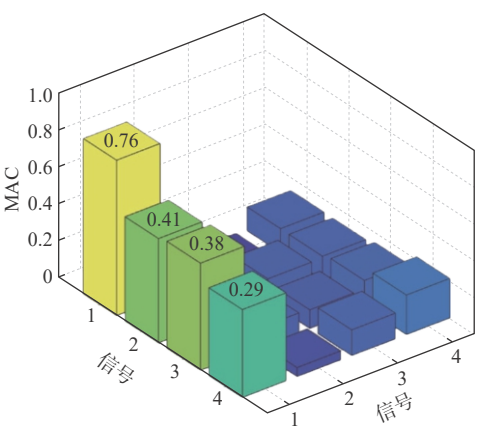


图 6 盲源分离 MAC 图

Fig. 6 MAC diagram of blind source separation

为检验信号的质量，对原始信号、加噪信号、增强信号进行软阈值、硬阈值、自适应阈值对比分析，详见表 1。结果显示，增强后的信号，软阈值相比硬阈值，SNR 提升了约 44.2%，RMSE 下降了约 28.8%，处理时间减少了约 1.4%。软阈值相比自适应阈值，SNR 提升了约 706%，RMSE 下降了约 16.7%，处理时间减少了约 3.0%。为更好地评估本文方法的优越性，进行显著性检验来确定这些差异是否统计上显著。表 1 可知其方差齐性概率为 0.016，小于 0.05，数据满足显著性检测。

结果显示，软阈值去噪算法的 SNR 和 RMSE 指标表现最好，处理效率也较高，而自适应阈值算法的性能较弱。结果显示，软阈值去噪算法处理后的增强信号具有较好的 SNR 和低的 RMSE，而且处理时间也比硬阈值和自适应阈值算法都要少。这

表 1　原始信号、加噪信号、增强信号参数对比

Table 1　Comparison of parameters of original, noisy, and enhanced signals

指标	SNR提升/μV			RMSE下降/μV			处理时间/s		
	原始信号	加噪信号	增强信号	原始信号	加噪信号	增强信号	原始信号	加噪信号	增强信号
软阈值	8.17	8.16	8.15	0.05	0.05	0.05	3.81	3.72	3.67
硬阈值	5.65	5.63	5.61	0.066	0.07	0.07	3.83	3.74	3.65
自适应阈值	1.08	1.06	1.01	0.12	0.12	0.12	3.82	3.73	3.69

个结果对于需要进行信号去噪的应用来说是非常重要的, 因为一个好的去噪算法可以使得处理后的信号更加清晰准确, 在后续的数据分析和决策过程中能够提供更可靠的依据。

3　结果与讨论

3.1　实验结果

为清晰对比软阈值的参数差异, 对软阈值去噪算法处理后的 3 种信号进行归一化处理, 设置区间为 [−1, 1], 使用下式进行处理:

$$x_{\text{scaled}} = 2(x - \min(x))/(\max(x) - \min(x)) - 1 \quad (27)$$

式中: x 为原始数据; x_{scaled} 为归一化后的数据。$\min(x)$ 和 $\max(x)$ 分别为原始数据中的最小值和最大值。

由图 7 可知, 增强后的信号具有较好的信噪比、较低的均方根误差和较短的处理时间。这表明所提出的去噪算法的有效性, 并且可以提高信号质量。软阈值下增强后的信号与原始信号相比, SNR 提升了约 0.12%, RMSE 下降了约 2.5%, 处理时间减少了约 3.9%。

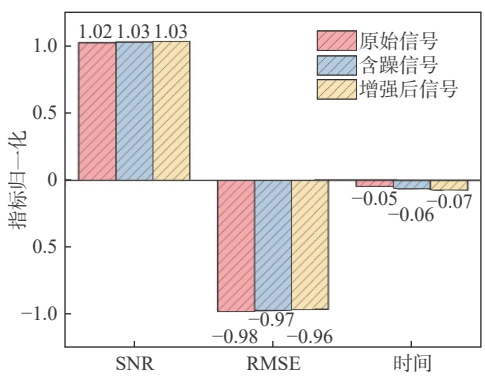


图 7　归一化参数对比

Fig. 7　Comparison of normalized parameters

3.2　讨　论

3.2.1　本文方法的有效性

通过多谱自适应小波和改进的盲源分离 USCA 的融合, 可以有效克服各自算法的缺陷, 提高信号的分离效果和准确性。多谱自适应小波能够增强信号, 在不同频率范围内提取有价值的信息, 为比较增强信号与原始信号之间的指标差异, 计算其 SNR 和 RMSE 值, 在 SNR 方面下降了 0.24%, 在处理时间上提升了 3.67%。盲源分离 USCA 则可从混合的观测信号中提取出干净的信号, 详见表 2。

表 2　盲源分离前后参数对比

Table 2　Comparison of parameters before and after blind source separation

指标	SNR提升/μV		RMSE下降/μV		处理时间/s	
	分离前	分离后	分离前	分离后	分离前	分离后
软阈值	7.90	7.96	10.76	7.65	4.18	4.04
硬阈值	5.43	5.51	14.30	10.15	4.25	4.12
自适应阈值	0.15	0.17	26.26	18.74	4.16	1.14

在软阈值下, 分离后信号的 SNR 提升 0.8%。RMSE 降低了 28.9%。时间消耗略有减少, 但变化不明显。在硬阈值下, 分离后信号的 SNR 提升了 1.5%。RMSE 降低 40.9%。时间消耗略有增加, 但变化不明显。在自适应阈值下, 分离后信号的 SNR 提升了 11.8%。RMSE 降低了 42.8%。时间消耗大幅减少, 降低了约 72.6%。综上所述, 在软阈值下和硬阈值下, 分离后信号的 SNR 性能提升分别为 0.8% 和 1.5%。而在自适应阈值下 SNR 提升了 11.8%, 但是 RMSE 的降低幅度较大, 为 42.8%。此外, 自适应阈值方法的时间消耗也是 3 种方法中最少的, 降低了约 72.6%。2 种算法联合使用可以综合利用它们的优点, 使分离后的信号更加清晰、准确, 便于后续处理和分析, 特别是在信号分离方面, 其效果显著。

3.2.2　局限性

盲源分离和小波阈值方法是一种常用的降噪技术, 但在实际应用中需要注意其局限性。首先, 该方法对于复杂信号场景的处理有限, 可能无法准确地识别和分离出各个源信号, 导致降噪效果不佳; 其次, 当信号中存在较高的噪声时, 小波阈值降噪方法可能无法有效去除噪声, 另外, 算法的鲁棒性和稳定性仍需进一步提高, 同时需要考虑参数的选择和算法实现方式等因素对降噪效果的影响, 此外, 盲源分离算法本身对数据量的要求较高, 需要较大的样本容量才能获得准确结果, 并且, 源信号独立分布的假设也是一个必须被满足的条件; 最

后，生理信号数据的特殊性也需要考虑，包括采集设备、测量环境等因素可能会对信号采集造成不同程度的影响，从而影响降噪效果和稳定性。因此，在实际应用中，需要根据具体场景和数据特点综合考虑以上局限性，选择适合的降噪方法和算法，并不断深入研究和探索，提高降噪算法的鲁棒性和稳定性，解决更复杂的生理信号降噪问题。

4 结 论

1）为降低生理信号中的噪声干扰，提高数据的质量和准确性。通过对盲源分离和小波阈值方法进行耦合研究，提出了 MAWD 方法和 USCA，二者融合对生理信号进行降噪处理。

2）解决了多通道信号中的相互干扰问题，并结合小波阈值增强信号以减少噪声干扰。结果表明，本文方法能够有效地去除生理信号中的噪声干扰，提高信号的信噪比和均方根误差，在自适应阈值下，分离后信号的 SNR 提升了 11.8%。RMSE 降低了 42.8%。

USCA 存在复杂度较大、融合效果依赖于信号类型、易分离出多余的信号源，以及对于强噪声干扰的信号仍存在分离难题等局限性。因此，未来需要进一步优化算法、探索更加有效的信号处理方法，以及适应更为广泛的生理信号类型，以提高降噪效果。为生理学领域的数据采集和分析提供了一种新的方法和思路。

参考文献（References）

[1] SHU L, XIE J Y, YANG M Y, et al. A review of emotion recognition using physiological signals[J]. Sensors, 2018, 18(7): 2074.

[2] LAL B, GRAVINA R, SPAGNOLO F, et al. Compressed sensing approach for physiological signals: a review[J]. IEEE Sensors Journal, 2023, 23(6): 5513-5534.

[3] ZHU J J, WEI Y X, FENG Y F, et al. Physiological signals-based emotion recognition via high-order correlation learning[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2019, 15(3s): 1-18.

[4] BELOUCHRANI A, ABED-MERAIM K, CARDOSO J F, et al. A blind source separation technique using second-order statistics[J]. IEEE Transactions on Signal Processing, 1997, 45(2): 434-444.

[5] FU Z Z, ZHANG B N, HE X R, et al. Emotion recognition based on multi-modal physiological signals and transfer learning[J]. Frontiers in Neuroscience, 2022, 16: 1000716.

[6] GOLDHACKER M, KECK P, IGEL A, et al. A multi-variate blind source separation algorithm[J]. Computer Methods and Programs in Biomedicine, 2017, 151: 91-99.

[7] CHEN X, WANG Z J, MCKEOWN M. Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods[J]. IEEE Signal Processing Magazine, 2016, 33(3): 86-107.

[8] XIONG F, CHEN D Y. CEEMDAN-IMFx-PCA-CICA: an improved single-channel blind source separation in multimedia environment for motion artifact reduction in ambulatory ECG[J]. Complex & Intelligent Systems, 2023, 9(3): 2555-2569.

[9] VON LÜHMANN A, BOUKOUVALAS Z, MÜLLER K R, et al. A new blind source separation framework for signal analysis and artifact rejection in functional Near-Infrared Spectroscopy[J]. NeuroImage, 2019, 200: 72-88.

[10] ROY V, SHUKLA S. Designing efficient blind source separation methods for EEG motion artifact removal based on statistical evaluation[J]. Wireless Personal Communications, 2019, 108(3): 1311-1327.

[11] SAUER J, STREPPEL M, PETERSEN E, et al. Blind source separation of inspiration and expiration in respiratory sEMG signals[J]. Physiological Measurement, 2022, 43(7): 075007.

[12] ROUSSEL J, RAVIER P, HARITOPOULOS M, et al. Decomposition of multi-channel intramuscular EMG signals by cyclostationary-based blind source separation[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2017, 25(11): 2035-2045.

[13] NEGRO F, MUCELI S, CASTRONOVO A M, et al. Multi-channel intramuscular and surface EMG decomposition by convolutive blind source separation[J]. Journal of Neural Engineering, 2016, 13(2): 026027.

[14] WEDEKIND D, KLEYKO D, OSIPOV E, et al. Robust methods for automated selection of cardiac signals after blind source separation[J]. IEEE Transactions on Biomedical Engineering, 2018, 65(10): 2248-2258.

[15] WAN P W, HAO B J, LI Z, et al. Time differences of arrival estimation of mixed interference signals using blind source separation based on wireless sensor networks[J]. IET Signal Processing, 2016, 10(8): 924-929.

[16] DONOHO D L, JOHNSTONE I M. Adapting to unknown smoothness via wavelet shrinkage[J]. Journal of the American Statistical Association, 1995, 90(432): 1200-1224.

[17] DAUBECHIES I. Orthonormal bases of compactly supported wavelets[J]. Communications on Pure and Applied Mathematics, 1988, 41(7): 909-996.

[18] ZHAO H M, ZUO S Y, HOU M, et al. A novel adaptive signal processing method based on enhanced empirical wavelet transform technology[J]. Sensors, 2018, 18(10): 3323.

[19] JUTTEN C, HERAULT J. Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture[J]. Signal Processing, 1991, 24(1): 1-10.

[20] XIE Y, ZOU T, YANG J, et al. SA-UCBSS: sparsity-based adaptive underdetermined convolutive blind source separation[J]. Knowledge-Based Systems, 2024, 212: 112224.

[21] LI Q, YANG J H, CHEN X, et al. The decomposition of surface EMG signals based on blind source separation of convolved mixtures[J]. Conference Proceedings, 2005, 2005: 5912-5915.

[22] HU J, GUO T, HAN C. Stochastic linearized generalized alternating direction method of multipliers: expected convergence rates and large deviation properties[J]. Mathematical Structures in Computer Science, 2023, 66(3): 1-18.

[23] YAN X T, BOUDRIAS M H, MITSIS G D. Removal of transcrani-

al alternating current stimulation EEG artifacts using blind source separation and wavelets[J]. IEEE Transactions on Biomedical Engineering, 2022, 69(10): 3183-3192.

[24] ORTIZ-ECHEVERRI C J, SALAZAR-COLORES S, RODRÍGUEZ-RESÉNDIZ J, et al. A new approach for motor imagery classification based on sorted blind source separation, continuous wavelet transform, and convolutional neural network[J]. Sensors, 2019, 19(20): 4541.

[25] LI X, ZHENG Y, ZHANG J, et al. Finite SNR diversity-multiplexing trade-off in hybrid ABCom/RCom-assisted NOMA networks[J]. IEEE Transactions on Mobile Computing, 2024, 23(10): 9108-9119.

[26] HASSAN N, RAMLI D A. Sparse component analysis (SCA) based on adaptive time: frequency thresholding for underdetermined blind source separation (UBSS)[J]. Sensors, 2023, 23(4): 2060.

[27] CASTAÑO USUGA F A, GISSEL C, HERNÁNDEZ A M. Motion artifact reduction in electrocardiogram signals through a redundant denoising independent component analysis method for wearable health care monitoring systems: algorithm development and validation[J]. JMIR Medical Informatics, 2022, 10(11): e40826.

[28] METSOMAA J, SARVAS J, ILMONIEMI R J. Blind source separation of event-related EEG/MEG[J]. IEEE Transactions on Biomedical Engineering, 2017, 64(9): 2054-2064.

[29] CHEN Y, CHENG Y N, LIU H L. Application of improved wavelet adaptive threshold de-noising algorithm in FBG demodulation[J]. Optik, 2017, 132: 243-248.

[30] MA S, ZHANG H J, MIAO Z Y. Blind source separation for the analysis sparse model[J]. Neural Computing and Applications, 2021, 33(14): 8543-8553.

[31] XU X W, LIANG Y, HE P, et al. Adaptive motion artifact reduction based on empirical wavelet transform and wavelet thresholding for the non-contact ECG monitoring systems[J]. Sensors, 2019, 19(13): 2916.

Physiological signal denoising method based on multi-spectrum adaptive wavelet and blind source separation

WANG Zhenyu[1, 2], XIANG Zerui[1, 3, *], ZHI Jinyi[1, 3], DING Tiecheng[1], ZOU Rui[1]

(1. School of Design, Southwest Jiaotong University, Chengdu 610031, China;
2. College of Design and Innovation, Tongji University, Shanghai 200092;
3. Institute of Design and Research for Man-Machine-Environment Engineering System, Southwest Jiaotong University, Chengdu, 611730, China)

Abstract: In order to improve the quality and reliability of physiological signals, blind source separation and wavelet threshold methods were coupled to propose a multi-spectrum adaptive wavelet signal enhancement method, which was combined with an improved blind source separation method for denoising. To evaluate the effectiveness of the proposed method, the signal-to-noise ratio (SNR) and root mean square error (RMSE) indicators were calculated by using three wavelet transform methods: soft threshold, hard threshold, and adaptive threshold. The results show that the proposed method has strong applicability under the soft threshold, and compared with that under a hard threshold, the enhanced signal under a soft threshold has an SNR improvement of about 44.2%, RMSE reduction of about 28.8%, and a time reduction of about 1.4%. Compared with the adaptive threshold, SNR is improved by about 706%; RMSE is reduced by about 16.7%, and time is reduced by about 3.0%. The original, noisy, and enhanced signals are analyzed and normalized by using a soft threshold to compare the differences in various parameters under a soft threshold. The results show that the enhanced signal has a better SNR, lower RMSE, and shorter processing time. Compared with the original signal under the soft threshold, SNR is improved by about 0.12%; RMSE is reduced by about 2.5%, and time is reduced by about 3.9%, which further verifies the effectiveness of the proposed algorithm and improves the signal quality.

Keywords: multi-spectrum adaptive wavelet; blind source separation; wavelet transform; denoising method; physiological signal

Received: 2023-04-14; **Accepted**: 2023-06-16; **Published Online**: 2023-07-19 11:01

URL: link.cnki.net/urlid/11.2625.V.20230718.1851.003

Foundation items: Second Batch of Collaborative Education Projects by the Ministry of Education in 2022 (220705329291641); New Interdisciplinary Cultivation Fund of Southwest Jiaotong University (YG2022006)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0084

新型复合式构型直升机气动干扰特性分析

梁家辉，张夏阳，招启军*

(南京航空航天大学 直升机动力学全国重点实验室，南京 210016)

摘　　要：综合共轴双旋翼直升机高悬停效率和操纵效率的特点及双推力螺旋桨复合式直升机高速前飞性能的优势，将共轴双旋翼应用到双推力螺旋桨复合式直升机上。以X3复合式直升机为基准，将单旋翼改为共轴双旋翼。为研究共轴双旋翼对复合式直升机气动特性的影响，建立全机快速配平方法，并在此基础上针对不同构型直升机开展气动特性分析。结果表明：相较于单旋翼构型，共轴双旋翼构型直升机空气动力学特性具有良好的对称性；在保持较好的高速前飞性能的情况下，共轴双旋翼复合构型可以显著提升悬停和低速飞行性能，悬停效率提升6.8%，100 km/h前飞速度下总需用功率降低23.1%；低速状态下共轴双旋翼对螺旋桨的气动干扰程度明显低于单旋翼构型。

关 键 词：共轴双旋翼；双推力螺旋桨；复合式直升机；旋翼动量源；气动干扰

中图分类号：V211.52; TB553

文献标志码：A　　**文章编号**：1001-5965(2025)03-0922-11

常规单旋翼带尾桨式直升机受自身构型影响，在达到一定飞行速度后，旋翼前行桨叶和后行桨叶分别会产生激波和气流分离，使得常规构型直升机的最大平飞速度难以突破350 km/h[1]。复合式直升机的出现提高了直升机的飞行速度，改善了大速度飞行时的空气动力性能。目前常见的一类复合构型是在机身两侧加装机翼和辅助推进装置[2]，该构型的特点是旋翼提供升力，向前飞的推力及平衡旋翼的反扭矩由两侧螺旋桨提供，机翼在飞行中产生升力逐渐使旋翼卸载，从而使后行桨叶迎角减小，推迟了后行桨叶的失速，最终提高前飞速度。此外，当飞行速度达到某一临界值时，旋翼可以通过降低转速来减小阻力，有利于提升直升机的最大前飞速度。这一构型的复合式直升机具有良好的高速前飞性能，但是悬停和低速性能却不具有优势。一是旋翼桨盘载荷比较高，如著名的欧洲直升机公司的X3，其桨盘载荷高达64 kg/m^2[3]，使得悬停效率较低；二是为平衡旋翼反扭矩，两侧螺旋桨的推力大小甚至方向不一样[4]，使得机身左右两侧的气动环境非对称，对直升机的操纵不利。共轴双旋翼直升机以美国西科斯基公司的X-2高速直升机[5]为代表，两副共轴安放且反向旋转的旋翼能使反扭矩相互平衡。该构型直升机的特点是悬停效率较高，且由于上下旋翼反转，空气动力保持对称，具有较高的操纵效率，但配平复杂。由AVX飞机公司和L3技术公司针对美军“未来攻击侦察飞机”竞赛项目所联合设计的AVX-L3共轴复合式直升机(coaxial compound helicopter, CCH)[6]创新地采用了复合共轴布局，是一款带有共轴双旋翼和两副涵道风扇推进螺旋桨的复合直升机，而AVX公司指出共轴双旋翼能使CCH在直升机模式飞行时具有较少动力需求，悬停能耗将降低15%～20%，巡航能耗降低10%。此外，朱清华等[7]发明了一种涵道螺旋桨推力共轴双旋翼直升机，该发明将涵道螺旋桨作为推力机构与共轴双旋翼结合，可以提高共轴双旋翼直升机的前飞速度，实现高速前飞。基于此，本

收稿日期：2023-02-28；**录用日期**：2023-07-04；**网络出版时间**：2023-07-11 11:04
网络出版地址：link.cnki.net/urlid/11.2625.V.20230711.0914.001
基金项目：国家自然科学基金(12032012,12102186)；国家重点实验室基金(61422202201)；江苏省自然科学基金(BK20200433)
* **通信作者**. E-mail：zhaoqijun@nuaa.edu.cn

引用格式：梁家辉，张夏阳，招启军. 新型复合式构型直升机气动干扰特性分析[J]. 北京航空航天大学学报，2025，51(3)：922-932.
LIANG J H，ZHANG X Y，ZHAO Q J. Aerodynamic interference characteristics of a new compound configuration helicopter[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：922-932 (in Chinese).

文在双推力螺旋桨复合式直升机的基础上将单旋翼改为共轴双旋翼, 采用 CFD 方法对全机流场进行模拟, 并同改动前的直升机展开气动特性和功率特性的对比分析。

国内外关于共轴双旋翼直升机气动特性的研究已经取得不俗进展。2008 年, 西科斯基公司的 Bagai[8] 根据广义旋翼性能 (generalized rotor performance, GRP) 程序分别应用均匀入流和基于固定尾迹的非均匀入流模型对 XH-59A 和 X2 旋翼性能进行了对比分析, 并表明这种方法并不适用于实际的高速直升机共轴式旋翼干扰分析。2009 年, Kenyon 等[9] 采用涡系传输模型对小速度下共轴旋翼构型直升机的旋翼/尾推桨干扰进行了计算, 指出共轴刚性旋翼与螺旋桨之间存在着严重的气动干扰。2011 年, Xu 和 Ye[10] 通过求解三维非定常欧拉方程, 模拟了共轴双旋翼直升机前飞的非定常流动, 结果表明上下 2 个旋翼由于气动相互作用性能都变差, 且上旋翼对下旋翼的影响大于下旋翼对上旋翼的影响, 下旋翼的下洗气流速度远大于上旋翼的下洗气流速度。2012 年, 李文浩[11] 结合动量源模型和 Euler 方程, 建立了一个适用于悬停及高速前飞状态的复合式高速直升机共轴双旋翼/机身干扰流场数值模拟方法, 得出在相同桨叶外形和旋翼实度条件下, 共轴式双旋翼的性能优于等效的单旋翼, 悬停状态时下旋翼性能比上旋翼性能差等结论。2014 年, 张银[12] 建立了共轴刚性旋翼直升机的平衡分析模型, 对不同飞行状态下的直升机干扰流场进行了计算, 并分析不同参数对旋翼/推进螺旋桨/机身干扰流场的影响, 得到小速度前飞时旋翼对尾部推进桨干扰明显, 而随前飞速度增加干扰减弱; 推进螺旋桨对旋翼流场干扰特性影响较小等结论。2018 年, 吴裕平等[13] 以类 S97 构型的高速直升机为研究对象, 建立全机平衡方程并进行了共轴双旋翼、尾推桨的功率占比分析, 表明全机功率分配比例在低速、高速飞行时差距很大, 悬停和低速以共轴刚性旋翼为主, 高速时以尾部推进桨为主。

综上所述, 共轴双旋翼直升机上下旋翼之间及旋翼和机身之间气动干扰的研究已成熟, 但共轴双旋翼对复合式直升机机翼和螺旋桨气动干扰的研究尚未开展。此外, 复合式直升机气动布局复杂且各部件之间的气动干扰问题突出, 传统的气动力模型难以精确计算各部件之间的气动干扰特性, 使得配平和气动特性分析精度降低。鉴于此, 基于简化的配平模型, 采用结合动量源的 CFD 方法对新构型直升机和单旋翼复合式直升机进行配平, 在此基础上展开气动特性分析并对比二者气动性能的差异及气动干扰机理。为方便描述, 后文以单旋翼构型代指双推力螺旋桨复合式直升机, 以共轴双旋翼构型代指本文提到的共轴双旋翼加双推力螺旋桨新型复合式直升机。

1　计算方法与验证

1.1　配平模型和配平方法

直升机稳定飞行时, 全机的力和力矩处于配平状态。单旋翼构型前飞状态下的动力主要由两侧螺旋桨提供, 考虑到机身姿态角度较小[14-15], 建立简化的配平模型[13], 假定滚转力矩和俯仰力矩的配平通过副翼调整和平尾舵偏自动满足。简化配平方程中各部分气动力如式 (1) 和图 1 所示。

$$\begin{cases} \sum F_x = 0 : T_{\mathrm{LP}} + T_{\mathrm{RP}} - D_{\mathrm{F}} - D_{\mathrm{W}} = 0 \\ \sum F_z = 0 : T_{\mathrm{R}} + L_{\mathrm{W}} - W = 0 \\ \sum M_z = 0 : (T_{\mathrm{RP}} - T_{\mathrm{LP}})L - M_{\mathrm{R}} = 0 \end{cases} \tag{1}$$

式中: T_{LP}和T_{RP}分别为机身左、右侧螺旋桨的推力; T_{R}和M_{R}分别为旋翼拉力和反扭矩; L_{W}和D_{W}为机翼升阻力; W为机身重力; D_{F}为机身阻力; L为螺旋桨旋转中心到机身纵向对称面的距离。

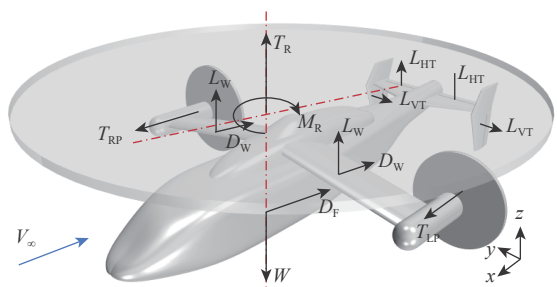


图 1　复合式直升机受力分析

Fig. 1　Force analysis of compound helicopter

对共轴双旋翼构型的配平方程进行修改:

$$\begin{cases} \sum F_x = 0 : T_{\mathrm{LP}} + T_{\mathrm{RP}} - D_{\mathrm{F}} - D_{\mathrm{W}} = 0 \\ \sum F_z = 0 : T_{\mathrm{R1}} + T_{\mathrm{R2}} + L_{\mathrm{W}} - W = 0 \\ \sum M_z = 0 : M_{\mathrm{R1}} - M_{\mathrm{R2}} = 0 \end{cases} \tag{2}$$

式中: T_{R1}和T_{R2}分别为上旋翼和下旋翼的拉力, M_{R1}和M_{R2}为对应的反扭矩。此外, 共轴双旋翼构型机身两侧的螺旋桨推力相同, 即$T_{\mathrm{RP}} = T_{\mathrm{LP}}$。

采用基于 CFD 方法的配平方法进行全机配平计算。首先, 对各部件进行气动力建模并将其代入式 (1)(共轴双旋翼构型为式 (2)) 进行配平求解并将结果作为初始配平量; 然后, 输入初始配平结果进行全机 CFD 分析, 由于 CFD 配平过程中需要不断对旋翼和螺旋桨的操纵进行更新, 针对旋翼和螺旋桨采用高效的动量源方法模拟气动特性, 配平流程如图 2 所示。

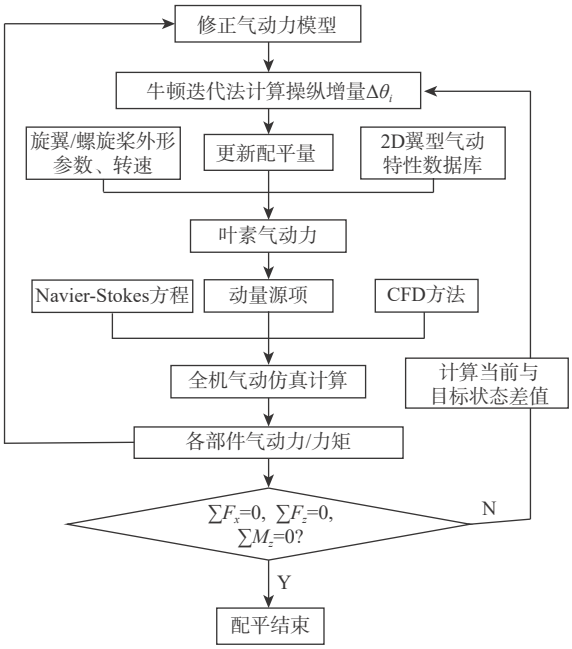


图 2 基于动量源模型的配平流程

Fig. 2 Trim process based on momentum source model

共轴双旋翼构型的配平难点在于上下旋翼的反扭矩平衡，采用总距差动的配平策略，迭代过程中的气动力通过 CFD 方法求解。设总距平均值为$\bar{\varphi}$，总距差动$\Delta\varphi$，上下旋翼总距输入分别为$\varphi_u=\bar{\varphi}-\Delta\varphi$，$\varphi_l=\bar{\varphi}+\Delta\varphi$。配平具体步骤如下：

步骤 1 给定配平初值：$\bar{\varphi}=\varphi_0$，$\Delta\varphi=0$。

步骤 2 根据配平目标$\left|T^t-(T_u+T_l)\right|\leqslant\varepsilon_1$对$\bar{\varphi}$进行迭代求解，并输出配平后的$\bar{\varphi}$和$\Delta\varphi$。不等式中$T^t$为目标旋翼拉力，$T_u$和$T_l$分别为上下旋翼拉力，$\varepsilon_1$为拉力配平容差。

步骤 3 判断在步骤 2 所输出的总距操纵下是否满足$|M_u-M_l|\leqslant\varepsilon_2$，若满足，则配平结束；若不满足，则进行步骤 4。不等式中M_u和M_l分别为上下旋翼反扭矩，ε_2为反扭矩配平容差。

步骤 4 根据配平目标$|M_u-M_l|\leqslant\varepsilon_2$对$\Delta\varphi$进行迭代求解，并输出配平后的$\bar{\varphi}$和$\Delta\varphi$。

步骤 5 判断在步骤 4 所输出的总距操纵下是否满足$\left|T^t-(T_u+T_l)\right|\leqslant\varepsilon_1$，若满足，则配平结束；若不满足，则返回到步骤 2，直到满足所有配平目标为止。

1.2 数值计算模型

以 X3 复合式直升机为单旋翼构型的原准机，进一步将单旋翼改为共轴双旋翼以作为共轴双旋翼构型的研究对象。为避免旋翼尺寸对研究结果造成影响，保持共轴双旋翼构型的旋翼直径、实度、转速、桨叶外形与单旋翼构型一致。2 种构型直升机的旋翼基本参数如表 1 和表 2 所示。

表 1 单旋翼构型原准机[3]

Table 1 Prototype of single-rotor configuration [3]

半径R/m	转速/(rad·s^{-1})	桨叶片数	桨叶根切/m	弦长/m	翼型
6.3	36.55	5	1.512	0.385	NACA6412

表 2 共轴双旋翼构型

Table 2 Coaxial-rotor configuration

旋翼参数	半径R/m	转速/(rad·s^{-1})	桨叶片数	上下旋翼间距H/R	旋翼参数	桨叶根切/m	弦长/m	翼型
数值	6.3	36.55	3+3	0.16	数值	1.512	0.321	NACA6412

单旋翼构型的几何模型如图 1 所示，共轴双旋翼构型的几何模型如图 3 所示。机身和计算域网格采用非结构网格，网格类型选择多面体网格，该网格对复杂外形具有良好的几何适应性，并且由于具有多个相邻单元，对流场中的速度和压强梯度能进行更好的计算。图 4 为机身纵向对称面及螺旋桨所在平面的网格视图：对旋翼和螺旋桨所在区域进行加密以有效添加动量源项，同时在机身后方设置尾流加密区以提高前飞时流场的计算精度。机身设置壁面无滑移边界条件，计算域的进口根据来流速度设置为速度入口边界条件，出口设置为压力出口边界条件。采用包含动量源项的 RANS 方程[16]作为流场求解的控制方程并选择 S-A 湍流模型。

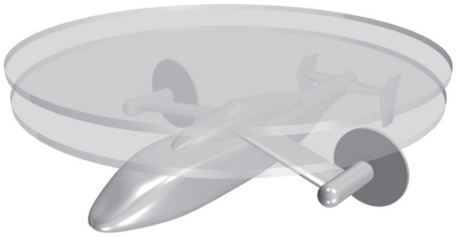


图 3 共轴双旋翼构型几何模型

Fig. 3 Geometric model of coaxial-rotor configuration

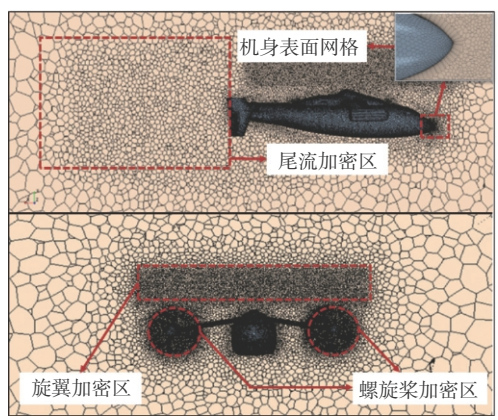


图 4 机身纵向对称面和螺旋桨平面网格

Fig. 4 Grids of longitudinal symmetry plane of fuselage and propeller plane

1.3　算例验证

1.3.1　动量源方法验证

本文将旋翼和螺旋桨处理为等效盘, 用动量源方法模拟真实旋翼/螺旋桨在流场中的作用[17-20]。选用文献 [21] 中的经典旋翼模型验证动量源方法的有效性, 计算值与试验值的对比结果如图 5 所示。

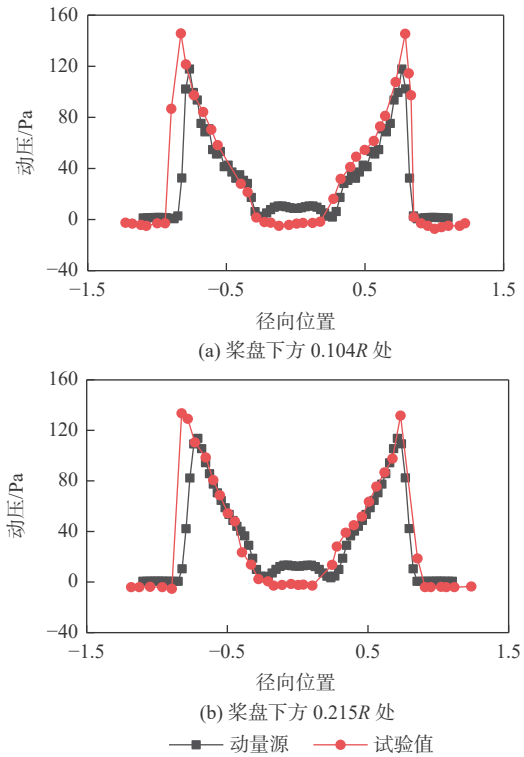


图 5　悬停旋翼的计算结果对比

Fig. 5　Comparison of calculation results of hovering rotor

通过对比旋翼桨盘下方的动压分布可知, 计算值与试验值吻合较好, 表明本文所采用的动量源方法能够用于旋翼悬停流场的数值模拟。

进一步选用 NASA 的 2MRTS(2-meter rotor test system)[22] 前飞旋翼试验模型验证动量源方法对旋翼前飞流场模拟的有效性。图 6 为前进比为 0.15 时旋翼桨叶不同剖面在各个方位角下的诱导速度对比。无量纲诱导速度定义为旋翼桨盘诱导速度与桨尖速度的比值。由图 6 可知旋翼桨盘诱导速度的计算值与试验值吻合良好, 表明动量源方法能够有效模拟旋翼前飞流场。

1.3.2　机身流场计算验证

选择 ROBIN 机身作为验证算例, 采用本文建立的数值方法模拟机身在高速前飞时的流场特性。该算例的具体参数和试验数据取自文献 [23]。图 7 为机身凸台段和尾梁段上横截面的表面压强系数与试验值的对比。可以看出机身表面压强系数计算值与试验值大体上吻合较好, 说明本文所采用的 CFD 方法能有效地对机身流场进行数值模拟。

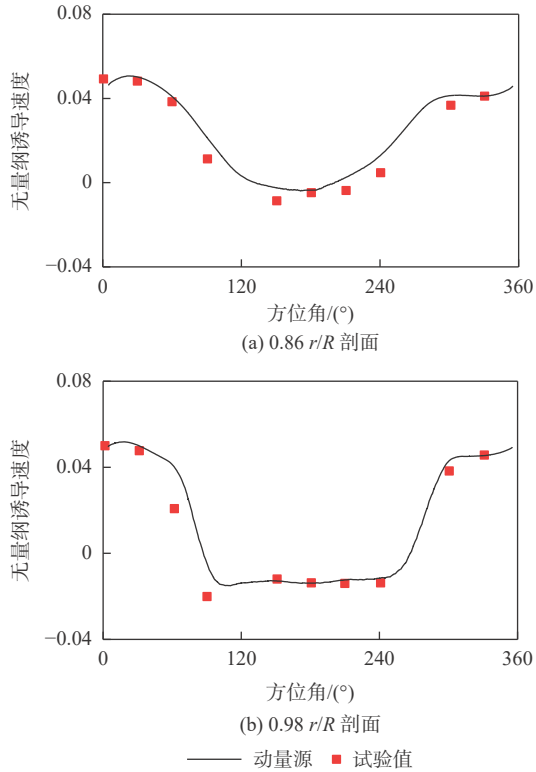


图 6　前飞旋翼的计算结果对比

Fig. 6　Comparison of calculation results of forward flight rotor

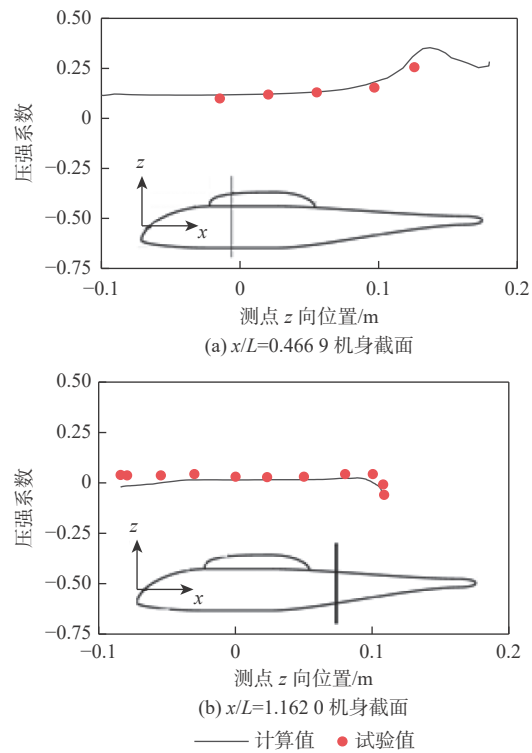


图 7　机身流场计算结果与试验值对比

Fig. 7　Comparison of calculation results of fuselage flow field with test results

2　结果分析

针对悬停和前飞 2 个典型状态, 对不同前飞速

度下的单旋翼构型和共轴双旋翼构型展开流场计算，并系统开展气动特性对比分析。

2.1 典型飞行状态气动特性分析

2.1.1 悬停状态

为比较2种构型旋翼的悬停性能，图8给出了共轴双旋翼的上、下2个旋翼及单旋翼构型的旋翼悬停效率。图中共轴双旋翼上、下2个旋翼拉力之和等于单旋翼拉力，在该配平状态下2种构型旋翼的下洗流对机身和机翼产生的垂向载荷差别很小，在配平误差范围内。通过对比共轴双旋翼2个旋翼，可知下旋翼拉力低于上旋翼拉力且下旋翼性能比上旋翼性能差，与文献[11]的结论一致。引入文献[24]的计算方法对共轴双旋翼的悬停效率(figure of merit, FM)进行分析：

$$F_{\mathrm{M}}=\frac{\kappa_{\mathrm{int}}(C_{\mathrm{Tu}}^{3/2}+C_{\mathrm{Tl}}^{3/2})}{2C_{\mathrm{P}}} \tag{3}$$

式中：κ_{int}为上、下旋翼功率干扰因子；C_{Tu}和C_{Tl}分别为上、下旋翼的拉力系数；C_{P}为整副旋翼的功率系数。计算结果表明共轴双旋翼的悬停效率为0.79，相较于单旋翼的0.74提高了6.8%。

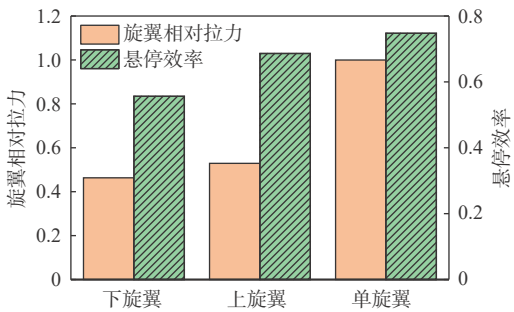


图8 单旋翼构型与共轴双旋翼构型悬停效率对比

Fig. 8 Comparison of hovering efficiency between single-rotor configuration and coaxial-rotor configuration

图9为悬停状态下机翼表面压强分布云图，单旋翼构型中的上方机翼(图中左侧)处于螺旋桨下游和旋翼桨叶后行侧区域，可以看出在机翼外侧(图中红色虚线区域)出现了大片高压区，下方机翼(图中右侧)处于螺旋桨上游和旋翼桨叶前行侧区域，该机翼前缘的低压区范围更大；而共轴双旋翼构型的机翼表面压强分布则基本对称。因为悬停状态下单旋翼构型的旋翼反扭矩需要两侧螺旋桨产生等大反向的推力来平衡，两侧机翼分别处于螺旋桨滑流的上游和下游区域，桨盘上下方的压强不同使得两者的气动环境有很大差异；而共轴双旋翼构型的旋翼反扭矩能够自平衡，螺旋桨处于关闭状态，对两侧机翼基本不存在干扰，故二者的气动环境具有对称性。

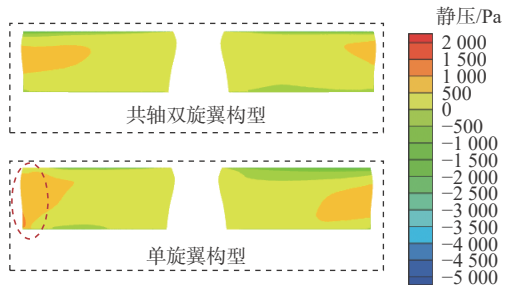
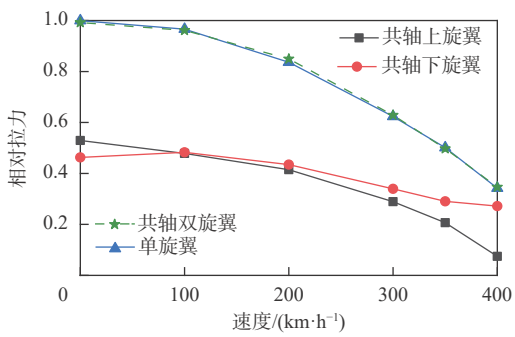


图9 悬停状态机翼上表面压强对比

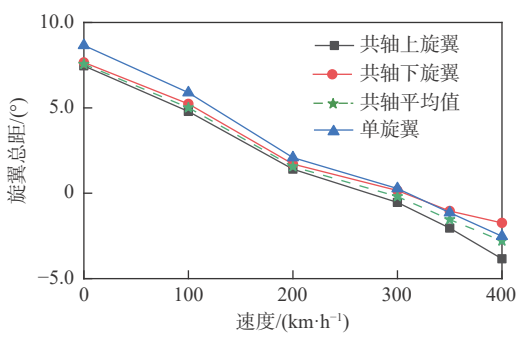
Fig. 9 Comparison of upper surface pressure of wing in hovering

2.1.2 前飞状态

复合式直升机机翼随前飞速度的增加逐渐产生升力并使旋翼卸载。根据配平要求，在共轴双旋翼构型直升机的流场计算中始终保持其旋翼拉力随前飞速度的变化与单旋翼构型旋翼拉力的变化一致。图10为共轴双旋翼上、下旋翼拉力和总距随前飞速度的变化规律。低速前飞时下旋翼会受到上旋翼尾流的影响，当二者总距相同时，下旋翼桨叶的实际迎角要小于上旋翼，因此，为平衡旋翼反扭矩，下旋翼的总距要大于上旋翼，但其产生的拉力仍小于上旋翼；由于上旋翼的尾迹偏斜角随前飞速度增大而增大，其对下旋翼的气动干扰减弱进而上、下2个旋翼的拉力相接近，但此时的总距仍有差异。随着前飞速度进一步增大，上、下旋翼的总距差异会扩大。这是因为在高速飞行时(V>300 km/h)机翼将产生较大的升力，其上表面的低压区会对旋翼产生气动干扰。通过图11给出的有/无机翼



(a) 旋翼拉力



(b) 旋翼总距

图10 旋翼拉力和总距随前飞速度的变化

Fig. 10 Variation of rotor thrust and collective pitch with forward flight speed

状态下旋翼横向剖面压强云图(图 (a) 和图 (b) 旋翼的操纵量相同)可以看出, 机翼对下旋翼的气动干扰较为严重, 在机翼的影响下, 下旋翼的桨盘下表面压强降低, 进而使得拉力降低。因此, 为了使共轴双旋翼反扭矩平衡, 需要增大下旋翼的总距, 这导致上、下旋翼总距出现较大差异。图 10(a) 反映了由于上、下旋翼总距差异的扩大所导致的上、下旋翼拉力差异的扩大。

复合式直升机的机翼处在旋翼下方受旋翼气动干扰严重, 为比较不同构型的机翼气动性能, 图 12(a) 给出了 2 种构型直升机的机翼升力和升阻比随前飞速度的变化规律。由图可知, 机翼在低速时受旋翼下洗流的影响较大升力为负, 产生了“增重效应”[25]; 2 种构型直升机的机翼升力随前飞速度的变化基本相同, 但共轴双旋翼构型的机翼升阻比要高于单旋翼构型。此外, 由图 12(b) 可知, 共轴双旋翼构型两侧机翼升力分布比较均匀, 但单旋翼构型两侧机翼升力有较大差异, 其机翼升力分配不均会使机体产生附加滚转力矩, 进而对全机的平衡和操纵不利。

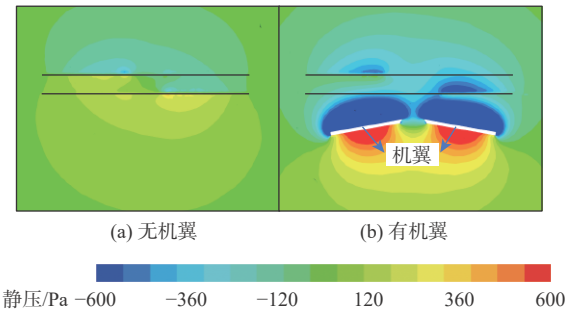


图 11　旋翼桨盘横向剖面压强对比, V=400 km/h

Fig. 11　Pressure comparison of transverse profile of rotor disk, V = 400 km/h

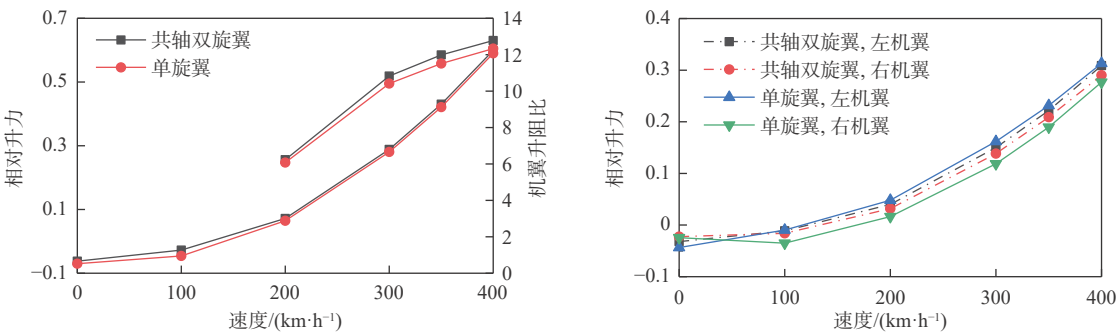


图 12　机翼气动特性随前飞速度的变化

Fig. 12　Variation of aerodynamic characteristics of wing with forward flight speed

2 种构型直升机前进的动力都来自螺旋桨, 但二者螺旋桨工况不同: 单旋翼构型在低速飞行时螺旋桨除要提供向前的推进力外还需要平衡旋翼反扭矩, 因此, 两侧螺旋桨的推力大小不相同, 而高速飞行时的旋翼反扭矩可通过垂尾产生侧向力来平衡, 两侧螺旋桨的推力大小相同; 共轴双旋翼构型的螺旋桨由于不需要平衡旋翼反扭矩, 故推力大小始终相等。图 13 为 2 种构型直升机配平状态下螺旋桨总距的变化情况。低速飞行时单旋翼构型的螺旋桨总距相差很大, 并且左侧螺旋桨为了产生向后的推力总距为负值; 而共轴双旋翼构型两侧螺旋桨的总距则始终保持一致, 并且随前飞速度的变化呈规律的线性分布, 从操纵角度来说该构型相对有利。

图 14 为 2 种构型直升机的螺旋桨推进效率对

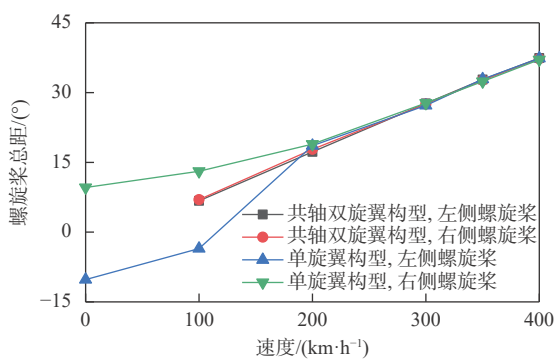


图 13　螺旋桨总距随前飞速度的变化

Fig. 13　Variation of propeller collective pitch with forward flight speed

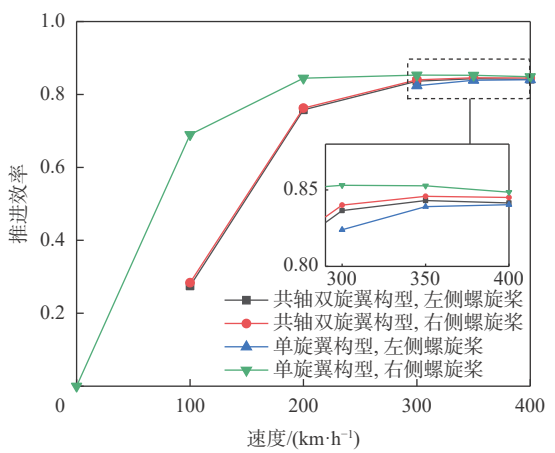


图 14　螺旋桨推进效率随前飞速度的变化

Fig. 14　Variation of propeller propulsion efficiency with forward flight speed

比，所采用的推进效率η的计算式为

$$\eta=\frac{TV_0}{P}=\frac{C_T}{C_P}\lambda \tag{4}$$

式中：V_0为前飞速度；$\lambda=V_0/n_sD$为螺旋桨的前进比；C_T和C_P分别为螺旋桨的拉力和功率系数。

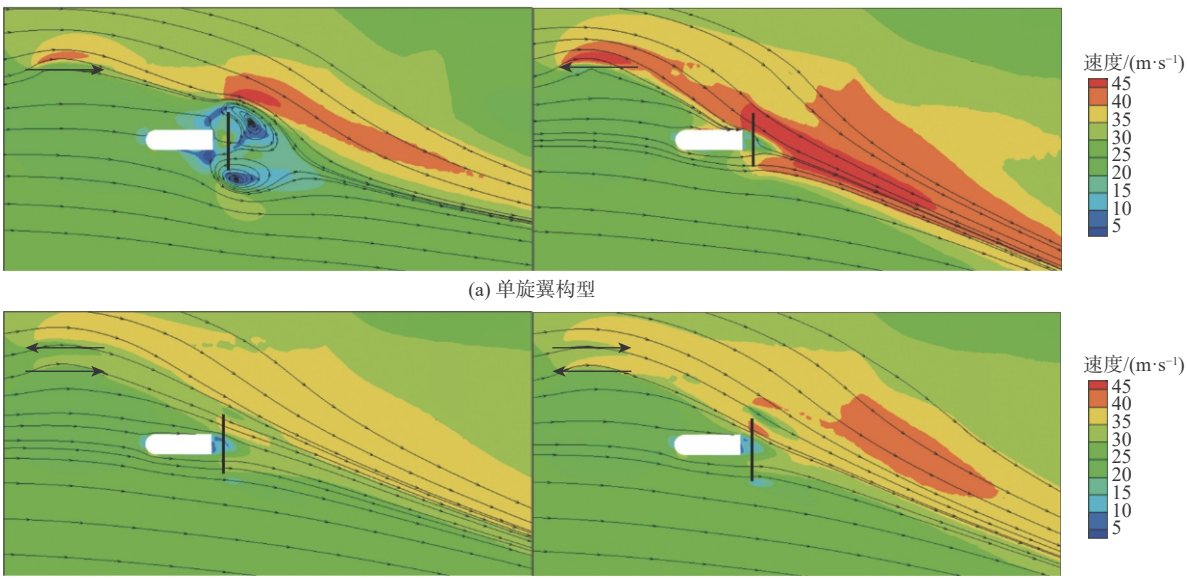
因为机体阻力随前飞速度增加而增加，所以配平所需的螺旋桨推力增加进而推进效率增大，但在达到一定速度后，螺旋桨桨尖出现激波使叶片阻力增大，螺旋桨功率增大进而效率降低。另外，由于共轴双旋翼构型的旋翼对两侧螺旋桨的气动干扰对称，故二者的推进效率基本相同；而单旋翼构型虽然在大速度前飞时两侧螺旋桨推力一致，但旋翼下洗流的不对称性会对两侧螺旋桨产生不同程度的气动干扰，二者的推进效率仍有较大差异。

2.2 旋翼-螺旋桨气动干扰特性分析

为比较2种构型直升机在不同飞行状态下旋翼对螺旋桨的气动干扰情况，本节给出螺旋桨纵向剖面和桨盘平面的流场分布云图。

图15为前飞速度为100 km/h时螺旋桨纵向剖面气流分布情况。图中箭头是旋翼桨叶运动方向，黑色短线是螺旋桨所在位置。旋翼下洗流的干扰使螺旋桨滑流由水平方向向下发生偏折，图15(a)中气流呈现明显的不对称性，左侧的螺旋桨出现了涡环状态[26-27]并形成大范围的气流紊乱；对比图15(b)可知两侧螺旋桨气动环境基本对称，这是共轴双旋翼构型气动布局的优势。

图16为前飞速度为100 km/h和350 km/h时2种

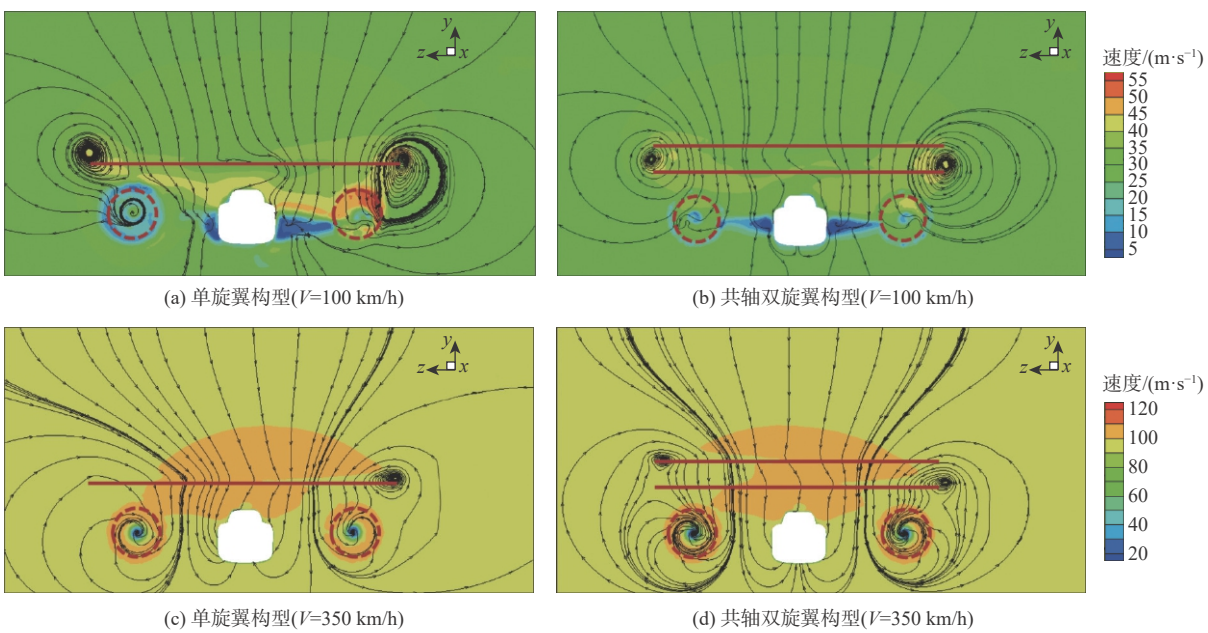


(a) 单旋翼构型

(b) 共轴双旋翼构型

图15 两侧螺旋桨纵向剖面流场分布(V=100 km/h)

Fig. 15 Flow field distribution in longitudinal profile of propellers (V = 100 km/h)



(a) 单旋翼构型(V=100 km/h)

(b) 共轴双旋翼构型(V=100 km/h)

(c) 单旋翼构型(V=350 km/h)

(d) 共轴双旋翼构型(V=350 km/h)

图16 螺旋桨桨盘平面流场分布

Fig. 16 Flow field distribution of propeller disk plane

构型直升机的螺旋桨桨盘平面流场分布情况。图中红色虚线是螺旋桨所在位置, 红色实线是旋翼所在位置, 视图方向为从机尾向机头看。当飞行速度较低时, 旋翼尾迹在所示平面上形成的大尺度旋涡使得螺旋桨附近流场比较紊乱, 表明此时旋翼对螺旋桨的气动干扰作用明显; 而当飞行速度较高时, 经过螺旋桨桨盘的流线呈现出绕旋转中心的稳定旋涡状, 且旋涡方向与螺旋桨旋向一致, 表明此时螺旋桨的下洗流是稳定的, 进一步说明旋翼对螺旋桨的干扰很小。此外, 由图 16(b) 和图 16(d) 可知共轴双旋翼构型两侧螺旋桨附近的流线分布从低速到高速相对于机身基本对称, 也说明共轴双旋翼对螺旋桨的气动干扰始终对称。

为进一步比较 2 种构型直升机对螺旋桨的干扰特性, 进行孤立螺旋桨的气动分析, 计算中保持孤立螺旋桨总距与全机流场下螺旋桨配平总距一致。图 17 为 2 种构型直升机的螺旋桨推力与孤立螺旋桨推力的对比结果。在低速飞行状态下, 单旋翼构型的螺旋桨推力变化幅度大于共轴双旋翼构型; 因为在旋翼拉力相等的情况下, 单旋翼构型的旋翼下洗流速度大于共轴双旋翼构型的下洗流速度(见图 16(a) 和图 16(b)); 而旋翼下洗流会增加螺旋桨的空气流量进而增大螺旋桨的推力, 由于单旋翼对空气流量的增幅较大, 所以该构型的螺旋桨推力变化也较大。在中高速飞行状态下 (V>200 km/h), 全机流场中的螺旋桨推力与孤立螺旋桨推力的差异随前飞速度提升而逐渐增大。这是由于机翼上下表面压差增大导致翼尖涡增强, 而推进式螺旋桨处于机翼尖端的正后方, 进而涡系对螺旋桨产生较大的气动干扰。图 17(c) 对比了高速飞行时 2 种构型直升机的螺旋桨推力变化率:

$$\Delta T = \frac{T_{\cdot P} - T'_{\cdot P}}{T'_{\cdot P}} \times 100\% \tag{5}$$

式中: $T_{\cdot P}$为全机流场中的螺旋桨推力(·表示 L 或 R, LP 表示机身左侧螺旋桨, RP 表示机身右侧螺旋桨); $T'_{\cdot P}$为孤立螺旋桨推力。

由图 17(c) 可知, 当 V=300 km/h 时, 单旋翼构型直升机的两侧螺旋桨推力变化率有很大差异。通过分析图 17(a) 可知, 此时左侧螺旋桨推力仍小于右侧螺旋桨推力, 又由于单旋翼对螺旋桨还存在部分非对称的气动干扰, 所以二者推力变化率有很大不同。当飞行速度达到 V=400 km/h 时, 2 种构型直升机对应同一侧的螺旋桨推力变化率相似, 一是因为高速飞行时 2 种构型直升机的螺旋桨工况相同且产生的推力大小相同; 二是因为旋翼的气动干扰很小而机翼的气动干扰情况类似, 螺旋桨所处的气动环境也相似, 所以 2 种构型直升机的螺旋桨推力变化率相差不大。整体来看, 共轴双旋翼构型的螺旋桨推力变化率随速度变化比较均匀且两侧螺旋桨推力变化率相差不大, 表明该构型对螺旋桨的气动干扰有良好的对称性; 而单旋翼构型两侧螺旋桨推力变化率有较大差异, 对称性不好, 对机身的稳定性和操纵性不利。

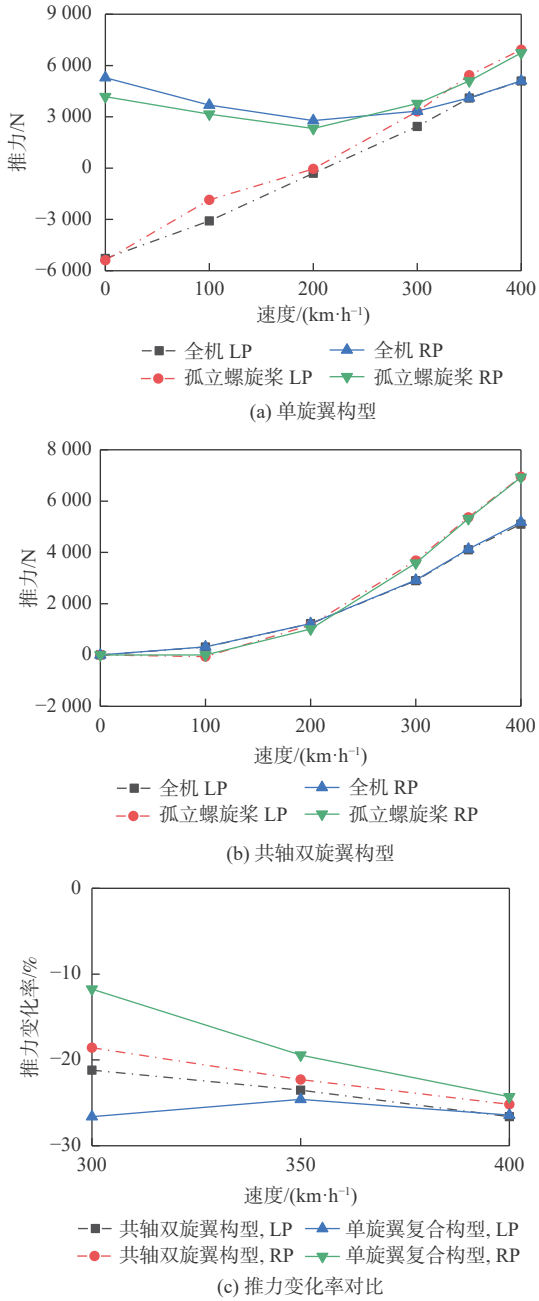


图 17　螺旋桨推力与孤立螺旋桨的对比

Fig. 17　Comparison of propeller thrust with isolated propeller thrust

2.3　全机需用功率分析

复合式构型直升机需用功率主要包括旋翼和螺旋桨两部分(不考虑传动损失和附件损耗), 需用功率是评价直升机飞行性能的重要指标。图 18 为 2 种构型直升机在飞行包线内不同前飞速度下的需

用功率。由图 18 可知，2 种构型直升机的需用功率随前飞速度先减小后增大，且共轴式双旋翼构型在悬停和低速飞行状态下的需用功率明显低于单旋翼构型，悬停时前者的需用功率仅占后者的 74.3%。如图 19 所示，共轴双旋翼构型功率较低的原因是低速条件下该构型的旋翼诱导功率较低使得旋翼需用功率较低；另一方面，螺旋桨的推力处在较低水平，对应螺旋桨功率也低于单旋翼构型。此外，图 19 表明高速飞行状态下 2 种构型的螺旋桨功率相持平，因为此时二者的螺旋桨都主要为克服前飞阻力做功；随着飞行速度进一步提升，共轴双旋翼构型的总需用功率呈现出高于单旋翼构型的趋势。复合式直升机高速前飞时旋翼需用功率以型阻功率为主，后者主要与桨叶激波阻力有关。为进一步分析共轴双旋翼阻力较大的原因，图 20 给出

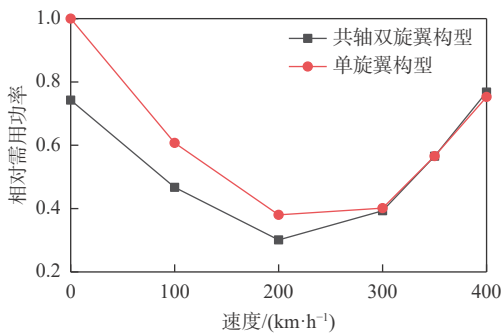


图 18 全机需用功率对比

Fig. 18 Comparison of total required power

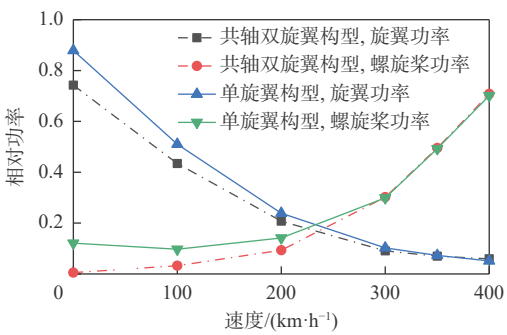


图 19 需用功率组分对比

Fig. 19 Comparison of required power components

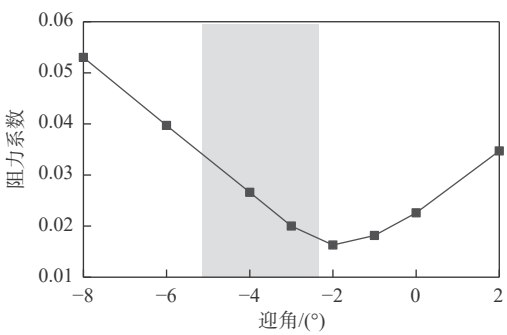


图 20 旋翼翼型阻力特性曲线

Fig. 20 Drag characteristic curve of rotor airfoil

了前行桨叶靠近桨尖位置处剖面翼型阻力特性曲线，来流马赫数为 0.85。由图 20 可知，由于高速飞行时旋翼总距为负值，而共轴双旋翼桨叶来流迎角小于单旋翼构型（见图 10(b)），因此，产生了较大的激波阻力。

3 结 论

本文针对共轴双旋翼加双推进螺旋桨新型复合式构型直升机，采用基于动量源的 CFD 方法进行了全机配平和流场数值模拟；对比分析了新型复合式构型直升机与传统单旋翼复合式构型直升机在不同前飞速度下旋翼、螺旋桨和机翼的气动特性及三者之间的气动干扰特性，得到以下结论：

1）共轴双旋翼构型具有良好的悬停性能，在旋翼拉力和实度相同的条件下，共轴双旋翼构型的悬停效率高出单旋翼构型 6.8%；

2）在低速飞行时，单旋翼构型直升机一侧螺旋桨产生负推力，螺旋桨效率低，甚至会陷入涡环状态，对飞行安全不利，而共轴双旋翼构型没有这一现象；

3）相较于单旋翼构型，共轴双旋翼构型的旋翼对螺旋桨和机翼的气动干扰更小，且在全速域范围内旋翼的气动干扰相对机身纵向对称面具有良好的对称性；

4）低速前飞时，共轴双旋翼构型的需用功率显著低于单旋翼构型，当 V=100 km/h 时，前者相较于后者需用功率降低了 23.1%；高速前飞时，共轴双旋翼构型的需用功率大于单旋翼构型，这与共轴双旋翼阻力较大有关。

参考文献（References）

[1] EDI P, YUSOFF N, YAZID A A, et al. New design concept of compound helicopter[C]// Proceedings of the 4th WSEAS International Conference on APPLIED and THEORETICAL MECHANICS. Athens: WSEAS Press, 2008: 101-105.

[2] ORCHARD M, NEWMAN S. The fundamental configuration and design of the compound helicopter[J]. Journal of Aerospace Engineering, 2003, 217(6): 297-315.

[3] 赵寅宇, 黎鑫, 史勇杰, 等. 双拉力螺旋桨构型复合式高速直升机旋翼/螺旋桨干扰流场分析[J]. 南京航空航天大学学报, 2017, 49(2): 154-164.
ZHAO Y Y, LI X, SHI Y J, et al. Analysis of rotor/propeller interference flow field of double-pull propeller configuration composite high-speed helicopter[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2017, 49(2): 154-164(in Chinese).

[4] FERGUSON K, THOMSON D. F1ight dynamics investigation of compound helicopter configurations[J]. Journal of Aircraft, 2015, 52(1): 1-12.

[5] BLACKWELL R, MILLOTT T. Dynamics design characteristics of

the Sikorsky X2 technology™ demonstrator aircraft: 64-2008-000174[R]. Montreal: The American Helicopter Society, 2008: 1-13.

[6] 于琦. 美军下一代直升机的需求与竞争[J]. 兵器知识, 2020(12): 51-55.
YU Q. Demand and competition of American next generation helicopter[J]. Ordnance Knowledge, 2020(12): 51-55(in Chinese).

[7] 朱振华, 朱清华, 申遂愿. 一种涵道螺旋桨推力共轴双旋翼直升机: CN109665096A[P]. 2019-04-23.
ZHU Z H, ZHU Q H, SHEN S Y. A coaxial twin rotor helicopter with ducted thrust propeller: CN109665096A[P]. 2019-04-23(in Chinese).

[8] BAGAI A. Aerodynamic design of the X2 technology demonstrator™ main rotor blade: 64-2008-000227[R]. Montreal: The American Helicopter Society, 2008: 1-16.

[9] KENYON A R, BROWN R E, DURAISAMY K. Interactional aerodynamics and acoustics of a hingeless coaxial helicopter with an auxiliary propeller in forward flight[J]. Aeronautical Journal, 2009, 113(1140): 65-78.

[10] XU H Y, YE Z Y. Numerical simulation of unsteady flow around forward flight helicopter with coaxial rotors[J]. Chinese Journal of Aeronautics, 2011, 24(1): 1-7.

[11] 李文浩. 复合式高速直升机旋翼/机身气动干扰特性的 CFD 分析[D]. 南京: 南京航空航天大学, 2012.
LI W H. CFD analysis of rotor/fuselage aerodynamic interference characteristics of compound high speed helicopter[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012(in Chinese).

[12] 张银. 复合式共轴刚性旋翼直升机气动干扰及飞行特性分析[D]. 南京: 南京航空航天大学, 2014.
ZHANG Y. Aerodynamic interference and flight characteristics analysis of compound coaxial rigid rotor helicopter[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2014(in Chinese).

[13] 吴裕平, 习娟, 范俊. 共轴刚性旋翼高速直升机配平及旋翼系统气动特性研究[J]. 直升机技术, 2018, 2(1): 1-7.
WU Y P, XI J, FAN J. Research on balancing and rotor system aerodynamic characteristics of coaxial rigid rotor high speed helicopter[J]. Journal of Helicopter Technology, 2018, 2(1): 1-7 (in Chinese).

[14] 段赛玉, 陈铭. 复合式共轴直升机飞行动力学数学模型研究[J]. 飞机设计, 2011, 31(3): 1-6.
DUAN S Y, CHEN M. Research on flight dynamics mathematical model of compound coaxial helicopter[J]. Aircraft Design, 2011, 31(3): 1-6(in Chinese).

[15] 王倩楠. 复合推力高速直升机气动干扰特性分析[D]. 南京: 南京航空航天大学, 2021.
WANG Q N. Analysis of aerodynamic interference characteristics of compound thrust high-speed helicopter[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2021(in Chinese).

[16] BLAZEK J. Computational fluid dynamics: Principles and applications, second edition[M]. Amsterdam: Elsevier, 2005: 16-18.

[17] RAJAGOPALAN R G, LIM C K. Laminar flow analysis of a rotor in hover[J]. Journal of the American Helicopter Society, 1991, 36(1): 12-23.

[18] RAJAGOPALAN R G, MATHUR S R. Three dimensional analysis of a rotor in forward flight[J]. Journal of the American Helicopter Society, 1993, 38(3): 14-25.

[19] RAJAGOPALAN R G, BERG D E, KLIMAS P C. Development of a three-dimensional model for the Darrieus rotor and its wake[J]. Journal of Propulsion and Power, 1995, 11(2): 185-195.

[20] RAJAGOPALAN R G. A procedure for rotor performance, flowfield and interference——A perspective[C]//Proceedings of the 38th Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2000.

[21] MCKEE J W, NAESETH R L. Experimental investigation of the drag of flat plates and cylinders in the slipstream of a hovering rotor: NASA-TN-4239[R]. Washington, D. C. : NASA, 1958: 1-43.

[22] JOE W E, SUSAN L A, RICHARD H S. Inflow measurement made with a laser velocimeter on a helicopter model in forward flight. Volume 1: Rectangular planform blades at an advance ratio of 0.15: NASA-TM-100541 [R]. Washington, D. C. : NASA, 1998: 1-323.

[23] FREEMAN C E, MINECK R E. Fuselage measurements of a helicopter wind-tunnel model with a 3.15-meter diameter single rotor: NASA-TM-80051[R]. Washington, D. C. : NASA, 1979: 1-170.

[24] LEISHMAN J G, SYAL M. Figure of merit definition for coaxial rotors[J]. Journal of the American Helicopter Society, 2008, 53(3): 290-300.

[25] LEISHMAN J G. Principles of helicopter aerodynamics[M]. Cambridge: Cambridge University Press, 2006: 304-313.

[26] TOROPOV M Y, STEPANOV S Y. Modeling of helicopter flight imitation in the vortex ring state[J]. Russian Aeronautics, 2016, 59(4): 517-522.

[27] 王涌钦, 余新, 陈仁良, 等. 双螺旋桨推进复合式直升机操纵分配与最优过渡路线设计[J]. 南京航空航天大学学报, 2022, 54(2): 211-218.
WANG Y Q, YU X, CHEN R L, et al. Redundant control and optimal transition route design of compound helicopter[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2022, 54(2): 211-218(in Chinese).

Aerodynamic interference characteristics of a new compound configuration helicopter

LIANG Jiahui，ZHANG Xiayang，ZHAO Qijun*

(National Key Laboratory of Helicopter Aeromechanics，Nanjing University of Aeronautics and Astronautics，Nanjing 210016，China)

Abstract：By combining the characteristics of high hovering performance and control efficiency of coaxial-rotor helicopters and the advantage of high-speed forward flight performance of double-thrust-propeller compound configuration helicopters, the coaxial rotor was applied to the double-thrust-propeller compound configuration helicopter. The single rotor was changed to a coaxial rotor based on the configuration of the X3 compound helicopter. To study the influence of coaxial rotors on the aerodynamic characteristics of compound configuration helicopters, a fast trim method was established. On this basis, the aerodynamic characteristics of different configuration helicopters were analyzed. The results indicate that the aerodynamic characteristics of the coaxial-rotor helicopter have good symmetry compared with the single-rotor configuration. In the case of maintaining good high-speed forward flight performance, the coaxial-rotor configuration can significantly improve the hovering and low-speed flight performance. When the hovering efficiency is increased by 6.8%, and the forward flight speed is 100 km/h, the total required power is reduced by 23.1%. The aerodynamic interference of the coaxial rotor to the propeller at low speeds is greatly lower than that of the single rotor.

Keywords：coaxial rotor；double-thrust-propeller；compound helicopter；rotor momentum source；aerodynamic interference

Received：2023-02-28；**Accepted**：2023-07-04；**Published Online**：2023-07-11 11：04

URL：link.cnki.net/urlid/11.2625.V.20230711.0914.001

Foundation items：National Natural Science Foundation of China (12032012,12102186); State Key Laboratory Foundation of China (61422202201); Natural Science Foundation of Jiangsu Province (BK20200433)

* **Corresponding author**. E-mail：zhaoqijun@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0090

基于位置约束与注意力的低空无人机障碍物检测方法

唐友军[1]，缪存孝[1,*]，张贺[1]，李玉峰[1]，叶文[2]
(1. 北京科技大学 机械工程学院，北京 100083；2. 中国计量科学研究院，北京 100029)

摘 要：无人机（UAV）在低空领域广泛应用于电力巡检、搜救、侦察等任务，对飞行过程中的障碍物进行提前检测是完成既定任务的安全保障。为满足无人机低空飞行时对障碍物的检测精度要求及位置回归精度要求，提出一种基于位置约束与注意力改进的低空无人机障碍物检测方法。分析位置回归损失函数的不足并基于此提出分离尺度损失与融合方向约束的损失函数对回归过程进行优化；改进注意力机制CBAM提出双重注意力机制以强化特征抑制干扰，提高检测性能。实验结果表明：本文方法在mAP上提高了2.28%，在mAP@0.5:0.95上提高了2.7%，在检测精度和位置回归精度上都表现出了更好的低空障碍物检出性能。

关 键 词：低空无人机；障碍物检测；位置回归损失函数；双重注意力机制；位置回归损失函数
中图分类号：V249
文献标志码：A **文章编号**：1001-5965(2025)03-0933-10

无人机因高机动性等优势，被广泛应用于军民领域，如作物病虫防治、植被检测、战场地势勘察、侦察与反侦察等；在日常生活中，无人机应用于物流运输、气象监测等[1-4]。随着低空空域的开放，无人机的应用也得到进一步的发展，电力巡线、人员搜救、视频拍摄等低空领域都能见到无人机的身影[5-6]。然而野外低空飞行时，无人机的安全事故频发，最主要的原因是在无人机的飞行路线上存在障碍物，无人机无法对障碍物进行准确的识别预警，或对障碍物的方位信息判断有误，导致无法快速执行避障操作而直接相撞，造成严重事故。野外低空空域中无人机的主要障碍物包括树木、电力线塔与电线杆等3类，对障碍物的识别与定位对于无人机安全飞行格外重要。其中，对主要障碍物的准确检测与障碍物在图像中的准确位置回归是融合定位避障的基础。传统检测方法无法满足实时性与精度的要求，基于深度学习的目标检测方法具有精度高、速度快的特点，是实现无人机障碍物检测最重要的手段之一[7]。

基于深度学习的目标检测算法分为两阶段算法和单阶段算法。两阶段算法的代表有R-CNN[8]、SPPNet[9]、Fast R-CNN[10]及Faster R-CNN[11]等，算法检测精度较高但检测速度十分有限。单阶段目标检测算法兼顾速度和精度，代表算法有YOLO[12]、SSD[13]和RetinaNet[14]等。其中，YOLO算法具有较高的检测精度和实时性，在工业上应用广泛。YOLOv5平衡速度与精度，相比于其他复杂的网络对机载嵌入式设备的资源消耗更低，更适用于无人机对飞行过程障碍物的检测与预警。因此，本文针对无人机低空飞行下的障碍物检测场景对YOLOv5进行优化。

低空飞行时，主要障碍物尺度差异明显，背景复杂，导致检测精确率下降及位置回归不精确等问题。在回归过程中，位置回归损失函数对预测框的准确回归格外重要。Girshick[10]提出Smooth L1损失函数，位置回归时将边界信息进行独立求和；Yu等[15]

收稿日期：2023-02-28；录用日期：2023-06-25；网络出版时间：2023-08-02 10:15
网络出版地址：link.cnki.net/urlid/11.2625.V.20230802.0905.001
*通信作者. E-mail：miaocx@ustb.edu.cn

引用格式：唐友军，缪存孝，张贺，等. 基于位置约束与注意力的低空无人机障碍物检测方法[J]. 北京航空航天大学学报，2025，51(3)：933-942. TANG Y J，MIAO C X，ZHANG H，et al. A low-altitude UAV obstacle detection method based on position constraint and attention mechanism[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：933-942 (in Chinese).

提出IOU损失函数，关联回归时的边界信息，但交并比为0时无法反映回归的相关性；Rezatofighi等[16]提出GIOU损失函数，用真实框与预测框之间的最小包络区域进行回归优化，但存在包含情况时无法提供优化方向；Zheng等[17]提出DIOU损失函数，考虑预测框与真实框的中心距，随后提出CIOU损失函数考虑了宽高比。无人机低空飞行时背景复杂，针对复杂背景下的无人机目标检测，奉志强等[4]通过改进注意力机制的全连接层与连接方式，提高复杂背景下的特征捕获能力；李子豪等[18]通过在颈部网络中融入自适应协同注意力机制对权重自适应加权并对网络进行轻量化设计，在减少模型参数的情况下精度可明显提升；彭继慎等[19]通过在主干引入注意力机制提高检测精度，采用深度可分离卷积与模型压缩方式提高检测速度；张瑞倩等[20]通过引入空洞卷积模块以增大感受野，提高了复杂背景下目标的检测召回率和精度。

本文针对低空无人机飞行过程中对障碍物检测的精度要求及位置回归精度要求，分析障碍物位置回归时的不足并引入双通道注意力机制，基于此提出一种基于位置约束与注意力改进的低空无人机障碍物检测方法，仿真与实际实验验证算法在低空障碍物检测时的有效性。

1 障碍物检测方法总体框架

本文提出一个适用于无人机低空飞行时对主要障碍物进行检测的模型，如图1所示。首先，在位置回归损失函数中，对原函数在障碍物检测方面的不足进行分析并改进，将原函数进行尺度损失分离并融合方向约束，提出新的位置回归损失L_{loc}。其次，为使网络更加关注目标特征减弱对冗余干扰的响应，基于注意力的优化思想提出双重注意力机制(dual attention convlution block, DACB)，基于此构建特征强化网络对主干特征和颈部网络中提取的特征进行增强，提高检测性能。

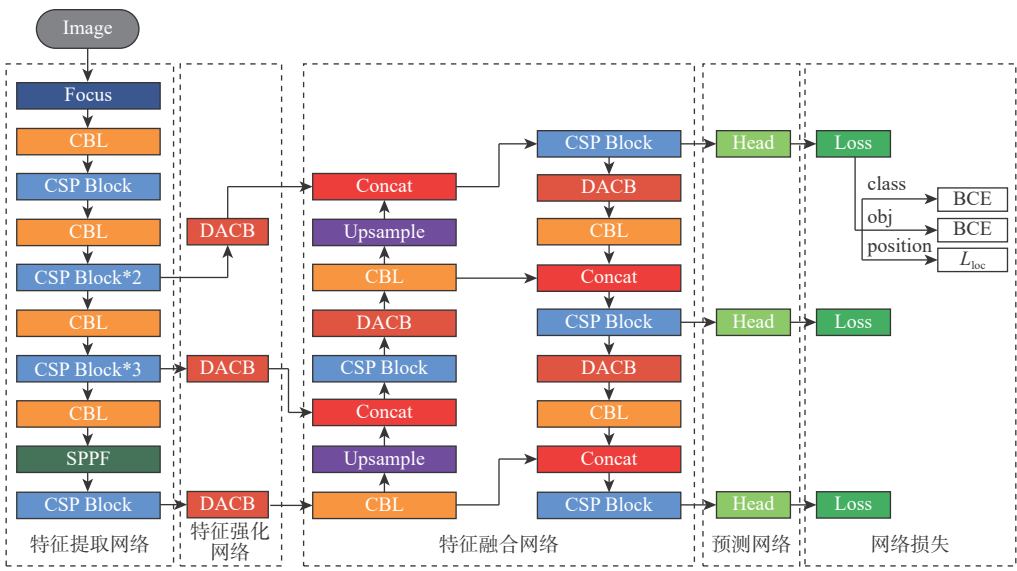


图1 低空无人机障碍物检测模型

Fig. 1 Low-altitude UAV obstacle detection model

网络基本结构与YOLOv5算法保持一致。在DACB的残差连接中采用两次卷积操作的卷积核大小为3，步长为1，Padding为1，归一化后的激活函数为Sigmoid。

2 基于位置约束与注意力改进的低空无人机障碍物检测方法

2.1 分离尺度损失与融合方向约束的损失函数优化

在YOLOv5算法中，位置回归损失函数CIOU定义为

$$L_{\mathrm{CIOU}}=1-L_{\mathrm{CIOU}}=1-r_{\mathrm{IOU}}+\frac{d^2}{c^2}+\alpha v \tag{1}$$

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

式中：L_{CIOU}为CIOU损失函数值；r_{IOU}为交并比；d为真实框与预测框两者中心的欧氏距离；c为真实框与预测框最小闭包区域的对角线长度；α为长宽比一致的权衡函数；v为衡量真实框与预测框宽高比一致性的参数；w^{gt}、h^{gt}、w、h分别为真实框的宽度、真实框的高度、预测框的宽度、预测框的高度。

但是在野外低空空域所涉及的3类主要障碍物中，电线杆、电力线塔与树木三者之间真实框的宽高比具有显著差异，预测框对宽高方向的优化更

加敏感。通过对低空无人机障碍物数据集 LAOD 与通用数据集 VOC 的真实框进行聚类, 如图 2 所示。

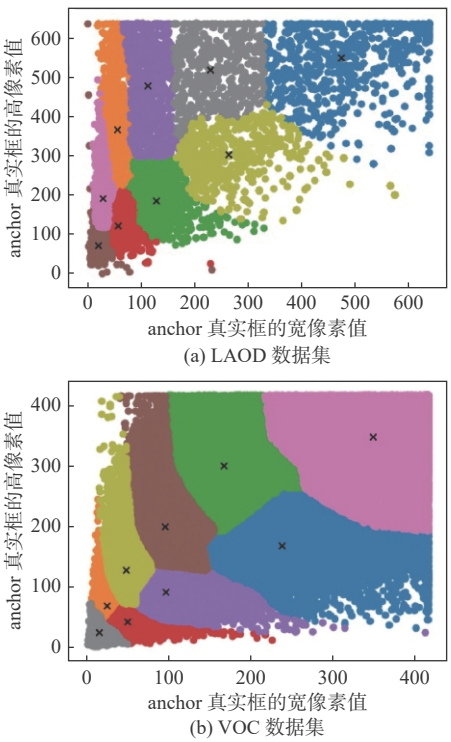


图 2　LAOD 数据集与 VOC 数据集聚类结果

Fig. 2　Clustering results of LAOD dataset and VOC dataset

由图 2 可知, LAOD 相比于 VOC 在宽高比的分布上具有更大的差异与波动, 因此, 在宽高的优化上更加敏感。

在式 (2) 中, v 用于衡量预测框回归时的宽高比一致性。但是根据式 (2) 进行位置回归时并不能代表真实的宽高变化。

首先, 在预测框的回归过程中, 如果预测框和真实框的宽高比呈线性比例, 即满足 $w=kw^{gt}$ 且 $h=kh^{gt}$ (k 为常数), $v=0$, CIOU 中对宽高的惩罚项便不再起作用, 函数退化。

其次, 由 v 对 w 和 h 的偏导数:

$$\frac{\partial v}{\partial w}=\frac{8h}{\pi^2(w^2+h^2)}\left(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w}{h}\right) \tag{3}$$

$$\frac{\partial v}{\partial h}=-\frac{8w}{\pi^2(w^2+h^2)}\left(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w}{h}\right) \tag{4}$$

显然:

$$\frac{\partial v}{\partial w}\cdot\frac{\partial v}{\partial h}<0 \tag{5}$$

而在优化过程中, 宽 w 与高 h 的更新方式为

$$w\leftarrow w-r\frac{\partial v}{\partial w} \tag{6}$$

$$h\leftarrow h-r\frac{\partial v}{\partial h} \tag{7}$$

式中: r 为学习率。

结合式 (5)～式 (7) 可知, 障碍物预测框的宽与高在优化时, 优化方向相反。其中一个值在增大的时候, 另外一个值必然在减小, 不能同时增大或同时减小, 显然不符合优化的实际情况。且当预测框的宽和高均分别小于(或大于)真实框的宽高时, 无法优化到理想情况。因此, 平衡因子并不能很好地优化宽高, 预测框失去收敛方向, 导致定位不准, 进而会影响最终检测与定位精度。但是低空无人机障碍物检测过程对预测框在图像中的定位要求较高, 精确的位置回归有利于机载相机与其他传感器(如雷达等)的精确融合, 并以此为基础进行避障。

针对此问题, 首先, 将预测框的宽高进行独立回归, 去掉位置回归损失函数中的宽高比损失 αv, 而直接对宽高进行惩罚, 设计宽高惩罚 l_{sca} 为

$$l_{sca}=\frac{\rho^2(w,w^{gt})}{C_w^2}+\frac{\rho^2(h,h^{gt})}{C_h^2} \tag{8}$$

式中: ρ 为预测框与真实框的宽高差; C_w 和 C_h 分别为真实框与预测框最小闭包区域的宽和高。

由式 (8) 可知, 在预测框回归的过程中, 宽度与高度方向的优化彼此独立, 互不干扰。当预测框与真实框准确拟合的时候, 宽高方向的损失具有最小值, 具有更好的预测框位置回归性能。

其次, 考虑预测框与真实框之间的相对方位问题, 施加方位约束, 使得位置回归更加快速准确。在一般的位置回归过程中, 预测框在优化的时候不具有方向性, 回归过程会在方位上不断漂移, 进而也可能导致模型变差、位置回归不准确等问题。

因此, 考虑在回归过程中对预测框施加方位约束并定义角度损失, 角度损失计算方式为

$$\Lambda=1-2\sin^2\left(\arcsin x-\frac{\pi}{4}\right) \tag{9}$$

式中:

$$x=\begin{cases}\dfrac{c_{hh}}{\sigma}=\sin\alpha & \alpha<\dfrac{\pi}{4}\\[2mm] \dfrac{c_{ww}}{\sigma}=\sin\beta & \text{其他}\end{cases} \tag{10}$$

$$\sigma=\sqrt{(b_{cx}^{gt}-b_{cx})^2+(b_{cy}^{gt}-b_{cy})^2} \tag{11}$$

$$c_{hh}=\max(b_{cy}^{gt},b_{cy})-\min(b_{cy}^{gt},b_{cy}) \tag{12}$$

$$c_{ww}=\max(b_{cx}^{gt},b_{cx})-\min(b_{cx}^{gt},b_{cx}) \tag{13}$$

式中: c_{hh} 和 c_{ww} 分别为以真实框 B^{gt} 与预测框 B 中心连线为对角线所构成的矩形的高和宽; c_{hh} 和 c_{ww} 分别为真实框与预测框的中心高度差和宽度差; σ 为两框中心距。角度损失的计算示意图如图 3 所示。引入角度约束后, 模型会试着先将预测框往更靠近真实框的方向(横向或竖向)靠近, 并沿

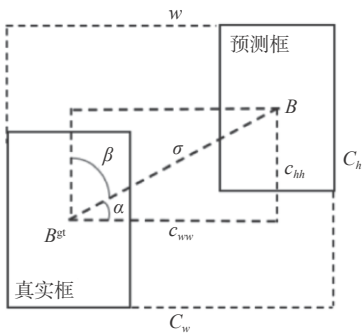


图3 角度损失计算

Fig. 3 Angle loss calculation

着这个方向对方位和其他惩罚不断地优化。如果$\alpha<\pi/4$，则不断优化α，否则对其余角β进行不断优化。

对于角度损失，假设$0<\alpha<\pi/4$，那么：

$$\frac{\mathrm{d}\Lambda}{\mathrm{d}\alpha}=-4\sin\left(\alpha-\frac{\pi}{4}\right)\cos\left(\alpha-\frac{\pi}{4}\right)>0 \tag{14}$$

由式(14)可知：

1）首先，当α被不断优化变小(α逐渐为0)，角度损失也在减小，符合优化方向。当$\alpha=0$时，$\Lambda=0$取到最小值；$\alpha=\pi/4$时，$\Lambda=1$取到最大值。

2）其次，

$$\frac{\mathrm{d}^2\Lambda}{\mathrm{d}\alpha^2}=-4\cos 2\left(\alpha-\frac{\pi}{4}\right)<0 \tag{15}$$

由式(14)和式(15)可知，角度损失是关于α的递增凸函数，那么当方位偏移较大时，角度损失也较大，此时方位约束可起到限制预测框游荡的作用(即使方位难以被约束，角度损失仍能起到限制游荡的作用，即方位约束在优化时都可起到限制回归过程的自由度的作用)。随着其余惩罚的作用及优化过程的进行，角度被不断优化，角度优化速度加快，角度损失加速下降。当方位偏差较小时即方位上拟合较好时，角度损失的贡献加速下降，最终可在方位上被精确约束，使得位置回归更加精确。

3)最后，方位被施加约束后，除了回归的过程，回归到最后的时候，当其余约束降到低值的时候，即使方位上仍存在一定偏移，如图4所示，此时仍具有一定的角度损失，可以使得预测框往更加靠近真实框的方向回归，使得位置回归更加精确。因此，可证明方向约束的有效性。

综上，本文针对野外低空障碍物的尺度特点及高位置回归精度的要求，提出一种分离尺度损失与融合方向约束的位置回归损失函数L_{loc}，表达式为

$$L_{\mathrm{loc}}=1-r_{\mathrm{IOU}}+\frac{d^2}{c^2}+l_{\mathrm{sca}}+\lambda\Lambda \tag{16}$$

式中：λ为平衡加权系数，为了平衡角度损失与其他损失的贡献，在实验中取为0.1。

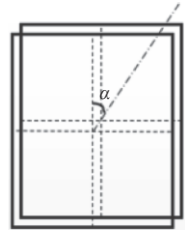


图4 角度损失说明

Fig. 4 Angle loss specification

2.2 基于残差的注意力特征优化

低空飞行时，无人机面对的背景复杂多变，视野中存在诸多无关的干扰信息。因此，考虑在网络中构建注意力机制，使网络更加关注于对目标特征的学习，抑制对无关干扰的响应。

CBAM[20](convolutional block attention module)是一种串行混合注意力机制，分别在通道和空间的维度上对特征进行重新细化调整。但是由于这种直接相连的串行结构，原始的特征图经过通道注意力调整后直接作为空间注意力的输入，该输入反映的特征好坏直接影响后续对空间方向的细化调整，在低空无人机的障碍物检测领域中，这种连接的直接干扰可能会使得注意力机制的提升效果不充分[21]。本文考虑利用ResNet的残差连接思想断开这种直接连接方式，并对通道注意力机制进行改进，基于此提出一种DACB对特征进行整合与优化。

DACB的设计如图5所示。

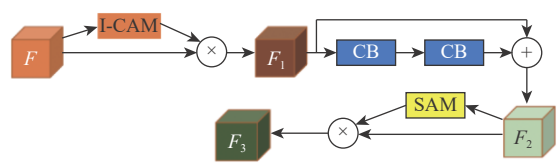


图5 DACB示意图

Fig. 5 Schematic diagram for DACB

首先，对输入特征图F，依靠I-CAM(Improved-CAM)建立通道方向的依赖关系。I-CAM是对CBAM中通道注意力CAM优化而来的，结构如图6所示。

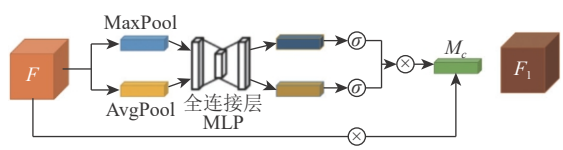


图6 I-CAM示意图

Fig. 6 Schematic diagram for I-CAM

与CBAM注意力机制中的CAM不同的是，经过共享全连接层(multilayer perceptron, MLP)输出的2个注意力权重并非直接相加再激活，而是各自激活之后再相乘。对于某个通道，这种相乘的方式可以更加突出两部分权重中都较高的位置，而对于两部分权重中仅有一处较高或两处均较低的位置，

这种方式可以起到更加明显的抑制作用。

I-CAM 整个过程表示为

$$M_C = \sigma(\mathrm{MLP}(\mathrm{MaxPool}(F))) \otimes \sigma(\mathrm{MLP}(\mathrm{MaxPool}(F))) \tag{17}$$

$$F_1 = F \otimes M_C \tag{18}$$

式中: M_C 为通道权重; σ 为 Sigmoid 激活函数; F_1 为经过通道注意力输出的特征图; MaxPool 为最大池化操作。

其次, 为避免 I-CAM 和空间注意力机制 SAM 直接串联带来的影响, 使用残差连接进行改进, 进行特征整合, 生成的中间预测 F_2 是在经过通道方向优化的特征图上进行的进一步特征提取的特征图, 是对特征的进一步细化。该过程表示为

$$F_2 = f_{\mathrm{CB}}(f_{\mathrm{CB}}(F_1)) + F_1 \tag{19}$$

式中: f_{CB} 为卷积块操作。

最后, 通过空间注意力机制 SAM 对中间预测进行空间方向的权重重新分配。如图 7 所示, 该过程可表示为

$$F_3 = \sigma(\mathrm{Conv}([\mathrm{MaxPool}(F_2); \mathrm{AvgPool}(F_2)])) \otimes F_2 \tag{20}$$

式中: F_3 为 DACB 最终输出的特征图; Conv 为卷积操作; AvgPool 为平均池化操作; M_s 为空间注意力模块的输出。

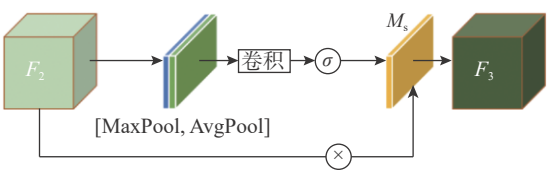


图 7　SAM 示意图

Fig. 7　Schematic diagram for SAM

将提出的 DACB 融合在颈部网络中对提取的特征进行细化, 并在特征提取网路之后基于 DACB 构建特征强化网络。

针对低空无人机的障碍物检测, 实验发现 DACB 对右下角干扰特征的抑制效果要优于 CBAM 机制, 可减弱冗杂特征的响应, 如图 8 所示。



(a) DACB 热力图　　(b) CBAM 热力图

图 8　热力图对比

Fig. 8　Heat map comparison

3　实验与分析

3.1　实验数据与训练方式

为验证算法的有效性, 构建包含低空无人机飞行过程中的主要障碍物的数据集(low altitude UAV obstacle dataset, LAOD), 数据来源于北方天途航空公司低空无人机飞行实验中所摄的图像, 飞行速度为 4～10 m/s, 拍摄设备为大疆 Mavic2 无人机与 OI-G35SH 云台吊舱, 并结合了电网巡检图像及百度图片, 共 2 033 张图像, 包含低空飞行时的主要障碍物树木、电线杆、电力线塔等 3 类, 具有不同尺度与背景等分布特点, 并利用几何变换、随机裁剪等将数据集扩充至 5 481 张样本, 并利用 LabelImg 完成数据集标注。

取其中 80% 作为训练集, 输入图像的尺寸统一调整为 640×640。基于迁移学习思想进行训练, 加载 ImageNet 预训练主干权重初始化网络主干权重参数, 再在障碍物数据集 LAOD 上进行训练。采用 SGD 对网络进行优化, 动量参数为 0.937, 权重衰减系数为 5×10^{-4}。学习率调整方式为预热与余弦退火方式, 初始学习率 0.000 5, 最小学习率 0.000 05。共训练 150 轮, 前 50 轮为冻结训练, 后 100 轮为解冻训练, 且前 100 轮开启 Mosaic 数据增强, 后 50 轮关闭 Mosaic 数据增强。

3.2　评价指标

采用平均精确度均值 mAP、召回率 R 及 F1 分数对算法进行综合评价。

对于预测框的回归精度, 选取 mAP@0.5:0.95 与 mAP@0.75 指标进行对比。计算方式分别为

$$\mathrm{mAP@0.5:0.75} = \frac{1}{10}(\mathrm{mAP@0.5} + \mathrm{mAP@0.55} + \cdots + \mathrm{mAP@0.95}) \tag{21}$$

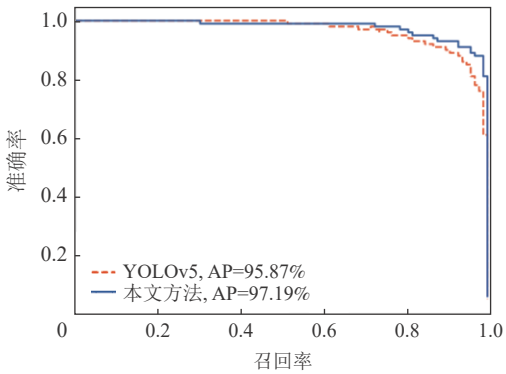
$$\mathrm{mAP@0.75} = \frac{1}{c}\sum_{i=1}^{c}\mathrm{AP@0.75}_i \tag{22}$$

式中: mAP@0.5:0.95 为在不同交并比阈值下的综合表现, 其值越高代表模型的预测框回归能力越强, 即预测框与真实框的拟合越精确, 是算法位置回归性能的准确表现[21-23]。mAP@0.75 为较高交并比阈值下较为严格的 mAP 值, 也可以用于反映回归的位置精确程度, 因此, 可用于辅助判定回归精度。

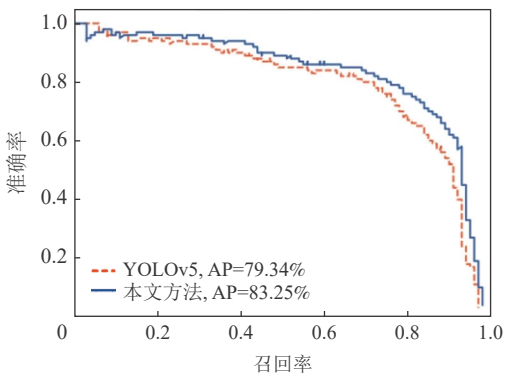
3.3　仿真实验与结果分析

实验 1: 先将改进后的算法与原生 YOLOv5 模型进行对比实验, 综合对比不同的评价指标。

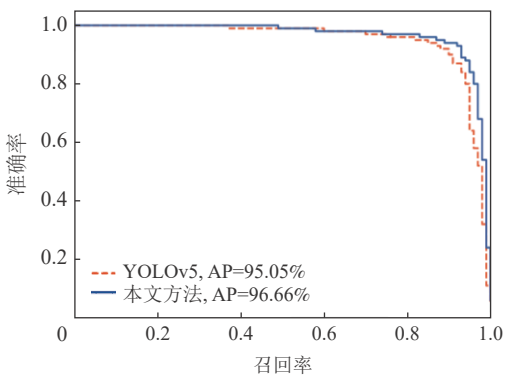
1) 平均精确度(AP)对比如图 9 所示, 图中曲线为 P-R 曲线。



(a) 电力线塔类



(b) 电线杆类



(c) 树木类

图 9 平均精确度(AP)对比

Fig. 9 AP comparison

AP 指标用于衡量单类目标的检测性能，在 *P*-*R* 曲线中表现为曲线与横轴所围成的面积。由图 9 可知，由于电力线塔和树木属于大目标，且特征更加明显不易受到干扰，因此，原模型也具有较高的 AP 值，而电线杆类特征不如其他两类明显，容易受到背景环境或遮挡等的干扰，AP 值不如其余两类。本文的改进算法通过改进位置回归损失与引入双重注意力机制特征优化，使得特征学习更加准确，位置回归更加精确，在 3 类目标的检测性能上均有所提升。在电线杆类的提升最为明显，AP 由 79.34% 提升至 83.25%，提升了 3.91%，因此，优化的算法在保持相同召回率的同时能够保持更高的检测精度，且在较高召回率的情况下仍然能保持一定精度。

2）mAP、召回率 *R*、F1 值对比如表 1 所示，mAP@0.5:0.95 和 mAP@0.75 指标对比如表 1 所示。

表 1 性能对比

Table 1 Performance comparison

模型	mAP/%	*R*/%	F1/%	mAP@0.5:0.95/%	mAP@0.75/%
YOLOv5	90.09	80.21	84.86	64.8	71.1
本文方法	92.37	84.66	88.35	67.5	73.7

优化算法的 mAP 由原模型的 90.09% 提升到 92.37%，提高了 2.28%。召回率提高了 4.45%，F1 分数提高了 3.49%。

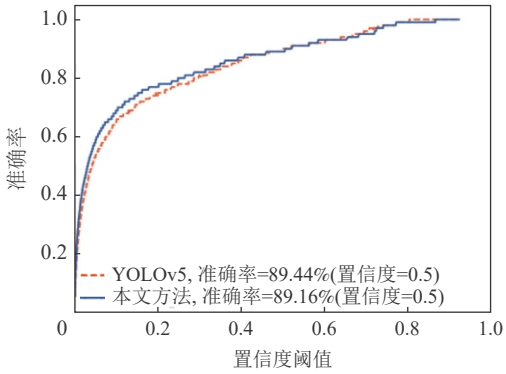
在 mAP@0.5:0.95 和 mAP@0.75 指标上分别提高了 2.7% 和 2.6%。说明优化后的 YOLOv5 位置回归更加精确，具有更好的定位性能，证明了本文方法在低空无人机的障碍物目标检测的有效性。

3）精确率的对比如图 10 所示。

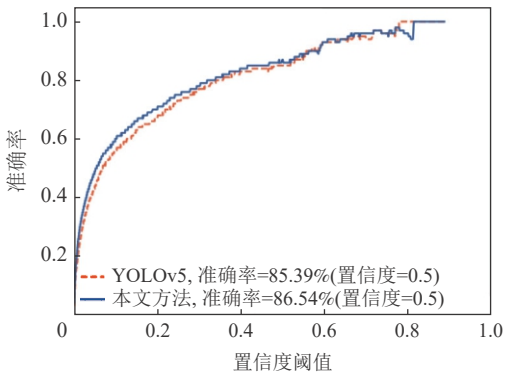
如图 10 所示，当置信度阈值取 0.5 时，本文方法在电力线塔类的精确率达到 89.16%，略低于原模型性能 0.28%，但电线杆类和树木类相较于原模型分别提高了 1.15% 和 0.75%，因此，本文方法的障碍物检出性能更高。

实验 2：本文方法分析及消融实验，主要对比 mAP、mAP@0.5:0.95 及 *R*。如表 2 所示，在低空无人机的障碍物检测中，本文提出的分离尺度约束与融合方向约束的位置损失函数，在反映预测框的位置回归能力的 mAP@0.5:0.95 指标中，相较于采用 CIOU 模型提高了 0.3%，相较于采用 GIOU 模型提高了 0.4%，说明本文提出的损失函数的优化方式即使在严格的交并比要求下，也可提高预测框的位置回归性能。本文提出的双通道注意力机制在原算法上的 mAP 提升 1.32% 并且优于 CBAM 的效果。本文提出的模型相较于原算法 mAP 提升了 2.28%，mAP@0.5:0.95 提升了 2.7%，位置回归能力的提升与双重注意力对特征的强化和对干扰的抑制提高了电线杆类的检测精度，因此，改进后的模型召回率提升了 4.45%。以电线杆类为例，图 11 为不同的改进下电线杆类的 *P*-*R* 曲线，可知在不同阶段电线杆类的 AP 均有不同程度的提升。由于使用残差连接会引入卷积运算，因此模型大小增加了约 18.2 MB。

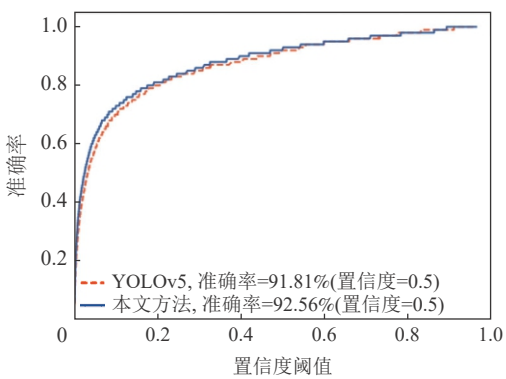
实验 3：为进一步验证本文模型在低空无人机障碍物检测上的性能，对其他主流的目标检测算法在同样的障碍物数据集 LAOD 上进行对比实验。如表 3 所示，本文方法在 mAP 指标上超过 CenterNet、SSD 等算法，与 Faster R-CNN 算法基本持平，但 Faster R-CNN 属于两阶段算法，检测速度过慢，无法达到无人机障碍物的快速检测要求。在 mAP@0.5:0.95 和 mAP@0.75 指标上本文算法也优于其他



(a) 电力线塔类



(b) 电线杆类



(c) 树木类

图 10　精确率对比

Fig. 10　Precision comparison

表 2　消融实验

Table 2　Ablation experiment

模型	mAP/%	mAP@0.5:0.95/%	*R*/%	参数量/MB
YOLOv5(CIOU)	90.09	64.8	80.21	46.149
YOLOv5+GIOU	89.89	64.7	79.54	46.149
YOLOv5+L_{loc}	90.56	65.1	80.55	46.149
YOLOv5+DACB	91.41	66.1	82.73	64.345
YOLOv5+CBAM	91.06	65.6	81.56	46.641
YOLOv5+L_{loc}+DACB(本文)	92.37	67.5	84.66	64.345

算法, 具有良好的位置回归性能, 为融合定位避障等算法打下良好基础。图 12 为不同算法在电线杆类障碍物上的 *P-R* 曲线对比。

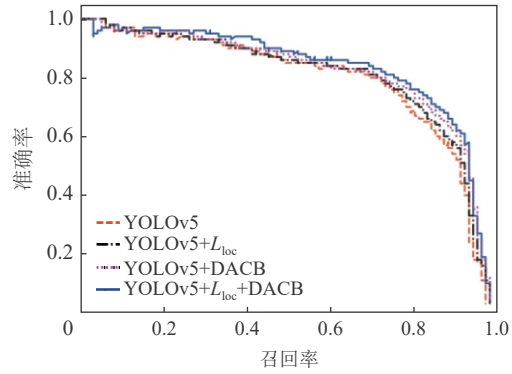


图 11　电线杆类消融实验

Fig. 11　Telegraph pole ablation experiment

表 3　主流目标检测算法性能对比

Table 3　Performance comparison of mainstream target detection algorithms

模型	mAP/%	mAP@0.5:0.95/%	mAP@0.75/%	参数量/MB
CenterNet	78.09	51.0	53.6	32.665
SSD	86.86	60.6	68.2	23.879
Faster R-CNN	92.01	64.2	72.3	136.730
本文方法	92.37	67.5	73.7	64.345

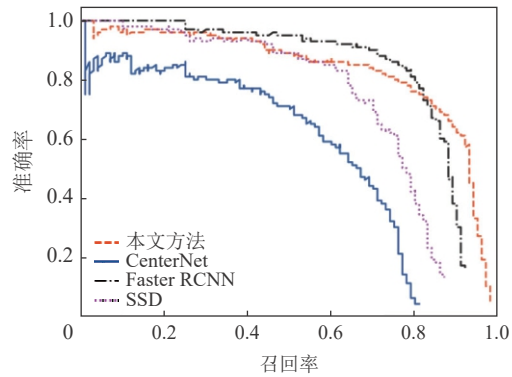


图 12　不同算法的 *P-R* 曲线

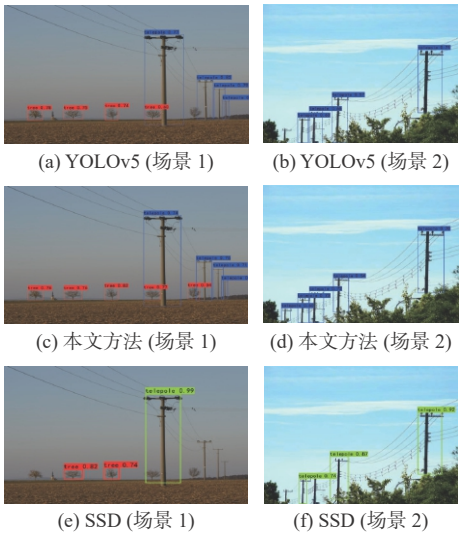
Fig. 12　*P-R* curves of different algorithms

3.4　检测效果对比

分别选取典型场景进行实际检测效果的对比, 如图 13 所示。

在场景 1 中, 本文方法将树木与电线杆障碍物全部检出, 检测精度更高, 位置回归也更加精确, 而 YOLOv5 算法存在漏检情况, 并未将故障全部检出, 且准确率也不如本文方法, 而 SSD 存在严重漏检情况且位置回归不精确。在场景 2 中, 即使存在干扰, 本文方法对电线杆的检测结果也不存在漏检情况, 位置回归准确, 且在遮挡情况下也检测出了电线杆障碍物。检测结果说明了本文方法对漏检情况的改善, 提高了检测精度与召回率, 位置回归更加合理, 为后续融合定位奠定良好基础。

为验证本文方法在实际低空飞行场景的有效性, 在大六轴无人机 M6FC 上搭载检测平台, 机载



(a) YOLOv5 (场景 1) (b) YOLOv5 (场景 2)

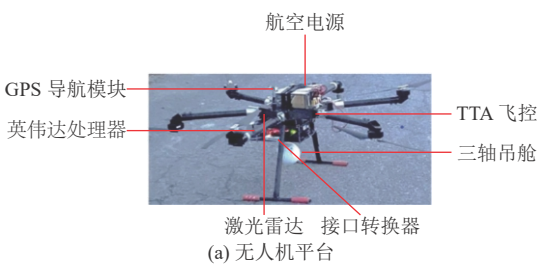
(c) 本文方法 (场景 1) (d) 本文方法 (场景 2)

(e) SSD (场景 1) (f) SSD (场景 2)

图 13 实际检测效果对比

Fig. 13 Comparison of actual detection results

包括 OI-G35SH 吊舱、Livox 激光雷达、英伟达 TX2 处理器等硬件，于北京市昌平区埝头工业园进行实验，无人机实验平台及检测结果如图 14 所示，可以看出本文方法在无人机低空飞行时，即使存在较强的背景干扰，仍然能够对周围的障碍物电线杆与树木检测出来，为低空无人机的安全飞行提供保障。



(a) 无人机平台



(b) 检测结果

图 14 无人机平台和检测结果

Fig. 14 UAV platform and detection result

4 结 论

1）本文针对低空无人机障碍物的尺度分布特点，分析位置回归损失函数的不足并基于此提出一种分离尺度损失与融合方向约束的位置回归损失函数对障碍物的预测框回归过程进行优化。

2）为了减少不同优化维度间的直接影响，本文在混合注意力机制中嵌入残差特征整合模块并改进通道注意力，以构建双重注意力机制 DACB 对特征进行优化，抑制干扰信息，提升检测性能。

3）进行仿真验证，本文方法在 mAP 上提高了 2.28%，在 mAP@0.5:0.95 上提高了 2.7%；搭建无人机平台进行实际飞行场景验证，证明了方法的有效性。

参考文献（References）

[1] 王恒涛, 张上, 陈想, 等. 轻量化无人机航拍目标检测算法[J]. 电子测量技术, 2022, 45(19): 167-174.
WANG H T, ZHANG S, CHEN X, et al. Target detection algorithm of lightweight UAV aerial photography[J]. Electronic Measurement Technology, 2022, 45(19): 167-174(in Chinese).

[2] 戴颖成, 陈知明, 刘峰, 等. 基于无人机红外影像的森林火灾燃烧点检测方法[J]. 中南林业科技大学学报, 2022, 42(9): 102-114.
DAI Y C, CHEN Z M, LIU F, et al. Forest fire detection method based on UAV infrared images and deep learning[J]. Journal of Central South University of Forestry & Technology, 2022, 42(9): 102-114 (in Chinese).

[3] 宋谱怡, 陈红, 苟浩波. 改进 YOLOv5s 的无人机目标检测算法[J]. 计算机工程与应用, 2023, 59(1): 108-116.
SONG P Y, CHEN H, GOU H B. Improving UAV object detection algorithm for YOLOv5s[J]. Computer Engineering and Applications, 2023, 59(1): 108-116(in Chinese).

[4] 奉志强, 谢志军, 包正伟, 等. 基于改进 YOLOv5 的无人机实时密集小目标检测算法[J]. 航空学报, 2023, 44(7): 327106.
FENG Z Q, XIE Z J, BAO Z W, et al. Real-time dense small object detection algorithm for UAV based on improved YOLOv5[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(7): 327106(in Chinese).

[5] 谢华, 苏方正, 尹嘉男, 等. 低空无人机飞行冲突分类探测与差异解脱方法研究[J]. 安全与环境学报, 2023, 23(9): 3131-3142.
XIE H, SU F Z, YIN J N, et al. Research on classified detection and differential resolution method for UAV flight conflicts in low altitude airspace[J]. Journal of Safety and Environment, 2023, 23(9): 3131-3142(in Chinese).

[6] 赵玥萌, 刘会刚. 基于优化 YOLOv4 算法的低空无人机检测与跟踪[J]. 激光与光电子学进展, 2022, 59(12): 397-406.
ZHAO Y M, LIU H G. Detection and tracking of low-altitude unmanned aerial vehicles based on optimized YOLOv4 algorithm[J]. Laser & Optoelectronics Progress, 2022, 59(12): 397-406(in Chinese).

[7] 史雨馨, 朱继杰, 凌志刚. 基于特征增强 YOLOv4 的无人机检测算法研究[J]. 电子测量与仪器学报, 2022, 36(7): 16-23.
SHI Y X, ZHU J J, LING Z G. Research on UAV detection method based on feature enhanced YOLOv4 algorithm[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(7): 16-23 (in Chinese).

[8] GIRSHICKR, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[9] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

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

[11] 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.

[12] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 936-942.

[13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[M]. Cham: Springer, 2016: 412-415.

[14] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 936-944.

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

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

[17] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[18] 李子豪, 王正平, 贺云涛. 基于自适应协同注意力机制的航拍密集小目标检测算法[J]. 航空学报, 2023, 44(13): 327944.
LI Z H, WANG Z P, HE Y T. Aerial-photography dense small target detection algorithm based on adaptive cooperative attention mechanism[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(13): 327944(in Chinese).

[19] 彭继慎, 孙礼鑫, 王凯, 等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法[J]. 仪器仪表学报, 2021, 42(10): 161-170.
PENG J S, SUN L X, WANG K, et al. ED-YOLO power inspection UAV obstacle avoidance target detectionalgorithm based on model compression[J]. Chinese Journal of Scientific Instrument, 2021, 42(10): 161-170(in Chinese).

[20] 张瑞倩, 邵振峰, PORTNOV A, 等. 多尺度空洞卷积的无人机影像目标检测方法[J]. 武汉大学学报 (信息科学版), 2020, 45(6): 895-903.
ZHANG R Q, SHAO Z F, PORTNOV A, et al. Multi-scale dilated convolutional neural network for object detection in UAV images[J]. Geomatics and Information Science of Wuhan University, 2020, 45(6): 895-903(in Chinese).

[21] 王美华, 吴振鑫, 周祖光. 基于注意力改进 CBAM 的农作物病虫害细粒度识别研究[J]. 农业机械学报, 2021, 52(4): 239-247.
WANG M H, WU Z X, ZHOU Z G. Fine-grained identification research of crop pests and diseases based on improved CBAM via attention[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 239-247(in Chinese).

[22] 陈宇梁, 董绍江, 孙世政, 等. 改进 YOLOv5s 的弱光水下生物目标检测算法[J]. 北京航空航天大学学报, 2024, 50(2): 499-507.
CHEN Y L, DONG S J, SUN S Z, et al. Improved YOLOv5s low-light underwater biological target detection algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(2): 499-507(in Chinese).

[23] 毛杨坤, 段现银, 林昕, 等. 基于目标检测的选区激光熔融成形过程熔池与飞溅监测[J]. 机械工程学报, 2023, 59(9): 335-348.
MAO Y K, DUAN X Y, LIN X, et al. Melt pool and spatter monitoring in selective laser melting forming process based on target detection[J]. Journal of Mechanical Engineering, 2023, 59(9): 335-348 (in Chinese).

A low-altitude UAV obstacle detection method based on position constraint and attention mechanism

TANG Youjun[1], MIAO Cunxiao[1,*], ZHANG He[1], LI Yufeng[1], YE Wen[2]

(1. School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China;
2. National Institute of Metrology, China, Beijing 100029, China)

Abstract: Unmanned aerial vehicles (UAV) are widely used in low-altitude areas for power inspection, search and rescue, reconnaissance, and other tasks. The detection of obstacles in advance during flight aims to ensure the completion of established tasks. In order to meet the requirements of obstacle detection accuracy and position regression accuracy of UAV flying at low altitudes, a low-altitude UAV obstacle detection method based on improved position constraints and attention mechanism was proposed. The deficiency of position regression loss function was analyzed. On this basis, the loss function of separation scale loss and fusion direction constraint was proposed to optimize the regression process. The improved attention mechanism CBAM proposed a dual attention mechanism to strengthen the feature suppression interference and improve the detection performance. The experimental results show that the proposed method improves by 2.28% on mAP and 2.7% on mAP@0.5:0.95, showing better detection performance of low-altitude obstacles in terms of both detection accuracy and position regression accuracy.

Keywords: low-altitude UAV; obstacle detection; position regression loss function; dual attention mechanism; position regression loss function

Received: 2023-02-28; **Accepted**: 2023-06-25; **Published Online**: 2023-08-02 10:15
URL: link.cnki.net/urlid/11.2625.V.20230802.0905.001

* **Corresponding author**. E-mail: miaocx@ustb.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0111

基于学习算法的结构大变形预测及气动弹性分析

陈乔，安朝*，谢长川，杨超
(北京航空航天大学 航空科学与工程学院，北京 100191)

摘　　要：大柔性飞行器在气动载荷作用下会产生较大的结构变形，动力学特性发生明显改变，准确的结构变形预测对大柔性飞行器设计及气动弹性仿真具有重要意义。几何非线性有限元等全阶结构模型仿真效率低，已有的非线性结构降阶模型（ROM）具有较高的仿真效率，但建立降阶模型的过程中需要大量样本数据。基于学习算法建立考虑几何非线性因素的大柔性结构静变形预测模型，利用均方根误差（RMSE）对该预测模型进行性能评估，论证几类学习算法在结构大变形预测中的适用性。结合结构大变形预测模型与曲面涡格法（VLM）提出一种新的几何非线性静气动弹性分析方法，兼顾计算精度与效率。采用所提方法计算单梁式机翼静气动弹性变形，对比仿真结果与风洞试验结果，表明所提方法计算精度及效率高，实际应用价值较大。

关　键　词：几何非线性；机器学习；气动弹性；涡格法；结构大变形
中图分类号：V215.3
文献标志码：A　　**文章编号**：1001-5965(2025)03-0943-10

与小展弦比机翼相比，大展弦比机翼具有更好的空气动力学特性，由此获得的升阻比增益能够明显改善高空长航时无人机和大型运输机等大柔性飞行器的飞行性能[1-2]。随着现代飞行器对飞行性能要求的逐渐提高及复合材料使用比例的大幅提升，大展弦比机翼柔性不断增大，在飞行过程中承受气动载荷时变形大，带来显著的几何非线性效应，严重影响飞行器的稳定性、动力学响应特性、气动弹性特性和飞行性能[3-5]。传统的线性结构理论和相关的气动弹性分析方法不适用于大展弦比机翼的变形计算及气动弹性特性分析。同时气动载荷所具有的随动力特性也进一步增加了结构分析难度。结构大变形的准确计算对大柔性飞行器设计具有重要意义[6]。

非线性有限元方法作为全阶模型的代表具有理论发展成熟，计算准确度高，模型适用性高的优势。在实际工程应用中，非线性有限元方法自由度数高，低计算效率限制了其在工程复杂模型中的应用，尤其是应用在气动弹性或优化设计等需要多轮次迭代的分析过程中时，高计算资源消耗是不可接受的。大柔性结构的几何非线性降阶模型逐渐受到国内外研究人员的广泛关注。

几何非线性降阶模型的提出可以追溯到 Shaw 和 Pierre 的开创性工作，其出发点在于利用非线性模态表征大柔性结构的动力学特性，对于复杂有限元模型适用性不佳[7-8]。近年来发展的基于隐式缩聚思路建立的大变形结构降阶模型保留了有限元法适用于大型或复杂结构的优点，在工程分析及气动弹性研究中得到了一定应用[9-11]。该方法用结构广义坐标的多项式形式表征非线性刚度，通过指定载荷计算结构变形，利用回归分析方法求解非线性刚度系数。这套求解技术无须修改现有的非线性有限元代码，复杂模型适用性好。由于采用传统回归分析技术，为达到计算精度要求，降阶模型的构

收稿日期：2023-03-09；**录用日期**：2023-04-27；**网络出版时间**：2023-05-16 10:32
网络出版地址：link.cnki.net/urlid/11.2625.V.20230515.1706.001
基金项目：国家自然科学基金 (12102027)；北京市科学技术协会青年人才托举工程 (BYESS2023345)
* **通信作者**. E-mail：ac@buaa.edu.cn

引用格式：陈乔，安朝，谢长川，等. 基于学习算法的结构大变形预测及气动弹性分析 [J]. 北京航空航天大学学报，2025，51（3）：943-952.
CHEN Q，AN C，XIE C C，et al. Large deformation prediction and geometric nonlinear aeroelastic analysis based on machine learning algorithm [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（3）：943-952 （in Chinese）.

造需要大量“载荷-变形”样本，而计算样本需要大量时间，这一缺陷使得该类方法在实际应用中受到一定限制[12]。Mignolet等[13]在综述文章中详细介绍了该类方法。

机器学习算法可以在物理模型未知的情况下有效获取非线性预测模型，与传统的回归建模方法相比，机器学习算法在样本相对较少的情况下预测能力更好。近年来，包括人工神经网络（artificial neural network, ANN）、支持向量回归（support vector regression, SVR）及随机森林（random forest, RF）等机器学习算法因良好的适用性及泛化能力在航空航天等工程领域特别是在流体力学建模领域得到了广泛应用[14-16]。在考虑几何非线性的大柔性结构动力学建模方法，机器学习应用研究还比较少。虽然相较于流动力学，结构动力学资源耗费相对较小，但在非线性气动弹性及优化分析等需要大量迭代运算的分析中，计算效率同样会成为重要的限制因素。

几何非线性气动弹性研究中的气动力计算方面，二维气动力模型是早期应用比较广泛的高效计算模型，同时其建模也比较方便。Patil等[17]采用几何精确梁理论及片条理论研究了几何非线性因素对于机翼气动弹性特性的影响。Xie等[18]采用片条理论及结构非线性有限元方法分析了复杂机翼模型的非线性气动弹性特性。基于二维气动力模型的分析框架使得研究人员可以了解和认识非线性气动弹性的基本规律。随着数值算法及理论的不断完善，基于势流理论的三维气动力方法被引入非线性气动弹性计算中。涡格法是气动弹性分析中常用的三维气动力方法，Murua等[19]将精确梁理论和非定常涡格法结合，给出了线化模型。Demasi和Livne[20]采用基于壳单元的有限元模型及涡格法解决了联结翼布局飞机的静/动气动弹性问题，经过几十年的发展，涡格法作为一种有效实用的方法在气动弹性分析中得到了广泛应用。

本文将人工神经网络、支持向量回归、随机森林回归及自适应增强回归算法应用于大柔性机翼结构模型的大变形分析计算中。首先，将非线性有限元软件计算结果作为训练数据集，然后，利用上述4种机器学习算法建立了结构静变形的预测模型。通过与非线性有限元计算结果的比较，验证了预测模型的准确性，并比较了不同机器学习算法的性能。最后，结合曲面涡格法和结构预测模型，建立了大柔性机翼静气动弹性的分析框架。阐明了本文结构预测模型在气动弹性应用中的有效性和准确性。

1 理论方法

1.1 结构动力学

在空间坐标系中，运动物体中每个质点的空间位置可用一组坐标来表示。初始时刻质点坐标为$X_i(i=1,2,3)$，质点的坐标随时间变化，任意t时刻质点坐标可表示为$x_i=x_i(X_j,t)(i=1,2,3)$，质点的位移分量可表达为

$$u_i=x_i(X_j,t)-X_i \tag{1}$$

几何非线性问题是典型的“大位移-小应变-小应力”问题，材料未超过弹性极限，基于Green应变张量E_{ij}及第2类Piola-Kirchhoff应力张量S_{ij}度量，结构域Ω_0内平衡方程可写为

$$\frac{\partial}{\partial X_k}(F_{ij}S_{jk})+\rho_0 b_i^0=\rho_0\ddot{u}_i \qquad \boldsymbol{X}\in\Omega_0 \tag{2}$$

$$S_{ij}=C_{ijkl}E_{kl} \tag{3}$$

式中：F_{ij}为变形梯度张量；C_{ijkl}为材料本构张量；ρ_0为结构密度；b_i^0为单元上的体作用力加速度。式(2)使用了Einstein求和约定。将式(2)改写为其等效积分弱形式，并假定位移场分量$u_i(\boldsymbol{X},t)$满足形式：

$$u_i(\boldsymbol{X},t)=\sum_{n=1}^{M}q_n(t)U_i^{(n)}(\boldsymbol{X}) \tag{4}$$

式中：$U_i^{(n)}(\boldsymbol{X})$为基函数向量；$q_n(t)$为待定参数即广义坐标；$M$为基函数截断阶数。利用Galerkin方法求解微分方程近似解，经代数运算后可得结构动力学方程为[12-13]

$$M_{mn}\ddot{q}_n+K_{mn}^{(1)}q_n+K_{mnl}^{(2)}q_nq_l+K_{mnlp}^{(3)}q_nq_lq_p=F_m \tag{5}$$

式中：$m=1,2,\cdots,M$；M_{mn}为广义质量系数；$K_{mn}^{(1)}$为广义线性刚度系数；$K_{mnl}^{(2)}$为对应广义坐标二次多项式的非线性刚度系数；$K_{mnlp}^{(3)}$为对应广义坐标三次多项式的非线性刚度系数。Galerkin方法并不限制基函数向量的具体形式，但结构线性模态具有易计算及易获取的优势，同样满足前述基函数向量的要求，采用结构线性模态作为上述非线性降阶函数合理有效。设质量归一化的结构线性模态集为$\boldsymbol{\Phi}=\{\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_m\}$，物理坐标与广义模态坐标之间的转化关系为

$$\boldsymbol{x}(t)=\boldsymbol{\Phi}\boldsymbol{q}(t) \tag{6}$$

广义空间非线性结构动力学方程可重写为

$$\ddot{q}_m+\bar{K}_mq_m+K_{mnl}^{(2)}q_nq_l+K_{mnlp}^{(3)}q_nq_lq_p=\bar{F}_m \tag{7}$$

式中：$\bar{K}_m$为结构线性模态的广义刚度项；$\bar{F}_m$为模态广义力。忽略时变项，可得

$$\bar{K}_mq_m+K_{mnl}^{(2)}q_nq_l+K_{mnlp}^{(3)}q_nq_lq_p=\bar{F}_m \tag{8}$$

在经典的几何非线性结构降阶模型构建中, 确定非线性刚度系数$K_{mnl}^{(2)}$、$K_{mnlp}^{(3)}$是核心工作。利用非线性有限元法求解结构在不同载荷工况作用下的位移响应得到“载荷-变形”样本, 结合最小二乘回归法可以求解未知的非线性刚度系数。将非线性刚度系数解出后, 可以通过式 (6) 和式 (7) 预测结构在特定载荷作用下的位移响应。

1.2　机器学习算法

传统的几何非线性结构降阶模型受限于指定的多项式形式及最小二乘回归方法往往需要较多的样本才能取得满意的精度, 为得到性能更好的结构大变形预测模型, 本文采用机器学习算法对式 (7) 中的非线性环节进行建模。给定数据集公式为

$$G=\{(\boldsymbol{x}_1,\boldsymbol{y}_1),(\boldsymbol{x}_2,\boldsymbol{y}_2),\cdots,(\boldsymbol{x}_S,\boldsymbol{y}_S)\} \tag{9}$$

式中: $\boldsymbol{x}_i=(x_{i1},x_{i2},\cdots,x_{id})^{\mathrm{T}}\in\mathbf{R}^{\mathbf{d}}$为特征空间中维数为$d$的第$i$个训练样本的输入参数向量; x_{ij}为第i个样本中的第j个特征参数即输入参数; $\boldsymbol{y}_i\in\mathbf{R}$为样本的输出参数向量; S为样本数。

1.2.1　支持向量回归

支持向量回归是一种高效的回归算法, 在非线性函数建模当中应用广泛。支持向量回归算法源于支持向量机(support vector machine, SVM), 由Cortes 和 Vapnik[21] 首先引入系统机器学习及模式识别领域, 并且发展迅速。

在非线性支持向量回归算法中, 将特征空间中的变量和样本响应值作为训练的数据, 希望获得对于每个训练点$\boldsymbol{x}$, 偏离目标相应的值都不大于ε的目标函数$f(\boldsymbol{x})$。给定非线性核函数$K(\boldsymbol{x},\boldsymbol{x}')=\langle\Phi(\boldsymbol{x}),\Phi(\boldsymbol{x}')\rangle$, 其中, Φ是从样本空间到特征参数空间的映射关系函数, 非线性支持向量回归的解可以描述为

$$f(\boldsymbol{x})=b+\sum_{i}^{S}(\alpha_i-\bar{\alpha}_i)K(\boldsymbol{x},\boldsymbol{x}_i) \tag{10}$$

式中: b为距离参数; α_i和$\bar{\alpha}_i$为引入的拉格朗日乘子。

支持向量回归算法原理如图 1 所示。其中, ω 为支持向量回归算法中的超平面参数; ξ_i,$\bar{\xi}_i$为支持向量回归算法中的距离松弛变量。

1.2.2　随机森林回归

随机森林回归算法通过引入回归树进行预测, 是一种基于决策树算法的集成学习方法。随机森林回归算法在构造每个回归树时, 利用训练样本的输入特征值随机序列构成回归树的分支, 使用回归树输出的平均值计算样本的目标响应值估计点。算法关键参数是回归树的数量。较多的数据能够增加计算精度, 但降低了计算效率。假设预测模型包含R个回归树, 则预测模型的最终输出为

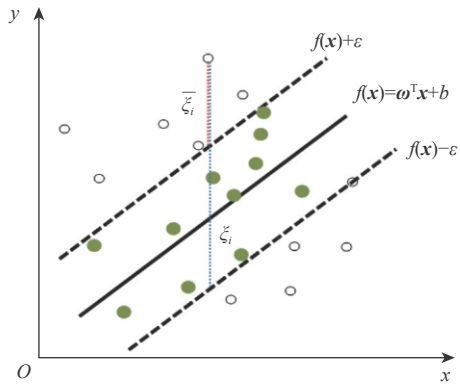


图 1　支持向量回归原理

Fig. 1　Schematic diagram of SVR

$$H(x)=\frac{1}{R}\sum_{i=1}^{R}h_i(x) \tag{11}$$

式中: h_i为样本中第i个回归树的输出。

1.2.3　自适应增强回归

自适应增强算法(AdaBoost)是由 Freund 和 Schapire[22] 提出的最早的实用增强算法之一。通过迭代训练对弱训练器进行组合, 提高集合后的整体预测精度, 构建更好的预测模型。自适应增强算法基于权向量从可用的样本数据中创建训练集, 建立输入与输出之间的假设关系, 将该假设关系应用于所有样本数据后, 通过迭代构造优化预测模型。每次算法迭代时, 前一轮生成预测器中预测误差较大的样本的权值将提高, 误差较小的样本的权值将降低, 直至预定整体误差及最大迭代次数L满足要求。

1.2.4　人工神经网络

人工神经网络是受到生物神经系统启发所构建的机器学习算法, 具有强大的非线性拟合能力, 能够通过现有样本数据的训练识别出数据中隐含的非线性映射关系[23]。人工神经网络由大量人工神经元经广泛互联组成, 该系统包括输入层、一个或多个中间隐含层和输出层, 用来模拟脑神经系统的结构和功能。人工神经网络可以看成以人工神经元为节点, 用有向加权弧连接起来的有向图。人工神经网络的结构如图 2 所示。

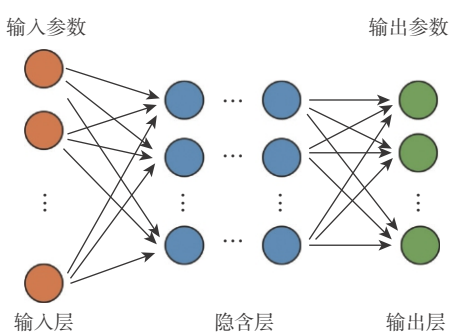


图 2　神经网络结构

Fig. 2　Structure of artificial neural networks

1.3 结构大变形预测

基于上述各类机器学习算法，构建考虑几何非线性因素的机翼结构大变形预测模型。将模态空间中的广义力定义为输入参数，将结构大变形在模态空间中的广义坐标定义为输出参数。预测数据集为

$$G^r=\{(\tilde{\boldsymbol{f}}_1,q_{r,1}),(\tilde{\boldsymbol{f}}_2,q_{r,2}),\cdots,(\tilde{\boldsymbol{f}}_S,q_{r,S})\}\qquad r=1,2,\cdots,m \tag{12}$$

式中：m为模态空间阶数；S为数据样本数。对于m阶模态空间，需要的预测模型个数也是m。

样本数据质量对于机器学习算法预测准确度具有重要影响。本方法样本数据包括给定的外力及相应的结构变形，这与传统的非线性结构降阶模型一致[12-13]。在选择样本中的给定外力时，有2个因素十分重要：气动力的空间分布特征；期望的结构响应幅值。本文将结构模态振型的加权叠加作为数据样本中的给定外力：

$$\tilde{\boldsymbol{f}}=\sum_{r=1}^{m}a_r\lambda_r\boldsymbol{\varphi}_r \tag{13}$$

$$\lambda_r=\frac{2i_r}{g_r-1}-1\qquad i_r=0,1,\cdots,g_r-1 \tag{14}$$

式中：a_r为每个所选模态振型的加权系数；g_r为每个所选模态振型的样本网格密度，较大的样本网格密度可以获得较宽的采样空间范围。给定外力后，结构变形基于非线性有限元软件进行求解。需要注意的是，在基于有限元软件获取样本数据的过程中，给定的外力均作为随动力，与真实的气动力特性相似，提高了在气动弹性分析中的适用性。

将样本数据集分为训练数据集和测试数据集。训练数据集用于结构大变形预测模型建立，试验数据集用于预测模型的性能评估。结构大变形预测模型的建立过程如图3所示。建立并验证结构大变形预测模型后，在给定任意力输入的情况下，预测模型可以快速计算广义坐标下的机翼结构大变形，基于式(2)中给出的广义坐标与物理坐标关系即可以恢复其物理空间位移。

1.4 静气动弹性分析

静气动弹性问题主要针对大柔性机翼结构变形与气动载荷耦合作用开展分析研究。几何非线性气动弹性分析中，结构变形导致气动载荷与结构动力学特性发生变化，几何非线性因素不能忽略。在得到结构大变形预测模型后，使用曲面涡格法计算气动力载荷，基于三维曲面样条插值方法传递气动与结构界面信息[24]，采用松耦合迭代思路计算静气动弹性变形。

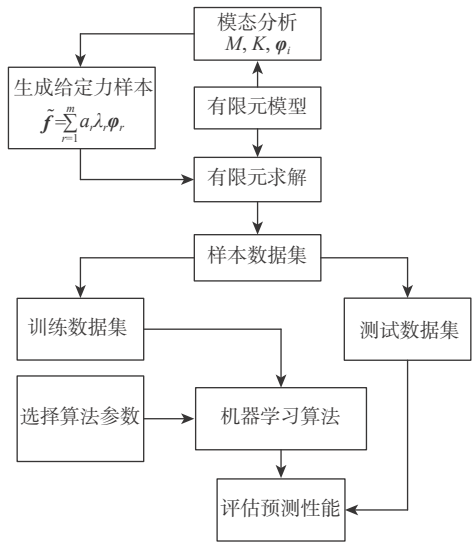


图3 结构大变形预测模型建立流程

Fig. 3 Procedure of large deformation prediction model establishment

1.4.1 曲面涡格法

曲面涡格法是计算机翼结构大变形气动载荷的有效方法。机翼气动面可以划分为包含涡环的面元。涡环由涡线组成，前段涡线位于面元1/4弦线上，点设为气动力载荷作用点。控制点位于面元3/4弦线中点，边界条件需在控制点处满足。利用毕奥-萨伐尔定理可以计算任意位置的诱导速度。

机翼上附着涡涡量的分布由几何精确边界条件确定：

$$(\nabla\Psi+\boldsymbol{v})\cdot\boldsymbol{n}=0 \tag{15}$$

式中：Ψ为势函数；$\boldsymbol{v}$为边界运动速度向量；$\boldsymbol{n}$为当地法向量。涡环沿来流方向在后缘脱落，由尾涡面模拟，附着涡及尾涡划分模型如图4所示。在计算气动载荷时，一般尾涡数量给定为10倍弦向附着涡网格数即可满足精度要求。

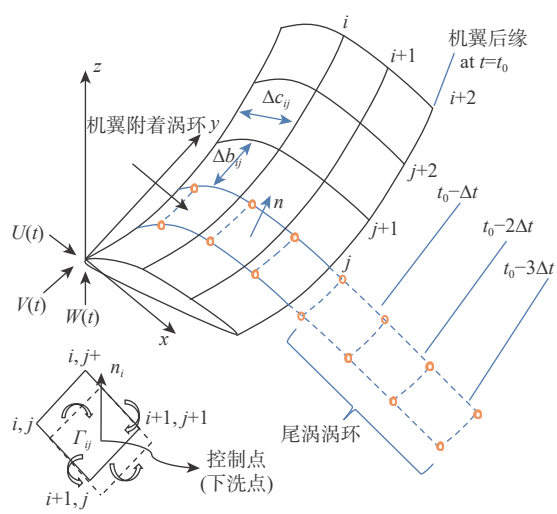


图4 曲面涡格法模型

Fig. 4 Model of unsteady vortex lattice method

1.4.2　曲面样条插值

曲面样条插值方法适用于考虑结构大变形因素的气动与结构界面信息交换。变形后的结构通常是三维的, 而未变形结构可以是一维、二维或三维结构。当确定n个结构网格及相应变形向量$\boldsymbol{U}_S$后, m个气动网格的变形向量可表达为

$$\boldsymbol{U}_A = \boldsymbol{G}\boldsymbol{U}_S \tag{16}$$

式中: $\boldsymbol{G}$为插值矩阵; 下标A为气动相关变量; 下标S为结构相关变量。结合虚功原理可以得到气动力载荷向量$\boldsymbol{F}_A$与结构外力向量$\boldsymbol{F}_S$的插值关系为

$$\delta\boldsymbol{U}_A^{\mathrm{T}}\boldsymbol{F}_A = \delta\boldsymbol{U}_S^{\mathrm{T}}\boldsymbol{F}_S \tag{17}$$

式中: $\delta\boldsymbol{U}_A$和$\delta\boldsymbol{U}_S$为任意虚位移矢量。

1.4.3　静气动弹性迭代计算

结合曲面涡格法, 曲面样条插值及基于学习算法的结构大变形预测模型, 本文给出可以在任意给定工况条件下的大柔性机翼几何非线性静气动弹性计算流程。计算采用松耦合迭代思想, 迭代过程如图 5 所示。

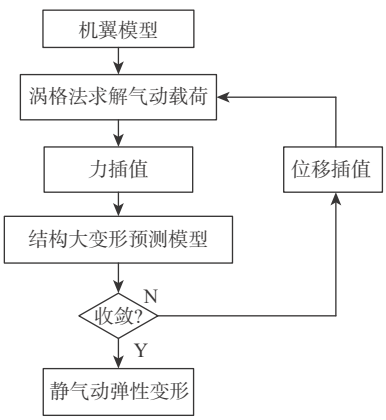


图 5　静气动弹性分析流程

Fig. 5　Iterative process of static aeroelastic analysis

2　结构模型及数据

本文给出某大柔性机翼模型, 计算几何非线性静气动弹性变形, 结合风洞试验数据验证本文预测模型的准确性和有效性。

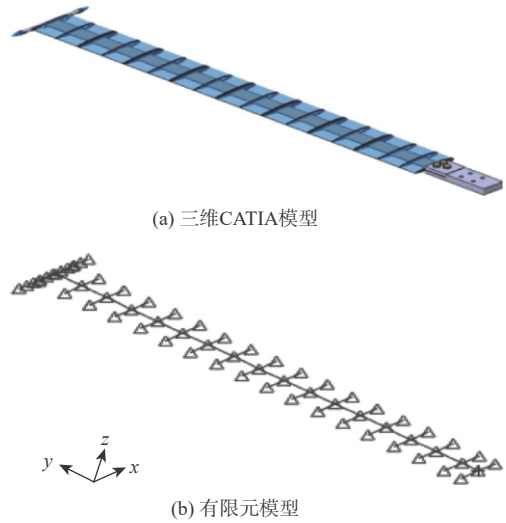
2.1　机翼模型

大柔性机翼模型设计参数如表 1 所示。有限元模型、CATIA 模型及实际模型如图 6 和图 7 所示。该模型主梁采用矩形截面直梁。矩形截面尺寸为 35 mm×1.5 mm。边界条件为翼根固支。

表 1　大柔性机翼模型设计参数

Table 1　Design parameters of the wing model

翼展/mm	质量/kg	翼根弦长/mm	翼尖弦长/mm	翼型	展弦比	弹性轴位置	弹性模量/GPa
1 000.0	0.613	100.0	100.0	NACA0015	10.0	50%机翼弦长	219



(a) 三维CATIA模型

(b) 有限元模型

图 6　大柔性机翼模型

Fig. 6　Model of large flexible wing

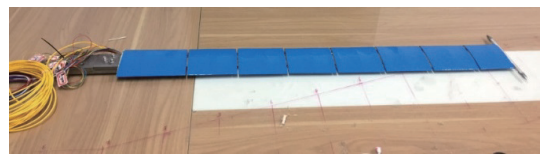


图 7　实际机翼模型

Fig. 7　Real model of large flexible wing

表 2 为大柔性机翼模型前五阶线性模态有限元分析结果及地面试验结果, 有限元分析结果与地面试验结果吻合较好。机翼第 1 阶结构模态频率约为 1.18 Hz, 结构柔性大, 具有明显的几何非线性特征。

表 2　前五阶机翼模态

Table 2　The first 5 modes of wing

模态	类型	有限元分析结果/Hz	地面试验结果/Hz
1	垂直一弯	1.179	1.18
2	垂直二弯	7.724	7.81
3	垂直三弯	22.19	22.35
4	一阶扭转	22.95	23.07
5	水平一弯	27.47	27.63

2.2　学习算法样本数据

使用式(12)和式(13)构建样本载荷后, 在有限元软件 MSC.Nastran 中将样本载荷作为随动力加载于大柔性机翼模型, 计算结构变形并转换至模态空间给出广义坐标。本文选择垂直一弯模态$\boldsymbol{\varphi}_1$和一阶扭转模态$\boldsymbol{\varphi}_2$生成样本载荷, 网格密度为 5, 样本数据集总数为 168。结构变形预测模型包含除垂直三弯模态外的前四阶模态, 选择相应广义力$\tilde{\boldsymbol{f}}_1$、$\tilde{\boldsymbol{f}}_2$、$\tilde{\boldsymbol{f}}_3$、$\tilde{\boldsymbol{f}}_4$作为学习算法的输入参数, 将广义模态坐标q_1、q_2、q_3、q_4作为输出参数。选取合适的加权因子a_1和a_2, 样本数据集翼尖垂向变形分布如图 8 所示, 最大变形量约为 220mm, 达到翼展长度的

22%。样本集中广义模态坐标q_1、q_2、q_3、q_4分布如图 9～图 12 所示。广义模态坐标的符号与模态振型具体的方向有关。

2.3 风洞试验

针对几何非线性静气动弹性问题开展风洞试验研究。将本文算例中的大柔性机翼模型根部固定于自身刚性足够大的支撑系统，同时模型迎角可

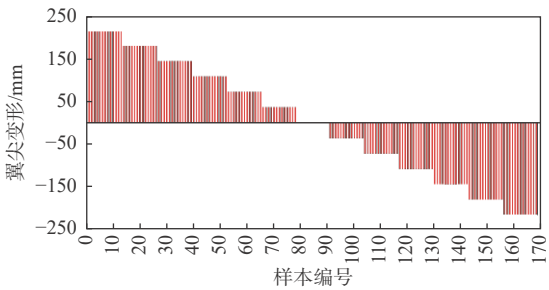


图 8 样本集中的翼尖垂向变形分布

Fig. 8 Distribution of wing tip deformations with samples

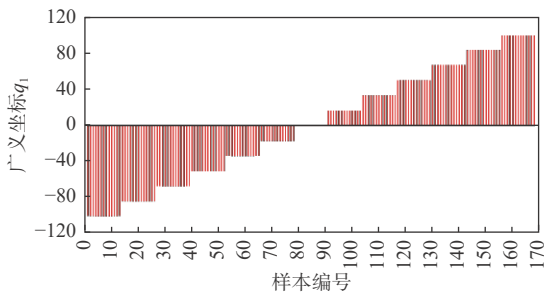


图 9 样本集中的第 1 阶模态广义坐标分布

Fig. 9 Distribution of 1st generalized coordinate with samples

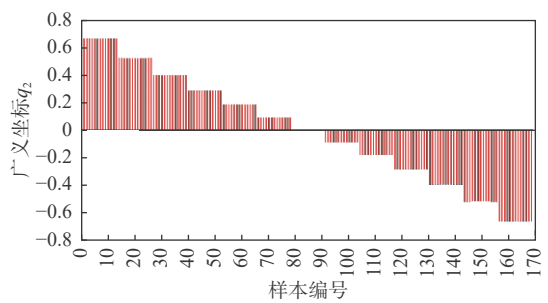


图 10 样本集中的第 2 阶模态广义坐标分布

Fig. 10 Distribution of 2nd generalized coordinate with samples

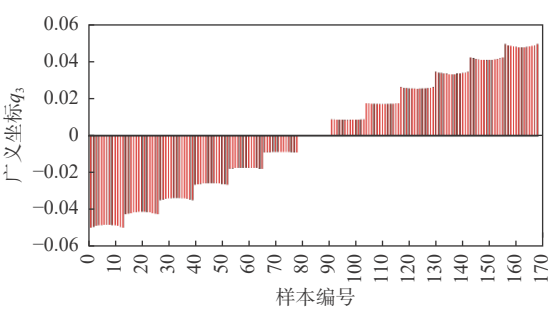


图 11 样本集中的第 3 阶模态广义坐标分布

Fig. 11 Distribution of 3rd generalized coordinate with samples

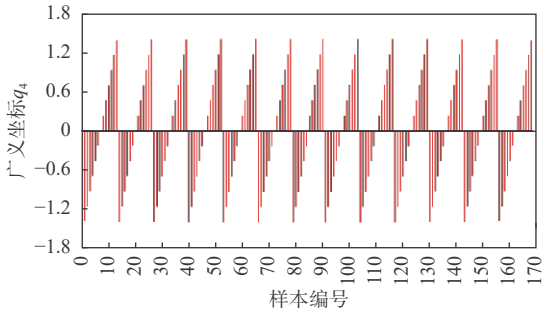


图 12 样本集中的第 4 阶模态广义坐标分布

Fig. 12 Distribution of 4th generalized coordinate with samples

以从 0°变化至 5°。针对大柔性结构变形测量难题，将光纤光栅传感器安装在沿机翼展向不同位置以测量应变值的分布，测量得到应变值后通过数值积分及应变位移恢复技术计算结构大变形。固定在风洞中的真实机翼模型如图 13 所示。

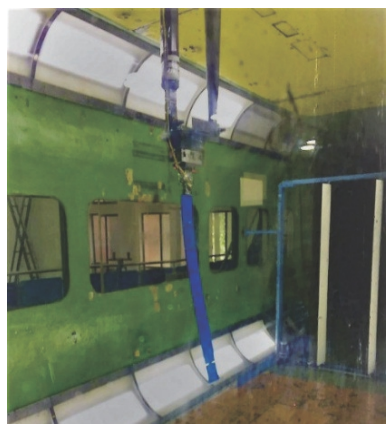


图 13 风洞试验中的机翼模型

Fig. 13 Wing model in wind tunnel test

3 分析结果

利用本文引入的 4 种机器学习算法建立机翼模型大变形预测模型，将 75% 的样本数据集作为训练数据集以构建预测模型（126 个样本），其余 25% 作为测试数据集用于测试预测模型性能（42 个样本）。使用均方根误差（root mean square error, RSME）作为性能评估指标：

$$R_M = \sqrt{\frac{1}{S}\sum_{i=1}^{S}(x_i - y_i)^2} \tag{18}$$

式中：x_i和y_i分别为样本值和预测值；S为样本总数。

3.1 结果大变形预测结果

本文选择经典的网格搜索技术来确定机器学习算法中的参数，在确定参数的同时也考虑了计算效率的影响。在支持向量回归算法中，选择高斯核函数 $K(\boldsymbol{x},\boldsymbol{x}_i)=\exp(-\gamma\|\boldsymbol{x}-\boldsymbol{x}_i\|^2)$，参数 γ 为 1/16；随机森林回归算法中，回归树的数量R设置为 70；自适应增强算法中，迭代的次数设置为 100；人工神经网

络算法中, 隐含层的个数选择为 10。图 14～图 17 给出了测试集广义坐标预测结果。黑色实线表示测试集样本输出参数值, 符号表示机器学习算法预测结果。为方便起见, 测试集样本编号的顺序按照预测值由小到大重新排列。

广义坐标预测结果表明, 自适应增强算法和人工神经网络算法的总体预测性能优于支持向量回归算法和随机森林算法。其中, 自适应增强算法在精度上整体表现更好。从表 3 的均方根误差计算结果来看, 自适应增强算法和人工神经网络算法的

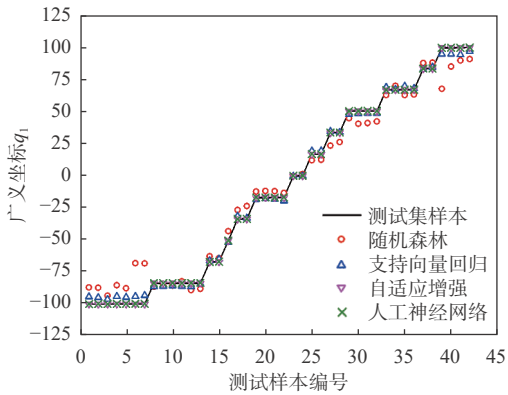


图 14　第 1 阶广义坐标预测结果

Fig. 14　Prediction results of 1st generalized coordinate

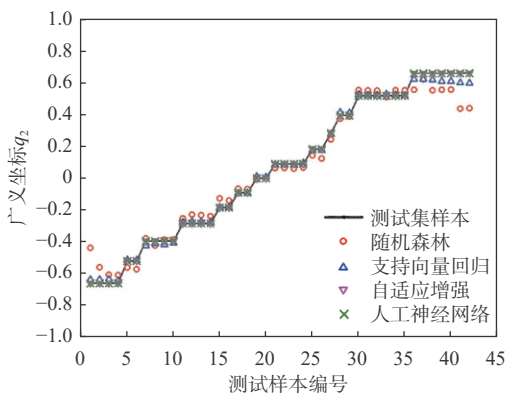


图 15　第 2 阶广义坐标预测结果

Fig. 15　Prediction results of 2nd generalized coordinate

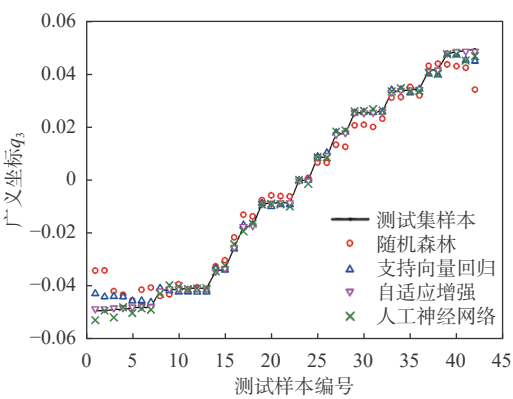


图 16　第 3 阶广义坐标预测结果

Fig. 16　Prediction results of 3rd generalized coordinate

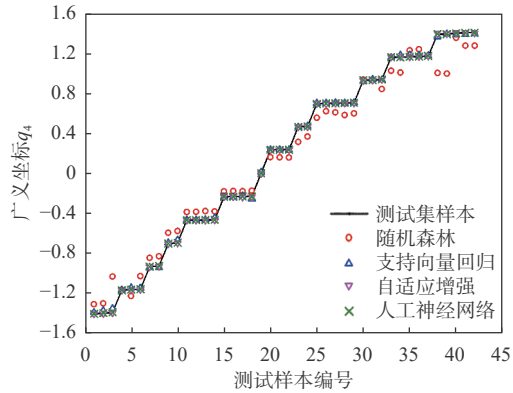


图 17　第 4 阶广义坐标预测结果

Fig. 17　Prediction results of 4th generalized coordinate

预测值与实际值吻合较好。结构大变形广义坐标与物理坐标之间存在直接转换关系, 因此, 广义坐标预测精度能够反映机翼实际结构变形的预测精度。

训练数据集样本的数量对机器学习算法的性能有重要影响。修改训练数据集样本数量占全部数据集样本数量的百分比, 研究结构大变形预测模型性能的变化。表 4～表 7 为不同训练数据集样本

表 3　机器学习算法预测结果

Table 3　Prediction results of machine learning algorithms

算法	均方根误差			
	第1阶模态	第2阶模态	第3阶模态	第4阶模态
支持向量回归	0.799 0	0.008 9	5.65×10^{-4}	0.013 5
随机森林	3.106 1	0.023 2	0.002 1	0.018 7
自适应增强	0.039 5	1.67×10^{-4}	2.97×10^{-4}	0.002 5
人工神经网络	0.090 3	7.57×10^{-4}	0.001 3	0.001 3

表 4　第 1 阶模态广义坐标不同训练样本数量下预测结果

Table 4　Prediction results of 1st generalised coordinate under different percent of training samples

算法	均方根误差				
	10%	25%	50%	75%	90%
支持向量回归	5.994 7	2.726 2	1.234 2	0.799 0	0.182 2
随机森林	18.767 8	11.446 9	3.724 8	3.106 1	1.831 5
自适应增强	15.99	13.775 4	0.045 9	0.039 5	0.022
人工神经网络	16.03	0.029 1	1×10^{-3}	0.090 3	5.69×10^{-4}

表 5　第 2 阶模态广义坐标不同训练样本数量下预测结果

Table 5　Prediction results of 2nd generalised coordinate under different percent of training samples

算法	均方根误差				
	10%	25%	50%	75%	90%
支持向量回归	0.065 9	0.021 5	0.01	0.008 9	0.002 8
随机森林	0.140 7	0.081 7	0.024 6	0.023 2	0.012 3
自适应增强	0.143 5	0.076 2	0.026 7	1.65×10^{-4}	9.293×10^{-5}
人工神经网络	0.008	7.288×10^{-4}	7.34×10^{-4}	7.57×10^{-4}	5.08×10^{-4}

表 6 第 3 阶模态广义坐标不同训练样本数量下预测结果

Table 6 Prediction results of 3rd generalised coordinate under different percent of training samples

算法	均方根误差				
	10%	25%	50%	75%	90%
支持向量回归	0.002 7	5.15×10^{-4}	5.95×10^{-4}	5.65×10^{-4}	1.49×10^{-4}
随机森林	0.008 9	0.005 8	0.005 1	0.002 1	6.06×10^{-4}
自适应增强	0.007 7	0.005 8	2.95×10^{-4}	2.97×10^{-4}	1.55×10^{-4}
人工神经网络	0.003 6	0.003	0.002 5	0.001 3	4.714×10^{-4}

表 7 第 4 阶模态广义坐标不同训练样本数量下预测结果

Table 7 Prediction results of 4th generalised coordinate under different percent of training samples

算法	均方根误差				
	10%	25%	50%	75%	90%
支持向量回归	0.063 2	0.056 8	0.031 4	0.013 5	0.005 6
随机森林	0.354 9	0.208 9	0.051 2	0.018 7	0.008 1
自适应增强	0.375 3	0.108 7	0.041 7	0.002 5	0.001 7
人工神经网络	0.080 4	0.013 5	0.001 6	0.001 3	8.37×10^{-4}

百分比下预测模型的预测结果。图 18 为不同训练数据集样本百分比下第 1 阶模态广义坐标的预测结果，其他坐标变化趋势相似。可以看出，当训练数据集样本的百分比较低，如 10% 时，支持向量回归算法的性能优于其他 3 种算法。随着训练数据样本的增加，人工神经网络和自适应增强算法的性能迅速提高，并且人工神经网络算法需要更少的训练集样本数量来保持相同的精度。当训练数据集样本足够多时，自适应增强算法的性能最好。与其他 3 种算法相比，随机森林算法在所有训练数据集样本百分比范围内的性能都较差。

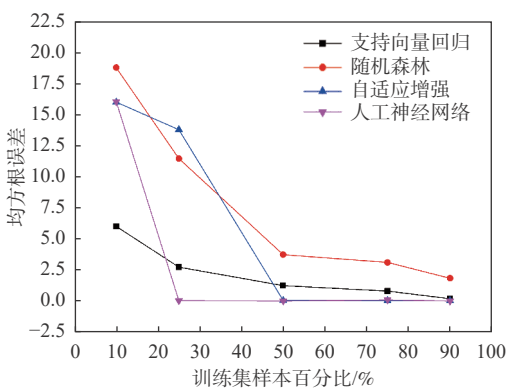


图 18 不同训练样本数量下第 1 阶广义坐标预测结果

Fig. 18 Prediction results of 1st generalised coordinate under different percent of training samples

3.2 静气动弹性分析结果

采用支持向量回归和人工神经网络算法构建静态气动弹性变形预测模型，将 75% 的样本数据集作为训练数据集样本。基于本文前述的静气动弹性迭代分析过程，建立气动模型如图 19 所示。在给定迎角（angle of attack, AOA）和来流风速条件下计算大柔性机翼模型几何非线性静气动弹性变形。同时，给出采用 MSC.Nastran 软件计算得到的线性分析结果及风洞试验结果。图 20 和图 21 分别为风速 10～18 m/s 时，3°迎角和 4°迎角下机翼模型的静气动弹性变形。预测模型的求解结果与风洞试验结果吻合较好。线性方法的分析结果与非线性预测和试验结果有明显差异，可以发现线性分析结果大于非线性预测结果及风洞试验结果，且差异随风速增大而不断增大。考虑几何非线性因素

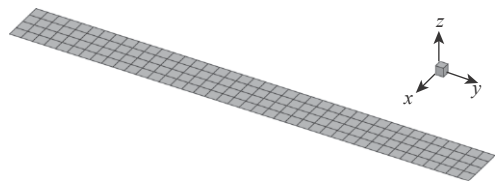


图 19 气动面模型

Fig. 19 Aerodynamics model

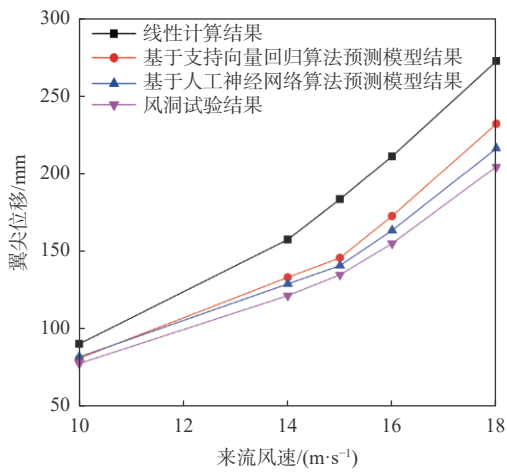


图 20 3°迎角下静气动弹性分析结果

Fig. 20 Static aeroelastic analysis results under 3°AOA

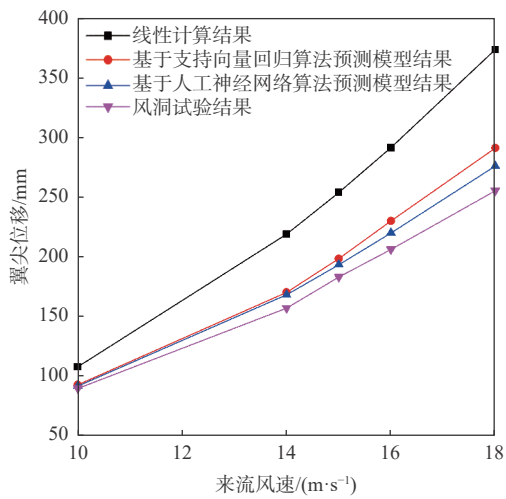


图 21 4°迎角下静气动弹性分析结果

Fig. 21 Static aeroelastic analysis results under 4° AOA

对大柔性机翼气动弹性分析具有重要意义。图 22 为 3°迎角及 18 m/s 速度下的气动压强分布结果。该分析方法对于考虑非线性气动弹性因素的气动载荷计算同样适用。

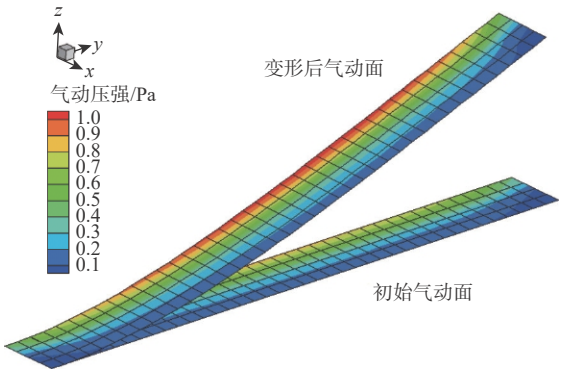


图 22　气动压力分布

Fig. 22　Aerodynamics pressure distribution

4　结　论

1) 本文利用支持向量回归、随机森林、自适应增强和人工神经网络 4 种机器学习算法建立大柔性机翼结构几何非线性静变形预测模型。以模态空间广义力为输入参数, 以模态空间变形广义坐标为输出参数。基于非线性有限元软件获得数据样本。

2) 本文利用均方根误差评估不同的模型预测性能。算法预测结果表明, 人工神经网络和自适应增强算法总体性能优于随机森林和支持向量回归算法。如果训练样本数量较少, 推荐使用支持向量回归算法。如果训练样本数量足够多, 则推荐使用自适应增强算法。综合考虑不同的训练样本数量, 人工神经网络算法的性能更加稳定。不推荐使用随机森林算法。

3) 本文建立了基于结构大变形预测模型和曲面涡格法的大柔性机翼静气动弹性分析方法。对比风洞试验结果和线性分析方法结果, 本文方法在计算几何非线性静气动弹性变形时准确有效。

参考文献（References）

[1] PATIL M J, HODGES D H, CESNIK C E S. Nonlinear aeroelasticity and flight dynamics of high-altitude long-endurance aircraft[J]. Journal of Aircraft, 2001, 38(1): 88-94.

[2] EATON A J, HOWCROFT C, COETZEE E B, et al. Numerical continuation of limit cycle oscillations and bifurcations in high-aspect-ratio wings[J]. Aerospace, 2018, 5(3): 78.

[3] 刘燚. 大柔性飞机气动弹性稳定性及动响应分析[D]. 北京: 北京航空航天大学, 2016.
LIU Y. Aeroelastic stability and response analysis of very fleible aircraft[D]. Beijing: Beihang University, 2016(in Chinese).

[4] TANG D M, DOWELL E H. Experimental and theoretical study on aeroelastic response of high-aspect-ratio wings[J]. AIAA Journal, 2001, 39(8): 1430-1441.

[5] FRULLA G, CESTINO E, MARZOCCA P. Critical behaviour of slender wing configurations[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2010, 224(5): 587-600.

[6] AN C, YANG C, XIE C C, et al. Gust load alleviation including geometric nonlinearities based on dynamic linearization of structural ROM[J]. International Journal of Aerospace Engineering, 2019: 3207912.

[7] SHEN Y C, VIZZACCARO A, KESMIA N, et al. Comparison of reduction methods for finite element geometrically nonlinear beam structures[J]. Vibration, 2021, 4(1): 175-204.

[8] 黄行蓉, 刘久周, 李琳. 基于非线性模态的复杂系统动力学特性分析方法[J]. 北京航空航天大学学报, 2019, 45(7): 1337-1348.
HUANG X R, LIU J Z, LI L. Dynamic characteristics analysis method of complex systems based on nonlinear mode[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(7): 1337-1348 (in Chinese).

[9] MCEWAN M I, WRIGHT J R, COOPER J E, et al. A combined modal/finite element analysis technique for the dynamic response of a nonlinear beam to harmonic excitation[J]. Journal of Sound and Vibration, 2001, 243(4): 601-624.

[10] HOLLKAMP J J, GORDON R W, SPOTTSWOOD S M. Nonlinear modal models for sonic fatigue response prediction: A comparison of methods[J]. Journal of Sound and Vibration, 2005, 284(3/4/5): 1145-1163.

[11] AN C, YANG C, XIE C C, et al. Flutter and gust response analysis of a wing model including geometric nonlinearities based on a modified structural ROM[J]. Chinese Journal of Aeronautics, 2020, 33(1): 48-63.

[12] 安朝. 大柔性飞行器大柔性飞行器结构建模方法及气动弹性分析研究[D]. 北京: 北京航空航天大学, 2020.
AN C. Structure modeling methodology and aeroelastic analysis of large flexible aircraft[D]. Beijing: Beihang University, 2020(in Chinese).

[13] MIGNOLET M P, PRZEKOP A, RIZZI S A, et al. A review of indirect/non-intrusive reduced order modeling of nonlinear geometric structures[J]. Journal of Sound and Vibration, 2013, 332(10): 2437-2460.

[14] YANG X Y, PANG S, SHEN W, et al. Aero engine fault diagnosis using an optimized extreme learning machine[J]. International Journal of Aerospace Engineering, 2016: 7892875.

[15] HUANG M S, LEI Y Z, LI X F, et al. Damage identification of bridge structures considering temperature variations-based SVM and MFO[J]. Journal of Aerospace Engineering, 2021, 34(2), 04020113-1: 18.

[16] 尹明朗, 寇家庆, 张伟伟. 一种高泛化能力的神经网络气动力降阶模型[J]. 空气动力学学报, 2017, 35(2): 205-213.
YIN M L, KOU J Q, ZHANG W W. A reduced-order aerodynamic model with high generalization capability based on neural network[J]. Acta Aerodynamica Sinica, 2017, 35(2): 205-213(in Chinese).

[17] PATIL M, HODGES D, CESNIK C. Limit cycle oscillations in high-aspect-ratio wings[J]. Journal of Fluids and Structures, 2002, 15(1): 107-132.

[18] XIE C C, LENG J Z, YANG C. Geometrical nonlinear aeroelastic stability analysis of a composite high-aspect-ratio wing[J]. Shock and Vibration, 2008, 15(3/4): 325-333.

[19] MURUA J, PALACIOS R, GRAHAM J M R. Applications of the unsteady vortex-lattice method in aircraft aeroelasticity and flight dynamics[J]. Progress in Aerospace Sciences, 2012, 55: 46-72.

[20] DEMASI L, LIVNE E. Dynamic aeroelasticity of structurally nonlinear configurations using linear modally reduced aerodynamic generalized forces[J]. AIAA Journal, 2009, 47(1): 70-90.

[21] CORTES C, VAPNIK V. Support vector networks[J]. Machine Learning, 1995, 20(3): 273-297.

[22] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139.

[23] 王年华, 鲁鹏, 常兴华, 等. 基于人工神经网络的非结构网格尺度控制方法[J]. 力学学报, 2021, 53(10): 2682-2691.
WANG N H, LU P, CHANG X H, et al. Unstructured mesh size control method based on artificial neural network[J]. Chinese Journal of Theoretical and Applied Mechanics, 2021, 53(10): 2682-2691 (in Chinese).

[24] XIE C C, AN C, LIU Y, et al. Static aeroelastic analysis including geometric nonlinearities based on reduced order model[J]. Chinese Journal of Aeronautics, 2017, 30(2): 638-650.

Large deformation prediction and geometric nonlinear aeroelastic analysis based on machine learning algorithm

CHEN Qiao, AN Chao*, XIE Changchuan, YANG Chao

(School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China)

Abstract: Large flexible aircraft possess large structural deformation with aerodynamic loads, which makes the dynamic characteristics change obviously. Structural deformation predictions are vital to large flexible aircraft design and aeroelastic simulations. Full-order models like the finite element method have low simulation efficiency. Traditional reduced-order model (ROM) obtains high simulation efficiency but requires large amounts of sample data participating in model building. This study investigates machine learning algorithms to build a prediction model to calculate the static deformation of flexible structures considering geometric nonlinearities. The performance of the prediction models is compared under evaluation with root mean square error (RMSE). It is shown that several machine learning techniques can be applied to the prediction of large deformations. Moreover, a new static aeroelastic analysis method is proposed with a large deformation prediction model and non-planar vortex lattice method (VLM) with high accuracy and efficiency In the end, a single-beam flexible wing model is used, and the prediction model and wind tunnel test's static aeroelastic response are contrasted. The results demonstrate that the proposed method has good performance and great practical application value.

Keywords: geometric nonlinearity; machine learning; aeroelasticity; vortex lattice method; structural large deformation

Received: 2023-03-09; **Accepted**: 2023-04-27; **Published Online**: 2023-05-16 10:32

URL: link.cnki.net/urlid/11.2625.V.20230515.1706.001

Foundation items: National Natural Science Foundation of China (12102027); Young Elite Scientist Sponsorship Program by Beijing Association for Science and Technology (BYESS2023345)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0128

基于监督对比学习的无线电引信干扰识别方法

钱鹏飞[1]，秦高林[3]，陈齐乐[1]，郝新红[1,2,*]

(1. 北京理工大学 机电动态控制重点实验室，北京 100081；2. 北京理工大学 唐山研究院，唐山 063099；
3. 华北水利水电大学，郑州 450045)

摘 要：连续波调频多普勒引信在战场上容易受到干扰，从而导致弹药早炸失去毁伤能力。为了提高调频多普勒引信对信息型干扰的抗干扰能力，实现多种干扰信号与目标回波的区分，提出一种基于监督对比学习的目标与干扰信号分类识别方法。该方法首先通过残差网络和自注意力机制搭建了主干网络；然后利用引入标签的方式改进了对比学习损失函数，实现了监督对比学习；最后采用中频信号搭建数据集，通过监督对比学习的方式来训练网络，从而实现对目标与干扰信号的分类和识别。仿真结果表明：该方法能够实现多种干扰种类与目标回波的识别，并且识别率能够达到98.7%。在低信噪比环境下的识别效果更为出色，在信噪比为−18 dB的环境下，仍然能有91.81%的识别率，相比普通残差网络的86.12%的识别率更高。

关 键 词：调频多普勒引信；电子对抗；深度神经网络；监督对比学习；信号识别
中图分类号：TJ43+4.1
文献标志码：A **文章编号**：1001-5965(2025)03-0953-09

随着电子干扰技术的发展，信息型干扰具有干扰效率高等优点，逐渐成为主要的干扰手段。面对复杂的信息型干扰，连续波调频多普勒无线电引信很容易受到干扰进而导致早炸，极大地降低了弹药造成的毁伤性能。因此，让连续波调频多普勒引信能够识别出目标与干扰信号，从而提升自身的抗干扰性能，具有重要研究意义。

对于信号与干扰的识别，目前已经有学者进行了相关的研究，并取得了研究成果。文献[1-3]提取了包括信号频域熵等多维特征在支持向量机和模糊C均值算法中进行训练和分类，对目标信号进行聚类划分。文献[4-5]对引信检波进行傅里叶变换后根据频谱局部比值差异性作为特征进行区分。但这种方式仅对扫频式干扰有效。文献[6]利用差频信号和多普勒信号构建信号的二维特征对粗糙地面回波信号和DRFM干扰进行分类识别。文献[7]利用混频后的输出信号奇异谱熵和香农熵对目标与干扰做区分，提出以支持向量机作为分类器利用特征进行分类识别。但这种方法的正确率受到人为特征提取的影响大，对于其他干扰类型不具备适用性。文献[8]利用加入CBAM注意力机制来提高普通残差网络的性能，将时频图导入网络实现对雷达信号调制类型的识别。文献[9]分别对二维特征图像和以为特征序列进行了分析与识别，实现了对雷达信号的分类识别。文献[10]利用卷积网络实现了对8种雷达信号的时频图像的识别，并取得了不错的识别率。但时频图的计算量大，需要消耗大量的物理资源。

传统的支持向量机需要人为提取特征，对于多种类复杂信号的分类，特征提取较为困难，特征提取的好坏决定着分类效果。而卷积神经网络有着自主提取特征的优势，并且采用监督对比学习的方式能够进一步提高网络的分类效果。因此本文提出一种基于神经网络对多种信息型干扰进行识别的方法。该方法采用采样后的中频时域信号作为数据集，通过残差网络和自注意力机制搭建主干网

收稿日期：2023-03-16；录用日期：2023-06-30；网络出版时间：2023-07-19 10:59
网络出版地址：link.cnki.net/urlid/11.2625.V.20230718.1657.002
*通信作者. E-mail：haoxinhong@bit.edu.cn

引用格式：钱鹏飞，秦高林，陈齐乐，等. 基于监督对比学习的无线电引信干扰识别方法[J]. 北京航空航天大学学报，2025，51(3)：953-961.
QIAN P F，QIN G L，CHEN Q L，et al. A recognition method of radio fuze signal based on supervised contrastive learning[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：953-961 (in Chinese).

络，利用加入标签改进对比学习熵实现有监督对比学习来训练网络，最后利用训练好的网络来实现分类识别。

1 干扰响应过程

调频多普勒引信是一种利用差频信号与距离之间的关系为基础原理进行定距的引信，相对于连续波多普勒引信具有定距精度高、炸点散布小等特点。差频信号的频谱是以调制频率的整数倍为谐波分量所构成的，并在各次谐波周围出现以多普勒频率为边带的频谱特性。由于各次谐波的峰值与弹目距离存在对应关系，含有距离信息，多普勒频率含有速度信息，因此可以通过二次混频提取出谐波的多普勒包络来进行目标识别[11]。如图1所示。

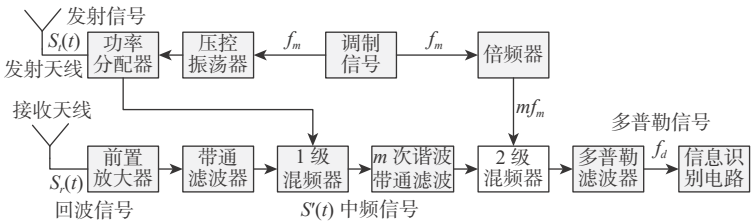


图1 引信结构图

Fig. 1 Fuze structure

随着电子干扰设备的发展，多种复杂的信息型干扰能够使调频多普勒引信受到影响，从而导致误判。相较于回波，干扰信号的能量往往更大，所以干扰信号能够在远距离进入到引信中，产生较强的多普勒信号，从而导致误判造成早炸。本文将以三角波调频信号作为发射信号，瞄准式调幅干扰信号作为干扰为例，分析干扰的作用过程。三角波调频信号通常可以表示为

$$S_t(t)=\begin{cases}A\cos\left(2\pi(f_c+(4n+1)\Delta f)t-\pi\beta t^2\right)\\ \qquad nT\leqslant t\leqslant nT+\dfrac{T}{2}\\ A\cos\left(2\pi(f_c-(4n+3)\Delta f)t+\pi\beta t^2\right)\\ \qquad nT+\dfrac{T}{2}\leqslant t\leqslant nT+T\end{cases}\tag{1}$$

式中：A表示幅度；f_c表示载波频率；Δf表示单边调频带宽；T表示调制周期；$\beta=4\Delta f/T$表示调频斜率。为了后续推导公式的简洁可以将发射信号表示为

$$\begin{cases}S_t(t)=A\cos(2\pi f_c t)\sum\limits_{m=-\infty}^{\infty}a_m e^{j2\pi m f_m t}\\ a_m=\left(\cos\left(\dfrac{\pi(\mu+m)^2}{4\mu}\right)(C(a)+C(b))-\right.\\ \left.\sin\left(\dfrac{\pi(\mu+m)^2}{4\mu}\right)(S(a)+S(b))\right)\Big/\sqrt{2\mu}\end{cases}\tag{2}$$

式中：a_m是发射信号的傅里叶系数；$\mu=\Delta f/f_m$表示调频指数；$C(a)$，$C(b)$和$S(a)$，$S(b)$表示涅菲尔积分。

瞄准式调幅干扰信号可以表示为

$$S_j(t)=(1+B\cos(2\pi f_j t))\cos(2\pi f_C t)\tag{3}$$

式中：B表示干扰信号的幅度；f_j表示干扰信号的调制频率；f_C表示干扰信号的载频。干扰信号通过接收天线，经过带通滤波器，经过一次混频后得到：

$$S_{tj}(t)=S_t(t)\cdot S_j(t)=A(1+B\cos(2\pi f_j t))\cdot\cos(2\pi f_c t)\cos(2\pi f_C t)\sum_{m=-\infty}^{\infty}a_m e^{j2\pi m f_m t}\tag{4}$$

混频后的信号经过低通滤波滤除高频信号后，得到中频信号，将中频信号经过M次谐波带通滤波器后得到：

$$S'_{tj}(t)=A(1+B\cos(2\pi f_j t))\cos(2\pi(f_c-f_C)t)\sum_{m=M}a_m e^{j2\pi m f_m t}\tag{5}$$

式中：M表示能够进入带通滤波器的谐波数。再与第M次谐波进行二次混频经过低通滤波得到多普勒信号：

$$S''_{tj}(t)=Aa_m(1+B\cos(2\pi f_j t))\cos(2\pi(f_c-f_C)t)\tag{6}$$

从式(6)可知，当干扰信号的中心频率对准信号的中心频率即$f_c=f_C$，调制信号的频率接近于多普勒频率$f_j=f_d$时，干扰信号能够进入到多普勒滤波器中，对引信造成干扰。

根据以上理论推导后的结果进行设置参数，搭建仿真模型，检波端输出如图2所示。

根据图2可知，按照理论推导的条件设置参数能够让干扰信号进入到检波端输出信号中，对引信

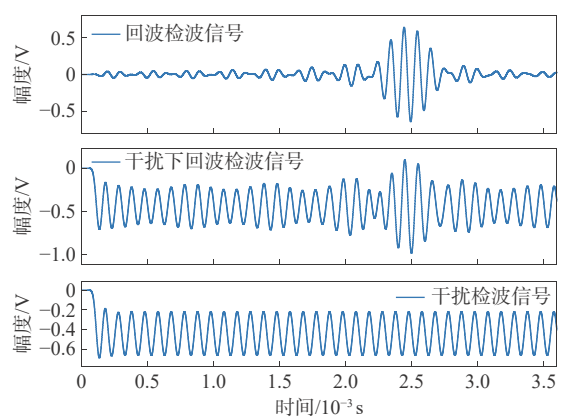


图2 检波输出图

Fig. 2 Detection output

的检波信号造成影响。

2　基于监督对比学习的目标与干扰的识别方法

从第 1 节可知不同干扰和目标的中频信号由于含有丰富且复杂的时域特征难以人为提取, 而神经网络可以通过训练的方式自主提取出复杂特征, 文献 [12] 通过神经网络实现了对不同复杂时域信号之间的区分, 证明了神经网络对复杂时域波形区分有着不错的性能。因此, 本文提出一种利用监督对比学习网络来对目标回波和各类干扰的中频信号进行区分, 从而实现目标与干扰的识别的方法。该方法通过用目标回波和不同干扰的中频信号构成训练集, 采用监督的对比学习策略训练网络, 最后利用训练好的网络实现对目标与干扰的识别。

2.1　监督对比学习框架

传统的自监督对比学习主要是用来解决无标签的分类任务, 依靠数据增强来实现描述相似和不同事物特征的一种方法。在此基础之上, 谷歌团队提出了一种新的有监督对比学习的方法[12]。该方法综合了对比学习的聚类特性和有监督学习的训练优势, 实现了对有标签的分类任务的有效分类, 相比于自监督对比学习, 通过加入标签的方式优化监督对比损失函数来进行训练, 降低训练难度; 并且相比于传统利用交叉熵的有监督学习方式, 得益于监督对比熵的聚类特性, 网络具有更强的泛用性和鲁棒性。

基于文献 [13] 所提出的框架, 本文提出了一种基于监督对比学习的目标与干扰的识别方法。网络框架主要由四部分构成。

(1) 数据增强的数据集: 通过随机噪声、随机平移、信号遮挡等对数据集中的数据进行数据增强, 获得一对正样本用于训练。

(2) 基于残差网络和自注意力机制的编码网络: 本文针对复杂时域信号的特征提取, 选用的是由残差网络和自注意力机制层共同构成的编码网络, 网络首先采用残差神经网络提取出丰富多样的特征向量来解决时域特征复杂的问题; 再通过自注意力机制通过特征之间的相关性将特征进行分析, 从而获取最佳分类特征。

(3) 投影网络: 将编码网络中的特征向量空间投影到对比空间中, 本文是采用含有一个隐藏层的多层感知机网络来实现的。

(4) 分类器: 作用是利用已经训练好的编码网络得到的特征向量来完成分类任务, 本文采用一个结构简单的全连接网络来实现分类。

其中, 监督对比学习的主要核心内容在于利用监督对比损失函数通过特殊的训练方式来训练网络。训练方式如图 3 所示。

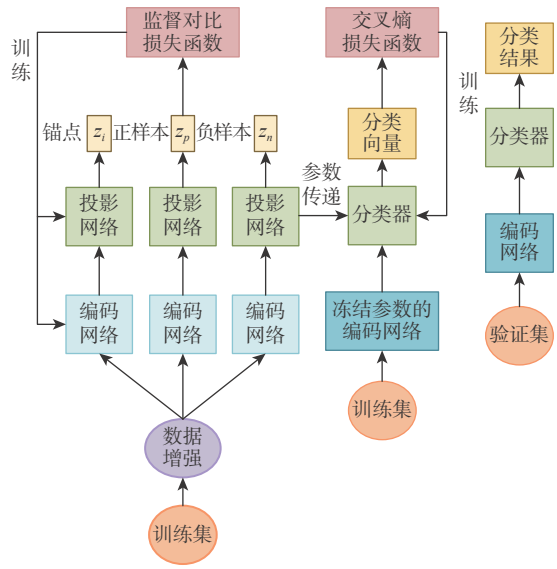


图 3　监督对比学习的训练方式

Fig. 3　Training mode of supervised contrastive learning

在训练阶段, 首先利用增广后的训练集导入由投影网络和编码网络构成的网络结构中, 通过监督对比损失函数, 来对编码网络和投影网络进行训练。然后保留编码器网络的结构和参数, 将投影网络换成分类器, 利用交叉熵损失函数来训练分类器, 最终完成训练; 在验证阶段, 采用编码网络和分类器构成的网络结构对验证集的数据进行分类识别。

2.2　主干网络

本文的网络结构是通过由残差网络和自注意力机制共同构成的编码网络和由全连接网络实现的分类器共同组成的, 如图 4 所示。

网络首先采用残差神经网络提取出丰富多样的特征向量来解决多种干扰的分类特征复杂的问题; 再通过自注意力机制将残差神经网络中提取出来的特征进行分析, 对每一个特征向量融入与其他特征向量之间的相关性, 从而获取最佳分类特征; 最后将特征导入结构简单的全连接分类器进行分类, 最终实现对目标与干扰信号的识别。该网络相较于经典的 ResNet18 等网络的结构更为精简。网络的结构直接影响了网络运算效率和所需的计算量。因此, 该网络结构更加符合干扰识别的实时性要求。

2.2.1　残差网络

残差网络和传统的卷积神经网络(convolutional neural networks, CNN)一样, 都是利用不同参数的卷

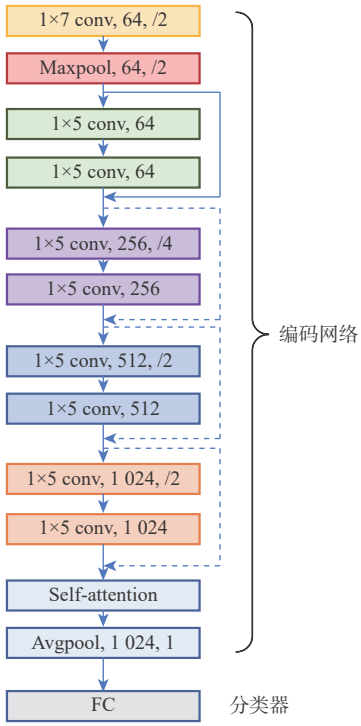


图4 网络结构

Fig. 4 Network structure

积核来提取出局部的特征信息，再通过层数的叠加将特征信息进一步细化和综合上层局部的信息最终产生更具有语义的特征块。卷积核的大小决定了感知范围，并且利用权重共享原则来检测到不同位置的相同特征。因此可以通过卷积核来提取时域信息，如图5所示。

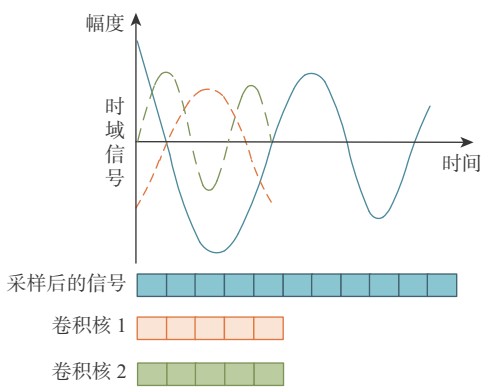


图5 卷积原理

Fig. 5 Principle of convolution

卷积核提取时域特征的原理就是利用不同参数的卷积核对时域的信号进行卷积，提取出与卷积块的相关性特征，其中的相关性与卷积核的参数有关。提供给下层进行进一步提取，最终基于含有不同参数的卷积核得到丰富的信号特征，反映了波峰波谷的形状和之间的相对位置关系等更具有语义的特征信息。

与残差网络相比，卷积神经网络的优势在于网络更有利于将训练过程中的损失值从输出反传给输入，大大简化了学习的难度，提高了网络性能。残差块是残差网络的基本单元，残差块的基本结构如图6所示。

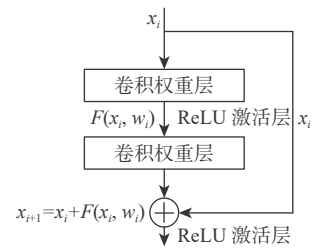


图6 残差块结构图

Fig. 6 Residual block structure

残差的公式表达为

$$x_{i+1} = x_i + F(x_i, w_i) \tag{7}$$

式中：x_i表示残差块输入值；x_{i+1}表示残差块输出值；$F(x_i, w_i)$表示残差块内部运算的函数。经过不同层之后，可以得到任意两层之间的表达式为

$$x_I = x_i + \sum_{n=i}^{I-1} F(x_i, w_i) \tag{8}$$

式中：x_I表示任一层的输出值。通过式(8)可以发现，任意两层之间都具有残差的特性。对于反向传播，由此可得

$$\frac{\partial \varepsilon}{\partial x_i} = \frac{\partial \varepsilon}{\partial x_I}\frac{\partial x_I}{\partial x_i} = \frac{\partial \varepsilon}{\partial x_I}\left(1 + \frac{\partial}{\partial x_i}\sum_{n=i}^{I-1} F(x_i, w_i)\right) \tag{9}$$

式中：ε表示损失值。当$\frac{\partial}{\partial x_i}\sum_{n=i}^{I-1} F(x_i, w_i)$趋近于0时，卷积神经网络出现梯度消失的现象，而残差网络$\frac{\partial \varepsilon}{\partial x_i} = \frac{\partial \varepsilon}{\partial x_I}$则跳过权重层进行传递，实现任意两层之间的传递。

2.2.2 自注意力机制

从残差网络可以得到大量的特征向量，而特征向量之间存在一定的相关性信息，所以为了提高网络性能，采用自注意力机制可以将特征向量之间的相关性信息和特征向量本身的信息共同构成新的特征向量。自注意力机制网络可以通过训练的方式，利用特征向量之间的相关信息来调整网络参数，从而选出对分类结果重要的特征向量，来构成新的特征向量，达到优化特征向量的目的[14]。自注意力机制的原理如图7所示。

首先通过输入的特征向量与构建三组参数矩阵相乘，变成三组新的向量空间，里面不仅含有输入的特征向量的信息，还含有参数矩阵存储的参数信息。三组参数可以用向量形式和矩阵形式分别表示如下。

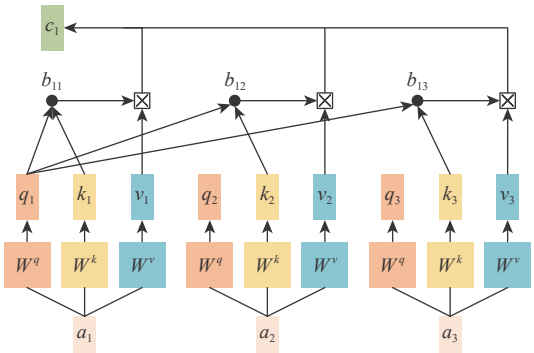


图 7　自注意力机制原理

Fig. 7　Principle of self-attention mechanism

向量形式:

$$\begin{cases} \boldsymbol{q}_i = \boldsymbol{W}^q \cdot \boldsymbol{a}_i \\ \boldsymbol{k}_i = \boldsymbol{W}^k \cdot \boldsymbol{a}_i \\ \boldsymbol{v}_i = \boldsymbol{W}^v \cdot \boldsymbol{a}_i \end{cases} \tag{10}$$

矩阵形式:

$$\begin{cases} \boldsymbol{Q} = \boldsymbol{W}^q \cdot \boldsymbol{I} \\ \boldsymbol{K} = \boldsymbol{W}^k \cdot \boldsymbol{I} \\ \boldsymbol{V} = \boldsymbol{W}^v \cdot \boldsymbol{I} \end{cases} \tag{11}$$

式中: $\boldsymbol{I}$是由 n个特征向量 $\boldsymbol{a}_i$构成的矩阵, 即输入矩阵。

然后用本身向量的 $\boldsymbol{q}_i$向量与其他向量的 $\boldsymbol{k}_j$向量求积, 得到含有相关信息的值 b_{ij}。

向量形式:

$$\boldsymbol{b}_i = \boldsymbol{K}^{\mathrm{T}} \cdot \boldsymbol{q}_i \tag{12}$$

矩阵形式:

$$\boldsymbol{B} = \boldsymbol{K}^{\mathrm{T}} \cdot \boldsymbol{Q} \tag{13}$$

式中: 向量 $\boldsymbol{b}_i$中的元素是由 $\boldsymbol{b}_{ij} = \boldsymbol{k}_j^{\mathrm{T}} \cdot \boldsymbol{q}_i$构成的, 里面包含了相关信息, 这样矩阵 $\boldsymbol{B}$中每一个值就得到了两个特征向量的相关性。将矩阵 $\boldsymbol{B}$归一化后得到 $\boldsymbol{B}'$, 再将含有原信息的矩阵 $\boldsymbol{V}$与含有相关性信息的矩阵 $\boldsymbol{B}'$相乘, 就得到了既含有原始特征向量信息又含有相关性信息的输出矩阵 $\boldsymbol{C}$。

向量形式:

$$\boldsymbol{c}_i = \boldsymbol{V} \cdot \boldsymbol{b}_i' \tag{14}$$

矩阵形式:

$$\boldsymbol{C} = \boldsymbol{V} \cdot \boldsymbol{B}' \tag{15}$$

因此通过自注意力机制能够通过加入特征向量之间的相关性信息来构成新的特征向量, 并且受到 3 个参数矩阵里的参数值的影响。通过训练参数来筛选出更具有分类意义的特征向量构成新的特征向量来进行分类识别。

2.3　监督对比损失函数

对于传统的有监督的训练方式, 交叉熵损失函数是应用最为广泛的损失函数, 它被用来度量标签分布和经验回归分布之间的 KL 散度。交叉熵表明两个概率分布的相似程度, 刻画了通过经验回归分布来表达标签分布的困难程度, 交叉熵越小, 表示两个概率分布越接近, 预测结果越符合真实标签。交叉熵损失函数可以表示为

$$L = \frac{1}{N}\sum_N H(p,q) = -\frac{1}{N}\sum_N \log\left(\exp(z_{\text{class}}) \Big/ \sum_k^C \exp(z_k)\right) = -\frac{1}{N}\sum_N \left(z_{\text{class}} - \log\sum_k^C \exp(z_k)\right) \tag{16}$$

式中: N表示一次训练中含有 N组数据, 训练中的 batch size, 对于一个输入都会得到一个分类向量 $\boldsymbol{z}$, 它里面包含分别对于每一类的预测值, z_{class}表示正确标签下向量 $\boldsymbol{z}$中对应的值, z_k表示第 k类在 $\boldsymbol{z}$中的预测值。虽然交叉熵能够满足大部分的分类任务, 但是只通过交叉熵训练出的网络参数缺乏鲁棒性, 聚类效果不明显, 抗噪能力弱。

传统的对比学习是一种自监督的学习方式, 为了在没有标签的情况下学习出事物的特征, 通过将本身的数据集增广, 把同源的样本作为正样本, 而其他的所有全部作为负样本来进行学习, 从而拉近锚点与正样本的距离来达到学习的目的。所以具有很优秀的聚类特性。对于样本数据 $\{\boldsymbol{x}_n, \boldsymbol{y}_n\}_{n=1}^N$, 通过随机数据增强获得扩展数据集 $\{\hat{\boldsymbol{x}}_k, \hat{\boldsymbol{y}}_k\}_{k=1}^{2N}$, 其中 $\boldsymbol{x}_n \in \{\boldsymbol{x}_n, \boldsymbol{y}_n\}_{n=1}^N$的随机数据增强所得的 $\hat{\boldsymbol{x}}_{2n-1}, \hat{\boldsymbol{x}}_{2n} \in \{\hat{\boldsymbol{x}}_k, \hat{\boldsymbol{y}}_k\}_{k=1}^{2N}$为同源正样本, 而其他 $2N-2$个数据为负样本, 对于数据集 $I = \{1, 2, \cdots, 2N\}$, 对比学习损失函数定义为

$$I^{\text{self}} = \sum_{i\in I} I_i^{\text{self}} = -\sum_{i\in I} \log \frac{\exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_{j(i)}/\tau)}{\sum_{a\in A(i)} \exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_a/\tau))} \tag{17}$$

式中: $\boldsymbol{z}_i$是增广后的数据 $\{\hat{\boldsymbol{x}}_k, \hat{\boldsymbol{y}}_k\}_{k=1}^{2N}$中的任意样本; $\boldsymbol{x}_i$是经过编码器网络和投影网络得到的向量, 锚点的输出向量; $\boldsymbol{x}_{j(i)}$是在增广数据集中的 $\boldsymbol{x}_i$正样本; $\boldsymbol{z}_{j(i)}$是正样本的输出向量; $A(i) = I\backslash\{i\}$ 是 $I = \{1, 2, \cdots, 2N\}$除了锚点 i以外的元素构成的集合; $\boldsymbol{z}_a$为除了锚点以外的输出向量; $\tau \in R^+$是一个温度参数, 用来调整式 (17)。

因此, 式 (17) 表示对于每一个锚点 $\boldsymbol{x}_i$, 都能得到一个 I_i^{self}损失值, 它是由正样本与锚点的内积构成分子, 其他样本与锚点的内积构成分母所得到的值。当与正样本的近似程度相近时, I_i^{self}越接近于 0, 反之变大。

由此可见, 该损失函数只考虑了同源正样本的近似程度, 而对于有标签的任务, 需要拉近与所有同标签的距离, 实现分类。所以, 监督对比学习的方式, 就是将式 (17) 的正样本扩大为所有同标

签的统称为正样本，利用标签来进行训练，让具有相同标签的数据更加聚集。监督对比损失函数定义为

$$\Gamma^{\sup}=\sum_{i\in I}\Gamma_i^{\sup}=-\sum_{i\in I}\frac{-1}{|P(i)|}\sum_{p\in P(i)}\log\frac{\exp(z_i\cdot z_{\mathrm{p}}/\tau)}{\sum_{a\in A(i)}\exp(z_i\cdot z_a/\tau))} \tag{18}$$

式中：$p\in p(i)$表示正样本的序号构成的集合，即$\boldsymbol{x}_{\mathrm{p}}$表示集合中$\{\hat{\boldsymbol{x}}_k,\hat{\boldsymbol{y}}_k\}_{k=1}^{2N}$与$\boldsymbol{x}_i$有同标签$\boldsymbol{y}_{\mathrm{p}}=\boldsymbol{y}_i$的正样本。其他参数与自对比监督函数中的参数具有同样意义。通过式 (18) 可知，锚点与同标签的相似程度越接近，分子越大，该损失函数的值越接近于 0，反之则变大。因此通过该损失训练网络可以拉近同标签样本之间的距离。从而达到聚类的效果，有助于分类器的分类。

3 仿真实验

为了验证论文所提出的方法的有效性，进行了回波与干扰区分的仿真实验。仿真所选的引信体制是三角波调频多普勒引信，干扰包括瞄准式干扰、扫频式干扰和噪声压制式干扰，其中瞄准式干扰和扫频式干扰选用 3 种不同的调制类型：调频、调幅和调相，噪声压制式干扰选用的是噪声调幅干扰。真实战场环境复杂且多变，为了让仿真数据尽可能地接近于真实数据，仿真所选的干扰类型是目前战场上常见的干扰类型，并且通过理论分析和仿真实验来确保其参数设置的有效性；另外噪声对其的影响不容忽视，因此仿真是在不同信噪比的环境下进行的，从而模拟现实复杂的电磁波环境。

3.1 网络训练

信号的仿真通过 simulink 来搭建不同干扰下的模型来实现。实验所用的发射波形是三角波调频信号，载波 3 GHz，调制频率 100 kHz，调制带宽 50 MHz；场景模拟的弹速为 500 m/s，从 50 m 开始仿真并加入干扰；干扰信号根据论文 [15-16] 设定参数，载波频率与发射波形的载波频率相等，调制周期选用多普勒信号所对应的周期，确保干扰能够进入多普勒滤波内。根据以上构建出的仿真模型，仿真出的差频数据经过采样后，通过 To Workspace 模块将数据导入 Matlab 工作区打包为 mat 文件进行数据集的存储。

网络模型的搭建是通过 PyTorch 编写代码来实现的，采用 Scipy 库中的函数读取.mat 文件，将数据导入经过预处理以及数据增强分为训练集、测试集和验证集。数据集由回波和 7 类干扰共同组成，每一类含有 2 000 组数据，总共 16 000 组数据。其中，按照类别随机抽取出 60% 构成训练集，20% 构成测试集，20% 构成验证集。训练集用于网络的训练，训练出合适的参数进行特征提取与分类；测试集用于在网络训练的过程中观察网络的训练效果，从而保存分类性能好的模型参数用于分类；验证集用于训练完成后对网络性能进行验证，模拟已经训练好的网络对信号的识别。通过 dateset 和 dateloader 对数据进行打包和预处理，完成数据集的构建。再通过 torch 库搭建网络，根据第 2 节的监督对比损失函数的原理编写损失函数，利用训练集通过对比监督损失函数训练构建好的编码网络，交叉熵训练分类器，完成训练。

3.2 仿真结果

将上述训练好的网络，用验证集进行验证。通过上述实验方法进行实验，对处于−18～9 dB 信噪比环境下的信号与干扰的区分，该方法整体可以达到 98.7% 的准确率；另外增加对比实验，所选用的网络模型一致，但没有用有监督对比学习的方式，而采用交叉熵训练所有参数，可以达到 95.41% 的准确率。对于每一类的区分准确度，本文用混淆矩阵的形式来表示，结果如图 8 所示。

从图 8 可知，采用监督对比学习的方式训练网

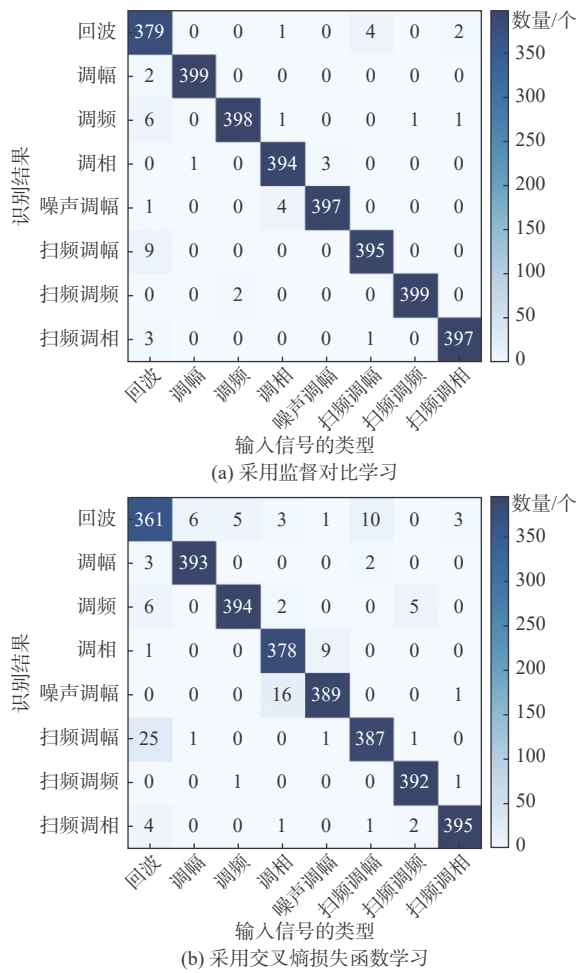


图 8 混淆矩阵图

Fig. 8 Confusion matrix

络, 有助于提高识别率, 增强网络性能, 有利于信号与干扰的区分。另外, 将网络得到的特征向量通过 t-SNE 分别降成二维, 一维进行可视化画图, 二维散点图聚类结果如图 9 所示, 横纵坐标分别代表两个不同的维度, 散点图中的点代表高维特征向量在此二维空间的投影。

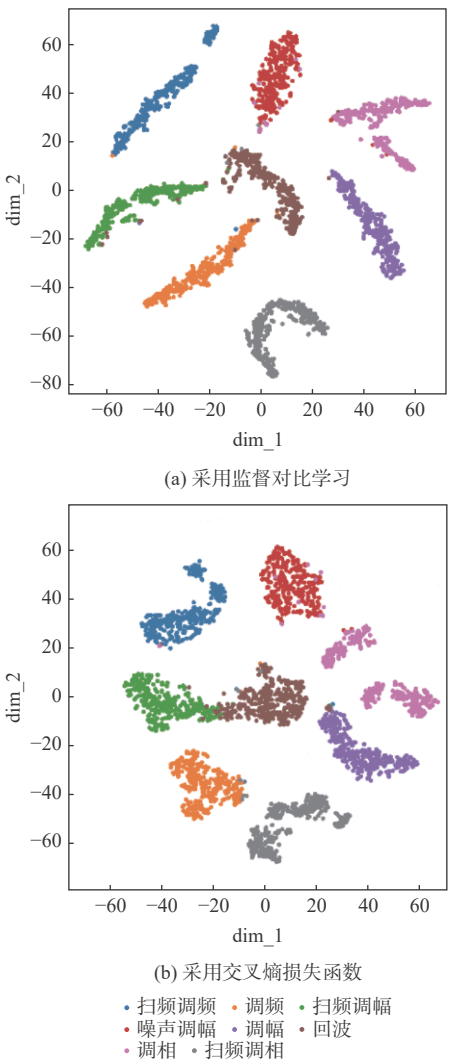


图 9　二维散点图

Fig. 9　2D scatterplots

由图 9 可知, 采用监督对比学习训练出的特征向量聚类效果更明显, 更有利于分类。一维箱型图如图 10 所示。

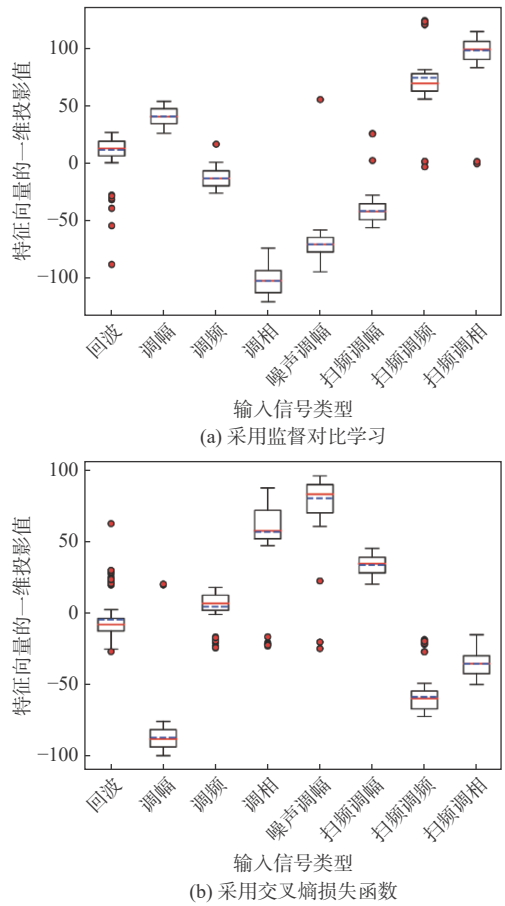


图 10　箱型图

Fig. 10　Box plot

由图 10 可知, 采用监督对比学习训练出的特征向量区分度更高, 有利于提升分类效果, 更具有鲁棒性, 对于复杂环境下干扰与信号的区分更具有稳健性。

为了验证该方法的不同信噪比环境下的抗噪能力, 在−18～9 dB 各信噪比环境下, 对使用本文提出的主干网络分别采用对比监督损失函数和交叉熵损失函数训练进行实验。另外, 由于文献 [8] 所采用的残差网络相比于文献 [10] 中的卷积网络有着易于训练、训练效果好等优点, 因此, 论文采用残差网络作为对比实验。具体的识别率如图 11 所示。

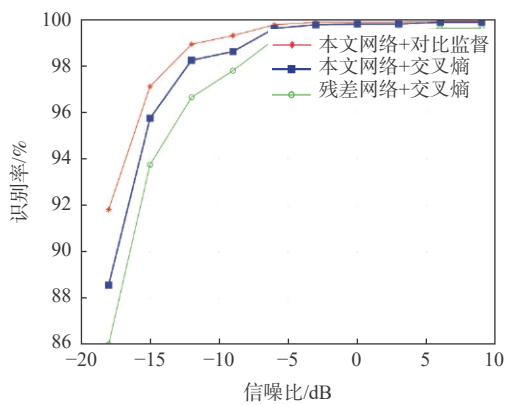


图 11　不同信噪比下的识别率

Fig. 11　Recognition rate under different SNRs

由图 11 可知, 该方法在信噪比为−10 dB 及以上的环境下, 识别率超过 99%。在信噪比为−18 dB 的环境下, 识别率仍然能有 91.81%。相比于直接用交叉熵进行训练的网络, 该方法抗噪能力更佳, 更能够满足低信噪比环境下对干扰与信号进行识别区分的要求。

4 结 论

本文针对目前已有的调频多普勒引信与干扰区分方法中，特征不具备泛用性，区分种类少，低信噪比环境识别率低等问题，以三角波调频多普勒体制为例，提出了一种基于监督对比学习神经网络对信号和其他多种不同类别的干扰进行分类识别的方法。通过仿真实验验证了方法的有效性，并得出以下结论：

1）该方法能够实现信号与多种干扰类型之间的区分，并且有着较高的分类准确率。主干网络结构可以通过训练的方式自适应地提取出有效的特征向量。

2）基于监督对比学习的方式能够产生聚类效果明显的特征向量，进一步提升分类效果和识别率。

3）对于低信噪比的环境下，该方法仍然能实现有效的区分。在信噪比为−18 dB 的环境下，识别率超过 90%。显著优于直接用交叉熵训练的网络。

参考文献（References）

[1] 代健, 晏祺, 闫晓鹏, 等. 基于模糊 c-均值增量更新的脉冲多普勒引信干扰与目标信号识别[J]. 兵工学报, 2018, 39(9): 1711-1718.
DAI J, YAN Q, YAN X P, et al. Recognition of jamming and target signal for pulse Doppler fuze based on FCM algorithm with incremental update[J]. Acta Armamentarii, 2018, 39(9): 1711-1718(in Chinese).

[2] 代健, 李泽, 郝新红, 等. 基于目标联合特征提取的脉冲多普勒引信抗干扰方法[J]. 兵工学报, 2019, 40(2): 225-233.
DAI J, LI Z, HAO X H, et al. Anti-jamming method for pulse Doppler fuze based on joint feature extraction of target signal[J]. Acta Armamentarii, 2019, 40(2): 225-233(in Chinese).

[3] DAI J, HAO X H, LI Z, et al. Adaptive target and jamming recognition for the pulse Doppler radar fuze based on a time-frequency joint feature and an online-updated naive Bayesian classifier with minimal risk[J]. Defence Technology, 2022, 18(3): 457-466.

[4] 张彪, 闫晓鹏, 栗苹, 等. 基于支持向量机的无线电引信抗扫频式干扰研究[J]. 兵工学报, 2016, 37(4): 635-640.
ZHANG B, YAN X P, LI P, et al. Research on anti-frequency sweeping jamming of radio fuze based on support vector machine[J]. Acta Armamentarii, 2016, 37(4): 635-640(in Chinese).

[5] 李志强. 连续波多普勒无线电引信目标信号识别方法研究[D]. 北京: 北京理工大学, 2014.
LI Z Q. Research on target signal identification method of continuous wave Doppler radio fuze[D]. Beijing: Beijing Institute of Technology, 2014(in Chinese).

[6] 郝新红, 杜涵宇, 陈齐乐. 调频引信粗糙面目标与干扰信号识别[J]. 北京航空航天大学学报, 2019, 45(10): 1946-1955.
HAO X H, DU H Y, CHEN Q L. Rough surface target and jamming signal recognition of FM fuze[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(10): 1946-1955(in Chinese).

[7] 黄莹, 郝新红, 孔志杰, 等. 基于熵特征的调频引信目标与干扰信号识别[J]. 兵工学报, 2017, 38(2): 254-260.
HUANG Y, HAO X H, KONG Z J, et al. Recognition of target and jamming signal for FM fuze based on entropy features[J]. Acta Armamentarii, 2017, 38(2): 254-260(in Chinese).

[8] 吴礼洋, 吴鹏程, 刘超, 等. 基于注意力机制增强残差网络的雷达信号调制类型识别[J]. 兵工学报, 2023, 44(8): 2310-2318.
WU L Y, WU P C, LIU C, et al. Radar signal modulation type recognition based on attention mechanism enhanced residual network[J]. Journal of Ordnance Engineering, 2023, 44(8): 2310-2318(in Chinese).

[9] 瞿奇哲. 基于深度学习的雷达信号识别[D]. 成都: 电子科技大学, 2021.
QU Q Z. Radar signal recognition based on deep learning [D]. Chengdu: University of Electronic Science and Technology of China, 2021(in Chinese).

[10] 刘国满, 聂旭娜. 一种基于卷积神经网络的雷达干扰识别算法[J]. 北京理工大学学报, 2021, 41(9): 990-998.
LIU G M, NIE X N. A radar jamming recognition algorithm based on convolutional neural network[J]. Transactions of Beijing Institute of Technology, 2021, 41(9): 990-998(in Chinese).

[11] 崔占忠, 宋世和, 徐立新. 近炸引信原理 [M]. 第 3 版. 北京: 北京理工大学出版社, 2009: 76-100.
CUI Z Z, SONG S H, XU L X. Principle of proximity fuze [M]. 3 ed. Beijing: Beijing Institute of Technology Press, 2009: 76 -100(in Chinese).

[12] KHOSLA P, TETERWARK P, WANG C, et al. Supervised contrastive learning[J]. Advances in Neural Information Processing Systems, 2020, 33: 18661-18673.

[13] WANG Q, DU P F, YANG J Y, et al. Transferred deep learning based waveform recognition for cognitive passive radar[J]. Signal Processing, 2019, 155: 259-267.

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.

[15] 左环宇, 郝新红, 岳凯. 调频多普勒引信抗调幅干扰性能分析[J]. 北京航空航天大学学报, 2017, 43(10): 2163-2170.
ZUO H Y, HAO X H, YUE K. Anti-AM jamming performance analysis of FM Doppler fuze[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(10): 2163-2170(in Chinese).

[16] 李泽, 闫晓鹏, 栗苹, 等. 扫频式干扰对调频多普勒引信的干扰机理研究[J]. 兵工学报, 2017, 38(9): 1716-1722.
LI Z, YAN X P, LI P, et al. Jamming mechanism of frequency sweep jamming to FM Doppler fuze[J]. Acta Armamentarii, 2017, 38(9): 1716-1722(in Chinese).

A recognition method of radio fuze signal based on supervised contrastive learning

QIAN Pengfei[1]，QIN Gaolin[3]，CHEN Qile[1]，HAO Xinhong[1, 2, *]

(1. Science and Technology on Electromechanical Dynamic Control Laboratory，Beijing Institute of Technology，Beijing 100081，China；
2. Tangshan Research Institute，Beijing Institute of Technology，Tangshan 063099，China；
3. North China University of Water Resources and Electric Power，Zhengzhou 450045，China)

Abstract: Frequency modulated continuous wave (FMCW) Doppler fuze is easy to be interfered with on the battlefield, resulting in an early explosion and loss of damage ability. To improve the anti-jamming ability of FMCW Doppler fuze against information-based jamming and realize the distinction between multiple jamming signals and target echoes, this paper proposed a method of target and jamming signal classification and recognition based on supervised contrastive learning. Firstly, the backbone network was constructed by residual network and self-attention mechanism. Then, the contrastive learning loss function was improved by introducing labels, and supervised contrastive learning was realized. Finally, an intermediate frequency signal was used to build the dataset, and the network was trained by supervised comparative learning, so as to realize the classification and recognition of the target and jamming signal. The simulation results show that this method can realize the recognition of multiple jamming types and target echoes, and the recognition rate can reach 98.7%. In the low signal-to-noise ratio (SNR) environment, the recognition effect is better. In the SNR environment of −18 dB, the recognition rate is still 91.81%, which is higher than the 86.12% recognition rate of ordinary residual networks.

Keywords: frequency modulated Doppler fuze; electronic countermeasures; deep neural network; supervised contrastive learning; signal recognition

Received：2023-03-16；**Accepted**：2023-06-30；**Published Online**：2023-07-19 10：59
URL：link.cnki.net/urlid/11.2625.V.20230718.1657.002

* **Corresponding author**. E-mail：haoxinhong@bit.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0135

轴压铝合金蜂窝圆筒整体稳定性研究

魏家睿[1,2,3]，吴琼[1,2,*]，吴伟[4]，李群智[5]，时红伟[5]，张虞[1,2]

(1. 北京航空航天大学 虚拟现实技术与系统国家重点实验室，北京 100191；2. 北京航空航天大学江西研究院景德镇分院，景德镇 333000；3. 中国航天科工集团第九总体设计部，武汉 430040；4. 北京空天技术研究所，北京 100074；5. 北京空间飞行器总体设计部，北京 100094)

摘 要：金属蜂窝圆筒结构常因轴压整体屈曲导致承载能力下降，对于其弹性屈曲问题目前已有大量研究，但在塑性屈曲问题方面的理论研究工作较少。为解决上述问题，基于蜂窝圆筒等效模型和Hamilton原理，推导了铝合金蜂窝圆筒承受轴压载荷作用时弹、塑性屈曲临界载荷的理论计算公式，并进行了轴压仿真分析和实验验证，利用理论公式讨论了蜂窝圆筒各尺寸参数对临界载荷的影响。结果表明：所提理论公式和仿真模型均能准确预测铝合金蜂窝圆筒的临界载荷，理论公式计算效率更高；圆筒质量变化±10%时，影响临界载荷的因素依次为面板厚度、圆筒半径、芯子厚度、圆筒高度；蜂窝圆筒的临界载荷与芯子厚度、面板厚度、圆筒半径近似呈线性正相关，与圆筒高度关系较弱。

关 键 词：蜂窝圆筒；整体稳定性；临界载荷；弹塑性；Hamilton原理；有限元
中图分类号：TH122
文献标志码：A **文章编号**：1001-5965(2025)03-0962-11

金属圆筒结构具有优良的力学特性，被广泛用于海洋工程、建筑工程和航空航天工程中[1-3]。而蜂窝夹层结构凭借其比强度高、比刚性大等特征[4]，常被用来实现结构的轻量化设计。若在金属圆筒中运用蜂窝夹层形成蜂窝圆筒结构，则有望获得质量更轻、承载能力更强的轻量化结构形式。由铝合金制成的蜂窝圆筒具备轻质高强度和抗冲击性好等优异性能，目前在实际工程领域得到了大量运用，包括运载火箭的尾段和整流罩、深潜器的耐压外壳及高速车辆的抗冲击防护结构[5-7]。对于金属夹层圆筒结构，受轴压载荷作用是常见的载荷工况，通常会发生整体屈曲导致结构的承载能力下降，因此，圆筒的屈曲临界载荷是评价结构承载能力的重要指标，有必要针对轴压金属蜂窝圆筒的整体稳定性问题进行深入研究。

对于圆筒在轴压作用下的弹性屈曲问题，已进行了许多研究工作[8-11]，建立了不同尺寸圆柱壳在不同形式轴压载荷作用下的弹性屈曲分析方法。若圆柱壳材料为极易发生塑性变形的金属材料，则发生弹性屈曲前材料可能已进入塑性变形阶段，从而发生塑性屈曲。针对圆柱壳的轴压塑性屈曲问题也开展了一些研究，但在理论研究方面远不及弹性屈曲透彻，尤其是J_2流动理论和J_2形变理论这2种弹塑性本构关系理论的选择上仍存在争议[12]。Legendre等[13]基于J_2流动理论构建材料的本构方程，利用塑性分叉理论推导了低径厚比铝合金圆柱壳的塑性屈曲模式和临界载荷的表达式。Evkin等[14]基于能垒法和数值分析结果提出了弹塑性轴向压缩圆柱壳的局部屈曲估算公式。Rajabiehfard等[15]利用实验和数值方法研究了铝合金圆柱壳在轴向

收稿日期：2023-03-23；**录用日期**：2023-07-07；**网络出版时间**：2023-08-03 09:37
网络出版地址：link.cnki.net/urlid/11.2625.V.20230802.1344.004
基金项目：国家自然科学基金(51875024)；民用航天技术预先研究项目(D020304)
* **通信作者**. E-mail：wuqiong@buaa.edu.cn

引用格式：魏家睿，吴琼，吴伟，等. 轴压铝合金蜂窝圆筒整体稳定性研究[J]. 北京航空航天大学学报，2025，51(3)：962-972.
WEI J R，WU Q，WU W，et al. Study on global stability of aluminum alloy honeycomb cylinder under axial compression[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：962-972 (in Chinese).

冲击下的动态塑性屈曲行为。张靖华和郑武[16]基于 J_2 形变理论和 Hamilton 原理, 研究了功能梯度材料圆柱壳的轴压弹塑性屈曲问题, 并讨论了几何参数和材料参数对临界载荷和弹塑性分界面的影响。

对于夹层复合结构的弹性屈曲问题, 相关学者以夹层板和夹层圆筒为研究对象已提出了多种研究方法。Hou 等[17]基于 Hamilton 原理和代表性体积单元法讨论了 3 种低密度芯层结构蜂窝板的弹性屈曲模式和临界载荷。Zhang 等[18]基于虚位移原理推导了六边形蜂窝板的屈曲控制方程, 并采用纳维法进行求解。赵杰等[19]利用实验和有限元法分析了含缺陷钎焊蜂窝铝板在平压和侧压工况下的屈曲特性。王宝芹等[20]基于多尺度数值分析方法研究了复合材料蜂窝板在面内载荷作用下的稳定性。Zhang 等[21]借助经典弹性稳定性理论对方形蜂窝圆筒的屈曲模式进行推导, 并采用仿真和实验方法加以验证。Pan 等[22]基于一阶剪切变形理论得到了理想泡沫夹芯玻璃钢圆筒的局部屈曲理论解。

针对夹层复合结构的塑性屈曲问题, 目前主要采用实验和有限元法进行研究, 同时有部分学者对蜂窝夹层板进行了初步理论研究[23-24], 但鲜有针对蜂窝夹层圆筒塑性屈曲问题的理论研究工作。本文利用三明治夹芯板理论对蜂窝夹层结构进行等效, 根据 J_2 形变理论构建面板材料的弹塑性本构模型, 结合 Hamilton 原理和变分法进行硬夹芯铝合金蜂窝圆筒在轴压载荷作用下的弹、塑性屈曲理论分析, 并分别采用仿真法与实验法进行对比分析和验证, 利用临界载荷表达式分析了蜂窝圆筒尺寸参数对临界载荷的影响。

1　蜂窝圆筒等效模型

铝合金蜂窝圆筒由内外金属面板和蜂窝夹芯组成, 如图 1 所示。图中: R 为中曲面到圆筒轴心的距离, H 为圆筒高度, b 为蜂窝芯子厚度, 蜂窝圆筒的内、外面板厚度相等, 单侧面板厚度为 d。定义 x 轴为圆筒母线方向, θ 轴为环向, $x\theta$ 面为中曲面, z 轴垂直于 $x\theta$ 面, 后续理论推导均基于该坐标系。

蜂窝圆筒内蜂窝胞元的排列方式如图 2 所示。其中, 沿圆筒母线方向(即前文定义的 x 方向)的胞壁厚度为 $2t$, 其余胞壁厚度为 t。厚度为 $2t$ 的蜂窝胞壁均指向圆筒的中轴线。

提取蜂窝单胞进行分析, 靠近内面板的一侧定义为内胞面, 靠近外面板的一侧定义为外胞面, 如图 3 所示。由于蜂窝胞元的边长很小, 仍可近似将曲面蜂窝芯的各蜂窝边视作直边, 但外胞面所在曲

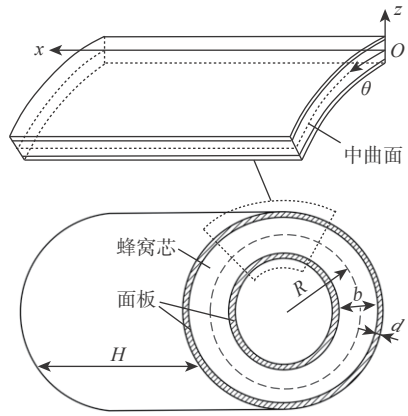


图 1　蜂窝圆筒示意图

Fig. 1　Schematic diagram of honeycomb cylinder

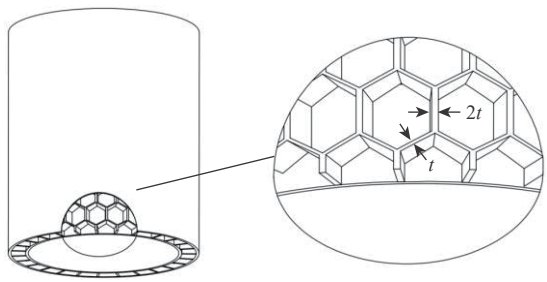


图 2　蜂窝胞元排列方式

Fig. 2　Honeycomb cell arrangement

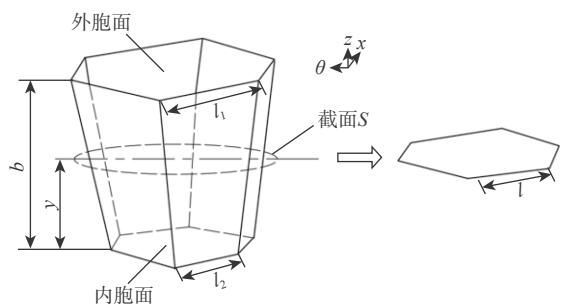


图 3　蜂窝单胞结构示意图

Fig. 3　Schematic diagram of honeycomb single cell structure

面直径大于内胞面, 因此, 外胞面的胞壁边长大于内胞面。讨论蜂窝芯为正六边形蜂窝的情况, 同一胞面内各胞壁边长相等, 定义 l_1 与 l_2 分别为外胞面和内胞面的胞壁边长。截面 S 为蜂窝胞元上的任意截面, 其胞壁边长为 l, 内胞面之间的距离为 y。

根据图 3 中的几何关系, 令 $\Delta l=l_1-l_2$, 可将 l 表示为

$$l=\frac{\Delta l}{b}y+l_2 \tag{1}$$

若 y 取 $b/2$, 截面 S 位于夹层中曲面, 令此时截面 S 的胞壁边长为 l_0。根据图 2 中蜂窝胞元的排列方式, 利用几何相似性将 l_1、l_2 用 l_0、R、b、d 表示为

$$\begin{cases} l_1=\dfrac{2R-2d}{2R-b-2d}l_0 \\ l_2=\dfrac{2R-2b-2d}{2R-b-2d}l_0 \end{cases} \tag{2}$$

由于蜂窝胞元结构的尺寸远小于整体结构的外形尺寸，为提高后续仿真分析的计算效率，同时简化后续理论分析过程，需要将蜂窝夹层结构进行等效替代。为便于引入面板的弹塑性本构模型，利用三明治夹芯板理论[25]对蜂窝夹层结构进行等效，保持结构外形尺寸和面板材料参数不变，将蜂窝芯等效为厚度不变的正交各向异性均质实心层，如图4所示。



(a) 等效前 (b) 等效后

图4 三明治夹芯等效法示意图

Fig. 4 Schematic diagram of sandwich core equivalent method

基于力学等效和能量等效原则得到以截面 S 为基础的芯层等效公式[26]：

$$\begin{cases} E_{ex}=\dfrac{4E_{\mathrm{c}}}{\sqrt{3}}\left(\dfrac{bt}{y\Delta l+bl_2}\right)^3\left[1-\dfrac{5}{3}\left(\dfrac{bt}{y\Delta l+bl_2}\right)^2\right] \\ E_{e\theta}=\dfrac{4E_{\mathrm{c}}}{\sqrt{3}}\left(\dfrac{bt}{y\Delta l+bl_2}\right)^3\left[1-3\left(\dfrac{bt}{y\Delta l+bl_2}\right)^2\right] \\ E_{ez}=\dfrac{8\sqrt{3}}{9}E_{\mathrm{c}}\dfrac{bt}{y\Delta l+bl_2} \\ G_{ex\theta}=\dfrac{4\sqrt{3}}{5}E_{\mathrm{c}}\left(\dfrac{bt}{y\Delta l+bl_2}\right)^3 \\ G_{exz}=\dfrac{\sqrt{3}G_{\mathrm{c}}t\left(l_1^2+l_2^2+4l_0^2\right)}{12l_0^3} \\ G_{e\theta z}=\dfrac{\sqrt{3}G_{\mathrm{c}}t\left(l_1^2+l_2^2+4l_0^2\right)}{18l_0^3} \\ \rho_{\mathrm{e}}=\dfrac{8\sqrt{3}(l_1+l_2)\rho_{\mathrm{c}}t}{3\left(l_1^2+l_2^2\right)+12l_0^2} \\ \mu_{ex\theta}=1-\dfrac{8}{3}\left(\dfrac{bt}{y\Delta l+bl_2}\right)^2 \\ \mu_{e\theta x}=1-4\left(\dfrac{bt}{y\Delta l+bl_2}\right)^2 \end{cases} \tag{3}$$

式中：E_{ei} 为蜂窝芯在 i 方向的等效弹性模量；G_{eij} 为蜂窝芯在 i、j 坐标轴间的等效剪切模量；μ_{eij} 为蜂窝芯在 i 方向拉(压)应力引起 j 方向缩(伸)的等效泊松比；ρ_{e} 为蜂窝芯的等效密度；E_{c}、G_{c} 和 ρ_{c} 分别为蜂窝芯材料的弹性模量、剪切模量和密度。

由蜂窝芯的结构特征可知，μ_{ezx} 和 $\mu_{ez\theta}$ 即为蜂窝芯材料的泊松比 μ_{c}，而对正交各向异性材料有

$$\begin{cases} \mu_{exz}=\dfrac{E_{ex}}{E_{ez}}\mu_{ezx}=\dfrac{E_{ex}}{E_{ez}}\mu_{\mathrm{c}} \\ \mu_{e\theta z}=\dfrac{E_{e\theta}}{E_{ez}}\mu_{ez\theta}=\dfrac{E_{e\theta}}{E_{ez}}\mu_{\mathrm{c}} \end{cases} \tag{4}$$

后续将针对3种尺寸铝合金蜂窝圆筒进行稳定性研究，蜂窝圆筒的尺寸参数如表1所示。

表1 3种铝合金蜂窝圆筒的尺寸参数

Table 1 Dimensional parameters of three aluminum alloy honeycomb cylinders

圆筒编号	b/mm	d/mm	R/mm	H/mm	t/mm	l_0/mm
圆筒1	12.6	1.2	692.5	1 100	0.2	6
圆筒2	22.6	1.2	687.5	1 100	0.2	6
圆筒3	37.6	1.2	1 140	1 100	0.2	6

蜂窝圆筒采用铝合金制造，其中面板材料为6A02，蜂窝芯材料为3003，材料的弹性参数取值如表2所示。

表2 面板和蜂窝芯材料的弹性参数

Table 2 Elastic parameters of panel and honeycomb core materials

材料名称	弹性模量/MPa	泊松比	密度/(kg·m^{-3})
6A02	63 000	0.33	2 700
3003	69 000	0.33	2 730

计算过程中，取截面 S 为夹层中曲面，将表1中蜂窝圆筒的几何参数和表2中蜂窝芯的材料参数代入式(2)～式(4)，求得3种蜂窝圆筒的蜂窝芯等效参数，如表3所示。

表3 蜂窝圆筒的蜂窝芯等效参数

Table 3 Equivalent parameters of honeycomb core of honeycomb cylinder

等效参数	圆筒1	圆筒2	圆筒3
E_{ex}/MPa	5.692 6	5.536 1	5.535 0
$E_{e\theta}$/MPa	5.895 5	5.906 2	5.906 2
E_{ez}/MPa	3 533.1	3 526.6	3 526.6
$G_{ex\theta}$/MPa	3.489 6	3.448 2	3.447 9
G_{exz}/MPa	984.811	973.845	973.763
$G_{e\theta z}$/MPa	500.343	501.247	501.254
$\mu_{ex\theta}$	0.997	0.997	0.997
μ_{exz}	5.32×10^{-4}	5.18×10^{-4}	5.18×10^{-4}
$\mu_{e\theta z}$	5.51×10^{-4}	5.53×10^{-4}	5.53×10^{-4}
ρ_{c}/(kg·m^{-3})	139.31	138.67	138.66

2 轴压稳定性理论推导

2.1 基本假设

推导蜂窝圆筒的轴压稳定性方程时，采用如下假设：

1）由于面板很薄，假定面板处于薄膜应力状态。

2）夹芯为硬夹芯，考虑夹芯平行于 $x\theta$ 面的应力分量。

3）忽略面板和夹芯中应力分量 σ_z 和应变分量 ε_z。

4）考虑横向剪切的作用，变形后原本垂直于夹芯中面的直线仍为直线，但不再垂直于变形后的中面。

2.2　几何方程

基于上述基本假设，将夹芯各点的位移表示为

$$\begin{cases} u_2(x,\theta,z)=u_0(x,\theta)+z\alpha(x,\theta) \\ v_2(x,\theta,z)=v_0(x,\theta)+z\beta(x,\theta) \\ w_2(x,\theta,z)=w(x,\theta) \end{cases} \tag{5}$$

式中：u_2、v_2、w_2 分别为夹芯各点沿 x、θ、z 方向的位移；u_0、v_0、w 分别为夹层结构中曲面上各点沿 x、θ、z 方向的位移；α 为 x 方向截面绕 θ 轴转角的平均值；β 为 θ 方向截面绕 x 轴转角的平均值。

考虑夹芯横向剪切的影响，则 α 和 β 表示为

$$\begin{cases} \alpha(x,\theta)=-\dfrac{\partial w(x,\theta)}{\partial x}+\gamma_{xz2} \\ \beta(x,\theta)=\dfrac{v}{R}-\dfrac{1}{R}\cdot\dfrac{\partial w(x,\theta)}{\partial\theta}+\gamma_{\theta z2} \end{cases} \tag{6}$$

式中：γ_{ij} 为 ij 面内的切应变；v 表示结构上任意一点沿 θ 方向的位移。

外面板和内面板的位移为

$$\begin{cases} u_{1,3}=u_0\pm\dfrac{b+d}{2}\alpha \\ v_{1,3}=v_0\pm\dfrac{b+d}{2}\beta \\ w_{1,3}=w \end{cases} \tag{7}$$

式中：u_1、v_1、w_1 和 u_3、v_3、w_3 分别为外面板和内面板各点沿 x、θ、z 方向的位移。

蜂窝圆筒轴向加载属于轴对称问题，所有形变分量和位移分量都与 θ 无关，环向位移 v 也为 0，利用轴对称条件获得夹芯和内、外面板的几何方程。

夹芯的几何方程为

$$\begin{cases} \varepsilon_{x2}=\dfrac{\mathrm{d}u_0}{\mathrm{d}x}+z\dfrac{\mathrm{d}\alpha}{\mathrm{d}x},\varepsilon_{\theta2}=\dfrac{w}{R} \\ \gamma_{x\theta2}=0,\gamma_{xz2}=\dfrac{\mathrm{d}w}{\mathrm{d}x}+\alpha,\gamma_{\theta z2}=0 \end{cases} \tag{8}$$

外面板和内面板的几何方程为

$$\begin{cases} \varepsilon_{x1,3}=\dfrac{\mathrm{d}u_0}{\mathrm{d}x}\pm\dfrac{b+d}{2}\cdot\dfrac{\mathrm{d}\alpha}{\mathrm{d}x} \\ \varepsilon_{\theta1,3}=\dfrac{w}{R} \\ \gamma_{x\theta1,3}=0 \end{cases} \tag{9}$$

式中：ε_i 为 i 方向的正应变；γ_{ij} 为 ij 面内的切应变；下标 k (k=1,2,3) 分别表示外面板、夹芯、内面板。

2.3　物理方程

由式 (8) 可知，夹芯层仅存在 x、θ 方向的正应变和 xz 面内的切应变，假设夹芯层始终处于弹性阶段，利用 Hooke 定律得到等效夹芯层的本构方程：

$$\begin{cases} \sigma_{x2}=\dfrac{E_{ex}(\varepsilon_{x2}+\mu_{e\theta x}\varepsilon_{\theta2})}{1-\mu_{ex\theta}\mu_{e\theta x}} \\ \sigma_{\theta2}=\dfrac{E_{e\theta}(\varepsilon_{\theta2}+\mu_{ex\theta}\varepsilon_{x2})}{1-\mu_{ex\theta}\mu_{e\theta x}} \\ \tau_{xz2}=G_{exz}\gamma_{xz2} \end{cases} \tag{10}$$

式中：σ_i 为 i 方向的正应力；τ_{ij} 为 ij 面内的切应力。

面板材料为铝合金，在轴压过程中易产生塑性变形，将其单向应力-应变曲线用双线性硬化模型表示，如图 5 所示。图中：OA 段为弹性阶段，其斜率为面板材料的弹性模量 E_f，A 点所对应的应变 ε_s 为发生屈服时产生的应变，应力 σ_s 为面板材料的屈服强度，取 σ_s 为 80 MPa。AM 段为塑性阶段，其斜率 E_h 表示面板材料塑性阶段的强化模量，取 E_h 为 800 MPa。假定材料当前的应力-应变状态位于 M 点，定义 M 点的应力为 σ_m，应变为 ε_m，割线模量为 E_m。

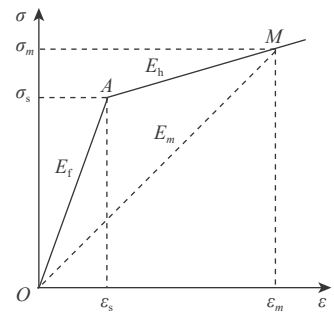


图 5　双线性硬化模型示意图

Fig. 5　Schematic diagram of bilinear hardening model

令面板材料的泊松比为 μ_f，基于单向应力-应变曲线，利用 J_2 形变理论，将内、外面板塑性阶段的本构方程表示为[27]

$$\begin{cases} \sigma_{x1,3}=q_1\varepsilon_{x1,3}+q_2\varepsilon_{\theta1,3} \\ \sigma_{\theta1,3}=q_1\varepsilon_{\theta1,3}+q_2\varepsilon_{x1,3} \\ q_1=\dfrac{4E_mE_\mathrm{f}^2}{[3E_\mathrm{f}-E_m(1-2\mu_\mathrm{f})][E_\mathrm{f}+E_m(1-2\mu_\mathrm{f})]} \\ q_2=\dfrac{2E_mE_\mathrm{f}[E_\mathrm{f}-E_m(1-2\mu_\mathrm{f})]}{[3E_\mathrm{f}-E_m(1-2\mu_\mathrm{f})][E_\mathrm{f}+E_m(1-2\mu_\mathrm{f})]} \end{cases} \tag{11}$$

若式 (11) 中的割线模量 E_m 取弹性模量 E_f，则式 (11) 退化为弹性本构方程，因此，可用式 (11) 表示面板整个弹塑性阶段的本构方程。

2.4　整体稳定性方程

考虑两端简支的铝合金蜂窝圆筒，端面受到轴向压力载荷 P 的作用，载荷均匀分布，令 N 为单位长度上的轴压载荷，则轴压载荷 P 表示为

$$P=2\pi RN \tag{12}$$

令 K 为圆筒总动能，U 为圆筒总应变能，W 为外力做的总功，根据 Hamilton 原理：

$$\delta\int_{T_0}^{T_1}(K-U+W)\mathrm{d}T=0 \tag{13}$$

芯子的应变能 U_2 为

$$U_2=\iiint_V\frac{1}{2}(\sigma_{x2}\varepsilon_{x2}+\sigma_{\theta2}\varepsilon_{\theta2}+\tau_{xz2}\gamma_{xz2})\mathrm{d}V \tag{14}$$

式中：V 表示体积积分。

内、外面板的应变能 $U_{1,3}$ 为

$$U_{1,3}=\iiint_V \frac{1}{2}(\sigma_{x1,3}\varepsilon_{x1,3}+\sigma_{\theta1,3}\varepsilon_{\theta1,3})\mathrm{d}V \tag{15}$$

将芯子和内、外面板的应变能求和，并沿 z 方向积分，得到圆筒总应变能表达式为

$$\begin{cases} U=\iint_S \left[D_1\left(\dfrac{\mathrm{d}u_0}{\mathrm{d}x}\right)^2+D_2\dfrac{w}{R}\cdot\dfrac{\mathrm{d}u_0}{\mathrm{d}x}+D_3\left(\dfrac{\mathrm{d}\alpha}{\mathrm{d}x}\right)^2+ D_4\dfrac{w^2}{R^2}+D_5\left(\dfrac{\mathrm{d}w}{\mathrm{d}x}+\alpha\right)^2\right]\mathrm{d}S \\ D_1=q_1d+\dfrac{E_{\mathrm{e}x}b}{2(1-\mu_{\mathrm{e}x\theta}\mu_{\mathrm{e}\theta x})} \\ D_2=2q_2d+\dfrac{\mu_{\mathrm{e}\theta x}E_{\mathrm{e}x}b+\mu_{\mathrm{e}x\theta}E_{\mathrm{e}\theta}b}{2(1-\mu_{\mathrm{e}x\theta}\mu_{\mathrm{e}\theta x})} \\ D_3=q_1d\dfrac{(b+d)^2}{4}+\dfrac{E_{\mathrm{e}x}b^3}{24(1-\mu_{\mathrm{e}x\theta}\mu_{\mathrm{e}\theta x})} \\ D_4=q_1d+\dfrac{E_{\mathrm{e}\theta}b}{2(1-\mu_{\mathrm{e}x\theta}\mu_{\mathrm{e}\theta x})} \\ D_5=\dfrac{1}{2}G_{\mathrm{e}xz}b \end{cases} \tag{16}$$

将式 (16) 积分号内的表达式记作 F，同时令

$$\begin{cases} w_x=\dfrac{\mathrm{d}w}{\mathrm{d}x} \\ u_{0x}=\dfrac{\mathrm{d}u_0}{\mathrm{d}x} \\ \alpha_x=\dfrac{\mathrm{d}\alpha}{\mathrm{d}x} \end{cases} \tag{17}$$

将式 (17) 代入式 (16) 并进行变分和化简：

$$\delta U=\iint_S \left\{\left[\frac{\partial F}{\partial w}-\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial w_x}\right)\right]\delta w+\left[\frac{\partial F}{\partial \alpha}-\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial \alpha_x}\right)\right]\delta\alpha-\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_{0x}}\right)\delta u_0\right\}\mathrm{d}S \tag{18}$$

圆筒屈曲时外力做的总功为

$$W=\frac{1}{2}\iint_S\left[N\left(\frac{\partial w}{\partial x}\right)^2\right]\mathrm{d}S \tag{19}$$

对外力做的总功进行相同的变分：

$$\delta W=-\iint_S N\frac{\partial^2 w}{\partial x^2}\delta w\mathrm{d}S \tag{20}$$

将轴向加载过程视为准静态，忽略系统的总动能，将式 (18) 和式 (20) 代入式 (13)，可得

$$\iint_S\left\{\left[\frac{\partial F}{\partial w}-\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial w_x}\right)+N\frac{\partial^2 w}{\partial x^2}\right]\delta w+\left[\frac{\partial F}{\partial \alpha}-\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial \alpha_x}\right)\right]\delta\alpha-\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_{0x}}\right)\delta u_0\right\}\mathrm{d}S=0 \tag{21}$$

因此有

$$\begin{cases} \dfrac{\partial F}{\partial w}-\dfrac{\partial}{\partial x}\left(\dfrac{\partial F}{\partial w_x}\right)+N\dfrac{\partial^2 w}{\partial x^2}=0 \\ \dfrac{\partial F}{\partial \alpha}-\dfrac{\partial}{\partial x}\left(\dfrac{\partial F}{\partial \alpha_x}\right)=0 \\ -\dfrac{\partial}{\partial x}\left(\dfrac{\partial F}{\partial u_{0x}}\right)=0 \end{cases} \tag{22}$$

对式 (22) 进行求导和消元，得到铝合金蜂窝圆筒的轴压稳定性方程：

$$-2D_3\frac{\mathrm{d}^4\alpha}{\mathrm{d}x^4}+\left(\frac{2D_3D_4}{D_5R^2}-\frac{D_2^2D_3}{2R^2D_1D_5}\right)\frac{\mathrm{d}^2\alpha}{\mathrm{d}x^2}+\left(\frac{D_2^2}{2R^2D_1}-\frac{2D_4}{R^2}\right)\alpha+N\left(\frac{D_3}{D_5}\cdot\frac{\mathrm{d}^4\alpha}{\mathrm{d}x^4}-\frac{\mathrm{d}^2\alpha}{\mathrm{d}x^2}\right)=0 \tag{23}$$

2.5 临界载荷求解

对于两端简支的铝合金蜂窝圆筒，在 x=0 和 x=L 处有 w=0 和 M=0，M 表示弯矩。不妨令

$$w=A\sin\left(\frac{n\pi x}{H}\right)\qquad n=0,1,2,\cdots \tag{24}$$

根据式 (22) 求导后的形式，可设

$$\alpha=B\cos\left(\frac{n\pi x}{H}\right)\qquad n=0,1,2,\cdots \tag{25}$$

式中：A 和 B 为待定系数。

结合式 (12)、式 (23)～式 (25)，可得

$$P=\frac{-2\pi R}{\dfrac{D_3}{D_5}\cdot\dfrac{n^4\pi^4}{H^4}+\dfrac{n^2\pi^2}{H^2}}\left[\left(\frac{D_2^2}{2R^2D_1}-\frac{2D_4}{R^2}\right)+\left(\frac{2D_3D_4}{D_5R^2}-\frac{D_2^2D_3}{2R^2D_1D_5}\right)\left(-\frac{n^2\pi^2}{H^2}\right)-2D_3\frac{n^4\pi^4}{H^4}\right] \tag{26}$$

将式 (26) 求出的轴压载荷记作 P_{a}，P_{a} 的表达式中含有变量 E_m，由图 5 可知，当割线模量为 E_m 时，对应的应力和应变分别为 σ_m 和 ε_m，据此可将圆筒所受的轴压载荷表示为

$$P=4\pi Rd\sigma_m+2\pi RbE_{\mathrm{e}x}\varepsilon_m \tag{27}$$

将式 (27) 求出的轴压载荷记作 P_{b}，当圆筒处于临界屈曲状态时，E_m 的取值应使 P_{a} 和 P_{b} 的数值相等。求解过程中，取 E_m=E_{f} 并代入式 (26)，求得临界载荷的弹性解 P_{E}，将图 5 中 A 点对应的应力、应变数值代入式 (27)，求得圆筒弹性阶段的最大载荷 P_{Emax}。若 P_{E}<P_{Emax}，圆筒发生弹性屈曲，P_{E} 即为铝合金蜂窝圆筒的实际临界载荷。若 P_{E}>P_{Emax}，圆筒发生塑性屈曲，取 E_m 的初始值为 E_{f}，每次迭代计算前使当前 E_m 值减小 ΔE_m，经过 i 次迭代后 E_m 的取值变为 $E_{\mathrm{f}}-i\Delta E_m$。每次迭代过程中，将 E_m 的数值代入式 (26) 和式 (27) 分别计算 P_{a} 和 P_{b}，直至 P_{a} 和 P_{b} 的相对误差在 1% 以内，取当前的 P_{a} 值为铝合金蜂窝圆筒的临界载荷。计算过程中，若 ΔE_m 取值过大，可能使 E_m 的数值无法接近精确解，导致计算无法收敛，若 ΔE_m 取值过小，又会使计算效率变低。综合考虑上述问题后，取 ΔE_m 为 1 MPa。令 M 点所对应的圆筒端面轴向位移量为 D_m，而 D_m 与 ε_m 的关系为

$$\varepsilon_m=\frac{D_m}{H} \tag{28}$$

σ_m 与 ε_m 的关系为

$$\begin{cases}\sigma_m = E_{\mathrm{f}}\varepsilon_m & \varepsilon_m \leqslant \varepsilon_{\mathrm{s}} \\ \sigma_m = \sigma_{\mathrm{s}} + E_{\mathrm{h}}(\varepsilon_m - \varepsilon_{\mathrm{s}}) & \varepsilon_m > \varepsilon_{\mathrm{s}}\end{cases} \tag{29}$$

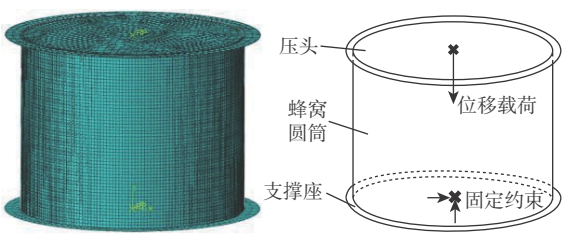
将式 (28) 和式 (29) 代入式 (27)，求出圆筒轴压载荷与 D_{m} 的关系：

$$\begin{cases} P = 2\pi R\dfrac{D_m}{H}(2dE_{\mathrm{f}} + bE_{ex}) & \dfrac{D_m}{H} \leqslant \varepsilon_{\mathrm{s}} \\ P = 4\pi Rd\left[\sigma_{\mathrm{s}} + E_{\mathrm{h}}\left(\dfrac{D_m}{H} - \varepsilon_{\mathrm{s}}\right)\right] + 2\pi RbE_{\mathrm{ex}}\dfrac{D_m}{H} & \dfrac{D_m}{H} > \varepsilon_{\mathrm{s}} \end{cases} \tag{30}$$

迭代求出临界载荷和 M 点的应力、应变数值后，利用式 (30) 绘制由初始状态到发生屈曲的过程中圆筒轴压载荷随端部位移的变化情况。

3　有限元分析

表 1 中 3 种铝合金蜂窝圆筒的轴压仿真分析均可由图 6 所示的有限元分析模型实现，该模型由铝合金蜂窝圆筒及两侧的压头和支撑座组成，利用三明治夹芯板理论对蜂窝圆筒进行等效，即芯子采用 C3D8R 实体单元构建，内、外面板采用 S4R 单元构建。芯子与内、外面板采用绑定接触，等效芯子在厚度方向划分 4 层网格。将压头和支撑座视作刚体，参考钢和铝之间的摩擦系数，取蜂窝圆筒端部与压头、支撑座接触面的摩擦系数为 0.17[28]。对支撑座施加固定约束，在压头处施加 100 mm 位移载荷，利用弧长法求解，提取加载过程中支撑座的最大支反力，即为圆筒仿真模型的轴压临界载荷。



(a) 网格划分　　(b) 施加载荷

图 6　蜂窝圆筒轴压有限元分析模型

Fig. 6　Finite element analysis model of honeycomb cylinder under axial compression

4　轴压实验

针对表 1 中 3 种尺寸铝合金蜂窝圆筒进行轴压实验，每种蜂窝圆筒制造 2 个相同的平行试样进行 2 次重复实验。依次将蜂窝圆筒安装于压力机上，并利用最大载荷为 6 300 kN 的 YW27-630-3800-2500 型压力机进行加载，如图 7 所示。实验过程中，蜂窝圆筒上方压头沿圆筒轴向施加载荷，加载速率为 5 mm/min，记录压头载荷随压头位移变化曲线，直至曲线开始下降，取实验平均曲线的峰值载荷为圆筒的轴压临界载荷。



(a) 实验设备　　(b) 蜂窝圆筒安装情况

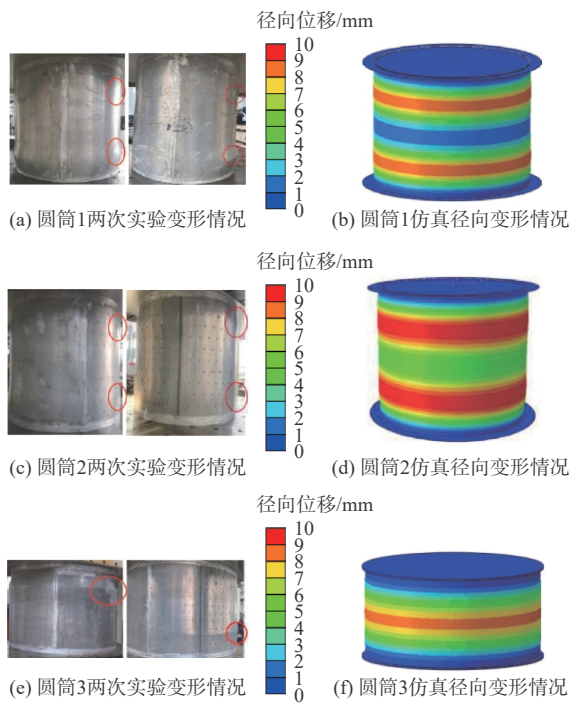
图 7　蜂窝圆筒轴压实验现场

Fig. 7　Honeycomb cylinder axial compression test site

5　结果分析

5.1　临界载荷结果对比

提取蜂窝圆筒仿真模型发生屈曲时的径向变形云图，并与轴压实验发生屈曲时的变形情况进行对比，如图 8 所示。可知，轴压实验和仿真分析的屈曲模式具有较好的一致性，即圆筒 1 和圆筒 2 均以两侧端部附近出现 2 个鼓包的形式发生屈曲，圆筒 3 以 1 个鼓包的形式发生屈曲。但实验与仿真的区别在于：轴压实验中，圆筒 1 和圆筒 2 的 2 个鼓包并不对称，圆筒 3 的鼓包位置不位于圆柱壳的中部，认为产生差异的原因是实际蜂窝圆柱壳结构



(a) 圆筒1两次实验变形情况　　(b) 圆筒1仿真径向变形情况

(c) 圆筒2两次实验变形情况　　(d) 圆筒2仿真径向变形情况

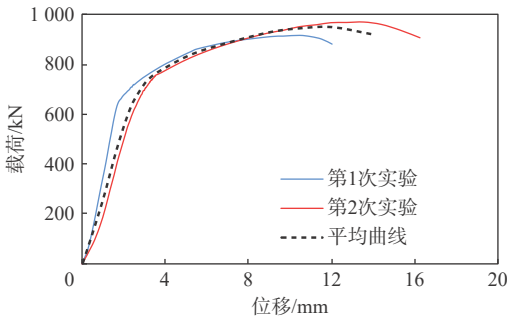
(e) 圆筒3两次实验变形情况　　(f) 圆筒3仿真径向变形情况

图 8　蜂窝圆筒轴压屈曲模式示意图

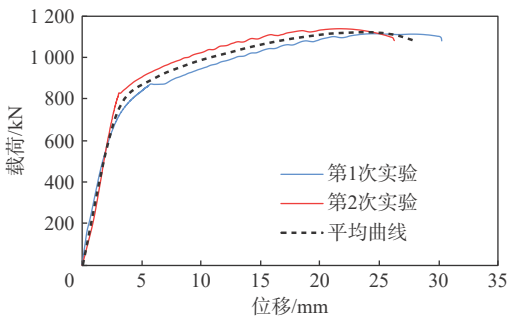
Fig. 8　Schematic diagram of honeycomb cylinder buckling mode under axial compression

存在缺陷。

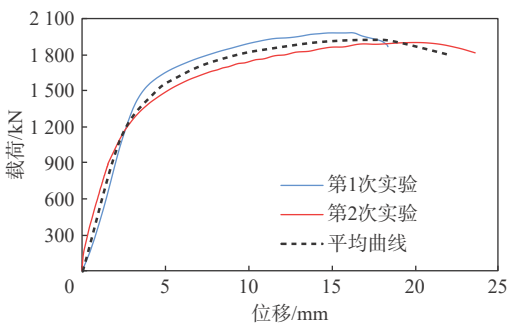
绘制3种蜂窝圆筒的轴压实验载荷-位移曲线，如图9所示。每种圆筒进行2次实验，得到2条载荷-位移实验曲线，由于圆筒缺陷等因素的差异，2条实验曲线的峰值载荷及峰值载荷对应的位移均不相同，无法直接通过代数求平均值的方式获得平均曲线，因此，在曲线图中取2条实验曲线所夹的1条曲线作为实验平均曲线。



(a) 圆筒1轴压载荷-位移曲线



(b) 圆筒2轴压载荷-移曲线



(c) 圆筒3轴压载荷-位移曲线

图9 蜂窝圆筒轴压实验载荷-位移曲线

Fig. 9 Load-displacement curves of honeycomb cylindrical axial compression test

将实验、有限元仿真和理论计算得到的压头载荷-位移曲线进行汇总，如图10所示。对于圆筒1，实验、有限元仿真和理论计算曲线分别在压头位移量12.1 mm、16.1 mm、23.4 mm时达到峰值载荷，数值分别为974.6 kN、921.4 kN、1 013.1 kN；对于圆筒2，实验、有限元仿真和理论计算曲线分别在压头位移量24.3 mm、38.2 mm、38.2 mm时达到峰值载荷，数值分别为1 132.8 kN、1 114.5 kN、1 144.3 kN；对于圆筒3，实验、有限元仿真和理论计算曲线分

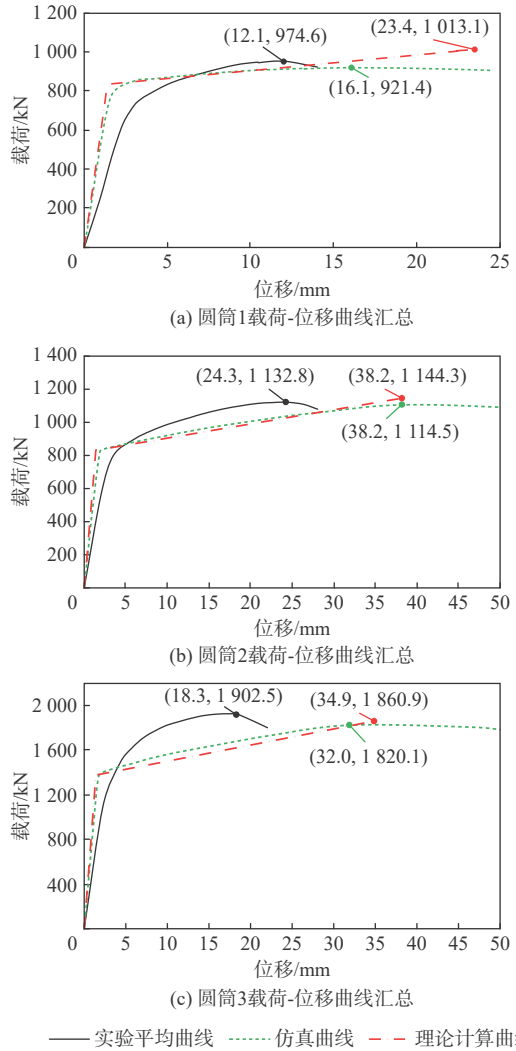


图10 蜂窝圆筒载荷-位移曲线汇总

Fig. 10 Summary of honeycomb cylinder load-displacement curves

别在压头位移量18.3 mm、32.0 mm、34.9 mm时达到峰值载荷，数值分别为1 902.5 kN、1 820.1 kN、1 860.9 kN。

由图10可知，当压头位移较小时，蜂窝圆筒处于弹性变形阶段，载荷-位移曲线斜率较大，随后进入塑性变形阶段，载荷-位移曲线斜率明显降低，直至载荷增加至峰值载荷后结构发生塑性屈曲，载荷开始下降。从图中还能看出，仿真曲线和理论计算曲线一致性较好，但与实验曲线相比存在一定差异，主要体现在实验曲线弹性阶段的斜率更小而塑性阶段的斜率更大。认为产生上述差异的原因是：①实际蜂窝芯结构多孔、非均质的特点降低了其与面板、压头之间的连接刚度，弹性阶段斜率较小；②蜂窝圆筒内部的结构缺陷使部分结构进入塑性变形阶段时仍有另一部分处于弹性变形阶段，实验曲线塑性阶段的斜率更高。

将铝合金蜂窝圆筒轴压临界载荷的理论计算

结果、仿真结果和实验结果进行汇总，如图 11 所示。将理论计算和仿真结果分别与实验结果进行对比，对于圆筒 1、圆筒 2、圆筒 3，理论计算结果与实验结果的相对误差分别为 3.95%、1.02%、2.19%，仿真结果与实验结果的相对误差分别为 5.46%、1.62%、4.33%。对于上述 3 种蜂窝圆筒，采用理论计算方法预测的平均相对误差为 2.39%，而采用仿真分析方法预测的平均相对误差为 3.80%。综上所述，认为利用理论公式计算和仿真分析手段均能准确预测铝合金蜂窝圆筒的实际临界载荷，相较之下，理论计算方法能够获得更精确的结果。

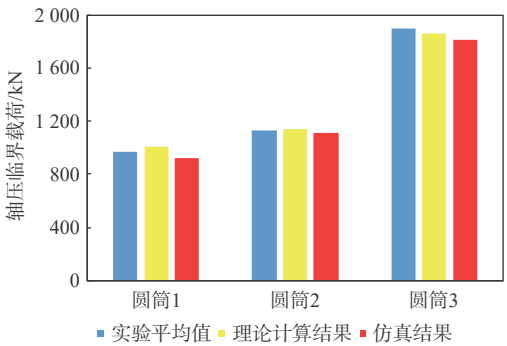


图 11　蜂窝圆筒轴压临界载荷结果汇总

Fig. 11　Summary of honeycomb cylinder axial compression critical load results

求解所用的计算机 CPU 为 Intel Core i5-9300H，通过有限元结果文件读取仿真分析用时并求得 3 种蜂窝圆筒有限元分析的平均时间约 35 min，通过理论公式迭代程序中插入的计时器读取理论计算用时并求得 3 种蜂窝圆筒理论计算所需平均时间为 1.43 s。对比可知，理论公式求解临界载荷的速度远高于仿真分析。

5.2　尺寸参数对临界载荷的影响

为比较 4 种尺寸参数对临界载荷的影响程度，以蜂窝圆筒质量变化率为衡量标准，依次改变 4 种尺寸参数使蜂窝圆筒产生相同的质量变化率，利用求解速度更快的理论公式分析芯子厚度 b、面板厚度 d、圆筒半径 R、圆筒高度 H 对临界载荷的影响。令 ρ_f 为面板材料的密度，则蜂窝圆筒的质量 m 表示为

$$m=\rho_f 2\pi RH2d+\rho_e 2\pi RbH \tag{31}$$

分别使表 1 中 3 种蜂窝圆筒的质量改变±5%与±10%，通过式 (31) 计算不同质量条件下各尺寸参数的数值，代入理论公式计算相应的临界载荷，整理得到几何参数变化对临界载荷的影响，如图 12 所示。

计算各尺寸参数变化使质量变化率由−10%～+10% 时临界载荷的变化率，如表 4 所示。可知，在

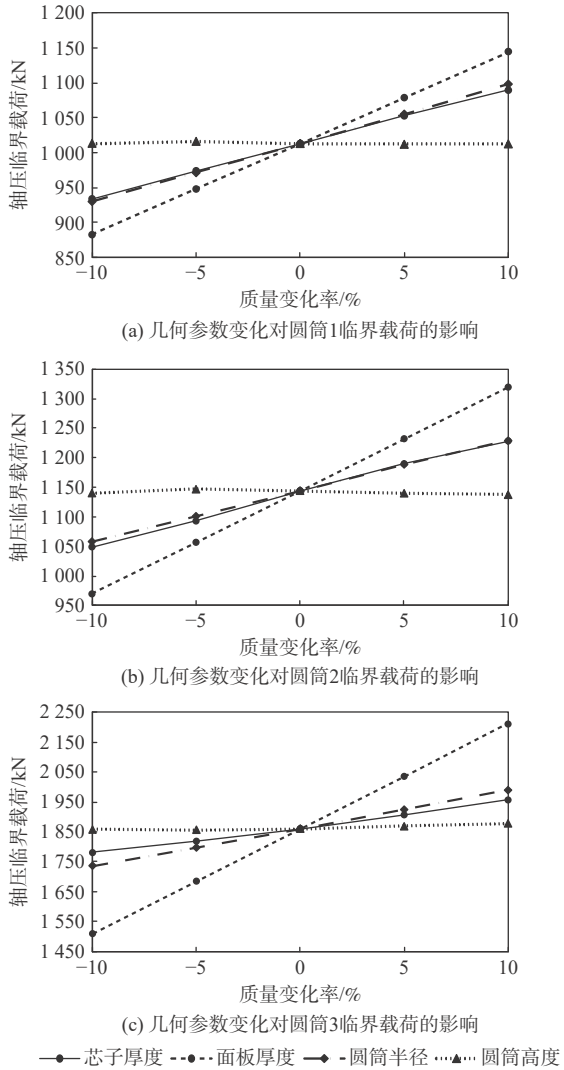


图 12　几何参数变化对 3 种蜂窝圆筒临界载荷的影响

Fig. 12　Influence of geometric parameters on critical loads of three kinds of honeycomb cylinders

尺寸参数变化引起相等质量变化率的条件下，面板厚度、圆筒半径、芯子厚度和圆筒高度使临界载荷产生的平均变化率依次降低，因此，认为各参数对临界载荷的影响程度也按上述顺序依次降低。其中，圆筒高度使临界载荷产生的平均变化率接近 0，说明该参数变化基本不会使临界载荷发生明显变化。

将各尺寸参数与临界载荷的线性相关程度用相关系数 r 衡量，令 q 为圆筒的质量变化率，P_c 为轴压临界载荷，下标 i 从 1～5 表示质量变化率从 −10%～+10% 所对应的数据，上横线表示参数的平均值，则相关系数 r 表示为

$$r=\frac{\sum_{i=1}^{5}(q_i-\bar{q})\left(P_{ci}-\bar{P}_c\right)}{\sqrt{\sum_{i=1}^{5}(q_i-\bar{q})^2}\sqrt{\sum_{i=1}^{5}\left(P_{ci}-\bar{P}_c\right)^2}} \tag{32}$$

表4 质量变化率由−10%～+10%时临界载荷的变化率

Table 4 Critical load change rate when mass change rate is from −10% to +10%

参数	临界载荷变化率/%			平均值/%
	圆筒1	圆筒2	圆筒3	
芯子厚度	+16.7	+17.1	+9.88	+14.6
面板厚度	+29.8	+36.1	+46.6	+37.5
圆筒半径	+18.1	+16.1	+14.6	+16.3
圆筒高度	−0.06	−0.25	+1.12	+0.27

利用式(32)求出3种蜂窝圆筒尺寸参数与临界载荷的相关系数，如表5所示。可知，芯子厚度、面板厚度、圆筒半径与临界载荷的平均相关系数接近1，说明三者与临界载荷近似为线性正相关关系。圆筒高度与临界载荷的平均相关系数绝对值小于0.4，同时相关系数既有正值也有负值，而前述分析表明，圆筒高度基本不使临界载荷发生明显变化，因此，认为改变圆筒高度会使临界载荷在极小区间内上下波动，两者相关性较弱。

表5 蜂窝圆筒尺寸参数与临界载荷的相关系数

Table 5 Correlation coefficient between dimension parameters and critical load of honeycomb cylinder

参数	相关系数			平均值
	圆筒1	圆筒2	圆筒3	
芯子厚度	0.999 9	0.999 0	0.998 6	0.999 2
面板厚度	0.999 9	0.999 9	0.999 9	0.999 9
圆筒半径	0.999 9	0.999 9	0.999 9	0.999 9
圆筒高度	−0.545 7	0.648 5	0.926 8	0.343 2

6 结 论

本文对轴压铝合金蜂窝圆柱壳稳定性问题进行了理论研究、仿真分析和实验验证，结论如下：

1）本文提出的稳定性理论计算公式和有限元仿真模型均能准确预测铝合金蜂窝圆筒轴压实验的临界载荷，相较之下，理论公式的准确性更高、求解速度更快。

2）当尺寸参数变化使圆筒质量变化率为±10%时，面板厚度、圆筒半径、芯子厚度、圆筒高度对临界载荷的影响程度依次降低，而圆筒高度的改变基本不会使临界载荷发生明显变化。

3）当尺寸参数变化使圆筒质量变化率为±10%时，芯子厚度、面板厚度、圆筒半径与临界载荷近似为线性正相关，而圆筒高度与临界载荷的相关性较弱。

虽然理论公式和仿真模型能够准确预测轴压实验临界载荷的数值，但载荷-位移曲线仍与实验曲线有一定差异，后续仍需要考虑圆柱壳结构缺陷等因素完善理论计算公式和仿真模型。

参考文献（References）

[1] HUANG S Z, JIN Z Y, CHEN Y. Underwater blast resistance of double cylindrical shells with circular tube stiffeners[J]. Ocean Engineering, 2021, 238: 109691.

[2] TENG J G, HU Y M. Behaviour of FRP-jacketed circular steel tubes and cylindrical shells under axial compression[J]. Construction and Building Materials, 2007, 21(4): 827-838.

[3] LAN X K, FENG S S, HUANG Q, et al. A comparative study of blast resistance of cylindrical sandwich panels with aluminum foam and auxetic honeycomb cores[J]. Aerospace Science and Technology, 2019, 87: 37-47.

[4] CAI L C, ZHANG D Y, ZHOU S H, et al. Investigation on mechanical properties and equivalent model of aluminum honeycomb sandwich panels[J]. Journal of Materials Engineering and Performance, 2018, 27(12): 6585-6596.

[5] 骆洪志, 郭彦明, 吴会强. 直径五米大型箭体结构设计与优化[J]. 深空探测学报 (中英文), 2021, 8(4): 380-388.
LUO H Z, GUO Y M, WU H Q. Structural design and optimization of *Φ*5 m diameter large arrow body[J]. Journal of Deep Space Exploration, 2021, 8(4): 380-388(in Chinese).

[6] LI Q, YANG D Q, MAO X. Pressure-resistant cylindrical shell structures comprising graded hybrid zero Poisson's ratio metamaterials with designated band gap characteristics[J]. Marine Structures, 2022, 84: 103221.

[7] TIWARI G, KHAIRE N. Ballistic performance and energy dissipation characteristics of cylindrical honeycomb sandwich structure[J]. International Journal of Impact Engineering, 2022, 160: 104065.

[8] ALMROTH B O, BURNS A B, PITTNER E V. Design criteria for axially loaded cylindrical shells[J]. Journal of Spacecraft and Rockets, 1970, 7(6): 714-720.

[9] 祝恩淳, MANDAL P, CALLADINE C R. 轴压圆柱薄壳的屈曲分析[J]. 土木工程学报, 2001, 34(3): 18-22.
ZHU E C, MANDAL P, CALLADINE C R. Analysis of buckling of thin cylindrical shells under axial compression[J]. China Civil Engineering Journal, 2001, 34(3): 18-22(in Chinese).

[10] ANSARI Q M, TRINH L C, ZUCCO G, et al. Effect of elastic support on the linear buckling response of quasi-isotropic cylindrical shells under axial compression[J]. Engineering Structures, 2021, 244: 112796.

[11] LAI A D, JIA J F, ZHOU Z H, et al. Homotopic analysis for post-buckling of cylindrical shells with local thickness defects[J]. Acta Astronautica, 2022, 193: 44-55.

[12] SHAMASS R, ALFANO G, GUARRACINO F. A numerical investigation into the plastic buckling paradox for circular cylindrical shells under axial compression[J]. Engineering Structures, 2014, 75: 429-447.

[13] LEGENDRE J, LE GROGNEC P, DOUDARD C, et al. Analytical, numerical and experimental study of the plastic buckling behavior of thick cylindrical tubes under axial compression[J]. International Journal of Mechanical Sciences, 2019, 156: 494-505.

[14] EVKIN A, LYKHACHOVA O. Energy barrier method for estima-

tion of design buckling load of axially compressed elasto-plastic cylindrical shells[J]. Thin-Walled Structures, 2021, 161: 107454.

[15] RAJABIEHFARD R, DARVIZEH A, ALITAVOLI M, et al. Experimental and numerical investigation of dynamic plastic behavior of tube with different thickness distribution under axial impact[J]. Thin-Walled Structures, 2016, 109: 174-184.

[16] 张靖华, 郑武. 功能梯度圆柱壳的弹塑性屈曲[J]. 兰州理工大学学报, 2021, 47(5): 167-172.
ZHANG J H, ZHENG W. Elastoplastic buckling of functionally graded cylindrical shells[J]. Journal of Lanzhou University of Technology, 2021, 47(5): 167-172(in Chinese).

[17] HOU X H, ZHOU S Q, CHENG Z W, et al. Buckling of regular and auxetic honeycombs under a general macroscopic stress state in symplectic system[J]. Applied Mathematical Modelling, 2022, 109: 318-340.

[18] ZHANG C W, EYVAZIAN A, ALKHEDHER M, et al. Modified couple stress theory application to analyze mechanical buckling behavior of three-layer rectangular microplates with honeycomb core and piezoelectric face sheets[J]. Composite Structures, 2022, 292: 115582.

[19] 赵杰, 张娟, 蒋晓琴, 等. 不同压缩工况下焊接缺陷对钎焊铝蜂窝板承载能力的影响[J]. 机械工程材料, 2020, 44(12): 97-102.
ZHAO J, ZHANG J, JIANG X Q, et al. Influence of welding defects on bearing capacity of brazed aluminum honeycomb panel under different compression conditions[J]. Materials for Mechanical Engineering, 2020, 44(12): 97-102(in Chinese).

[20] 王宝芹, 王沫楠, 刘长喜. 基于多尺度方法的蜂窝夹层复合材料结构轴向压缩稳定性[J]. 复合材料学报, 2020, 37(3): 601-608.
WANG B Q, WANG M N, LIU C X. Stability of honeycomb sandwich composite structure under axial compression based on multi-scale method[J]. Acta Materiae Compositae Sinica, 2020, 37(3): 601-608(in Chinese).

[21] ZHANG Z J, WANG Y J, HUANG L, et al. Mechanical behaviors and failure modes of sandwich cylinders with square honeycomb cores under axial compression[J]. Thin-Walled Structures, 2022, 172: 108868.

[22] PAN D R, CHEN L, ZHAO Q L, et al. Local buckling theoretical calculation method of the FRP foam sandwich cylinder under axial compression[J]. Composite Structures, 2020, 246: 112371.

[23] 王颖坚, 徐颖. 复合材料夹层板的塑性屈曲[J]. 北京大学学报 (自然科学版), 1993, 29(1): 48-56.
WANG Y J, XU Y. Plastic buckling of composite sandwich plates[J]. Acta Scicentiarum Naturalum Universitis Pekinesis, 1993, 29(1): 48-56(in Chinese).

[24] 刘强, 黄争鸣. 金属蜂窝材料的弹塑性屈曲临界应力值[J]. 力学季刊, 2008, 29(4): 515-520.
LIU Q, HUANG Z W. Elasto-plastic buckling critical stress of metal honeycomb[J]. Chinese Quarterly of Mechanics, 2008, 29(4): 515-520(in Chinese).

[25] 李贤冰, 温激鸿, 郁殿龙, 等. 蜂窝夹层板力学等效方法对比研究[J]. 玻璃钢/复合材料, 2012(增刊 1): 11-15.
LI X B, WEN J H, YU D L, et al. The comparative study of equivalent mechanical methods on honeycomb sandwich plate[J]. Composites Science and Engineering, 2012(Sup 1): 11-15(in Chinese).

[26] 沈元星, 侯文彬. 基于双曲面蜂窝夹层结构等效模型的研究及数值分析[J]. 计算力学学报, 2022, 39(2): 209-215.
SHEN Y X, HOU W B. Equivalent models research and numerical analysis based on double curved honeycomb sandwich structure[J]. Chinese Journal of Computational Mechanics, 2022, 39(2): 209-215 (in Chinese).

[27] 陈明祥. 弹塑性力学[M]. 北京: 科学出版社, 2007: 278-281.
CHEN M X. Elasticity and plasticity[M]. Beijing: Science Press, 2007: 278-281(in Chinese).

[28] 闻邦椿. 机械设计手册 (第 1 卷)[M]. 5 版. 北京: 机械工业出版社, 2010: 25-26.
WEN B C. Machine design handbook (Vol. 1)[M]. 5th ed. Beijing: China Machine Press, 2010: 25-26(in Chinese).

Study on global stability of aluminum alloy honeycomb cylinder under axial compression

WEI Jiarui[1, 2, 3], WU Qiong[1, 2, *], WU Wei[4], LI Qunzhi[5], SHI Hongwei[5], ZHANG Yu[1, 2]

(1. State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China;
2. Jingdezhen Branch of Jiangxi Research Institute of Beihang University, Jingdezhen 333000, China;
3. The 9th Overall Design Department of China Aerospace Science and Industry Corporation, Wuhan 430040, China;
4. Beijing Aerospace Technology Institute, Beijing 100074, China;
5. Beijing Institute of Spacecraft System Engineering, Beijing 100094, China)

Abstract: Metal honeycomb cylinder structures frequently encounter a decrease in load-carrying capacity due to axial compression buckling. Extensive research has been conducted on their elastic buckling behavior. However, there are few theoretical studies on their plastic buckling behavior. To address this issue, a theoretical formula was derived for the elastic and plastic buckling critical load of the aluminum alloy honeycomb cylinder under axial compression, drawing upon the equivalent model of the honeycomb cylinder and the Hamilton principle. The simulation analysis and experimentations on the above problem were carried out. Finally, the influence of the honeycomb cylinder dimensional parameters on the critical load was discussed using the theoretical formula. The results show that both the theoretical formula and the simulation model can accurately predict the critical load of the aluminum alloy honeycomb cylinder. In comparison, the theoretical formula is calculated faster. When the cylinder mass changes within the range of ±10%, the influence degree of panel thickness, cylinder radius, core thickness and cylinder height on the critical load decreases successively. The honeycomb cylinder's critical load exhibits a minor association with the cylinder height, but it exhibits an approximately linear positive correlation with the cylinder radius, core thickness, and panel thickness.

Keywords: honeycomb cylinder; global stability; critical load; elasto-plastic; Hamilton principle; finite element

Received: 2023-03-23; **Accepted**: 2023-07-07; **Published Online**: 2023-08-03 09:37

URL: link.cnki.net/urlid/11.2625.V.20230802.1344.004

Foundation items: National Natural Science Foundation of China (51875024); Civil Space Technology Pre-Research Program (D020304)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0137

基于深浅层特征融合的孪生网络目标跟踪算法

王子康，姚文进*，薛尚捷，司婷波
(南京理工大学 机械工程学院，南京 210094)

摘 要：深度学习相关的目标跟踪算法在利用深浅层特征融合时，未考虑分类分支与回归分支的差异性，两分支均使用同一融合特征，不能同时满足各自分支的不同任务要求。依据分类分支与回归分支的不同任务要求与深浅层特征之间的关系，提出了一种新的特征融合方式用于视觉目标跟踪。将骨干网络中不同特征层的通道数按比例进行微调，分别形成适合分类分支与回归分支的融合特征。为验证所提特征融合方式的有效性，在基于SiamCAR算法的基础上进行优化，改变特征提取与融合方式，在UAV123、GOT-10K、LaSOT数据集上提高了2%～3%的精度。实验结果证明：所提特征融合方式是有效的，同时框架整体以75帧/s的实时运行速率实现了良好的跟踪性能。

关 键 词：目标跟踪；孪生网络；分类；回归；深浅层特征融合；分类与IoU联合训练
中图分类号：TP391.4
文献标志码：A **文章编号**：1001-5965(2025)03-0973-12

目标跟踪是基本的视觉任务。在仅提供任意目标第1帧位置的情况下，推断视频序列后续帧中的对应目标位置。目标跟踪在军用、民用和安全等领域都有广泛应用，如视频监控、智慧医疗、人机交互、自动驾驶、精确制导、巡飞弹追踪等。

目标跟踪算法可分为传统方法、基于相关滤波的方法、基于深度学习的方法。深度学习算法中，基于孪生网络的目标跟踪算法在保证实时性与准确度平衡的同时，还能实现较高的准确度，因而受到了研究者的广泛关注。SINT[1]使用孪生网络进行目标跟踪，跟踪过程中不对模板进行更新。SiamFC[2]使用一种全卷积孪生网络，利用孪生网络学习目标对象与候选图像之间的相似度，将跟踪问题转化为完整图像上的目标搜索问题。因孪生网络的结构简单且跟踪性能较好，大量孪生系列跟踪器被提出，实现了令人满意的性能。其中，SiamRPN[3]的孪生区域候选网络是具有代表性的，其引入的区域候选网络由一个用于前后景估计的分类网络和一个用于回归的锚框微调网络组成。SiamRPN++[4]在SiamRPN基础上引入了更深层的ResNet50[5]网络，替换了SiamRPN中的AlexNet[6]网络，使用均匀分布的中心采样方式，缓解了因破坏严格平移不变性所带来的影响。SiamCAR[7]、Ocean[8]等算法将无锚框方法用于目标跟踪中，取代了区域候选的方法。无锚框方法降低了网络的计算量，也提高了回归精度。本文算法使用修改后的ResNet50作为骨干网络来提取特征，同时使用无锚框方法进行逐像素地端到端分类与回归。

本文算法基于视觉，视觉对应的图像特征中，2个重要性质是不变性和等变性。目标跟踪中，目标分类需要不变特征，任务目标是学习语义信息；目标定位需要等变特征，任务目标是确定位置和目标尺度变化。对跟踪器来说，同时学习不变性和等变性至关重要。随着卷积神经网络(convolutional neural network, CNN)的加深，物体会逐渐丢失部分和整体的关系，因此，深层网络得到的主要是语义特征(不变性特征)，浅层网络得到的主要是空间特征(等变性特征)。研究者将特征融合应用在目标跟踪

收稿日期：2023-03-23；录用日期：2023-05-19；网络出版时间：2023-06-09 16:07
网络出版地址：link.cnki.net/urlid/11.2625.V.20230608.1150.001
*通信作者. E-mail：njyaowj@njust.edu.cn

引用格式：王子康，姚文进，薛尚捷，等. 基于深浅层特征融合的孪生网络目标跟踪算法[J]. 北京航空航天大学学报，2025，51(3)：973-984. WANG Z K，YAO W J，XUE S J，et al. Siamese network object tracking algorithm based on deep and shallow feature fusion[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：973-984 (in Chinese).

领域，通过特征融合充分利用CNN的深浅层特征。

目标跟踪领域中，SiamRPN++[4]将ResNet50中的3层网络特征进行层间聚合，利用深浅层特征的融合来提高跟踪器性能。SiamCAR利用后3层的特征，按1:1:1的通道数比例进行特征拼接，从而得到融合特征。蒲磊等[9]提出基于高层语义嵌入的孪生网络跟踪算法，有效融合了深层特征和浅层特征，达到了优化浅层特征的目的。以上算法均简单利用了深浅层特征的融合，忽略了分类与回归的不同任务属性。分类与回归对融合特征所涵盖的空间特征与语义特征的要求不同，因此，本文通过调整输入分类与回归分支的融合特征来优化网络结构。

特征融合的方法分为早融合与晚融合。使用早融合的有FCN[10]、Inside-Outside Net[11]、HyperNet[12]等，使用晚融合的有FPN[13]、DenseNet[14]等。早融合中的基本融合方法有逐点相加和拼接[15]，晚融合中包含各种复杂的特征融合方式。逐点相加属于拼接的特殊形式，其他特殊融合方式相较于拼接更为复杂。为在结构简单的情况下验证本文算法的有效性，选取拼接的特征融合方式。

随着质量评估分支在检测器中不断取得良好的增益效果，跟踪器的研究者也开始从中借鉴经验，来提高目标跟踪的定位精度，同时减弱分类分支与回归分支在训练和推理阶段的弱相关性产生的消极影响。FCOS[16]中，检测器通过添加额外的质量评估分支（IoU分支[17]、中心度分支）显著提升了检测精度。SiamCAR选择中心度分支作为质量评估分支，以提高跟踪器的定位精度。Li等[18]提出了分类与IoU联合训练的分类分支，与IoU分支和中心度分支进行比较，验证得到分类与IoU联合训练分支更具有优越性，并根据分支特点，设计了广义焦点损失(generalized focal loss, GFL)函数（含质量焦点损失（quality focal loss, QFL）函数和分布焦点损失（distribution focal loss, DFL）函数）。本文将用于目标检测的分类与IoU联合训练的分类分支进行微调后，迁移到目标跟踪框架中，进一步提高目标跟踪的性能。

本文的主要贡献如下：

1）提出了一种新的特征融合方法，该方法根据分类与回归的特性，对不同特征层的通道进行调整，利用深度互相关和特征拼接的方式进行特征融合，更充分地利用了深浅层特征所涵盖的空间特征与语义特征。

2）将目标检测领域中提出的分类与IoU联合训练的分类分支迁移到目标跟踪领域，进一步提升了目标跟踪的性能。

3）提出了一个基于深浅层特征融合的无锚框全卷积孪生网络框架，该框架相对简单且有较高的性能，同时有较好的跟踪实时性。

1 相关工作

因为本文工作基于CNN，所以主要介绍CNN方面的相关工作。

跟踪器的研究人员致力于从不同方面提高跟踪精度与实时速度，如特征提取[19]、模板更新[20]、分类器设计[21]、边界框回归[22]等。其中，模板更新提高了跟踪器精度，但严重破坏了跟踪实时性。实际应用中，模板更新不能真正实现实时更新，因此，本文工作中未引入模板更新的方法。

得益于深度学习的发展，深度卷积特征被广泛应用。研究表明，基于孪生网络的在线训练与离线跟踪方法实现了准确度和实时性之间的平衡。作为相关工作，SiamFC首次使用一种全卷积孪生网络架构，利用孪生网络学习目标对象与候选图像之间的相似度，将跟踪问题视为在完整图像上搜索目标的问题。DSiam[23]提出了一个动态孪生网络，在损害速度的情况下，实现了较好的跟踪准确度。SA-Siam[24]建立了一个双支路孪生网络，包括外观分支与语义分支2个分支，分别训练，用来保证特征的异质性，从而得到较高的跟踪准确度。为解决尺度变化问题，这些孪生网络都采用了多尺度搜索，并且导致了耗时的问题，破坏了实时性。

受目标检测中的区域候选网络[25]所启发，SiamRPN在孪生网络输出结果上进行候选区域提取，为候选区域联合训练一个分类分支与回归分支。SiamRPN++使用ResNet50作为骨干网络来优化网络架构，在进行模型训练时，在搜索区域上随机移动训练目标的位置来消除中心偏置的消极影响。

为了提取候选区域，锚框被引入到基于区域候选网络的跟踪器中。锚框不仅充分利用了深度特征，也避免了重复计算，可以加快跟踪进程。但跟踪器对锚框的相关超参数非常敏感，需要仔细地调参和丰富的经验来获得理想性能。此外，锚框的尺寸和比例是固定的，还有启发式调参，相关跟踪器在大的形状改变和姿态变化下会遇到跟踪困难。为解决锚框问题，一系列无锚框的跟踪器框架被提出，如ECO[26]、SiamCAR、Ocean等。这些无锚框跟踪器在避免上述问题的情况下，以更简单的结构和更快的速度逐渐实现了比锚框跟踪器更好的性能。

随着CNN的不断发展，逐渐产生了很多深层CNN结构，如ResNet、GoogleNet[27]、VGG[28]、Inception[29]、

MobileNetV2[30] 等。网络越深, 不同层之间的特征差异越大。不同特征层所含的空间特征与语义特征区别较大, 多层特征的融合变得有意义。深浅层特征的融合结合空间特征与语义特征的优点, 因此, 衍生了大量基于特征融合的网络架构, 如 FCN、Inside-outside net、FPN、DenseNet、SiamRPN、SiamCAR 等。本文根据分类分支与回归分支的特性, 在调整通道数比例的基础上, 采用拼接的方法进行深浅层特征层融合。

为提高定位精度, NMS[31]、IoU-Net[32]、MSR-CNN[33]、FCOS、IoU-aware[17] 等利用 IoU 分支或中心度分支作为单独分支来进行定位质量估计。这种额外的分支引入会导致训练和测试之间的不平衡, 产生不可靠的定位质量估计。PISA[34] 和 IoU-balance[35] 等基于定位质量估计, 给予分类损失不同的权重, 然而权重策略是隐性和有限的, 并未改变目标的分类损失最优值。为解决定位质量估计的问题, Li 等[18] 提出了分类与 IoU 联合训练的分类分支, 与 IoU 分支、中心度分支及 IoU 权重、中心度权重进行比较, 验证得到分类与 IoU 联合训练的分类分支更具有优越性, 同时设计了分类分支对应的 QFL 函数。Li 等[18] 优化了回归分支的损失函数, 针对回归分支提出了 DFL 函数, 同时结合 IoU 损失函数, 将分类分支与回归分支对应的损失函数相结合, 得到 GFL 函数。本文引入了分类与 IoU 联合训练的分类分支, 同时采用 GFL 函数来消除削弱训练阶段和推断阶段不平衡所导致的消极影响, 提高跟踪准确度和实时性。

2　基于深浅层特征融合的全卷积孪生网络框架

本文将跟踪任务分为分类与回归 2 个子问题, 以逐像素的无锚框方式解决问题。如图 1 所示, 框架由 3 部分组成: 特征提取子网络、特征融合模块、分类-回归子网络。特征提取子网络包括模板分支和搜索分支, 使用修改后的 ResNet50 作为骨干网络来提取特征, 分别对模板图像和搜索图像进行卷积, 得到对应特征层。通过 1×1 卷积和深度互相关操作得到不同通道数的特征层, 将所得用于分类和回归的第 5 个和第 6 个融合特征层, 输入后续的特征融合模块。特征融合模块采用特征拼接的方式进行特征融合, 分别得到用于分类分支与回归分支的 2 个融合特征层, 分别输入后一模块的分类分支与回归分支。分类-回归子网络由分类与 IoU 联合训练的分类分支和回归分支组成。

2.1　特征提取子网络与特征融合模块

利用修改后的 ResNet50 构建特征提取子网络。子网络包括 2 个分支: 模板分支将模板图像 $\boldsymbol{Z}$ 作为输入, 搜索分支将搜索图像 $\boldsymbol{X}$ 作为输入。2 个分支共享同一 CNN 作为骨干网络, 输出特征图 $\boldsymbol{\varphi}(z)$ 和 $\boldsymbol{\varphi}(x)$。将 $\boldsymbol{\varphi}(z)$ 作为 $\boldsymbol{\varphi}(x)$ 的卷积核来执行互相关操作, 以此融合 2 个分支的信息, 得到响应图 $\boldsymbol{R}$。在后续的子网络中, 通过解码响应图 $\boldsymbol{R}$ 来获得目标的定位信息和尺度信息, 响应图 $\boldsymbol{R}$ 应包含丰富的信息内容。

传统互相关操作只能形成单通道的压缩响应图, 缺乏有用特征和重要信息, 受 SiamRPN++所启发, 本文使用深度互相关操作来生成多通道响应图:

$$\boldsymbol{R}=\boldsymbol{\varphi}(x)\star\boldsymbol{\varphi}(z) \tag{1}$$

式中: "$\star$" 表示深度互相关操作, 属于一种变型的轻量化互相关操作。

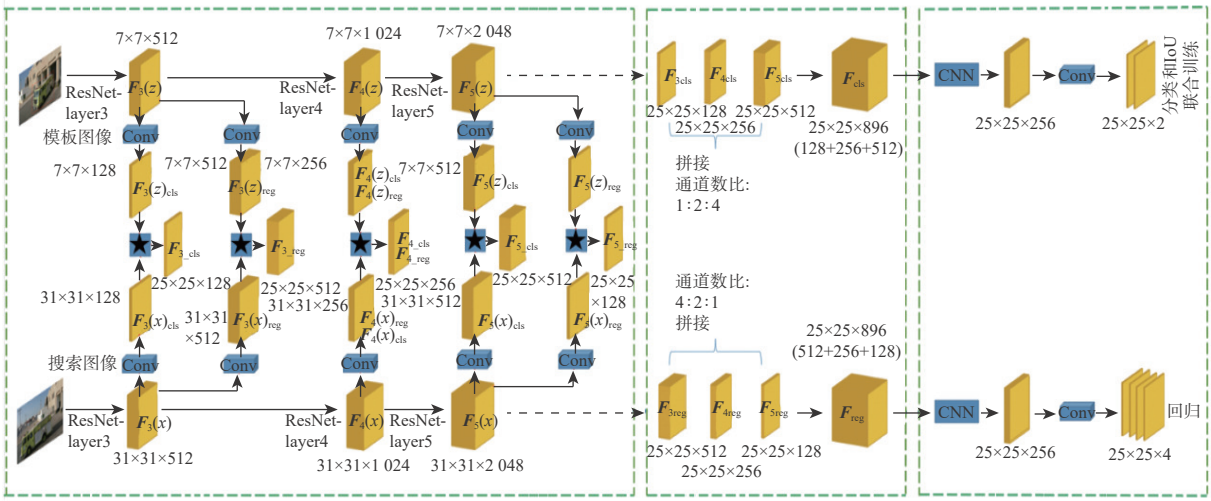
生成的响应图 $\boldsymbol{R}$ 与 $\boldsymbol{\varphi}(x)$、$\boldsymbol{\varphi}(z)$ 有相同的通道数, 包含大量的空间信息与语义信息, 深度互相关操作如图 2 所示。

浅层网络生成的浅层特征包含边缘、棱角、颜色、形状等信息, 富含空间信息, 更适合目标定位, 但因经过的卷积少, 所以语义性低, 噪声多。深层网络提取的特征距离输出较近, 含有更多抽象信息, 即语义信息, 更擅长目标分类, 但因其分辨率低, 细节感知差, 缺乏空间感知力。随着深层 CNN 的出现与应用, 本文可通过深浅层特征融合优化特征图, 获得更好的跟踪准确度。

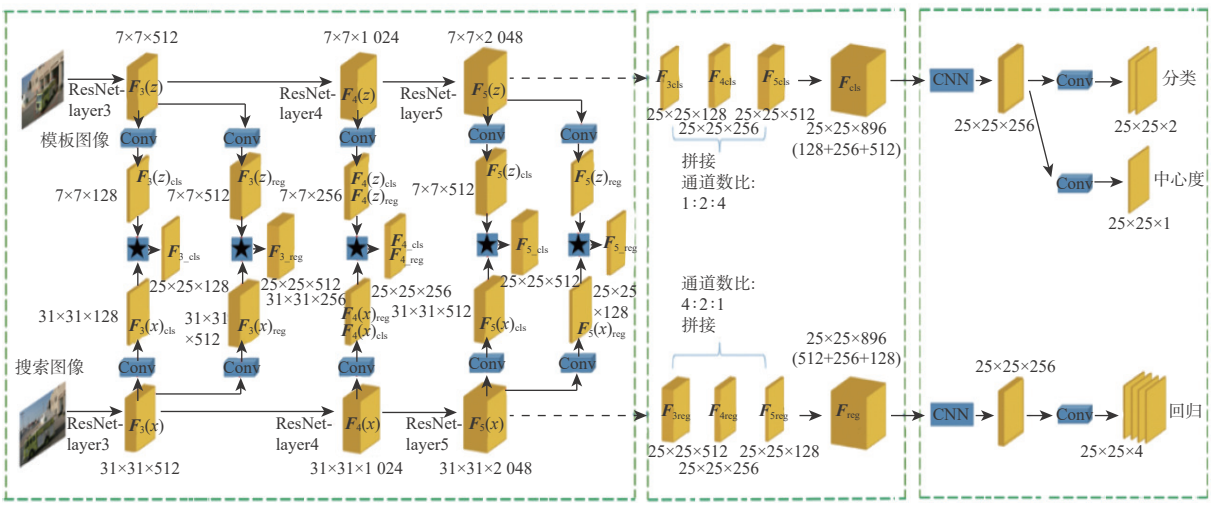
目标跟踪领域中, 多框架利用深浅层特征融合, 但用于分类和回归分支的融合特征相同, 没有根据任务要求来细化区分融合特征。分类分支需要更多语义信息, 回归分支需要更多空间信息, 因此, 本文不再使用相同的融合特征用于分类和回归分支。根据分类与回归分支的特点, 分别生成适合分类分支与回归分支的不同融合特征。

本文选用修改后 ResNet50 的后 3 个残差块的最后一层特征图来进行深浅层特征融合, 将其记为 $\boldsymbol{F}_3$、$\boldsymbol{F}_4$、$\boldsymbol{F}_5$, 模板分支对应 $\boldsymbol{F}_3(z)$、$\boldsymbol{F}_4(z)$、$\boldsymbol{F}_5(z)$, 搜索分支对应 $\boldsymbol{F}_3(x)$、$\boldsymbol{F}_4(x)$、$\boldsymbol{F}_5(x)$。分类分支需要更多语义特征, 回归分支需要更多空间特征。用于分类分支的 3 个特征图, 其通道数比例由少到多。用于回归分支的 3 个特征图, 其通道数比例由多到少。本文使用 1×1 卷积来调整 $\boldsymbol{F}_3(z)$、$\boldsymbol{F}_4(z)$、$\boldsymbol{F}_5(z)$ 与 $\boldsymbol{F}_3(x)$、$\boldsymbol{F}_4(x)$、$\boldsymbol{F}_5(x)$ 的通道数, 得到用于分类分支的 $\boldsymbol{F}_3(z)_{\mathrm{cls}}$、$\boldsymbol{F}_4(z)_{\mathrm{cls}}$、$\boldsymbol{F}_5(z)_{\mathrm{cls}}$ 及 $\boldsymbol{F}_3(x)_{\mathrm{cls}}$、$\boldsymbol{F}_4(x)_{\mathrm{cls}}$、$\boldsymbol{F}_5(x)_{\mathrm{cls}}$, 同时得到用于回归分支的 $\boldsymbol{F}_3(z)_{\mathrm{reg}}$、$\boldsymbol{F}_4(z)_{\mathrm{reg}}$、$\boldsymbol{F}_5(z)_{\mathrm{reg}}$ 及 $\boldsymbol{F}_3(x)_{\mathrm{reg}}$、$\boldsymbol{F}_4(x)_{\mathrm{reg}}$、$\boldsymbol{F}_5(x)_{\mathrm{reg}}$。

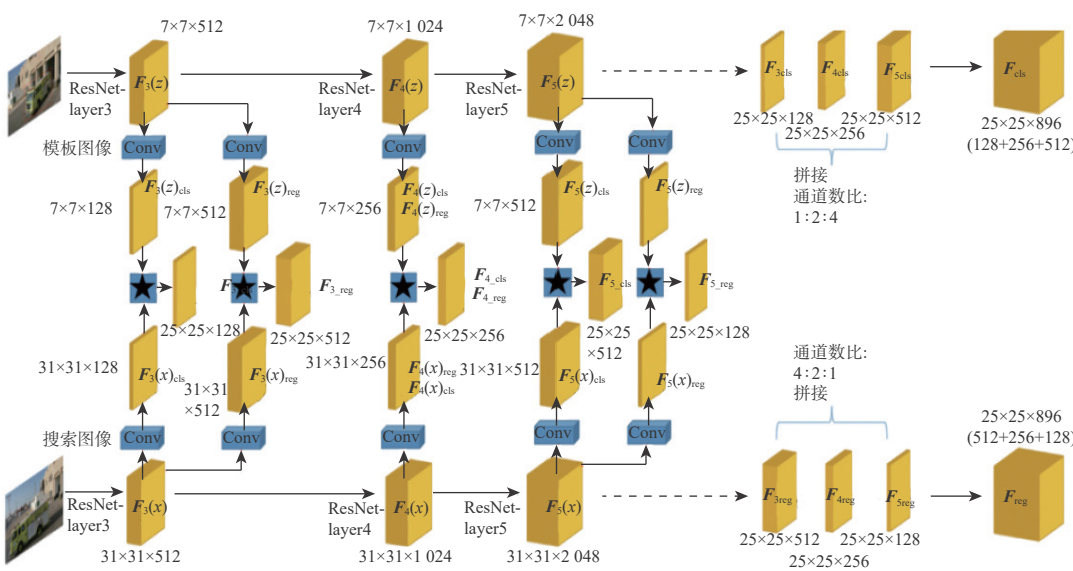
深浅层特征图通道数比例调参过程中, 本文选取 1∶2∶4(分类) 与 4∶2∶1(回归), 1∶3∶9(分类) 与 9∶3∶1



(a) 基于分类与IoU联合训练的算法框架 (本文最终框架)



(b) 基于中心度分支的算法框架 (本文对比框架)



(c) 特征提取子网络与特征融合模块的放大

图 1 基于深浅层特征融合的全卷积孪生网络框架示意图

Fig. 1 Siamese fully-convolutional network framework based on deep and shallow feature fusion

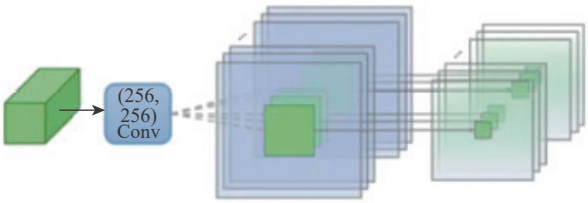


图 2　深度互相关操作

Fig. 2　Depth-wise cross correlation operation

(回归), 1∶1∶2(分类) 与 2∶1∶1(回归) 这 3 组参数与 1∶1∶1(分类、回归) 的特征融合方式进行比较。根据跟踪效果, 最终选取 1∶2∶4(分类) 与 4∶2∶1(回归) 作为通道数比例参数。将模板分支对应的 $\boldsymbol{F}_3(z)_{\text{cls}}$、$\boldsymbol{F}_4(z)_{\text{cls}}$、$\boldsymbol{F}_5(z)_{\text{cls}}$、$\boldsymbol{F}_3(z)_{\text{reg}}$、$\boldsymbol{F}_4(z)_{\text{reg}}$、$\boldsymbol{F}_5(z)_{\text{reg}}$ 与回归分支对应的 $\boldsymbol{F}_3(x)_{\text{cls}}$、$\boldsymbol{F}_4(x)_{\text{cls}}$、$\boldsymbol{F}_5(x)_{\text{cls}}$、$\boldsymbol{F}_3(x)_{\text{reg}}$、$\boldsymbol{F}_4(x)_{\text{reg}}$、$\boldsymbol{F}_5(x)_{\text{reg}}$ 进行深度互相关, 得到 $\boldsymbol{F}_{3_\text{cls}}$、$\boldsymbol{F}_{3_\text{reg}}$、$\boldsymbol{F}_{4_\text{cls}}$、$\boldsymbol{F}_{4_\text{reg}}$、$\boldsymbol{F}_{5_\text{cls}}$、$\boldsymbol{F}_{5_\text{reg}}$ 这 6 个特征图。具体的函数表达式如式 (2)～式 (7) 所示:

$$\boldsymbol{F}_{3_\text{cls}} = \boldsymbol{F}_3(x)_{\text{cls}} \star \boldsymbol{F}_3(z)_{\text{cls}} \tag{2}$$

$$\boldsymbol{F}_{4_\text{cls}} = \boldsymbol{F}_4(x)_{\text{cls}} \star \boldsymbol{F}_4(z)_{\text{cls}} \tag{3}$$

$$\boldsymbol{F}_{5_\text{cls}} = \boldsymbol{F}_5(x)_{\text{cls}} \star \boldsymbol{F}_5(z)_{\text{cls}} \tag{4}$$

$$\boldsymbol{F}_{3_\text{reg}} = \boldsymbol{F}_3(x)_{\text{reg}} \star \boldsymbol{F}_3(z)_{\text{reg}} \tag{5}$$

$$\boldsymbol{F}_{4_\text{reg}} = \boldsymbol{F}_4(x)_{\text{reg}} \star \boldsymbol{F}_4(z)_{\text{reg}} \tag{6}$$

$$\boldsymbol{F}_{5_\text{reg}} = \boldsymbol{F}_5(x)_{\text{reg}} \star \boldsymbol{F}_5(z)_{\text{reg}} \tag{7}$$

式中: $\boldsymbol{F}_{3_\text{cls}}$、$\boldsymbol{F}_{4_\text{cls}}$、$\boldsymbol{F}_{5_\text{cls}}$、$\boldsymbol{F}_{3_\text{reg}}$、$\boldsymbol{F}_{4_\text{reg}}$、$\boldsymbol{F}_{5_\text{reg}}$ 为框架特征提取子网络得到的 6 个用于后续特征融合的特征图。

特征融合模块将所得的分类分支对应的 3 个特征图 $\boldsymbol{F}_{3_\text{cls}}$、$\boldsymbol{F}_{4_\text{cls}}$、$\boldsymbol{F}_{5_\text{cls}}$ 和回归分支对应的 3 个特征图 $\boldsymbol{F}_{3_\text{reg}}$、$\boldsymbol{F}_{4_\text{reg}}$、$\boldsymbol{F}_{5_\text{reg}}$ 进行融合。本文使用特征拼接的方式将特征图三三融合, 分别得到用于分类分支和回归分支的融合特征图 $\boldsymbol{F}_{\text{cls}}$ 与 $\boldsymbol{F}_{\text{reg}}$。

2.2　分类-回归子网络

分类-回归子网络包括分类与 IoU 联合训练的分类分支和回归分支, 共同作用完成目标边界框的位置预测, 进行目标跟踪。

每个在融合特征图 $\boldsymbol{F}_{\text{cls}}$ 与 $\boldsymbol{F}_{\text{reg}}$ 上的位置点 (i,j) 都可以映射回输入的搜索区域上, 标记为 (x,y)。基于区域候选网络的跟踪器将搜索区域上的对应位置当做多尺度锚框的中心, 将这些锚框作为回归目标边界框的参考依据。本文算法采用无锚框方法, 在每个位置精准地分类与回归目标边界框。通过全卷积操作, 以端到端的方式进行相关训练, 避免调参过程, 减少人类干预。

跟踪任务被分为 2 个子任务: 预测每个位置类别的分类任务和计算目标边界框的回归任务。将特征融合所得的 $\boldsymbol{F}_{\text{cls}}$ 输入分类分支, 得到用于目标分类的分类特征图 $\boldsymbol{A}^{\text{cls}}_{w\times h\times 2}$, 其中, w 和 h 分别为图像的高度和宽度。将特征融合所得的 $\boldsymbol{F}_{\text{reg}}$ 输入分类分支, 得到用于目标定位的回归特征图 $\boldsymbol{A}^{\text{reg}}_{w\times h\times 4}$。$\boldsymbol{A}^{\text{cls}}_{w\times h\times 2}$ 上的每个点 (i,j) 都包括一个二维向量, 代表搜索图像区域上对应位置的前景与后景的分数。类似地, $\boldsymbol{A}^{\text{reg}}_{w\times h\times 4}$ 上的每个点 (i,j) 都包括一个四维向量 (l, t, r, b), 代表搜索图像区域上对应位置到边界框 4 个边的距离。

为提高目标的定位精度, 本文引入了质量评估部分。NMS、IoU-Net、MSR-CNN、FCOS、IoU-aware 等为了提高定位精度, 利用 IoU 分支或中心度分支作为单独分支来进行定位质量估计, 这种额外的分支引入会导致训练和测试之间的不平衡, 产生不可靠的定位质量估计。PISA、IoU-balance 等基于定位质量估计, 给予分类损失不同的权重, 然而权重策略是隐性和有限的, 并未改变目标的分类损失最优值。

为解决上述定位质量估计的问题, Li 等[18] 提出了分类与 IoU 联合训练的分类分支, 与多种分支方法进行比较, 验证得到分类与 IoU 联合训练的分类分支更具优越性, 同时设计了分类分支对应的 QFL 函数。受 QFL 函数启发, 本文选择分类与 IoU 联合训练的分类分支代替原分类分支。周丽芳等[36] 将 QFL 函数应用在其提出的 IoU 质量评估分支上, 并验证了可行性。

Li 等[18] 在引入分类与 IoU 联合训练分支和与之对应的 QFL 函数的同时, 也优化了回归分支的损失函数。针对回归分支提出了 DFL 函数, 同时结合 IoU 损失函数 IoUloss 来训练回归分支。DFL 函数可以在任意灵活的分布情况下, 使网络快速聚焦学习目标边界框连续性分布的概率值, 从而更好地模拟目标边界框灵活的实际分布, 提供更加准确的边界框位置。Li 等[18] 将分类分支与回归分支对应的损失函数 (QFL、DFL、IoUloss) 相结合, 得到分支结构的 GFL 函数。本文受 Li 等[18] 提出的 DFL 函数所启发, 引入 DFL 函数来训练回归分支。

在本文提出的框架基础上, 将中心度分支、IoU 分支、中心度权重、IoU 权重的质量评估方法和分类与 IoU 联合训练的方法进行性能对比, 进一步确定分类与 IoU 联合训练方法的优越性。

1) 分类分支损失函数。本文采用 Li 等[18] 提出的 QFL 函数来训练分类分支, 具体的函数表达式如下:

$$\text{QFL}(\sigma) = -|y-\sigma|^{\beta}[(1-y)\ln(1-\sigma)+y\ln\sigma] \tag{8}$$

式中：y为质量评估分支标签；σ为质量评估分支预测结果；β为控制比例常数。

2）回归分支损失函数。回归分支所用的损失函数一般为IoU损失函数，受到Li等[18]所提GFL函数的启发，本文引入DFL函数来优化IoU损失函数。

遵循Tian等[16]所提方法，将对应位置到边界框4条边的偏置距离作为回归目标。传统的边界框回归模型将回归标签作为狄拉克分布$\delta(x-y)$，其满足$\int_{-\infty}^{+\infty}\delta(x-y)x\mathrm{d}x=1$，恢复$y$的积分形式如下：

$$y=\int_{-\infty}^{+\infty}\delta(x-y)x\mathrm{d}x \tag{9}$$

根据分析，Li等[18]直接学习没有引入任何其他先验条件的、基本的一般分布$P(x)$，来替代狄拉克分布[37]和高斯分布[38]。给定标签y的范围，最小值为y_0，最大值为y_n（$y_0\leqslant y\leqslant y_n$，$n\in N^+$），可以从模型中得到估计值$y^*$（$y^*$也满足$y_0\leqslant y^*\leqslant y_n$）：

$$y^*=\int_{-\infty}^{+\infty}P(x)x\mathrm{d}x=\int_{y_0}^{y_n}P(x)x\mathrm{d}x \tag{10}$$

为与CNN相对应，引入间隔Δ，将连续的区域分为散点表征，用间隔Δ将范围(y_0,y_n)转换成集合$\{y_0,y_1,\cdots,y_{n-1},y_n\}$，$\Delta=y_{i+1}-y_i,\forall i\in[0,n-1]$。离散分布的性质$\sum_{i=0}^{n}P(y_i)=1$，则估计回归值$y^*$可以表示为

$$y^*=\sum_{i=0}^{n}P(y_i)y_i \tag{11}$$

因为边界框的学习仅适用于正样本，无类不平衡的危险，所以，Li等[18]简单地将QFL函数中完整的交叉熵部分应用到DFL函数的定义中：

$$\mathrm{DFL}(S_i,S_{i+1})=-(y_i+1-y)\ln S_i+(y-y_i)\ln(S_i+1) \tag{12}$$

式中：$S_i=\dfrac{y_{i+1}-y}{y_{i+1}-y_i}$，$S_{i+1}=\dfrac{y-y_i}{y_{i+1}-y_i}$，且有关系$y=\sum_{i=0}^{n}P(y_i)y_i=S_iy_i+S_{i+1}y_{i+1}=\dfrac{y_{i+1}-y}{y_{i+1}-y_i}y_i+\dfrac{y-y_i}{y_{i+1}-y_i}y_{i+1}=y$。

回归分支中采用的IoU损失函数表达式为

$$I_{\mathrm{IoUloss}}=-\ln\left(\frac{\mathrm{Intersection}(A,B)}{\mathrm{Union}(A,B)}\right) \tag{13}$$

式中：Intersection(A,B)为预测框A与目标真实框B的交集；Union(A,B)为预测框A与目标真实框B的并集。

分支结构的总损失函数为QFL函数、DFL函数、IoU损失函数的权重和：

$$L=\frac{1}{N_{\mathrm{pos}}}\sum_z L_{\mathrm{Q}}+\frac{1}{N_{\mathrm{pos}}}\sum_z 1_{\{c_z^*>0\}}(\lambda_0L_{\mathrm{B}}+\lambda_1L_{\mathrm{D}}) \tag{14}$$

式中：L_{Q}对应QFL函数；L_{D}对应DFL函数；L_{B}对应IoU损失函数；N_{pos}为正样本的数量；λ_0取为2，λ_1取为1/4，λ_0和λ_1分别为用来平衡L_{B}和L_{D}的权重；$1\{c_z^*>0\}$为指示函数，如果$c_z^*>0$，则为1，否则为0。

在训练过程中，通过上述总损失函数来优化对应网络框架的参数，得到高性能的跟踪器。

3 实 验

3.1 实施细节

3.1.1 实验环境

本文实验使用Ubuntu 20.04系统，CPU型号为i7-10700，GPU为2张RTX2080Ti，深度学习框架为Pytorch1.12.1，Python版本为3.9。

3.1.2 训练细节

本文将修改后的ResNet50网络在ImageNet数据集上进行预训练，将预训练后的ResNet50在ImageNetVID、ImageNetDET、COCO、GOT-10k上进行30个Epoch的迭代训练。将批量大小设置为32，模板图像大小设置为127×127，搜索图像大小设置为255×255。选用随机梯度下降的方法来优化模型，初始学习率设置为0.001。前10轮，在训练分类分支与回归分支时，特征提取子网络的参数被冻结。10轮之后，解冻ResNet-50的后3个残差块加入训练。整个训练过程约为68 h。为公平起见，在GOT-10k、LaSOT数据集上所进行的测试，使用官方网站提供的训练集来单独训练本文的网络框架。

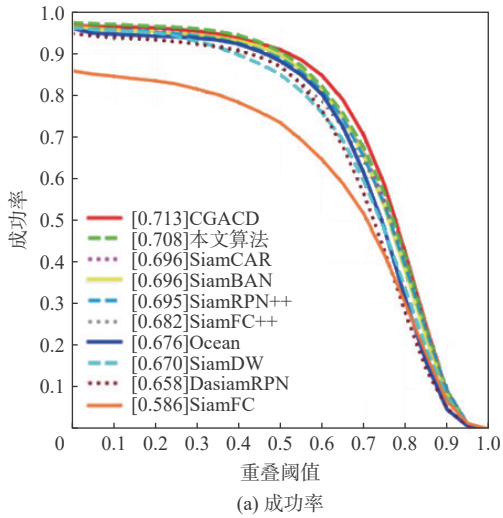
3.1.3 测试细节

在测试期间，本文使用离线跟踪策略，仅提供模板分支上初始帧的目标位置。在整个跟踪过程中，特征提取子网络的模板分支被预先计算且固定。当前帧的搜索图像被当做搜索分支的输入。将训练得到的网络框架在OTB-100[39]、UAV123、GOT-10k[40]、LaSOT[41]4个数据集上进行测试。本文算法在4个数据集上进行测试时的运行速率大约为75帧/s，实时运行速率良好。

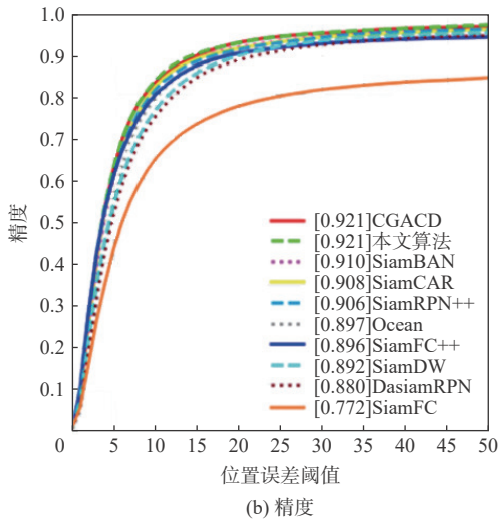
3.2 OTB-100数据集上的实验结果

OTB-100数据集是目标跟踪领域中比较经典的测试集，本文使用精度和成功率这2个评价指标来测试本文框架。OTB-100数据集包含11个属性：光照变化、遮挡、运动模糊、平面内旋转、平面外旋转、低像素、背景干扰、出视野、尺度变化、形变、快速运动。

本文算法与CGACD[42]、DaSiamRPN[43]、Ocean[8]、SiamCAR[7]、SiamFC[2]、SiamFC++[44]、SiamRPN++[4]、SiamDW[45]、SiamBAN[46]等9个先进的跟踪器做了对比。如图3所示，本文框架的成功率得分为0.708，排名第2，精度为0.921，排名第2，成功率与精度均



(a) 成功率



(b) 精度

图 3　OTB-100 数据集上不同跟踪器的实验结果

Fig. 3　Experimental results of different trackers on the OTB-100 dataset

比 SiamCAR 高, 证明了本文算法的有效性。排名靠前的性能也体现了本文框架的优越性。

3.3　GOT-10k 数据集上的实验结果

GOT-10k 数据集是大型目标跟踪数据集, 覆盖 560 类常见移动目标, 包含 10 000 多个视频序列。由于训练过程都由官方提供的数据集进行, 保证了测试的公平性。数据集提供的评价指标包括平均重叠率和成功率。平均重叠率表示估计边界框和真实框之间的平均重叠率, SR0.5 表示重叠率超过 0.5 的成功跟踪帧的比例。

本文在 GOT-10k 数据集上进行测试, 与 ECO[26]、SiamFC[2]、DSiam[23]、SPM[47]、SiamRPN++[4]、ATOM[22]、SiamCAR[7]、Ocean[8]、DiMP[48] 等 9 个跟踪器的结果进行比较。如表 1 所示, 本文算法在平均重叠率、SR0.5 上的分数表现很好, 使用中心度分支的本文算法(Ourscenterness)与 SiamCAR 只有深浅层特征融合方式不同, 却比 SiamCAR 分别提高了 0.015 与 0.021, 证明了本文特征融合方式的有效性, 优于传统的深浅层特征融合方式。本文算法作为完整的跟踪算法, 精度和成功率的性能相比 Ourscenterness 提高了 1%, 证明了分类与 IoU 联合训练分支替换分类分支与中心度分支的有效性, 也表明了本文算法可以兼容其他可提高框架性能的模块。本文算法在表 1 的跟踪器性能中, 平均重叠率排名第 2, SR0.5 排名第 2, 实现了优秀的跟踪性能。因为在 GOT-10k 数据集上, 跟踪器使用相同的训练集和测试集, 所以, 跟踪结果比在其他基准上的结果更具说服力。

表 1　GOT-10k 数据集上不同跟踪器的实验结果

Table 1　Experimental results of different trackers on the GOT-10K dataset

跟踪器	平均重叠率	SR0.5
ECO	0.299	0.303
SiamFC	0.325	0.328
DSiam	0.417	0.461
SPM	0.513	0.593
SiamRPN++	0.517	0.616
ATOM	0.556	0.634
SiamCAR	0.569	0.670
Ocean	0.592	0.695
DiMP	0.611	0.712
Ourscenterness	0.584	0.691
本文算法	0.593	0.698

3.4　LaSOT 数据集上的实验结果

LaSOT 数据集是一个用于单目标跟踪的数据集, 数据集包括 352 万以上的注释帧和 1 400 以上的视频序列, 其包含 70 个类别, 每个类别包括 20 个跟踪序列。如此庞大的测试集对于跟踪算法是一个巨大挑战。评估指标采用归一化精确度、精确度、一次性评估成功率。

将测试结果与 ECO[26]、SiamFC[2]、SiamRPN++[4]、SiamCAR[7]、ATOM[22]、DiMP-18、DiMP-50、Ocean[8] 这 8 个跟踪器进行比较。结果如表 2 所示, Ourscenternes 与 SiamCAR 相比, 一次性评估成功率和精确度分别提高了 0.025 和 0.032, 进一步论证了本文特征融合方式的有效性, 优于传统的深浅层特征融合方式。本文算法作为完整的跟踪框架, 一次性评估成功率和精确度分别比 Ourscenterness 进一步提高了 0.012 和 0.011, 证明了本文算法框架的兼容性, 可以方便地修改添加框架结构。本文算法在表 2 中的跟踪器中排名靠前, 均排名第 3, 表明了本文算法框架对于视觉跟踪有很好的泛化性。

3.5　UAV123 数据集上的实验结果

UAV123 数据集是从低航空视角捕获的视频数

表2 LaSOT数据集上不同跟踪器的实验结果

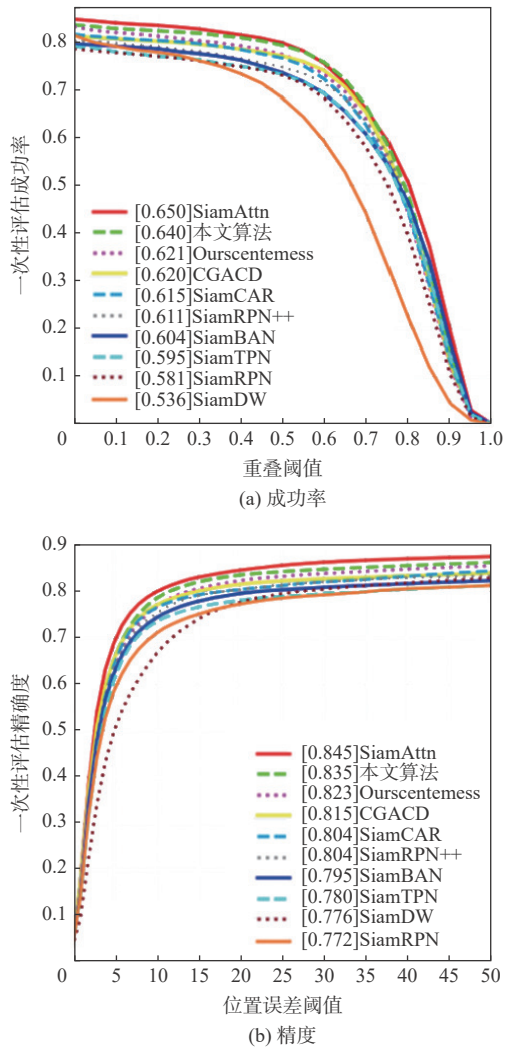
Table 2 Experimental results of different trackers on the LaSOT dataset

跟踪器	一次性评估成功率	精确度
ECO	0.324	0.301
SiamFC	0.336	0.339
SiamRPN++	0.496	0.491
SiamCAR	0.507	0.510
ATOM	0.514	0.505
DiMP-18	0.537	0.541
DiMP-50	0.558	0.564
Ocean	0.555	0.566
Ourscenterness	0.532	0.542
本文算法	0.544	0.553

据，其包括123个视频序列，有超过11万帧的图片，适合于无人机任务类型的数据测试。数据集中主要面对的挑战有快速移动、大尺度变化、大光照变化、遮挡等，有较强的挑战性。评估指标为一次性评估的精确度和成功率。

将本文框架在UAV123数据集上进行测试，并与SiamAttn[49]、SiamCAR[7]、CGACD[42]、SiamRPN++[4]、SiamBAN[46]、SiamTPN[50]、SiamRPN[3]、SiamDW[45]这8个跟踪器进行比较。如图4所示，基于中心度分支的本文算法框架(Ourcenterness)，比SiamCAR的指标分数分别高了0.006和0.019，表明了本文提出的特征融合方法的有效性和优越性。用分类与IoU联合训练的分类分支替代分类分支与中心度分支后，本文算法的指标分数分别提高了0.9%和1%，证明了分类与IoU联合训练分支对于本文框架的有效性，表明本文框架结构的兼容性，可以通过加入其他模块来进一步提升算法性能。

本文以UAV123数据集为基础进行了消融实验。①第1个消融实验。通过改变特征融合时的3个特征图通道比例数来选取一组合适的参数。本文共选取了3组参数：1:2:4(分类)、4:2:1(回归)；1:3:9(分类)、9:3:1(回归)；1:1:2(分类)、2:1:1(回归)。如表3所示，第1组参数所得结果优于其他组参数，所以本文选择第1组参数来作为特征融合时的通道比例。②第2个消融实验。以本文框架为基础，分别修改分类-回归子网络的分支结构，判断哪种分支结构更加适合本框架。本文共有5组结构：①分类分支、回归分支、中心度分支；②分类分支、回归分支、IoU分支；③分类分支、回归分支、中心度权重；④分类分支、回归分支、IoU权重；⑤分类与IoU联合训练的分支、回归分支。根据表4结果可知，Li等[18]提出的分类与IoU联合训练的分类分支更适合于本框架结构。因此，本文框架的分支结构依据第5组结构进行设计。



(a) 成功率

(b) 精度

图4 UAV123数据集上不同跟踪器的实验结果

Fig. 4 Experimental results of different trackers on the UAV123 dataset

表3 不同通道比例的实验结果

Table 3 Experimental results of different channel ratios

通道比例	一次性评估成功率	一次性评估精确度
1:2:4(分类)，4:2:1(回归)	0.640	0.835
1:3:9(分类)，9:3:1(回归)	0.617	0.819
1:1:2(分类)，2:1:1(回归)	0.629	0.824

表4 不同分支结构的实验结果

Table 4 Experimental results of different branch structure

分支结构	一次性评估成功率	一次性评估精确度
分类分支、回归分支、中心度分支	0.621	0.823
分类分支、回归分支、IoU分支	0.633	0.830
分类分支、回归分支、中心度权重	0.612	0.813
分类分支、回归分支、IoU权重	0.616	0.819
分类与IoU联合训练分支、回归分支	0.640	0.835

3.6　数据集实验结果分析

根据 LaSOT、GOT-10k、UAV123 数据集上的各自性能指标, 对本文算法进行综合性分析。将含中心度分支的框架与 SiamCAR 的数据指标进行相比, 发现性能均有所提升, 综合证明了本文的深浅层特征融合方式优于目标跟踪领域中传统直接的深浅层特征融合方式。使用分类与 IoU 联合训练分支的本文最终框架与含中心度分支的本文框架相比, 性能指标均得到了进一步提升, 证明了 Li 等[18]提出的分类与 IoU 联合训练分支对于本文算法框架的有效性和本文框架对于其他模块的兼容性。

本文选择 SiamCAR 作为对比模板进行了性能验证, 但该方法不仅限于 SiamCAR, 其还具有良好的迁移性和普适性, 也可以用于其他目标跟踪算法。可以置换其他算法的深浅层特征融合部分, 在不牺牲跟踪速度的情况下提升相应跟踪算法的性能; 也可以在不使用深浅层特征融合方法的目标跟踪算法框架中引入该融合方法, 在牺牲一定跟踪速度的情况下来进一步提升跟踪性能。

实验结果中, 性能优于本文算法的都有着更低的实时跟踪速度, 且架构更加复杂, 引入其他模块的难度更高。跟踪领域会不断出现优化各类算法性能且可迁移的模块方法, 但较复杂的算法框架很难兼容这些模块。而本文算法架构相对简单, 性能较高且跟踪速率高, 有较大的性能提升空间容量, 可以作为基础框架来加入其他复杂模块, 在保证实时跟踪速率的情况下来继续提升性能。本文算法在相对简单的框架结构下, 实现了良好的跟踪性能和跟踪速率, 可通过引入特征合并模块、注意力模块等来进一步提升跟踪性能。

3.7　实验结果定性分析

为对比本文算法与其他算法的性能差异, 选择 OTB-100 数据集的测试结果进行定性分析。测试结果如图 5 所示, 4 个视频序列包括形变、遮挡、快速移动、相似物干扰、尺度变化、平面外旋转、运动模糊、分辨率低等挑战。对此算法为 SRDCF[51]、SiamRPN[3]、SiamRPN++[4]、SiamDW[45]、Ocean[8]、SiamCAR[7] 算法。

在 Basketball 测试序列中, 存在相似物干扰、遮挡等挑战, 在发生相似物干扰的情况之后, 可以在第 500 帧中发现, 大部分算法跟丢, 只有 SiamDW、SRDCF 及本文算法还在继续跟踪目标。在 Bird1 测试序列中, 由第 212 帧可知, 在经过遮挡之后, 只有 Siam DW、SRDCF 及本文算法在继续跟踪目标, 从第 283 帧中可知, SRDCF 也跟丢目标, SiamDW 将要跟丢目标, 经过一系列形变和快速移动之后, 只有本文算法还在继续跟踪目标。在 BlurCar1 测试序列中, 面对运动模糊和分辨率低的挑战, 包括本文算法在内的孪生系列算法都成功跟踪目标, 只有非孪生系列的 SRDCF 算法跟丢目标, 说明了孪

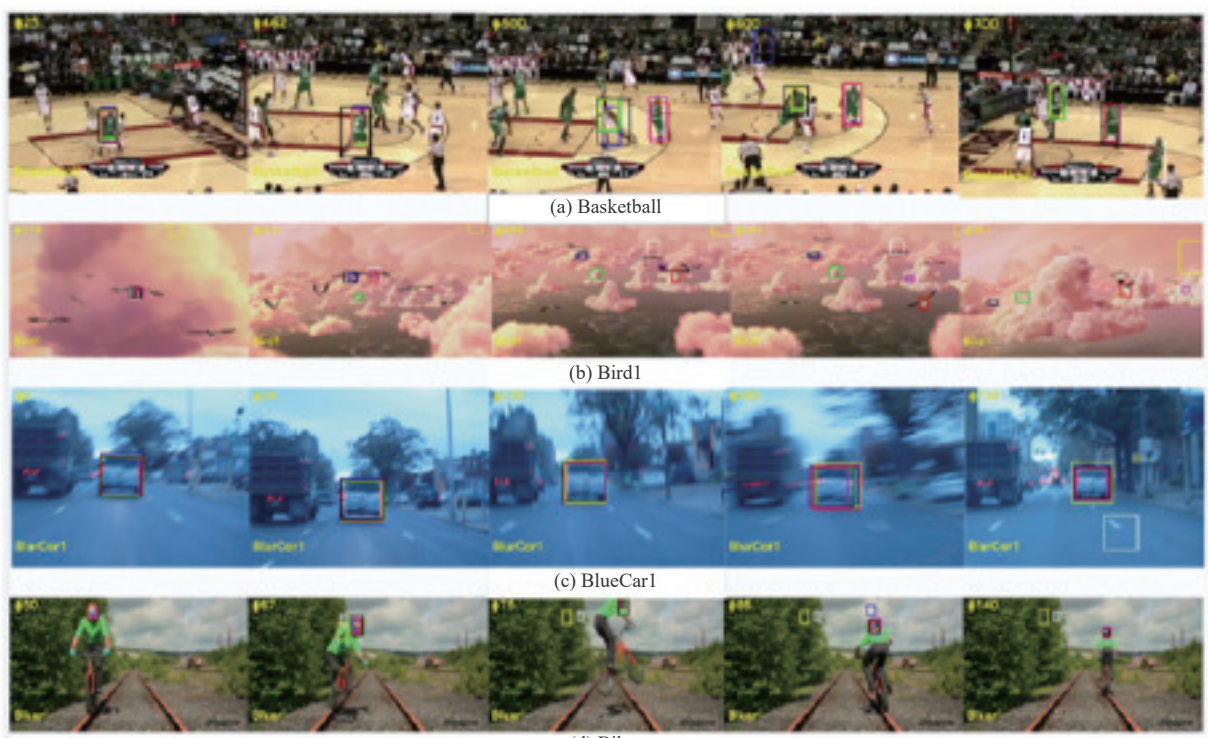


图 5　各类算法在不同视频下的跟踪结果

Fig. 5　Tracking results of various algorithms on different videos

生系列算法面对此类挑战的优越性。在Biker测试序列中，面对姿态变化、尺度变化、平面外旋转、快速移动等挑战，SRDCF、Ocean、SiamRPN都跟丢过目标，通过系列帧图像可知，成功跟踪的算法中，SiamDW和本文算法的效果最好，但是SiamDW算法利用了掩码，实时帧率不高。

4 结 论

本文提出了一种基于深浅层特征融合的全卷积孪生网络目标跟踪算法。

1）通过1×1卷积微调通道数，同时采用深度互相关和特征拼接的方法，结合分类分支与回归分支的特性，得到一种新的特征融合方法，并验证了方法的有效性，该方法可以迁移到其他跟踪算法框架中。

2）本文的分类分支引入了分类与IoU联合训练的方式，削弱了分类分支与回归分支的低相关性所造成的影响。

3）本文算法以75帧/s的帧率，在OTB-100、GOT-10k、LaSOT、UAV123这4个数据集上取得了较为先进的性能，证明了本文算法的有效性、实时性和泛化性。本框架较为简单，可以加入其他模块来进一步提升跟踪性能。

参考文献（References）

[1] TAO R, GAVVES E, SMEULDERS A W M. Siamese instance search for tracking[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1420-1429.

[2] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional Siamese networks for object tracking[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2016: 850-865.

[3] LI B, YAN J J, WU W, et al. High performance visual tracking with Siamese region proposal network[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8971-8980.

[4] LI B, WU W, WANG Q, et al. SiamRPN++: evolution of Siamese visual tracking with very deep networks[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4277-4286.

[5] 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.

[6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[7] GUO D Y, WANG J, CUI Y, et al. SiamCAR: Siamese fully convolutional classification and regression for visual tracking[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 6268-6276.

[8] ZHANG Z P, PENG H W, FU J L, et al. Ocean: Object-aware anchor-free tracking[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2020: 771-787.

[9] 蒲磊, 李海龙, 侯志强, 等. 基于高层语义嵌入的孪生网络跟踪算法[J]. 北京航空航天大学学报, 2023, 49(4): 792-803.
PU L, LI H L, HOU Z Q, et al. Siamese network tracking based on high level semantic embedding[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(4): 792-803(in Chinese).

[10] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[11] BELL S, ZITNICK C L, BALA K, et al. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 2874-2883.

[12] NÚÑEZ-GÓMEZ C, GARCIA-FONT V. HyperNet: A conditional k-anonymous and censorship resistant decentralized hypermedia architecture[J]. Expert Systems with Applications, 2022, 208: 118079.

[13] GUAN B, YAO J K, ZHANG G S, et al. Thigh fracture detection using deep learning method based on new dilated convolutional feature pyramid network[J]. Pattern Recognition Letters, 2019, 125: 521-526.

[14] OTÁLORA S, ATZORI M, ANDREARCZYK V, et al. Image magnification regression using DenseNet for exploiting histopathology open access content[C]//Proceedings of the Internetional Workshop on Ophthalmic Medical Image Analysis. Berlin: Springer, 2018: 148-155.

[15] LI K, ZOU C Q, BU S H, et al. Multi-modal feature fusion for geographic image annotation[J]. Pattern Recognition, 2018, 73: 1-14.

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

[17] WU S K, LI X P, WANG X G. IoU-aware single-stage object detector for accurate localization[J]. Image and Vision Computing, 2020, 97: 103911.

[18] LI X, WANG W H, WU L J, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection [EB/OL]. (2020-06-08)[2023-03-10]. https://doi.org/10.48550/arXiv.2006.04388.

[19] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.

[20] VALMADRE J, BERTINETTO L, HENRIQUES J, et al. End-to-end representation learning for correlation filter based tracking[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 5000-5008.

[21] ZHANG L, VARADARAJAN J, SUGANTHAN P N, et al. Robust visual tracking using oblique random forests[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recogni-

tion. Piscataway: IEEE Press, 2017: 5825-5834.

[22] DANELLJAN M, BHAT G, KHAN F S, et al. ATOM: accurate tracking by overlap maximization[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4655-4664.

[23] GUO Q, FENG W, ZHOU C, et al. Learning dynamic Siamese network for visual object tracking[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 1781-1789.

[24] HE A F, LUO C, TIAN X M, et al. A twofold Siamese network for real-time object tracking[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4834-4843.

[25] 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.

[26] DANELLJAN M, BHAT G, KHAN F S, et al. ECO: efficient convolution operators for tracking[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6931-6939.

[27] YOO H J. Deep convolution neural networks in computer vision: A review[J]. IEIE Transactions on Smart Processing and Computing, 2015, 4(1): 35-43.

[28] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//Proceedings of the 3rd International Conference on Learning Representations. San Diego: ICLR Press, 2015: 149801.

[29] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 1-9.

[30] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4510-4520.

[31] TYCHSEN-SMITH L, PETERSSON L. Improving object localization with fitness NMS and bounded IoU loss[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6877-6885.

[32] JIANG B R, LUO R X, MAO J Y, et al. Acquisition of localization confidence for accurate object detection[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 816-832.

[33] HUANG Z J, HUANG L C, GONG Y C, et al. Mask scoring R-CNN[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 6402-6411.

[34] CAO Y H, CHEN K, LOY C C, et al. Prime sample attention in object detection[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11580-11588.

[35] WU S K, YANG J R, WANG X G, et al. IoU-balanced loss functions for single-stage object detection[J]. Pattern Recognition Letters, 2022, 156: 96-103.

[36] 周丽芳, 刘金兰, 李伟生, 等. 基于 IoU 约束的孪生网络目标跟踪方法[J]. 北京航空航天大学学报, 2022, 48(8): 1390-1398.
ZHOU L F, LIU J L, LI W S, et al. Object tracking method based on IoU-constrained Siamese network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(8): 1390-1398(in Chinese).

[37] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6154-6162.

[38] CHOI J, CHUN D, KIM H, et al. Gaussian YOLOv3: an accurate and fast object detector using localization uncertainty for autonomous driving[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 502-511.

[39] WU Y, LIM J, YANG M H. Online object tracking: a benchmark [C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2013: 2411-2418.

[40] HUANG L H, ZHAO X, HUANG K Q. GOT-10k: a large high-diversity benchmark for generic object tracking in the wild[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5): 1562-1577.

[41] FAN H, LIN L T, YANG F, et al. LaSOT: a high-quality benchmark for large-scale single object tracking[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 5369-5378.

[42] DU F, LIU P, ZHAO W, et al. Correlation-guided attention for corner detection based visual tracking[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 6835-6844.

[43] ZHU Z, WANG Q, LI B, et al. Distractor-aware Siamese networks for visual object tracking[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 103-119.

[44] XU Y D, WANG Z Y, LI Z X, et al. SiamFC++: towards robust and accurate visual tracking with target estimation guidelines[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12549-12556.

[45] ZHANG Z P, PENG H W. Deeper and wider Siamese networks for real-time visual tracking[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4586-4595.

[46] CHEN Z D, ZHONG B N, LI G R, et al. Siamese box adaptive network for visual tracking[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 6667-6676.

[47] LAZEBNIK S, SCHMID C, PONCE J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories [C]//Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2006: 2169-2178.

[48] BHAT G, DANELLJAN M, VAN GOOL L, et al. Learning discriminative model prediction for tracking[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Pis-

cataway: IEEE Press, 2019: 6181-6190.

[49] YU Y C, XIONG Y L, HUANG W L, et al. Deformable Siamese attention networks for visual object tracking[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 6727-6736.

[50] LI Y, ZHANG X, WANG J, et al. Siamese transformer pyramid networks for real-time UAV tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 10001-10015.

[51] DANELLJAN M, HÄGER G, KHAN F S, et al. Learning spatially regularized correlation filters for visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 4310-4318.

Siamese network object tracking algorithm based on deep and shallow feature fusion

WANG Zikang，YAO Wenjin*，XUE Shangjie，SI Tingbo

(School of Mechanical Engineering，Nanjing University of Science and Technology，Nanjing 210094，China)

Abstract：Until recently, object tracking algorithms connected to deep learning have not considered the distinctions between regression and classification branches while employing shallow and deep feature fusion. Both branches used the same fusion feature, which could not meet the different task requirements of each branch well at the same time. According to the relationship between different task requirements of branches and features, a new feature fusion method was proposed for the object tracking algorithm. The channels of different feature layers in the backbone network were adjusted proportionally to form the fusion features suitable for the classification branch and regression branch. To compare and prove the effectiveness of the new feature fusion method, optimization was carried out on the basis of the SiamCAR algorithm. By changing the method of feature extraction and fusion, the accuracy of 2%−3% was improved on the three datasets of UAV123, GOT-10k and LaSOT. The experimental findings demonstrate the efficacy of the novel feature fusion technique, and the framework as a whole achieves good tracking performance at a real-time running speed of 75 frames per second.

Keywords：object tracking；Siamese network；classification；regression；deep and shallow feature fusion；joint training of classification and IoU

Received：2023-03-23；**Accepted**：2023-05-19；**Published Online**：2023-06-09 16：07

URL：link.cnki.net/urlid/11.2625.V.20230608.1150.001

* **Corresponding author**. E-mail：njyaowj@njust.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0144

单离子通道神经元电路的动力学分析方法

杨梓悦，吴静*，李乐桐

(北京航空航天大学 自动化科学与电气工程学院，北京 100191)

摘　　要：神经网络是由大量神经元通过突触联结而成的复杂网络系统，神经元的合理建模与分析是理解神经网络的功能及动力学特征的关键，对于推动脑科学与类脑科学研究具有重要价值。利用细胞膜电容和钾离子忆阻器分别表征神经元对电荷的存储和记忆特性，在Hodgkin-Huxley单离子通道神经元经典电路模型的基础上建立介观电路模型。使用经典电路理论和介观量子电路理论，在给予正弦激励下推导神经元细胞膜电压的响应表达式。计算结果表明：神经元细胞膜电压峰-峰值和迟滞回线的面积随激励频率的增大先增大后减小。在经典电路模型中，神经元细胞膜电压峰-峰值和迟滞回线面积达到最大值的频率与外部激励源的幅值有关，而在介观电路模型中，其仅与神经元电路参数有关，不依赖外部激励，更能表征神经元自身的特性。神经元的介观电路模型有益于揭示神经网络的动力学机理，推动脑科学理论体系的发展。

关　键　词：Hodgkin-Huxley神经元；介观电路；忆阻器；量子理论；动力学响应

中图分类号：TM133

文献标志码：A　　**文章编号**：1001-5965(2025)03-0985-07

对神经网络模型及其动力学的分析一直是生物学、医学、脑科学及类脑科学与技术等领域的研究热点，涉及的研究内容广泛，具有多学科交叉融合的研究特点。通过对神经网络的研究，一方面可以推动脑科学朝定量、精确和理论化的体系发展，推动临床医学对致病机理、诊断方法和治疗手段等的突破；另一方面可以了解脑功能实现机制，包括认知、学习、记忆等，探索建立人脑认知过程的结构体系。

神经网络是由大量神经元通过突触联结而成的复杂网络系统，其中，神经元模型是理解神经功能与兴奋性及神经网络动力学特征的关键[1-3]。神经元建模分为电导依赖型和非电导依赖型2种[4]。从经典电路理论的角度，前者主要利用生理参数进行建模，研究神经元及其网络的电学生理特性和动力学行为，后者基于前者，不再讨论神经元模型中电路参数的实际意义，而是使用数学模型来建模研究神经元输出，从而解释神经元传导的机制。1952年，Hodgkin和Huxley利用电压钳技术对枪乌贼的巨大轴突进行实验，提出了Hodgkin-Huxley(HH)模型[5]，模型计算结果与枪乌贼的巨大轴突实验结果吻合，从而奠定了现代神经元电生理模型的基础[3]。基于经典HH电路模型，文献[6-9]探索了该模型在神经网络中的集体行为，如相干、嵌合体状态和不相干等方面的应用。文献[10-12]研究了该模型在外部交直流激励源作用下对调节病理性疼痛等脑疾病的影响。文献[13]利用该模型对经颅磁声激励产生的感应电流与神经元模型放电活动响应特性之间的关系进行了仿真分析。可以认为，作为被广泛研究的一类电导依赖型神经元模型，HH模型从离子通道和门控蛋白的层面描述了神经元动作电位的形成机制，能够用于模拟大量神经元的电活

收稿日期：2023-03-24；**录用日期**：2023-11-16；**网络出版时间**：2023-11-27 14:05
网络出版地址：link.cnki.net/urlid/11.2625.V.20231123.1618.006
基金项目：北京市自然科学基金(Z220017)
* **通信作者**. E-mail：wujing06@buaa.edu.cn

引用格式：杨梓悦，吴静，李乐桐. 单离子通道神经元电路的动力学分析方法[J]. 北京航空航天大学学报，2025，51(3)：985-991.
YANG Z Y，WU J，LI L T. Dynamic analysis method of single ion channel neuron circuit[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：985-991 (in Chinese).

动，在神经科学和物理学之间建立了桥梁[14-15]。

在经典HH电路模型中，细胞膜电容表征了细胞膜的电荷存储特性，离子通道电导表征了神经元中的离子浓度变化。文献[16-17]利用忆阻器代替电导，进一步表征了离子通道电导与通道内离子浓度变化历史的关联。然而，脑神经元的大小为微米量级，属于介观尺度[18]。如要探讨介观神经系统的动力学特征，有必要对神经元的介观电路模型进行研究。现有量子化忆阻器及其在超导电路和集成量子光学中的实现，为神经元介观电路模型的建立与分析提供了一种新的思路[19-21]。

本文研究了经典的HH神经元电路模型，提出了神经元介观电路模型及其量子化分析方法，并对比正弦激励下经典电路模型和介观电路模型神经元细胞膜电压响应的异同，展示了介观电路模型在表征神经元动力学特征方面的适用性。

1 HH单离子通道神经元电路模型

1.1 利用忆阻器描述的经典电路模型

神经元电路模型主要考虑的是细胞膜电压的变化与外界激励之间的关系。神经元的细胞膜电压主要由神经元细胞膜内外带电离子浓度差来产生。当神经元感受到外界激励时，神经元细胞膜上的离子通道会输送带电离子来改变细胞膜内外离子浓度，从而使细胞膜电压发生变化。因此，在神经元电路模型中，通常以离子通道电导来表征细胞膜的导电能力，以细胞膜电容来表征细胞膜内外的电位差。离子通道电导和细胞膜电容是反映神经元活动的重要参数。为反映离子通道电导与通道内离子浓度变化历史的关联，采用忆阻器元件。

对于文献[7]中描述枪乌贼巨大轴突的经典HH神经元电路模型，当仅考虑单离子(K^+)通道时，模型可简化为如图1所示。其中，C表示细胞膜电容，C_g表示神经元与外部输入的耦合电容，E表示

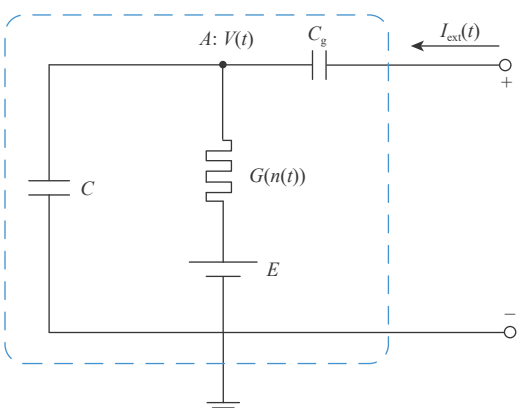


图1 HH单离子通道神经元的经典电路模型

Fig. 1 Classical circuit model of HH single ion channel neuron

该离子通道的静息电位，$G(n(t))$表示K^+通道对应的电导值，$n(t)$表示该离子通道的激活概率[7]。

$$\begin{cases} G(n(t)) = g_{\mathrm{K}}(n(t))^4 \\ \dfrac{\mathrm{d}n(t)}{\mathrm{d}t} = \alpha_n(t)(1-n(t)) - \beta_n(t)n(t) \end{cases} \tag{1}$$

式中：g_{K}为K^+通道的最大电导；$\alpha_n(t)$、$\beta_n(t)$可由细胞膜电压$V(t)$表示为

$$\begin{cases} \alpha_n(t) = 0.01(10 - V(t))/\left[\mathrm{e}^{(10-V(t))/10} - 1\right] \\ \beta_n(t) = 0.125\mathrm{e}^{-(V(t)/80)} \end{cases} \tag{2}$$

根据电压型忆阻器的电流$I_{\mathrm{G}}(t)$和电压$V(t)$的关系[17]：

$$\begin{cases} I_{\mathrm{G}}(t) = G(n(t))V(t) \\ \dfrac{\mathrm{d}n(t)}{\mathrm{d}t} = f(n(t), V(t)) \end{cases} \tag{3}$$

对于固定的状态变量$n(t)$，$G(n(t))\geqslant 0$，$f(n(t),V(t))$为单调函数，且满足$f(n(t),0)=0$。对比式(1)和式(3)可知，K^+通道电导可表征为状态变量为$n(t)$的电压型忆阻器。求解式(1)需要对输入电压的时间变量进行积分，表明K^+通道忆阻器的电导值与输入电压的当前值和先前值均有关。因此，在周期性激励源作用下，带有忆阻器的电路的伏安特性曲线为具有记忆特性的迟滞回线。

对于图1所示模型中的节点A，在幅值为I_0、角频率为ω的外部正弦激励源$I_{\mathrm{ext}}(t) = I_0\sin(\omega t)$作用下，根据基尔霍夫电流定律，可得

$$C\frac{\mathrm{d}V(t)}{\mathrm{d}t} + G(n(t))(V(t) - E) = I_0\sin(\omega t) \tag{4}$$

在已知细胞膜电容C、忆阻器电导$G(n(t))$、静息电位E及外部激励源参数的情况下，可以通过式(4)使用数值方法求解细胞膜电压$V(t)$。

1.2 介观电路模型及其量子化分析方法

1.2.1 介观电路模型

当电路的尺寸趋近于介观尺寸时，载流子的传输长度达到了相位相干长度，此时需要考虑电路的量子效应[22]。已有大量文献通过类比简谐振动的量子化方法，将电荷作为广义坐标实现了对介观RLC电路的量子化[23-25]。对于图1所示模型，根据忆阻器的定义，$G(t) = \mathrm{d}q(t)/\mathrm{d}\Phi(t)$，忆阻器的电导值与流经的磁通有关。为实现对介观忆阻器的量子化，借鉴文献[26]的方法，选节点磁通作为广义坐标，利用由电感和电容组成的Caldeira-Leggett谐振子电路模型等效忆阻器[27]，在此基础上建立HH单离子通道神经元的介观电路模型。具体做法为：将节点A作为坐标原点，以流经忆阻器的电流方向为x轴的正方向，将忆阻器用M ($M\to\infty$)个长度为Δx的Caldeira-Leggett谐振子电路等效替换。

如图2所示，第m个谐振子等效电路的节点磁

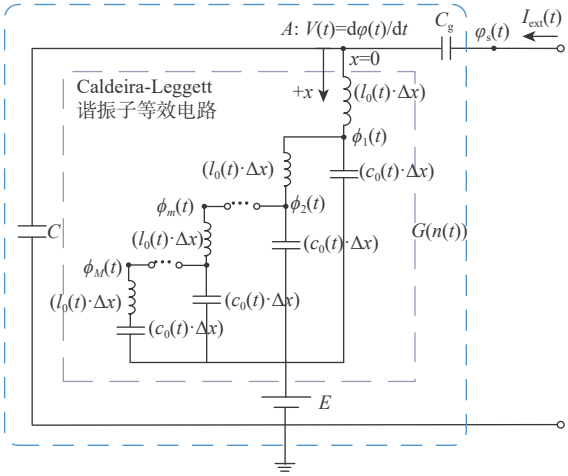


图 2　HH 单离子通道神经元的介观电路模型

Fig. 2　Mesoscopic circuit model of HH single ion channel neuron

通、单位长度电容和电感表示为 $\phi_m(t)$、$c_0(t)$ 和 $l_0(t)$, $\phi(t)$ 表示节点 A 处的磁通, $\phi_s(t)$ 为激励源处的节点磁通。

1.2.2　介观电路模型量子化分析方法

图 2 所示 HH 单离子通道神经元介观电路模型的量子化分析过程为: 首先, 求解表征神经元电路模型在时间维度上能量变化过程的拉格朗日量; 然后, 基于最小作用量原理, 使用欧拉-拉格朗日方程求解节点磁通变化表达式; 最后, 利用节点磁通表达式求出神经元细胞膜电压的响应及其他相关参数。

取 $x=0$ 处的 $\phi(t)$ 作为广义坐标, 则图 2 所示介观电路模型的拉格朗日量可表示为

$$L=-\frac{(\phi-\phi_1)^2}{2(l_0\Delta x)}+\sum_{m=1}^{\infty}\left[\frac{c_0\Delta x}{2}\left(\frac{\mathrm{d}\phi_1}{\mathrm{d}t}-E\right)^2-\frac{(\phi_{m+1}-\phi_m)^2}{2(l_0\Delta x)}\right]+\frac{C}{2}\left(\frac{\mathrm{d}\phi}{\mathrm{d}t}\right)^2+\frac{C_\mathrm{g}}{2}\left(\frac{\mathrm{d}\phi_\mathrm{s}}{\mathrm{d}t}-\frac{\mathrm{d}\phi}{\mathrm{d}t}\right)^2 \tag{5}$$

当 $\Delta x\to 0$ 时, 根据节点磁通 ϕ_m 的欧拉-拉格朗日方程 $\frac{\mathrm{d}}{\mathrm{d}t}\cdot\frac{\partial L}{\partial(\mathrm{d}\phi_m/\mathrm{d}t)}=\frac{\partial L}{\partial\phi_m}$, 可知

$$\frac{\mathrm{d}^2\phi_m}{\mathrm{d}t^2}=\frac{1}{l_0c_0}\cdot\frac{\partial^2\phi_m}{\partial x^2} \tag{6}$$

根据节点磁通 $\phi(t)$ 的欧拉-拉格朗日方程 $\frac{\mathrm{d}}{\mathrm{d}t}\cdot\frac{\partial L}{\partial(\mathrm{d}\phi/\mathrm{d}t)}=\frac{\partial L}{\partial\phi}$, 联立式 (5) 和式 (6), 可得

$$-I_\mathrm{ext}(t)+C\frac{\mathrm{d}^2\phi}{\mathrm{d}t^2}=\frac{1}{l_0}\cdot\frac{\partial\phi}{\partial x} \tag{7}$$

式中: $I_\mathrm{ext}(t)=C_\mathrm{g}\left(\mathrm{d}^2\phi_\mathrm{s}(t)/\mathrm{d}t^2-\mathrm{d}^2\phi(t)/\mathrm{d}t^2\right)$。$\phi(t)$ 可利用输入的波函数 $\varphi_\mathrm{in}(t)$ 表示为[26]

$$\frac{\partial\phi(t)}{\partial x}=\frac{1}{\sqrt{l_0c_0}}\left(2\frac{\mathrm{d}\phi_\mathrm{in}(t)}{\mathrm{d}t}-\frac{\mathrm{d}\phi(t)}{\mathrm{d}t}\right) \tag{8}$$

式中: $\phi_\mathrm{in}(t)=\sqrt{\hbar Z_0/(4\pi)}\int_0^\infty a_\mathrm{in}(\omega)\mathrm{e}^{-\mathrm{j}\omega t}/\sqrt{\omega}\mathrm{d}\omega$, $\hbar$ 为约化普朗克常量, $a_\mathrm{in}(\omega)$ 为输入量子湮灭算符, 离子忆阻器的阻抗为 $Z_0=\sqrt{l_0/c_0}$。

联立式 (7) 和式 (8), 可得

$$-I_\mathrm{ext}(t)+C\frac{\mathrm{d}^2\phi}{\mathrm{d}t^2}+\frac{1}{Z_0}\cdot\frac{\mathrm{d}\phi}{\mathrm{d}t}=\frac{2}{Z_0}\cdot\frac{\mathrm{d}\phi_\mathrm{in}}{\mathrm{d}t} \tag{9}$$

忽略节点磁通 $\phi(t)$ 的高频分量, 其量子化分解为

$$\phi(t)=\sqrt{\frac{\hbar Z_0}{4\pi}}\int_0^\infty\frac{\left(a_\mathrm{in}(\omega)\mathrm{e}^{-\mathrm{j}\omega t}+a_\mathrm{out}(\omega)\mathrm{e}^{-\mathrm{j}\omega t}\right)}{\sqrt{\omega}}\mathrm{d}\omega \tag{10}$$

式中: $a_\mathrm{out}(\omega)$ 为输出量子湮灭算符。

将外部输入电流 $I_\mathrm{ext}(t)$ 在频域内分解为

$$I_\mathrm{ext}(t)=\int_0^\infty\frac{1}{\sqrt{\omega}}\left(\mathcal{I}_\mathrm{ext}(\omega)\mathrm{e}^{-\mathrm{j}\omega t}+\mathcal{I}_\mathrm{ext}^*(\omega)\mathrm{e}^{\mathrm{j}\omega t}\right)\mathrm{d}\omega \tag{11}$$

联立式 (9)～式 (11), 则输出量子湮灭算符可用输入量子湮灭算符表示为

$$a_\mathrm{out}(\omega)=a_\mathrm{in}(\omega)\frac{\mathrm{j}-C\omega Z_0}{\mathrm{j}+C\omega Z_0}-\sqrt{\frac{4\pi Z_0}{\hbar}}\cdot\frac{\mathcal{I}_\mathrm{ext}(\omega)}{\omega\left(\mathrm{j}+C\omega Z_0\right)} \tag{12}$$

式中: $\mathcal{I}_\mathrm{ext}(\omega)=\frac{\sqrt{\omega}}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{j}\omega t}I_\mathrm{ext}(t)\mathrm{d}t$。

根据 $\langle\mathrm{d}\phi/\mathrm{d}t\rangle=V(t)$, 输出细胞膜电压 $V(t)$ 可表示为

$$V(t)=-\mathrm{j}\sqrt{\frac{\hbar Z_0}{4\pi}}\int_0^\infty\left[\langle a_\mathrm{in}(\omega)\rangle\left(1+\frac{\mathrm{j}-C\omega Z_0}{\mathrm{j}+C\omega Z_0}\right)+\sqrt{\frac{4\pi Z_0}{\hbar}}\cdot\frac{\mathcal{I}_\mathrm{ext}(\omega)}{\omega\left(\mathrm{j}+C\omega Z_0\right)}\right]\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}\omega \tag{13}$$

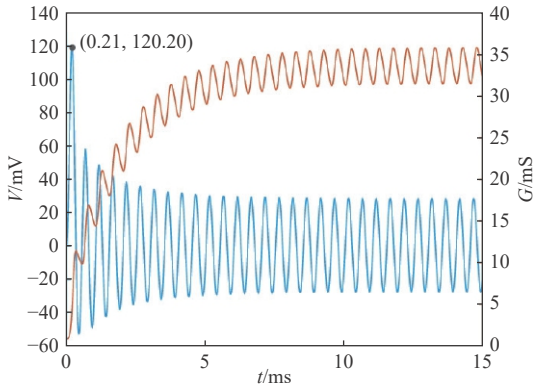
若选择量子化真空态, 则有 $\langle 0|a_\mathrm{in}(\omega)|0\rangle=\left\langle 0|a_\mathrm{in}^+(\omega)|0\right\rangle=0$, 故正弦激励下的输出细胞膜电压可表示为

$$V(t)=I_0Z_0\frac{\sin(\omega t)-\omega CZ_0\cos(\omega t)}{1+(\omega CZ_0)^2} \tag{14}$$

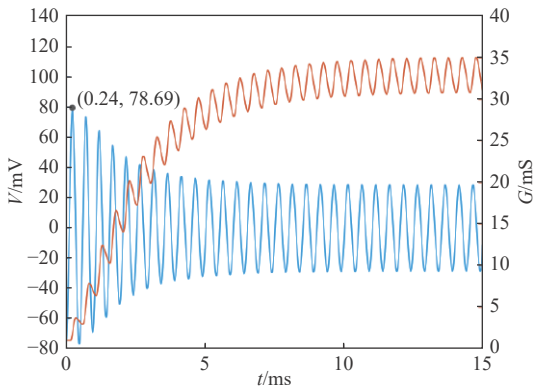
之所以选择正弦激励源, 是考虑到在线性系统中一些不容易分析和处理的信号, 都可以通过傅里叶变换、拉普拉斯变换、小波变换等, 变换成容易处理的正弦信号的叠加进行分析和处理。以上给出的介观电路模型及其量子化分析方法可推广到其他激励形式。

2　理论分析结果对比

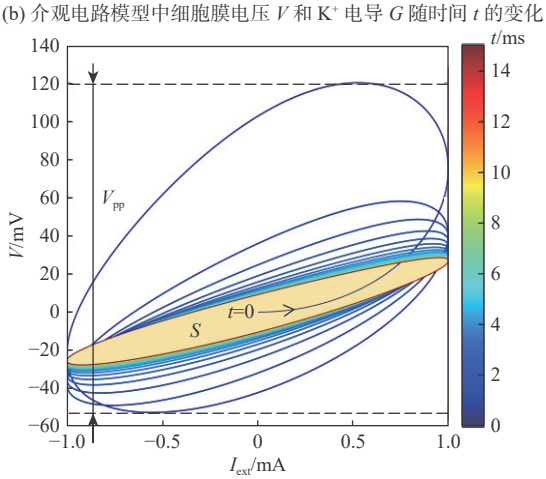
对于图 1 所示模型, 取枪乌贼实验参数 $g_\mathrm{K}=1.33$ S, $C=1$ μF。因为静息电位 E 的存在并不会影响到神经元的动力学响应, 为方便考虑取 $E=0$。分别利用式 (3) 和式 (14), 可以得到在不同幅值和频率的外部正弦激励源作用下, 经典电路模型和介观电路模型中的细胞膜电压 $V(t)$ 和 K^+ 电导 $G(t)$ 的响应。图 3 表示幅值为 1 mA、频率为 2 kHz 的正弦激励源作用下的 $V(t)$ 和 $G(t)$ 响应。由图 3(a) 和图 3 (b) 可



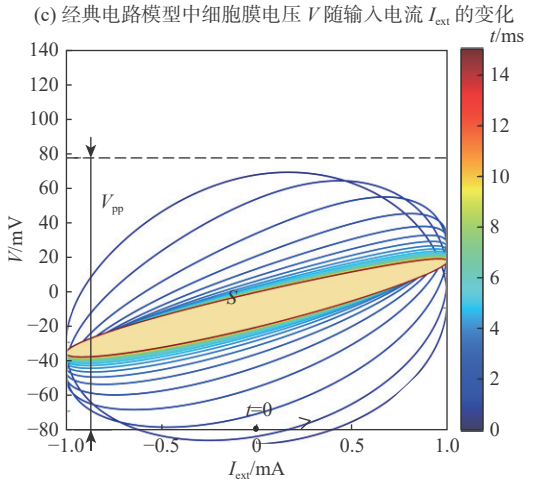
(a) 经典电路模型中细胞膜电压 V 和 K^+ 电导 G 随时间 t 的变化



(b) 介观电路模型中细胞膜电压 V 和 K^+ 电导 G 随时间 t 的变化



(c) 经典电路模型中细胞膜电压 V 随输入电流 I_{ext} 的变化



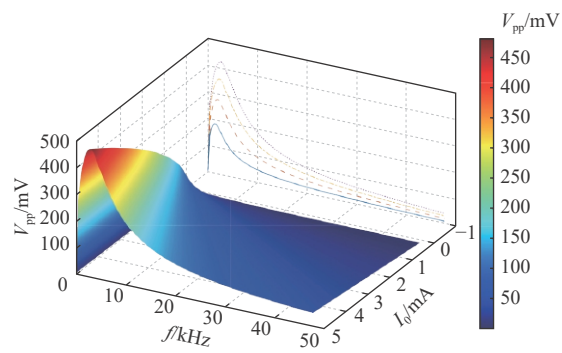
(d) 介观电路模型中细胞膜电压 V 随输入电流 I_{ext} 的变化

图 3 正弦激励下 HH 单离子通道神经元模型的动力学响应

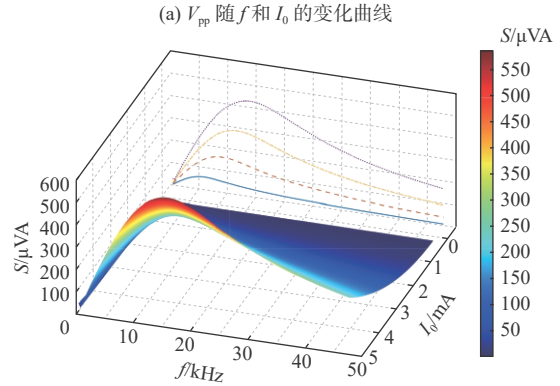
Fig. 3 Dynamic response of HH single ion channel neuron model under sinusoidal excitation

知，2 种电路模型下，K^+电导的幅值随时间逐渐增加，稳态值约 32 mS，细胞膜电压的幅值均随时间逐渐下降，最后达到正弦稳态值。在经典电路模型中，当 t = 0.21 ms 时，细胞膜电压会出现尖峰值 V_p = 120.20 mV；细胞膜电压峰-峰值 V_{pp} = 172.85 mV。在介观电路模型中，当 t = 0.24 ms 时，V_p = 78.69 mV；V_{pp} = 157.90 mV。而 V_{pp} 的值可反映神经元受到外部激励后的兴奋程度。图 3(c) 和图 3 (d) 分别为经典电路和介观电路模型下细胞膜电压随输入电流变化的曲线。2 种模型下，细胞膜电压的曲线都出现了迟滞回线特性，且当系统达到稳态时迟滞回线的面积 S 不再变化。S 可表示为神经元对外部激励的记忆持久性，面积越大，记忆持久性越大[20-21]。经典电路模型中，S = 32.16 μVA，而介观电路模型中，S = 31.56 μVA。可知，在介观电路模型下，枪乌贼神经元受到外部激励后的兴奋程度低于经典电路模型，但 2 种模型对外部激励的记忆持久性相近。因此，介观电路也可以用来表征神经元的部分动力学特征。

为进一步分析 V_{pp} 和 S 与输入激励的关系，计算 HH 单离子通道经典电路模型下 V_{pp} 和 S 随输入正弦激励的频率 f 和幅值 I_0 的变化曲线如图 4(a) 和图 4 (b) 所示；计算 HH 单离子通道介观电路模型



(a) V_{pp} 随 f 和 I_0 的变化曲线



(b) S 随 f 和 I_0 的变化曲线

I_0=1 mA I_0=2 mA I_0=3 mA I_0=4 mA

图 4 经典电路模型中细胞膜电压峰-峰值 V_{pp} 和稳态迟滞回线面积 S 随外部激励的频率 f 和幅值 I_0 的变化

Fig. 4 Variation of voltage peak-peak value of membrane V_{pp} and its steady-state hysteresis loop area S with frequency f and amplitude I_0 of external excitation in classical model

下 V_{pp} 和 S 随 f 和 I_0 的变化曲线如图 5(a) 和图 5 (b) 所示。在图 4 中, 当 I_0 = 1 mA, f 在覆盖 α 波、β 波、θ 波及 δ 波的低频段 (0～100 Hz), V_{pp} 和 S 均较小, 分别在 0～42.14 mV 和 0～1.60 μVA 范围内。随着外部输入激励幅值从 1 mA 增大到 2 mA、3 mA 和 4 mA, V_{pp} 分别在 1.60 kHz、1.92 kHz、2.24 kHz 和 2.56 kHz 处达到最大值 186 mV、279 mV、356 mV 和 422 mV, S 分别在 6.10 kHz、8.99 kHz、11.55 kHz 和 13.79 kHz 处达到最大值 59.82 μVA、160.90 μVA、288.70 μVA 和 429.90 μVA, 表明 V_{pp} 和 S 达到最大值时 f 会随着外部激励的幅值 I_0 的增强而变大。

对于图 5, 与经典电路模型的结果对比可知, 介观电路模型下 V_{pp} 和 S 同样会随 f 的增大呈现先增大后减小的特点。不同的是, 当 I_0 = 1 mA、f 在低频段 (0～100 Hz) 时, V_{pp} 和 S 分别在 0～897 mV 和 0～1.32 μVA 范围内, V_{pp} 比经典电路模型的范围更宽, S 与经典电路模型的范围接近, 后者略小于前者。随着外部输入激励幅值从 1 mA、2 mA、3 mA 增大到 4 mA, 使 V_{pp} 和 S 达到最大值的 f 不变, 即当 f = 0.15 kHz 时, V_{pp} 分别达到最大值 1 020 mV、1 980 mV、2 950 mV 和 3 920 mV; 当 f = 1.94 kHz 时, S 分别达到最大值 34 μVA、129.4 μVA、286.4 μVA 和 403.9 μVA。将该频率定义为神经元的本征频率 f_{eigen}, 主要由离子忆阻器的电导 $G(t)$ 和细胞膜电容 C 决定。由于 $G(t)$ 会随着 $V(t)$ 的改变而变化, 图 6 仅给出了 I_0 = 1 mA 时, f_{eigen} 随 C 的变化曲线。随着 C 的增加, f_{eigen} 急剧下降后缓慢减小。当 C > 23.36 μF 时, f_{eigen} < 100 Hz, 神经元受低频激励的影响较大; 当 C < 0.05 μF 时, f_{eigen} > 10 kHz, 神经元受高频激励的影响较大。

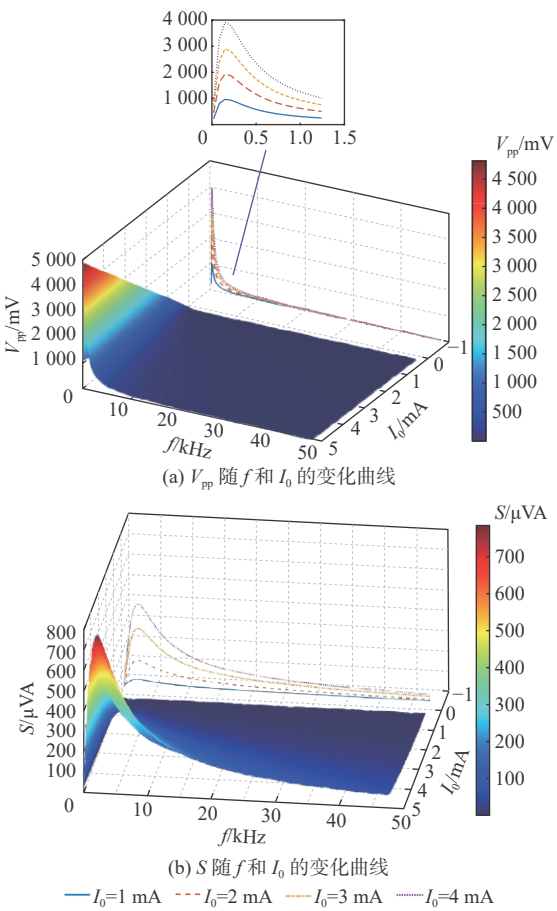


图 5　介观电路模型中细胞膜电压峰-峰值 V_{pp} 和稳态迟滞回线面积 S 随外部激励的频率 f 和幅值 I_0 的变化

Fig. 5　Variation of voltage peak-peak value of membrane V_{pp} and its steady-state hysteresis loop area S with frequency f and amplitude I_0 of external excitation in mesoscopic model

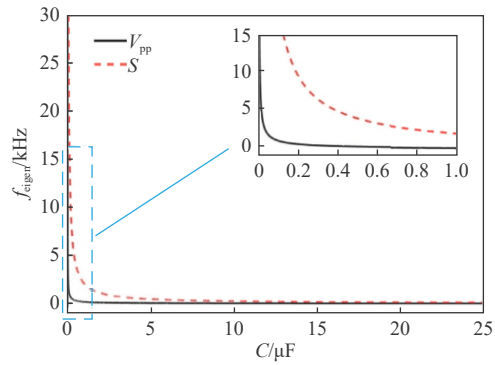


图 6　介观电路模型的本征频率 f_{eigen} 与细胞膜电容 C 的关系

Fig. 6　Dependence of intrinsic frequency f_{eigen} on membrane capacitance C in mesoscopic circuit model

3　结　论

本文研究了 HH 单离子通道神经元的经典电路模型, 并在此基础上引入了介观电路的量子化分析方法, 建立了 HH 单离子通道神经元的介观电路模型。基于所建立的经典电路模型和介观电路模型, 对比研究了在外部正弦激励源作用下细胞膜电压和 K^+电导的响应特性, 所得主要结论如下:

1) 经典模型下细胞膜电压峰-峰值及稳态迟滞回线面积均会随激励源频率的增大先增大后减小, 会随着激励源幅值的增大而增大。与经典电路模型相比, 在常见脑电波频段 (0～100 Hz), 利用介观电路模型获得的细胞膜电压峰-峰值较大; 频率较高时, 如达到数千赫兹时, 细胞膜电压峰-峰值较小。

2) 由 2 种电路模型得到的细胞膜电压迟滞回线面积接近, 介观电路模型的结果略小。

3) 在经典电路模型中, 细胞膜电压峰-峰值和迟滞回线面积达到最大值的频率会随外部激励源幅值的增大而增大, 而在介观电路模型中, 该频率与外部激励源的幅值无关, 当离子忆阻器电导和膜电容等参数变化时, 该频率才会变化, 可将其视为电路的本征频率。

根据枪乌贼巨型轴突的实验结果[15], 神经元在

受到外部刺激后的部分电学生理特征可以利用细胞膜电压峰-峰值及迟滞回线面积表征[26]，因此，神经元的经典电路模型和介观电路模型都可为其生理特征的预测提供理论支撑。介观电路模型相对于经典电路模型存在着本征频率，该频率直接影响神经元电路的灵敏度和响应能力。由神经元自身电路参数确定的本征频率能在一定程度上反映个体神经元的固有特性。

本文方法为研究神经网络的电学生理特性和动力学行为提供了一种新思路。基于理论模型得到的结果和结论后续仍需通过生物实验和电路实验进行验证。

参考文献（References）

[1] 张帅, 崔琨, 史勋, 等. 经颅磁声电刺激参数对神经元放电模式的影响分析[J]. 电工技术学报, 2019, 34(18): 3741-3749.
ZHANG S, CUI K, SHI X, et al. Effect analysis of transcranial magneto-acousto-electrical stimulation parameters on neural firing patterns[J]. Transactions of China Electrotechnical Society, 2019, 34(18): 3741-3749(in Chinese).

[2] 张帅, 许家悦, 李梦迪, 等. 基于皮层神经元模型的经颅磁声电刺激神经网络放电活动仿真分析[J]. 电工技术学报, 2021, 36(18): 3851-3859.
ZHANG S, XU J Y, LI M D, et al. Simulation of the discharge activity of neural network under transcranial magneto-acousto-electrical stimulation based on cortical neuron model[J]. Transactions of China Electrotechnical Society, 2021, 36(18): 3851-3859(in Chinese).

[3] 朱海军, 尹晓楠, 丁冲, 等. 高频即时磁刺激对神经兴奋性与电压门控钠钾离子通道的影响[J]. 电工技术学报, 2021, 36(18): 3787-3798.
ZHU H J, YIN X N, DING C, et al. Effects of high-frequency immediate magnetic stimulation on neuronal excitability and voltage-gated Na^+ and K^+ channels[J]. Transactions of China Electrotechnical Society, 2021, 36(18): 3787-3798(in Chinese).

[4] 徐泠风, 李传东, 陈玲. 神经元模型对比分析[J]. 物理学报, 2016, 65(24): 5-16.
XU L F, LI C D, CHEN L. Contrastive analysis of neuron model[J]. Acta Physica Sinica, 2016, 65(24): 5-16(in Chinese).

[5] HODGKIN A L, HUXLEY A F. Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo[J]. The Journal of Physiology, 1952, 116(4): 449-472.

[6] ANDREEV A V, FROLOV N S, PISARCHIK A N, et al. Chimera state in complex networks of bistable Hodgkin-Huxley neurons[J]. Physical Review E, 2019, 100(2-1): 022224.

[7] BAYSAL V, SARAÇ Z, YILMAZ E. Chaotic resonance in Hodgkin-Huxley neuron[J]. Nonlinear Dynamics, 2019, 97(2): 1275-1285.

[8] BOSSY M, FONTBONA J, OLIVERO H. Synchronization of stochastic mean field networks of Hodgkin-Huxley neurons with noisy channels[J]. Journal of Mathematical Biology, 2019, 78(6): 1771-1820.

[9] CARLU M, CHEHAB O, DALLA PORTA L, et al. A mean-field approach to the dynamics of networks of complex neurons, from nonlinear integrate-and-fire to Hodgkin-Huxley models[J]. Journal of Neurophysiology, 2020, 123(3): 1042-1051.

[10] KHODASHENAS M, BAGHDADI G, TOWHIDKHAH F. A modified Hodgkin-Huxley model to show the effect of motor cortex stimulation on the trigeminal neuralgia network[J]. The Journal of Mathematical Neuroscience, 2019, 9(1): 4.

[11] LIU D, ZHAO S, LUO X Y, et al. Unidirectional synchronization of Hodgkin-Huxley neurons with prescribed performance under transcranial magneto-acoustical simulation[J]. Frontiers in Neuroscience, 2019, 13: 1061.

[12] ZHANG X J, GU H G, WU F Q. Memristor reduces conduction failure of action potentials along axon with Hopf bifurcation[J]. The European Physical Journal Special Topics, 2019, 228(10): 2053-2063.

[13] 张帅, 史勋, 尹宁, 等. 基于H-H神经元模型的经颅磁声刺激对神经元放电活动的影响[J]. 高电压技术, 2019, 45(4): 1124-1130.
ZHANG S, SHI X, YIN N, et al. Effects of transcranial magneto-acoustical stimulation on neuronal firing activities based on H-H neuron model[J]. High Voltage Engineering, 2019, 45(4): 1124-1130 (in Chinese).

[14] HODGKIN A L, HUXLEY A F. The components of membrane conductance in the giant axon of Loligo[J]. The Journal of Physiology, 1952, 116(4): 473-496.

[15] HODGKIN A L, HUXLEY A F. A quantitative description of membrane current and its application to conduction and excitation in nerve[J]. Bulletin of Mathematical Biology, 1990, 52(1-2): 25-71.

[16] CHUA L. Memristor, Hodgkin-Huxley, and edge of chaos[J]. Nanotechnology, 2013, 24(38): 383001.

[17] CHUA L, SBITNEV V, KIM H. Hodgkin-Huxley axon is made of memristors[J]. International Journal of Bifurcation and Chaos, 2012, 22(3): 1230011.

[18] EBELING W. Stochastic processes in physics and chemistry[J]. Zeitschrift Für Physikalische Chemie, 1995, 188(1-2): 310-311.

[19] PFEIFFER P, EGUSQUIZA I L, DI VENTRA M, et al. Quantum memristors[J]. Scientific Reports, 2016, 6: 29507.

[20] SALMILEHTO J, DEPPE F, DI VENTRA M, et al. Quantum memristors with superconducting circuits[J]. Scientific Reports, 2017, 7: 42044.

[21] SANZ M, LAMATA L, SOLANO E. Invited Article: Quantum memristors in quantum photonics[J]. APL Photonics, 2018, 3(8): 080801.

[22] NEOUIOUA B, BENAMIRA F, BENBITOUR M A. Quantum fluctuations of mesoscopic RLC circuit with sources and time-dependant resistances[J]. Modern Physics Letters B, 2015, 29(15): 1550077.

[23] 陈斌, 李有泉, 沙健, 等. 介观电路中电荷的量子效应[J]. 物理学报, 1997, 46(1): 129-133.
CHEN B, LI Y Q, SHA J, et al. Quantum effects of charge in the mesoscopic circuit[J]. Acta Physica Sinica, 1997, 46(1): 129-133(in Chinese).

[24] 康金平, 蔡绍洪, 张玉强. 介观电路系统的量子化及其量子涨落分析[J]. 江南大学学报 (自然科学版), 2011(1): 86-92.
KANG J P, CAI S H, ZHANG Y Q. Analysis of quantization and quantum fluctuations of mesoscopic circuit system[J]. Journal of

Jiangnan University (Natural Science Edition), 2011(1): 86-92(in Chinese).

[25] 孔令杰. 有源介观二阶 RLC 串并联电路的量子效应[J]. 河北师范大学学报 (自然科学版), 2015, 39(2): 117-122.
KONG L J. Quantum effects of active mesoscopic second order RLC series and parallel circuit[J]. Journal of Hebei Normal University (Natural Science Edition), 2015, 39(2): 117-122(in Chinese).

[26] GONZALEZ-RAYA T, CHENG X H, EGUSQUIZA I N L, et al. Quantized single-ion-channel Hodgkin-Huxley model for quantum neurons[J]. Physical Review Applied, 2019, 12(1): 014037.

[27] CALDEIRA A O, LEGGETT A J. Quantum tunnelling in a dissipative system[J]. Annals of Physics, 1983, 149(2): 374-456.

Dynamic analysis method of single ion channel neuron circuit

YANG Ziyue，WU Jing*，LI Letong

(School of Automation Science and Electrical Engineering，Beihang University，Beijing 100191，China)

Abstract：By using a cellular membrane capacitor and a potassium ion memristor to characterize the charge storage and memory function of a neuron, this paper modified the classical circuit model and got the mesoscopic circuit model of the Hodgkin-Huxley single ion channel neuron. Based on the classical circuit theory and quantum theory of mesoscopic circuits, the response of membrane voltage of neurons under the sinusoidal excitation was derived. The calculation results show that with the increase in the frequency of the excitation source, the voltage peak-peak value of the membrane of the neuron and its steady-state hysteresis loop area first increase and then decrease. In the classical circuit model, the frequency at which the voltage peak-peak value and the hysteresis loop area reach the maximum depends on the amplitude of the external excitation source, while in the mesoscopic circuit model, it only depends on the circuit parameters of the neuron and is independent of the external excitation, which can better describe the characteristics of neurons. The mesoscopic circuit model of neurons is beneficial to reveal the dynamic mechanism of neural networks and promote the development of brain science theory systems.

Keywords：Hodgkin-Huxley neuron; mesoscopic circuits; memristors; quantum theory; dynamic response

Received：2023-03-24；**Accepted**：2023-11-16；**Published Online**：2023-11-27 14：05

URL：link.cnki.net/urlid/11.2625.V.20231123.1618.006

Foundation item：Beijing Municipal Natural Science Foundation (Z220017)

* **Corresponding author**. E-mail：wujing06@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0149

基于改进NSGA-Ⅱ算法的不规则集装器装载优化

文淦[1]，袁雷峰[2]，王旭东[3]，毕军[1,*]，王永兴[1]

(1. 北京交通大学 交通运输学院，北京 100044；2. 中国民航信息集团有限公司，北京 101300；
3. 天信达信息技术有限公司，北京 101300)

摘　　要：为贴合机舱内部弧形形状，实际集装器(ULD)多为不规则形状，实现机场货站不规则集装器的高效率装载，有利于提升航空货物运输的整体效率。在考虑货物容积限制等一系列实际装载约束下，建立以集装器容积利用率最大、装载货物数量最多和载重利用率最大为目标的不规则集装器多目标装载优化模型，并设计改进NSGA-Ⅱ算法对模型求解。以空间三向分割等原则构建启发式算法；采用NSGA-Ⅱ算法，进行交叉、变异、选择等操作，生成帕累托前沿集合；选出容积利用率最大的最优装载方案。实验结果表明：改进NSGA-Ⅱ算法在强异构货物装载条件下，仍能保持93.84%的高容积利用率；相比于初始NSGA-Ⅱ算法，所提的改进NSGA-Ⅱ算法能使容积利用率提升13%，表明改进NSGA-Ⅱ算法在不规则集装器装载问题上有良好的适应性，可为机场货站不规则集装器装载提供参考。

关 键 词：机场货站；集装器装载；不规则集装器；多目标优化；NSGA-Ⅱ算法
中图分类号：V353; TP311
文献标志码：A　　**文章编号**：1001-5965(2025)03-0992-13

航空集装器(unit load device, ULD)的装载是机场货站操作流程中的重要环节，其装载效率直接影响机场货站内货物的运转。不规则集装器在大型机场货站内装载效率较低，已经成为阻碍航空货运发展的痛点，人工凭经验装载的方式已不满足航空货运的发展需求，对于航空货运整体效率的提升影响较大。因此，如何提升机场货站内不规则集装器的容积利用率等是航空货运发展过程中亟待解决的关键问题。

装载问题根据优化目标的不同，主要分为单目标问题和多目标问题，针对不同类型的问题，国内外学者均提出了相应的解决方法，其中，以集装箱容积利用率最大为优化目标的研究最为广泛。例如，Bischoff和Ratcliff[1]考虑货物装载方向、尺寸、负载稳定性等限制约束，以容积利用率最大化为目标提出2种性能互补的启发式方法；Liu等[2]在以集装箱容积利用率最大为目标的前提下标准化订单货物的包装，建立树搜索结构并进行搜索得到货物的装箱方案；廖云峰等[3]以装载容积利用率最大为优化目标，提出一种结合块装载的启发式优化算法；吕雪菊和倪静[4]通过半径多样化的小生境遗传算法求解以集装箱体积利用率最大为目标的集装器模型，能够解决强异构货物的装载问题。上述研究虽然能够较好地优化集装箱的容积利用率，但在实际装载中，其他目标对于集装箱的装载优化同样重要。近年来，针对多目标函数装载问题研究逐渐兴起。赵桂红等[5]以集装箱体积利用率最大和载重利用率最大为目标函数建立模型，并提出混合遗传粒子群算法进行求解，但实际上大部分多目标优化还是将多目标转化为单目标进行求解，并且缺乏

收稿日期：2023-03-29；录用日期：2023-04-27；网络出版时间：2023-05-16 10:37
网络出版地址：link.cnki.net/urlid/11.2625.V.20230515.1746.002
基金项目：国家重点研发计划(2018YFC0706005)
*通信作者. E-mail：jbi@bjtu.edu.cn

引用格式：文淦，袁雷峰，王旭东，等. 基于改进NSGA-Ⅱ算法的不规则集装器装载优化[J]. 北京航空航天大学学报，2025，51(3)：992-1004.
WEN G，YUAN L F，WANG X D，et al. Loading optimization of irregular unit load device based on improved NSGA-Ⅱ algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：992-1004 (in Chinese).

考虑不规则情况下的装载优化。

受限于实际装载规则的限制及求解的复杂性, 大部分学者均将不规则集装器简化为规则形状进行研究。Li 等[6] 提出了一种新的大规模邻域搜索启发式算法, 对于解决货物装载计划问题具有一定的实用性。Mikram 等[7] 根据集装器装载随时间变化的特点, 基于 ARIMA 模型和指数平滑模型预测货物的装载需求, 更好地优化货运集装器的管理。陈丙成和李艳华[8] 针对货物在航空运输中的装载优化问题, 采用混合遗传算法进行优化求解。但近年来也有少量学者在不规则集装器装载问题上提出了解决办法。张长勇和刘佳瑜[9] 以实际多箱型不规则集装箱为研究对象, 提出混合遗传算法获得装载方案。张长勇和翟一鸣[10] 针对经过“飘板”处理的不规则集装器, 提出了一种满足具体装箱要求的改进遗传算法, 能够实现弱异构货物的装载。

目前, 对于集装器装载问题的研究仍以规则集装器为主, 而实际航空货运中, 集装器大多是不规则的, 因此, 装载受到的限制更多、复杂度更高。本文在充分考虑唯一性约束、不重叠约束等实际约束条件的基础上, 建立兼顾集装器容积利用率、装载货物数量和集装器载重利用率的不规则集装器装载多目标优化模型, 并结合基于最深处装载的三空间分割启发式算法与 NSGA-Ⅱ算法, 设计改进 NSGA-Ⅱ算法对模型进行求解。通过实际机场货站货物装载数据进行计算分析, 表明本文算法能够在强弱异构货物装载条件下, 高质量完成不规则集装器的装载。

1　问题描述

机场货站不规则集装器装载优化问题可以描述为: 在实际约束条件下, 将不同种类的货物按照一定的规则装入到集装器中。其中, 常用的航空集装器类型包括不规则集装器(如进行“收脚”处理的 PMC(Q4)、PMC(Q5), 进行“飘板”处理的 PMC(QF)、AKE)和规则长方体(如 PMC(Q6))等。使得满足各类约束和限制下的集装器容积利用率、装载货物数量和载重利用率最大, 以提升机场货站装载效率。图 1 为不规则集装器装载到机舱内的情况。

在实际场景中, 机场货站集装器装载一般需要考虑多方面因素, 包括集装器容积利用率、装载货物数量、载重利用率等, 因此, 该问题是典型的多目标优化问题。多目标优化问题通常难以获得满足所有目标均为最优的解, 一种有效的解决方法是找到一整套非支配解的良好表示[11]。

为保证货物在装载及运输过程中的安全性和



图 1　航空集装器装载示意图

Fig. 1　Loading diagram of ULD

完整性, 需要考虑多方面的约束, 包括方向约束、体积约束、质量约束、不重叠约束、稳定性约束、支撑限制约束、货物装载顺序约束等。

2　模型构建

2.1　基本假设

鉴于航空运输背景下不规则集装器装载问题的特点, 同时为便于建模, 给出如下基本假设:

1) 货物均为规则物体且质量分布均匀。打包处理后的不规则形状货物较少且对装载影响较小, 因此统一为规则均匀物体。

2) 不考虑货物挤压变形的影响。在装载过程中, 货物之间难免会有挤压, 但是影响很小, 故本文不做考虑。

3) 不考虑货物中途换位。货物在实际运输过程中, 除特殊情况外, 一般不会进行移动。

2.2　模型建立

如式(1)所示, 模型决策变量为 η_i^j, 其中, 当第 i 个货物装载到第 j 个集装器内时, 决策变量为 1, 表明货物装入该集装器内; 反之, 决策变量为 0, 则表明第 i 个货物没有装入第 j 个集装器。

$$\eta_i^j=\begin{cases}1 & \text{第}i\text{个货物装入第}j\text{个集装器}\\ 0 & \text{否则}\end{cases}\quad \forall i\in I,\forall j\in J \tag{1}$$

建立多目标优化模型, 其中, 模型的目标函数包括集装器最大容积利用率、最大装载货物数量和最大载重利用率, 具体如式(2)～式(4)所示。

集装器最大容积利用率为

$$P_{\mathrm{v}}=\max\frac{\sum_{i\in I}v_i\eta_i^j}{v_j}\qquad \forall j\in J \tag{2}$$

集装器内货物装载最大数量为

$$N=\max\sum_{i\in I}\eta_i^j\qquad \forall j\in J \tag{3}$$

集装器最大载重利用率为

$$P_{\mathrm{m}}=\max\frac{\sum_{i\in I}m_i\eta_i^j}{m_j}\qquad\forall j\in J \tag{4}$$

式中：P_{v}为集装器的容积利用率；N为集装器内所装载的货物数量；P_{m}为集装器的载重利用率；v_i为第i件货物的体积；v_j为第j个集装器的体积；m_i为第i件货物的质量；m_j为第j个集装器的最大载重量；I为所有货物的集合；J为所有待装载集装器的集合。

为适应装载过程中的不同需求，提高模型普适性，本文提出的3个装载目标均为实际装载中十分重要的参考标准。容积利用率反映了集装器内货物的装载情况，值越高说明集装器内的空间得到了越充分的利用；货物装载数量反映了所需装载货物的消耗情况，及时将货物尽可能多地进行装载也是集装器装载过程的主要目标之一。而不同的装载方案可能有较为接近的容积利用率，但有的方案的装载数量可能更多而更加有利于装载；同样的，装载方案的货物装载数量相近，但有的方案容积利用率更大，也更有利于装载。因此，在装载中，2个目标均需要考虑。而载重利用率反映了集装器的载重情况，各集装器的载重情况与航空器的重心平衡等有较大关系，涉及飞行耗油与安全，随着重货、大货不断增多，对于载重利用率的考虑也是必要的。

为保证集装器装载的规范性与可行性，模型的约束条件包括货物的唯一性限制、容积限制和载重限制等，具体如式(5)～式(12)所示。

1）货物装载唯一性约束。任意货物能装载且只能装载到一个集装器中：

$$\sum_{j\in J}\eta_i^j=1\qquad\forall i\in I,\forall j\in J,\eta_i^j\in[0,1] \tag{5}$$

2）容积限制约束。所装载货物的总体积不超过集装器的最大可利用容积：

$$\sum_{i\in I}v_i\eta_i^j\leqslant v_j\qquad\forall j\in J \tag{6}$$

3）载重限制约束。货物总质量不超过集装器的最大载重量限制：

$$\sum_{i\in I}m_i\eta_i^j\leqslant m_j\qquad\forall j\in J \tag{7}$$

4）货物不重叠约束。货物两两之间不能相交重合，保证装入集装器的货物的独立性：

$$\begin{cases}\left|\left(x_i^j+\dfrac{1}{2}l_i\right)-\left(x_k^j+\dfrac{1}{2}l_k\right)\right|\geqslant\dfrac{l_i+l_k}{2}\\ \left|\left(y_i^j+\dfrac{1}{2}w_i\right)-\left(y_k^j+\dfrac{1}{2}w_k\right)\right|\geqslant\dfrac{w_i+w_k}{2}\\ \left|\left(z_i^j+\dfrac{1}{2}h_i\right)-\left(z_k^j+\dfrac{1}{2}h_k\right)\right|\geqslant\dfrac{h_i+h_k}{2}\\ \forall i,k\in I,\forall j\in J\end{cases} \tag{8}$$

式中：(x_i^j,y_i^j,z_i^j)表示货物放置位置的参考坐标，即第i件货物在第j个集装器中放置的左后下坐标；l_i和l_k分别为2个相邻货物i和货物k的长度；w_i和w_k分别为货物i和货物k的宽度；h_i和h_k分别为货物i和货物k的高度。

5）货物装载尺寸约束。货物装载中的长、宽、高不能超过集装器的尺寸要求：

$$\begin{cases}0\leqslant x_i^j+l_i\leqslant l_j\\ 0\leqslant y_i^j+w_i\leqslant w_j\\ 0\leqslant z_i^j+h_i\leqslant h_j\end{cases}\qquad\forall i\in I,\forall j\in J \tag{9}$$

式中：l_j、w_j、h_j分别为第j个集装器内壁的长、宽、高。

6）稳定性约束。在货物装载过程中不能出现货物悬空放置的情况，尽可能避免货物因底部支撑不足导致稳定性丧失，进而发生货物损坏的情况：

$$\begin{aligned}T_i^k=&\{l_i+l_k-[\max(x_i+l_i,x_k+l_k)-\min(x_i,x_k)]\}\cdot\\&\left\{w_i+w_k-\left[\max(y_i+w_i,y_k+w_k)-\min(y_i,y_k)\right]\right\}\cdot\\&\frac{1}{l_iw_i}T_i^k>\frac{1}{2}\qquad\forall i,k\in I\end{aligned} \tag{10}$$

式中：T_i^k为货物i和货物k的接触率。

7）方向约束。所装载货物需要与集装器的长或宽平行，不能斜放：

$$\begin{cases}l_{xi}+l_{yi}+l_{zi}=1\\ w_{xi}+w_{yi}+w_{zi}=1\\ h_{xi}+h_{yi}+h_{zi}=1\\ l_{xi}+w_{xi}+h_{xi}=1\\ l_{yi}+w_{yi}+h_{yi}=1\\ l_{zi}+w_{zi}+h_{zi}=1\\ \left[\left(x_i^{j'}-x_i^j\right)-l_i\right]\left[\left(y_i^{j'}-y_i^j\right)-w_i\right]=0\end{cases}\qquad\forall i\in I,\forall j\in J \tag{11}$$

式中：l_{xi}、l_{yi}、l_{zi}表示第i件货物的长是否平行于x、y、z坐标轴中的某个轴，取值为0或1；w_{xi}、w_{yi}、w_{zi}表示第i件货物的宽是否平行于x、y、z坐标轴中的某个轴，取值为0或1；h_{xi}、h_{yi}、h_{zi}表示第i件货物的高是否平行于x、y、z坐标轴中的某个轴，取值为0或1。$(x_i^{j'},y_i^{j'},z_i^{j'})$表示第$i$件货物在第$j$个集装器中的右前上坐标。

8）特殊货物隔离约束。在航空货物中存在一些特殊货物或危险品货物，需要进行特殊处理，将该类货物与其他货物分开装载，既保证运输过程中货物的安全，同时也保证了整个运输过程的安全：

$$\eta_i^j+\eta_k^j=1\qquad\forall i,k\in I,\forall j\in J \tag{12}$$

3 求解算法设计

本文研究的机场货站不规则集装器装载问题属于多目标优化问题，同样也是NP-Hard问题，其复杂性主要体现在航空货物种类丰富，属于强异构

类型, 同时, 集装器为不规则形状, 求解难度增加。由于 NSGA-Ⅱ算法在求解多目标问题上的优越性[12], 兼顾良好的分布性及收敛速度快的特点, 有学者还提出了专门针对多目标求解的粒子群优化算法[13], 还有学者提出丢弃不良解的方法用于进化算法的多目标优化, 能够在帕累托最优前沿附近找到更好的解决方法[14]。因此, 尝试将该算法用于求解不规则集装器装载问题。

同时, 为适应模型众多约束条件和限制, 提升不规则集装器装载优化效率, 在原算法的基础上加入基于最深处装载的三空间分割启发式算法, 并针对编码、交叉和变异操作进行改进。

3.1　启发式算法

由于装载问题为 NP-Hard 问题, 应用传统算法很难求得合适的解, 通常采用启发式算法进行求解。例如, 运用基于列生成技术的启发式算法, 解决带旋转的货物装载问题[15]; 采用进化启发式算法, 主要考虑如何减少容器的使用[16]。

根据实际机场货站不规则集装器装载策略, 设计基于最深处装载的三空间分割启发式算法, 其基本原则是以集装器装载空间三部分分割为基础, 在高度限制条件下, 不断向空间内最深处(即货物的三维坐标最小处)装载。而算法的核心在于空间的三向分割、不同位置的高度限制及货物的最深处装载原则。

不规则集装器内部空间受限较多, 需要对货物放置过程进行规则限制。以 PMC(Q5) 集装器为例, 由于进行了“收角”处理, 在装载过程中需要对部分区域的货物进行限高处理:

$$H_{\text{position}}+h_{\text{goods}}\leqslant h_{\text{PMC(Q5)_long}}-\frac{\left(w_{\text{position}}+w_{\text{goods}}-w_{\text{PMC(Q5)_high}}\right)\left(h_{\text{PMC(Q5)_long}}-h_{\text{PMC(Q5)_short}}\right)}{w_{\text{PMC(Q5)_deep}}-w_{\text{PMC(Q5)_high}}} \tag{13}$$

式中: H_{position} 和 w_{position} 分别为货物装载位置的高度和宽度; h_{goods} 和 w_{goods} 分别为货物的高度和宽度; $h_{\text{PMC(Q5)_long}}$ 和 $h_{\text{PMC(Q5)_short}}$ 分别为 PMC(Q5) 集装器的长边高度和短边高度; $w_{\text{PMC(Q5)_high}}$ 和 $w_{\text{PMC(Q5)_deep}}$ 分别为 PMC(Q5) 集装器的顶部宽度和底部宽度。

在装载过程中, 剩余空间的不同也会导致货物装载位置的不同。假设在集装器空间的(0,0,0)位置放入第 1 个货物, 货物的长、宽、高分别为 l_1、w_1、h_1, 根据空间分割规则, 将剩余空间分为前向空间、侧向空间及顶部空间。图 2 为集装器内装载空间的三向分割, 对应的装载位置分别为$(0, w_1, 0)$、$(l_1, 0, 0)$、$(0, 0, h_1)$, 第 2 个货物便可以选择 3 个位置中的任意一个位置进行装载。根据货物最深处装载的原则, 先向 x 坐标和 y 坐标最小的位置进行装载, 即选择$(0, 0, h_1)$位置。第 2 个货物装载完成后便占据了一个可装载位置, 并新产生了 3 个可用的装载位置, 分别为$(l_2, 0, h_1)$、$(0, w_2, h_1)$、$(0, 0, h_1+h_2)$。以此类推, 后续货物的装载均按照该原则进行。由于货物具有长、宽、高 3 个方向, 共有 6 种摆放方式, 将其依次编码为 0～5, 如图 3 所示。

基于以上货物装载规则及算法核心设计, 图 4 为进行货物装载的启发式算法流程。首先, 输入相关参数信息, 并对装载空间进行初始化; 然后, 迭代货物、位置(空间)列表, 依次将货物进行装载判断, 包括危险品判断、是否可以倒置判断、集装器装载

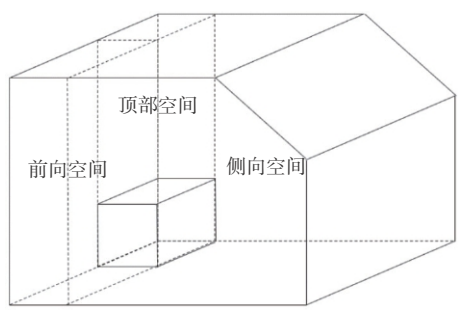


图 2　三空间分割示意图

Fig. 2　Schematic diagram three-space division

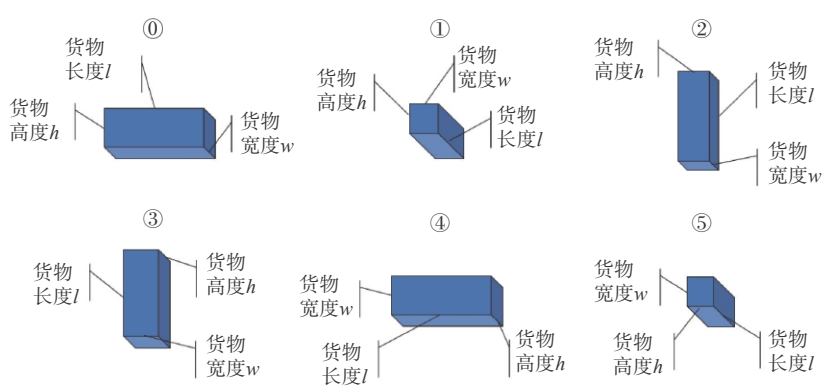


图 3　货物的 6 种摆放方式

Fig. 3　6 ways of placing goods

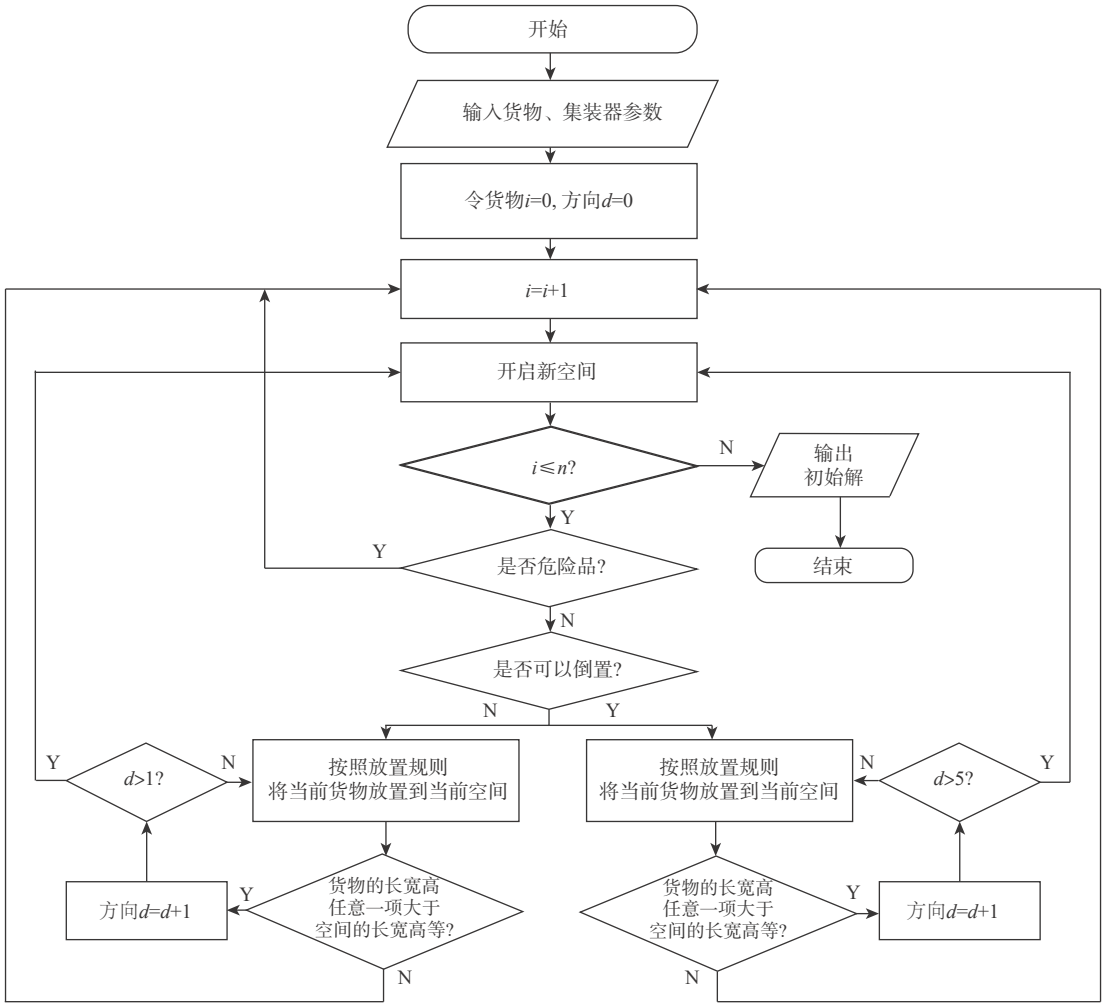


图4 启发式算法流程

Fig. 4 Heuristic algorithm flow

约束判断等；若满足各类要求及规则，即视为装载成功，反之则不能装载。其中，装载新空间的开启部分是基于算法核心设计中的三空间分割规则对装载后的集装器空间进行分割；放置规则判断部分是针对不规则集装器高度限制及最深处装载规则进行货物装载判断，以便于达到最优装载状态。

3.2 改进NSGA-Ⅱ算法

针对机场货站不规则集装器装载的特殊性，对基本遗传算法的步骤进行改进，主要体现在染色体编码、交叉和变异操作阶段。

3.2.1 染色体编码操作

在解决机场货站集装器装载问题中，解决方案指是否及如何将货物装入集装器中，以优化集装器容积利用率及载重利用率等多个目标。因此，通过特殊的二倍体染色体代表每个货物个体的2个主要特征。根据最终的求解目标，将货物按照体积和质量从大到小进行排列，即体积或质量越大的货物，顺序越靠前；由于货物共有6种装载方向，则装载编号依次为0～5，而货物方向则从初始开始均为正向装载，即初始化为0，若不满足装载要求，按照装载规则及方向限制进行适当旋转，得到货物的方向编号。如图5所示，由于编码方式不同，染色体1的基因编码是根据装载方案中货物的顺序进行排列的顺序号，染色体2是在允许的旋转类别基础上，每个货物的旋转参数值，其中，染色体2的旋转参数值与染色体1的货物顺序一一对应。

本文针对实际装载过程中机场集装器的装载，考虑2种不同的货物旋转方式，分别为2种方向的旋转及6种方向的旋转。

3.2.2 交叉操作

为适应二倍染色体，使用改进的基于顺序的交叉操作。从2个父代个体 P_1 和 P_2 产生2个子代个体 C_1 和 C_2。如图6所示，从中选择2个随机位置 i 和 j 开始，即 $i=2$ 和 $j=4$ 的位置，父染色体中 $P_1(i)\sim P_1(j)$ 之间的染色体被复制到子染色体中 $C_1(i)\sim C_1(j)$ 上，同样的，$P_2(i)\sim P_2(j)$ 被复制到 $C_2(i)\sim C_2(j)$ 上。从 $j+1$ 中循环扫描 C_1，填充 P_2 中的剩余值，类似地，用 P_1 中的值填充 C_2。不断重复

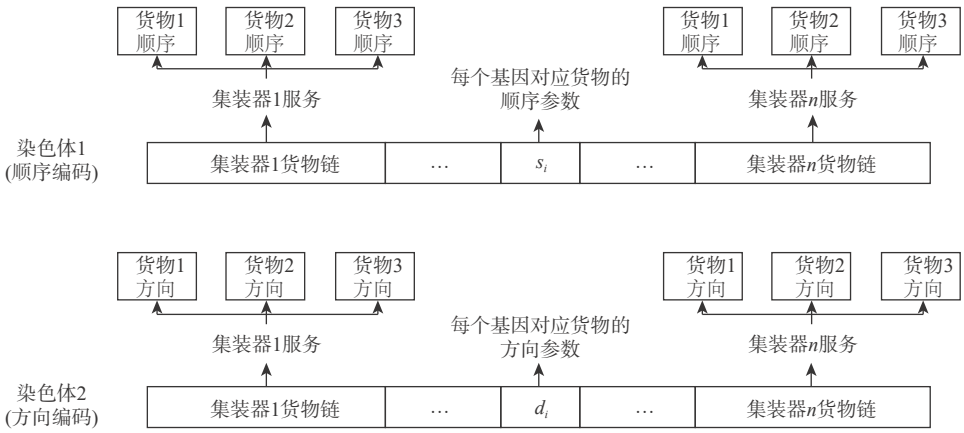


图 5　二倍染色体编码操作

Fig. 5　Encoding operation of diploid chromosome



图 6　二倍染色体交叉操作

Fig. 6　Crossover operation of diploid chromosome

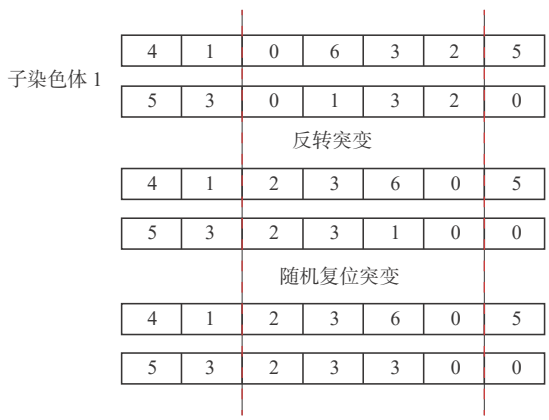


图 7　二倍染色体变异操作

Fig. 7　Variation operation of diploid chromosome

以上过程, 完成染色体的交叉操作。

3.2.3　变异操作

变异操作作为算法产生新个体的有效途径, 需要将经过交叉重组产生的个体进行两步突变[17]。首先, 对染色体基因以 p_{m1} 的概率进行反转突变; 然后, 以 p_{m2} 的概率进行随机复位突变, 即在可选择的基因编码中随机分配整数值给所选择的基因。如图 7 所示, 以子代个体 C_1 为例, 选择染色体中 2 个随机位置 i 和 j, 并将个体 $P(i)\cdots P(j)$ 中的部分反转。进一步地, 以 p_{m2} 的概率选择第 2 染色体中的每个基因, 并基于旋转类别随机重置为不同的旋转值。

图 7 为双向旋转的两步突变过程, 在父代染色体上随机选择起始和终止的基因片段, 对所选择的基因片段先进行前后反转, 再对该基因片段内的基因以 p_{m2} 的概率选择突变的基因, 最终完成变异操作。

3.3　改进 NSGA-Ⅱ算法步骤

将第 3.1 节和第 3.2 节所提到的优化策略和要素设计结合起来, 加入非支配排序、拥挤度排序及精英选择策略得到改进的 NSGA-Ⅱ算法。图 8 为本文提出的改进 NSGA-Ⅱ算法流程。算法主要分为 4 个阶段, 分别为输入阶段、交叉变异阶段、排序阶段及输出阶段。首先, 根据三空间分割等规则生成满足约束条件的初始装载方案(生成初始种群); 然后, 将生成的初始种群进行交叉、变异, 将得到的满足约束条件的装载方案(交叉变异生成的个体)与初始种群混合, 并将初始种群和交叉变异产生的个体进行排序筛选得到最优的一组帕累托解集; 最后, 验证迭代次数是否满足要求, 若不满足, 则把排序筛选得到的解集作为下一次迭代的初始种群再进行交叉变异排序, 若满足, 则输出帕累托解集。由于算法针对初始解进行了适应于不规则集装器装载的算法设计, 初始解的求解质量较高, 交叉操作的运行正常, 不会导致算法长时间陷入循环当中。

输入与交叉变异阶段已进行了详细阐述, 排序阶段的主要操作为: 首先, 针对初始种群与交叉变异产生的个体, 利用快速非支配排序法进行排序; 然后, 通过计算种群的拥挤度, 进行排序与选择, 即如果 2 个个体的非支配排序不同, 取排序号较小的

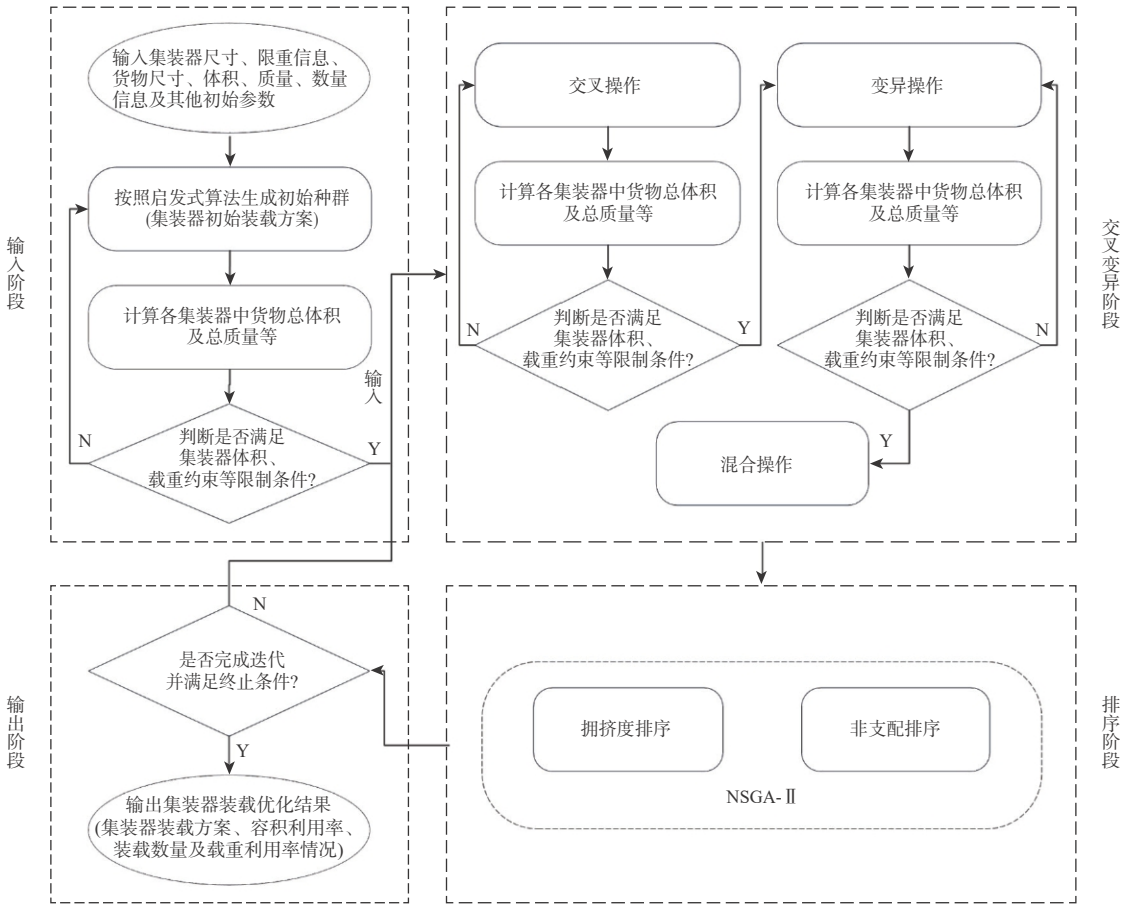


图 8 改进 NSGA-Ⅱ算法流程

Fig. 8 Improved NSGA-Ⅱ algorithm flow

个体，反之，若排序号相同，则取周围较不拥挤的个体。

4 算例分析

针对本文提出的机场货站不规则集装器装载多目标优化模型，利用改进 NSGA-Ⅱ算法进行求解，从最终得到的帕累托解集中筛选出最有利于实际装载的满意解，并将其作为集装器装载方案。

4.1 算例场景

本文以天信达信息技术有限公司所提供的实际数据为基础，数据主要包括某机场货站内某天的实际运单货物及其可使用的集装器作为本文的研究对象。其中，货物随数量和种类的增加依次从弱异构简单货物到强异构复杂货物，求解复杂度逐渐增加，而可使用集装器型号为 PMC(Q5)，具体信息如表 1 所示。为更加客观地展示本文所构建的集装器装载模型和算法性能，从弱异构简单货物和强异构复杂货物装载、改进 NSGA-Ⅱ算法和 NSGA-Ⅱ算法 2 个方面进行实验比较。

4.2 结果分析

在 Python 3.7 的环境下，实验所使用的计算机处理器型号为：Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz 1.80 GHz。算法参数设置如表 2 所示。

表 1 PMC(Q5) 集装器信息

Table 1 PMC (Q5) container information

参数	型号及数值
机型	B777-F
集装器类型	PMC
板型	Q5
自重/kg	114
最大载重量/kg	6 804
高度/cm	300
底板尺寸/(cm×cm)	244×318
底板可装货物尺寸/(cm×cm)	234×308
收脚高度/cm	短边一侧收角193
顶部尺寸/(cm×cm)	117×308

4.2.1 弱异构货物与强异构货物装载对比

根据实际机场货站运单数据，提取其中的货物数据，分别设置 4 种不同的货物数据，依次从弱异构简单货物装载到强异构复杂货物装载问题。货物种类(件数)依次为 8 类(288 件)、16 类(330 件)、24 类(487 件)和 51 类(1 070 件)，由于篇幅限制，截取部分货物详细信息，如表 3～表 6 所示。

表 2　算法参数设置

Table 2　Algorithm parameter setting

参数	数值
最大迭代次数	400
种群规模	48
交叉率	0.8
突变概率1	0.2
突变概率2	0.02
可旋转方向数	6
锦标赛规模	2

基于包含 8 类、16 类、24 类和 51 类货物的数据, 代入模型, 利用改进 NSGA-Ⅱ算法对各算例进行求解。

如图 9 所示, 通过不断优化, 可以得到包含不同种类货物装载方案的帕累托前沿。可以看出, 计算出的解集在解空间分布较均匀, 且大多集中在比较理想的位置。图 10 中曲线表示算法寻找最优解的过程, 随着迭代次数增加, 整体呈现不断向好的优化方向, 收敛速度快。

表 3　货物信息（8 类）

Table 3　Sheet of goods information (Category 8)

件数	质量/kg	尺寸/(cm×cm×cm)	体积/cm³	是否允许倒置	是否危险品	是否允许分批
141	17.52	60×44×44	116 160	是	否	否
32	7.88	43×38×32	52 288	是	否	否
57	5.5	38×30×32	36 480	是	否	否
⋮	⋮	⋮	⋮	⋮	⋮	⋮
3	4.31	34×28×30	28 560	是	否	否
1	3.42	30×27×28	22 680	是	否	否

表 4　货物信息（16 类）

Table 4　Sheet of goods information (Category 16)

件数	质量/kg	尺寸/(cm×cm×cm)	体积/cm³	是否允许倒置	是否危险品	是否允许分批
19	19.49	67×47×41	129 109	是	否	否
15	12.84	61×41×34	85 034	是	否	否
51	4.2	39×34×21	27 846	是	否	否
⋮	⋮	⋮	⋮	⋮	⋮	⋮
9	16.55	41×41×48	80 688	是	否	否
12	17.19	57×49×30	83 790	是	否	否

表 5　货物信息（24 类）

Table 5　Sheet of goods information (Category 24)

件数	质量/kg	尺寸/(cm×cm×cm)	体积/cm³	是否允许倒置	是否危险品	是否允许分批
11	9.52	54×54×37	107 892	否	否	否
1	13.29	74×55×37	150 590	是	否	否
5	6.77	43×42×48	86 688	是	否	否
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	7.75	39×28×30	32 760	是	否	否
50	11.56	40×40×40	64 000	是	否	否

表 6　货物信息（51 类）

Table 6　Sheet of goods information (Category 51)

件数	质量/kg	尺寸/(cm×cm×cm)	体积/cm³	是否允许倒置	是否危险品	是否允许分批
11	9.52	54×54×37	107 892	否	否	否
1	13.29	74×55×37	150 590	是	否	否
5	6.77	43×42×48	86 688	是	否	否
⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	8.54	55×37×28	56 980	是	否	否
10	15.1	67×47×32	100 768	是	否	否

不同场景下集装器装载方案的帕累托最优解及其相关指标如表 7～表 10 所示。由于在机场货站实际装载过程中, 集装器的容积利用率最大为首要目标, 按照容积利用率进一步筛选出最优解。表 11～表 14 为各类货物条件下的装载方案及其评价指标。其中, 货物位置即其在集装器内装载的三维坐标, 以此来确定货物的装载位置, 货物方向即为其装载时的旋转情况, 参数可参考图 3 的货物旋转参数。从表中可以看出, 集装器容积利用率均大于 83%, 并且在装载 51 类、1 070 件货物的强异构货物装载算例中, 容积利用率达到了 93.84%, 装载效果良好, 可以很好地解决航空货运中强异构货物的装载需求。

图 11 为集装器装载的可视化展示, 更加直观地反映了货物的装载情况, 其中 x、y、z 轴分别代表集装器的宽、长和高。从图 11 中可以看出, 本文算法对弱异构和强异构货物的装载均呈现良好的表现, 甚至随着货物种类和数量的增加, 装载的效果更佳, 也证明了本文算法能在强异构货物装载情况下仍能够表现出优良的表现, 并且所有数据为实际航空货运数据, 进一步表明算法具有较强的实用性和普适性, 能够适应不同种类的货物装载及不同场景的装载情况。

4.2.2　算法性能分析

为验证本文算法在求解不规则集装器装载问题上的优越性, 基于上述强异构复杂货物数据, 利用初始 NSGA-Ⅱ算法进行求解, 并将求解结果与本文算法求解结果进行比较。初始 NSGA-Ⅱ算法所使用的进化算子包括交叉算子、变异算子、非支配排序法、共享半径的统计比较标准和精英选择策略, 与本文算法的主要区别在于初始解的求解及使用拥挤度代替共享半径。将相同的强异构复杂货

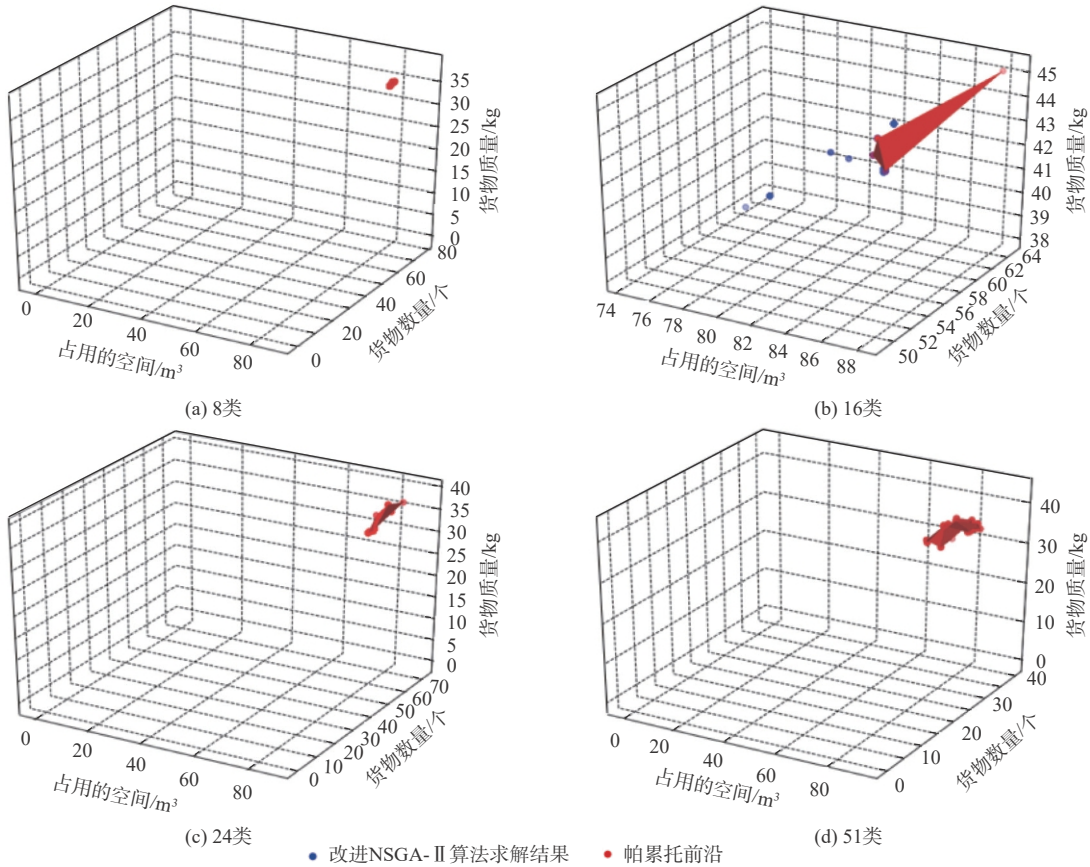


图 9 改进 NSGA-Ⅱ算法求得帕累托前沿

Fig. 9 Pareto frontier obtained by improved NSGA-Ⅱ algorithm

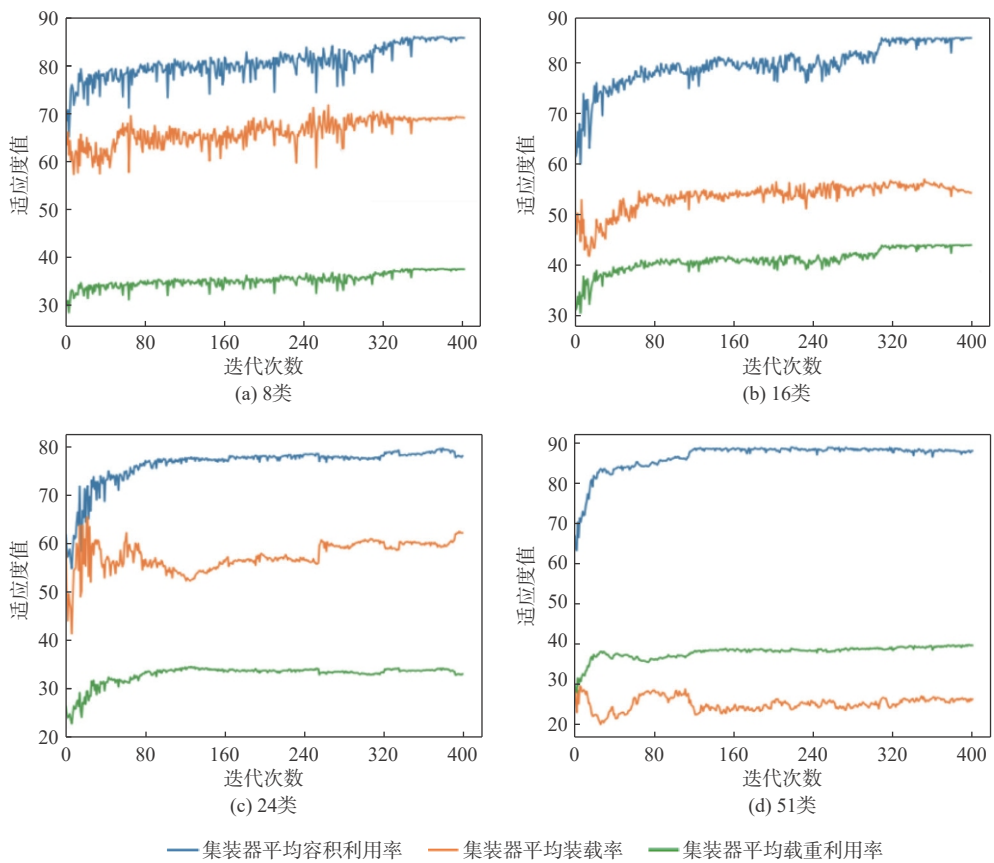


图 10 指标变化趋势曲线

Fig. 10 Curves of index change trend

表 7　8 类货物情况下的帕累托解集

Table 7　Pareto solution set for 8 types of goods

非支配解编号	装载货物数量/个	集装器容积利用率/%	集装器载重利用率/%
1	222	83.52	36.46
2	207	86.30	37.67
3	199	86.43	37.73
4	189	85.31	37.24

表 8　16 类货物情况下的帕累托解集

Table 8　Pareto solution set for 16 types of goods

非支配解编号	装载货物数量/个	集装器容积利用率/%	集装器载重利用率/%
1	205	74.13	38.58
2	204	74.93	38.29
3	189	85.08	44.63
4	184	90.27	45.59

表 9　24 类货物情况下的帕累托解集

Table 9　Pareto solution set for 24 types of goods

非支配解编号	装载货物数量/个	集装器容积利用率/%	集装器载重利用率/%
1	254	83.62	38.13
2	342	71.26	31.18
3	342	80.15	32.93
4	255	83.51	38.20

表 10　51 类货物情况下的帕累托解集

Table 10　Pareto solution set for 51 types of goods

非支配解编号	装载货物数量/个	集装器容积利用率/%	集装器载重利用率/%
1	212	93.48	41.28
2	415	76.25	32.34
3	181	92.67	41.48
4	228	93.84	40.99

表 11　最优装载方案（8 类）

Table 11　Optimal loading scheme (Category 8)

参数	数值
货物位置	[[0, 0, 0], [0, 0, 32], [0, 0, 64], ⋯, [98, 88, 264], [158, 220, 176], [158, 264, 0]]
货物顺序	[207, 206, 205, ⋯, 274, 113, 114]
货物方向	[4, 6, 4, ⋯, 4, 2, 5]
装载货物个数/个	199
集装器容积利用率/%	86.43
集装器载重利用率/%	37.73

表 12　最优装载方案（16 类）

Table 12　Optimal loading scheme (Category 16)

参数	数值
货物位置	[[0, 0, 0], [0, 0, 36], [0, 0, 72], ⋯, [171, 240, 150], [0, 274, 244], [171, 240, 180]]
货物顺序	[86, 85, 277, ⋯, 322, 306, 305]
货物方向	[3, 2, 1, ⋯, 1, 1, 4]
装载货物个数/个	184
集装器容积利用率/%	90.27
集装器载重利用率/%	45.59

表 13　最优装载方案（24 类）

Table 13　Optimal loading scheme (Category 24)

参数	数值
货物位置	[[0, 0, 0], [0, 0, 26], [0, 0, 52], ⋯, [106, 227, 244],[106,267,122],[106,267,168]]
货物顺序	[307,306,305, ⋯, 352,351,350]
货物方向	[1, 1, 3, ⋯, 3, 1, 4]
装载货物个数/个	254
集装器容积利用率/%	83.62
集装器载重利用率/%	38.13

表 14　最优装载方案（51 类）

Table 14　Optimal loading scheme (Category 51)

参数	数值
货物位置	[[0, 0, 0], [0, 0, 26,], [0, 0, 52], ⋯, [106, 220, 264],[106,264,264],[0,248,264]]
货物顺序	[216, 217, 218, ⋯, 1 057, 1 058, 1 056]
货物方向	[2, 2, 1, ⋯, 3, 5, 3]
装载货物个数/个	228
集装器容积利用率/%	93.84
集装器载重利用率/%	40.99

物数据(24类货物)输入2种算法中进行计算,对比算法求解的帕累托前沿及参数变化曲线变化情况,如图12和图13所示。

表15为2种算法在24类、487件货物的算例

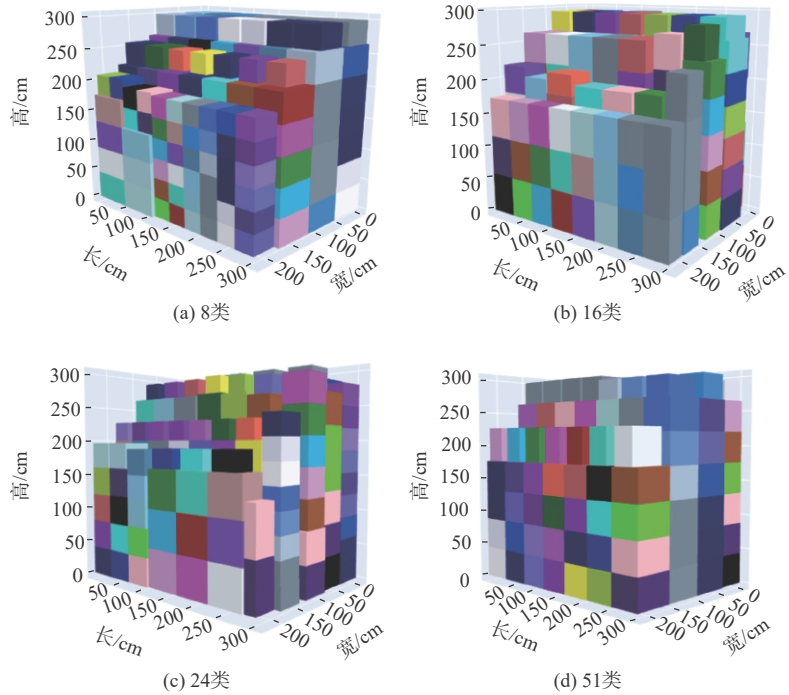


图11 货物装载示意图

Fig. 11 Schematic diagram of cargo loading

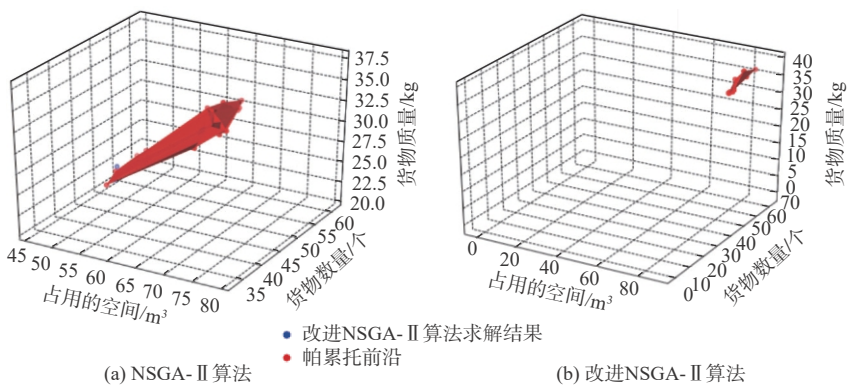


图12 2类算法求得帕累托前沿

Fig. 12 Pareto frontier obtained by two types of algorithms

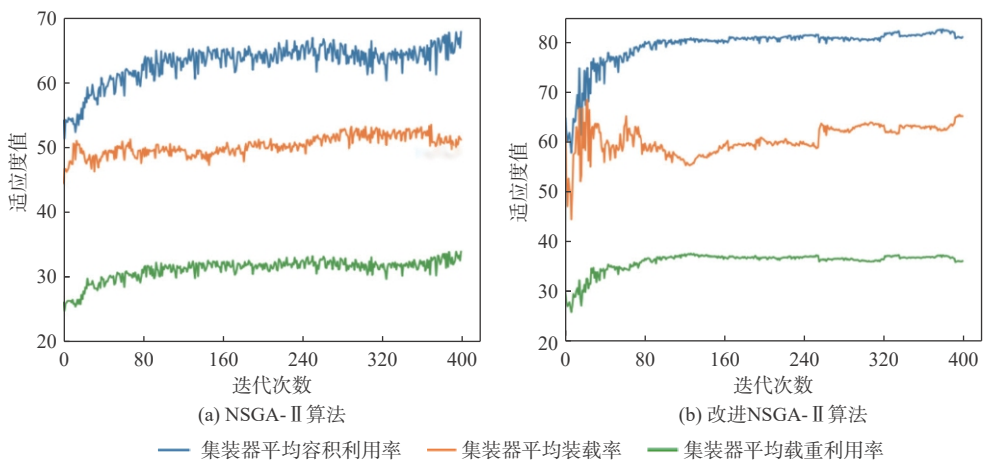


图13 2类算法指标变化趋势曲线

Fig. 13 Index change trend curves by two types of algorithms

表 15　NSGA-Ⅱ算法与改进 NSGA-Ⅱ算法求解结果

Table 15　Solution results of NSGA-Ⅱ algorithm and improved NSGA-Ⅱ algorithm

算法	装载货物数量/个	集装器容积利用率/%	集装器载重利用率/%
NSGA-Ⅱ算法	257	70.56	34.25
改进NSGA-Ⅱ算法	254	83.62	38.13

条件下的装载情况, 本文算法求解得到的装载结果明显优于 NSGA-Ⅱ算法。在集装器的容积利用率方面, 本文算法为 83.62%, 而 NSGA-Ⅱ算法只有 70.56%, 在载重利用率上, 本文算法也较优。表明本文算法相较于 NSGA-Ⅱ算法在复杂强异构货物装载上有更佳的表现, 有利于不规则集装器的装载, 提升机场货站内不规则集装器的装载效率。

5　结　论

1) 在多种现实约束限制下, 建立了不规则集装器多目标装载优化模型, 最大化利用集装器容积、货物装载数量及集装器载重量, 并设计了一种改进 NSGA-Ⅱ算法进行求解。

2) 对于 NSGA-Ⅱ算法的改进主要为: 总结货物装载规则, 设计基于最深处装载的三空间分割启发式算法, 对初始解求解进行优化; 采用二倍染色体表示货物的顺序和旋转值; 使用改进的顺序交叉运算符和两步突变运算符, 使其更利于得到最优装载方案; 加入非支配排序、拥挤度排序及精英选择策略, 进一步提高改进 NSGA-Ⅱ算法的优化效果, 在解集规模相同的情况下, 解集的分布更加集中且收敛速度快。

3) 利用机场货站内的实际运单数据, 设计算例并进行实验分析, 在更接近实际装载情况的复杂强异构货物装载条件下, 保持 93.84% 的高容积利用率。在相同算例下, 比初始 NSGA-Ⅱ算法求解结果高 13%。表明本文算法能够很好地解决机场货站内不规则集装器装载问题。

4) 本文只考虑了“收角”类型的不规则集装器装载问题, 但实际机场货站内的不规则集装器还包括“飘板”类型, 在未来的研究中需要继续考虑这类不规则集装器的装载优化。

参考文献（References）

[1] BISCHOFF E E, RATCLIFF M S W. Issues in the development of approaches to container loading[J]. Omega, 1995, 23(4): 377-390.

[2] LIU S, SHANG X Q, CHENG C J, et al. Heuristic algorithm for the container loading problem with multiple constraints[J]. Computers & Industrial Engineering, 2017, 108: 149-164.

[3] 廖云峰, 单鸿涛, 赵文洁. 多约束三维装箱问题的启发式优化算法[J]. 软件导刊, 2022, 21(5): 101-104.
LIAO Y F, SHAN H T, ZHAO W J. Heuristic optimization algorithm for multi-constrained threedimensional container loading problem[J]. Software Guide, 2022, 21(5): 101-104(in Chinese).

[4] 吕雪菊, 倪静. 基于自适应空间划分策略的三维装箱问题[J]. 系统工程, 2020, 38(4): 95-102.
LV X J, NI J. Three dimensional container loading problem based on adaptive spatial partition strategy[J]. Systems Engineering, 2020, 38(4): 95-102(in Chinese).

[5] 赵桂红, 杨文君, 孟超, 等. 基于混合遗传粒子群算法的民航宽体客机腹舱装载优化[J]. 数学的实践与认识, 2021, 51(11): 47-57.
ZHAO G H, YANG W J, MENG C, et al. Optimization of belly cabin loading of civil wide-body passenger aircraft based on hybrid genetic particle swarm optimization[J]. Mathematics in Practice and Theory, 2021, 51(11): 47-57(in Chinese).

[6] LI Y Z, TAO Y, WANG F. A compromised large-scale neighborhood search heuristic for capacitated air cargo loading planning[J]. European Journal of Operational Research, 2009, 199(2): 553-560.

[7] MIKRAM M, RHANOUI M, YOUSFI S, et al. Unit load devices (ULD) demand forecasting in the air cargo for optimal cost management[J]. Journal of Automation, Mobile Robotics and Intelligent Systems, 2019, 14(3): 71-80.

[8] 陈丙成, 李艳华. 基于混合遗传算法的多航空集装器装箱优化设计[J]. 包装工程, 2021, 42(5): 181-186.
CHEN B C, LI Y H. Optimal design of packing for multi unit load device based on hybrid genetic algorithm[J]. Packaging Engineering, 2021, 42(5): 181-186(in Chinese).

[9] 张长勇, 刘佳瑜. 基于混合遗传算法的多箱型集装箱装载问题分析[J]. 北京航空航天大学学报, 2022, 48(5): 747-755.
ZHANG C Y, LIU J Y. Multi-box container loading problem based on hybrid genetic algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(5): 747-755(in Chinese).

[10] 张长勇, 翟一鸣. 面向不规则集装箱的货物装载优化问题研究[J]. 计算机应用与软件, 2022, 39(11): 238-244.
ZHANG C Y, ZHAI Y M. Optimization of cargo loading for irregular containers[J]. Computer Applications and Software, 2022, 39(11): 238-244(in Chinese).

[11] ARAYA I, MOYANO M, SANCHEZ C. A beam search algorithm for the biobjective container loading problem[J]. European Journal of Operational Research, 2020, 286(2): 417-431.

[12] 于明正, 徐斌, 陈佳. 基于双层启发式遗传算法的三维装箱问题[J]. 科学技术与工程, 2020, 20(5): 2042-2047.
YU M Z, XU B, CHEN J. 3D packing problem based on double-layer heuristic genetic algorithm[J]. Science Technology and Engineering, 2020, 20(5): 2042-2047(in Chinese).

[13] FIGUEIREDO E M N, LUDERMIR T B, BASTOS-FILHO C J A. Many objective particle swarm optimization[J]. Information Sciences, 2016, 374: 115-134.

[14] DEB K, PRATAP A, AGARWAL S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-Ⅱ[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182-197.

[15] MAHVASH B, AWASTHI A, CHAUHAN S. A column generation-based heuristic for the three-dimensional bin packing problem with rotation[J]. The Journal of the Operational Research Society, 2018, 69(1): 78-90.

[16] BORGULYA I. A hybrid evolutionary algorithm for the offline bin

packing problem[J]. Central European Journal of Operations Research, 2021, 29(2): 425-445.

[17] WANG H F, CHEN Y J. A hybrid genetic algorithm for 3D bin packing problems[C]//Proceedings of the 2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications. Piscataway: IEEE Press, 2010: 703-707.

Loading optimization of irregular unit load device based on improved NSGA-Ⅱ algorithm

WEN Gan[1], YUAN Leifeng[2], WANG Xudong[3], BI Jun[1,*], WANG Yongxing[1]

(1. School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China;
2. China Travelsky Holding Company, Beijing 101300, China;
3. Infosky Technology Co., Ltd., Beijing 101300, China)

Abstract: The unit load device (ULD) is commonly shaped irregularly to fit the arc shape inside the cabin. This results in efficient loading of irregular ULD at the airport cargo station, which helps to increase the overall efficiency of air freight transportation. Considering a series of actual loading constraints such as cargo volume constraints, a multi-objective loading optimization model for irregular ULD is established with the goal of maximizing the utilization rate of volume, loading cargo quantity, and load utilization rate. Moreover, a heuristic algorithm is constructed based on the principles of three-dimensional space division; the NSGA-Ⅱ algorithm is used to perform operations such as crossover, mutation, and selection. Finally, the algorithm generates a Pareto frontier set, and selects the optimal loading scheme with the greatest volume utilization. The experimental results show that the improved NSGA-Ⅱ algorithm still maintain a high volume utilization rate of 93.84% under strong heterogeneous cargo loading conditions; compared to the initial NSGA-Ⅱ algorithm, the proposed algorithm can increase the volume utilization rate by 13%, indicating that the improved NSGA-Ⅱ algorithm has good adaptability on irregular ULD loading. The research results can provide a reference for irregular ULD loading at airport cargo stations.

Keywords: airport cargo station; unit load device loading; irregular unit load device; muti-objective optimization; NSGA-Ⅱ algorithm

Received: 2023-03-29; **Accepted**: 2023-04-27; **Published Online**: 2023-05-16 10:37

URL: link.cnki.net/urlid/11.2625.V.20230515.1746.002

Foundation item: National Key Research and Development Program of China (2018YFC0706005)

* **Corresponding author**. E-mail: jbi@bjtu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0152

前倾转变体无人机固定时间轨迹跟踪控制

窦磊，李新凯*，张宏立，王昊，杨加秀
(新疆大学 电气工程学院，乌鲁木齐 830017)

摘　　要：针对具有时变扰动的前倾转变体无人机轨迹跟踪控制问题，提出了一种基于浸入与不变(I&I)理论和固定时间收敛理论的非奇异终端滑模控制方案。融合动态尺度因子设计一种基于I&I的时变扰动观测器；构造一种分段式的固定时间非奇异终端滑模面，消除滑模面的奇异性，并使系统状态在固定时间内收敛到平衡点附近小邻域内，且收敛时间的上界与系统的初始状态无关；基于Lyapunov稳定性理论证明系统的全局固定时间稳定性，并给出其收敛时间的上界。在2种实验场景下验证了所提控制方案的有效性和优越性。与传统的控制方法相比，所提控制方案使系统的收敛速度更快，抗扰动能力更好。

关 键 词：变体无人机；浸入与不变；动态尺度因子；非奇异终端滑模面；固定时间收敛
中图分类号：V221$^+$.3; TB553
文献标志码：A　　**文章编号**：1001-5965(2025)03-1005-13

近年来，凭借自身监测和监视的优势，无人机在环境监测、农业植保、电力巡检等领域的应用受到人们的广泛关注。变体无人机作为其中的一个分支，因其可切换飞行模式而适应动态的飞行环境，在复杂的环境中表现出良好的稳定性和灵活性，从而更具有潜在的应用前景[1-3]。但变体无人机飞行模式的切换涉及机械结构的变换，因而其数学模型相比于固定构型的传统无人机更加复杂，这对于设计控制器的高效性和稳定性提出了更高的要求。前倾转变体无人机[4]融合了传统多旋翼和固定翼无人机的特性，具有可快速垂直起降、可悬停任务侦察、可切换飞行模式适应动态环境等显著优点，但由于其欠驱动、强耦合等特性，使得变体无人机在面对外部扰动，特别是时变扰动时非常敏感，同时，还要抑制在飞行模式切换时产生的内部扰动，这些都大大增加了控制的难度。

浸入与不变(immersion and invariance, I&I)理论不需要构造Lyapunov函数，而是一种几何降阶的方法，是将对象系统映射到某一可能更低阶的系统，并要求该系统具有预先设定的性能[5]。设计基于I&I理论的观测器来估计系统的某些参数，其优势在于系统参数的估计值不取决于控制律但却无限接近于真实值，从而受到人们的广泛关注[5-6]。文献[7]针对倾斜三旋翼无人机的外界干扰问题，基于I&I设计观测器来估计干扰，提高了姿态控制的鲁棒性。文献[8]基于I&I，为具有线性、非线性及时变增益的扩展状态观测器的设计和分析提供了统一的几何框架。近年来，动态尺度因子被提出，并与I&I理论相融合，估计方程中的非线性项[9]。文献[10]针对角速度缺少的刚体姿态跟踪控制问题，基于I&I并融合动态尺度因子设计了一种新型的观测器来估计角速度。文献[11]针对惯性参数未知的欧拉-拉格朗日系统，基于I&I并融合动态尺度因子提出了自适应控制框架，动态尺度因子的引入克服了I&I方法固有的可积性障碍。不可否认，动态尺度因子的引入提高了I&I观测器参数估计的准确性。然而，目前国内外大多数的研究是针对多旋翼无人机，对于涉及飞行模式切换的变体无人

收稿日期：2023-03-31；**录用日期**：2023-05-26；**网络出版时间**：2023-06-28 14:30
网络出版地址：link.cnki.net/urlid/11.2625.V.20230627.1733.002
基金项目：国家自然科学基金(62263030)；新疆维吾尔自治区自然科学基金青年项目(2022D01C86)
* **通信作者**. E-mail：lxk@xju.edu.cn

引用格式：窦磊，李新凯，张宏立，等. 前倾转变体无人机固定时间轨迹跟踪控制[J]. 北京航空航天大学学报，2025，51(3)：1005-1017. DOU L，LI X K，ZHANG H L，et al. Fixed time trajectory tracking control of forward-tilting morphing aerospace vehicle[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：1005-1017 (in Chinese).

机，其应用研究还比较少。

随着变体无人机控制对收敛时间的要求逐渐提高，许多学者开始在如何提高收敛时间方面进行研究。其中，有限时间控制方法[12-14]可使闭环系统的状态变量在某一时间内收敛，但时间界限取决于系统的初始状态。为克服这一局限性，文献[15]在2011年提出了固定时间稳定性概念，其系统状态可在固定时间内收敛，且收敛时间的上界与系统的初始状态无关，仅取决于设计的控制器参数。文献[16]在文献[15]所提概念的基础上，提出了一类新型的固定时间稳定系统。文献[17-18]基于以上系统设计了固定时间收敛滑模面及相应的控制律，分别研究了二阶多智能体和倒立摆系统的轨迹跟踪问题，取得了较好的控制效果。在此基础上，文献[19]设计了一种非奇异终端滑模面，通过倒立摆仿真的对比试验，说明其控制方案优于文献[15-18]。以上文献设计的固定时间控制器虽然保证了系统的固定时间收敛，但其重点是在理论研究上设计收敛速度更快的滑模面，仅对简单系统进行了仿真验证。同时，一些学者将固定时间控制理论应用到数学模型较为复杂的系统，得到了满意的控制效果。文献[20-21]将反步控制和文献[16]中的固定时间稳定理论相结合，保证了收敛时间的上界与系统的初始状态无关，分别研究了吸气式高超声速飞行器和无人水面飞行器的固定时间轨迹跟踪问题。文献[22]同样以文献[16]中的固定时间稳定理论为基础，针对自动驾驶汽车的路径约束问题，使车辆的视距项和角度误差项在固定时间内收敛到平衡点附近的小集合。但是，文献[16]所提出的固定时间理论收敛速度较慢，而文献[17-18]虽然提出了收敛速度更快的固定时间理论，但是同样仅对简单系统进行了仿真验证。因此，以前倾转变体无人机为研究对象，利用文献[17-18]所提出的固定时间稳定理论，设计一种固定时间轨迹跟踪控制策略值得探究。

综上，本文以前倾转变体无人机为研究目标，针对外部时变扰动和飞行模式切换时产生的内部扰动问题，基于固定时间控制理论，结合终端滑模控制方法、I&I理论及动态尺度技术，设计了基于I&I的时变扰动观测器和分段式的固定时间非奇异终端滑模控制器。该控制器的特点是：使系统状态在固定时间内收敛到平衡点附近小邻域内，且收敛时间的上界与系统的初始状态无关，同时，由飞行模式切换产生内部扰动对系统的影响可以被有效抑制。

本文的主要贡献如下：

1）考虑了具有时变扰动的前倾转变体无人机的固定时间轨迹跟踪控制问题。

2）提出了基于I&I并融合动态尺度因子的时变扰动观测器，进一步提高了对时变扰动的估计精度。

3）设计了一种分段式的固定时间非奇异终端滑模面，消除了终端滑模面的奇异性，进一步提高了系统状态的固定时间收敛性能，保证了其收敛时间的上界与初始状态无关。

1 问题描述及预备知识

1.1 前倾转变体无人机数学模型

前倾转变体无人机在飞行过程中的机身受力分析及坐标系框架建立如图1所示。图中：$\mathrm{W}:(O_{\mathrm{w}},x_{\mathrm{w}},y_{\mathrm{w}},z_{\mathrm{w}})$和$\mathrm{B}:(O_{\mathrm{b}},x_{\mathrm{b}},y_{\mathrm{b}},z_{\mathrm{b}})$分别表示世界坐标系和机身坐标系，$x_{\mathrm{w}}$为正北，$y_{\mathrm{w}}$为正西，$z_{\mathrm{w}}$为垂直于地心的反向，$x_{\mathrm{b}}$为正前，$y_{\mathrm{b}}$为正左，$z_{\mathrm{b}}$为垂直于机身的正上方；$F_i(i=1,2,3,4)$为4个旋翼的出力；$F_{\mathrm{L}}^l$和$F_{\mathrm{D}}^l(l=1,2)$分别为左右侧机翼受到的升力和阻力；$\theta_{\mathrm{f}}$为机身与左右侧机翼间的夹角。

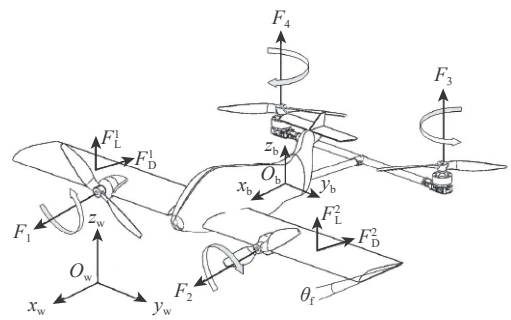


图1 前倾转变体无人机受力分析及坐标系建立

Fig. 1 Force analysis and coordinate system establishment of forward-tilting morphing aerospace vehicle

前倾转变体无人机的垂直起降飞行、垂直水平过渡飞行、水平高速飞行这3种飞行模式的切换是通过前翼倾转的动作实现的，如图2所示。

前倾转变体无人机的动力学方程为

$$\begin{cases} m\ddot{x}=\cos\theta\cos\psi\cos\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}-(\cos\phi\sin\theta\cos\psi+\sin\phi\sin\psi)\cdot\\ \qquad(\sin\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}+u_{\mathrm{r_{sum}}})+W_x\\ m\ddot{y}=\cos\theta\sin\psi\cos\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}-(\cos\phi\sin\theta\sin\psi-\sin\phi\cos\psi)\cdot\\ \qquad(\sin\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}+u_{\mathrm{r_{sum}}})+W_y\\ m\ddot{z}=\sin\theta\cos\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}+\cos\phi\cos\theta(\sin\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}+u_{\mathrm{r_{sum}}})-\\ \qquad mg-W_z\\ I_{xx}\ddot{\phi}=(I_{yy}-I_{zz})\dot{\theta}\dot{\psi}-I_{\mathrm{r}}\dot{\theta}w_q+l_{\mathrm{s}}(\sin\theta_{\mathrm{f}}u_{\mathrm{f_{dif}}}-u_{\mathrm{r_{dif}}})-\\ \qquad\vartheta\cos\theta_{\mathrm{f}}u_{\mathrm{f_{dif}}}\\ I_{yy}\ddot{\theta}=(I_{zz}-I_{xx})\dot{\phi}\dot{\psi}+I_{\mathrm{r}}(\dot{\phi}w_q+\dot{\psi}w_{\mathrm{r}})+l_{\mathrm{l}}(\sin\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}+u_{\mathrm{r_{sum}}})+\\ \qquad I_{yy}W_{\mathrm{t}}\\ I_{zz}\ddot{\psi}=(I_{xx}-I_{yy})\dot{\phi}\dot{\theta}-I_{\mathrm{r}}\dot{\theta}w_{\mathrm{r}}+l_{\mathrm{s}}\cos\theta_{\mathrm{f}}u_{\mathrm{f_{dif}}}+\\ \qquad\vartheta(\sin\theta_{\mathrm{f}}u_{\mathrm{f_{dif}}}+u_{\mathrm{r_{dif}}}) \end{cases} \tag{1}$$

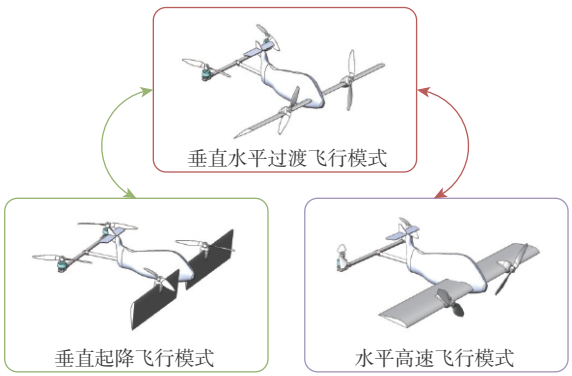


图 2　前倾转变体无人机飞行模式切换

Fig. 2　Flight mode switching of forward-tilting morphing aerospace vehicle

式中：$\theta_{\mathrm{f}} \in [0,\pi/2]$；$w_q = \sin\theta_{\mathrm{f}} w_1 - \sin\theta_{\mathrm{f}} w_2 + w_3 - w_4$；$u_{\mathrm{f_{sum}}} = k(w_1^2 + w_2^2)$；$u_{\mathrm{r_{sum}}} = k(w_3^2 + w_4^2)$；$u_{\mathrm{f_{dif}}} = k(w_1^2 - w_2^2)$；$u_{\mathrm{r_{dif}}} = k(w_3^2 - w_4^2)$；$w_r = \cos\theta_{\mathrm{f}} w_1 - \cos\theta_{\mathrm{f}} w_2$；$w_i (i = 1,2,3,4)$为4个电机的转速；$k$为转速-力系数；$m$为机身总质量；$g$为重力加速度；$\boldsymbol{\xi} = [x,y,z]^{\mathrm{T}}$和$\boldsymbol{\eta} = [\phi,\theta,\psi]^{\mathrm{T}}$分别为机身在世界坐标系中的位置和姿态矩阵，$\phi$、$\theta$和$\psi$分别为横滚角、俯仰角和偏航角；$I_{xx}$、$I_{yy}$、$I_{zz}$分别为机身沿各惯性轴的惯性矩；$I_{\mathrm{r}}$为转动部件绕$z_{\mathrm{b}}$轴的转动惯量；$l_{\mathrm{l}}$、$l_{\mathrm{s}}$分别为左右机翼沿$x_{\mathrm{b}}$、$y_{\mathrm{b}}$轴方向到重心的距离；$\vartheta$为转矩/力的比率。

W_x、W_y、W_z分别为机翼沿x_{w}、y_{w}、z_{w}轴的空气动力学参数，表示为

$$\begin{bmatrix} W_x \\ W_y \\ W_z \end{bmatrix} = \boldsymbol{R}_{\mathrm{wb}} \begin{bmatrix} F_{\mathrm{D}}^1(\theta_{\mathrm{f}},v_x,v_z) + F_{\mathrm{D}}^2(\theta_{\mathrm{f}},v_x,v_z) \\ 0 \\ F_{\mathrm{L}}^1(\theta_{\mathrm{f}},v_x,v_z) + F_{\mathrm{L}}^2(\theta_{\mathrm{f}},v_x,v_z) \end{bmatrix} \tag{2}$$

式中：$\boldsymbol{R}_{\mathrm{wb}}$为机身到世界坐标系的转换矩阵，表示为

$$\boldsymbol{R}_{\mathrm{wb}} = \begin{bmatrix} \cos\theta\cos\psi & \sin\phi\sin\theta\cos\psi - \cos\phi\sin\psi & \cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi \\ \cos\theta\sin\psi & \sin\phi\sin\theta\sin\psi + \cos\phi\cos\psi & \cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta \end{bmatrix} \tag{3}$$

$W_{\mathrm{t}} = l_{\mathrm{l}}(F_{\mathrm{L}}^1(\theta_{\mathrm{f}},v_x,v_z) + F_{\mathrm{L}}^2(\theta_{\mathrm{f}},v_x,v_z))$是机翼受到的升力和阻力在$y_{\mathrm{b}}$轴方向产生的转矩。机翼产生的升力$F_{\mathrm{L}}^i(\theta_{\mathrm{f}},v_x,v_z)$和阻力$F_{\mathrm{D}}^i(\theta_{\mathrm{f}},v_x,v_z)$为

$$\begin{bmatrix} F_{\mathrm{D}}^i \\ 0 \\ F_{\mathrm{L}}^i \end{bmatrix} = \boldsymbol{R}(\theta_{\mathrm{f}}) \begin{bmatrix} -\frac{1}{2} C_{\mathrm{D}} \theta_{\mathrm{f}} \rho A v_{\mathrm{f}}^2 \\ 0 \\ -\frac{1}{2} C_{\mathrm{L}} \theta_{\mathrm{f}} \rho A v_{\mathrm{f}}^2 \end{bmatrix} \tag{4}$$

式中：$i = 1,2$分别表示左、右侧机翼；$\boldsymbol{R}(\theta_{\mathrm{f}})$为沿$y$轴方向将机翼上的力垂直分解到机身轴上的旋转矩阵；$C_{\mathrm{D}}$和$C_{\mathrm{L}}$分别为机翼的阻力系数和升力系数；$\rho$为空气密度；$A$为机翼总面积；$v_{\mathrm{f}} = (v_x^2 + v_z^2)^{1/2}$为气流速度，$v_x$和$v_z$分别为机身在$x_{\mathrm{b}}$和$z_{\mathrm{b}}$轴方向上的速度。

前倾转变体无人机的动力学方程可转化为以下非线性系统形式：

$$\begin{cases} \dot{\boldsymbol{X}}_1 = \dot{\boldsymbol{X}}_2 \\ \dot{\boldsymbol{X}}_2 = \boldsymbol{f} + \boldsymbol{B}\boldsymbol{\varsigma} + \boldsymbol{d} \end{cases} \tag{5}$$

式中：$\boldsymbol{X}_1 = [\boldsymbol{\xi}_{3\times1}, \boldsymbol{\eta}_{3\times1}]^{\mathrm{T}}$；$\boldsymbol{X}_2 = [\dot{\boldsymbol{\xi}}_{3\times1}, \dot{\boldsymbol{\eta}}_{3\times1}]^{\mathrm{T}}$；$\boldsymbol{\varsigma} = [\boldsymbol{\tau}_{3\times1}, \boldsymbol{u}_{3\times1}]^{\mathrm{T}} = [\tau_1,\tau_2,\tau_3,u_1,u_2,u_3]^{\mathrm{T}}$；$\boldsymbol{B} = \mathrm{diag}\{\boldsymbol{b}_{\xi_{3\times3}}, \boldsymbol{b}_{\eta_{3\times3}}\} = \mathrm{diag}\{1/m,1/m,1/m,1/I_{xx},1/I_{yy},1/I_{zz}\}$；$\boldsymbol{d} = [\boldsymbol{d}_{\xi_{3\times1}}^{\mathrm{T}}, \boldsymbol{d}_{\eta_{3\times1}}^{\mathrm{T}}]^{\mathrm{T}} = [d_1, d_2, d_3, d_4, d_5, d_6]^{\mathrm{T}}$。

$$\boldsymbol{f} = \begin{bmatrix} \boldsymbol{f}_{\xi_{3\times1}} \\ \boldsymbol{f}_{\eta_{3\times1}} \end{bmatrix} \begin{bmatrix} W_x/m \\ W_y/m \\ g + W_z/m \\ (I_{yy} - I_{zz})\dot{\theta}\dot{\psi}/I_{xx} - I_{\mathrm{r}}\dot{\theta}w_q/I_{xx} \\ (I_{zz} - I_{xx})\dot{\phi}\dot{\psi}/I_{yy} + I_{\mathrm{r}}(\dot{\phi}w_q + \dot{\psi}w_r)/I_{yy} + W_t \\ (I_{xx} - I_{yy})\dot{\phi}\dot{\theta}/I_{zz} - I_{\mathrm{r}}\dot{\theta}w_r/I_{zz} \end{bmatrix}$$

假设 1　前倾转变体无人机在飞行过程中，机身与左右侧机翼的夹角θ_{f}相同，并且左机翼产生的升阻力与右机翼产生的升阻力相同。

假设 2　时变扰动$\boldsymbol{d}(t)$是连续可导的，且满足$\|\boldsymbol{d}(t)\| \leqslant \Delta$，其中，$\Delta$为 Lipschitz 常数，是时变扰动的上界。

假设 3　姿态子系统的状态变量ϕ、θ和ψ是完全反馈的，且满足条件范围$-\pi/2 < \theta < \pi/2$和$-\pi/2 < \phi < \pi/2$。

1.2　控制目标

在满足假设 1 和假设 3 及时变扰动$\boldsymbol{d}(t)$满足假设 2 的情况下，本文的控制目标是设计固定时间非奇异终端滑模控制律，从而实现前倾转变体无人机的快速高精度轨迹跟踪控制，并使系统的跟踪误差在固定时间内收敛。

本文采用内外环控制策略，外环是位置子系统的控制，内环是姿态子系统的控制，前倾转变体无人机系统的整体控制策略示意图如图 3 所示。

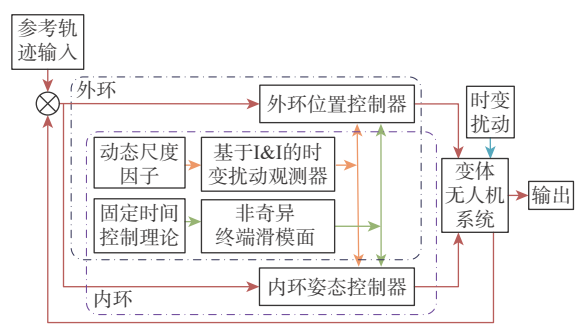


图 3　本文方案的控制策略

Fig. 3　Control strategy of the proposed scheme

根据预设轨迹，将位置外环的实际控制输出转换为姿态内环的虚拟控制输入，进而可计算出相应的理想姿态角。转换后的虚拟控制输入为

$$\begin{cases}\tau_1 = m(u_{\mathrm{f_{sum}}}+u_{\mathrm{r_{sum}}})-W_x=\cos\theta\cos\psi\cos\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}-\\ \quad(\cos\phi\sin\theta\cos\psi+\sin\phi\sin\psi)(\sin\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}+u_{\mathrm{r_{sum}}})\\ \tau_2 = m(u_{\mathrm{f_{sum}}}+u_{\mathrm{r_{sum}}})-W_y=\cos\theta\sin\psi\cos\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}-\\ \quad(\cos\phi\sin\theta\sin\psi-\sin\phi\cos\psi)(\sin\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}+u_{\mathrm{r_{sum}}})\\ \tau_3 = m(u_{\mathrm{f_{sum}}}+u_{\mathrm{r_{sum}}})-mg-W_z=\\ \quad-\sin\theta\cos\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}-\cos\phi\cos\theta(\sin\theta_{\mathrm{f}}u_{\mathrm{f_{sum}}}+u_{\mathrm{r_{sum}}})\end{cases} \tag{6}$$

通过反解，总推力大小和理想的姿态角由式(6)可得

$$\begin{cases}u_1=\sqrt{\tau_1^2+\tau_2^2+\tau_3^2}=\sqrt{v_1^2+v_2^2}\\ \phi_{\mathrm{d}}=\arcsin(-\lambda_1/v_2)\\ \theta_{\mathrm{d}}=\arcsin(-v_1\tau_3-\cos\phi_{\mathrm{d}}v_2\lambda_2/\lambda_2^2+\tau_3^2)\end{cases} \tag{7}$$

式中：$\lambda_1=\tau_1\sin\psi-\tau_2\cos\psi$；$\lambda_2=\tau_1\cos\psi+\tau_2\sin\psi$；$v_1=\lambda_2\cos\theta-\tau_3\sin\theta$；$v_2=-(\lambda_2\sin\theta+\tau_3\cos\theta)/\cos\phi$。

将 $\lambda_2^2+\tau_3^2=v_1^2+\cos\phi^2v_2^2$ 和 $\tau_1^2+\tau_2^2+\tau_3^2=v_1^2+v_2^2$ 联立，可解算出v_1、v_2。

1.3 基本定义与引理

考虑如下非线性系统：

$$\begin{cases}\dot{x}(t)=F(t,x)\\ x(0)=x_0\end{cases} \tag{8}$$

式中：$x\in\mathbf{R}^n$表示系统状态；$F(t,x):\mathbf{R}^+\times\mathbf{R}^n\to\mathbf{R}^n$表示连续的非线性函数。假设系统的平衡点为原点。

定义 1 若系统(8)的解$x(t,x_0)$于有限时间内到达平衡点，且系统(8)的平衡点$x=0$是渐近稳定的，则原点是系统的有限时间收敛平衡点，有

$$\begin{cases}\lim\limits_{x\to T(x_0)}x(t,x_0)\to 0\\ x(t,x_0)=0\qquad\forall t>T(x_0)\end{cases} \tag{9}$$

式中：$T(x_0)$为收敛时间。

定义 2 若系统(8)的平衡点$x=0$是有限时间稳定的，且$T(x_0)$有实数上界$T_{\max}>0$，则原点是系统的固定时间收敛平衡点，$T(x_0)\leqslant T_{\max}$，$\forall x_0\in\mathbf{R}^n$。

定义 3 定义$\mathrm{sign}^a y=\mathrm{sign}(y)|y|^a$，其中，$\mathrm{sign}(\cdot)$为符号函数，且$\mathrm{sign}(0)=0$。

定义 4 定义$\Gamma(x)=(\mathrm{e}^x-x)/(\mathrm{e}^{-x}+x)$，其导数为$\dot{\Gamma}(x)=2+x\mathrm{e}^x-x\mathrm{e}^{-x}-\mathrm{e}^x-\mathrm{e}^{-x}$。

引理 1[15] 考虑如下非线性系统：

$$\begin{cases}\dot{y}=-l_1\mathrm{sign}^{m_1}y-l_2\mathrm{sign}^{m_2}y\\ y(0)=y_0\end{cases} \tag{10}$$

式中：$\mathrm{sign}^{m_i}y=\mathrm{sign}(y)|y|^{m_i}$，$i=1,2$，$l_1>0$，$l_2>0$，$m_1>1$，$0<m_2<1$。

则系统(10)是固定时间稳定的，其收敛时间T有上界$T_{\max}$，满足：

$$T<T_{\max}=\frac{1}{l_1(m_1-1)}+\frac{1}{l_2(1-m_2)} \tag{11}$$

引理 2[23] 若有 Lyapunov 函数$V(x)$满足条件$\dot{V}(x)\leqslant -l_1V^{m_1}(x)-l_2V^{m_2}(x)$，其中，$l_1>0$，$l_2>0$，$m_1>1$，$0<m_2<1$，则系统是固定时间稳定的，其收敛时间$T$有上界$T_{\max}$，满足：

$$T<T_{\max}=\frac{1}{l_3(m_4-1)}+\frac{1}{l_3(1-m_4)} \tag{12}$$

引理 3[18] 考虑如下非线性系统：

$$\begin{cases}\dot{y}=-l_1\mathrm{sign}^{k_1}y-l_2\mathrm{sign}^{k_2}y\\ y(0)=y_0\end{cases} \tag{13}$$

式中：$\mathrm{sign}^{k_i}y=\mathrm{sign}(y)|y|^{k_i}$，$i=1,2$，$l_1>0$，$l_2>0$，$k_1=\dfrac{m_1+1}{2}+\dfrac{m_1-1}{2}\mathrm{sign}(|y|-1)$，$m_1>1$，$k_2=\dfrac{m_2+1}{2}+\dfrac{1-m_2}{2}\mathrm{sign}(|y|-1)$，$0<m_2<1$。

则系统(13)是固定时间稳定的，其收敛时间T有上界$T_{\max}$，满足：

$$T<T_{\max}=\frac{1}{l_2(m_1-1)}\ln\left(1+\frac{l_2}{l_1}\right)+\frac{1}{l_1(1-m_2)}\ln\left(1+\frac{l_1}{l_2}\right) \tag{14}$$

证明 1 由系统(13)可得

$$\dot{y}=\begin{cases}-l_1\mathrm{sign}^{m_1}y-l_2y & |y|>1\\ -l_1y-l_2\mathrm{sign}^{m_2}y & |y|\leqslant 1\end{cases} \tag{15}$$

用变量z代换，可得

$$\begin{cases}z=\begin{cases}\ln|y|+1 & |y|>1\\ |y|^{1-m_2} & |y|\leqslant 1\end{cases}\\ z(0)=z_0\end{cases} \tag{16}$$

则新系统为

$$\dot{z}=\begin{cases}-l_1\mathrm{e}^{(m_1-1)(z-1)}-l_2 & z>1\\ -(1-m_2)l_1z-(1-m_2)l_2 & 0<z\leqslant 1\end{cases} \tag{17}$$

系统的收敛时间上界$T_{\max}$为

$$\begin{aligned}T_{\max}&=\lim_{y_0\to+\infty}T(y_0)=\lim_{z_0\to+\infty}T(z_0)=\\ &\lim_{z_0\to+\infty}\left(\int_0^1\frac{1}{(1-m_2)(l_1z+l_2)}\mathrm{d}z+\int_1^{z_0}\frac{1}{l_1\mathrm{e}^{(m_1-1)(z-1)}+l_2}\mathrm{d}z\right)=\\ &\frac{1}{l_1(1-m_2)}\ln\left(1+\frac{l_1}{l_2}\right)+\lim_{z_0\to+\infty}\left(\int_1^{z_0}\frac{1}{l_1\mathrm{e}^{(m_1-1)(z-1)}+l_2}\mathrm{d}z\right)=\\ &\frac{1}{l_1(1-m_2)}\ln\left(1+\frac{l_1}{l_2}\right)+T_0\end{aligned} \tag{18}$$

令$\rho=\mathrm{e}^{(m_1-1)(z-1)}$，$\rho(0)=\rho_0$，则可得

$$\begin{aligned}T_0&=\lim_{\rho_0\to+\infty}\left(\int_1^{\rho_0}\frac{1}{\rho(l_1\rho+l_2)}\mathrm{d}\rho\right)=\\ &\lim_{\rho_0\to+\infty}\left(\frac{1}{l_2(m_1-1)}\int_1^{\rho_0}\left(\frac{1}{\rho}-\frac{l_1}{l_1\rho+l_2}\right)\mathrm{d}\rho\right)=\\ &\frac{1}{l_2(m_1-1)}\ln\left(\frac{l_1+l_2}{l_1}\right)\end{aligned} \tag{19}$$

综上, 系统的收敛时间上界$T_{\max}$为

$$T_{\max}=\frac{1}{l_1(1-m_2)}\ln\left(1+\frac{l_1}{l_2}\right)+\frac{1}{l_2(m_1-1)}\ln\left(1+\frac{l_2}{l_1}\right) \tag{20}$$

证毕

注 1　由于$\ln(1+l_1/l_2)\leqslant l_1/l_2$且$\ln(1+l_2/l_1)\leqslant l_2/l_1$, 系统 (13) 的收敛时间上界小于系统 (10) 的收敛时间上界。

2　基于 I&I 理论的时变扰动观测器

本节以位置子系统为例, 基于 I&I 理论设计了时变扰动观测器, 并引入了动态尺度因子进一步提高观测器估计扰动的精度。

考虑系统 (5) 中的位置子系统, 定义位置误差为: $\boldsymbol{e}_1=\boldsymbol{\xi}-\boldsymbol{\xi}_{\mathrm{d}}$, $\dot{\boldsymbol{e}}_1=\dot{\boldsymbol{\xi}}-\dot{\boldsymbol{\xi}}_{\mathrm{d}}$, 进而可得速度误差为: $\boldsymbol{v}_e=\dot{\boldsymbol{e}}_1=\boldsymbol{v}-\dot{\boldsymbol{\xi}}_{\mathrm{d}}$。虚拟控制量可设计为

$$\begin{cases}\boldsymbol{c}_{\xi}=-\boldsymbol{h}_{\xi}\arctan\boldsymbol{e}_1\\ h_{\xi_i}>0\end{cases} \tag{21}$$

则可得虚拟速度误差为

$$\boldsymbol{e}_2=\boldsymbol{v}_e-\boldsymbol{c}_{\xi}=\dot{\boldsymbol{e}}_1-\boldsymbol{c}_{\xi} \tag{22}$$

式中: $\boldsymbol{e}_1=[e_{11},e_{12},e_{13}]^{\mathrm{T}}$; $\boldsymbol{e}_2=[e_{21},e_{22},e_{23}]^{\mathrm{T}}$。

将式 (5) 代入式 (22) 中, 可得位置系统误差动态模型:

$$\begin{cases}\dot{\boldsymbol{e}}_1=\boldsymbol{e}_2-\boldsymbol{h}_{\xi}\arctan\boldsymbol{e}_1\\ \dot{\boldsymbol{e}}_2=\boldsymbol{f}_{\xi}+\boldsymbol{b}_{\xi}\boldsymbol{\tau}+\boldsymbol{d}_{\xi}-\ddot{\boldsymbol{\xi}}_{\mathrm{d}}-\dot{\boldsymbol{c}}_{\xi}\end{cases} \tag{23}$$

式中: $\dot{c}_{\xi_i}=-\dfrac{h_{\xi_i}\dot{e}_{1i}}{1+e_{1i}^2}, i=1,2,3$。

基于 I&I 理论, 前倾转变体无人机的外部扰动的估计误差流型可定义为

$$\begin{cases}\tilde{d}_i=\hat{d}_i-\gamma_i d_i\\ \gamma_i=\gamma_i(e_{2i})=\Gamma(e_{2i})\end{cases} \tag{24}$$

式中: $\hat{d}_i$为扰动的估计值; γ_i为自定义的补偿函数, 用来弥补$\hat{d}_i$和d_i之间的误差。

对$\tilde{d}_i$求导可得

$$\dot{\tilde{d}}_i=\dot{\hat{d}}_i-\frac{\partial\gamma_i}{\partial e_{2i}}\dot{e}_{2i}d_i=\dot{\hat{d}}_i-\dot{\Gamma}(e_{2i})\frac{\hat{d}_i-\tilde{d}_i}{\gamma_i}\dot{e}_{2i} \tag{25}$$

为了确保外部扰动估计误差流型的吸引力, 时变扰动观测器可设计为

$$\dot{\hat{d}}_i=\gamma_i^{-1}(\hat{d}_i-\tilde{d}_i\exp(\dot{e}_{2i}))\dot{\Gamma}(e_{2i})\dot{e}_{2i} \tag{26}$$

将式 (26) 代入式 (25) 可得

$$\dot{\tilde{d}}_i=\gamma_i^{-1}\tilde{d}_i(1-\exp(\dot{e}_{2i}))\dot{\Gamma}(e_{2i})\dot{e}_{2i} \tag{27}$$

为了提高扰动估计的误差精度, 引入动态尺度因子r_i, 尺度估计误差定义为

$$z_i=\frac{\tilde{d}_i}{r_i(t)} \tag{28}$$

对z_i求导得到动态形式为

$$\dot{z}_i=\frac{\dot{\tilde{d}}_i r_i-\tilde{d}_i\dot{r}_i}{r_i^2(t)}=\gamma_i^{-1}(1-\exp(\dot{e}_{2i}))\dot{\Gamma}(e_{2i})\dot{e}_{2i}z_i-\frac{\dot{r}_i}{r_i}z_i \tag{29}$$

r_i的动态形式构建如下:

$$\dot{r}_i=k_{r_i}r_i e_{2i}^2 \tag{30}$$

式中: $r(0)=0$, $\dot{r}\geqslant 0$, $k_{r_i}>0$且$r(t)\geqslant 1$, $t\geqslant 0$。

为了证明z_i的有界性, 考虑 Lyapunov 函数:

$$V_{z_i}=\frac{1}{2}z_i^2 \tag{31}$$

其导数为

$$\dot{V}_{z_i}=z_i\dot{z}_i=\frac{\dot{\tilde{d}}_i r_i-\tilde{d}_i\dot{r}_i}{r_i^2(t)}=\gamma_i^{-1}(1-\exp(\dot{e}_{2i}))\dot{\Gamma}(e_{2i})\dot{e}_{2i}z_i^2-\frac{\dot{r}_i}{r_i}z_i^2 \tag{32}$$

当$t\geqslant 0$时, $(1-\exp(\dot{e}_{2i}))\dot{e}_{2i}\leqslant 0$, 故$\dot{V}_{z_i}\leqslant 0$, 尺度估计误差$z_i$有界, 且渐进收敛到原点。

由于$\tilde{d}_i=r_i z_i$, 当r_i有界时, $\tilde{d}_i$可收敛到 0, 考虑 Lyapunov 函数:

$$V_{r_i}=r_i\mathrm{e}^{-r_i} \tag{33}$$

式中: V_{r_i}连续, $V_{r_i}(0)=0$且$V_{r_i}>0$。

对V_{r_i}求导, 将式 (30) 代入可得

$$\dot{V}_{r_i}=(\mathrm{e}^{-r_i}-r_i\mathrm{e}^{-r_i})\dot{r}_i=k_{r_i}(1-r_i)\mathrm{e}^{-r_i}r_i e_{2i}^2 \tag{34}$$

式中: $(1-r_i)\mathrm{e}^{-r_i}\leqslant 0$, $\dot{r}_i(t)\geqslant 0$, $V_{r_i}\leqslant 0$, 故r_i有界。

因此, 尺度估计误差z_i及扰动误差$\tilde{d}_i$可渐进收敛到原点, 即: $z_i\to 0$, $\tilde{d}_i\to 0$。

3　固定时间非奇异终端滑模控制器

本节基于固定时间收敛理论提出了一种新型的非奇异终端滑模面, 并以位置子系统为例, 设计了位置控制器, 并对系统状态进行了固定时间稳定性证明。

为了消除终端滑模面的奇异性及加快收敛速率, 本文设计了一种分段式的滑模面:

$$s_i=e_{2i}+\alpha_1\mathrm{sign}^{k_1}e_{1i}+\beta_1\varphi(e_{1i}) \tag{35}$$

式中:

$$\varphi(e_{1i})=\begin{cases}k_2 e_{1i}^2\mathrm{sign}(e_{1i}) & |e_{1i}|<\varepsilon\\ \mathrm{sign}^{k_2}e_{1i} & |e_{1i}|\geqslant\varepsilon\end{cases} \tag{36}$$

其中: $\alpha_1>0$, $\beta_1>0$, ε为一个尽量小的正常数, $k_1=\dfrac{m_1+1}{2}+\dfrac{m_1-1}{2}\mathrm{sign}(|e_{1i}|-1)$, $m_1>1$, $k_2=\dfrac{m_2+1}{2}+\dfrac{1-m_2}{2}\mathrm{sign}(|e_{1i}|-1)$, $0<m_2<1$。

式 (35) 求导可得

$$\dot{s}_i=\begin{cases}\dot{e}_{2i}+\alpha_1k_1|e_{1i}|^{k_1-1}\dot{e}_{1i}+2\beta_1k_2e_{1i}\dot{e}_{1i}\mathrm{sign}(e_{1i}) & |e_{1i}|<\varepsilon\\ \dot{e}_{2i}+\alpha_1k_1|e_{1i}|^{k_1-1}\dot{e}_{1i}+\beta_1k_2|e_{1i}|^{k_2-1}\dot{e}_{1i} & |e_{1i}|\geqslant\varepsilon\end{cases} \tag{37}$$

由于当$e_{1i}=0$时，式(37)中没有负幂次项，消除了滑模面的奇异性。

选择滑模趋近律为

$$\dot{s}_i=-\alpha_2\mathrm{sign}^{g_1}s_i-\beta_2\mathrm{sign}^{g_2}s_i-\mu s_i\mathrm{sign}(s_i)-\tilde{d}_i/\gamma_i \tag{38}$$

式中：$\alpha_2>0$，$\beta_2>0$，$\mu>0$，$g_1=\dfrac{n_1+1}{2}+\dfrac{n_1-1}{2}\mathrm{sign}(|s_i|-1)$，$n_1>1$，$g_2=\dfrac{n_2+1}{2}+\dfrac{1-n_2}{2}\mathrm{sign}(|s_i|-1)$，$0<n_2<1$。

基于本文提出的固定时间非奇异终端滑模面，可设计控制器为

$$\begin{aligned}\tau_i=-b_{\xi_i}^{-1}\left[\left(f_{\xi_i}+\hat{d}_{\xi_i}/\gamma_i-\ddot{\xi}_{\mathrm{d}i}-\dot{c}_{\xi_i}\right)+\alpha_1k_1|e_{1i}|^{k_1-1}\dot{e}_{1i}+\right.\\ \left.\beta_1\dot{\varphi}(e_{1i})+\alpha_2\mathrm{sign}^{g_1}s_i+\beta_2\mathrm{sign}^{g_2}s_i+\mu s_i\mathrm{sign}(s_i)\right]\end{aligned} \tag{39}$$

定理1 针对非线性系统(23)，采用本文所设计的滑模面(35)，选择控制律(39)，则：①滑模面s_i有界且固定时间收敛；②系统状态可在固定时间T内收敛到原点的任意小邻域内，且收敛时间有仅与设计参数相关的上界：

$$T<T_{\max}=T_1+T_2 \tag{40}$$

$$\begin{aligned}T_1=&\frac{1}{2\beta_2\left(\dfrac{n_1+1}{2}-1\right)}\ln\left(1+\frac{2\beta_2}{2^{\frac{n_1+1}{2}}\alpha_2}\right)+\\ &\frac{1}{2\alpha_2\left(1-\dfrac{n_2+1}{2}\right)}\ln\left(1+\frac{2\alpha_2}{2^{\frac{n_2+1}{2}}\beta_2}\right)\end{aligned} \tag{41}$$

$$\begin{aligned}T_2=&\frac{1}{2\beta_1\left(\dfrac{m_1+1}{2}-1\right)}\ln\left(1+\frac{2\beta_1}{2^{\frac{m_1+1}{2}}\alpha_1}\right)+\\ &\frac{1}{2\alpha_1\left(1-\dfrac{m_2+1}{2}\right)}\ln\left(1+\frac{2\alpha_1}{2^{\frac{m_2+1}{2}}\beta_1}\right)\end{aligned} \tag{42}$$

证明2 1）选取Lyapunov函数为

$$V_{s_i}=\frac{1}{2}s_i^2 \tag{43}$$

根据式(37)和式(38)，对V_{s_i}求导，可得

$$\begin{aligned}\dot{V}_{s_i}=s_i\dot{s}_i=s_i[&f_{\xi_i}+b_{\xi_i}\tau_i+d_{\xi_i}-\ddot{\xi}_{\mathrm{d}i}-\dot{c}_{\xi_i}+\alpha_1k_1|e_{1i}|^{k_1-1}(e_{2i}-\\ &h_{\xi_i}\arctan(e_{1i}))+\beta_1\dot{\varphi}(e_{1i})]=s_i(-\alpha_2\mathrm{sign}^{g_1}s_i-\\ &\beta_2\mathrm{sign}^{g_2}s_i-\mu s_i\mathrm{sign}(s_i)-\tilde{d}_i/\gamma_i)\end{aligned} \tag{44}$$

为了证明s_i的有界性，考虑Lyapunov函数：

$$W_i=V_{s_i}+V_{z_i} \tag{45}$$

根据式(31)和式(43)，对W_i求导，可得

$$\begin{aligned}\dot{W}_i=&s_i(-\alpha_2\mathrm{sign}^{g_1}s_i-\beta_2\mathrm{sign}^{g_2}s_i-\mu s_i\mathrm{sign}(s_i)-\\ &\tilde{d}_i/\gamma_i)+\gamma_i^{-1}(1-\exp(\dot{e}_{2i}))\dot{\Gamma}(e_{2i})\dot{e}_{2i}z_i^2-\frac{\dot{r}_i}{r_i}z_i^2=\\ &-\alpha_2\mathrm{sign}^{g_1+1}s_i-\beta_2\mathrm{sign}^{g_2+1}s_i-\mu s_i^2\mathrm{sign}(s_i)-\\ &\gamma_i^{-1}r_iz_is_i-[k_{r_i}e_{2i}^2-\gamma_i^{-1}(1-\exp(\dot{e}_{2i}))\dot{\Gamma}(e_{2i})\dot{e}_{2i}]z_i^2\leqslant\\ &-\alpha_2|s_i|^{g_1+1}-\beta_2|s_i|^{g_2+1}-\mu s_i^2-\gamma_i^{-1}r_iz_is_i-\\ &[k_{r_i}e_{2i}^2-\gamma_i^{-1}(1-\exp(\dot{e}_{2i}))\dot{\Gamma}(e_{2i})\dot{e}_{2i}]z_i^2\leqslant\\ &-\alpha_2|s_i|^{g_1+1}-\beta_2|s_i|^{g_2+1}-[s_i\quad z_i]\,\boldsymbol{Q}\,[s_i\quad z_i]^{\mathrm{T}}\end{aligned} \tag{46}$$

式中：$t\in[0,\infty)$，$\boldsymbol{Q}=\begin{bmatrix}\mu & \dfrac{\gamma_i^{-1}r_i}{2}\\ \dfrac{\gamma_i^{-1}r_i}{2} & k_{r_i}e_{2i}^2-\gamma_i^{-1}(1-\exp(\dot{e}_{2i}))\dot{\Gamma}(e_{2i})\dot{e}_{2i}\end{bmatrix}$

正定，$\dot{W}_i\leqslant-\alpha_2|s_i|^{g_1+1}-\beta_2|s_i|^{g_2+1}=-\alpha_2\left|2V_{s_i}\right|^{(g_1+1)/2}-\beta_2\left|2V_{s_i}\right|^{(g_2+1)/2}\leqslant0$。

由此可知，W_i有界，由于z_i已证明渐进稳定性，$\tilde{d}_i\to0$，故由式(44)可得

$$\begin{aligned}\dot{V}_{s_i}=&s_i(-\alpha_2\mathrm{sign}^{g_1}s_i-\beta_2\mathrm{sign}^{g_2}s_i-\mu s_i\mathrm{sign}(s_i))\leqslant\\ &-\alpha_2|s_i|^{g_1+1}-\beta_2|s_i|^{g_2+1}-\mu s_i^2\leqslant-\alpha_2\left|2V_{s_i}\right|^{(g_1+1)/2}-\\ &\beta_2\left|2V_{s_i}\right|^{(g_2+1)/2}\leqslant0\end{aligned} \tag{47}$$

由引理2可知，滑模面s_i有界且固定时间收敛。

2）根据引理3，系统状态可在固定时间内收敛到滑模面$s_i=0$，其收敛时间上界T_1满足式(41)。

系统状态在滑模面上收敛至平衡点附近小邻域内，此时滑动模态满足：

$$\dot{e}_{1i}=-\alpha_{1i}\mathrm{sign}^{k_1}e_{1i}-\beta_{1i}\varphi(e_{1i})-h_{\xi_i}\arctan e_{1i} \tag{48}$$

选取Lyapunov函数为

$$E_i=\frac{1}{2}e_{1i}^2 \tag{49}$$

当$|e_{1i}|\geqslant\varepsilon$时，由式(36)可得

$$\dot{e}_{1i}=-\alpha_1\mathrm{sign}^{k_1}e_{1i}-\beta_1\mathrm{sign}^{k_2}e_{1i}-h_{\xi_i}\arctan e_{1i} \tag{50}$$

对式(48)求导可得

$$\begin{aligned}\dot{E}_i=e_{1i}\dot{e}_{1i}=e_{1i}(&-\alpha_1\mathrm{sign}^{k_1}e_{1i}-\beta_1\mathrm{sign}^{k_2}e_{1i}-\\ &h_{\xi_i}\arctan e_{1i})\leqslant-\alpha_1|e_{1i}|^{k_1+1}-\beta_1|e_{1i}|^{k_2+1}=\\ &-\alpha_1|2E_i|^{(k_1+1)/2}-\beta_1|2E_i|^{(k_2+1)/2}\leqslant0\end{aligned} \tag{51}$$

系统状态e_{1i}在固定时间内收敛至平衡点附近极小区间$|e_{1i}|<\varepsilon$，其收敛时间上界T_2满足式(42)。

$$|e_{2i}|\leqslant\alpha_1|e_{1i}|^{k_1}+\beta_1|e_{1i}|^{k_2}<\alpha_1\varepsilon^{k_1}+\beta_1\varepsilon^{k_2} \tag{52}$$

可得系统(23)的收敛时间$T<T_{\max}=T_1+T_2$。

当$|e_{1i}|<\varepsilon$时，由式(36)可得

$$\dot{e}_{1i}=-\alpha_1\mathrm{sign}^{k_1}e_{1i}-\beta_1k_2e_{1i}^2\mathrm{sign}(e_{1i})-h_{\xi_i}\arctan e_{1i} \tag{53}$$

对式(49)求导可得

$$\begin{aligned}\dot{E}_i=e_{1i}\dot{e}_{1i}=e_{1i}(&-\alpha_1\mathrm{sign}^{k_1}e_{1i}-\beta_1k_2e_{1i}^2\mathrm{sign}(e_{1i})-\\ &h_{\xi_i}\arctan e_{1i})\leqslant-\alpha_1|e_{1i}|^{k_1+1}-\beta_1k_2e_{1i}^2=\\ &-\alpha_1|2E_i|^{\frac{k_1+1}{2}}-\beta_1k_2|2E_i|^2\leqslant0\end{aligned} \tag{54}$$

$$e_{2i}=-\alpha_1\mathrm{sign}^{k_1}e_{1i}-\beta_1k_2e_{1i}^2\mathrm{sign}(e_{1i}) \tag{55}$$

$$|e_{2i}|\leqslant\alpha_1|e_{1i}|^{k_1}+\beta_1\left|k_2e_{1i}^2\mathrm{sign}(e_{1i})\right|\leqslant\alpha_1\varepsilon^{k_1}+\beta_1k_2\varepsilon^2 \tag{56}$$

系统状态e_{1i}将在$|e_{1i}|<\varepsilon$区间内迅速到达平衡点。

若系统状态在滑动模态时有$|e_{1i}|<\varepsilon$, 说明系统状态收敛至平衡点附近小邻域内, 随后会迅速收敛至平衡点, 即系统状态可在固定时间收敛至平衡点。

综上所述, $|e_{1i}|<\varepsilon$为系统全局固定时间收敛域, 其收敛时间上界为$T_{\max}$, 满足$T<T_{\max}=T_1+T_2$。

证毕

注 2　系统收敛时间的上界仅与设计参数α_1、β_1、m_1、m_2、α_2、β_2、n_1、n_2和ε有关, 可通过增大α_1和β_1、减小m_2来调节系统状态的收敛速率, 同时可通过增大α_2和β_2、减小n_2来调节滑模面的收敛速率, 使系统在收敛快速性和稳态跟踪精度方面达到一个较为均衡的结果。

4　仿真结果与分析

为验证本文所提固定时间控制方案的有效性和优越性, 设计了 2 种实验场景。第 1 种实验场景是模拟垂直起降模式下的前倾转变体无人机在时变扰动下的飞行, 通过复杂的预设轨迹验证方法的有效性; 第 2 种实验场景是模拟前倾转变体无人机的某个飞行模式变化周期, 包括地面起飞、平滑加速、模式切换、水平飞行、平面盘旋、水平飞行, 综合测试其在飞行过程中的机身稳定性、抗干扰性、飞行模式切换过程的平滑性和水平飞行模式时的可操控性等特性。

前倾转变体无人机的机身物理参数设定为$m=3.0$ kg, $\rho=1.293$ kg/m^3, $A=0.08$ m^2, $I_{xx}=I_{yy}=1.75$ N·s^2/rad, $I_{zz}=3.5$ N·s^2/rad, $l_l=l_s=0.28$ m。

4.1　实验场景 1

设置前倾转变体无人机的预设轨迹为

$$\begin{cases}x=10(1-\cos t\mathrm{e}^{0.05t})\\ y=5\sin(2t)\mathrm{e}^{0.05t}\\ z=10(1-\mathrm{e}^{0.05t})\\ \phi=0\\ \theta=0\\ \psi=\pi/6\end{cases} \tag{57}$$

设置前倾转变体无人机受到的时变扰动为

$$\boldsymbol{d}=\begin{bmatrix}0.8/(1+\mathrm{e}^{-t})\\ 0.5\sin(\pi t)\\ 0.2\sin(3t)+0.1\cos(5t)\\ 0.2/(1+\mathrm{e}^{-t})-0.5\sin(0.5\pi t)\\ 0.5\tanh(\sqrt{2}t)\\ 0.3\cos(2t+1)\end{bmatrix} \tag{58}$$

时变扰动观测器的初值均设置为 0, 本文方法在 2 种实验场景中的控制增益选取为: $h_{\xi_i}=100$, $m_{1i}=1.7$, $m_{2i}=0.3$, $n_{1i}=1.8$, $n_{2i}=0.56$, $\mu_i=0.1$, $\varepsilon=0.01$, $i=1,2,\cdots,6$, $\alpha_{1i}=\beta_{1i}=\alpha_{2i}=\beta_{2i}=4$, $i=1,2,3$, $\alpha_{14}=\beta_{14}=\alpha_{24}=\beta_{24}=8$, $\alpha_{1i}=\beta_{1i}=\alpha_{2i}=\beta_{2i}=1.5$, $i=5,6$。

为验证本文所提固定时间控制方案的有效性, 考虑如下 3 种不同初始状态, 选取的 3 种初始状态如表 1 所示。

表 1　3 种初始状态的选取

Table 1　Selection of three initial states　　m

状态	$x(0)$	$y(0)$	$z(0)$	$\phi(0)$	$\theta(0)$	$\psi(0)$
1	1.5	−0.5	−1.5	0	0	0
2	−2.5	−0.5	−1.5	0	0	π/6
3	2	−0.5	1	0	0	π/3

图 4 为基于固定时间非奇异终端滑模控制方法的前倾转变体无人机预设轨迹与在不同初始状态下实际轨迹的三维跟踪曲线对比。可以看到, 无人机从 3 个位置出发, 在某一时间内与预设轨迹重合。图 5 选取了位置子系统的跟踪误差在不同初始状态下的收敛对比情况。通过对比观察, 位置子系统的跟踪误差可以在不同的初始状态下达到固定时间内收敛, 说明本文设计的固定时间非奇异终端滑模控制律对在垂直起降模式下的前倾转变体无人机进行轨迹跟踪任务时, 系统的收敛时间上界与初始状态无关, 因此, 本文所提固定时间控制方案是有效的。

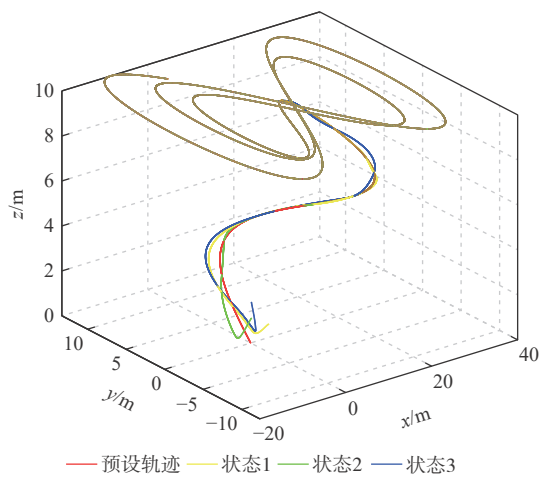


图 4　不同初始状态下的三维轨迹跟踪效果

Fig. 4　Three-dimensional trajectory tracking effects in different initial states

4.2　实验场景 2

实验场景 2 总体的飞行实验设计为: 前倾转变体无人机在地面起飞, 开始在xOz平面内以 14 m/s 的速度平滑飞行, 同时在 8 s 时机翼开始倾转, 至 10 s 时完成倾转, 完成过渡切换模式, 进入到水平飞行模式, 10 s 后继续以平飞模式加速到 15 s, 达到接

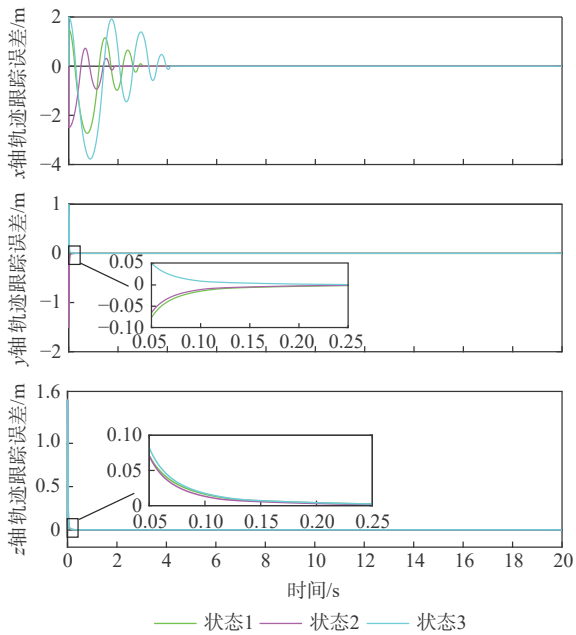


图5 位置子系统的轨迹跟踪误差收敛情况

Fig. 5 Trajectory tracking error convergence of position subsystem

近20 m的高度，无人机在xOy面内跟踪一个“∞”形轨迹，随后继续沿x_w轴飞行至35 s，结束整个飞行实验。

设置前倾转变体无人机的预设轨迹为

$$x=\begin{cases}40\ln\left(1+\mathrm{e}^{0.5t-5}\right)-0.268\,6 & t\in[0,15)\\ R\sin(t-15)+P & t\in[15,15+2\pi)\\ R\sin(t-15-2\pi)+P & t\in[15+2\pi,15+4\pi)\\ 40\ln\left(1+\mathrm{e}^{0.5(t-4\pi)-5}\right)-0.268\,6 & t\in[15+4\pi,35]\end{cases}\tag{59}$$

$$y=\begin{cases}0 & t\in[0,15)\\ R\cos(t-15)-R & t\in[15,15+2\pi)\\ R-R\cos(t-15-2\pi) & t\in[15+2\pi,15+4\pi)\\ 0 & t\in[15+4\pi,35]\end{cases}\tag{60}$$

$$z=20\left(1-\mathrm{e}^{0.7t}\right)\qquad t\in[0,35]\tag{61}$$

前倾转变体无人机初始位置为世界坐标系原点，$R=20\mathrm{e}^{2.5}/(\mathrm{e}^{2.5}+1)$，$P=40\ln\left(1+\mathrm{e}^{2.5}\right)-0.268\,6$。

为展示本文方法对不同类型轨迹跟踪的通用性，系统受到的时变扰动及控制增益与实验场景1中保持一致。

线性化机翼的升阻力系数C_L和C_D随机身与机翼的夹角θ_f的变化[24]设置为：$\theta_f=[0°,90°]$，$C_L=0.1\theta_f+0.3$，$C_D=0.002\theta_f+0.004$，当切换为水平飞行模式时，机身俯仰角θ设置为0°，升力F_L可由式(2)～式(4)得出，进而计算出C_L和C_D，得到θ_f的实时状态变化，同时θ_f的改变又实时反馈到F_L，从而形成了调节的闭环。

实验场景2通过2种对比方案对本文方法的有效性进行验证。第1种方案是基于传统I&I方法设计扰动观测器，并结合文献[25]中的传统终端滑模控制方案，对比本文方法所设计的观测器对时变扰动的估计性能。第2种方案是对比本文方法所设计的固定时间非奇异终端滑模控制器的有效性。文献[26]针对机动目标拦截问题提出了固定时间控制方案，结果表明，其制导律有效提高了对未知目标的拦截性能，文献[27]针对四旋翼无人机的轨迹跟踪问题提出了终端滑模控制方案，在保证飞行稳定性的基础上更好地跟踪了预定的飞行路径。因此，第2种对比方案采用本文方法设计的扰动观测器，并结合文献[26-27]中具有与本文方法类似的控制器形式进行对比。

图6和图7为4种方法对预设轨迹的跟踪对比效果。图6说明4种方法在轨迹跟踪的过程中均达到了较高的跟踪精度，但是从图7可以看出，文献[25,27]在3个位置轨迹跟踪误差的收敛速度方面都慢于本文方法，特别是在15 s跟踪“∞”轨迹时发生了较大的振幅波动，文献[26]也在15 s跟踪“∞”轨迹时y轴轨迹跟踪误差抖振加剧。这是由于无人机在15 s切换轨迹时快速改变推力大小造成的。此外，文献[26]虽然在x轴和z轴轨迹跟踪误差的收敛速度方面与本文方法相接近，但在10 s飞行模式切换后，z轴轨迹跟踪误差开始产生小幅抖振，这是由于无人机在10 s飞行模式切换后，关键形变量θ_f对机身造成明显的综合扰动导致的。

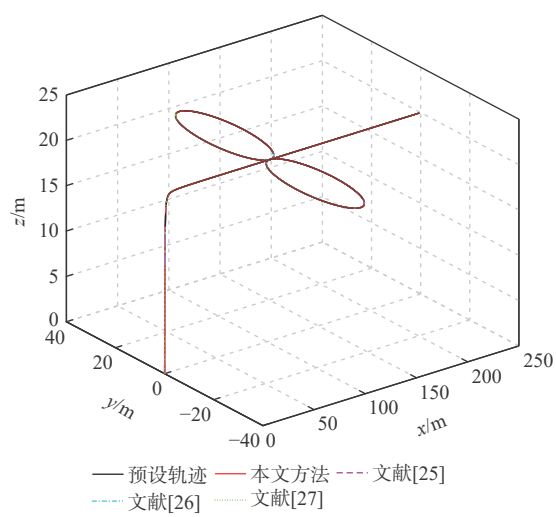


图6 4种方法的三维轨迹跟踪效果

Fig. 6 Three-dimensional trajectory tracking effects of four methods

图8和图9分别为解算出的期望姿态角及4种方法对期望姿态角的轨迹跟踪对比效果。文献[25]在3个姿态角轨迹跟踪误差的收敛速度方面都慢

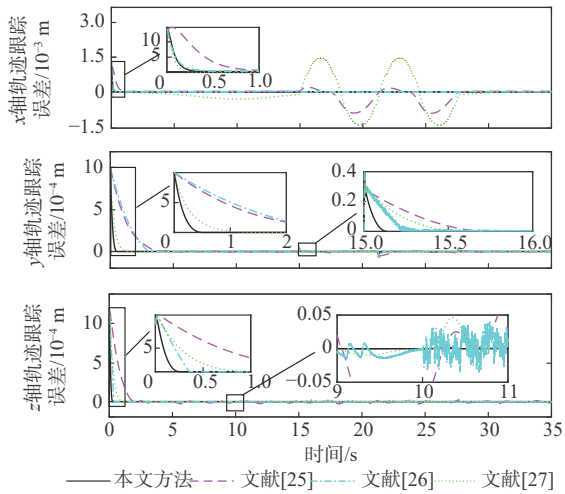


图 7　4 种方法的位置轨迹跟踪误差

Fig. 7　Position trajectory tracking errors of four methods

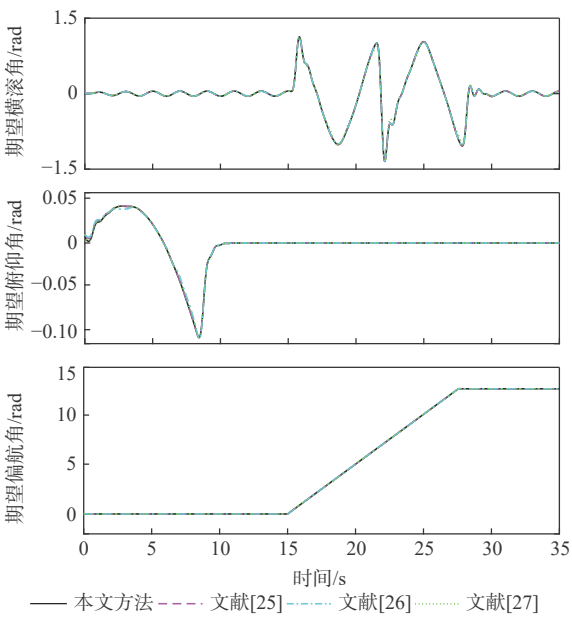


图 8　4 种方法解算出的期望姿态角

Fig. 8　Desired attitude angles calculated by four methods

于本文方法, 并且横滚角和偏航角轨迹跟踪误差在 8 s 切换飞行模式时及在 15 s 跟踪“∞”轨迹, 入圆和出圆时都出现了较大的振幅波动, 而文献 [26-27] 在跟踪 3 个姿态角的轨迹方面各有优劣, 文献 [26] 在 8 s 切换飞行模式时, 俯仰角轨迹跟踪误差产生了小幅抖振, 文献 [27] 则在俯仰角轨迹跟踪误差的初始阶段出现了较大的振幅, 并一直伴随着小幅抖振。

为了测试时变扰动模型更为复杂的情况下对轨迹跟踪效果的影响, 设计了一组复合扰动, 根据预设轨迹设置时间节点, 分别加入常值扰动、周期性扰动、参数变化周期性扰动及发散型扰动, 同时为了更贴合实际扰动模型, 还加入了信噪比 5% 的高斯噪声, 设计的多源时变扰动d_x如图 10 所示。

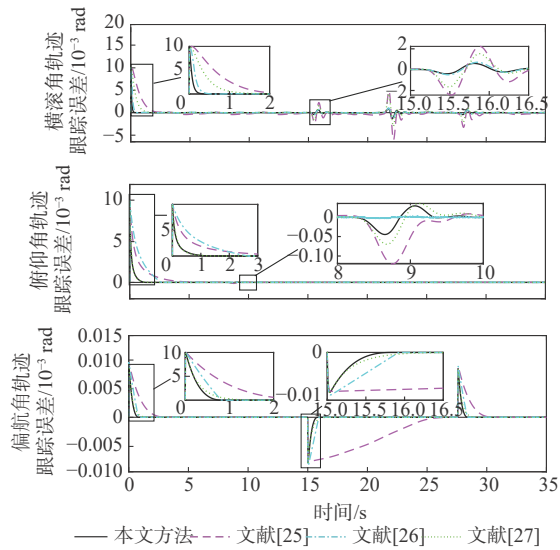


图 9　4 种方法的姿态轨迹跟踪误差

Fig. 9　Attitude trajectory tracking errors of four methods

$$d_x=\begin{cases}0.5 & t\in[0,15)\\ 0.5\cos(\pi t) & t\in[15,15+2\pi)\\ 0.8\cos(2\pi t) & t\in[15+2\pi,15+4\pi)\\ -0.5\cos te^{0.001t}+0.5 & t\in[15+4\pi,35]\end{cases}\tag{62}$$

图 11 为在时变扰动模型复杂情况下 x 轴的轨迹跟踪误差效果。可以看出, 本文方法及文献 [26-27] 在时变扰动d_x下表现出相似的收敛情况, 但文献 [25] 的轨迹跟踪误差发生了较大的改变, 说明其在复杂时变扰动的情况下, 对机身产生了较大的影响。

综上所述, 本文方法虽然也在 15 s 跟踪“∞”轨迹, 入圆和出圆时轨迹跟踪误差出现了较小的幅值波动, 但是波动之后的收敛速度都明显优于其他方法, 并且在 8 s 切换飞行模式时和 10 s 飞行模式切换后没有出现抖振, 同时抑制了外部的时变扰动和飞行模式切换产生的内部扰动对系统造成的影响, 整体表现出较好的位置和姿态轨迹跟踪性能。

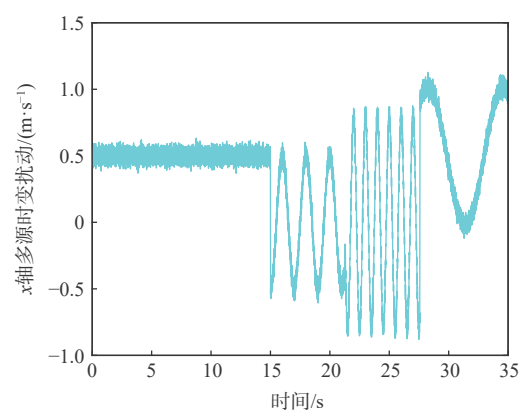


图 10　多源时变扰动

Fig. 10　Multi-source time-varying disturbances

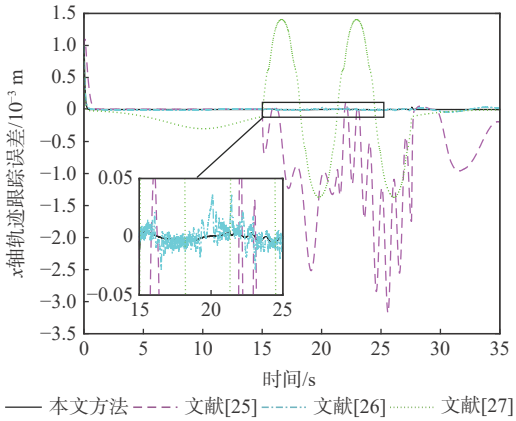


图 11 多源时变扰动下的轨迹跟踪误差

Fig. 11 Trajectory tracking error under multi-source time-varying disturbances

图 12 和图 13 分别为 4 种方法的系统虚拟控制输入信号和控制输入信号的对比情况。图 12 说明了文献 [26-27] 在前倾转变体无人机整个飞行实验的过程中，特别是在 10 s 飞行模式切换后和在 15 s 跟踪“∞”轨迹，入圆和出圆时，虚拟控制输入信号都产生了较大幅度的抖振，可能造成执行器饱和故障。图 13 说明文献 [27] 在 15 s 跟踪“∞”轨迹时，控制的输入信号 u_1 出现了较大的幅值波动，迫使变体无人机迅速平衡自身，出力损耗较大，而文献 [26] 在 8 s 飞行模式开始切换时，控制的输入信号 u_3 就一直处于较大的出力状态。综合分析可知，文献 [25] 和本文方法在控制信号的稳定性和幅值大小方面都要优于其他 2 种方法。

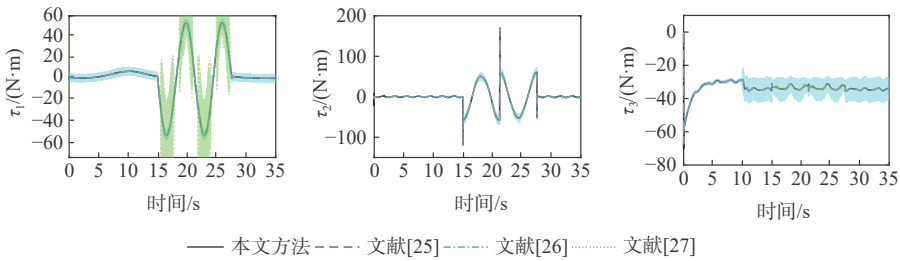


图 12 4 种方法的虚拟控制输入信号响应

Fig. 12 Virtual control input signal responses of four methods

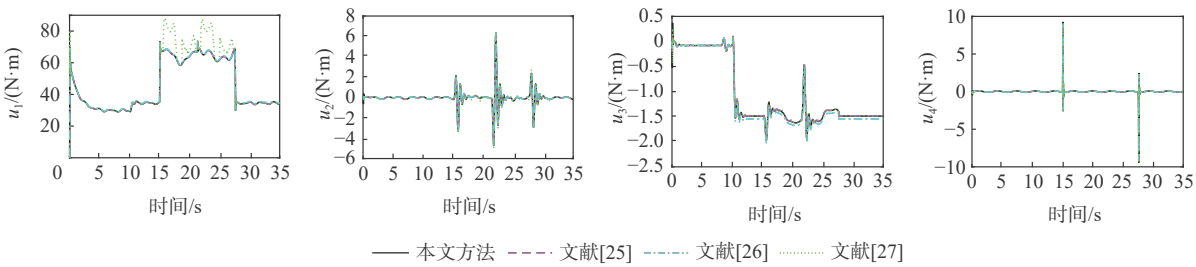


图 13 4 种方法的实际控制输入信号响应

Fig. 13 Actual control input signal responses of four methods

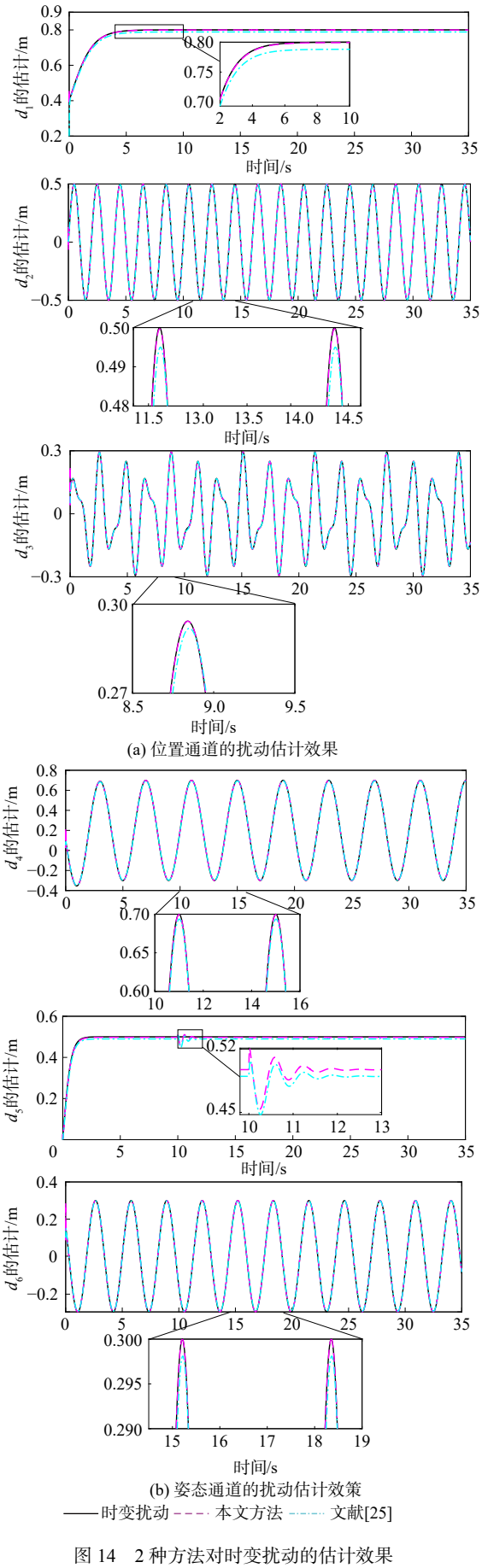
图 14 为本文方法和文献 [25] 对时变扰动估计的对比情况。本文方法相较于传统 I&I 观测器在扰动的估计方面表现出更高的精度，在复合 I&I 基础上加入动态尺度因子的观测器具有更优异的估计性能。另外，从 d_5 的估计效果来看，在 10 s 飞行模式切换后，观测器对扰动的估计产生了较大的误差，这是由于模式切换过程对机身产生了较大的综合扰动所导致的，一段时间后观测器对扰动的估计效果得到明显改善。

图 15 和图 16 分别展示了 6 个控制通道动态尺度因子 r_i 和动态尺度误差 z_i 的自适应收敛过程。可以看出，r_i 和 z_i 均可以在一定时间后收敛到稳定值，同时，动态尺度因子 r_i 的加入可以更准确地对时变扰动进行估计和补偿。

在整个飞行实验过程中，前倾转变体无人机在 8 s 时前翼开始倾转，到 10 s 时切换到水平飞行模式，之后机身与机翼的夹角 θ_f 会随着机身飞行速度的改变而动态调整，从而为机翼提供升力和阻力，使机身在水平飞行模式下保持稳定。θ_f 在垂直起降模式下设置为 90°，8 s 开始进入垂直水平过渡切换模式，切换曲线设置为

$$\theta_f = 90 - 43.5(2 - 2e^{-3(t-8)}) \qquad t \in [8,10) \tag{63}$$

图 17 和图 18 分别展示了整个飞行实验中 θ_f 的动态变化过程及机身在 x、y、z 轴方向的速度变化过程。图 17 说明文献 [26] 在 10 s 飞行模式切换后，θ_f 的动态调整出现了小幅高频抖振，而本文方法及文献 [25,27] 在平飞模式时 θ_f 能够保持得更平稳，飞行稳定性更好。图 18 说明在飞行速度大于



(a) 位置通道的扰动估计效果

(b) 姿态通道的扰动估计效策

图 14　2 种方法对时变扰动的估计效果

Fig. 14　Estimation effects of two methods on time-varying disturbances

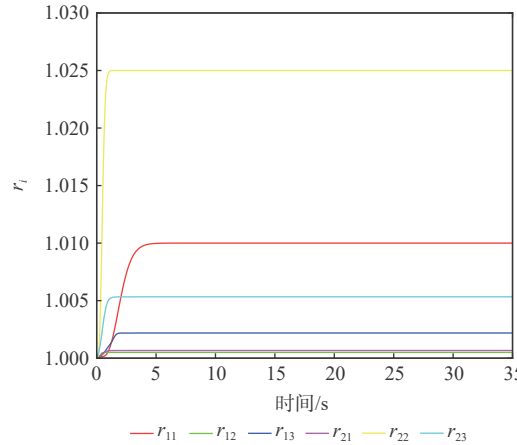


图 15　6 个控制通道动态尺度因子r_i的自适应过程

Fig. 15　Adaptive process of dynamic scale factor r_i of six control channels

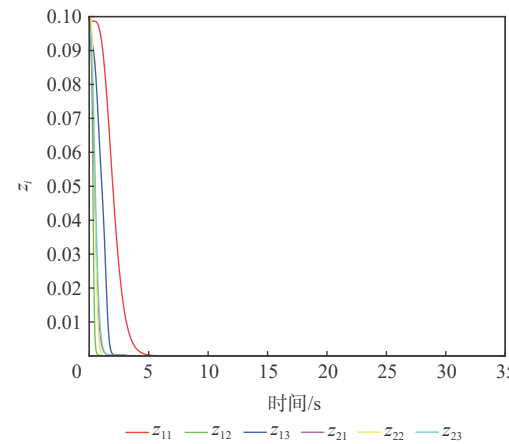


图 16　6 个控制通道动态尺度误差z_i的自适应过程

Fig. 16　Adaptive process for dynamic scale error z_i of six control channels

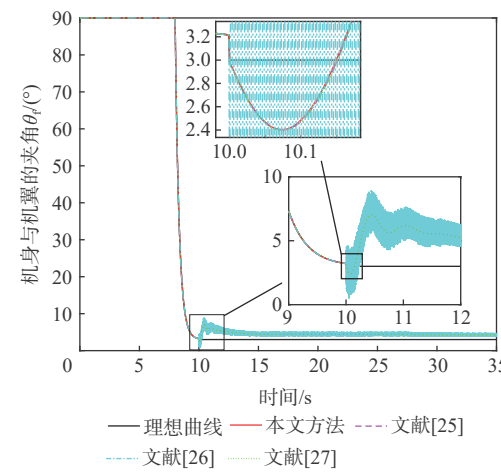


图 17　4 种方法的机翼与机身夹角θ_f自适应响应过程

Fig. 17　Adaptive response process of angle between wing and fuselage of four methods

10 m/s 之后, 无人机由垂直起降模式切换到水平飞行模式, 而θ_f随着飞行速度的变化趋于平稳而收敛到稳定值。

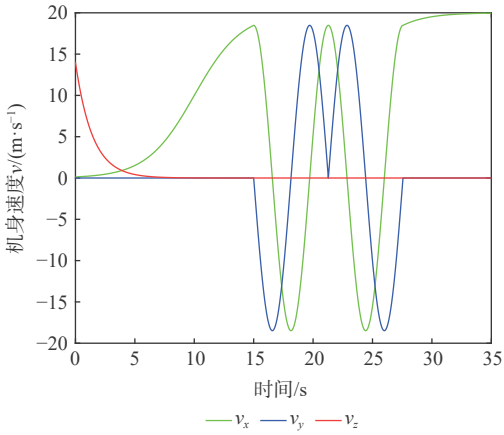


图18 机身速度v的变化过程

Fig. 18 Variation of vehicle speed v

注3 从图8可以看出,前倾转变体无人机在10 s设置俯仰角θ为0°,但从图17可知,此时θ_f发生了较大的变化,说明切换飞行模式对机身产生了内部扰动,但由于控制器的抑制作用,θ_f之后会根据机身飞行速度的改变而动态调整,从而为机翼提供所需的力来保持θ设定的目标,并保证了机身的稳定飞行。

5 结 论

本文针对具有时变扰动的前倾转变体无人机,提出了基于I&I的时变扰动观测器,并设计了分段式的固定时间非奇异终端滑模控制器,结论如下:

1)基于I&I并融合动态尺度因子设计的扰动观测器,对时变扰动有较好的估计效果,仿真实验表明,动态尺度因子的引入进一步提高了观测器估计扰动的精度。

2)基于固定时间收敛理论设计的非奇异终端滑模控制器,可以保证系统状态的固定时间收敛速度,且收敛时间仅决定于预设的控制器参数,仿真实验表明,分段非线性函数的引入在解决奇异性的同时,保证了系统的固定时间收敛性。

3)理论和仿真实验表明,本文方法可以使前倾转变体无人机在固定时间内完成多种复杂的轨迹跟踪任务,且系统具有较好的收敛快速性及抗干扰性,特别是对于外部时变扰动和飞行模式切换产生的内部扰动,都有较好的抑制能力。

参考文献(References)

[1] 刘重, 何玉庆, 谷丰, 等. 四倾转旋翼无人机无源控制与飞行实验[J]. 控制理论与应用, 2021, 38(8): 1287-1298.
LIU Z, HE Y Q, GU F, et al. Passivity-based control and flight experiment of quad-tilt rotor unmanned aerial vehicle[J]. Control Theory & Applications, 2021, 38(8): 1287-1298(in Chinese).

[2] 陈卓英, 陈怀民, 王永康. 飞翼布局矢量推力无人机容错控制研究[J]. 西北工业大学学报, 2022, 40(5): 990-996.
CHEN Z Y, CHEN H M, WANG Y K. Study on tolerance control of flying-wing layout vectored thrust UAV[J]. Journal of Northwestern Polytechnical University, 2022, 40(5): 990-996(in Chinese).

[3] ZHANG J M, LIU Y B, GAO L, et al. Bioinspired drone actuated using wing and aileron motion for extended flight capabilities[J]. IEEE Robotics and Automation Letters, 2022, 7(4): 11197-11204.

[4] 李新凯, 张宏立, 范文慧. 非匹配扰动下变体无人机预设性能控制[J]. 航空学报, 2022, 43(2): 325008.
LI X K, ZHANG H L, FAN W H. Prescribed performance control for morphing aerospace vehicle under mismatched disturbances[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(2): 325008(in Chinese).

[5] ASTOLFI A, ORTEGA R. Immersion and invariance: a new tool for stabilization and adaptive control of nonlinear systems[J]. IEEE Transactions on Automatic Control, 2003, 48(4): 590-606.

[6] ARBO M H, GRØTLI E I, GRAVDAHL J T. On the globally exponentially convergent immersion and invariance speed observer for mechanical systems[C]//Proceedings of the 2017 American Control Conference. Piscataway: IEEE Press, 2017: 3294-3299.

[7] YU L, HE G, ZHAO S L, et al. Immersion and invariance-based sliding mode attitude control of tilt tri-rotor UAV in helicopter mode[J]. International Journal of Control, Automation and Systems, 2021, 19(2): 722-735.

[8] HOSSEINI-PISHROBAT M, KEIGHOBADI J, PIRASTEHZAD A, et al. Immersion and invariance-based extended state observer design for a class of nonlinear systems[J]. International Journal of Robust and Nonlinear Control, 2021, 31(13): 6233-6254.

[9] KARAGIANNIS D, SASSANO M, ASTOLFI A. Dynamic scaling and observer design with application to adaptive control[J]. Automatica, 2009, 45(12): 2883-2889.

[10] XIA D D, YUE X K, WEN H W. Output feedback tracking control for rigid body attitude via immersion and invariance angular velocity observers[J]. International Journal of Adaptive Control and Signal Processing, 2020, 34(12): 1812-1830.

[11] XIA D D, YUE X K. Dynamic scaling-based adaptive control without scaling factor: with application to Euler-Lagrange systems [J]. International Journal of Robust and Nonlinear Control, 2021, 31(10): 4531-4552.

[12] BHAT S P, BERNSTEIN D S. Geometric homogeneity with applications to finite-time stability[J]. Mathematics of Control, Signals and Systems, 2005, 17(2): 101-127.

[13] YU J P, SHI P, ZHAO L. Finite-time command filtered backstepping control for a class of nonlinear systems[J]. Automatica, 2018, 92: 173-180.

[14] HOU H Z, YU X H, XU L, et al. Finite-time continuous terminal sliding mode control of servo motor systems[J]. IEEE Transactions on Industrial Electronics, 2019, 67(7): 5647-5656.

[15] POLYAKOV A. Nonlinear feedback design for fixed-time stabilization of linear control systems[J]. IEEE Transactions on Automatic Control, 2012, 57(8): 2106-2110.

[16] ZUO Z Y. Non-singular fixed-time terminal sliding mode control of non-linear systems[J]. IET Control Theory & Applications, 2015, 9(4): 545-552.

[17] HUANG Y, JIA Y M. Fixed-time consensus tracking control for

second-order multi-agent systems with bounded input uncertainties via NFFTSM[J]. IET Control Theory & Applications, 2017, 11(16): 2900-2909.

[18] TIAN Y, CAI Y L, DENG Y F. A fast nonsingular terminal sliding mode control method for nonlinear systems with fixed-time stability guarantees[J]. IEEE Access, 2020, 8: 60444-60454.

[19] LI H J, CAI Y L. Fixed-time non-singular terminal sliding mode control with globally fast convergence[J]. IET Control Theory & Applications, 2022, 16(12): 1227-1241.

[20] WANG X, GUO J, TANG S J, et al. Fixed-time disturbance observer based fixed-time back-stepping control for an air-breathing hypersonic vehicle[J]. ISA Transactions, 2019, 88: 233-245.

[21] SONG S, PARK J H, ZHANG B Y, et al. Event-triggered adaptive practical fixed-time trajectory tracking control for unmanned surface vehicle[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2020, 68(1): 436-440.

[22] JIN X, DAI S L, LIANG J J. Fixed-time path-following control of an autonomous vehicle with path-dependent performance and feasibility constraints[J]. IEEE Transactions on Intelligent Vehicles, 2021, 8(1): 458-468.

[23] ZOU A M, KUMAR K D, DE RUITER A H J. Fixed-time attitude tracking control for rigid spacecraft[J]. Automatica, 2020, 113: 108792.

[24] 李新凯, 张宏立, 范文慧. 基于时变障碍李雅普诺夫函数的变体无人机有限时间控制[J]. 自动化学报, 2022, 48(8): 2062-2074.
LI X K, ZHANG H L, FAN W H. Finite-time control for morphing aerospace vehicle based on time-varying barrier Lyapunov function [J]. Acta Automatica Sinica, 2022, 48(8): 2062-2074(in Chinese).

[25] NI J K, LIU L, LIU C X, et al. Fast fixed-time nonsingular terminal sliding mode control and its application to chaos suppression in power system[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2016, 64(2): 151-155.

[26] ZHANG Y, TANG S J, GUO J. Adaptive terminal angle constraint interception against maneuvering targets with fast fixed-time convergence[J]. International Journal of Robust and Nonlinear Control, 2018, 28(8): 2996-3014.

[27] NEKOUKAR V, MAHDIAN DEHKORDI N. Robust path tracking of a quadrotor using adaptive fuzzy terminal sliding mode control [J]. Control Engineering Practice, 2021, 110: 104763.

Fixed time trajectory tracking control of forward-tilting morphing aerospace vehicle

DOU Lei，LI Xinkai*，ZHANG Hongli，WANG Hao，YANG Jiaxiu

(School of Electrical Engineering，Xinjiang University，Urumqi 830017，China)

Abstract：In view of the trajectory tracking problem of the forward-tilting morphing aerospace vehicle with time-varying disturbance, a non-singular terminal sliding mode control scheme based on immersion and invariance (I&I) theory and fixed time convergence theory was proposed. Firstly, a time-varying disturbance observer based on I&I was designed by combining dynamic scale factors. Secondly, a segmented fixed-time non-singular terminal sliding surface was constructed, which eliminated the singularity of the sliding mode surface and made the system state converge to any small neighborhood of the equilibrium point within a fixed time, and the upper bound of the convergence time had nothing to do with the initial state of the system. Finally, based on the Lyapunov stability theory, the global fixed-time stability of the system was proven, and the upper bound of its convergence time was given. The effectiveness and superiority of the proposed control scheme were verified in two experimental scenarios. Compared with the traditional control method, the control scheme proposed in this paper made the system converge faster and has better anti-disturbance ability.

Keywords：morphing aerospace vehicle; immersion and invariance; dynamic scale factors; non-singular terminal sliding mode surface; fixed-time convergence

Received：2023-03-31；**Accepted**：2023-05-26；**Published Online**：2023-06-28 14：30

URL：link.cnki.net/urlid/11.2625.V.20230627.1733.002

Foundation items：National Natural Science Foundation of China (62263030); Youth Project of Natural Science Foundation of Xinjiang Uygur Autonomous Region (2022D01C86)

* **Corresponding author**. E-mail：lxk@xju.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0156

基于多轮次分布式拍卖的异构多任务分配算法

吕晔[1]，周锐[1,*]，李兴[1]，刘志恒[2]，邸斌[3]

(1. 北京航空航天大学 自动化科学与电气工程学院，北京 100191；2. 南京电子技术研究所，南京 210039；
3. 中国人民解放军 32806 部队，北京 100091)

摘　　要：针对多机协同探测、攻击、评估等不同作战任务的分布式决策需要，建立了一种面向异构且具有时序约束任务的多轮次分布式拍卖算法。所提算法较为灵活地解决了不同能力无人机(UAV)之间具有复杂约束的分布式协同任务分配问题。在分布式任务分配框架中，每架无人机可实现分布式异步计算和异步通信，且能确保较好的求解效率和准确性。任务分配原则以就近分配为主，同时兼顾目标价值和时序约束，实现了在有限资源情况下对重要目标的优先分配，且使总的行程最短。仿真结果表明：在给定任意无人机和目标数量及无人机作战能力配置的情况下，所提算法均能得到较好的分配结果。

关 键 词：多机协同；异构任务分配；分布式拍卖算法；任务规划；时序约束
中图分类号：V219
文献标志码：A　　**文章编号：**1001-5965(2025)03-1018-10

多无人机(unmanned aerial vehicle, UAV)的任务分配问题是一类典型的含约束的组合优化问题[1-2]，其意义在于通过合理的分配使得有限的机载资源取得最大化的作战成果。任务分配问题的求解算法从数据的分布方式可分为集中式与分布式2类[3-7]，求解任务分配问题的集中式方法可分为最优化方法和启发式方法。穷举法、混合整数规划法和图论法属于最优化方法[2]。混合整数规划法通过建立约束条件及目标函数的方式进行求解，图论法将参与分配的智能体与待分配目标的特征表示出来，并在目标与智能体之间用图的方式建立匹配[2]。启发式方法可以在计算精度与计算时间之间进行调节，各种智能类算法如遗传算法、群智能算法等都属于启发式方法。近年来，国内外学者对集中式任务分配问题从不同的角度提出了具体的建模思路[8-16]。冉华明等[17]针对空战中协同干扰任务分配问题，综合考虑了敌我双方飞机的干扰功率、干扰频率、干扰样式及作战性能，建立了总的优势函数，并利用遗传算法进行求解。魏政磊等[18]针对传统算法对动态任务分配问题求解速度慢和精度不高的问题，提出了一种自适应灰狼优化算法。Zhen等[19]利用改进的合同网算法处理一对多和多对一的任务分配问题，并综合考虑了任务时序及无人机异构负载等约束。国外学者也有诸多研究，Kurdi等[20]受细菌在菌落中觅食行为的启发，将其总结并应用到大规模无人机集群任务分配中，得到了相比传统任务分配算法更加快速稳定的效果。集中式方法计算结构简单、容易实现，能够真正得到全局最优解，但若中心节点故障则会导致全局任务失败，同时中心节点的通信量巨大，在受干扰的作战环境中可能不易实现。与集中式方法相比，分布式计算方式具有去中心化的特点，灵活性较好，但是算法设计复杂、技术难度大。对于分布式方法，国内外学者也有相应的研究。Chen等[9]针对分布式

收稿日期：2023-04-03；录用日期：2023-07-15；网络出版时间：2023-08-03 08:47
网络出版地址：link.cnki.net/urlid/11.2625.V.20230802.1250.003
基金项目：国家自然科学基金(62003365)
*通信作者. E-mail：zhr@buaa.edu.cn

引用格式：吕晔，周锐，李兴，等. 基于多轮次分布式拍卖的异构多任务分配算法[J]. 北京航空航天大学学报，2025，51(3)：1018-1027.
LYU Y，ZHOU R，LI X，et al. A heterogeneous multi-task assignment algorithm based on multi-round distributed auction[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：1018-1027 (in Chinese).

任务分配场景中的通信拓扑动态变化及高实时性等要求, 提出了一种建立在拍卖算法基础上的动态分组分配算法, 在计算量一定的情况下提升了解的质量。

目前, 分布式任务分配算法多以基于市场机制的拍卖算法(auction algorithm)为主[21]。拍卖算法是模拟市场拍卖机制的一种方法, 其由参与方、竞拍品、收益函数和加价策略 4 个部分组成, 其求解速度快、操作规则清晰灵活, 目前已经广泛应用在多无人机的任务分配问题中[22-24]。拍卖算法最早用于解决一对一的任务分配问题, 后 Zavlanos 等[25] 将传统的集中式拍卖算法改进为基于协商的分布式拍卖算法, 并同时证明了基于协商的一致性协议收敛性及分布式拍卖算法能够在有限的迭代步内收敛至最优解。近年来, 国内外学者也对分布式拍卖算法进行了诸多的改进与演化。颜骥等[26] 利用扩展的一致性包算法(consensus-based bundle algorithm, CBBA)研究多智能体系统的多目标多任务分配问题, 在保持 CBBA 算法特性的同时考虑了任务之间的时序关系。邸斌等[24] 将无人机所具备的不同能力视为不同的智能体, 利用分布式拍卖算法实现了具有任务时序约束的分布式拍卖问题。廖沫和陈宗基[27] 提出了一种基于多 Agent 分布式协同拍卖的动态目标分配算法, 能够在一定时间约束的条件下给出较好的分配结果。

目前的任务分配研究对于任务具有时序约束和机载资源的异构性等问题, 研究仍较为薄弱, 而且未特别重视实际工程应用中由于分布式计算机制带来的异步计算和通信延时问题可能对算法求解效率和准确性所造成的影响。本文针对资源异构及任务具有时序约束的多任务分配问题, 设计了一种基于异步分布式拍卖算法的异构多轮次分布式拍卖算法, 较好地解决了异构无人机之间的任务分配问题。

1　多任务分配问题描述

1.1　具有任务时序约束的多任务分配问题

本文任务分配问题的根本意义在于: 通过合理的分配, 使得有限的作战资源取得最大化的作战收益。其本质是一类带有约束的组合优化问题。大部分目标分配问题可以统一描述为如下形式: 将数量为N_{t}个目标分配给N_{r}个我方智能体, 使得总的效益J^*最大化。数学表述形式如下:

$$J^* = \max \sum_{i=1}^{N_{\mathrm{r}}} \sum_{j=1}^{N_{\mathrm{t}}} V_{ij} x_{ij} \tag{1}$$

式中: V_{ij}表示将目标t_j分配给智能体a_i所获得的收益值; x_{ij}表示是否将目标T_j分配给智能体a_i, $x_{ij} \in \{0,1\}$, 一般的约束形式如下:

$$\sum_{j=1}^{N_{\mathrm{t}}} x_{ij} \leqslant L_i^{\mathrm{R}} \qquad \forall i, L_i^{\mathrm{R}} \geqslant 1 \tag{2}$$

$$\sum_{i=1}^{N_{\mathrm{r}}} x_{ij} \leqslant L_j^{\mathrm{T}} \qquad \forall j, L_j^{\mathrm{T}} \geqslant 1 \tag{3}$$

式(2)表示智能体a_i所分配到的目标总数不应超过其最大分配能力限制L_i^{R}, 式(3)表示任意一个目标T_j最多可以被L_j^{T}个智能体共同执行。

对于异构任务的多任务分配问题, 每个目标包含一个到多个可以被执行的任务, 每架无人机也同样包含执行不同种类任务的能力。因此, 解决这类多任务分配问题的思路不再是将无人机与目标匹配, 而是将无人机执行每种任务的能力与目标的任务相对应。假设一般情况下每个目标可被执行的任务分为探测 D、攻击 A、评估 V, 对任意目标 j所需执行的任务分为$T_j = \{T_j^D, T_j^A, T_j^V\}$。显然, 这 3 种任务的执行应当满足一定的先后顺序, 即任务间的时序约束, 除探测任务外, 其余 2 项任务需要上一阶段的任务已经被某一无人机执行后才能进行。而参与分配的无人机u_i可能具备这 3 项能力中的一项或多项:

$$\begin{cases} U_i^k \in \{U_i^1, U_i^2, U_i^3\} \\ U_i^1, U_i^2, U_i^3 \in \{0,1\}, \forall i \end{cases} \tag{4}$$

式中: U_i^k表示无人机执行某一任务的能力, 其中, $U_i^1, U_i^2, U_i^3 \in \{0,1\}, \forall i$分别表示无人机执行探测 D、攻击 A 与毁伤评估 V 的能力, $U_i^k = 0$表示该无人机不具备执行该项任务的能力, 1 表示该无人机具备执行该项任务的能力。

因此, 可以将含有时序约束的多任务分配问题目标函数总结为如式(5)所示的数学描述形式, 式(6)为无人机能力限制约束条件:

$$J^* = \max \sum_{i=1}^{N_{\mathrm{r}}} \sum_{j=1}^{N_{\mathrm{t}}} \sum_{k=1}^{N_k} \delta_j(k) V_{ij}^k x_{ij}^k \tag{5}$$

$$\sum_{j=1}^{N_{\mathrm{t}}} x_{ij}^k \leqslant U_i^k \qquad \forall i, k \in \{D, A, V\} \tag{6}$$

式中: N_k为不同种类任务的总数; V_{ij}^k为将第 j个目标的k项任务分配给无人机U_i获得的收益; x_{ij}^k表示是否将第 j个目标的k项任务分配给无人机u_i; $\delta_j(k)$为目标 j的$k-1$项任务是否已经被分配, $\delta_j(k) \in \{0,1\}$, $k \geqslant 2$。

1.2　无人机网络拓扑

在分布式计算方式下, 无人机之间的通信关系常通过图论中的节点与边来表示。因此, 可以将整

个分布式系统抽象为一个图$G=\{N,E\}$。其中，$N=\{h_1,h_2,\cdots,h_i\}$表示所有节点的集合，E表示节点间所有边的集合，若h_i、h_j间存在一条边，则数对$(i,j)\in E$。在本文所讨论的拓扑关系中，不考虑单向通信的情况，以及节点之间存在多重边或某一顶点到自身的环等复杂情形。因此，图G为一个简单无向图。邻接矩阵$\boldsymbol{A}_{dj}=(\tilde{a}_{ij})_{n\times n}$常被用来表示节点间的连通关系，若节点$h_i$与节点$h_j$可以通信，则$a_{ij}=1$，否则值为0。因此，对于无向图来说，其邻接矩阵为对称阵。

图1为本文实验所使用的无人机通信拓扑网络。图中，每个圆代表一个无人机节点，数字代表无人机节点编号，2个节点之间的连线表示通信连通，即2个节点之间可以进行双向通信。用通信跳数$\tilde{b}_{ij}$作为任意2个点n_i、n_j间拓扑距离的衡量。例如，图2中节点n_2与n_4之间的通信跳数$\tilde{b}_{2,4}=2$，节点n_1与n_7之间的通信跳数$\tilde{b}_{1,7}=7$。本文实验使用如图1中所有节点依次顺序相连的通信拓扑网络，代表了无人机网络在保证连通情况下最弱的连通条件，用于检验算法的正确性。

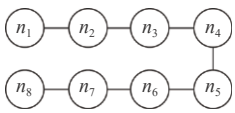


图1 无人机通信拓扑关系

Fig. 1 UAV communication topological relationships

2 异构多轮次分布式拍卖算法

2.1 异步分布式拍卖算法

异步分布式拍卖算法由传统的拍卖算法演化得来，由于在分布式计算下每一节点无法获得全局数据，以及可能由于异步计算所导致的数据滞后，异步分布式拍卖算法在拍卖算法的基础上引入了共识环节，用于实现个体对某些关键数据达成共识。

个体的每次计算可以细分为3个步骤来进行：①对全局最高出价人与报价表的更新；②依据自身的出价情况重新选择竞拍目标并更新报价；③检查迭代退出条件。

步骤1 更新全局报价表。更新全局最高出价人与报价表。每个计算节点在本地均存有对所有目标的最高出价人表与报价表，由于算法异步的原因，各节点所存数据可能不一致，在进行竞拍计算之前需要与周围邻居节点通信，并依据最大一致性原则对两表数据进行更新。

$$p_{ij}(t+1)=\max_{\forall_c\in N_i}\{p_{ij}(t),p_{cj}(t)\} \tag{7}$$

$$b_{ij}(t+1)=\max_{z\in Z}\{b_{zj}(t)\} \tag{8}$$

$$Z=\arg\max_{i,\forall c\in N_i}\{p_{ij}(t),p_{cj}(t)\} \tag{9}$$

式(7)表示对节点本地全局报价表的更新，p_{ij}为无人机u_i记录的对第j个目标的报价，N_i为所有邻居节点的集合。式(8)和式(9)表示对最高出价人表的更新，b_{ij}为该节点记录的第j个目标的最高出价人，其中，Z为可能记录了目标j的最高出价人的邻居节点的集合，可能存在多个邻居节点记录的对目标j报价相同的情况。式(7)与式(9)中c表示无人机u_i的某一邻居索引值。拍卖过程开始后，u_i先从邻居处接收估价信息，再根据最大一致性策略对所有目标重新定价。值得说明的是，如式(8)所示，由于异步计算可能导致存在多个节点对某一目标记录的最高出价相同而最高出价人不同的情况，需要在对b_i的更新过程中引入一种打破平局的机制，也即在缺乏全局信息的情况下，多个节点的出价达成平局时，具有最大索引的节点将赢得此次竞拍。

步骤2 更新竞拍目标及报价。重新选择竞拍目标并更新报价。在经过步骤1后，可以假定此时更新后的最高出价人表与报价表已经与全局数据达成共识。在此阶段，无人机u_i开始检查其对目标的出价是否被其余节点超过。如果被超过，u_i按照利润最大的原则重新选择目标，并增加对该目标的报价：

$$T_i(t+1)=\arg\max_{1\leqslant j\leqslant n}\{V_{ij}-p_{ij}(t+1)\} \tag{10}$$

$$b_{iT_i}(t+1)=i \tag{11}$$

$$p_{iT_i}(t+1)=p_{iT_i}(t+1)+\gamma \tag{12}$$

式(10)为重新选择能获得最大利润的目标，其中，T_i为无人机重新选择竞拍的目标；V_{ij}为无人机选择该目标所获得的收益。式(11)和式(12)分别为对本地最高出价人表和报价表的更新，其中，γ为报价增量，由最大利润P_i与次大利润P_i'之差和最小报价增量常数ε组成，计算式为

$$\gamma_i=P_i-P_i'+\varepsilon \tag{13}$$

式中：$\varepsilon>0$为最小报价增量常数，其大小取值与解的质量和计算复杂度相关。

步骤3 检查退出条件。检查报价表与最高出价人表与上一次计算是否相同。若连续D_{net}次相同(D_{net}为网络图直径，代表网络内最大通信跳数)，则认为算法收敛，可退出计算流程。为使异步算法对通信网络延迟具有一定的鲁棒性，实际使用中应使得连续检查次数大于D_{net}次。若检查次数过大，又会降低计算效率，应当对具体的应用问题进行折中

考量。

假设通信网络拓扑是连通的, 算法可以终止于有限次数的迭代, 迭代次数上限为

$$N^{\mathrm{up}} = D_{\mathrm{net}}(m-1)\cdot \max_{i=1,\cdots,m} \mathrm{ceil}\left(\frac{\max\limits_{j=1,\cdots,n}(V_{ij}) - \min\limits_{j=1,\cdots,n}(V_{ij})}{\varepsilon}\right) \tag{14}$$

式中: ceil(·)表示对数据进行向上取整运算。

2.2　异构多轮次分布式拍卖算法流程

异构多轮次分布式拍卖算法主要面向具有不同能力的异构无人机与具有时间顺序约束的多种任务的分配问题。在 2 轮分布式拍卖算法之间引入对目标的可用性协商阶段, 用于实现所有无人机对下一轮次需要进行分配的任务达成共识。

异构多轮次分布式拍卖算法通过引入虚拟目标的方式实现了无人机数量大于目标数量的任务分配问题($m > n$), 避免了无人机间的无休止竞拍而导致的算法无法收敛。通过在可用性协商阶段对虚拟目标的收益进行共识协商, 使得虚拟目标的收益值小于任意一架无人机对所有真实目标的最小收益。该举措既能保证虚拟目标不会干预真实目标的分配, 也能够避免对虚拟目标设置的收益过小导致算法收敛慢的问题, 是兼顾算法结果最优性与计算效率的解决方案。

每个计算节点所运行的总的计算流程如算法 1 所示, 该计算流程可分为 6 个计算阶段, 每 2 个计算流程进行一项任务的分配, 共需分配 3 个任务。当最后一个阶段的计算结束后, 算法自动停止。$\mathrm{AlreadyFlag}_i$表示该节点计算流程的停止标志, τ^i_{stage}为任务阶段, 对于仅有 3 个分配任务的计算流程来说, 任务阶段的取值范围为$\tau^i_{\mathrm{stage}} \in \{1,2,\cdots,6\}$。

算法 1　异构多轮次分布式拍卖算法。

1. while $\mathrm{AlreadyFlag_i}$ = false do
2. 　if $\tau^i_{\mathrm{stage}} \in \{1,2,\cdots,6\}$ then
3. 　　Consensus()
4. 　else
5. 　　while $C_i \leqslant \eta\Delta$ do //任务分配阶段
6. 　　for $j \leftarrow 1$ to n do
7. 　　$p_{ij}(t+1) = \max\limits_{\forall c\in N_i}(p_{ij}(t), p_{cj}(t))$
8. 　　$b_{ij}(t+1) = \max\limits_{z\in Z}(b_{zj}(t))$
9. 　　$Z = \arg\max\limits_{\forall c\in N_i}(p_{ij}(t), p_{cj}(t))$
10. 　　end
11. 　　if $p_{i\hat{T}^k_i}(t) \leqslant p_{i\hat{T}^k_i}(t+1)\&U^k_i$ = true then
12. 　　$\hat{T}^k_i = \arg\max\limits_{1\leqslant j\leqslant n}(V^k_{ij} - p_{ij}(t+1))$
13. 　　$b_{i\hat{T}^k_i}(t+1) \leftarrow i$
14. 　　$p_{i\hat{T}^k_i}(t+1) = p_{i\hat{T}^k_i}(t) + \gamma_i$
15. 　　end
16. 　　若p_i与b_i无变化, 计数器C_i自增 1; 否则C_i清零
17. 　end
18. 　　τ^i_{stage}与k自增 1, $T^k_i \leftarrow \hat{T}^k_i$
19. 　end
20. end

算法流程中, $\hat{T}^k_i$表示在第 k 项任务分配过程中, 无人机 u_i 当前时刻选定的目标, T^k_i表示最终分配的目标。可用性协商阶段的作用是使网络内的所有节点对本阶段所有目标的有效性达成共识, 并在此基础上调整自身对所有目标的收益, 为下一阶段的任务分配计算做数据准备。由于任务分配问题中的任务具有时序约束的特性, 也即只有上一阶段被分配的目标才能继续作为有效目标参与下一阶段的分配计算, 否则在下一阶段中该目标不应当被所有智能体优先竞拍, 因此, 需要被所有节点标记为虚拟目标并调小其收益, 通过拓扑网络共识使其小于全局内任意有效目标的收益。

算法 2　可用性协商阶段 Consensus()。

1. 　if τ^i_{stage} then
2. 　　while $C_i \leqslant \eta\Delta$ do
3. 　　if $A^{k-1}_i = 1\&\boldsymbol{\Psi}^{k-1}_i(T^{k-1}_i) = 1$ then
　　//该无人机具备执行此项任务的能力且其分配到的目标为非虚拟目标
4. 　　$\hat{\boldsymbol{\Psi}}^k_i(T^{k-1}_i) \leftarrow 1$
5. 　　end
　　//基于最大一致性原则对所有目标的有效性进行共识
6. 　　for $j \leftarrow 1$ to n do
7. 　　$\hat{\boldsymbol{\Psi}}^k_i(j) = \max_{l\in N_i}(\hat{\boldsymbol{\Psi}}^k_i(j), \hat{\boldsymbol{\Psi}}^k_l(j))$
8. 　　end
　　//基于最小一致性原则对所有有效目标的最小收益共识
9. 　　$v^i_{\min} = \min\limits_{l\in N_i}(v^i_{\min}, v^l_{\min})$
　　//更新计数器
10. 　若$v^i_{\min}$与$\hat{\boldsymbol{\Psi}}^k_i$无变化, C_i自增 1, 否则清零
11. 　$\boldsymbol{\Psi}^k_i \leftarrow \hat{\boldsymbol{\Psi}}^k_i, \tau^i_{\mathrm{stage}}$自增 1; 根据共识结果重新调整$\boldsymbol{V}^k_i$

如算法 2 所示, 可用性协商阶段计算流程可总结为 3 个主要步骤: ①每个节点依据自身上一阶段分配结果对有效目标进行标记; ②通过通信网络各智能体分别基于最大一致性与最小一致性原则对所有目标的有效性和最小收益进行共识; ③检查退出条件, 并对收益进行调整。

在算法流程中, 某一无人机u_i自身保存的变量

有$\boldsymbol{U}_i=[U_i^1,U_i^2,U_i^3]$为自身的能力数据，$T_i^{k-1}\in\{1,\cdots,n\}$为上一次分配阶段的分配结果，$\boldsymbol{\Psi}_i^{k-1}=[\Psi_{i,1}^{k-1},\Psi_{i,2}^{k-1},\cdots,\Psi_{i,n}^{k-1}]$表示上一次协商对所有目标的有效性共识结果。算法的输出为$\boldsymbol{\Psi}_i^k$，代表本次有效性的共识结果，以及在此基础上对所有目标重新调整后的收益$\boldsymbol{V}_i^k=[V_{i,1}^k,V_{i,2}^k,\cdots,V_{i,n}^k]$。算法需检查自身分配所得目标的有效性，满足自身具备对应能力及该目标已经为有效目标：

$$\hat{\boldsymbol{\Psi}}_i^k(T_i^{k-1})=\begin{cases}1 & U_i^{k-1}与\boldsymbol{\Psi}_i^{k-1}(T_i^{k-1})同时为真\\ 0 & 其他\end{cases} \tag{15}$$

式(15)表示某节点先依据分配结果及自身能力对本地目标可用性向量$\hat{\boldsymbol{\Psi}}_i^k$进行标记，再通过拓扑网络对其余目标的有效性进行共识，基于最大一致性原则，若目标T_z被某一无人机u_j标记为有效目标，则该节点同样将其标记为有效目标：

$$\hat{\boldsymbol{\Psi}}_i^k(T_z)=\begin{cases}1 & \hat{\boldsymbol{\Psi}}_j^k(T_z)=1,\forall j\in\{1,2,\cdots,m\}\\ 0 & 其他 \quad z\in\{1,\cdots,n\}\end{cases} \tag{16}$$

式(15)和式(16)中，$\hat{\boldsymbol{\Psi}}_i^k$的长度与上一轮次参与拍卖的目标总数相对应，每个节点仅需对自身分配得到的目标的可用性负责，其余目标的可用性均从通信网络获得，无需考虑其是被哪一节点所标记。

参考上述最大一致性共识原则，每个节点同时也需要对所有有效目标的全局最小收益$v_{\min}$共识。具体计算过程如下式所示，每轮迭代中均取自己所能通信到的范围内的最小值：

$$v_{\min}=\min_{\forall c\in N_i}\{v_{\min}^i,v_{\min}^c\} \tag{17}$$

在对全部目标的有效性及全局最小收益达成共识后，所有节点需要对本地存储的目标的收益进行调整，保证虚拟目标的收益不大于任何一个有效目标的收益值。对于无人机u_i来说，具体的收益调整原则如下：

$$\boldsymbol{V}_i^k(T_j^{\mathrm{V}})<v_{\min}-\varepsilon \qquad \forall T_j^{\mathrm{V}}\in\boldsymbol{T}^{\mathrm{V}} \tag{18}$$

式中：$v_{\min}$为全局有效目标收益集合中的最小值；$\boldsymbol{T}^{\mathrm{V}}$为所有虚拟目标的集合。

为防止多个虚拟目标的收益值冲突，本文算法所使用的更新规则如下：

$$\boldsymbol{V}_i^k(T_j^{\mathrm{V}})=v_{\min}-\varepsilon-I_j \tag{19}$$

式中：I_j为目标j的全局索引。

调整完本地收益后，保存对所有目标的可用性共识结果：

$$\boldsymbol{\Psi}_i^k\leftarrow\hat{\boldsymbol{\Psi}}_i^k \qquad \forall i \tag{20}$$

某一节点数据改变需经过通信链路依次流经所有节点，为保证异步算法对网络通信延迟的鲁棒性，当$\hat{\boldsymbol{\Psi}}_i^k$与$v_{\min}$在连续多数个周期内无变化，则允许该阶段退出，$\Delta$为通信网络半径。本阶段算法使用计数器$C_i$控制流程，若$\hat{\boldsymbol{\Psi}}_i^k$与$v_{\min}$相较上一周期无变化，$C_i$自增1，若有变化则$C_i$清零。直到$C_i>\eta\Delta$时结束该流程。注意，$\eta$取至少大于2的整数，否则不具备对网络延迟的鲁棒性。

整体的可用性协商阶段可以用图2所示流程概括。首先，对本地目标可用性数据进行一次更新；然后，利用通信网络对目标可用性表与全局有效目标最小收益进行多次迭代共识，直至共识结束后，各节点调整自身本地目标收益并开始下一轮拍卖。

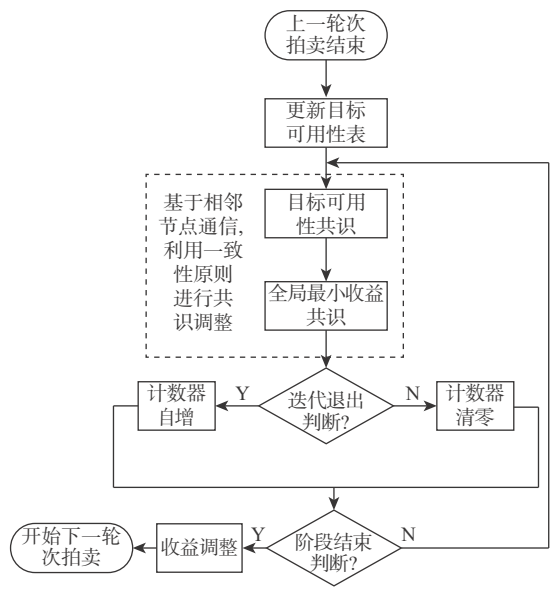


图2 可用性协商阶段流程

Fig. 2 Flowchart for the availability negotiation phase

上述3个主要计算步骤完成后即可进入拍卖阶段，直至完成所有轮次的可用性协商与竞拍流程后，算法终止，每个节点得到各自分配结果。其中，竞拍阶段与异构多轮次分布式拍卖算法计算流程相同，不再赘述。

3 仿真结果与分析

分别针对目标数大于无人机节点数及目标数小于无人机节点数2种情况设计了2组实验，共有8架能力异构的无人机参与分配，采用就近分配原则使得分配结果为总的作战行程最短。无人机位置与目标位置随机设定，分别在长宽均为50 m的矩形二维区域内随机分布，且分布区域无重叠。收益值由距离优势与目标重要度组成，距离收益值设置为无人机与目标之间距离的负值，表示距离越大收益值越小，使得每个无人机更倾向于优先分配距离自较近的目标，同时兼顾目标的重要度等级，使

重要目标优先分配。使用 $\tilde{P}_i=[x_i,y_i]$ 与 $\tilde{Q}_j=[x_j,y_j]$ 分别表示无人机 u_i 与目标 T_j 的坐标向量, 具体收益计算式如下:

$$V_{ij}=(V_{ij}^{\mathrm{D}})^{\lambda^1}(V_j^{\mathrm{T}})^{\lambda^2}\lambda^1+\lambda^2=1 \tag{21}$$

$$\begin{cases} V_{ij}^{\mathrm{D}}=-\left\|\tilde{P}_i-\tilde{Q}_j\right\| \\ V_j^{\mathrm{T}}=1/\tilde{n}_j,\tilde{n}_j\in\{1,2,3,\cdots\} \end{cases} \tag{22}$$

式中: $\tilde{n}$ 为目标的重要度等级; λ^1 和 λ^2 为二者分配权重。

本文算法属于异步分布式算法, 为模拟在真实应用场景中计算调用与通信在各节点之间不同步情况, 本文仿真实验采用各节点被均匀随机调用且保证在任意时间间隔内各节点被调用次数最大差值不大于通信网络半径:

$$\left|\xi_i^t-\xi_j^t\right|<\Delta \qquad \forall 1\leqslant i,j\leqslant m,\forall t\in[t_{\min},\infty) \tag{23}$$

式中: ξ_i^t 为在 t 时间间隔内该节点被调用的总次数; $t_{\min}$ 为最小调用时间间隔。

3.1　想定场景 1

仿真初始条件为: 无人机数量为 8, 目标数量为 10, 各目标重要程度相等。表 1 为各无人机能力设置。

表 1　无人机探测、攻击、毁伤评估能力设置（场景 1）

Table 1　UAV detection, attack, and damage assessment capability setting (Scenario 1)

能力	1号无人机	2号无人机	3号无人机	4号无人机	5号无人机	6号无人机	7号无人机	8号无人机
探测	√	×	√	√	×	√	√	√
攻击	√	√	×	√	√	×	√	√
毁伤评估	×	√	×	×	√	×	√	×

注: √表示具备能力, ×不具备能力。

图 3～图 5 及表 2 为场景 1 的 3 个任务阶段分配结果。在场景 1 中, 除 2 号与 5 号无人机以外, 共有 6 架无人机具有探测的能力, 因此, 在第 1 阶段的分配中, 共有 3,4,6,7,8,9 号 6 个目标被执行了探测任务。在第 2 阶段的分配中, 3,4,6,7,8,9 号目标作为有效目标参与分配, 其余目标被标记为虚拟目标, 依据可用性协商阶段的最小一致性原则, 被所有无人机调整收益为全局最小值。

在第 2 阶段分配中, 1,2,4,5,7,8 号无人机具有执

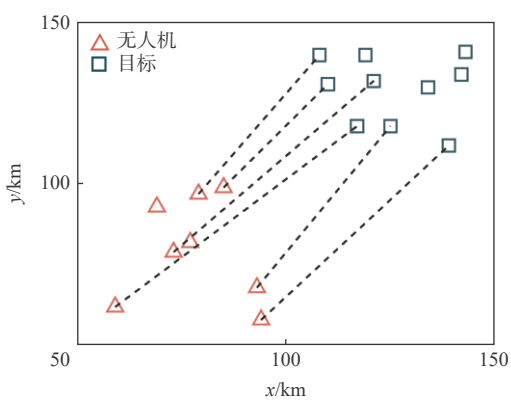


图 3　场景 1 探测任务分配结果示意图

Fig. 3　Schematic diagram of the task allocation results of the detection in scenario 1

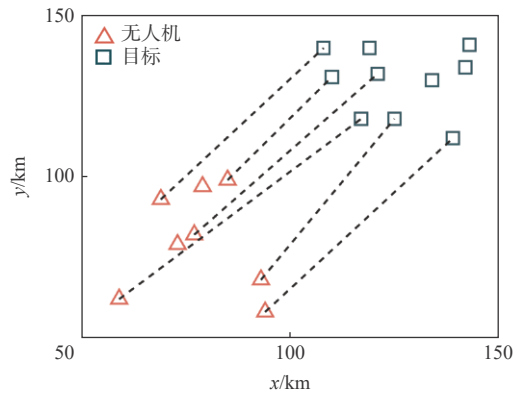


图 4　场景 1 攻击任务分配结果示意图

Fig. 4　Schematic diagram of the task allocation results of the attack in scenario 1

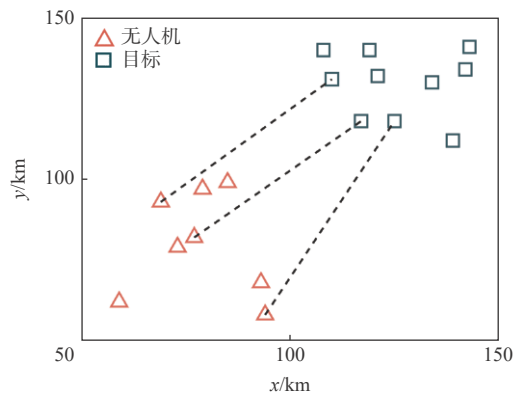


图 5　场景 1 毁伤评估任务分配结果示意图

Fig. 5　Schematic diagram of the task assignment results of the damage assessment in scenario 1

表 2　各无人机探测、攻击与毁伤评估任务对各目标的分配结果（场景 1）

Table 2　Summary table of the allocation results of each UAV detection, attack and evaluation mission to each target (Scenario 1)

无人机编号	目标1	目标2	目标3	目标4	目标5	目标6	目标7	目标8	目标9	目标10
1								*D/A*		
2				*V*			*A*			
3			*D*							
4				*D/A*						
5			*A*			*V*				
6							*D*			
7								*V*	*D/A*	
8						*D/A*				

行攻击任务的能力，共同对3,4,6,7,8,9号目标进行分配。

在第3阶段分配中，2,5,7号无人机具有执行毁伤评估任务的能力，共同对3,4,6,7,8,9号目标进行分配。

无论是哪一阶段的任务分配均未出现无人机数量大于目标数量的情况，在第2个实验中，通过修改实验条件，将出现$m>n$的任务分配，以验证该算法下的无人机具有自行添加虚拟目标、避免拍卖出现无休止竞争的能力。

表3为场景2无人机探测、攻击、毁伤评估能力设置，图6～图8及表4为分配结果。想定情景2出现8架无人机分配6个目标的情况，检验算法

表3 无人机探测、攻击、毁伤评估能力设置（场景2）

Table 3 UAV detection, attack, and damage assessment capability setting (Scenario 2)

能力	UAV 1	UAV 2	UAV 3	UAV 4	UAV 5	UAV 6	UAV 7	UAV 8
探测	√	√	√	√	√	√	×	√
攻击	√	×	√	×	√	×	√	√
毁伤评估	√	×	×	×	×	√	×	×

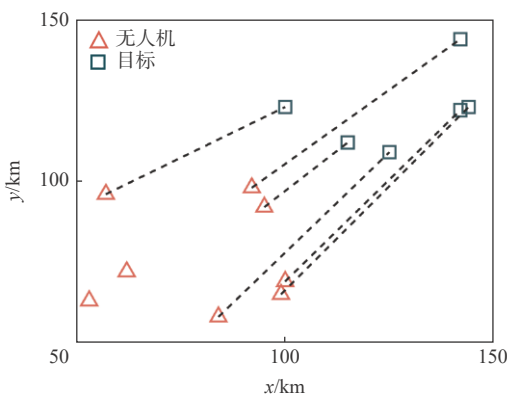


图6 场景2探测任务分配结果示意图

Fig. 6 Schematic diagram of the task allocation results of the detection in scenario 2

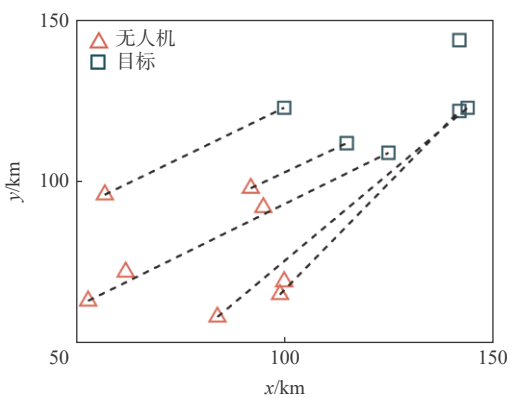


图7 场景2攻击任务分配结果示意图

Fig. 7 Schematic diagram of attack task allocation results in scenario 2

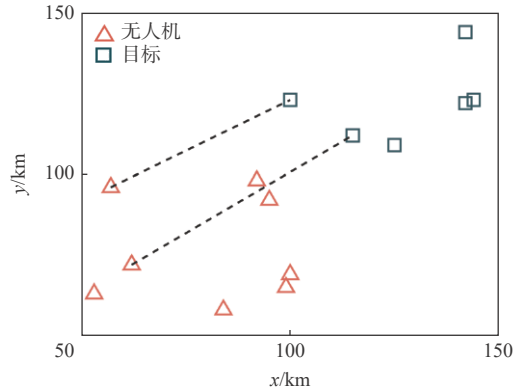


图8 场景2毁伤评估任务分配结果示意图

Fig. 8 Schematic diagram of the task allocation results of the destruction assessment in scenario 2

表4 各无人机探测、攻击与评估任务对各目标的分配结果（场景2）

Table 4 Summary table of the allocation results of each UAV detection, attack and evaluation mission to each target (Scenario 2)

无人机编号	目标1	目标2	目标3	目标4	目标5	目标6
UAV1		*D*/*A*/*V*				
UAV2					*D*	
UAV3			*D*	*A*		
UAV4				*D*		
UAV5	*A*					*D*
UAV6				*V*		
UAV7						*A*
UAV8	*D*				*A*	

具有自行添加虚拟目标的能力。在攻击任务分配阶段，1,3,5,7,8号无人机分别分配到2,4,1,6,5号目标，因此，这5个目标作为有效目标参与评估任务的分配。其中，1,6号无人机具有执行评估任务能力，共有2架，1,6号无人机需从5个目标中选出2个目标，分配结果为：分别分配到2,4号目标可使总行程最短。由于作战资源不足，导致其余目标流拍。

实验结果中存在有目标未被分配的现象，从算法的设计与具有时序约束的任务分配问题的角度来看，主要有2个方面的原因：①该目标的上一阶段任务未被分配，鉴于任务具有时序约束的关系，因此，本阶段目标的任务不能被分配。例如，未被执行探测任务的目标不能被执行攻击任务，未被攻击的目标也无法对其进行毁伤评估。②无人机群体的能力不能覆盖掉所有的目标，因此，分配结果是优先执行能使群体收益更大的目标，反映在实验结果中就是距离无人机群体更近的目标会被优先分配。

3.2　想定场景 2

本文在想定场景 1 中验证了目标数量大于无人机数量的分配效果, 而在想定场景 2 中, 将仿真初始条件中的目标数量从 10 降低为 6, 算法需自行添加虚拟目标并调整其收益值, 可用以检验算法对由于无人机数量大于目标数量导致的无休止竞拍问题的解决效果。8 架无人机的能力设置如表 3 所示。

从图 9 可以看出, 在想定场景 2 中, 运行在 1 号无人机的算法先添加了 2 个虚拟目标(虚拟 1、虚拟 2), 并将其收益值调整为全局最小, 由于目标 3 未被执行攻击任务, 在评估任务分配前算法将目标 3 设定为虚拟目标, 并调整其收益值为全局最小。

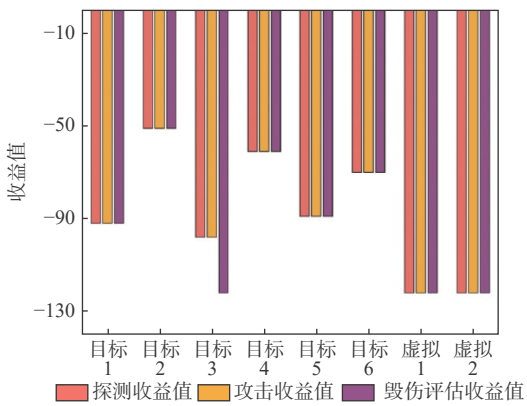


图 9　无人机 1 对目标的收益调整过程

Fig. 9　The process of adjusting the earnings of UAV1 to the target

3.3　对通信延迟的鲁棒性测试

在实际工程应用中, 通信延迟和异步计算往往是分布式计算机制无法避免的问题。本节对本文算法的异步计算和通信延迟的鲁棒性进行测试, 检验算法计算结果的正确性及收敛效率。在关于计数器的描述中, η为控制计数器容量的变量, 计数器容量越大, 能够容忍通信延迟的鲁棒性也就越大, 但是效率也会降低。实验通过改变η来改变算法的收敛效率, 并进行多组实验, 实验结果表明, 增大η可以增大算法对通信延时的鲁棒性, 使得算法在应对随机的异步计算时有更大的概率得到正确的计算结果, 但是会导致算法收敛效率的降低。

重新想定一个 5 架无人机对 5 个目标的简易分配场景, 任何一架无人机均具备探测、攻击和毁伤评估 3 项能力。实验中, 随机调用各节点进行单步计算, 模拟异步计算与通信延时。任一时刻仅有一个节点被调用, 且各节点被调用的概率相同, 且在任意一段时间内无需满足式(23)所示的最大总调用次数之差约束。分别设置$\eta=4,6,8$进行 3 组实验, 3 组实验初始条件均相同, 仅节点调用次序存在随机性差异, 统计 3 轮分配均得到正确结果的各节点总调用次数。具体实验结果如图 10 所示。

从图 10 的实验结果中可以看出, 增大η会使所有节点经过更多次计算才能得到结果。减小η确实能提高计算效率, 减少平均收敛时间, 但代价是更可能导致错误的计算结果。如下实验对算法得到正确或错误计算概率进行分析, 实验结果与结论基本一致。

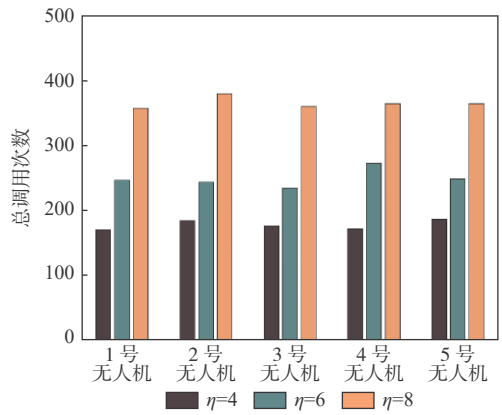


图 10　不同鲁棒性条件设置下算法收敛性能测试

Fig. 10　Algorithm convergence performance test under different robustness conditions

保持上述实验条件不变, 分别取$\eta=2,4,6,8,10$进行 5 组实验, 每组总共进行 10 次重复实验, 统计每组实验结果的正确率, 具体实验结果如图 11 所示。

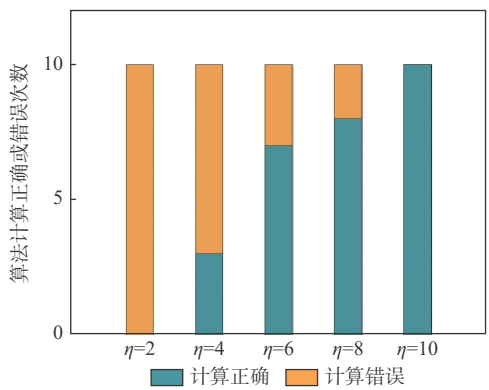


图 11　各组实验计算结果正确率统计

Fig. 11　Statistics on the correct rate of calculation results for each group of experiments

由图 11 可以看出, 随着η取值不断增大, 算法所计算结果的正确率也在不断提高, 在当前实验条件下, $\eta=10$时算法可以具有 100% 的正确率, $\eta=2$时算法不能应对通信延时, 正确率为 0。可以推测, 当实验条件更加恶劣时, 需要进一步提高η取值, 同时也需要付出更长时间收敛的代价。

4　结　论

1) 本文算法创新性地在 2 轮次拍卖中引入了

可用性协商阶段，在一定程度上较为灵活地解决了分布式计算机制下具有时序约束的多任务分配问题。

2）本文算法对分布式计算中导致的通信延时问题具有一定的鲁棒性，能够保证在任意拓扑条件下算法的收敛，计算量小具有较大的工程价值。

设计更加完善的分布式异常处理机制，支持个体的动态加入与退出，是今后工程应用与本文需要进一步研究的方向。

参考文献（References）

[1] 陈侠, 魏晓明, 徐光延. 基于分布式邀请拍卖算法的多机协同空战决策[J]. 系统仿真学报, 2014, 26(6): 1257-1261.
CHAN X, WEI X M, XU G Y. Cooperative air combat decision making for multiple UCAV based on decentralized invite auction algorithm[J]. Journal of System Simulation, 2014, 26(6): 1257-1261 (in Chinese).

[2] 贾高伟, 王建峰. 无人机集群任务规划方法研究综述[J]. 系统工程与电子技术, 2021, 43(1): 99-111.
JIA G W, WANG J F. Research review of UAV swarm mission planning method[J]. Systems Engineering and Electronics, 2021, 43(1): 99-111(in Chinese).

[3] WANG J F, JIA G W, XIN H B, et al. Research on dynamic task allocation method of heterogeneous multi-UAV based on consensus based bundle algorithm[C]//Proceeding of the 2020 Chinese Automation Congress. Piscataway: IEEE Press, 2020: 2214-2219.

[4] LIU P F, WANG B, LIU W J, et al. Multi-task allocation of multi-UAV coalition based on improved quantum genetic algorithm[C]//Proceeding of the 2021 40th Chinese Control Conference. Piscataway: IEEE Press, 2021: 1802-1807.

[5] Y YU J Y, GUO J S, ZHANG X F, et al. A novel tent-levy fireworks algorithm for the UAV task allocation problem under uncertain environment[J]. IEEE Access, 2022, 10: 102373-102385.

[6] CHEN Y H, CHEN J C, DU C L. Allocation of multi-UAVs timing-dependent tasks based on completion time[C]//Proceeding of the 2022 WRC Symposium on Advanced Robotics and Automation. Piscataway: IEEE Press, 2022: 71-76.

[7] TANG Y F, DOU L Q, ZHANG R L. A novel deep learning scheme for cooperative task allocation of large-scale UAVs[C]//Proceeding of the 2022 41st Chinese Control Conference. Piscataway: IEEE Press, 2022: 1866-1871.

[8] ZHANG W, LI Z X, WU W H, et al. A bilevel task allocation method for heterogeneous multi-UUV recovery system[J]. Ocean Engineering, 2023, 274: 114057.

[9] CHEN X Y, ZHANG P, DU G L, et al. A distributed method for dynamic multi-robot task allocation problems with critical time constraints[J]. Robotics and Autonomous Systems, 2019, 118: 31-46.

[10] XUE J B, ZHANG H, HU Q C. A low-cost communication security scheme of UAV based on multi-round reverse auction model[J]. Pervasive and Mobile Computing, 2022, 85: 101652.

[11] SHI J Q, TAN L, LIAN X F, et al. A multi-unmanned aerial vehicle dynamic task assignment method based on bionic algorithms[J]. Computers and Electrical Engineering, 2022, 99: 107820.

[12] KARELS V C G, VEELENTURF L P, VAN WOENSEL T. An auction for collaborative vehicle routing: models and algorithms[J]. EURO Journal on Transportation and Logistics, 2020, 9(2): 100009.

[13] RINALDI M, PRIMATESTA S, GUGLIERI G, et al. Auction-based task allocation for safe and energy efficient UAS parcel transportation[J]. Transportation Research Procedia, 2022, 65: 60-69.

[14] JIN L, LI S, LA H M, et al. Dynamic task allocation in multi-robot coordination for moving target tracking: a distributed approach[J]. Automatica, 2019, 100: 75-81.

[15] ZHANG M, LI W, WANG M M, et al. Helicopter-UAVs search and rescue task allocation considering UAVs operating environment and performance[J]. Computers & Industrial Engineering, 2022, 167: 107994.

[16] NEDJAH N, DE MENDONÇA R M, DE MACEDO MOURELLE L. PSO-based distributed algorithm for dynamic task allocation in a robotic swarm[J]. Procedia Computer Science, 2015, 51: 326-335.

[17] 冉华明, 周锐, 董卓宁, 等. 空战中协同干扰、探测、攻击任务分配[J]. 北京航空航天大学学报, 2015, 41(5): 911-918.
RAN H M, ZHOU R, DONG Z N, et al. Cooperative jamming-detecting-attacking task allocation in air combat[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(5): 911-918(in Chinese).

[18] 魏政磊, 赵辉, 黄汉桥, 等. 基于SAGWO算法的UCAVs动态协同任务分配[J]. 北京航空航天大学学报, 2018, 44(8): 1651-1664.
WEI Z L, ZHAO H, HUANG H Q, et al. Dynamic UCAVs cooperative task allocation based on SAGWO algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(8): 1651-1664(in Chinese).

[19] ZHEN Z Y, WEN L D, WANG B L, et al. Improved contract network protocol algorithm based cooperative target allocation of heterogeneous UAV swarm[J]. Aerospace Science and Technology, 2021, 119: 107054.

[20] KURDI H, ALDAOOD M F, AL-MEGREN S, et al. Adaptive task allocation for multi-UAV systems based on bacteria foraging behaviour[J]. Applied Soft Computing, 2019, 83: 105643.

[21] 陈侠, 刘永泰. 多UAV攻击移动目标的协同任务分配与航迹规划[J]. 火力与指挥控制, 2020, 45(9): 35-40.
CHEN X, LIU Y T. Cooperative task assignment and track planning for multi-UAV attack mobile targets[J]. Fire Control & Command Control, 2020, 45(9): 35-40(in Chinese).

[22] SHI J K, YANG Z, ZHU J W. An auction-based rescue task allocation approach for heterogeneous multi-robot system[J]. Multimedia Tools and Applications, 2020, 79(21-22): 14529-14538.

[23] MOORE B J, PASSINO K M. Distributed task assignment for mobile agents[J]. IEEE Transactions on Automatic Control, 2007, 52(4): 749-753.

[24] 邸斌, 周锐, 丁全心. 多无人机分布式协同异构任务分配[J]. 控制与决策, 2013, 28(2): 274-278.
DI B, ZHOU R, DING Q X. Distributed coordinated heterogeneous task allocation for unmanned aerial vehicles[J]. Control and Decision, 2013, 28(2): 274-278(in Chinese).

[25] ZAVLANOS M M, SPESIVTSEV L, PAPPAS G J. A distributed auction algorithm for the assignment problem[C]//Proceeding of the 47th IEEE Conference on Decision and Control. Piscataway: IEEE Press, 2008: 1212-1217.

[26] 颜骥, 李相民, 刘波. 考虑时序约束的多智能体协同任务分配[J].

控制与决策, 2015, 30(11): 1999-2003.
YAN J, LI X M, LIU B. Multi-agents cooperative task allocation with precedence constrains[J]. Control and Decision, 2015, 30(11): 1999-2003(in Chinese).
[27] 廖沫, 陈宗基. 基于多 Agent 分布协同拍卖的动态目标分配算法[J]. 北京航空航天大学学报, 2007, 33(2): 180-183.
LIAO M, CHEN Z J. Dynamic target assignment method based on multi-Agent decentralized cooperative auction[J]. Journal of Beijing University of Aeronautics and Astronautics, 2007, 33(2): 180-183 (in Chinese).

A heterogeneous multi-task assignment algorithm based on multi-round distributed auction

LYU Ye[1], ZHOU Rui[1,*], LI Xing[1], LIU Zhiheng[2], DI Bin[3]

(1. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China;
2. Nanjing Research Institute of Electronics Technology, Nanjing 210039, China;
3. Unit 32806 of the PLA, Beijing 100091, China)

Abstract: The algorithm solves the distributed cooperative task assignment problem with complex constraints among unmanned aerial vehicles (UAVs) with different capabilities in a more flexible way. Better solution accuracy and efficiency can be ensured by each UAV in the distributed task allocation framework by enabling distributed asynchronous computation and asynchronous communication. In order to prioritize significant targets with limited resources and the quickest total journey time, the task assignment principle is based on proximity assignment while taking the target value and schedule limitations into consideration. Simulation results show that this heterogeneous multi-round distributed auction algorithm can obtain good assignment results for any given number of UAVs and targets and UAV combat capability configuration.

Keywords: multi-machine collaboration; heterogeneous task assignment; distributed auction algorithms; task planning; timing constraints

Received: 2023-04-03; **Accepted**: 2023-07-15; **Published Online**: 2023-08-03 08:47
URL: link.cnki.net/urlid/11.2625.V.20230802.1250.003
Foundation item: National Natural Science Foundation of China (62003365)
* **Corresponding author**. E-mail: zhr@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0164

基于注意力机制的GRU-IKF场面滑行轨迹预测模型

刘雨生[1]，汤新民[1,2,*]，任宣铭[1]

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

摘 要：为解决机场场面滑行冲突、等待时间长等运行问题，保证场面安全的同时提高服务水平，增加机场吞吐量，针对机器学习模型性能依赖于良好数据集的现状，提出一种基于注意力机制、融合门控循环单元（GRU）和改进卡尔曼滤波算法（IKF）的场面航空器滑行轨迹预测模型。使用3个独立的门控循环单元网络来捕获航空器未来时刻的运动状态和时间上的依赖性，并引入注意力机制加强提取数据差异性特征的能力，学习输入到输出的映射关系；与改进后的扩展卡尔曼滤波器融合，将神经网络输出的结果整合到状态预测和更新过程，以提高预测轨迹序列的准确性。利用禄口机场航空器真实滑行轨迹对所提模型的有效性进行验证，仿真结果表明：所提模型能够对场面航空器滑行轨迹进行有效准确的预测，总体均方误差约为0.001 28，相较于单一循环神经网络（RNN）、长短时记忆（LSTM）网络及GRU模型，均方根误差（RMSE）分别减小72.9%，54.7%和39.9%，预测耗时40 ms，可以准确、快速预测滑行轨迹，为降低机场场面管理系统运行负荷提供帮助。

关 键 词：航空运输；改进卡尔曼滤波；GRU-Attention神经网络；机场场面运行；轨迹预测
中图分类号：V351.11; TB183
文献标志码：A **文章编号**：1001-5965(2025)03-1028-09

场面轨迹预测一直是机场智能化的研究重点，根据航空器历史运动轨迹，捕捉目标运动特征，估计航空器的未来位置，判断其运动趋势，对动态路由规划、航班次序调整、风险预警、重要节点安排等运行问题均具有关键作用，有助于减轻机场场面管理的运作负荷，提高整体安全水平和运行效率。

轨迹预测的核心是seq2seq，实质属于时间序列的预测问题，近几年，其主要分为以卡尔曼滤波算法为主的传统方法和各类神经网络为代表的机器学习模型。卡尔曼滤波是一种应用非常广泛且有效的目标跟踪和预测算法[1-5]，通常适用于不考虑控制输入或外部因素变化的轨迹数据（适用于参数随时间变化的动态系统）。文献[6]等提出了双层卡尔曼滤波对目标状态进行预测和修正的方法，仿真结果表明该方法可有效提高定位精度；文献[7]提出了模型转移概率改进自适应交互式多模型-拓展卡尔曼滤波（interacting multiple model-extended Kalman filter, IMM-EKF）算法，模型预测的峰值误差显著降低，实现了高精度的目标跟踪；文献[8]利用恒定转弯速率运动模型来预测目标未来运动趋势，并提出一种自适应容积卡尔曼滤波更新跟踪状态，引入亲和力模型来评估轨迹之间的相似性，实现了可靠的数据关联，提高了模型预测的准确性。上述方法适用于运动状态相对稳定，运动过程相对单一的轨迹，但对于复杂多变的航空器滑行轨迹预测精度有限，难以适应不同的运动状态和不确定性因素[9]。

基于数据驱动的模型可从大量原始轨迹数据中学习轨迹本质特征，拟合任意复杂度的函数，在

收稿日期：2023-04-04；**录用日期**：2023-06-02；**网络出版时间**：2023-06-16 17:06
网络出版地址：link.cnki.net/urlid/11.2625.V.20230616.0925.001
基金项目：国家重点研发计划(2021YFB1600500)；国家自然科学基金(61773202，52072174)；中国航空无线电电子研究所航空电子系统综合技术国防科技重点实验室基金(6142505180407)；中国民航管理干部学院民航通用航空运行重点实验室开放基金(CAMICKFJJ-2019-04)
*通信作者. E-mail：tangxinmin@nuaa.edu.cn

引用格式：刘雨生，汤新民，任宣铭. 基于注意力机制的GRU-IKF场面滑行轨迹预测模型[J]. 北京航空航天大学学报，2025，51(3)：1028-1036.
LIU Y S，TANG X M，REN X M. Prediction of aircraft surface trajectory based on the GRU-IKF model with attention mechanism[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(3)：1028-1036 (in Chinese).

飞行轨迹预测方面得到了广泛应用[10]。文献 [11] 基于动力学模型构造运动轨迹数据, 采用粒子群算法对长短时记忆网络(long short-term memory, LSTM)模型自动寻优, 并通过自适应增强算法改进模型框架, 提高预测精度; 文献 [12] 用杂交粒子群算法优化改进后的 Elman 网络参数, 并提出飞行坐标单独预测方法以提高精度, 收敛速度更快, 预测时间更短; 文献 [13] 提出一种基于深度学习的新型四维轨迹预测混合架构, 使用卷积神经网络捕获轨迹的空间维度特征, 通过 LSTM 网络挖掘其时间序列特征, 实现了对四维轨迹的准确预测; 文献 [14] 根据飞机爬升、巡航和下降等动态特性, 提出基于物理约束的 LSTM 网络轨迹预测方法, 利用空间聚类进行数据分割处理, 构建时空轨迹集验证了算法的可行性。由于扩展卡尔曼滤波(extended Kalman filter, EKF)利用一阶雅可比矩阵拟合非线性目标的精度较低, 神经网络模型对非线性的映射能力更强, 但数据样本集的优劣对预测结果影响巨大, 因此, 本文利用门控循环单元(gated recurrent unit, GRU)建立轨迹预测模型, 捕获序列数据中的动态信息, 并引入注意力机制强化提取数据多样性的能力, 增强轨迹关键信息对轨迹预测的影响; 考虑到系统的动态特性, 将神经网络输出结果应用于改进后卡尔曼滤波(improved Kalman filter, IKF) 预测和更新方程, 对轨迹输出序列进行实时调整, 提高预测精度; 提出了一种基于注意力机制的 GRU-IKF 模型的轨迹预测模型。

1　理论基础

1.1　GRU-attention 多层神经网络

1.1.1　GRU 网络

航空器滑行轨迹具有时变性, 常见的线性函数对坐标位置进行处理和预测非常困难, 目前广泛应用循环神经网络(recurrent neural networks, RNN)处理此类复杂的时间序列数据, 将当前时刻的输出和状态作为输入传递到下一时刻, 这种链型结构可以获取时间上的依赖性, 保持各时刻间的数据关系, 但随着计算深度增大会出现梯度消失、梯度爆炸的问题[15], 存在一定局限性, 因此, 在 RNN 基础上进一步产生了许多优秀的演化模型, 如 LSTM 网络和 GRU 模型, 通过加入记忆单元来解决长期依赖的问题, 利用门控机制保留尽可能多的轨迹信息, 避免梯度爆炸; 由于地面滑行轨迹数据相对较少, 不适合大规模训练, 且 GRU 相较于 LSTM 网络结构更为简单, 参数少, 收敛快, 因此, 本文选择 GRU 算法进行轨迹预测, 其内部结构如图 1 所示。

GRU 将 $t-1$ 时刻的隐藏状态和当前的轨迹信息作为输入来获得 t 时刻的轨迹信息, 通过更新门和重置门控制当前时刻受前一时刻航空器状态的

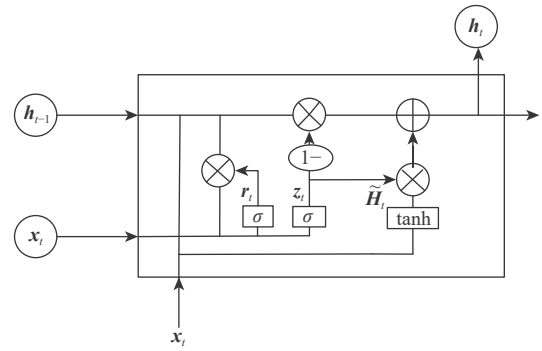


图 1　GRU 单元结构

Fig. 1　Structure of GRU unit

影响程度并遗忘航空器历史信息, 前者输出为 $\boldsymbol{z}_t$, 其大小与航空器状态信息成正比; 后者输出为 $\boldsymbol{r}_t$, 数值与丢弃信息的数量呈反比。相应的前向传播如下:

$$\boldsymbol{z}_t=\sigma(\boldsymbol{W}_z\boldsymbol{x}_t+\boldsymbol{U}_z\boldsymbol{h}_{t-1}+\boldsymbol{b}_z) \tag{1}$$

$$\boldsymbol{r}_t=\sigma(\boldsymbol{W}_r\boldsymbol{x}_t+\boldsymbol{U}_r\boldsymbol{h}_{t-1}+\boldsymbol{b}_r) \tag{2}$$

$$\tilde{\boldsymbol{h}}_t=\tanh(\boldsymbol{W}_h\boldsymbol{x}_t+\boldsymbol{U}_h(\boldsymbol{h}_{t-1}\odot\boldsymbol{r}_t)+\boldsymbol{b}_h) \tag{3}$$

$$\boldsymbol{h}_t=(\boldsymbol{1}-\boldsymbol{z}_t)\odot\boldsymbol{h}_{t-1}+\boldsymbol{z}_t\odot\tilde{\boldsymbol{h}}_t \tag{4}$$

式中: $\boldsymbol{x}_t$ 为输入向量; $\boldsymbol{h}_t$ 为输出隐藏状态; σ 为 sigmoid 激活函数; tanh为双曲正切激活函数; $\odot$ 表示点积; $\boldsymbol{W}$、$\boldsymbol{U}$ 为权重矩阵; $\boldsymbol{b}$ 为对应的偏置向量。

1.1.2　注意力机制

由于航空器轨迹受到诸多因素共同作用, 且不同变量对输出的影响程度不一, 可能对预测结果产生影响。因此, 本文引用注意力机制[16] 为不同时间步的输入分配权重来改善轨迹预测效果。

GRU-ATT(GRU-attention)多层神经网络架构图如图 2 所示, 若当前时刻为 $t+1$, 将此前 t 时刻 GRU 网络的隐藏状态输出 $\boldsymbol{h}_t$ 输入注意力机制层中, 得到相应的各时刻航空器滑行轨迹特征注意力权重 $\boldsymbol{\alpha}_t$, 重要的隐藏状态向量会被赋予更高权重; 最后, 对输出向量进行加权求和, 输出表示航空器轨迹特征将来全局变化趋势的隐藏层状态值 s。

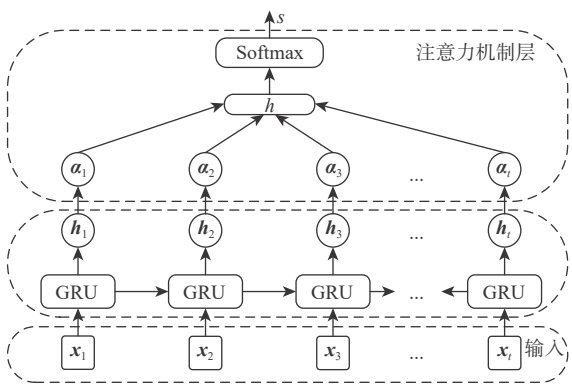


图 2　GRU-ATT 多层神经网络结构

Fig. 2　Structure of GRU-ATT multilayer neural network

1.2 改进扩展卡尔曼滤波

扩展卡尔曼滤波算法是一种最优的自回归数据处理算法，实时状态估计和预测性能较强。设非线性离散随机系统模型的状态方程及观测方程为

$$\boldsymbol{X}_k=\boldsymbol{A}_{k,k-1}\boldsymbol{X}_{k-1}+\boldsymbol{w}_{k-1} \tag{5}$$

$$\boldsymbol{Y}_k=\boldsymbol{H}_k\boldsymbol{X}_k+\boldsymbol{v}_k \tag{6}$$

式中：$\boldsymbol{X}_k$和$\boldsymbol{Y}_k$分别为k时刻的状态向量和量测向量；$\boldsymbol{A}_{k,k-1}$为$k-1$时刻到k时刻的状态转移矩阵；$\boldsymbol{H}_k$为量测矩阵；随机系统噪声$\boldsymbol{w}_{k-1}$和观测噪声$\boldsymbol{v}_k$为零均值互不相关的高斯白噪声。常规的卡尔曼滤波方程如下：

1）状态预测

$$\widehat{\boldsymbol{x}}'_{k,k-1}=\boldsymbol{A}_{k,k-1}\widehat{\boldsymbol{x}}_{k-1} \tag{7}$$

$$\widehat{\boldsymbol{P}}'_{k,k-1}=\boldsymbol{A}_{k,k-1}\widehat{\boldsymbol{P}}'_{k-1}\boldsymbol{A}_{k,k-1}^{\mathrm{T}}+\widehat{\boldsymbol{Q}}_{k-1} \tag{8}$$

2）卡尔曼增益

$$\boldsymbol{K}_k=\widehat{\boldsymbol{P}}'_{k,k-1}\boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{H}_k\widehat{\boldsymbol{P}}'_{k,k-1}\boldsymbol{H}_k^{\mathrm{T}}+\widehat{\boldsymbol{R}}_k)^{-1} \tag{9}$$

3）状态估计

$$\widehat{\boldsymbol{x}}_k=\widehat{\boldsymbol{x}}'_{k,k-1}+\boldsymbol{K}_k(\widehat{\boldsymbol{y}}_k-\boldsymbol{H}_k\widehat{\boldsymbol{x}}'_{k,k-1}) \tag{10}$$

$$\widehat{\boldsymbol{P}}_k=(\boldsymbol{I}-\boldsymbol{K}_k\boldsymbol{H}_k)\widehat{\boldsymbol{P}}'_{k,k-1} \tag{11}$$

式中：$\widehat{\boldsymbol{x}}'_{k,k-1}$为状态预测向量；$\widehat{\boldsymbol{x}}_{k-1}$，$\widehat{\boldsymbol{x}}_k$为$k-1$和$k$时刻的状态更新向量；$\boldsymbol{K}_k$为滤波器增益矩阵；$\widehat{\boldsymbol{P}}'_{k,k-1}$和$\widehat{\boldsymbol{P}}_k$分别为预测协方差和状态更新协方差矩阵；$\widehat{\boldsymbol{Q}}_{k-1}$和$\widehat{\boldsymbol{R}}_k$分别为过程噪声方差矩阵和量测噪声方差矩阵；$\widehat{\boldsymbol{P}}'_{k-1}$为$k-1$时刻最优滤波值的协方差矩阵；$\boldsymbol{I}$为单位矩阵。

由于标准扩展卡尔曼滤波受到系统模型和噪声统计模型不准确、误差累积等因素影响，无法对系统真实状态进行实时跟踪，精度降低。本文在协方差的更新方程式中引入一个对角渐消因子矩阵（见式(13)和式(14)），使得输出的残差序列在各个时刻都保持正交，能够有效改善预测效果。

$$\widehat{\boldsymbol{P}}'_{k,k-1}=\boldsymbol{D}_{k+1}\boldsymbol{A}_{k,k-1}\widehat{\boldsymbol{P}}'_{k-1}\boldsymbol{A}_{k,k-1}^{\mathrm{T}}+\widehat{\boldsymbol{Q}}_{k-1} \tag{12}$$

$$\boldsymbol{D}_{k+1}=\mathrm{diag}(\boldsymbol{\lambda}_1(k+1),\boldsymbol{\lambda}_2(k+1),\cdots,\boldsymbol{\lambda}_n(k+1)) \tag{13}$$

$\boldsymbol{D}_{k+1}$为多重渐消因子的对角矩阵，其可以通过式(14)求解：

$$\boldsymbol{\lambda}_n(k+1)=\begin{cases}a_k\boldsymbol{C}_k & a_k\boldsymbol{C}_k\geqslant 1\\ 1 & a_k\boldsymbol{C}_k<1\end{cases} \tag{14}$$

式中：a_k为标量因子，$a_k\geqslant 1$，当某分量发生突变时会增加以加速渐消因子的求取；$\boldsymbol{C}_k$表示如下：

$$\boldsymbol{C}_k=\frac{\mathrm{tr}(\boldsymbol{N}_k)}{\sum_{i=1}^{n}a_i\boldsymbol{M}_k} \tag{15}$$

式中：$\mathrm{tr}(\cdot)$为矩阵的迹；$\boldsymbol{N}_k=\boldsymbol{V}_{0,k}-\boldsymbol{H}_k\boldsymbol{Q}\boldsymbol{H}_k^{\mathrm{T}}-l_k\widehat{\boldsymbol{R}}_k$，$l_k\geqslant 1$，$l_k$为弱化因子，$\boldsymbol{V}_{0,k}$为均方误差；$\boldsymbol{M}_k=\boldsymbol{H}_k\boldsymbol{P}_{k-1}\boldsymbol{H}_k^{\mathrm{T}}$，改进后的扩展卡尔曼滤波，通过调节渐消因子，对预测协方差矩阵进行调整，增加当前观测值对滤波结果的修正作用，用不同的速率淡入淡化每个数据通道，使增益矩阵的调整更加有效，系统鲁棒性更强，估算结果更加准确、稳定。

2 本文模型

2.1 算法模型

从1.2节分析可知，卡尔曼滤波算法需要运动状态的线性方程，而随机系统噪声和观测噪声都是均值为**0**的高斯白噪声，对于非线性机动目标，其运动状态往往是不确定的，系统噪声和观测噪声的分布也是先验未知的。考虑到EKF拟合非线性只有一阶精度而神经网络在非线性近似上具有一定优势，本文提出卡尔曼滤波器融合神经网络，用于非线性目标跟踪的网络模型。具体来说，划分3种GRU-ATT网络模块来模拟卡尔曼滤波中未确定的运动状态$\boldsymbol{A}^{(\mathrm{GRU-ATT})_A}$、系统噪声$\widehat{\boldsymbol{Q}}_t^{(\mathrm{GRU-ATT})_Q}$和观测噪声$\widehat{\boldsymbol{R}}_t^{(\mathrm{GRU-ATT})_R}$。从训练数据中学习捕获输入输出之间复杂的状态转移关系，能够找到3个网络的最佳系数并将其应用于卡尔曼滤波的预测和更新过程计算，以获得更准确的目标状态预测。

这种结构的优点是不需要运动模型和先验分布，即可进行非线性机动目标的在线跟踪，突破线性高斯的限制，随着观测值的不断细化从而有效提高预测精度。轨迹预测方程如下：

$$\boldsymbol{X}_k=\boldsymbol{A}_{k,k-1}^{(\mathrm{GRU-ATT})_A}\boldsymbol{X}_{k-1}+\boldsymbol{w}_{k-1}\qquad \boldsymbol{w}\sim(\boldsymbol{0},\widehat{\boldsymbol{Q}}_{k-1}^{(\mathrm{GRU-ATT})_Q}) \tag{16}$$

$$\boldsymbol{Y}_k=\boldsymbol{H}_k\boldsymbol{X}_k+\boldsymbol{v}_k\qquad \boldsymbol{v}\sim(\boldsymbol{0},\widehat{\boldsymbol{R}}_k^{(\mathrm{GRU-ATT})_R}) \tag{17}$$

本文模型主要分为2个阶段，预测阶段需要将航空器滑行历史数据输入GRU-ATT神经网络模型，用$k-1$时刻的轨迹坐标来预测k时刻航空器的状态，但由于基于历史数据的预测模型未考虑到航空器在滑行过程中遇到的其他因素，因此，需要将第1阶段的输出结果作为第2阶段输入，结合改进的卡尔曼滤波对其进行实时动态调整，获得最终状态预测值，从而保证结果的准确性。卡尔曼滤波的预测方程和更新如下：

$$\widehat{\boldsymbol{x}}'_{k,k-1}=\boldsymbol{A}_{k,k-1}^{(\mathrm{GRU-ATT})_A}(\widehat{\boldsymbol{x}}_{k-1}) \tag{18}$$

$$\widehat{\boldsymbol{P}}'_{k,k-1}=\boldsymbol{D}_{k+1}\boldsymbol{A}_{k,k-1}^{(\mathrm{GRU-ATT})_A}\widehat{\boldsymbol{P}}'_{k-1}\boldsymbol{A}_{k,k-1}^{(\mathrm{GRU-ATT})_A\,\mathrm{T}}+\widehat{\boldsymbol{Q}}_{k-1}^{(\mathrm{GRU-ATT})_Q} \tag{19}$$

预测步将前一时刻的状态估计值$\widehat{\boldsymbol{x}}_{k-1}$与神经网络的输出结合，得到下一时刻的航空器状态，并更新协方差矩阵：

$$\boldsymbol{K}_k=\widehat{\boldsymbol{P}}'_{k,k-1}(\widehat{\boldsymbol{P}}'_{k,k-1}+\widehat{\boldsymbol{R}}_k^{(\mathrm{GRU-ATT})_R})^{-1} \tag{20}$$

$$\widehat{\boldsymbol{x}}_k=\widehat{\boldsymbol{x}}'_{k,k-1}+\boldsymbol{K}_k(\widehat{\boldsymbol{y}}_k-\widehat{\boldsymbol{x}}'_{k,k-1}) \tag{21}$$

$$\widehat{\boldsymbol{P}}_k=(\boldsymbol{I}-\boldsymbol{K}_k)\widehat{\boldsymbol{P}}'_k \tag{22}$$

更新步利用卡尔曼增益矩阵$\boldsymbol{K}_k$结合状态量测

向量$\widehat{\boldsymbol{y}}_k$，滤波器输出的$\widehat{\boldsymbol{x}}_k$和$\widehat{\boldsymbol{P}}_k$将作为下一时刻的输入，持续对航空器进行跟踪预测。本文模型轨迹预测结构图 3 所示，通过 3 个 GRU-ATT 模块进一步划分学习任务，分别生成状态预测值$\widehat{\boldsymbol{x}}'_{k,k-1}$和噪声矩阵$\widehat{\boldsymbol{R}}_k^{(\mathrm{GRU-ATT})_R}$，$\widehat{\boldsymbol{Q}}_{k-1}^{(\mathrm{GRU-ATT})_Q}$的估计值，然后通过式 (23)～式 (25) 计算卡尔曼增益修正预测值，减少当前值对预测的影响，以获得下一时刻的最优预测，整个预测过程中每个时间步均按照以上步骤进行迭代，最终获得完整的跟踪效果。算法内部结构如图 3 所示。

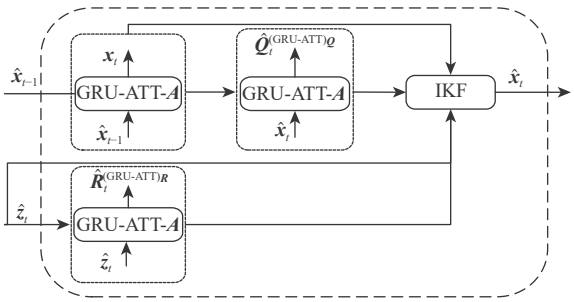


图 3　本文模型内部架构图

Fig. 3　Internal architecture of the proposed model

考虑到机场滑行道的限制，将滑行道尺寸约束条件引入模型（见式（23）），即预测位置不能超出滑行道边界。

$$\frac{|(x_0-x_i)(y_i-y_{i+1})-(x_i-x_{i+1})(y_0-y_i)|}{\sqrt{(x_i-x_{i+1})^2+(y_i-y_{i+1})^2}}<(1/2)r \tag{23}$$

式中：(x_0,y_0)和(x_i,y_i)分别为预测值和滑行道中线位置坐标；r为滑行道宽度。

2.2　构造时间序列样本集

考虑到使用非参数模型对航空器坐标按照坐标轴分量进行独立预测的方式，相较于将航空器全部位置信息作为整体输入预测模型的方式具备更高的精度[12]，因此，本文将航空器滑行的轨迹数据按照坐标轴独立提取建立数据集，预测航空器下一时刻的经纬度坐标分量，最后将经纬度组合构成航空器在预测时刻的滑行位置。当预测纬度坐标X时，需利用前n个时刻的轨迹数据，预测t_{n+1}时刻的坐标值。其函数映射关系为

$$\widehat{\boldsymbol{X}}_{n+1}=g(\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_n) \tag{24}$$

$$\boldsymbol{P}_i=(x_i,\varphi_i,v_i)\qquad i\in[1,n] \tag{25}$$

式中：$\boldsymbol{P}_i$为航空器在t_i时刻的纬度坐标x_i、航向φ_i及地速v_i，$\widehat{\boldsymbol{X}}_{n+1}$为飞机在$t_{n+1}$时刻的纬度坐标。采用时间窗口滑动预测坐标分量值，将连续前n个时刻航空器轨迹数据作为输入，$n+1$时刻的位置信息作为输出，具体如下：

$$\begin{bmatrix}\boldsymbol{P}_1 & \boldsymbol{P}_2 & \cdots & \boldsymbol{P}_n\\ \boldsymbol{P}_2 & \boldsymbol{P}_3 & \cdots & \boldsymbol{P}_{n+1}\\ \vdots & \vdots & & \vdots\\ \boldsymbol{P}_m & \boldsymbol{P}_{m+1} & \cdots & \boldsymbol{P}_{m+n}\end{bmatrix}=\begin{bmatrix}\boldsymbol{X}_{n+1}\\ \boldsymbol{X}_{n+2}\\ \vdots\\ \boldsymbol{X}_{m+n+1}\end{bmatrix} \tag{26}$$

2.3　数据归一化

由于滑行轨迹各维度的数据单位并不统一，为解决数据的不可比较性，消除不同量纲和数量级的影响，本文选择对输入的经纬度、速度、航向进行最值归一化处理，以提高预测的准确性和模型的鲁棒性，加快神经网络的收敛速度，表达式为

$$x'=\frac{x-x_{\min}}{x_{\max}-x_{\min}} \tag{27}$$

其中：x为原始样本数据，$x_{\max}$和$x_{\min}$为样本最大值和最小值；x'为经过归一化处理后的样本数据。

样本集建立后，根据神经网络模型结构展开轨迹预测，并利用改进的卡尔曼滤波算法对预测值进行实时调整，实现航空器滑行轨迹持续跟踪。本文模型流程如图 4 所示。

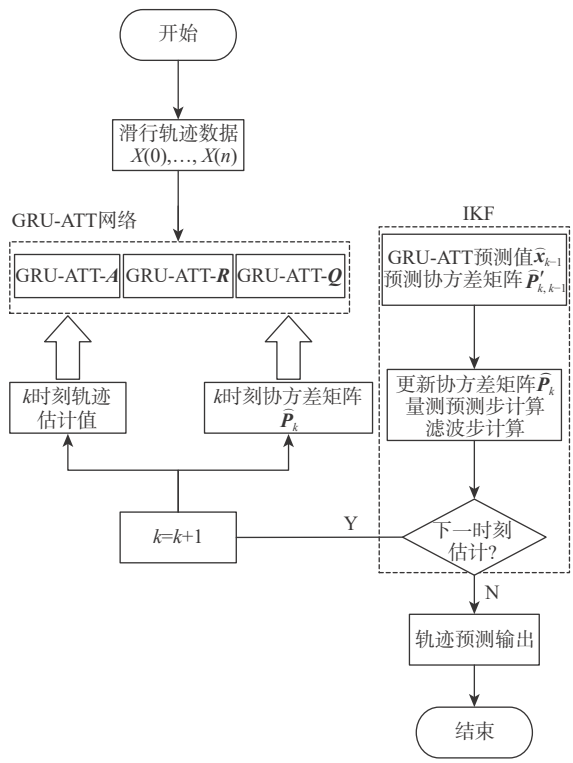


图 4　本文模型实施流程

Fig. 4　Flow chart of the proposed model

3　仿真实验及结果分析

本文将航空器真实滑行轨迹数据进行预测，根据场面轨迹预测结果图和预测误差进行数据分析，并与 LSTM 网络、RNN 等多种神经网络模型，以及传统轨迹预测方法进行实验对比，以验证本文模型可有效提高场面滑行轨迹预测的精度，具备一定普适性和优越性。

3.1　数据描述

为验证本文模型的可行性，从南京禄口机场 ADS-B 设备采集 2022 年 5 月 31 日上午 9 时—中午 12 时的真实航空器轨迹，筛选出 30 条地面滑行轨迹，采样

数据为非等间隔，约为 1 s，报文包含经度、纬度、时间、航空器呼号、速度、高度、航向等特征；生成样本集时，将经纬度转换为直角坐标，并将其归一化以简化计算，消除奇异样本数据导致的不良影响。

本文采用经纬度、地速及航向 4 个轨迹指标作为轨迹预测模型的输入，选取前 9 个时刻样本作为模型的初始输入，窗口长度为 5，滑动单位为 1，取 80% 的轨迹点作为神经网络的训练集，剩余 20% 的轨迹数据作为模型的测试集，训练采用 Adam 训练策略，学习率设定为 0.001；仿真运行环境为 Intel(R) Core(TM) i5-10400F CPU@2.90 GHz，内存为 8 GB，win10 系统的 Pytorch 平台。为提高预测精度采用坐标独立预测，与其他模型进行误差对比。部分轨迹数据如表 1 所示。

表 1　部分轨迹数据

Table 1　Part of trajectory data

时刻	纬度/(°)	经度/(°)	地速/(km·h^{-1})	航向/(°)
07:52:55	31.705 31	118.277 71	63.222 000 122 07	58.0
07:52:56	31.705 67	118.833 51	63.736 000 061 03	58.0
07:52:57	31.705 80	118.833 73	63.736 000 061 04	58.0

3.2　对比实验

由于运动模型难以精确建立，使得 EKF 等传统方法的预测步精度较低，本文通过 GRU-ATT 模块，提高了传统算法中预测步精度，通过不同模型对轨迹数据进行预测，并将结果进行对比，以验证本文模型的准确性。

对比真实和预测的轨迹图，由图 5 可直观看出 RNN 模型预测效果最差，GRU 比 LSTM 网络预测效果稍好，且 GRU-ATT 模型相较于单一神经网络模型的预测准确度更高，表明注意力机制可有效提取关键信息，提高精度，但这些模型整体轨迹点突变较多，并存在较大偏离，这是由于深度学习模型受数据集质量影响较大，数据缺失或异常导致的非等间隔数据集会降低预测效果；卡尔曼滤波的轨迹预测效果稍好于单一神经网络模型，而引入 IKF 后，多模型的预测结果均有一定改善，克服了对数据优劣的依赖性，整体轨迹偏离较小，本文模型预测效果最佳，二者预测轨迹与真实轨迹几乎重叠，实际滑行轨迹与预测值坐标位置点误差偏移很小，表明本文模型能够良好地完成轨迹预测任务，反映航空器真实滑行状态，为进一步验证本文模型的优越性，绘制不同模型的预测误差曲线。

仿真预测误差如图 6 和图 7 所示，横轴为时间

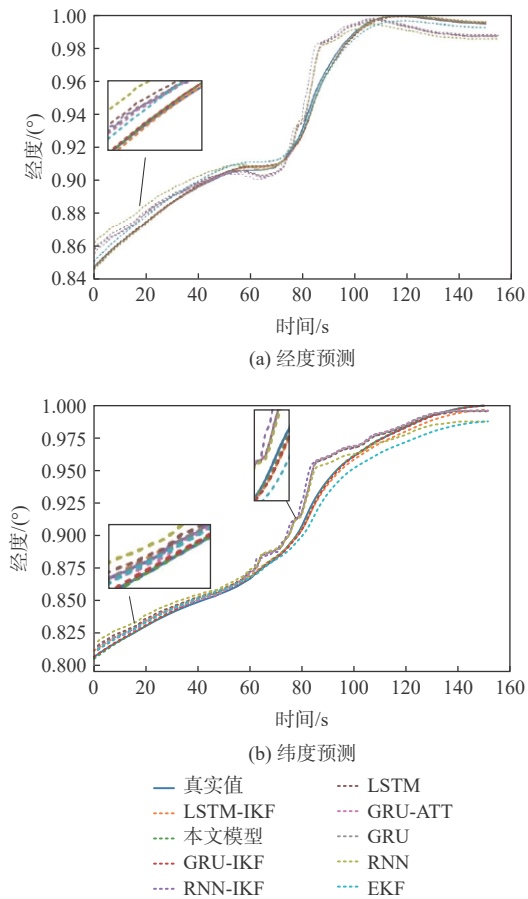


图 5　真实数据预测结果

Fig. 5　Prediction results of real data

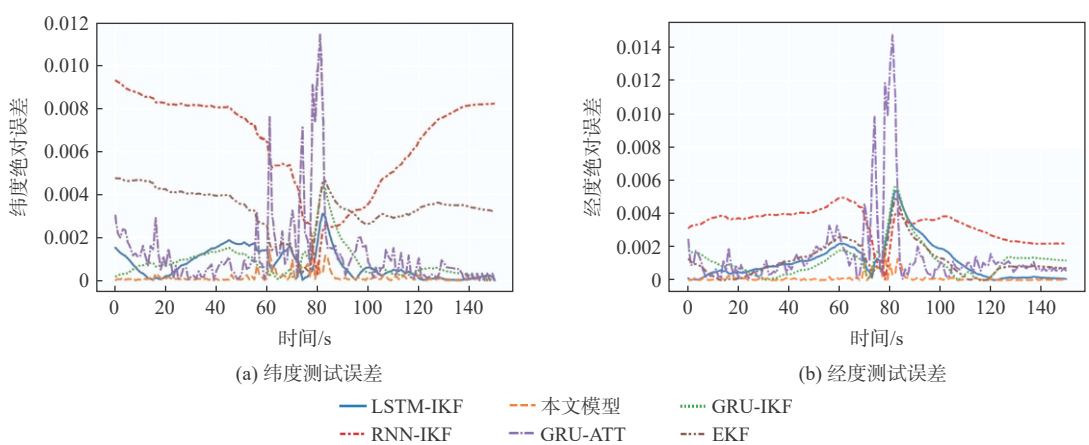
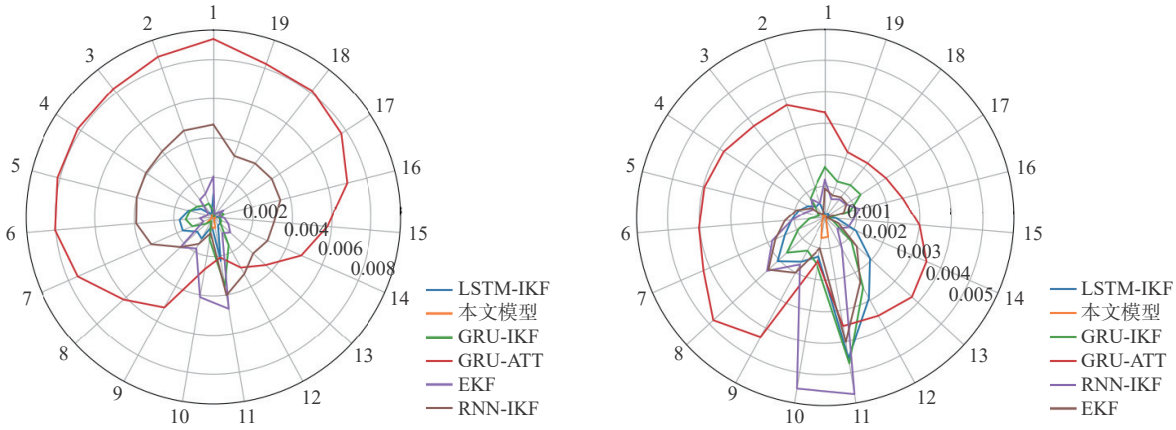


图 6　测试误差随时间变化曲线

Fig. 6　Prediction error curves with time



(a) 纬度预测误差　　(b) 经度预测误差

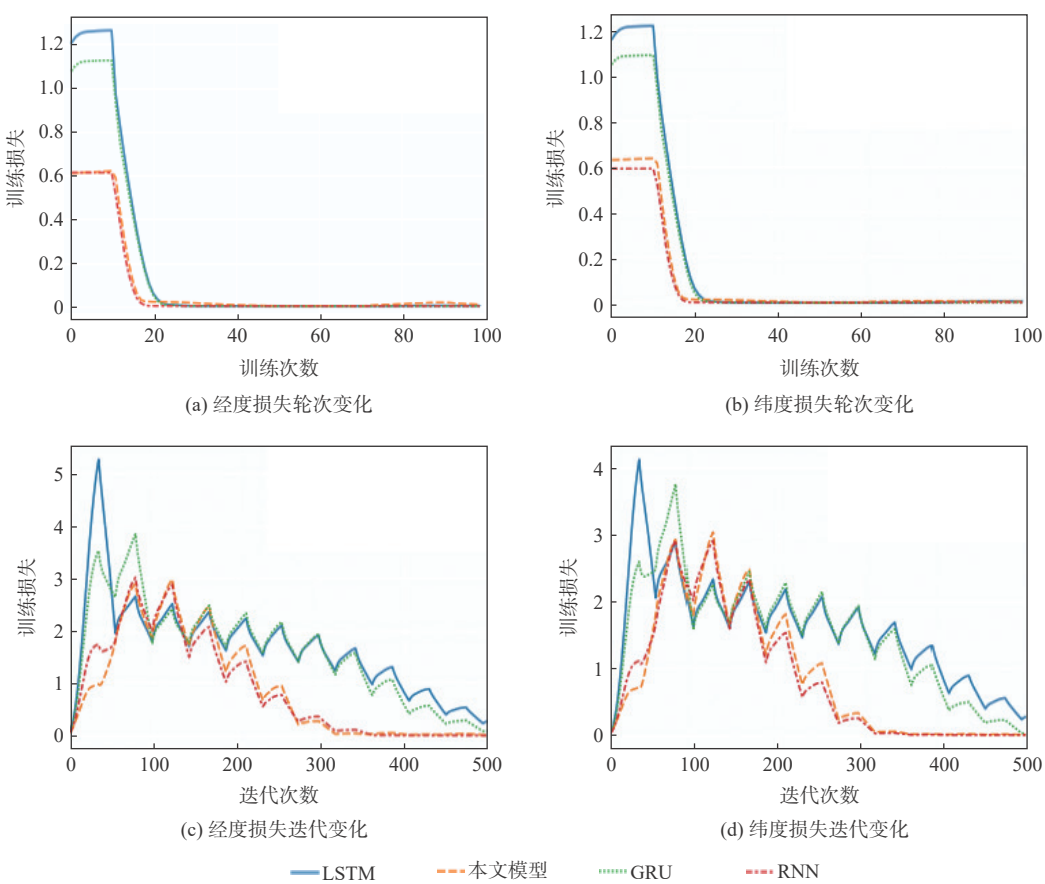
图 7　测试误差随训练轮次变化曲线

Fig. 7　Test error curves with train epoch

轨迹点，纵轴为归一化后真实值与预测值的绝对误差，可见，GRU-ATT 模型波动起伏明显，误差最大；RNN、GRU 和 LSTM 网络模型引入 IKF 修正后预测效果均有所改善；相比之下，本文模型对目标轨迹的预测误差更小，这表示本文模型处理时序数据的性能更强，学习的历史信息更全面，抗干扰能力更强，预测轨迹更接近于真实滑行轨迹。

本文将不同模型进行滑行轨迹预测时损失函数随训练回合的变化曲线绘制如图 8 所示，从图中的损失变化可见，LSTM 网络模型由于结构复杂，训练轮次较多，收敛最慢，RNN 模型收敛速度最快，图 8(c)～图 8(d) 能更直观看到训练损失随迭代次数的具体变化情况，可见本文提出的模型在训练和测试时，收敛速度与 RNN 模型相差无几，损失值小，可以有效提高预测效率。

为验证模型是否满足轨迹预测的实时性要求，



(a) 经度损失轮次变化　　(b) 纬度损失轮次变化

(c) 经度损失迭代变化　　(d) 纬度损失迭代变化

图 8　损失变化曲线

Fig. 8　Loss variation curve of real data

将本文模型与其他模型的单步预测时间进行对比，结果如图 9 所示。由图可知，所有模型的单步运行时间相差不大，本文模型的单步运行时间约为 0.04 s，虽然比其他模型较长，但是相较于 ADS-B 数据 0.5 s 的传输间隔，本文模型的单步运行时间仍可满足实时跟踪的需求。

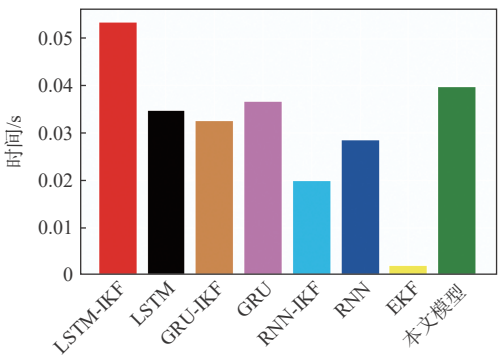


图 9 预测时间对比

Fig. 9 Comparison of prediction time

整体滑行轨迹预测效果如图 10，可见预测值与真实轨迹点基本重合，本文模型可有效完成轨迹预测任务。

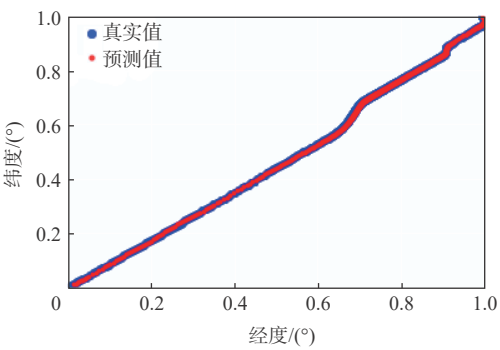


图 10 轨迹预测结果

Fig. 10 Results of trajectory prediction

3.3 评价指标

为进一步体现本文模型的优越性，分析本文模型的预测效果，衡量实际轨迹与预测值的误差精度，本文采用均方根误差（root mean square error，RMSE）作为评价指标：

$$e_{\mathrm{RMSE}}=\sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k-p_k)^2} \tag{28}$$

式中：n为样本长度；x_k为实际轨迹坐标分量；p_k为预测值。

从表 2 可以直观看出，本文模型均方根误差最小，RMSE 分别比单一的 RNN、LSTM 及 GRU 模型降低 72.9%，54.7% 和 39.9%，预测精度和模型拟合效果均有所提升，说明基于注意力机制的 GRU-IKF 模型可以深入捕获航空器输入轨迹信息，改良了运动目标轨迹预测的性能。

表 2 多模型轨迹预测结果的 RMSE

Table 2 RMSE of multi-model trajectory prediction results

模型	RMSE		
	经度	纬度	总体
RNN	0.003 04	0.006 44	0.004 74
LSTM	0.002 56	0.003 10	0.002 83
GRU	0.001 42	0.002 64	0.002 13
本文模型	0.001 55	0.001 01	0.001 28

4 结 论

本文针对场面滑行轨迹预测问题，将神经网络和卡尔曼滤波进行结合，并引入注意力机制提取轨迹数据的关键特征，得出结论如下：

1）针对场面航空器滑行跟踪，提出基于注意力机制的 GRU-IKF 轨迹预测模型，采用 GRU 神经网络结合注意力机制，可有效提取滑行轨迹的关键信息，获取目标运动过程和高斯噪声；改进后的卡尔曼滤波方程，相较于传统的扩展卡尔曼滤波，模型鲁棒性更强，通过对神经网络模型输出的轨迹预测值进行动态修正，克服了神经网络模型对于数据集质量的依赖及传统方法运动方程难以建立的缺点，实现了对航空器运动状态的准确估计。

2）对神经网络模型进行更深层次划分，采用 3 个模块对学习任务进一步分解，降低任务的复杂程度，在卡尔曼滤波的预测和更新过程将 3 个网络有机结合，相较于通过一个完整的 GRU-ATT 学习过程，这种安排能够使每个模块的学习任务更简单。与 LSTM 网络等神经网络及传统模型相比，本文模型对场面滑行轨迹展开预测时，收敛速度有所提高，表明本文模型在轨迹预测方面具备一定优越性，整体效果也有所提升。

3）通过数据驱动的方式，建立航空器运动状态空间模型，捕获输入和输出的映射关系，并将其应用于卡尔曼滤波的关键参数的估计。传统方法中，这些参数通常是由固定的模型决定的，或者通过经验获得，这就限制了其在非线性和非高斯目标跟踪中的应用。神经网络与卡尔曼滤波的融合有助于滤波更新过程中结合 GRU 网络得到的估计和观测结果，预测和滤波精度更高、鲁棒性更好。

参考文献（References）

[1] ZHANG X W, YU W Z. Research on the application of Kalman filter algorithm in aircraft trajectory analysis[C]//Proceedings of the 7th International Conference on Intelligent Computing and Signal Processing. Piscataway: IEEE Press, 2022: 196-199.

[2] 吕波, 王超. 改进的扩展卡尔曼滤波在航空器 4D 航迹预测算法中的应用[J]. 计算机应用, 2021, 41(S1): 277-282.

LU B, WANG C. Application of improved extended Kalman filter in 4D flight path prediction algorithm of aircraft[J]. Journal of Computer Applications, 2021, 41(S1): 277-282(in Chinese).

[3] SCHIMPF N, KNOBLOCK E J, WANG Z, et al. Flight trajectory prediction based on hybrid-recurrent networks[C]//Proceedings of the IEEE Cognitive Communications for Aerospace Applications Workshop. Piscataway: IEEE Press, 2021: 1-6.

[4] PANG Y T, LIU Y M. Conditional generative adversarial networks (CGAN) for aircraft trajectory prediction considering weather effects[C]//Proceedings of the AIAA Scitech 2020 Forum. Reston: AIAA, 2020.

[5] 陈明强, 傅嘉赟. 基于无迹卡尔曼滤波的飞行航迹预测方法研究[J]. 计算机仿真, 2021, 38(6): 27-30.
CHEN M Q, FU J Y. Research on flight path prediction method based on untraced Kalman filter[J]. Computer Simulation, 2021, 38(6): 27-30(in Chinese).

[6] 刘浩然, 覃玉华, 邓玉静, 等. 基于双层修正无迹卡尔曼的水下节点定位算法[J]. 仪器仪表学报, 2020, 41(3): 142-149.
LIU H R, QIN Y H, DENG Y J, et al. An underwater node localization algorithm based on double layer modified unscented Kalman filter[J]. Chinese Journal of Scientific Instrument, 2020, 41(3): 142-149(in Chinese).

[7] 尹聚祺, 杨震, 罗亚中, 等. 空间机动目标跟踪的改进自适应 IMM 算法[J]. 系统工程与电子技术, 2021, 43(12): 3658-3666.
YIN J Q, YANG Z, LUO Y Z, et al. Improved adaptive IMM algorithm for space maneuvering target tracking[J]. Systems Engineering and Electronics, 2021, 43(12): 3658-3666(in Chinese).

[8] GUO G, ZHAO S J. 3D multi-object tracking with adaptive cubature Kalman filter for autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(1): 512-519.

[9] QIAO S J, HAN N, ZHU X W, et al. A dynamic trajectory prediction algorithm based on Kalman filter[J]. Acta Electronica Sinica, 2018, 46(2): 418-423.

[10] LIU W S, LIANG X, ZHENG M H. Dynamic model informed human motion prediction based on unscented Kalman filter[J]. IEEE/ASME Transactions on Mechatronics, 2022, 27(6): 5287-5295.

[11] 谢磊, 丁达理, 魏政磊, 等. AdaBoost-PSO-LSTM 网络实时预测机动轨迹[J]. 系统工程与电子技术, 2021, 43(6): 1651-1658.
XIE L, DING D L, WEI Z L, et al. Real time prediction of maneuver trajectory by AdaBoost-PSO-LSTM network[J]. Systems Engineering and Electronic Technology, 2021, 43(6): 1651-1658(in Chinese).

[12] 王新, 杨任农, 左家亮. 基于 HPSO-TPFENN 的目标机轨迹预测[J]. 西北工业大学学报, 2019, 37(3): 612-620.
WANG X, YANG R N, ZUO J L. Trajectory prediction of target aircraft based on HPSO-TPFENN neural network[J]. Journal of Northwestern Polytechnical University, 2019, 37(3): 612-620(in Chinese).

[13] MA L, TIAN S. A hybrid CNN-LSTM model for aircraft 4D trajectory prediction[J]. IEEE Access, 2020, 8: 134668-134680.

[14] SHI Z Y, XU M, PAN Q. 4-D flight trajectory prediction with constrained LSTM network[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(11): 7242-7255.

[15] 吉瑞萍, 张程祎, 梁彦, 等. 基于 LSTM 的弹道导弹主动段轨迹预报[J]. 系统工程与电子技术, 2022, 44(6): 1968-1976.
JI R P, ZHANG C Y, LIANG Y, et al. Trajectory prediction of boost-phase ballistic missile based on LSTM[J]. Systems Engineering and Electronics, 2022, 44(6): 1968-1976(in Chinese).

[16] 杨春伟, 刘炳琪, 王继平, 等. 基于注意力机制的高超声速飞行器 LSTM 智能轨迹预测[J]. 兵工学报, 2022, 43(S2): 78-86.
YANG C W, LIU B Q, WANG J P, et al. LSTM intelligent trajectory prediction for hypersonic vehicles based on attention mechanism[J]. Acta Armamentarii, 2022, 43(S2): 78-86(in Chinese).

Prediction of aircraft surface trajectory based on the GRU-IKF model with attention mechanism

LIU Yusheng[1]，TANG Xinmin[1, 2, *]，REN Xuanming[1]

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

Abstract：The prediction of aircraft taxiing trajectory helps to solve operational problems such as taxiing conflicts and long waiting times at airports, ensuring surface safety while improving service levels and increasing airport throughput. A model is proposed to predict the taxiing trajectory of a surface aircraft based on an attention mechanism that combines gated recurrent units (GRU) and an improved Kalman filter algorithm (IKF). This addresses the current situation where the performance of machine learning models depends on good data sets. In order to better extract data discrepancy features and learn input-to-output mapping relationships, three independent networks of gated recurrent units are first used to capture the future moment motion states and temporal dependencies of the aircraft. An enhanced extended Kalman filter is then fused with the neural network outputs to integrate them into the state prediction and update process, ultimately improving the predicted trajectory sequence accuracy. Finally, the validity of the model was verified using real aircraft taxi trajectories at Lukou Airport. The simulation results show that the proposed model can effectively and accurately predict aircraft taxi trajectories at the field with an overall mean square error of approximately 0.00128. Compared with the single recurrent neural network (RNN), long and short-term memory network (LSTM) and GRU model, the root mean square error (RMSE) is reduced by 72.9%, 54.7% and 39.9% respectively, and the prediction time is 40 ms, which could accurately and quickly predict the taxiing trajectory and provide assistance in reducing the operating load of the airport surface management system.

Keywords：air transportation; improved Kalman filtering; GRU-attention multi-layer neural network; surface operation of airport; trajectory prediction

Received：2023-04-04；**Accepted**：2023-06-02；**Published Online**：2023-06-16 17：06

URL：link.cnki.net/urlid/11.2625.V.20230616.0925.001

Foundation items：National Key Research and Development Program (2021YFB1600500); National Natural Science Foundation of China (61773202，52072174); National Defense Science and Technology Key Laboratory Foundation of Avionics System Integrated Technology of China Institute of Aeronautical Radio Electronics (6142505180407); Civil Aviation General Aviation Operation Key Laboratory Foundation of China Civil Aviation Management Cadre Institute (CAMICKFJJ-2019-04)

*** Corresponding author**. E-mail：tangxinmin@nuaa.edu.cn

《北京航空航天大学学报》征稿简则

《北京航空航天大学学报》是北京航空航天大学主办的以航空航天科学技术为特色的综合性自然科学学术期刊(月刊)。本刊以反映航空航天领域研究成果与动态、促进学术交流、培养科技人才和推动科技成果向社会生产力转化为办刊宗旨。本刊为中国自然科学技术核心期刊,并被 Ei Compendex 等国内外权威文献检索系统收录。本刊向国内外公开发行,为进一步提高办刊质量和所刊出文章的学术水平,特制定本简则。

1 论文作者及内容

1.1 本刊面向海内外所有学者。

1.2 主要刊载与航空航天科学技术有关的材料科学与工程、飞行器设计与制造、宇航科学与工程、信息与电子技术、控制技术和自动化工程、流体力学和动力工程、计算机科学与应用技术、可靠性工程与失效分析等领域的研究文章。航空航天科学技术民用方面以及具有航空航天工程背景的应用数学、应用物理、应用力学和工程管理等方面的文章也在本刊优先考虑之列。

2 来稿要求

2.1 论文应具有创新性、科学性、学术性和可读性。

2.2 论文为原创作品,尚未公开发表过,并且不涉及泄密问题。若发生侵权或泄密问题,一切责任由作者承担。

2.3 主题明确,数据可靠,图表清晰,逻辑严谨,文字精练,公式符号标准规范,标点符号正确。

2.4 文稿撰写顺序:中文题名(一般不超过 20 个汉字),作者中文姓名、单位、所在城市、邮政编码(作者顺序始终不得更改),中文摘要(包括目的、方法、结果及结论),中文关键词(5 ~ 8 个),中图分类号,引言,正文,参考文献,英文题名,作者英文姓名、单位、所在城市、邮政编码、国别,英文摘要,英文关键词。首、末页下角注明基金项目名称及编号,作者信息。

2.5 作者请登录本刊网页进行在线投稿。

3 稿件的审核、录用与版权

3.1 来稿须经专家两审和主编、编委讨论后决定刊用与否。

3.2 若来稿经审查后认定不宜在本刊发表,将及时告知作者。如果在投稿满 3 个月后仍未收到本刊任何通知,作者有权改投其他刊。在此之前,请勿一稿多投,否则一切后果自负。

3.3 来稿一经作者签字并在本刊刊出,即表明所有作者都已经认可其版权转至本刊编辑部。本刊在与国内外文献数据库或检索系统进行交流及合作时,不再征询作者意见。

邮寄地址:100191 北京市海淀区学院路 37 号 《北京航空航天大学学报》编辑部
办公地点:北京航空航天大学办公楼 405,407 房间
电　　话:(010)82316698,82316981,82315499,82317448
E-mail: jbuaa@ buaa. edu. cn
http://bhxb. buaa. edu. cn
http://www. buaa. edu. cn

《北京航空航天大学学报》
第五届编辑委员会

主　任（主　编）：赵沁平

（以下按姓氏笔画为序）

副主任（副主编）：丁希仑　王少萍　孙志梅　李秋实　李焕喜　杨嘉陵
苗俊刚　相　艳　徐立军　钱德沛　曹晋滨

编　委：马殿富　王　琪　王　聪　邓小燕　王青云　王荣明　刘　宇
刘　红　江　洁　刘　强　闫　鹏　朱天乐　刘铁钢　齐铂金
陈万春　邹正平　苏东林　杨世春　沈成平　邱志平　宋知人
杨树斌　张晓林　杨晓奕　杨继萍　李惠峰　吴新开　张瑞丰
杨照华　宋凝芳　周　锐　林宇震　林贵平　战　强　姚仰平
胡庆雷　赵秋红　段海滨　赵巍胜　席　平　郭　宏　徐　洁
徐世杰　郭洪波　康　锐　翟　锦　熊华钢

编辑部主任：张　嵘

编　　辑：孙　芳　万玉敏　王　茜　卞欢欢　李亚泰

北京航空航天大学学报

Beijing Hangkong Hangtian Daxue Xuebao

（原《北京航空学院学报》）

（月刊　1956 年创刊）

第51卷 第3期　2025年3月

主管单位　中华人民共和国工业和信息化部
主办单位　北京航空航天大学
主　　编　赵沁平
编辑出版　《北京航空航天大学学报》编辑部
邮　　编　100191
地　　址　北京市海淀区学院路37号
印　　刷　北京科信印刷有限公司
发　　行　《北京航空航天大学学报》编辑部
发行范围　国内外发行
联系电话　(010) 82316698 82316981
82315499 82317448
电子信箱　jbuaa@ buaa.edu.cn

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)

(Monthly, Started in 1956)

Vol.51 No.3　March 2025

Administrated by　Ministry of Industry and Information Technology of the People's Republic of China
Sponsored by　Beijing University of Aeronautics and Astronautics (BUAA)
(Beijing 100191, P. R. China)
Chief Editor　ZHAO Qinping
Edited and Published by　Editorial Board of JBUAA
Printed by　Beijing Kexin Printing Co., Ltd.
Distributed by　Editorial Board of JBUAA
Telephone　(010) 82316698 82316981
82315499 82317448
E-mail　jbuaa@buaa.edu.cn
http://bhxb.buaa.edu.cn

中国标准连续出版物号：ISSN 1001-5965 / CN 11-2625/V

ISSN 1001-5965

9 771001 596250 03>

国内定价：50.00 元／期

微信公众号　超星电子书　QQ群